UNIVERSITAT DE BARCELONA

Development of model-driven approaches for metabolic flux analysis and anticancer drug discovery

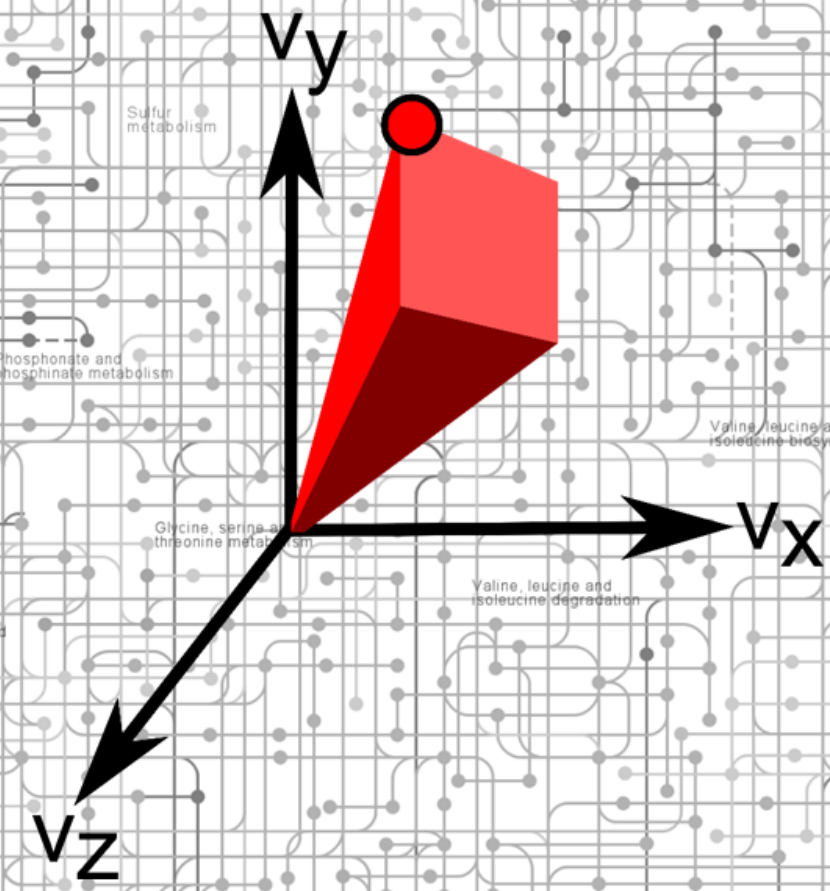
Carles Foguet Coll

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Development of model-driven approaches for metabolic flux analysis and anticancer drug discovery



Carles Foguet

2019

UNIVERSITAT DE BARCELONA

Development of model-driven approaches for metabolic flux analysis and anticancer drug discovery

Memòria presentada per Carles Foguet Coll per optar al grau de doctor per la Universitat de Barcelona

Carles Foguet Coll
Doctorand

Marta Cascante Serratosa
Codirectora

Pedro de Atauri Carulla
Codirector

Doctoral Program in Biotechnology

Faculty of Biology, Department of Biochemistry and Molecular Biomedicine

2019

Alas, five years is far too short a time to do a Ph.D. among such excellent and admirable people. I don't know half of you half as well as I should like.

Abstract

Metabolism is a hallmark of life and underlies most biological processes in both health and disease. For instance, dysregulation of liver metabolism underlies multifactorial disorders such as diabetes or obesity. Similarly, cancer progression involves a reprogramming of metabolism to support unchecked proliferation, metastatic spread and other facets of the cancer phenotype. Hence, the study of metabolism is of great biomedical interest.

The metabolic phenotype emerges from the complex interactions of metabolites, enzymes, and the signaling cascades regulating their expression and thus must be studied following a holistic approach. With this aim, Systems Biology formulates the interactions between the molecular components of metabolism as a set of mathematical expressions, termed metabolic models, and uses them as a framework to integrate multiple layers of data (e.g., transcriptomics, proteomics and metabolomics) and simulate the emergent metabolic phenotype. The Systems Biology toolbox for the analysis of metabolism consists of several complementary model-based approaches, each with its strengths and limitations. For instance, constraint-based modeling can predict flux distributions at a genome-scale, whereas kinetic modeling and ^{13}C metabolic flux analysis (^{13}C MFA) can more accurately model central carbon metabolism.

As part of this Ph.D. thesis, we have expanded this toolbox through the development of new model-based approaches for computing both detailed metabolic maps of central carbon metabolism and genome-scale flux maps. With this aim, we developed HepatoDyn, a highly detailed kinetic model of hepatocyte metabolism capable of dynamic ^{13}C MFA and used it to characterize the negative effects of fructose in hepatic metabolic function. Similarly, we also developed Iso2Flux, a novel steady-state ^{13}C MFA software, and parsimonious ^{13}C MFA, a new ^{13}C MFA algorithm that can integrate transcriptomics to trace flux through large metabolic networks. Even more, we developed r^{2}MTA a constraint-based modeling algorithm to robustly identify the optimal interventions to induce a transition towards a therapeutically desirable metabolic state. Finally, we also developed a workflow for integrating transcriptomics, metabolomics, gene dependencies, and ^{13}C MFA to predict genome-scale flux maps.

Furthermore, we apply the systems biology toolbox, using both newly developed and existing tools, to the genome-scale analysis of the molecular drivers underlying cancer stem cells and metastasis in prostate and colorectal cancer, respectively. We identify putative therapeutic interventions against both phenotypes paving the way for a new generation of anticancer drugs.

Table of contents

1. Introduction

1. Introduction

1.1. Metabolism: a hallmark of life

Life is characterized by the capacity for growing, reproducing and responding to stimuli which contribute to maintaining living organisms in an ordered state and away from equilibrium. Hence, life must be coupled with a set of processes that dissipate energy to the environment. For metazoan, and for any non-photosynthetic or chemosynthetic organisms, life is coupled with the breakdown of complex molecules from the extracellular environment into simpler molecules with lower chemical potential energy. Furthermore, growth and reproduction require taking molecules from the extracellular environment and transforming them, in an ordered and coordinated manner, into the different macromolecular structures that form cells, the basic units of life[1]. Therefore, underlying life there is the metabolism, a large network of interconnected chemical reactions and processes occurring in a highly ordered, regulated and coordinated manner.

The metabolic phenotype of a given organism is defined by both metabolite concentrations and metabolic fluxes. Metabolites are small organic molecules that are the bulk of the reactants in metabolism, whereas metabolic fluxes are the rates at which substrates are converted into products in chemical reactions or transported across a cellular membrane. Metabolite concentrations offer a static snap-shot like view, whereas metabolic fluxes provide a dynamic view of metabolism[2]

To study metabolism, metabolic reactions are often grouped into metabolic pathways. Metabolic pathways are sets of interconnected reactions that fulfill a given role. Furthermore, metabolic pathways can be generally classified as anabolic or catabolic[3].

Catabolic pathways are pathways that break down complex molecules from the media or cellular storage and store their released energy in high energy phosphate bonds of phosphate nucleotides, mainly by phosphorylating adenosine diphosphate (ADP) into adenosine triphosphate (ATP). Catabolic pathways can also be a source of reductive power as many reactions are coupled with the oxidation of nicotinamide adenine dinucleotides (NAD and NADP) and flavin adenine dinucleotide (FAD), which can then act as electron donors[a].

[a] Through this work the terms NADH, NAPDH and FADH will be used to refer to the reduced from of the electron carriers. Conversely, NAD, NADP and FAD will be used to refer to their oxidized forms.

Conversely, anabolic pathways are pathways that use metabolic energy (ATP) and reductive power (mainly NADPH) generated by catabolic pathways to synthesize biomass building blocks (e.g., amino acids, lipids and nucleotides) and assemble them into macromolecules (e.g., proteins, DNA and RNA) and cellular supra-molecular structures (e.g., membranes).

1.2. The principles of metabolic regulation

In order to maintain metabolic homeostasis (i.e., maintain relatively constant composition) and adapt to a diverse external environment, both metabolite concentrations and metabolic fluxes must be tightly regulated[4]. In this regard, enzymes and the interactions of metabolites with enzymes, are the primary level of metabolic regulation.

Enzymes, usually proteins or protein complexes[a], act as catalysts that facilitate reactions by lowering their activation energy. Enzymes achieve this by binding substrates to an active site, a tridimensional structure with amino acid residues which stabilize the transient state of the reaction. Enzyme-mediated catalysis is highly effective, it can increase reactions rate more than twelve orders of magnitude, and as such most metabolic reactions only occur at non-negligible rates through the catalytic activity of the appropriate enzymes. Thus, most metabolic reactions are enzyme-catalyzed reactions[3].

Metabolic regulation can be analyzed at two levels, at the molecular level to understand how molecules (i.e., metabolites and enzymes) influence the rate through individual reactions and at the systemic level, to study the metabolic phenotype (i.e., metabolite concentrations and reaction fluxes) that emerges from the interactions of the individual molecular regulatory components.

1.2.1. Molecular regulation of metabolism

In enzyme-catalyzed reactions one or more substrates bind to the enzyme's active site forming the enzyme-substrate complex, then the substrate(s) react to the product(s) of the reaction which are then released freeing the active site to further catalytic cycles. Hence, the rate through an enzyme-catalyzed reaction primarily depends on the concentrations of metabolites, the amount of enzyme and the kinetic properties of such enzyme. The Kinetic properties of an enzyme determine the affinity of substrates for the active site and the number of catalytic cycles that the enzyme can undergo per unit of time (i.e., the turnover number). Such properties can

[a] Exceptionally enzymes can also be RNA molecules termed riboenzymes

be modulated through the binding of specific metabolites (activators and inhibitors) and by posttranslational modifications of the enzyme[5].

1.2.1.1. Modulation of enzyme-catalyzed reactions by metabolite concentrations

In short time scales, the rate through enzyme-catalyzed reaction is primarily modulated through metabolite concentrations. In the framework of metabolic control analysis (MCA), the sensitivity of the rate through an enzyme-catalyzed reaction to the concentration of a metabolite is defined as its elasticity coefficient[a]. Elasticity coefficients depend on the concentration of the metabolite of interest, the concentration of any other metabolites participating or regulating the reaction and the kinetic properties of the enzyme[4,6].

Substrates are the primary modulator of the rate through enzyme-catalyzed reactions which generally[b] have positive elasticity coefficients to them (i.e., increasing the concentration of substrates tends to increase reaction rates). However, enzyme-catalyzed reactions can asymptotically approach saturation at high substrate concentrations. Saturation describes a state when all enzyme active sites are occupied by the substrates, and any further increase of substrate concentrations will not have any effect on the reaction rate. By approaching saturating concentrations, the elasticity of a reaction rate to the concentration of the substrate decrease towards zero[5].

Additionally, the rate of an enzyme-catalyzed reaction can be modulated by inhibitors or activators. Inhibitors can bind reversibly to the enzyme active site and form unreactive abortive complexes. This type of inhibition can be competitive, if the inhibitor competes with the substrate for binding at the active site, uncompetitive if the inhibitor binds only to the enzyme-substrate complex, and mixed if the inhibitor binds to both. Furthermore, metabolites can also bind to modulatory sites that trigger changes in the enzyme conformation resulting in either increased (activators) or decreased (inhibitors) affinity for substrates or turnover number; this is termed allosteric regulation. Inhibitors (allosteric or non-allosteric) have negative elasticity coefficients, whereas activators have positive elasticity coefficients.

Moreover, for every reaction, the rate of the forward reaction coexists with the rate of the reverse reaction[c]. Hence, for enzyme-catalyzed reactions both the reverse and forward reaction

---

[a] The elasticity coefficient for the effect of metabolite *S* on the rate of the enzyme-catalyzed reaction *R* (ε_S^R) is formally defined as the fractional change in rate of the isolated enzyme for a fractional change in the concentration of metabolite S.

[b] An exception is enzymes with substrate inhibition where the elasticity can be negative.

[c] Chemical equilibrium is reached when the rate through the forward and reverse reaction are equalized.

compete for the enzyme active site and substrates of the forward reaction will act as inhibitors (negative elasticity) of the reverse reaction whereas products will act as inhibitors of the forward reaction. However, for reactions that in the physiological range of concentrations are far from chemical equilibrium, the reverse rate can often be considered negligible. Those reactions are usually referred to as irreversible.

Finally, enzymes which are organized as multiple subunits (each one with its active site) may display cooperativity. Cooperativity occurs when the binding of substrate to the active site at one of the subunits trigger changes in the conformation of the remaining subunits which result in increased (positive cooperativity) or decreased (negative cooperativity) affinity for the substrate(s).

1.2.1.2. Modulation of enzyme-catalyzed reactions at the gene expression and posttranslational level

Gene expression defines the process through which genes are expressed into functional proteins. In eukaryotes, gene expression is primarily regulated at the level of transcription through transcription factors, proteins which bind to specific DNA sequences upstream or downstream of the target gene and modulate its transcription by facilitating or inhibiting the recruitment of the transcription machinery and modulating chromatin packaging of DNA. Processing of the resulting transcripts by splicing machinery can result in the inclusion/exclusion of multiple combinations of exons (i.e., coding regions of the transcript). The processed transcript is then translated into proteins, folded into functional conformation and directed to their targeted cellular compartment. Eventually, proteins are degraded by a proteolytic enzymatic complex known as the proteasome, which can result in different proteins having varying half-lives[3].

Regulation of the gene expression of enzymes plays a crucial role in modulating the rate through enzyme-catalyzed reactions. Indeed, at a constant concentration of substrates and activators/inhibitors, the rate through an enzyme-catalyzed reaction is generally assumed to be directly proportional to the enzyme's concentration. Hence, gene expression can regulate the rate through an enzyme-catalyzed reaction by modulating the concentration of the enzyme. Additionally, for any given reaction, there can be multiple enzymes with different kinetic properties (e.g., substrate affinity or turnover number) capable of catalyzing it. Such enzymes are termed isoenzymes and can arise either from different genes or alternative splicing of transcripts from a single gene. Changes in transcription, splicing and translation usually require

fairly long times to take effect (hours and even days) and hence they are generally used for long term regulation of enzyme-catalyzed reactions[4].

On intermediate time scales, enzyme-catalyzed reactions can also be regulated by posttranslational modifications (e.g., phosphorylation, acetylation, glycosylation, hydroxylation, and ubiquitination) of enzymes. These modifications can have a wide range of effects. For instance, ubiquitination of an enzyme can rapidly reduce its concentrations by triggering their proteasome-mediated degradation. They can also modulate the availability of an enzyme in a given compartment by regulating its cellular localization. Finally, they can induce shifts between multiple conformation states of the enzyme where each state can have different affinity, turnover number or degree of cooperativity[7].

1.2.1.3. Carrier-mediated transport processes

Cellular membranes are generally impermeable to polar molecules including most of the metabolites. Therefore, metabolites are usually transported across membranes through transmembrane proteins termed carriers or transporters. The binding of metabolites to specific sites of the carrier on either side of the lipid bilayer triggers a conformational change that transports metabolites across the membrane. As such, carriers share many properties with enzyme-catalyzed reactions (e.g., carriers have metabolite affinities and turnover numbers akin to enzyme's kinetic properties, can become saturated, can be subject to inhibition or activation both by metabolites or posttranslational modifications). Therefore carrier-mediated transport processes are generally regulated by the same molecular mechanisms as enzyme-catalyzed reactions[8].

1.2.2. System-wide regulation of metabolism

Enzyme-catalyzed reactions and transport processes do not exist in isolation. Instead, they are part of a vast network of interconnected reactions where each metabolite can react or regulate (i.e., activate or inhibit) one or more reactions. Potentially, this allows molecular perturbations (e.g., variations in extracellular metabolite concentrations and enzyme amounts) to propagate across the system and have system-wide effects. Hence, metabolic fluxes, metabolite concentrations, and their dynamics emerge from the collective interaction between the molecular regulation mechanism of specific enzymes. Likewise, the capacity to maintain homeostasis and switch metabolic phenotype in response to stimuli also emerge only at the system level.

1.2.2.1. Metabolic steady-state and short term-homeostasis

Metabolic perturbations are primarily propagated through changes in metabolite concentrations. An increase in the concentration of any given metabolite will always lead to at least two complementary effects, thermodynamic inhibition of reactions producing such metabolite (i.e., reactions producing it move closer to thermodynamic equilibrium) and increased flux through reactions consuming such metabolite. Hence, accumulation of any given metabolite will simultaneously inhibit the flux through reactions that produce it while increasing the flux through reactions that consume it. Likewise, depletion of any given metabolite will have the opposite effect and result in increased flux through reactions producing it and decreased flux through reactions consuming it. Thus, metabolic systems tend to evolve towards a state where inputs and output fluxes are balanced for each metabolite, and both metabolite concentrations and reaction fluxes remain constant or quasi-constant in time. This is termed the metabolic steady state.

However, some metabolic steady states can potentially be incompatible with life. Firstly, enzymes approaching saturating concentration of reactants lose the capacity to increase their flux in response to increased concentrations of substrates. Furthermore, for reactions that are highly favorable thermodynamically, product inhibition might not become relevant until large metabolite concentrations of products are reached. This could hypothetically lead to metabolite accumulations high enough to cause osmotic rupture or toxicity in the cell.

Fortunately, this is rarely the case as living organisms have evolved towards having a complement of enzymes with kinetic properties that tend to facilitate metabolic homeostasis and prevent toxic metabolite accumulations. Firstly, most enzymes tend to have affinities such as that saturation is not generally reached for physiological concentrations ranges and hence can respond to increased concentration of substrates. Secondly, many enzymes have substrate cooperativity and consequently have enhanced response to small variations in metabolite concentrations compared to non-cooperative enzymes[a]. Additionally, a significant number of enzymes and pathways are regulated by negative feedback loops (Figure 1). Under this regulatory loop, reactions are inhibited by either one or more of their direct products

[a] Non-cooperative enzymes have maximum elasticity when substrate concentration tends towards zero while enzymes with positive cooperativity achieve maximum elasticity at a positive non-zero concentration. At a range around such concentration, they have an enhanced capacity to respond to changes in substrate concentration compared to a non-cooperative enzyme.

.

(reinforcing thermodynamic inhibition) or by one or more of the end products of the pathway. Negative feedback inhibition contributes to preventing metabolite accumulation and improves homeostasis by allowing fluxes through pathways to be determined by the metabolic demand of their products rather than by the availability of substrates. Similarly, some enzymes are allosterically activated by metabolites upstream in the pathway and this can also contribute to preventing the accumulation of intermediates of the pathway[4,6].

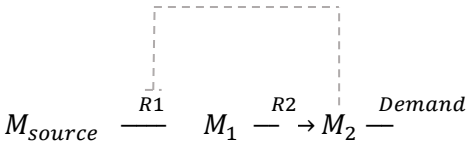


Figure 1. Example of a negative feedback loop. M_{source}, M_1 and M_2 are metabolites and R1 and R2 are enzyme-catalyzed reactions. Accumulation of M_2 inhibits R_1 Under this regulatory loop, the flux through the pathway (R1 and R2) will be regulated by the demand of its end product (M_2)

1.2.2.2. Metabolic reprogramming

On the medium and long time scales, metabolic homeostasis is also maintained through changes in enzyme abundance and posttranslational modifications of enzymes as the result of complex signaling cascades. The process is often termed metabolic reprogramming and can significantly alter the metabolic phenotype of the system of study. For instance, the insulin and glucagon signaling cascades contribute to maintaining glucose and energy homeostasis at an organism level by modulating the activity of a large number of enzymes of central carbon metabolism[9]. Metabolic reprogramming often goes beyond mere metabolic homeostasis. For instance, metabolic reprogramming underlies the switch from quiescent to proliferative state in both health and disease[10,11].

It was initially thought that each pathway had a rate-limiting enzyme, whose activity fully determined the flux through the pathway, and hence metabolic reprogramming could be achieved by modulating this single step. However, it is now accepted that control over concentration and fluxes is distributed across several enzymes in the metabolic network. In the framework of MCA, control over the flux of a pathway is evaluated through flux control

coefficients[a]. The flux control map of a network emerges from the network topology (i.e., stoichiometry and activation/inhibition loops), thermodynamics and the kinetic properties and concentration of enzymes[4,6,12].

1.3. Signaling pathways and transcription factors modulating metabolism

With metabolic control distributed among enzymes in a pathway(s), coordinately modulating the activity of multiple enzymes (i.e., multisite modulation) is generally the most efficient way to change flux distributions. Additionally, multisite modulation also results in less variation in metabolite concentrations[b4,13].

Accordingly, the changes in expression and posttranslational modifications of enzymes in response to external or internal stimuli are coordinated by a complex and highly interconnected network of signaling pathways (e.g., PI3K-AKT-mTOR, RAS-MAPK and cAMP-PKA pathways) and transcriptions factors (e.g., FOXO, MYC and SREBP). As metabolism underlies many cellular functions, most of such pathways regulate key cellular processes such as growth and proliferation or adaptation to cellular stresses. An overview of some of the signaling pathways and transcription factors regulating metabolism is provided in the following subsections.

1.3.1. PI3K-AKT-mTOR pathway

The PI3K-AKT-mTOR is a signaling pathway that regulates a wide array of cellular functions including cell proliferation, survival, motility, and metabolism. The PI3K-AKT-mTOR pathway is primarily activated by the binding of signaling peptides (e.g., insulin, insulin-like growth factor or epidermal growth factor), to specific tyrosine kinase receptors which, trigger the recruitment of phosphoinositide-3-kinase (PI3K) to the plasma membrane. PI3K catalyzes the phosphorylation of the phospholipid phosphatidylinositol 4,5 bisphosphate (PIP2) to form 3,4,5-trisphosphate (PIP3). Conversely, phosphatase and tensin homolog (PTEN) catalyzes the conversion to of PIP3 to PIP2 and contributes to preventing over-activation of the pathway. In turn, PIP3 promotes the activation of the AKT kinase which phosphorylates a wide array of proteins such as tuberous sclerosis protein (TSC)2 or Glycogen synthase kinase 3β (GSK3β)[14,15].

---

[a] The flux control coefficient of enzyme E on reaction R (C_E^R) is formally defined as the fractional change in the steady-state flux through of reaction R for a fractional change in the activity of enzyme E. Flux control coefficients are connected to elasticities through the connectivity theorem.

[b] Indeed, following the Universal Method, it is theoretically possible to modulate the flux of given pathway with no effect on metabolite concentrations by multisite modulation of the all activities of all the enzymes involved. To apply the universal method in a linear pathway, all enzyme activities should be increased equally. If the pathway has ramifications, increases of enzyme activities upstream of such ramifications will depend on the relative flux through each ramification.

The TSC1-TSC2 dimer constitutively inhibits the mammalian target of rapamycin complex (mTORC)1 and AKT phosphorylation of TSC2 releases such inhibition[16]. Additionally, the mTORC1 is regulated by nutrient availability as it is activated by leucine, arginine, glutamine, and α-ketoglutarate[17]. mTORC1 activity promotes protein synthesis by phosphorylating proteins regulating the translation of RNA such as eukaryotic translation initiation factor 4E binding proteins (4EBPs)[18] and ribosomal S6 kinases (S6Ks)[19] as well as promoting ribosome biogenesis[20]. Moreover, mTORC1 activity selectively enhances the translation of transcripts with certain 5' untranslated regions, including components of the translation machinery and genes promoting cell survival and proliferation such as the MYC protooncogene[21–23].

By modulating the activity of several downstream transcription factors (e.g., MYC, HIF, ATF4, FOXO, and SREBP), the PI3K-AKT-mTORC1 pathway can promote the expression of amino acid and glucose transporters, glycolytic enzymes, and enzymes involved in amino acid, lipid and nucleotide biosynthesis[22,24–31]. Furthermore, mTORC1 also promotes oxidative metabolism by inducing mitochondrial biogenesis through the upregulation of the transcription and translation of nuclear-encoded mitochondrial related genes[32–34]. Even more, it can also enhance glycolysis and glycogen and nucleotide synthesis by modulating posttranslational modifications of key activities in such pathways[34–36]. Hence, the PI3K-AKT-mTORC1 pathway, in response to both nutrient availability and mitogenic/pro-anabolic signaling, coordinately upregulates both biosynthetic pathways and the catabolic pathways (i.e., glycolysis and oxidative metabolism) needed to support them[30,37].

In addition to mTORc1, there is a second mTORC complex, mTORC2[a]. This complex can be activated both by direct PIP3 interaction[38] and by AKT mediated phosphorylation[39]. Interestingly, mTORC2 activity also can enhance AKT activity creating a potential positive feedback loop of PI3K-AKT-mTOR signaling[40]. Little is known of mTORC2 functions; however, it might be involved in promoting cell migration, cell cycle progression, and ribosome biogenesis[41,42].

1.3.2. RAS-MAPK pathway

In addition to the PI3K-AKT-mTOR pathway, the RAS and mitogen-activated protein kinase (MAPK) pathway also plays a key role in the regulation of many facets of the cellular phenotype. In this pathway, binding of growth factors to tyrosine kinase receptors or G-protein-coupled receptors activates RAS, which triggers a set of successive kinase activations leading to the

[a] Despite the name and unlike mTORC1, mTORC2 is not targeted by rapamycin

MAPKs. MAPKs phosphorylate a wide variety of effectors, including transcription factors, and promote proliferation survival and migration[43,44].

Additionally, the RAS-MAPK pathway can promote the activation of the PI3K-AKT-mTOR pathway by activating PI3K[45], repressing PTEN expression[46], inhibiting TSC2[47] or phosphorylating the mTORC1 complex. Furthermore, both pathways share common targets and display cross inhibition to prevent excessive mitogenic signaling[48–50].

The RAS-MAPK pathway can optimize metabolism for growth and proliferation both by activating the PI3K-AKT-mTOR pathway and by posttranslationally upregulating the transcription factors MYC [51] and HIFα[52–54]. However, in a context-specific manner, the pathway can also promote gluconeogenesis and fatty acid oxidation[55].

1.3.3. cAMP-PKA signaling

Cyclic AMP (cAMP) is a signaling molecule synthesized from ATP through adenylate cyclase. cAMP levels increase in response to the activation of adenylate cyclase by G-protein coupled receptors such as glucagon or beta-adrenergic receptors[56]. Conversely, cAMP signaling is inhibited by phosphodiesterases, which catalyze its degradation. Indeed, insulin signaling has been found to antagonize cAMP-mediated signaling by enhancing phosphodiesterase activities[57]. The primary effector of cAMP is cAMP-dependent protein kinase A (PKA) a promiscuous kinase with a large number of potential substrates[58].

At the metabolic level, cAMP-PKA signaling promotes glycolysis and glycogenolysis and inhibits lipogenesis and gluconeogenesis. This is achieved through a combination of posttranslational modifications of enzymes of glucose and glycogen metabolism[59–63] and by modulating transcription factors regulating the expression of enzymes associated to glycolysis, gluconeogenesis and lipogenesis (e.g., FOXO, CREB, and SREBP)[64–67].

Furthermore, PKA signaling is also involved in the regulation of MAPK signaling by phosphorylating components of the kinase cascade[68,69].

1.3.4. Sterol regulatory element-binding proteins (SREBP)

Sterol regulatory element-binding proteins (SREBP-1c, SREBP-1a, and SREBP-2) induce the expression of several genes involved in lipogenesis. The different SREBP family members have distinct specificity with SREBP-1c preferentially activating genes involved in fatty acid and triacylglycerol synthesis, SREBP-2 preferentially activating genes involved in cholesterol synthesis and SREBP-1a activating both sets of genes strongly[70]. Physiologically SREBP-1c is

found in lipogenic organs such as liver, SREBP-1a is found in rapidly proliferating such as cancer cells and SREBP-2 is ubiquitously expressed[64,71].

SREBP-1a, SREBP-1c, and SREBP-2 are translated as transmembrane proteins bound to the endoplasmic reticulum (ER). In this regard, they are activated by their translocation to the Golgi apparatus, which is inhibited by high levels of cholesterol in the ER, followed by two proteolytic processes, which release the N-terminal transcription activation domain[72].

The PI3K-AKT-mTOR pathway promotes SREBP transcription and its posttranslational activation[24,29]. Conversely, cAMP signaling though PKA reduces SREBP transcription and activation[64,65].

1.3.5. Forkhead box class O (FOXO)

The Forkhead box class O (FOXO) is a family of transcription factors that modulate the expression of genes involved in apoptosis, cell cycle arrest, oxidative response, as well as gluconeogenetic enzymes[73,74]. FOXOs proteins are primarily regulated through posttranslational modifications, such as AKT-mediated phosphorylation, which causes FOXO to be sequestered in the cytosol and subsequently degraded[25]. Conversely, PKA-mediated phosphorylation can enhance FOXO protein stability and nuclear localization[67]. Furthermore, FOXO proteins can also be activated by AMPK in response to low ATP levels[75] and by acetylation/deacetylation in response to oxidative stress[76,77]. Hence it is a transcription factor that contributes to protecting cells from oxidative and energy stress and maintaining a quiescent state under lack of mitogenic signaling.

1.3.6. Activating transcription factor 4(ATF4)

ATF4 is a transcription factor that promotes both amino acid homeostasis and adaptation to metabolic stress. In this regard, ATF4 translation is activated by stress conditions such as amino acid deprivation, hypoxia or the accumulation of misfolded proteins in the endoplasmic reticulum[78]. Additionally, mTORC1 activates ATF4 by enhancing its mRNA stability and its translation[30]. Similarly, AKT activity has also been reported to enhance its transcription[31], translation[28] and activity[27].

ATF4 regulates amino acid availability by activating the transcription of genes involved in carrier-mediated uptake of amino acids, amino acid synthesis and folate metabolism[30,79–81]. Furthermore, elevated ATF4 activity can also lead to autophagy[82], further increasing amino acid availability.

1.3.7. MYC

MYC is a transcription factor that regulates a wide array of cellular function and controls the transcription of a large number of genes. Indeed, while MYC appears to preferentially promote the transcription of specific genes, it has the potential to bind to the promoter or enhancer region of any gene being actively transcribed when present in large concentrations[83]. While the set of genes regulated by MYC can be context-specific, it generally encompasses genes that promote and support growth and proliferation such as genes associated with the gene expression machinery (e.g., genes involved in RNA transcription stability and genes that promote ribosome biogenesis and protein synthesis), mitochondria biogenesis and cell cycle progression [84,85]. Furthermore, to support rapid proliferation, MYC also upregulates many metabolic pathways such as glycolysis, nucleotide metabolism, folate metabolism, and glutaminolysis[10,22,86].

MYC activity is heavily regulated at the transcriptional, translational and posttranslational levels. For instance, MYC transcription is promoted by growth-promoting signaling pathways such as WNT, Notch JAK and STAT3[84]. Moreover, MTORC1 signaling enhances the translation of MYC RNA[21,22]. Finally, at the posttranslational level, MYC stability is regulated through phosphorylation at different sites. Briefly, phosphorylation by MAPK or cyclin-dependent kinases (CDK) enhances its stability whereas phosphorylation by GSK3β primes MYC for degradation[51].

1.3.8. Hypoxia-inducible factors (HIFs)

Hypoxia-inducible factors (HIFs) are a family of heterodimeric transcription factors that mediate the adaptation to hypoxia. For instance, they upregulate genes that promote glycolysis coupled to lactate production such as glucose and lactate transporters, glycolytic enzymes, and pyruvate dehydrogenase kinase 1[87,88].

The HIF heterodimer is mainly regulated through the HIF-α subunit and the posttranslational level. HIFα is hydroxylated on proline residues by prolyl hydroxylase enzymes (PHDs) and subsequently ubiquitinated, leading to proteasome-mediated degradation[89]. PHDs activity requires O_2 and α-ketoglutarate, substrates of the reaction, and is inhibited by succinate, product of the PHD reaction, and by fumarate[90]. Hence, in low O_2 concentrations, which are associated with the accumulation of succinate[91], HIFα is stabilized, forms a functional HIF heterodimer and triggers a metabolic switch towards anaerobic glycolysis.

Nevertheless, HIF is also regulated by other mechanisms. For instance, mTORC1 promotes both HIF-α transcription and translation[92]. Furthermore, it is also regulated through

phosphorylation[93]. For example, GSK3β phosphorylation reduces HIF-α stability in a PHD independent manner[26,94]. Likewise, MAPKs mediated phosphorylations have been reported to enhance HIFα's nuclear localization[54], activity[52], and stability[53].

1.3.9. Carbohydrate-response element-binding protein (ChREBP)

Carbohydrate-response element-binding protein (ChREBP) is a transcription factor activated by metabolites which mainly increase in response to glucose or fructose uptake, such as glucose-6-phosphate, xylulose-5-phosphate or fructose-2,6-bisphosphate. Conversely, it is suppressed by metabolites that might increase during fasting (e.g., AMP and ketone bodies) and by cAMP signaling (PKA- mediated inhibition). ChREBP induces the expression of glycolytic, lipogenic and pentose phosphate enzymes. ChREBP is mainly expressed in liver, adipose tissue and small intestine[66,95,96].

1.3.10. P53

P53, termed the guardian of the genome, is a protein that under cellular stress, primarily DNA damage, induces cell-cycle arrest, enhances DNA repair and, should the stress persist, apoptosis[97,98]. P53 action is mainly associated with its transcription factor activity, but it can also modulate stress response through other mechanisms such as P53-protein interactions[99,100]. It is primarily regulated at the posttranslational level through phosphorylation and other modifications which enhance its stability and affinity for DNA in response to cellular stress[97,101].

P53 activity negatively regulates the PI3K-AKT-MTOR signaling pathway at multiple levels. For instance, it enhances PTEN expression, inhibits mTOR and S6K, promotes dephosphorylation of 4EBPs and activates the transcription of FOXO[102]. Such regulation is part of a negative feedback loop as P53 translation and stability is enhanced by PI3K-AKT-mTOR signaling[103]. At the metabolic level, P53 inhibits both glycolysis and the pentose phosphate pathway[104].

1.4. Hepatocytes: the master regulators of metabolism

In mammals, the capacity to maintain metabolic homeostasis without a constant input of nutrients from the environment emerges from the interaction between multiple organs and tissues, chief among them the liver. Histologically, the liver is primarily comprised of hepatocytes, which account for roughly 80% of total liver volume[9].

Hepatocytes are organized in hexagonal lobes with a liver arteriole at the center and a triad of bile ducts, hepatic arteriole, and a branch of the portal vein at each lobe corner. Indeed, unlike other organs and tissues, which receive most of the blood from arteries and arterioles, most of the blood supply of the hepatocytes comes from the portal vein, which collects the blood from

the intestine. Hence, in the feed state, blood rich in nutrients absorbed from the intestine flows through hepatocytes. In such state, hepatocytes metabolize glucose and other substrates into glycogen (a polysaccharide of glucose).

During fasting, hepatocytes metabolism is reprogrammed to release glucose and help maintain homeostasis of glucose in the blood. This is achieved by breaking down the stored glycogen and by activating gluconeogenesis (i.e., the synthesis of glucose from amino acids and other substrates like lactate). Furthermore, fatty acid oxidation is enhanced, allowing both to produce the ATP and reductive potential needed for gluconeogenesis and the acetyl-CoA needed for the synthesis of ketone bodies. Ketone bodies are used as substrates by organs that cannot metabolize lipids such as the brain where the uptake of lipids is blocked by the hematoencephalic barrier[9].

Due to their paramount importance to metabolic homeostasis, the metabolism of hepatocytes is tightly regulated both by nutrient availability and hormone-mediated signaling cascades. The main hormones regulating hepatocyte metabolism are insulin and glucagon[a]. Although the pancreas secretes both hormones, they have opposite effects. Insulin activates the PI3K-AKT-mTORC1 pathway and conditions hepatocytes metabolism for the fed state, whereas glucagon acts through the cAMP-PKA pathway and conditions hepatocytes for the fasting state.

1.4.1. Regulation of hepatic glycolysis and gluconeogenesis in short time scales

Glucose metabolism in hepatocytes is characterized by its capacity to switch between glycolysis and glycogenesis, and gluconeogenesis and glycogenolysis in response to both glucose availability and the balance between insulin and glucagon signaling. In short time scales, this plasticity emerges both from liver-specific isoforms with substrate affinities and regulatory loops that allow them to act as sensors of nutrient availability and from posttranslational modifications induced by insulin and glucagon signaling (Figure 2).

1.4.1.1. Glut2 and glucokinase

GLUT2 and glucokinase (GK) are the facilitated glucose transporter and the hexokinase isoenzyme expressed in the liver, respectively. They are both characterized by low affinities for glucose and, hence, the rate through this reaction is highly dependent on glucose concentration, allowing them to act as sensors of glucose levels in the blood[105,106].

[a] To a lesser extent, liver metabolism is also regulated by adrenaline, noradrenaline and cortisol.

Additionally, GK is inhibited by the binding of the GK regulatory protein (GKRP). Such inhibition is released by high glucose concentrations, further enhancing the sensitivity of GK to glucose availability. Additionally, the interaction between GK and GKRP is facilitated by fructose 6-phosphate, which effectively acts as an indirect inhibitor of GK. Such inhibition facilitates the release of glucose when fructose 6-phosphate accumulates as a result of glycogenolysis or gluconeogenesis[63,107,108].

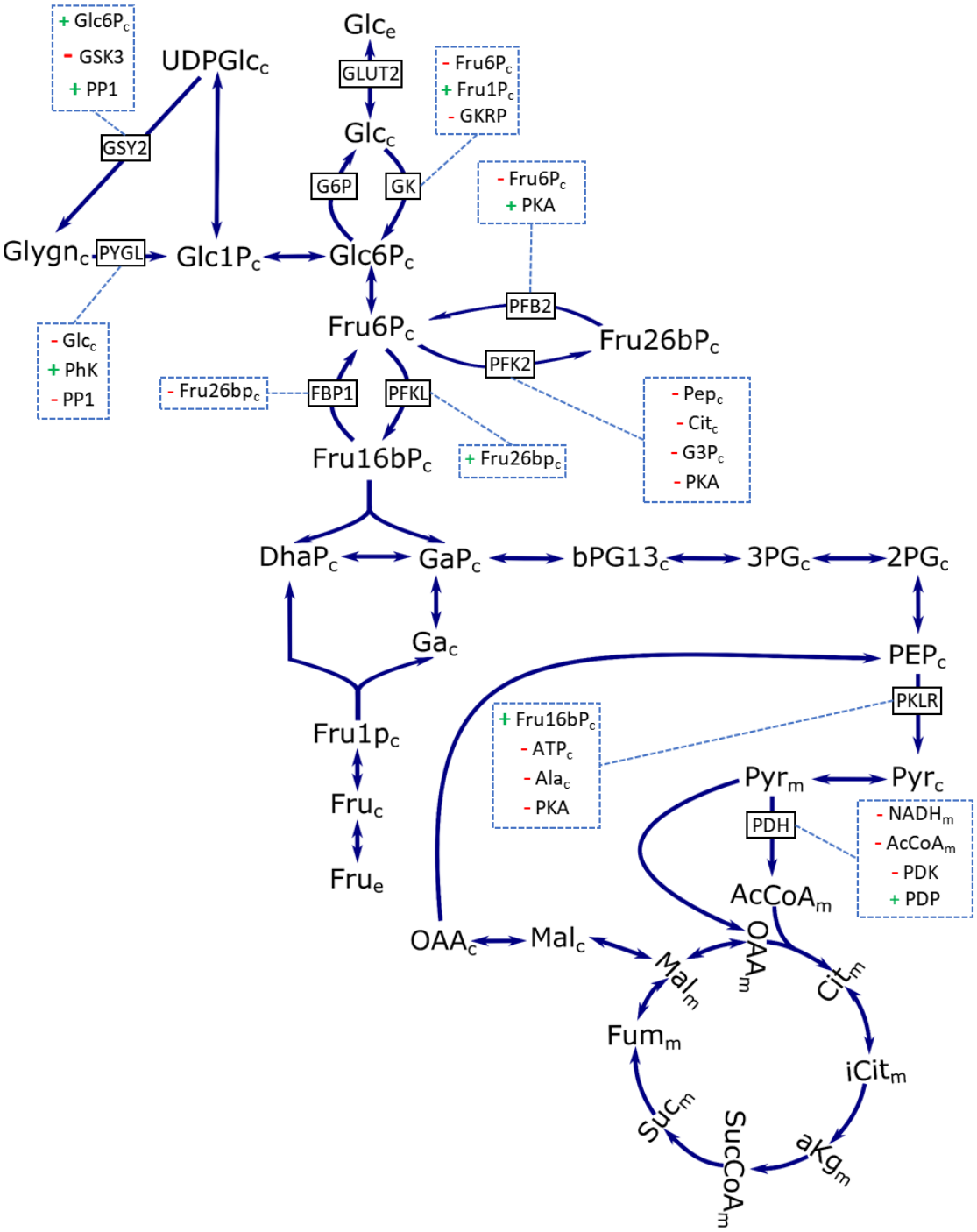


Figure 2 (previous page). Regulation of liver metabolism in short and medium time scales. Energy and redox cofactors have been omitted for simplicity. 2PG: 2-Phosphoglycerate. 3PG: 3-Phosphoglycerate. AcCoA: Acetyl-CoA. αKG: α-Ketoglutarate. Ala: alanine. bPG13: 1,3-Bisphosphoglycerate. Cit: Citrate. DhaP: Dihydroxyacetone phosphate. Fru: Fructose, Fru1P: Fructose 1-phosphate. Fru16bP: Fructose 1,6-bisphosphate. Fru6P: Fructose 6-phosphate. Fum: Fumarate. G3P: glyceraldehyde 3-phosphate Ga: Glyceraldehyde. GaP: Glyceraldehyde-3-Phosphate. Glc: Glucose. Glc1P: Glucose 1-phosphate. Glc6P: Glucose 6-phosphate. Glygn: Glycogen. iCit: Isocitrate. Mal: Malate. OAA: Oxaloacetate. PEP: Phosphoenolpyruvate. Pyr: Pyruvate. Suc: Succinate. SucCoa: Succinyl-CoA. UDPGlc: Uridine diphosphate glucose. The subscripts e, c, and m denote the extracellular, cytosolic and mitochondrial compartments, respectively.

1.4.1.2. Regulation of glycogen synthase and glycogen phosphorylase

Glycogen synthase and glycogen phosphorylase catalyze the synthesis and degradation of glycogen, respectively. The liver isoenzyme of glycogen synthase (GYS2) is activated by glucose 6-phosphate, whereas liver glycogen phosphorylase (PYGL) is allosterically, inhibited by glucose[109]. Due to the sensor-like role of GLUT2/GK, effectively this enhances glycogenesis at high glucose blood concentrations (fed state) and glycogenolysis at low glucose concentrations (fasting state).

At the posttranslational level, GYS2 is inhibited by GSK3α mediated phosphorylation which is inhibited by the insulin signaling cascade[35]. Conversely, PYGL is activated by phosphorylation mediated by phosphorylase kinase (PhK), which is activated by the cAMP-PKA signaling cascade[61].

Both GYS2 and PYGL are dephosphorylated by protein phosphatase-1 (PP1) which is post-translationally activated by the insulin signaling cascade, although the mechanisms of such activation are not yet fully elucidated, and inhibited by PKA[109–111]. Additionally, the dephosphorylation of GYS2 by PP1 is allosterically inhibited by phosphorylated PYGL, further contributing to maintaining GYS2 in the inactive state when PYGL is active[112,113].

1.4.1.3. Regulation of phosphofructokinase-1 and fructose-1,6-bisphosphatase and pyruvate kinase in liver

In the liver, phosphofructokinase-1 (PFKL) and fructose-1,6-bisphosphatase (FBP1), are primarily regulated through fructose 2-6 bisphosphate which is an allosteric activator of PFKL and allosteric inhibitor of FBP1[9,63].

Fructose 2-6 bisphosphate is synthesized and degraded by the 6-phosphofructo-2-kinase (PFK2) and the fructose bisphosphatase-2 (FBP2) activities, respectively, both of which are localized in a single bifunctional enzyme. In hepatocytes and in short time scales, fructose 2-6 bisphosphate concentrations, are primarily regulated by fructose 6-phosphate, which is the substrate of PFK2

and inhibits FBP2, and by citrate, phosphoenolpyruvate and glycerol 3-phosphate which inhibit PFK2. In intermediate timescales, PFK2/FBP2 is regulated by phosphorylation by PKA which enhances FBP2 while inhibiting PFK2[60,62,63].

Regulation of the liver isoenzyme of pyruvate kinase (PKLR) is highly connected to the regulation of PFKL/FBP1 as it is allosterically activated by fructose 1-6 bisphosphate[114]. Furthermore, PKLR is inhibited by PKA-mediated phosphorylation, ATP and alanine. As alanine is a major gluconeogenetic substrate, this contributes to activating gluconeogenesis when alanine is present in significant quantities in blood[59,60,63,115].

1.4.1.4. Pyruvate dehydrogenase

Pyruvate dehydrogenase (PDH) is also tightly regulated as pyruvate is gluconeogenic, whereas its product (i.e., acetyl-CoA) it is not. Firstly, it has product inhibition by both acetyl-CoA and NADH[116]. This leads to the inhibition of PDH during the fasting state as the activation of β-oxidation leads to both high NADH and Acetyl-CoA concentrations. Secondly, PDH is also inhibited by phosphorylation by PDH kinases (PDKs). PDKs are allosterically activated by acetyl-CoA and NADH further reinforcing the product inhibition of PDH. PDK mediated inhibition is countered by PDH phosphatases (PDP), which dephosphorylates PDH[117].

1.4.2. Fructose metabolism in the liver

Like glucose, fructose is absorbed in the intestine and flows through the portal vein to the liver where is metabolized by hepatocytes. Fructose metabolism in hepatocytes consists of phosphorylation of fructose to fructose 1-phosphate by fructokinase and the split of this metabolite by the liver aldolase isoform (aldolase B, ALDOB) into dihydroxyacetone-phosphate and glyceraldehyde, with the latter metabolite being phosphorylated by triokinase into glyceraldehyde 3-phosphate[118].

As it has been covered in a previous section, GK is characterized by both a low affinity for glucose and heavily regulated to modulate the glycolytic flow and prevent excessive accumulation of glycolytic intermediates. However, this is not the case with fructose metabolism where fructokinase has both a high affinity for fructose and lacks any significant allosteric regulation or product inhibition[119]. Indeed, fructose metabolism appears to be primarily controlled by the fructose concentration in blood[120].

Additionally, fructose uptake also promotes glucose uptake by inhibiting the fructose-6-phosphate mediated binding of GKRP to GK[108]. Hence, a small amount of fructose in the diet can have a beneficial effect by enhancing the clearing of glucose from blood during the feeding state[107]. However, largely due to the lack of regulation in fructokinase, diets with excessive

fructose are associated with increased risk to metabolic diseases such as nonalcoholic fatty liver disease, insulin resistance or obesity[121–124]. Hence, considering the prevalence of fructose-rich diets and the social cost of the aforementioned diseases, it is of great interest to develop tools to characterize the effects of fructose on hepatic metabolic function.

1.4.3. Long term regulation of hepatic glucose metabolism

In the long term, hepatic glucose metabolism is primarily regulated through FOXO, CREB, and ChREBP. FOXO transcription factors activate the expression of gluconeogenetic enzymes (e.g., Phosphoenolpyruvate carboxykinase and glucose 6-phosphatase) and PDK4[125]. FOXO is repressed by insulin signaling during the fed state[126]. Similarly, the expression of key gluconeogenic genes is also induced by CREB in response to cAMP signaling[127,128]. Conversely, ChREBP induces the expression of glycolytic genes (e.g., PKLR and GLUT2) as well as genes associated with lipogenesis. It is activated by intermediaries that increase during feeding state and suppressed by ketone bodies and by PKA during fasting[127,129,130].

1.5. Cancer: the biomedical challenge of the 21st century

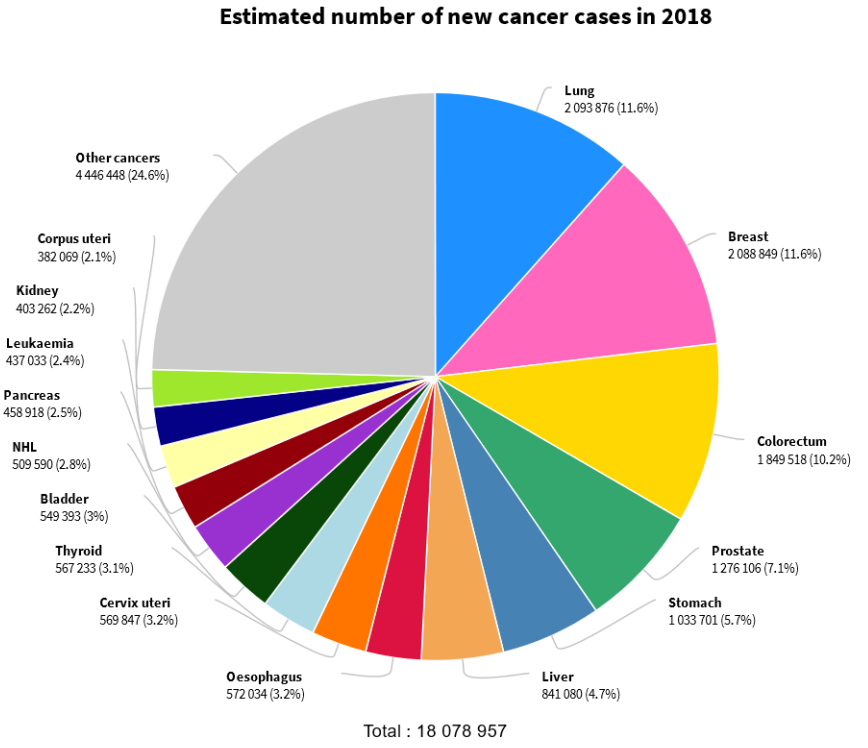
The term cancer is attributed to Hippocrates which coined the term to describe diseases characterized by abnormal tissue growth that resulted in the formation of ulcerating lumps, which reportedly had a morphology reminiscent of that of a crab (*karkinos*). Prior to Hippocrates, mentions of cancer-like diseases can be found in ancient Egyptian writings and fossil evidence indicates that cancer predates written history and indeed humanity[a131].

Nowadays, the term cancer is generally used to define a large group of multifactorial diseases associated with unregulated growth of a cell population with the potential to invade and metastasize. A mass of cancer cells growing in a particular site is termed malignant tumor. Cancers cells have the capacity to invade the primary tumor site by acquiring both a motile phenotype and the capacity to remodel the extracellular matrix (ECM) encapsulating the tumor. Following the invasion of local tissue, cancer cells can escape the primary tumor site by breaking into the blood and lymph vessels and forming secondary tumors in distant tissues and organs[132].

Cancer is a disease with both a strong prevalence and mortality. According to the most recent data published by The Global Cancer Observatory, over 18 million patients were diagnosed with cancer in 2018 while over 9.5 deaths were associated to the disease during the same period (Figure 3). Furthermore, it is expected that cancer incidence will continue to rise in the following

[a] The earliest evidence of cancer comes from masses identified in dinosaur fossils through radiologic surveys[597].

decades, and by 2040 there will be over 29 million yearly new diagnoses of cancer[133,134]. Hence, acquiring a better understanding of the mechanisms underlying cancer and the development of effective therapeutic strategies is one of the major biomedical challenges of the 21st century.



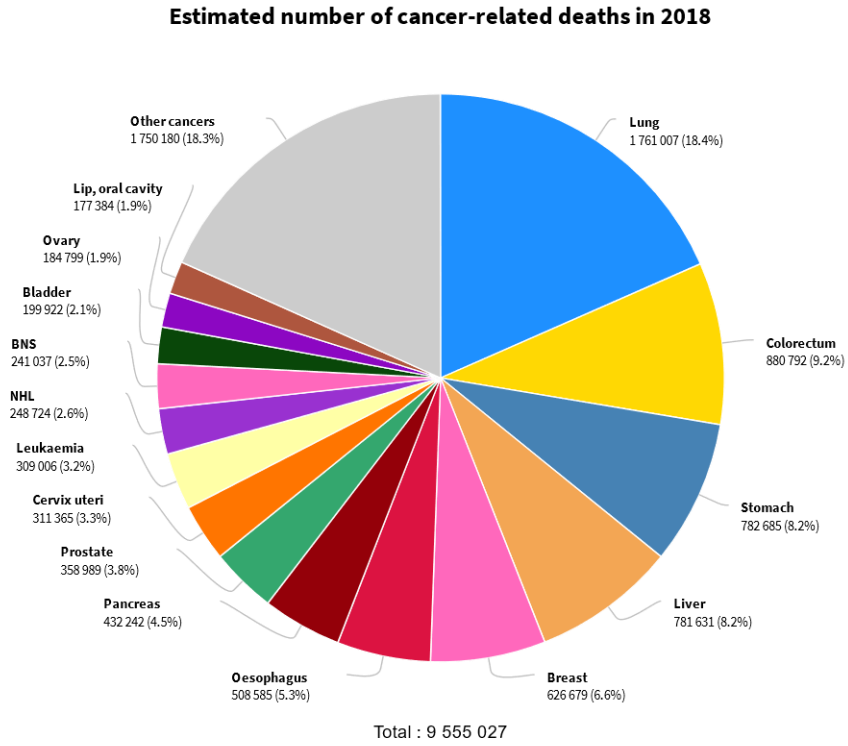


Figure 3. Estimated cancer incidence and mortality in 2018 (worldwide, both sexes, all ages all cancers). Data obtained from GLOBOCAN 2018 (Global Cancer observatory).

1.5.1. Hallmarks of Cancer

In addition to the capacity to invade and metastasize, cancer cells have a set of unique features that define them. Following the influential review by Douglas Hanahan and Robert A. Weinberg, these characteristics are referred to as the hallmarks of cancer[135]. Initially, the hallmarks were defined as six acquired capabilities (i.e., the functional capability of self-sufficiently in growth signals, insensitivity to anti-growth signals, evading apoptosis, limitless replicative potential, angiogenesis and tissue invasion and metastasis) driven by genome instability[135]. Later, the hallmarks were expanded to include tumor-associated inflammation, the capacity to avoid immune destruction, and metabolic reprogramming[136]. Additionally, although it is not one of the eight hallmarks of cancer, Epithelial-Mesenchymal Transition (EMT) it is also recognized as a key step for the acquisition of invasiveness in cancers of epithelial origin[137,138].

1.5.1.1. Altered signaling pathways in cancer

In health, cell proliferation, migration and metabolism are tightly regulated through a series of signaling cascades in order to maintain tissue homeostasis. As a result of genetic instability and environmental cues, cancer cells gain the ability to overactivate pathways that promote growth and proliferation. This is achieved by overexpressing or acquiring gain of function mutations for positive regulators of mitogenic pathways or downregulating or acquiring loss of function mutations for negative regulators of growth and proliferation.

Case in point, the PI3K-AKT-mTOR pathway is overactivated in many cancers. Such activation can arise through many potential alterations in the signaling molecules involved in the pathway, such as overexpression of insulin-like growth factor[139], insulin-like growth factor receptors[140], gain of function mutations in PI3K subunits[141], loss of function of PTEN[142], gain of function mutations in AKT[143], loss of function mutations in TSC1 and TSC2[143], and activating mutations of mTOR[144]. Likewise, the MAPK pathway is frequently activated by mutation in RAS and other components of the kinase signaling cascade[43]. Dysregulation also occurs at the level of transcription factors regulated by such pathways. For instance, MYC can be overactivated independently of the above-mentioned pathways through amplification, translocation to a chromosomal location under the control of a powerful promoter or through mutations that increase its half-life[145].

Conversely, pathways or factors that promote DNA repair or trigger apoptosis in response to cellular stresses are frequently inactivated in cancer. For instance, P53 is inactivated in most tumors[146].

1.5.1.2. Genetic instability and clonal selection

Cancer progression is mainly driven by the accumulation of beneficial genetic alterations which tend to activate or potentiate oncogenes (i.e., genes which promote cancer progression) or inactivate or attenuate tumor suppressor genes. Such genetic alterations range from nucleotide-level mutations (e.g., single nucleotide variations and small insertions) to large chromosomal level alterations (i.e., copy number variation, translocations, inversions, tandem duplications and others)[147]. Such genomic instability is driven, among other factors, by the increased proliferation rate and the loss of tumor suppressor genes such as P53[97,98,148] or DNA mismatch repair genes[149].

Additionally, inactivation of tumor suppressor genes and activation of oncogenes can also be achieved through epigenetic variations. Epigenetic variations are potentially heritable changes in gene regulation that do not involve changes in the nucleotide sequence of DNA. They include changes in DNA methylation or covalent modification of DNA-bound histones, both of which regulate the DNA regions that are open to transcription. Genes involved in these processes are frequently mutated in cancer[150,151].

The genome instability and the high rate of epigenetic changes in cancer cells lead to a large degree of inter- and intratumor heterogeneity. Tumors generally consist of a diverse group of clone subpopulations with distinct genetic and epigenetic information. Though selective pressure, clones with mutations or stable epigenetic changes that are advantageous for cancer progression are positively selected whereas those clones with unfavorable phenotype are negatively selected allowing tumor populations to progress towards the most efficient phenotype for cancer progression[152,153].

In this regard, cancer cell lines derived from a single patient, such as the SW480/SW620/LiM2[154,155] or PC3/PC3M/PC3S models[156,157], can be used to study the role of distinct tumor populations at different stages of progression and identify putative targets against the subpopulations that contribute more to therapy resistance and metastatic spread.

1.5.1.2.1. Cancer stem cells (CSCs)

Among the many cell populations found within of the bulk of the tumor, the study of cancer stem cells (CSCs) populations is of great therapeutic interest as they are strongly associated with therapy resistance and metastatic spread. CSCs are subpopulations of cancer characterized by a high proliferative and self-renewal potential, and the capacity to autonomously establish primary tumors and metastasis[158,159]. Likewise, CSCs are also endowed with increased resistance

to cell death and cellular stress and hence can survive therapeutic interventions that might kill the bulk of the tumor[160–164].

Similar to stem cells found in non-tumoral tissues, CSCs have the capacity to generate non-CSC populations through asymmetric divisions and epigenetic changes[165]. There is also evidence that, in some circumstances, differentiation might be reverted in response to environmental cues or the acquisition of mutations that promote stemness[166–168].

Cancer cell stemness is regulated through a variety of signaling pathways that promote proliferation, cell survival and contribute to maintaining pluripotency. For instance, cancer stemness has been associated with signaling through PI3K/Akt/mTOR[169], MAPK[170,171], focal adhesion kinase[172] , NOTCH[173], Sonic Hedgehog[174] and the inhibition of the HIPO signaling pathway[175]. Similarly, and largely regulated by the above-mentioned signaling pathways, the transcription factors MYC[176,177], KLF4[178,179], BMI1[180,181] and TAZ[175,182], among others, have also been associated with CSCs.

1.5.1.3. Epithelial-mesenchymal transition (EMT)

EMT is a process by which epithelial cells lose stable cell to cell adhesions and a shift from apical-basal to front-rear polarity, leading to migratory mesenchymal cells[137,138].

Under normal physiological conditions, EMT has a crucial role in embryonic development and wound healing and tissue repair processes[183]. However, EMT is also hijacked by cancers of epithelial origin in order to gain the capacity to invade the primary site, extravasate into the blood or lymphatic vessels, and form metastatic lesions to distant tissues or organs[136]. Furthermore, EMT has also been reported to endow cancer cells with other selective advantages such as drug resistance[184] or resistance to cell death[185]. However, it is also beginning to emerge that a complete EMT can also lead to both reduced cell proliferation and tumorigenic potential [157,186–191]. For such reason, it has been suggested that cancer progression can often involve a partial or intermediate EMT where expression of mesenchymal and epithelial markers coexist[137,138,192–194].

EMT is supported by significant changes in gene expression including the cadherin switch (i.e., the downregulation of E-cadherin and the upregulation N-cadherin), downregulation of claudin and occludins (i.e., the components of tight junctions) and changes in the cytoskeleton (e.g. upregulation vimentin and downregulation of cytokeratin). Further supporting motility, cells undergoing EMT acquire the capacity to remodel the ECM through changes in the expression of ECM proteins (e.g., upregulation of fibronectin and type I/II collagen, and downregulation of laminins) and the expression of ECM-degrading metalloproteases[183,184]. Such changes are

coordinated by transcription factors such as the members of the Snail family, ZEB1 or TWIST1[137,138].

1.5.1.3.1. Snail

Snail is a family of transcription factors involved in the repression of genes associated with the epithelial phenotype (chiefly among them E-Cadherin) and the induction of EMT. Structurally, members of the Snail family have a conserved DNA binding domain, which binds specifically to sequences found in the promoter regions of epithelial genes like E-cadherin and claudins, and a conserved SNAG domain which can recruit corepressors[195]. In humans, there are three members of the Snail family, SNAI1, SNAI2 (also known as Slug) and SNAI3 (also known as Smuc). The different family members mainly differ in the central domain of the protein. For instance, in SNAI1 the central domain contains a destruction box and nuclear export sequence whereas in SNAI2 the central region contains the Slug domain that can recruit corepressors and regulates the activity of the SNAG domain[195–197].

At the transcription level, both SNAI1 and SNAI2 are activated by TGF-β signaling [198,199], while SNAI1 transcription has also been reported to be activated by TNF-α/NF-κ and MAPK signaling[200,201]. Furthermore, SNAI1 and SNAI2 are part of an autoregulatory loop where transcription of the former is repressed by SNAI1[202] and possibly SNAI2[203,204], and the latter is repressed by SNAI1[205]. Additionally, Krüppel-like factor 4 (KLF4), a transcription factor associated with pluripotency [206], can repress the transcription of both SNAI1 and SNAI2[207–209].

SNAI1 is also extensively regulated at the posttranslational level. For instance, GSK-3β mediated phosphorylation of the nuclear export sequence, and the destruction box of SNAI1 triggers its nuclear export and proteasome-mediated degradation[210]. Similarly, Protein kinase D1(PRKD1) mediated phosphorylation of SNAG domain can lead to an inactive SNAI1/DNA complex[211] or trigger SNAI1 nuclear export and ubiquitination in a context-dependent manner[212,213]. On the other hand, phosphorylation at sites at the DNA binding domain of SNAI1 by the motility promoting p21-activated kinase 1[214] or Large Tumor Suppressor Kinase 2 (LATS2) of the Hippo pathway[215] facilitates SNAI1 retention in the nucleus preventing proteasome-mediated degradation. Likewise, TNF-α/NF-κ promotes SNAI1 stability by inducing the transcription of COP9 signalosome complex subunit 2 which disrupts the interaction of GSK-3β and ubiquitin ligases with SNAI1[216].

SNAI2 is also regulated at the posttranslational level. For instance, GSK-3β mediated phosphorylation has been linked to increased SNAIL2 nuclear export and degradation[217]. Similarly, TNF-α/NF-κB signaling has been reported to increase its stability[218].

1.5.1.4. Metabolic reprogramming in cancer

Metabolism underlies most cellular processes and, as such, it plays a crucial role in cancer progression. First and foremost, the acquisition of the capacity for rapid proliferation must be supported by a metabolic phenotype capable of supplying both the building blocks (e.g., amino acids, lipids and nucleotides) and the metabolic energy and reductive potential needed for the S phase of the cellular cycle. As such, most mitogenic signaling pathways overactivated in cancer (e.g., PI3K-AKT-mTOR and MYC) trigger a metabolic reprogramming that optimizes the metabolic phenotype for rapid proliferation[10].

Furthermore, metabolic reprogramming can also promote signaling and epigenetic changes that support cancer progression by modulating the concentration of critical metabolites, such as 2-hydroxyglutarate, succinate or fumarate, which regulate posttranslational modification of proteins (e.g., HIF or DNA associated histones) and DNA[90,137,219].

Even more, cancer associated metabolic reprogramming can also underlie, the acquisition of metastatic potential[220], genetic instability[221,222], drug resistance[223,224] or immune escape[225,226].

1.5.1.4.1. Glycolysis in cancer: the Warburg effect

One of the most striking features of cancer cell metabolism is the propensity to use aerobic glycolysis (i.e., the fermentation of glucose into lactate), in detriment to glucose oxidation through the TCA cycle. This phenomenon is named Warburg effect after Otto Warburg, who 90 years ago described that slices of hepatoma incubated with glucose generated lactate at a rate 70 times higher than healthy liver tissue[227].

1.5.1.4.1.1. What are the advantages of the Warburg effect?

Fermentation of glucose to lactate yields only 2 ATP per molecule, whereas oxidizing glucose through the TCA cycle can yield approximately 31 molecules of ATP per glucose[228]. While it was initially thought that the Warburg effect could be attributed to defective mitochondria function, indeed Warburg himself was a strong proponent of this hypothesis, this was disproven by the evidence that most cancer cells have functional mitochondria[229,230].

However, cancer almost invariably converges towards the Warburg effect, suggesting that it offers a selective advantage to cell proliferation and cancer progression. Indeed, the Warburg effect is also active in many non-cancerous cell populations with high proliferative capacity[11]. In this regard, it has been hypothesized that aerobic glycolysis can support the metabolic demands associated with a fast proliferation rate better than the oxidative metabolism of glucose. Firstly,

a high rate of aerobic glycolysis facilitates the synthesis of biomass building blocks and reductive power from glycolytic intermediaries[231,232]. Secondly, as aerobic glycolysis is enzymatically less complex than glucose respiration, the latter might support faster ATP and biomass building blocks synthesis when the cost of synthesizing and maintaining enzymes and the limited solvent capacity of the cell are considered[233–235]. Furthermore, the Warburg effect might provide an enhanced capacity to adapt to perturbations in the supply of both oxygen and glucose[236].

An additional advantage of the Warburg effect is that it can produce ATP independently of mitochondrial respiration and hence reduce the production of reactive oxygen species (ROS). In this regard, detachment from the extracellular matrix is associated with increased formation of ROS, which can lead to a specific form of apoptosis termed anoikis[237,238]. Hence, the Warburg effect is theorized to promote metastasis by enabling cancer cells to evade anoikis and survive in circulation[220].

Finally, other advantages of the Warburg effect for cancer cells might include inhibiting the immune response by depriving immune subpopulations of glucose and through acidification of the tumor microenvironment[225,226]. Similarly, acidification of the tumor microenvironment can also promote a gene expression program associated with invasiveness and metastasis in cancer cells[239,240].

1.5.1.4.1.2. Transcriptional drivers of the Warburg effect

Both MYC and HIF have been reported to induce the transcription of glucose transporters (GLUT1 and GLUT3)[241], glycolytic enzymes (e.g. HK2, GPI, PFK-P/M, ALDOA/C, TPI1, GAPDH, PGK1, ENO1, PKM2 and LDHA/B[87,88,242–245]) and lactate transporters (i.e. MCT1,MCT2 and MCT4 [246,247])(Figure 4). Conversely, P53 can suppress the expression of glucose transporters GLUT1 and GLUT4[104] and HK2[248]. Hence, overactivation of MYC or HIF and the loss of P53 transcriptionally promote the Warburg effect.

1.5.1.4.1.3. Isoenzymes supporting Warburg effect

As part of tumor progression and MYC and HIF signaling, cancer cells also express isoenzymes with kinetic properties favorable for the Warburg effect. For instance, the glucose transporters (i.e., GLUT1 and GLUT3) and hexokinase isoenzyme (i.e., HK2) expressed in cancer cells have a high affinity for glucose and thus facilitate both glucose uptake and usage[a249,250].

[a] This is in stark contrast with liver, where glucokinase and GLUT2 have low glucose affinity and serve as glucose sensors.

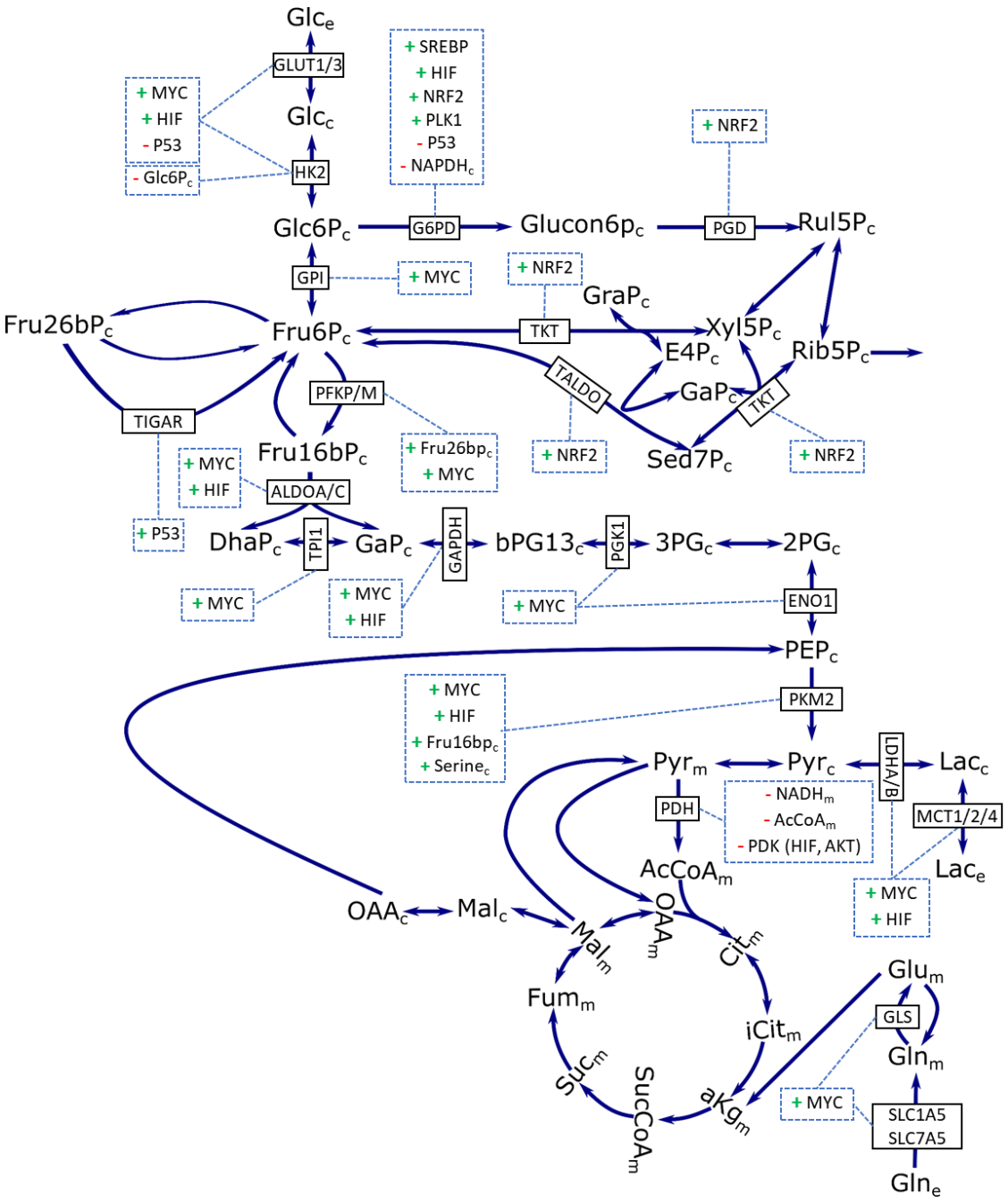


Figure 4. Regulation of glycolysis, pentose phosphate pathway, and glutaminolysis in cancer. Energy and redox cofactors have been omitted for simplicity. 2PG: 2-Phosphoglycerate. 3PG: 3-Phosphoglycerate. AcCoA: Acetyl-CoA. aKG: α-Ketoglutarate. bPG13: 1,3-Bisphosphoglycerate. Cit: Citrate. DhaP: Dihydroxyacetone phosphate. Fru26bP: Fructose 2,6-bisphosphate Fru16bP: Fructose 1,6-bisphosphate. Fru6P: Fructose 6-phosphate. Fum: Fumarate. GaP: Glyceraldehyde-3-Phosphate. Glc: Glucose. Glc6P: Glucose 6-phosphate. Gln: Glutamine. Glu: Glutamate. Glucon6P: Gluconate 6-phosphate. iCit: Isocitrate. Lac: Lactate. Mal: Malate. OAA: Oxaloacetate. PEP: Phosphoenolpyruvate. Pyr: Pyruvate. Rib5P: Ribose 5-phosphate. Rul5P: Ribulose 5-phosphate. Sed7P: Sedoheptulose 7-phosphate. Suc: Succinate. SucCoa: Succinyl-CoA. The subscripts e, c, and m denote the extracellular, cytosolic and mitochondrial compartments, respectively.

Another isoenzyme with a critical role in Warburg effect, and indeed in cancer progression, is PKM2 which is a product of alternative splicing of the Pyruvate Kinase M gene. Unlike other isoenzymes, PKM2 can exist either in dimeric or tetrameric form. The later has a high affinity for phosphoenolpyruvate, the substrate of pyruvate kinase, compared to the later which is nearly inactive in physiological conditions. The tetrameric form is stabilized by fructose 1-6 bisphosphate which acts as a strong allosteric activator of PKM2[251]. Additionally, PKM2 is also allosterically activated by serine, which is synthesized from glycolytic intermediates[252]. Together, those activations result in a regulatory loop that promotes increased accumulation of glycolytic intermediaries hence providing metabolic precursors for anabolic pathways such as nucleotide or serine synthesis[79,253]. Additionally, it is worth mentioning that in response to certain stimuli such hypoxia or MAPK signaling PKM2 can migrate to the nucleus where it can enhance the expression of MYC[254] and interact with HIF to increase the expression of HIF target genes[88], further promoting Warburg effect.

1.5.1.4.1.4. Fructose 2-6-bisphosphate in cancer

Like in hepatocytes, the isoforms of phosphofructokinase expressed in cancer, mainly PFKM and PFKP, are allosterically activated by fructose 2-6-bisphosphate and hence regulated by PFK2/PFBP2 activity. In this regard, most cancer cells have high expression of the PFK/FBPase 3 isoenzyme, which has a high ratio of PFK2 to FBP2 activity, leading to increased fructose 2-6-bisphosphate concentrations and hence a larger glycolytic flux[255]. Expression of PFKFB3 is driven both by HIF[87,256] and PI3K-AKT-mTORC1 signaling[257,258]. However, expression of PFK2/PFBP2 isoenzymes with less kinase activity has also been reported as a mechanism to redirect flux through the pentose phosphate pathway and protect against oxidative stress[259,260].

Conversely, P53 activity lowers the concentration of fructose 2-6-bisphosphate by activating the transcription of TP53-induced glycolysis and apoptosis regulator (TIGAR), an enzyme which catalyzes the dephosphorylation of fructose 2-6 bisphosphate[261]. To a lesser extent, TIGAR can also catalyze the dephosphorylation of fructose 1-6 bisphosphate further suppressing the glycolytic flux[262].

1.5.1.4.1.5. Pyruvate Dehydrogenase kinases in cancer

PDKs play a key role in the Warburg effect by inhibiting PDH and redirecting pyruvate towards lactate production. In this regard, HIF has been reported to enhance the transcription of PDK1/3[263,264] while P53 has been reported to repress the transcription of PDK2[265]. Additionally,

AKT can enter the mitochondria, in a transport process that is enhanced under hypoxia, and phosphorylate PDK1 enhancing its activity[266].

1.5.1.4.2. Pentose phosphate pathway

The pentose phosphate pathway (PPP) has a crucial role in cancer cell metabolism, producing both the ribose phosphate backbones for nucleotide synthesis and the NADPH needed for anabolic pathways and redox homeostasis. As such, it is frequently upregulated in cancer[267,268]. Indeed, in proliferating cells the oxidative branch of PPP is generally assumed to be the primary contributor to NAPDH production, accounting for between 30-50% of its total production[269].

The PPP consists of two branches, the irreversible oxidative branch and a reversible non-oxidative branch (Figure 4). In short time scales, the flux through the oxidative branch is regulated by glucose-6 phosphate concentration and by NADPH demand[270]. Conversely, the flux (and direction) of the non-oxidative branch will depend on the demand of ribose-phosphate relative to the activity of the oxidative branch as well as the concentration of fructose 6-phosphate and glyceraldehyde 3-phosphate[267]. Hence, the Warburg effect, by increasing the concentrations of glycolytic intermediates, can facilitate the flux through both the oxidative and non-oxidative branch of the pentose phosphate pathway[253,259].

At the transcription level, the PI3K-AKT-mTORC1 pathway can promote expression of Glucose-6-phosphate dehydrogenase in an SREBP-dependent manner[24,29]. Furthermore, enzymes in both the oxidative and non-oxidative branches of PPP (Glucose-6-phosphate dehydrogenase, Phosphogluconate Dehydrogenase, Transketolase, Transaldolase) are also upregulated by the Nuclear factor erythroid2-related factor 2 (NRF2)[271–273] which can be activated by oxidative stress or PI3K/AKT signaling[274,275]. Additionally, ChREBP has also been shown to upregulate PPP, but the effects appear to be restricted to hepatocellular carcinoma[66,276,277].

Finally, at the posttranslational level, Glucose-6-phosphate dehydrogenaseis inhibited by direct interaction with P53[99] and activated through phosphorylation by the Polo-like kinase 1(PLK1), which is frequently upregulated in cancer[278].

1.5.1.4.3. Lipid metabolism

Lipid metabolism plays a crucial role in cancer progression by regulating the supply of structural lipids, signaling processes and energy and redox balances[279].

Firstly, most cancer cell display increased lipogenesis in order to meet the demand of structural lipids (e.g. phospholipids and cholesterol) needed to support a fast proliferative rate. In this regard, enzymes of the fatty acid synthesis pathways (e.g., ATP-citrate lyase, Acetyl-CoA

carboxylase, and fatty acid synthase) are overexpressed in most cancer types[280]. This upregulation appears to be primarily mediated by SREBP activity driven by PI3K-AKT-mTORC signaling[24,29,281,282]. Additionally, ATP-citrate lyase, a key enzyme in the *de novo* fatty acid synthesis pathway, is activated by AKT-mediated phosphorylation[283,284].

Lipid biosynthetic pathways can also modulate lipid composition in the cell and regulate a wide array of cellular processes. For instance, ChREBP-induced stearoyl CoA desaturase has been shown to promote metastasis and invasion in colorectal cancer by increasing the ratio of monounsaturated fatty acids leading to the downregulation of PTEN[285]. Likewise, changes in sphingolipid metabolism have been shown to underlie EMT[286] , metastatic potential[287] and apoptosis[288].

β-oxidation is also active in many cancers and has been reported to be important for cancer cell survival and proliferation[279]. Firstly, β-oxidation can serve as a source of reductive power and metabolic energy, and provide metabolic flexibility under stress[289–291]. Additionally, β-oxidation might also contribute to cancer progression by reducing the concentration of lipids with antiproliferative or lipotoxic effects[291], promoting genetic instability through ROS generation[221] or, inhibiting apoptosis[109].

1.5.1.4.4. Folate Metabolism

Folate metabolism is centered around tetrahydrofolate (THF), which is obtained from folate (i.e., vitamin B9) through two consecutive reactions catalyzed by dihydrofolate reductase (DHFR)[a]. THF can act as a carrier of one-carbon units (1C) with different oxidation states (methyl, methylene, methenyl, and formyl). This allows C-THF to act as a 1C donor for purine synthesis, deoxythymidine synthesis, and homocysteine re-methylation. The latter is part of the methionine cycle which produces S-adenosyl methionine, the principal methyl donor in the cell.

The primary source of 1C are the reactions catalyzed by serine hydroxymethyltransferases (SHMT1/2), which catalyze the cleavage of serine into glycine[b], and the glycine cleavage system. Both reactions use THF as a carbon acceptor and produce N^5,N^{10} methylene-THF(CH2-THF). However, there are other sources of 1C, such as histidine catabolism or formate. Enzymes such as methylenetetrahydrofolate dehydrogenase (MTHFD*1*/2/2l) and methylenetetrahydrofolate reductase (MTHFR) catalyze redox reactions that interconvert CH2-THF to N^{10}-formyl

[a] DHFR activity is also necessary to regenerate THF from the dihydrofolate generated by deoxythymidine synthase

[b] The reaction is reversible and can also serve to synthetize serine from glycine

tetrahydrofolate (CHO-THF) and to N^5-methyl tetrahydrofolate (CH3-THF). In this regard, folate metabolism can also play a key role in redox homeostasis as several reactions in the pathway are coupled to NAD(P)H synthesis[269,292]. Furthermore, it can also serve as an important source of ATP as the release of formate from N^{10}-formyl tetrahydrofolate (CHO-THF) is coupled to ATP phosphorylation[293,294]. Finally, it is worth mentioning that folate metabolism has parallel branches in the cytosol and mitochondria leading to some redundancy in the activities in each branch[295,296](Figure 5).

Folate metabolism is often over-activated in cancer cells playing a key role supporting *de novo* nucleotide synthesis, epigenetic regulation and energy and redox balance[269,293,294,297–299]. In this regard, the Warburg effect and the expression of PKM2 promote folate metabolism by favoring *de novo* serine synthesis from the glycolytic intermediate 3-phosphoglycerate[253,300]. Additionally, MYC and mTORC1-ATF4, have been reported to activate the transcription of enzymes involved in serine *de novo* synthesis(phospho-glycerate dehydrogenase, phosphoserine aminotransferase 1, and phosphoserine phosphatase), MTHFD2, SHMT1/2 and glycine decarboxylase (GLDC, part of the glycine cleavage system)[79–81,301–304].

1.5.1.4.5. Nucleotide synthesis

To support their rapid proliferation rate, cancer cells often have an increased rate of nucleotide and deoxynucleotide synthesis. Firstly, MYC overactivation promotes the transcription of enzymes of the purine and pyrimidine biosynthesis pathways[305,306]. Additionally, S6K, activated by mTORC1, phosphorylates and activates CAD (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase) the enzyme which catalyzes the first three reactions on the *de novo* pyrimidine synthesis pathway[36]. Moreover, increased nucleotide synthesis is also supported by overactivation of the PPP and folate metabolism, which provide the ribose backbones and 1C necessary for the pathway[80,268].

Dysregulation of nucleotide metabolism might also affect genome instability. For instance, alterations in the urea cycle can result in an excess of carbamoyl-phosphate leading to an excess of synthesis of pyrimidines. The resulting imbalance between purine and pyrimidine synthesis increases the frequency of transversion mutations where a pyrimidine is added in place of a purine[222].

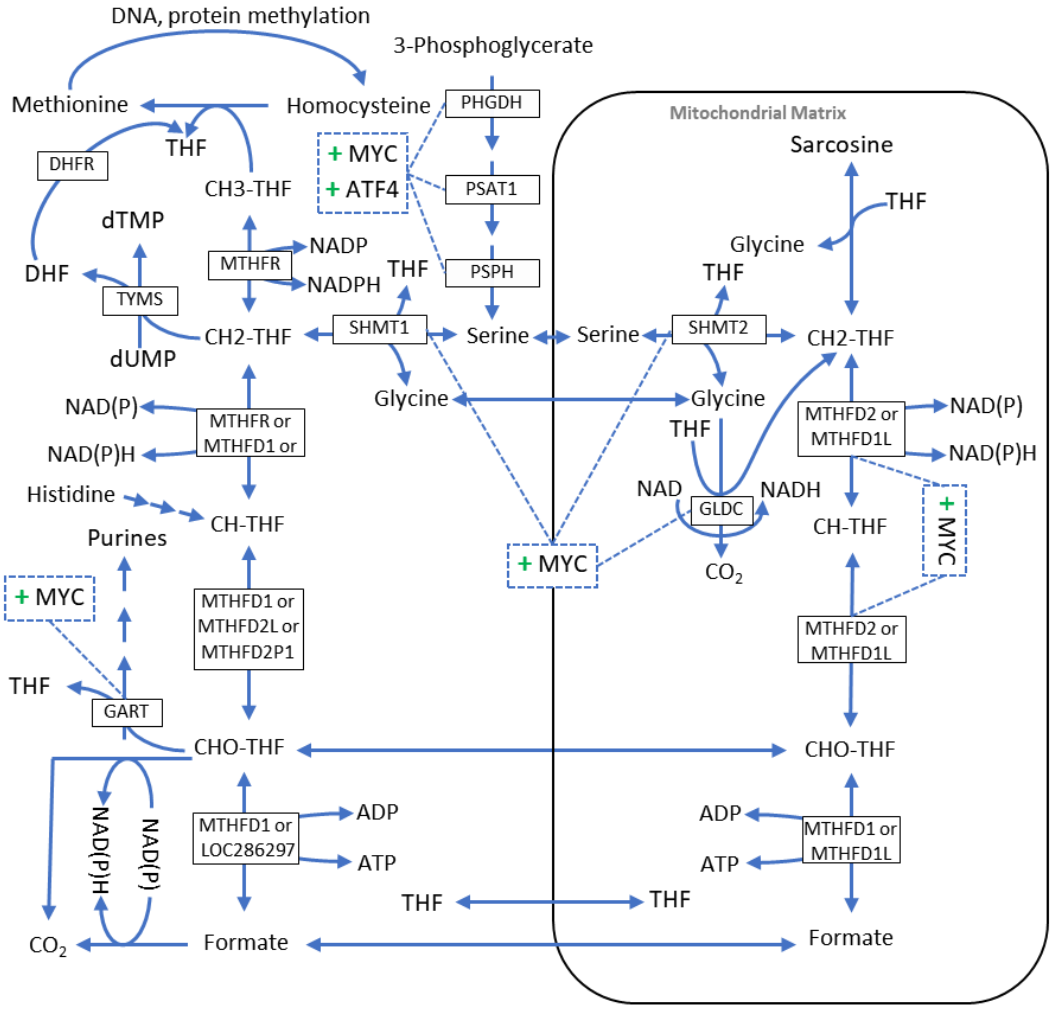


Figure 5. Graphical representation of folate metabolism. The genes catalyzing reactions of interest are indicated inside of white boxes. DHFR: Dihydrofolate Reductase. GART: Trifunctional Purine Biosynthetic Protein Adenosine-3. LOC286297: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 1 Like Pseudogene. MTHFD1: Methylenetetrahydrofolate Dehydrogenase, Cyclohydrolase And Formyltetrahydrofolate Synthetase1. MTHFD1L: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent)1 Like. MTHFD2: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 2, Methenyltetrahydrofolate Cyclohydrolase. MTHFD2L: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent)2 Like. MTHFD2P1: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent)2, Methenyltetrahydrofolate Cyclohydrolase Pseudogene 1. MTHFR: Methylenetetrahydrofolate Reductase. PHGDH: phospho-glycerate dehydrogenase*. PSAT1: Phosphoserine aminotransferase PSPH: Phosphoserine Phosphatase. SHMT1/2: Serine Hydroxymethyltransferase1/2. TYMS: Thymidylate Synthetase.*

1.5.1.4.6. Amino acid metabolism and glutathione metabolism

Overactivation of signaling pathways that promote proliferation, such as PI3K-AKT-mTORC or MYC, lead to increased demand for amino acids for protein synthesis, and to a lesser extent, nucleotide synthesis. To meet such demand, AKT-mTORC1-ATF4 and MYC signaling increase the

supply of amino acids by activating the expression of both amino acid carriers and enzymes from biosynthetic pathways of non-essential amino acids[30,303,307–309].

Additionally, MYC also plays a critical role in glutamine metabolism, activating the transcription of both glutamine carriers(i.e., SLC1A5 and SLC7A5) and glutaminase (GLS)[308,310,311](Figure 4). Glutamine supports rapid proliferation by acting as donor of amine groups in the synthesis of non-essential amino acids and nucleotides, replenishing TCA cycle intermediates lost to biosynthetic pathways as well as acting as a catabolic substrate[10,312]. Because of its role replenishing TCA cycle intermediates, glutaminolysis has also been reported to support redox homeostasis by facilitating NADPH production through the NADPH-dependent malic enzyme which oxidizes the TCA-cycle-derived malate to pyruvate[313].

Even more, glutaminolysis-derived glutamate can support the synthesis of glutathione[314,315], a tripeptide synthesized from glutamate, cysteine and glycine. Glutathione plays a key role in the defense against oxidative stress, is widely associated with resistance to both chemotherapy and radiotherapy[316,317], and often found in higher concentration in cancer cells than in healthy tissues[318]. As part of its antioxidant function, glutathione acts as an electron donor for redox reactions and is reduced to glutathione-disulfide. Glutathione-disulfide can be recycled to glutathione in a NAPDH-dependent manner through the glutathione reductase activity[223,319].

However, even if glutamate contributes to glutathione synthesis, cysteine is assumed to be the limiting substrate for glutathione production[318] and, as such, cysteine supply plays a key role in protecting against oxidative stress. In blood, as in culture media, cysteine is mostly found as cystine (L-dicysteine)[320,321] which is transported inside the cell primarily through the cystine/glutamate antiporter system Xc- (coded by the genes SLC7A11 and SLC3A2)[322] and to a lesser extent, through the cystine/neutral amino antiport acid system $b^{0,+}$ (coded by the genes SLC7A9 and SLC3A1)[323]. In the cytoplasm, cystine is reduced to cysteine primarily by reacting with glutathione[324–326] and to a lesser extent by the thioredoxin reductase system[327,328]. However, cysteine can also be produced from methionine, the other sulfur-containing amino acid, through the transsulfuration pathway.

Transcription of the Xc- components and the enzymes of the transsulfuration pathway are induced both by NRF2 and ATF4[28,329,330]. Similarly, transcription of enzymes involved in glutathione synthesis and glutathione reductase are activated by NRF2 and NF-κB[331,332].

1.5.1.4.7. Targeting cancer metabolism

Because metabolic reprogramming underlies critical aspects of cancer progression, it also represents a set of unique and powerful therapeutic opportunities: cancer cells can be

selectively killed, or their malignancy can be severely limited by identifying and targeting the key players of cancer metabolic reprogramming.

For instance, as part of such reprogramming cancer cells might become dependent on specific pathways. Case in point, many cancer cell populations, particularly those driven by RAS or MYC overactivation, are highly dependent on glutamine and are vulnerable to therapies targeting glutamine metabolism[224,333–335]. Similarly, cancer cells can also become dependent on the system Xc- for cystine import and hence its inhibition has been found to inhibit proliferation[336–340] and tumorigenic potential[341–343] in a variety of cancer types.

Additionally, due to overactivation of folate metabolism, antifolates such as Methotrexate, or Pemetrexed have proven to be highly effective against many cancer types[344–348]. Similarly, some cancer cell populations are vulnerable to inhibition or silencing of specific activities of folate metabolism such as SHMT1, SHMT2, MTHFD1 or MTHFD2[304,349–351].

Finally, targeting features of CSC specific metabolic reprogramming has proven to be an effective strategy to reduce stemness o sensitize cancer cells to chemotherapy or radiotherapy. For instance, inhibiting the Warburg effect, oxidative phosphorylation, glutaminolysis or lipid metabolism have been shown to reduce the tumorigenic potential of prostate CSCs[287,291,333].

1.6. The Systems Biology toolbox for metabolic analysis

In previous sections, it has been established that the metabolic phenotype emerges from the interaction of metabolites, enzymes and signaling pathways. Hence, analyzing a metabolic system through a reductionist approach (i.e., analyzing individual measurements of metabolites or enzymes in isolation) might fail to encapsulate the emerging properties of the system.

Systems Biology is the discipline of biology that studies biological systems in a holistically manner (i.e., as a whole) to study the properties that emerge from the complex interaction of its components and, as such, has been widely applied to the study of metabolism[352,353]. To achieve this, Systems Biology uses mathematical representations of metabolism termed metabolic models. Such models serve as a framework where multiple layers of data (e.g., metabolomics, transcriptomics, proteomics or enzyme kinetic properties) are integrated to simulate different properties of the metabolic phenotype such as metabolic fluxes, metabolites concentrations or the response to a metabolic perturbation. Metabolic models can be broadly classified into two categories, constraint-based or kinetic models[353]. Complementing both modeling approaches, there is ^{13}C MFA (Metabolic Flux Analysis), a set of techniques designed to integrate ^{13}C resolved metabolomics in the framework of either kinetic modeling or constraint-based modeling[354–356].

1.6.1. Mathematical notation of metabolic networks: the stoichiometric matrix

Both kinetic models and constrain-based models use the stoichiometric matrix to represent metabolic networks in mathematical notation. The stoichiometric matrix (*s*), is a matrix of *n* rows and *m* columns where *n* is the number of metabolites in the network and *m* is the number of reactions and transport processes. In this matrix, $s_{i,x}$ represents the stoichiometric coefficient for the metabolite *i* in reaction/transport process *x*. If the metabolite *i* does not participate in the reaction or transport process *x* the coefficient will be 0 (Figure 6).

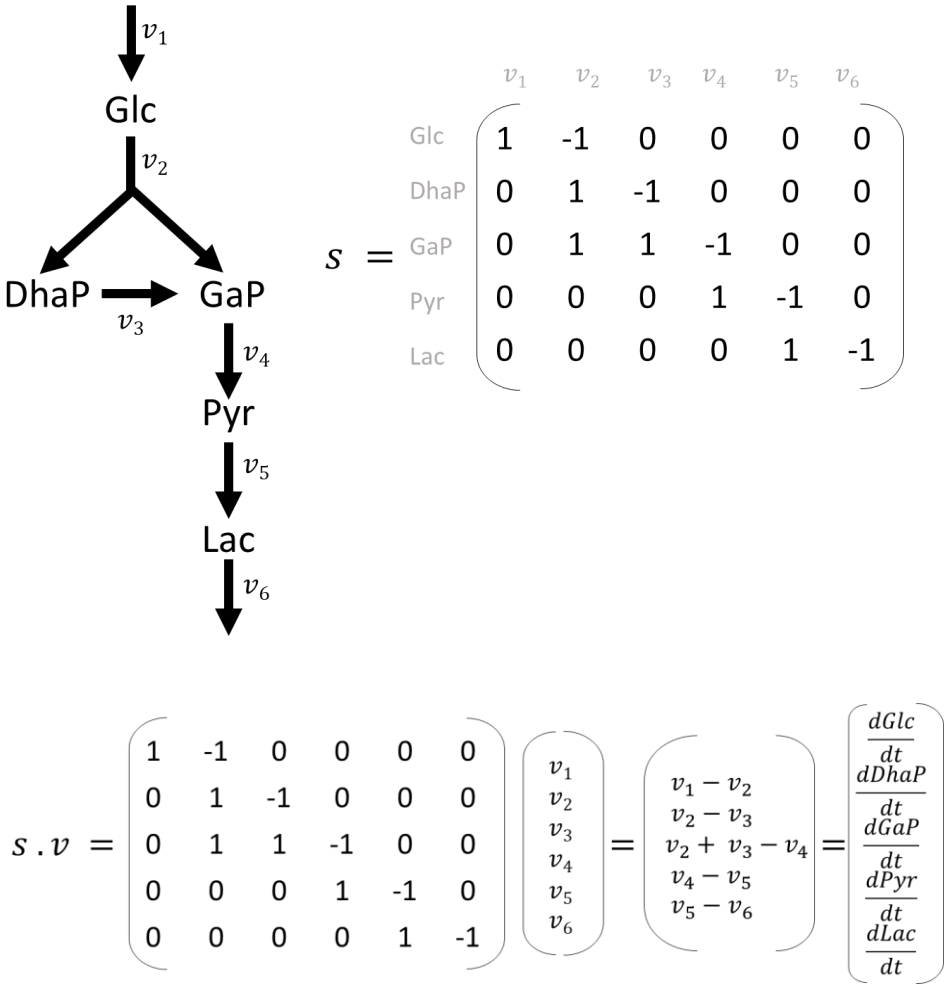


Figure 6. Example of the stoichiometric matrix and the balance equations for a simplified network of Glycolysis. Glc: Glucose. DhAP: Dihydroxyacetone phosphate. GaP: Glyceraldehyde 3-phosphate. Pyr: Pyruvate. Lac: Lactate.

The advantage of using such notation is that the mass balance equations for metabolites (i.e., how metabolites concentration would change in time) (Figure 6) in the network can be written as follows:

$$\frac{dM}{dt} = s.v$$

where

M is a vector of metabolites concentrations.

v is the vector of reaction fluxes.

1.6.2. Kinetic Models

Kinetic models, also known as dynamic models, are metabolic models capable of simulating the evolution of metabolite concentrations and metabolic fluxes in time[357]. As covered in previous sections, the rate through an enzyme-catalyzed reaction depends on the concentration of substrates, products, activators and inhibitors, and the properties and concentration of the enzyme catalyzing it. In kinetic models, this dependency is simulated with mathematical expressions referred to as kinetic laws or rate equations which compute the rate through an enzyme-catalyzed reaction as a function of metabolite concentrations and the amount and kinetic parameters of the enzyme[5]. Substituting the rate equations in the mass balance equation results on a system of ordinary of differential equations (ODE) for metabolite concentrations in time(*t*).

$$\frac{dc[t]}{dt} = s \cdot v(c[t], p)$$

Where *v* is a vector of reaction fluxes, which is computed using rate equations as a function of the vector of metabolite concentrations (c[t]), and enzyme concentrations and properties (defined in the vector kinetic parameters *p*). In such a system, *v* should have units of concentration per time. Solving this system of ODEs from a set of starting metabolite concentrations (c[0]) allows simulating the evolution in time of metabolite concentration and by extension metabolic fluxes.

The system defined above is valid for kinetic models with a single cellular compartment. However, if the kinetic model encompasses multiple cellular or extracellular compartments with different volume, the ODE systems should be corrected to account for this. This need arises because any concentration is a function of volume; hence, concentrations are not conserved when a metabolite moves to a cellular compartment with a different volume. An approach to account for this is to use absolute units for metabolic fluxes (e.g. mmol·min^{-1}) and divide each ODE by the volume (V) of the compartment where the metabolite is located.

$$\frac{dc[t]}{dt} = s \cdot v(c[t], p) \cdot \frac{1}{V}$$

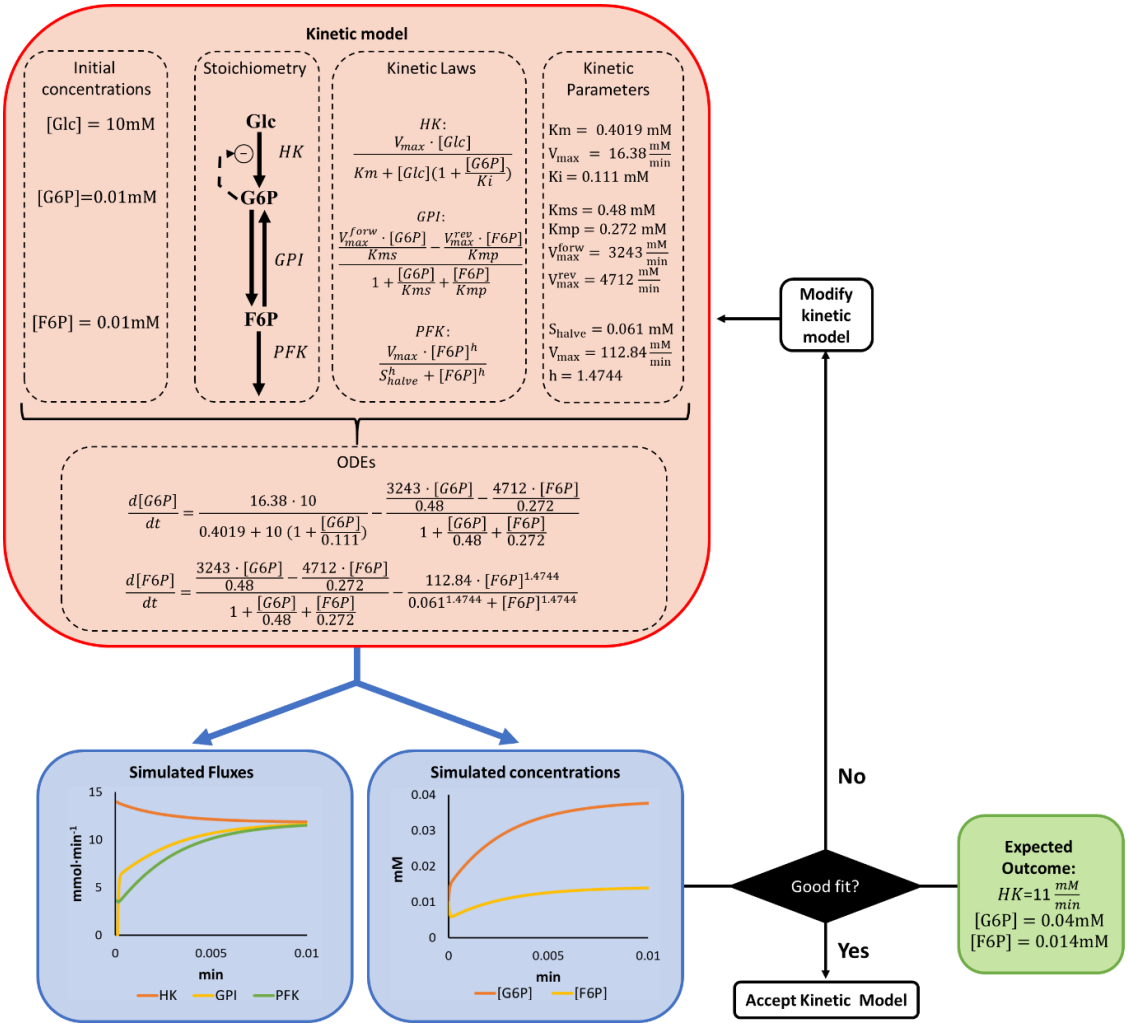


Figure 7. Example of a kinetic model of upper glycolysis build by Puigjaner et al.[358]. The network has three metabolites (Glc: Glucose, G6P: Glucose 6-phosphate, F6P: Fructose 6-phosphate) connected by three reactions (HK: Hexokinase, GPI: Glucose 6-phosphate isomerase, PFK: phosphofructokinase). From network stoichiometry, the parametrized kinetic laws are combined to build a system of ODEs, with each equation describing the dependent dynamic of a metabolite concentration. Starting with initial metabolomic values, solving the system of ODEs simulates time courses for metabolite concentrations and reaction fluxes. The model structure and parameters can be iteratively modified until the simulation is a good fit for the measured concentrations and extracellular fluxes (expected outcome).

1.6.2.1. Kinetic laws

The accuracy of a kinetic model is highly dependent on the quality of its kinetic laws and their parametrization. The simplest and oldest form of kinetic laws is the mass action law where the rate through reactions is computed as the product of a rate constant and the concentration of the substrate raised to the power of their order of reaction[359]. For a mono substrate reaction, the mass action law can be written as:

$$v = k \cdot S^n$$

Where

k is a rate constant

S is the concentration of the substrate of the reaction

n it is the order of the reaction[a].

However, mass action has some severe limitations when it comes to simulating enzyme-catalyzed reactions. For instance, it does not simulate saturation of the enzyme active site, nor can it simulate the effect of activators, inhibitors or cooperativity[b].

This led to the development of mechanistic kinetic laws that implicitly consider enzyme concentration, the turnover number, the affinity of the enzyme for substrates and products as well as the effects of activators and inhibitors. The most iconic of such kinetic laws is the Michaelis-Menten equation which describes the kinetics properties of a mono substrate irreversible[c] enzyme-catalyzed reaction with no activators nor inhibitors[5].

$$v = \frac{e0 \cdot k_{cat} \cdot S}{S + Km_S} = \frac{Vmax \cdot S}{S + Km_S}$$

Where,

$e0$ is the enzyme concentration

k_{cat} is the turnover number or catalytic constant of the enzyme.

Km_S , the Michaelis constant, is the concentration at which 50% of the total enzyme catalytic sites are occupied by S. Such parameter is indicative of the affinity of the substrate for the enzyme active site.

$Vmax$ is the maximal rate through the reaction which is the product of $e0$and K_{cat}.

---

[a] For a simple reaction, the order of reaction for a metabolite can generally be approximated to its stochiometric coefficient.

[b] Some of these limitations would be partially addressed with the development of power laws, which are based on a similar principle as mass action. Power laws raise metabolite concentrations to positive or negative coefficients representing the sensitivity (elasticity) of the reaction rate to the concentration of the metabolites participating or regulating the reaction[598].

[c] *In vitro*, it can also be used to simulate the initial velocity (i.e. the velocity without any product of the reaction present in the medium) of reversible reactions

Similar kinetic laws describe mono-substrate enzyme-catalyzed reactions with cooperativity, reversibility or inhibitions (Table 1).

Table 1: Kinetic laws for mono-substrate reactions. h hill cooperativity coefficient. K_{halve} is the equivalent constant to Km_S for the Hill equation. I is the concentration of inhibitor. K_{ic} and K_{iu} are the inhibition constants for competitive and uncompetitive inhibition. V^{rev} is the maximal rate of the reverse reaction and Km_p is the Michaelis constant for the product of the reaction (P). V^{rev} and Km_p are connected to Vmax and Km_S and the equilibrium constant (Keq) though the Haldane relationship.

Name	Kinetic Law
Hill equation (substrate cooperativity)	$v = \frac{Vmax \cdot S^h}{K_{halve}^h + S^h}$
Competitive inhibition	$v = \frac{Vmax \cdot S}{S + Km_S \cdot (1 + \frac{I}{K_{ic}})}$
Uncompetitive inhibition	$v = \frac{VmaxV \cdot S}{S(1 + \frac{I}{K_{iu}}) + Km_S}$
Mixed inhibition	$v = \frac{Vmax \cdot S}{S \cdot \left(1 + \frac{I}{K_{iu}}\right) + Km_S \cdot (1 + \frac{I}{K_{ic}})}$
Reversible Michaelis-Menten	$v = \frac{\frac{Vmax \cdot S}{Km_S} - \frac{Vmax^{rev} \cdot P}{Km_p}}{1 + \frac{S}{Km_S} + \frac{P}{Km_p}} = \frac{\frac{Vmax}{Km_S} \cdot \left(S - \frac{P}{Keq}\right)}{1 + \frac{S}{Km_S} + \frac{P}{Km_p}}$

Likewise, for bi-substrates and bi-product reactions, several kinetic laws define the kinetic properties based on the order in which substrates bind and products are released from the enzyme active site. In that regard, there are three basic types of mechanisms[5](Table 2):

- Compulsory-order ternary-complex mechanism: Both substrates bind to the catalytic site in a predetermined order forming a ternary complex.
- Random-order ternary-complex: Both substrates bind to the catalytic site forming a ternary complex, but they can bind in any order.
- Substituted enzyme mechanism: A substrate binds to the enzyme and is released after transferring a chemical moiety to the enzyme, next the second substrate binds and accepts the chemical moiety from the enzyme.

Table 2: Kinetic law for bi-substrate reactions. For simplicity, only the kinetic law for irreversible reactions are provided. A and B are the concentrations of substrates. For compulsory-order ternary-complex and substituted enzyme reaction mechanisms, A is the first substrate to bind to the active site. K_{iA} , K_{iB} are constants specific to bi-substrate reactions.

Mechanism	Kinetic Law
Compulsory-order ternary-complex	$v = \frac{\frac{Vmax}{K_{iA} \cdot Km_b} \cdot A \cdot B}{1 + \frac{A}{K_{iA}} + \frac{K_{mA} \cdot B}{K_{iA} \cdot Km_B} + \frac{A \cdot B}{K_{iA} \cdot Km_B}}$
Random-order ternary-complex	$v = \frac{\frac{Vmax}{K_{iA} \cdot Km_B} \cdot A \cdot B}{1 + \frac{A}{K_{iA}} + \frac{B}{K_{iB}} + \frac{A \cdot B}{K_{iA} \cdot Km_B}} = \frac{\frac{Vmax}{K_{iB} \cdot Km_A} \cdot A \cdot B}{1 + \frac{A}{K_{iA}} + \frac{B}{K_{iB}} + \frac{A \cdot B}{K_{iB} \cdot Km_A}}$
Substituted enzyme	$v = \frac{\frac{Vmax}{K_{iA} \cdot Km_b} \cdot A \cdot B}{\frac{A}{K_{iA}} + \frac{K_{mA} \cdot B}{K_{iA} \cdot Km_B} + \frac{A \cdot B}{K_{iA} \cdot Km_B}}$

Enzyme catalyzed reactions with more than two substrates, or products require more complex kinetic laws. For such cases, Liebermeister and Klipp developed an approach termed convenience kinetics that can be applied for any number of substrates and products. This approach assumes a random order mechanism where the affinities of individual reactants do not depend on other reactants already bound to the enzyme[360]. The kinetic law can be written as follows for n_s substrates and n_p products:

$$v = \frac{Vmax \cdot \frac{1}{\prod_i^{n_s} Km_{S_i}} \cdot \left(\prod_i^{n_s} S_i - \frac{\prod_j^{n_p} P_j}{Keq} \right)}{\prod_i^{n_s} \left(1 + \frac{S_i}{Km_{S_i}} \right) + \prod_j^{n_p} \left(1 + \frac{P_j}{Km_{P_j}} \right) - 1}$$

Where,

S and *P* are vectors of concentrations for substrate and products, respectively.

Km_{S_i} and Km_{P_j} are vectors defining the Michaelis constants for the substrates and product, respectively.

It is worth noting that for an irreversible reaction ($P_j = 0$) with two substrates, the expression is equivalent to a Random-order ternary-complex kinetic law where $K_{iS} = Km_S$.

However, for enzymes with complex kinetic mechanisms, there are instances when none of the pre-defined kinetic laws can be adequate. In such instances, a kinetic law can be derived from the kinetic mechanism following the method of King and Altman[5]. For instance, transketolase has a complex reaction mechanism because the enzyme can catalyze multiple reversible reactions, and as such specific kinetic laws for such reactions must be derived using King and Altman[361]. Similar to transketolase, ALDOC can catalyze the reversible cleavage of both fructose

1,6-bisphosphate and fructose-1 phosphate and hence products and substrates for both reactions compete for the same active site. As part of this Ph.D. thesis, we derived a kinetic law to properly simulate this behavior[362].

1.6.2.2. Parametrizing kinetic laws

In order for a model to be functional, the parameters in kinetic laws need to be given numerical values. Over the years, many of such measurements (particularly Km and Ki) for enzymes of central carbon metabolism have been measured *in vitro*. Such information can be readily accessed from BRENDA[363] a vast repository of kinetic information which has been extracted both manually and through text mining from the literature[363]. Similarly, equilibrium constants for metabolic reactions under physiological conditions, which are critical parameters in reversible reactions, can be found both on the literature[364] and dedicated databases[365].

However, $Vmax$ or any equivalent parameters associated enzyme concentration, cannot be generally extracted from the literature as enzyme amounts are dependent on the condition-specific gene expression program. *Vmax* can be estimated from *in vitro* measures of enzyme specific activities in cellular or tissue extracts from the condition of study[358], but it is not feasible to do such measurements for all activities in a sizeable kinetic model. Hence, most of these parameters will generally need to be fitted by identifying the set of parameters that minimizes the difference between simulated and measured metabolite concentrations and fluxes. In this regard, parameter fitting can significantly be enhanced by simulating several conditions and by integrating ^{13}C resolved metabolomics. However, very few kinetic models have the capacity to integrate ^{13}C data to the fitting procedure[361,366], requiring instead many metabolites and flux measurements to achieve meaningful fittings[63,367].

1.6.2.3. Kinetic models of liver metabolism

Due to its role as the master regulator of metabolism[a], over the years, multiple kinetic models of hepatocytes have been developed. However, many of them have a relatively small scope containing only a small number of reactions and often being limited to a single pathway[368–370].

In the last decade, this trend has begun to change with a model covering most of hepatic central carbon metabolism[366] and a highly detailed model of the glycolysis, gluconeogenesis and glycogen metabolism pathways[63]. However, a limitation of both models was that neither of them was able to dynamically simulate the energy and redox metabolism as they considered both ATP, and equivalent metabolites such as GTP or UTP, and NAPD(P)H to be constant. Hence,

[a] See section 1.4

neither model would be able to simulate any perturbation that severely perturbates energy metabolism. This would be partially addressed in 2018 with HEPATOKIN1[367] a large-scale kinetic mode hepatic metabolism capable of dynamically modeling energy and redox dynamics.

However, the existing models of hepatocyte central metabolism[63,366,367] have a limited capacity to simulate fructose metabolism as they do not consider the activator effect of fructose 1-phosphate on GK nor the interdependence of fructose 1,6-bisphosphate and fructose 1-phosphate cleavage emerging from the fact that both reactions are catalyzed by ALDOB.

1.6.3. Constraint-based modeling

As it has been established in the previous section, kinetic models are excellent platforms to integrate all the multiple layers of molecular regulation that lead to the emergence of metabolic phenotype. However, they are limited by the complexity to build and parameterize kinetics laws for large networks beyond central metabolism where little kinetic information might be available.

Furthermore, most systems evolve towards a metabolic steady state where metabolites do not accumulate nor deplete; hence the fluxes leading (inputs) and emerging (outputs) from any given metabolite must be balanced, and steady-state flux distributions will be largely determined by network stoichiometry[12]. This is the basis of constraint-based modeling, which uses the network stoichiometry to simulate steady-state flux distributions[371–373]. Additional constraints in the form of upper and lower bounds for fluxes can also be added. Such bounds can be used to integrate thermodynamic constraints (i.e., reversibility of reactions), the availability of metabolites in the extracellular media and experimental flux measurements. Hence, a constraint-based model can be formulated as follows:

$S \cdot v = 0$, $lb < v < ub$

Where:

v is a vector of steady-state fluxes

lb and *ub* are vectors with the upper and lower bounds for fluxes

This allows formulating the metabolic network as a set of linear equations, that depends only on flux values and require no prior knowledge of metabolite concentrations, enzymes amount or kinetic and mechanistic properties of enzymes. This makes constraint-based modeling particularly suited to studying large metabolic networks such as GSMM (Genome-Scale

Metabolic Models). Such linearity also allows solving and analyzing constraint-based models using linear programming approaches.

1.6.3.1. The validity of the steady-state assumption

As described above, the steady-state assumption is the basis of constraint-based modeling. Therefore, it is crucial to define the extent of conditions where this assumption, and by extension constraint-based modeling, are applicable.

The metabolic steady state is defined as the state where all metabolic fluxes and metabolite concentrations remain constant in time. Hence, such state can only be truly achieved if extracellular metabolite concentrations remain constant. *In vitro,* this can be achieved in cell culture systems where the extracellular medium is continuously replaced by fresh medium but not for the most common cell culture systems. However, due to the significant disparity between intracellular and extracellular volume in cell cultures, it can generally be assumed that intracellular fluxes and concentrations quickly reach a metabolic steady state with respect to varying external concentrations. Likewise, changes in enzyme concentrations mediated by gene regulatory networks also occur at a much slower time scales than the intracellular metabolic transition towards steady state. Hence, with a few exceptions (e.g., metabolites with slow dynamics or oscillating systems), at any given point in time, steady state can be generally assumed for intracellular metabolites and fluxes. This is known as the quasi-steady-state assumption[374,375].

Additionally, cells growing in exponential phase with standard culture medium can generally be assumed to be close to a metabolic steady state both for extracellular and intracellular fluxes provided that no nutrient from the medium is close to being depleted. This can be easily validated by measuring growth rates and extracellular fluxes at multiple time points to confirm exponential growth.

Going beyond the quasi-steady-state assumption, starting from the principle that no metabolite can accumulate to excessive concentrations in the cell, then for extended time periods every metabolite should be produced, on average at the same rate at which is consumed. Therefore, it becomes apparent than the average flux distribution will always be steady-state balanced[375].

Finally, it is worth noting that in proliferating systems, there can be a small deviation from steady-state associated with proliferation-induced dilution. In other words, metabolites must be slightly overproduced to compensate for the increased intracellular volume associated with cellular proliferation in order to maintain constant concentrations. However, for most metabolites, such deviation is generally assumed to be negligible. Nevertheless, such dilution

can also be corrected through the integration of condition-specific metabolomic measurements[a375].

1.6.3.2. Genome-Scale Metabolic Models (GSMMs)

The metabolism of any given organism consists mainly of enzyme-catalyzed reactions, which are the product of the expression of genes coded in the genome of the organism. Hence the metabolic potential of any given organism, the metabolic reactions, and pathways that can potentially be active, is determined by the nucleotide sequence of its genome. Then, for any organism with a sequenced and sufficiently annotated genome, it is feasible to obtain a metabolic network with all possible metabolic reactions for such organism. The resulting network is known as a Genome-Scale Metabolic Model (GSMM). A GSMM represents the entire known metabolic reaction network and its gene-associations for a given organism[372,376]. Specifically, GSMMs contain the information on:

- The metabolic network: a set of reactions with defined stoichiometry. Reactions and metabolites can be distributed across several subcellular compartments and processes transporting metabolites across compartments are also included. The network stoichiometry is encapsulated into the stoichiometric matrix.
- Reaction reversibility: whether reactions are reversible (i.e., they can proceed forward or reverse) in the phycological range of concentrations for substrates and products.
- Gene-protein reaction rules (GPR): a set of rules which indicate the association between genes and reactions (i.e., which gene products are necessary or sufficient to catalyze an enzyme-catalyzed reaction). GPR are a set of Boolean rules between genes where the Boolean operator AND is used to define that two gene products are necessary (i.e., must form a protein complex) to allow the catalytic activity and the operator OR is used to indicate that any of the two isoenzymes or protein complexes are sufficient to allow catalytic activity(Figure 8). The GPR are used to simulate the effect of gene knockouts (KOs) and provide the framework for integrating gene expression data (transcriptomics and proteomics) into GSMMs[376].

[a]See section 1.6.3.3.4

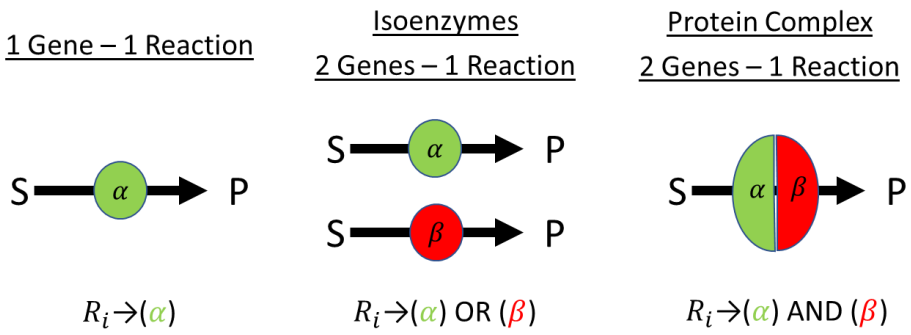


Figure 8. Example of the elemental GPR rules

Additionally, GSMM can also contain annotations that facilitate interpreting the results such as metabolite names, metabolite chemical formulas, reaction names, or reaction pathways.

1.6.3.2.1. Genome-scale metabolic reconstruction

A genome-scale metabolic reconstruction, the process of building a GSMM, begins from a well-annotated genome. From gene annotations, metabolic genes are identified using gene ontology terms or keywords and then mapped to reactions using enzyme commission numbers[377] or biochemical databases like KEGG[378] or BRENDA[363]. This process generates an initial draft of the network and can be automated[379,380]. The initial draft must then be curated using biochemical databases and available literature to correct for erroneous or missing annotations[381]. The curation involves the following steps:

- Manual curation of pathways to ensure that pathways reported in databases and the literature for the target organism are present in the network.
- The balance of charge and mass. First metabolite formulas are corrected to consider the ionization state of metabolites in physiological pH. Then protons, water or any other missing metabolites are added to reactions to ensure that both the charge and mass are conserved between substrates and products in all reactions in the model.
- Adding reactions that are not catalyzed by enzymes (spontaneous reactions).
- Assignment of reactions to a cellular compartment (e.g., cytosol or mitochondrial matrix). This can be done based on literature, gene ontology annotations for cellular localization or using computational tools capable of predicting the most likely localization of an enzyme-based on its amino acid sequence[382].
- Assignment of reaction reversibility. This is done based on the literature or by computing the Gibbs free energy from the formation energies of substrates and products[383].
- Removal of gaps and dead ends in the metabolic network (i.e., metabolites that cannot be produced or consumed).

Finally, the reconstruction is validated by simulating metabolic functions for the target organism reported in the literature. If validation fails, the model is further curated until the model is able to simulate the reported functions.

Over a hundred GSMMs have been built for different species, ranging from Archea to Mammals[384–388] and among them several reconstructions of human metabolism[389–393]. If a previous reconstruction for a given organism is available, it is frequent to build new GSMMs starting from existing reconstructions. For instance, concerning human GSMMs, both Recon2[390] and the Human Metabolic Reaction database (HMR)[391] were built starting from Recon1[389]. Likewise, Recon2.2[392] was constructed from Recon2 and Recon3D[393] was built from Recon2.2 and HMR[391]. Similarly, GSMMs can be built starting from reconstructions build for other species provided that the genome from any of the species with existent models has high enough homology with the genome of the target organism[379,380].

1.6.3.2.2. Biomass reaction

A key part of a GSMMs is the biomass reaction, an artificial reaction that represents the demand of metabolic energy and building blocks (i.e., amino acids, nucleotides, lipids and carbohydrates) needed to produce the macromolecules necessary for growth and proliferation. The coefficients of building blocks are derived from the relative abundance of each component in the target organism[394]. For instance, Recon2.2 has the following biomass reaction:

0.505626 L-alanine + 0.35926 L-argininium + 0.279425 L-asparagine + 0.352607 L-aspartate + 20.704451 ATP + 0.020401 cholesterol + 0.011658 cardiolipin + 0.039036 CTP + 0.046571 L-cysteine + 0.013183 dATP + 0.009442 dCTP + 0.009898 dGTP + 0.013091 dTTP + 0.275194 D-Glucose 6-phosphate + 0.325996 L-glutamine + 0.385872 L-glutamate + 0.538891 Glycine + 0.036117 GTP + 20.650823 Water + 0.126406 L-histidine + 0.286078 L-isoleucine + 0.545544 L-leucine + 0.592114 L-lysinium + 0.153018 L-methionine + 0.023315 1-phosphatidyl-1D-myo-inositol + 0.154463 Phosphatidylcholine + 0.055374 phosphatidylethanolamine + 0.002914 phosphatidylglycerol + 0.259466 L-phenylalanine + 0.412484 L-proline + 0.005829 phosphatidylserine + 0.392525 L-serine + 0.017486 sphingomyelin betaine + 0.31269 L-threonine + 0.013306 L-tryptophan + 0.159671 L-tyrosine + 0.053446 UTP + 0.352607 L-valine → 20.650823 ADP + 20.650823 proton + 20.650823 hydrogenphosphate

It assumes that the dry weight of a human cell consists of 70.6% protein, 1.4% DNA, 5.8% RNA, 7.1% carbohydrates and 9.7% lipids. The relative abundance of amino acids and nucleotides was derived from analyzing the nucleotide and amino acid sequence of open reading frames in humans. Lipid composition was taken from the literature. ATP requirements, other than the

small amount needed as nucleotides, were derived from the estimated cost of protein synthesis (i.e., 4 ATP per peptidyl bond and an average protein length of 333 amino acids)[390].

Biomass composition can vary between cell types and conditions. Therefore, this equation is only an approximation that, ideally, should be modified to accommodate the specific requirements of the cell or tissue being modeled. However, given the complexity of measuring biomass composition, this is not always feasible.

1.6.3.3. Using constraint-based modeling to simulate flux distributions at a genome-scale

In GSMMs there are more reactions than metabolites[389–393] and even with directionality constraints, the steady-state constraint is generally not enough to obtain a unique valid steady-state flux distribution. Instead, there is a large space of feasible solutions that satisfy both the steady-state and directionality constraints. This has led to the development of several approaches that can be applied to narrow the solution space and select the biologically relevant solutions for any given condition. For instance, flux balance analysis (FBA) and its derivative parsimonious FBA (pFBA) can be applied to identify the solutions that maximize or minimize a given set of fluxes representing a biological objective (e.g., proliferation).

Additionally, a GSMM describes all the metabolic reactions that are possible in a given organism. However, in any given condition and cell type, only a subset of metabolic pathways is active. Hence, the solution space can be greatly restricted by building a condition-specific GSMM. This can be achieved by integrating metabolomics and gene expression measurements (transcriptomics or proteomics) in the framework of a generic GSMM such as Recon2.2[392](Figure 9).

Together, FBA-like approaches coupled with the integration of multiple layers of omics data in the framework of GSMMs allow computing metabolic flux distributions and simulating the effect of metabolic perturbations at a genome-scale. An overview of the constraint-based modeling algorithms that can be used for this purpose is provided in the following sections.

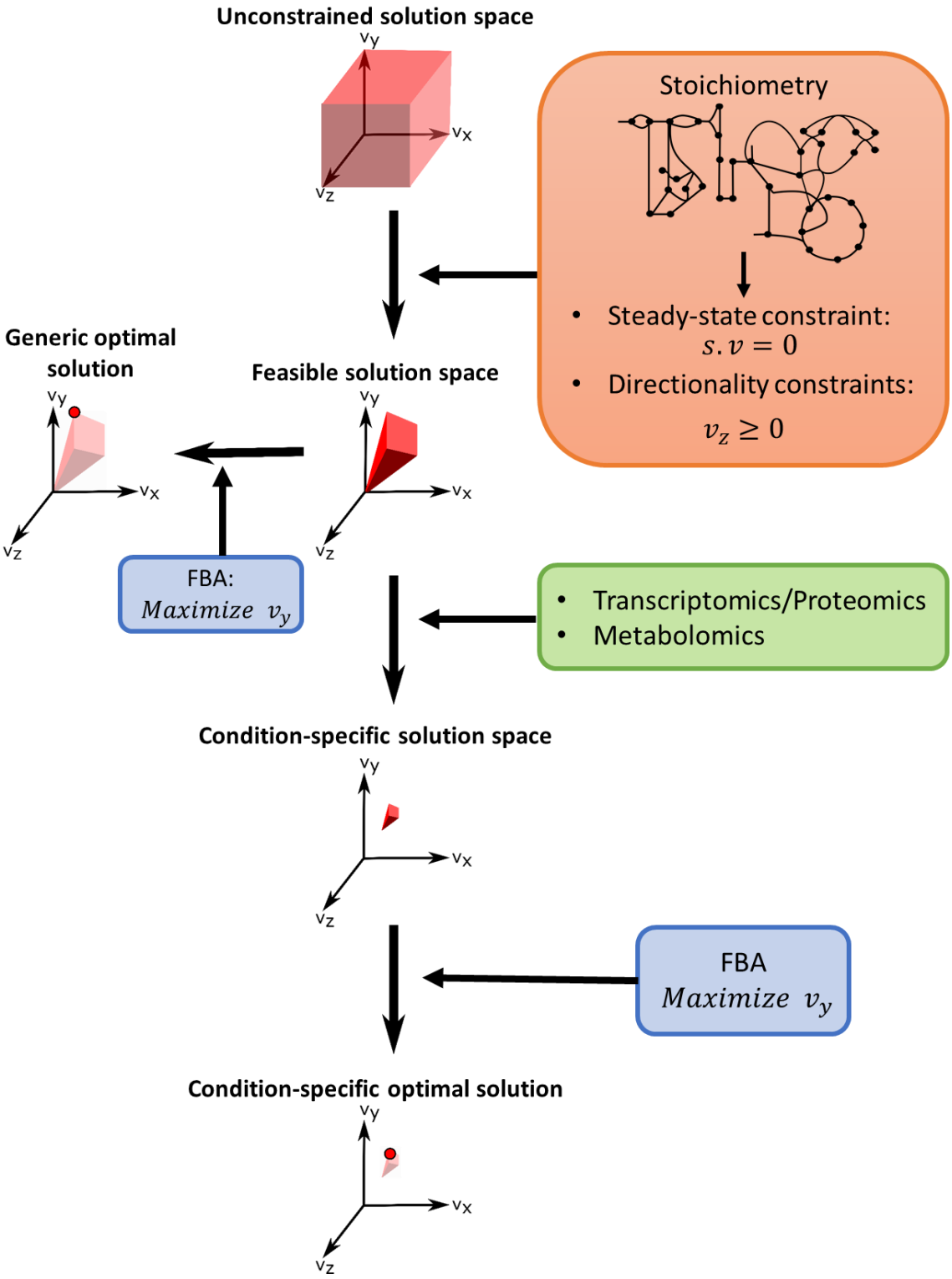


Figure 9. Condition-specific constraint-based modeling. From an infinite space of possible flux (v) solutions, a feasible solution space which contains possible steady-state solutions is obtained by applying the steady-state constraint (s.v=0) and defining the directionality of reactions. From such space, a generic optimal solution can be obtained through FBA by optimizing a biologically desirable phenotype such as proliferation. Alternatively, a condition-specific solution space can be obtained by integrating condition-specific omics like transcriptomics, proteomics or metabolomics. Finally, FBA can be performed in the solution space to select the condition-specific optimal solution(s).

1.6.3.3.1. Flux balance analysis (FBA)

FBA is the most widespread application of constraint-based modeling. FBA is based on the assumptions that living organisms have evolved towards maximum efficiency and hence will tend to have flux distributions that optimize a desirable trait. Accordingly, FBA identifies the steady-state flux distribution that maximizes or minimizes a set of reaction fluxes which represent a biological objective deemed desirable for the conditions of study. For highly proliferating models, such as cancer cells, the optimized attribute is generally proliferation and growth. This is represented by maximizing the flux through the biomass reaction. For non-proliferating systems, ATP production can be optimized (i.e., maximize energetic efficiency). Alternatively, tissue-specific fluxes can be maximized or minimized (e.g., maximizing ammonia detoxification or bile acid synthesis to simulate liver metabolism)[371–373,395].

The FBA optimization can be formulated as follows:

$$FBA_{opt} = \max\ (v \cdot c)$$

Subject to:

$$s \cdot v = 0$$

$$lb \leq v \leq ub$$

where:

c is a vector defining the maximization or minimization weight for each reaction flux. Positive coefficients indicate that the reaction will be maximized, whereas negative objective coefficients indicate that a reaction will be minimized. Reactions with an objective coefficient of 0 will neither be maximized or minimized.

FBA_{opt} is the optimal value of the FBA objective.

FBA has two significant limitations. The first limitation lays on the difficulty of defining an appropriate set of fluxes to be maximized for the condition and cell or tissue under study. While it can be assumed that most living systems aim towards an optimal phenotype, this metabolic phenotype can sometimes be complex to simulate through a set of fluxes to be maximized or minimized. The second limitation is that there is usually a space of solutions that can be optimal. Such space emerges both from having redundant pathways and from pathways the activity of

which does not affect the objective. This solution space can either be evaluated using flux variability analysis(FVA)[396] or, alternatively, reduced using pFBA[397].

1.6.3.3.2. Flux variability analysis (FVA)

FVA evaluates the possible range of variation for each flux under FBA[396]. It consists of solving two optimization problems for each flux of interest. One optimization problem finds the maximum feasible flux value(vmax_i^{FVA}):

$$\mathrm{vmax}_i^{FVA} = \max\ v_i$$

$$subject\ to\ s.v\ =\ 0\,, \qquad lb \le v \le ub, \qquad v \cdot c \ge FBA_{opt} \cdot (1 - T^{FBA})$$

And the second optimization problem finds the minimal feasible flux value (vmin_i^{FVA})

$$\mathrm{vmin}_i^{FVA} = \min\ v_i$$

$$subject\ to\ s.v\ =\ 0\,, \qquad lb \le v \le ub, \qquad v \cdot c \ge FBA_{opt} \cdot (1 - T^{FBA})$$

Where:

T^{FBA}: Is the tolerance of the primary FBA objective (i.e., how much can the solution deviate from the optimal FBA solution).

vmin_i^{FVA}: is the minimum flux value allowed for flux *i* with tolerance T^{FBA} ;

vmax_i^{FVA}: is the minimum flux value allowed for flux *i* if tolerance T^{FBA};

The result of FVA is the minimum (vmin_i^{FVA}) and maximum (vmax_i^{FVA}) value allowed for each flux, which provides a direct estimation of the FBA solution space with the current objective (c). When all objective coefficients are set to 0 or T is set to 1, FVA can be used to compute the range of feasible fluxes with a set of constraints (i.e. lb , ub). In this instance, FVA can also be referred to as flux spectrum analysis (FSA)[398].

1.6.3.3.3. Parsimonious flux balance analysis (pFBA)

pFBA[397] consists on running a second optimization in the optimal FBA solution space to select the solution(s) that allow achieving the optimal FBA objective with less total reaction flux. pFBA assumes that biological systems will tend towards solutions that minimize total reaction flux as they will generally require less total enzyme amount to be achieved. Hence, pFBA selects the most efficient solution(s) within the space of optimal FBA solutions.

As the objective of pFBA is to minimize absolute flux values, the model must be first converted into an irreversible GSMM where all reactions have a positive value. This is achieved by splitting

each reversible reaction into a forward and backward reaction. Then pFBA can be performed on the irreversible GSMM as follows:

$$\min \sum_i v_i^{irrev}$$

$$subject\ to:$$

$$v^{irrev} \cdot c^{irrev} = FBA_{opt}$$

$$sS^{irrev}.v^{irrev} = 0, lb^{irrev} \leq v^{irrev} \leq ub^{irrev}$$

Where,

$s^{irrev}, v^{irrev}, lb^{irrev}, c^{irrev}$ are the stoichiometric matrix, steady-state flux distributions, lower bounds, upper bounds and objective coefficients for the irreversible model respectively.

1.6.3.3.4. Integrating metabolomics

Metabolomics, understood as concentration measurements for extracellular or intercellular metabolites, can be integrated into GSMMs to build condition-specific GSMMs consistent with such measurements. Extracellular metabolomics measured at different time points can be used to constrain the extracellular fluxes (i.e., rates of uptake of secretion for extracellular metabolites) defined in the GSMM. For non-proliferating systems such rates can be estimated as follows[224]:

$$v_{M_{ex}} = \frac{M_1 - M_0}{N \cdot (t_1 - t_0)}$$

Where:

M_1 and M_0 are metabolite abundances measured at time points t_1 and t_0, respectively;

N is the total cell number.

$v_{M_{ex}}$ is the estimated rate at which the metabolite M is produced/consumed per cell.

For proliferating systems, the rate at which metabolites are consumed or produced per cell depends on the growth rate and can be computed as follows:

$$\mu = \frac{ln\left(\frac{N_1}{N_0}\right)}{t_1 - t_0}$$

$$v_{M_{ex}} = \frac{M_1 - M_0}{N_1 - N_0} \cdot \mu$$

Where

μ is the growth rate

N_1 and N_0 are the cell numbers measured at time points t_1 and t_0, respectively.

Furthermore, in proliferating systems, metabolomics measured in the cellular pellet can be integrated with the growth rate to account for the dilution associated with proliferation[a]. In the framework of constraint-based modeling, this can be represented by adding a sink reaction (i.e., a reaction consuming the measured metabolites). Such reaction constraints the model to have net production of the measured metabolite to maintain the concentrations of such metabolite constant while proliferating[375].

$$v_{M_{sink}} = M \cdot \mu$$

Where:

$v_{M_{sink}}$ is the flux through the sink reaction for metabolite M.

1.6.3.3.5. Integrating gene expression

As described in section 1.2.1.2, a key component of metabolic regulation emerges from the enzymes expressed in a given condition. Hence, a crucial part of building condition-specific GSMMs is integrating gene expression measurements (e.g., transcriptomics or proteomics). In the framework of GSMM, such measures can easily be mapped to metabolic reactions using GPR[399,400]. Then, algorithms, such as iMAT or GIMME, can be applied to identify the flux distribution(s) most consistent with the gene expression profile in a given condition[392,400–404].

1.6.3.3.5.1. Integrative Metabolic Analysis Tool (iMAT)

iMAT (Integrative Metabolic Analysis Tool)[399,401] is an algorithm that seeks to identify the map of active/inactive enzyme-catalyzed reactions most consistent with gene expression data. iMAT classifies reactions as either lowly expressed or highly expressed if they have a gene expression value above a "high expression threshold" or below a "low expression threshold", respectively. The iMAT optimization maximizes both:

a) The number of reactions catalyzed by highly expressed enzymes (*hex*) which are active
b) The number of reactions catalyzed by lowly expressed enzymes (*lex*) which are inactive

[a] See section 1.6.3.1

Enzymes are defined as active if they carry flux above a given threshold (ε^{hex}) and inactive if they carry flux below a certain threshold (ε^{lex}).

This can be formulated as the following mixed linear integer problem:

$$imat_{opt} = \max\left(\sum_{i \in hex} (b_i^{hex+} + b_i^{hex-}) + \sum_{i \in lex} b_i^{lex}\right)$$

Subject to:

$$s \cdot v = 0$$

$$lb \leq v \leq ub$$

$$v_i + b_i^{hex+}(lb_i - \varepsilon) \geq lb_i \ , \qquad i \in hex$$

$$v_i + b_i^{hex-}(ub_i + \varepsilon) \leq ub_i \ , \qquad i \in hex$$

$$lb_i\left(1 - b_i^{lex}\right) \leq v_i \leq ub_i\left(1 - b_i^{lex}\right) \ , \qquad i \in lex$$

$$b_i^{hex}, b_i^{lex} \in \{0,1\}$$

where,

$imat_{opt}$ is the optimal value of the iMAT objective

b_i^{hex+} and b_i^{hex-} are the binary variable associated with highly expressed reaction *i*. b_i^{hex+} is associated with positive flux values and can only be 1 if v_i is larger than ε. Conversely, b_i^{hex-} can only be 1 if v_i is lower than -ε.

b_i^{lex} is the binary variable associated with lowly expressed reaction *i*. It can only be 1 if v_i is 0.

iMAT has several inherent limitations. The first limitation is that the resulting flux distribution is overly dependent on ε. Additionally, the binary approach used to classify reactions as highly expressed or lowly expressed is also highly dependent on the gene expression thresholds used and is unable to integrate differences of gene expression above the high expression threshold or below the low gene expression threshold. Because of this, iMAT will only be able to generate differentiated reconstructions for cell types with profoundly different gene expression profile (e.g., from different tissues). This is because in conditions with similar gene expression profiles it is unlikely that many (if any) enzymes will switch from being highly expressed to being lowly

expressed and only a handful of enzymes will move below or above the high or low expression threshold.

1.6.3.3.5.2. Gene Inactivity Moderated by Metabolism and Expression (GIMME)

The GIMME[400] (and its derivative GIM^{3}E[405]) is an algorithm that applies the flux minimization approach of pFBA using gene expression data to give greater weight to the minimization of fluxes through reactions catalyzed by lowly expressed enzymes. The minimization weight for each reaction is defined as follows:

$$w_i = \max\left(ge^{th} - ge_i, 0\right)$$

where,

ge_i is the gene expression value of reaction *i*.

ge^{th} is the gene expression value below which reactions will be penalized. To maximize the information extracted from gene expression data such a threshold can be set at the maximum expression value found in metabolic reactions[405].

w_i is the minimization weight for reaction *i*

Then, the GIMME solution is obtained by solving the following linear programming problem.

$$\text{GIMME}_{opt} = \min \sum_i w_i \cdot v_i^{irrev}$$

$$subject\ to:$$

$$v^{irrev} \cdot c^{irrev} = FBA_{opt} \cdot (1 - T^{FBA})$$

$$S^{irrev} . v^{irrev} = 0 , lb^{irrev} \leq v^{irrev} \leq ub^{irrev}$$

Where,

GIMME_{opt} is the optimal value of the GIMME objective.

Additionally, a similar approach to FVA can be applied to estimate the range of flux solutions compatible with a given tolerance for the GIMME objective[405].

Compared to iMAT, GIMME has less reliance on arbitrary thresholds and is generally a better choice for building quantitative flux maps. However, it will also struggle at generating differential flux distributions in conditions with similar gene expression profiles.

1.6.3.4. GSMMs in cancer drug discovery

Analyzing GSMMs in the framework of constraint-based modelling has a wide array of biotechnological and biomedical applications[406,407]. Indeed, cancer-specific GSMM have received particular interest as platforms for identifying putative metabolic targets that can be used in therapies against cancer[371,408]. In this regard, two approaches can be taken: 1) identifying metabolic vulnerabilities that selectively kill cancer cells[404,409–411] or 2) identifying metabolic features that can be targeted to revert cancer progression[403,412].

1.6.3.4.1. Identifying cancer metabolic vulnerabilities: Essential and synthetic lethal genes

Metabolic vulnerabilities of cancer cell are closely associated with essential and synthetic lethal gene pairs. Essential genes are defined as genes whose function is required for cell survival in a given condition. Conversely, a set of genes are synthetic lethal if the isolated loss of function on any of them is compatible with cell viability, but the simultaneous mutation of all genes in the set is lethal[376]. Hence, cancer-specific essential and synthetic lethal genes represent vulnerabilities that can be targeted to selectively kill cancer cells[376,409]. Furthermore, they can also provide a greater understanding of cancer biology and its dependencies[413,414].

With the goal of identifying such dependencies, systematic knockdown of genes have been performed for some cancer cell lines, although limited to a subset of genes[413,414]. To date, the most comprehensive screening is project DRIVE (deep RNAi interrogation of viability effects in cancer) in which 7837 human genes were targeted in 398 cancer cell lines[414]. Synthetic lethal gene pair screenings have also been performed; however they are limited to a small set of the total possible gene combinations[415–417]. In this regard, SynLethDB a database that gathers reported synthetic lethal information from the literature has reported only a total of 16196 synthetic lethal gene pairs for human cell lines[418].

In this regard, GSMMs have emerged as *in silico* platforms where putative metabolic essential and synthetic lethal genes can be identified without requiring large scale *in vitro* gene screenings. For instance, Folger *et al*. used a cancer-specific GSMM reconstruction that identified 52 potential drug targets, many of which were targets of known anticancer drugs[409]. Likewise, Agren *et al*. built patient-specific GSMMs of hepatocellular carcinoma and identified 101 potential drug targets[410].

To predict essential and synthetic lethal genes using GSMMs, GPR rules are used to predict which reactions will be blocked when a given set of genes is inactivated (gene KO), and constraint-based modeling is used to evaluate its effect on growth and proliferation (i.e., biomass reaction). There are primarily three constraint-based modeling approaches that can be used to simulate

the effect of gene KO in the framework of GSMM: FBA, Minimization of metabolic adjustment (MOMA) and minimal cut set analysis (MCS).

1.6.3.4.1.1. Simulating gene KOs with FBA

A gene KO can be simulated by running FBA with the reactions predicted to be inactive ($KO^{reactions}$) constrained to have a flux value of 0.

$$FBA_{opt}^{KO} = \max\ (v \cdot c)$$

Subject to:

$$v_i = 0\,, \qquad i \in\ KO^{reactions}$$

$$S \cdot v = 0, lb \leq v \leq ub$$

Where:

FBA_{opt}^{KO} is the optimal value of the biological objective after inducing the KO

A gene is considered essential if FBA_{opt}^{KO} is lower than a predefined threshold (e.g. 10% of FBA_{opt})

Simulating a KO using FBA assumes that the target cell will reprogram the metabolism to find the optimal flux distribution after the KO. The essential genes and synthetic lethal predicted under this method are vulnerabilities inherent of the condition-specific GSMM used and do not consider the initial flux state nor the complexity of readjusting the metabolism after the gene KO.

1.6.3.4.1.2. Simulating gene KO with Minimization of Metabolic Adjustment (MOMA)

Minimization of metabolic adjustment (MOMA) is a method to simulate the effect of a gene KO assuming that when subjected to gene KO, metabolic fluxes will readjust to a state that requires the minimal redistribution with respect to the flux configuration prior to the perturbation (i.e., wild type flux distribution)[419]. This can be achieved by solving a quadratic linear programming problem that seeks the flux distribution that minimizes the distance between the wild type flux distribution and the flux distribution after inactivating the reactions associated to the gene KO:

$$Min \sum_i \left(v_i^{wt} - v_i\right)^2$$

Subject to:

$$v_i = 0\,, \qquad i \in KO^{reactions}$$

$$S \cdot v = 0\,,\, 0 < lb < v < ub$$

Where v_i^{wt} is a reference (i.e., wild type) steady-state flux distribution before the gene KO. Such distribution can be obtained by any constraint-based modeling technique such as FBA, pFBA or GIMME.

Unlike when gene KO are simulated with FBA, flux distributions simulated with MOMA are not necessarily optimal. The basis of MOMA is that although the assumption of optimality for a wild-type cell is justifiable due to natural or clonal selection, this may not be valid for and induced KO not yet subjected to long-term natural or clonal selection[419].

Predicting essential genes or synthetic lethal sets with MOMA will always provide at least the same essential genes or synthetic lethal predicted by FBA plus additional essential or synthetic lethal genes that will depend on the initial flux distribution. Hence, the accuracy of such predictions will significantly depend on the reference flux distribution used.

1.6.3.4.1.3. Minimal Cut Sets analysis (MCS)

MCS analysis is an alternative to FBA and MOMA to predicting essential genes and synthetic lethal sets. MCS are minimal sets of reactions or genes whose removal blocks a target reaction[420,421]. Essential genes and synthetic lethal gene sets can be considered as MCSs targeting a reaction(s) required for survival or proliferation (e.g., biomass reaction).

MCSs can be computed by formulating the dual problem of a constraint-based problem and computing its "elementary flux modes" (i.e., the minimal set of non-zero variables that allow a valid solution). As per the Farkas Lemma theorem if there exists a solution for the primal problem there can be no solution to the dual problem. Hence the "elementary flux modes" of the dual problem will be MCSs of the constraint-based problem[420–422].

MCS analysis predicts only the essential and synthetic lethal genes that emerge from the stoichiometry of the network. Hence, unlike FBA or MOMA it cannot predict essential or synthetic lethal genes that emerge from limited enzyme capacity (i.e., maximum flux boundaries), limited availability of substrates or measured metabolic fluxes. However, MCS analysis is much more computationally efficient at finding synthetic lethal sets with more than two members than a systematic search using FBA or MOMA since the number of combinations to evaluate with the latter methods growths exponentially with set size.

1.6.3.4.2. Targeting metabolic features associated with cancer progression: The metabolic transformation algorithm

The metabolic transformation algorithm (MTA) is an algorithm that can be used to identify the best therapeutic interventions to switch from a source state to a desirable metabolic state[423]. As such, one of its many applications is to identify therapeutic interventions that can attenuate an aggressive cancer phenotype (e.g., a CSC phenotype) by inducing a metabolic switch to the metabolic phenotype of a less aggressive cell population[412]. Unlike essential and synthetic lethal genes, such targets might not necessarily kill the analyzed cancer cells; however, by attenuating their aggressiveness, they might increase their sensitivity to radiotherapy or chemotherapy.

MTA achieves this by integrating transcriptomics to identify pathways coordinately upregulated or downregulated at the gene expression level and then identifying the therapeutic interventions (i.e., gene KOs) that facilitate such switch. More specifically, the list of reactions that should increase (Rf) or decrease (Rb) to switch from the source state to target state are obtained by mapping differentially expressed genes between both conditions to GSMMs reactions using the GPR. Then, the switch from the source condition (with a reference flux distribution v_i^{ref}) to the target condition (v_i^{MTA}) is simulated by maximizing the number of reactions that switch in the desired direction while minimizing the variation in reactions that are not differentially expressed(Ru). To evaluate the capacity of a gene KO to facilitate the switch, such optimization is performed with the reactions associated with such gene inactivated.

$$\min\left((1-\alpha)\sum_{i\in Ru}\left(v_i^{ref}-v_i^{MTA}\right)^2+\frac{\alpha}{2}\sum_{i\in Rf}y_i+\frac{\alpha}{2}\sum_{i\in Rb}y_i\right)$$

Subject to:

$$s\cdot v_i^{MTA}=0$$

$$lb<v^{MTA}<ub$$

$$v_i^{MTA}=0\,,\qquad i\in KO^{reactions}$$

$$v_i^{MTA}-y_i^F\cdot\left(v_i^{ref}+\varepsilon\right)-y_i\cdot lb_i\geq 0\,,i\in Rf$$

$$y_i^F+y_i=1,i\in Rf$$

$$v_i^{MTA}-y_i^B\cdot\left(v_i^{ref}-\varepsilon\right)-y_i\cdot ub_i\leq 0\,,i\in Rb$$

$$y_i^B+y_i=1,i\in Rb$$

$$y_i^F, y_i^B, y_i \in \{0,1\}$$

Where,

α is the weight given to the reactions that should increase or decrease and α-1 the weight given to reactions that should remain unchanged.

$y_i$is a binary variable that is 1 when upregulated/downregulated reactions are not changing in the target direction.

ε is the minimum flux change to consider that a reaction has moved in the desired reaction.

Such optimization is repeated for all genes of interest, which are then ranked based on the capacity to facilitate the switch form the target state to the source state[423].

MTA was further refined by Valcárcel *et al.* who increased the robustness of the technique by also considering the worst-case scenario and MOMA to score the potential targets[403]. The worst-case scenario consists of running MTA inverting the list of upregulated/downregulated genes and allows to determine the capacity to move "away" from the target state with a given gene KO. They termed their approach robust MTA (rMTA). By integrating both MOMA and the worst-case scenario in the scoring functions, rMTA can more robustly identify the targets that have a greater propensity to facilitate the transition towards the target state.

Compared to other approaches that integrate gene expression like GIMME or iMAT, the benefit of MTA and rMTA is that they can integrate differential gene expression data between the conditions of study (i.e., the list of genes that are over or underexpressed) and hence can better represent the more subtle variations between conditions.

Nevertheless, a potential limitation of both MTA and rMTA is that there can potentially be more than one optimal solution to the MTA optimization problem. The multiplicity of solutions arises from the $\frac{\alpha}{2}\sum_{i\in Rf} y_i + \frac{\alpha}{2}\sum_{i\in Rb} y_i$ component of the problem as any upregulated/downregulated reactions with a variation above/below ε will contribute equally to the objective function regardless of their flux value. This can potentially affect the ranking of potential gene targets, leading to some genes being over-scored.

1.6.4. ^{13}C resolved metabolomics and ^{13}C MFA

As it has been established in the previous sections, the accuracy of both kinetic modeling and constraint-based modeling is largely dependent on the integrated experimental data. For

instance, metabolite concentrations in the extracellular media taken at multiple time points can be used to constrain the rates of uptake and secretion of metabolites (i.e., extracellular fluxes) of the system under study. However, intracellular fluxes cannot be so readily estimated from metabolite concentrations alone as they are not informative of the fluxes leading to and emerging from them[2,424,425]. This limitation led to the development of stable isotope-resolved metabolomics (SIRM).

In SIRM, a biological system is fed with one or more metabolic substrates labeled with stable heavy isotopes (e.g., ^{13}C, ^{15}N or ^{2}H). These labeled molecules, usually referred to as tracers, are metabolized by the system of interest through different metabolic pathways and propagate at a time, flux and pathway-dependent manner, generating characteristic labeling patterns. Hence, the labeling pattern of a given metabolite can provide information about the fluxes through the pathways leading to such metabolite. In this regard, ^{13}C resolved metabolomics (i.e., using tracers enriched with ^{13}C) are widely used to infer the flux through central carbon metabolism (Figure 10)[355,426].

The propagation of ^{13}C, or any other stable isotopes, from tracers to metabolites is quantified by isotope-sensitive analytical techniques, namely nuclear magnetic resonance (NMR)[427,428] and mass spectrometry (MS). Whereas NMR quantifies the abundance of positional isotopomers, MS quantifies the relative abundance of isotopologues (i.e., isomers with a specific number of ^{13}C substitutions)[429]. ^{13}C isotopologues for a given metabolite or metabolite fragment are usually denoted as Mi where *i* is the number of ^{13}C substitutions.

Additionally, an essential step for ^{13}C resolved metabolomics is to correct for natural heavy isotope enrichment. The need for such step emerges because certain isotopes occur naturally with significant frequency (e.g., 4.3% for ^{34}S and 1.1% for ^{13}C) and their effect might be indistinguishable from the artificially enriched ^{13}C from the tracer[a]. This correction can be performed with dedicated software[430–432].

[a]With NMR or very high resolution MS[599], ^{13}C can be distinguished from other naturally occurring isotopes (e.g., ^{15}N or ^{34}S). However, data should still be corrected for natural ^{13}C enrichment.

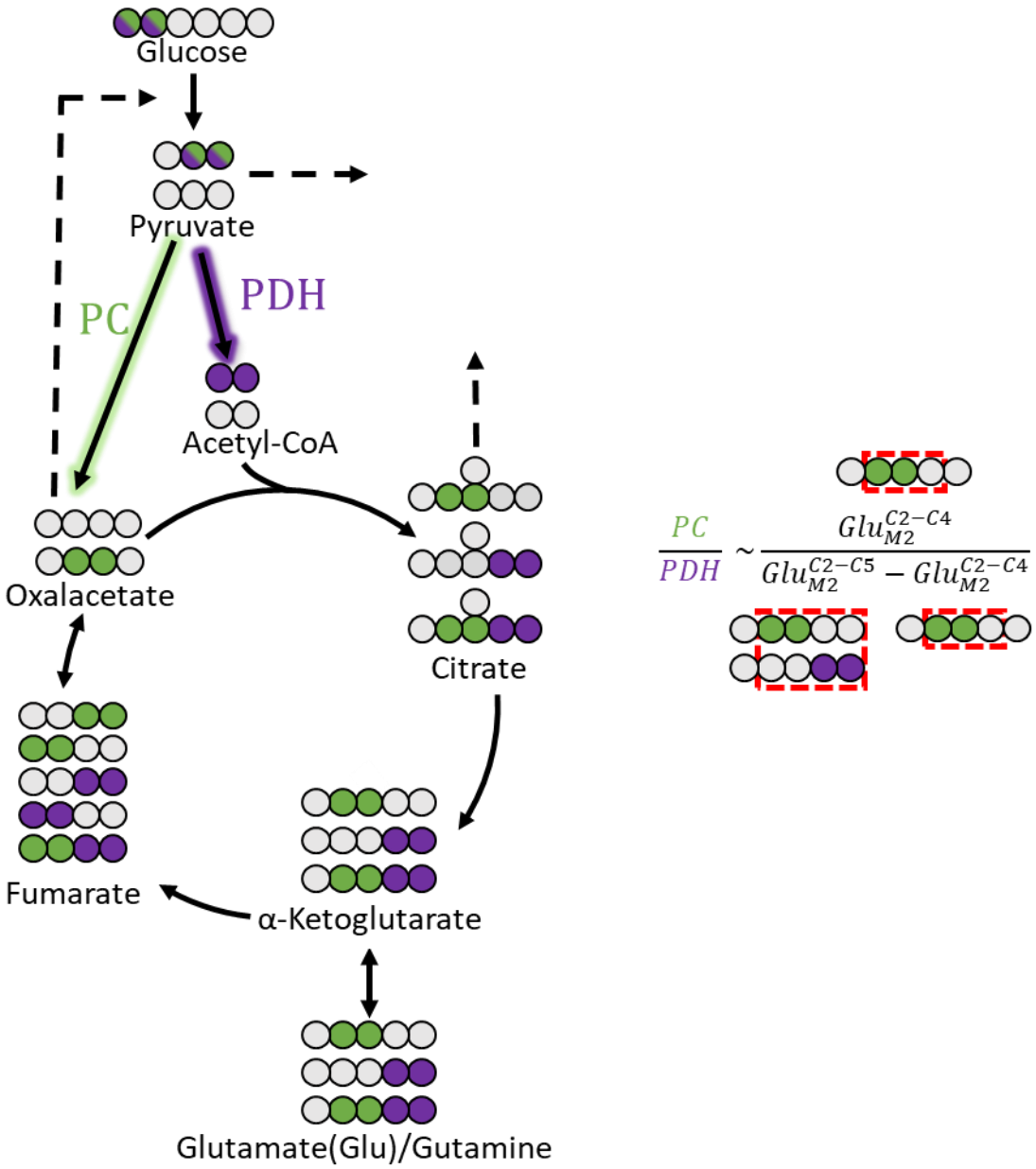


Figure 10. Different pathways differentially propagate ^{13}C. Example of a simple model of ^{13}C propagation from [1,2-^{13}C]-glucose and how the flux through pyruvate carboxylase (PC) or pyruvate dehydrogenase (PDH) results on differentiated label pattern on TCA cycle intermediates. In this regard, measured isotopologues in glutamate fragments can be informative of the ratio between PC and PDH.

Label patterns, corrected for natural enrichment, can be analyzed to trace intracellular metabolic fluxes. Generally, the analysis falls into three categories: substrate contribution, pathway activity analysis and ^{13}C MFA[355,425,426]. While the first two approaches make local flux predictions integrating a small set of label patterns, ^{13}C MFA is a more powerful approach that can integrate large sets of measured label patterns and extracellular fluxes to identify network-wide flux distributions.

1.6.4.1. ^{13}C MFA

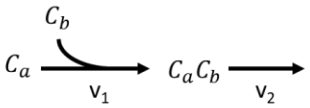
^{13}C MFA integrates ^{13}C resolved-metabolomics data in the framework of Systems Biology to estimate the underlying flux distribution. More precisely, ^{13}C MFA uses a mathematical model to identify the intracellular flux distributions in a given metabolic network most consistent with the measured label patterns and extracellular fluxes[355,426]. At its core, a ^{13}C MFA model comprises a metabolic model coupled to a label propagation model. The metabolic model consists of a set of metabolite balance equations built from the stoichiometry of the metabolic network. The label propagation model consists of a set of positional isotopomer/isotopologue balance equations built from integrating the stoichiometry of the metabolic network with the relevant carbon mappings (i.e., how carbon atoms are rearranged from substrates to products) (Figure 11). The balance equations can be built either around positional isotopomers or around isotopologues using the EMU (elementary metabolite unit) approach. The latter identifies the minimum amount of information needed to simulate the experimentally measured isotopologue fractions[355,424,426].

There are several variants of ^{13}C-based MFA depending on whether metabolic steady state or isotopic steady state (i.e., a state where label patterns are stable in time) are assumed.

In stationary ^{13}C MFA, often merely called ^{13}C MFA, the system is solved assuming both metabolic steady-state and isotopic steady state (i.e., both metabolite concertation and isotopologue fractions are constant in time). Under this approach, isotopologue or positional isotopomers balances are solved as a system of non-linear equations, using a steady-state flux distribution and a set of labeled substrates as input, to obtain steady-state isotopologue/positional isotopomer distributions[355,424,426]. An advantage of this approach is that is it does not need an estimation of the size of metabolite pools (i.e., metabolite concentrations) and can work directly with fractional isotopologue/positional isotopologue abundances (Figure 12).

On the other hand, isotopically nonstationary ^{13}C MFA assumes a metabolic steady-state but not an isotopic steady-state. In this analysis, isotopologue distributions are simulated by solving isotopologue/positional isotopomer balances as a system of ordinary differential equations (ODEs) from the initial time to the experimental endpoint. Such a system depends not only on flux distribution but also on the size of each metabolite pool, which also needs to be estimated through iteration[354,426,433].

Toy metabolic network



Positional isotopomers concentrations

$$[^{12}C_a] \quad [^{13}C_a] \quad [^{12}C_b] \quad [^{13}C_b] \quad [^{13}C_a\,^{13}C_b] \quad [^{12}C_a\,^{12}C_b] \quad [^{12}C_a\,^{13}C_b] \quad [^{13}C_a\,^{12}C_b]$$

Isotopologues concentrations

$$\left[M_0^{C_a}\right] = [^{12}C_a] \qquad \left[M_1^{C_a}\right] = [^{13}C_a] \qquad \left[M_0^{C_b}\right] = [^{12}C_b] \qquad \left[M_1^{C_b}\right] = [^{13}C_b]$$

$$\left[M_0^{C_aC_b}\right] = [^{12}C_a\,^{12}C_b] \qquad \left[M_1^{C_aC_b}\right] = [^{13}C_a\,^{12}C_b] + [^{12}C_a\,^{13}C_b] \qquad \left[M_2^{C_aC_b}\right] = [^{13}C_a\,^{13}C_b]$$

Metabolite concentrations

$$[C_a] = [^{12}C_a] + [^{13}C_a] = \left[M_0^{C_a}\right] + \left[M_1^{C_a}\right]$$

$$[C_b] = [^{12}C_b] + [^{13}C_b] = \left[M_0^{C_b}\right] + \left[M_1^{C_b}\right]$$

$$[C_aC_b] = [^{12}C_a\,^{12}C_b] + [^{13}C_a\,^{12}C_b] + [^{12}C_a\,^{13}C_b] + [^{13}C_a\,^{13}C_b] = \left[M_0^{C_aC_b}\right] + \left[M_1^{C_aC_b}\right] + \left[M_2^{C_aC_b}\right]$$

Positional isotopologue balances for metabolite C_aC_b

$$\frac{d[^{12}C_a\,^{12}C_b]}{dt} = \frac{[^{12}C_a]\cdot[^{12}C_b]}{[C_a]\cdot[C_b]}\cdot v_1 - \frac{[^{12}C_a\,^{12}C_b]}{[C_aC_b]}\cdot v_2$$

$$\frac{d[^{12}C_a\,^{13}C_b]}{dt} = \frac{[^{12}C_a]\cdot[^{13}C_b]}{[C_a]\cdot[C_b]}\cdot v_1 - \frac{[^{12}C_a\,^{13}C_b]}{[C_aC_b]}\cdot v_2$$

$$\frac{d[^{13}C_a\,^{12}C_b]}{dt} = \frac{[^{13}C_a]\cdot[^{12}C_b]}{[C_a]\cdot[C_b]}\cdot v_1 - \frac{[^{13}C_a\,^{12}C_b]}{[C_aC_b]}\cdot v_2$$

$$\frac{d[^{13}C_a\,^{13}C_b]}{dt} = \frac{[^{13}C_a]\cdot[^{13}C_b]}{[C_a]\cdot[C_b]}\cdot v_1 - \frac{[^{13}C_a\,^{13}C_b]}{[C_aC_b]}\cdot v_2$$

Isotopologue balances equation for metabolite C_aC_b

$$\frac{d\left[M_0^{C_aC_b}\right]}{dt} = \frac{\left[M_0^{C_a}\right]\cdot\left[M_0^{C_b}\right]}{[C_a]\cdot[C_b]}\cdot v_1 - \frac{\left[M_0^{C_aC_b}\right]}{[C_aC_b]}\cdot v_2$$

$$\frac{d\left[M_1^{C_aC_b}\right]}{dt} = \frac{\left[M_1^{C_a}\right]\cdot\left[M_0^{C_b}\right]}{[C_a]\cdot[C_b]}\cdot v_1 + \frac{\left[M_0^{C_a}\right]\cdot\left[M_1^{C_b}\right]}{[C_a]\cdot[C_b]}\cdot v_1 - \frac{\left[M_1^{C_aCb}\right]}{[C_aC_b]}\cdot v_2$$

$$\frac{d\left[M_2^{C_aC_b}\right]}{dt} = \frac{\left[M_1^{C_a}\right]\cdot\left[M_1^{C_b}\right]}{[C_a]\cdot[C_b]}\cdot v_1 - \frac{\left[M_2^{C_aCb}\right]}{[C_aC_b]}\cdot v_2$$

Figure 11. Example of ^{13}C balance equations in a toy metabolic network. In this toy metabolic network, two mono-carbon metabolites (C_a and C_b) are condensed into a bi-carbon metabolite (C_a-C_b) through a reaction with a flux v_1. Metabolite C_a-C_b is removed from the system at a rate of v_2. C_a and C_b are assumed to be a constant input. ^{13}C MFA problems are formulated with either isotopologue balances or positional isotopomer balances, here both are shown for comparison. ''[ ]'' is used to denote concentration and [M_x] to indicate the concentration of isotopologues with x number of ^{13}C substitutions.

Stationary ^{13}C MFA

Metabolic steady state (constraint based-modelling)

$$S \cdot v = 0$$

Isotopologue fractions

$$M_0^{C_a} = \frac{[M_0^{C_a}]}{[C_a]} \qquad M_1^{C_b} = \frac{[M_1^{C_b}]}{[C_b]} \qquad M_1^{C_a} = \frac{[M_1^{C_a}]}{[C_a]} \qquad M_0^{C_b} = \frac{[M_0^{C_b}]}{[C_b]}$$

$$M_0^{C_aC_b} = \frac{[M_0^{C_aC_b}]}{[C_aC_b]} \qquad M_1^{C_a-C_b} = \frac{[M_1^{C_aC_b}]}{[C_aC_b]} \qquad M_2^{C_a-C_b} = \frac{[M_2^{C_aC_b}]}{[C_aC_b]}$$

$$M_0^{C_a} + M_1^{C_a} = 1 \qquad M_0^{C_b} + M_0^{C_b} = 1 \qquad M_0^{C_aC_b} + M_1^{C_aC_b} + M_2^{C_aC_b} = 1$$

Isotopologue balances equation for metabolite C_aC_b

$$\frac{d\, M_0^{C_aC_b}}{dt} = M_0^{C_a} \cdot M_1^{C_b} \cdot v_1 - M_0^{C_aC_b} \cdot v_2 = 0$$

$$\frac{d\, M_1^{C_aC_b}}{dt} = M_1^{C_a} \cdot M_0^{C_b} \cdot v_1 + M_0^{C_a} \cdot M_1^{C_b} \cdot v_1 - M_1^{C_aC_b} \cdot v_2 = 0$$

$$\frac{d\, M_2^{C_aC_b}}{dt} = M_1^{C_a} \cdot M_1^{C_b} \cdot v_1 - M_2^{C_aC_b} \cdot v_2 = 0$$

Figure 12. Stationary ^{13}C MFA. An example of stationary ^{13}C MFA for the toy model of Figure 11. For simplicity, the problem is formulated only with isotopologue fractions. Working in isotopic steady state allows working with isotopologue fractions (M_x) rather than isotopologue concentrations.

This approach allows working with systems that have yet to reach isotopic steady-state and to estimate fluxes with greater precision than stationary ^{13}C MFA. This can be relevant because isotopic steady-state can require a significant amount of time to be reached for metabolites with large pools or with small fluxes from the tracer. However, to take advantage of it, label patterns must be measured at multiple time points before isotopic steady-state is reached, which results in an increased experimental workload[354,426,433]. Because both stationary ^{13}C MFA and isotopically nonstationary ^{13}C MFA work under metabolic steady-state, constraint-based modeling is used to compute steady-state flux distributions in both approaches.

Conversely, dynamic ^{13}C MFA assumes neither isotopic nor metabolic steady-state and uses a kinetic model, that simulates the evolution of metabolite concentrations and metabolic fluxes in time, coupled to a system of ODEs for isotopologue/positional isotopomer balances (Figure 13)[355,434,435]. Compared to the other two ^{13}C MFA approaches, dynamic ^{13}C MFA is able to

simulate systems outside of the metabolic steady state. A limitation of dynamic ^{13}C MFA is that it is far more computationally complex to solve than nonstationary ^{13}C MFA. A second limitation to dynamic ^{13}C MFA lays on the complexity of building and parameterizing an adequate set of kinetic laws as well as on the limited offer of software capable of dynamic MFA analyses[354,436].

Dynamic ^{13}C MFA

Kinetic model

$$v_1 = \frac{\frac{V}{K_{iC_a} \cdot Km_{C_b}} \cdot [C_a] \cdot [C_b]}{1 + \frac{[C_a]}{K_{iC_a}} + \frac{[C_b]}{K_{iC_b}} + \frac{[C_a] \cdot [C_b]}{K_{iC_a} \cdot Km_{C_b}}} \qquad v_2 = \frac{V \cdot [C_a C_b]}{[C_a C_b] + Km_{C_a C_b}}$$

$$\frac{d[C_a]}{dt} = -v_1 \qquad \frac{d[C_b]}{dt} = -v_1 \qquad \frac{d[C_a C_b]}{dt} = v_1 - v_2$$

ODE equations for isotopologues of metabolite C_aC_b

$$\frac{d\left[M_0^{C_a C_b}\right]}{dt} = \frac{\left[M_0^{C_a}\right] \cdot \left[M_0^{C_b}\right]}{[C_a] \cdot [C_b]} \cdot v_1 - \frac{\left[M_0^{C_a C_b}\right]}{[C_a C_b]} \cdot v_2$$

$$\frac{d\, M_1^{C_a C_b}}{dt} = \frac{\left[M_1^{C_a}\right] \cdot \left[M_0^{C_b}\right]}{[C_a] \cdot [C_b]} \cdot v_1 + \frac{\left[M_0^{C_a}\right] \cdot \left[M_1^{C_b}\right]}{[C_a] \cdot [C_b]} \cdot v_1 - \frac{\left[M_1^{C_a Cb}\right]}{[C_a C_b]} \cdot v_2$$

$$\frac{d\, M_2^{C_a C_b}}{dt} = \frac{\left[M_1^{C_a}\right] \cdot \left[M_1^{C_b}\right]}{[C_a] \cdot [C_b]} \cdot v_1 - \frac{\left[M_2^{C_a Cb}\right]}{[C_a C_b]} \cdot v_2$$

Figure 13. Dynamic ^{13}C MFA. An example of dynamic ^{13}C MFA for the toy model of Figure 11. As an example, reactions 1 and 2 are assumed to have random order ternary-complex and a Michaelis Menten kinetic laws, respectively. The kinetic model and the ODE for isotopologue balances are solved simultaneously as a single system of ODEs.

Regardless of the type of analysis, the label propagation model takes a set of fluxes generated by the metabolic model and simulates the expected label patterns associated with such flux distribution for the tracer used in the experiment; this is referred as the MFA forward problem. However, in ^{13}C MFA the goal is to estimate the flux distributions from a set of measured label patterns; this is referred to as the MFA inverse problem. As analytical solutions to this inverse problem are only possible for straightforward systems, the solution of an inverse ^{13}C MFA problem is usually determined by an iterative fitting procedure. In this procedure, fluxes and other model parameters are iteratively varied seeking the set that minimizes the difference between simulated and experimental isotopologue abundances[355,424,426]. More specifically, a ^{13}C MFA optimization can be formulated as follows:

$$X_{opt} = min \sum_j \left(\frac{E_j - Y_j(p)}{\sigma_j} \right)^2$$

where,

p is the vector of free fluxes or kinetic model parameters that are iteratively fit;

X_{opt} is the optimal value of the ^{13}C MFA objective;

E_j is the experimentally quantified fraction for isotopologue *j*;

$Y_j(v)$ is the simulated isotopologue fraction for isotopologue *j* with *p*.

However, there is usually not a unique solution to an MFA problem. Instead, there is a space of valid flux solutions and parameter values that can be consistent with the measured isotopologue abundances. This space of solutions can be evaluated by computing the confidence intervals for fitted parameters or flux values[437]. The width of the confidence intervals for any given flux will depend on the complexity of the metabolic network, the measured extracellular fluxes, the metabolites and time points where label patterns have been measured and the tracer(s) used in the experiment.

Indeed, in large metabolic networks and when a small set of ^{13}C data is integrated, the confidence intervals can be too wide to draw meaningful conclusions about the underlying flux distributions[438–440]. This can be overcome by simplifying the metabolic network, which can potentially create an arbitrary bias, or integrating data from parallel experiments, each using a tracer tailored to a specific part of the network[441]. However, and although the latter approach is quite powerful, there is need of an approach to select the most likely solutions in of undetermined parts of the network without requiring additional measurements as large metabolic networks might need an unattainable large number of parallel experiments to be fully determined. To address this issue, as part of this Ph.D., we developed parsimonious ^{13}C MFA an approach to select the best solutions in an undetermined ^{13}C MFA solution space[a442].

[a] See section 4.2

2. Objectives

2. Objectives

The general objective of this Ph.D. thesis is to expand the Systems Biology toolbox by developing new model-based approaches for metabolic flux analysis and to apply them to the analysis of the molecular drivers underlying hepatic fructose intolerance and cancer progression.

In order two accomplish the above-mentioned general objective, the following specific objectives were defined:

1. To develop a kinetic model of hepatocyte metabolism in the framework of dynamic ^{13}C MFA and to apply it to characterize the short-term effects of high fructose concentrations in hepatic metabolism (Chapter 1).
2. To develop new a ^{13}C MFA software with an algorithm capable of integrating both ^{13}C and gene expression data to identify the underlying flux distribution (Chapter 2).
3. To characterize the molecular and metabolic drivers of the CSC phenotype in prostate cancer and to develop a new approach to identify putative metabolic targets to revert it (Chapter 3).
4. To develop a workflow for multiomics data integration in the framework of GSMMS and to apply it to identify metabolic vulnerabilities underlying the metastatic phenotype in colorectal cancer (Chapter 4).

3. Report of the supervisors

3. Report of the supervisors

The work developed by Carles Foguet Coll in this Ph.D. thesis has resulted in eight publications, and two manuscripts that are currently submitted to be published in international scientific journals. Carles Foguet is a principal author of four of the six articles that integrate this thesis.

All the results here published have not been presented in any other thesis. However, as described underneath, Josep Tarragó has performed the experimental characterization and target validation of the colorectal cancer model and such results will be included in his Ph.D. thesis.

Following is the list of all the articles included in the thesis, indicating the impact factor of the journal in which the article was published or has been submitted and specifying the tasks that Carles Foguet has performed in each one of them:

Chapter 1:

Carles Foguet, Silvia Marin, Vitaly A Selivanov, Eric Fanchon, Wai-Nang Paul Lee, Joan J Guinovart, Pedro de Atauri, Marta Cascante. **HepatoDyn: A Dynamic Model of Hepatocyte Metabolism That Integrates ^{13}C Isotopomer Data**. *PLoS computational biology* 2016, 12(4):e1004899

Metrics from Clarivate Analytics (2016): Article; Impact Factor 4,542; Q1 (11 of 78) in BIOCHEMICAL RESEARCH METHODS; **D1** (4 of 57) in MATHEMATICAL & COMPUTATIONAL BIOLOGY

URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004899

In this work, Carles Foguet developed a kinetic model of hepatic metabolism, including detailed kinetic laws and the ability to simulate ^{13}C propagation. With this model, he simulated the metabolism of hepatocytes in the presence of high fructose concentrations. Carles Foguet is the principal author of this paper, analyzed the data, coded the model and wrote the manuscript.

Chapter 2:

Kristian Peters, James Bradbury, Sven Bergmann, Marco Capuccini, Marta Cascante, Pedro de Atauri, Timothy MD Ebbels, **Carles Foguet**, Robert Glen, Alejandra Gonzalez-Beltran, Ulrich L Günther, Evangelos Handakas, Thomas Hankemeier, Kenneth Haug, Stephanie Herman, Petr Holub, Massimiliano Izzo, Daniel Jacob, David Johnson, Fabien Jourdan, Namrata Kale, Ibrahim Karaman, Bita Khalili, Payam Emami Khonsari, Kim Kultima, Samuel Lampa, Anders Larsson, Christian Ludwig, Pablo Moreno, Steffen Neumann, Jon Ander Novella, Claire O'Donovan, Jake TM Pearce, Alina Peluso, Marco Enrico Piras, Luca Pireddu, Michelle AC Reed, Philippe Rocca-Serra, Pierrick Roger, Antonio Rosato, Rico Rueedi, Christoph Ruttkies, Noureddin Sadawi, Reza M Salek, Susanna-Assunta Sansone, Vitaly Selivanov, Ola Spjuth, Daniel Schober, Etienne A Thévenot, Mattia Tomasoni, Merlijn van Rijswijk, Michael van Vliet, Mark R Viant, Ralf JM Weber, Gianluigi Zanetti, Christoph Steinbeck. **PhenoMeNal: Processing and analysis of Metabolomics data in the Cloud**. *GigaScience* 2018, 8(2):giy149

Metrics from Clarivate Analytics (2019): Article; Impact Factor: 4,688; **Q1** (13 of 69) in MULTIDISCIPLINARY SCIENCES

URL: https://academic.oup.com/gigascience/article/8/2/giy149/5232984

Carles Foguet contributed to this work by developing the Iso2Flux software, implementing it into PhenoMeNal and participating in the development of the

start-to-end fluxomics workflow. He also contributed to the writing of the manuscript.

Payam Emami Khoonsari, Pablo Moreno, Sven Bergmann, Joachim Burman, Marco Capuccini, Matteo Carone, Marta Cascante, Pedro De Atauri, **Carles Foguet**, Alejandra Gonzalez-Beltran, Thomas Hankemeier, Kenneth Haug, Sijin He, Stephanie Herman, David Johnson, Namrata Kale, Anders Larsson, Steffen Neumann, Kristian Peters, Luca Pireddu, Philippe Rocca-Serra, Pierrick Roger, Rico Rueedi, Christoph Ruttkies, Noureddin Sadawi, Reza M Salek, Susanna-Assunta Sansone, Daniel Schober, Vitaly Selivanov, Etienne A Thévenot, Michael Van Vliet, Gianluigi Zanetti, Christoph Steinbeck, Kim Kultima, Ola Spjuth. **Interoperable and scalable data analysis with microservices: Applications in Metabolomics**. *Bioinformatics* 2019, 1-9
Metrics from Clarivate Analytics (2018): Journal Article; Impact Factor 4,531; Q1 (8 of 79) in BIOCHEMICAL RESEARCH METHODS; Q1 (25 of 162) in BIOTECHNOLOGY & APPLIED MICROBIOLOGY; **D1** (4 of 59) in MATHEMATICAL & COMPUTATIONAL BIOLOGY
URL: https://ora.ox.ac.uk/objects/uuid:9ab846f0-5110-421f-90d1-baf235d0ede3

Carles Foguet contributed to this work by developing the Iso2Flux software, implementing it into PhenoMeNal and participating in the development of the start-to-end fluxomics workflow, which is referred to as demonstrator 4 in the paper. He also contributed to the writing of the manuscript.

Carles Foguet, Anusha Jayaraman, Silvia Marin, Vitaly A Selivanov, Pablo Moreno, Ramon Messeguer, Pedro de Atauri, Marta Cascante. **p^{13}CMFA: Parsimonious ^{13}C metabolic flux analysis**. *PLoS computational biology* 2019, 15(9):e1007310
Metrics from Clarivate Analytics (2018): Article; Impact Factor 4,428; Q1 (9 of 79) in BIOCHEMICAL RESEARCH METHODS; **D1** (5 of 59) in MATHEMATICAL & COMPUTATIONAL BIOLOGY
URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007310

In this paper, Carles Foguet presents p^{13}CMFA, a new ^{13}C MFA algorithm that uses flux minimization to address the problem of solution multiplicity when estimating flux distributions based on ^{13}C propagation. Carles Foguet is the principal author of this paper. Carles, conceived p^{13}CMFA, implement it, performed the simulations and validations and wrote the manuscript.

Chapter 3:

Carles Foguet, Cristina Balcells, Timothy M. Thomson, Pedro de Atauri, Marta Cascante. **Identification of the molecular drivers of prostate cancer stem cells.** *Cells, submitted*
Metrics from Clarivate Analytics (2018): Journal Article; Impact Factor 5,656; Q1 (40 of 193) in CELL BIOLOGY

In this paper, Carles Foguet analyzed RNA-SEQ data to identify the molecular drivers of prostate cancer stem cells. To study the metabolic reprogramming underlying cancer stem cells, he developed a new algorithm termed r^{2}MTA. Carles Foguet is the principal author of this paper and processed and analyzed the RNA-SEQ data, developed and conceived r^{2}MTA and wrote the manuscript.

Chapter 4:

Josep Tarragó-Celada, **Carles Foguet**, Jordi Perarnau, Xavier Hernández-Alias, Silvia Marin, Míriam Tarrado-Castellarnau, Pedro de Atauri, Marta Cascante. **Cysteine and folate metabolism are major vulnerabilities of the metastatic subpopulations of colorectal cancer**. *Cell Metabolism, submitted*
CF and JT contributed equally to this manuscript.
Metrics from Clarivate Analytics (2018): Journal Article; Impact Factor 22,415; D1 (5 of 193) in CELL BIOLOGY; D1 (3 of 145) in ENDOCRINOLOGY & METABOLISM

This is a multidisciplinary work combining both a thorough experimental characterization of colorectal cancer cell lines with genome-scale computational modeling. Carles Foguet and Josep Tarragó are coauthors of this manuscript and contributed equally to this work. Carles Foguet performed the genome-scale simulations, the statistical analysis of targeted metabolomics and conceived and implemented the multiomics integration workflow. Josep Tarragó performed the metabolic and phenotypic characterization of the cell lines under study, the target validation and the associated statistical analysis. Both Carles Foguet and Josep Tarragó contributed equally to the writing of the manuscript.

Additionally, the first and last page of four additional publications are included in the appendix. These are four reviews in which three of them Carles is a principal author.

Igor Marín De Mas, Esther Aguilar, Anusha Jayaraman, Ibrahim H Polat, Alfonso Martín-Bernabé, Rohit Bharat, **Carles Foguet**, Enric Milà, Balázs Papp, Josep J Centelles, Marta Cascante. **Cancer cell metabolism as new targets for novel designed therapies**. *Future medicinal chemistry* 2014, 6(16):1791-1810
Metrics from Clarivate Analytics (2014): Review; Impact Factor (2014): 3,744; Q1 (8 of 59) in CHEMISTRY, MEDICINAL
URL: https://www.future-science.com/doi/abs/10.4155/fmc.14.119

Carles Foguet participated in the writing of the manuscript.

Alfonso Martín-Bernabé, Cristina Balcells, Josep Tarragó-Celada, **Carles Foguet**, Sandrine Bourgoin-Voillard, Michel Seve, Marta Cascante. **The importance of post-translational modifications in systems biology approaches to identify therapeutic targets in cancer metabolism**. *Current Opinion in Systems Biology* 2017, 3:161-169
CF and AM-B, CB and JT-C contributed equally to this review
URL: https://www.sciencedirect.com/science/article/pii/S2452310017300082

Carles Foguet is a principal author of this review together with AM-B, CB and JT-C. Carles was fully involved with all aspects of the review, including coordination with the other authors and the writing of the manuscript.

Cristina Balcells, **Carles Foguet**, Josep Tarragó-Celada, Pedro de Atauri, Silvia Marin, Marta Cascante. **Tracing metabolic fluxes using mass spectrometry: stable isotope-resolved metabolomics in health and disease**. *TrAC Trends in Analytical Chemistry* 2019, available online, *in press corrected proof*
CF and CB and JT-C contributed equally to this review
Metrics from Clarivate Analytics (2018): Review; Impact Factor 8,428; **D1** (3 of 84) in CHEMISTRY, ANALYTICAL

URL: https://www.sciencedirect.com/science/article/abs/pii/S0165993618302139

Carles Foguet is a principal author of this review together with CB and JT-C. Carles was fully involved all aspects of the review, including coordination with the other authors and the writing of the manuscript.

Effrosyni Karakitsou, **Carles Foguet**, Pedro de Atauri, Kim Kultima, Payam Emami Khoonsari, Vitor AP Martins dos Santos, Edoardo Saccenti, Antonio Rosato, Marta Cascante. **Metabolomics in Systems Medicine: an overview of methods and applications**. *Current Opinion in Systems Biology* 2019, 15: 91-99

CF and EK contributed equally to this review
URL: https://www.sciencedirect.com/science/article/pii/S2452310018301227

Carles Foguet is a principal author of this review together with EK. Carles was fully involved with all aspects of the review, including coordination with the other authors and the writing of the manuscript.

Marta Cascante
(Supervisor)

Pedro de Atauri
(Supervisor)

4. Summary of the results

4. Summary of the results

As outlined in the objectives, this work covers both the development of new tools and workflows and its application to the study of hepatic and cancer biology. It is divided into several chapters, each centered around a set of publications.

Chapter 1 focuses on the development of HepatoDyn a highly detailed kinetic model of hepatic metabolism for dynamic ^{13}C MFA analyses and its application to the study of fructose metabolism in hepatocytes.

Chapter 2 focuses on the development of Iso2Flux and parsimonious ^{13}C MFA ($p^{13}CMFA$). Iso2Flux is an open-source software for stationary ^{13}C MFA that we developed as part of this Ph.D. and that we implemented into PhenoMeNal (Phenome and Metabolome aNalysis), an e-infrastructure for metabolomics analysis. $p^{13}CMFA$ is a novel algorithm that allows integrating both ^{13}C measurements and transcriptomics to select the best solution in an undetermined solution space.

Next, Chapter 4 focuses on the characterization of the metabolic and signaling drivers of the CSC phenotype in prostate cancer and their association with EMT. On this subject, in order to characterize in the genome-scale the metabolic phenotype of CSC and identify therapeutic interventions to target it, we developed an enhanced version of rMTA[403a] that we termed r^2MTA.

Lastly, Chapter 5 focuses on the genome-scale characterization of the metabolic phenotype of metastatic and non-metastatic colorectal colon cancer cell lines and the identification of metabolic vulnerabilities underlying the metastatic phenotype. With this aim, a novel workflow for the integration of large multiomics data sets into GSMMs was developed.

A summary of each of the chapters, each centered around a set of publications, is provided in the following sections.

[a] See section 1.6.3.4.2

4.1. HepatoDyn: A Dynamic Model of Hepatocyte Metabolism that Integrates ^{13}C Isotopomer Data

Dynamic ^{13}C MFA combines kinetic modeling, with its unique capabilities to integrate the molecular regulation of enzyme-catalyzed reactions and dynamically simulate the metabolism, with the power to trace central carbon metabolism fluxes using ^{13}C MFA[a]. However, because of the complexity of building and parameterizing condition-specific kinetic models, dynamic ^{13}C MFA use has not been as widespread as other ^{13}C MFA variants. Here we aim to address this by developing HepatoDyn, a large kinetic model of hepatocyte central carbon metabolism capable of performing dynamic ^{13}C MFA. Furthermore, as a proof of concept, we applied HepatoDyn to the study of the effects of high fructose concentrations on hepatocyte metabolism.

4.1.1. HepatoDyn, a highly detailed model of hepatocyte metabolism

HepatoDyn was the first model of hepatocyte core metabolism capable of dynamically simulating energy and redox metabolism. It consists of 88 reactions and 81 metabolites distributed into three compartments (extracellular, cytosolic and mitochondrial). Additionally, channeling of hexose phosphates to glycogen is also included in the form of two separate pools of hexose phosphates, as previously described[366] (Figure 14).

HepatoDyn kinetic laws were based on existing laws described in the literature[5,63,368], and modified, when necessary, to incorporate the effects of activators and inhibitors relevant to the hepatic metabolic regulation. The kinetic laws for the reactions catalyzed by ALDOB were the exception and were build using the King and Altman method[5]. This was necessary to adequately simulate the complexity of ALDOB, which can catalyze the reversible cleavage of both fructose 1,6-bisphosphate and fructose 1-phosphate and several half-reactions.

In total, the model has 470 parameters associated with kinetic laws. Of such parameters, 55 parameters representing enzymes activities (Vmax or equivalent) were fitted to the experimental data while the remaining parameters were assigned based on an extensive literature search assisted by the Brenda database[363].

[a] See section 1.6.4.1

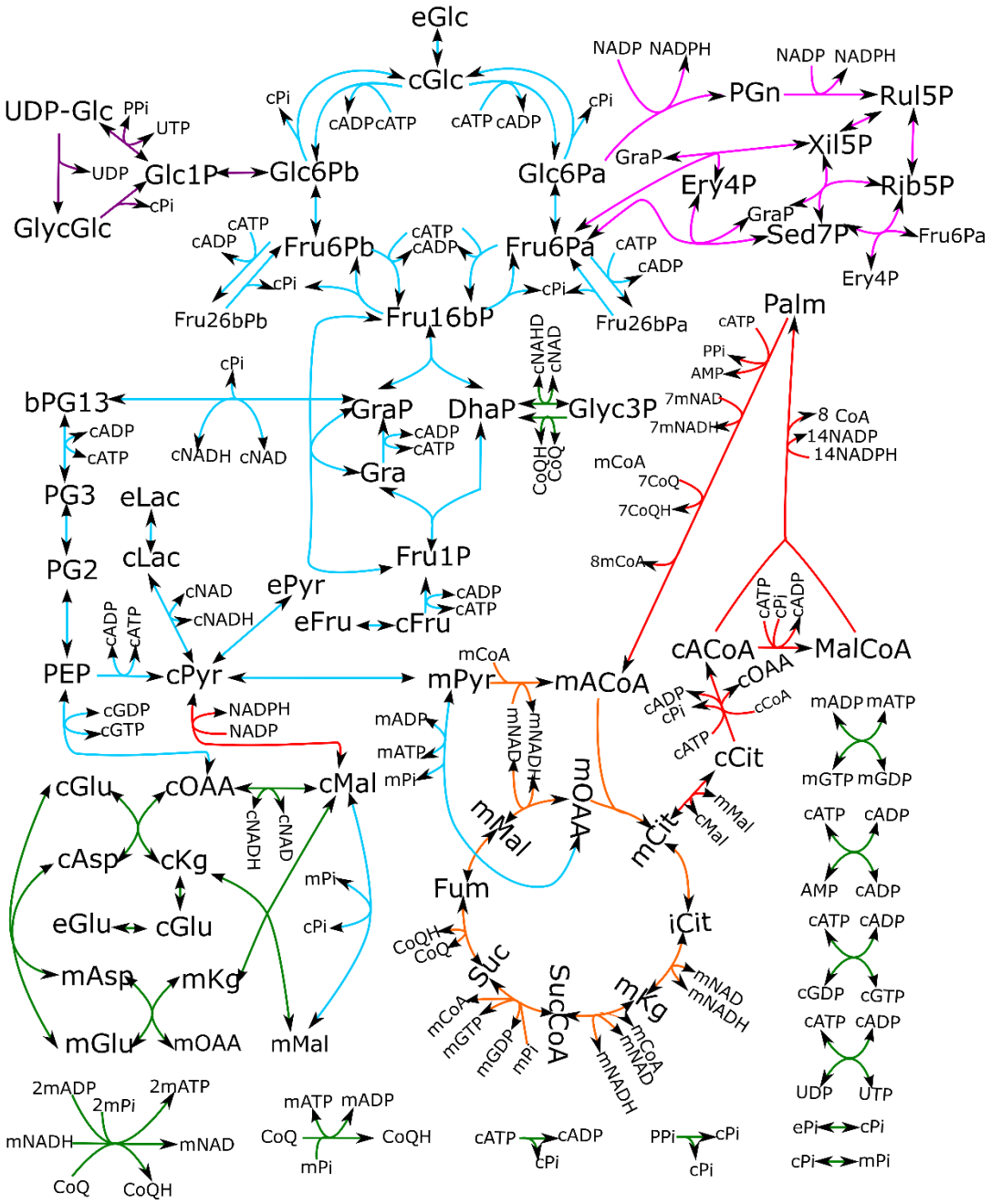


Figure 14. Graphical representation of the metabolic network used in HepatoDyn. In this representation, reactions associated with the glycolytic and gluconeogenic pathways are colored in blue, reactions associated with glycogen metabolism are colored in purple, reactions associated with the PPP are colored in pink, reactions associated with the TCA cycle are colored in orange, reactions associated with fatty acid metabolism are colored in red and other reactions associated with the redox and energy metabolism are colored in green.

4.1.2. Large fructose concentrations induce ATP depletion in hepatocytes

HepatoDyn was applied to the study the short-term response of hepatocytes to incubation with 20mM glucose supplemented by either 3mM fructose or 20mM fructose. Remarkably, while glycogen content increased at a fast rate with 3mM fructose, there was almost no glycogen accumulation with 20mM fructose (Figure 15.A). Furthermore, isotopologue analysis indicated

that with 20mM fructose, unlike the first condition, almost no ^{13}C from labeled glucose was propagated to lactate (Figure 15.B).

HepatoDyn predicted that this phenotype was caused by an ATP and phosphate depletion induced by an accumulation of fructose 1-phosphate at high fructose concentrations (Figure 16). This behavior had been previously reported[443,444] but had never been extensively modeled. This phenomenon occurs because, unlike glucose, the rate of fructose uptake is mostly unregulated and controlled chiefly by the extracellular concentration of fructose due to the low affinity of the fructose transporter[120,444]. Furthermore, unlike GK, fructokinase is not part of any negative feedback loop. Hence, high fructose concentrations can potentially lead to an excessive and possibly toxic accumulation of fructose 1-phosphate. Thus, the low glycogen synthesis and the almost non-existent propagation of ^{13}C from glucose to lactate observed with 20 mM fructose can be attributed to the depletion of ATP and phosphate which leads to reduced glycogen synthesis and GK and PFKL activities.

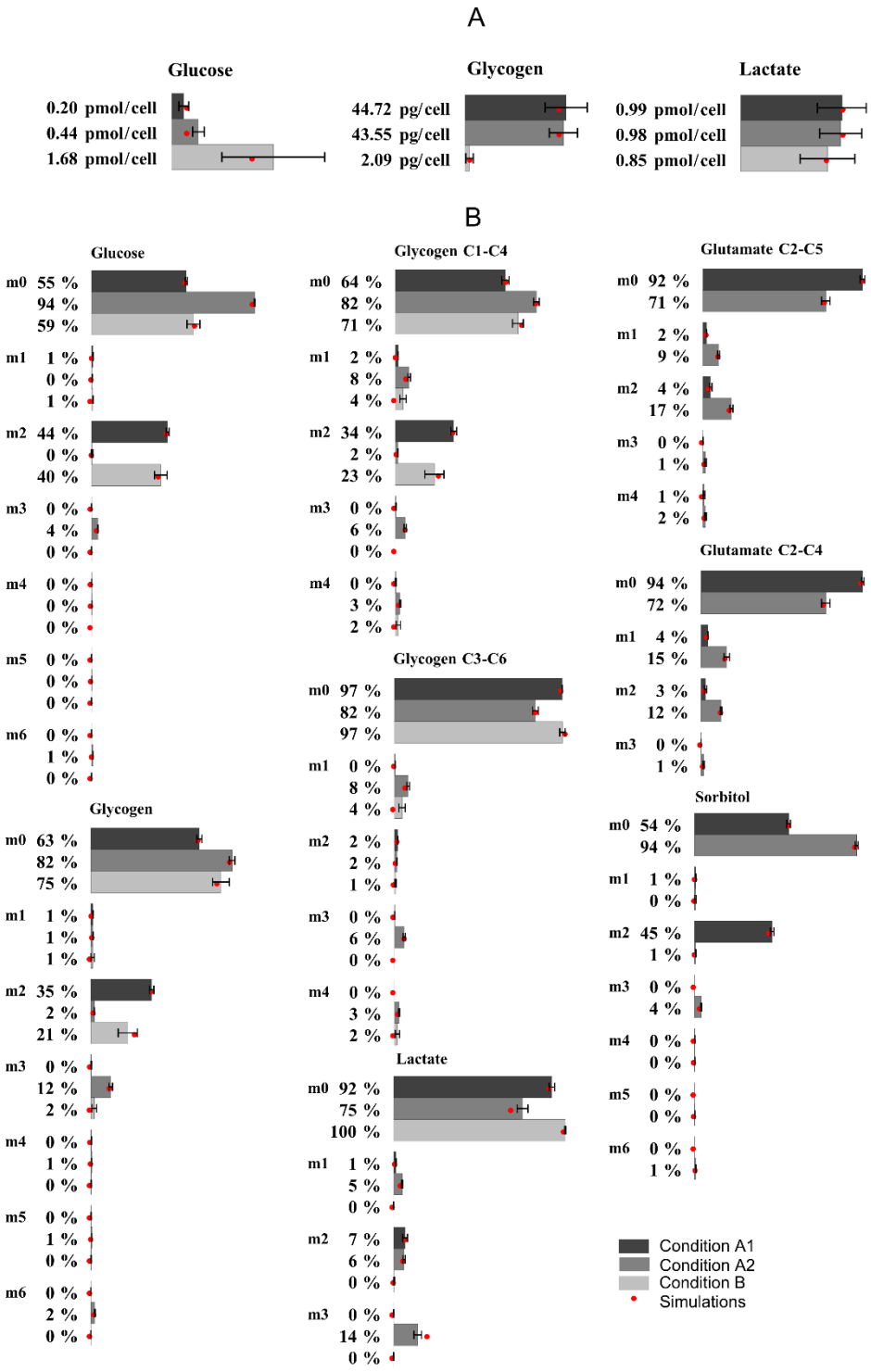


*Figure 15. Bar graphs representing the experimentally determined metabolite productions **(A)** and isotopologue fractions **(B)**. Measurements were taken after incubating hepatocytes for 2 hours with 20mM glucose 50% enriched in [1,2-$^{13}C_2$]-glucose and 3mM fructose (condition A1), 20 mM glucose and 3 mM fructose 50% enriched in [U-$^{13}C_6$]-fructose (condition A2) and 20mM glucose 50% enriched in [1,2-$^{13}C_2$]-glucose and 20mM fructose (condition B). The red dot indicates the value fractions simulated by HepatoDyn using the best fit parameter set.*

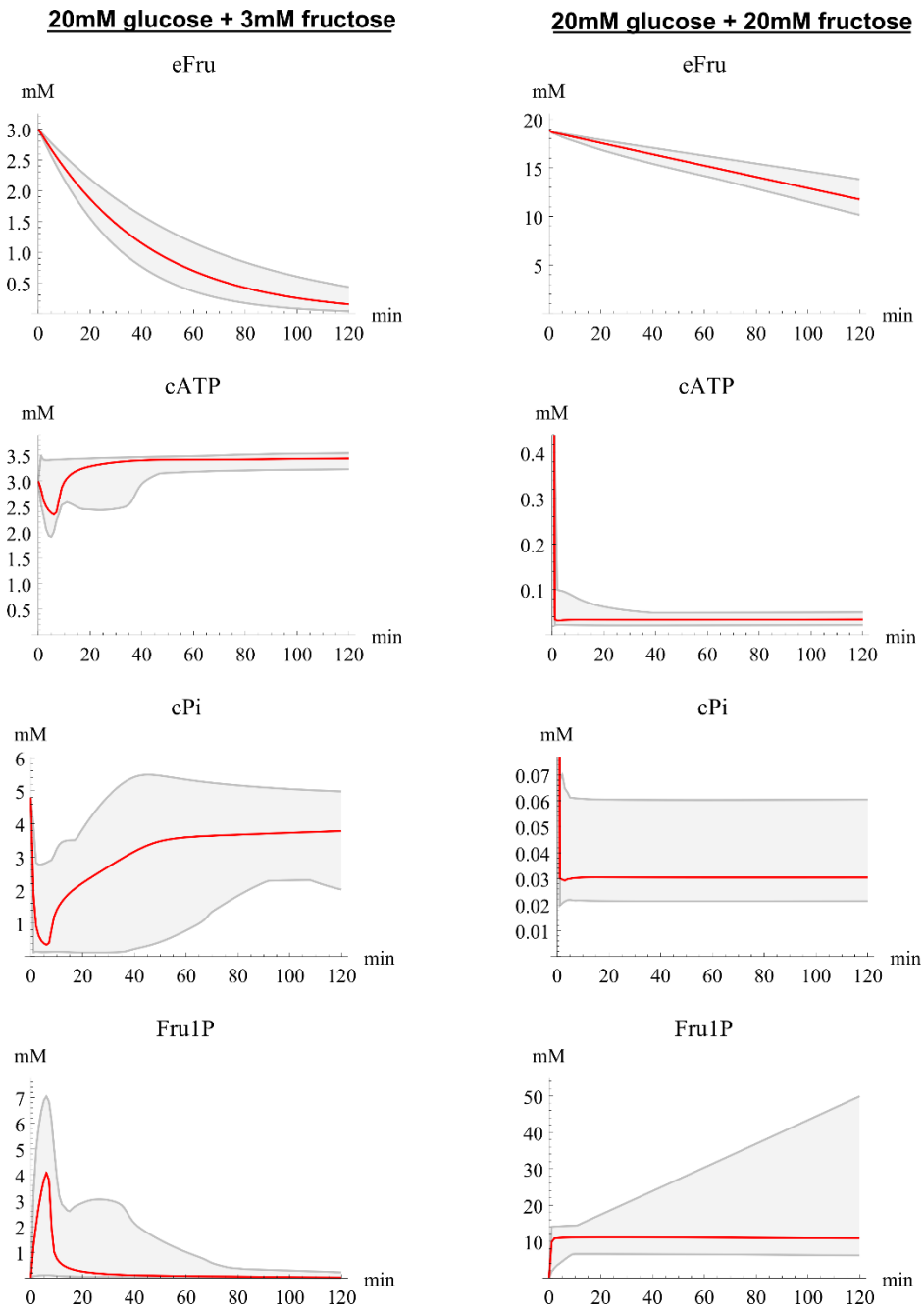


Figure 16. Plot of the simulated concentrations over time for extracellular fructose (eFru), fructose 1-phosphate (Fru1P), cytosolic phosphate (cPi) and cytosolic ATP (cATP). Specifically, the simulated concentrations in hepatocytes incubated with 20 mM glucose and 3 mM fructose (conditions A1/ A2) or 20 mM glucose and 20 mM fructose (condition B) are shown. The red plot indicates the values predicted with the best fit parameter set, and the grey area indicates the estimated range of variations taking parameters sets within the 95% confidence intervals.

4.2. Iso2Flux and p^{13}CMFA

As described in the introduction, ^{13}C MFA identifies the flux distribution that leads to a set of experimentally measured ^{13}C label patterns. Hence, an inherent limitation of ^{13}C MFA is that fluxes cannot be uniquely determined when there are several possible flux combinations that can lead to a set of experimentally measured ^{13}C label patterns[a]. Indeed, when ^{13}C MFA is used in large metabolic networks or with small data sets this can lead to a large part of the network being unidentifiable by ^{13}C MFA[438,439,445]. Furthermore, many of the ^{13}C MFA solutions can involve large fluxes through futile cycles, which are artifacts of the optimization process as *in vivo* enzyme activities cannot support such large flux values.

In this chapter, we present p^{13}CMFA a new ^{13}C MFA approach that uses flux minimization to identify the most enzymatically efficient solution in an undetermined ^{13}C MFA solution space (Figure 17). Flux minimization has been widely used in the context of constraint-based modeling at the genome-scale with approaches like pFBA[397] or GIMME[400,405b] and here we export the concept to ^{13}C MFA with great success.

Under p^{13}CMFA, flux through cycles and pathways are minimized to the minimum amount needed to account for the experimental ^{13}C measurements. Furthermore, this approach can integrate gene expression data to give greater weight to the minimization of fluxes through enzymes with low gene expression allowing, for the first time, seamless integration of ^{13}C MFA with gene expression data. The latter allows selecting those pathways with stronger gene expression evidence in instances where multiple pathways can result in similar label patterns.

Having developed the concept of p^{13}CMFA, we needed to implement it into a ^{13}C MFA software to test its feasibility. However, most available software for ^{13}C MFA was not open-source nor could not be easily modified to implement p^{13}CMFA. Hence, in order to prove the feasibility of p^{13}CMFA, we developed Iso2Flux, a new open-source ^{13}C MFA software that would be compatible with p^{13}CMFA. Furthermore, Iso2Flux was implemented into the framework of the PhenoMeNal e-infrastructure[446,447] allowing ^{13}C MFA to be run from the cloud and as part of multi-step data analysis workflows.

[a] See section 1.6.4.1
[b] See sections 1.6.3.3.3 and 1.6.3.3.5.2

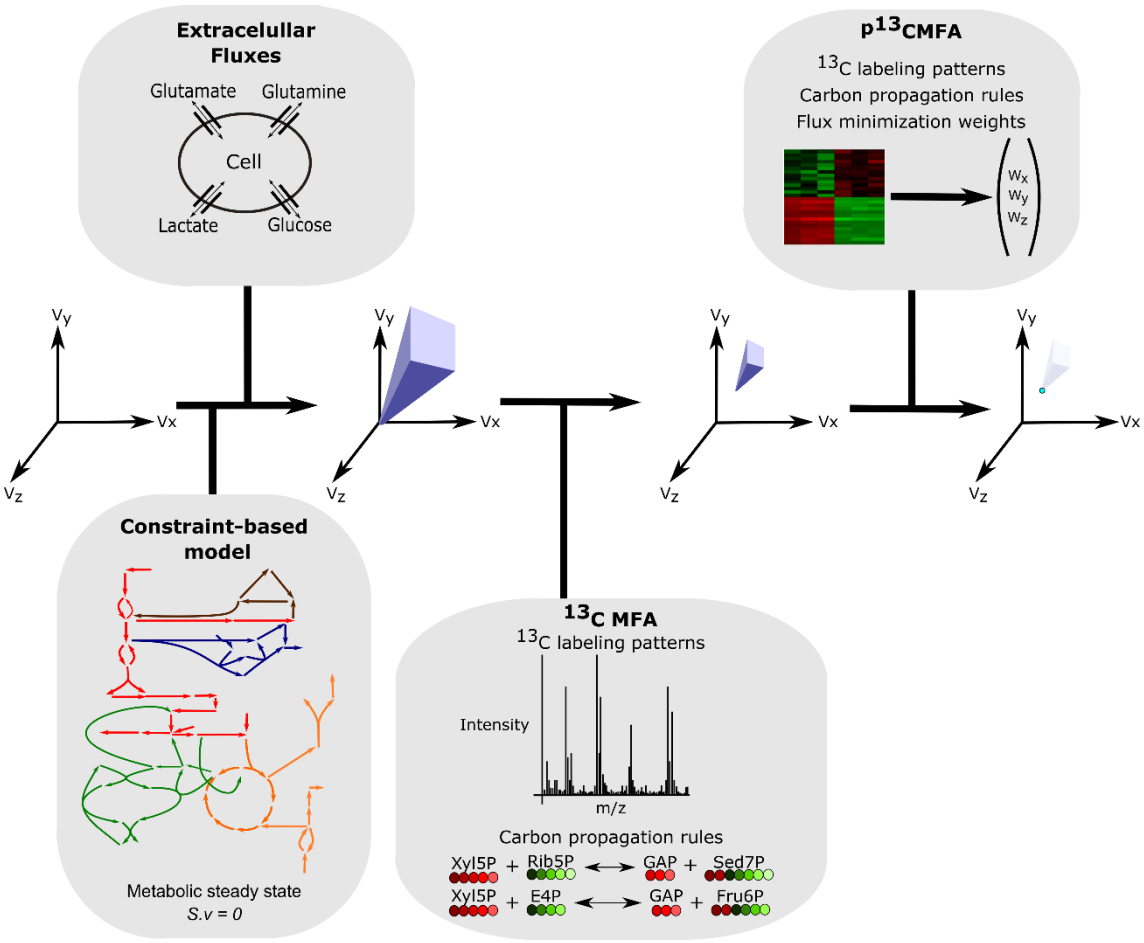


Figure 17. The conceptual basis of p[13]CMFA: From an infinite space of flux (v) solutions, a space of feasible solutions is obtained through the integration of stoichiometric and thermodynamic constraints (in the form of a constraint-based model) and the measured extracellular fluxes. Applying [13]C MFA to integrate experimental [13]C data can further reduce the solution space to those flux distributions that are consistent with such data. Through flux minimization, p[13]CMFA can identify the optimal flux distribution that lies on the edge of the [13]C MFA solution space. Such minimization can be weighted according to the gene expression evidence for each enzyme.

4.2.1. Example of p[13]CMFA usage

As an example of a potential application of p[13]CMFA, we applied it to analyze metabolic flux distribution in human umbilical vein endothelial cells. The analyzed data set was quite small and [13]C MFA was unable to significantly constraint the flux ranges emerging from the stoichiometric and thermodynamic constraints and the measured extracellular fluxes (Figure 18). For instance, despite integrating measurements of [13]C enrichment in ribose, it was not possible to conclude the relative contribution of the oxidative and non-oxidative branch of PPP to ribose synthesis.

Nevertheless, p[13]CMFA, together with transcriptomics data from the literature[448], can be applied to obtain the most enzymatically efficient flux distribution in the conditions of the study. In this regard, we can now estimate that in PPP, the non-oxidative branch contributes to roughly 60% of the net ribose synthesis.

To evaluate the contribution of ^{13}C MFA to the p^{13}CMFA solution, GIMME (i.e., flux minimization weighted by gene expression without integrating ^{13}C data) was also performed (Figure 18). The resulting flux distribution had some notable differences such as that GIMME did not predict any activity in the oxidative branch of the PPP and predicted a higher flux through PDH than p^{13}CMFA. This confirms that p^{13}CMFA benefits from ^{13}C data to predict significantly different flux maps than those derived from flux minimization alone.

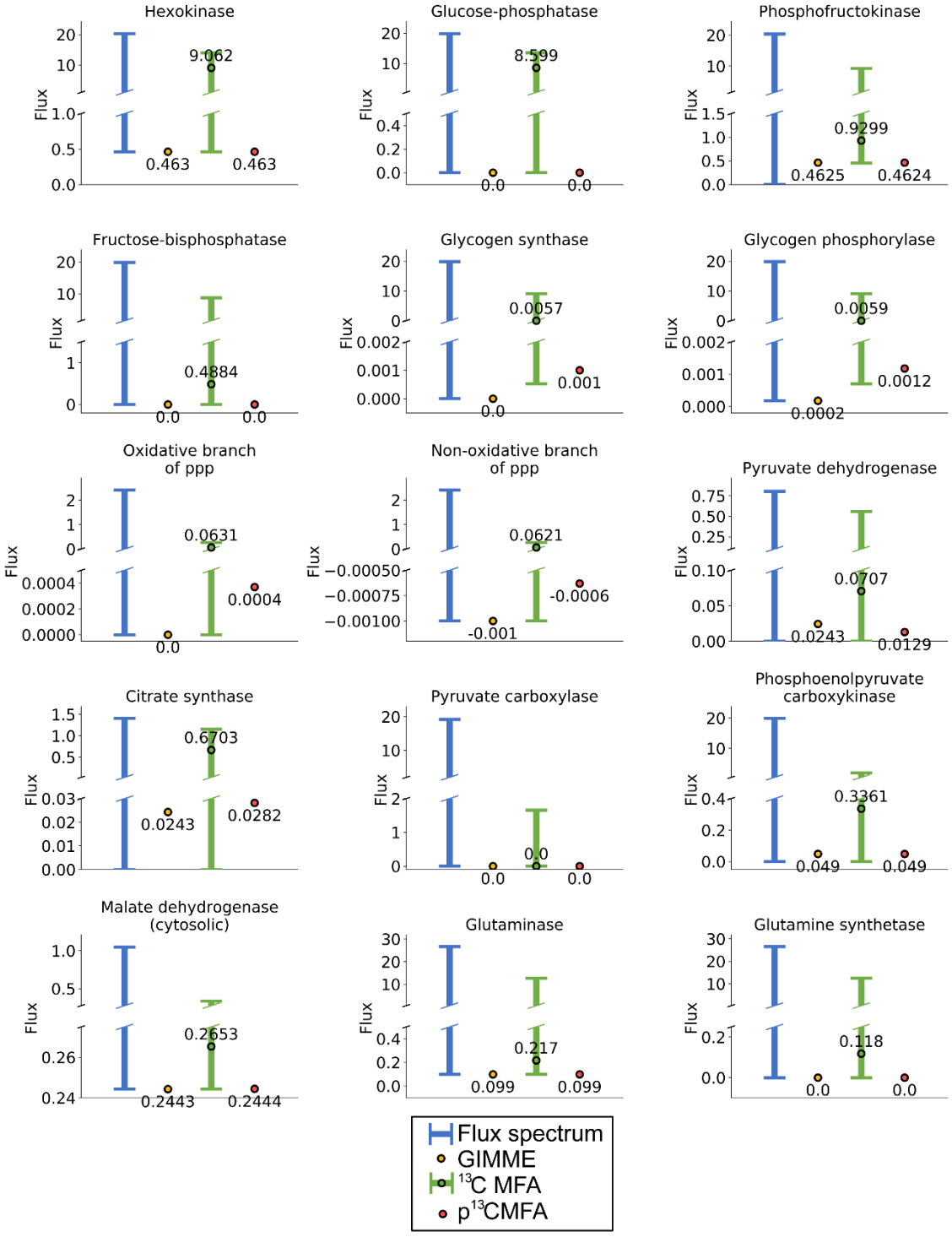


Figure 18 (previous page). Flux spectrum, GIMME solutions, ^{13}C MFA flux ranges, and p^{13}CMFA solutions for some key reaction fluxes. Flux spectrum represents the feasible flux ranges considering only the stoichiometric and thermodynamic constraints and the measured extracellular fluxes. GIMME flux values are obtained when total reaction flux is minimized weighted by gene expression without integrating ^{13}C data. For ^{13}C MFA, the flux values obtained after the ^{13}C MFA optimization and the range of the 95% confidence intervals for such values are shown. The p^{13}CMFA flux values are obtained when total reaction flux is minimized within the ^{13}C MFA solution space. Fluxes are expressed in μmol·h^{-1}·million-cells^{-1}.

4.2.2. p^{13}CMFA outperforms pFBA, GIMME and ^{13}C MFA

To validate the p^{13}CMFA method, we used as a reference a study of the colon cancer cell line HCT 116 published by Tarrado-Castellarnau *et al.*[224] whereby integrating a large set of experimental measurements ^{13}C MFA had been able to estimate the central carbon metabolism flux map with a high degree of accuracy.

From this large data set, we selected a partial data set consisting of 7 experimental flux measurements and 4 sets of isotopologue fractions. The partial data set was used to run pFBA, GIMME, ^{13}C MFA, and p^{13}CMFA with and without integrating gene expression data (p^{13}CMFA+ge and p^{13}CMFA-ge, respectively). The accuracy of the resulting flux distributions was evaluated by comparing them to the reference flux distribution computed from the whole set of experimental measurements[224]. The results showed that p^{13}CMFA-ge resulted in significantly more accurate predictions than both pFBA and ^{13}C MFA alone. Similarly, p^{13}CMFA+ge was able to achieve substantially more accurate flux predictions than ^{13}C MFA and GIMME (Figure 19).

4.2.3. Iso2Flux

p^{13}CMFA was implemented in Iso2Flux, an open-source stationary ^{13}C MFA software we developed as part of this Ph.D. thesis (https://github.com/cfoguet/iso2flux/releases/tag/0.7.2).

Iso2Flux uses the EMU framework[424] to build systems of isotopologue balance equations under the assumption of isotopic steady state[a]. Such a system, which is compiled into C to enhance performance, is solved to predict the isotopologue fractions associated with a given steady-state flux distribution. Next, Iso2Flux uses the self-adaptive differential evolution (SADE) algorithm from PyGMO (Python Parallel Global Multiobjective Optimizer) to iteratively find the optimal solution of the ^{13}C MFA and p^{13}CMFA problems.

The formats of the inputs and outputs of Iso2Flux have been designed with the ease of user accessibility in mind. For instance, the metabolic network can be defined either as a XLSX/CSV file or in the Systems Biology Markup Language (SBML) format, which is a standard format for

[a] See figure 12 in section 1.6.4.1

biological models[449]. Similarly, in the file defining the carbon propagation rules (i.e., how ^{13}C propagate between the substrates and products of reactions in the network) only the reactions with a change in carbon backbone need to be defined as the remaining reactions will be automatically added. Furthermore, Iso2flux can also detect inconsistencies in the carbon propagation rules and metabolic network and highlight them to the user.

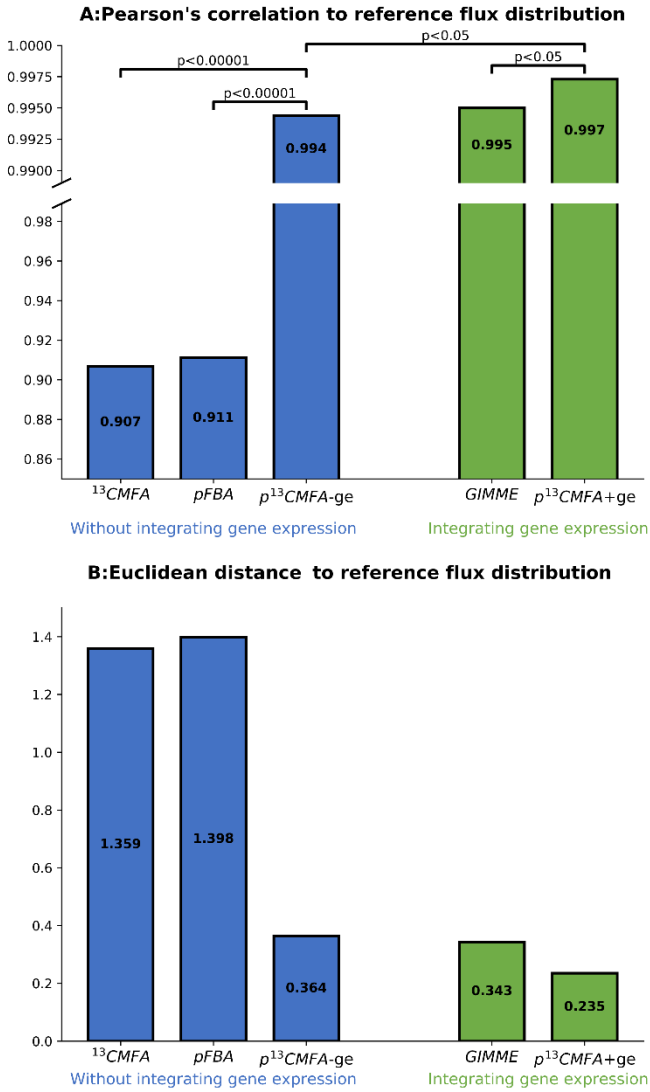


Figure 19. Comparison of the predictive power of ^{13}C MFA, pFBA, GIMME, and p^{13}CMFA. (A) Pearson's correlation coefficients between the reference flux distribution and the flux maps obtained from ^{13}C MFA (optimal solution), pFBA, GIMME, and p^{13}CMFA. p^{13}CMFA was applied both with and without integrating gene expression data (p^{13}CMFA+ge and p^{13}CMFA-ge, respectively). The statistical significance of the difference between correlation coefficients was evaluated using the Fisher r-to-z transformation[450]. (B) Euclidian distances between the reference flux distribution and the flux maps obtained from ^{13}C MFA (optimal solution), pFBA, GIMME, and p^{13}CMFA.

4.2.4. Iso2Flux in the framework of PhenoMeNal

PhenoMeNal (https://portal.phenomenal-h2020.eu/home) is a newly developed cloud e-infrastructure that provides a large suite of standardized and interoperable metabolomics data processing tools[446,447]. The PhenoMeNal cloud research environment can be run in commercial public cloud providers (e.g., Amazon web services, Google cloud platform and Microsoft Azure) or private clouds and uses a container orchestrator to deploy containerized computational tools. A container for given computational tool contains all the libraries and configuration files to execute the tool autonomously and is akin to a lightweight virtual machine optimized for running the tool in question[451]. Once the PhenoMeNal cloud research environment is deployed, it can be accessed through Galaxy which provides a web interface where data can be uploaded and analyzed using the tools in the PhenoMeNal repository[446]. Additionally, Galaxy supports workflows, automated and multi-step analyses where several tools are run sequentially[452].

Iso2Flux was chosen as the stationary ^{13}C MFA software for the Phenomenal e-infrastructure. As such, a standalone Docker container for Iso2flux was developed (https://github.com/phnmnl/container-iso2flux) and it was implemented to the PhenoMeNal runtime for Galaxy (https://github.com/phnmnl/container-galaxy-k8s-runtime). Additionally, the inputs of Iso2Flux were made interoperable with other tools available in PhenoMeNal. In this regard, Iso2flux was implemented as part of the start-to-end fluxomics workflow (Figure 20). This workflow consists of the following containerized tools executed in succession in the framework of Galaxy:

- RAMID; it identifies metabolites of interest and quantifies isotopologue fractions starting from raw chromatograms-mass spectra files.
- MIDCOR; it corrects the measured isotopologue fractions for naturally occurring isotopes[432].
- Iso2Flux; it performs stationary ^{13}C MFA.
- Escher for Fluxomics; it allows to visualize the resulting flux distribution predicted by ^{13}C MFA on top of a pre-defined metabolic map[453].

This workflow was used as an example of PhenoMeNal's potential to connect separately developed tools covering subsequent steps of metabolomics analysis. In this regard, this workflow was demonstrated in several PhenoMeNal-themed workshops such as CloudMET 2017 or "PhenoMeNal Gateway Workshop: Taking metabolomics data processing to the cloud".

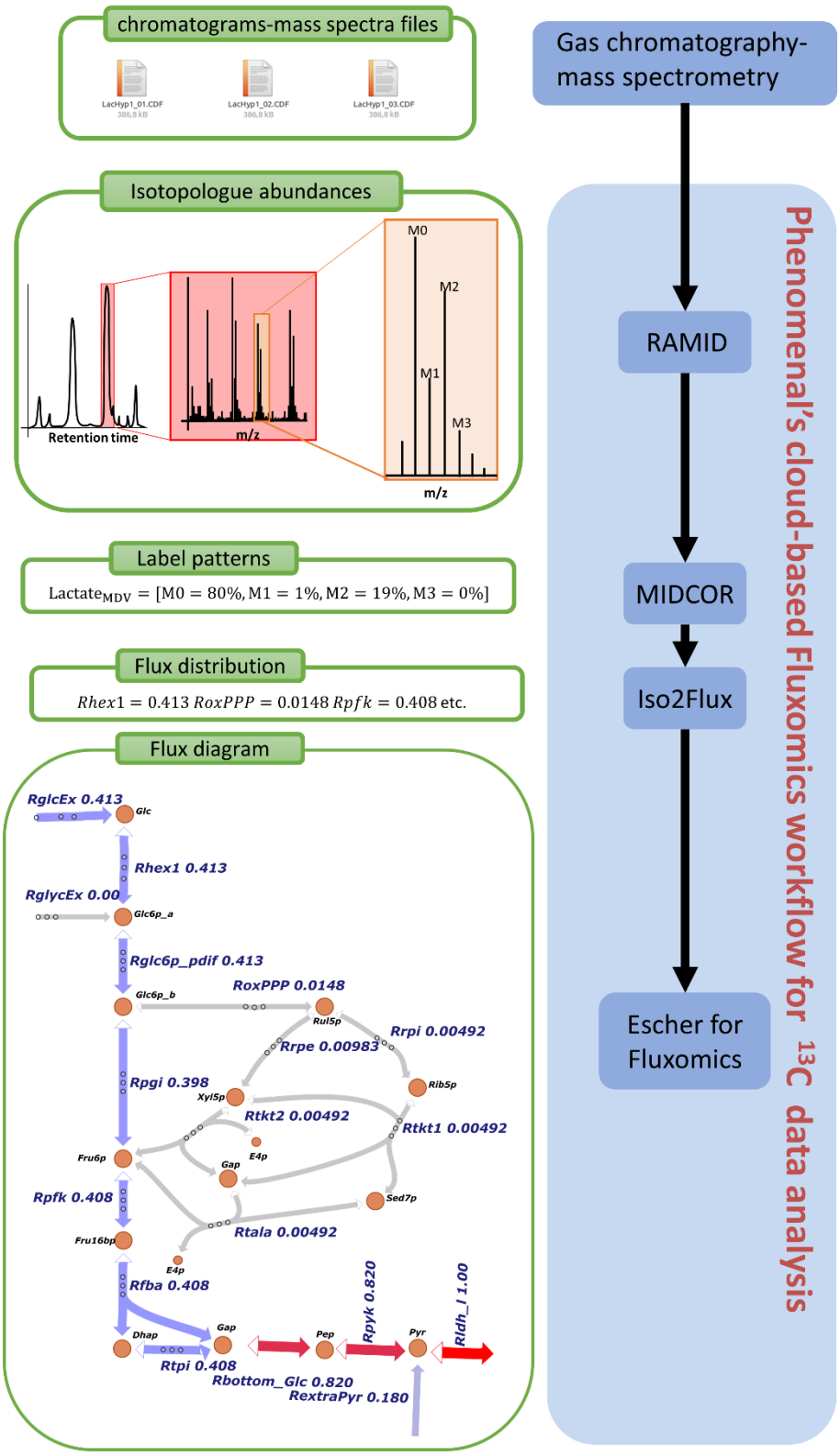


Figure 20. PhenoMeNal's start-to-end fluxomics workflow. This workflow begins with raw chromatograms-mass spectra files, corrects natural enrichment, performs ^{13}C stationary MFA using Iso2Flux and, plots the predicted flux distribution.

4.3. Identification of the molecular drivers of prostate cancer stem cells

It is widely accepted that in prostate cancer, and indeed most cancer types, both metastatic potential, and the acquisition of resistance to chemotherapy and radiotherapy are associated with the CSC[a] population[158,159]. It was originally believed that the acquisition of the CSC phenotype required a complete EMT[454–456]. However, it is now beginning to emerge that the CSC phenotype can be acquired independently of EMT[157,457]. Even more, in some cellular models, EMT has been reported to inhibit CSC traits such as metastatic, tumorigenic and proliferative potential[157,186–191].

In this regard, the dual cell model PC3M/PC3S, derived from the PC3 metastatic prostate cancer cell line, has emerged as a model to characterize the EMT independent CSC phenotype in prostate cancer[157]. PC3M have an epithelial-like and a CSC-like phenotype with a high proliferation rate, and high capacity to form spheroids *in vitro* and establish tumors and metastasis *in vivo.* Conversely, PC3S are mesenchymal-like and have significantly reduced proliferative and tumorigenic potential but they are more invasive than PC3M[157]. Interestingly, the CSC phenotype appears to be strongly associated with the epithelial-like phenotype as overexpressing EMT factors, such as murine Snai1, in PC3M was shown to reduce both its proliferation rate, anchorage-independent growth, and tumorigenic potential[157].

In this chapter, we use RNA sequencing to characterize the PC3M, PC3S, and a PC3M derived cell line overexpressing the EMT factor Snai1 (PC3M-Snai1) at a genome-scale and characterize the relationship between EMT, metabolic reprogramming and CSC phenotype in prostate cancer. In this regard, to analyze metabolic reprogramming, we have developed an enhanced version of the rMTA algorithm[403b] that we have termed r^{2}MTA.

4.3.1. Overexpressing Snai1 partially reverts the PC3M gene expression program

Gene set enrichment analysis (GSEA)[458] was performed to identify sets of genes coordinately upregulated or downregulated between PC3M and PC3S and upon Snai1 overexpression in PC3M. In this regard, compared to PC3S, PC3M displayed a strong upregulation in the gene expression signatures of MYC and mTORC1, both of which have been extensively associated with the CSC phenotype[176,177,459–461]. Consistent with well-known effects of both MYC and mTORC1 overactivation[84,462,463], PC3M also displayed significantly upregulated gene sets associated to gene expression machinery. Remarkably, increasing the dosage of SNAI1 through

[a] CSC is the abbreviation of cancer stem cells which have been covered in section 1.5.1.2.1
[b] See section 1.6.3.4.2

overexpression of murine Snai1 significantly downregulated most of such gene signatures confirming that CSC phenotype is strongly inhibited by EMT in the PC3M/PC3S model[157,333].

4.3.2. The CSC phenotype of PC3M is associated with partial EMT

Consistent with their more epithelial phenotype, PC3M strongly overexpressed several epithelial markers such as E-cadherin (CDH1), Claudin1 (CLDN1) and the epithelial cell adhesion molecule (EPCAM) while strongly underexpressing mesenchymal markers such as Fibronectin 1 (FN1) and Secreted Protein Acidic And Cysteine Rich (SPARC) compared to PC3S. However, PC3M also displayed a strong expression of many mesenchymal markers, such as Vimentin (VIM) or N-cadherin (CDH2), suggesting that PC3M might have a partial EMT phenotype. Supporting this notion, PC3M also had a strong expression of several EMT transcription factors such as SNAI1, SNAI2, and TWIST1(Figure 21.A).

Increasing the dosage of SNAI1 through Snai1 overexpression (PC3M-Snai1) resulted in an increased EMT gene expression program as seen by a strong downregulation of the transcription of epithelial markers (i.e., CDH1, CLDN1, EPCAM) and a strong upregulation of mesenchymal markers (i.e., FN1, SPARC) consistent with the previous work[157]. This suggested that the partial EMT underlying the PC3M phenotype is mediated by a delicate balance between pro-EMT factors (such as SNAI1, SNAI2 and LATS2) and anti-EMT factors and that increasing the dosage of the former tilts the scale in favor of EMT. As putative anti-EMT factors playing a key role in such delicate balance, we identified KLF4[207,208,464,465] and PRKD1[211–213,466].

4.3.3. PC3M overexpress mesenchyme-associated genes associated with metastatic spread

Closely associated with partial EMT, PC3M overexpressed several genes related to the mesenchymal phenotype (i.e., part of the gene set ‘‘Hallmark EMT”[467]) compared to PC3S (Figure 21.B). Many of such genes have been associated with poor prognosis and invasive and metastatic potential in a variety of cancer types[468,469,478–487,470,488–497,471–477]. Indeed, several of such genes are involved in the remodeling of the extracellular matrix, adhesion, and migration and they might confer PC3M the sufficient invasion, extravasation and intravasation capabilities necessary to be able to metastasize autonomously.

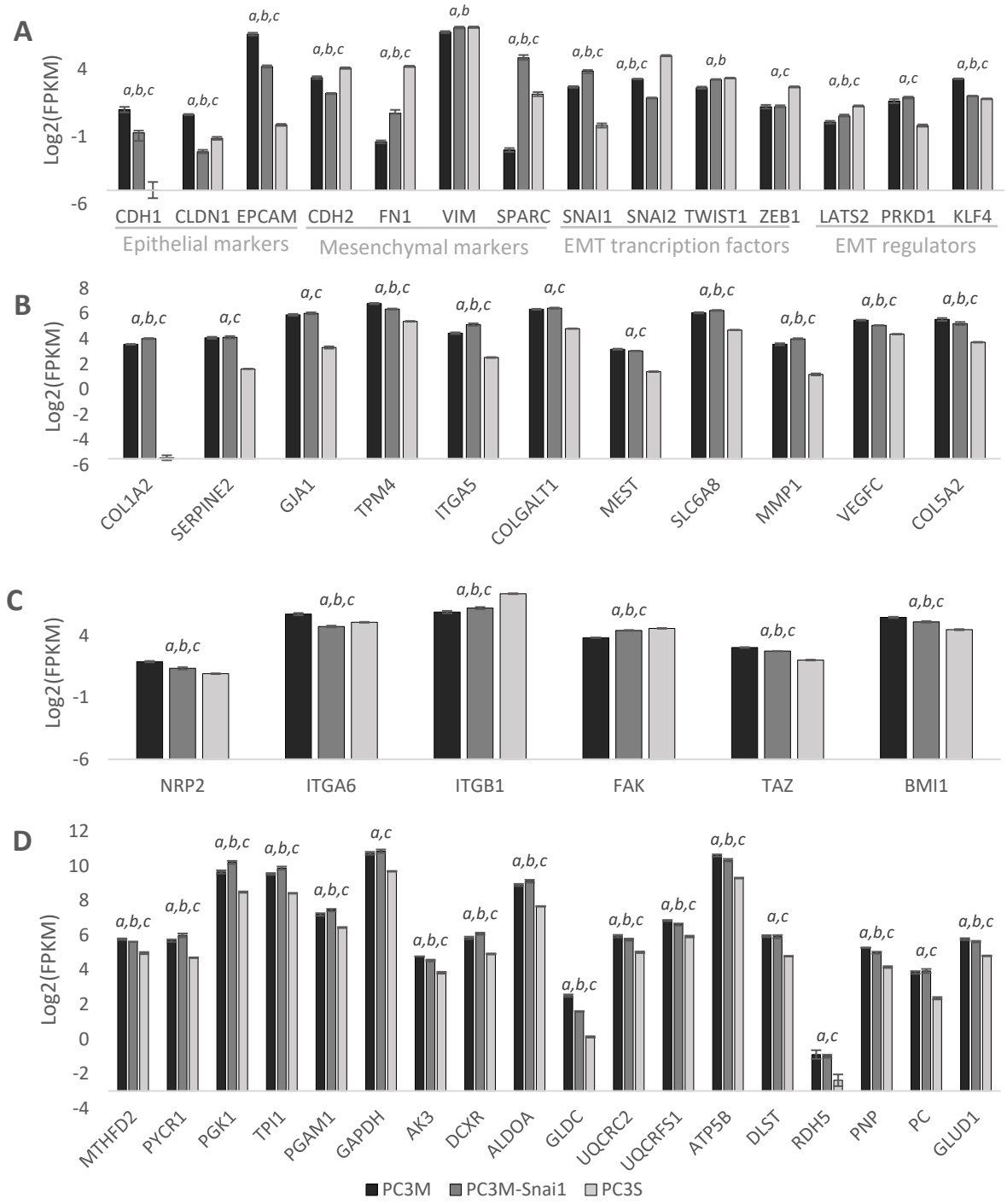


Figure 21. Transcript levels. ***(A)*** *Expression of epithelial and mesenchymal biomarkers, EMT transcription factors and EMT regulators.* ***(B)*** *Top EMT associated genes overexpressed in PC3M.* ***(C)*** *Expression of genes associated to VEGFC/NRP2 signaling.* ***(D)*** *Metabolic genes identified as putative targets against CSC metabolism with r^2MTA. Expression levels are indicated as FPKM Fragments (Per Kilobase of transcript per Million mapped reads) in a Log2 scale. a,b and c denote a statistically significant difference (FDR<0.05) between PC3M and PC3S, PC3M and PC3M-Snai1 and PC3M-Snai1 and PC3S, respectively.*

Additionally, the vascular endothelial growth factor C (VEGFC), part of the set of EMT associated genes overexpressed in PC3M, is of particular interest as VEGFC signaling through the Neuropilin 2 (NRP2) receptor has been associated to with the CSC phenotype in breast[498–500] and prostate cancer[501,502]. More in detail, upon VEGFC binding, NRP2 has been reported to interact laterally with integrins (i.e., ITGA5, ITGA6 or ITGB1) and activate focal adhesion kinase (FAK) signaling[499,503,504] leading to the activation of the transcription factors TAZ[500] and BMI1[499,501]. Remarkably, VEGFC, NRP2, TAZ and BMI1 gene expression patterns appear to be correlated with stemness (PC3M>PC3M-Snai1>PC3S) (Figure 21.C). Therefore, we theorize that the VEGFC might play a key role in maintaining the PC3M CSC phenotype.

4.3.4. Identifying key players of the CSC phenotype

Given that Snai1 overexpression significantly inhibits key aspects of the CSC phenotype of PC3M, we posited that key players of such phenotype could be identified in the subset of genes strongly downregulated both in PC3S compared to PC3M and in PC3M-SNAI1 compared to PC3M. To narrow the list of candidates, copy number variation (CNA) and survival analysis were used to select those genes that were more frequently amplified in prostate cancer and whose amplification or mutation correlated with poor clinical prognosis. Overall, 9 potential candidates were identified (Figure 22). Most of such genes have been extensively associated with features of CSC such as metastatic and invasive potential, cell proliferation, apoptosis or drug resistance[505,506,515–517,507–514]. We hypothesized that expression of such genes is linked to partial-EMT status and might account for the loss of the CSC phenotype upon Snai1 overexpression.

4.3.5. r^{2}MTA

As outlined in the introduction[a], rMTA[403] is an algorithm that identifies putative metabolic targets to switch towards the desired metabolic state. With this aim, it simulates metabolic transitions between two states by integrating transcriptomics in the framework of GSMMs to identify metabolic pathways coordinately upregulated or downregulated.

Here, we developed r^{2}MTA an enhanced version of the rMTA algorithm[403], which improves its robustness by making the following improvements:

- Adding flux minimization with GIMME[400b] as a pre-processing step to reduce flux loops

[a] See section 1.6.3.4.2
[b] See section 1.6.3.3.5.2

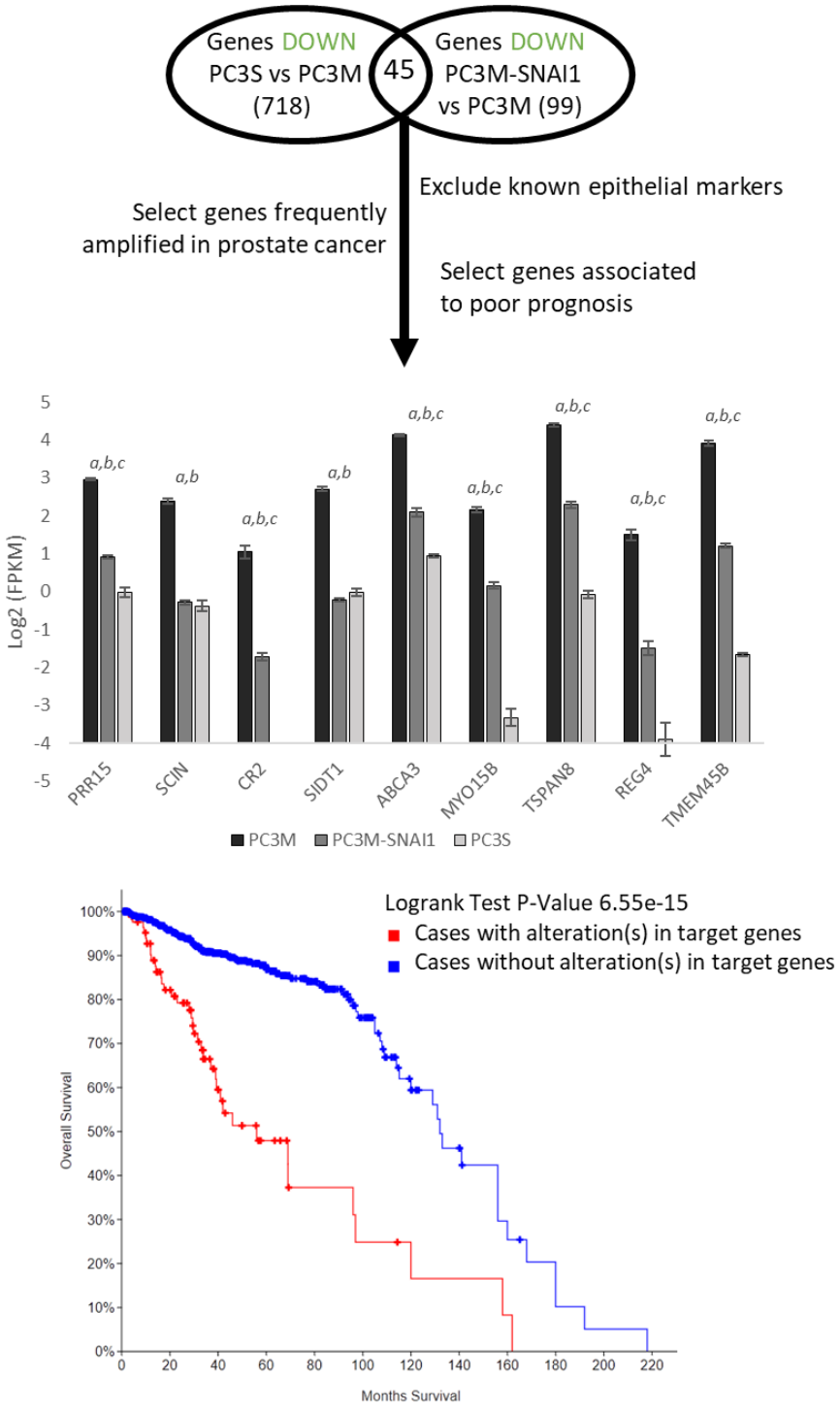


Figure22: Key players of the CSC phenotype associated to partial EMT. Such genes are upregulated in PC3M vs PC3S, downregulated upon Snai1 overexpression and are characterized by being frequently amplified in prostate cancer with such alterations associated to poor prognosis. PRR15(Proline-Rich Protein 15), SCIN (Scinderin), CR2(Complement C3d Receptor 2) SIDT1(SID1 Transmembrane Family Member 1), ABCA3 (ATP Binding Cassette Subfamily A Member 3), MYO15B (Myosin XVB), TSPAN8 (Tetraspanin 8), REG4(Regenerating Family Member 4),TMEM45B(Transmembrane Protein 45B)

- Integrating differentially consumed/produced metabolites.

- Adding the differentially expressed reactions to the quadratic component of the optimization to reduce the r^2MTA solution space towards a unique solution.
- Using the base MTA scores (the capacity to switch between states without any gene KO) to normalize the score given to each gene KO.

This allows both normalizing the transformation score and correcting for possible biases towards the best- or worst-case scenarios in the MTA optimization.

4.3.6. Metabolic reprogramming in the PC3M/PC3S/PC3M-Snai1 models

Using r^2MTA, we integrated the RNA-SEQ data together with previously published metabolomics data for PC3M and PC3S[333] to characterize the metabolic differences between PC3M/PC3S and PC3M-Snai1/PC3M. r^2MTA predicted that PC3M displayed an upregulated glycolysis, PPP, TCA cycle, oxidative phosphorylation and cholesterol, amino acid and folate metabolism (Figure 23). Conversely, PC3S have increased capacity on the nucleotide interconversion system as they have higher expression in some key nucleotide kinases and ribonucleotide reductase enzymes.

Conversely, central carbon and amino acid metabolism were not significantly affected by Snai1 overexpression. However, cholesterol metabolism and nucleotide interconversion pathways were downregulated and upregulated, respectively, bringing them closer to PC3S. Additionally, glutathione metabolism was also strongly downregulated in line with what is reported in the literature[518]. Overall, r^2MTA indicated that the metabolic gene expression program underlying the CSC-like PC3M metabolic phenotype is not strongly associated with the EMT status.

4.3.7. Identifying metabolic targets against CSC metabolism

With r^2MTA we identified putative metabolic targets that can be inhibited to cause a metabolic shift from a PC3M-like metabolic phenotype to a PC3S-like phenotype. As expected, the pathways upregulated in PC3M compared to PC3S are heavily enriched in putative targets to transform PC3M to PC3S (Figure 21D). Namely, folate metabolism (MTHFD2 and GLDC), proline metabolism (PYCR1), glycolysis (PGK1, TPI1, PGAM1, GAPDH), oxidative phosphorylation (UQCRC2, UQCRFS1, ATP5B) and glutamate metabolism (GLUD1).

Indeed, multiple targets were validated by our previous characterization of the PC3M model. For instance, oligomycin (an inhibitor of ATP synthase), BPTES (an inhibitor of glutaminase which is upstream of glutamate dehydrogenase in the glutaminolysis pathway) and 2-deoxyglucose(a glycolytic inhibitor) significantly reduced spheroid formation in PC3M[333]. Even more, both BPTES and 2-deoxyglucose reduced the proliferation to a greater extent in PC3M than PC3S[333]. Additionally, Metformin, reported to inhibit oxidative phosphorylation[519], and 2-deoxyglucose[520]

have been reported to reduce resistance to chemotherapy and radiotherapy in prostate cancer, traits which are often associated with the CSC subpopulation.

Concerning some of the other top-scoring targets, MTHFD2 has been found to promote tumorigenesis and CSC-like properties in lung cancer [521] and to support rapid proliferation and a lowly differentiated state in acute myeloid leukemia[304]. Similarly, PYCR1 (Pyrroline-5-Carboxylate Reductase 1), which catalyzes the last step of proline synthesis, has been associated to increased proliferation and apoptotic resistance in a variety of cancer types[522–524], including prostate cancer[525].

Remarkably, GLDC, identified as a putative metabolic target, was significantly downregulated upon Snai1 overexpression. This suggests that although folate metabolism might not be coordinately downregulated at the gene expression level as a result of SNAI1 overexpression, it might be significantly reduced through the downregulation of GLDC. Interestingly, GLDC has often been associated with tumorigenesis and pluripotency[526–528].

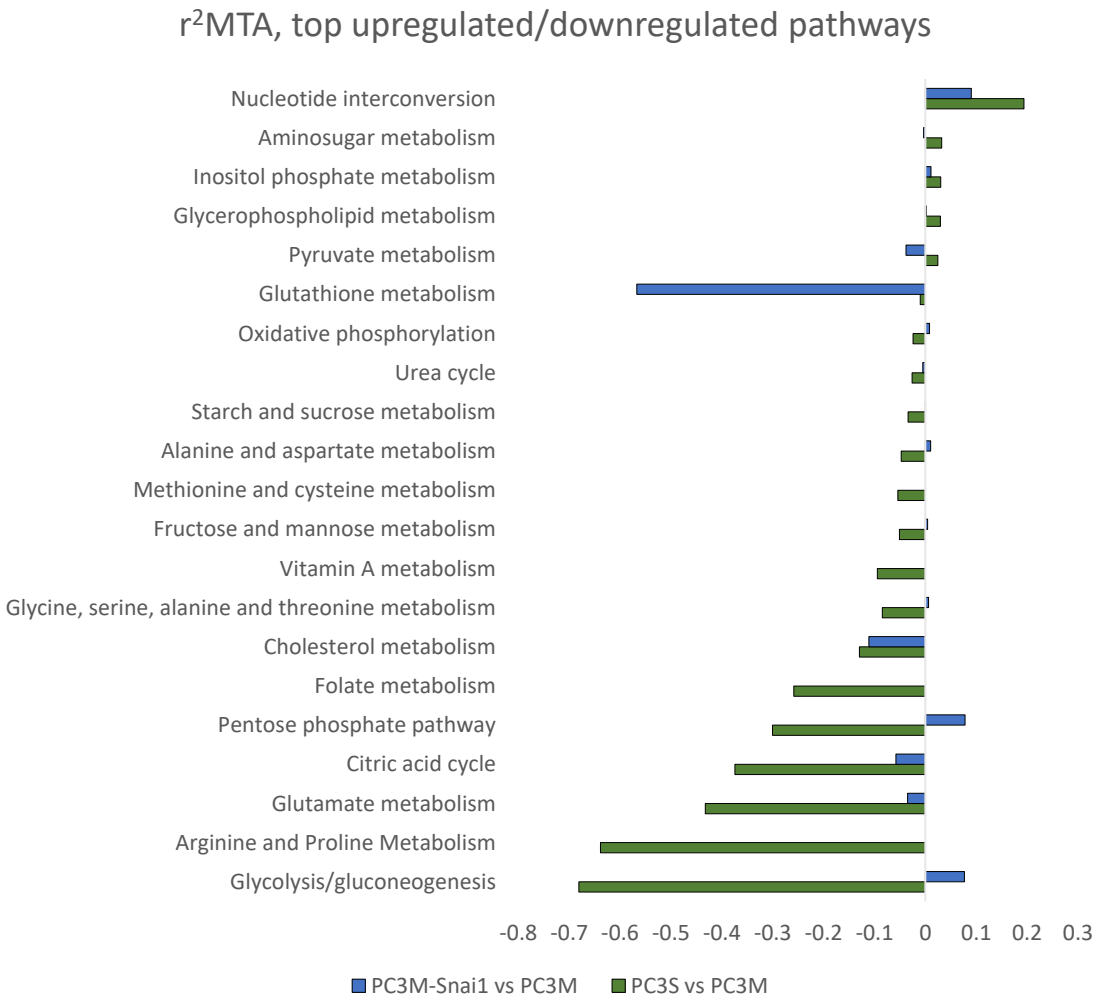


Figure 23(previous page). Top pathways upregulated and downregulated between PC3S and PC3M and, PC3M-Snai1 and PC3M. The sum of reaction flux variations as predicted by r^2MTA are plotted. Positive values represent pathways upregulated in PC3S or PC3M-Snai1 compared to PC3M while negative values represent pathways downregulated in PC3S or PC3M-Snai1 compared to PC3M.

Because folate metabolism was found to be strongly upregulated in PC3M compared to PC3S, and indeed two putative metabolic targets were identified in folate metabolism (i.e., MTHFD2 and GLDC), we theorized that the PC3M population could be differentially vulnerable to antifolates. In this regard, we evaluated the therapeutic potential of methotrexate and pemetrexed, two FDA-approved antifolates[345], on PC3M and PC3S(Figure 24). Both drugs were shown to have stronger growth inhibitory effects on PC3M than PC3S.

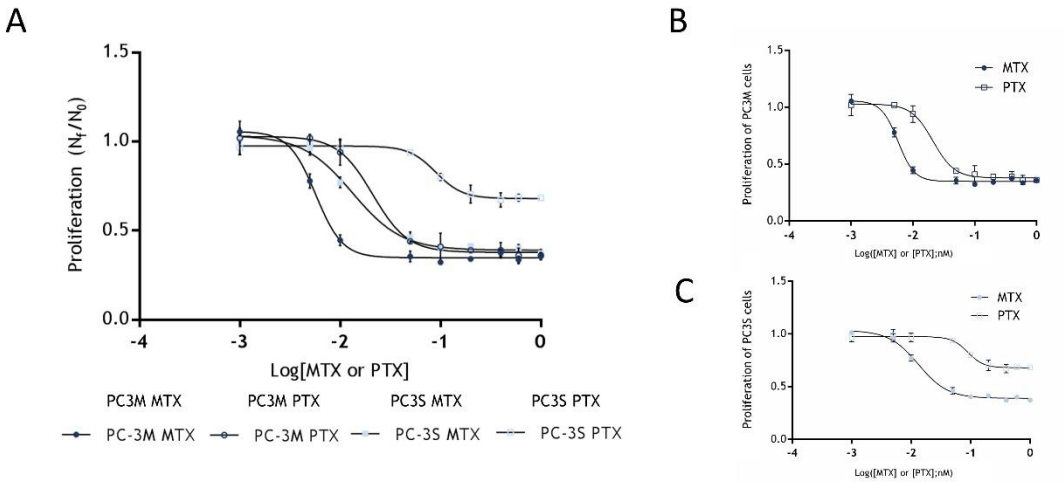


Figure 24. Effect of folate analogues methotrexate (MTX) and pemetrexed (PTX) on the proliferation of PC3M and PC3S cells. ***(A)*** *Effect of MTX on the cell proliferation of PC3M (solid circles) and PC3S (solid squares); and PTX on PC3M (open circles) and PC3S (open squares). Cell viability was assessed after 72 h incubation with either MTX or PTX (nM). Values represent mean ± SD of n=3.* ***(B)*** *Comparison of the effect of MTX and PTX on PC3M.* **(C)** *Comparison of the effect of MTX and PTX on PC3S.*

4.4. Cysteine and folate metabolism are major vulnerabilities of the metastatic subpopulations of colorectal cancer

Metabolic reprogramming has begun to emerge as a key feature of the acquisition of the metastatic phenotype by endowing metastatic cell populations with the capacity to cope with oxidative stress[220,529], adapt to varying substrate availabilities[530,531] and rapidly expand into macroscopic lesions[530,531]. However, such metabolic reprogramming might lead to the emergence of metabolic vulnerabilities that can be used to selectively kill metastatic cell populations[220,531].

With the aim of identifying such vulnerabilities, we performed a thorough metabolic characterization analysis of three $KRAS^{G12V}$-mutant[532] and same-patient derived cell lines with increasing metastatic potential: SW480, SW620, and SW620-LiM2. SW480 were isolated from the primary tumor, SW620 from a lymph node metastases of the same patient[533] and LiM2 were a metastatic-enriched derivative from SW620[155]. We determined that phenotypically, SW480 are mesenchymal-like whereas SW620 and LIM2 are epithelial-like. However, SW620 and particularly LiM2 also have significant expression of mesenchymal markers. Additionally, SW620 and LiM2 are endowed with increased proliferation and spheroid formation capacity.

To integrate the data generated from the metabolic characterizations, as well as data in publicly available databases, we developed a novel multiomics data integration workflow in the framework of constraint-based modeling. With such a workflow, we generated cell line specific GSMMs and used it to identify putative metabolic targets against the metastatic populations.

4.4.1. A workflow for multiomics data integration in the genome-scale

As part of the metabolic characterization of the cell lines of study, we generated an extensive heterogeneous data set. Such set included growth rates, rates of uptake and secretions of glucose, lactate, and amino acids, oxygen consumption rate and respiration parameters, ^{13}C resolved metabolomics and targeted metabolomics. Additionally, the set was expanded by including data from public datasets. For instance, transcriptomics for the cell lines of study was obtained from the literature[155,534,535]. Additionally, the effect of silencing metabolic genes on the viability of SW620 and SW480 viability was obtained from the Project DRIVE database[414].

To integrate such heterogeneous dataset, we developed the following workflow combining multiple constraint-based modeling algorithms:

1. Building a flux map of central metabolism. Stationary ^{13}C MFA was used to integrate ^{13}C resolved metabolomics, growth rates, rates of metabolite uptake and secretion, oxygen

consumption rate, and respiration parameters to estimate the range of fluxes through central metabolism. The resulting flux ranges were used to constrain the human GSMM model Recon2.2[392].

2. Integration of targeted metabolomics. Measures of intracellular concentration of amino acid and biogenic amines were used to constrain the cell line specific GSMMs. This was achieved by adding a sink reaction for each measured metabolite representing the required production of such metabolite to maintain its concentration constant in a proliferating system (i.e., to compensate the dilution associated with proliferation)[a]. Additionally, lipidomics measurements were used to customize the biomass reaction of each cell line.
3. Minimal cut set (MCS) analysis. Based on the essential metabolic genes reported in project DRIVE[414] for SW620 and SW480, the reactions that should be inactive for such genes to be essential were determined using MCS analysis[b420,422]. Such reactions were subsequently inactivated in the model. For LiM2, we implemented all the MCSs shared between SW480 and SW620 and the MCS in SW620 where none of the genes were significantly overexpressed.
4. Integration of transcriptomics data. GIMME[400,405] was used to integrate transcriptomics data and restrict the maximum flux through reactions and pathways based on their gene expression evidence[c].
5. Selection of the most representative flux samples. Flux sampling was used to compute possible flux combinations consistent with the above-integrated data. Each flux sample was then used to systematically simulate the effect of all gene KOs on biomass production and rank them based on their consistency with the gene dependency data from Project DRIVE. The goal of this step is both to minimize the false positives of gene essentiality and to integrate partial dependencies on genes that while not essential might have a significant impact on cell viability. For each cell line, the top 100 ranked flux samples were selected as representative flux distributions.

In order to identify the targets with minimal side effects on non-tumoral cell populations, a condition-specific GSMM was also built for the NCM460 cell line, derived from healthy colon mucosal epithelium[536]. This was achieved using a similar workflow as the above described but omitting step 3 and 5.

---

[a] See section 1.6.3.3.4
[b] See section 1.6.3.4.1.3
[c] See section 1.6.3.3.5.2

4.4.2. The metastatic cell lines display increased Warburg effect glutaminolysis and oxidative metabolism

We determined that the metastatic cell lines (SW620 and LiM2) had a stronger Warburg effect than SW480, as they consumed significantly more glucose and produced more lactate. Remarkably, the cell-line specific flux maps show that metastatic cell lines also have more flux through PDH (Figure 25). While it might seem paradoxical that the metastatic cell lines have both stronger Warburg effect and increased PDH activity, it must be noted that flux through PDH was quite low compared to lactate production ($\sim 10\%$). Such increase of PDH flux appears to be mediated by a decrease in PDH phosphorylation as SW620 and LiM2 had significantly less fraction of phosphorylated PDH than SW480. Similarly, glutaminolysis was also upregulated in the metastatic cell lines, which had increased glutamine consumption and glutamate production (Figure 25).

Additionally, the metastatic cell lines were also shown to have increased mitochondrial function (Figure 25), and the mitochondrial fuel test determined that such function was highly flexible and could easily adapt to variable substrate availability. Hence, increased mitochondrial function might endow the metastatic cell lines with enhanced capacity to maintain a constant supply of ATP under variable, and potentially hostile, metabolic environments[537].

In this regard, increased Warburg effect, glutaminolysis and mitochondrial functions are consistent with the known effects of MYC[85,86,308,310]. Because MYC is overexpressed in the metastatic cell lines, such oncogene might play a key role in supporting the metabolic reprogramming underlying the metastatic phenotype.

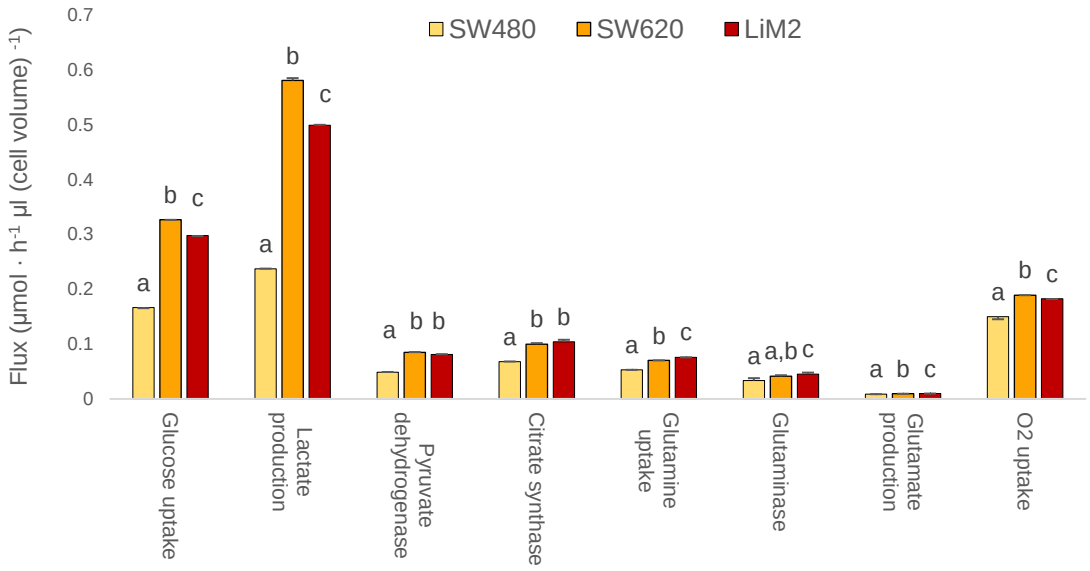


Figure 25(previous page). Metabolic fluxes predicted by the multiomics integration workflow for some key reactions and transport processes. Cell lines with a distinct letter (a,b,c) had no overlap in the 100 flux samples selected through multiomics integration workflow (i.e., the largest flux sample value in one condition was lower than the higher flux sample value in the other condition).

4.4.3. Putative metabolic targets

The cell line-specific GSMMs were used to identify essential or synthetic lethal gene pairs that could selectively inhibit growth in the metastatic cell lines SW620 and LiM2. Excluding the targets predicted to significantly impair the proliferation of normal colonic epithelial cells, represented by the cell line NCM460, 6 single putative gene targets and 9 putative synthetic lethal pairs were identified. We decided to focus our attention on the MTHFD1 single target and on synthetic lethal combinations inhibiting both the Xc- (SLC7A11 and SLC3A2) and $b^{0,+}$ (SLC7A9 and SLC3A1) cystine transporters.

4.4.4. Metastatic cell lines are vulnerable to Xc- inhibition

The synthetic lethal pairs targeting both Xc- and $b^{0,+}$ emerge because the condition-specific GSMM predict that SW620 and LiM2 are cysteine/cystine auxotroph. This emerges from the MCS analysis that predicts that essentiality of GSR in SW620 reported in project DRIVE[414] emerges from insufficient activity in both thioredoxin dependent cystine reduction and the cystathionine β-synthase activity in the transsulfuration pathway(Figure 26A). This hypothesis was validated by cystine deprivation (Figure 26B) which had significantly more effect on the proliferation of the metastatic cell lines (SW620 and LiM2) than in SW480.

Concerning the Xc- and $b^{0,+}$ systems, the former is generally assumed to play a greater role in cancer and it is often overexpressed in cancer cells[316,336,338,342,538]. Even more, its expression has been shown to be consistently higher in cell lines with mutant KRAS[343], such as the SW480-SW620-LiM2 models, and its activity and stability enhanced through the interaction with CD44v6[340] which has been described as marker of metastatic potential and CSC phenotype in colorectal cancer[539]. Additionally, cell line specific flux maps predicted significantly more flux through the Xc- system than $b^{0,+}$ system across all analyzed cell lines(25C). Hence, we theorized that inhibition of Xc- alone might be sufficient to inhibit growth in the metastatic cell lines.

With this aim, we evaluated effects of Erastin, an irreversible inhibitor the Xc- system[317,540,541], in the proliferation and viability of SW480, SW620, LiM2 and the non-tumoral colon cell line NCM460. As expected, Erastin had significantly more growth inhibitory effects on the metastatic cell lines (Figure 26D) which also had their viability significantly reduced by the drug(figure 26E).

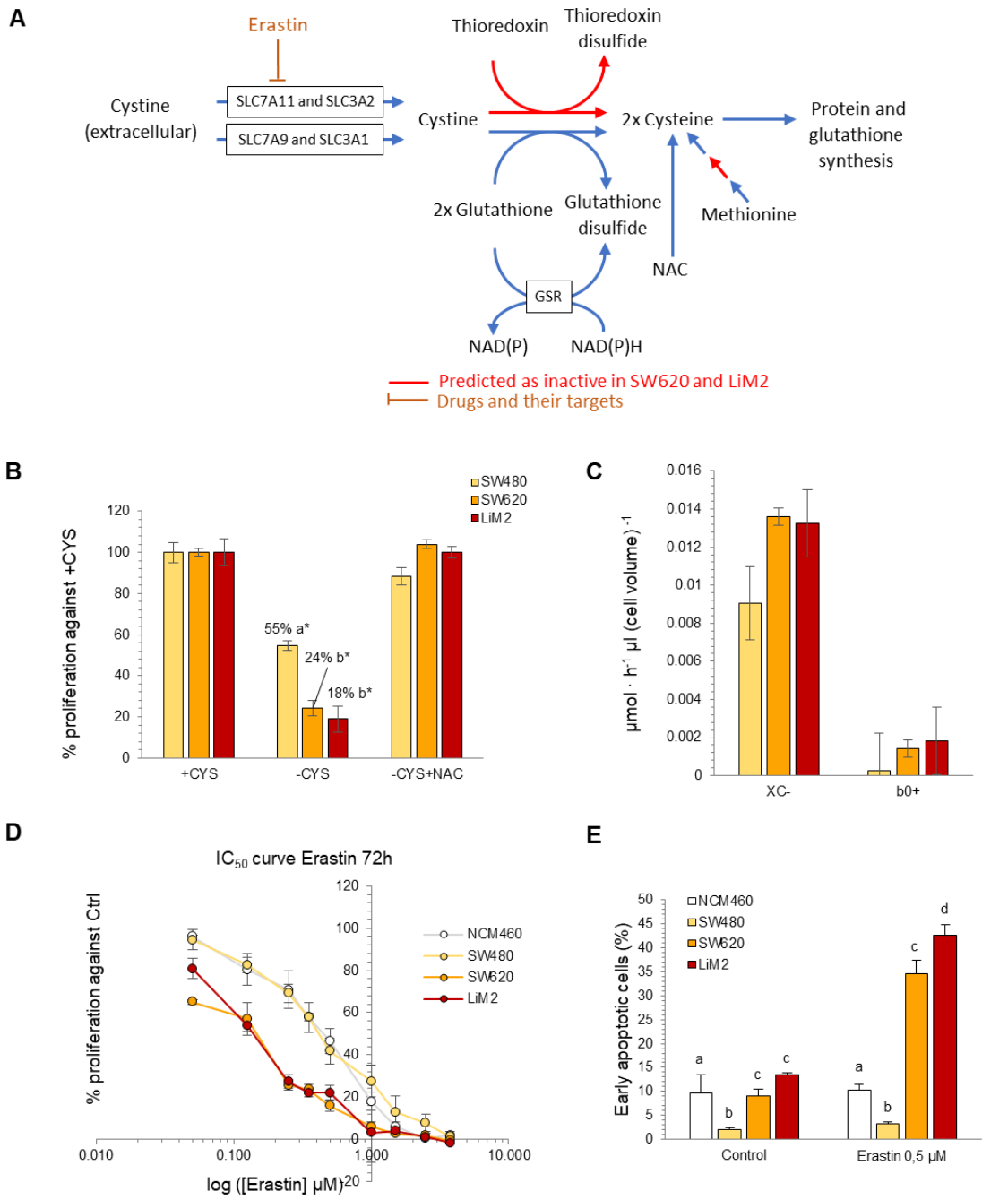


Figure 26. Metastatic cells are dependent on cystine uptake and vulnerable to system Xc- inhibition. ***(A)*** *Graphical representation of cysteine and glutathione metabolism. GSR: Glutathione-Disulfide Reductase. SLC3A1: B(0,+)-Type Amino Acid Transport Protein. SLC3A2: Solute Carrier Family 3 (Activators Of Dibasic And Neutral Amino Acid Transport), Member 2. SLC7A11: Amino Acid Transport System Xc. SLC7A9: B(0,+)-Type Amino Acid Transporter 1.* ***(B)*** *Cell proliferation measured by DNA content using HO33342 under control conditions (12,5 mM Glc and 4 mM Gln), under cysteine deprivation and adding N-acetylcysteine.* * *t Student test for –CYS or -CYS+NAC vs Control conditions, p<0.05.* ***(C)*** *Predicted fluxes through the Xc- and* $b^{0,+}$ *system.* ***(D)*** IC_{50} *curve of the system* Xc^{-} *inhibitor Erastin assessed by cell proliferation measured by DNA content using HO33342 under various Erastin concentrations after 72h incubation.* ***(E)*** *Percentage of apoptotic cells measured by imaging with Incucyte® adding a caspase-3/7 green reagent with various Erastin concentrations after 72h incubation. One-way ANOVA was performed for the factor "cell line", and Scheffe's test was used for multiple comparisons. Groups sharing the same letter don't show a significant difference with α=0.05.*

4.4.5. The metastatic cell lines are vulnerable to folate metabolism inhibitors

In the project DRIVE database[414], it was reported that SW620 but not SW480 were highly dependent on the expression of MTHFD1 which codes for an enzyme catalyzing three steps of the cytosolic branch of folate metabolism (Figure 27A). We confirmed that such dependency was related to MTHFD1 enzymatic activity by evaluating the effect of the drug LY345899 on the cell lines under study. LY345899 is an inhibitor of both MTHDF1 and MTHFD2 but with a significantly lower Ki for the former[542,543]. Such inhibitor was shown to selectively inhibit the proliferation of SW620 and LiM2 confirming the dependency of the metastatic cell lines on the MTHDF1 enzymatic activity(Figure 27B).

MCS analysis predicted that such dependency emerged because the CHO-THF generated into the mitochondria cannot be transported to the cytosol to compensate for MTHDF1 deficiency (Figure 27C). CHO-THF is the substrate of GART activity, which is necessary for *de novo* synthesis of purines. Hence, the model predicted that in the metastatic cell lines folate metabolism was to some extent uncoupled between the cytosol and the mitochondria.

On this subject, SHMT1, thymidylate synthase, and MTHFD1 have been reported to translocate to the nucleus during the S phase of the cell cycle in order to promote thymidylate synthesis[544–546]. Similarly, although purine synthesis is assumed to occur primarily in the cytoplasm[547], recent evidence suggests that several key enzymes of the pathway, such as GART, are partially localized in the nucleus where the pathway can be active in an MTHFD1-dependent manner[548]. Hence, we theorize that the dependency of MTHFD1 and the reduced coupling between the cytosolic and mitochondrial branches of the folate cycle could possibly be attributed to increased nuclear localization of the folate pathway to the nucleus in the metastatic cell lines. Indeed, in some cellular models, inhibition of SHMT1 or MTHFD1 has been reported to induce cell cycle arrest and compromise cell viability by reducing thymidylate synthesis and thus leading to uracil incorporation into DNA[349,549–551].

In addition to their dependency of MTHFD1, we also predicted that the metastatic cell lines (i.e., SW620 and LiM2) had a significantly more active folate metabolism than SW480 and hence could be more vulnerable to antifolates. In this context, Methotrexate, Pemetrexed, and Lometrexol were evaluated and shown to have more growth inhibitory effects in the metastatic cell lines. Additionally, SHIN2[350], a chemical inhibitor of both SHMT1/2, was also tested. Remarkably it had similar growth inhibitory effect on all colorectal cancer cell lines (SW480 and SW620 and LiM2). However, SHIN2 was a selective drug for colorectal cancer cells as it was shown to have less growth inhibitory effect on the immortalized colon epithelial cell line

NCM460. Additionally, SHIN2 significantly induced more apoptosis in the metastatic cell lines (Figure 27E).

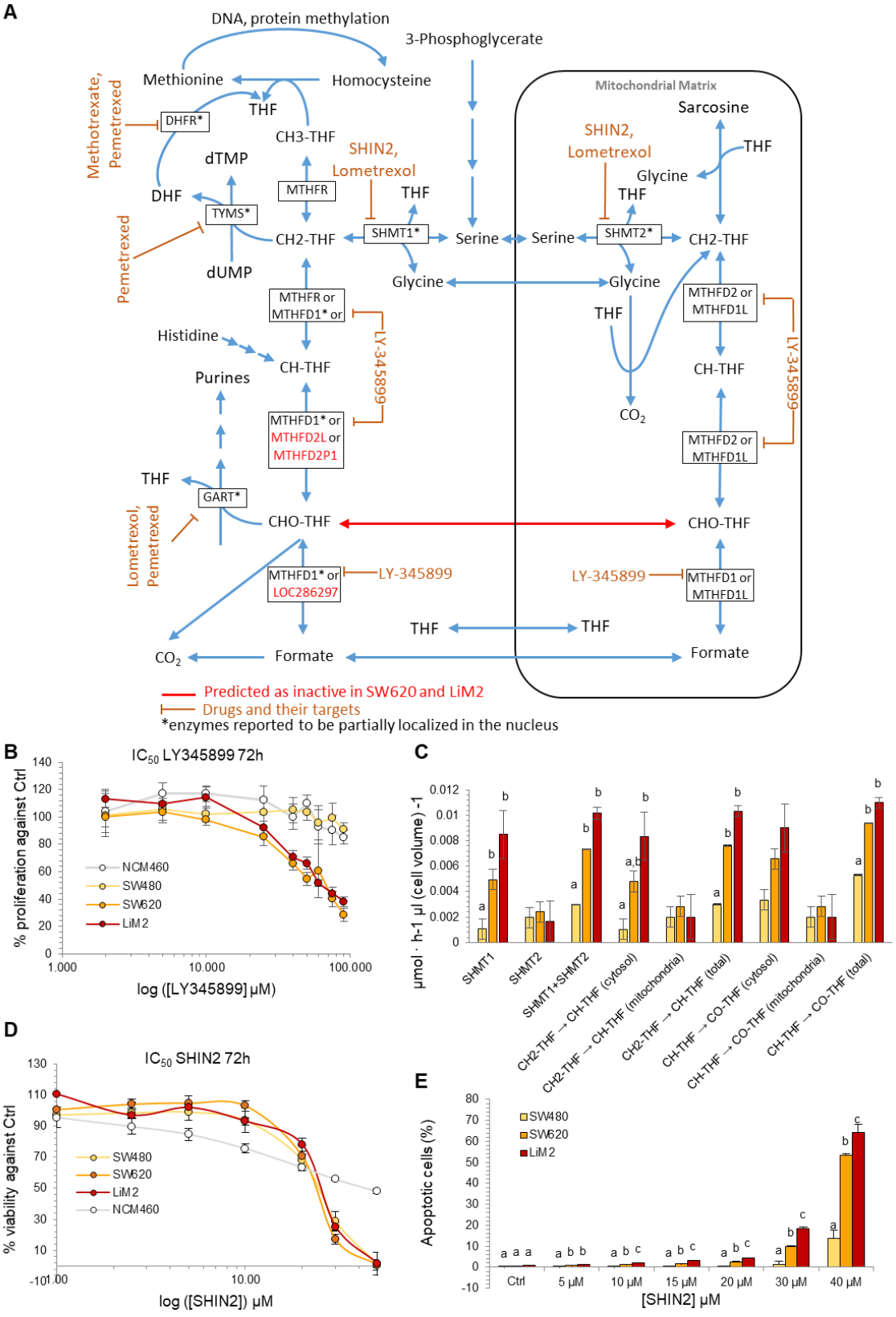


Figure 27(previous page). Metastatic cells are vulnerable to MTHFD1 and SHMT inhibition ***(A)*** *Graphical representation of folate metabolism. CH2-THF: N^5,N^{10} methylene-THF. CH3-THF: N^5-methyl tetrahydrofolate. CHO-THF: N^{10}-formyl tetrahydrofolate. CH-THF: N^5,N^{10}-methenyl tetrahydrofolate. THF: Tetrahydrofolate. DHFR: Dihydrofolate Reductase. GART: Trifunctional Purine Biosynthetic Protein Adenosine-3. LOC286297: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 1 Like Pseudogene. MTHFD1: Methylenetetrahydrofolate Dehydrogenase, Cyclohydrolase And Formyltetrahydrofolate Synthetase 1. MTHFD1L: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent)1 Like. MTHFD2: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 2, Methenyltetrahydrofolate Cyclohydrolase. MTHFD2L: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent)2 Like. MTHFD2P1: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent)2, Methenyltetrahydrofolate Cyclohydrolase Pseudogene 1. MTHFR: Methylenetetrahydrofolate Reductase. SHMT1/2: Serine Hydroxymethyltransferase1/2. TYMS: Thymidylate Synthetase.* ***(B)*** *IC_{50} curve of the MTHFD1 inhibitor LY345899 determined by cell proliferation measured by DNA content using HO33342 under various LY345899 concentrations after 72h incubation.* ***(C)*** *Predicted flux values for different reactions of cytosolic, mitochondrial folate and metabolism. Cell lines with a distinct letter (a,b) had no overlap in the 100 flux samples selected through the multiomics integration workflow.* ***(D)*** *IC_{50} curve of the SHMT1/2 inhibitor SHIN2 assessed by cell proliferation measured by DNA content using HO33342 under various SHIN2 concentrations after 72h incubation.* ***(E)*** *Percentage of apoptotic cells measured by imaging with Incucyte® adding a caspase-3/7 green reagent with various SHIN2 concentrations after 72h incubation. A one-way ANOVA was performed for the factor "cell line", and Scheffe's test was used for multiple comparisons. Groups sharing the same letter don't show a significant difference with α=0.05.*

4.4.6. High citrate synthase and PDH activities protect the metastatic cell lines from carnosine accumulation

We determined that the metastatic cell lines (SW620 and Lim2) had significantly lower intracellular carnosine levels than SW480(Figure 28.A). Carnosine is a naturally occurring dipeptide of β-alanine and histidine which has been reported to inhibit cancer cell proliferation *in vitro* and *in vivo*[552–559]. The limiting metabolite for carnosine synthesis is generally assumed to be β-alanine[560] which can be synthesized from aspartate decarboxylation, spermine oxidation, uracil catabolism and from propionyl-CoA[378]. Because in the SW480 cell line both carnosine and aspartate concentration were significantly higher than other cell lines, it becomes apparent that aspartate decarboxylation was the predominant pathway fueling β-alanine synthesis.

In order to identify the drivers of carnosine synthesis, the condition-specific GSMMs were used to systematically analyze the effect of gene KO on carnosine accumulation. According to such simulations, the most efficient intervention to increase carnosine synthesis was the inhibition of citrate synthase or PDH. Simulations predicted that inhibiting either PDH or citrate synthase would lead to an accumulation of aspartate which could be redirected towards β-alanine synthesis and fuel carnosine production (Figure 28B). Indeed, the flux through citrate synthase and PDH was significantly higher in SW620 and LIM2 than SW480 (Figure 25), suggesting that

high activity of such enzymes contributed to preventing accumulation of carnosine in the metastatic cell lines.

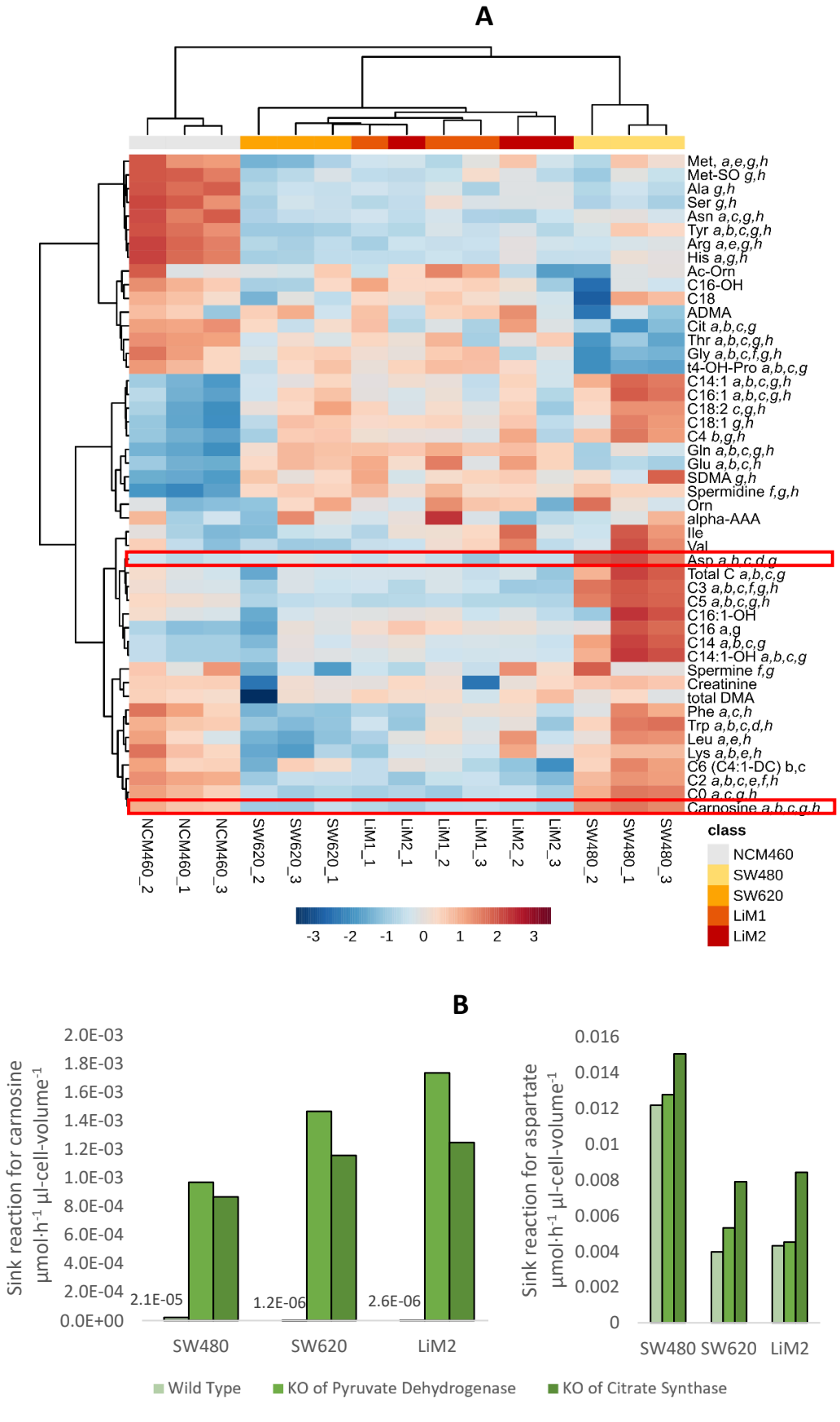


Figure 28(previous page). Targeted Metabolomics. ***(A)*** *Heatmap for intracellular content of amino acid, biogenic amines, and acyl-carnitines. Concentrations were normalized by protein and mean-centered and divided by the standard deviation of each metabolite. Hierarchical clustering for samples and features was performed using the Ward algorithm and Euclidian distance. ANOVA was used to determine statistically significant features, and Fisher's least significant difference method was used to evaluate statistically significant differences (a: statistically significant difference between SW480-SW620 , b: SW480-LiM1 , c: SW480-LiM2, d: SW620-LiM1, e: SW620- LiM2, f:LiM1 - LiM2, g: NCM460-SW480, h: NCM460-SW620). Cx:y denotes an acylcarnitine with x carbons and y instaurations in the acyl moiety.* ***(B)*** *Simulated flux through the sink reactions for carnosine and aspartate in the WT or under KO of citrate synthase and pyruvate dehydrogenase. Sink reactions simulate the required intracellular production of a metabolite to maintain its concentration constant under proliferation. The flux through a sink reaction is proportional to the intracellular concentration of the metabolite and the proliferation rate of the cell line.*

5. Global discussion

5. Global discussion

Systems Biology uses metabolic models (i.e., computable sets of assumptions and hypothesis of metabolism) to integrate multiple layers of data and analyze the underlying metabolic phenotype[352,353]. In this regard, there exist various complementary modeling approaches, each one with its advantages and limitations. Kinetic modeling[357] and ^{13}C MFA[425] (which can be combined as part of dynamic ^{13}C MFA[354]) can excel at predicting central carbon metabolism flux maps with great detail. Conversely, constraint-based modeling in the framework of GSMMs excels at simulating metabolism at a genome-scale by integrating large multiomics data sets[371,372,408]. As part of this Ph.D. thesis, we expanded the Systems Biology toolbox by developing new model-based approaches for flux simulations both for central carbon metabolism and genome-scale, as well as new approaches aiming to bridge both modeling layers.

On this subject, we developed HepatoDyn[362], a highly detailed kinetic model of hepatocyte central carbon metabolism capable of performing ^{13}C dynamic MFA[561]. Compared to existing kinetic models[63,366–370], HepatoDyn stands out for combining a large scale, a detailed modeling of fructose metabolism, the capacity to dynamically simulate redox and energy metabolism (e.g., ATP/APD/Pi, NAD/NADH) and the capacity of integrating ^{13}C resolved metabolomics. The abovementioned features allow simulating hepatic central carbon metabolism with an unprecedented level of detail. As a proof of concept of the capabilities of the model, we applied it to study the effects of high fructose concentration on hepatic function. We determined that high doses of fructose can severely impair the metabolic function of hepatocytes by causing an ATP depletion[443,444] that impairs both the capacity to synthesize glycogen and clear glucose from blood. This phenomenon has a strong dynamic component and is dependent on both the kinetic properties of enzymes and on ATP metabolism, and hence it perfectly encapsulates the need for a model with HepatoDyn unique capabilities. Furthermore, such a phenomenon is also of high biomedical interest as fructose-induced ATP depletion is reported to increase the susceptibility to nonalcoholic fatty liver disease[562]. Even more, modeling fructose metabolism has biomedical interest beyond hepatocytes as certain cancer cell populations, such as colorectal cancer metastases in liver[563] and lung adenocarcinoma cell lines[564], have also been reported to have high fructose uptake. Hence, future works could use kinetic modeling to explore the feasibility of therapeutic interventions aimed at inducing fructose mediated-ATP depletion in such cancer populations.

Although dynamic ^{13}C MFA is a powerful modeling technique, it is also quite complex and ideally requires several metabolic conditions (e.g., separate incubations with 3mM and 20mM fructose) or measures at several time points. For such reason, the more accessible stationary ^{13}C MFA is often used for flux analyses at metabolic steady state and at incubations times where most metabolites have reached isotopic steady state[425]. As part of this Ph.D., we developed Iso2Flux, an open-source stationary ^{13}C MFA software, and implement it into the cloud in the framework of PhenoMeNal[446,447]. We believe that this will contribute to increasing the usage of ^{13}C MFA by enabling researchers and clinicians without a computational background to perform ^{13}C MFA.

However, when ^{13}C MFA is applied either with a small set of experimental data or with a large metabolic network, the resulting solution space can be too wide to draw meaningful conclusions about the underlying flux distribution. Indeed, such limitation has traditionally led to qualitative or semiquantitative approaches, such as pathway activity analysis, being used when only small sets of ^{13}C measurements are available[425,426,565–567]. To address such limitations and drawing from our experience working with GSMMs (i.e., chapter 3 and 4) where we extensively used the algorithm GIMME[400,405] to reduce the solution space, we developed p^{13}CMFA which uses flux minimization to identify the best solution in an undetermined ^{13}C MFA solution space[442]. Indeed, as p^{13}CMFA can be weighted by gene expression, it also allows for the first-time ever seamless integration of ^{13}C resolved metabolomics with transcriptomics data. Furthermore, we demonstrated that, when only a limited set of measurements is available, p^{13}CMFA can yield more accurate flux predictions than both ^{13}C MFA and GIMME. We aim to make p^{13}CMFA available in the next release of PhenoMeNal to facilitate its widespread use by the scientific community.

Moving beyond the current implementation of p^{13}CMFA, an exciting prospect might be to implement this approach to dynamic ^{13}C MFA. Indeed, working with HepatoDyn, we noticed that a significant number of fitted enzyme activities had a substantial degree of uncertainty. Hence minimizing total enzyme activity, coupled to gene expression and enzyme activity measurements, could be an effective approach to reduce such uncertainty. Furthermore, with atom mappings now available on a genome-scale for humans GSMMs[393] and some early successes in bacteria[439], likely, ^{13}C MFA will progressively move from central carbon metabolism to the genome-scale. Because the undetermined parts of the ^{13}C MFA solution space increase with the size and complexity of the metabolic network, p^{13}CMFA might prove to be essential to support this change of scale. We note however that this will require either a significantly more efficient implementation of p^{13}CMFA or a leap in computing power as the complexity of solving the p^{13}CMFA non-linear problem increases with the size of the network.

Nevertheless, until the possibility of genome-scale ^{13}C MFA or p^{13}CMFA becomes a reality, genome-scale flux maps can be simulated using constraint-based modeling approaches in the framework of GSMMs. In this context, we developed a workflow for integrating large multiomics data sets to build condition-specific GSMMs which can be then used to identify putative essential and synthetic lethal genes. Remarkably, a key step of such workflow is constraining the GSMM using central carbon metabolism flux maps computed from ^{13}C MFA, thus partially combining both modeling layers. In the context of GSMM, we also developed r^{2}MTA which can integrate both transcriptomics and rates of metabolite uptake and secretion to characterize the metabolic differences between two conditions and identify targets to induce a transition towards the biomedically desirable state. Indeed, a promising future prospect is to combine both approaches and enable r^{2}MTA to integrate large multiomics data sets, including central carbon metabolism flux maps estimated with ^{13}C MFA, to define the flux distribution in both the source and target metabolic state and identify the putative drivers of the transition between both states.

The above-mentioned tools were used to study the metabolic phenotype underlying CSC and metastasis in prostate and colorectal cancer, respectively. In this regard, the studied cellular models, SW480/SW620/LiM2 and PC3S/PC3M had some striking similarities. For instance, each cellular model was single-patient derived, and the most aggressive populations (i.e., LiM2 and PC3M) had been obtained by selecting the subpopulations with enhanced capacity to metastasize in animal models[155,157]. In this context, PC3M, LiM2, and to a slightly lesser extent SW620, were endowed with a strong tumorigenic capacity *in vivo*[155,157]. Similarly, SW620/LiM2 and PC3M were endowed with a faster proliferative rate and capacity to form spheroids than their less aggressive counterparts (i.e., SW480 and PC3S). However, an important distinction is that PC3M and PC3S were obtained from single-cell cloning[157] and hence each cell line can be assumed to represent a mostly homogeneous population whereas this is not necessarily the case with the LiM2, SW620 or SW480 cell lines[154]. In this context, because PC3M had a high expression of pluripotency markers (e.g., MYC, KLF4, and SOX2) and were shown to be capable of differentiating into a PC3S-like population *in vitro* and *in vivo*[157] they can be considered as a cellular model of CSCs in prostate cancer[157,287,291,333]. Conversely, although a CSC subpopulation within the bulk of the cell lines likely plays a crucial role in their tumorigenic potential, with the available experimental evidence, SW620/LiM2 are more accurately defined as a cellular model of metastatic populations in colorectal cancer[155].

Another common feature between both cellular models was that the most tumorigenic populations (i.e., SW620/LiM2 and PC3M) had a more epithelial-like phenotype than the less

aggressive populations (i.e., SW480 and PC3S) which were more invasive. In this context, it is beginning to emerge that metastatic or CSC properties can be associated with the expression of epithelial markers such as E-cadherin[186,188–191]. Indeed, a previous work determined that the tumorigenic potential of PC3M could be decreased by inducing a complete EMT through genetic modifications[157]. In this regard, it has been proposed that both the metastatic and CSC phenotypes might be supported by an intermediate EMT where the expression of both mesenchymal an epithelial markers coexist[137,138,192–194]. Indeed, using RNA-SEQ analysis we determined that PC3M most likely displayed a partial EMT phenotype maintained by a balance between pro-EMT factors (e.g., SNAI1, SNAI2, and LATS2) and anti-EMT factors such as KLF4 or PRKD1. Even more, we determined that forcing a higher degree of EMT to PC3M through Snai1 overexpression partially reversed a significant number of the gene programs differentially expressed between the PC3M and PC3S cellular models. Indeed, when Snai1 was overexpressed several genes associated with the CSC phenotype (i.e., genes that promote proliferation, chemotherapy resistance or facilitate invasiveness and metastatic spread) were strongly downregulated. Such genes included secretory proteins with potential autocrine and paracrine roles (e.g., VEGFC and REG4), and surface proteins (e.g., ITGA5, NRP2, TMEM4B, and TSPAN8) and could potentially be involved in signaling cascades that endow the PC3M cell line with the CSC phenotype[498,500,501,508,511,568–571].

Concerning the SW480/SW620/LiM2 models, western blot analysis showed that both SW620 and LiM2 had strong expression of some mesenchymal markers such as VIM, CDH2 and, FN1 with the expression of such markers being higher in LiM2. Hence it seems likely that SW620/LiM2 also display a partial EMT phenotype and future works should explore whether inducing a complete EMT in such populations could inhibit their metastatic capacity. Surprisingly, none of the putative key players of partial-EMT that we identified in the PC3M/PC3S model (e.g. SNAI1, SNAI2, LATS2, KLF4 and PRKD) were differentially expressed at the transcriptional level between SW480/SW620/LiM2[155,534] suggesting that either they are regulated at the post-transcriptional level, or that the putative partial-EMT is modulated through other factors in the colorectal cellular models.

At the metabolic level, PC3M/PC3S/PC3M-SNAI1 were analyzed with r^2MTA with the goal of identifying therapeutic targets that can revert the CSC-like metabolic phenotype (PC3M-like) to a non-CSC metabolic phenotype (PC3S-like). Conversely, SW480/SW620/LiM2 were analyzed with the multiomics integration workflow, which was used to generate cell-line specific flux maps and identify metabolic targets that could selectively kill SW620 and LiM2. In this regard,

the colon epithelial cell line NCM460 was also analyzed to select those targets with minimal side effects on non-tumoral proliferating cell populations.

We identified some commonalities in both the prostate and colorectal cellular models. For instance, the Warburg effect and glutaminolysis were upregulated in both PC3M compared to PC3S and in SW620/LiM2 compared to SW480. In this context, the Warburg effect is widely reported to support rapid proliferation, reduce ROS production and promote invasiveness, traits which can potentially support both metastasis and the CSC phenotype[220,231,232,235,333,572–576]. Similarly, enhanced glutaminolysis can also support rapid proliferation and protect cells from oxidative stress[10,312–315] and has been associated with metastatic potential and the CSC phenotype[333,577–583]. Remarkably, MYC is reported to promote both the Warburg effect and glutaminolysis[85,308,310] and it has increased expression in both PC3M and SW620/LiM2 compared to PC3S and SW480, respectively. This suggests that MYC might play a key role in the metabolic reprogramming supporting both the metastatic and CSC populations in the cell models of study.

In addition to the Warburg effect and glutaminolysis, both PC3M and SW620/LiM2 displayed more activity of folate metabolism than their less aggressive counterparts (i.e., PC3S and SW480). In this context, folate metabolism can also support both the metastatic and CSC populations by promoting nucleotide synthesis, epigenetic regulation and, energy and redox balance[269,293,294,297–299]. Indeed, we determined that such overactivation of folate metabolism led to PC3M and SW620/LiM2 being selectively vulnerable to antifolates compared to PC3S and SW480/NCM460, respectively.

However, while SW620 and LiM2 relied primarily on the cytosolic branch of folate metabolism, PC3M relied on the mitochondrial branch. Such dependencies represent a vulnerability as SW620/LiM2 could be selectively killed by inhibiting the cytosolic activity MTHFD1 while we predicted that the CSC-like phenotype of PC3M could be potentially attenuated by inhibiting the mitochondrial activities MTHFD2 and GLDC. We note that there is ample literature supporting that MTHFD1, MTHFD2 or GLDC can be targeted to inhibit cancer cell proliferation or tumorigenic potential in a variety of cancer types[292,304,521,526,528,551,584,585], hence different cancer cell populations might be differentially dependent on the cytosolic or mitochondrial branches of folate metabolism.

Additionally, in the colorectal metastasis model, we identified cysteine metabolism as a specific vulnerability of the metastatic cell lines. Through the multiomics integration workflow, we determined that in the metastatic cell lines the transsulfuration pathway is unable to produce enough cysteine to meet their demands for protein and glutathione synthesis and hence are

dependent on cystine uptake from the extracellular media. In this context, chemical inhibition of the Xc- cystine/glutamate antiport system was shown to selectively inhibit the proliferation and viability of the metastatic populations (i.e., SW620 and LiM2). Indeed, such treatment would synergize well with electrophilic chemotherapeutic agents (e.g. cisplatin)[316,317] or radiotherapy[586,587] as resistance to such therapeutic interventions is mostly glutathione-dependent.

Concerning the root cause of the dependency on cystine uptake in the colorectal cancer model, increased glutathione synthesis might promote metastatic spread by countering the oxidative stress generated as a result of the detachment from the extracellular matrix[318,588–591]. Indeed, inhibition of the Xc- system has been reported to cause a form cell death triggered by excessive oxidative stress as result of glutathione depletion[540,541,592]. Hence, we theorize that, as part of the metabolic adaptation underlying the acquisition of metastasis, SW620 and LiM2 might have upregulated glutathione metabolism and then subsequently become dependent on it to maintain redox balance and proliferate. In this context, it is worth noting that both SW620 and LiM2 had a more active oxidative metabolism than SW480 and could be expected to have a higher rate of ROS generation under basal conditions[593]. Regarding the prostate cancer cellular models, glutathione metabolism did not appear to be differentially active between PC3M and PC3S. Even more, PC3M were shown to have lower glutathione levels than PC3S[333]. Hence, targeting the Xc- system is not likely to be a selective therapeutic approach against PC3M.

Additionally, in the colorectal cancer model, we identified that the metastatic cell lines had significantly less intracellular concentration of the antiproliferative peptide carnosine[552–557,559] than SW480. Similarly, previous work had determined that PC3M were shown to have less content of long fatty acids with antiproliferative properties than PC3S[291]. Indeed, the proliferation of PC3M could be selectively inhibited by blocking fatty acid oxidation leading to the accumulation of such fatty acids[291]. Similarly, MTHFD2 and PYCR1, identified as targets against the PC3M phenotype using r^2MTA, have been suggested to support cancer proliferation by contributing to maintaining low concentrations of 4-carboxamide ribonucleotide and pyrroline-5-carboxylate, respectively, which are described to have anti-proliferative effects[521,523,524]. Hence, we theorized that inducing carnosine synthesis could prove to be an effective therapeutic strategy against SW620 and LiM2. With this aim, we performed a systematic KO of all genes within the cell line specific GSMMs and evaluated their capacity to induce carnosine synthesis. We identified that carnosine synthesis is most likely driven by the GAD1-catalyzed aspartate decarboxylation[378,594] and that it could be enhanced by increasing aspartate concentration through inhibition of either PDH or citrate synthase. In this context,

future works should evaluate the therapeutic potential of citrate synthase inhibitors (e.g. Suramin[595,596]) which could increase carnosine content in cancer cell populations expressing both GAD1 and carnosine synthase. On this subject, such inhibitors would have little effect on the PC3M/PC3S models as they displayed low expression of such enzymes. Additionally, the analysis here performed highlights that simulating gene KO in the framework of cancer-specific GSMM can go beyond identifying essential and synthetic lethal genes necessary for biomass production and also identify therapeutic interventions aimed at inducing the synthesis of a metabolite with antiproliferative or pro-apoptotic properties.

Notably, both in the prostate and colorectal cancer models we also determined that variations in metastatic and tumorigenic potential could occur largely independent of metabolic reprogramming. For instance, inducing a complete EMT in PC3M was shown to significantly reduce their proliferation and tumorigenic potential[157] and indeed we saw that many gene signatures and genes putatively associated with CSC were significantly downregulated. However, through r^2MTA we determined that such process was not associated with a significant metabolic shift towards a PC3S-like metabolic phenotype. Hence, we posit that the metabolic reprogramming supporting the PC3M-like CSC metabolic phenotype is mostly independent of EMT status. However, we note that SNAI1 overexpression in PC3M resulted in a significant downregulation of glutathione metabolism, including the Xc- system, hence their reduced tumorigenic potential could be partially attributed to a reduced capacity to cope with oxidative stress[237].

Similarly, there were only subtle differences between the metabolic phenotype of SW620 and LiM2 even though the latter were derived through *in vivo* selection of the SW620 clones with more metastatic potential. We theorize that the metabolic phenotype of SW620 is already largely optimized for metastasis and that the increased metastatic potential in LiM2 is largely attributed to non-metabolic changes in line with the previous findings[155]. Indeed, this similarity can be exploited therapeutically as both populations were shown to share the same metabolic dependency on cystine uptake and folate metabolism and hence could be targeted by the same drugs.

Overall, the work presented here provides a greater understanding of the metabolic reprogramming supporting the metastatic and CSC phenotypes in colorectal and prostate cancer. In this regard, the metabolic vulnerabilities emerging from such metabolic reprogramming and identified as part of this Ph.D. pave the way for developing targeted therapies against such cancer types. Future works should explore the effectiveness of the

identified targets in animal models and in other cancer types with the ultimate goal of identifying biomarkers that can be used to select the patients that would respond favorably to such therapies. Furthermore, the Systems Biology tools and techniques developed here can potentially be used to build patient- and disease-specific metabolic networks and determine the best therapeutic strategy for each individual patient, hence contributing to ushering the age of personalized medicine.

6. Conclusions

6. Conclusions

1) HepatoDyn is a highly detailed kinetic model capable of simulating hepatocyte central carbon metabolism with an unprecedented level of detail.
2) Exposure to high fructose concentrations can severely impair the metabolism of hepatocytes by causing a depletion of ATP, a phenomenon that can be studied with HepatoDyn.
3) ^{13}C MFA can be easily run in the cloud and as part of multi-step workflows with Iso2Flux, a novel ^{13}C MFA software implemented into the PhenoMeNal e-infrastructure.
4) ^{13}C resolved metabolomics and transcriptomics data can be seamlessly integrated through p^{13}CMFA to trace metabolic fluxes across metabolic networks.
5) r^{2}MTA is a new algorithm for robustly identifying the metabolic pathways coordinately upregulated between two conditions and identifying the optimal therapeutic interventions to drive the transition towards the biomedically desirable condition.
6) Expression of putative molecular drivers of prostate CSCs (e.g., MYC, VEGFC, REG4, ABCA3, and TSPAN8) is associated with a partial EMT phenotype in the PC3M cellular model.
7) The prostate PC3M CSC-like phenotype is also supported by a metabolic gene expression program that is largely uncoupled from partial-EMT and which can be targeted by antifolates.
8) Multiple layers of experimental data (e.g., rates of uptake and secretions of metabolites, respiration parameters, ^{13}C resolved metabolomics, targeted metabolomics, transcriptomics, and gene dependencies) can be integrated to build condition-specific GSMMs using a newly developed workflow.
9) The colorectal cancer metastatic cell lines SW620 and LiM2 can be selectively killed by targeting either cystine uptake or folate metabolism.
10) Inducing the synthesis of the antiproliferative peptide carnosine by inhibiting citrate synthase is a putative therapeutic strategy against cancer cell populations expressing GAD1 and carnosine synthase.

7. References

7. References

1. Van Esch, J. H., Klajn, R. & Otto, S. Chemical systems out of equilibrium. *Chem. Soc. Rev.* **46**, 5474–5475 (2017).

2. Zamboni, N., Saghatelian, A. & Patti, G. J. Defining the metabolome: size, flux, and regulation. *Mol. Cell* **58**, 699–706 (2015).

3. Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry*. (2007).

4. Fell, D. *Understanding the control of metabolism*. *Biochemical Education* (Portland Pr, 1997). doi:10.1016/S0307-4412(97)87557-7

5. Cornish-Bowden, A. *Fundamentals of Enzyme Kinetics*. (Portland Press Ltd., 1995).

6. Cornish-Bowden, A. *Metabolic Control Analysis in Theory and Practice*. *Advances in Molecular and Cell Biology* **11**, (1995).

7. Martín-Bernabé, A. *et al.* The importance of post-translational modifications in systems biology approaches to identify therapeutic targets in cancer metabolism. *Curr. Opin. Syst. Biol.* **3**, 161–169 (2017).

8. Alberts, B. *et al. Molecular Biology of the cell*. (Garland Science, 2008).

9. Frayn, K. N. *Metabolic Regulation: A Human Perspective*. (2010).

10. Tarrado-Castellarnau, M., de Atauri, P. & Cascante, M. Oncogenic regulation of tumor metabolic reprogramming. *Oncotarget* **7**, 62726–62753 (2016).

11. Abdel-Haleem, A. M. *et al.* The Emerging Facets of Non-Cancerous Warburg Effect. *Front. Endocrinol. (Lausanne).* **8**, 1–7 (2017).

12. Klipp, E., Liebermeister, W. & Wierling, C. Inferring Dynamic Properties of Biochemical Reaction Networks from Structural Knowledge. *Genome Informatics* **15**, 125–137 (2004).

13. Kacser, H. & Acerenza, L. A universal method for achieving increases in metabolite production. *Eur. J. Biochem.* **216**, 361–7 (1993).

14. Pons-Tostivint, E., Thibault, B. & Guillermet-Guibert, J. Targeting PI3K Signaling in Combination Cancer Therapy. *Trends in cancer* **3**, 454–469 (2017).

15. Petersen, M. C. & Shulman, G. I. Mechanisms of Insulin Action and Insulin Resistance. *Physiol. Rev.* **98**, 2133–2223 (2018).

16. Huang, J. & Manning, B. D. A complex interplay between Akt, TSC2 and the two mTOR complexes. *Biochem. Soc. Trans.* **37**, 217–22 (2009).

17. Yao, Y., Jones, E. & Inoki, K. Lysosomal Regulation of mTORC1 by Amino Acids in Mammalian Cells. *Biomolecules* **7**, 1–18 (2017).

18. Gingras, A. C. *et al.* Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev.* **13**, 1422–37 (1999).

19. Holz, M. K., Ballif, B. A., Gygi, S. P. & Blenis, J. mTOR and S6K1 mediate assembly of the translation preinitiation complex through dynamic protein interchange and ordered phosphorylation events. *Cell* **123**, 569–80 (2005).

20. Iadevaia, V., Liu, R. & Proud, C. G. MTORC1 signaling controls multiple steps in ribosome biogenesis. *Semin. Cell Dev. Biol.* **36**, 113–120 (2014).

21. Shahbazian, D. *et al.* Control of Cell Survival and Proliferation by Mammalian Eukaryotic Initiation Factor 4B. *Mol. Cell. Biol.* **30**, 1478–1485 (2010).

22. Csibi, A. *et al.* The mTORC1/S6K1 pathway regulates glutamine metabolism through the eIF4B-dependent control of c-Myc translation. *Curr. Biol.* **24**, 2274–80 (2014).

23. Nandagopal, N. & Roux, P. P. Regulation of global and specific mRNA translation by the mTOR signaling pathway. *Translation* **3**, e983402 (2015).

24. Düvel, K. *et al.* Activation of a metabolic gene regulatory network downstream of mTOR complex 1. *Mol. Cell* **39**, 171–183 (2010).

25. Tzivion, G., Dobson, M. & Ramakrishnan, G. FoxO transcription factors; Regulation by AKT and 14-3-3 proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **1813**, 1938–1945 (2011).

26. Flügel, D., Görlach, A., Michiels, C. & Kietzmann, T. Glycogen synthase kinase 3 phosphorylates hypoxia-inducible factor 1alpha and mediates its destabilization in a VHL-independent manner. *Mol. Cell. Biol.* **27**, 3253–65 (2007).

27. Cao, H. *et al.* Critical role of AKT protein in myeloma-induced osteoclast formation and osteolysis. *J. Biol. Chem.* **288**, 30399–30410 (2013).

28. Lewerenz, J. *et al.* Phosphoinositide 3-kinases upregulate system xc(-) via eukaryotic initiation factor 2α and activating transcription factor 4 - A pathway active in glioblastomas and epilepsy. *Antioxid. Redox Signal.* **20**, 2907–22 (2014).

29. Eid, W. *et al.* mTORC1 activates SREBP-2 by suppressing cholesterol trafficking to lysosomes in mammalian cells. *Proc. Natl. Acad. Sci.* **114**, 7999–8004 (2017).

30. Park, Y., Reyna-Neyra, A., Philippe, L. & Thoreen, C. C. mTORC1 Balances Cellular Amino Acid Supply with Demand for Protein Synthesis through Post-transcriptional Control of ATF4. *Cell Rep.* **19**, 1083–1090 (2017).

31. Gwinn, D. M. *et al.* Oncogenic KRAS Regulates Amino Acid Homeostasis and Asparagine Biosynthesis via ATF4 and Alters Sensitivity to L-Asparaginase. *Cancer Cell* **33**, 91-107.e6 (2018).

32. Cunningham, J. T. *et al.* mTOR controls mitochondrial oxidative function through a YY1-PGC-1alpha transcriptional complex. *Nature* **450**, 736–40 (2007).

33. Morita, M. *et al.* mTORC1 Controls Mitochondrial Activity and Biogenesis through 4E-BP-Dependent Translational Regulation. *Cell Metab.* **18**, 698–711 (2013).

34. Takenaka, N., Araki, N. & Satoh, T. Involvement of the protein kinase Akt2 in insulin-stimulated Rac1 activation leading to glucose uptake in mouse skeletal muscle. *PLoS One* **14**, e0212219 (2019).

35. MacAulay, K. *et al.* Glycogen Synthase Kinase 3α-Specific Regulation of Murine Hepatic Glycogen Metabolism. *Cell Metab.* **6**, 329–337 (2007).

36. Ben-Sahra, I., Howell, J. J., Asara, J. M. & Manning, B. D. Stimulation of de novo pyrimidine synthesis by growth signaling through mTOR and S6K1. *Science (80-.).* **339**, 1323–8 (2013).

37. Lien, E. C., Lyssiotis, C. A. & Cantley, L. C. Metabolic Reprogramming by the PI3K-Akt-mTOR Pathway in Cancer. in *Metabolism in Cancer* (eds. Cramer, T. & Schmitt, C.) **207**, 39–72 (Springer International Publishing, 2016).

38. Liu, P. *et al.* PtdIns(3,4,5)P3-Dependent Activation of the mTORC2 Kinase Complex. *Cancer Discov.* **5**, 1194–209 (2015).

39. Humphrey, S. J. *et al.* Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2. *Cell Metab.* **17**, 1009–20 (2013).

40. Yang, G., Murashige, D. S., Humphrey, S. J. & James, D. E. A Positive Feedback Loop between Akt

and mTORC2 via SIN1 Phosphorylation. *Cell Rep.* **12**, 937–943 (2015).

41. Kliegman, J. I. *et al.* Chemical genetics of rapamycin-insensitive TORC2 in S. cerevisiae. *Cell Rep.* **5**, 1725–1736 (2013).

42. Saxton, R. A. & Sabatini, D. M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **168**, 960–976 (2017).

43. Burotto, M., Chiou, V. L., Lee, J.-M. & Kohn, E. C. The MAPK pathway across different malignancies: A new perspective. *Cancer* **120**, 3446–3456 (2014).

44. Dorard, C., Vucak, G. & Baccarini, M. Deciphering the RAS/ERK pathway in vivo . *Biochem. Soc. Trans.* **45**, 27–36 (2017).

45. Suire, S., Hawkins, P. & Stephens, L. Activation of phosphoinositide 3-kinase gamma by Ras. *Curr. Biol.* **12**, 1068–75 (2002).

46. Ciuffreda, L. *et al.* The mitogen-activated protein kinase (MAPK) cascade controls phosphatase and tensin homolog (PTEN) expression through multiple mechanisms. *J. Mol. Med.* **90**, 667–679 (2012).

47. Ma, L., Chen, Z., Erdjument-Bromage, H., Tempst, P. & Pandolfi, P. P. Phosphorylation and functional inactivation of TSC2 by Erk: Implications for tuberous sclerosis and cancer pathogenesis. *Cell* **121**, 179–193 (2005).

48. Zimmermann, S., Moelling, K., Sarraf, P. & Spiegelman, B. Phosphorylation and Regulation of Raf by Akt (Protein Kinase B). *Science (80-.).* **286**, 1741–1744 (1999).

49. Yu, C. F., Liu, Z.-X. & Cantley, L. G. ERK negatively regulates the epidermal growth factor-mediated interaction of Gab1 and the phosphatidylinositol 3-kinase. *J. Biol. Chem.* **277**, 19382–8 (2002).

50. Mendoza, M. C., Er, E. E. & Blenis, J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem. Sci.* **36**, 320–8 (2011).

51. Farrell, A. S. & Sears, R. C. MYC degradation. *Cold Spring Harb. Perspect. Med.* **4**, 1–15 (2014).

52. Sodhi, A., Montaner, S., Miyazaki, H. & Gutkind, J. S. MAPK and akt Act cooperatively but independently on hypoxia inducible factor-1α in rasV12 upregulation of VEGF. *Biochem. Biophys. Res. Commun.* **287**, 292–300 (2001).

53. Kwon, S. J., Song, J. J. & Lee, Y. J. Signal Pathway of Hypoxia-Inducible Factor-1 Phosphorylation and its Interaction with von Hippel-Lindau Tumor Suppressor Protein During Ischemia in MiaPaCa-2 Pancreatic Cancer Cells. *Clin. Cancer Res.* **11**, 7607–7613 (2005).

54. Mylonis, I., Chachami, G., Paraskeva, E. & Simos, G. Atypical CRM1-dependent nuclear export signal mediates regulation of hypoxia-inducible factor-1alpha by MAPK. *J. Biol. Chem.* **283**, 27620–7 (2008).

55. Lawan, A. & Bennett, A. M. Mitogen-Activated Protein Kinase Regulation in Hepatic Metabolism. *Trends Endocrinol. Metab.* **28**, 868–878 (2017).

56. William F., S. G protein regulation of adenylate cyclase. *Trends Pharmacol. Sci.* **20**, 66–73 (1999).

57. Jeon, Y. H. *et al.* Phosphodiesterase: overview of protein structures, potential therapeutic applications and recent progress in drug development. *Cell. Mol. Life Sci.* **62**, 1198–1220 (2005).

58. Shabb, J. B. Physiological substrates of cAMP-dependent protein kinase. *Chem. Rev.* **101**, 2381–411 (2001).

59. Titanji, V. P., Zetterqvist, O. & Engstroöm, L. Regulation in vitro of rat liver pyruvate kinase by phosphorylation-dephosphorylation reactions, catalyzed by cyclic-AMP dependent protein kinases and a histone phosphatase. *Biochim. Biophys. Acta* **422**, 98–108 (1976).

60. Pilkis, S. J., El-Maghrabi, M. R. & Claus, T. H. Hormonal regulation of hepatic gluconeogenesis and

glycolysis. *Annu. Rev. Biochem.* **57**, 755–83 (1988).

61. Brushia, R. J. & Walsh, D. A. Phosphorylase kinase: the complexity of its regulation is reflected in the complexity of its structure. *Front. Biosci.* **4**, D618-41 (1999).

62. Okar, D. A. *et al.* PFK-2 / FBPase-2 : maker and breaker of the essential biofactor fructose-2 , 6-bisphosphate. **26**, 30–35 (2001).

63. König, M., Bulik, S. & Holzhütter, H. G. Quantifying the contribution of the liver to glucose homeostasis: A detailed kinetic model of human hepatic glucose metabolism. *PLoS Comput. Biol.* **8**, (2012).

64. Shimano, H. & Sato, R. SREBP-regulated lipid metabolism: Convergent physiology-divergent pathophysiology. *Nat. Rev. Endocrinol.* **13**, 710–730 (2017).

65. Yellaturu, C. R. *et al.* Posttranslational processing of SREBP-1 in rat hepatocytes is regulated by insulin and cAMP. *Biochem. Biophys. Res. Commun.* **332**, 174–180 (2005).

66. Iizuka, K. The Role of Carbohydrate Response Element Binding Protein in Intestinal and Hepatic Fructose Metabolism. *Nutrients* **9**, 181 (2017).

67. Wu, Y. *et al.* Novel mechanism of FOXO1 phosphorylation in glucagon signaling in control of glucose homeostasis. *Diabetes* **67**, 2167–2182 (2018).

68. Takahashi, M., Li, Y., Dillon, T. J. & Stork, P. J. S. Phosphorylation of Rap1 by cAMP-dependent protein kinase (PKA) creates a binding site for KSR to sustain ERK activation by cAMP. *J. Biol. Chem.* **292**, 1449–1461 (2017).

69. Gold, M. G., Gonen, T. & Scott, J. D. Local cAMP signaling in disease at a glance. *J. Cell Sci.* **126**, 4537–4543 (2013).

70. Horton, J. D., Goldstein, J. L. & Brown, M. S. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J. Clin. Invest.* **109**, 1125–31 (2002).

71. Nickels, J. T. New links between lipid accumulation and cancer progression. *J. Biol. Chem.* **293**, 6635–6636 (2018).

72. Brown, M. S. & Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331–40 (1997).

73. Xing, Y. qi *et al.* The regulation of FOXO1 and its role in disease progression. *Life Sci.* **193**, 124–131 (2018).

74. Farhan, M. *et al.* FOXO signaling pathways as therapeutic targets in cancer. *Int. J. Biol. Sci.* **13**, 815–827 (2017).

75. Zou, J. *et al.* Metformin inhibits estrogen-dependent endometrial cancer cell growth by activating the AMPK-FOXO1 signal pathway. *Cancer Sci.* **107**, 1806–1817 (2016).

76. Brunet, A. *et al.* Stress-Dependent Regulation of FOXO Transcription Factors by the SIRT1 Deacetylase. *Science (80-.).* **303**, 2011–2015 (2004).

77. Yang, Y. *et al.* Acetylation of FoxO1 activates Bim expression to induce apoptosis in response to histone deacetylase inhibitor depsipeptide treatment. *Neoplasia* **11**, 313–24 (2009).

78. Wortel, I. M. N., van der Meer, L. T., Kilberg, M. S. & van Leeuwen, F. N. Surviving Stress: Modulation of ATF4-Mediated Stress Responses in Normal and Malignant Cells. *Trends Endocrinol. Metab.* **28**, 794–806 (2017).

79. Ye, J. *et al.* Pyruvate kinase M2 promotes de novo serine synthesis to sustain mTORC1 activity and cell proliferation. *Proc. Natl. Acad. Sci.* **109**, 6904–6909 (2012).

80. Ben-Sahra, I., Hoxhaj, G., Ricoult, S. J. H., Asara, J. M. & Manning, B. D. mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle. *Science (80-.).* **351**, 728–

733 (2016).

81. Khan, N. A. *et al.* mTORC1 Regulates Mitochondrial Integrated Stress Response and Mitochondrial Myopathy Progression. *Cell Metab.* **26**, 419-428.e5 (2017).

82. B'Chir, W. *et al.* The eIF2α/ATF4 pathway is essential for stress-induced autophagy gene expression. *Nucleic Acids Res.* **41**, 7683–7699 (2013).

83. Lin, C. Y. *et al.* Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **151**, 56–67 (2012).

84. Kress, T. R., Sabò, A. & Amati, B. MYC: connecting selective transcriptional control to global RNA production. *Nat. Rev. Cancer* **15**, 593–607 (2015).

85. Dang, C. V. *et al.* The c-Myc target gene network. *Semin. Cancer Biol.* **16**, 253–264 (2006).

86. Goetzman, E. S. & Prochownik, E. V. The role for myc in coordinating glycolysis, oxidative phosphorylation, glutaminolysis, and fatty acid metabolism in normal and neoplastic tissues. *Front. Endocrinol. (Lausanne).* **9**, (2018).

87. Benita, Y. *et al.* An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res.* **37**, 4587–4602 (2009).

88. Luo, W. *et al.* Pyruvate kinase M2 is a PHD3-stimulated coactivator for hypoxia-inducible factor 1. *Cell* **145**, 732–44 (2011).

89. Masson, N., Willam, C., Maxwell, P. H., Pugh, C. W. & Ratcliffe, P. J. Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation. *EMBO J.* **20**, 5197–206 (2001).

90. Koivunen, P. *et al.* Inhibition of Hypoxia-inducible Factor (HIF) Hydroxylases by Citric Acid Cycle Intermediates. *J. Biol. Chem.* **282**, 4524–4532 (2007).

91. Zhang, J. *et al.* Accumulation of Succinate in Cardiac Ischemia Primarily Occurs via Canonical Krebs Cycle Activity. *Cell Rep.* **23**, 2617 (2018).

92. Dodd, K. M., Yang, J., Shen, M. H., Sampson, J. R. & Tee, A. R. mTORC1 drives HIF-1α and VEGF-A signalling via multiple mechanisms involving 4E-BP1, S6K1 and STAT3. *Oncogene* **34**, 2239–50 (2015).

93. Kietzmann, T., Mennerich, D. & Dimova, E. Y. Hypoxia-Inducible Factors (HIFs) and Phosphorylation: Impact on Stability, Localization, and Transactivity. *Front. Cell Dev. Biol.* **4**, 1–14 (2016).

94. Flügel, D., Görlach, A. & Kietzmann, T. GSK-3β regulates cell growth, migration, and angiogenesis via Fbw7 and USP28-dependent degradation of HIF-1α. *Blood* **119**, 1292–301 (2012).

95. Iizuka, K. The transcription factor carbohydrate-response element-binding protein (ChREBP): A possible link between metabolic disease and cancer. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1863**, 474–485 (2017).

96. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

97. Joerger, A. C. & Fersht, A. R. The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. *Annu. Rev. Biochem.* **85**, 375–404 (2016).

98. Chen, J. The Cell-Cycle Arrest and Apoptotic Functions of p53 in Tumor Initiation and Progression. *Cold Spring Harb. Perspect. Med.* **6**, a026104 (2016).

99. Jiang, P. *et al.* p53 regulates biosynthesis through direct inactivation of glucose-6-phosphate dehydrogenase. *Nat. Cell Biol.* **13**, 310–316 (2011).

100. Fernandez-Fernandez, M. R. & Sot, B. The relevance of protein-protein interactions for p53

function: The CPE contribution. *Protein Eng. Des. Sel.* **24**, 41–51 (2011).

101. Kruse, J.-P. & Gu, W. Modes of p53 regulation. *Cell* **137**, 609–22 (2009).

102. Hasty, P., Sharp, Z. D., Curiel, T. J. & Campisi, J. mTORC1 and p53: clash of the gods? *Cell Cycle* **12**, 20–5 (2013).

103. Astle, M. V. *et al.* AKT induces senescence in human cells via mTORC1 and p53 in the absence of DNA damage: Implications for targeting mTOR during malignancy. *Oncogene* **31**, 1949–1962 (2012).

104. Schwartzenberg-Bar-Yoseph, F., Armoni, M. & Karnieli, E. The Tumor Suppressor p53 Down-Regulates Glucose Transporters ***GLUT1*** and ***GLUT4*** Gene Expression. *Cancer Res.* **64**, 2627–2633 (2004).

105. Vandercammen, A. & Van Schaftingen, E. Competitive inhibition of liver glucokinase by its regulatory protein. *Eur. J. Biochem.* **200**, 545–51 (1991).

106. Thorens, B. & Mueckler, M. Glucose transporters in the 21st Century. *Am. J. Physiol. Endocrinol. Metab.* **298**, E141-5 (2010).

107. Kyeong, H.-H., Seo, M.-H., Choi, J. M., Kim, E. & Kim, H.-S. Molecular basis for the role of glucokinase regulatory protein as the allosteric switch for glucokinase. *Proc. Natl. Acad. Sci.* **110**, 10171–10176 (2013).

108. Anderka, O. *et al.* Biophysical Characterization of the Interaction between Hepatic Glucokinase and Its Regulatory Protein. *J. Biol. Chem.* **283**, 31333–31340 (2008).

109. Han, H.-S., Kang, G., Kim, J. S., Choi, B. H. & Koo, S.-H. Regulation of glucose metabolism from a liver-centric perspective. *Exp. Mol. Med.* **48**, e218 (2016).

110. Aggen, J. B., Nairn, A. C. & Chamberlin, R. Regulation of protein phosphatase-1. *Chem. Biol.* **7**, R13-23 (2000).

111. Ragolia, L. & Begum, N. Protein phosphatase-1 and insulin action. *Mol. Cell. Biochem.* **182**, 49–58 (1998).

112. Pautsch, A. *et al.* Molecular recognition of the protein phosphatase 1 glycogen targeting subunit by glycogen phosphorylase. *J. Biol. Chem.* **283**, 8913–8 (2008).

113. Alemany, S. & Cohen, P. Phosphorylase a is an allosteric inhibitor of the glycogen and microsomal forms of rat hepatic protein phosphatase-1. *FEBS Lett.* **198**, 194–202 (1986).

114. Faustova, I. & Järv, J. Interaction of Non-Phosphorylated Liver Pyruvate Kinase with Fructose 1,6-Bisphosphate and Peptides that Mimic the Phosphorylatable N-terminus of the Enzyme. *Protein Pept. Lett.* **20**, 1200–1203 (2013).

115. Llorente, P., Marco, R. & Sols, A. Regulation of Liver Pyruvate Kinase and the Phosphoenolpyruvate Crossroads. *Eur. J. Biochem.* **13**, 45–54 (1970).

116. Bremer, J. Pyruvate dehydrogenase, substrate specificity and product inhibition. *Eur. J. Biochem.* **8**, 535–40 (1969).

117. Jeoung, N. H. Pyruvate dehydrogenase kinases: Therapeutic targets for diabetes and cancers. *Diabetes Metab. J.* **39**, 188–197 (2015).

118. Mayes, P. A. Intermediary metabolism of fructose. *Am. J. Clin. Nutr.* **58**, 754S-765S (1993).

119. Asipu, A., Hayward, B. E., O'Reilly, J. & Bonthron, D. T. Properties of normal and mutant recombinant human ketohexokinases and implications for the pathogenesis of essential fructosuria. *Diabetes* **52**, 2426–32 (2003).

120. Sestoft, L. & Fleron, P. Determination of the kinetic constants of fructose transport and phosphorylation in the perfused rat liver. *Biochim. Biophys. Acta* **345**, 27–38 (1974).

121. Tappy, L. & Lê, K.-A. Metabolic effects of fructose and the worldwide increase in obesity. *Physiol. Rev.* **90**, 23–46 (2010).

122. Papandreou, D. & Andreou, E. Role of diet on non-alcoholic fatty liver disease: An updated narrative review. *World J. Hepatol.* **7**, 575–82 (2015).

123. Baena, M. *et al.* Fructose, but not glucose, impairs insulin signaling in the three major insulin-sensitive tissues. *Sci. Rep.* **6**, 26149 (2016).

124. Pereira, R. M. *et al.* Fructose Consumption in the Development of Obesity and the Effects of Different Protocols of Physical Exercise on the Hepatic Metabolism. *Nutrients* **9**, (2017).

125. Chowdhury, F. *et al.* Regulation of pyruvate dehydrogenase kinase isoform 4 (PDK4) gene expression by glucocorticoids and insulin. *Mol. Cell. Endocrinol.* **315**, 159–167 (2009).

126. Lee, S. & Dong, H. H. FoxO integration of insulin signaling with glucose and lipid metabolism. *J. Endocrinol.* **233**, R67–R79 (2017).

127. Oh, K.-J., Han, H.-S., Kim, M.-J. & Koo, S.-H. CREB and FoxO1: two transcription factors for the regulation of hepatic gluconeogenesis. *BMB Rep.* **46**, 567–74 (2013).

128. Wang, Y., Viscarra, J., Kim, S. J. & Sul, H. S. Transcriptional regulation of hepatic lipogenesis. *Nat. Rev. Mol. Cell Biol.* **16**, 678–689 (2015).

129. Henriksson, E. *et al.* SIK2 regulates CRTCs, HDAC4 and glucose uptake in adipocytes. *J. Cell Sci.* **128**, 472–486 (2014).

130. Song, Y. *et al.* Structural Insights into the CRTC2–CREB Complex Assembly on CRE. *J. Mol. Biol.* **430**, 1926–1939 (2018).

131. Faguet, G. B. A brief history of cancer: age-old milestones underlying our current knowledge database. *Int. J. cancer* **136**, 2022–36 (2015).

132. Pecorino, L. *Molecular biology of cancer : mechanisms, targets, and therapeutics*. (Oxford University Press, 2012).

133. Bray, F. & Møller, B. Predicting the future burden of cancer. *Nat. Rev. Cancer* **6**, 63–74 (2006).

134. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).

135. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).

136. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

137. Thomson, T. M. M. *et al.* Metabolic Plasticity and Epithelial-Mesenchymal Transition. *J. Clin. Med.* **8**, 967 (2019).

138. Saitoh, M. Involvement of partial EMT in cancer progression. *J. Biochem.* **164**, 257–264 (2018).

139. Bol, D. K., Kiguchi, K., Gimenez-Conti, I., Rupp, T. & DiGiovanni, J. Overexpression of insulin-like growth factor-1 induces hyperplasia, dermal abnormalities, and spontaneous tumor formation in transgenic mice. *Oncogene* **14**, 1725–1734 (1997).

140. Weber, M. M. *et al.* Overexpression of the insulin-like growth factor I receptor in human colon carcinomas. *Cancer* **95**, 2086–2095 (2002).

141. Ligresti, G. *et al.* PIK3CA mutations in human solid tumours. *Cell cycle* **8**, 1352–1358 (2009).

142. Álvarez-Garcia, V., Tawil, Y., Wise, H. M. & Leslie, N. R. Mechanisms of PTEN loss in cancer: It's all about diversity. *Semin. Cancer Biol.* (2019). doi:10.1016/j.semcancer.2019.02.001

143. Yi, K. H. & Lauring, J. Recurrent *AKT* mutations in human cancers: functional consequences and effects on drug sensitivity. *Oncotarget* **7**, (2016).

144. Grabiner, B. C. *et al.* A Diverse Array of Cancer-Associated MTOR Mutations Are Hyperactivating and Can Predict Rapamycin Sensitivity. *Cancer Discov.* **4**, 554–563 (2014).

145. Kalkat, M. *et al.* MYC deregulation in primary human cancers. *Genes (Basel).* **8**, 2–30 (2017).

146. Wang, X. & Sun, Q. *TP53* mutations, expression and interaction networks in human cancers. *Oncotarget* **8**, 624–643 (2016).

147. Lee, J.-K., Choi, Y.-L., Kwon, M. & Park, P. J. Mechanisms and Consequences of Cancer Genome Instability: Lessons from Genome Sequencing Studies. *Annu. Rev. Pathol. Mech. Dis.* **11**, 283–312 (2016).

148. Mackay, H. L. *et al.* Genomic instability in mutant p53 cancer cells upon entotic engulfment. *Nat. Commun.* **9**, 3070 (2018).

149. Dowty, J. G. *et al.* Cancer risks for MLH1 and MSH2 mutation carriers. *Hum. Mutat.* **34**, 490–7 (2013).

150. Bennett, R. L. & Licht, J. D. Targeting Epigenetics in Cancer. *Annu. Rev. Pharmacol. Toxicol.* **58**, 187–207 (2017).

151. Yegnasubramanian, S., De Marzo, A. M. & Nelson, W. G. Prostate Cancer Epigenetics: From Basic Mechanisms to Clinical Implications. *Cold Spring Harb. Perspect. Med.* **9**, a030445 (2019).

152. Turajlic, S., Sottoriva, A., Graham, T. & Swanton, C. Resolving genetic heterogeneity in cancer. *Nat. Rev. Genet.* (2019). doi:10.1038/s41576-019-0114-6

153. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–13 (2012).

154. Leibovitz, A. *et al.* Classification of human colorectal adenocarcinoma cell lines. *Cancer Res.* **36**, 4562–9 (1976).

155. Urosevic, J. *et al.* Colon cancer cells colonize the lung from established liver metastases through p38 MAPK signalling and PTHLH. *Nat. Cell Biol.* **16**, 685–94 (2014).

156. Kaighn, M. E., Narayan, K. S., Ohnuki, Y., Lechner, J. F. & Jones, L. W. Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Invest. Urol.* **17**, 16–23 (1979).

157. Celià-Terrassa, T. *et al.* Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *J. Clin. Invest.* **122**, 1849–68 (2012).

158. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).

159. Kuşoğlu, A. & Biray Avcı, Ç. Cancer stem cells: A brief review of the current status. *Gene* **681**, 80–85 (2019).

160. Armstrong, C. M. & Gao, A. C. Drug resistance in castration resistant prostate cancer: resistance mechanisms and emerging treatment strategies. *Am. J. Clin. Exp. Urol.* **3**, 64–76 (2015).

161. Deng, Q. & Tang, D. G. Androgen receptor and prostate cancer stem cells: biological mechanisms and clinical implications. *Endocr. Relat. Cancer* **22**, T209-20 (2015).

162. Safa, A. R. Resistance to Cell Death and Its Modulation in Cancer Stem Cells. *Crit. Rev. Oncog.* **21**, 203–219 (2016).

163. Prieto-Vila, M., Takahashi, R.-U., Usuba, W., Kohama, I. & Ochiya, T. Drug Resistance Driven by Cancer Stem Cells and Their Niche. *Int. J. Mol. Sci.* **18**, (2017).

164. Luo, M. & Wicha, M. S. Targeting Cancer Stem Cell Redox Metabolism to Enhance Therapy Responses. *Semin. Radiat. Oncol.* **29**, 42–54 (2019).

165. Yoo, Y. D. & Kwon, Y. T. Molecular mechanisms controlling asymmetric and symmetric self-renewal of cancer stem cells. *J. Anal. Sci. Technol.* **6**, (2015).

166. Marjanovic, N. D., Weinberg, R. A. & Chaffer, C. L. Cell plasticity and heterogeneity in cancer. *Clin. Chem.* **59**, 168–179 (2013).

167. Poli, V., Fagnocchi, L. & Zippo, A. Tumorigenic cell reprogramming and cancer plasticity: Interplay between signaling, microenvironment, and epigenetics. *Stem Cells Int.* **2018**, (2018).

168. Granit, R. Z. *et al.* Regulation of Cellular Heterogeneity and Rates of Symmetric and Asymmetric Divisions in Triple-Negative Breast Cancer. *Cell Rep.* **24**, 3237–3250 (2018).

169. Xia, P. & Xu, X.-Y. PI3K/Akt/mTOR signaling pathway in cancer stem cells: from basic research to clinical application. *Am. J. Cancer Res.* **5**, 1602–9 (2015).

170. Roy, S. *et al.* Role of p38 MAPK in disease relapse and therapeutic resistance by maintenance of cancer stem cells in head and neck squamous cell carcinoma. *J. Oral Pathol. Med.* **47**, 492–501 (2018).

171. Xu, M. *et al.* Role of p38γ MAPK in regulation of EMT and cancer stem cells. *Biochim. Biophys. acta. Mol. basis Dis.* **1864**, 3605–3617 (2018).

172. Kolev, V. N. *et al.* Inhibition of FAK kinase activity preferentially targets cancer stem cells. *Oncotarget* **8**, 51733–51747 (2017).

173. Venkatesh, V. *et al.* Targeting Notch signalling pathway of cancer stem cells. *Stem cell Investig.* **5**, 5 (2018).

174. Sari, I. N. *et al.* Hedgehog Signaling in Cancer: A Prospective Therapeutic Target for Eradicating Cancer Stem Cells. *Cells* **7**, (2018).

175. Park, J. H., Shin, J. E. & Park, H. W. The Role of Hippo Pathway in Cancer Stem Cell Biology. *Mol. Cells* **41**, 83–92 (2018).

176. Yang, A. *et al.* MYC Inhibition Depletes Cancer Stem-like Cells in Triple-Negative Breast Cancer. *Cancer Res.* **77**, 6641–6650 (2017).

177. Zhang, H.-L., Wang, P., Lu, M.-Z., Zhang, S.-D. & Zheng, L. c-Myc maintains the self-renewal and chemoresistance properties of colon cancer stem cells. *Oncol. Lett.* **17**, 4487–4493 (2019).

178. Li, Y., Xian, M., Yang, B., Ying, M. & He, Q. Inhibition of KLF4 by Statins Reverses Adriamycin-Induced Metastasis and Cancer Stemness in Osteosarcoma Cells. *Stem cell reports* **8**, 1617–1629 (2017).

179. Yu, F. *et al.* Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion. *Oncogene* **30**, 2161–72 (2011).

180. Srinivasan, M. *et al.* Downregulation of Bmi1 in breast cancer stem cells suppresses tumor growth and proliferation. *Oncotarget* **8**, 38731–38742 (2017).

181. Vora, P. *et al.* Bmi1 regulates human glioblastoma stem cells through activation of differential gene networks in CD133+ brain tumor initiating cells. *J. Neurooncol.* **143**, 417–428 (2019).

182. Cordenonsi, M. *et al.* The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. *Cell* **147**, 759–72 (2011).

183. Cannito, S. *et al.* Epithelial-mesenchymal transition: from molecular mechanisms, redox regulation to implications in human health and disease. *Antioxid. Redox Signal.* **12**, 1383–430 (2010).

184. Zeisberg, M. & Neilson, E. G. Biomarkers for epithelial-mesenchymal transitions. *J. Clin. Invest.* **119**, 1429–37 (2009).

185. Vega, S. *et al.* Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev.* **18**, 1131–1143 (2004).

186. Ocaña, O. H. *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal

transition inducer Prrx1. *Cancer Cell* **22**, 709–24 (2012).

187. Pattabiraman, D. R. *et al.* Activation of PKA leads to mesenchymal-to-epithelial transition and loss of tumor-initiating ability. *Science* **351**, aad3680 (2016).

188. Hugo, H. J. *et al.* Epithelial requirement for in vitro proliferation and xenograft growth and metastasis of MDA-MB-468 human breast cancer cells: oncogenic rather than tumor-suppressive role of E-cadherin. *Breast Cancer Res.* **19**, 86 (2017).

189. Zheng, X. *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).

190. Reichert, M. *et al.* Regulation of Epithelial Plasticity Determines Metastatic Organotropism in Pancreatic Cancer. *Dev. Cell* **45**, 696-711.e8 (2018).

191. Padmanaban, V. *et al.* E-cadherin is required for metastasis in multiple models of breast cancer. *Nature* **324**, 297–314 (2019).

192. Beerling, E. *et al.* Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep.* **14**, 2281–2288 (2016).

193. Shibue, T. & Weinberg, R. A. EMT, CSCs, and drug resistance: The mechanistic link and clinical implications. *Nat. Rev. Clin. Oncol.* **14**, 611–629 (2017).

194. Lo, H. C. & Zhang, X. H.-F. EMT in Metastasis: Finding the Right Balance. *Dev. Cell* **45**, 663–665 (2018).

195. Wang, Y., Shi, J., Chai, K., Ying, X. & Zhou, B. P. The Role of Snail in EMT and Tumorigenesis. *Curr. Cancer Drug Targets* **13**, 963–972 (2013).

196. Molina-Ortiz, P. *et al.* Characterization of the SNAG and SLUG domains of Snail2 in the repression of E-cadherin and EMT induction: modulation by serine 4 phosphorylation. *PLoS One* **7**, e36132 (2012).

197. Villarejo, A., Cortés-Cabrera, A., Molina-Ortíz, P., Portillo, F. & Cano, A. Differential role of Snail1 and Snail2 zinc fingers in E-cadherin repression and epithelial to mesenchymal transition. *J. Biol. Chem.* **289**, 930–41 (2014).

198. Aomatsu, K. *et al.* TGF-β Induces Sustained Upregulation of SNAI1 and SNAI2 through Smad and Non-Smad Pathways in a Human Corneal Epithelial Cell Line. *Investig. Opthalmology Vis. Sci.* **52**, 2437 (2011).

199. Naber, H. P. H., Drabsch, Y., Snaar-Jagalska, B. E., ten Dijke, P. & van Laar, T. Snail and Slug, key regulators of TGF-β-induced EMT, are sufficient for the induction of single-cell invasion. *Biochem. Biophys. Res. Commun.* **435**, 58–63 (2013).

200. Barberà, M. J. *et al.* Regulation of Snail transcription during epithelial to mesenchymal transition of tumor cells. *Oncogene* **23**, 7345–7354 (2004).

201. Grotegut, S., von Schweinitz, D., Christofori, G. & Lehembre, F. Hepatocyte growth factor induces cell scattering through MAPK/Egr-1-mediated upregulation of Snail. *EMBO J.* **25**, 3534–45 (2006).

202. Peiró, S. *et al.* Snail1 transcriptional repressor binds to its own promoter and controls its expression. *Nucleic Acids Res.* **34**, 2077–84 (2006).

203. Nakamura, R. *et al.* Reciprocal expression of Slug and Snail in human oral cancer cells. *PLoS One* **13**, e0199442 (2018).

204. Chen, Y. & Gridley, T. The SNAI1 and SNAI2 proteins occupy their own and each other's promoter during chondrogenesis. *Biochem. Biophys. Res. Commun.* **435**, 356–60 (2013).

205. Sundararajan, V. *et al.* SNAI1 recruits HDAC1 to suppress SNAI2 transcription during epithelial to mesenchymal transition. *Sci. Rep.* **9**, 1–9 (2019).

206. Ghaleb, A. M. & Yang, V. W. Krüppel-like factor 4 (KLF4): What we currently know. *Gene* **611**, 27–37 (2017).

207. Liu, Y.-N. *et al.* Critical and Reciprocal Regulation of KLF4 and SLUG in Transforming Growth Factor -Initiated Prostate Cancer Epithelial-Mesenchymal Transition. *Mol. Cell. Biol.* **32**, 941–953 (2012).

208. Yori, J. L. *et al.* Krüppel-like Factor 4 Inhibits Tumorigenic Progression and Metastasis in a Mouse Model of Breast Cancer. *Neoplasia* **13**, 601-IN5 (2015).

209. Li, Q. *et al.* Suppression of epithelial-mesenchymal transition in hepatocellular carcinoma cells by Krüppel-like factor 4. *Oncotarget* **7**, 29749–60 (2016).

210. Zhou, B. P. *et al.* Dual regulation of Snail by GSK-3beta-mediated phosphorylation in control of epithelial-mesenchymal transition. *Nat. Cell Biol.* **6**, 931–40 (2004).

211. Bastea, L. I., Döppler, H., Balogun, B. & Storz, P. Protein kinase D1 maintains the epithelial phenotype by inducing a DNA-bound, inactive SNAI1 transcriptional repressor complex. *PLoS One* **7**, (2012).

212. Du, C., Zhang, C., Hassan, S., Biswas, M. H. U. & Balaji, K. C. Protein kinase D1 suppresses epithelial-to-mesenchymal transition through phosphorylation of snail. *Cancer Res.* **70**, 7810–7819 (2010).

213. Zheng, H. *et al.* PKD1 phosphorylation-dependent degradation of SNAIL by SCF-FBXO11 regulates epithelial-mesenchymal transition and metastasis. *Cancer Cell* **26**, 358–373 (2014).

214. Yang, Z. *et al.* Pak1 phosphorylation of snail, a master regulator of epithelial-to-mesenchyme transition, modulates snail's subcellular localization and functions. *Cancer Res.* **65**, 3179–84 (2005).

215. Zhang, K. *et al.* Lats2 kinase potentiates Snail1 activity by promoting nuclear retention upon phosphorylation. *EMBO J.* **31**, 29–43 (2012).

216. Wu, Y. & Zhou, B. P. TNF-alpha/NF-kappaB/Snail pathway in cancer cell migration and invasion. *Br. J. Cancer* **102**, 639–44 (2010).

217. Kim, J. Y. *et al.* Functional regulation of Slug/Snail2 is dependent on GSK-3β-mediated phosphorylation. *FEBS J.* **279**, 2929–2939 (2012).

218. Liu, S. *et al.* Stabilization of Slug by NF-κB is Essential for TNF-α -Induced Migration and Epithelial-Mesenchymal Transition in Head and Neck Squamous Cell Carcinoma Cells. *Cell. Physiol. Biochem.* **47**, 567–578 (2018).

219. Zhou, Z., Ibekwe, E. & Chornenkyy, Y. Metabolic Alterations in Cancer Cells and the Emerging Role of Oncometabolites as Drivers of Neoplastic Change. *Antioxidants (Basel, Switzerland)* **7**, (2018).

220. Lu, J. The Warburg metabolism fuels tumor metastasis. *Cancer Metastasis Rev.* **38**, 157–164 (2019).

221. Havas, K. M. *et al.* Metabolic shifts in residual breast cancer drive tumor recurrence. *J. Clin. Invest.* **127**, 2091–2105 (2017).

222. Lee, J. S. *et al.* Urea Cycle Dysregulation Generates Clinically Relevant Genomic and Biochemical Signatures. *Cell* **174**, 1559-1570.e22 (2018).

223. Traverso, N. *et al.* Role of glutathione in cancer progression and chemoresistance. *Oxid. Med. Cell. Longev.* **2013**, 972913 (2013).

224. Tarrado-Castellarnau, M. *et al.* De novo MYC addiction as an adaptive response of cancer cells to CDK4/6 inhibition. *Mol. Syst. Biol.* **13**, 940 (2017).

225. Chang, C.-H. *et al.* Metabolic Competition in the Tumor Microenvironment Is a Driver of Cancer Progression. *Cell* **162**, 1229–41 (2015).

226. Huber, V. *et al.* Cancer acidity: An ultimate frontier of tumor immune escape and a novel target of immunomodulation. *Semin. Cancer Biol.* **43**, 74–89 (2017).

227. Otto, A. M. Warburg effect(s)—a biographical sketch of Otto Warburg and his impacts on tumor metabolism. *Cancer Metab.* **4**, 5 (2016).

228. Salway, J. G. *Metabolism at a Glance*. (Blackwell publishing, 2008).

229. Vyas, S., Zaganjor, E. & Haigis, M. C. Mitochondria and Cancer. *Cell* **166**, 555–566 (2016).

230. Porporato, P. E., Filigheddu, N., Pedro, J. M. B. S., Kroemer, G. & Galluzzi, L. Mitochondrial metabolism and cancer. *Cell Res.* **28**, 265–280 (2018).

231. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–33 (2009).

232. Locasale, J. W. & Cantley, L. C. Metabolic flux and the regulation of mammalian cell growth. *Cell Metab.* **14**, 443–51 (2011).

233. Vazquez, A., Liu, J., Zhou, Y. & Oltvai, Z. N. Catabolic efficiency of aerobic glycolysis: The Warburg effect revisited. *BMC Syst. Biol.* **4**, 58 (2010).

234. Schuster, S., de Figueiredo, L. F., Schroeter, A. & Kaleta, C. Combining metabolic pathway analysis with evolutionary game theory. Explaining the occurrence of low-yield pathways by an analytic optimization approach. *BioSystems* **105**, 147–153 (2011).

235. Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R. & Ruppin, E. Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect. *PLoS Comput. Biol.* **7**, 1–8 (2011).

236. Epstein, T., Gatenby, R. A. & Brown, J. S. The Warburg effect as an adaptation of cancer cells to rapid fluctuations in energy demand. *PLoS One* **12**, e0185085 (2017).

237. Kamarajugadda, S. *et al.* Manganese superoxide dismutase promotes anoikis resistance and tumor metastasis. *Cell Death Dis.* **4**, e504 (2013).

238. Li, A. E. *et al.* A role for reactive oxygen species in endothelial cell anoikis. *Circ. Res.* **85**, 304–310 (1999).

239. Rozhin, J. *et al.* Pericellular pH Affects Distribution and Secretion of Cathepsin B in Malignant Cells. *Cancer Res.* **54**, 6517–6525 (2013).

240. Rohani, N. *et al.* Acidification of Tumor at Stromal Boundaries Drives Transcriptome Alterations Associated with Aggressive Phenotypes. *Cancer Res.* **79**, 1952–1966 (2019).

241. Ancey, P., Contat, C. & Meylan, E. Glucose transporters in cancer – from tumor cells to the tumor microenvironment. *FEBS J.* **285**, 2926–2943 (2018).

242. Menssen, A. & Hermeking, H. Characterization of the c-MYC-regulated transcriptome by SAGE: Identification and analysis of c-MYC target genes. *Proc. Natl. Acad. Sci.* **99**, 6274–6279 (2002).

243. O'Connell, B. C. *et al.* A large scale genetic analysis of c-Myc-regulated gene expression patterns. *J. Biol. Chem.* **278**, 12563–73 (2003).

244. Kim, J. *et al.* Evaluation of myc E-box phylogenetic footprints in glycolytic genes by chromatin immunoprecipitation assays. *Mol. Cell. Biol.* **24**, 5923–36 (2004).

245. David, C. J., Chen, M., Assanah, M., Canoll, P. & Manley, J. L. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* **463**, 364–8 (2010).

246. Gan, L. *et al.* Metabolic targeting of oncogene MYC by selective activation of the proton-coupled monocarboxylate family of transporters. *Oncogene* **35**, 3037–3048 (2016).

247. Ullah, M. S., Davies, A. J. & Halestrap, A. P. The plasma membrane lactate transporter MCT4, but

not MCT1, is up-regulated by hypoxia through a HIF-1alpha-dependent mechanism. *J. Biol. Chem.* **281**, 9030–7 (2006).

248. Mathupala, S. P., Heese, C. & Pedersen, P. L. Glucose catabolism in cancer cells. The type II hexokinase promoter contains functionally active response elements for the tumor suppressor p53. *J. Biol. Chem.* **272**, 22776–80 (1997).

249. Macheda, M. L., Rogers, S. & Best, J. D. Molecular and cellular regulation of glucose transporter (GLUT) proteins in cancer. *J. Cell. Physiol.* **202**, 654–62 (2005).

250. Anderson, M., Marayati, R., Moffitt, R. & Yeh, J. J. Hexokinase 2 promotes tumor growth and metastasis by regulating lactate production in pancreatic cancer. *Oncotarget* **8**, 56081–56094 (2016).

251. Ashizawa, K., Willingham, M. C., Liang, C. M. & Cheng, S. Y. In vivo regulation of monomer-tetramer conversion of pyruvate kinase subtype M2 by glucose is mediated via fructose 1,6-bisphosphate. *J. Biol. Chem.* **266**, 16842–6 (1991).

252. Chaneton, B. *et al.* Serine is a natural ligand and allosteric activator of pyruvate kinase M2. *Nature* **491**, 458–462 (2012).

253. Mazurek, S. Pyruvate kinase type M2: A key regulator of the metabolic budget system in tumor cells. *Int. J. Biochem. Cell Biol.* **43**, 969–980 (2011).

254. Yang, W. *et al.* ERK1/2-dependent phosphorylation and nuclear translocation of PKM2 promotes the Warburg effect. *Nat. Cell Biol.* **14**, 1295–304 (2012).

255. Bartrons, R. *et al.* Fructose 2,6-Bisphosphate in Cancer Cell Metabolism. *Front. Oncol.* **8**, 331 (2018).

256. Minchenko, A. *et al.* Hypoxia-inducible factor-1-mediated expression of the 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase-3 (PFKFB3) gene. Its possible role in the Warburg effect. *J. Biol. Chem.* **277**, 6183–7 (2002).

257. Duran, J. *et al.* Pfkfb3 is transcriptionally upregulated in diabetic mouse liver through proliferative signals. *FEBS J.* **276**, 4555–68 (2009).

258. Feng, Y. & Wu, L. mTOR up-regulation of PFKFB3 is essential for acute myeloid leukemia cell survival. *Biochem. Biophys. Res. Commun.* **483**, 897–903 (2017).

259. Ros, S. *et al.* Functional Metabolic Screen Identifies 6-Phosphofructo-2-Kinase/Fructose-2,6-Biphosphatase 4 as an Important Regulator of Prostate Cancer Cell Survival. *Cancer Discov.* **2**, 328–343 (2012).

260. Strohecker, A. M. *et al.* Identification of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase as a novel autophagy regulator by high content shRNA screening. *Oncogene* **34**, 5662–76 (2015).

261. Bensaad, K. *et al.* TIGAR, a p53-Inducible Regulator of Glycolysis and Apoptosis. *Cell* **126**, 107–120 (2006).

262. Li, H. & Jogl, G. Structural and biochemical studies of TIGAR (TP53-induced glycolysis and apoptosis regulator). *J. Biol. Chem.* **284**, 1748–1754 (2009).

263. Kim, J., Tchernyshyov, I., Semenza, G. L. & Dang, C. V. HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia. *Cell Metab.* **3**, 177–185 (2006).

264. Lu, C.-W., Lin, S.-C., Chen, K.-F., Lai, Y.-Y. & Tsai, S.-J. Induction of pyruvate dehydrogenase kinase-3 by hypoxia-inducible factor-1 promotes metabolic switch and drug resistance. *J. Biol. Chem.* **283**, 28106–14 (2008).

265. Contractor, T. & Harris, C. R. p53 negatively regulates transcription of the pyruvate dehydrogenase kinase Pdk2. *Cancer Res.* **72**, 560–567 (2012).

266. Chae, Y. C. *et al.* Mitochondrial Akt Regulation of Hypoxic Tumor Reprogramming. *Cancer Cell* **30**, 257–272 (2016).

267. Benito, A. *et al.* Glucose-6-phosphate dehydrogenase and transketolase modulate breast cancer cell metabolic reprogramming and correlate with poor patient outcome. *Oncotarget* **8**, 106693–106706 (2017).

268. Jin, L. & Zhou, Y. Crucial role of the pentose phosphate pathway in malignant tumors. *Oncol. Lett.* **17**, 4213–4221 (2019).

269. Fan, J. *et al.* Quantitative flux analysis reveals folate-dependent NADPH production. *Nature* **510**, 298–302 (2014).

270. Kuehne, A. *et al.* Acute Activation of Oxidative Pentose Phosphate Pathway as First-Line Response to Oxidative Stress in Human Skin Cells. *Mol. Cell* **59**, 359–371 (2015).

271. Thimmulappa, R. K. *et al.* Identification of Nrf2-regulated genes induced by the chemopreventive agent sulforaphane by oligonucleotide microarray. *Cancer Res.* **62**, 5196–203 (2002).

272. Mitsuishi, Y. *et al.* Nrf2 redirects glucose and glutamine into anabolic pathways in metabolic reprogramming. *Cancer Cell* **22**, 66–79 (2012).

273. Xu, I. M.-J. *et al.* Transketolase counteracts oxidative stress to drive cancer development. *Proc. Natl. Acad. Sci.* **113**, E725–E734 (2016).

274. Rada, P. *et al.* SCF/{beta}-TrCP promotes glycogen synthase kinase 3-dependent degradation of the Nrf2 transcription factor in a Keap1-independent manner. *Mol. Cell. Biol.* **31**, 1121–33 (2011).

275. Chowdhry, S. *et al.* Nrf2 is controlled by two distinct β-TrCP recognition motifs in its Neh6 domain, one of which can be modulated by GSK-3 activity. *Oncogene* **32**, 3765–3781 (2013).

276. Ribback, S. *et al.* Oncogene-dependent addiction to carbohydrate-responsive element binding protein in hepatocellular carcinoma. *Cell Cycle* **17**, 1496–1512 (2018).

277. Wang, H. *et al.* Myc and ChREBP transcription factors cooperatively regulate normal and neoplastic hepatocyte proliferation in mice. *J. Biol. Chem.* **293**, 14740–14757 (2018).

278. Ma, X. *et al.* Polo-like kinase 1 coordinates biosynthesis during cell cycle progression by directly activating pentose phosphate pathway. *Nat. Commun.* **8**, 1506 (2017).

279. Kuo, C. Y. & Ann, D. K. When fats commit crimes: Fatty acid metabolism, cancer stemness and therapeutic resistance. *Cancer Commun.* **38**, 1–12 (2018).

280. Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* **7**, 763–77 (2007).

281. Guo, D., Bell, E. H., Mischel, P. & Chakravarti, A. Targeting SREBP-1-driven lipid metabolism to treat cancer. *Curr. Pharm. Des.* **20**, 2619–26 (2014).

282. Wen, Y.-A. *et al.* Downregulation of SREBP inhibits tumor growth and initiation by altering cellular metabolism in colon cancer. *Cell Death Dis.* **9**, 265 (2018).

283. Migita, T. *et al.* ATP Citrate Lyase: Activation and Therapeutic Implications in Non-Small Cell Lung Cancer. *Cancer Res.* **68**, 8547–8554 (2008).

284. Wang, D., Yin, L., Wei, J., Yang, Z. & Jiang, G. ATP citrate lyase is increased in human breast cancer, depletion of which promotes apoptosis. *Tumor Biol.* **39**, 101042831769833 (2017).

285. Ran, H. *et al.* Stearoyl-CoA desaturase-1 promotes colorectal cancer metastasis in response to glucose by suppressing PTEN. *J. Exp. Clin. Cancer Res.* **37**, 54 (2018).

286. Levade, T., Andrieu-Abadie, N., Micheau, O., Legembre, P. & Ségui, B. Sphingolipids modulate the epithelial–mesenchymal transition in cancer. *Cell Death Discov.* **1**, 3–4 (2015).

287. Camacho, L. *et al.* Acid ceramidase as a therapeutic target in metastatic prostate cancer. *J. Lipid Res.* **54**, 1207–1220 (2013).

288. Dadsena, S. *et al.* Ceramides bind VDAC2 to trigger mitochondrial apoptosis. *Nat. Commun.* **10**, 1832 (2019).

289. Camarda, R. *et al.* Inhibition of fatty acid oxidation as a therapy for MYC-overexpressing triple-negative breast cancer. *Nat. Med.* **22**, 427–32 (2016).

290. Lue, H.-W. *et al.* Metabolic reprogramming ensures cancer cell survival despite oncogenic signaling blockade. *Genes Dev.* **31**, 2067–2084 (2017).

291. Marín de Mas, I. *et al.* Model-driven discovery of long-chain fatty acid metabolic reprogramming in heterogeneous prostate cancer cells. *PLoS Comput. Biol.* **14**, e1005914 (2018).

292. Piskounova, E. *et al.* Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature* **527**, 186–91 (2015).

293. Meiser, J. *et al.* Serine one-carbon catabolism with formate overflow. *Sci. Adv.* **2**, 1–11 (2016).

294. Tedeschi, P. M. *et al.* Contribution of serine, folate and glycine metabolism to the ATP, NADPH and purine requirements of cancer cells. *Cell Death Dis.* **4**, e877-12 (2013).

295. Tibbetts, A. S. & Appling, D. R. Compartmentalization of Mammalian Folate-Mediated One-Carbon Metabolism. *Annu. Rev. Nutr.* **30**, 57–81 (2010).

296. Ducker, G. S. *et al.* Reversal of Cytosolic One-Carbon Flux Compensates for Loss of the Mitochondrial Folate Pathway. *Cell Metab.* **23**, 1140–1153 (2016).

297. Kim, Y.-I. Nutritional epigenetics: impact of folate deficiency on DNA methylation and colon cancer susceptibility. *J. Nutr.* **135**, 2703–9 (2005).

298. Locasale, J. W. Serine, glycine and the one-carbon cycle: cancer metabolism is full circle. *Nat. Rev. Cancer* **13**, 572–583 (2013).

299. Newman, A. C. & Maddocks, O. D. K. One-carbon metabolism in cancer. *Br. J. Cancer* **116**, 1499–1504 (2017).

300. Mazurek, S., Boschek, C. B., Hugo, F. & Eigenbrodt, E. Pyruvate kinase type M2 and its role in tumor growth and spreading. *Semin. Cancer Biol.* **15**, 300–308 (2005).

301. Nikiforov, M. A. *et al.* A Functional Screen for Myc-Responsive Genes Reveals Serine Hydroxymethyltransferase, a Major Source of the One-Carbon Unit for Cell Metabolism. *Mol. Cell. Biol.* **22**, 5793–5800 (2002).

302. Vazquez, A., Tedeschi, P. M. & Bertino, J. R. Overexpression of the Mitochondrial Folate and Glycine-Serine Pathway: A New Determinant of Methotrexate Selectivity in Tumors. *Cancer Res.* **73**, 478–482 (2013).

303. Sun, L. *et al.* cMyc-mediated activation of serine biosynthesis pathway is critical for cancer progression under nutrient deprivation conditions. *Cell Res.* **25**, 429–444 (2015).

304. Pikman, Y. *et al.* Targeting MTHFD2 in acute myeloid leukemia. *J. Exp. Med.* **213**, 1285–1306 (2016).

305. Liu, Y.-C. *et al.* Global Regulation of Nucleotide Biosynthetic Genes by c-Myc. *PLoS One* **3**, e2722 (2008).

306. Barfeld, S. J. *et al.* Myc-dependent purine biosynthesis affects nucleolar stress and therapy response in prostate cancer. *Oncotarget* **6**, 12587–602 (2015).

307. Liu, W. *et al.* Reprogramming of proline and glutamine metabolism contributes to the proliferative and metabolic responses regulated by oncogenic transcription factor c-MYC. *Proc. Natl. Acad. Sci.* **109**, 8983–8988 (2012).

308. Bhutia, Y. D., Babu, E., Ramachandran, S. & Ganapathy, V. Amino Acid transporters in cancer and their relevance to 'glutamine addiction': novel targets for the design of a new class of anticancer drugs. *Cancer Res.* **75**, 1782–8 (2015).

309. Yue, M., Jiang, J., Gao, P., Liu, H. & Qing, G. Oncogenic MYC Activates a Feedforward Regulatory Loop Promoting Essential Amino Acid Metabolism and Tumorigenesis. *Cell Rep.* **21**, 3819–3832 (2017).

310. Gao, P. *et al.* c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature* **458**, 762–5 (2009).

311. Zhao, X., Petrashen, A. P., Sanders, J. A., Peterson, A. L. & Sedivy, J. M. SLC1A5 glutamine transporter is a target of MYC and mediates reduced mTORC1 signaling and increased fatty acid oxidation in long-lived *Myc* hypomorphic mice. *Aging Cell* **18**, e12947 (2019).

312. Yang, L., Venneti, S. & Nagrath, D. Glutaminolysis: A Hallmark of Cancer Metabolism. *Annu. Rev. Biomed. Eng.* **19**, 163–194 (2017).

313. Son, J. *et al.* Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature* **496**, 101–5 (2013).

314. Boysen, G. *et al.* Glutaminase inhibitor CB-839 increases radiation sensitivity of lung tumor cells and human lung tumor xenografts in mice. *Int. J. Radiat. Biol.* **95**, 436–442 (2019).

315. Sappington, D. R. *et al.* Glutamine drives glutathione synthesis and contributes to radiation sensitivity of A549 and H460 lung cancer cell lines. *Biochim. Biophys. Acta* **1860**, 836–43 (2016).

316. Ma, M.-Z. *et al.* Xc- inhibitor sulfasalazine sensitizes colorectal cancer to cisplatin by a GSH-dependent mechanism. *Cancer Lett.* **368**, 88–96 (2015).

317. Sato, M. *et al.* The ferroptosis inducer erastin irreversibly inhibits system xc- and synergizes with cisplatin to increase cisplatin's cytotoxicity in cancer cells. *Sci. Rep.* **8**, 1–9 (2018).

318. Estrela, J. M., Ortega, A., Mena, S., Sirerol, J. A. & Obrador, E. Glutathione in metastases: From mechanisms to clinical applications. *Crit. Rev. Clin. Lab. Sci.* **53**, 253–267 (2016).

319. Bansal, A. & Simon, M. C. Glutathione metabolism in cancer progression and treatment resistance. *J. Cell Biol.* **217**, 2291–2298 (2018).

320. Chawla, R. K. *et al.* Plasma cysteine, cystine, and glutathione in cirrhosis. *Gastroenterology* **87**, 770–6 (1984).

321. Crawhall, J. C., Lietman, P. S., Schneider, J. A. & Seegmiller, J. E. Cystinosis. Plasma cystine and cysteine concentrations and the effect of D-penicillamine and dietary treatment. *Am. J. Med.* **44**, 330–9 (1968).

322. Bridges, R. J., Natale, N. R. & Patel, S. A. System xc^- cystine/glutamate antiporter: an update on molecular pharmacology and roles within the CNS. *Br. J. Pharmacol.* **165**, 20–34 (2012).

323. Chillarón, J. *et al.* Pathophysiology and treatment of cystinuria. *Nat. Rev. Nephrol.* **6**, 424–434 (2010).

324. Tietze, F., Bradley, K. H. & Schulman, J. D. Enzymic reduction of cystine by subcellular fractions of cultured and peripheral leukocytes from normal and cystinotic individuals. *Pediatr. Res.* **6**, 649–58 (1972).

325. States, B. & Segal, S. Interrelationship of glutathione–cystine transhydrogenase and glutathione reductase in developing rat intestine. *Biochem. J.* **132**, 623–631 (1973).

326. Bannai, S. & Kitamura, E. Transport interaction of L-cystine and L-glutamate in human diploid fibroblasts in culture. *J. Biol. Chem.* **255**, 2372–6 (1980).

327. Holmgren, A. Bovine thioredoxin system. Purification of thioredoxin reductase from calf liver and

thymus and studies of its function in disulfide reduction. *J. Biol. Chem.* **252**, 4600–6 (1977).

328. Mandal, P. K. *et al.* System xc- and thioredoxin reductase 1 cooperatively rescue glutathione deficiency. *J. Biol. Chem.* **285**, 22244–22253 (2010).

329. Habib, E., Linher-Melville, K., Lin, H.-X. & Singh, G. Expression of xCT and activity of system xc− are regulated by NRF2 in human breast cancer cells in response to oxidative stress. *Redox Biol.* **5**, 33–42 (2015).

330. Sbodio, J. I., Snyder, S. H. & Paul, B. D. Regulators of the transsulfuration pathway. *Br. J. Pharmacol.* **176**, 583–593 (2019).

331. Harvey, C. J. *et al.* Nrf2-regulated glutathione recycling independent of biosynthesis is critical for cell survival during oxidative stress. *Free Radic. Biol. Med.* **46**, 443–53 (2009).

332. Lu, S. C. Regulation of glutathione synthesis. *Mol. Aspects Med.* **30**, 42–59 (2009).

333. Aguilar, E. *et al.* Metabolic Reprogramming and Dependencies Associated with Epithelial Cancer Stem Cells Independent of the Epithelial-Mesenchymal Transition Program. *Stem Cells* **34**, 1163–1176 (2016).

334. Bernfeld, E. & Foster, D. A. Glutamine as an Essential Amino Acid for KRas-Driven Cancer Cells. *Trends Endocrinol. Metab.* **xx**, 1–12 (2019).

335. Choi, Y. K. & Park, K. G. Targeting glutamine metabolism for cancer treatment. *Biomol. Ther.* **26**, 19–28 (2018).

336. Gout, P. W., Kang, Y. J., Buckley, D. J., Bruchovsky, N. & Buckley, A. R. Increased cystine uptake capability associated with malignant progression of Nb2 lymphoma cells. *Leukemia* **11**, 1329–37 (1997).

337. Doxsee, D. W. *et al.* Sulfasalazine-induced cystine starvation: potential use for prostate cancer therapy. *Prostate* **67**, 162–71 (2007).

338. Lo, M., Ling, V., Wang, Y. Z. & Gout, P. W. The xc- cystine/glutamate antiporter: a mediator of pancreatic cancer growth with a role in drug resistance. *Br. J. Cancer* **99**, 464–72 (2008).

339. Lo, M., Ling, V., Low, C., Wang, Y. Z. & Gout, P. W. Potential use of the anti-inflammatory drug, sulfasalazine, for targeted therapy of pancreatic cancer. *Curr. Oncol.* **17**, 9–16 (2010).

340. Ishimoto, T. *et al.* CD44 Variant Regulates Redox Status in Cancer Cells by Stabilizing the xCT Subunit of System xc- and Thereby Promotes Tumor Growth. *Cancer Cell* **19**, 387–400 (2011).

341. Chen, R.-S. *et al.* Disruption of xCT inhibits cancer cell metastasis via the caveolin-1/β-catenin pathway. *Oncogene* **28**, 599–609 (2009).

342. Shin, S.-S. *et al.* Participation of xCT in melanoma cell proliferation in vitro and tumorigenesis in vivo. *Oncogenesis* **7**, 86 (2018).

343. Lim, J. K. M. *et al.* Cystine/glutamate antiporter xCT (SLC7A11) facilitates oncogenic RAS transformation by preserving intracellular redox balance. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 9433–9442 (2019).

344. Chabner, B. A. & Roberts, T. G. Chemotherapy and the war on cancer. *Nat. Rev. Cancer* **5**, 65–72 (2005).

345. Walling, J. From methotrexate to pemetrexed and beyond. A review of the pharmacodynamic and clinical properties of antifolates. *Invest. New Drugs* **24**, 37–77 (2006).

346. Ross, K. C. *et al.* Methotrexate sensitizes drug-resistant metastatic melanoma cells to BRAF V600E inhibitors dabrafenib and encorafenib. *Oncotarget* **9**, 13324–13336 (2018).

347. Patel, J. D. *et al.* Pemetrexed Continuation Maintenance Phase 3 Trials in Nonsquamous, Non–Small-Cell Lung Cancer: Focus on 2-Year Overall Survival and Continuum of Care. *Clin. Lung Cancer*

19, e823–e830 (2018).

348. Mrugala, M. M. *et al.* Phase II Study of Systemic High-dose Methotrexate and Intrathecal Liposomal Cytarabine for Treatment of Leptomeningeal Carcinomatosis From Breast Cancer. *Clin. Breast Cancer* (2019). doi:10.1016/j.clbc.2019.04.004

349. Paone, A. *et al.* SHMT1 knockdown induces apoptosis in lung cancer cells by causing uracil misincorporation. *Cell Death Dis.* **5**, e1525 (2014).

350. Ducker, G. S. *et al.* Human SHMT inhibitors reveal defective glycine import as a targetable metabolic vulnerability of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11404–11409 (2017).

351. Ding, K., Jiang, J., Chen, L. & Xu, X. Methylenetetrahydrofolate Dehydrogenase 1 Silencing Expedites the Apoptosis of Non-Small Cell Lung Cancer Cells via Modulating DNA Methylation. *Med. Sci. Monit.* **24**, 7499–7507 (2018).

352. Kitano, H. Systems biology: a brief overview. *Science* **295**, 1662–4 (2002).

353. Nielsen, J. Systems Biology of Metabolism. *Annu. Rev. Biochem.* **86**, 245–275 (2017).

354. Selivanov, V. A., Marin, S., Lee, P. W. N. & Cascante, M. Software for dynamic analysis of tracer-based metabolomic data: estimation of metabolic fluxes and their statistical analysis. *Bioinformatics* **22**, 2806–12 (2006).

355. Niedenführ, S., Wiechert, W. & Nöh, K. How to measure metabolic fluxes: a taxonomic guide for (13)C fluxomics. *Curr. Opin. Biotechnol.* **34**, 82–90 (2015).

356. DeWaal, D. *et al.* Hexokinase-2 depletion inhibits glycolysis and induces oxidative phosphorylation in hepatocellular carcinoma and sensitizes to metformin. *Nat. Commun.* **9**, 446 (2018).

357. Saa, P. A. & Nielsen, L. K. Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnol. Adv.* **35**, 981–1003 (2017).

358. Puigjaner, J. *et al.* Comparison of control analysis data using different approaches: modelling and experiments with muscle extract. *FEBS Lett.* **418**, 47–52 (1997).

359. Voit, E. O., Martens, H. A. & Omholt, S. W. 150 years of the mass action law. *PLoS Comput. Biol.* **11**, e1004012 (2015).

360. Liebermeister, W. & Klipp, E. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Model.* **3**, 41 (2006).

361. Selivanov, V. A. *et al.* Rapid simulation and analysis of isotopomer distributions using constraints based on enzyme mechanisms: an example from HT29 cancer cells. *Bioinformatics* **21**, 3558–64 (2005).

362. Foguet, C. *et al.* HepatoDyn: A Dynamic Model of Hepatocyte Metabolism That Integrates 13C Isotopomer Data. *PLoS Comput. Biol.* **12**, e1004899 (2016).

363. Placzek, S. *et al.* BRENDA in 2017: New perspectives and new tools in BRENDA. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkw952

364. Reich, J. & Selkov, E. *Energy Metabolism of the Cell. A Theoretical Treatise*. (1981).

365. Flamholz, A., Noor, E., Bar-Even, A. & Milo, R. eQuilibrator--the biochemical thermodynamics calculator. *Nucleic Acids Res.* **40**, D770–D775 (2012).

366. de Mas, I. M. *et al.* Compartmentation of glycogen metabolism revealed from 13C isotopologue distributions. *BMC Syst. Biol.* **5**, 175 (2011).

367. Berndt, N. *et al.* HEPATOKIN1 is a biochemistry-based model of liver metabolism for applications in medicine and pharmacology. *Nat. Commun.* **9**, 2386 (2018).

368. Sabate, L., Franco, R., Canela, E. I., Centelles, J. J. & Cascante, M. A model of the pentose phosphate pathway in rat liver cells. *Mol. Cell. Biochem.* **142**, 9–17 (1995).

369. Maher, A. D. *et al.* Mathematical modelling of the urea cycle. A numerical investigation into substrate channelling. *Eur. J. Biochem.* **270**, 3953–61 (2003).

370. Mogilevskaya, E., Demin, O. & Goryanin, I. Kinetic model of mitochondrial Krebs cycle: unraveling the mechanism of salicylate hepatotoxic effects. *J. Biol. Phys.* **32**, 245–71 (2006).

371. Yizhak, K., Chaneton, B., Gottlieb, E. & Ruppin, E. Modeling cancer metabolism on a genome scale. *Mol. Syst. Biol.* **11**, 817 (2015).

372. O'Brien, E. J., Monk, J. M. & Palsson, B. O. Using genome-scale models to predict biological capabilities. *Cell* **161**, 971–987 (2015).

373. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–8 (2010).

374. Song, H. & Ramkrishna, D. When is the Quasi-Steady-State Approximation Admissible in Metabolic Modeling ? When Admissible , What Models are Desirable ? 7976–7985 (2009).

375. Reimers, A. M. & Reimers, A. C. The steady-state assumption in oscillating and growing systems. *J. Theor. Biol.* **406**, 176–186 (2016).

376. de Mas, I. M. *et al.* Cancer cell metabolism as new targets for novel designed therapies. *Future Med. Chem.* **6**, 1791–1810 (2014).

377. Bairoch, A. The ENZYME database in 2000. *Nucleic Acids Res.* (2002). doi:10.1093/nar/28.1.304

378. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

379. Agren, R. *et al.* The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. *PLoS Comput. Biol.* **9**, (2013).

380. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).

381. Thiele, I. & Palsson, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).

382. Meinken, J. & Min, J. Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report. *Comput. Mol. Biol.* **2**, 1–7 (2012).

383. Noor, E., Haraldsdóttir, H. S., Milo, R. & Fleming, R. M. T. Consistent Estimation of Gibbs Energy Using Component Contributions. *PLoS Comput. Biol.* **9**, e1003098 (2013).

384. Benedict, M. N., Gonnerman, M. C., Metcalf, W. W. & Price, N. D. Genome-scale metabolic reconstruction and hypothesis testing in the methanogenic archaeon Methanosarcina acetivorans C2A. *J. Bacteriol.* (2012). doi:10.1128/JB.06040-11

385. Blais, E. M. *et al.* Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat. Commun.* (2017). doi:10.1038/ncomms14250

386. Heavner, B. D., Smallbone, K., Barker, B., Mendes, P. & Walker, L. P. Yeast 5 - an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. *BMC Syst. Biol.* (2012). doi:10.1186/1752-0509-6-55

387. Mintz-Oron, S. *et al.* Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc. Natl. Acad. Sci.* (2011). doi:10.1073/pnas.1100358109

388. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of Escherichia coli metabolism-

2011. *Mol. Syst. Biol.* (2011). doi:10.1038/msb.2011.65

389. Duarte, N. C. *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci.* **104**, 1777–1782 (2007).

390. Thiele, I. *et al.* A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **31**, 419–425 (2013).

391. Agren, R. *et al.* Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* **8**, e1002518 (2012).

392. Swainston, N. *et al.* Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* **12**, (2016).

393. Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **36**, 272–281 (2018).

394. Feist, A. M. & Palsson, B. O. The biomass objective function. *Curr. Opin. Microbiol.* **13**, 344–349 (2010).

395. Gille, C. *et al.* HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol. Syst. Biol.* **6**, 411 (2010).

396. Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489 (2010).

397. Lewis, N. E. *et al.* Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6**, 390 (2010).

398. Llaneras, F. & Picó, J. An interval approach for dealing with flux distributions and elementary modes activity patterns. *J. Theor. Biol.* **246**, 290–308 (2007).

399. Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. & Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* **26**, 1003–1010 (2008).

400. Becker, S. A. & Palsson, B. O. Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Comput. Biol.* **4**, e1000082 (2008).

401. Zur, H., Ruppin, E. & Shlomi, T. iMAT: An integrative metabolic analysis tool. *Bioinformatics* **26**, 3140–3142 (2010).

402. Jensen, P. A. & Papin, J. A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* **27**, 541–547 (2011).

403. Valcárcel, L. V, Torrano, V., Tobalina, L., Carracedo, A. & Planes, F. J. rMTA: Robust Metabolic Transformation Analysis. *Bioinformatics* (2019). doi:10.1093/bioinformatics/btz231

404. Yizhak, K. *et al.* Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. *Elife* **3**, 1–23 (2014).

405. Schmidt, B. J. *et al.* GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics* **29**, 2900–8 (2013).

406. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol.* **20**, 1–18 (2019).

407. Fouladiha, H. & Marashi, S. A. Biomedical applications of cell- and tissue-specific metabolic network models. *J. Biomed. Inform.* **68**, 35–49 (2017).

408. Nilsson, A. & Nielsen, J. Genome scale metabolic modeling of cancer. *Metab. Eng.* **43**, 103–112 (2017).

409. Folger, O. *et al.* Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **7**, 501–501 (2014).

410. Agren, R. *et al.* Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* **10**, 721–721 (2014).

411. Ghaffari, P. *et al.* Identifying anti-growth factors for human cancer cell lines through genome-scale metabolic modeling. *Sci. Rep.* **5**, 1–10 (2015).

412. Auslander, N. *et al.* An integrated computational and experimental study uncovers FUT 9 as a metabolic driver of colorectal cancer . *Mol. Syst. Biol.* **13**, 956 (2017).

413. Cowley, G. S. *et al.* Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci. Data* **1**, 1–12 (2014).

414. McDonald, E. R. *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170**, 577-592.e10 (2017).

415. Vizeacoumar, F. J. *et al.* A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. *Mol. Syst. Biol.* **9**, 696–696 (2014).

416. Wong, A. S. L. *et al.* Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc. Natl. Acad. Sci.* **113**, 2544–2549 (2016).

417. Shen, J. P. *et al.* Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* **14**, 573–576 (2017).

418. Guo, J., Liu, H. & Zheng, J. SynLethDB: Synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. *Nucleic Acids Res.* **44**, D1011–D1017 (2016).

419. Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci.* **99**, 15112–15117 (2002).

420. Apaolaza, I. *et al.* An in-silico approach to predict and exploit synthetic lethality in cancer metabolism. *Nat. Commun.* **8**, 1–9 (2017).

421. Ballerstein, K., von Kamp, A., Klamt, S. & Haus, U. U. Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics* **28**, 381–387 (2012).

422. Apaolaza, I., Valcarcel, L. V. & Planes, F. J. gMCS: fast computation of genetic minimal cut sets in large networks. *Bioinformatics* 1–3 (2018). doi:10.1093/bioinformatics/bty656

423. Yizhak, K., Gabay, O., Cohen, H. & Ruppin, E. Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat. Commun.* **4**, 2632 (2013).

424. Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* **9**, 68–86 (2007).

425. Buescher, J. M. *et al.* A roadmap for interpreting (13)C metabolite labeling patterns from cells. *Curr. Opin. Biotechnol.* **34**, 189–201 (2015).

426. Antoniewicz, M. R. A guide to 13C metabolic flux analysis for the cancer biologist. *Exp. Mol. Med.* **50**, 19 (2018).

427. Lane, A. N., Fan, T. W. M., Xie, Z., Moseley, H. N. B. & Higashi, R. M. Isotopomer analysis of lipid biosynthesis by high resolution mass spectrometry and NMR. *Anal. Chim. Acta* **651**, 201–208 (2009).

428. Chong, M. *et al.* Combined Analysis of NMR and MS Spectra (CANMS). *Angew. Chemie - Int. Ed.* **56**, 4140–4144 (2017).

429. Paul Lee, W.-N., Wahjudi, P. N., Xu, J. & Go, V. L. Tracer-based metabolomics: concepts and practices. *Clin. Biochem.* **43**, 1269–77 (2010).

430. Fernandez, C. A., Des Rosiers, C., Previs, S. F., David, F. & Brunengraber, H. Correction of 13C mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.* **31**, 255–62 (1996).

431. Millard, P., Letisse, F., Sokol, S. & Portais, J.-C. IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics* **28**, 1294–6 (2012).

432. Selivanov, V. A. *et al.* MIDcor, an R-program for deciphering mass interferences in mass spectra of metabolites enriched in stable isotopes. *BMC Bioinformatics* **18**, 88 (2017).

433. Cheah, Y. E. & Young, J. D. Isotopically nonstationary metabolic flux analysis (INST-MFA): putting theory into practice. *Curr. Opin. Biotechnol.* **54**, 80–87 (2018).

434. Selivanov, V. A., Marin, S., Lee, P. W. N. & Cascante, M. Software for dynamic analysis of tracer-based metabolomic data: Estimation of metabolic fluxes and their statistical analysis. *Bioinformatics* **22**, 2806–2812 (2006).

435. Wahl, S. A., Nöh, K. & Wiechert, W. 13C labeling experiments at metabolic nonstationary conditions: An exploratory study. *BMC Bioinformatics* **9**, 152 (2008).

436. Wahl, S. A., Nöh, K. & Wiechert, W. 13C labeling experiments at metabolic nonstationary conditions: An exploratory study. *BMC Bioinformatics* **9**, 152 (2008).

437. Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab. Eng.* **8**, 324–37 (2006).

438. Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab. Eng.* **8**, 324–337 (2006).

439. Gopalakrishnan, S. & Maranas, C. D. ^{13}C metabolic flux analysis at a genome-scale. *Metab. Eng.* **32**, 12–22 (2015).

440. Crown, S. B., Long, C. P. & Antoniewicz, M. R. Optimal tracers for parallel labeling experiments and 13C metabolic flux analysis: A new precision and synergy scoring system. *Metab. Eng.* **38**, 10–18 (2016).

441. Antoniewicz, M. R. Parallel labeling experiments for pathway elucidation and (13)C metabolic flux analysis. *Curr. Opin. Biotechnol.* **36**, 91–7 (2015).

442. Foguet, C. *et al.* p13CMFA: Parsimonious 13C metabolic flux analysis. *PLOS Comput. Biol.* **15**, e1007310 (2019).

443. Woods, H. F., Eggleston, L. V & Krebs, H. A. The cause of hepatic accumulation of fructose 1-phosphate on fructose loading. *Biochem. J.* **119**, 501–10 (1970).

444. Masson, S., Henriksen, O., Stengaard, A., Thomsen, C. & Quistorff, B. Hepatic metabolism during constant infusion of fructose; comparative studies with 31P-magnetic resonance spectroscopy in man and rats. *Biochim. Biophys. Acta* **1199**, 166–74 (1994).

445. Ahn, W. S., Crown, S. B. & Antoniewicz, M. R. Evidence for transketolase-like TKTL1 flux in CHO cells based on parallel labeling experiments and (13)C-metabolic flux analysis. *Metab. Eng.* **37**, 72–78 (2016).

446. Emami Khoonsari, P. *et al.* Interoperable and scalable data analysis with microservices: Applications in Metabolomics. *Bioinformatics* 1–9 (2019). doi:10.1093/bioinformatics/btz160

447. Peters, K. *et al.* PhenoMeNal: processing and analysis of metabolomics data in the cloud. *Gigascience* **8**, 1–12 (2019).

448. Weigand, J. E., Boeckel, J.-N., Gellert, P. & Dimmeler, S. Hypoxia-Induced Alternative Splicing in Endothelial Cells. *PLoS One* **7**, e42697 (2012).

449. Keating, S. M. & Le Novère, N. Supporting SBML as a Model Exchange Format in Software Applications. in *Methods in molecular biology (Clifton, N.J.)* **1021**, 201–225 (2013).

450. Weaver, B. & Wuensch, K. L. SPSS and SAS programs for comparing Pearson correlations and OLS regression coefficients. *Behav. Res. Methods* **45**, 880–895 (2013).

451. Silver, A. Software simplified. *Nature* **546**, 173–174 (2017).

452. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10 (2016).

453. King, Z. A. *et al.* Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLOS Comput. Biol.* **11**, e1004321 (2015).

454. Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–15 (2008).

455. Hennessy, B. T. *et al.* Characterization of a Naturally Occurring Breast Cancer Subset Enriched in Epithelial-to-Mesenchymal Transition and Stem Cell Characteristics. *Cancer Res.* **69**, 4116 (2009).

456. Blick, T. *et al.* Epithelial Mesenchymal Transition Traits in Human Breast Cancer Cell Lines Parallel the CD44hi/CD24lo/- Stem Cell Phenotype in Human Breast Cancer. *J. Mammary Gland Biol. Neoplasia* **15**, 235–252 (2010).

457. Sánchez-Cid, L. *et al.* MicroRNA-200, associated with metastatic breast cancer, promotes traits of mammary luminal progenitor cells. *Oncotarget* **8**, 83384–83406 (2017).

458. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 1–17 (2009).

459. Yang, C. *et al.* Downregulation of cancer stem cell properties via mTOR signaling pathway inhibition by rapamycin in nasopharyngeal carcinoma. *Int. J. Oncol.* **47**, 909–17 (2015).

460. Bahmad, H. F. *et al.* The Akt/mTOR pathway in cancer stem/progenitor cells is a potential therapeutic target for glioblastoma and neuroblastoma. *Oncotarget* **9**, 33549–33561 (2018).

461. Deng, J. *et al.* Inhibition of PI3K/Akt/mTOR signaling pathway alleviates ovarian cancer chemoresistance through reversing epithelial-mesenchymal transition and decreasing cancer stem cell marker expression. *BMC Cancer* **19**, 618 (2019).

462. Iadevaia, V., Zhang, Z., Jan, E. & Proud, C. G. mTOR signaling regulates the processing of pre-rRNA in human cells. *Nucleic Acids Res.* **40**, 2527–39 (2012).

463. Morita, M. *et al.* mTOR coordinates protein synthesis, mitochondrial activity and proliferation. *Cell Cycle* **14**, 473–80 (2015).

464. Tiwari, N. *et al.* Klf4 Is a Transcriptional Regulator of Genes Critical for EMT, Including Jnk1 (Mapk8). *PLoS One* **8**, (2013).

465. Li, X. *et al.* Klf4 reduces stemness phenotype, triggers mesenchymal-epithelial transition (MET)-like molecular changes, and prevents tumor progression in nasopharygeal carcinoma. *Oncotarget* **8**, 93924–93941 (2017).

466. Ganju, A. *et al.* Protein kinase D1 regulates subcellular localisation and metastatic function of metastasis-associated protein 1. *Br. J. Cancer* **118**, 587–599 (2018).

467. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2016).

468. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54 (2003).

469. Li, J., Ding, Y. & Li, A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World J. Surg. Oncol.* **14**, 1–5 (2016).

470. Stanbrough, M. *et al.* Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res.* **66**, 2815–2825 (2006).

471. Vaillant, C. *et al.* Serpine2/PN-1 is required for proliferative expansion of pre-neoplastic lesions and malignant progression to medulloblastoma. *PLoS One* **10**, 1–22 (2015).

472. Shen, Y., Wang, X., Xu, J. & Lu, L. SerpinE2, a poor biomarker of endometrial cancer, promotes the proliferation and mobility of EC cells. *Cancer Biomarkers* **19**, 271–278 (2017).

473. Wang, K. *et al.* Prognostic significance of SERPINE2 in gastric cancer and its biological function in SGC7901 cells. *J. Cancer Res. Clin. Oncol.* **141**, 805–812 (2015).

474. Buchholz, M. *et al.* SERPINE2 (protease nexin I) promotes extracellular matrix production and local invasion of pancreatic tumors in vivo. *Cancer Res.* **63**, 4945–4951 (2003).

475. Bergeron, S. *et al.* The serine protease inhibitor serpinE2 is a novel target of ERK signaling involved in human colorectal tumorigenesis. *Mol. Cancer* **9**, 1–15 (2010).

476. Wu, J.-I. & Wang, L.-H. Emerging roles of gap junction proteins connexins in cancer metastasis, chemoresistance and clinical application. *J. Biomed. Sci.* **26**, 1–14 (2019).

477. Stoletov, K. *et al.* Role of connexins in metastatic breast cancer and melanoma brain colonization. *J. Cell Sci.* **126**, 904–913 (2013).

478. Oliveira, R. *et al.* Contribution of gap junctional communication between tumor cells and astroglia to the invasion of the brain parenchyma by human glioblastomas. *BMC Cell Biol.* **6**, 1–17 (2005).

479. Behrens, J., Kameritsch, P., Wallner, S., Pohl, U. & Pogoda, K. The carboxyl tail of Cx43 augments p38 mediated cell migration in a gap junction-independent manner. *Eur. J. Cell Biol.* **89**, 828–838 (2010).

480. Ryszawy, D. *et al.* Functional links between Snail-1 and CX43 account for the recruitment of CX43-positive cells into the invasive front of prostate cancer. *Carcinogenesis* **35**, 1920–1930 (2014).

481. Zhang, A. *et al.* Connexin 43 expression is associated with increased malignancy in prostate cancer cell lines and functions to promote migration. *Oncotarget* **6**, 11640–51 (2015).

482. Zheng, W., Jiang, C. & Li, R. Integrin and gene network analysis reveals that ITGA5 and ITGB1 are prognostic in non-small-cell lung cancer. *Onco. Targets. Ther.* **9**, 2317–27 (2016).

483. Sawada, K. *et al.* Loss of E-cadherin promotes ovarian cancer metastasis via α5-integrin, which is a therapeutic target. *Cancer Res.* **68**, 2329–2339 (2008).

484. Gong, C. *et al.* MIR 17 inhibits ovarian cancer cell peritoneal metastasis by targeting ITGA5 and ITGB1. *Oncol. Rep.* **36**, 2177–2183 (2016).

485. Ju, J. A. *et al.* Hypoxia Selectively Enhances Integrin α 5 β 1 Receptor Expression in Breast Cancer to Promote Metastasis . *Mol. Cancer Res.* **15**, 723–734 (2017).

486. Zhang, X. *et al.* MicroRNA-26a promotes anoikis in human hepatocellular carcinoma cells by targeting alpha5 integrin. *Oncotarget* **6**, 2277–89 (2015).

487. Feng, L., Ma, J., Ji, H., Liu, Y. & Hu, W. miR-330-5p suppresses glioblastoma cell proliferation and invasiveness through targeting ITGA5. *Biosci. Rep.* **37**, BSR20170019 (2017).

488. Nakanishi, H. *et al.* Loss of imprinting of PEG1/MEST in lung cancer cell lines. *Oncol. Rep.* **12**, 1273–8 (2004).

489. Pedersen, I. S. *et al.* Frequent loss of imprinting of PEG1/MEST in invasive breast cancer. *Cancer Res.* **59**, 5449–51 (1999).

490. Boström, P. *et al.* MMP-1 expression has an independent prognostic value in breast cancer. *BMC Cancer* **11**, 1–8 (2011).

491. An, H. J. *et al.* The prognostic role of tissue and serum MMP-1 and TIMP-1 expression in patients with non-small cell lung cancer. *Pathol. Res. Pract.* **212**, 357–364 (2016).

492. Sunami, E. MMP-1 is a Prognostic Marker for Hematogenous Metastasis of Colorectal Cancer. *Oncologist* **5**, 108–114 (2004).

493. Ma, F. *et al.* MiR-361-5p inhibits glycolytic metabolism, proliferation and invasion of breast cancer by targeting FGFR1 and MMP-1. *J. Exp. Clin. Cancer Res.* **36**, 1–12 (2017).

494. Tang, M. li, Bai, X. jun, Li, Y., Dai, X. jing & Yang, F. MMP-1 Over-expression Promotes Malignancy and Stem-Like Properties of Human Osteosarcoma MG-63 Cells In Vitro. *Curr. Med. Sci.* **38**, 809–817 (2018).

495. Katkoori, V. R. *et al.* Prognostic significance and gene expression profiles of p53 mutations in microsatellite-stable Stage III colorectal adenocarcinomas. *PLoS One* **7**, (2012).

496. Zeng, X.-T., Liu, X.-P., Liu, T.-Z. & Wang, X.-H. The clinical significance of COL5A2 in patients with bladder cancer. *Medicine (Baltimore).* **97**, e0091 (2018).

497. Cheon, D.-J. *et al.* A collagen-remodeling gene signature regulated by TGF-β signaling is associated with metastasis and poor survival in serous ovarian cancer. *Clin. Cancer Res.* **20**, 711–23 (2014).

498. Wang, C. A., Harrell, J. C., Iwanaga, R., Jedlicka, P. & Ford, H. L. Vascular endothelial growth factor C promotes breast cancer progression via a novel antioxidant mechanism that involves regulation of superoxide dismutase 3. *Breast Cancer Res.* **16**, 1–17 (2014).

499. Goel, H. L. *et al.* GLI1 regulates a novel neuropilin-2/α6β1 integrin based autocrine pathway that contributes to breast cancer initiation. *EMBO Mol. Med.* **5**, 488–508 (2013).

500. Elaimy, A. L. *et al.* VEGF-neuropilin-2 signaling promotes stem-like traits in breast cancer cells by TAZ-mediated repression of the Rac GAP β2-chimaerin. *Sci. Signal.* **11**, (2018).

501. Goel, H. L. *et al.* VEGF/neuropilin-2 regulation of Bmi-1 and consequent repression of IGF-IR define a novel mechanism of aggressive prostate cancer. *Cancer Discov.* **2**, 906–21 (2012).

502. Muders, M. H., Zhang, H., Wang, E., Tindall, D. J. & Datta, K. Vascular endothelial growth factor-C protects prostate cancer cells from oxidative stress by the activation of mammalian target of rapamycin complex-2 and AKT-1. *Cancer Res.* **69**, 6042–6048 (2009).

503. Cao, Y. *et al.* Neuropilin-2 promotes extravasation and metastasis by interacting with endothelial α5 integrin. *Cancer Res.* **73**, 4579–4590 (2013).

504. Goel, H. L., Pursell, B., Standley, C., Fogarty, K. & Mercurio, A. M. Neuropilin-2 regulates 6 1 integrin in the formation of focal adhesions and signaling. *J. Cell Sci.* **125**, 497–506 (2012).

505. Greco, C. *et al.* E-cadherin/p120-catenin and tetraspanin Co-029 cooperate for cell motility control in human colon carcinoma. *Cancer Res.* **70**, 7674–7683 (2010).

506. Kobunai, T., Watanabe, T. & Fukusato, T. REG4, NEIL2, and BIRC5 gene expression correlates with gamma-radiation sensitivity in patients with rectal cancer receiving radiotherapy. *Anticancer Res.* **31**, 4147–53 (2011).

507. Elhassan, M. O., Christie, J. & Duxbury, M. S. Homo sapiens Systemic RNA Interference-defective-1 Transmembrane Family Member 1 (SIDT1) Protein Mediates Contact-dependent Small RNA Transfer and MicroRNA-21-driven Chemoresistance. *J. Biol. Chem.* **287**, 5267–5277 (2012).

508. Katsuno, Y. *et al.* Coordinated expression of REG4 and aldehyde dehydrogenase 1 regulating tumourigenic capacity of diffuse-type gastric carcinoma-initiating cells is inhibited by TGF-β. *J. Pathol.* **228**, 391–404 (2012).

509. Overbeck, T. R. *et al.* ABCA3 phenotype in non-small cell lung cancer indicates poor outcome. *Oncol.* **93**, 270–278 (2017).

510. Zhao, L. *et al.* TMEM45B promotes proliferation, invasion and migration and inhibits apoptosis in pancreatic cancer cells. *Mol. Biosyst.* **12**, 1860–1870 (2016).

511. Li, Y. *et al.* Silencing Transmembrane Protein 45B (TNEM45B) Inhibits Proliferation, Invasion, and Tumorigenesis in Osteosarcoma Cells. *Oncol. Res.* **25**, 1021–1026 (2017).

512. Cajigas-Du Ross, C. K. *et al.* RNA sequencing reveals upregulation of a transcriptomic program associated with stemness in metastatic prostate cancer cells selected for taxane resistance. *Oncotarget* **9**, 30363–30384 (2018).

513. Lai, X. *et al.* Loss of scinderin decreased expression of epidermal growth factor receptor and promoted apoptosis of castration-resistant prostate cancer cells. *FEBS Open Bio* **8**, 743–750 (2018).

514. Qiao, X. *et al.* Scinderin is a novel transcriptional target of BRMS1 involved in regulation of hepatocellular carcinoma cell apoptosis. *Am. J. Cancer Res.* **8**, 1008–1018 (2018).

515. Shen, K., Yu, W., Yu, Y., Liu, X. & Cui, X. Knockdown of TMEM45B inhibits cell proliferation and invasion in gastric cancer. *Biomed. Pharmacother.* **104**, 576–581 (2018).

516. El Kharbili, M. *et al.* Tspan8-β-catenin positive feedback loop promotes melanoma invasion. *Oncogene* **38**, 3781–3793 (2019).

517. Wang, Z., Sun, H., Provaznik, J., Hackert, T. & Zöller, M. Pancreatic cancer-initiating cell exosome message transfer into noncancer-initiating cells: the importance of CD44v6 in reprogramming. *J. Exp. Clin. Cancer Res.* **38**, 1–20 (2019).

518. Mezencev, R., Matyunina, L. V., Jabbari, N. & McDonald, J. F. Snail-induced epithelial-to-mesenchymal transition of MCF-7 breast cancer cells: systems analysis of molecular changes and their effect on radiation and drug sensitivity. *BMC Cancer* **16**, 236 (2016).

519. Mayer, M. J., Klotz, L. H. & Venkateswaran, V. Metformin and prostate cancer stem cells: A novel therapeutic target. *Prostate Cancer Prostatic Dis.* **18**, 303–309 (2015).

520. Rae, C., Sey, C. H. & Mairs, R. J. Radiosensitization of Prostate Cancer Cells by 2-Deoxyglucose. *Madridge J. Oncog.* **2**, 30–34 (2018).

521. Nishimura, T. *et al.* Cancer stem-like properties and gefitinib resistance are dependent on purine synthetic metabolism mediated by the mitochondrial enzyme MTHFD2. *Oncogene* **38**, 2464–2481 (2019).

522. Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**, 346–50 (2011).

523. Ye, Y., Wu, Y. & Wang, J. Pyrroline-5-carboxylate reductase 1 promotes cell proliferation via inhibiting apoptosis in human malignant melanoma. *Cancer Manag. Res.* **10**, 6399–6407 (2018).

524. Cai, F. *et al.* Pyrroline-5-carboxylate reductase 1 promotes proliferation and inhibits apoptosis in non-small cell lung cancer. *Oncology Letters* **15**, 731–740 (2018).

525. Zeng, T. *et al.* Knockdown of PYCR1 inhibits cell proliferation and colony formation via cell cycle arrest and apoptosis in prostate cancer. *Med. Oncol.* **34**, 1–9 (2017).

526. Zhang, W. C. *et al.* Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell* **148**, 259–72 (2012).

527. Kang, P. J. *et al.* Glycine decarboxylase regulates the maintenance and induction of pluripotency via metabolic control. *Metab. Eng.* **53**, 35–47 (2019).

528. Alptekin, A. *et al.* Glycine decarboxylase is a transcriptional target of MYCN required for neuroblastoma cell proliferation and tumorigenicity. *Oncogene* 1–17 (2019). doi:10.1038/s41388-019-0967-3

529. Elia, I., Doglioni, G. & Fendt, S. M. Metabolic Hallmarks of Metastasis Formation. *Trends Cell Biol.* **28**, 673–684 (2018).

530. Lehuédé, C., Dupuy, F., Rabinovitch, R., Jones, R. G. & Siegel, P. M. Metabolic Plasticity as a Determinant of Tumor Growth and Metastasis. *Cancer Res.* **76**, 5201–8 (2016).

531. Pascual, G., Domínguez, D. & Benitah, S. A. The contributions of cancer cell metabolism to metastasis. *Dis. Model. Mech.* **11**, (2018).

532. Yeh, J. J. *et al.* KRAS/BRAF mutation status and ERK1/2 activation as biomarkers for MEK1/2 inhibitor therapy in colorectal cancer. *Mol. Cancer Ther.* **8**, 834–43 (2009).

533. Hewitt, R. E. *et al.* Validation of a model of colon cancer progression. *J. Pathol.* **192**, 446–454 (2000).

534. Provenzani, A. *et al.* Global alterations in mRNA polysomal recruitment in a cell model of colorectal cancer progression to metastasis. *Carcinogenesis* **27**, 1323–33 (2006).

535. Sheffer, M. *et al.* Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7131–6 (2009).

536. Moyer, M. P., Manzano, L. A., Merriman, R. L., Stauffer, J. S. & Tanzer, L. R. NCM460, a normal human colon mucosal epithelial cell line. *In Vitro Cell. Dev. Biol. Anim.* **32**, 315–7 (1996).

537. Cannino, G., Ciscato, F., Masgras, I., Sánchez-Martín, C. & Rasola, A. Metabolic Plasticity of Tumor Cell Mitochondria. *Front. Oncol.* **8**, 333 (2018).

538. Savaskan, N. E. *et al.* Small interfering RNA–mediated xCT silencing in gliomas inhibits neurodegeneration and alleviates brain edema. *Nat. Med.* **14**, 629–632 (2008).

539. Todaro, M. *et al.* CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. *Cell Stem Cell* **14**, 342–356 (2014).

540. Dixon, S. J. *et al.* Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell* **149**, 1060–72 (2012).

541. Dixon, S. J. *et al.* Pharmacological inhibition of cystine–glutamate exchange induces endoplasmic reticulum stress and ferroptosis. *Elife* **3**, (2014).

542. Gustafsson, R. *et al.* Crystal Structure of the Emerging Cancer Target MTHFD2 in Complex with a Substrate-Based Inhibitor. *Cancer Res.* **77**, 937–948 (2017).

543. Schmidt, A. *et al.* Structures of three inhibitor complexes provide insight into the reaction mechanism of the human methylenetetrahydrofolate dehydrogenase/cyclohydrolase. *Biochemistry* **39**, 6325–35 (2000).

544. Anderson, D. D. & Stover, P. J. SHMT1 and SHMT2 are functionally redundant in nuclear de novo thymidylate biosynthesis. *PLoS One* **4**, e5839 (2009).

545. MacFarlane, A. J. *et al.* Nuclear localization of de novo thymidylate biosynthesis pathway is required to prevent uracil accumulation in DNA. *J. Biol. Chem.* **286**, 44015–22 (2011).

546. Anderson, D. D., Woeller, C. F., Chiang, E. P., Shane, B. & Stover, P. J. Serine hydroxymethyltransferase anchors de Novo thymidylate synthesis pathway to nuclear lamina for DNA synthesis. *J. Biol. Chem.* **287**, 7051–7062 (2012).

547. An, S., Kumar, R., Sheets, E. D. & Benkovic, S. J. Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. *Science* **320**, 103–6 (2008).

548. Sdelci, S. *et al.* MTHFD1 interaction with BRD4 links folate metabolism to transcriptional regulation. *Nat. Genet.* **51**, 990–998 (2019).

549. Macfarlane, A. J., Perry, C. A., McEntee, M. F., Lin, D. M. & Stover, P. J. Shmt1 heterozygosity impairs folate-dependent thymidylate synthesis capacity and modifies risk of Apc(min)-mediated intestinal cancer risk. *Cancer Res.* **71**, 2098–107 (2011).

550. Field, M. S., Kamynina, E., Watkins, D., Rosenblatt, D. S. & Stover, P. J. Human mutations in methylenetetrahydrofolate dehydrogenase 1 impair nuclear de novo thymidylate biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 400–405 (2015).

551. Kamynina, E. A. *et al.* Arsenic trioxide targets MTHFD1 and SUMO-dependent nuclear de novo thymidylate biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E2319–E2326 (2017).

552. Shen, Y. *et al.* Carnosine inhibits the proliferation of human gastric cancer SGC-7901 cells through both of the mitochondrial respiration and glycolysis pathways. *PLoS One* **9**, (2014).

553. Cheng, J.-Y. *et al.* Profiling and targeting of cellular mitochondrial bioenergetics: inhibition of human gastric cancer cell growth by carnosine. *Acta Pharmacol. Sin.* **40**, 938–948 (2019).

554. Bao, Y. *et al.* Carnosine Inhibits the Proliferation of Human Cervical Gland Carcinoma Cells ThrCarnosine Inhibits the Proliferation of Human Cervical Gland Carcinoma Cells Through Inhibiting Both Mitochondrial Bioenergetics and Glycolysis Pathways and Retarding Cell Cyc. *Integr. Cancer Ther.* **17**, 80–91 (2018).

555. Fouad, A. A., Qutub, H. O., Al Rashed, A. S. & Al-Melhim, W. N. Therapeutic effect of carnosine in rat model of experimental liver carcinogenesis. *Environ. Toxicol. Pharmacol.* **56**, 10–14 (2017).

556. Ding, M., Jiao, G., Shi, H. & Chen, Y. Investigations on in vitro anti-carcinogenic potential of L-carnosine in liver cancer cells. *Cytotechnology* **70**, 163–167 (2018).

557. Iovine, B. *et al.* The anti-proliferative effect of L-carnosine correlates with a decreased expression of hypoxia inducible factor 1 alpha in human colon cancer cells. *PLoS One* **9**, e96755 (2014).

558. Iovine, B., Guardia, F., Irace, C. & Bevilacqua, M. A. L-carnosine dipeptide overcomes acquired resistance to 5-fluorouracil in HT29 human colon cancer cells via downregulation of HIF1-alpha and induction of apoptosis. *Biochimie* **127**, 196–204 (2016).

559. Lee, J. *et al.* l-carnosine induces apoptosis/cell cycle arrest via suppression of NF-κB/STAT1 pathway in HCT116 colorectal cancer cells. *Vitr. Cell. Dev. Biol. - Anim.* **54**, 505–512 (2018).

560. Sale, C., Saunders, B. & Harris, R. C. Effect of beta-alanine supplementation on muscle carnosine concentrations and exercise performance. *Amino Acids* **39**, 321–333 (2010).

561. Foguet, C. *et al.* HepatoDyn: A Dynamic Model of Hepatocyte Metabolism That Integrates13C Isotopomer Data. *PLoS Comput. Biol.* **12**, e1004899 (2016).

562. Abdelmalek, M. F. *et al.* Higher dietary fructose is associated with impaired hepatic adenosine triphosphate homeostasis in obese individuals with type 2 diabetes. *Hepatology* **56**, 952–960 (2012).

563. Bu, P. *et al.* Aldolase B-Mediated Fructose Metabolism Drives Metabolic Reprogramming of Colon Cancer Liver Metastasis. *Cell Metab.* **27**, 1249-1262.e4 (2018).

564. Weng, Y. *et al.* SLC2A5 promotes lung adenocarcinoma cell growth and metastasis by enhancing fructose utilization. *Cell Death Discov.* **4**, (2018).

565. Dong, W., Keibler, M. A. & Stephanopoulos, G. Review of metabolic pathways activated in cancer cells as determined through isotopic labeling and network analysis. *Metabolic Engineering* **43**, 113–124 (2017).

566. Bruntz, R. C., Lane, A. N., Higashi, R. M. & Fan, T. W. M. Exploring cancer metabolism using Stable isotope-resolved metabolomics (SIRM). *J. Biol. Chem.* **292**, 11601–11609 (2017).

567. Balcells, C. *et al.* Tracing metabolic fluxes using mass spectrometry: Stable isotope-resolved metabolomics in health and disease. *TrAC Trends Anal. Chem.* (2019). doi:10.1016/j.trac.2018.12.025

568. Vanderlaag, K. *et al.* Regenerating islet-derived family member, 4 modulates multiple receptor tyrosine kinases and mediators of drug resistance in cancer. *Int. J. Cancer* **130**, 1251–1263 (2012).

569. Pan, S. J. *et al.* Over-expression of tetraspanin 8 in malignant glioma regulates tumor cell progression. *Biochem. Biophys. Res. Commun.* **458**, 476–482 (2015).

570. Yue, S., Mu, W., Erb, U. & Zöller, M. The tetraspanins CD151 and Tspan8 are essential exosome components for the crosstalk between cancer initiating cells and their surrounding. *Oncotarget* **6**, (2015).

571. Hu, R. *et al.* TMEM45B, up-regulated in human lung cancer, enhances tumorigenicity of lung cancer cells. *Tumour Biol.* **37**, 12181–12191 (2016).

572. Palorini, R. *et al.* Energy Metabolism Characterization of a Novel Cancer Stem Cell-Like Line 3AB-OS. *J. Cell. Biochem.* **115**, 368–379 (2014).

573. Shen, Y.-A., Wang, C.-Y., Hsieh, Y.-T., Chen, Y.-J. & Wei, Y.-H. Metabolic reprogramming orchestrates cancer stem cell properties in nasopharyngeal carcinoma. *Cell Cycle* **14**, 86–98 (2015).

574. Kolesnik, D. L. *et al.* Effect of dichloroacetate on Lewis lung carcinoma growth and metastasis. *Exp. Oncol.* **37**, 126–9 (2015).

575. Du, J. *et al.* PDK1 promotes tumor growth and metastasis in a spontaneous breast cancer model. *Oncogene* **35**, 3314–3323 (2016).

576. Chen, C.-L. *et al.* NANOG Metabolically Reprograms Tumor-Initiating Stem-like Cells through Tumorigenic Changes in Oxidative Phosphorylation and Fatty Acid Metabolism. *Cell Metab.* **23**, 206–19 (2016).

577. Shelton, L. M., Huysentruyt, L. C. & Seyfried, T. N. Glutamine targeting inhibits systemic metastasis in the VM-M3 murine tumor model. *Int. J. cancer* **127**, 2478–85 (2010).

578. Yang, L. *et al.* Metabolic shifts toward glutamine regulate tumor growth, invasion and bioenergetics in ovarian cancer. *Mol. Syst. Biol.* **10**, 728 (2014).

579. Rodrigues, M. F. *et al.* Enhanced OXPHOS, glutaminolysis and β-oxidation constitute the metastatic phenotype of melanoma cells. *Biochem. J.* **473**, 703–15 (2016).

580. Dornier, E. *et al.* Glutaminolysis drives membrane trafficking to promote invasiveness of breast cancer cells. *Nat. Commun.* **8**, 2255 (2017).

581. Liao, J. *et al.* Regulation of stem-like cancer cells by glutamine through β-catenin pathway mediated by redox signaling. *Mol. Cancer* **16**, 51 (2017).

582. Jin, L. *et al.* The PLAG1-GDH1 Axis Promotes Anoikis Resistance and Tumor Metastasis through CamKK2-AMPK Signaling in LKB1-Deficient Lung Cancer. *Mol. Cell* **69**, 87-99.e7 (2018).

583. Lu, V. *et al.* Mitochondrial metabolism and glutamine are essential for mesoderm differentiation of human pluripotent stem cells. *Cell Res.* **29**, 596–598 (2019).

584. Dong, C. *et al.* Loss of FBP1 by snail-mediated repression provides metabolic advantages in basal-like breast cancer. *Cancer Cell* **23**, 316–331 (2013).

585. Lin, H. *et al.* MTHFD2 Overexpression Predicts Poor Prognosis in Renal Cell Carcinoma and is Associated with Cell Proliferation and Vimentin-Modulated Migration and Invasion. *Cell. Physiol. Biochem.* **51**, 991–1000 (2018).

586. Cobler, L., Zhang, H., Suri, P., Park, C. & Timmerman, L. A. xCT inhibition sensitizes tumors to γ-radiation via glutathione reduction. *Oncotarget* **9**, 32280–32297 (2018).

587. Pan, X. *et al.* Erastin decreases radioresistance of NSCLC cells partially by inducing GPX4-mediated ferroptosis. *Oncol. Lett.* **17**, 3001–3008 (2019).

588. Andreassen, K., Mortensen, B., Winberg, J.-O. & Huseby, N.-E. Increased resistance towards oxidative stress accompanies enhancement of metastatic potential obtained by repeated in vivo passage of colon carcinoma cells in syngeneic rats. *Clin. Exp. Metastasis* **19**, 623–9 (2002).

589. Estrela, J. M. *et al.* γ-glutamyl transpeptidase overexpression increases metastatic growth of B16

melanoma cells in the mouse liver. *Hepatology* **35**, 74–81 (2002).

590. Ortega, A. L. *et al.* Tumor cytotoxicity by endothelial cells: Impairment of the mitochondrial system for glutathione uptake in mouse B16 melanoma cells that survive after in vitro interaction with the hepatic sinusoidal endothelium. *J. Biol. Chem.* **278**, 13888–13897 (2003).

591. Gal, K. Le *et al.* Antioxidants can increase melanoma metastasis in mice. *Sci. Transl. Med.* **7**, 1–8 (2015).

592. Sehm, T. *et al.* Sulfasalazine impacts on ferroptotic cell death and alleviates the tumor microenvironment and glioma-induced brain edema. *Oncotarget* **7**, 36021–36033 (2016).

593. Murphy, M. P. How mitochondria produce reactive oxygen species. *Biochem. J.* **417**, 1–13 (2009).

594. Porter, T. G. & Martin, D. L. Non-steady-state kinetics of brain glutamate decarboxylase resulting from interconversion of the apo- and holoenzyme. *Biochim. Biophys. Acta* **874**, 235–44 (1986).

595. Salvarrey, M. S. & Cazzulo, J. J. Citrate synthase from Crithidia fasciculata: Inhibition by adenine nucleotides and suramin. *Comp. Biochem. Physiol. Part B Comp. Biochem.* **72**, 165–168 (1982).

596. Xu, M. *et al.* Mitochondrial enzymes and citrate transporter contribute to the aluminium-induced citrate secretion from soybean (Glycine max) roots. *Funct. Plant Biol.* **37**, 285 (2010).

597. Rothschild, B. M., Tanke, D. H., Helbling, M. & Martin, L. D. Epidemiologic study of tumors in dinosaurs. *Naturwissenschaften* **90**, 495–500 (2003).

598. Curto, R., Sorribas, A. & Cascante, M. Comparative characterization of the fermentation pathway of Saccharomyces cerevisiae using biochemical systems theory and metabolic control analysis: Model definition and nomenclature. *Math. Biosci.* **130**, 25–50 (1995).

599. Lane, A. N., Higashi, R. M. & Fan, T. W. NMR and MS-based Stable Isotope-Resolved Metabolomics and Applications in Cancer Metabolism. *TrAC - Trends Anal. Chem.* (2018). doi:S0165993618302036

8. Publications

8.1 Chapter 1

HepatoDyn: A Dynamic Model of Hepatocyte Metabolism That Integrates ^{13}C Isotopomer Data

Carles oguet, Silvia Marin, Vitaly A Selivanov, Eric anchon, ai- ang aul ee, oan uinovart, edro de Atauri, Marta Cascante.

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

HepatoDyn: A Dynamic Model of Hepatocyte Metabolism That Integrates ^{13}C Isotopomer Data

Carles Foguet[1,2], Silvia Marin[1,2], Vitaly A. Selivanov[1,2], Eric Fanchon[3], Wai-Nang Paul Lee[4], Joan J. Guinovart[1,5], Pedro de Atauri[1,2]*, Marta Cascante[1,2]*

1 Department of Biochemistry and Molecular Biomedicine, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain, **2** Institute of Biomedicine of Universitat de Barcelona (IBUB) and Associated Unit to CSIC, Barcelona, Spain, **3** UGA – CNRS, TIMC-IMAG UMR 5525, Grenoble, France, **4** Department of Pediatrics, Los Angeles Biomedical Research Institute, Torrance, California, United States of America, **5** Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

* pde_atauri@ub.edu (PdA); martacascante@ub.edu (MC)

CrossMark
click for updates

OPEN ACCESS

Citation: Foguet C, Marin S, Selivanov VA, Fanchon E, Lee W-NP, Guinovart JJ, et al. (2016) HepatoDyn: A Dynamic Model of Hepatocyte Metabolism That Integrates ^{13}C Isotopomer Data. PLoS Comput Biol 12(4): e1004899. doi:10.1371/journal.pcbi.1004899

Editor: Nathan E Lewis, University of California San Diego, UNITED STATES

Received: December 24, 2015

Accepted: April 5, 2016

Published: April 28, 2016

Copyright: © 2016 Foguet et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: CF was supported by European commission as part of the Biohealth Computing Erasmus Mundus Program, by "La Caixa" foundation as part of the program "Becas de la Caixa para estudios de doctorado en universidades españolas" and by an internship grant funded by "Centre National de la Recherche Scientifique". MC was supported by MINECO-European Commission FEDER funds – "Una manera de hacer Europa" (SAF2014-56059-R), by the European Commission (Etherpaths, FP7-KBBE-222639) and by the Agència de Gestió d'Ajuts

Abstract

The liver performs many essential metabolic functions, which can be studied using computational models of hepatocytes. Here we present HepatoDyn, a highly detailed dynamic model of hepatocyte metabolism. HepatoDyn includes a large metabolic network, highly detailed kinetic laws, and is capable of dynamically simulating the redox and energy metabolism of hepatocytes. Furthermore, the model was coupled to the module for isotopic label propagation of the software package IsoDyn, allowing HepatoDyn to integrate data derived from ^{13}C based experiments. As an example of dynamical simulations applied to hepatocytes, we studied the effects of high fructose concentrations on hepatocyte metabolism by integrating data from experiments in which rat hepatocytes were incubated with 20 mM glucose supplemented with either 3 mM or 20 mM fructose. These experiments showed that glycogen accumulation was significantly lower in hepatocytes incubated with medium supplemented with 20 mM fructose than in hepatocytes incubated with medium supplemented with 3 mM fructose. Through the integration of extracellular fluxes and ^{13}C enrichment measurements, HepatoDyn predicted that this phenomenon can be attributed to a depletion of cytosolic ATP and phosphate induced by high fructose concentrations in the medium.

Author Summary

Despite the key role of hepatocytes in carbohydrate and lipid homeostasis, available dynamic models of hepatocyte metabolism tend to be limited to a single pathway and/or are based on assumptions of constant concentrations of key metabolites involved in redox and energy metabolism (ATP, NAD, NADPH etc.). Furthermore, most dynamic models are unable to integrate information from ^{13}C based experiments. ^{13}C based experiments allow us to infer the relative activity of alternative pathways and hence are highly useful

Universitaris i de Recerca (AGAUR) – Generalitat de Catalunya (2014SGR1017). MC acknowledges the support received through the prize "ICREA Academia" for excellence in research, funded by ICREA foundation – Generalitat de Catalunya. WNPL was partially supported by NIH/NCCAM Center for Excellence in Pancreatic Diseases grant P01 AT003960-01. JJG was supported by MINECO-European Commission FEDER funds – "Una manera de hacer Europa" (SAF2014-54525-P). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

for indicating flux distributions. To overcome these limitations, we developed HepatoDyn, a dynamic model of hepatic metabolism. HepatoDyn uses a large metabolic network including key pathways such as glycolysis, the Krebs cycle, the pentose phosphate pathway and fatty acid metabolism, and dynamically models the concentrations of metabolites involved in the redox and energy metabolism of hepatocytes. In addition, the model was coupled to the label propagation module of the package IsoDyn, allowing it to integrate data from ^{13}C based experiments to assist in the parametrization process. These features make HepatoDyn a powerful tool for studying the dynamics of hepatocyte metabolism.

Introduction

No other organ performs as many physiological functions as the liver. The liver is responsible for detoxification, bile acid and blood proteins synthesis, plays a key role in the inflammatory response and, above all, it is a key regulator of glucose and lipid homeostasis in blood. Most of its functions and properties can be linked to hepatocytes, the most abundant cell type in liver, and therefore hepatocytes are often used as a model to study liver function and pathologies [1]. Accordingly, computational modelling of hepatocyte metabolism has received a great deal of interest.

Recently, genome scale metabolic reconstructions based on stoichiometric modelling techniques have been successfully used to model hepatocyte metabolism [2–4]. However, stoichiometric models provide a static picture of metabolism based on mass balance equations and the assumption that the system is under a strict steady state. In these models each reaction step is described by only one parameter, its steady state flux [5]. The alternative is to use dynamic metabolic models, usually referred to as kinetic models. They are based on building a system of ordinary differential equations (ODEs), with kinetic laws describing transport and chemical transformations for each reaction-step and parameters describing biochemical and biophysical constraints. Kinetic modelling has two main advantages over stoichiometric based modelling; firstly, it is capable of performing dynamic simulations, that is to say, it can predict the variation in metabolite concentrations and fluxes over time outside of the steady state. Secondly, it can follow the global effects of constraints emerging from the specific kinetic properties of enzymes, post-translational modifications and regulatory circuits, thus revealing the complex regulation of the system. Over the years, multiple kinetics models of hepatocyte metabolism have been developed [6–11]. The main limitation of kinetic models is that they are complex to build and parametrize. Due to this complexity, kinetic models of hepatocyte metabolism available in the literature contain only a small number of reactions and, with some exceptions [11], are often limited to a single pathway. Furthermore, with the exception of some models focused on mitochondria [8, 9], most of them assume a constant redox and energy state, which limits their application. In fact, despite the huge interest in hepatocyte metabolism, there are no models capable of adequately modelling the effects of the energy and redox dynamics on hepatocyte core metabolism. Additionally, while ^{13}C experiments have proven their usefulness in studying the metabolism of hepatocyte under metabolic steady state [12–24], there was only one kinetic model of hepatocyte capable of integrating ^{13}C data [10].

In this work, we present HepatoDyn (**Hepato**cyte **Dyn**amics) a model of hepatocyte core metabolism capable of simulating the redox (NAD/NADH, NADP/NADPH, etc.) and energy (ATP/ADP/AMP, etc.) dynamics. The model includes glycolysis, gluconeogenesis, glycogen metabolism, the pentose phosphate pathway, the Krebs cycle and fatty acid metabolism as well as reactions associated with energy and redox metabolism (respiratory chain, malate/aspartate

shuttle, glycerol phosphate shuttle, etc.). To our knowledge, no model of such size capable of dynamic redox and energy metabolism simulations exists in the literature. Furthermore, the model was coupled to the module for isotopic label propagation of the software package IsoDyn [25, 26]. This enables HepatoDyn to integrate data from ^{13}C based experiments to assist in the parametrization process, regardless of whether experimental measurements correspond to an isotopic steady state. The latter is a key feature because the levels of isotopic label enrichment are often a non-steady phenomenon with long transition times [27]. Therefore, HepatoDyn is a very powerful tool capable of taking advantage of both the constraints derived from a detailed tissue-specific kinetic model and data derived from ^{13}C based experiments to simulate hepatocytes.

In the last decades there has been a significant increase in fructose in our diets [28] and accordingly there is great interest in studying the potential effects of fructose in the metabolism [29–32]. To date, fructose-rich diets have been associated with many adverse metabolic conditions, such as nonalcoholic fatty liver disease, insulin resistance and obesity [28, 33, 34], most of which are directly or indirectly related to abnormal hepatocyte function. Therefore, we used HepatoDyn to study the short-term response of hepatocyte metabolism to different concentrations of fructose.

Materials and Methods

Experimental Methods

Materials. [1,2-$^{13}C_2$]D-glucose (>99% enriched) and [U-$^{13}C_6$]D-fructose (>99% enriched) were purchased from Isotec (Miamisburg, OH, USA), and other reagents used were from Sigma-Aldrich Company (St. Louis, MO, USA).

Animals. 180–200 g male Wistar rats were used. They were maintained in a 12 h:12 h light-dark cycle with free access to standard laboratory rat chow pellets (Panlab) and water. Animals were deprived of food 24 h prior to hepatocyte isolation. Experiments were conducted according to guidelines accepted by the University Animal Care and Use Committee. Appropriate measures were taken to minimize pain or discomfort in the animals.

Preparation of cells and incubation. Suspensions of isolated parenchymal liver cells were prepared from 24-h starved animals as described previously [35]. Cell suspensions were incubated at 37°C with gassing and continuous shaking (160 strokes/min) for 2 h with Krebs–Ringer bicarbonate buffer of pH 7.4 containing glucose and fructose. At the end of the incubations, cells were centrifuged and the incubation media and cell pellets were obtained.

Measurement of metabolites. Glycogen content from cell pellets and glucose and lactate concentrations in incubation media were determined as described previously [19].

Gas chromatography/ mass spectrometry sample processing and analysis. Incubation media were processed for isolation of lactate, glucose, and glutamate using previously established methods [36, 37]. To analyse fructose isotopologues distribution, lyophilized incubation medium was treated with 0.5 N sodium borohydride in methanol for 2 h at room temperature, causing both fructose and glucose to be transformed to sorbitol. The resulting sorbitol was then isolated by ion exchange chromatography as described for glucose [37]. Glycogen was isolated from cell pellets as described previously [19]. Once isolated, glucose from the medium or from hydrolysed glycogen, as well as lactate, glutamate and sorbitol were derivatized for gas chromatography/mass spectrometry (GC/MS) analysis [36, 38, 39]. In the case of sorbitol, it was derivatized to its hexaacetate derivative according to a modification of the method described by Wolfe [40]. A mass selective detector HP 5973 equipment coupled to a gas chromatograph HP 6890 was used for all the metabolites as described elsewhere [36, 38, 39]. The GC/MS method for sorbitol analysis was the same as that for glucose analysis. Chemical ionization was used to obtain the molecular ion (C1-C6) of the glycogen or medium glucose

molecules at m/z 328, and the same for the lactate molecule (C1-C3) at m/z 328 and sorbitol molecule (C1-C6) at m/z 375. Electron impact ionization was used to characterize the isotopologue fractions of C1-C4 (m/z 242) and C3-C6 (m/z 187) glycogen glucose fragments, as well as C2-C4 (m/z 152) and C2-C5 (m/z 198) glutamate fragments.

Spectral data were corrected using regression analysis to extract natural ^{13}C enrichment from results [41]. Measurement of ^{13}C label distribution determined the different relative distribution percentages of the isotopologues, m0 (without any ^{13}C labels), m1 (with one ^{13}C), m2 (with two ^{13}C), etc.

Building the model

A metabolic network, including those pathways deemed necessary to accurately and dynamically simulate the core metabolism of rat hepatocytes in the study conditions, was constructed based on pathways that have been reported in the literature to be active in hepatocytes [42, 43].

Each reaction in the metabolic network was assigned a kinetic law. Kinetic laws describe the dependence of each reaction flux on metabolite concentrations. They take into account the affinity of substrates and products, the reaction mechanism and the effect of activators and inhibitors on reaction fluxes. The kinetic laws used were mostly derived from existing kinetic laws described in the literature [6, 11, 44]. The exceptions were the kinetic laws for aldolase activity, which catalyses eight related elementary reactions, which were built as described in the Supplementary Material (S1 Text).

Kinetic laws are integrated with the metabolic network topology, described by the stoichiometric matrix (N), to build a system of ordinary differential equations (ODEs) that predict the evolution of metabolite concentrations, and by extension the evolution of reaction fluxes, over time. Because fluxes are provided in units of mmol *per* cell *per* minute, but ODEs are solved in units of mmol *per* litre *per* minute, in order to build the ODEs, the cell number and the volume of the compartment at which each metabolite is located must also be taken into account. Therefore, the system of ODEs can be written as:

$$\frac{dc[t]}{dt} = N \cdot j(c[t], p) \cdot \frac{ncell}{vol} \quad (1)$$

Where j is a vector of reaction fluxes, which is a function of the vector of metabolite concentrations (c[t]) in mM, and a vector model parameter (p) as defined by the kinetic laws used in the model, *ncell* is the cell number and *vol* is a vector containing the volumes of the compartment at which each metabolite is localized in litres.

In reversible reactions, forward and reverse reaction rates are computed separately with different kinetic laws, albeit sharing most of the parameters. Additionally, the fluxes of invisible reactions, that is to say, reactions that can propagate labelled carbons even though they do not change the overall concentrations of metabolites, are also computed [10]. This is necessary in order to fully simulate the propagation of ^{13}C.

To simulate the propagation of ^{13}C through the metabolic network, fluxes are decomposed into isotopomer fluxes. Then, an ODE system is built using the algorithms from IsoDyn [25, 26]. The resulting ODE accounts for concentrations of all isotopomers, isomers with ^{13}C substitution in specific carbon positions [24]. To avoid unnecessary complexity, isotopomers are not simulated for those metabolites where, according to the defined metabolic network, ^{13}C from labelled substrates cannot be propagated. The process is briefly summarized in Fig 1.

The system of differential equations for metabolite and isotopomer concentrations is solved to predict metabolic fluxes, metabolic concentrations and isotopomer concentrations from the initial time to the defined end time.

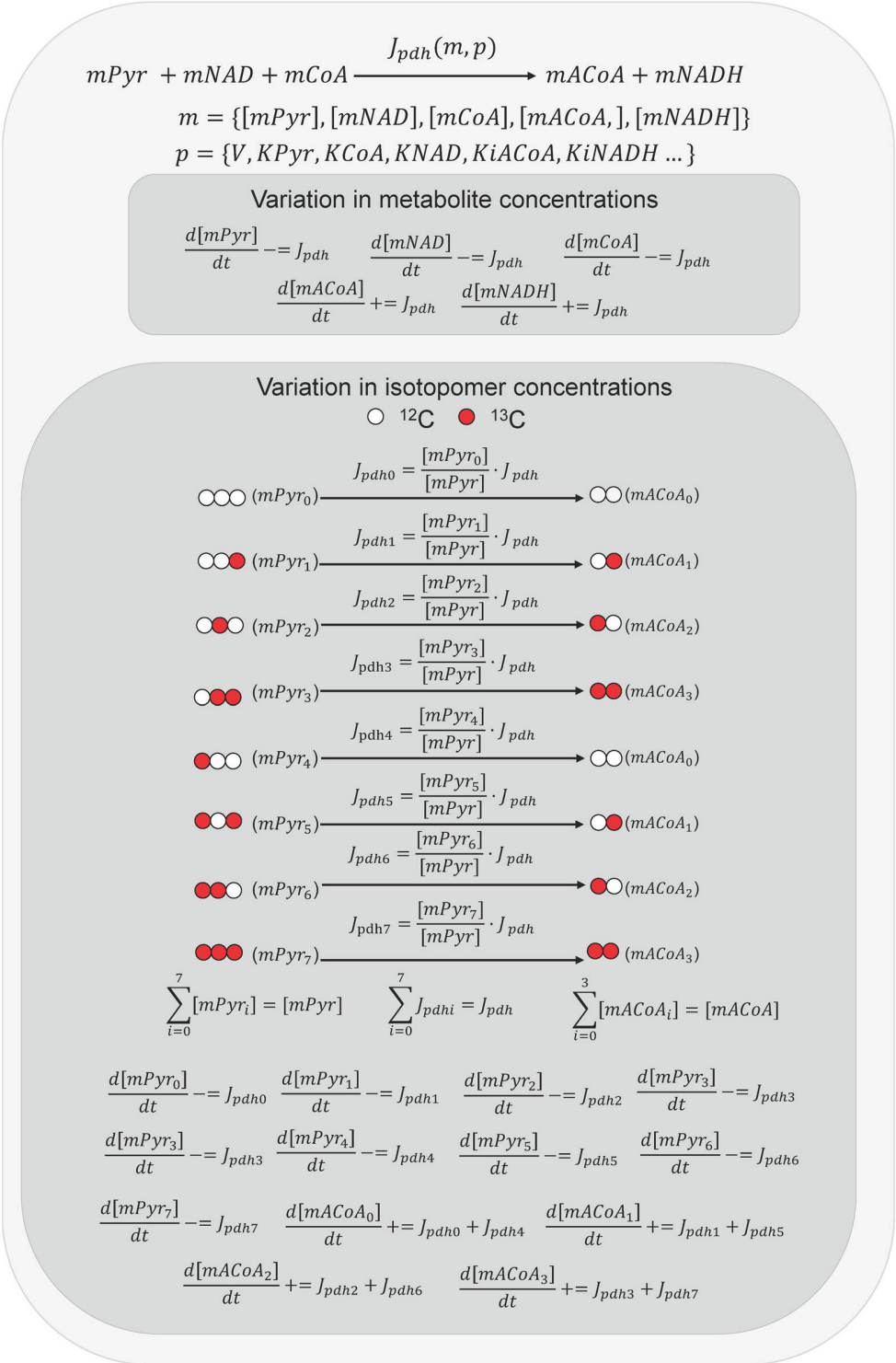


Fig 1. Example of how ODEs are automatically built for isotopomers and metabolites consumed or produced by the pyruvate dehydrogenase catalysed reaction (PDH). PDH irreversibly transforms mitochondrial pyruvate (mPyr), NAD (mNAD), and coenzyme A (mCoA) into mitochondrial acetyl-CoA (mACoA) and NADH (mNADH). The system of differential equations is solved taking into account all equations for total concentrations of metabolites and for concentrations of isotopomers. From the previous

step in the simulation, the PDH flux (J_{pdh}) is computed, which is a function of the concentrations of the reactants and products (m) and the kinetic parameters of PDH (p). For the ODEs describing the concentration of metabolites the computed value is added (+ =) and subtracted (- =) for products and substrates, respectively. For the ODE describing a particular isotopomer, the flux value is scaled according to the relative abundance of the isotopomer for the substrate ($mPyr_i$) and the resulting scaled flux (J_{PDHi}) is added (+ =) and subtracted (- =) to $d[mACoA_i]/dt$ and $d[mPyr_i]/dt$, respectively. Isotopomers are not simulated for CoA, NAD or NADH because it is assumed that ^{13}C from labelled substrates does not propagate to such metabolites.

doi:10.1371/journal.pcbi.1004899.g001

Model predictions are for isotopomers but experimental measurements refer to isotopologues (or mass isotopomers), isomers with a specific number of ^{13}C substitutions [24]. Thus, the resulting concentrations of isotopomers are converted into fractions of isotopologues, by adding up all isotopomers that correspond to each isotopologue and dividing by the total concentration of each metabolite (S1 Fig). The fractions of such isotopologues can then be compared with the experimental measurements obtained with GC coupled to MS.

Parameterization

Kinetic parameters representing enzyme activity (V_{max} or equivalent) were fitted to the experimental data. For this process V_{max} from the reverse reaction rate in reversible reactions are assumed to be a function of the V_{max} of the forward reaction and of the equilibrium constant as described by the Haldane relationship [44]. To further reduce the number of parameters fitted, enzyme activities catalysing sequential reactions with no ramifications (the so called reactions chains) were fitted as a group. This is because in reactions chains the flux through the whole chain could be determined by any of the enzyme activities involved and consequently most of the activities of enzymes constituting the chain would be unidentifiable. Furthermore, other activities known to be unidentifiable are not fitted, such as the activities of reactions that are known to operate in rapid equilibrium in physiological conditions (glucose phosphate isomerase, triose phosphate isomerase, enolase, etc.). The remaining parameters of the kinetic model were assigned based on an extensive literature search, completed with data from Brenda [45] and UniProt [46] databases.

The fitting algorithm, a variant of the basic simulated annealing algorithm [47], seeks the set of m parameters (E_z) that minimizes the objective function. The objective function (X^2) is the square deviation between the n experimentally measured values (Y_i) and simulated values (Z_i) for both isotopologue fractions and total metabolite concentrations, normalized by the experimental standard deviation (σ_i). To prevent a bias generated by very low standard deviations, a minimum threshold of 0.01 was used. Additionally, parameter sets where any metabolite reached concentrations greater than 50 mM were discarded.

$$X^2 = \sum_{i=1}^{n} \left(\frac{Y_i - Z_i(E_1, E_2, \ldots, E_m)}{\sigma_i} \right)^2 \quad (2)$$

Consequently, the fitting algorithm seeks the set of enzyme activities that minimize the difference between experimentally measured and simulated isotopologue fractions and metabolite concentrations in the experimental conditions considered.

Identifiability analysis

The fitting procedure provides one set of fitted parameters, which minimizes the objective function, and is referred to as the best fit parameter set. However, other sets of parameter values might result in similar or equal objective function values and are therefore as valid as the best fit. The range of acceptable variation in parameters was evaluated through an

identifiability analysis. Identifiability is a property that indicates whether unknown model parameters can be determined from the available experimental data. It depends both on the structure of the model and the quality and amount of experimental data. A parameter is defined as identifiable if the confidence interval for its estimated value at a given significance level is finite [48, 49].

If we define $X^2(\theta_i)$ as the optimized square deviation if parameter i is fixed to a value of θ_i and the remaining parameters being fitted θ_j are readjusted to minimize the square deviation

$$X^2(\theta_i) = \min_{\theta_{j\neq i}}[x^2(\theta_j)] \quad (3)$$

then if experimental errors are assumed to follow a normal distribution, for a parameter i, the confidence interval can be defined as:

$$\{\theta_i | X^2(\theta_i) - X^2_{bf} < \Delta\alpha\} \;\; with \;\; \Delta\alpha = X^2(\alpha, 1) \quad (4)$$

where X^2_{bf} is the best fit square deviation (optimized with no fixed parameters) and $\Delta\alpha$ is the significance threshold associated with a given significance level (α) with a Chi Square distribution with one degree of freedom. Accordingly, the upper and lower limit of the confidence intervals for a given parameter are estimated by respectively increasing and decreasing the value of the parameter until the square deviation difference obtained when optimizing the remaining parameters exceeds the threshold ($\Delta\alpha$) [48].

Additionally, intervals for system dependent variables (fluxes, metabolite concentrations and isotopologue fractions at different time points) are estimated from the maximum and minimum parameter values of confidence intervals generated during the identifiability analysis.

Results

HepatoDyn: A kinetic model capable of integrating ^{13}C based measurements

We present HepatoDyn, the first detailed model of hepatocyte core metabolism capable of dynamically simulating energy and redox metabolism. It consists of 88 reactions and 81 metabolites distributed into three compartments (extracellular, cytosolic and mitochondrial). A schematic representation of the model can be found in Fig 2 and a complete list of metabolites, reactions and compartments can be found in S1, S2 and S3 Tables, respectively.

Each reaction has an associated kinetic law and the model has a total of 470 parameters associated to kinetic laws (S4 Table). 55 of these parameters correspond to enzyme activities that were fitted to experimental data, taking parameter groups (S5 Table) into account this results in 29 independent parameters that were fitted to experimental data. To the greatest extent possible, the kinetic laws and their parameters were specific to the enzyme isoforms active in the liver.

It is worth noting, that while most of the reactions included in HepatoDyn are also present in genome scale reconstructions of hepatocyte metabolism [2–4], HepatoDyn includes complete kinetic laws and regulatory loops, which allow for dynamic and regulatory studies. Nevertheless, HepatoDyn also has 2 reactions that are absent in genome scale reconstructions of hepatocyte. Specifically, the reactions aldolase 3 (Fru16bP + Gra ↔ Fru1P + GraP) and transketolase 3 (Fru6Pa + Rib5P ↔ E4P + Sed7P). Those reactions emerge because the enzymes aldolase and transketolase allow multiple combinations of substrates and products. Additionally, HepatoDyn also incorporates the channelling of hexose phosphates to glycogen in the form of two separate pools of hexose phosphates, *a* and *b*, as previously described in the literature [10].

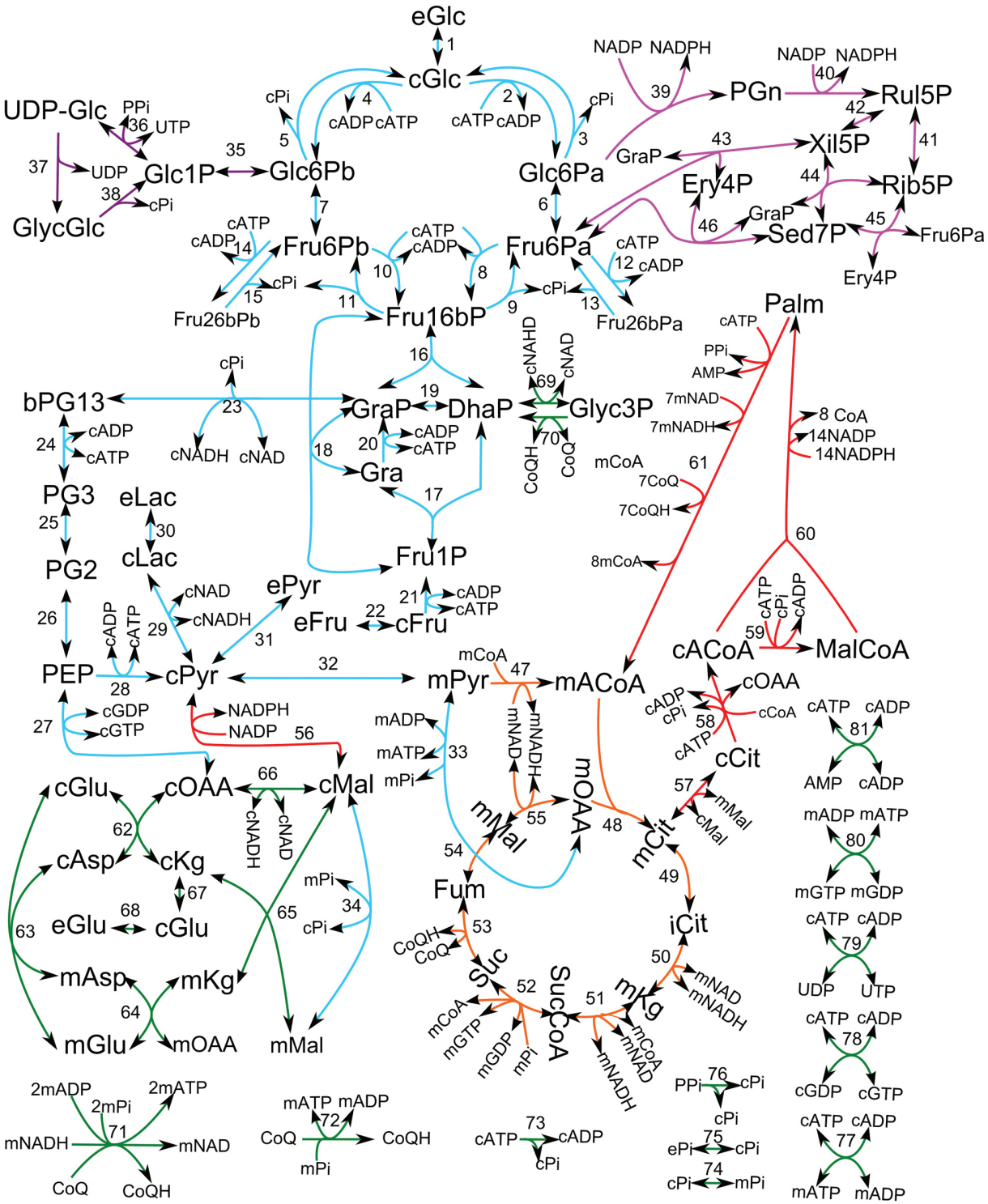


Fig 2. Schematic representation of the metabolic network used in the model. In this representation, reactions associated with the glycolytic and gluconeogenic pathways are coloured in blue, reactions associated with glycogen metabolism are coloured in purple, reactions associated with the pentose phosphate pathway are coloured in pink, reactions associated with the Krebs cycle are coloured in orange, reactions associated with fatty acid metabolism are coloured in red and other reactions associated with redox and energy metabolism are coloured in green. Specifically, the reactions id of each reaction represented are 1:glctr, 2: gka, 3 g6pasea, 4: gkb, 5: g6paseb, 6: gpia, 7: gpib, 8: pfkla1, 9: fbasea1, 10: pfklb1, 11: fbaseb1, 12: pfkla2, 13: fbasea2, 14: pfklb2, 15: fbaseb2, 16: aldo1, 17: aldo2, 18: aldo3, 19: tim, 20: trik, 21: fruhk, 22: frutr, 23: gapdh, 24: pgk, 25: pgm, 26: eno, 27: pepck, 28: pk, 29: ldh, 30:

lactr, 31: pyrtr, 32: mpyrtr, 33: pc, 34: dic, 35: pglm, 36: ugt, 37: gs, 38: gp, 39: g6pdh, 40: pgndh, 41: rpi, 42: rul5pepi, 43: tk1, 44: tk2, 45: tk3, 46: ta, 47: pdh, 48: cs, 49: aco, 50: idh, 51: kdh, 52: scs, 53: sdh, 54: fh, 55: mmdh, 56: malic, 57: citmtr, 58: citly, 59: acoacar, 60: fasyn, 61: box, 62: aatc, 63: aspglumtrans, 64: aatm, 65: malkgmtrans, 66: cmdh, 67: transa, 68: glutr, 69: glyc3pcdh, 70: glyc3pmdh, 71: nadhdh, 72: coqhoxi, 73: atpase, 74: pimtr, 75: pitr, 76: ppase, 77: atpmtrans, 78: cndk1, 79: cndk2, 80: mndk and 81 adk. Invisible reactions are not shown for clarity. The full lists of metabolites and reactions can be found on S1 and S2 Tables respectively.

doi:10.1371/journal.pcbi.1004899.g002

The kinetic model, fully parametrized, can be found in SBML format in the Supplementary Material (S1 XML and S2 XML).

In addition, HepatoDyn is capable of simulating the propagation of ^{13}C from isotopically labelled substrates to metabolic intermediaries and products. This allows HepatoDyn to integrate isotopologue enrichment measurements from ^{13}C based experiments greatly enhancing the predictive capabilities of the model.

HepatoDyn is provided in the Supplementary Material as a C++ program (S1 Software).

Analysing the effects of fructose on hepatocyte metabolism using HepatoDyn

The liver has a high capacity to metabolize fructose, it is estimated that up to 50% of fructose ingested is metabolized by hepatocytes [50]. Fructose metabolism in hepatocytes consists of phosphorylation of fructose to fructose 1-phosphate by fructokinase and the split of this metabolite by the liver aldolase isoform (aldolase B) into dihydroxyacetone-phosphate and glyceraldehyde, with the latter metabolite being phosphorylated by triokinase into glyceraldehyde 3-phosphate. Because fructose enters at the level of triose phosphate, bypassing the highly regulated glucokinase and phosphofructokinase steps of glycolysis, fructose uptake is largely unregulated. Consequently, the limiting step in fructose metabolism is assumed to be fructose uptake by hepatocytes, which is heavily dependent on the extracellular concentration of fructose due to the low affinity of the proteins mediating fructose transport into hepatocytes, GLUT2 and other carriers like GLUT8 [51–53].

As a proof of concept of the capabilities of HepatoDyn, we applied it to study the short term response of hepatocytes to incubation with 20mM glucose supplemented by either 3mM fructose or 20mM fructose. These concentrations were chosen because our experimental data showed that hepatocytes responded quite differently to them. While incubation with 20mM glucose supplemented with 3mM fructose resulted on a rapid glycogen accumulation, incubation with 20mM glucose supplemented with 20mM fructose resulted on almost no glycogen accumulation (Fig 3.A). While it has been reported that supplementation with low concentrations of fructose favours glycogen accumulation [19, 29, 54], the fact that supplementation with high fructose concentrations inhibits glycogen accumulation was not known. Furthermore, isotopologue analysis indicated that in the second condition, unlike the first condition, almost no ^{13}C from labelled glucose was propagated to lactate (Fig 3.B). In both conditions lactate and glucose were produced from fructose at a similar rate. Hence it was an interesting case of study.

Specifically, HepatoDyn was used to integrate experimental measurements derived from rat hepatocytes incubated for 2 h with the following media: 20 mM glucose 50% enriched in [1,2-$^{13}C_2$]-glucose and 3 mM fructose (condition A1), 20 mM glucose and 3 mM fructose 50% enriched in [U-$^{13}C_6$]-fructose (condition A2) and 20 mM glucose 50% enriched in [1,2-$^{13}C_2$]-glucose and 20 mM fructose (condition B). The experimental data for condition A1 had been published previously [19]. This integration was achieved using the experimental measurements of extracellular concentrations and isotopologue fractions as input to fit the 29 independent parameters associated to enzyme activities in the model assuming that the

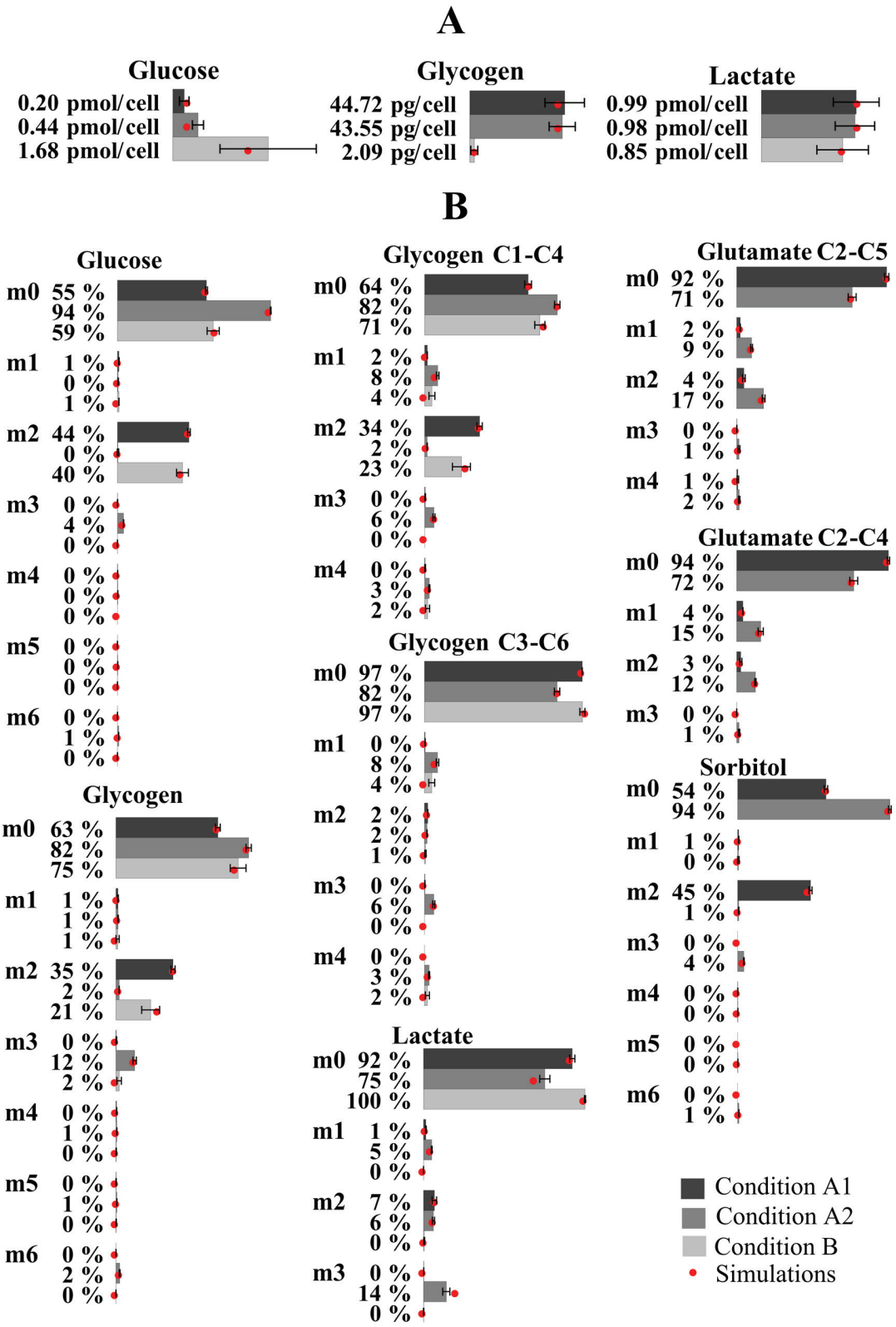


Fig 3. Bar graphs representing the experimentally determined metabolite productions (3.A) and isotopologue fractions (3.B) in experimental conditions. Measurements were taken after incubating hepatocytes for 2 hours with 20 mM glucose 50% enriched in [1,2-$^{13}C_2$]-glucose and 3 mM fructose (condition A1), 20 mM glucose and 3 mM fructose 50% enriched in [U-$^{13}C_6$]-fructose (condition A2) and 20 mM glucose 50% enriched in [1,2-$^{13}C_2$]-glucose

and 20 mM fructose (condition B). The red dot indicates the value fractions simulated by HepatoDyn using the best fit parameter set. Results of the isotopologue fractions are reported as m0, m1, m2, etc. where m0, m1, m2. . . indicate the number of ^{13}C atoms in the isotopologue fractions of a given metabolite.

doi:10.1371/journal.pcbi.1004899.g003

enzyme activities, normalized by cell number (S2 Fig), were equivalent in the three conditions. Consequently, the fitting algorithm identifies a single set of parameters that allows reproduction of the three experimental conditions. It is worth noting that because conditions A1 and A2 only differ in the labelling pattern of substrates, the predicted fluxes and concentrations values will be the same in both conditions. The resulting values of the fitted parameters can be found in S6 Table. The resulting metabolites concentrations for condition A1/A2 and condition B can be found on S3 and S4 Figs respectively. The resulting fluxes for condition A1/A2 and condition B can be found on S5 and S6 Figs respectively. The resulting isotopologue fractions for key metabolites in condition A1, A2 and B can be found on S7, S8 and S9 Figs respectively. A comparison between the experimentally measured metabolite concentrations and isotopologue fractions and those simulated by the model with the best fit parameter set can be found in Fig 3.

High concentrations of fructose have been shown *in vivo* and *in vitro* to result in the depletion of ATP and phosphate in hepatocytes [52, 55]. This occurs due to an accumulation of fructose 1-phosphate caused by the elevated fructokinase activity [52, 55]. This phenomenon was predicted by HepatoDyn. The model predicted that a persistent cytosolic ATP and phosphate depletion would occur with an extracellular concentration of 20 mM fructose (Fig 4). This is mainly caused by an accumulation of fructose 1-phosphate, although the depletion can also be partially attributed to the accumulation of some other phosphorylated metabolites. In this context, the low glycogen synthesis observed at 20mM glucose supplemented with 20 mM fructose can be attributed to the depletion of cytosolic ATP and phosphate. Likewise, the almost non-existent propagation of ^{13}C from glucose to lactate under this condition can mainly be attributed to the low glucokinase and phosphofructokinase activities caused by ATP depletion. Conversely, at 20mM glucose supplemented with 3 mM fructose, a persistent accumulation of fructose 1-phosphate does not occur. Accordingly, under this condition, ATP and phosphate are not persistently depleted (Fig 4).

Identifiability

Overall, 25 of the 29 independent parameters were identifiable with at least 95% confidence. This remarkable degree of identifiability can be attributed to the numerous feedback regulations through the redox and energy balances (ATP/ADP, NADH/NAD, etc.), the use ^{13}C data and the integration of data from multiple metabolic conditions.

Concerning the non-identifiable parameters, the non-identifiability of the aldolase activity and the activities involved in the lactate production and malate aspartate shuttle reaction chains can be attributed to the fact that the reactions associated to those pathways are predicted to be close to the equilibrium in experimental conditions, hence the system is fairly insensitive to the value of the enzyme activities associated to them. On the other hand, the non-identifiability of the citrate synthase activity arises because in our model the flux through the citrate synthase reaction can depend solely on the two activities upstream, pyruvate dehydrogenase and β-oxidation, which catalyse the production of acetyl-CoA, the substrate of citrate synthase.

Compared to parameters, fluxes and to a lesser extent concentrations, show a much narrower range of variation (S3, S4, S5 and S6 Figs). This can serve as an indication of robustness,

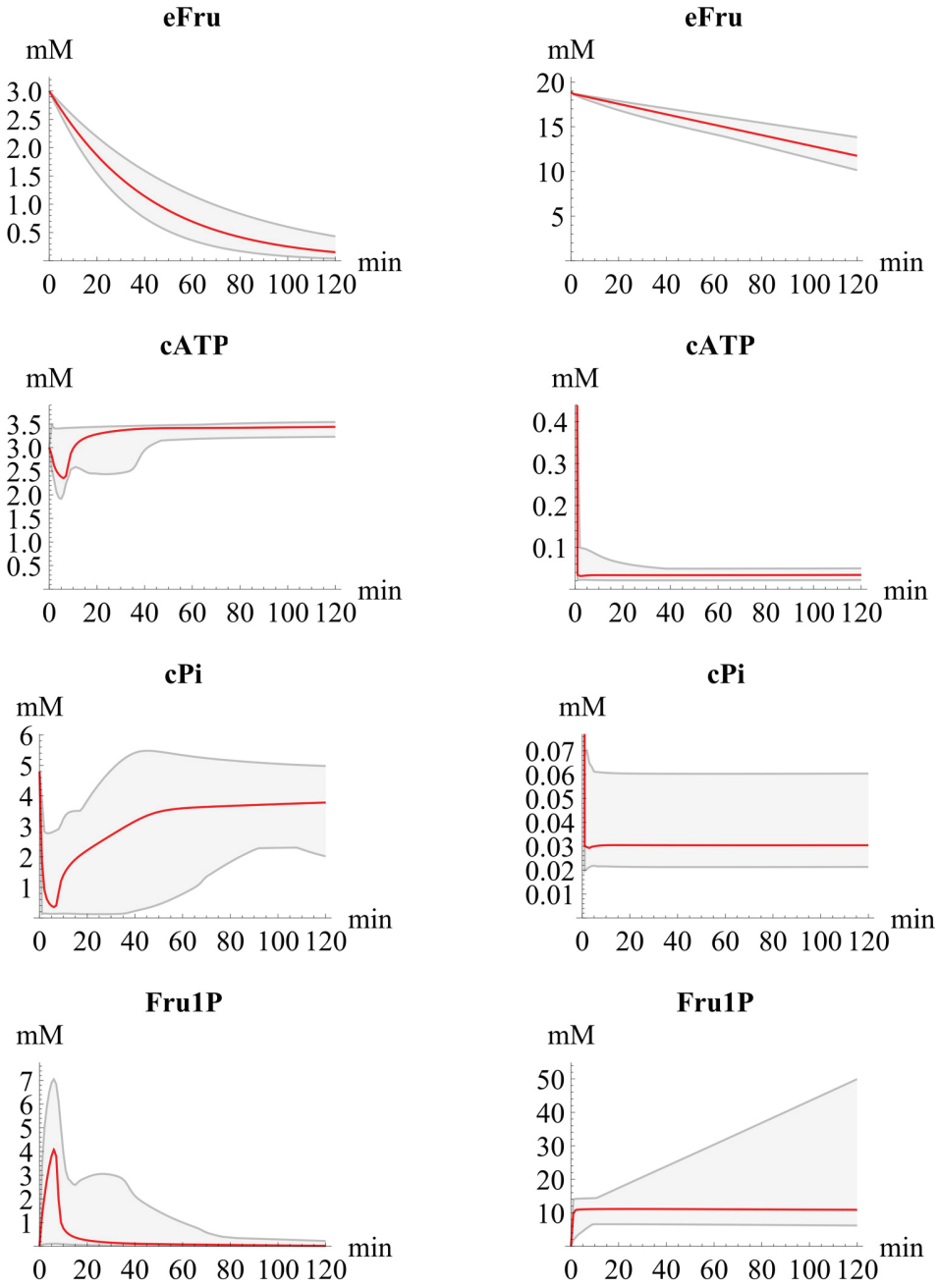


Fig 4. Plot of the simulated concentrations over time for extracellular fructose (eFru), fructose 1-phosphate (Fru1P), cytosolic phosphate (cPi) and cytosolic ATP (cATP). Specifically, the simulated concentrations in hepatocytes incubated with 20 mM glucose and 3 mM fructose (conditions A1 and A2, described in the main text) or 20 mM glucose and 20 mM fructose (condition B, described in the main text) are shown. The red plot indicates the values predicted with the best fit parameter set and the grey area indicates the estimated range of variations taking parameter sets within the 95% confidence intervals derived from the identifiability analysis.

doi:10.1371/journal.pcbi.1004899.g004

the capacity of the system to maintain its functional properties in the face of external and internal perturbations and uncertainty [56].

Interestingly, fluxes associated with the pentose phosphate pathway and fatty acid synthesis have fairly low upper bounds in both conditions (incubation with 3 mM fructose and 20 mM glucose and incubation with 20 mM fructose and 20 mM glucose). This is consistent with hepatocytes extracted from fasted rats, as they can be expected to have low activity in fatty acid synthesis, and thus only need to generate a small amount of reductive potential (NADPH) to maintain cell functions. However, with longer incubation times, an increase in the fatty acid synthesis and pentose phosphate pathway activities and fluxes should be observed as fructose is known to increase the expression of key lipogenic enzymes in hepatocytes[28, 57, 58].

It is also worth noting that the identifiability analysis further reinforces the notion that hexose phosphate metabolism in hepatocytes is compartmentalized into two different pools as previously reported [10]. This is because most of enzyme activities present in both hexose pools have a lower bound above 0 in the confidence interval, suggesting that the separation of hexose phosphates into two separate pools must be taken into account to adequately simulate the experimental conditions. If there was no compartmentalization, all activities present in both pools would have a lower bound of 0 because they would be made redundant by the activities in the other pool.

Discussion

Metabolic modelling is based on applying constraints to limit the space of feasible solutions for system variables, such as reaction fluxes and metabolite concentrations. Constraints can arise from different components of the model including reaction stoichiometry and kinetic laws, and from the experimental measurements integrated by the model. Consequently, the use of a highly complete metabolic network, including the fundamental balances affecting redox and energy metabolism (ATP/ADP, NAD/NADH, etc.), serve as an important set of constraints. Furthermore, the inclusion of highly detailed kinetic laws and parameters derived from the literature further constrains the solution space. For instance, important constraints that emerge from kinetic laws are regulatory circuits, such as fructose 6-phosphate inhibiting glucokinase or fructose-1-phosphate disrupting such inhibition [59–61]. Other important constraints that emerge from the kinetic laws are thermodynamics constraints, which are in the form of equilibrium constants. Finally, integrating ^{13}C based data provides additional constraints such as labelling enrichments which provide information on ratios among fluxes through alternative metabolic pathways. While numerous kinetic models of hepatocytes exist in the literature [6–11], HepatoDyn is the first that is capable of integrating all the aforementioned constraints in a single model.

As a proof of concept of the capabilities of the model, we applied HepatoDyn to study the metabolic effects of high fructose concentrations on rat hepatocytes. Experimental data showed that hepatocytes behaved quite differently depending on whether they were incubated with 20mM Glucose supplemented with either 3 mM fructose or 20 mM fructose. Using HepatoDyn, we managed to find a physiological explanation for this behaviour, which involved the rapid and persistent depletion of cytosolic ATP and phosphate at 20 mM fructose, which was in accordance with information reported in the literature [52, 55]. This phenomenon has a strong dynamic component, is dependent on the kinetic properties of enzymes and on the balances involved in energy metabolism. Additionally, it may be relevant for understanding the potential adverse effects of fructose-rich diets. This is because ATP depletion impairs protein synthesis and induces inflammatory and prooxidative changes and thus, in a fructose-rich diet,

this depletion might result in increased susceptibility of hepatocytes to injury leading to adverse hepatic conditions such as nonalcoholic fatty liver disease [62].

Furthermore, HepatoDyn has countless applications that go beyond studying the effects of fructose. For instance, HepatoDyn can be used to study liver centric metabolic diseases such as diabetes. Given that HepatoDyn is capable of dynamically simulating the redox and energetic state of hepatocytes, it can be used to better understand the mechanism of action of anti-diabetic drugs like metformin which target the energetic and redox metabolism [63] as well as identifying new drug targets. HepatoDyn can also be used to study the relative contribution of different reactions to redox and energy balances in different conditions. Therefore, potential applications of HepatoDyn can be to analyse the ATP consumption or production associated to different pathways or the relative contribution of the glycerol phosphate shuttle and the malate aspartate shuttle to the transfer of reducing equivalents between the cytosol and the mitochondrial matrix. Last, but not least, new reactions can easily be added to HepatoDyn provided kinetic mechanisms and kinetic information such as affinity constants or inhibition constants are known for the enzymes catalysing those reactions. Likewise, through the modification of reactions and kinetic laws specific to hepatocytes, HepatoDyn can be adapted to other cell types.

Supporting Information

S1 Fig. Example of how concentrations of pyruvate's isotopomers ($[Pyr_i]$) are converted to isotopologue fractions ($fPyr_{mz}$). This is achieved by adding up all isotopomers that correspond to each isotopologue of pyruvate and dividing by the total concentration of pyruvate.
(TIFF)

S2 Fig. Bar graph representing the concentration of cells in each experimental condition described in the main text (A1, A2 and B).
(TIFF)

S3 Fig. Plots of the simulated concentrations over time for hepatocytes incubated with 3 mM fructose and 20 mM glucose (conditions A1 and A2, described in the main text). The red plot indicates the values predicted with the best fit parameter set and the grey area indicates the estimated range of variations taking parameter sets within the 95% confidence intervals derived from the identifiability analysis.
(TIFF)

S4 Fig. Plots of the simulated concentrations over time for hepatocytes incubated with 20 mM fructose and 20 mM glucose (condition B, described in the main text). The red plot indicates the values predicted with the best fit parameter set and the grey area indicates the estimated range of variations taking parameter sets within the 95% confidence intervals derived from the identifiability analysis.
(TIFF)

S5 Fig. Plots of the simulated fluxes over time for hepatocytes incubated with 3 mM fructose and 20 mM glucose (conditions A1 and A2, described in the main text). The red plot indicates the values predicted with the best fit parameter set and the grey area indicates the estimated range of variations taking parameter sets within the 95% confidence intervals derived from the identifiability analysis.
(TIFF)

S6 Fig. Plots of the simulated fluxes over time for hepatocytes incubated with 20 mM fructose and 20 mM glucose (condition B, described in the main text). The red plot indicates the

values predicted with the best fit parameter set and the grey area indicates the estimated range of variations taking parameter sets within the 95% confidence intervals derived from the identifiability analysis.
(TIFF)

S7 Fig. Plots of the simulated isotopologue fractions over time for Glucose, Glycogen, Sorbitol, Lactate and Glutamate for hepatocytes incubated with 20 mM glucose 50% enriched in [1,2-$^{13}C_2$]-glucose and 3 mM fructose (condition A1, described in the main text). The isotopologue fractions of sorbitol refer to sorbitol derived from simulated Glucose and simulated Fructose. The red plot indicates the values predicted with the best fit parameter set and the grey area indicates the estimated range of variations taking parameter sets within the 95% confidence intervals derived from the identifiability analysis. Isotopologue fractions are reported as m0, m1, m2, etc. where m0, m1, m2. . . indicate the number of ^{13}C atoms in the fraction.
(TIFF)

S8 Fig. Plots of the simulated isotopologue fractions over time for Glucose, Glycogen, Sorbitol, Lactate and Glutamate for hepatocytes incubated with 20 mM glucose and 3 mM fructose 50% enriched in [U-$^{13}C_6$]-fructose (condition A2, described in the main text). The isotopologue fractions of sorbitol refer to sorbitol derived from simulated Glucose and simulated Fructose. The red plot indicates the values predicted with the best fit parameter set and the grey area indicates the estimated range of variations taking parameter sets within the 95% confidence intervals derived from the identifiability analysis. Isotopologue fractions are reported as m0, m1, m2, etc. where m0, m1, m2. . . indicate the number of ^{13}C atoms in the fraction.
(TIFF)

S9 Fig. Plots of the simulated isotopologue fractions over time for Glucose, Glycogen, Sorbitol, Lactate and Glutamate for hepatocytes incubated with 20 mM glucose 50% enriched in [1,2-$^{13}C_2$]-glucose and 20 mM fructose (condition B, described in the main text). The isotopologue fractions of sorbitol refer to sorbitol derived from simulated Glucose and simulated Fructose. The red plot indicates the values predicted with the best fit parameter set and the grey area indicates the estimated range of variations taking parameter sets within the 95% confidence intervals derived from the identifiability analysis. Isotopologue fractions are reported as m0, m1, m2, etc. where m0, m1, m2. . . indicate the number of ^{13}C atoms in the fraction.
(TIFF)

S1 Table. Metabolites included in the model. This table describes all metabolites included in the model and provides abbreviation, full name, initial concentrations used as initial values for simulations and whether they are dependent variables or assumed as constants. For initial concentrations, "A1, "A2" and "B" superscripts refer to values specific for conditions A1 & A2 and B, respectively, as described in the main text.
(PDF)

S2 Table. Reactions included in the model. This table describes all the reactions included in the metabolic network used in the model and indicates abbreviation, full name and stoichiometry. In reaction stoichiometry, ↔ denotes reversible reactions and → denotes irreversible reactions.
(PDF)

S3 Table. Compartments included in the model. This table describes the three compartments included in the model and their volumes.
(PDF)

S4 Table. Parameters of the kinetic model. This table describes all the parameters of the kinetic model sorted by the reaction to which they are associated.
(PDF)

S5 Table. Parameter groups. This table describes the parameters representing enzyme activities that are fitted as a group, indicating the relative value of the enzyme activities associated to each group.
(PDF)

S6 Table. Parameters fitted to experimental conditions. This table describes the predicted values for all parameters that have been fitted to the experimental data. In addition to the best fit value, the confidence intervals for 95% confidence according to identifiability analysis are also shown.
(PDF)

S1 Text. Kinetic laws used for the aldolase reaction. This text describes the kinetic laws used for the aldolase reactions and how they have been constructed.
(PDF)

S1 XML. Kinetic model 3mM Fructose. SBML version of the kinetic model described in Methods with experiment specific variables set to match those of the experiment in which hepatocytes were incubated with 20 mM glucose and 3 mM fructose (conditions A1 and A2 described in the main text).
(XML)

S2 XML. Kinetic model 20mM Fructose. SBML version of the kinetic model described in Methods with experiment specific variables set to match those of the experiment in which hepatocytes were incubated with 20 mM glucose and 20 mM fructose (condition B described in the main text).
(XML)

S1 Software. HepatoDyn. The software HepatoDyn (that integrates both the kinetic model and the label propagation model) and its source code. Further information about the software is provided in the readme file packaged with the software.
(ZIP)

Acknowledgments

The authors thank Anna Adrover for her excellent technical assistance, International Science Editing (ISE) for language editing and the reviewers for their comments and suggestions.

Author Contributions

Conceived and designed the experiments: MC PdA SM. Performed the experiments: SM. Analyzed the data: CF MC SM WNPL. Contributed reagents/materials/analysis tools: MC JJG. Wrote the paper: CF EF JJG MC PdA SM VAS WNPL. Developed the algorithms: CF MC VAS PdA. Provided technical advice: EF JJG MC PdA SM VAS WNPL. Directed the work: MC PdA. Performed the simulations CF.

References

1. Orman MA, Mattick J, Androulakis IP, Berthiaume F, Ierapetritou MG. Stoichiometry Based Steady-State Hepatic Flux Analysis: Computational and Experimental Aspects. Metabolites. 2012; 2(1):268–91. doi: 10.3390/metabo2010268 PMID: 24957379

2. Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Mol Syst Biol. 2010; 6:401. Epub 2010/09/09. doi: 10.1038/msb.2010.56 msb201056 [pii]. PMID: 20823844; PubMed Central PMCID: PMC2964116.

3. Gille C, Bolling C, Hoppe A, Bulik S, Hoffmann S, Hubner K, et al. HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. Mol Syst Biol. 2010; 6:411. Epub 2010/09/09. msb201062 [pii] doi: 10.1038/msb.2010.62 PMID: 20823849; PubMed Central PMCID: PMC2964118.

4. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nat Commun. 2014; 5:3083. Epub 2014/01/15. doi: 10.1038/ncomms4083 ncomms4083 [pii]. PMID: 24419221.

5. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat Biotech. 2010; 28(3):245–8. doi: 10.1038/nbt.1614. PMID: 20212490

6. Sabate L, Franco R, Canela EI, Centelles JJ, Cascante M. A model of the pentose phosphate pathway in rat liver cells. Mol Cell Biochem. 1995; 142(1):9–17. Epub 1995/01/12. PMID: 7753046.

7. Maher AD, Kuchel PW, Ortega F, de Atauri P, Centelles J, Cascante M. Mathematical modelling of the urea cycle. A numerical investigation into substrate channelling. Eur J Biochem. 2003; 270(19):3953–61. Epub 2003/09/27. 3783 [pii]. PMID: 14511377.

8. Berthiaume F, MacDonald AD, Kang YH, Yarmush ML. Control analysis of mitochondrial metabolism in intact hepatocytes: effect of interleukin-1beta and interleukin-6. Metab Eng. 2003; 5(2):108–23. Epub 2003/07/10. S1096717603000107 [pii]. PMID: 12850133.

9. Mogilevskaya E, Demin O, Goryanin I. Kinetic Model of Mitochondrial Krebs Cycle: Unraveling the Mechanism of Salicylate Hepatotoxic Effects. Journal of Biological Physics. 2006; 32(3–4):245–71. doi: 10.1007/s10867-006-9015-y PMID: 19669466

10. Marin de Mas I, Selivanov VA, Marin S, Roca J, Oresic M, Agius L, et al. Compartmentation of glycogen metabolism revealed from C-13 isotopologue distributions. BMC systems biology. 2011; 5. 175 doi: 10.1186/1752-0509-5-175 WOS:000301740700001. PMID: 22034837

11. Konig M, Bulik S, Holzhutter HG. Quantifying the contribution of the liver to glucose homeostasis: a detailed kinetic model of human hepatic glucose metabolism. PLoS Comput Biol. 2012; 8(6):e1002577. PMID: 22761565. doi: 10.1371/journal.pcbi.1002577

12. Beylot M, Soloviev MV, David F, Landau BR, Brunengraber H. Tracing hepatic gluconeogenesis relative to citric acid cycle activity in vitro and in vivo. Comparisons in the use of [3-13C]lactate, [2-13C]acetate, and alpha-keto[3-13C]isocaproate. J Biol Chem. 1995; 270(4):1509–14. Epub 1995/01/27. PMID: 7829478.

13. Jones JG, Naidoo R, Sherry AD, Jeffrey FM, Cottam GL, Malloy CR. Measurement of gluconeogenesis and pyruvate recycling in the rat liver: a simple analysis of glucose and glutamate isotopomers during metabolism of [1,2,3-(13)C3]propionate. FEBS Lett. 1997; 412(1):131–7. Epub 1997/07/21. S0014-5793(97)00764-3 [pii]. PMID: 9257705.

14. Jucker BM, Lee JY, Shulman RG. In vivo 13C NMR measurements of hepatocellular tricarboxylic acid cycle flux. J Biol Chem. 1998; 273(20):12187–94. Epub 1998/06/20. PMID: 9575166.

15. Yarmush DM, MacDonald AD, Foy BD, Berthiaume F, Tompkins RG, Yarmush ML. Cutaneous burn injury alters relative tricarboxylic acid cycle fluxes in rat liver. J Burn Care Rehabil. 1999; 20(4):292–302. Epub 1999/07/30. PMID: 10425591.

16. Lee K, Berthiaume F, Stephanopoulos GN, Yarmush DM, Yarmush ML. Metabolic flux analysis of postburn hepatic hypermetabolism. Metab Eng. 2000; 2(4):312–27. Epub 2000/12/20. doi: 10.1006/mben.2000.0160 S1096-7176(00)90160-5 [pii]. PMID: 11120643.

17. Choi IY, Wu C, Okar DA, Lange AJ, Gruetter R. Elucidation of the role of fructose 2,6-bisphosphate in the regulation of glucose fluxes in mice using in vivo (13)C NMR measurements of hepatic carbohydrate metabolism. Eur J Biochem. 2002; 269(18):4418–26. Epub 2002/09/17. 3125 [pii]. PMID: 12230553.

18. Otto M, Breinholt J, Westergaard N. Metformin inhibits glycogen synthesis and gluconeogenesis in cultured rat hepatocytes. Diabetes Obes Metab. 2003; 5(3):189–94. Epub 2003/04/12. 263 [pii]. PMID: 12681026.

19. Marin S, Lee WN, Bassilian S, Lim S, Boros LG, Centelles JJ, et al. Dynamic profiling of the glucose metabolic network in fasted rat hepatocytes using [1,2-13C2]glucose. The Biochemical journal. 2004; 381(Pt 1):287–94. Epub 2004/03/23. doi: 10.1042/BJ20031737 BJ20031737 [pii]. PMID: 15032751; PubMed Central PMCID: PMC1133787.

20. Hausler N, Browning J, Merritt M, Storey C, Milde A, Jeffrey FM, et al. Effects of insulin and cytosolic redox state on glucose production pathways in the isolated perfused mouse liver measured by integrated 2H and 13C NMR. The Biochemical journal. 2006; 394(Pt 2):465–73. Epub 2005/11/18. BJ20051174 [pii] doi: 10.1042/BJ20051174 PMID: 16288601; PubMed Central PMCID: PMC1408677.

21. Maier K, Hofmann U, Reuss M, Mauch K. Identification of metabolic fluxes in hepatic cells from transient 13C-labeling experiments: Part II. Flux estimation. Biotechnol Bioeng. 2008; 100(2):355–70. Epub 2007/12/21. doi: 10.1002/bit.21746 PMID: 18095336.

22. Maier K, Hofmann U, Bauer A, Niebel A, Vacun G, Reuss M, et al. Quantification of statin effects on hepatic cholesterol synthesis by transient (13)C-flux analysis. Metab Eng. 2009; 11(4–5):292–309. Epub 2009/06/27. doi: 10.1016/j.ymben.2009.06.001 S1096-7176(09)00043-3 [pii]. PMID: 19555774.

23. Schoors S, De Bock K, Cantelmo AR, Georgiadou M, Ghesquiere B, Cauwenberghs S, et al. Partial and Transient Reduction of Glycolysis by PFKFB3 Blockade Reduces Pathological Angiogenesis. Cell Metabolism. 2014; 19(1):37–48. doi: 10.1016/j.cmet.2013.11.008 WOS:000329431200007. PMID: 24332967

24. Cascante M, Benito A, Marin de Mas I, Centelles JJ. Fluxomics. In: Orseic M, Vidal-Puig A, editors. A Systems Biology Approach to Study Metabolic Syndrome: Springer; 2014. p. 237–50.

25. Selivanov VA, Marin S, Lee PWN, Cascante M. Software for dynamic analysis of tracer-based metabolomic data: estimation of metabolic fluxes and their statistical analysis. Bioinformatics. 2006; 22 (22):2806–12. doi: 10.1093/bioinformatics/btl484 PMID: 17000750

26. Selivanov VA, Vizán P, Mollinedo F, Fan TWM, Lee PWN, Cascante M. Edelfosine-induced metabolic changes in cancer cells that precede the overproduction of reactive oxygen species and apoptosis. BMC Systems Biology. 2010; 4(1):135. doi: 10.1186/1752-0509-4-135 PMID: 20925932

27. Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, et al. A roadmap for interpreting C metabolite labeling patterns from cells. Curr Opin Biotechnol. 2015; 34C:189–201. Epub 2015/03/04. S0958-1669(15)00022-1 [pii] doi: 10.1016/j.copbio.2015.02.003 PMID: 25731751.

28. Tappy L, Le KA. Metabolic effects of fructose and the worldwide increase in obesity. Physiol Rev. 2010; 90(1):23–46. Epub 2010/01/21. doi: 10.1152/physrev.00019.2009 90/1/23 [pii]. PMID: 20086073.

29. Fillat C, Gomez-Foix AM, Guinovart JJ. Stimulation of glucose utilization by fructose in isolated rat hepatocytes. Arch Biochem Biophys. 1993; 300(2):564–9. Epub 1993/02/01. S0003986183710787 [pii]. PMID: 8382026.

30. Rigau T, Rivera M, Palomo MJ, Fernandez-Novell JM, Mogas T, Ballester J, et al. Differential effects of glucose and fructose on hexose metabolism in dog spermatozoa. Reproduction. 2002; 123(4):579–91. Epub 2002/03/27. PMID: 11914120.

31. Lopes A, Vilela TC, Taschetto L, Vuolo F, Petronilho F, Dal-Pizzol F, et al. Evaluation of the effects of fructose on oxidative stress and inflammatory parameters in rat brain. Mol Neurobiol. 2014; 50 (3):1124–30. Epub 2014/04/03. doi: 10.1007/s12035-014-8676-y PMID: 24691544.

32. Fernandez-Novell JM, Ramio-Lluch L, Orozco A, Gomez-Foix AM, Guinovart JJ, Rodriguez-Gil JE. Glucose and fructose have sugar-specific effects in both liver and skeletal muscle in vivo: a role for liver fructokinase. PLoS ONE. 2014; 9(10):e109726. Epub 2014/10/21. doi: 10.1371/journal.pone.0109726 PONE-D-14-34365 [pii]. PMID: 25330076; PubMed Central PMCID: PMC4201455.

33. Varma V, Boros L, Nolen G, Chang C-W, Wabitsch M, Beger R, et al. Metabolic fate of fructose in human adipocytes: a targeted 13C tracer fate association study. Metabolomics. 2014:1–16. doi: 10.1007/s11306-014-0716-0 PMID: 25972768

34. Papandreou D, Andreou E. Role of diet on non-alcoholic fatty liver disease: An updated narrative review. World J Hepatol. 2015; 7(3):575–82. Epub 2015/04/08. doi: 10.4254/wjh.v7.i3.575 PMID: 25848481; PubMed Central PMCID: PMC4381180.

35. Fernandez-Novell JM, Bellido D, Vilaro S, Guinovart JJ. Glucose induces the translocation of glycogen synthase to the cell cortex in rat hepatocytes. The Biochemical journal. 1997; 321 (Pt 1):227–31. Epub 1997/01/01. PMID: 9003423; PubMed Central PMCID: PMC1218058.

36. Tserng KY, Gilfillan CA, Kalhan SC. Determination of carbon-13 labeled lactate in blood by gas chromatography/mass spectrometry. Anal Chem. 1984; 56(3):517–23. Epub 1984/03/01. PMID: 6711822.

37. Katz J, Lee WN, Wals PA, Bergner EA. Studies of glycogen synthesis and the Krebs cycle by mass isotopomer analysis with [U-13C]glucose in rats. J Biol Chem. 1989; 264(22):12994–3004. Epub 1989/08/05. PMID: 2753898.

38. Szafranek J, Pfaffenberger CD, Horning EC. The mass spectra of some per-O-acetylaldononitriles. Carbohydr Res. 1974; 38:97–105. Epub 1974/12/01. S0008-6215(00)82341-1 [pii]. PMID: 4447950.

39. Kaiser FE, Gehrke CW, Zumwalt RW, Kuo KC. Amino acid analysis. Hydrolysis, ion-exchange cleanup, derivatization, and quantitation by gas-liquid chromatography. J Chromatogr. 1974; 94(0):113–33. Epub 1974/07/17. PMID: 4844608.

40. Wolfe RR. Tracers in metabolic research: Radioactive and stable isotope/ mass spectrometry methods. New York: Alan R., Liss, USA 1984.

41. Lee WN, Byerley LO, Bergner EA, Edmond J. Mass isotopomer analysis: theoretical and practical considerations. Biol Mass Spectrom. 1991; 20(8):451–8. Epub 1991/08/01. doi: 10.1002/bms.1200200804 PMID: 1768701.

42. Salway JG. Metabolism at a glance. 3rd ed. Malden, Mass.: Blackwell Pub.; 2004. 125 p. p.
43. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, et al. Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res. 2009; 37(Database issue):D619–22. Epub 2008/11/05. doi: 10.1093/nar/gkn863 gkn863 [pii]. PMID: 18981052; PubMed Central PMCID: PMC2686536.
44. Cornish-Bowden A. Fundamentals of enzyme kinetics. London: Portland Press Ltd; 1995. xiii, 230 p. p.
45. Chang A, Schomburg I, Placzek S, Jeske L, Ulbrich M, Xiao M, et al. BRENDA in 2015: exciting developments in its 25th year of existence. Nucleic Acids Res. 2015; 43(Database issue):D439–46. Epub 2014/11/08. doi: 10.1093/nar/gku1068 gku1068 [pii]. PMID: 25378310.
46. UniProt: a hub for protein information. Nucleic Acids Res. 2015; 43(Database issue):D204–12. Epub 2014/10/29. doi: 10.1093/nar/gku989 gku989 [pii]. PMID: 25348405.
47. Sumam B, Kumar P. A survey of simulated annealing as a tool for single and multiobjective optimization. Journal of the Operational Research Society. 2006;(57):1143–60.
48. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmuller U, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics. 2009; 25(15):1923–9. doi: 10.1093/bioinformatics/btp358 PMID: 19505944
49. Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J. Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics. 2014; 30(10):1440–8. doi: 10.1093/bioinformatics/btu006 PMID: 24463185
50. Mayes PA. Intermediary metabolism of fructose. Am J Clin Nutr. 1993; 58(5 Suppl):754S–65S. Epub 1993/11/01. PMID: 8213607.
51. Sestoft L, Fleron P. Determination of the kinetic constants of fructose transport and phosphorylation in the perfused rat liver. Biochim Biophys Acta. 1974; 345(1):27–38. Epub 1974/04/12. 0005-2736(74)90242-9 [pii]. PMID: 4365063.
52. Masson S, Henriksen O, Stengaard A, Thomsen C, Quistorff B. Hepatic metabolism during constant infusion of fructose; comparative studies with 31P-magnetic resonance spectroscopy in man and rats. Biochim Biophys Acta. 1994; 1199(2):166–74. Epub 1994/03/02. 0304-4165(94)90112-0 [pii]. PMID: 8123666.
53. Debosch BJ, Chen Z, Saben JL, Finck BN, Moley KH. Glucose transporter 8 (GLUT8) mediates fructose-induced de novo lipogenesis and macrosteatosis. J Biol Chem. 2014; 289(16):10989–98. Epub 2014/02/13. doi: 10.1074/jbc.M113.527002 M113.527002 [pii]. PMID: 24519932; PubMed Central PMCID: PMC4036240.
54. Ciudad CJ, Carabaza A, Guinovart JJ. Glycogen synthesis from glucose and fructose in hepatocytes from diabetic rats. Arch Biochem Biophys. 1988; 267(2):437–47. Epub 1988/12/01. 0003-9861(88)90049-5 [pii]. PMID: 3145717.
55. Woods HF, Eggleston LV, Krebs HA. The cause of hepatic accumulation of fructose 1-phosphate on fructose loading. The Biochemical journal. 1970; 119(3):501–10. Epub 1970/09/01. PMID: 5500310; PubMed Central PMCID: PMC1179380.
56. Kitano H. Systems Biology: A Brief Overview. Science. 2002; 295(5560):1662–4. doi: 10.1126/science.1069492 PMID: 11872829
57. Caton PW, Nayuni NK, Khan NQ, Wood EG, Corder R. Fructose induces gluconeogenesis and lipogenesis through a SIRT1-dependent mechanism. J Endocrinol. 2011; 208(3):273–83. Epub 2011/01/08. doi: 10.1530/JOE-10-0190 JOE-10-0190 [pii]. PMID: 21212096.
58. Haas JT, Miao J, Chanda D, Wang Y, Zhao E, Haas ME, et al. Hepatic insulin signaling is required for obesity-dependent expression of SREBP-1c mRNA but not for feeding-dependent expression. Cell Metab. 2012; 15(6):873–84. Epub 2012/06/12. doi: 10.1016/j.cmet.2012.05.002 S1550-4131(12)00195-7 [pii]. PMID: 22682225; PubMed Central PMCID: PMC3383842.
59. van Schaftingen E, Vandercammen A, Detheux M, Davies DR. The regulatory protein of liver glucokinase. Adv Enzyme Regul. 1992; 32:133–48. Epub 1992/01/01. PMID: 1496915.
60. Vandercammen A, Van Schaftingen E. Competitive inhibition of liver glucokinase by its regulatory protein. Eur J Biochem. 1991; 200(2):545–51. Epub 1991/09/01. PMID: 1889417.
61. Beck T, Miller BG. Structural Basis for Regulation of Human Glucokinase by Glucokinase Regulatory Protein. Biochemistry. 2013; 52(36):6232–9. doi: 10.1021/bi400838t PMID: 23957911
62. Abdelmalek MF, Lazo M, Horska A, Bonekamp S, Lipkin EW, Balasubramanyam A, et al. Higher dietary fructose is associated with impaired hepatic adenosine triphosphate homeostasis in obese individuals with type 2 diabetes. Hepatology. 2012; 56(3):952–60. Epub 2012/04/03. doi: 10.1002/hep.25741 PMID: 22467259; PubMed Central PMCID: PMC3406258.
63. Bridges HR, Jones AJ, Pollak MN, Hirst J. Effects of metformin and other biguanides on oxidative phosphorylation in mitochondria. The Biochemical journal. 2014; 462(3):475–87. Epub 2014/07/16. doi: 10.1042/BJ20140620 BJ20140620 [pii]. PMID: 25017630; PubMed Central PMCID: PMC4148174.

8.2 Chapter 2.

PhenoMeNal: Processing and analysis of Metabolomics data in the Cloud

Kristian eters, ames radbury, Sven ergmann, Marco Capuccini, Marta Cascante, edro de Atauri, Timothy MD Ebbels, Carles oguet, et al.

Interoperable and scalable data analysis with microservices: Applications in Metabolomics

ayam Emami Khoonsari, ablo Moreno, Sven ergmann, oachim urman, Marco Capuccini, Matteo Carone, Marta Cascante, edro De Atauri, Carles oguet, et al.

p^{13}CMFA: Parsimonious ^{13}C metabolic flux analysis

Carles oguet, Anusha ayaraman, Silvia Marin, Vitaly A Selivanov, ablo Moreno, Ramon Messeguer, edro de Atauri, Marta Cascante.

GigaScience, 8, 2019, 1–12

doi: 10.1093/gigascience/giy149
Advance Access Publication Date: 7 December 2018
Technical Note

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

TECHNICAL NOTE

PhenoMeNal: processing and analysis of metabolomics data in the cloud

Kristian Peters[1,*], James Bradbury[2,*], Sven Bergmann[3,4], Marco Capuccini[5,6], Marta Cascante[7], Pedro de Atauri[7], Timothy M. D. Ebbels[9], Carles Foguet[7], Robert Glen[9,10], Alejandra Gonzalez-Beltran[11], Ulrich L. Günther[12], Evangelos Handakas[9], Thomas Hankemeier[14], Kenneth Haug[15], Stephanie Herman[6,16], Petr Holub[17], Massimiliano Izzo[11], Daniel Jacob[18], David Johnson[11,19], Fabien Jourdan[20], Namrata Kale[15], Ibrahim Karaman[21], Bita Khalili[3,4], Payam Emami Khonsari[16], Kim Kultima[16], Samuel Lampa[6], Anders Larsson[6,22], Christian Ludwig[23], Pablo Moreno[15], Steffen Neumann[1,24], Jon Ander Novella[6,22], Claire O'Donovan[15], Jake T.M. Pearce[9], Alina Peluso[9], Marco Enrico Piras[25], Luca Pireddu[25], Michelle A.C. Reed[12], Philippe Rocca-Serra[11], Pierrick Roger[26], Antonio Rosato[27], Rico Rueedi[3,4], Christoph Ruttkies[1], Noureddin Sadawi[8,9], Reza M. Salek[15], Susanna-Assunta Sansone[11], Vitaly Selivanov[7], Ola Spjuth[6], Daniel Schober[1], Etienne A. Thévenot[26], Mattia Tomasoni[3,4], Merlijn van Rijswijk[13,28], Michael van Vliet[14], Mark R. Viant[2,29], Ralf J. M. Weber[2,29], Gianluigi Zanetti[25] and Christoph Steinbeck[30,*]

[1]Leibniz Institute of Plant Biochemistry, Stress and Developmental Biology, Weinberg 3, 06120 Halle (Saale), Germany, [2]School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom, [3]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland, [4]Swiss Institute of Bioinformatics, Lausanne, Switzerland, [5]Division of Scientific Computing, Department of Information Technology, Uppsala University, Sweden, [6]Department of Pharmaceutical Biosciences, Uppsala University, Box 591, 751 24 Uppsala, Sweden, [7]Department of Biochemistry and Molecular Biomedicine,

Received: 6 September 2018; **Revised:** 19 October 2018; **Accepted:** 20 November 2018

© The Author(s) 2018. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Universitat de Barcelona; Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Instituto de Salud Carlos III (ISCIII), Spain, [8]Department of Computer Science, College of Engineering, Design and Physical Sciences, Brunel University, London, UK, [9]Department of Surgery & Cancer, Imperial College London, South Kensington, London, SW7 2AZ, United Kingdom, [10]Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB21EW, United Kingdom, [11]Oxford e-Research Centre, Department of Engineering Science, University of Oxford, 7 Keble Road, OX1 3QG, Oxford, United Kingdom., [12]Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom, [13]Netherlands Metabolomics Center, Leiden, 2333 CC, Netherlands, [14]Division of Systems Biomedicine and Pharmacology, Leiden Academic Centre for Drug Research (LACDR), Leiden University, Leiden, 2333 CC, The Netherlands, [15]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom, [16]Department of Medical Sciences, Clinical Chemistry, Uppsala University, 751 85 Uppsala, Sweden, [17]BBMRI-ERIC, Graz, Austria, [18]INRA, University of Bordeaux, Plateforme Métabolome Bordeaux-MetaboHUB, 33140 Villenave d'Ornon, France, [19]Department of Informatics and Media, Uppsala University, Box 513, 751 20 Uppsala, Sweden, [20]INRA - French National Institute for Agricultural Research, UMR1331, Toxalim, Research Centre in Food Toxicology, Toulouse, France, [21]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, St. Mary's Campus, Norfolk Place, W2 1PG, London, United Kingdom, [22]National Bioinformatics Infrastructure Sweden, Uppsala University, Uppsala, Sweden, [23]Institute of Metabolism and Systems Research (IMSR), University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom, [24]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany, [25]Distributed Computing Group, CRS4, Pula, Italy, [26]CEA, LIST, Laboratory for Data Analysis and Systems' Intelligence, MetaboHUB, Gif-Sur-Yvette F-91191, France, [27]Magnetic Resonance Center (CERM) and Department of Chemistry, University of Florence and CIRMMP, 50019 Sesto Fiorentino, Florence, Italy, [28]ELIXIR-NL, Dutch Techcentre for Life Sciences, Utrecht, 3503 RM, Netherlands, [29]Phenome Centre Birmingham, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom, [30]Cheminformatics and Computational Metabolomics, Institute for Analytical Chemistry, Lessingstr. 8, 07743 Jena, Germany, [31] and [32]

*Correspondence address. Kristian Peters, Leibniz Institute of Plant Biochemistry, Stress and Developmental Biology, Weinberg 3, 06120 Halle (Saale), Germany. E-mail: kpeters@ipb-halle.de http://orcid.org/0000-0002-4321-0257; James Bradbury, E-mail: j.bradbury@bham.ac.uk; Christoph Steinbeck, E-mail: christoph.steinbeck@uni-jena.de http://orcid.org/0000-0001-6966-0814

Abstract

Background: Metabolomics is the comprehensive study of a multitude of small molecules to gain insight into an organism's metabolism. The research field is dynamic and expanding with applications across biomedical, biotechnological, and many other applied biological domains. Its computationally intensive nature has driven requirements for open data formats, data repositories, and data analysis tools. However, the rapid progress has resulted in a mosaic of independent, and sometimes incompatible, analysis methods that are difficult to connect into a useful and complete data analysis solution. **Findings:** PhenoMeNal (Phenome and Metabolome aNalysis) is an advanced and complete solution to set up Infrastructure-as-a-Service (IaaS) that brings workflow-oriented, interoperable metabolomics data analysis platforms into the cloud. PhenoMeNal seamlessly integrates a wide array of existing open-source tools that are tested and packaged as Docker containers through the project's continuous integration process and deployed based on a kubernetes orchestration framework. It also provides a number of standardized, automated, and published analysis workflows in the user interfaces Galaxy, Jupyter, Luigi, and Pachyderm. **Conclusions:** PhenoMeNal constitutes a keystone solution in cloud e-infrastructures available for metabolomics. PhenoMeNal is a unique and complete solution for setting up cloud e-infrastructures through easy-to-use web interfaces that can be scaled to any custom public and private cloud environment. By harmonizing and automating software installation and configuration and through ready-to-use scientific workflow user interfaces, PhenoMeNal has succeeded in providing scientists with workflow-driven, reproducible, and shareable metabolomics data analysis platforms that are interfaced through standard data formats, representative datasets, versioned, and have been tested for reproducibility and interoperability. The elastic implementation of PhenoMeNal further allows easy adaptation of the infrastructure to other application areas and 'omics research domains.

Keywords: metabolomics; data analysis; e-infrastructures; NMR; mass spectrometry; computational workflows; galaxy; cloud computing; standardization; statistics

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

Findings

Background

The field of metabolomics has seen remarkable progress over the last decade and has enabled fascinating discoveries in many different research areas. Metabolomics is the study of small molecules in organisms that can reveal detailed insights into metabolic biochemistry, e.g., changes in concentrations of specific molecules, metabolic fluxes between cells or compartments, identification of molecules that are involved in the pathogenesis of a disease, and the study of the biochemical phenotype of animals, plants, and even soil microorganisms [1–3].

The principal metabolomics technologies of mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) typically generate large datasets that require computationally intensive analyses [4]. Biomedical investigations can involve large cohorts with many thousands of metabolite profiles and can produce hundreds of gigabytes of data [5–8]. With such large datasets, processing becomes impracticable and unmanageable on commodity hardware. Cloud computing can offer a solution by enabling the outsourcing of calculations from local workstations to scalable cloud data centers, with the possibility to allocate thousands of central processing unit (CPU) cores simultaneously. Furthermore, cloud computing allows for resources to be instantiated on-demand (CPUs, random access memory, network, storage) and allows access to computational tools in the form of microservices that can dynamically grow or shrink.

MS and NMR data processing usually involves selection of parameters (that are often specific to the analytical instrumentation), algorithmic peak detection, peak alignment and grouping, annotation of putative compounds, and extensive statistical analyses [9, 10]. Many open-source tools have been developed that address these different steps in data processing and analysis. These tools, however, usually come with their own software dependencies, resource requirements, and scripting languages. As a consequence, configuring and running them is often complicated, especially for researchers who are untrained in computer science [4]. Furthermore, many tools require users to input parameters that can significantly affect results and performance, and reporting of these parameters is not always clear [11].

A number of infrastructures and integration efforts have been initiated in the past five years, including metabolomics data repositories with a global scope [6, 12], platforms for reproducible workflow analysis [13, 14], as well as initiatives to integrate and coordinate data standards [15]. Simultaneously, multiple networks of service centers such as the international Phenome Centers [16] and MetaboHub [17] have formed with the goal to facilitate the acquisition, processing, and analysis of metabolomics data [6–8] at ever increasing scales.

Currently, several web-based metabolomics data processing platforms are available. XCMSOnline provides a platform based on XCMS for downstream data analysis, visualization, data sharing, and access to Metlin to facilitate metabolite identification and pathway analysis [18]. MetaboAnalyst presents a wide variety of data processing and analysis tools including statistical analysis, time-series analysis, functional analysis, and pathway analysis [19]. Workflow4Metabolomics is based on Galaxy and provides various metabolomics processing workflows, including NMR [13, 20]. These common tools for analyzing metabolomics data provide web-based graphical user interfaces (GUIs) with different functionality.

Here, we present PhenoMeNal (Phenome and Metabolome aNalysis), a unique, easy-to-use, complete, robust, and performant cloud e-infrastructure that provides a large suite of standardized and interoperable metabolomics data processing tools as a complete data analysis solution. In contrast to current metabolomics processing platforms, PhenoMeNal provides Infrastructure-as-a-Service (IaaS) and seamlessly integrates a wide array of existing open-source tools.

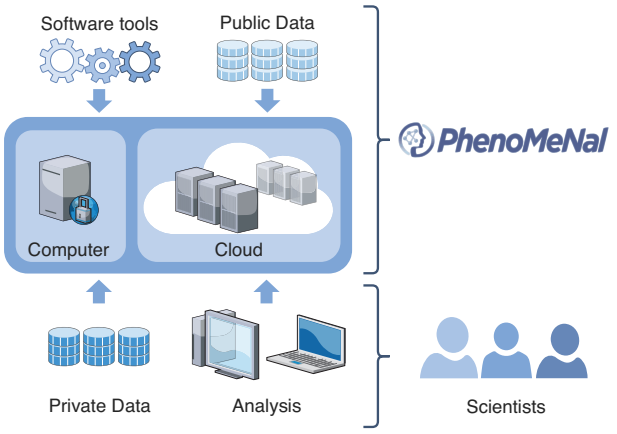


Figure 1: Conceptual design of the PhenoMeNal cloud e-infrastructure, which brings compute to the data for any large number of data scientists.

A major advantage over other platforms is that PhenoMeNal make it possible to instantiate many different services in the cloud and provides a number of standardized, automated, and published analysis workflows in the user interfaces Galaxy, Jupyter, Luigi, and Pachyderm (Fig. 1). Moreover, the PhenoMeNal e-infrastructure can be easily deployed onto public and private cloud environments and can be configured elastically to fit into any cloud-based environment, thus enabling scalable and cost-effective high-performance metabolomics data analysis in a way that hides the technical complexity from the user. PhenoMeNal further facilitates reproducible analyses through automated, sharable, and citable workflows.

Overview

The features of the PhenoMeNal e-infrastructure are encapsulated as a cloud research environment (CRE). The PhenoMeNal CRE can be instantiated on major commercial public cloud providers, including Amazon web services (AWS) and Google cloud platform (GCP), as well as OpenStack-based private clouds and in custom environments. Technical complexity is hidden from the users, simplifying setting up the cloud infrastructure for administrators (Fig. 2).

From a web-based portal, users can deploy the CRE, which includes several web services and software tools (Fig. 2). Data can be processed directly in the e-infrastructure without the need to install additional software. Scientific workflows can be executed via user-friendly web-based platforms such as Galaxy, as well as programmatic interfaces and notebooks. Each service has been supplied with a rich source of documentation and training material to assist researchers.

The PhenoMeNal Portal

The PhenoMeNal Portal [21] allows users to deploy, manage, and delete PhenoMeNal CREs simply through a web interface. Deployments to major commercial cloud platforms (AWS and GCP) as well as OpenStack, an open-source cloud platform, can be made using an easy-to-follow wizard (Fig. 2). OpenStack deploy-

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

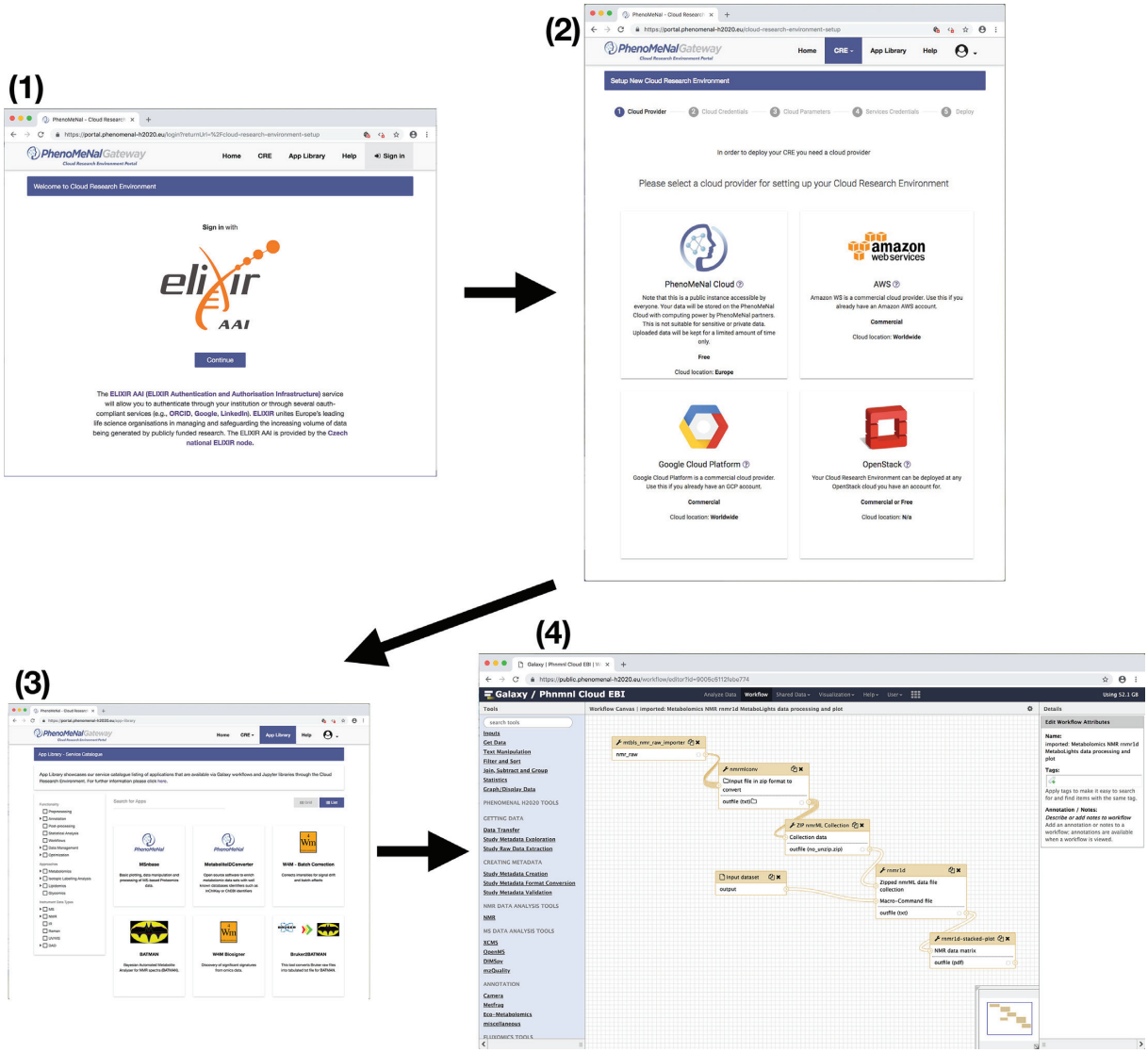


Figure 2: Screenshots of creating and using the PhenoMeNal cloud e-infrastructure. First, log in with ELIXIR to the cloud research environment (CRE) portal. Second, select a public or private cloud provider. After entering cloud credentials and setting up parameters in the dedicated portal, the deployment of the PhenoMeNal e-infrastructure into the cloud environment can be made. Third, in the PhenoMeNal Portal app library there are several services ready to be deployed and used in the set-up infrastructure. Fourth, dedicated web services such as Galaxy are readily available in the cloud e-infrastructure. All steps can be operated from an easy-to-use web interface that is accessible from any standard web browser.

ments can be deployed behind clinical firewalls, which is especially pertinent when dealing with sensitive (i.e., patient) data.

The PhenoMeNal public instance allows users to test-run a CRE without the need to deploy on a cloud platform. It can be deployed and accessed through the portal. Once credentials for users have been generated, analyses can be run through a Galaxy instance containing the tools and workflows present in any deployed CRE. The portal also includes user and developer documentation, workflow tutorials, and links to training videos.

Scientific workflows

A scientific workflow is a set of computational steps that are carried out to process and analyze data [22]. Usually, a workflow is comprised of several linked software tools that are each executed during a particular step of the workflow. In order to manage and automate scientific workflows, in PhenoMeNal the well-established dedicated workflow management system Galaxy can be deployed, which presents the user with an easy-to-use graphical user interface as well as providing a programmatic interface [20, 23]. Galaxy facilitates collaborative exchange, reproducibility, and traceability of data analysis by enabling users to share entire workflows and analysis histories [24]. In addition to Galaxy, programmatic executable notebooks (Jupyter) and the workflow tools exposed as programmatic interfaces Luigi and Pachyderm are also supported [25].

In order to cover typical use cases in metabolomics and to illustrate the usage and applicability of given analytical pipelines and software tools, five representative scientific workflows are available in the PhenoMeNal Galaxy (Table 1), each having different computational demands and purposes. More than 250 individual modules have been integrated in Galaxy (see the subsection Scientific Workflows in the Methods section).

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

Table 1: List of workflows that are representative for their respective metabolomics domains (identification in NMR, Fluxomics, Annotation, and identification in MS and eco-metabolomics)

Workflow name	Description	Reference
1D NMR	Processes 1D NMR experiments from raw data to a data matrix required for visualization and statistical analysis, building on nmrML and NMRProcFlow. The automatic workflow is based on the MTBLS1 dataset, describing urinary changes in type 2 diabetes in humans.	[26, 27, 28]
Fluxomics	Quantifies steady-state fluxes following ^{13}C metabolic flux analysis. The workflow was first based on the analysis of the MTBLS412 dataset with 13C tracer data of human umbilical vein endothelial cells under hypoxia.	[29, 30]
LC-MS/MS	Processes, quantifies, and annotates/identifies features in mass spectra using MetFrag — a tool that annotates molecules from compound databases of tandem mass spectrometry (MS/MS) spectra. The workflow is based on MTBLS558.	[31, 32, 33]
Univariate and Multivariate Statistics	Applies univariate and multivariate statistical analysis and illustrates how datasets may be explored, enabling the identification of variables of interest and the construction of predictive models. The workflow is based on MTBLS404.	[13, 34]
Eco-Metabolomics	Implementation of a resource demanding metabolomics use case in ecology, used in large field experiments to describe interactions between different species of organisms in remarkable detail. The workflow is based on MTBLS520.	[35]
ISA-Create-Validate-Upload	A workflow to create Investigation, Study, and Assay data model framework-compliant metadata files based on study design information, augmented with semantic markup as source, implementing UK Phenome center naming conventions. Following validation, the workflow also allows visualization of overall study design and deposition to EMBL-EBI.	

Software tools

The Portal App Library [36] shows the software tools packaged in PhenoMeNal that are available through the CRE deployment (Fig. 2). The range of software tools available covers several metabolomics domains, making PhenoMeNal relevant for use in a wide range of data analysis scenarios. The domains covered include clinical metabolomics, plant metabolomics, fluxomics, and eco-metabolomics. Data from both targeted and untargeted analysis can be analyzed for metabolite profiling and fingerprinting approaches [1, 2]. NMR and MS (liquid chromatography coupled with mass spectrometry, gas chromatography coupled with mass spectrometry, direct infusion mass spectrometry) data can be processed.

PhenoMeNal also provides tools for data management (e.g., via the Investigation, Study, and Assay data model framework [ISA] format and application programming interface [API]), metabolite feature detection (e.g., XCMS, CAMERA, nmrProcFlow), metabolite identification (MetFrag, BATMAN, MetaboMatching), and (bio)statistics (e.g., univariate, multivariate, and power analyses) (Supplementary Table S1). Tools can be filtered for functionality, approaches, and instrument (data) types to readily find the most appropriate software tools. Some tools that implement specific functionality (e.g., Rnmr1D, which performs baseline correction of NMR spectra as part of nmrProcFlow) are available through dedicated Galaxy modules or through software containers (Supplementary Table S1).

Study design

PhenoMeNal was designed to use standardized protocols and software tools and to comply with state-of-the-art dedicated specifications and data formats across the entire project. Development was geared toward implementation of open standards for tracking provenance of both data and metadata generated by clinical phenotyping projects. In PhenoMeNal, the ISA model and specifications were implemented using the ISA format to generate, annotate, validate, and deposit experimental metadata information of datasets and studies to public repositories such as MetaboLights [37, 38]. ISA-based metadata tracking is used for the different analysis pipelines that are specific to the distinct metabolomics domains. PhenoMeNal reached native support for the ISA format by developing a dedicated Galaxy composite data type. Such component affords direct recognition of the ISA format by the Galaxy environment, thus ensuring seamless integration with the downstream workflow component.

Data deposition

PhenoMeNal encourages the metabolomics data repository MetaboLights as a primary source of data deposition [39]. Private and public datasets are supported, as are download and upload to MetaboLights. If the storage in a data repository such as MetaboLights is not possible, data can be stored locally or in the cloud e-infrastructure. Access to the data is strictly controlled and secured. To support data deposition, ISA-based Galaxy modules are available making it possible to publish and disseminate scientific results in standard compliant ways.

Reproducibility

One of the challenges of cloud computing is that analyses need to be run continuously and successfully in different environments [40]. Specifically, it has to be ensured that, given the same input, workflows and tools produce identical results regardless of the underlying environment [4, 40]. When these requirements are fulfilled, end users can be confident that their data will be analyzed correctly. PhenoMeNal has implemented three major testing strategies to ensure technical reproducibility using a continuous integration framework [41]. Tests were implemented for the infrastructure components, individual software containers, and data involved in computational workflows.

Sustainability

PhenoMeNal is part of a number of initiatives (BioMedBridges, COSMOS, and ELIXIR) to foster the role of metabolomics and to harmonize experimental data and metadata usage [15, 42]. Col-

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

laborations were established with EGI [43] and Indigo Datacloud [44] infrastructure providers and initiatives [45, 46] to ensure that PhenoMeNal uses technologies that are well supported and ensure their widespread usage, continuity, and further development. For example, the development of KubeNow and contributions to the Galaxy and Workflow4Metabolomics community are essential for PhenoMeNal [47]. Core development will continue on GitHub and is fostered by collaborations with tool developers.

Dependencies on specific technologies and frameworks were avoided by focusing on open standards such as ISA-Tab/ISA-JSON, mzML and nmrML, and widely accepted software [48]. By being able to deploy PhenoMeNal on multiple types of cloud environments, lock-ins to specific computing resource providers are avoided. PhenoMeNal implemented continuous integration and delivery, validated by extensive testing and with clear maintenance responsibilities (see Methods section).

Privacy and security

With human or animal material, the collection, storage, and analysis of metabolomics data introduce a number of constraints due to ethical, legal, and social implications (ELSI) [49]. In particular, data initially derived from human clinical studies may be identifiable and will require consent for use, usually for a defined objective, such as diagnosis, or be related to a particular disease study. Where data is identifiable or pseudonymized, users can deploy PhenoMeNal on local secure resources, thus avoiding the export of data. In this scenario, access to the e-infrastructure should be strictly controlled through local access and authorization. It is recommended that clinical data be fully anonymized before analysis in PhenoMeNal [49, 50].

The PhenoMeNal portal provides substantial guidance to enable users to comply with ELSI and general data protection regulation (GDPR) requirements. Users must register in order to use the individual parts of the e-infrastructure. PhenoMeNal was implemented to use secured and encrypted transport and network communications.

Documentation and training materials

Extensive user documentation and tutorials are provided via the PhenoMeNal Wiki page [51]. The Wiki includes detailed developer resources including information about the PhenoMeNal release schedule; guidelines for tool, workflow, and portal developers; continuous integration; and testing. Further documentation is also provided detailing, creating, and managing PhenoMeNal CREs and tutorials for the Galaxy modules and preconfigured workflows, as well as Galaxy tours that provide step-by-step guidance for inexperienced users.

Community engagement

The PhenoMeNal project is open source and is hosted on GitHub [52]. Developers can contribute tools to PhenoMeNal and are encouraged to do so. To add a tool to PhenoMeNal, it must be containerized using Docker and then integrated into the build process. Detailed documentation is available in the project's Wiki for developers who wish to add their tools to PhenoMeNal.

Collaborations with other projects have been actively encouraged during the development of PhenoMeNal, including Workflow4Metabolomics [13] and the developers of both nmrML and nmrProcFlow [26]. These collaborations are essential to fostering greater standardization within PhenoMeNal and to increasing compatibility with other metabolomics data processing infrastructures.

Availability

Information on how to access PhenoMeNal can be found at the project's website [53]. The GitHub repository hosts the source code of all development projects [52]. The project container-galaxy-k8s-runtime contains all of the developments regarding Galaxy. The Wiki containing documentation is also hosted on GitHub [51]. The PhenoMeNal Portal can be reached at [21]. The public instance of Galaxy is accessible at [54]. Source code and documentation are available under the terms of the Apache 2.0 license. Integrated open-source projects are available under the respective licensing terms.

Conclusions

PhenoMeNal has succeeded in increasing the robustness and coverage of representative metabolomics data processing in scientific cloud e-infrastructures. The presented cloud e-infrastructure covers a wide range of analysis pipelines including data generation and download, data pre- and post-processing, (bio)statistics, and result deposition in data repositories. A large effort has been made to introduce lower-level changes to cloud e-infrastructures (e.g., the cloud deployment software KubeNow) to meet the demands of the biomedical domain. Furthermore, Galaxy has been enriched with metabolomics data standards, in particular, the ISA format for study metadata and mzML and nmrML for acquired data files, as well as support for Kubernetes. PhenoMeNal has fostered the visibility of new metabolomics tools and has enabled the development of more sophisticated data analysis workflows. Our efforts were also guided by feedback from real-life test scenarios collected at workshops with users from the clinical domain.

PhenoMeNal constitutes a keystone solution in cloud platforms available for metabolomics data analysis. The platform was designed to deliver optimal performance and functionality for typical use cases in the metabolomics domain. While the needs of clinicians and researchers in the biomedical and biochemical domains have been targeted, PhenoMeNal is not limited to a specific domain as the cloud infrastructure, tools, and workflows can be adapted to other use cases as demonstrated with the inclusion of the eco-metabolomics workflow. The technological advancements can be reused in other scientific cloud environments and could be integrated with solutions from other 'omics domains in the future.

Methods

Cloud e-infrastructure

The PhenoMeNal CRE is designed as a microservice architecture, with services being implemented as virtual machine images and software containers. Containers are used to provide microservices for metabolomics data analysis tools and also long-running services such as workflow management systems. A container orchestrator runs containers on top of the scalable infrastructure. The orchestrator takes a group of machines that act as a distributed cluster and receives requests for tools as well as service executions. PhenoMeNal implements various layers to providea container orchestrator on top of either bare metal hardware or IaaS given by a cloud provider [55] (Supplementary Fig. S1).

During the setup process and while PhenoMeNal is deployed, data storage and CPU limits can be configured and dynamically scaled to fit any cloud environment. Deployments can be made

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

to GCE, AWS, and OpenStack-based private clouds from the PhenoMeNal portal. Deployments are also supported from the command line to Microsoft Azure [56], the European Science Cloud [57], and local servers (bare metal) [58]; we provide step-by-step instructions for these solutions.

PhenoMeNal provides IaaS for three different cloud environments:

"local cloud": local workstations or bare metal clusters where data are not allowed to leave the facility.
"public cloud": the flexible use of commercial cloud providers such as GCE and AWS.
"shared cloud": using OpenStack—a free and open-source software platform for cloud computing, ideal for custom environments and research networks.

Software tools

The PhenoMeNal portal has an application library that allows users to deploy tools as microservices into the cloud infrastructure (Fig. 2, Supplementary Table S1). The portal is packaged into frontend and backend engines on top of Kubernetes.

Most software tools in PhenoMeNal are compiled from source code and use a variety of programming languages. Linux versions of software tools and user interfaces such as Galaxy are supported in dedicated encapsulated Docker containers that are implemented as minimum-sized microservices. PhenoMeNal currently hosts 100 such projects in its GitHub repository [59] (Supplementary Table S1). Projects are indicated by the trailing Àcontainer-À name and include a ruleset to build and run the containerized tools, as well as datasets for testing and other necessary files.

PhenoMeNal provides tutorials for developers who want to integrate their tools into our e-infrastructure [60].

Scientific workflows

In PhenoMeNal, a number of options are available for running reproducible and standardized workflows (Table 1).

Galaxy

The Galaxy workflow management system is widely regarded as one of the most popular scientific workflow platforms [20, 61]. It provides a user-friendly web-based GUI to make it easy for the end user to configure and run individual modules and entire workflows without programming experience. Command-line tools and scripts are encapsulated into modules that are launched via the web interface. Galaxy also supports more powerful features such as programmatic access through a REST API and helper libraries to access the running instance of Galaxy [62].

PhenoMeNal has been able to adapt Galaxy for use with a microservices-based architecture [31]. To this end, modules are encapsulated into Docker containers that can be flexibly launched within the cloud e-infrastructure. Galaxy is available in all deployed PhenoMeNal CREs and contains more than 250 modules that have been implemented as part of PhenoMeNal.

Six representative metabolomics Galaxy workflows have been fully integrated into PhenoMeNal (Table 1), and more workflows (mzQuality, NMR-BATMAN) are available for testing.

Jupyter

Jupyter, which started its history as the IPython notebook, is the most popular among the tools commonly referred to as executable notebooks or computational notebooks [63]. Jupyter lets users combine executable code with results from code executions such as text, tables, and figures. Usually, Jupyter notebooks are enriched with extended information that explains what the code does. As a result, they are often used for training material and for tutorials. Also, computational notebooks can, to some extent, be used as a way to document code executions and to make executions more reproducible [64].

Luigi and pachyderm

Luigi is a Python workflow programming library that was originally developed by the company Spotify. It manages pipelines of computations primarily on "big data" systems such as Hadoop and Apache Spark but also supports local execution [63, 64]. Luigi is a very flexible library that facilitates building complex pipelines of batch jobs handling dependency resolution, workflow management, and visualization.

Similarly, Pachyderm makes it possible to process distributed data and to keep track of the data from every stage of the analysis pipeline [25]. With Pachyderm, it is possible to track the provenance of results and to accurately reproduce scientific workflows. Luigi and Pachyderm are well suited for complex scientific tasks and are easy to use from the python environment in Jupyter notebooks without additional integration tooling needed.

In PhenoMeNal, we have extended Galaxy, Jupyter, Luigi, and Pachyderm in such a way that they can be orchestrated throughout the cloud infrastructure together with the data analysis tools themselves [31]. Six important metabolomics workflows have been fully integrated into PhenoMeNal (Table 1), and more (mzQuality, NMR-BATMAN) are available for testing.

Reproducibility

Three strategies are realized to ensure technical reproducibility. They are implemented in the continuous integration (CI) software development framework Jenkins [41] which is accessible at [65]. These strategies are implemented as tests in our Jenkins and a tutorial guide is available at [66].

- Infrastructure testing: Procedures were implemented to ensure that each individual component (e.g., the deployment process of software containers, resource management, APIs/application binary interfaces [ABIs]) within the infrastructure is interacting correctly with the other components.
- Container testing: Verification that tools, which are packaged into software containers, build and run correctly in the infrastructure. Dependencies within one container and across several interdependent containers are tested.
- Data testing: The output of tools, which process demonstration data, is checked against a data set that is known to contain the expected result. This is being done for both individual tools and for several tools running in a workflow using the workflow testing tool for Galaxy called wft4galaxy [67].

Standardization

PhenoMeNal has implemented several dedicated Galaxy modules that directly retrieve and store ISA-Tab data set descriptors from and to MetaboLights, and can convert between other formats. Native Galaxy composite data types to support ISA-Tab and ISA-JSON have also been integrated, building upon the ISA API [38, 48]. The ISA data type allows for the upload of an ISA-Tab archive (a zip file containing the ISA set of files and raw

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

Table 2: Overview of the most important FAIR criteria and implementations suggested for PhenoMeNal data, tools and workflows

	Data	Tools	Workflows
(F)indability	Indexing in domain relevant databases (e.g., MetaboLights)	Indexing in domain relevant software repositories (e.g., the PhenoMeNal App Library, GitHub)	Indexing in workflow management systems such as Galaxy (e.g., PhenoMeNal, W4M), or libraries such as [69]
	Rich descriptions of metadata (e.g., ISA-Tab)	Tool descriptions follow the EDAM ontology	Persistent identifier (e.g., W4M ID, DOI) and intuitive naming patterns
(A)ccessibility	Data access and rights management based on e.g., data use ontology (DUO)	Accessible open-source licenses	Access to workflow systems can be configured to be shared or restricted
(I)nteroperability	Standard formats for experimental metadata (ISA-Tab/ISA-JSON)	Standardized tool descriptions	Standardized workflow format (e.g., Galaxy GA format, Common Workflow Language CWL)
	Domain specific standards for raw data (e.g., mzML, nmrML)	Containerization of software tools	Execution in various software environments (e.g., through the use of containers)
	OboFoundry vocabularies and established domain ontologies to annotate data	EDAM ontology to annotate tools	Workflow annotation ontologies (e.g., Ontology of workflow motifs for annotating workflow specifications [70])
(R)eusability	Deposition in data repositories (e.g., MetaboLights) and data indexing sites (e.g., OmicsDI)	Rich documentation and usage guides	Rich documentation and tutorials (e.g., Galaxy tours)

data when available), which is displayed to the users as a single Galaxy history data set. The integrated Galaxy modules include a MetaboLights downloader and uploader (for ingestion and submission), an ISAcreate module for the creation of ISA compliant archives, modules to explore study metadata through queries on study factors, ISA-Tab "slicing" where queries are used to select subsets of data files of interest, as well as format conversion (export to ISA-JSON and Workflow4Metabolomics [W4M]) and study metadata validation (Supplemental Table S1).

PhenoMeNal also advanced the specification of the nmrML standard data format [27] and contributed a dedicated composite data type for nmrML to Galaxy. nmrML is used extensively throughout the NMR 1D workflow and conversion from raw format into nmrML is supported via dedicated Galaxy modules (Table 1).

Throughout the entire analysis pipeline, modules of computational workflows were designed to accept standard formats such as mzML, XML or CSV whenever possible.

Standardized APIs/ABIs are being used for the programmatic interfaces as well as for deploying services. To this end, modern and standardized programming, scripting and meta languages were selected such as Go, HCL, Python, Shell, XML and YAML that are widely used in cloud computing.

Reusability

In an ongoing effort, PhenoMeNal is actively advancing the criteria for good data management and stewardship based on findability, accessibility, interoperability and reusability (FAIR) for good data management and stewardship [68] to be applied not only to data, but also to software tools and computational workflows (Table 2).

Privacy

PhenoMeNal supports fully anonymized data, which cannot be traced back to individuals in any way [50] and treats pseudonymized data as identifiable. As pseudonymized data are anonymous to the investigator, third parties may be able to link pseudonymized data back to identifiable individuals through mappings such as a hash or code [49]. In these cases, e.g., in a hospital environment, users must deploy PhenoMeNal within a private cloud or bare metal cluster behind their institution's firewall.

PhenoMeNal provides guidance on ethical and technical frameworks to regulate and secure the use of private or sensitive data [49, 50]. It is possible to combine data and metadata within an ELSI compliant framework [50] and in such cases users can follow the example of the European Genome Phenome Archive (EGA) [71]. In public installations of PhenoMeNal, the ELIXIR policy on privacy has been implemented within a technically secure environment to process data [42].

Security

Open-source tools are used throughout the entire e-infrastructure. This promotes community efforts to discover and resolve bugs and security issues. The container build process is steered by the continuous integration (CI) service Jenkins, which continuously builds the containers and generates reports. On success and through authentication, container images are pushed to the PhenoMeNal container registry, which is publicly available but read-only. Cloud provider credentials are not stored in the cloud but only on the deployer host. The Kubernetes cluster running the Jenkins-CI and the container registry, as well as the portal, runs on a CoreOS container, which is a self-updatable, cluster-aware system with most portions being read-only. It reboots nodes sequentially to avoid lack of availability.

KubeNow is a key component that initializes the cloud infrastructure and configures access to it via Cloudflare [72], providing dynamic Domain Name Services (DNS) and encryption for all network communication. The flexible implementation of PhenoMeNal allows the user to decide to not use Cloudflare, in

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

which case encryption is disabled. KubeAdm, which manages the setup of Kubernetes, is not reachable at runtime by default. The only way to access it is by having access to the private key stored on the computer on which it was launched. PhenoMeNal only allows access to standard ports (ssh, http, https, and port 44 for the Galaxy Downloader) and implements a cloud-specific firewall for all supported cloud providers.

Microservices are designed to be launched on-demand and terminated after completed analysis. If security issues are reported for the microservices, tool, or dependencies or if incremental security patches are available, new builds are automatically triggered in the CI system and developers and the release manager are notified to take additional actions if required. Images are built on a daily basis and tested for deployment to avoid security patches from introducing any abnormality in the deployment process.

User resources

There are many user resources for both PhenoMeNal users and developers in the form of documentation, tutorials, and training videos. The PhenoMeNal Wiki [51] contains detailed documentation on all aspects of PhenoMeNal, including general user guides, workflow and tool tutorials, developer documentation, and general information on topics such as security and the e-infrastructure landscape. The PhenoMeNal portal contains help pages generated from the Wiki [73], which are categorized as User Documentation, Developer Documentation, and Workflow Tutorials. Interactive Galaxy tours are directly integrated in Galaxy [74]. Training videos are available at the project's YouTube page [75].

Availability of source code and requirements

Project name: PhenoMeNal,
Project home page: http://phenomenal-h2020.eu
Operating system(s): Platform independent
Programming language: Go, HCL, Java, JavaScript, Python, R, Shell, XML, YAML
Other requirements: Linux, Docker, Kubernetes, Terraform, Ansible, Helm
License: MIT license for all code written by the PhenoMeNal project. Individual, Open Source Foundation approved licenses for all containerized tools.
RRID:SCR_016605

Availability of supporting data

The following MetaboLights datasets are integrated into PhenoMeNal and are used to demonstrate the cloud integration and reproducibility of Galaxy workflows: MTBLS1 (NMR1D), MTBLS404 (Uni- and multivariate statistics), MTBLS412 (Fluxomics), MTBLS520 (Eco-Metabolomics), MTBLS558 (MetFrag). Datasets are available at https://www.ebi.ac.uk/metabolights. Snapshots of the code and additional supporting data are available in the *GigaScience* repository, GigaDB [76].

Additional files

Supplemental Figure 1: PhenoMeNal implements various layers to provision containers on top of the e-infrastructure.

Supplemental Table 1: List of external software tools that were incorporated into PhenoMeNal.

Abbreviations

ABI: application binary interface; API: application programming interface; AWS: Amazon web services; CI: continuous integration; CPU: central processing unit; CRE: cloud research environment; ELSI: ethical, legal, and social implications; FAIR: criteria for good data management and stewardship based on findability, accessibility, interoperability, and reusability; GCP: Google cloud platform; GUI: graphical user interface; IaaS: Infrastructure-as-a-Service; ISA: Investigation, Study, and Assay data model framework; MS: mass spectrometry; NMR: nuclear magnetic resonance (spectroscopy); PhenoMeNal: Phenome and Metabolome aNalysis; W4M: Workflow4Metabolomics.

Competing interests

The authors declare that they have no competing interests.

Declarations

Human-derived samples in the datasets MTBLS404 and MTBLS412 were processed according to ELSI guidelines.

Author contributions

Writing original draft: K.P. and J.B.
Conceptualization: C.S.
Supervision: R.G., U.L.G., K.H., S.N., A.R., M.vR., C.S., O.S., P.R.-S., R.W.
Project administration: N.K.
Technical lead: P.M.
Review and editing: J.B., M.C., M.Cap., M.Cas., P.dA., T.M.D.E., R.G., A.G.-B., K.H., S.H., D.Ja., D.Jo., F.J., K.K., N.K., P.E.K., A.L., S.L., P.M., S.N., C.O.D., K.P., L.P., M.E.P., M.A.C.R., P.R.-S., P.R.-M., A.R., R.R., C.R., M.vR., N.S., R.M.S., S.-A.S., D.S., O.S., V.S., E.A.T., M.T., T.H., M.vV., M.R.V., R.J.M.W., G.Z., C.S.
Software: J.B., M.Cap., M.Cas., P.dA., A.G.-B., U.L.G., K.H., S.H., D.Jo., F.J., P.E.K., A.L., C.L., P.M., S.N., C.O.D., K.P., L.P., M.E.P., M.A.C.R., P.R.-S., P.R.-M., A.R., R.R., C.R., M.vR., N.S., R.M.S., S.-A.S., O.S., V.S., E.A.T., M.T., T.H., M.vV., R.J.M.W., G.Z.
External software: S.B., C.F., E.H., S.H., M.I., D.Ja., B.K., I.K., K.K., P.E.K., S.L., J.A.N., J.T.M.P., A.P., L.P., R.R.
Data curation: K.H., S.-A.S., P.R.-S.
Funding acquisition: R.G., U.L.G., K.H., S.N., A.R., M.vR., C.S., O.S., P.R.-S., R.W.
Support was provided by the Nordice-Infrastructure Collaboration (NeIC) via the Glenna2 and Tryggve2 projects. All authors contributed to, read, and approved the final manuscript.

Funding

The project was funded by the European Commission PhenoMeNal (grant EC654241). The consortium members include J.B., M.Cap., M.Cas., P.dA., T.M.D.E., R.G., A.G.-B., K.H., M.I., D.Jo., F.J., N.K., P.E.K., A.L., P.M., S.N., C.O.D., K.P., L.P., M.A.C.R., P.R.-S., P.R.-M., A.R., R.R., C.R., T.H., M.vR., M.vV., N.S., R.M.S., S.-A.S., D.S., O.S., V.S., E.A.T., M.T., M.R.V., and R.J.M.W. C.S. received funding from the European Commission PhenoMeNal (grant EC654241).

References

1. Gowda GN, Zhang S, Gu H, et al. Metabolomics-based methods for early disease diagnostics. Expert Rev Mol Diagn 2008;**8**:617–33.

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

2. Bundy JG, Davey MP, Viant MR. Environmental metabolomics: a critical review and future perspectives. Metabolomics 2009;**5**:3–21.
3. Peters K, Worrich A, Weinhold A, et al. Current challenges in plant eco-metabolomics. Int J Mol Sci 2018;**19**:1385.
4. Weber RJM, Lawson TN, Salek RM, et al. Computational tools and workflows in metabolomics: an international survey highlights the opportunity for harmonisation through Galaxy. Metabolomics 2017;**13**:12.
5. Joyce AR, Palsson BØ. The model organism as a system: integrating "omics" data sets. Nat Rev Mol Cell Biol 2006;**7**:198–210.
6. Haug K, Salek RM, Conesa P, et al. MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res 2013;**41**:D781–6.
7. Lindon JC, Nicholson JK. The emergent role of metabolic phenotyping in dynamic patient stratification. Expert Opin Drug Metab Toxicol 2014;**10**:915–9.
8. Sumner LW, Hall RD. Metabolomics across the globe. Metabolomics 2013;**9**:258–64.
9. Rosato A, Tenori L, Cascante M, et al. From correlation to causation: analysis of metabolomics data using systems biology approaches. Metabolomics Off J Metabolomic Soc 2018;**14**:37.
10. Vignoli A, Ghini V, Meoni G, et al. High-throughput metabolomics by 1D NMR. Angew. Chem. Int. Ed., 2018, **57**, 2–29, doi:10.1002/anie.201804736.
11. Goodacre R, Broadhurst D, Smilde AK, et al. Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics 2007;**3**:231–41.
12. Sud M, Fahy E, Cotter D, et al. Metabolomics Workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Res 2016;**44**:D463–70.
13. Giacomoni F, Le Corguille G, Monsoor M, et al. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. Bioinformatics 2015;**31**:1493–5.
14. Haug K, Salek RM, Steinbeck C. Global open data management in metabolomics. Curr Opin Chem Biol 2017;**36**:58–63.
15. Salek RM, Neumann S, Schober D, et al. COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access. Metabolomics 2015;**11**:1587–97.
16. IPCN. International Phenome Centre Network. http://phenomenetwork.org. 2018. Accessed 25 Oct 2018.
17. French Ministry of Research, Higher Education and the National Agency for Science. MetaboHUB. http://www.metabohub.fr/metabohub.html. 2018. Accessed 25 Oct 2018.
18. Tautenhahn R, Patti GJ, Rinehart D, et al. XCMS Online: a web-based platform to process untargeted metabolomic data. Anal Chem 2012;**84**:5035–9.
19. Chong J, Soufan O, Li C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res 2018;**46**:W486–94.
20. Afgan E, Baker D, van den Beek M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 2016;**44**:W3–10.
21. PhenoMeNal: The PhenoMeNal Portal. https://portal.phenomenal-h2020.eu. 2018. Accessed 25 Oct 2018.
22. Hoffa C, Mehta G, Freeman T, et al. On the Use of Cloud Computing for Scientific Workflows. 2008 IEEE Fourth Int Conf EScience. Indianapolis, IN, USA: IEEE; 2008 [cited 2018 Sep 3]. p. 640–5. Available from: http://ieeexplore.ieee.org/document/4736878/.
23. Digan W, Countouris H, Barritault M, et al. An architecture for genomics analysis in a clinical setting using Galaxy and Docker. GigaScience 2017;**6**:1–9.
24. Goecks J, Nekrutenko A, Taylor J, Galaxy Team T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 2010;**11**:R86.
25. Novella JA, Khoonsari PE, Herman S, et al. Container-based bioinformatics with Pachyderm, Wren J . editor. Bioinformatics 2018, 1–8; doi:10.1093/bioinformatics/bty699/5068160.
26. Jacob D, Deborde C, Lefebvre M, et al. NMRProcFlow: a graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. Metabolomics 2017;**13**:36.
27. Schober D, Jacob D, Wilson M, et al. nmrML: a community supported open data standard for the description, storage, and exchange of NMR Ddta. Anal Chem 2018;**90**:649–56.
28. Salek RM, Maguire ML, Bentley E, et al. A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. Physiol Genomics 2007;**29**:99–108.
29. Buescher JM, Antoniewicz MR, Boros LG, et al. A roadmap for interpreting 13 C metabolite labeling patterns from cells. Curr Opin Biotechnol 2015;**34**:189–201.
30. Niedenführ S, Wiechert W, Nöh K. How to measure metabolic fluxes: a taxonomic guide for 13 C fluxomics. Curr Opin Biotechnol 2015;**34**:82–90.
31. Emami Khoonsari P, Moreno P, Bergmann S, et al. Interoperable and scalable data analysis with microservices: Applications in Metabolomics, Journal: bioRxiv. 2018, **bioRxiv:213603**, 1–29 bioRxiv doi:10.1101/213603.
32. Ruttkies C, Schymanski EL, Wolf S, et al. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J Cheminformatics 2016;**8**:3. http://www.jcheminf.com/content/8/1/3.
33. Herman S, Khoonsari PE, Tolf A. et al. Integration of magnetic resonance imaging and protein and metabolite CSF measurements to enable early diagnosis of secondary progressive multiple sclerosis. Theranostics 2018;**8**:4477–90.
34. Thévenot EA, Roux A, Xu Y. et al. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. J Proteome Res 2015;**14**:3322–35.
35. Peters K, Gorzolka K, Bruelheide H, et al. Computational workflow to study the seasonal variation of secondary metabolites in nine different bryophytes. Sci Data 2018;**5**:180179.
36. PhenoMeNal. The Portal App Library. https://portal.phenomenal-h2020.eu/app-library. 2018. Accessed 25 Oct 2018.
37. Rocca-Serra P, Salek RM, Arita M, et al. Data standards can boost metabolomics research, and if there is a will, there is a way. Metabolomics. 2016;**12**:14.
38. Smith B, Ashburner M, Rosse CThe OBI Consortium,, et al., The OBI Consortium, The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 2007;**25**:1251–5.
39. Steinbeck C, Conesa P, Haug K, et al. MetaboLights: towards a new COSMOS of metabolomics data management. Metabolomics 2012;**8**:757–60.
40. Gil Y, Deelman E, Ellisman M, et al. Examining the challenges of scientific workflows. Computer 2007;**40**:24–32.
41. Moutsatsos IK, Hossain I, Agarinis C, et al. Jenkins-CI, an

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

open-source continuous integration system, as a scientific data and image-processing platform. SLAS Discov Adv Life Sci RD 2017;**22**:238–49.
42. van Rijswijk M, Beirnaert C, Caron C, et al. The future of metabolomics in ELIXIR. F1000Research 2017;**6**:1649.
43. EGI Foundation. EGI: Advanced Computing for Research. https://www.egi.eu. 2018. Accessed 25 Oct 2018.
44. INIGO Datacloud. INtegrating Distributed data Infrastructures for Global ExplOitation. https://www.indigo-datacloud.eu. 2018. Accessed 25 Oct 2018.
45. Viljoen M, Dutka L, Kryza B, et al. Towards European Open Science Commons: the EGI Open Data Platform and the EGI DataHub. Procedia Comput Sci 2016;**97**:148–52.
46. Salomoni D, Campos I, Gaido L, et al. INDIGO-DataCloud: a Platform to Facilitate Seamless Access to E-Infrastructures, J Grid Computing, 2018, **16**, 381–408. ArXiv160309536 Cs. doi: 10.1007/s10723-018-9453-3.
47. Capuccini M, Larsson A, Carone M, et al. On-demand virtual research environments using microservices, 10.1093/bioinformatics/bty699/5068160, **arXiv:1805.06180**, 1–31. ArXiv180506180 Cs. 2018; http://arxiv.org/abs/1805.06180.
48. Rocca-Serra P, Brandizi M, Maguire E, et al. ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. Bioinformatics 2010;**26**:2354–6.
49. Sariyar M, Schluender I, Smee C, et al. Sharing and reuse of sensitive data and samples: supporting researchers in identifying ethical and legal requirements. Biopreservation Biobanking 2015;**13**:263–70.
50. Heatherly R, Rasmussen LV, Peissig PL, et al. A multi-institution evaluation of clinical profile anonymization. J Am Med Inform Assoc 2016;**23**:e131–7.
51. PhenoMeNal. Wiki. https://github.com/phnmnl/phenomenal-h2020/wiki. 2018. Accessed 25 Oct 2018.
52. PhenoMeNal. GitHub Project Repository. https://github.com/phnmnl/. 2018. Accessed 25 Oct 2018.
53. PhenoMeNal. Phenome and Metabolome aNalysis. https://phenomenal-h2020.eu. 2018. Accessed 25 Oct 2018.
54. PhenoMeNal. Public Galaxy Instance. https://public.phenomenal-h2020.eu. 2018. Accessed 25 Oct 2018.
55. Mell PM, Grance T. The NIST definition of cloud computing. In: Gaithersburg MD . National Institute of Standards and Technology; 2011. Report No.: NIST SP 800-145. Available from: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.
56. PhenoMeNal. Deploy on Microsoft Azure. https://github.com/phnmnl/phenomenal-h2020/wiki/Deploy-on-Microsoft-Azure. 2018. Accessed 25 Oct 2018.
57. PhenoMeNal. Deploy on European Open Science Cloud (EOSC). https://github.com/phnmnl/phenomenal-h2020/wiki/Deploy-on-European-Open-Science-Cloud-(EOSC). 2018. Accessed 25 Oct 2018.
58. PhenoMeNal. Deploy on a local server (bare metal). https://github.com/phnmnl/phenomenal-h2020/wiki/Deploy-on-a-local-server-(bare-metal). 2018. Accessed 25 Oct 2018.
59. Phnmnl GitHub https://github.com/phnmnl/?q=container.
60. PhenoMeNal. How to make your software tool available through PhenoMeNal. https://github.com/phnmnl/phenomenal-h2020/wiki/How-to-make-your-software-tool-available-through-PhenoMeNal. 2018. Accessed 25 Oct 2018.
61. Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat Rev Genet 2012;**13**:667–72.
62. Sloggett C, Goonasekera N, Afgan E. BioBlend: automating pipeline analyses within Galaxy and CloudMan. Bioinformatics 2013;**29**:1685–6.
63. Thomas K, Benjamin R-K, Fernando P, et al. Jupyter Notebooks - a publishing format for reproducible computational workflows. Stand Alone. 2016;87–90.
64. Lampa S, Alvarsson J, Spjuth O. Towards agile large-scale predictive modelling in drug discovery with flow-based programming design principles. J Cheminformatics 2016;**8**:67.
65. PhenoMeNal. Jenkins-CI Instance. http://phenomenal-h2020.eu/jenkins/. 2018. Accessed 25 Oct 2018.
66. PhenoMeNal. Jenkins Guide. https://github.com/phnmnl/phenomenal-h2020/wiki/Jenkins-Guide. 2018. Accessed 25 Oct 2018.
67. Piras ME, Pireddu L, Zanetti G. wft4galaxy: a workflow testing tool for galaxy. Bioinformatics 2017;**33**:3805–7.
68. Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;**3**:160018.
69. myExperiment www.myexperiment.org. Accessed 25 Oct 2018
70. Cohen-Boulakia S, Belhajjame K, Collin O, et al. Scientific workflows for computational reproducibility in the life sciences: status, challenges and opportunities. Future Gener Comput Syst 2017;**75**:284–98.
71. Lappalainen I, Almeida-King J, Kumanduri V, et al. The European genome-phenome archive of human data consented for biomedical research. Nat Genet 2015;**47**:692–5.
72. Cloudflare Inc. Cloudflare. https://www.cloudflare.com/. 2018. Accessed 25 Oct 2018.
73. PhenoMeNal. Portal Help. https://portal.phenomenal-h2020.eu/help. 2018. Accessed 25 Oct 2018.
74. PhenoMeNal. Interactive Galaxy Tours. https://public.phenomenal-h2020.eu/tours. 2018. Accessed 25 Oct 2018.
75. PhenoMeNal. The PhenoMeNal YouTube page. https://www.youtube.com/channel/UCXGAvsVNQk-aUpckjRC8Ang. 2018. Accessed 25 Oct 2018.
76. Peters K, Bradbury J, Bergmann S, et al. Supporting data for "PhenoMeNal: Processing and analysis of Metabolomics data in the Cloud." GigaScience Database 2018. http://dx.doi.org/10.5524/100528.
77 Brikman Y. Terraform: Writing Infrastructure as Code. Sebastopol: O'Reilly Media; 2017. Available from: http://public.eblib.com/choice/publicfullrecord.aspx?p=4822376.
78 Hanwell MD, de Jong WA, Harris CJ. Open chemistry: RESTful web APIs, JSON, NWChem and the modern web application. J Cheminformatics. 2017;**9**:55.
79 Newman S. Building microservices: designing fine-grained systems. First Edition. Beijing Sebastopol, CA: O'Reilly Media; 2015.
80 Erl T (Ed.). SOA with REST: principles, patterns & constraints for building enterprise solutions with REST. Upper Saddle River, NJ: Prentice Hall; 2012.
81 Bandrowski A, Brinkman R, Brochhausen M, et al. The Ontology for Biomedical Investigations. PLoS One 2016;**11**:e0154556.
82 Sansone S-A, Rocca-Serra P, Field D, et al. Toward interoperable bioscience data. Nat Genet 2012;**44**:121–6.
83 Sansone S-A, Schober D, Atherton HJ, et al. Metabolomics standards initiative: ontology working group work in progress. Metabolomics 2007;**3**:249–56.
84 Dyke SOM, Philippakis AA, Rambla De Argila, J et al. Consent Codes: upholding standard data use conditions. PLoS Genet 2016;**12**:e1005772.

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

85 Selivanov VA, Benito A, Miranda, A et al. MIDcor, an R-program for deciphering mass interferences in mass spectra of metabolites enriched in stable isotopes. BMC Bioinformatics 2017;**18**:88.

86 Hao J, Liebeke M, Astle W, et al. Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. Nat Protoc 2014;**9**:1416–27.

87 Rinaudo P, Boudah S, Junot C, et al. biosigner: a new method for the discovery of significant molecular signatures from omics data. Front Mol Biosci 2016;**3**:26.

88 Kuhl C, Tautenhahn R, Böttcher C, et al. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem 2012;**84**:283–9.

89 Dührkop K, Shen H, Meusel M, et al. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc Natl Acad Sci 2015;**112**:12580–5.

90 Southam AD, Weber RJM, Engel J, et al. A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics. Nat Protoc 2017;**12**:255–73.

91 King ZA, Dräger A, Ebrahim A, et al. Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. PLOS Comput Biol 2015;**11**:e1004321.

92 Cottret L, Frainay C, Chazalviel M, et al. MetExplore: collaborative edition and exploration of metabolic networks. Nucleic Acids Res 2018;**46**:W495–502.

93 Libiseller G, Dvorzak M, Kleb U, et al. IPO: a tool for automated optimization of XCMS parameters. BMC Bioinformatics 2015;**16**:118.

94 González-Beltrán A, Neumann S, Maguire E, et al. The Risa R/Bioconductor package: integrative data analysis from experimental metadata and back again. BMC Bioinformatics 2014;**15**:S11.

95 Sansone S-A, Rocca-Serra P, Field D, et al. Toward interoperable bioscience data. Nat Genet 2012;**44**:121–6.

96 Selivanov VA, Vizán P, Mollinedo F, et al. Edelfosine-induced metabolic changes in cancer cells that precede the overproduction of reactive oxygen species and apoptosis. BMC Syst Biol 2010;**4**:135.

97 Perez F, Granger BE. IPython: a system for interactive scientific computing. Comput Sci Eng 2007;**9**:21–9.

98 Ludwig C, Günther UL. MetaboLab - advanced NMR data processing and analysis for metabolomics. BMC Bioinformatics 2011;**12**:366.

99 Wohlgemuth G, Haldiya PK, Willighagen E, et al. The Chemical Translation Service–a web-based tool to improve standardization of metabolomic reports. Bioinformatics 2010;**26**:2647–8.

100 Rueedi R, Mallol R, Raffler J, et al. Metabomatching: using genetic association to identify metabolites in proton NMR spectroscopy. PLOS Comput Biol 2017;**13**:e1005839.

101 Helmus JJ, Jaroniec CP. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. J Biomol NMR 2013;**55**:355–67.

102 Mohamed A, Nguyen CH, Mamitsuka H. NMRPro: an integrated web component for interactive processing and visualization of NMR spectra. Bioinformatics 2016;**32**:2067–8.

103 Sturm M, Bertsch A, Gröpl C, et al. OpenMS – an open-source software framework for mass spectrometry. BMC Bioinformatics 2008;**9**:163.

104 Blaise BJ, Correia G, Tin A, et al. Power analysis and sample size determination in metabolic phenotyping. Anal Chem 2016;**88**:5179–88.

105 Scheubert K, Hufsky F, Petras D, et al. Significance estimation for large scale metabolomics annotations by spectral matching. Nat Commun 2017;**8**, 1–24. doi:10.1038/s41467-017-01318-5.

106 Chambers MC, Maclean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 2012;**30**:918–20.

107 Lewis IA, Schommer SC, Markley JL. rNMR: open source software for identifying and quantifying metabolites in NMR spectra. Magn Reson Chem 2009;**47**:S123–6.

108 Rodriguez N, Thomas A, Watanabe L, et al. JSBML 1.0: providing a smorgasbord of options to encode systems biology models: Table 1. Bioinformatics 2015;**31**:3383–6.

109 Benton HP, Wong DM, Trauger SA, et al. XCMS ^{2} : processing tandem mass spectrometry data for metabolite identification and structural characterization. Anal Chem 2008;**80**:6382–9.

Downloaded from https://academic.oup.com/gigascience/article-abstract/8/2/giy149/5232984 by Universitat de Barcelona user on 26 September 2019

Bioinformatics, 2019, 1–9
doi: 10.1093/bioinformatics/btz160
Advance Access Publication Date: 9 March 2019
Original Paper

OXFORD

Downloaded from https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btz160/5372675 by guest on 02 May 2019

Systems biology

Interoperable and scalable data analysis with microservices: applications in metabolomics

Payam Emami Khoonsari[1], Pablo Moreno[2], Sven Bergmann[3,4], Joachim Burman[5], Marco Capuccini[6,7], Matteo Carone[7], Marta Cascante[8,9], Pedro de Atauri[8,9], Carles Foguet[8,9], Alejandra N. Gonzalez-Beltran[10], Thomas Hankemeier[11], Kenneth Haug[2], Sijin He [2], Stephanie Herman[1,7], David Johnson[10], Namrata Kale[2], Anders Larsson[7,12], Steffen Neumann[13,14], Kristian Peters[13], Luca Pireddu[15], Philippe Rocca-Serra[10], Pierrick Roger[16], Rico Rueedi[3,4], Christoph Ruttkies[13], Noureddin Sadawi[17], Reza M. Salek[18], Susanna-Assunta Sansone[10], Daniel Schober[13], Vitaly Selivanov[8,9], Etienne A. Thévenot[16], Michael van Vliet[11], Gianluigi Zanetti[15], Christoph Steinbeck[2,19], Kim Kultima[1] and Ola Spjuth[7,*]

[1]Department of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden, [2]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK, [3]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland, [4]Swiss Institute of Bioinformatics, Lausanne, Switzerland, [5]Department of Neuroscience, Uppsala University, Uppsala, Sweden, [6]Department of Information Technology, Uppsala University, Uppsala, Sweden, [7]Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden, [8]Department of Biochemistry and Molecular Biomedicine, and Institute of Biomedicine (IBUB), Faculty of Biology, Universitat de Barcelona (IBUB), Barcelona, Spain, [9]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD) and Metabolomics Node at INB-Bioinfarmatics Platform, Instituto de Salud Carlos III (ISCIII), Madrid, Spain, [10]Oxford e-Research Centre, Department of Engineering Science, University of Oxford, Oxford, UK, [11]Division of Analytical Biosciences, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands, [12]National Bioinformatics Infrastructure Sweden, Uppsala University, Uppsala, Sweden, [13]Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, Halle, Germany, [14]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany, [15]CRS4: Center for Advanced Studies, Research and Development in Sardinia, Distributed Computing Group, Pula, Italy, [16]CEA, LIST, Laboratory for Data Analysis and Systems' Intelligence, MetaboHUB, Gif-sur-Yvette, France, [17]Faculty of Medicine, Department of Surgery & Cancer, Imperial College London, London, UK, [18]International Agency for Research on Cancer, 69372 Lyon CEDEX 08, France and [19]Institute for Inorganic and Analytical Chemistry, Friedrich-Schiller-University, Jena, Germany

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on August 22, 2018; revised on February 25, 2019; editorial decision on March 1, 2019; accepted on March 8, 2019

Abstract

Motivation: Developing a robust and performant data analysis workflow that integrates all necessary components whilst still being able to scale over multiple compute nodes is a challenging task. We introduce a generic method based on the microservice architecture, where software tools are encapsulated as Docker containers that can be connected into scientific workflows and executed using the Kubernetes container orchestrator.

© The Author(s) 2019. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Results: We developed a Virtual Research Environment (VRE) which facilitates rapid integration of new tools and developing scalable and interoperable workflows for performing metabolomics data analysis. The environment can be launched on-demand on cloud resources and desktop computers. IT-expertise requirements on the user side are kept to a minimum, and workflows can be re-used effortlessly by any novice user. We validate our method in the field of metabolomics on two mass spectrometry, one nuclear magnetic resonance spectroscopy and one fluxomics study. We showed that the method scales dynamically with increasing availability of computational resources. We demonstrated that the method facilitates interoperability using integration of the major software suites resulting in a turn-key workflow encompassing all steps for mass-spectrometry-based metabolomics including preprocessing, statistics and identification. Microservices is a generic methodology that can serve any scientific discipline and opens up for new types of large-scale integrative science.
Availability and implementation: The PhenoMeNal consortium maintains a web portal (https://portal.phenomenal-h2020.eu) providing a GUI for launching the Virtual Research Environment. The GitHub repository https://github.com/phnmnl/ hosts the source code of all projects.
Contact: ola.spjuth@farmbio.uu.se
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Biology is becoming data-intensive as high throughput experiments in genomics or metabolomics are rapidly generating datasets of massive volume and complexity (Marx, 2013; Schadt *et al.*, 2010), posing a fundamental challenge on large scale data analytics.

Currently, the most common large-scale computational infrastructures in science are shared High-Performance Computing (HPC) systems. Such systems are usually designed primarily to support computationally intensive batch jobs—e.g. for the simulation of physical processes—and are managed by specialized system administrators. This model leads to rigid constraints on the way these resources can be used. For instance, the installation of software must undergo approval and may be restricted, which contrasts with the needs in the analysis where a multitude of software components of various versions—and their dependencies—are needed, and where these need to be continuously updated.

Cloud computing offers a compelling alternative to shared HPC systems, with the possibility to instantiate and configure on-demand resources such as virtual computers, networks and storage, together with operating systems and software tools. Users only pay for the time the virtual resources are used, and when they are no longer needed they can be released and incur no further costs for usage or ownership. For scientists, this constitutes a shift from owning computer hardware, to starting up Infrastructure-as-a-Service (IaaS) nodes with virtual machines on cloud resources, with the explicit need to then install all necessary software for the analysis which in many cases constitutes a demanding and time-consuming task (Langmead and Nellore, 2018). Along with infrastructure provisioning, software provisioning—i.e. installing and configuring software for users—has also advanced. Consider, for instance, containerization (Silver, 2017), which allows entire applications with their dependencies to be packaged, shipped and run on a computer but isolated from one another in a way analogous to virtual machines, yet much more efficiently. Containers are more compact, and since they share the same operating system kernel, they are fast to start and stop and incur little overhead in execution. These traits make them an ideal solution to implement lightweight *microservices*, a software engineering methodology in which complex applications are divided into a collection of smaller, loosely coupled components that communicate over a network (Newman, 2015). Microservices share many properties with traditional always-on web services found on the Internet, but microservices are generally smaller, portable and can be started on-demand within a separate computing environment. Another important feature of microservices is that they have a technology-agnostic communication protocol, and hence can serve as building blocks that can be combined and reused in multiple ways (da Veiga Leprevost *et al.*, 2017).

Microservices are highly suitable to run in elastic cloud environments that can dynamically grow or shrink on demand, enabling applications to be scaled-up by simply starting multiple parallel instances of the same service. However, to achieve effective scalability a system needs to be appropriately sectioned into microservice components and the data to be exchanged between the microservices needs to be defined for maximum efficiency—both being challenging tasks.

One of the omics fields that faces challenges by data growth is metabolomics which measures the occurrence, concentrations and changes of small molecules (metabolites) in organisms, organs, tissues, cells and cellular compartments. Metabolite abundances are assayed in the context of environmental or dietary changes, disease or other conditions (Nicholson and Wilson, 2003). Metabolomics is, as most other omics technologies, characterized by the use of high-throughput experiments performed using a variety of spectroscopic methods such as Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) that produce large amounts of data (Montenegro-Burke *et al.*, 2017). With increasing data size and number of samples, the analysis process becomes intractable for desktop computers due to requirements on compute cores, memory, storage, etc. As a result, large-scale computing infrastructures have become important components in scientific projects (Liew *et al.*, 2016). Moreover, making use of such complex computing resources in an analysis workflow presents its own challenges, including achieving efficient job parallelism and scheduling as well as error handling (Suplatov *et al.*, 2016). In addition, configuring the necessary software tools and chaining them together into a complete re-runnable analysis workflow commonly requires substantial IT-expertise, while creating portable and fault-tolerant workflows with a robust audit trail is even more difficult. Metabolomics has already benefited from cloud-based systems enabling the users certain

Downloaded from https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btz160/5372675 by guest on 02 May 2019

preprocessing and main downstream analysis on e.g. MS data. Examples of such systems are XCMS ONLINE (Warth *et al.*, 2017), MetaboAnalyst (Xia *et al.*, 2012), Chorus (chorusproject.org) and The Metabolomics Workbench (Sud *et al.*, 2016) (www.metabolomicsworkbench.org) which provide tools that scale with computational demands.

In this manuscript, we present a method that uses components for data analysis encapsulated as microservices and connected into computational workflows to provide complete, ready-to-run, reproducible data analysis solutions that can be easily deployed on desktop computers as well as public and private clouds. Our work contrasts to previously reported research environments, sometimes termed Virtual Research Environments (Allan, 2009; Candela *et al.*, 2013), Scientific Gateways (Lawrence *et al.*, 2015) and Virtual Labs (Waldrop, 2013), in that it encompasses the complete setup of the computational infrastructure and frameworks to run analysis in a wide range of environments; however our approach requires virtually no involvement in the setup and no special IT skills from the user. The methodology provides a framework for rapid and efficient integration of new tools and developing scalable, and interoperable workflows, supporting multiple workflow engines such as Galaxy (Goecks *et al.*, 2010) and Luigi (https://github.com/spotify/luigi). We validate the method on four metabolomics studies and show that it enables scalable and interoperable data analysis.

2 Materials and methods

2.1 Microservices

A detailed description of the methods is present in Supplementary Method S1. Briefly, in order to construct a microservice architecture for metabolomics we used Docker (Merkel, 2014) (https://www.docker.com/) containers to encapsulate software tools. Tools are developed as open source and are available in a public repository such as GitHub (https://github.com/), and the PhenoMeNal project containers are built and tested on a Jenkins continuous integration (CI) server (http://phenomenal-h2020.eu/jenkins/). Containers are assembled in different branches using the git versioning system. Builds originating from the development branch of each container repository give rise to container images tagged as 'development'; builds coming from the master branches result in release images. In order for a container be pushed to the container registry, it must pass a testing criteria which is defined by the developer of the tool. All published containers are thus available for download and can be used in any microservice architecture. Data is exchanged between services by passing references to a shared local file system. The CI system constitutes a key part of the methodology, as it ensures that containers are continuously successfully packaged, versioned, tested and that adequate reporting measures are in place to handle any errors in this process over time.

2.2 Virtual Research Environment (VRE)

We developed a Virtual Research Environment (VRE) which uses Kubernetes (https://kubernetes.io/) for orchestration of the containers, including initialization and scaling of jobs based on containers, abstractions to file system access for running containers, exposure of services, as well as rescheduling of failed jobs and long running services. Kubernetes was chosen over other frameworks such as Docker Swarm because of its larger momentum and that it is more widely used in production environments. Docker also provides Kubernetes as part of their Enterprise solutions (and even now the community ones). To enable convenient instantiation of a complete virtual infrastructure, we developed KubeNow (https://github.com/kubenow/KubeNow) (Capuccini *et al.*, 2018) which includes instantiation of compute nodes, shared file system storage, networks, configure DNS, operating system, container implementation and orchestration tools, including Kubernetes, on a local computer or server. In order to deploy applications, we used two main classes of services: long-lasting services, and compute jobs. Long-lasting services were used for applications such as the user interface whereas compute jobs were used to perform temporary functions in data processing. The VRE includes Galaxy, Luigi workflow engine and Jupyter notebook as user-facing services. In the PhenoMeNal CI system, the VRE is instantiated and tested on all supported cloud providers nightly in order to ensure a working system over time.

2.3 Demonstrators

We validated our method in the field of metabolomics using four demonstrators. Demonstrators 1 and 2 showcase scalability and interoperability of our microservice-based architecture whereas Demonstrators 3 and 4 exemplify flexibility to account for new application domains, showing the architecture is domain-agnostic.

Demonstrator 1: Scalability of microservices in a cloud environment. The objective of this analysis was to demonstrate the computational scalability of an existing workflow on a large dataset [Metabolomics data have been deposited to the EMBL-EBI MetaboLights database (Haug *et al.*, 2013) with the identifier MTBLS233 (Ranninger *et al.*, 2016). The complete dataset can be accessed here https://www.ebi.ac.uk/metabolights/MTBLS233]. The experiment includes 528 mass spectrometry samples from whole cell lysates of human renal proximal tubule cells that were pre-processed through a five-step workflow (consisting of peak picking, feature finding, linking, file filtering and exporting) using the OpenMS software (Sturm *et al.*, 2008). This preprocessing workflow was reimplemented using Docker containers and run using the Luigi workflow engine. Scalability of concurrent running tools (on 40 Luigi workers, each worker receives tasks from the scheduler and executes them) was measured using weak scaling efficiency (WSE), where the workload assigned to each worker stays constant and additional workers are used to solve a larger total problem.

Demonstrator 2: Interoperability of microservices. The objective of this analysis was to demonstrate interoperability as well as to present a real-world scenario in which patients' data are processed using a microservices-based platform. We used a dataset consisting of 37 clinical cerebrospinal fluid (CSF) samples including thirteen relapsing-remitting multiple sclerosis (RRMS) patients and 14 secondary progressive multiple sclerosis (SPMS) patients as well as 10 non-multiple sclerosis controls. 26 quality controls (19 blank and 7 dilution series samples) were also added to the experiment. In addition, 8 pooled CSF samples containing MS/MS data were included in the experiment for improving identification [Metabolomics data have been deposited to the EMBL-EBI MetaboLights database with the identifier MTBLS558. The complete dataset can be accessed here https://www.ebi.ac.uk/metabolights/MTBLS558]. The samples were processed and analyzed on the Galaxy platform (Goecks *et al.*, 2010), running in a VRE behind the Uppsala University Hospital firewall to be compliant with local ELSI (Ethics, Legal, Social implications) regulations.

Demonstrator 3: 1D NMR-analysis workflow. The purpose of this demonstrator was to highlight the fact that the microservice architecture is indeed domain-agnostic and is not limited to a particular assay technology. This NMR-based metabolomics study was originally performed by Salek *et al.* (2007) on urine of type 2

Downloaded from https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btz160/5372675 by guest on 02 May 2019

diabetes mellitus (T2DM) patients and controls [Metabolomics data have been deposited to the EMBL-EBI MetaboLights database with the identifier MTBLS1. The complete dataset can be accessed here https://www.ebi.ac.uk/metabolights/MTBLS1]. In total, 132 samples (48 T2DM and 84 controls) were processed using a Galaxy workflow performing conversion, preprocessing, multivariate data analysis and result visualization.

Demonstrator 4: Start-to-end fluxomics workflow. The purpose of this demonstrator was to show the integrated use of separately developed tools covering subsequent steps of the study of metabolic fluxes based on ^{13}C stable isotope-resolved metabolomics (SIRM) (Buescher *et al.*, 2015; King *et al.*, 2015; Niedenführ *et al.*, 2015). Here we implemented the analysis of flux distributions in HUVEC cells under hypoxia [Metabolomics data have been deposited to the EMBL-EBI MetaboLights database with the identifier MTBLS412. The complete dataset can be accessed here https://www.ebi.ac.uk/metabolights/MTBLS412], from raw mass spectra contained in netCDF files, using a workflow implemented in Galaxy including reading and extraction of the data, correcting the evaluated mass spectra for natural isotopes and computing steady-state distribution of ^{13}C label as function of steady-state flux distribution.

2.4 Availability and implementation

The PhenoMeNal consortium maintains a web portal (https://portal.phenomenal-h2020.eu) providing a GUI for launching VREs using KubeNow (Capuccini *et al.*, 2018) on a selection of the largest public cloud providers, including Amazon Web Services, Microsoft Azure and Google Cloud Platform, or on private OpenStack-based installations. The Wiki containing documentation is also hosted on GitHub https://github.com/phnmnl/phenomenal-h2020/wiki. The PhenoMeNal Portal can be reached at https://portal.phenomenal-h2020.eu. The public instance of Galaxy is accessible at https://public.phenomenal-h2020.eu. The containers provisioned by PhenoMeNal comprise tools built as open source software that are available in a public repository such as GitHub, and are subject to continuous integration testing. The containers that satisfy testing criteria are pushed to a public container repository, and containers that are included in stable VRE releases are also pushed to Biocontainers (da Veiga Leprevost *et al.*, 2017). The GitHub repository https://github.com/phnmnl/hosts the source code of all development projects. Source code and documentation are available under the terms of the Apache 2.0 license. Integrated open source projects are available under the respective licensing terms. The Demonstrators can be obtained from: Demonstrator 1: https://github.com/phnmnl/MTBLS233-Jupyter; Demonstrator 2: https://public.phenomenal-h2020.eu/u/phenoadmin/w/metabolomics-lcmsms-processing-quantification-annotation-identification-and-statistics-1; Demonstrator 3: https://public.phenomenal-h2020.eu/u/phenoadmin/w/metabolomics-nmr-rnmr1d-metabolights-data-processing-and-plot; Demonstrator 4: https://public.phenomenal-h2020.eu/u/phenoadmin/w/fluxomics-stationary-13c-ms-iso2flux-with-visualization

3 Results

We developed a VRE based on a microservices architecture encapsulating a large suite of software tools for performing metabolomics data analysis (see Supplementary Table S1). Scientists can interact with the microservices programmatically via an Application Programming Interface (API) or via a web-based graphical user interface (GUI), as illustrated in Figure 1. To connect microservices into computational workflows, the two frameworks Galaxy (Goecks *et al.*, 2010) and Luigi (https://github.com/spotify/luigi) were adapted to execute jobs on Kubernetes. Galaxy is a web-based interface for individual tools and allows users to share workflows, analysis histories and result datasets. Luigi on the other hand focuses on scheduled execution, monitoring, visualization and the implicit dependency resolution of tasks (Leipzig, 2017). These basic infrastructure services, together with the Jupyter notebook (Kluyver *et al.*, 2016) interactive programming environment, are deployed as long-running services in the VRE, whereas the other analysis tools are deployed as transient compute jobs to be used on-demand. System and client applications were developed for launching the VRE on desktop computers, public and private cloud providers, automating all steps required to instantiate the virtual infrastructures.

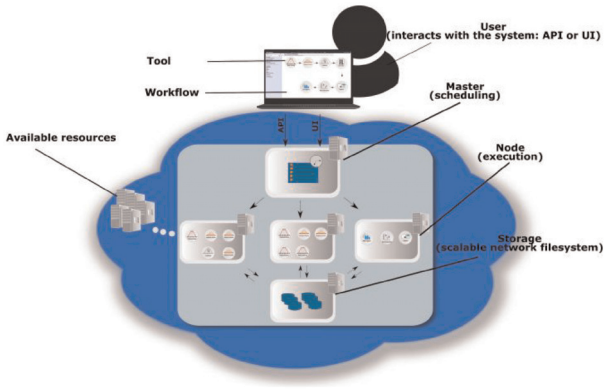


Fig. 1. Overview of the components in a microservices-based framework. Complex applications are divided into smaller, focused and well-defined (micro-) services. These services are independently deployable and can communicate with each other, which allows to couple them into complex task pipelines, i.e. data processing workflows. The user can interact with the framework programmatically via an Application Program Interface (API) or via a graphical user interface (GUI) to construct or run workflows of different services, which are executed independently. Multiple instances of services can be launched to execute tasks in parallel, which effectively can be used to scale analysis over multiple compute nodes. When run in an elastic cloud environment, virtual resources can be added or removed depending on the computational requirements

Demonstrator 1: Scalability of microservices in a cloud environment. The Diagram of scalability-testing on the metabolomics dataset is illustrated in Figure 2. The analysis resulted to WSE of 88% with an execution time of approximately four hours (online methods, Supplementary Fig. S2), compared with the ideal case of 100% where linear scaling is achieved if the run time stays constant while the workload is increased. In addition, the final result of the workflow (online methods, Supplementary Fig. S3) was identical to that presented by the original MTBLS233 study (Ranninger *et al.*, 2016) in negative ionization mode. However, in the positive ionization mode, one *m/z* feature was found in a different group (*m/z* range 400–1000) than it was originally reported by Ranninger *et al.* (*m/z* range 200–400).

Demonstrator 2: Interoperability of microservices. We developed a start to end workflow for pre-processing and statistical analysis of LC-MS metabolomics data (Fig. 3). The workflow allows seamless integration of six major metabolomics data analysis components (26 steps) each was already implemented in independent software suites: noise reduction and filtering [OpenMS (Rost *et al.*, 2016)], quantification, alignment and matching [XCMS

Downloaded from https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btz160/5372675 by guest on 02 May 2019

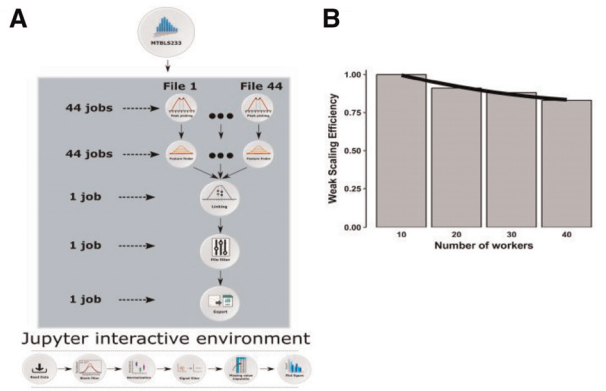


Fig. 2. Diagram of scalability-testing on a metabolomics dataset (MetaboLights ID: MTBLS233) in Demonstrator 1 to illustrate the scalability of a microservice approach. A) The preprocessing workflow is composed of 5 OpenMS tasks that were run in parallel over the 12 groups in the dataset using the Luigi workflow system. The first two tasks, peak picking (528 tasks) and feature finding (528 tasks), are trivially parallelizable, hence they were run concurrently for each sample. The subsequent feature linking task needs to process all of the samples in a group at the same time, therefore 12 of these tasks were run in parallel. In order to maximize the parallelism, each feature linker container (microservice) was run on 2 CPUs. Feature linking produces a single file for each group, that can be processed independently by the last two tasks: file filter (12 tasks) and text exporter (12 tasks), resulting in total of 1092 tasks. The downstream analysis consisted of 6 tasks that were carried out in a Jupyter Notebook. Briefly, the output of preprocessing steps was imported into R and the unstable signals were filtered out. The missing values were imputed and the resulting number of features were plotted. B) The weak scaling efficiency plot for Demonstrator 1. Given the full MTBLS233 dataset, the preprocessing was run on 40 Luigi workers. Then for 1/4, 2/4, 3/4 of MTBLS233, the analysis was run again on 10, 20 and 30 workers respectively. For each run, we measured the processing time T10, T20, T30 and T40, and we computed the WSEn = T10/Tn for n = 10, 20, 30, 40. The WSE plot shows scalability up to 40 CPUs, where we achieved ~88% scaling efficiency. The running time for the full dataset (a total of 1092 tasks) on 40 workers was ~4 hours

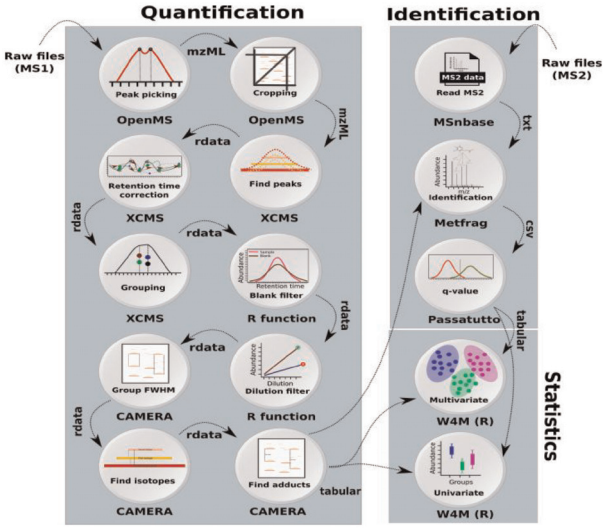


Fig. 3. Overview of the workflow used to process multiple-sclerosis samples in Demonstrator 2, where a workflow was composed of the microservices using the Galaxy system. The data was centroided and limited to a specific mass over charge (*m*/*z*) range using OpenMS tools. The mass traces quantification and retention time correction was done via XCMS (Smith *et al.*, 2006). Unstable signals were filtered out based on the blank and dilution series samples using an in-house function (implemented in R). Annotation of the peaks was performed using CAMERA (Kuhl *et al.*, 2012). To perform the metabolite identification, the tandem spectra from the MS/MS samples in mzML format were extracted using MSnbase and passed to MetFrag. The MetFrag scores were converted to q-values using Passatutto software. The result of identification and quantification were used in 'Multivariate' and 'Univariate' containers from Workflow4Metabolomics (Giacomoni *et al.*, 2015) to perform Partial Least Squares Discriminant Analysis (PLS-DA)

(Smith *et al.*, 2006)], filtering of biological non-relevant signals (R), annotation of signals [CAMERA (Kuhl *et al.*, 2012)], identification [MetFrag (Wolf *et al.*, 2010)], statistics [Workflow4Metabolomics (Giacomoni *et al.*, 2015)]. The result of the workflow (multivariate analysis) showed a clear difference in the metabolic constitution between the three disease groups of RRMS, SPMS and non-multiple sclerosis controls (Fig. 4A). In addition, the univariate analysis resulted in a total of three metabolites being significantly altered ($p < 0.05$) between multiple sclerosis subtypes and control samples, namely alanyltryptophan and indoleacetic acid with higher and linoleoyl ethanolamide with lower abundance in both RRMS and SPMS compared to controls (Fig. 4B).

Demonstrators 3 and 4: Domain agnosticity (NMR and fluxomics workflows). We developed a workflow for analysis of 1D NMR data. The workflow consisted of automatic downloading NMR vendor data (and metadata) from MetaboLights database followed by format standardization, spectral processing and statistical analysis. We processed a NMR dataset (demonstrator 3) resulting to quantification of a total of 726 features which were used to perform Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA). This resulted in a clear separation between T2DM and controls (Fig. 5), similar to that of previous findings (Salek *et al.*, 2007). Lastly, we designed a workflow for analyzing metabolite metabolic fluxes. The workflow integrated four main steps including data extraction, data correction, calculation of flux distribution and visualization. Using this workflow (Fig. 6), we achieved detailed description of the magnitudes of the fluxes through the reactions accounting for glycolysis and pentose phosphate pathway.

4 Discussion

Implementing the different tools and processing steps of a data analysis workflow as separate services that are made available over a network was in the spotlight in the early 2000s (Foster, 2005) as service-oriented architectures (SOA) in science. At that time, web services were commonly deployed on physical hardware and exposed and consumed publicly over the internet. However, it soon became evident that this architecture did not fulfill its promises as it was hard to scale from a computational and maintainability perspective. In addition, the web services were not portable and mirroring them was complicated (if at all possible). Furthermore, API changes and frequent services outage made it frustrating to connect them into functioning computational workflows. Ultimately, the ability to replicate an analysis on local and remote hardware (such as a computer cluster) was very difficult due to heterogeneity in the computing environments.

At first sight microservices might seem similar to above mentioned SOA web services, but microservices can with great benefit be executed in virtual environments (abstracting over OS and hardware architectures) in such a way that they are only instantiated and executed on-demand, and then terminated when they are no longer needed. This makes such virtual environments inherently portable and they can be launched on demand on different platforms

Downloaded from https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btz160/5372675 by guest on 02 May 2019

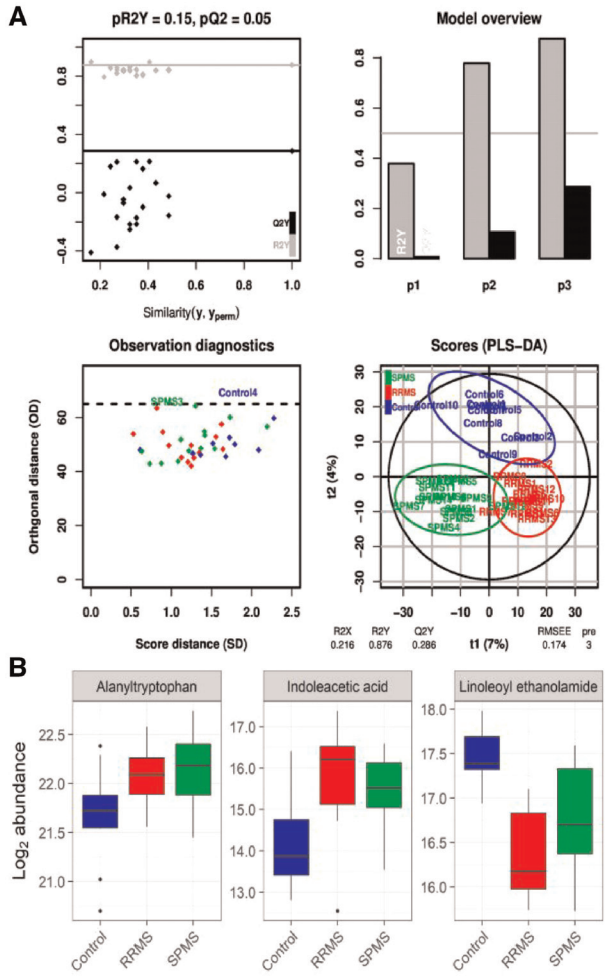


Fig. 4. The results from analysis of multiple sclerosis data in Demonstrator 2, presenting new scientifically useful biomedical knowledge. **A**) The PLS-DA results suggest that the metabolite distribution in the RRMS and SPMS samples are different to controls. **B**) Three metabolites were identified as differentially regulated between multiple sclerosis subtypes and control samples, namely Alanyltryptophan and Indoleacetic acid with higher and Linoleoyl ethanolamide with lower abundance in both RRMS and SPMS compared to controls. Abbr., RRMS: relapsing-remitting multiple sclerosis, SPMS: secondary progressive multiple sclerosis

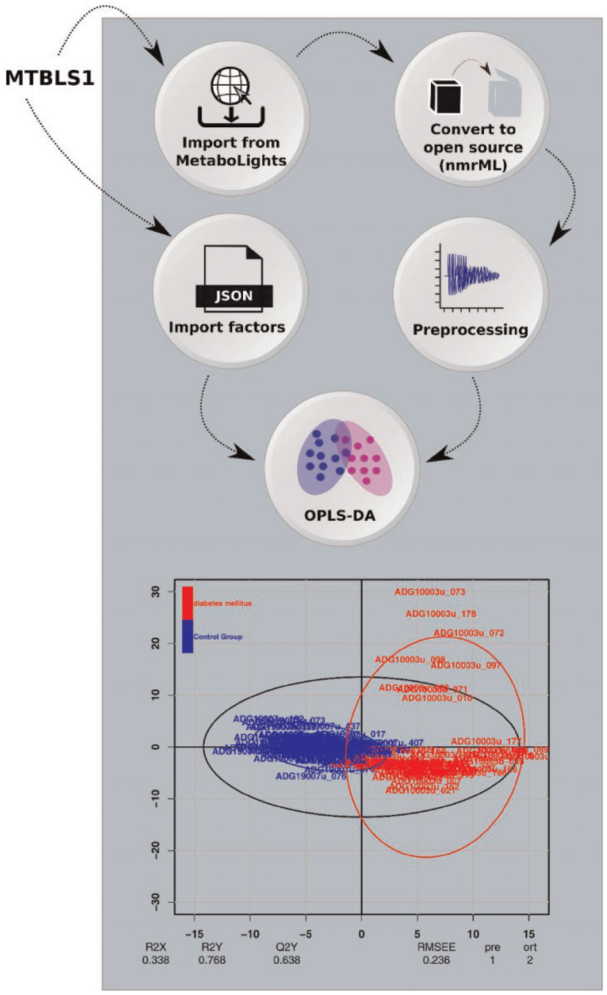


Fig. 5. Overview of the NMR workflow in Demonstrator 3. The raw NMR data and experimental metadata (ISATab) was automatically imported from the Metabolights database and converted to open source nmrML format. The preprocessing was performed using the rnmr1d package part of nmrprocflow tools. All study factors were imported from MetaboLights and were fed to the multivariate node to perform an OPLS-DA

(e.g. a laptop, a powerful physical server or an elastic cloud environment). A key aspect is that workflows of microservices are still executed identically, agnostic of the underlying hardware platform. Container-based microservices provide a wide flexibility in terms of versioning, allowing the execution of newer and older versions of each container as needed for reproducibility. Since all software dependencies are encompassed within the container, which is versioned, the risk of workflow failure due to API changes is minimized. An orchestration framework such as Kubernetes further allows for managing errors in execution and transparently handles the restarting of services. Hence, technology has caught up with service-oriented science, and microservices have taken the methodology to the next level, alleviating many of the previous problems related to scalability, portability and interoperability of software tools. This is advantageous in the context of omics analysis, which produces multidimensional datasets reaching beyond gigabytes, on into terabytes, leading to ever-increasing demand on processing performance (Marx, 2013; Schadt *et al.*, 2010). However, containerization does not address how services communicate with each other, but this has to be implemented inside the container itself. Traditional web services addressed this by standardizing the messaging protocol and public-facing interfaces (e.g. SOAP and WSDL) (Stockinger *et al.*, 2008), while in a containerized environment Representational State Transfer (REST) (Fielding, 2000) or passing files by reference to a shared file system is more common. In Demonstrator 1, we showed that microservices enable highly efficient and scalable data analyses by executing individual modules in parallel, and that they effectively harmonize with on-demand elasticity of the cloud computing paradigm. The reached scaling efficiency of ~88% indicates remarkable performance achieved on generic cloud providers. Furthermore, although our results in positive ionization model was slightly different to that of Ranninger *et al.* (2016), the results of our analysis were replicable regardless of the platform used to perform the computations.

In addition to the fundamental demand for high performance, the increased throughput and complexity of omics experiments has

Downloaded from https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btz160/5372675 by guest on 02 May 2019

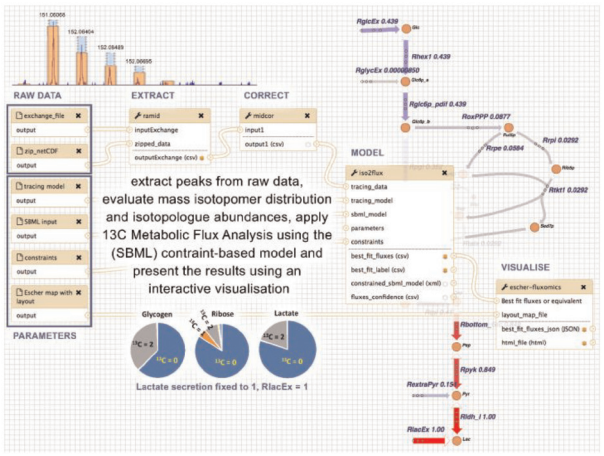


Fig. 6. Overview of the workflow for fluxomics, with Ramid, Midcor, Iso2Flux and Escher-fluxomics tools supporting subsequent steps of the analysis. The example refers to HUVEC cells incubated in the presence of [1,2-$^{13}C_2$]glucose and label (^{13}C) propagation to glycogen, RNA ribose and lactate measured by mass spectrometry. Ramid reads the raw netCDF files, corrects baseline and extracts the peak intensities. The resulting peak intensities are corrected (natural abundance, overlapping peaks) by Midcor, which provides isotopologue abundances. Isotopologue abundances, together with a model description (SBML model, tracing data, constraints), are used by Iso2Flux to provide flux distributions through glycolysis and pentose-phosphate pathways, which are shown as numerical values associated to a metabolic scheme of the model by the Escher-fluxomics tool

led to a large number of sophisticated computational tools (Berger *et al.*, 2013), which in turn necessitates integrative workflow engines (Atkinson *et al.*, 2017; Di Tommaso *et al.*, 2017; Liew *et al.*, 2016). In order to integrate new tools in such workflow engines, compatibility of the target environment, tools and APIs needs to be considered (Di Tommaso *et al.*, 2017). Containerization facilitates this by providing a platform-independent virtual environment for developing and running the individual tools. However, the problem of compatibility between tools/APIs and data formats remains and needs to be tackled by international consortia (Wilkinson *et al.*, 2016). Our methodology the currently non-trivial task of instantiating the complete microservice environment through a web portal that allows for convenient deployment of the VRE on public cloud providers. Moreover, using this web portal, microservices and VREs can be deployed on a trusted private cloud instance or a local physical server on an internal network, such as within a hospital network, allowing for levels of isolation and avoiding transfer of data across untrusted networks which often are requirements in the analysis of sensitive data. This was exemplified in Demonstrator 2, where a complete start-to-end workflow was run on the Galaxy platform on a secure server at Uppsala University Hospital, Sweden, leading to the identification of novel disease fingerprints in the CSF metabolome of RRMS and SPMS patients. It is worth mentioning that the selected metabolites were part of the tryptophan metabolism (alanyltryptophan and indoleacetic acid) and endocannabinoids (linoleoyl ethanolamide), both of which have been previously implicated in multiple sclerosis (Amirkhani *et al.*, 2005; Baker and Pryce, 2008; Centonze *et al.*, 2007; Lim *et al.*, 2017; Lovelace *et al.*, 2016; Zamberletti *et al.*, 2012). However, since the cross-validated predictive performance ($Q2Y = 0.286$) is not much higher than some of the models generated after random permutation of the response (Fig. 4A), the quality of the model needs to be confirmed in a future study on an independent cohort of larger size.

The microservice architecture is domain-agnostic and not limited to a particular assay technology, i.e. mass spectrometry. This was showcased in Demonstrator 3 and 4, where an automated 1 D NMR workflow and calculation of flux distributions (derived from the application of stable isotope resolved metabolomics) were performed. In Demonstrator 3, we showed that the pattern of the metabolite expression is different between type 2 diabetic and healthy controls, and that a large number of metabolites contribute to such separation. In Demonstrator 4, we showed a high rate of glycolysis in cells cultured in hypoxia, which is consistent with the one expected for endothelial cells (Iyer *et al.*, 1998) and with how these cells maintain energy in low oxygen environments and without oxidative phosphorylation (Eelen *et al.*, 2015; Polet and Feron, 2013). These two examples further show that complex workflows can be applied with minimal effort on other studies (i.e. simply by providing a MetaboLights accession number), leading to the capability to reanalyze data and compare the results with the original publication findings. Furthermore, it demonstrates the value of standardised dataset descriptions such as nmrML (Schober *et al.*, 2017) and ISA format (Rocca-Serra *et al.*, 2016; Sansone *et al.*, 2012) for representing NMR based studies, as well as the potential of the VRE to foster reproducibility. Furthermore, the data processing steps are trackable and replicable as each container/tool is versioned for a specific release and data processing steps and the corresponding parameters are taken care of by the workflow engine. In addition, the cli KubeNow is using speciffic pinned versions of all dependant software and all versions of software is stored in the user config dir created by the init-command. The specific version of KubeNow used is saved in user config directory.

While microservices are not confined to metabolomics and generally applicable to a large variety of applications, there are some important implications and limitations of the method. Firstly, tools need to be containerized in order to operate in the environment. This is however not particularly complex, and an increasing number of developers provide containerized versions of their tools on public container repositories such as Dockerhub or Biocontainers (da Veiga Leprevost *et al.*, 2017). Secondly, uploading data to a cloud-based system can take a considerable amount of time, and having to re-do this every time a VRE is instantiated can be time-consuming. This can be alleviated by using persistent storage on a cloud resource, but the availability of such storage varies between different cloud providers. Further, the storage system can become a bottleneck when many services try to access a shared storage. We observe that using a distributed storage system with multiple storage nodes can drastically increase performance, and the PhenoMeNal VRE comes with a distributed storage system by default. When using a workflow system to orchestrate the microservices, stability and scalability are inherently dependent on the workflow system's job runner. Workflow execution is dependent on the underlying workflow engine, and we observed that a large number of outputs can make the Galaxy engine unresponsive, whereas the Luigi engine did not have these shortcomings. With clouds and microservices maturing, workflow systems will need to evolve and further embrace the new possibilities of these infrastructures. It is important to note that microservices do not overcome the incompatibility between tools with respect to using different data formats, and code resolving such incompatibility is still needed. However, using a shared platform makes such bridging components easier to maintain and makes them reusable. There remains great challenges in establishing interoperable and agreed-upon standards and data formats that are widely accepted and implemented by tools, as well as achieving complete support for the FAIR principles (Wilkinson *et al.*, 2016). Further,

Downloaded from https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btz160/5372675 by guest on 02 May 2019

not all research can be easily pipelined, for example exploratory research might be better carried out in an ad-hoc manner than with workflows and the overhead this implies. A Jupyter Notebook as used in in Demonstrator 1 or embedded in Galaxy (Grüning *et al.*, 2017) constitutes a promising way to make use of microservices for interactive analysis. The serverless architecture, also called Functions as a Service (FaaS) architecture, is an interesting methodology when deployed with microservices as it allow developers to execute code in response to events without managing the underlying infrastructure. While serverless technologies have irrupted strongly in areas of software engineering closer to web development, this doesn't mean that their usage can be easily transferred to scientific workloads. This is due to the far more complex network of dependencies that scientific software will have compared to web applications, where large applications can be managed for instance through npm package resolutions only. On scientific software solutions one will commonly find dependencies in different programming languages, different underlying libraries and even sometimes on different incompatible versions of the same frameworks. This level of complexity is not resolvable today through server less approaches and requires more isolated approaches based on containers, such as the one presented here.

In summary, we showed that microservices allow for efficient horizontal scaling of analyses on multiple computational nodes, enabling the processing of large datasets. By applying a number of data [mzML (Martens *et al.*, 2011), nmrML] and metadata standards [ISA serializations for study descriptions (Rocca-Serra *et al.*, 2016; Sansone *et al.*, 2012)], we also demonstrated a high level of interoperability in the context of metabolomics, by providing completely automated start-to-end analysis workflows for mass spectrometry and NMR data. In addition, many of the state-of-the-art tools such as components of XCMS ONLINE (Warth *et al.*, 2017) and MetaboAnalyst (Xia *et al.*, 2012) can be incorporated in the workflows, providing more refined workflows. The ability to instantiate VREs close to large datasets, such as on local servers within a hospital for Demonstrator 2, makes it possible to use the VRE on sensitive data that is not allowed to leave the current environment for ELSI reasons. While the current PhenoMeNal VRE implementation uses Docker for software containers and Kubernetes for container orchestration, the microservice methodology is general and not restricted to these frameworks. Likewise, the choice of Luigi and Galaxy was here used to demonstrate the capabilities of workflow management microservices in cloud environments. In fact, our microservice architecture supports other major workflow engines such as Nextflow (Di Tommaso *et al.*, 2017) or Snakemake (Köster and Rahmann, 2012). Hence it is possible to use any of such workflow engines in our VRE and still produce reproducible results. In addition, despite some of our workflows were novel in the context of metabolomics (e.g. Demonstrator 2) and can be readily applied on other datasets, their main contribution in this work is to showcase scalability and interoperability of the microservices methodology. Finally, we emphasise that the presented methodology goes beyond metabolomics and can be applied to virtually any field, lowering the barriers for taking advantage of cloud infrastructures and opening up for large-scale integrative science.

Author contributions

KK, MAC, MC, PEK, SH contributed to Demonstrator 1. CR, KK, KP, PEK, SH, SN contributed to Demonstrator 2. KK designed the study in Demonstrator 2. JB performed collection of samples and characterization of the multiple sclerosis cohort. SH performed the mass spectrometry experiment in Demonstrator 2. DS, KP, PEK, PM, RMS, contributed to Demonstrator 3. AGB, CF, DJ, MCA, MVV, PDA, PM, PRS, SAS, TH and VS contributed to Demonstrator 4. GZ, LP, PEK and PM contributed to developments of Galaxy in Kubernetes. AL and MC contributed to the development of Luigi in Kubernetes. AL, MAC, MC and NS developed KubeNow. PM contributed to Galaxy-Kubernetes. EAT and PR contributed to containerizing of Workflow4Metabolomics tools. AGB, DJ, PRS and SAS contributed to ISA-API. DJ, EAT, KP, MVV, NS, OS, PEK, PM, PR, PRS, DS, RMS, RR and SB were involved in testing the containers and the VRE. PM, SIH and KH were involved in development and maintenance of the portal. MVV, PM and RMS contributed to the release. NK coordinated the PhenoMeNal project. CS conceived and managed the PhenoMeNal project. OS conceived and coordinated the study and e-infrastructure. All authors contributed to manuscript writing.

Funding

This research was supported by The European Commission's Horizon 2020 programme funded under grant agreement number 654241 (PhenoMeNal), The Swedish Research Council FORMAS, Uppsala Berzelii Technology Centre for Neurodiagnostics, Åke Wiberg Foundation and the Nordic e-Infrastructure Collaboration (NeIC) via the Glenna2 and Tryggve2 projects. We kindly acknowledge contributions by Daniel Jacob (INRA) and to cloud resources by SNIC Science Cloud, Embassy Cloud, c-Pouta and CityCloud. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest: none declared.

References

Allan,R.N. (2009) *Virtual Research Environments: From Portals to Science Gateways*. ChandosŁ Publishing, Oxford, UK.

Amirkhani,A. *et al.* (2005) Interferon-beta affects the tryptophan metabolism in multiple sclerosis patients. *Eur. J. Neurol.*, **12**, 625–631.

Atkinson,M. *et al.* (2017) Scientific workflows: past, present and future. *Future Gener. Comput. Syst.*, **75**, 216–227.

Baker,D. and Pryce,G. (2008) The endocannabinoid system and multiple sclerosis. *Curr. Pharm. Des.*, **14**, 2326–2336.

Berger,B. *et al.* (2013) Computational solutions for omics data. *Nat. Rev. Genet.*, **14**, 333–346.

Buescher,J.M. *et al.* (2015) A roadmap for interpreting (13)C metabolite labeling patterns from cells. *Curr. Opin. Biotechnol.*, **34**, 189–201.

Candela,L. *et al.* (2013) Virtual research environments: an overview and a research agenda. *Data Sci. J.*, **12**, GRDI75–GRDI81.

Capuccini,M. *et al.* (2018) On-Demand Virtual Research Environments using Microservices. arXiv [cs.DC].

Centonze,D. *et al.* (2007) The endocannabinoid system is dysregulated in multiple sclerosis and in experimental autoimmune encephalomyelitis. *Brain*, **130**, 2543–2553.

da Veiga Leprevost,F. *et al.* (2017) BioContainers: an open-source and community-driven framework for software standardization. *Bioinformatics*, **33**, 2580–2582.

Di Tommaso,P. *et al.* (2017) Nextflow enables reproducible computational workflows. *Nat. Biotechnol.*, **35**, 316–319.

Eelen,G. *et al.* (2015) Endothelial cell metabolism in normal and diseased vasculature. *Circ. Res.*, **116**, 1231–1244.

Fielding,R.T. (2000) *Architectural Styles and the Design of Network-Based Software Architectures*. Irvine: University of California, Irvine.

Foster,I. (2005) Service-oriented science. *Science*, **308**, 814–817.

Giacomoni,F. *et al.* (2015) Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*, **31**, 1493–1495.

Downloaded from https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btz160/5372675 by guest on 02 May 2019

Goecks,J. *et al.* (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.

Grüning,B.A. *et al.* (2017) Jupyter and Galaxy: easing entry barriers into complex data analyses for biomedical researchers. *PLoS Comput. Biol.*, **13**, e1005425.

Haug,K. *et al.* (2013) MetaboLights–an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.*, **41**, D781–D786.

Iyer,N.V. *et al.* (1998) Cellular and developmental control of O2 homeostasis by hypoxia-inducible factor 1alpha. *Genes Dev.*, **12**, 149–162.

King,Z.A. *et al.* (2015) Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput. Biol*, **11**, e1004321.

Kluyver,T. *et al.* (2016) Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides,F. and Scmidt,B. (eds), *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press, pp. 87–90.

Köster,J. and Rahmann,S. (2012) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, **28**, 2520–2522.

Kuhl,C. *et al.* (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.*, **84**, 283–289.

Langmead,B. and Nellore,A. (2018) Cloud computing for genomic data analysis and collaboration. *Nat. Rev. Genet.*, **19**, 325.

Lawrence,K.A. *et al.* (2015) Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community. *Concurr. Comput.*, **27**, 4252–4268.

Leipzig,J. (2017) A review of bioinformatic pipeline frameworks. *Brief. Bioinform.*, **18**, 530–536.

Liew,C.S. *et al.* (2016) Scientific workflows: moving across paradigms. *ACM Comput. Surv.*, **49**, 1–39.

Lim,C.K. *et al.* (2017) Kynurenine pathway metabolomics predicts and provides mechanistic insight into multiple sclerosis progression. *Sci. Rep.*, **7**, 41473.

Lovelace,M.D. *et al.* (2016) Current evidence for a role of the kynurenine pathway of tryptophan metabolism in multiple sclerosis. *Front. Immunol.*, **7**, 246.

Martens,L. *et al.* (2011) mzML—a community standard for mass spectrometry data. *Mol. Cell. Proteomics*, **10**, R110.000133.

Marx,V. (2013) Biology: the big challenges of big data. *Nature*, **498**, 255–260.

Merkel,D. (2014) Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux Journal*, 1–19.

Montenegro-Burke,J.R. *et al.* (2017) Data streaming for metabolomics: accelerating data processing and analysis from days to minutes. *Anal. Chem.*, **89**, 1254–1259.

Newman,S. (2015) *Building Microservices*. 'O'Reilly Media, Inc, Sebastopol, CA.

Nicholson,J.K. and Wilson,I.D. (2003) Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discov.*, **2**, 668–676.

Niedenführ,S. *et al.* (2015) How to measure metabolic fluxes: a taxonomic guide for 13 C fluxomics. *Curr. Opin. Biotechnol.*, **34**, 82–90.

Polet,F. and Feron,O. (2013) Endothelial cell metabolism and tumour angiogenesis: glucose and glutamine as essential fuels and lactate as the driving force. *J. Intern. Med.*, **273**, 156–165.

Ranninger,C. *et al.* (2016) Improving global feature detectabilities through scan range splitting for untargeted metabolomics by high-performance liquid chromatography-Orbitrap mass spectrometry. *Anal. Chim. Acta*, **930**, 13–22.

Rocca-Serra,P. *et al.* (2016) Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics*, **12**, 14.

Rost,H.L. *et al.* (2016) OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods*, **13**, 741–748.

Salek,R.M. *et al.* (2007) A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. *Physiol. Genomics*, **29**, 99–108.

Sansone,S.-A. *et al.* (2012) Toward interoperable bioscience data. *Nat. Genet.*, **44**, 121–126.

Schadt,E.E. *et al.* (2010) Computational solutions to large-scale data management and analysis. *Nat. Rev. Genet.*, **11**, 647–657.

Schober,D. *et al.* (2017) nmrML: a community supported open data standard for the description, storage, and exchange of NMR data. *Anal. Chem.*, **90**, 649–656.

Silver,A. (2017) Software simplified. *Nature*, **546**, 173–174.

Smith,C.A. *et al.* (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.*, **78**, 779–787.

Stockinger,H. *et al.* (2008) Experience using web services for biological sequence analysis. *Brief. Bioinform.*, **9**, 493–505.

Sturm,M. *et al.* (2008) OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinformatics*, **9**, 163.

Sud,M. *et al.* (2016) Metabolomics Workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.*, **44**, D463–D470.

Suplatov,D. *et al.* (2016) Parallel workflow manager for non-parallel bioinformatic applications to solve large-scale biological problems on a supercomputer. *J. Bioinform. Comput. Biol.*, **14**, 1641008.

Waldrop,M.M. (2013) Education online: the virtual lab. *Nature*, **499**, 268–270.

Warth,B. *et al.* (2017) Metabolizing data in the cloud. *Trends Biotechnol.*, **35**, 481–483.

Wilkinson,M.D. *et al.* (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*, **3**, 160018.

Wolf,S. *et al.* (2010) In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics*, **11**, 148.

Xia,J. *et al.* (2012) MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. *Nucleic Acids Res.*, **40**, W127–W133. p

Zamberletti,E. *et al.* (2012) The endocannabinoid system and schizophrenia: integration of evidence. *Curr. Pharm. Des.*, **18**, 4980–4990.

Downloaded from https://academic.oup.com/bioinformatics/advance-article-abstract/doi/10.1093/bioinformatics/btz160/5372675 by guest on 02 May 2019

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

p^{13}CMFA: Parsimonious ^{13}C metabolic flux analysis

Carles Foguet[1,2], Anusha Jayaraman[1‡], Silvia Marin[1,2‡], Vitaly A. Selivanov[1,2], Pablo Moreno[3], Ramon Messeguer[4], Pedro de Atauri[1,2]*, Marta Cascante[1,2]*

1 Department of Biochemistry and Molecular Biomedicine & Institute of Biomedicine of University of Barcelona, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain, **2** Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD) and Metabolomics node at Spanish National Bioinformatics Institute (INB-ISCIII-ES-ELIXIR), Instituto de Salud Carlos III (ISCIII), Madrid, Spain, **3** European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom, **4** LEITAT Technological Center, Health & Biomedicine Unit, Barcelona, Spain

‡ These authors contributed equally to this work.
* pde_atauri@ub.edu (PdA); martacascante@ub.edu (MC)

Check for updates

OPEN ACCESS

Citation: Foguet C, Jayaraman A, Marin S, Selivanov VA, Moreno P, Messeguer R, et al. (2019) p^{13}CMFA: Parsimonious ^{13}C metabolic flux analysis. PLoS Comput Biol 15(9): e1007310. https://doi.org/10.1371/journal.pcbi.1007310

Editor: Vassily Hatzimanikatis, Ecole Polytechnique Fédérale de Lausanne, SWITZERLAND

Received: October 23, 2018

Accepted: August 6, 2019

Published: September 6, 2019

Copyright: © 2019 Foguet et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by European Commission (EC-654241, FP7-PEOPLE-264780), MINECO-European Commission FEDER funds – "Una manera de hacer Europa" (SAF2017-89673-R; SAF2015-70270-REDT), Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) – Generalitat de Catalunya (2017SGR1033) and Instituto de Salud Carlos III (CIBEREHD, CB17/04/00023). CF acknowledges the support received

Abstract

Deciphering the mechanisms of regulation of metabolic networks subjected to perturbations, including disease states and drug-induced stress, relies on tracing metabolic fluxes. One of the most informative data to predict metabolic fluxes are ^{13}C based metabolomics, which provide information about how carbons are redistributed along central carbon metabolism. Such data can be integrated using ^{13}C Metabolic Flux Analysis (^{13}C MFA) to provide quantitative metabolic maps of flux distributions. However, ^{13}C MFA might be unable to reduce the solution space towards a unique solution either in large metabolic networks or when small sets of measurements are integrated. Here we present parsimonious ^{13}C MFA (p^{13}CMFA), an approach that runs a secondary optimization in the ^{13}C MFA solution space to identify the solution that minimizes the total reaction flux. Furthermore, flux minimization can be weighted by gene expression measurements allowing seamless integration of gene expression data with ^{13}C data. As proof of concept, we demonstrate how p^{13}CMFA can be used to estimate intracellular flux distributions from ^{13}C measurements and transcriptomics data. We have implemented p^{13}CMFA in Iso2Flux, our in-house developed isotopic steady-state ^{13}C MFA software. The source code is freely available on GitHub (https://github.com/cfoguet/iso2flux/releases/tag/0.7.2).

Author summary

^{13}C Metabolic Flux Analysis (^{13}C MFA) is a well-established technique that has proven to be a valuable tool in quantifying the metabolic flux profile of central carbon metabolism. When a biological system is incubated with a ^{13}C-labeled substrate, ^{13}C propagates to metabolites throughout the metabolic network in a flux and pathway-dependent manner. ^{13}C MFA integrates measurements of ^{13}C enrichment in metabolites to identify the flux distributions consistent with the measured ^{13}C propagation. However, there is often a range of flux values that can lead to the observed ^{13}C distribution. Indeed, either when the metabolic network is large or a small set of measurements are integrated, the range of

through "Becas de la Caixa para estudios de doctorado en universidades españolas" funded by the "La Caixa" foundation. CF also acknowledges the support from the Spanish National Bioinformatics Institute (INB-ISCIII-ES-ELIXIR). MC acknowledges the support received through the prize "ICREA Academia" for excellence in research, funded by ICREA foundation – Generalitat de Catalunya. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

valid solutions can be too wide to accurately estimate part of the underlying flux distribution. Here we propose to use flux minimization to select the best flux solution in the ^{13}C MFA solution space. Furthermore, this approach can integrate gene expression data to give greater weight to the minimization of fluxes through enzymes with low gene expression evidence in order to ensure that the selected solution is biologically relevant. The concept of using flux minimization to select the best solution is widely used in flux balance analysis, but it had never been applied in the framework of ^{13}C MFA. We have termed this new approach parsimonious ^{13}C MFA (p^{13}CMFA).

Introduction

Fluxomics is the omics field that analyses metabolic fluxes (i.e., reaction and transport rates in living cells) which are a close reflection of the metabolic phenotype. As such, quantitative tracking of metabolic fluxes is vital for deciphering the regulation mechanisms of metabolic networks subjected to perturbations, including disease states and drug-induced stress. However, unlike other omics data that can be quantified directly, the fluxome can only be estimated through an indirect interpretation of experimental data[1–3].

There are two main model-based approaches to quantifying metabolic fluxes, Flux Balance Analysis (FBA) and ^{13}C Metabolic Flux Analysis (^{13}C MFA). Both methods use stoichiometric, thermodynamic and experimental constraints to find the range of feasible fluxes across a metabolic network and then find the flux distributions within that space that optimize a given objective function. However, both techniques differ in the type of objective function optimized.

In FBA, the objective function is a set of fluxes to be minimized or maximized. These fluxes must represent a biological objective deemed desirable in the conditions of study (e.g., synthesis of biomass components for proliferating systems)[4]. A significant limitation of FBA is that the choice of objective(s) can significantly influence the predicted flux distributions.

In ^{13}C MFA, the objective function is to minimize the difference between simulated and measured ^{13}C enrichment in metabolites [5,6]. ^{13}C enrichment is quantified in metabolic products and intermediates after incubating samples with metabolic substrates labeled with ^{13}C (tracers) and provides information about how carbons are redistributed along metabolic pathways[7]. Compared to FBA, ^{13}C MFA has a greater capacity to elucidate the fluxes of central carbon metabolism. However, ^{13}C MFA is more complex to solve than FBA due to the non-linear nature of the ^{13}C MFA objective.

A significant limitation of FBA is that there is generally a wide range of optimal flux distributions[8]. This is not usually the case with ^{13}C MFA which can generally determine flux distributions with a high degree of accuracy. ^{13}C MFA achieves this by integrating large sets of measured isotopologue fractions from parallel experiments with tracers optimized for different parts of the network[9–16]. However, when ^{13}C MFA is used in large metabolic networks and with a limited set of measurements, it can also suffer from the same limitation as FBA and result on a wide interval of flux values for part of the metabolic network[5,17–19].

On FBA, an approach to reduce the range of optimal solutions consists in running a second optimization step on the optimal solution range. One of such methods is parsimonious FBA (pFBA)[20]. This approach, which follows the principle of parsimony or simplicity, consists on finding the optimal value of the primary objective function through FBA and then running a second optimization step where the sum of reaction fluxes is minimized while maintaining the optimal primary objective. The GIMME (and its derivative GIM^{3}E) algorithms[21,22] are

based on a similar principle as pFBA. Unlike standard pFBA, where all reactions fluxes are minimized with equal weight, GIMME integrates gene expression data to give greater weight to the minimization of fluxes through reactions catalyzed by lowly expressed enzymes. Different to FBA, for ^{13}C MFA, there is currently no approach that relies on a second optimization to reduce the solution space when experimental data is insufficient to constrain the system towards a unique solution.

In addition to model-based approaches (e.g., FBA or ^{13}C MFA), metabolic fluxes can also be analyzed through the direct or semidirect interpretation of ^{13}C data. This approach primarily consists of predicting the contribution of a labeled substrate to the synthesis of a given metabolite (nutrient contribution) and predicting the relative activity of pathways (pathway activity analysis). Pathway activity analysis assumes that the isotopologue fractions used as a surrogate for the pathways of interest are primarily generated through them. This assumption is generally based on the assertion that the pathways of interest are the most direct way to generate such fractions from the labeled substrate used in the experiment[2,7,23–25]. Unlike ^{13}C MFA, direct interpretation of ^{13}C data is generally not able to quantify network-wide flux maps. Instead, it provides a series of qualitative or semiquantitative flux predictions around each analyzed metabolite. Strategies that couple direct interpretation of ^{13}C data to regression and correlation analyses are widely applied to unveil the effect of an external perturbation, such as a therapeutic intervention, on central carbon metabolism[26–30].

Here we present parsimonious ^{13}C MFA (p^{13}CMFA), a new model-based approach to flux estimation. p^{13}CMFA first minimizes the difference between experimental and simulated ^{13}C enrichment in metabolites (^{13}C MFA) and then applies the flux minimization principle to select the best solution among the solutions that fit experimental ^{13}C data. Hence, p^{13}CMFA can be used to select the best flux map in instances where experimental ^{13}C measurements are not enough to fully constrain the ^{13}C MFA solution space. Furthermore, the minimization can be weighted by gene expression allowing seamless integration of ^{13}C with gene expression data (Fig 1).

We have implemented p^{13}CMFA in Iso2Flux, our in-house developed isotopic steady-state ^{13}C MFA software (https://github.com/cfoguet/iso2flux/releases/tag/0.7.2). As a proof of concept, we have applied it to the analysis of the metabolic flux distribution in HUVECs (Human umbilical vein endothelial cells) through the integration of a small set of ^{13}C enrichment measurements and transcriptomics data. Furthermore, we validated the predictive capacity of p^{13}CMFA using data from a published study of HTC116 cells where fluxes had been estimated with a high degree of confidence[14]. Using only a small subset of the measurements from such study, p^{13}CMFA was able to achieve significantly better flux predictions than both ^{13}C MFA and GIMME.

Results

Description of the p^{13}CMFA approach

p^{13}CMFA consists of two consecutive optimizations: first, the optimal solution to the ^{13}C MFA problem is identified (Eq 1); secondly, the weighted sum of reaction fluxes is minimized within the optimal solution space of ^{13}C MFA (Eq 2).

The ^{13}C MFA optimization (Eq 1) identifies the flux distribution that minimizes the difference between measured and simulated isotopologue fractions [5,7]:

$$X_{opt} = min \sum_{j} \left(\frac{E_j - Y_j(v)}{\sigma_j} \right)^2 \quad \text{(Eq1)}$$

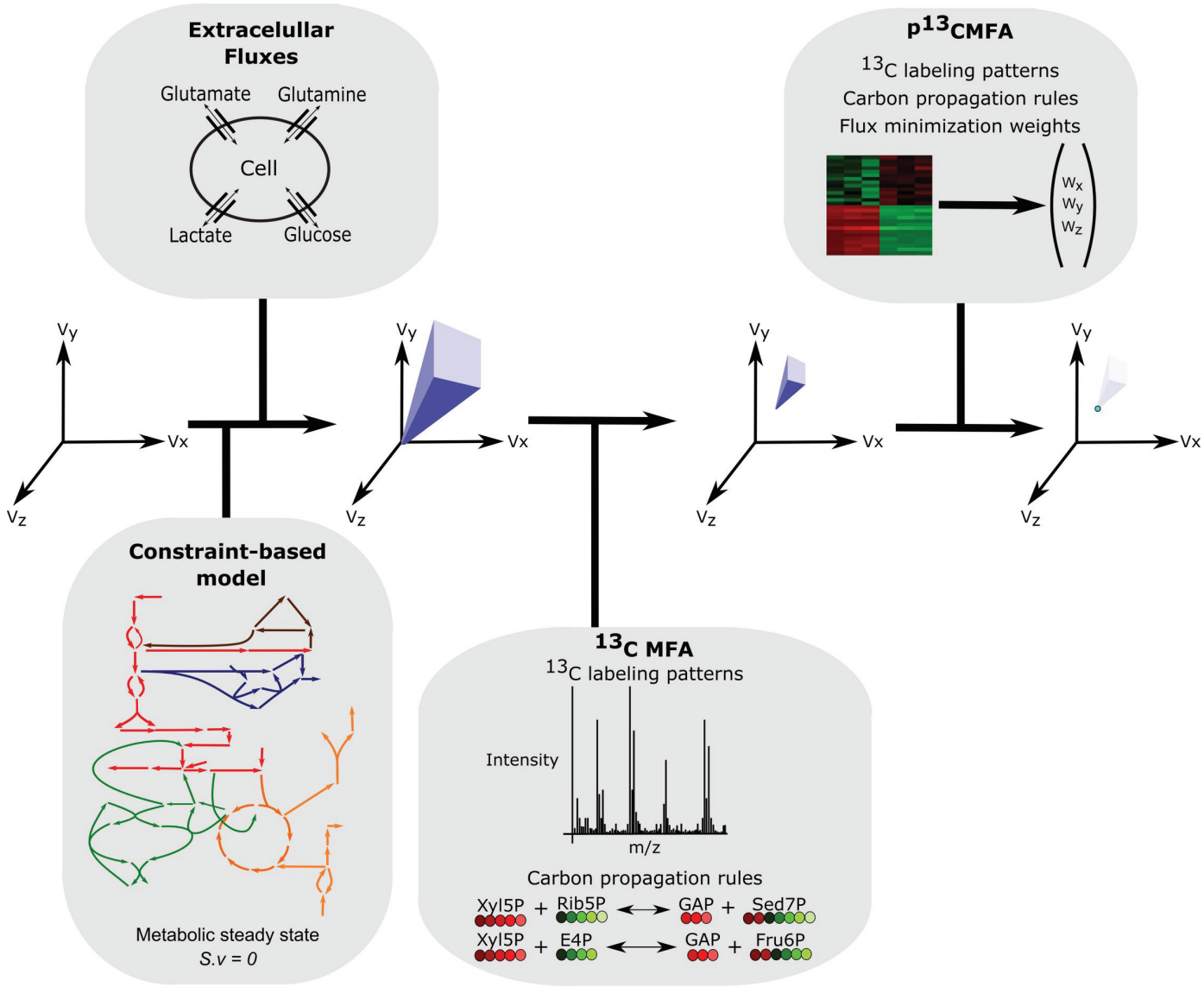


Fig 1. The conceptual basis of p[13]CMFA. From an infinite space of flux (v) solutions, a space of feasible solutions is obtained through the integration of stoichiometric and thermodynamic constraints (in the form of a constraint-based model) and the measured extracellular fluxes. Applying ^{13}C MFA to integrate experimental ^{13}C data can further reduce the solution space to those flux distributions that are consistent with such data. Through flux minimization, p^{13}CMFA can identify the optimal flux distribution that lies on the edge of the ^{13}C MFA solution space. Such minimization can be weighted according to the gene expression evidence for each enzyme.

https://doi.org/10.1371/journal.pcbi.1007310.g001

$$Subject\ to\ S.v = 0, \quad lb \leq v \leq ub$$

where,

v is a vector of flux values describing a valid steady-state flux distribution;

X_{opt} is the optimal value of the ^{13}C MFA objective;

E_j is the experimentally quantified fraction for isotopologue j;

$Y_j(v)$ is the simulated isotopologue fraction for isotopologue j with flux distribution v. Such simulation is performed by solving a complex non-linear system of equations built around isotopologues balances [1].

σ_j is the experimental standard deviation of the measurements of isotopologue *j*;

S is the stoichiometric matrix;

lb and *ub* are vectors defining the upper and lower bounds for flux values. Flux bounds can be used to integrate experimental flux measurements;

Either in large metabolic networks or when small sets of ^{13}C measurements are integrated, the ^{13}C MFA problem can be undetermined and there can be a wide range of possible solutions. Such indetermination emerges from cycles and alternative pathways in the metabolic network, which lead to many possible flux combinations that can result in the measured ^{13}C label patterns. Furthermore, many of the ^{13}C MFA solutions can involve large fluxes through futile cycles, which are usually artifacts of the optimization process as *in vivo* enzyme activities cannot support such large flux values. Therefore, to select the best solution among the many solutions that fit experimental ^{13}C data, p^{13}CMFA runs a second optimization where the weighted sum of fluxes is minimized (Eq 2):

$$\min \sum_i |v_i| \cdot w_i \qquad \text{(Eq2)}$$

$$\textit{subject to } S.v = 0,\ lb \le v \le ub, \quad \sum_j \left(\frac{E_j - Y_j(v)}{\sigma_j} \right)^2 \le X_{opt} + T$$

where:

w_i is the weight given to the minimization of flux through reaction *i*;

T is the maximum value that the ^{13}C MFA objective can deviate from its optimal value (primary objective tolerance) when fluxes are minimized;

The difference between the optimal ^{13}C MFA objective function value and the objective function value when the total reaction flux is minimized can be assumed to follow a Chi^2-distribution with one degree of freedom. Therefore, setting T to 3.84 gives a p^{13}CMFA solution within the 95% confidence intervals of ^{13}C MFA[5].

With p^{13}CMFA, the activity through cycles is minimized to the minimum amount needed to account for experimental measurements. Furthermore, gene expression measurements can be integrated to give greater weight to the minimization of fluxes through reactions catalyzed by lowly expressed enzymes. Then, in instances where multiple pathways can result in similar label patterns, those pathways with stronger gene expression evidence are selected. Hence, p^{13}CMFA reduces the solution space towards a unique solution without requiring a simplification of the metabolic network or additional ^{13}C measurements (Fig 1).

Example of p^{13}CMFA usage

As an example of a potential application of p^{13}CMFA, we applied it to analyze the metabolic flux distribution in HUVECs using a publicly available dataset not large enough to make meaningful flux predictions with conventional ^{13}C MFA.

In this study, available in the MetaboLights repository[31] (accession number MTBLS412), HUVECs were incubated in the presence of the tracer [1,2-$^{13}C_2$]-glucose, and the relative abundance of ^{13}C isotopologues was quantified in glycogen, ribose, lactate, and glutamate. The rates of production/consumption of glucose, glycogen, lactate, glutamate, and glutamine were also quantified. The data were integrated into a stoichiometric model of central metabolism which includes glycolysis, glycogen metabolism, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, fatty acid synthesis, and energy and redox metabolism (S1 ZIP).

To predict the flux distribution using conventional ^{13}C MFA, 95% confidence intervals were computed for each predicted flux value. From this analysis, the space of flux solutions

consistent with the measured ^{13}C enrichment was estimated. The resulting space of solution was still mostly undetermined and, in general, ^{13}C MFA was unable to significantly constraint the flux ranges emerging from the stoichiometric and thermodynamic constraints and the measured extracellular fluxes (Fig 2, S1 Table). For instance, despite integrating measurements of ^{13}C enrichment in ribose, it was not possible to conclude whether the oxidative branch of the pentose phosphate pathway contributed more to *de novo* ribose synthesis than the non-oxidative branch or vice versa.

Nevertheless, p^{13}CMFA can be applied to select the best solution in the ^{13}C MFA solution space. With this aim, transcriptomic data taken from the literature[32] were used to add additional penalties to the flux through lowly expressed enzymes. Indeed, by applying p^{13}CMFA, we can now conclude that, under the condition of the study, glucose is mostly directed towards lactate production except for a small part going to the TCA cycle through pyruvate dehydrogenase (Fig 2, Fig 3). Glutamine is mainly metabolized to glutamate or directed to glycolysis through the TCA cycle and phosphoenolpyruvate carboxykinase. In the PPP, the non-oxidative branch contributes to roughly 60% of the net ribose synthesis. Only the glycogen phosphorylase/glycogen synthase futile cycle is predicted to be active, while the remaining futile cycles (i.e., the hexokinase/glucose 6-phosphatase, phosphofructokinase/fructose bis-phosphatase, pyruvate carboxylase/phosphoenolpyruvate carboxykinase, and glutaminase/glutamine synthase cycles) are predicted to be inactive. Concerning redox metabolism, most of the reduced NAD+ (NADH) produced in the mitochondria is exported to the cytosol through the malate-aspartate shuttle, where it is used to reduce pyruvate to lactate.

To evaluate the contribution of ^{13}C MFA to the p^{13}CMFA solution, GIMME (i.e., flux minimization weighted by gene expression without integrating ^{13}C data) was also performed (Fig 2, S1 Table). Lacking ^{13}C data, GIMME does not predict any activity in the oxidative branch of the pentose phosphate pathway, nor on the glycogen phosphorylase/glycogen synthase futile cycle. Furthermore, GIMME predicts a significantly larger flux through pyruvate dehydrogenase than p^{13}CMFA. Interestingly, p^{13}CMFA predicts an increased activity of the TCA cycle compared to the GIMME solution. This increased activity is fueled by alternative sources of acetyl-CoA such as fatty acid oxidation or catabolism of ketogenic amino acids. Hence, p^{13}CMFA is able to take advantage of measured ^{13}C enrichments and predict significantly different flux maps than those derived from flux minimization alone.

Validation of the p^{13}CMFA approach

To validate the p^{13}CMFA method, we used data from a metabolic characterization of the colon cancer cell line HCT 116 published by Tarrado-Castellarnau *et al.* [14]. In this study, 25 direct flux measurements and 24 sets of isotopologue fractions, measured after incubation with either [1,2-$^{13}C_2$]-glucose or [U-$^{13}C_5$]-glutamine, had been integrated in the framework of ^{13}C MFA. With such a large set of experimental measurements, ^{13}C MFA had been able to estimate the flux through 62 reactions with a high degree of accuracy. In the same study, transcriptomics data were also collected.

From this large data set, we selected a partial data set consisting of 7 experimental flux measurements (the rates of uptake/secretion of glucose, lactate, glutamine, glutamate and, oxygen and the rate of protein and glycogen synthesis) and 4 sets of isotopologue fractions (isotopologue fractions in ribose, lactate, glutamate and glycogen measured after incubation with 1,2-$^{13}C_2$]-glucose). Those are the sets of isotopologues and fluxes that were analyzed in the HUVECs case study with the addition of the rate of protein synthesis and oxygen consumption which Tarrado-Castellarnau *et al.* described as key determinants of the metabolic phenotype of HCT 116 cells. The partial data set was used to apply pFBA, GIMME, ^{13}C MFA and

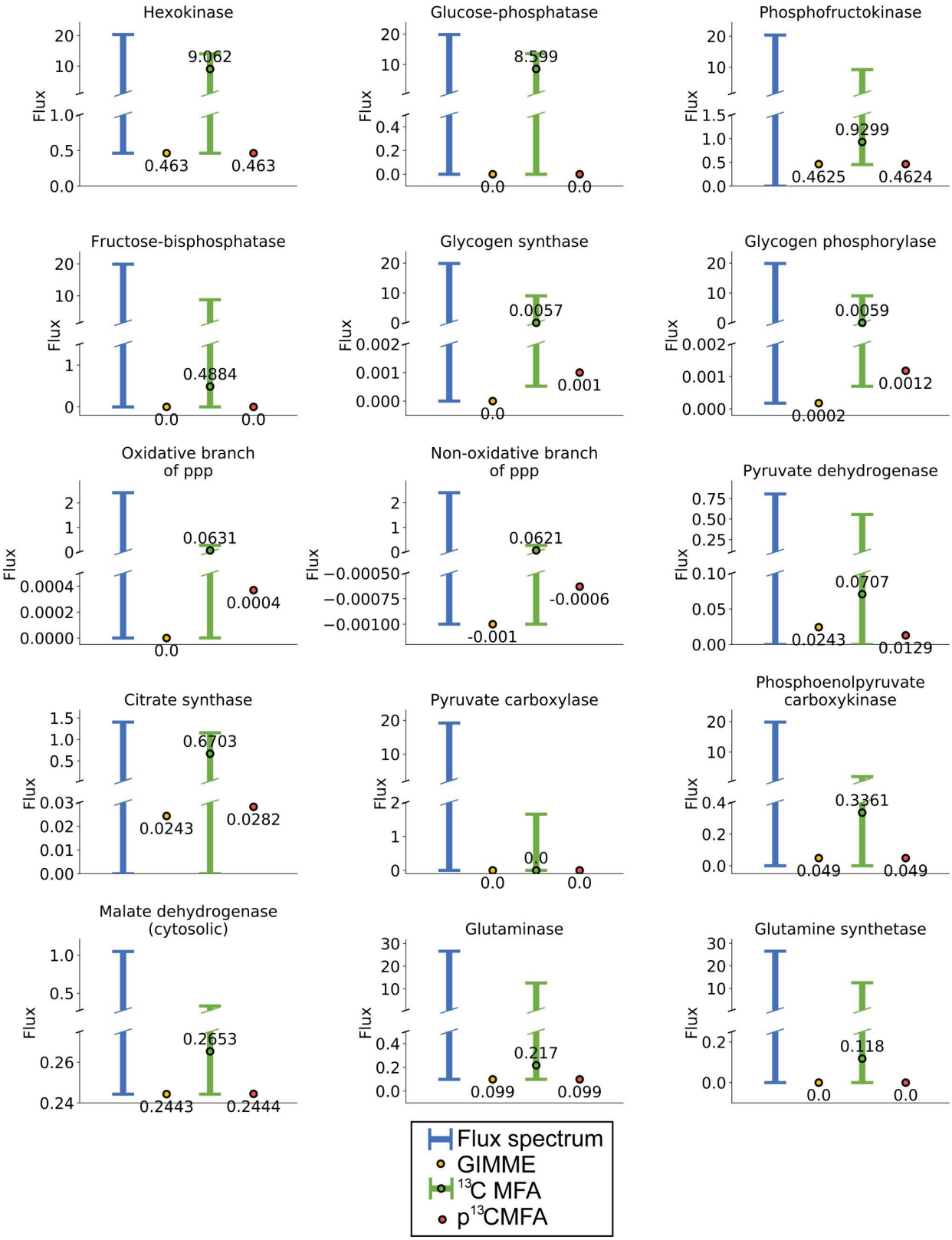


Fig 2. Flux spectrum, GIMME solutions, ^{13}C MFA flux ranges, and p^{13}CMFA solutions for some key reaction fluxes in the HUVECs case study. Flux spectrum represents the feasible flux ranges considering only the stoichiometric and thermodynamic constraints and the measured extracellular fluxes. GIMME flux values are obtained when total reaction flux is minimized weighted by gene expression without integrating ^{13}C data. For ^{13}C MFA,

the flux values obtained after the ^{13}C MFA optimization and the range of the 95% confidence intervals for such values are shown. The p^{13}CMFA flux values are obtained when total reaction flux is minimized within the ^{13}C MFA solution space. Fluxes are expressed in μmol·h^{-1}·million-cells^{-1}.

https://doi.org/10.1371/journal.pcbi.1007310.g002

p^{13}CMFA in the framework of the metabolic network defined by Tarrado-Castellarnau *et al.* [14] (S2 Zip). p^{13}CMFA was applied both with and without integrating gene expression data (p^{13}CMFA+ge and p^{13}CMFA-ge, respectively). Two complementary metrics, Pearson's correlation and Euclidian distance, were used to evaluate the similarity between the predicted flux distributions and the flux maps estimated by Tarrado-Castellarnau *et al.* using the full dataset [14] (Fig 4, S2 Table). The results show that p^{13}CMFA-ge yields a significantly more accurate flux prediction than both pFBA (i.e., flux minimization without integrating ^{13}C data), and ^{13}C MFA. Interestingly, while integrating gene expression significantly enhances the accuracy of p^{13}CMFA (p^{13}CMFA+ge compared to p^{13}CMFA-ge), such effect is less marked than the effect of adding gene expression to standard flux minimization (GIMME compared to pFBA). This is due to the fact that p^{13}CMFA-ge flux predictions have already a remarkable level of accuracy; hence, less information can be gained by adding transcriptomics data. Nevertheless, even if GIMME achieves flux predictions of similar accuracy to p^{13}CMFA-ge, p^{13}CMFA+ge results on flux predictions that are significantly more accurate than those obtained with GIMME. Hence, in instances were only a limited number of ^{13}C measurements are available, p^{13}CMFA is a valid method for obtaining accurate flux estimations, regardless of the availability of gene expression data.

Discussion

^{13}C MFA is a well-established technique and has proven to be an extremely valuable tool in quantifying metabolic fluxes[9–18]. However, to fully determine fluxes through a large metabolic network, parallel labeling experiments must be performed and ^{13}C propagation must be quantified in many metabolites in the network[19]. Indeed, when applying ^{13}C MFA either with a small set of experimental data or with a large metabolic network, part of the ^{13}C MFA solution space can be too wide to draw meaningful conclusions about the underlying flux distribution. This solution space can be reduced by removing degrees of freedom from the system, for instance, by removing reactions from the network or making reactions irreversible. However, this can introduce an arbitrary bias in the resulting flux distribution.

Here we describe p^{13}CMFA, a new approach for ^{13}C data integration which can overcome these limitations of ^{13}C MFA and estimate a realistic solution within an undetermined ^{13}C MFA solution space. This solution will be the flux distribution within the ^{13}C MFA solution space that minimizes the weighted sum of reaction fluxes. Thus, it will be the most enzymatically efficient solution. In that regard, p^{13}CMFA is partially based on a similar principle as pathway activity analysis (i.e., the assumption that specific fractions of isotopologues are primarily generated through the simplest combinations of pathways). However, unlike pathway activity analysis, p^{13}CMFA is able to integrate all quantified isotopologue fractions and flux measurements (e.g. rates of metabolite uptake and secretion) to generate network-wide flux maps consistent with such data. Furthermore, p^{13}CMFA is highly flexible; for instance, here we show that it can be used to seamlessly integrate gene expression data by giving higher weight to the minimization of the fluxes through lowly expressed enzymes.

As a proof of concept, we exemplified how p^{13}CMFA can be used to estimate flux distributions integrating only limited sets of ^{13}C measurements in a test case where traditional ^{13}C MFA was unable to provide a narrow solution space. Furthermore, we demonstrated that, when a limited set of measurements are integrated, p^{13}CMFA can yield more accurate flux predictions than both ^{13}C MFA and GIMME.

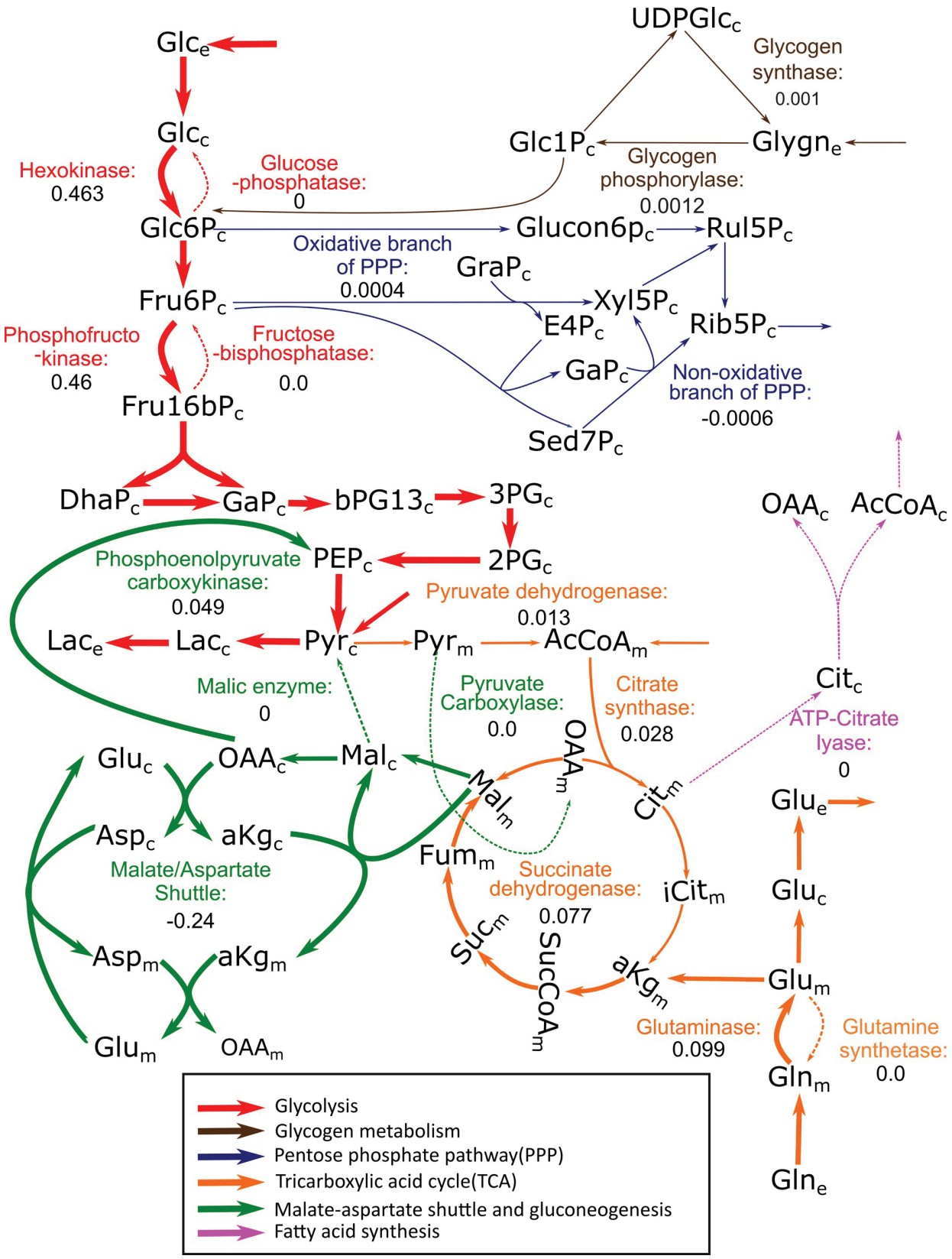


Fig 3. Schematic representation of the central carbon metabolism flux map obtained with p^{13}CMFA in the HUVECs case study. Reaction fluxes are indicated for some key reactions in μmol·h^{-1}·million-cells^{-1}. Arrows indicate net flux direction, and line width is representative of flux magnitude.

Reactions and metabolites of redox and energy metabolism have been omitted from this figure for clarity. 2PG: 2-Phosphoglycerate. 3PG: 3-Phosphoglycerate. AcCoA: Acetyl-CoA. aKG: α-Ketoglutarate. Asp: Aspartate. bPG13: 1,3-Bisphosphoglycerate. Cit: Citrate. DhaP: Dihydroxyacetone phosphate. Fru16bP: Fructose 1,6-bisphosphate. Fru6P: Fructose 6-phosphate. Fum: Fumarate. GaP: Glyceraldehyde-3-Phosphate. Glc: Glucose. Glc1P: Glucose 1-phosphate. Glc6P: Glucose 6-phosphate. Gln: Glutamine. Glu: Glutamate. Glucon6P: Gluconate 6-phosphate. Glygn: Glycogen. iCit: Isocitrate. Lac: Lactate. Mal: Malate. OAA: Oxaloacetate. PEP: Phosphoenolpyruvate. Pyr: Pyruvate. Rib5P: Ribose 5-phosphate. Rul5P: Ribulose 5-phosphate. Sed7P: Sedoheptulose 7-phosphate. Suc: Succinate. SucCoa: Succinyl-CoA. UDPGlc: Uridine diphosphate glucose. The subscripts *e*, *c*, and *m* denote the extracellular, cytosolic and mitochondrial compartments, respectively.

https://doi.org/10.1371/journal.pcbi.1007310.g003

p^{13}C MFA does not aim to be a replacement of ^{13}C MFA; instead, it seeks to supplement it by identifying the more straightforward solution in parts of the network that cannot be uniquely determined. In that regard, it can be used to quantitatively study flux distributions in instances where not enough information can be obtained with conventional ^{13}C MFA. Nor does it aim to replace the direct interpretation of ^{13}C data. The latter is still a suitable technique when the goal of the analysis is to compare the relative activity of well-established pathways across conditions or quantify substrate contributions rather than to generate complete flux maps.

^{13}C data has been widely used to assist in drug discovery. In this regard, tracer analysis coupled with regression and correlation analyses is frequently used to characterize drug response [26–29]. Such approach uses regression and correlation statistics with binary, numeric and visual analysis to integrate drug dosage, time points, as well as all necessary biological variables in order to diagnose disturbed stable isotope labeled matrices[29]. p^{13}CMFA could further expand the role of ^{13}C in drug discovery by allowing the integration of ^{13}C and transcriptomic data in the framework of genome-scale metabolic models. In the framework of such models, drug targets are identified by systematically simulating the effect of reactions or genes knock out to cell function[34]. This is usually attained by applying the ROOM[35] or MOMA[36] algorithms, which take a unique flux solution as input (wild-type flux distribution) to predict the most likely effect of a gene KO. Hence, p^{13}CMFA results could be potentially used as ROOM/MOMA inputs allowing to take full advantage of the flux information derived from both ^{13}C and transcriptomics data to predict new drug targets. With atom mappings now available on a genome-scale[37], the main obstacle to applying p^{13}CMFA at a genome-scale is the high computational complexity of solving the resulting non-linear problem which increases with the size of the network. Hence, the next challenge for p^{13}CMFA will be optimizing its implementation for genome-scale networks.

Methods

Flux spectrum

The flux spectrum[38] (i.e., the feasible range of fluxes for a given set of stoichiometric, thermodynamic and flux boundary constraints) was determined using flux variability analysis [8]. Under this approach, each flux is minimized (Eq 3) and maximized (Eq 4) subject to constraints to find the minimum (vmin_i^{FS}) and maximum (vmax_i^{FS}) feasible values for each flux:

$$\mathrm{vmin}_i^{FS} = \min v_i \quad \text{(Eq3)}$$

$$\textit{subject to } S.v = 0,\ lb \leq v \leq ub$$

$$\mathrm{vmax}_i^{FS} = \max v_i \quad \text{(Eq4)}$$

$$\textit{subject to } S.v = 0,\ lb \leq v \leq ub$$

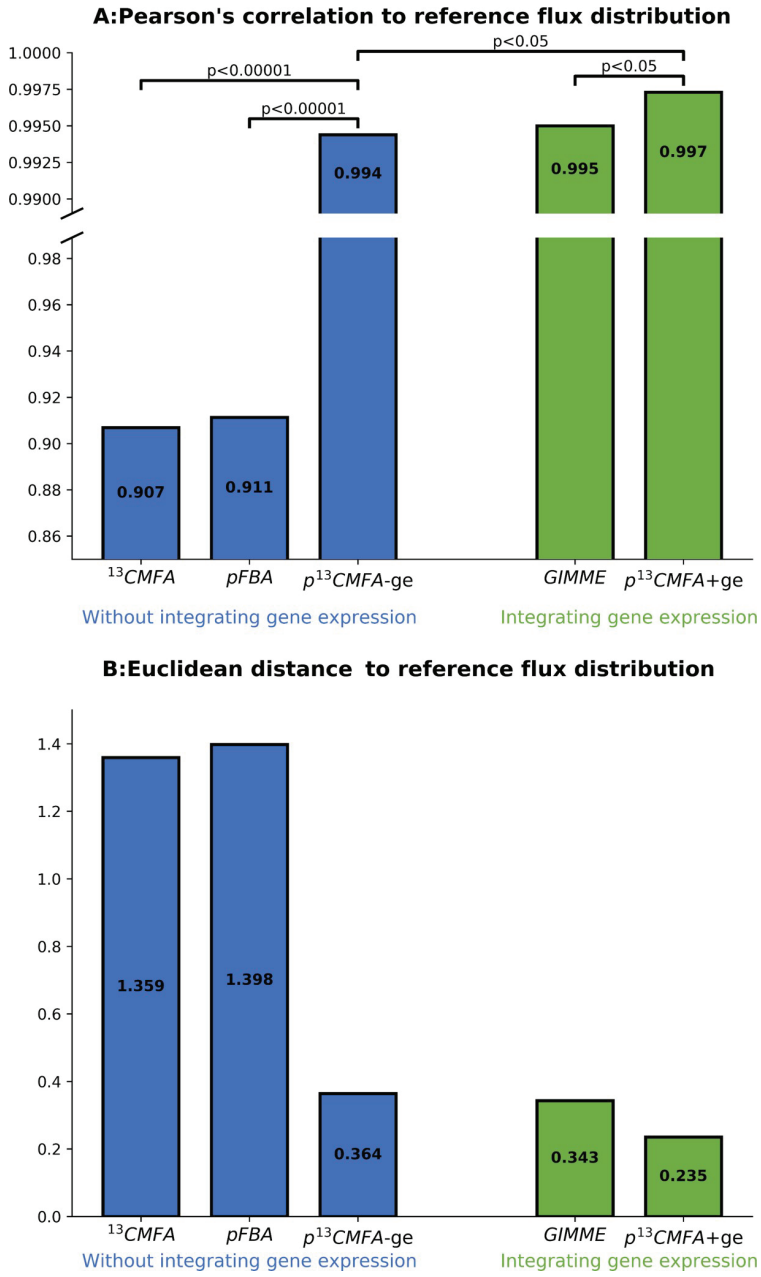


Fig 4. Comparison of the predictive power of ^{13}C MFA, pFBA, GIMME, and p^{13}CMFA. A: Pearson's correlation coefficients between the reference flux distribution and the flux maps obtained from ^{13}C MFA (optimal solution), pFBA, GIMME, and p^{13}CMFA. p^{13}CMFA was applied both with and without integrating gene expression data (p^{13}CMFA+ge and p^{13}CMFA-ge, respectively). The statistical significance of the difference between correlation coefficients was evaluated using the Fisher r-to-z transformation[33]. B: Euclidian distances between the reference flux distribution and the flux maps obtained from ^{13}C MFA (optimal solution), pFBA, GIMME, and p^{13}CMFA.

https://doi.org/10.1371/journal.pcbi.1007310.g004

^{13}C MFA confidence intervals

The ^{13}C MFA solution space is estimated by computing the confidence intervals for each flux. Such intervals are obtained by minimizing (Eq 5) and maximizing (Eq 6) each flux subject to

constraints[5].

$$vmin_i = \min v_i \quad \text{(Eq5)}$$

$$subject\ to\ S.v = 0,\ lb \le v \le ub,\ \sum_j \left(\frac{E_j - Y_j(v)}{\sigma_j}\right)^2 \le X_{opt} + T$$

$$vmax_i = \max v_i \quad \text{(Eq6)}$$

$$subject\ to\ S.v = 0,\ lb \le v \le ub,\ \sum_j \left(\frac{E_j - Y_j(v)}{\sigma_j}\right)^2 \le X_{opt} + T$$

where,

$vmin_i$: is the lower bound of the confidence interval for flux *i* with tolerance T;

$vmax_i$: is the upper bound of the confidence interval for flux *i* with tolerance T;

Provided that the same primary objective tolerance (T) is used in computing both the p^{13}CMFA solution and the ^{13}C MFA confidence intervals, the p^{13}CMFA solution will always fall within the boundaries of ^{13}C MFA confidence intervals ($vmin_i \le v_i \le vmax_i$).

GIMME and pFBA

To apply GIMME and pFBA, the sum of fluxes is minimized subject only to network stoichiometry and flux boundaries (Eq 7).

$$\min \textstyle\sum_i |v_i| \cdot w_i \quad \text{(Eq7)}$$

$$subject\ to\ S.v = 0,\ lb \le v \le ub$$

In GIMME, flux minimization weights are derived from gene expression measurements, whereas in pFBA all reactions are given the same minimization weight[20,22].

Transcriptomic analysis

Transcriptomic data of HUVECs and HCT 116 cells published by Weigand *et al.*[32] and Tarrado-Castellarnau[14] *et al.*, respectively, were obtained from the Gene Expression Omnibus repository[39]. A Robust Multichip Analysis gene-level normalization[40] was performed with the Oligo package for R[41].

Using gene-protein-reaction rules, normalized transcript intensities were mapped to each enzyme-catalyzed reaction or protein-facilitated transport process. The weight given to the minimization of fluxes was assigned according to the following equation:

$$w_i = 1 + \max(Th - ge_i, 0) \quad \text{(Eq8)}$$

where,

ge_i is the gene expression value assigned to reaction *i*;

Th is the gene expression threshold. Fluxes through reactions with gene expression levels below this threshold are given additional minimization weight;

Using the same criteria as GIM^{3}E[22], Th was set at the maximum gene expression value found in the set of genes mapped to the metabolic network (Eq 9):

$$Th = \max(ge) \quad \text{(Eq9)}$$

Using this threshold, the information gained from integrating available gene expression measurements is maximized. Other Th values were tested in the validation case study[14] and using the maximum gene expression as the threshold was found to yield the most accurate flux predictions (S3 Table).

p^{13}CMFA implementation

p^{13}CMFA was implemented in Iso2Flux, our in-house developed ^{13}C MFA software (https://github.com/cfoguet/iso2flux/releases/tag/0.7.2).

Iso2Flux computes steady-state flux distributions as the product of the null space of the stoichiometric matrix and the vector of free fluxes. Reversible reactions are split into forward and reverse reactions. For each reversible reaction, a turnover variable (t_i) is introduced defining the flux that is common to the forward (v_i^f) and reverse (v_i^r) reactions. These variables are used to assign values to the fluxes of the forward and reverse reactions as a function of the steady-state net flux (v_i).

$$v_i^f = t_i + \max(v_i, 0) \quad \text{(Eq10)}$$

$$v_i^r = t_i - \min(v_i, 0) \quad \text{(Eq11)}$$

Iso2flux uses the Elementary Metabolite Unit (EMU) framework[1] to build the ^{13}C propagation model. This framework is based on a highly efficient decomposition method that identifies the minimum amount of isotopologue transitions required to simulate the experimentally quantified isotopologues according to the defined carbon propagation rules. The isotopologue transitions are grouped into decoupled systems based on isotopologue size. Balance equations are built around each isotopologue fraction under the assumption of isotopic steady state (S1 Fig). Using the steady-state flux distribution as an input, systems of equations around isotopologues balances are solved sequentially starting with the smallest isotopologue size [1] using the fsolve function of the SciPy library (https://scipy.org/scipylib/index.html). Solving such system predicts the isotopologue distribution associated with a given steady-state flux distribution ($Y_j(v)$).

The self-adaptive differential evolution (SADE) algorithm from PyGMO (Python Parallel Global Multiobjective Optimizer, https://github.com/esa/pagmo2) was used to find the optimal solution of the ^{13}C MFA (Eq 1) and p^{13}CMFA (Eq 2) problems. SADE was parallelized using the generalized island-model paradigm. Under such implementation, SADE is run in parallel in different CPU processes (islands). After a given number of SADE iterations (generations), the best solutions (individuals) in each SADE process (island) are shared to parallel SADE processes (migrate to adjacent islands). To prevent bias from the starting solutions (starting populations), the islands are seeded through random sampling of all variables. Free fluxes variables are sampled using the optGpSampler implemented into COBRApy[42,43]. Turnover variables are sampled using the random.uniform function built into python. The algorithm was run with 7 islands, each with a population of 60, and with migrations between islands set to occur every 400 generations. For the analyzed ^{13}C MFA and p^{13}CMFA problems, repeated iterations of the algorithm were shown to reliably converge towards the same minimal objective function value.

Accommodating large metabolite pools

At the beginning of a ^{13}C experiment, all internal metabolites are unlabeled (m0). Progressively, these products are enriched in ^{13}C, with the subsequent decrease in m0. Isotopic steady state is quickly reached for small pools of metabolites but not necessarily for larger pools such as those of fatty acids, glycogen or metabolites present in large concentrations in the external medium[44]. For these larger pools, unlabeled isotopologues m0 are oversized and might not quickly decrease to the theoretical value that should be reached at steady-state.

However, it is possible to represent the effect of large pools in the framework of steady-state ^{13}C MFA through the addition of a virtual reaction. This reaction replaces labeled isotopologues by unlabeled isotopologues in metabolites with large pools. With p^{13}CMFA, the flux through this virtual reaction can be minimized. Effectively, this allows correcting steady-state ^{13}C simulations for large pools while identifying the solutions that require the minimum amount of correction.

Evaluating the significance of the difference between correlation coefficients

The statistical significance of the difference between correlation coefficients was evaluated using the Fisher r-to-z transformation[33]. Following this approach, Pearson's correlation coefficients (r) can be converted to a z-score (r'):

$$r' = \frac{1}{2} \cdot Ln\left(\frac{1+r}{1-r}\right) \quad \text{(Eq12)}$$

The variance of z (Sz) will depend only on the sample size (n):

$$S_z = \sqrt{\frac{1}{n-3}} \quad \text{(Eq13)}$$

From Eq 12 and Eq 13, the significance of the difference between two correlation coefficients (r1 and r2) can be evaluated by computing the z score corresponding to such difference (Eq 14) and its associated p-value.

$$Z = \frac{r'_1 - r'_2}{\sqrt{\frac{1}{n_{r1}-3} + \frac{1}{n_{r2}-3}}} \quad \text{(Eq14)}$$

Experimental methods

Human Umbilical Vein Endothelial Cells (HUVECs-pooled, Lonza) were maintained on 1% gelatin-coated flasks at 37˚C in a humidified atmosphere of 5% CO_2 and 95% air in MCDB131 (Gibco) medium, supplemented with the recommended quantity of endothelial growth medium (EGM) SingleQuots (Lonza), 10% fetal bovine serum (FBS) (Gibco), 2 mM glutamine (Gibco) and 0.1% Streptomycin (100 μg/mL)/Penicillin (100 units/mL) (S/P) (Gibco). 1×10^6 HUVECs were seeded in 1% gelatin-coated cell culture plates for 6h, and then the maintenance medium was replaced with the MCDB131 basal medium, supplemented with 2% FBS, 2 mM glutamine and 0.1% S/P and cells were incubated overnight for nutrient deprivation. After nutrient deprivation, the medium was replaced with a restricted medium containing MCDB131 medium supplemented with 2% FBS, 2 mM glutamine and 0.1% S/P with 10 mM of 50% [1,2-$^{13}C_2$]-glucose (Sigma-Aldrich) and cells were incubated for 40h in a humidified atmosphere with 5% CO_2 and 1% O_2 at 37˚C. Both at the beginning (t = 0h) and the end (t = 40h) of incubation, media and pellets were collected. On the one hand, media and cell

pellets were used for analyzing isotopologue abundances for glucose, lactate, glutamate, RNA ribose and glycogen. Raw data are publicly available in the MetaboLights repository at http://www.ebi.ac.uk/metabolights [31], with accession number MTBLS412. Isolation, derivatization and analysis details are described in MetaboLights. Glucose, lactate, glutamate, and glutamine concentrations were determined in media samples for estimation of secretion or uptake rates of these metabolites using spectrophotometric methods[45]. Also, the net rate of glycogen re-utilization into glucose was estimated by quantifying glycogen content at initial and final time points using [U-^{13}C-D_7]-glucose as recovery standard[46]. All biochemical data were normalized by cell number, and by incubation time (h). The resulting rates–expressed in micromoles of metabolite consumed/produced/transformed per hour per million cells (μmol·h^{-1}·million-cells^{-1})–were 0.463, 0.099, 0.050 and 1.169 for glucose uptake, glutamine uptake, glutamate secretion, and lactate secretion, respectively, and a net transformation of glycogen of 0.000175.

Supporting information

S1 Fig. Example of isotopologue balance equations in a toy metabolic network. In this toy metabolic network, two mono-carbon metabolites (C_a and C_b) are condensed into a bi-carbon metabolite (C_a-C_b) through a reaction with a flux v_1. Metabolite C_a-C_b is removed from the system at a rate of v_2. For each metabolite, isotopologue fractions (M_x) are defined as the relative abundance of the metabolite with *x* number of ^{13}C substitutions. Isotopologue balances for metabolite C_a-C_b are indicated. Under the assumption of isotopic steady state (i.e., isotopologue fractions are constant in time) and given v_1 and v_2, and a set of isotopologue fractions for C_a and C_b (assumed a constant input), the system can be solved to identify the steady-state isotopologue fractions for metabolite C_a-C_b.
(TIF)

S1 Table. Flux spectrum, GIMME, ^{13}C MFA and p^{13}CMFA flux solutions for all net reaction fluxes in the HUVECs case study. Fluxes are expressed in μmol·h^{-1}·million-cells^{-1}.
(XLSX)

S2 Table. Comparison between the reference flux map in HCT 116 cells and the flux maps computed from the partial data set using ^{13}C MFA, pFBA, GIMME, and p^{13}CMFA. Fluxes are indicated in μmol·h^{-1}·million-cells^{-1}.
(XLSX)

S3 Table. Comparison between the reference flux map in HCT 116 cells and the flux maps computed from the partial data set with p^{13}CMFA using different gene expression percentiles as thresholds for adding additional weight to flux minimization. Fluxes are indicated in μmol·h^{-1}·million-cells^{-1}.
(XLSX)

S1 ZIP. Files describing the metabolic network, carbon propagation rules, and experimental data used for the HUVECs case study. The files are inputs for running p^{13}CMFA on Iso2Flux.
(ZIP)

S2 ZIP. Files describing the metabolic network, carbon propagation rules, and experimental data used for the HCT 116 cells case study. The files are inputs for running p^{13}CMFA on Iso2Flux.
(ZIP)

Author Contributions

Conceptualization: Carles Foguet, Vitaly A. Selivanov, Pedro de Atauri, Marta Cascante.

Data curation: Anusha Jayaraman, Silvia Marin, Vitaly A. Selivanov, Pablo Moreno, Ramon Messeguer.

Formal analysis: Carles Foguet.

Funding acquisition: Marta Cascante.

Investigation: Anusha Jayaraman, Silvia Marin, Ramon Messeguer.

Methodology: Carles Foguet.

Project administration: Marta Cascante.

Software: Carles Foguet, Vitaly A. Selivanov, Pablo Moreno, Pedro de Atauri.

Supervision: Pedro de Atauri, Marta Cascante.

Writing – original draft: Carles Foguet, Silvia Marin, Pedro de Atauri.

Writing – review & editing: Anusha Jayaraman, Silvia Marin, Vitaly A. Selivanov, Pablo Moreno, Ramon Messeguer, Pedro de Atauri, Marta Cascante.

References

1. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. Metab Eng. 2007; 9: 68–86. https://doi.org/10.1016/j.ymben.2006.09.001 PMID: 17088092
2. Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, et al. A roadmap for interpreting (13)C metabolite labeling patterns from cells. Curr Opin Biotechnol. 2015; 34: 189–201. https://doi.org/10.1016/j.copbio.2015.02.003 PMID: 25731751
3. Zamboni N, Saghatelian A, Patti GJ. Defining the metabolome: size, flux, and regulation. Mol Cell. 2015; 58: 699–706. https://doi.org/10.1016/j.molcel.2015.04.021 PMID: 26000853
4. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010; 28: 245–8. https://doi.org/10.1038/nbt.1614 PMID: 20212490
5. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. Metab Eng. 2006; 8: 324–337. https://doi.org/10.1016/j.ymben.2006.01.004 PMID: 16631402
6. Niedenführ S, Wiechert W, Nöh K. How to measure metabolic fluxes: A taxonomic guide for 13C fluxomics. Curr Opin Biotechnol. 2015; 34: 82–90. https://doi.org/10.1016/j.copbio.2014.12.003 PMID: 25531408
7. Balcells C, Foguet C, Tarragó-Celada J, de Atauri P, Marin S, Cascante M. Tracing metabolic fluxes using mass spectrometry: Stable isotope-resolved metabolomics in health and disease. TrAC Trends Anal Chem. 2019; https://doi.org/10.1016/j.trac.2018.12.025
8. Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. BMC Bioinformatics. 2010; 11: 489. https://doi.org/10.1186/1471-2105-11-489 PMID: 20920235
9. Niklas J, Sandig V, Heinzle E. Metabolite channeling and compartmentation in the human cell line AGE1.HN determined by ^{13}C labeling experiments and13C metabolic flux analysis. J Biosci Bioeng. 2011; 112: 616–623. https://doi.org/10.1016/j.jbiosc.2011.07.021 PMID: 21865082
10. Walther JL, Metallo CM, Zhang J, Stephanopoulos G. Optimization of13C isotopic tracers for metabolic flux analysis in mammalian cells. Metab Eng. 2012; 14: 162–171. https://doi.org/10.1016/j.ymben.2011.12.004 PMID: 22198197
11. Metallo CM, Gameiro PA, Bell EL, Mattaini KR, Yang J, Hiller K, et al. Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. Nature. 2012; 481: 380–384. https://doi.org/10.1038/nature10602 PMID: 22101433
12. Grassian AR, Parker SJ, Davidson SM, Divakaruni AS, Green CR, Zhang X, et al. IDH1 mutations alter citric acid cycle metabolism and increase dependence on oxidative mitochondrial metabolism. Cancer Res. 2014; 74: 3317–3331. https://doi.org/10.1158/0008-5472.CAN-14-0772-T PMID: 24755473

13. Crown SB, Kelleher JK, Rouf R, Muoio DM, Antoniewicz MR. Comprehensive metabolic modeling of multiple 13C-isotopomer data sets to study metabolism in perfused working hearts. Am J Physiol Heart Circ Physiol. 2016; 311: H881–H891. https://doi.org/10.1152/ajpheart.00428.2016 PMID: 27496880

14. Tarrado-Castellarnau M, de Atauri P, Tarragó-Celada J, Perarnau J, Yuneva M, Thomson TM, et al. De novo MYC addiction as an adaptive response of cancer cells to CDK4/6 inhibition. Mol Syst Biol. 2017; 13: 940. https://doi.org/10.15252/msb.20167321 PMID: 28978620

15. Carinhas N, Koshkin A, Pais DAM, Alves PM, Teixeira AP. 13 C-metabolic flux analysis of human adenovirus infection: Implications for viral vector production. Biotechnol Bioeng. 2017; 114: 195–207. https://doi.org/10.1002/bit.26063 PMID: 27477740

16. DeWaal D, Nogueira V, Terry AR, Patra KC, Jeon S-M, Guzman G, et al. Hexokinase-2 depletion inhibits glycolysis and induces oxidative phosphorylation in hepatocellular carcinoma and sensitizes to metformin. Nat Commun. 2018; 9: 446. https://doi.org/10.1038/s41467-017-02733-4 PMID: 29386513

17. Gopalakrishnan S, Maranas CD. ^{13}C metabolic flux analysis at a genome-scale. Metab Eng. 2015; 32: 12–22. https://doi.org/10.1016/j.ymben.2015.08.006 PMID: 26358840

18. Foguet C, Marin S, Selivanov VA, Fanchon E, Lee W-NP, Guinovart JJ, et al. HepatoDyn: A Dynamic Model of Hepatocyte Metabolism That Integrates 13C Isotopomer Data. Lewis NE, editor. PLoS Comput Biol. 2016; 12: e1004899. https://doi.org/10.1371/journal.pcbi.1004899 PMID: 27124774

19. Crown SB, Long CP, Antoniewicz MR. Optimal tracers for parallel labeling experiments and 13C metabolic flux analysis: A new precision and synergy scoring system. Metab Eng. 2016; 38: 10–18. https://doi.org/10.1016/j.ymben.2016.06.001 PMID: 27267409

20. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol Syst Biol. 2010; 6: 390. https://doi.org/10.1038/msb.2010.47 PMID: 20664636

21. Becker SA, Palsson BO. Context-Specific Metabolic Networks Are Consistent with Experiments. Sauro HM, editor. PLoS Comput Biol. 2008; 4: e1000082. https://doi.org/10.1371/journal.pcbi.1000082 PMID: 18483554

22. Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson BØ, Hyduke DR. GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics. 2013; 29: 2900–8. https://doi.org/10.1093/bioinformatics/btt493 PMID: 23975765

23. Antoniewicz MR. A guide to 13C metabolic flux analysis for the cancer biologist. Exp Mol Med. 2018; 50: 19. https://doi.org/10.1038/s12276-018-0060-y PMID: 29657327

24. Dong W, Keibler MA, Stephanopoulos G. Review of metabolic pathways activated in cancer cells as determined through isotopic labeling and network analysis. Metabolic Engineering. 2017. pp. 113–124. https://doi.org/10.1016/j.ymben.2017.02.002 PMID: 28192215

25. Bruntz RC, Lane AN, Higashi RM, Fan TWM. Exploring cancer metabolism using Stable isotope-resolved metabolomics (SIRM). J Biol Chem. 2017; 292: 11601–11609. https://doi.org/10.1074/jbc.R117.776054 PMID: 28592486

26. Harrigan GG, Colca J, Szalma S, Boros LG. PNU-91325 increases fatty acid synthesis from glucose and mitochondrial long chain fatty acid degradation: A comparative tracer-based metabolomics study with rosiglitazone and pioglitazone in HepG2 cells. Metabolomics. 2006; 2: 21–29. https://doi.org/10.1007/s11306-006-0015-5 PMID: 24489530

27. Beger RD, Hansen DK, Schnackenberg LK, Cross BM, Fatollahi JJ, Lagunero FT, et al. Single valproic acid treatment inhibits glycogen and RNA ribose turnover while disrupting glucose-derived cholesterol synthesis in liver as revealed by the [U-13C6]-d-glucose tracer in mice. Metabolomics. 2009; 5: 336–345. https://doi.org/10.1007/s11306-009-0159-1 PMID: 19718458

28. Cantoria MJ, Boros LG, Meuillet EJ. Contextual inhibition of fatty acid synthesis by metformin involves glucose-derived acetyl-CoA and cholesterol in pancreatic tumor cells. Metabolomics. 2014; 10: 91–104. https://doi.org/10.1007/s11306-013-0555-4 PMID: 24482631

29. Boros LG, Beger RD, Meuillet EJ, Colca JR, Szalma S, Thompson PA, et al. Targeted 13C-Labeled Tracer Fate Associations for Drug Efficacy Testing in Cancer. Tumor Cell Metabolism. Vienna: Springer Vienna; 2015. pp. 349–372. https://doi.org/10.1007/978-3-7091-1824-5_15

30. Varma V, Boros LG, Nolen GT, Chang CW, Wabitsch M, Beger RD, et al. Metabolic fate of fructose in human adipocytes: a targeted ^{13}C tracer fate association study. Metabolomics. 2015; 11: 529–544. https://doi.org/10.1007/s11306-014-0716-0 PMID: 25972768

31. Haug K, Salek RM, Conesa P, Hastings J, de Matos P, Rijnbeek M, et al. MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res. 2013; 41: D781–6. https://doi.org/10.1093/nar/gks1004 PMID: 23109552

32. Weigand JE, Boeckel J-N, Gellert P, Dimmeler S. Hypoxia-Induced Alternative Splicing in Endothelial Cells. Preiss T, editor. PLoS One. 2012; 7: e42697. https://doi.org/10.1371/journal.pone.0042697 PMID: 22876330

33. Weaver B, Wuensch KL. SPSS and SAS programs for comparing Pearson correlations and OLS regression coefficients. Behav Res Methods. 2013; 45: 880–895. https://doi.org/10.3758/s13428-012-0289-7 PMID: 23344734

34. de Mas IM, Aguilar E, Jayaraman A, Polat IH, Martín-Bernabé A, Bharat R, et al. Cancer cell metabolism as new targets for novel designed therapies. Future Med Chem. 2014; 6: 1791–1810. https://doi.org/10.4155/fmc.14.119 PMID: 25574531

35. Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc Natl Acad Sci U S A. 2005; 102: 7695–700. https://doi.org/10.1073/pnas.0406346102 PMID: 15897462

36. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci. 2002; 99: 15112–15117. https://doi.org/10.1073/pnas.232349399 PMID: 12415116

37. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat Biotechnol. 2018; 36: 272–281. https://doi.org/10.1038/nbt.4072 PMID: 29457794

38. Llaneras F, Picó J. An interval approach for dealing with flux distributions and elementary modes activity patterns. J Theor Biol. 2007; 246: 290–308. https://doi.org/10.1016/j.jtbi.2006.12.029 PMID: 17292923

39. Edgar R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30: 207–210. https://doi.org/10.1093/nar/30.1.207 PMID: 11752295

40. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4: 249–64. https://doi.org/10.1093/biostatistics/4.2.249 PMID: 12925520

41. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinformatics. 2010; 26: 2363–2367. https://doi.org/10.1093/bioinformatics/btq431 PMID: 20688976

42. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol. 2013; 7: 74. https://doi.org/10.1186/1752-0509-7-74 PMID: 23927696

43. Megchelenbrink W, Huynen M, Marchiori E. optGpSampler: An Improved Tool for Uniformly Sampling the Solution-Space of Genome-Scale Metabolic Networks. Rogers S, editor. PLoS One. 2014; 9: e86587. https://doi.org/10.1371/journal.pone.0086587 PMID: 24551039

44. Selivanov VA, Meshalkina LE, Solovjeva ON, Kuchel PW, Ramos-Montoya A, Kochetov GA, et al. Rapid simulation and analysis of isotopomer distributions using constraints based on enzyme mechanisms: an example from HT29 cancer cells. Bioinformatics. 2005; 21: 3558–64. https://doi.org/10.1093/bioinformatics/bti573 PMID: 16002431

45. Benito A, Polat IH, Noé V, Ciudad CJ, Marin S, Cascante M. Glucose-6-phosphate dehydrogenase and transketolase modulate breast cancer cell metabolic reprogramming and correlate with poor patient outcome. Oncotarget. 2017; 8: 106693–106706. https://doi.org/10.18632/oncotarget.21601 PMID: 29290982

46. Vizán P, Sánchez-Tena S, Alcarraz-Vizán G, Soler M, Messeguer R, Pujol MD, et al. Characterization of the metabolic changes underlying growth factor angiogenic activation: identification of new potential therapeutic targets. Carcinogenesis. 2009; 30: 946–52. https://doi.org/10.1093/carcin/bgp083 PMID: 19369582

8.3 Chapter 3

Identification of the molecular drivers of prostate cancer stem cells

Carles oguet, Cristina alcells, Timothy M. Thomson, edro de Atauri, Marta Cascante.

Identification of the molecular drivers of prostate cancer stem cells

Carles Foguet[1], Cristina Balcells[2], Timothy M. Thomson[3], Pedro de Atauri[1], Marta Cascante[1*]

1 Department of Biochemistry and Molecular Biomedicine & Institute of Biomedicine of University of Barcelona, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain and Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD) and Metabolomics node at Spanish National Bioinformatics Institute (INB-ISCIII-ES-ELIXIR), Instituto de Salud Carlos III (ISCIII), Madrid, Spain

2 Department of Biochemistry and Molecular Biomedicine-Institute of Biomedicine (IBUB), University of Barcelona, Barcelona, Spain and Department of Materials Science and Physical Chemistry, University of Barcelona, Barcelona, Spain;

3 Department of Cell Biology, Molecular Biology Institute of Barcelona, Science Research Council, Barcelona, Spain, and Networked Center for Research in Liver and Digestive Diseases (CIBEREHD), Instituto de Salud Carlos III (ISCIII), Madrid, Spain

Correspondence: martacascante@ub.edu

1. Abstract:

Cancer stem cells (CSC) play a crucial role in cancer malignancy by supporting tumor growth, metastatic spread, and drug resistance. Initially, the concept of CSC was closely associated with Epithelial-Mesenchymal Transition (EMT) as the latter was assumed to be a necessary step to achieve a CSC phenotype. However, in the last years, evidence has begun to emerge that CSC can arise independently of EMT and that in a context-specific manner EMT can even reduce stemness. In this regard, the cell lines PC3M and PC3S, derived from the prostate cancer cell line PC3, have emerged as excellent models to study the complex relationship between EMT and CSCs in prostate cancer. In this dual model, PC3M have both an epithelial-like and CSC phenotype whereas PC3S have a non-CSC and mesenchymal phenotype. Even more, inducing EMT in PC3M significantly attenuates its CSC phenotype suggesting that it is partially associated with an epithelial-like gene expression program. With the aim to analyze the complex interplay between EMT and stemness in prostate cancer and identify its molecular drivers, here we use next-generation sequencing to characterize the transcriptome of PC3M, PC3S, and a PC3M derived cell line overexpressing the EMT factor Snai1. We determine that the CSC gene expression program of PC3M is likely supported by a partial EMT phenotype mediated by a

complex balance between pro-EMT and anti-EMT factors. We also identify several putative drivers of the CSC phenotype that are suppressed by a complete EMT such as TSPAN8, REG4 or VEGFC. Finally, we integrate the quantified transcriptome in the framework of a genome-scale reconstruction of human metabolism in order to characterize the metabolic drivers underlying CSC and its relationship with EMT. We observe that the CSC metabolic gene expression program is largely disassociated from EMT status as it is not significantly inhibited by SNAI overexpression. Nevertheless, we identify several putative metabolic targets which can induce a metabolic reprogramming from a CSC-like to a non-CSC metabolic phenotype such as MTHFD2 and PYCR1. Together, the work here presented provides a greater understanding of the CSC phenotype in prostate cancer and paves the way for developing therapies to target it selectively.

2. Introduction

With an estimated 1.3 million new cases diagnosed in 2018, prostate cancer is the second most frequent cancer type in men[1]. If prostate cancer is treated at an early stage, while it still localized to the prostate, the prognosis is good. However, once it progresses to the metastatic stage and spreads to other tissues and organs, patients have a median life expectancy of only three years[2].

It is widely accepted that in prostate cancer, and indeed most cancer types, both metastatic potential, and the acquisition of resistance to therapeutic interventions (such as androgen deprivation, chemotherapy or radiotherapy) are heavily associated to a small population of cancer stem cells (CSC) found within the bulk of the tumor. CSCs are characterized by a high proliferative and self-renewal potential, and the capacity to autonomously establish tumors[3,4]. Furthermore, CSCs can survive most therapeutic interventions and repopulate the tumor, hence leading to a relapse[5–8]. Thus, acquiring a better understanding of the molecular drivers of the CSC phenotype and developing therapeutic strategies aimed at disrupting it is of paramount importance to prevent both metastatic spread and disease recurrence.

It was originally believed that the acquisition of CSC phenotype required a complete Epithelial-mesenchymal transition (EMT)[9–11]. EMT involves the loss of cell to cell adhesion, a shift from apical-basal to front-rear polarity, and results on increased motility, invasiveness and drug resistance[12,13]. As such, EMT has traditionally been considered a hallmark of cancer progression[14]. However, it is now beginning to emerge that the CSC phenotype can be acquired independently of EMT[15,16]. Indeed EMT has been reported to inhibit CSC traits such as metastatic, tumorigenic and proliferative potential[15,17–19]. For such reason, it has been suggested that cancer progression, and in particular the CSC phenotype, often involves a partial or intermediate EMT where expression of mesenchymal and epithelial markers coexist[12,13,20,21].

In this context, the PC3M and PC3S cell lines have emerged as models to study the features of prostate CSCs and their complex interplay with EMT[15,22–24]. Such cell lines were derived by single-cell cloning from the PC3 cell line, a cell line of metastatic and androgen-independent prostate adenocarcinoma[15]. PC3M are a model of EMT-independent CSCs as they are characterized by an epithelial-like and CSC-like phenotype with high proliferation rate, and high capacity to form spheroids *in vitro* and establish tumors and metastasis *in vivo*. Conversely, while PC3S have a mesenchymal-like phenotype and significantly reduced proliferative and tumorigenic potential, they have enhanced motility and local invasiveness compared to PC3M[15]. Even more, inducing EMT in PC3M was shown to significantly attenuate its CSC phenotype. In this context, overexpressing EMT factors, such as murine Snai1, in PC3M was shown to reduce proliferation rate, anchorage-independent growth and tumorigenic potential[15].

Additionally, PC3M and PC3S can also be used as models to characterize the metabolic phenotype underlying prostate CSCs. PC3M are characterized by a more active Warburg effect, glutaminolysis, β-oxidation, and one-carbon metabolism than PC3S[23,24]. Furthermore, PC3M also display higher metabolic plasticity as shown by the capacity to adapt to varying substrate availabilities[23]. The latter is recognized as a CSC metabolic feature, which endows CSC with the capacity to adapt to a variable tumor microenvironment and colonize tissues with distinct substrate availabilities[3,25]. Nevertheless, targeting the metabolism of CSC, directly or indirectly, has proven to be an effective strategy to reduce stemness and viability in PC3M and other prostate CSCs models[5,22–24].

In this regard, Genome-Scale Metabolic Models (GSMMs), mathematical representations of the entire reaction complement in humans, have emerged as platforms where multi-omics data sets can be integrated to characterize the metabolic phenotype at a genome-scale[26,27]. Indeed, such an approach has been used to characterize the differential fatty acid metabolism in the PC3M-PC3S cellular models[24]. Additionally, with the development of new algorithms such as rMTA (robust metabolic transformation algorithm)[28], GSMM can now be used to identify the most efficient therapeutic interventions to induce a transition between two distinct metabolic states. This has paved the way to allow identifying metabolic targets that can selectively rewire the metabolic phenotype underlying CSC to a non-CSC metabolic phenotype.

Here, we use RNA sequencing to characterize PC3M, PC3S, and a PC3M derived cell line overexpressing the EMT factor Snai1 (PC3M-Snai1) at a genome-scale and analyze the complex relationship between EMT-status, metabolic reprogramming and the CSC phenotype in prostate cancer. Such analysis reveals that PC3M likely display a partial-EMT phenotype characterized by

expression of both mesenchymal and epithelial markers and that inducing full EMT through Snai1 overexpression partially reverts the CSC gene expression programs. In this context, we identify several putative drivers of the CSC phenotype that are suppressed by a complete EMT such as TSPAN8, REG4 or VEGFC. Using an enhanced version of rMTA, termed r^{2}MTA, we characterize the metabolic drivers underlying CSC and its relationship with EMT. Remarkably, the metabolic gene expression program was mostly independent of partial-EMT status as most metabolic features were not reverted to the non-CSC phenotype upon Snai1 overexpression. Nevertheless, we identified putative metabolic targets that can revert the CSC metabolic phenotype to a non-CSC phenotype such as MTHFD2 and PYCR1. We hypothesize that combining metabolic inhibitors against the identified targets with inhibitors targeting the partial-EMT associated CSC-drivers could be a highly effective therapy against the CSC population in androgen-independent prostate cancer.

3. Results

3.1. Overexpressing Snai1 partially reverts the PC3M gene expression program

Previous studies revealed that inducing EMT through overexpression of Snai1 largely inhibited the CSC-like phenotype of PC3M[15,23]. To characterize the effect of Snai1 overexpression in the gene expression program of PC3M, RNA-Seq was used to analyze the transcriptome of the PC3M, PC3S, and PC3M-Snai1 cell lines and identify differentially expressed genes. Likewise, Gene Set Enrichment Analysis (GSEA) was used to identify sets of functionally related genes differentially expressed across conditions[29].

Compared to PC3S, the PC3M cell line had a strong upregulation of gene sets (GSs) associated to MYC (i.e., “Hallmark: Myc Targets” and “Oncogenic signature: MYC overexpression”) and MTORC1 signaling (i.e., “Hallmark: MTORC1 signaling” and “Oncogenic signature:MTORC1”). As both MYC and MTORC1 promote cell proliferation, motility and chemoresistance, they have been extensively associated with the CSC phenotype[30–34]. Consistent with well-known effects of both MYC and MTORC1 overactivation[35–37], PC3M displayed significantly upregulated GSs associated to gene expression machinery such as RNA processing and transport (e.g., “GO: RNA splicing” and “KEGG: RNA transport”), ribosome biogenesis (e.g., “GO: ribosome biogenesis”) and mitochondrial gene expression machinery (e.g., “GO: mitochondrial gene expression”).

Interestingly, while globally PC3M-Snai1 were still remarkably similar to PC3M at the gene expression level (Figure 1), Snai1 overexpression partially downregulated most of the GSs upregulated in PC3M compared to PC3S, hence bringing PC3M phenotypically closer to PC3S. Of

the 119 GSs significantly (FDR<0.05) downregulated in PC3S compared to PC3M, overexpressing Snai1 in PC3M significantly (FDR<0.05) downregulated 33 GSs (Figure 2A). Such number increased to 71 GSs if less significantly enriched sets were also considered (FDR<0.25 for PC3M-Snai1 vs PC3M).

The GSs significantly downregulated by Snai1 overexpression included the GSs associated with MYC and MTORC1. While both MYC and MTORC1 are primarily regulated at the posttranscriptional level[38,39], there is also strong evidence that they also are upregulated at the transcript level in PC3M compared to PC3S. MYC was strongly overexpressed in PC3M, while several MTORC1 components were also differentially expressed (Figure 3A). Namely, MLST8, an activator of mTOR, was strongly overexpressed, whereas the inhibitor AKTLS1 was strongly downregulated in PC3M. To a lesser extent, mTORC1 inhibitors TSC1 and TSC2 were also downregulated in PC3M. Interestingly, MTORC inhibitor DEPTOR was also upregulated in PC3M. However, it is worth noting that DEPTOR has been described as a putative pluripotency factor[40] and to promote proliferation and inhibit apoptosis in a context-dependent manner[41]. Conversely, Snai1 overexpression appeared to primarily downregulate MYC and MTORC1 signaling at the posttranscriptional level as there were only small variations in the abundance of the transcripts of MYC and the MTORC components(Figure 3A).

Additionally, PC3M displayed a downregulated GSs associated to type I interferon response compared to PC3S, likely enabling them to escape immune detection[42] (Figure 2B, Text S1). We posit that downregulation of type I interferon signaling might be a key part of the metastatic and CSC phenotype, facilitating the immune escape of such population. Interestingly, a pilot study in prostate cancer where chemotherapy was co-administered with interferon-α obtained mixed results and no statistically significant improvement in overall survival[43].The fact that the CSC subpopulation, which is inherently more resistant to chemotherapy, might also have increased resistance to interferon signaling might partially account for this outcome. Nevertheless, a downregulated type I interferon can also open new therapeutic opportunities against a PC3M-like CSC. Namely, owing to the antiviral role of type I interferons, such CSC population could potentially be more vulnerable to oncolytic virus-based therapies[44].Remarkably, type I interferon signaling was not restored by overexpressing Snai1, suggesting that it is not linked to EMT.

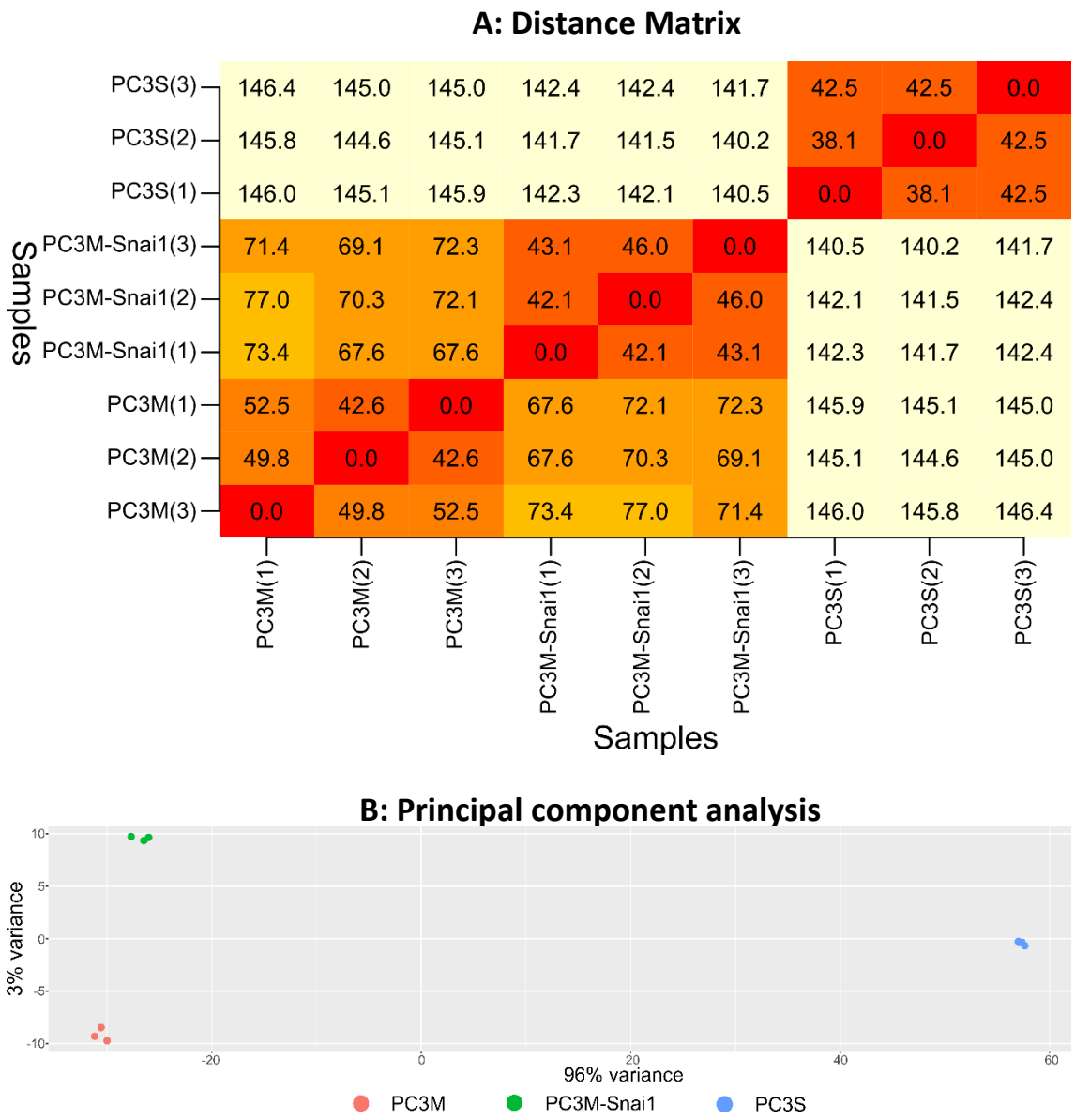


Figure 1 Similarity between PC3M, PC3M-Snai1, PC3S at the transcriptomic level. A Euclidian distance matrix between PC3M, PC3M-Snai1 and PC3S samples. B: Principal component analysis of PC3M, PC3M-Snai1 and PC3S samples.

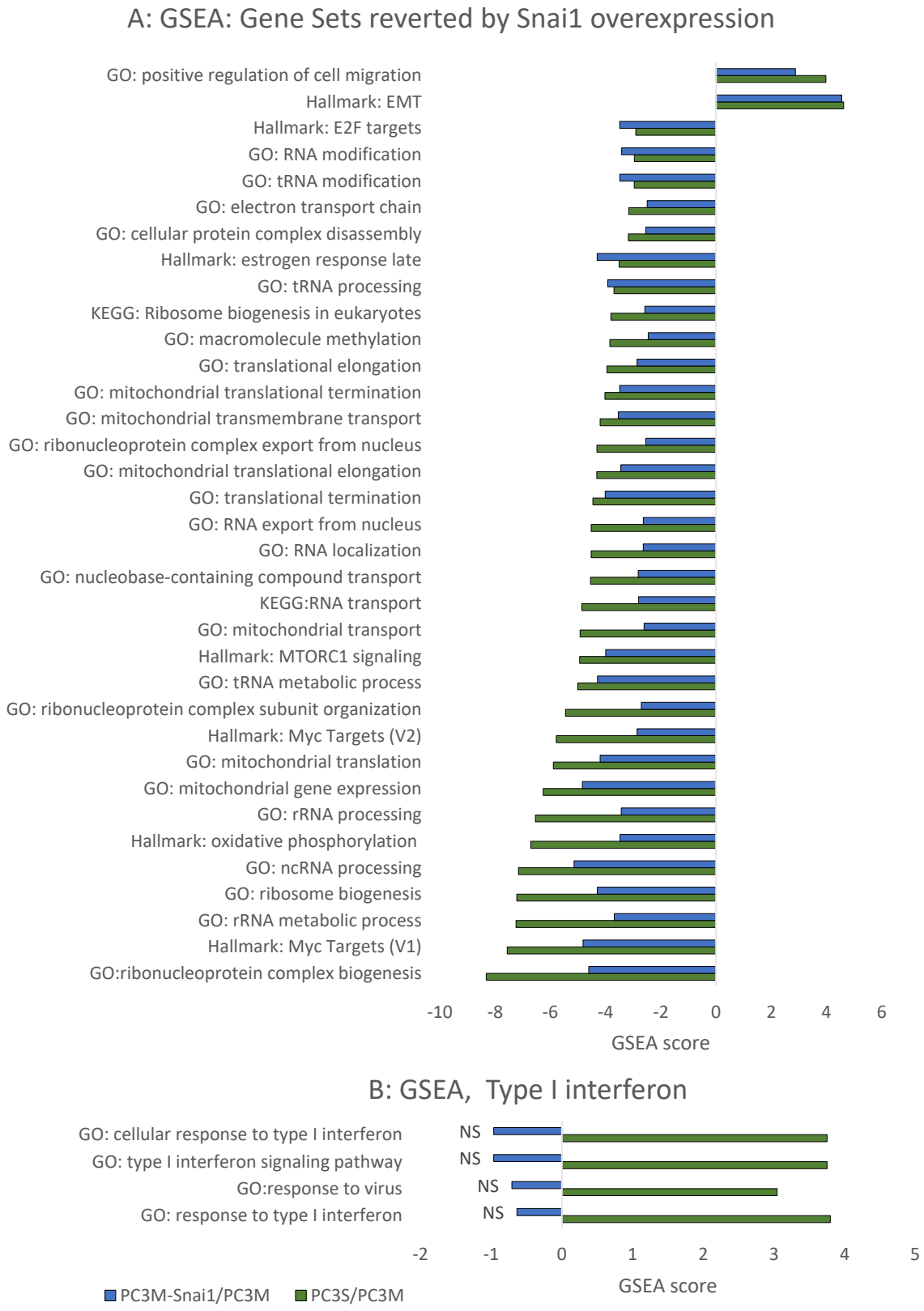


Figure 2. **A:** Gene set enrichment scores for the GSs significantly (FDR<0.05) upregulated/downregulated in PC3S compared to PC3M which significantly change in the same direction when Snai1 is overexpressed in PC3M. **B:** Gene set enrichment scores for GSs associated to type I interferon (NS: not significant, FDR>0.05)

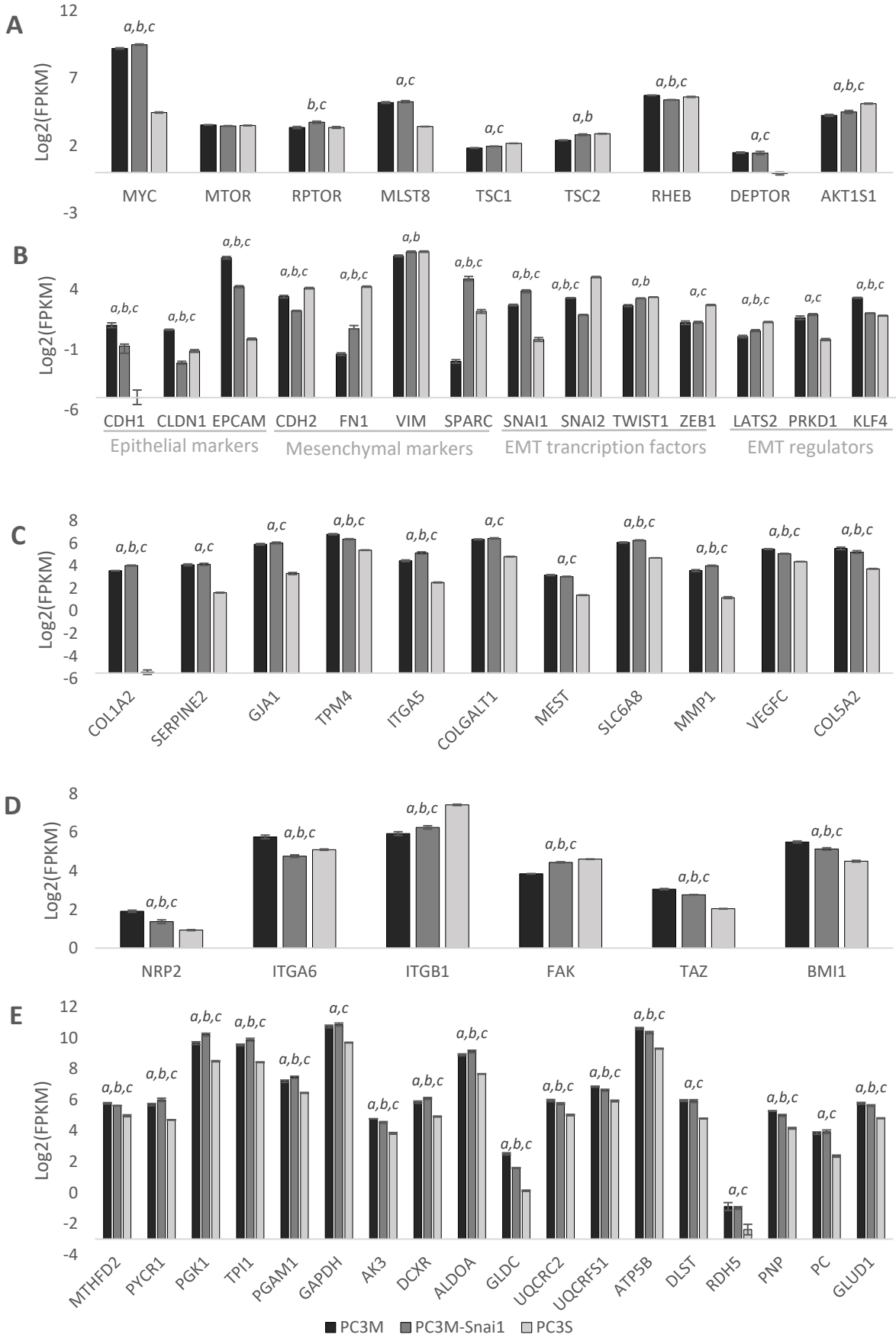

A
Log2(FPKM)
MYC
MTOR
RPTOR
MLST8
TSC1
TSC2
RHEB
DEPTOR
AKT1S1
B
CDH1
CLDN1
EPCAM
CDH2
FN1
VIM
SPARC
SNAI1
SNAI2
TWIST1
ZEB1
LATS2
PRKD1
KLF4
Epithelial markers
Mesenchymal markers
EMT trancription factors
EMT regulators
C
COL1A2
SERPINE2
GJA1
TPM4
ITGA5
COLGALT1
MEST
SLC6A8
MMP1
VEGFC
COL5A2
D
NRP2
ITGA6
ITGB1
FAK
TAZ
BMI1
E
MTHFD2
PYCR1
PGK1
TPI1
PGAM1
GAPDH
AK3
DCXR
ALDOA
GLDC
UQCRC2
UQCRFS1
ATP5B
DLST
RDH5
PNP
PC
GLUD1
PC3M
PC3M-Snai1
PC3S

Figure 3 Transcript levels of **A:** MYC, and MTORC1 components. **B:** Expression of epithelial and mesenchymal biomarkers, EMT transcription factors and EMT regulators. **C:** Top EMT associated genes overexpressed in PC3M. **D:** Expression of genes associated to VEGFC/NRP2 signaling **E:** Metabolic genes identified as putative targets against CSC metabolism with r^2MTA. Expression levels are indicated as FPKM (Fragments Per Kilobase of transcript per Million mapped reads) in a Log2 scale. *a,b* and *c* denote a statistically significant difference (FDR<0.05) between PC3M and PC3S, PC3M and PC3M-Snai1 and PC3M-Snai1 and PC3S, respectively.

3.2. PC3M display partial EMT modulated by SNAI1, SNAI2, KLF4, PRKD1 and LATS2

As previously reported, PC3M displayed a more epithelial phenotype and had less motility than PC3S[15]. Accordingly, PC3M had downregulated GSs associated with the EMT and cellular migration ("Hallmark: Epithelial-Mesenchymal Transition" and "GO: positive regulation of cell migration") compared to PC3S. As expected, overexpressing Snai1 upregulated the GSs associated with EMT and cell migration (Figure 2A).

In this regard, PC3M strongly overexpress several epithelial markers such as E-cadherin (CDH1), Claudin1 (CLDN1) and Epithelial Cell Adhesion Molecule (EPCAM) while they strongly underexpressed several mesenchymal markers such as Fibronectin 1 (FN1) and Secreted Protein Acidic And Cysteine Rich (SPARC) compared to PC3S. However, PC3M also displayed a strong expression of many mesenchymal markers, such as Vimentin (VIM) or N-cadherin (CHD2). In this regard, the simultaneous expression of epithelial and mesenchymal markers in the PC3M subpopulation is the definition of partial-EMT[13]. Supporting the notion of partial EMT, PC3M also had a strong expression of EMT transcription factors such as SNAI1, SNAI2, and TWIST1(Figure 3B).

Increasing the dosage of SNAI1 through Snai1 overexpression (PC3M-Snai1) resulted in an increased EMT as seen by a strong downregulation of the transcription of epithelial markers (i.e., CDH1, CLDN1, EPCAM) and a strong upregulation of mesenchymal markers (i.e., FN1, SPARC) consistent with previous works[15]. This suggests that the partial EMT underlying the PC3M phenotype is mediated by a delicate balance between pro-EMT factors (such as SNAI1) and anti-EMT factors and that increasing the dosage of the former tilts the scale in favor of EMT.

Krüppel-like factor 4 (KLF4) is a transcription factor associated with pluripotency, which can either promote EMT or inhibit EMT in a context-dependent manner[45]. Remarkably, KLF4 can act

as a direct transcriptional activator of epithelial genes, such as CDH1, and a repressor of both mesenchymal markers[46–48]and EMT transcription factors such as SNAI1 and SNAI2[49–51]. Furthermore, KLF4 transcription is repressed by SNAI1 and SNAI2[49,52] and hence KLF4 is downregulated in SNAI1/SNAI2 induced EMT. At the transcript level, KLF4 is strongly overexpressed in PC3M compared to PC3S and strongly downregulated upon Snai1 overexpression (Figure 3B).

Additionally, the activity of EMT factors, such as SNAI1, can also be regulated at the posttranslational level. In this regard, PRKD1 is a serine/threonine kinase that can lead to an inactive SNAI1/DNA complex[53] or trigger SNAI1 nuclear export and ubiquitination in a context-dependent manner[54,55]. Furthermore, PRKD1 can also inhibit EMT in a SNAI1-independent manner[56,57] and has been termed "the gatekeeper of the epithelial phenotype"[58]. Conversely, SNAI1 activity is enhanced by the LATS2 kinase, which facilities SNAI1 retention in the nucleus[59]. Remarkably, PRKD1 was strongly upregulated in PC3M while LATS2 was strongly upregulated in PC3S and upon Snai1 overexpression (Figure 3C).

We posit that high KLF4 and PRKD1 expression and low LATS2 expression serve as a counterweight to SNAI1 and SNAI2 activities in PC3M and play a crucial role in maintaining partial EMT and stemness. Indeed, the key role of KLF4 as a regulator of EMT in the PC3M/PC3S models was confirmed by the decreased expression of CDH1 and increased motility previously observed in PC3M upon KLF4 knockdown[15].

3.3. PC3M overexpress mesenchyme-associated genes associated with metastatic spread

Associated with the partial EMT phenotype, PC3M also overexpressed several genes related to the mesenchymal phenotype (i.e., part of the GS "Hallmark: EMT") compared to PC3S (Figure 3C). Close inspection of such genes revealed that many of them were involved in adhesion, migration and the remodeling of the extracellular matrix. Furthermore, most of these genes have been associated with poor prognosis and, invasive and metastatic potential in a variety of cancer types (Box 1). We posit that high expression of such genes might confer PC3M the invasion, extravasation and intravasation capabilities necessary to be able to establish distant metastases autonomously and by targeting such genes, we might be able to reduce the tumorigenic and metastatic potential of PC3M significantly. Of such genes, COL1A2, ITGA5, SLC6A8, and MMP1 were upregulated upon Snai1 overexpression, suggesting that they are at least partially controlled by SNA1. Remarkably, TPM4, VEGFC, and COL5A2 were significantly

downregulated by increased Snai1 dosage, indicating that they are not directly regulated by such EMT factor.

BOX1: Cancer-associated mesenchymal genes overexpressed in PC3M

COL1A2 (collagen, type I, alpha 2). It is associated with poor prognosis and metastatic potential in lung, breast, prostate, colorectal, uterus, ovary[60] and gastric cancer[61]. COL1A2 is also overexpressed in metastatic and androgen-independent prostate cancer[62]. Furthermore, silencing of COL1A2 in gastric cancer cell lines induced apoptosis and inhibited proliferation and invasion *in vitro*[63].

SERPINE2 (serpin peptidase inhibitor, clade E member 2). It is a serine protease inhibitor known to modulate a large number of biological processes[64]. It has been associated with poor prognosis and invasive and metastatic potential in medulloblastoma[65], endometrial cancer[66], gastric cancer[67], pancreatic cancer[68], melanoma[69] and colorectal cancer[70]. Additionally, in osteosarcoma it has been associated with drug resistance[71].

GJA1 (gap junction protein, alpha 1). A member of the connexin family, proteins that assemble into gap junctions. GJA1 is often overexpressed in metastatic cancer cell populations compared to primary tumor populations[72]. Its activity has been found to enhance cell migration and/or extravasation in breast cancer[73], glioma[74], melanoma[73], cervical cancer[75] and prostate cancer[76,77].

ITGA5 (integrin, alpha 5). ITGA5 has been associated with poor prognosis in non-small-cell lung cancer [78], ovarian cancer[79,80], breast cancer[81] and hepatocellular carcinoma[82]. Targeting ITGA5 has been found to reduce the formation of metastasis *in vivo* in ovarian[79] and breast cancer[81]. Likewise, ITGA5 silencing reduced the viability and invasiveness of glioblastoma cell lines[83].

MEST (mesoderm specific transcript): MEST is an imprinted gene, expressed only from the paternal allele, that is reported to promote neuronal migration[84]. MEST imprinting is frequently lost in breast and lung cancer[85,86].

MP1 (Matrix metallopeptidase 1). It has been associated with poor prognosis in breast [87], lung[88] and colorectal cancer[89]. In colorectal cancer cell lines, invasive potential was associated with MMP1 expression[90]. In an osteosarcoma cell line, overexpression of MMP1 increased proliferation rate and invasiveness as well as the expression of stem cell marker Pax-7[91].

SLC6A8(Solute Carrier Family 6 Member 8). It is a phosphocreatine transporter that has been found to be upregulated in colorectal cancer patients with p53 mutations and poor prognosis[92]. Likewise, in colorectal cancer, it has been suggested to promote liver metastasis by allowing the usage of phosphocreatine to fuel ATP production in hypoxia[93].

VEGFC(Vascular Endothelial Growth Factor C). A signaling protein which acts both on the tumor microenvironment and tumor cells(Wang and Tsai 2015).

COL5A (Collagen Type V Alpha 1 Chain). It is associated with poor prognosis in bladder[94] and ovarian cancer[95].

3.4. VEGFC/NRP2 signaling: a potential driver of the CSC phenotype in PC3M

VEGFC is one of the mesenchymal-associated genes overexpressed in PC3M and its expression has been reported to be activated by KLF4[96]. Even more, VEGFC signaling through the Neuropilin 2(NRP2) receptor has been associated with the CSC phenotype in breast[97–99] and prostate cancer[100,101]. Upon VEGFC binding, NRP2 can interact laterally with ITGA5, ITGA6 or ITGB1 and activate FAK (focal adhesion kinase) signaling[98,102,103]. VEGFC/NRP2-dependent FAK-mediated signaling has been reported to activate the TAZ[99] and BMI1[98,100] stemness transcription factors. Similarly, VEGFC/NRP2 has also been reported to activate mTORC1 in a FAK-dependent manner[100].

In our cell lines of study the VEGFC, NRP2, TAZ, and BMI1 gene expression patterns appear to be correlated with stemness (PC3M>PC3M-Snai1>PC3S) (Figure 3D). Additionally, even if FAK expression is downregulated in PC3M compared to PC3M-Snai1 and PC3S, it is nevertheless highly expressed in all subpopulations. Furthermore, integrins know to interact laterally with NRP2 are also highly expressed across the models of study. Therefore, we theorize that the VEGFC/NRP2 signaling cascade might play a key role in maintaining the CSC-like phenotype in the PC3M cell line.

In addition to the autocrine activity, VEGFC can also act on lymphatic endothelial cells where it promotes lymphangiogenesis and decreases the expression of endothelial-specific cadherin facilitating trans-endothelial migrations of cancer cells into the lymph vessels[104]. Accordingly, VEGFC expression in tumors enhances metastases to lymph nodes[104–107]. Indeed, a defining feature of PC3M compared to PC3S is their increased capacity to metastasize to lymph nodes in immunosuppressed mice[15].

3.5. Identifying key players of the CSC phenotype associated with partial-EMT

Given that Snai1 overexpression attenuates significant aspects of the PC3M phenotype, we posit that key players of the CSC phenotype can be identified in the subset of genes strongly downregulated (Log2 Fold Change <1.5) both in PC3S and PC3M-Snai1 compared to PC3M. A total of 45 genes were found matching these criteria and such number was further reduced to 38 by excluding genes that are epithelial markers. Too further narrow the list of candidates, copy number variation (CNA) and survival analysis were used to select those genes that were more frequently amplified in prostate cancer and whose amplification or mutation correlated with poor clinical prognosis. Overall, 9 potential candidates were identified (Figure 4, Figure S1,Box 2).

Most of such genes have been extensively associated with features of CSC such as metastatic and invasive potential, cell proliferation, apoptosis or drug resistance. Indeed, both REG4 and TSPAN8 have been shown to promote AKT-mediated signaling[108–112] and could partially account for the increased activity of MYC and MTORC1, downstream targets of AKT, inferred from GSEA.

Additionally, it is worth noting that both ABCA3 and TSPAN8 have been associated with exosome release. Even more, TSPAN8-expressing were exosomes reported to enhance metastatic potential and CSC properties of the recipient cancer cells[113,114]. Likewise, SIDT1 allows the contact-dependent transfer of small RNA between cells, and this has shown to allow the propagation of miRNA conferring chemoresistance[115]. Hence, expression of ABCA3, TSPAN8 and SIDT1 might confer a small population of PC3M-like CSC the capacity to enhance the malignancy of the bulk of the tumor population[115].

We hypothesize that expression of such genes is linked to partial-EMT status and accounts for the loss of CSC phenotype upon Snai1 overexpression. Remarkably, in colorectal cancer, TSPAN8 activity has been both reported to require the formation of a complex with proteins associated with the epithelial phenotype (i.e., EpCAM, Claudin 7 and CD44v6)[116] and to be inhibited by E-cadherin[117]. Hence, TSPAN8 activity could be mostly dependent on maintaining a partial-EMT gene expression program.

Finally, although it is not clear the role, if any, that CR2 might play in prostate cancer, differential CR2 expression could still potentially be used to selectively target the PC3M subpopulation through liposome-mediated drug delivery[118].

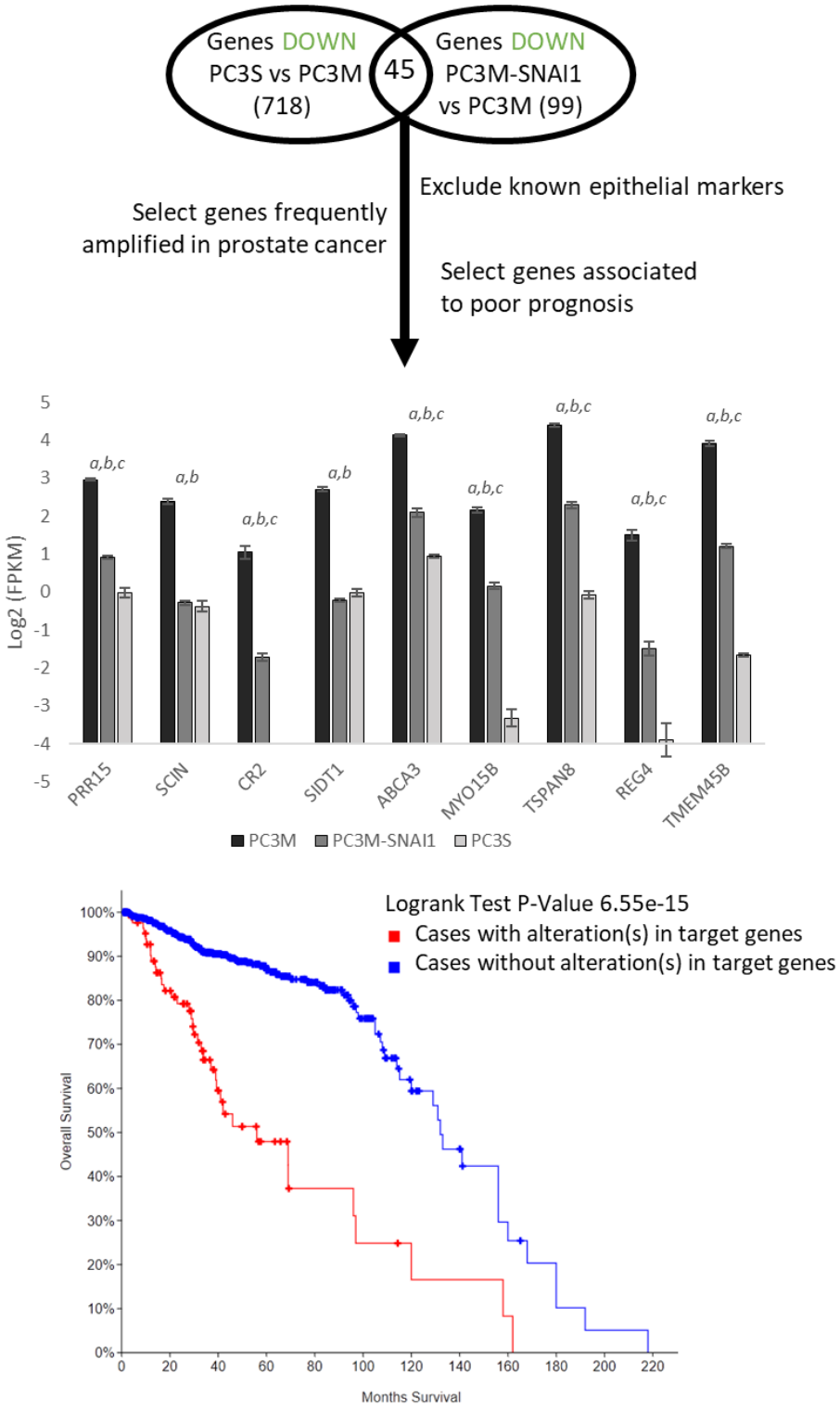


Figure 4 Key players of the CSC phenotype associated to partial EMT. Such genes are downregulated in PC3S vs PC3M, downregulated upon Snai1 overexpression and are characterized by being frequently amplified in prostate cancer with such alterations associated to poor prognosis. As a reference, a Kaplan-Meier plot for alterations in one or more of the putative targets are provided. Plots for individual targets can be found on figure S1.

Box 2: Putative drivers of the CSC phenotype associated with EMT

PRR15 (Proline-Rich Protein 15). Silencing PRR15 has been shown to reduce proliferation, viability, and invasion in trophoblast cell lines[119]. Additionally, in colorectal cancer, PRR15 expression has been found to correlate with primary tumor stage[120].

SCIN (Scinderin). SCIN is a regulator of actin organization. In gastric cancer, SCIN expression was associated with poor prognosis, and SCIN silencing reduced invasive and metastatic potential *in vivo*[121]. Silencing SCIN has also been found to reduce proliferation and increase apoptosis in prostate[122,123], lung [124] and breast cancer[125]. The studies in prostate cancer, performed in PC3 and DU145 cell lines, also indicated that silencing of SCIN decreased EGFR/MAPK signaling and inhibited tumorigenic potential in mice[122]. SCIN expression has also been linked to drug resistance in hepatocellular carcinoma[126] and bladder cancer[127].

CR2 (Complement C3d Receptor 2). This receptor is primarily found in lymphocytes where it recognizes antigens bound to complement 2. It also functions as a receptor for Epstein-Barr virus, which can promote nasopharyngeal carcinoma[128,129].

SIDT1(SID1 Transmembrane Family Member 1). It is a transmembrane channel-like protein that mediates contact-dependent small RNA transfer between cells[115], and uptake and intracellular transport of cholesterol[130]. It has been suggested to play a role in chemoresistance by allowing the contact-dependent transfer of a miRNA that enables chemoresistance between cells[115]. Furthermore, SDIT1 expression has been associated with poor prognosis in breast cancer[115] and a SNIP in the SIDT1 gene has been associated with increased risk of prostate cancer[131].

ABCA3(ATP Binding Cassette Subfamily A Member 3). It enhances the transport of phosphatidylcholine and cholesterol to intracellular vesicles and promotes lysosomes formation and exosome release[132–134]. It has been shown to promote resistance to chemotherapy in leukemia and non-small cell lung cancer[132,135–138]. ABCA3 is believed to promote resistance by enhancing lysosomal drug sequestration and drug efflux[132,138,139]

MYO15B(Myosin XVB) is frequently amplified in breast cancer cell lines and in the cell line MDA-MB-453 has been identified as a fusion protein with MAP3K3[140,141].

TSPAN 8(Tetraspanin 8) is a transmembrane protein, found both in the cellular membrane and exosomes and can interact laterally with other membrane proteins including integrins and cadherins[117,142]. TSPAN 8 has been associated with poor prognosis or metastatic and invasive potential in colorectal cancer[117], ovarian cancer[143], melanoma[111,112], pancreatic cancer[113], glioma[144,145], esophageal carcinoma[146] and hepatocellular carcinoma[147]. Additionally, it is upregulated in prostate cancer cell lines that have acquired drug resistance[148]. In melanoma, TSPAN8 was found to be part of a positive feedback loop with β-catenin[149] and increase AKT phosphorylation and invasive potential[111,112]. In glioma, TSPAN8 acts through FAK[144] and mTORC2 increasing AKT and PKCα phosphorylation[144,145]. Uptake of exosomes containing TSPAN8 by endothelial cells has been shown to promote angiogenesis[150,151]

Box 2: Putative drivers of the CSC phenotype associated with EMT (Continued)

REG4(Regenerating Family Member 4). Is a secretory protein member of the calcium-dependent lectin family. Overexpression of REG4 and high REG4 levels in serum have been associated with poor prognosis and metastatic potential in colorectal cancer[43,152,153], glioma[154], gastric cancer[155–157], prostate cancer[158], ovarian cancer[159] and pancreatic cancer[160]. REG4 expression has also been associated with resistance to chemotherapy or radiotherapy in colorectal cancer[161–163], pancreatic cancer[164] and gastric cancer[165,166]. Additionally, in gastric cancer REG4 was found to be overexpressed in the CSC subpopulation, and CSC phenotype was significantly reduced upon REG4 silencing[167]. Likewise, in prostate cancer REG4 overexpression has been suggested as a potential marker of metastatic and androgen-independent tumors[168]. Concerning the mechanism of action, REG4 acts by activating tyrosine kinase receptors such as EGFR, IR, and IGFR[108,109].The activation of such receptors by REG4 results on the activation of downstream signal transduction proteins (e.g. AKT, β-catenin, Jun or MAPK) resulting in increased cell proliferation, invasive capacity and increased apoptosis resistance[108–110,159,165].

TNEM45B(Transmembrane Protein 45B). It is a member of the transmembrane protein family. It has been reported to promote cell proliferation, invasion and migration, apoptosis resistance in gastric cancer, pancreatic cancer, lung cancer and osteosarcoma[169–172]. High expression of TMEM45B has been associated with poor prognosis in lung cancer[172]. Little is known of the mechanism of action of TNEM45B but it has been reported to activate the JAK2/STAT3[169] and β-Catenin pathway[170].

3.6. Metabolic reprogramming in the PC3M/PC3S/PC3M-Snai1 models

Concerning metabolism, GSEA shows that several GSs associated to metabolism were upregulated in PC3M compared to PC3S, namely GSs associated to glycolysis (“KEGG: Glycolysis/ Gluconeogenesis”) and amino acid metabolism (e.g., “GO: cellular amino acid metabolic process” and “Kegg Biosynthesis of amino acids”). Furthermore, PC3M also displayed upregulated GSs associated with oxidative phosphorylation (e.g., “Hallmark: oxidative phosphorylation” and “GO: respiratory electron transport chain”) and transport of substances into, out of or within a mitochondrion (e.g., “GO: mitochondrial transport”). Remarkably, only the GSs “Hallmark: oxidative phosphorylation” and “GO: mitochondrial transport” were significantly downregulated upon Snai1 overexpression(Figure 5.A).

However, GSEA can only offer a limited view of metabolic reprogramming as it does not explicitly consider the complex interdependencies of enzymes within a metabolic network that give rise to the metabolic phenotype. Such interdependencies can be studied using constraint-based modeling in the framework of GSMMs[26,27]. In that regard, we developed r^{2}MTA, a constrained-based algorithm that allows us to robustly characterize the metabolic transition between two conditions in the framework of GSMMs. Here, we integrated the RNA-SEQ data together with

previously published metabolomics data for PC3M and PC3S[23] to characterize the metabolic differences between PC3S/PC3M and PC3M-Snai1/PC3M.

Consistent with GSEA, most of the metabolic differences were identified only between the PC3M and PC3S cell lines. The main pathways of central carbon metabolism (i.e., glycolysis, pentose phosphate pathway, and tricarboxylic acid cycle) were strongly upregulated in PC3M compared to PC3S (Figure 5.B). Likewise, glutamate and proline metabolism were also strongly upregulated in PC3M. Furthermore, PC3M also displayed a highly upregulated folate metabolism fueled by cleavage of glycine and serine, with both glycine and serine uptake and de novo serine synthesis increasing in PC3M. Also consistent with GSEA, oxidative phosphorylation was also upregulated in PC3M. Interestingly, while PC3M overexpressed some nucleotide kinases (e.g., adenylate kinase 1 and adenylate kinase 3), they had less expression in some key nucleotide kinases activity (chiefly thymidine kinase 1) and ribonucleotide reductase enzymes (e.g., ribonucleotide reductase M2 B) resulting in less overall activity on the nucleotide interconversion subsystem.

Remarkably, glutathione metabolism was strongly downregulated upon Snai1 overexpression. This is in line with a previously reported downregulation of glutathione metabolism induced by Snai1 overexpression[173]. Additionally, Snai1 overexpression also downregulated the cholesterol metabolism and upregulated nucleotide interconversion (chiefly by upregulating thymidine kinase 1 and ribonucleotide reductase M2) pathways bringing PC3M metabolite closer to PC3S. However, most of central carbon and amino acid metabolism was not significantly affected by Snai1 overexpression. We theorize that as Snai1 overexpression has been reported to promote glycolytic and tricarboxylic acid cycle activity[174–176], it might contribute to masking the effects of decreased MYC and MTORC1 signaling on metabolism.

Together, GSEA and r^2MTA indicate that the metabolic gene expression program underlying the CSC-like PC3M metabolic phenotype was not strongly associated with partial-EMT.

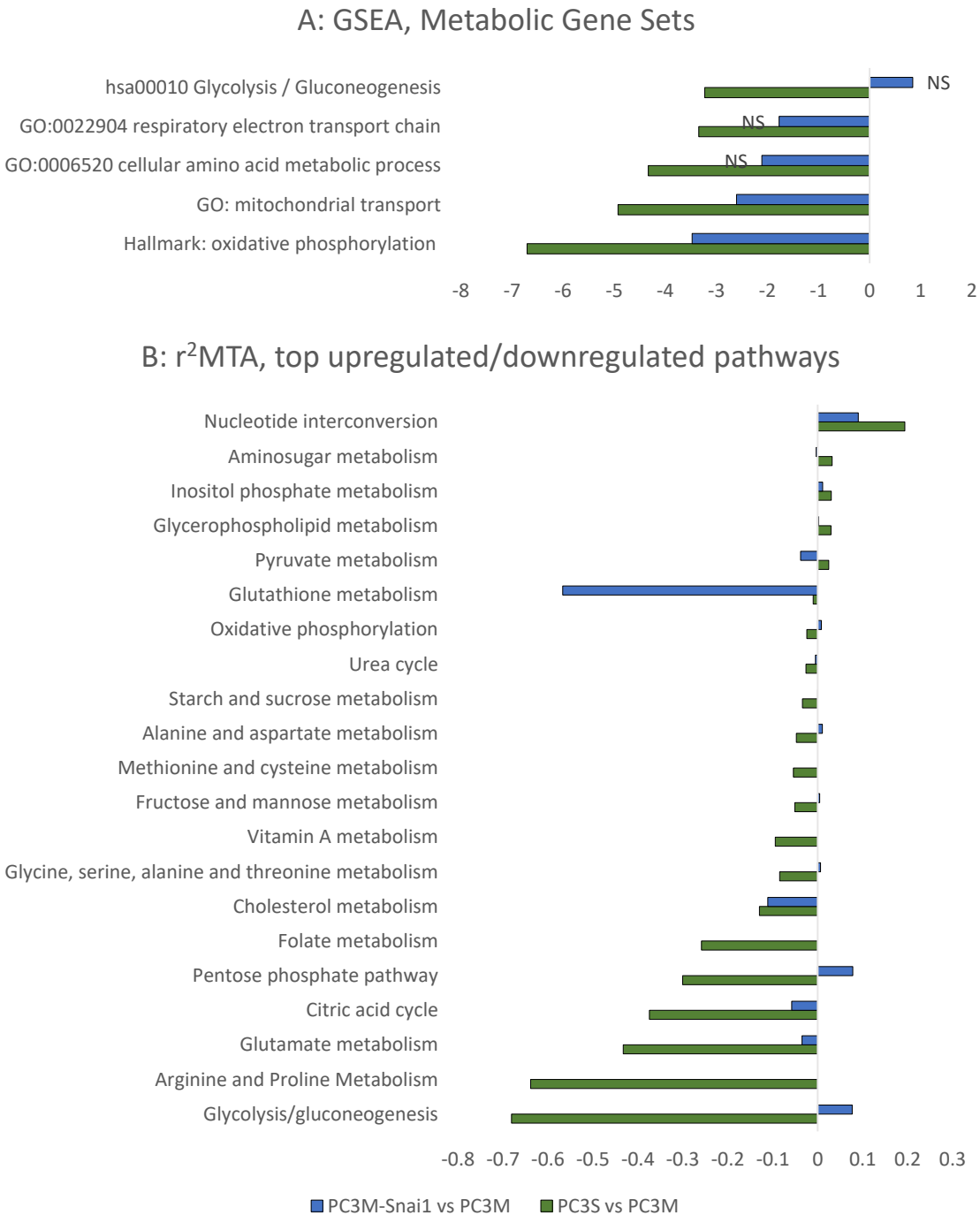


Figure 5: **A.** GSEA scores for metabolic genes sets. Non-significant (FDR>0.05 scores) are denoted with NS. **B.** Top pathways upregulated or downregulated between PC3S and PC3S and PC3M and ,PC3M-Snai1 and PC3M. The sum of reaction flux variations as predicted by r^2MTA are plotted. Positive values represent pathways upregulated in PC3S or PC3M-Snai1 compared to PC3M while negative values represent pathways downregulated in PC3S or PC3M-Snai1 compared to PC3M.

3.7. Identifying metabolic targets against CSC metabolism

Although the CSC metabolic phenotype was not reverted to a non-CSC state upon Snai1 overexpression, with r^2MTA we identified putative metabolic targets that can be inhibited to cause a metabolic shift from a PC3M-like metabolic phenotype to a PC3S-like phenotype. Following the approach of Valcárcel *et al*[28], only putative targets strongly overexpressed in PC3M (fold change>2) were considered. As expected, the pathways upregulated in PC3M compared to PC3S were heavily enriched in putative targets to transform PC3M to PC3S (Table 1, Figure 3E). Namely, folate metabolism (MTHFD2 and GLDC), proline metabolism (PYCR1), glycolysis (PGK1, TPI1, PGAM1, GAPDH), oxidative phosphorylation (UQCRC2, UQCRFS1, ATP5B, glutamate metabolism (GLUD1) and nucleotide interconversion (AK3).

Table 1: Top r^2MTA targets to transform the PC3M metabolic phenotype to a PC3S-like metabolic phenotype.

Gene	*Name*	*Log2(MTA Score)*
MTHFD2	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2; methenyltetrahydrofolate cyclohydrolase	11.43
PYCR1	pyrroline-5-carboxylate reductase 1	10.53
PGK1	phosphoglycerate kinase 1	8.70
TPI1	triosephosphate isomerase 1	8.68
PGAM1	phosphoglycerate mutase 1	8.63
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	8.56
AK3	adenylate kinase 3	8.48
DCXR	dicarbonyl and L-xylulose reductase	7.82
ALDOA	aldolase; fructose-bisphosphate A	6.55
GLDC	glycine decarboxylase	6.23
UQCRC2	ubiquinol-cytochrome c reductase core protein II	5.99
UQCRFS1	ubiquinol-cytochrome c reductase; Rieske iron-sulfur polypeptide 1	5.99
ATP5B	ATP synthase; H+ transporting; mitochondrial F1 complex; beta polypeptide	5.66
DLST	dihydrolipoamide S-succinyltransferase	5.23
RDH5	retinol dehydrogenase 5	4.66
PNP	purine nucleoside phosphorylase	4.60
PC	pyruvate carboxylase	4.57
GLUD1	glutamate dehydrogenase 1	4.37

Interestingly, multiple targets were validated by the previous metabolic characterization of the PC3M model. Namely, oligomycin (an inhibitor of ATP synthase), BPTES (an inhibitor of glutaminase which is upstream of glutamate dehydrogenase in the glutaminolysis pathway) and

2-deoxyglucose(a glycolytic inhibitor) greatly reduced spheroid formation in PC3M[23]. Furthermore, both BPTES and 2-deoxyglucose reduced proliferation to a higher extent in PC3M than PC3S[23]. Even more, Metformin (reported to inhibit oxidative phosphorylation[5]), and 2-deoxyglucose[177] have been reported to increase the response to chemotherapy and radiotherapy, in prostate cancer.

Concerning some of the other top-scoring targets, MTHFD2 expression is known to be induced by both MYC[178] and MTORC1[179] and correlates with poor prognosis in breast[180], lung[181], pancreatic[182] and renal[183] cancer. MTHFD2 is a key enzyme of the mitochondrial folate pathway which facilitates folate recycling and provides one-carbon units for purine and glycine synthesis[180]. In lung cancer, MTHFD2 was found to promote tumorigenesis and CSC-like properties by enhancing purine synthesis and by reducing the concentration of the purine metabolism intermediary 5-Aminoimidazole-4-carboxamide ribonucleotide, which can inhibit both cell proliferation and stemness[181]. Similarly, knockdown of MTHFD2 in acute myeloid leukemia cells decreased cellular proliferation and increased cell differentiation[178].

Regarding PYCR1, it is a mitochondrial enzyme that catalyzes the synthesis of proline from 1-pyrroline-5-carboxylate[184]. PYCR1 expression is induced by MYC[184] and correlates with poor prognosis in melanoma[185] and non-small cell lung cancer[186]. PYCR1 expression has been associated with increased proliferation and apoptotic resistance in melanoma[185], prostate cancer[187] and non-small cell lung cancer[186] cell lines and *in vivo* growth in breast cancer[188]. Remarkably, the effects of silencing PYCR1 in non-small cell lung cancer could not be mitigated through the addition of proline in the culture media[186]. In this regard, it has been suggested that PYCR1 might contribute to tumorigenesis by reducing the concentration of pyrroline-5-carboxylate[185,186], which is reported to inhibit proliferation and induce apoptosis[189], and by regulating the mitochondria redox balance[180,190]. Interestingly, both MTHFD2 and PYCR1, have a low or absent expression in most adult healthy tissues, including proliferating cell types associated to the digestive and immune systems, suggesting that targeting either MTHFD2 or PYCR1 could have little off-target effects[180].

Remarkably, one of the predicted targets, GLDC, was strongly downregulated upon Snai1 overexpression. This gene is part of the glycine cleavage system which generates one-carbon units for folate metabolism and it has been associated with tumorigenesis and pluripotency[191–193]. Hence, even if folate metabolism might not be coordinately downregulated at the gene expression level as result of increased Snai1 dosage, it might be significantly reduced through the downregulation of GLDC.

3.8. CSCs are selectively vulnerable to antifolates

Because folate metabolism was found to be strongly upregulated in PC3M compared to PC3S, and indeed two putative metabolic targets were identified in folate metabolism (i.e., MTHFD2 and GLDC), we theorized that the PC3M population could be differentially vulnerable to antifolates. In this regard, we evaluated the potential therapeutic effects of methotrexate and pemetrexed, two FDA-approved antifolates[194], on PC3M and PC3S(Figure 6). Both drugs were shown to have stronger growth inhibitory effects on PC3M than PC3S.

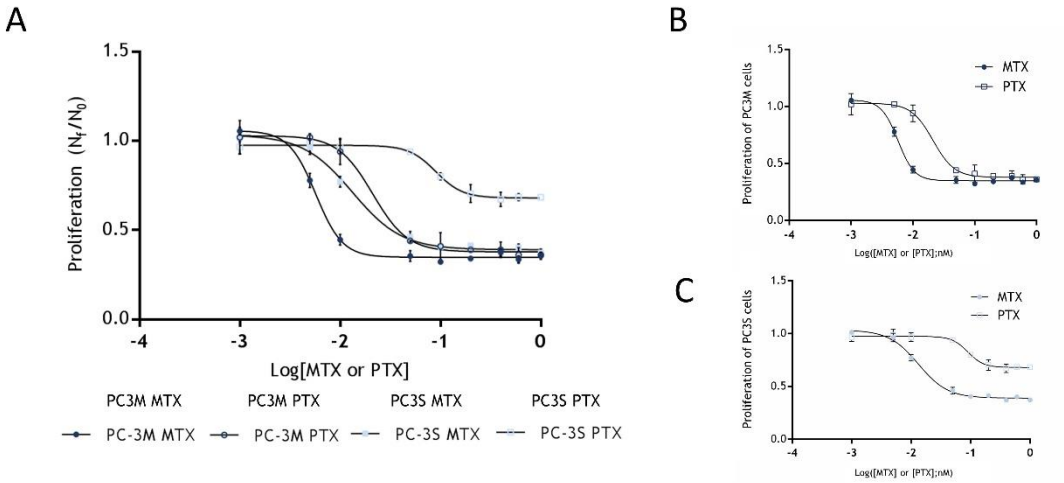


Figure 6. Effect of folate analogs methotrexate (MTX) and pemetrexed (PTX) on the proliferation of PC3M and PC3S cells. **A:** Effect of MTX on the cell proliferation of PC3M (solid circles) and PC3S (solid squares); and PTX on PC3M (open circles) and PC3S (open squares). Cell viability was assessed after 72 h incubation with either MTX or PTX (nM). Values represent mean ± SD of n=3. **B:** Comparison of the effect of MTX and PTX on PC3M. **C:** Comparison of the effect of MTX and PTX on PC3S.

4. Conclusions

Previous works established that both EMT or lack of thereof and metabolic reprogramming might play a key role in the CSC phenotype of prostate cancer[15,23,24,195]. Hence, to selectively target CSC populations in prostate cancer, it is of paramount importance to gain a better understanding of the complex interplay between EMT, metabolic reprogramming and CSC traits. With this aim, we used next-generation sequencing to analyze the gene expression program of the PC3M,PC3S, and PC3M-SNAI1 cell lines. In such model, PC3M had both an epithelial-like and CSC phenotype, which could be inhibited through SNAI1 overexpression (PC3M-SNAI1), whereas

PC3S had a mesenchymal-like non-CSC phenotype. Hence, by analyzing the differential gene expression and metabolic reprogramming between the abovementioned cell lines, the complex molecular associations between metabolism, EMT and CSC can be unraveled.

We determined that PC3M displayed a partial EMT phenotype and indeed overexpressed a significant number of mesenchymal associated genes compared to PC3S. Indeed, forcing a higher degree of EMT to PC3M through Snai1 overexpression partially reversed a significant number of the CSC-associated gene programs differentially expressed between the PC3M and PC3S cellular models. We posit that the balance between the anti-EMT (e.g., KLF4 and PRKD1) and pro-EMT factors (e.g., SNAI1, SNAI2, and LATS2) underlies the partial EMT phenotype in prostate cancer CSCs and that disrupting such balance can potentially be a valid therapeutic strategy.

We also identified several genes associated with CSC traits (i.e., genes that promote proliferation, chemotherapy resistance or facilitate invasiveness and metastatic spread) that were downregulated upon Snai1 overexpression, and hence, likely associated to partial EMT. Such genes include secretory proteins with potential autocrine and paracrine roles (e.g., VEGFC and REG4), and surface proteins (e.g., ITGA5, NRP2, TMEM4B and TSPAN 8) potentially involved in signaling cascades that endow the PC3M cell line with CSC phenotype[97,99,100,109,113,144,167,170,172]. Blocking such secretory and surface proteins with antibodies could be an effective strategy against CSC populations. Additionally, as both VEGC/NERP2 and TSPAN 8 signaling is mediated through the FAK[98,102,103,144], inhibiting FAK is also likely to be an effective strategy against the CSC population in prostate cancer. Indeed, inhibiting FAK has proven effective against CSC populations in both pancreatic and breast cancer[196–198]. Additionally, some of such genes are likely to be regulated directly or indirectly by KLF4, hence targeting this stem cell factor could also be a practical therapeutic approach[96].

Furthermore, we analyzed the metabolic reprogramming underlying the CSC phenotype at a genome-scale and identified that PC3M were endowed with a highly active central carbon metabolism, folate metabolism and glutamate and proline metabolism which likely supports their CSC-like phenotype. Globally, the metabolic gene expression program underlying CSC appeared to be mostly independent of partial EMT as it was not generally reverted by overexpressing Snai1. However, we identified several putative metabolic targets that could inhibit the CSC phenotype by targeting the underlying metabolic reprogramming. Two of the identified targets (i.e. MTHFD2 and GLDC) were activities of folate metabolism and were indirectly validated with methotrexate and pemetrexed, two antifolates that are already

approved for clinical use[194]. Such drugs were shown to have selective growth inhibitory effect on PC3M suggesting that they could be used to selectively target PC3M-like CSC populations in prostate cancer.

We posit that the partial-EMT associated genes and metabolic reprogramming identified in the PC3M-PC3S cellular model are two complementary components that lead to the emergence of a CSC-like phenotype in prostate cancer. Hence, we hypothesize that combining metabolic inhibitors against the identified targets (e.g., antifolates) with blocking antibodies or chemical inhibitors targeting the partial-EMT associated CSC-drivers could be a highly effective therapy against the CSC population in androgen-independent prostate cancer.

5. Methods

5.1. Transcriptomics analysis

5.1.1. RNA-Seq analysis

RNA-Seq analysis was performed on the cell lines PC3M-SNAI1, PC3M and PC3S, with both PC3M and PC3S stably transfected with an empty pBABE vector[15]. PC3M and PC3S and PC3M-SNAI1 cells were grown as a monolayer culture in RPMI-1640 medium supplemented with 2 mM L-glutamine, 10% fetal bovine serum (FBS) and 0.1% streptomycin/penicillin, under standard culture conditions (humidified air with 5% CO_2 at 37°C).

RNA was purified using the RNeasy Mini Kit following the manufacturer instructions. Total RNA was assayed for quantity and quality using Qubit® RNA HS Assay (Life Technologies) and RNA 6000 Nano Assay on a Bioanalyzer 2100.

From total RNA the RNASeq libraries were prepared using the TruSeq™ RNA Sample Prep Kit v2 (Illumina Inc.,) with minor modifications. Briefly, 0.5µg of total RNA was used as the input material for poly-A based mRNA enrichment with oligo-dT magnetic beads. After fragmentation the first and second strand cDNA were synthetized. The double stranded cDNA was end-repaired, 3´adenylated and the 3´-"T" nucleotide of the adapter was used for the Illumina indexed adapters ligation. The ligation product was enriched by 15 cycles of PCR.

Each library was sequenced using TruSeq SBS Kit v3-HS (Illumina, Inc), in paired-end mode with a read length of 2x76bp. Over 40 million paired-end reads for each sample were generated in a fraction of a sequencing lane on HiSeq2000 following the manufacturer's protocol. Images analysis, base calling and quality scoring of the run were processed using the manufacturer's

software Real Time Analysis (RTA 1.13.48, HCS 1.5.15.1) and followed by generation of FASTQ sequence files by CASAVA.

Quality control was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). STAR(Spliced Transcripts Alignment to a Reference)[199] was used to map to sequences to the GRCh37 assembly from Ensembl (Homo_sapiens.GRCh37.75.dna.primary_assembly.fa.gz) and obtain the number of reads per gene.

5.1.2. Differential expression analysis

Gene differential expression analysis was performed using the DESeq2 (differential gene expression analysis based on the negative binomial distribution)[200] package for R. To allow intergene comparisons, gene counts were converted to FPKM (Fragments Per Kilobase of transcript per Million mapped reads) using DESeq2 fpkm function.

5.1.3. Gene set enrichment analysis (GSEA)

GSEA was performed with the GAGE package for R[200]. The values for the Wald statistic computed by DESEQ2 for each gene were used as input for GAGE. This statistic depends on the fold change, variance and number of reads for each gene and hence it allows giving higher weight to increases or decreases in genes with low variance and a high number of reads[200].

A total of 2933 GS were evaluated to identify the gene expression programs significantly upregulated or downregulated between the PC3M and PC3S and PC3M-SNAI1 cellular models. They were obtained from gene ontology (GO) biological process annotations[201], KEGG[202] pathways annotations, and the Molecular Signatures Database[203]. From the later, oncogenic signatures (i.e., genes that are upregulated when known oncogene are perturbed) and hallmarks sets (curated sets that represent a given phenotype or feature) were evaluated[204].

To evaluate the capacity of Snai1 overexpression to revert the PC3M phenotype, only the GSs significantly enriched in PC3M vs PC3S were tested.

5.2. Copy number variation analysis

Copy number variation and survival analysis of the genes of interest was evaluated using the cBioPortal for Cancer Genomics[205]. Copy number variation was analyzed in 4018 prostate cancer samples. Survival analysis was performed on the subset of the aforementioned samples where patients progress had been tracked (i.e., 1121 patients).

5.3. r^{2}MTA

r^{2}MTA, an enhanced version of the previously published rMTA algorithm[28], was developed and used to characterize the metabolic transition between PC3M and PC3S and PC3M and PC3M-Snai1 at the genome-scale. Furthermore, r^{2}MTA was also used to identify putative metabolic targets that could switch the PC3M (source state) metabolic phenotype to the PC3S (target state) phenotype thus leading to the loss of the metabolic phenotype associated to CSCs.

r^{2}MTA was run in the framework of a specific GSMM of the PC3M-PC3S cellular models using a starting flux distribution derived from a PC3M-specific GSMM reconstruction (See "GSMM reconstruction"). To model each transition (PC3M/PC3S and PC3M/PC3M-Snai1) a list of upregulated and downregulated reactions was provided as input (See Section "Determining upregulated and downregulated reactions").

5.3.1. Mapping gene expression to reactions

Transcript abundances (expressed as FPKM) were mapped to Recon2.2[206], a network of all known metabolic reactions in humans. Recon2.2 describes the association of metabolic reactions to genes though a set of Booleans expressions referred to as gene protein reaction (GPR) rules. Using the standard GPR notation, the Boolean operator AND is used to define enzymatic complexes, and the Boolean operator OR is used to describe isoenzymes[207].

To map transcripts abundances, first OR operators were replaced by "+" operators and AND operators by MIN() operators in the GPR expressions. Then, GPR expressions for each reaction were evaluated, replacing geneIDs by their respective transcript abundances. Under such system, a reaction catalyzed by multiple isoenzymes will have a gene expression value equal to the sum of the transcript abundances of the isoenzymes while reactions catalyzed by protein complexes will be mapped to the minimum transcript abundance of the complex's components.

5.3.2. GSMM reconstructions

Previously published rates of metabolite uptake and secretion[23] were integrated into Recon2.2 as flux boundaries for exchange reactions with the extracellular media. Then, the average transcript abundances for PC3M and PC3S were mapped to reactions and integrated through the iMAT algorithm to build PC3M- and PC3S-specific genome-scale reconstructions[208].

iMAT maximizes both the number of highly expressed reactions active (i.e., carrying flux) and the number of lowly express reactions inactive (i.e. not carrying flux)[208]. A reaction was classified as highly expressed if it was associated to a gene expression level above the 80th percentile for metabolic genes and as lowly expressed if it was associated to a gene expression level below the

20th percentile for metabolic genes. Reactions were considered active if they had an absolute flux equal or greater than 0.01 pmol·cell^{-1}·h^{-1}. As there can be multiple iMAT optimal solutions, we systematically forced each highly/lowly expressed reaction to be inactive/active to identify those active/inactive reactions necessary for achieving the optimal iMAT solution.

Furthermore, to constraint the maximum flux through reactions, the GIMME algorithm was applied to the resulting iMAT models. GIMME optimizes biomass production and then performs a second optimization where fluxes through reactions are minimized with a weight that is a function of the gene expression value mapped to each reaction[209,210]. In our analysis, the minimization weight (w_i) of each reaction was defined as follows:

$$w_i = 1 + \max(Th - ge_i, 0) \quad (1)$$

where,

ge_i is the gene expression value mapped to reaction *i*

Th is the gene expression threshold bellow which reactions are given additional minimization weight. In this analysis, it was set to the 80th percentile of the gene expression for metabolic genes.

Following the GIM^{3}E approach[210], each flux was maximized and minimized to identify the ranges of feasible fluxes within the optimal GIMME solution with a tolerance of 99.95%.

Finally, a consensus model for the PC3M-PC3S cell lines was constructed by combining the flux boundaries for both the PC3M and PC3S specific GSMMs as follows:

$$ub_{PC3M-PC3S} = max(ub_{PC3M}, ub_{PC3S}, 0) \quad (2)$$

$$lb_{PC3M-PC3S} = min(lb_{PC3M}, lb_{PC3S}, 0) \quad (3)$$

Where,

ub_{PC3M}, ub_{PC3S}, $ub_{PC3M-PC3S}$ are the upper flux boundaries for the PC3M-specific model, PC3S-specific model, and PC3M-PC3S consensus model, respectively.

lb_{PC3M}, lb_{PC3S}, $lb_{PC3M-PC3S}$ are the lower flux boundaries for the PC3M-specific model, PC3S-specific model, and PC3M-PC3S consensus model, respectively.

5.3.3. Determining upregulated and downregulated reactions

One of the inputs of r^{2}MTA is a list of upregulated/downregulated reactions between the source state and target state computed from transcriptomics and metabolomics measurements.

Transcripts abundance (FPKM) for each replicate were mapped to reactions, and a T-test was used to determine statistical significance of the differences between reaction expression values across conditions. Reactions were considered to be differentially expressed if they had a fold change equal or greater to 0.25 and an FDR<0.05. Additionally, reactions catalyzed by gene products associated with more than 10 reactions were excluded.

Finally, the exchange reactions of metabolites produced/consumed at statically significantly different rates between the conditions of study[23] were also added at the list of upregulated/ downregulated reactions.

In total, 746 and 255 differentially expressed reactions were identified between PC3M and PC3S and PC3M and PC3M-Snai1, respectively.

5.3.4. Reference flux distribution and ε

Additionally, r^{2}MTA also requires as input a reference flux distribution (V_{ref}) for the source state, PC3M in the case of study, and ε, a vector describing the minimum variation to consider a reaction to be upregulated or downregulated[211].

The solution space of the PC3M-specific GSMM was sampled using the Artificially Centered hit-and-run (ACHR) algorithm implemented into COBRA[212,213]. ACHR was run with a thinning factor of 10000. In total, 250 flux samples were obtained.

V^{ref} was defined as the average of the 250 samples for each reaction flux. ε was defined as half the width of the 99.9 % confidence interval for the mean of each flux.

5.3.5. Predicting inactive reaction as a result of a gene KO

To simulate the effect of gene KOs, gene expression was mapped to reactions with the expression of the target gene set to 0. Any reaction with its mapped gene expression value reduced 10th fold or more compared to the wild type was assumed to be inactive as the consequence of the gene KO. The reactions identified as inactive were then constrained to a flux value of 0.

5.3.6. Formulation r^{2}MTA

r^{2}MTA and rMTA are based on the MTA (Metabolic transformation algorithm)[211]. MTA is built around a quadratic Mixed-Integer Quadratic Programming (MIQP) problem with two components. In the original MTA formulation, the quadratic component (QP) consisted of minimizing the variation between the initial and final state in reactions not differentially

expressed. The mixed linear integer programming (MILP) component maximized the number of upregulated/downregulated reactions with fluxes increases/decreases of at least a value of ε. The weight of the two components is defined by a parameter α[211]. By solving this problem with parts of the network blocked to simulate gene KOs, the best targets to switch from the source state to the target state are identified. In the original MTA implementation, gene KOs were then ranked based on the following scoring function:

$$\mathrm{TS}_g = \frac{\sum_{i \in Rssucces} \left| v_i^{ref} - v_i^{MTA} \right| - \sum_{i \in Runsuccess} \left| v_i^{ref} - v_i^{MTA} \right|}{\sum_{i \in Rs} \left| v_i^{ref} - v_i^{MTA} \right|} (4)$$

Where,

TS is the MTA score for the KO of gene *g.*

v^{ref} is the reference flux distribution (flux distribution of the source state).

v^{MTA} is the flux distribution resulting from the MTA optimization with the KO of gene g.

$Rssucces$ are the differentially expressed reactions that have changed in the desired direction (i.e., towards the target state) in v_i^{MTA}.

$Runsuccess$are the differentially expressed reactions that have not changed in the desired direction (i.e., they have shifted away from the target state).

Rs are the reactions that are not differentially expressed between the source and target state and hence should (ideally) remain unchanged in the MTA optimization.

MTA was further refined by Valcárcel *et al.*[28] who increased the robustness of the technique by also considering the worst-case scenario and the Minimization of Metabolic Adjustment (MOMA) to score the potential targets[28]. The worst-case scenario consists of running MTA inverting the list of upregulated/downregulated genes and allows to determine the capacity to move “away” from the target state with a given gene KO. MOMA is a method to simulate the effect of a gene KO assuming that when subjected to gene(s) KO, metabolic fluxes will switch to a state that requires the minimal variation from the flux distribution before the perturbation[214].

By integrating both MOMA and the worst-case scenario in the scoring functions, rMTA can more robustly identify the targets that have higher propensity to facilitate the transition to the target state.

Nevertheless, a limitation of both MTA and rMTA is that there can potentially be more than one optimal solution to the MTA optimization problem. Hence, the resulting MTA or rMTA score

might be dependent on the final solution selected by the MIPQP solver. The multiplicity of solutions arises from the MILP component of the problem as any upregulated/downregulated reactions with a variation above/bellow ε will contribute equally to the objective function regardless of their flux value. This is particularly noticeable in undetermined reactions that are part of cycles or loops which can potentially increase/decrease to infinite values if not restricted. Here we propose to overcome this limitation by two complementary approaches. Firstly, by computing v^{ref} and ε from a model that has been constrained with GIMME to minimize the presence of cycles and loops in the reference flux distribution. And secondly, by modifying the MTA optimization to also minimize the variation of differentially expressed reactions. This is formulated with the following MIQP:

$$\min\ (1-\alpha)\sum\left(v_i^{ref}-v_i^{MTA}\right)^2+\frac{\alpha}{2}\sum_{i\in Rf}y_i+\frac{\alpha}{2}\sum_{i\in Rb}y_i\ (5)$$

Subject to:

$$s\cdot v_i^{MTA}=0\ (6)$$

$$lb<v^{MTA}<ub\ (7)$$

$$v_i^{MTA}-y_i^F\cdot\left(v_i^{ref}+\varepsilon_i\right)-y_i\cdot lb_i\geq 0\ ,i\in Rf(8)$$

$$y_i^F+y_i=1,i\ \in Rf\ (9)$$

$$v_i^{MTA}-y_i^B\cdot\left(v_i^{ref}-\varepsilon_i\right)-y_i\cdot ub_i\leq 0\ ,i\in Rb(10)$$

$$y_i^B+y_i=1,i\ \in Rb\ (11)$$

$$y_i^F,y_i^B,y_i\in\{0,1\}\ (12)$$

Where,

Eq (5) is the optimization function where α is the weight given to the MILP component and $\alpha-1$ the weight given to the QP component of the optimization and $y_i$is a binary variable that is 1 when upregulated/downregulated reactions are not changing in the target direction.Rf and Rb are the reactions that should increase and decrease, respectively, to shift in the desired direction (i.e. away from the source state and towards the target state). Unlike previous implementations [28,211], upregulated/downregulated reactions are not excluded from the QP component.

Eq (6) is the steady-state constraint and *s* is the stoichiometric matrix.

Eq (7) defines the lower and upper bounds for reaction fluxes. In our analysis, the bounds for the consensus model for the PC3M-PC3S cell lines were used ($ub_{PC3M-PC3S}$, $lb_{PC3M-PC3S}$). To simulate gene KO, reactions inactivated as a result of the gene KO are constrained to 0.

Eq (8) to Eq (12) define the criteria for a reaction to shift in desired direction, namely that $v_i^{MTA} - v_i^{ref} \geq \varepsilon_i$ for Rf and $v_i^{ref} - v_i^{MTA} \geq \varepsilon_i$ for Rb.

Adding *Rf* and *Rb* reactions to the QP component of the optimization enhances the robustness of MTA by reducing the space of optimal solutions towards a single solution. However, a potential pitfall of such approach is that it could overprioritize the QP ($(v_i^{ref} - v_i^{MTA})^2$) over the MILP ($v_i^{MTA} - v_i^{ref} \geq \varepsilon_i$ for Rf and $v_i^{ref} - v_i^{MTA} \geq \varepsilon_i$ for Rb) component if the appropriate α value is not chosen. In our analysis, owing to the preceding flux minimization, most ε_i were significantly below 1, and we determined that with α set to 0.5 it was enough to prioritize the MILP component. Other α values (0.25 and 0.5) were tested, and highly similar results were obtained, proving that the results were not overly sensitive to the chosen α value.

Finally, to score potential targets, we used a new scoring function, based on the one used in rMTA[28], but also considering the base TS. We define as base TS the TS obtained when MTA is run in the base network without any gene KO for either the best- or worst-case scenarios.

$$r^2TS_g = \min\left(TS_g^{MOMA}, 0\right) \cdot \left(\frac{TS_g^{MTAbc}}{TS_{base}^{MTAbc}} - \frac{TS_g^{MTAwc}}{TS_{base}^{MTAwc}}\right) \cdot 1000 \quad (13)$$

Where,

r^2TS_g is the r²MTA score for gene *g*.

$TS_g{}^{MOMA}$ is the TS when the KO of gene *g* is simulated with MOMA.

$TS_g{}^{MTAbc}$ and $TS_g{}^{MTAwc}$ are the TSs when the MTA optimization (eq 5) is run with gene *g* KOed, for the best-case (bc) and worst-case (wc) scenarios, respectively.

$TS_{base}{}^{MTAbc}$ and $TS_{base}{}^{MTAwc}$ are the TSs when the MTA optimization (eq 5) is run without any gene KO, for the best-case (bc) and worst-case (wc) scenarios, respectively.

By Integrating the base scores, we are both normalizing the transformation score and correcting for possible biases towards the best- or worst-case scenarios in the MTA optimization.

5.4. Cell proliferation assay

Methotrexate (M9929-25MG) and pemetrexed (CDS024404-50MG) were purchased at Sigma Aldrich (Saint Louis, MO, USA) and dissolved to 20 mM in water. For the assessment of cell viability,

cells were cultured in 96 well plates for 24 hours prior to the addition of different compounds at different concentrations, in triplicate. Cells were incubated in the presence of the compounds for 72 hours. Then, cell proliferation was assessed by the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay, based on the ability of alive cells to cleave the tetrazolium ring of the MTT, quantitatively producing formazan (λ=550 nm)[215].

6. References

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).

2. Logothetis, C., Morris, M. J., Den, R. & Coleman, R. E. Current perspectives on bone metastases in castrate-resistant prostate cancer. *Cancer Metastasis Rev.* **37**, 189–196 (2018).

3. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).

4. Kuşoğlu, A. & Biray Avcı, Ç. Cancer stem cells: A brief review of the current status. *Gene* **681**, 80–85 (2019).

5. Mayer, M. J., Klotz, L. H. & Venkateswaran, V. Metformin and prostate cancer stem cells: A novel therapeutic target. *Prostate Cancer Prostatic Dis.* **18**, 303–309 (2015).

6. Armstrong, C. M. & Gao, A. C. Drug resistance in castration resistant prostate cancer: resistance mechanisms and emerging treatment strategies. *Am. J. Clin. Exp. Urol.* **3**, 64–76 (2015).

7. Deng, Q. & Tang, D. G. Androgen receptor and prostate cancer stem cells: biological mechanisms and clinical implications. *Endocr. Relat. Cancer* **22**, T209-20 (2015).

8. Prieto-Vila, M., Takahashi, R.-U., Usuba, W., Kohama, I. & Ochiya, T. Drug Resistance Driven by Cancer Stem Cells and Their Niche. *Int. J. Mol. Sci.* **18**, (2017).

9. Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–15 (2008).

10. Hennessy, B. T. *et al.* Characterization of a Naturally Occurring Breast Cancer Subset Enriched in Epithelial-to-Mesenchymal Transition and Stem Cell Characteristics. *Cancer Res.* **69**, 4116 (2009).

11. Blick, T. *et al.* Epithelial Mesenchymal Transition Traits in Human Breast Cancer Cell Lines Parallel the CD44hi/CD24lo/- Stem Cell Phenotype in Human Breast Cancer. *J. Mammary Gland Biol. Neoplasia* **15**, 235–252 (2010).

12. Thomson, T. M. M. *et al.* Metabolic Plasticity and Epithelial-Mesenchymal Transition. *J. Clin. Med.* **8**, 967 (2019).

13. Saitoh, M. Involvement of partial EMT in cancer progression. *J. Biochem.* **164**, 257–264 (2018).

14. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–

674 (2011).

15. Celià-Terrassa, T. *et al.* Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *J. Clin. Invest.* **122**, 1849–68 (2012).

16. Sánchez-Cid, L. *et al.* MicroRNA-200, associated with metastatic breast cancer, promotes traits of mammary luminal progenitor cells. *Oncotarget* **8**, 83384–83406 (2017).

17. Ocaña, O. H. *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* **22**, 709–24 (2012).

18. Pattabiraman, D. R. *et al.* Activation of PKA leads to mesenchymal-to-epithelial transition and loss of tumor-initiating ability. *Science* **351**, aad3680 (2016).

19. Hugo, H. J. *et al.* Epithelial requirement for in vitro proliferation and xenograft growth and metastasis of MDA-MB-468 human breast cancer cells: oncogenic rather than tumor-suppressive role of E-cadherin. *Breast Cancer Res.* **19**, 86 (2017).

20. Beerling, E. *et al.* Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep.* **14**, 2281–2288 (2016).

21. Shibue, T. & Weinberg, R. A. EMT, CSCs, and drug resistance: The mechanistic link and clinical implications. *Nat. Rev. Clin. Oncol.* **14**, 611–629 (2017).

22. Camacho, L. *et al.* Acid ceramidase as a therapeutic target in metastatic prostate cancer. *J. Lipid Res.* **54**, 1207–1220 (2013).

23. Aguilar, E. *et al.* Metabolic Reprogramming and Dependencies Associated with Epithelial Cancer Stem Cells Independent of the Epithelial-Mesenchymal Transition Program. *Stem Cells* **34**, 1163–1176 (2016).

24. Marín de Mas, I. *et al.* Model-driven discovery of long-chain fatty acid metabolic reprogramming in heterogeneous prostate cancer cells. *PLoS Comput. Biol.* **14**, e1005914 (2018).

25. Intlekofer, A. M. & Finley, L. W. S. Metabolic signatures of cancer cells and stem cells. *Nat. Metab.* **1**, (2019).

26. Yizhak, K., Chaneton, B., Gottlieb, E. & Ruppin, E. Modeling cancer metabolism on a genome scale. *Mol. Syst. Biol.* **11**, 817 (2015).

27. Nilsson, A. & Nielsen, J. Genome scale metabolic modeling of cancer. *Metab. Eng.* **43**, 103–112 (2017).

28. Valcárcel, L. V, Torrano, V., Tobalina, L., Carracedo, A. & Planes, F. J. rMTA: Robust Metabolic Transformation Analysis. *Bioinformatics* (2019). doi:10.1093/bioinformatics/btz231

29. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 1–17 (2009).

30. Yang, A. *et al.* MYC Inhibition Depletes Cancer Stem-like Cells in Triple-Negative Breast Cancer. *Cancer Res.* **77**, 6641–6650 (2017).

31. Zhang, H.-L., Wang, P., Lu, M.-Z., Zhang, S.-D. & Zheng, L. c-Myc maintains the self-renewal and chemoresistance properties of colon cancer stem cells. *Oncol. Lett.* **17**, 4487–4493 (2019).

32. Yang, C. *et al.* Downregulation of cancer stem cell properties via mTOR signaling pathway inhibition by rapamycin in nasopharyngeal carcinoma. *Int. J. Oncol.* **47**, 909–17 (2015).

33. Bahmad, H. F. *et al.* The Akt/mTOR pathway in cancer stem/progenitor cells is a potential therapeutic target for glioblastoma and neuroblastoma. *Oncotarget* **9**, 33549–33561 (2018).

34. Deng, J. *et al.* Inhibition of PI3K/Akt/mTOR signaling pathway alleviates ovarian cancer chemoresistance through reversing epithelial-mesenchymal transition and decreasing cancer stem cell marker expression. *BMC Cancer* **19**, 618 (2019).

35. Kress, T. R., Sabò, A. & Amati, B. MYC: connecting selective transcriptional control to global RNA production. *Nat. Rev. Cancer* **15**, 593–607 (2015).

36. Iadevaia, V., Zhang, Z., Jan, E. & Proud, C. G. mTOR signaling regulates the processing of pre-rRNA in human cells. *Nucleic Acids Res.* **40**, 2527–39 (2012).

37. Morita, M. *et al.* mTOR coordinates protein synthesis, mitochondrial activity and proliferation. *Cell Cycle* **14**, 473–80 (2015).

38. Farrell, A. S. & Sears, R. C. MYC degradation. *Cold Spring Harb. Perspect. Med.* **4**, 1–15 (2014).

39. Saxton, R. A. & Sabatini, D. M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **168**, 960–976 (2017).

40. Agrawal, P., Reynolds, J., Chew, S., Lamba, D. A. & Hughes, R. E. DEPTOR is a stemness factor that regulates pluripotency of embryonic stem cells. *J. Biol. Chem.* **289**, 31818–26 (2014).

41. Catena, V. & Fanciulli, M. Deptor: Not only a mTOR inhibitor. *J. Exp. Clin. Cancer Res.* **36**, 1–9 (2017).

42. Parker, B. S., Rautela, J. & Hertzog, P. J. Antitumour actions of interferons: Implications for cancer therapy. *Nat. Rev. Cancer* **16**, 131–144 (2016).

43. Li, Y. F. *et al.* Low dose of interferon-α improves the clinical outcomes of docetaxel in patients with castration-resistant prostate cancer: A pilot study. *Oncology Letters* **7**, 125–130 (2014).

44. Ebrahimi, S. *et al.* Interferon-Mediated Tumor Resistance to Oncolytic Virotherapy. *J. Cell. Biochem.* **118**, 1994–1999 (2017).

45. Ghaleb, A. M. & Yang, V. W. Krüppel-like factor 4 (KLF4): What we currently know. *Gene* **611**, 27–37 (2017).

46. Li, R. *et al.* A mesenchymal-to-Epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51–63 (2010).

47. Tiwari, N. *et al.* Klf4 Is a Transcriptional Regulator of Genes Critical for EMT, Including Jnk1 (Mapk8). *PLoS One* **8**, (2013).

48. Li, X. *et al.* Klf4 reduces stemness phenotype, triggers mesenchymal-epithelial transition (MET)-like molecular changes, and prevents tumor progression in nasopharygeal carcinoma. *Oncotarget* **8**, 93924–93941 (2017).

49. Liu, Y.-N. *et al.* Critical and Reciprocal Regulation of KLF4 and SLUG in Transforming Growth Factor -Initiated Prostate Cancer Epithelial-Mesenchymal Transition. *Mol. Cell. Biol.* **32**, 941–953 (2012).

50. Yori, J. L. *et al.* Krüppel-like Factor 4 Inhibits Tumorigenic Progression and Metastasis in a Mouse Model of Breast Cancer. *Neoplasia* **13**, 601-IN5 (2015).

51. Li, Q. *et al.* Suppression of epithelial-mesenchymal transition in hepatocellular carcinoma cells by Krüppel-like factor 4. *Oncotarget* **7**, 29749–60 (2016).

52. Bruyneel, E. *et al.* The Transcription Factor Snail Induces Tumor Cell Invasion through Modulation of the Epithelial Cell Differentiation Program. *Cancer Res.* **65**, 6237–6244 (2005).

53. Bastea, L. I., Döppler, H., Balogun, B. & Storz, P. Protein kinase D1 maintains the epithelial phenotype by inducing a DNA-bound, inactive SNAI1 transcriptional repressor complex. *PLoS One* **7**, (2012).

54. Du, C., Zhang, C., Hassan, S., Biswas, M. H. U. & Balaji, K. C. Protein kinase D1 suppresses epithelial-to-mesenchymal transition through phosphorylation of snail. *Cancer Res.* **70**, 7810–7819 (2010).

55. Zheng, H. *et al.* PKD1 phosphorylation-dependent degradation of SNAIL by SCF-FBXO11 regulates epithelial-mesenchymal transition and metastasis. *Cancer Cell* **26**, 358–373 (2014).

56. Du, C., Jaggi, M., Zhang, C. & Balaji, K. C. Protein kinase D1-mediated phosphorylation and subcellular localization of β-catenin. *Cancer Res.* **69**, 1117–1124 (2009).

57. Ganju, A. *et al.* Protein kinase D1 regulates subcellular localisation and metastatic function of metastasis-associated protein 1. *Br. J. Cancer* **118**, 587–599 (2018).

58. Storz, P. Protein kinase D1: Gatekeeper of the epithelial phenotype and key regulator of cancer metastasis? *Br. J. Cancer* **118**, 459–461 (2018).

59. Zhang, K. *et al.* Lats2 kinase potentiates Snail1 activity by promoting nuclear retention upon phosphorylation. *EMBO J.* **31**, 29–43 (2012).

60. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54 (2003).

61. Li, J., Ding, Y. & Li, A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World J. Surg. Oncol.* **14**, 1–5 (2016).

62. Stanbrough, M. *et al.* Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res.* **66**, 2815–2825 (2006).

63. Ao, R., Guan, L., Wang, Y. & Wang, J. N. Silencing of COL1A2, COL6A3, and THBS2 inhibits gastric cancer cell proliferation, migration, and invasion while promoting apoptosis through the PI3k-Akt signaling pathway. *J. Cell. Biochem.* **119**, 4420–4434 (2018).

64. Monard, D. SERPINE2/Protease Nexin-1 in vivo multiple functions: Does the puzzle make sense? *Semin. Cell Dev. Biol.* **62**, 160–169 (2017).

65. Vaillant, C. *et al.* Serpine2/PN-1 is required for proliferative expansion of pre-neoplastic lesions and malignant progression to medulloblastoma. *PLoS One* **10**, 1–22 (2015).

66. Shen, Y., Wang, X., Xu, J. & Lu, L. SerpinE2, a poor biomarker of endometrial cancer, promotes the proliferation and mobility of EC cells. *Cancer Biomarkers* **19**, 271–278 (2017).

67. Wang, K. *et al.* Prognostic significance of SERPINE2 in gastric cancer and its biological

function in SGC7901 cells. *J. Cancer Res. Clin. Oncol.* **141**, 805–812 (2015).

68. Buchholz, M. *et al.* SERPINE2 (protease nexin I) promotes extracellular matrix production and local invasion of pancreatic tumors in vivo. *Cancer Res.* **63**, 4945–4951 (2003).

69. Wu, Q. W. Serpine2, a potential novel target for combating melanoma metastasis. *Am. J. Transl. Res.* **8**, 1985–97 (2016).

70. Bergeron, S. *et al.* The serine protease inhibitor serpinE2 is a novel target of ERK signaling involved in human colorectal tumorigenesis. *Mol. Cancer* **9**, 1–15 (2010).

71. Mao, M. & Wang, W. SerpinE2 promotes multiple cell proliferation and drug resistance in osteosarcoma. *Mol. Med. Rep.* **14**, 881–887 (2016).

72. Wu, J.-I. & Wang, L.-H. Emerging roles of gap junction proteins connexins in cancer metastasis, chemoresistance and clinical application. *J. Biomed. Sci.* **26**, 1–14 (2019).

73. Stoletov, K. *et al.* Role of connexins in metastatic breast cancer and melanoma brain colonization. *J. Cell Sci.* **126**, 904–913 (2013).

74. Oliveira, R. *et al.* Contribution of gap junctional communication between tumor cells and astroglia to the invasion of the brain parenchyma by human glioblastomas. *BMC Cell Biol.* **6**, 1–17 (2005).

75. Behrens, J., Kameritsch, P., Wallner, S., Pohl, U. & Pogoda, K. The carboxyl tail of Cx43 augments p38 mediated cell migration in a gap junction-independent manner. *Eur. J. Cell Biol.* **89**, 828–838 (2010).

76. Ryszawy, D. *et al.* Functional links between Snail-1 and CX43 account for the recruitment of CX43-positive cells into the invasive front of prostate cancer. *Carcinogenesis* **35**, 1920–1930 (2014).

77. Zhang, A. *et al.* Connexin 43 expression is associated with increased malignancy in prostate cancer cell lines and functions to promote migration. *Oncotarget* **6**, 11640–51 (2015).

78. Zheng, W., Jiang, C. & Li, R. Integrin and gene network analysis reveals that ITGA5 and ITGB1 are prognostic in non-small-cell lung cancer. *Onco. Targets. Ther.* **9**, 2317–27 (2016).

79. Sawada, K. *et al.* Loss of E-cadherin promotes ovarian cancer metastasis via α5-integrin, which is a therapeutic target. *Cancer Res.* **68**, 2329–2339 (2008).

80. Gong, C. *et al.* MIR 17 inhibits ovarian cancer cell peritoneal metastasis by targeting ITGA5 and ITGB1. *Oncol. Rep.* **36**, 2177–2183 (2016).

81. Ju, J. A. *et al.* Hypoxia Selectively Enhances Integrin α 5 β 1 Receptor Expression in Breast Cancer to Promote Metastasis . *Mol. Cancer Res.* **15**, 723–734 (2017).

82. Zhang, X. *et al.* MicroRNA-26a promotes anoikis in human hepatocellular carcinoma cells by targeting alpha5 integrin. *Oncotarget* **6**, 2277–89 (2015).

83. Feng, L., Ma, J., Ji, H., Liu, Y. & Hu, W. miR-330-5p suppresses glioblastoma cell proliferation and invasiveness through targeting ITGA5. *Biosci. Rep.* **37**, BSR20170019 (2017).

84. Ji, L. *et al.* Defective neuronal migration and inhibition of bipolar to multipolar transition of migrating neural cells by Mesoderm-Specific Transcript, Mest, in the developing mouse neocortex. *Neuroscience* **355**, 126–140 (2017).

85. Nakanishi, H. *et al.* Loss of imprinting of PEG1/MEST in lung cancer cell lines. *Oncol. Rep.* **12**, 1273–8 (2004).

86. Pedersen, I. S. *et al.* Frequent loss of imprinting of PEG1/MEST in invasive breast cancer. *Cancer Res.* **59**, 5449–51 (1999).

87. Boström, P. *et al.* MMP-1 expression has an independent prognostic value in breast cancer. *BMC Cancer* **11**, 1–8 (2011).

88. An, H. J. *et al.* The prognostic role of tissue and serum MMP-1 and TIMP-1 expression in patients with non-small cell lung cancer. *Pathol. Res. Pract.* **212**, 357–364 (2016).

89. Sunami, E. MMP-1 is a Prognostic Marker for Hematogenous Metastasis of Colorectal Cancer. *Oncologist* **5**, 108–114 (2004).

90. Ma, F. *et al.* MiR-361-5p inhibits glycolytic metabolism, proliferation and invasion of breast cancer by targeting FGFR1 and MMP-1. *J. Exp. Clin. Cancer Res.* **36**, 1–12 (2017).

91. Tang, M. li, Bai, X. jun, Li, Y., Dai, X. jing & Yang, F. MMP-1 Over-expression Promotes Malignancy and Stem-Like Properties of Human Osteosarcoma MG-63 Cells In Vitro. *Curr. Med. Sci.* **38**, 809–817 (2018).

92. Katkoori, V. R. *et al.* Prognostic significance and gene expression profiles of p53 mutations in microsatellite-stable Stage III colorectal adenocarcinomas. *PLoS One* **7**, (2012).

93. Loo, J. M. *et al.* Extracellular metabolic energetics can promote cancer progression. *Cell* **160**, 393–406 (2015).

94. Zeng, X.-T., Liu, X.-P., Liu, T.-Z. & Wang, X.-H. The clinical significance of COL5A2 in patients with bladder cancer. *Medicine (Baltimore).* **97**, e0091 (2018).

95. Cheon, D.-J. *et al.* A collagen-remodeling gene signature regulated by TGF-β signaling is associated with metastasis and poor survival in serous ovarian cancer. *Clin. Cancer Res.* **20**, 711–23 (2014).

96. Choi, D. *et al.* ORAI1 Activates Proliferation of Lymphatic Endothelial Cells in Response to Laminar Flow Through Krüppel-Like Factors 2 and 4. *Circ. Res.* **120**, 1426–1439 (2017).

97. Wang, C. A., Harrell, J. C., Iwanaga, R., Jedlicka, P. & Ford, H. L. Vascular endothelial growth factor C promotes breast cancer progression via a novel antioxidant mechanism that involves regulation of superoxide dismutase 3. *Breast Cancer Res.* **16**, 1–17 (2014).

98. Goel, H. L. *et al.* GLI1 regulates a novel neuropilin-2/α6β1 integrin based autocrine pathway that contributes to breast cancer initiation. *EMBO Mol. Med.* **5**, 488–508 (2013).

99. Elaimy, A. L. *et al.* VEGF-neuropilin-2 signaling promotes stem-like traits in breast cancer cells by TAZ-mediated repression of the Rac GAP β2-chimaerin. *Sci. Signal.* **11**, (2018).

100. Goel, H. L. *et al.* VEGF/neuropilin-2 regulation of Bmi-1 and consequent repression of IGF-IR define a novel mechanism of aggressive prostate cancer. *Cancer Discov.* **2**, 906–21 (2012).

101. Muders, M. H., Zhang, H., Wang, E., Tindall, D. J. & Datta, K. Vascular endothelial growth factor-C protects prostate cancer cells from oxidative stress by the activation of mammalian target of rapamycin complex-2 and AKT-1. *Cancer Res.* **69**, 6042–6048 (2009).

102. Cao, Y. *et al.* Neuropilin-2 promotes extravasation and metastasis by interacting with

endothelial α5 integrin. *Cancer Res.* **73**, 4579–4590 (2013).

103. Goel, H. L., Pursell, B., Standley, C., Fogarty, K. & Mercurio, A. M. Neuropilin-2 regulates 6 1 integrin in the formation of focal adhesions and signaling. *J. Cell Sci.* **125**, 497–506 (2012).

104. Tacconi, C. *et al.* Vascular endothelial growth factor C disrupts the endothelial lymphatic barrier to promote colorectal cancer invasion. *Gastroenterology* **148**, 1438-1451.e8 (2015).

105. Skobe, M. *et al.* Induction of tumor lymphangiogenesis by VEGF-C promotes breast cancer metastasis. *Nat. Med.* **7**, 192–8 (2001).

106. Lund, A. W. *et al.* VEGF-C Promotes Immune Tolerance in B16 Melanomas and Cross-Presentation of Tumor Antigen by Lymph Node Lymphatics. *Cell Rep.* **1**, 191–199 (2012).

107. Eroglu, A., Ersoz, C., Karasoy, D. & Sak, S. Vascular endothelial growth factor (VEGF)-C, VEGF-D, VEGFR-3 and D2-40 expressions in primary breast cancer: Association with lymph node metastasis. *Adv. Clin. Exp. Med.* **26**, 245–249 (2017).

108. Bishnupuri, K. S. *et al.* Reg IV activates the epidermal growth factor receptor/Akt/AP-1 signaling pathway in colon adenocarcinomas. *Gastroenterology* **130**, 137–149 (2006).

109. Vanderlaag, K. *et al.* Regenerating islet-derived family member, 4 modulates multiple receptor tyrosine kinases and mediators of drug resistance in cancer. *Int. J. Cancer* **130**, 1251–1263 (2012).

110. Bishnupuri, K. S. *et al.* Reg4-induced mitogenesis involves Akt-GSK3β-β-Catenin-TCF-4 signaling in human colorectal cancer. *Mol. Carcinog.* **53**, 169–180 (2014).

111. Agaësse, G., Barbollat-Boutrand, L., El Kharbili, M., Berthier-Vergnes, O. & Masse, I. P53 targets TSPAN8 to prevent invasion in melanoma cells. *Oncogenesis* **6**, 1–7 (2017).

112. Berthier-Vergnes, O. *et al.* Gene expression profiles of human melanoma cells with different invasive potential reveal TSPAN8 as a novel mediator of invasion. *Br. J. Cancer* **104**, 155–165 (2011).

113. Yue, S., Mu, W., Erb, U. & Zöller, M. The tetraspanins CD151 and Tspan8 are essential exosome components for the crosstalk between cancer initiating cells and their surrounding. *Oncotarget* **6**, (2015).

114. Wang, Z., Sun, H., Provaznik, J., Hackert, T. & Zöller, M. Pancreatic cancer-initiating cell exosome message transfer into noncancer-initiating cells: the importance of CD44v6 in reprogramming. *J. Exp. Clin. Cancer Res.* **38**, 1–20 (2019).

115. Elhassan, M. O., Christie, J. & Duxbury, M. S. Homo sapiens Systemic RNA Interference-defective-1 Transmembrane Family Member 1 (SIDT1) Protein Mediates Contact-dependent Small RNA Transfer and MicroRNA-21-driven Chemoresistance. *J. Biol. Chem.* **287**, 5267–5277 (2012).

116. Kuhn, S. *et al.* A Complex of EpCAM, Claudin-7, CD44 Variant Isoforms, and Tetraspanins Promotes Colorectal Cancer Progression. *Mol. Cancer Res.* **5**, 553–567 (2007).

117. Greco, C. *et al.* E-cadherin/p120-catenin and tetraspanin Co-029 cooperate for cell motility control in human colon carcinoma. *Cancer Res.* **70**, 7674–7683 (2010).

118. Francian, A., Mann, K. & Kullberg, M. Complement C3-dependent uptake of targeted liposomes into human macrophages, B cells, dendritic cells, neutrophils, and MDSCs. *Int.*

J. Nanomedicine **12**, 5149–5161 (2017).

119. Gates, K. C., Goetzmann, L. N., Cantlon, J. D., Jeckel, K. M. & Anthony, R. V. Effect of proline rich 15-deficiency on trophoblast viability and survival. *PLoS One* **12**, 1–21 (2017).

120. Meunier, D. *et al.* Expression analysis of proline rich 15 (Prr15) in mouse and human gastrointestinal tumors. *Mol. Carcinog.* **50**, 8–15 (2011).

121. Liu, J. J. *et al.* Scinderin promotes the invasion and metastasis of gastric cancer cells and predicts the outcome of patients. *Cancer Lett.* **376**, 110–117 (2016).

122. Lai, X. *et al.* Loss of scinderin decreased expression of epidermal growth factor receptor and promoted apoptosis of castration-resistant prostate cancer cells. *FEBS Open Bio* **8**, 743–750 (2018).

123. Wang, D. *et al.* Suppression of SCIN inhibits human prostate cancer cell proliferation and induces G0/G1 phase arrest. *Int. J. Oncol.* **44**, 161–166 (2014).

124. Liu, H., Shi, D., Liu, T., Yu, Z. & Zhou, C. Lentivirus-mediated silencing of SCIN inhibits proliferation of human lung carcinoma cells. *Gene* **554**, 32–39 (2015).

125. Jian, W. *et al.* Scinderin-knockdown inhibits proliferation and promotes apoptosis in human breast carcinoma cells. *Oncology Letters* **16**, 3207–3214 (2018).

126. Qiao, X. *et al.* Scinderin is a novel transcriptional target of BRMS1 involved in regulation of hepatocellular carcinoma cell apoptosis. *Am. J. Cancer Res.* **8**, 1008–1018 (2018).

127. Miura, N. *et al.* Adseverin: A novel cisplatin-resistant marker in the human bladder cancer cell line HT1376 identified by quantitative proteomic analysis. *Mol. Oncol.* **6**, 311–322 (2012).

128. Sousa, H. *et al.* 5'UTR +24T>C CR2 is not associated with nasopharyngeal carcinoma development in the North Region of Portugal. *Oral Dis.* **22**, 280–284 (2016).

129. Fan, Q. *et al.* Functional polymorphism in the 5'-UTR of CR2 is associated with susceptibility to nasopharyngeal carcinoma. *Oncol. Rep.* **30**, 11–16 (2013).

130. Méndez-Acevedo, K. M., Valdes, V. J., Asanov, A. & Vaca, L. A novel family of mammalian transmembrane proteins involved in cholesterol transport. *Sci. Rep.* **7**, 7450 (2017).

131. Eeles, R. A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat. Genet.* **45**, 385–91, 391e1-2 (2013).

132. Chapuy, B. *et al.* ABC transporter A3 facilitates lysosomal sequestration of imatinib and modulates susceptibility of chronic myeloid leukemia cell lines to this drug. *Haematologica* **94**, 1528–1536 (2009).

133. Cheong, N. *et al.* ABCA3 is critical for lamellar body biogenesis in vivo. *J. Biol. Chem.* **282**, 23811–23817 (2007).

134. Aung, T. *et al.* Exosomal evasion of humoral immunotherapy in aggressive B-cell lymphoma modulated by ATP-binding cassette transporter A3. *Proc. Natl. Acad. Sci.* **108**, 15336–15341 (2011).

135. Overbeck, T. R. *et al.* ABCA3 phenotype in non-small cell lung cancer indicates poor outcome. *Oncol.* **93**, 270–278 (2017).

136. Overbeck, T. R. *et al.* Intracellular ATP-binding cassette transporter A3 is expressed in lung cancer cells and modulates susceptibility to cisplatin and paclitaxel. *Oncol.* **84**, 362–

370 (2013).

137. Steinbach, D. *et al.* ABCA3 as a possible cause of drug resistance in childhood acute myeloid leukemia. *Clin. Cancer Res.* **12**, 4357–4363 (2006).

138. Wulf, G. G. *et al.* ABC transporter ABCA3 is expressed in acute myeloid leukemia blast cells and participates in vesicular transport. *Haematologica* **89**, 1395–7 (2004).

139. Hirschmann-Jax, C. *et al.* A distinct 'side population' of cells with high drug efflux capacity in human tumor cells. *Proc. Natl. Acad. Sci.* **101**, 14228–14233 (2004).

140. Kao, J. *et al.* Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* **4**, (2009).

141. Fan, Y. *et al.* Amplification and over-expression of MAP3K3 gene in human breast cancer promotes formation and survival of breast cancer cells. *J. Pathol.* **232**, 75–86 (2014).

142. S., G. *et al.* Colocalization of the tetraspanins, CO-029 and CD151, with integrins in human pancreatic adenocarcinoma: Impact on cell motility. *Clin. Cancer Res.* **11**, 2840–2852 (2005).

143. Park, C. S. *et al.* Therapeutic targeting of tetraspanin8 in epithelial ovarian cancer invasion and metastasis. *Oncogene* **35**, 4540–4548 (2016).

144. Pan, S. J. *et al.* Over-expression of tetraspanin 8 in malignant glioma regulates tumor cell progression. *Biochem. Biophys. Res. Commun.* **458**, 476–482 (2015).

145. Pan, S. J. *et al.* Tetraspanin 8-rictor-integrin α3 complex is required for glioma cell migration. *Int. J. Mol. Sci.* **16**, 5363–5374 (2015).

146. Zhou, Z. *et al.* TM4SF3 promotes esophageal carcinoma metastasis via upregulating ADAM12m expression. *Clin. Exp. Metastasis* **25**, 537–548 (2008).

147. Fang, T. *et al.* Tetraspanin-8 promotes hepatocellular carcinoma metastasis by increasing ADAM12m expression. *Oncotarget* **7**, 40630 (2016).

148. Cajigas-Du Ross, C. K. *et al.* RNA sequencing reveals upregulation of a transcriptomic program associated with stemness in metastatic prostate cancer cells selected for taxane resistance. *Oncotarget* **9**, 30363–30384 (2018).

149. El Kharbili, M. *et al.* Tspan8-β-catenin positive feedback loop promotes melanoma invasion. *Oncogene* **38**, 3781–3793 (2019).

150. Nazarenko, I. *et al.* Cell surface tetraspanin Tspan8 contributes to molecular pathways of exosome-induced endothelial cell activation. *Cancer Res.* **70**, 1668–1678 (2010).

151. Bonnet, M. *et al.* Targeting the Tetraspanins with Monoclonal Antibodies in Oncology: Focus on Tspan8/Co-029. *Cancers (Basel).* **11**, 179 (2019).

152. Oue, N. *et al.* Serum concentration of Reg IV in patients with colorectal cancer: Overexpression and high serum levels of Reg IV are associated with liver metastasis. *Oncology* **72**, 371–380 (2008).

153. Zhu, X. *et al.* Overexpression of Reg4, alone or combined with MMP-7 overexpression, is predictive of poor prognosis in colorectal cancer. *Oncol. Rep.* **33**, 320–328 (2015).

154. Wang, Q. *et al.* Oncogenic reg IV is a novel prognostic marker for glioma patient survival. *Diagn. Pathol.* **7**, 1 (2012).

155. Tao, H. Q. *et al.* Evaluation of REG4 for early diagnosis and prognosis of gastric cancer. *Hum. Pathol.* **42**, 1401–1409 (2011).

156. Wang, H. *et al.* REG4 promotes peritoneal metastasis of gastric cancer through GPR37. *Oncotarget* **7**, (2016).

157. Zhang, N. *et al.* Expression of Reg IV and SOX9 and their correlation in human gastric cancer. *BMC Cancer* **18**, 1–11 (2018).

158. Ohara, S. *et al.* Reg IV is an independent prognostic factor for relapse in patients with clinically localized prostate cancer. *Cancer Sci.* **99**, 1570–1577 (2008).

159. Chen, S. *et al.* The role of the REG4 gene and its encoding product in ovarian epithelial carcinoma. *BMC Cancer* **15**, 471 (2015).

160. Saukkonen, K. *et al.* Prognostic and diagnostic value of REG4 serum and tissue expression in pancreatic ductal adenocarcinoma. *Tumor Biol.* **40**, 1–10 (2018).

161. Kobunai, T., Watanabe, T. & Fukusato, T. REG4, NEIL2, and BIRC5 gene expression correlates with gamma-radiation sensitivity in patients with rectal cancer receiving radiotherapy. *Anticancer Res.* **31**, 4147–53 (2011).

162. He, H. L. *et al.* Overexpression of REG4 confers an independent negative prognosticator in rectal cancers receiving concurrent chemoradiotherapy. *J. Surg. Oncol.* **110**, 1002–1010 (2014).

163. Violette, S. *et al.* Reg IV, a new member of the regenerating gene family, is overexpressed in colorectal carcinomas. *Int. J. Cancer* **103**, 185–193 (2003).

164. Nakagawa, H. *et al.* Serum REG4 level is a predictive biomarker for the response to preoperative chemoradiotherapy in patients with pancreatic cancer. *Pancreas* **38**, 791–798 (2009).

165. Jin, J. *et al.* Regenerating Family Member 4 (Reg4) Enhances 5-Fluorouracil Resistance of Gastric Cancer Through Activating MAPK/Erk/Bim Signaling Pathway. *Med. Sci. Monit.* **23**, 3715–3721 (2017).

166. Ying, L. S., Yu, J. L., Lu, X. X. & Ling, Z. Q. Enhanced RegIV expression predicts the intrinsic 5-fluorouracil (5-FU) resistance in advanced gastric cancer. *Dig. Dis. Sci.* **58**, 414–422 (2013).

167. Katsuno, Y. *et al.* Coordinated expression of REG4 and aldehyde dehydrogenase 1 regulating tumourigenic capacity of diffuse-type gastric carcinoma-initiating cells is inhibited by TGF-β. *J. Pathol.* **228**, 391–404 (2012).

168. Z., G. *et al.* Reg IV: A promising marker of hormone refractory metastatic prostate cancer. *Clin. Cancer Res.* **11**, 2237–2243 (2005).

169. Shen, K., Yu, W., Yu, Y., Liu, X. & Cui, X. Knockdown of TMEM45B inhibits cell proliferation and invasion in gastric cancer. *Biomed. Pharmacother.* **104**, 576–581 (2018).

170. Li, Y. *et al.* Silencing Transmembrane Protein 45B (TNEM45B) Inhibits Proliferation, Invasion, and Tumorigenesis in Osteosarcoma Cells. *Oncol. Res.* **25**, 1021–1026 (2017).

171. Zhao, L. *et al.* TMEM45B promotes proliferation, invasion and migration and inhibits apoptosis in pancreatic cancer cells. *Mol. Biosyst.* **12**, 1860–1870 (2016).

172. Hu, R. *et al.* TMEM45B, up-regulated in human lung cancer, enhances tumorigenicity of lung cancer cells. *Tumour Biol.* **37**, 12181–12191 (2016).

173. Mezencev, R., Matyunina, L. V., Jabbari, N. & McDonald, J. F. Snail-induced epithelial-to-mesenchymal transition of MCF-7 breast cancer cells: systems analysis of molecular changes and their effect on radiation and drug sensitivity. *BMC Cancer* **16**, 236 (2016).

174. Dong, C. *et al.* Loss of FBP1 by snail-mediated repression provides metabolic advantages in basal-like breast cancer. *Cancer Cell* **23**, 316–331 (2013).

175. Kim, N. H. *et al.* Snail reprograms glucose metabolism by repressing phosphofructokinase PFKP allowing cancer cell survival under metabolic stress. *Nat. Commun.* **8**, 14374 (2017).

176. Liu, M. *et al.* Snail-Overexpression Induces Epithelial-mesenchymal Transition and Metabolic Reprogramming in Human Pancreatic Ductal Adenocarcinoma and Non-tumorigenic Ductal Cells. *J. Clin. Med.* **8**, 1–18 (2019).

177. Rae, C., Sey, C. H. & Mairs, R. J. Radiosensitization of Prostate Cancer Cells by 2-Deoxyglucose. *Madridge J. Oncog.* **2**, 30–34 (2018).

178. Pikman, Y. *et al.* Targeting MTHFD2 in acute myeloid leukemia. *J. Exp. Med.* **213**, 1285–1306 (2016).

179. Ben-Sahra, I., Hoxhaj, G., Ricoult, S. J. H., Asara, J. M. & Manning, B. D. mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle. *Science (80-.).* **351**, 728–733 (2016).

180. Nilsson, R. *et al.* Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nat. Commun.* **5**, 1–10 (2014).

181. Nishimura, T. *et al.* Cancer stem-like properties and gefitinib resistance are dependent on purine synthetic metabolism mediated by the mitochondrial enzyme MTHFD2. *Oncogene* **38**, 2464–2481 (2019).

182. Noguchi, K. *et al.* The mitochondrial one-carbon metabolic pathway is associated with patient survival in pancreatic cancer. *Oncol. Lett.* **16**, 1827–1834 (2018).

183. Lin, H. *et al.* MTHFD2 Overexpression Predicts Poor Prognosis in Renal Cell Carcinoma and is Associated with Cell Proliferation and Vimentin-Modulated Migration and Invasion. *Cell. Physiol. Biochem.* **51**, 991–1000 (2018).

184. Liu, W. *et al.* Reprogramming of proline and glutamine metabolism contributes to the proliferative and metabolic responses regulated by oncogenic transcription factor c-MYC. *Proc. Natl. Acad. Sci.* **109**, 8983–8988 (2012).

185. Ye, Y., Wu, Y. & Wang, J. Pyrroline-5-carboxylate reductase 1 promotes cell proliferation via inhibiting apoptosis in human malignant melanoma. *Cancer Manag. Res.* **10**, 6399–6407 (2018).

186. Cai, F. *et al.* Pyrroline-5-carboxylate reductase 1 promotes proliferation and inhibits apoptosis in non-small cell lung cancer. *Oncology Letters* **15**, 731–740 (2018).

187. Zeng, T. *et al.* Knockdown of PYCR1 inhibits cell proliferation and colony formation via cell cycle arrest and apoptosis in prostate cancer. *Med. Oncol.* **34**, 1–9 (2017).

188. Possemato, R. *et al.* Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**, 346–50 (2011).

189. Maxwell, S. A. & Davis, G. E. Differential gene expression in p53-mediated apoptosis-resistant vs. apoptosis-sensitive tumor cell lines. *Proc. Natl. Acad. Sci.* **97**, 13009–13014 (2000).

190. Hollinshead, K. E. R. *et al.* Oncogenic IDH1 Mutations Promote Enhanced Proline Synthesis through PYCR1 to Support the Maintenance of Mitochondrial Redox Homeostasis. *Cell Rep.* **22**, 3107–3114 (2018).

191. Zhang, W. C. *et al.* Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell* **148**, 259–72 (2012).

192. Kang, P. J. *et al.* Glycine decarboxylase regulates the maintenance and induction of pluripotency via metabolic control. *Metab. Eng.* **53**, 35–47 (2019).

193. Alptekin, A. *et al.* Glycine decarboxylase is a transcriptional target of MYCN required for neuroblastoma cell proliferation and tumorigenicity. *Oncogene* 1–17 (2019). doi:10.1038/s41388-019-0967-3

194. Walling, J. From methotrexate to pemetrexed and beyond. A review of the pharmacodynamic and clinical properties of antifolates. *Invest. New Drugs* **24**, 37–77 (2006).

195. Skvortsov, S., Skvortsova, I. I., Tang, D. G. & Dubrovska, A. Concise Review: Prostate Cancer Stem Cells: Current Understanding. *Stem Cells* **36**, 1457–1474 (2018).

196. Thakur, R., Trivedi, R., Rastogi, N., Singh, M. & Mishra, D. P. Inhibition of STAT3, FAK and Src mediated signaling reduces cancer stem cell load, tumorigenic potential and metastasis in breast cancer. *Sci. Rep.* **5**, 10194 (2015).

197. Begum, A. *et al.* The extracellular matrix and focal adhesion kinase signaling regulate cancer stem cell function in pancreatic ductal adenocarcinoma. *PLoS One* **12**, e0180181 (2017).

198. Kolev, V. N. *et al.* Inhibition of FAK kinase activity preferentially targets cancer stem cells. *Oncotarget* **8**, 51733–51747 (2017).

199. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

200. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

201. The Gene Ontology Consortium *et al.* Gene Ontology : tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2011).

202. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

203. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

204. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2016).

205. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).

206. Swainston, N. *et al.* Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* **12**, (2016).

207. de Mas, I. M. *et al.* Cancer cell metabolism as new targets for novel designed therapies. *Future Med. Chem.* **6**, 1791–1810 (2014).

208. Zur, H., Ruppin, E. & Shlomi, T. iMAT: An integrative metabolic analysis tool. *Bioinformatics* **26**, 3140–3142 (2010).

209. Becker, S. A. & Palsson, B. O. Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Comput. Biol.* **4**, e1000082 (2008).

210. Schmidt, B. J. *et al.* GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics* **29**, 2900–8 (2013).

211. Yizhak, K., Gabay, O., Cohen, H. & Ruppin, E. Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat. Commun.* **4**, 2632 (2013).

212. De Martino, D., Mori, M. & Parisi, V. Uniform sampling of steady states in metabolic networks: heterogeneous scales and rounding. *PLoS One* **10**, e0122670 (2015).

213. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).

214. Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci.* **99**, 15112–15117 (2002).

215. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **65**, 55–63 (1983).

216. Hobeika, A. C., Subramaniam, P. S. & Johnson, H. M. IFNα induces the expression of the cyclin-dependent kinase inhibitor p21 in human prostate cancer cells. *Oncogene* **14**, 1165–1170 (1997).

217. Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434**, 772–7 (2005).

218. Ning, S., Pagano, J. S. & Barber, G. N. IRF7: activation, regulation, modification and function. *Genes Immun.* **12**, 399–414 (2011).

219. Fung, K. Y. *et al.* Interferon-ε protects the female reproductive tract from viral and bacterial infection. *Science* **339**, 1088–92 (2013).

220. de Weerd, N. A., Samarajiwa, S. A. & Hertzog, P. J. Type I interferon receptors: biochemistry and biological functions. *J. Biol. Chem.* **282**, 20053–7 (2007).

221. Snijders, A. M. *et al.* An interferon signature identified by RNA-sequencing of mammary tissues varies across the estrous cycle and is predictive of metastasis-free survival. *Oncotarget* **5**, (2014).

222. Linsley, P. S., Speake, C., Whalen, E. & Chaussabel, D. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. *PLoS One* **9**, (2014).

223. Katlinski, K. V. *et al.* Inactivation of Interferon Receptor Promotes the Establishment of Immune Privileged Tumor Microenvironment. *Cancer Cell* **31**, 194–207 (2017).

224. Fukasawa, M. *et al.* Microarray analysis of promoter methylation in lung cancers. *J. Hum. Genet.* **51**, 368–374 (2006).

225. Zhao, Y. *et al.* Overexpression of Interferon Regulatory Factor 7 (IRF7) Reduces Bone Metastasis of Prostate Cancer Cells in Mice. *Oncol. Res. Featur. Preclin. Clin. Cancer Ther.* **25**, 511–522 (2016).

226. Ahtiainen, L. *et al.* Defects in innate immunity render breast cancer initiating cells permissive to oncolytic adenovirus. *PLoS One* **5**, (2010).

227. Castiello, L. *et al.* Disruption of IFN-I Signaling Promotes HER2/Neu Tumor Progression and Breast Cancer Stem Cells. *Cancer Immunol. Res.* **6**, 658–670 (2018).

228. Legrier, M.-E. E. *et al.* Activation of IFN/STAT1 signalling predicts response to chemotherapy in oestrogen receptor-negative breast cancer. *Br. J. Cancer* **114**, 177–87 (2016).

229. Lan, Q. *et al.* Type I interferon/IRF7 axis instigates chemotherapy-induced immunological dormancy in breast cancer. *Oncogene* **38**, 2814–2829 (2019).

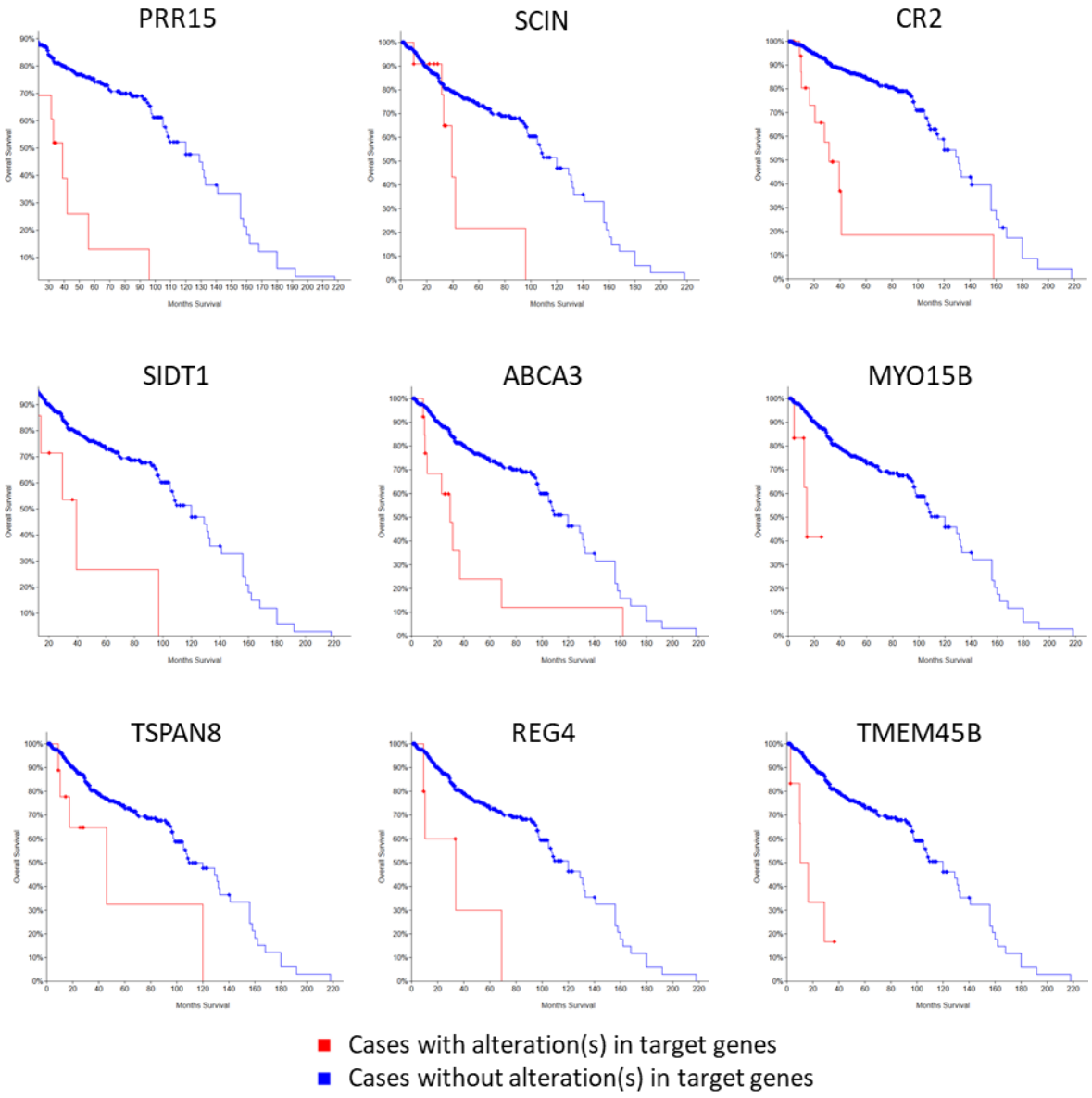


Figure S1: Kaplan-Meier plots for alterations in the putative targets associated to partial EMT. Such genes are frequently amplified in prostate cancer and such alterations are significantly associated to worst overall survival (Logrank Test P-Value < 0.05).

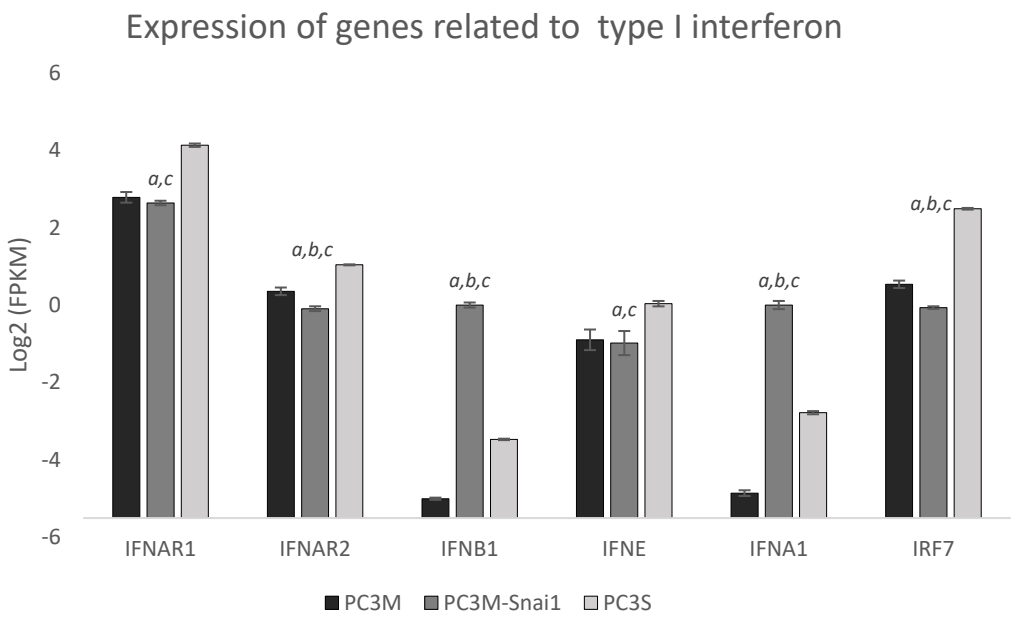
Text S1: Type I interferon Response

The GSs associated to type I interferon response (i.e., "GO: type I interferon signaling pathway," "GO: cellular response to type I interferon",) are strongly downregulated in PC3M compared to PC3S and, remarkably, such functions are not restored upon Snai1 overexpression. Type Interferon I signaling increases the immunogenic potential of cancer cells by enhancing the oncogenic-antigen presentation. In addition to the immunogenic role, type I interferon signaling has also been reported to inhibit proliferation in and trigger apoptosis in cancer cells[42]. For instance, interferon α has been shown to inhibit proliferation in a prostate cancer cell line by increasing the concentration of cell cycle inhibitor protein p21[216].

Although type I interferons are primarily produced by dendritic cells, they can be produced by most cell types, including cancer cells. Transcription of type I interferons is induced mainly by the IRF (Interferon Regulatory Factor) family of transcription factors, with IRF7 assumed to be master transcription factor[217,218]. The notable exception appears to be interferon ε, which appears to be regulated in a hormone-dependent manner[219].Type I interferons signaling acts through IFNAR(interferon alpha receptor) 1 and 2 heterodimers which are regulated both at the transcriptional and posttranscriptional level[220].

Low activity of type I interferon signaling is a hallmark of many cancer types and has been associated with poor prognosis and metastatic potential[221]. For instance, the loss of the interferon gene cluster was associated with poor prognosis in melanoma [222]. Similarly, in colorectal cancer, downregulation of IFNAR1 was shown to promote tumor growth and correlated with poor prognosis[223]. Likewise, IRF7 is often found to be hypermethylated in lung cancer cell lines[224] and its overexpression has been found to reduce metastasis in prostate cancer [225]. Furthermore, reduced type I interferon signaling has been associated with CSC phenotype in breast cancer[226]. Indeed, inactivating IFNAR1 in mice has been shown to enhance tumor progression and increase the proportion of CSC in breast tumors [227]. Type I interferon signaling also plays a key role in response to chemotherapy. In breast cancer, IFN-inducible genes were overexpressed in the xenografts that responded to therapy [228] while the activity of the IRF7/IFN-β/IFNAR axis was also found to correlate with longer distant metastasis-free survival after chemotherapy[229].

Reduced type I interferon signaling in PC3M appears to be driven by low expression of both IFNAR1 and IFNAR2 and reduced type interferon I production. It is worth noting, that the expression of the canonical type I interferons, IFNA1(Interferon α1) and IFNB (Interferon β1) is quite low, even in PC3S, and the predominant interferon appears to be IFNE (interferon ε).



Transcript levels of genes related to type I interferon. Expression levels are indicated as FPKM (Fragments Per Kilobase of transcript per Million mapped reads) in a Log2 scale. *a,b* and *c* denote a statistically significant difference (FDR<0.05) between PC3M and PC3S, PC3M and PC3M-Snai1 and PC3M-Snai1 and PC3S, respectively.

8.4 Chapter 4

Cysteine and folate metabolism are major vulnerabilities of the metastatic subpopulations of colorectal cancer

osep Tarrag -Celada, Carles oguet, ordi erarnau, avier ern ndez-Alias, Silvia Marin, M riam Tarrado-Castellarnau, edro de Atauri, Marta Cascante.

Cysteine and folate metabolism are major vulnerabilities of the metastatic subpopulations of colorectal cancer

Josep Tarragó-Celada[1*], Carles Foguet[1,2*], Jordi Perarnau[1], Xavier Hernández-Alias[1], Silvia Marin[1,2], Míriam Tarrado-Castellarnau[1,2], Pedro de Atauri[1,2], Marta Cascante[1,2]

1 Department of Biochemistry and Molecular Biomedicine & Institute of Biomedicine of University of Barcelona, Faculty of Biology, Universitat de Barcelona, Barcelona, Spain

2 Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD) and Metabolomics node at Spanish National Bioinformatics Institute (INB-ISCIII-ES-ELIXIR), Instituto de Salud Carlos III (ISCIII), Madrid, Spain

* both authors contributed equally to this work

Correspondence: martacascante@ub.edu

Summary

With most of the cancer-related deaths arising from the metastatic spread, it is of great biomedical importance to develop new therapeutic approaches against metastasis. In this regard, the acquisition of metastatic potential could be supported by a process of metabolic reprogramming that might result in metabolic vulnerabilities that can be therapeutic targets. With this aim, here we perform a thorough metabolic characterization of a primary colorectal cancer cell line (SW480), a cell line derived from the lymph node metastasis of the same patient (SW620) and a metastatically enriched derivative of the latter (SW620-LiM2). Using a novel genome-scale multiomics integration approach, we determined that the metastatic cell lines are selectively vulnerable to the inhibition of cystiine import and folate metabolism. Furthermore, we find that the metastatic cell lines had a reduced concentration of the antiproliferative peptide carnosine which can be putatively increased through the inhibition of citrate synthase. Together, the work presented here could contribute to the development of new therapies capable of selectively stopping the metastatic spread of colorectal cancer by targeting metabolic vulnerabilities of the metastatic populations.

Keywords

Colorectal cancer; Metastasis; Metabolism; Genome-scale

Introduction

Colorectal cancer is the third most common cancer type with an estimated 1.8 million new cases diagnosed yearly and the second cancer in mortality with over 800.000 deaths per year(Bray et al., 2018). Patients with early stages of colorectal cancer can be successfully treated by removing the primary tumor. Indeed, if diagnosed and surgically removed at an early stage the 5-year survival rate for patients is of approximately 90%(Kuipers et al., 2015). However, at late stages of tumor progression, cancer cells might have already disseminated from the primary site in the colon and extravasated into the blood or lymph vessels. While some of these cells might die in circulation, others may attach to distant organs and tissues and either remain dormant as micro-metastasis or proliferate and form clinically manifested metastasis(Oskarsson et al., 2014; Yamamoto et al., 2016). At that stage, colorectal cancer cannot be generally treated by surgery and with current therapies, the 5-year survival rate is less than 10%(Kuipers et al., 2015). Thus, it is of paramount importance to develop effective therapeutic strategies against colorectal cancer metastasis.

Recently, metabolic reprogramming has begun to emerge as an enabling feature of metastatic spread. Firstly, detachment from the extracellular matrix is associated with the increased formation of reactive oxygen species (ROS), which can lead to a specific form of apoptosis termed anoikis (Kamarajugadda et al., 2013; Li et al., 1999). In this regard, metastatic cell populations are able to survive in circulation by adapting metabolism to reduce ROS production and enhance antioxidant pathways(Elia et al., 2018; Lu, 2019). Secondly, metastatic populations are generally endowed with a degree of metabolic plasticity that allows them to survive and proliferate under the variable supply of substrates and oxygen encountered as part of the metastatic process(Lehuédé et al., 2016; Pascual et al., 2018). Finally, in order to form macroscopic metastasis from a small number of seeding cells, metastatic cell populations must be endowed with a metabolic phenotype capable of supplying the building blocks (e.g., nucleotides, amino acids, lipids) and ATP to support rapid cell proliferation(Tarrado-Castellarnau et al., 2016). However, the metabolic adaptation to promote metastasis can also lead to vulnerabilities and dependencies that can be exploited to selectively target metastatic cancer cells(Lu, 2019; Pascual et al., 2018).

With the goal of identifying metabolic vulnerabilities underlying the metastatic phenotype in colorectal cancer, here we performed a thorough metabolic characterization analysis of four cell lines derived from the same patient and with increasing metastatic potential: SW480, SW620, SW620-LiM1, and SW620-LiM2. SW480 and SW620 are $KRAS^{G12V}$-mutant cell lines derived from the primary tumor and a lymph node metastasis, respectively, from a 50-year-old patient with a Duke's type B colorectal cancer(Hewitt et al., 2000; Yeh et al., 2009) while LiM1 and LiM2 are two metastatic-enriched derivatives from SW620(Urosevic et al., 2014). Briefly, LiM1 was isolated from liver metastasis cells resulting from the inoculation of SW620 into the portal circulation of immunodeficient mice, and LiM2 from a second round of *in vivo* selection from LiM1. In this regard, after the initial characterization determined that phenotypically LiM1 and LiM2 were largely indistinguishable we decided to focus our efforts on SW480, SW620, and LiM2.

To integrate the generated data, together with publicly available data of the cell lines of study, we developed a novel genome-scale data integration workflow to build cell line-specific genome-scale metabolic models (GSMMs) from multiple layers of data (e.g. extracellular flux measurements, metabolomics, ^{13}C stable resolved metabolomics, respiration parameters and genetic dependencies). Furthermore, to identify targets with minimal off-target effect on non-tumoral colon cell populations, a condition-specific GSMMs was also built for the NCM460 cell line, derived from a non-tumoral colon mucosal epithelium(Moyer et al., 1996). The resulting cell line specific GSMMs were used to determine the metabolic vulnerabilities that can be used to selectively target metastatic cell populations in colorectal cancer.

Results

Characterization of the metastatic phenotype

We first performed a general characterization of the colorectal cancer cell lines in terms of cell growth, invasion, and metastatic capacities. The primary tumor-derived SW480 was the slowest-growing cell line, followed by the cell lines LiM2, LiM1 and SW620 (Figure 1A). Even more, cell volume had an indirect correlation with cell proliferation in line with the notion that smaller cells tend to proliferate faster than larger cells (Dolfi et al., 2013) (Figure 1B). Additionally, the metastatic cell lines had higher 3D growth capacity than SW480 (Figure 1C). Conversely, SW480 had a higher migratory capacity than the metastatic cell models (Figure 1D).

In this regard, SW480 had less expression of E-cadherin and γ-catenin than the metastatic cell lines, while they had more expression of β-catenin (Figure 1E). The higher motility together with

the loss of E-cadherin and γ-catenin, essential proteins for cell junctions in the epithelial lineage of the colon mucosa (Aktary et al., 2017; Schnoor, 2015), seem to indicate that SW480 had already undergone the epithelial-mesenchymal transition (EMT), which allows primary tumor cells to detach from the primary site (Mittal, 2018). Conversely, SW620 cells showed some epithelial properties (e.g., high E-cadherin expression and low expression of β-catenin and fibronectin). Although the LiM2 phenotype is similar to SW620, they also seemed to have more mesenchymal properties such as a small but significant ability to migrate and some mesenchymal markers enhanced with respect to SW620 such as fibronectin or vimentin.

Remarkably, the NF-κB factor, which plays a key role in cell migration, extracellular matrix degradation and EMT (Jana et al., 2017; Yan et al., 2010) was more expressed in the metastatic cell lines. However, the PI3K/AKT pathway, which can also promote EMT (Grille et al., 2003; Yang et al., 2019), was more active (i.e., higher fraction of P-AKT) in SW480 (Figure 1F). Finally, the MYC protein expression was higher in the metastatic cell lines (Figure 1G) and could also contribute to their increased proliferation rate and metastatic capacity (Cho et al., 2010; Yin et al., 2017).

Workflow for multiomics data integration

As part of the metabolic characterization of the cell lines under study, multiple layers of data were collected such as growth rates, rates of uptake and secretions of glucose, lactate, and amino acids, oxygen consumption rate, respiration parameters, ^{13}C resolved metabolomics, and targeted metabolomics. In this regard, due to the disparity of cell volume, flux measurements were normalized by cellular size to be comparable between cell lines (Dolfi et al., 2013). Additionally, transcriptomics for all cell lines of study were obtained from the literature (Provenzani et al., 2006; Sheffer et al., 2009; Urosevic et al., 2014). Similarly, from the Project DRIVE (deep RNAi interrogation of viability effects in cancer) database(McDonald et al., 2017) the dependencies of both SW620 and SW480 on a large number of metabolic genes were obtained. Together this created a large dataset of multiomics data that could be highly informative of the underlying flux phenotype.

Hence, we developed a workflow to integrate such complex data set and build cell-line specific GSMMs starting from the generic human GSMM Recon2.2(Swainston et al., 2016). The workflow, applied for SW480, SW620, and LiM2, consisted of the following steps:

1. Building a flux map of central metabolism. Stationary ^{13}C MFA(Antoniewicz, 2018) was used to integrate ^{13}C resolved metabolomics, growth rates, rates of metabolite uptake and secretion, oxygen consumption rate and respiration parameters to estimate the range of

fluxes through central metabolism. The resulting flux ranges were used to constrain the human GSMM model Recon2.2(Swainston et al., 2016).

2. Integration of targeted metabolomics. Measures of intracellular concentration of amino acid and biogenic amines were used to constrain the GSMMs to produce such metabolites at the rate required to maintain their levels at steady state while proliferating(Reimers and Reimers, 2016). Additionally, lipidomics measurements were used to customize the biomass reaction of each cell line.
3. Minimal cut set (MCS) analysis. Based on the essential metabolic genes reported in the project DRIVE database(McDonald et al., 2017) for SW620 and SW480, the reactions that should be inactive for such genes to be essential were determined using MCS analysis(Apaolaza et al., 2017, 2018). Such reactions were subsequently inactivated in the model. For LiM2, we implemented all the MCSs shared between SW480 and SW620 and the MCS in SW620 where none of the genes were significantly overexpressed.
4. Integration of transcriptomics data. The GIMME algorithm(Becker and Palsson, 2008; Schmidt et al., 2013) was used to integrate transcriptomics data and restrict the maximum flux through reactions and pathways based on their gene expression evidence.
5. Selection of the most representative flux samples. Flux sampling was used to compute possible flux combinations consistent with the above-integrated data. Each flux sample was then used to systematically simulate the effect of gene KOs on biomass production and ranked based on its consistency with the gene dependency data from Project DRIVE. The goal of this step is both to minimize the false positives of gene essentiality and to integrate partial dependencies on genes that while not-essential might have a significant impact on cell viability. For each cell line, the top 100 ranked flux samples were selected as representative flux distributions.

To build a specific GSMM for NCM460, a similar workflow was used but step 3 and 5 were omitted.

The metastatic cell lines display increased Warburg effect, pyruvate dehydrogenase activity, and glutaminolysis

We determined that the metastatic cell lines had a stronger Warburg effect than SW480, as they consumed more glucose and produced more lactate (Figure 2A-B). Remarkably, the metastatic cell lines also had less phosphorylation of Pyruvate Dehydrogenase (PDH), (Figure 2C) which suggested that they had this enzyme more active. This was confirmed from the cell line-specific flux maps, whiched show higher flux through PDH in SW620 and LiM2 than in SW480 (Figure

2G). In this regard, P-AKT, which was increased in SW480(Figure 1F), has been reported to phosphorylate PDK1 and enhance its activity (Chae et al., 2016) and could theoretically account for a decreased PDK activity in SW620 and LiM2. While it might seem paradoxical that the metastatic cell lines had both stronger Warburg effect and PDH activity, it must be noted that flux through PDH was quite low compared to lactate production ($\sim 10\%$, Figure 2C).

Regarding glutamine metabolism, we observed that the metastatic cells consumed more glutamine and produced more glutamate (Figure 2E-F). Glutaminase (GLS) and Glutamate Dehydrogenase 1 (GLUD1) expressions were also upregulated in the metastatic cell lines (Figure 2D), confirming a higher glutamine metabolism. Accordingly, condition-specific flux maps predicted increased glutaminase activity in the metastatic cell lines (Figure 2G).

The metastatic cell lines have increased respiration and respiratory flexibility

In line with the previously observed metabolic changes, mitochondrial function was also altered in the metastatic cell lines, which displayed an increased oxygen consumption rate (OCR) compared to SW480 (Figure 2H). We also performed Mito Fuel Assays using UK5099 (mitochondrial pyruvate carrier inhibitor), BPTES (GLS inhibitor) and etomoxir (CPT1 inhibitor) in order to characterize the capacity, dependency, and flexibility of mitochondrial respiration to glucose, glutamine and fatty acids. The metastatic cell lines showed similar capacity but lower dependency and higher flexibility for the three substrates (Figure 2I-K) than the SW480 cell line. Hence, even if the metastatic cell lines displayed increased PDH flux under basal conditions, they were not dependent on such activity for respiration.

Metabolomic profile

Intracellular Metabolite profiling of the cell lines was performed using the Absolute IDQ p180 kit (Biocrates Life Sciences AG) (Figure 3). Glutamate and glutamine were found to be significantly increased in the metastatic cell lines (SW620, LiM1, and LiM2) in accordance with their high rate of glutamine uptake and glutaminolysis. Similarly, increased glycine concentrations in the metastatic cell lines could be indicative of the increased activity of serine hydroxymethyltransferase and thus enhanced folate metabolism.

Conversely, the metastatic cell lines displayed a decreased concentration of the essential amino acids tryptophan, threonine, and tyrosine, which might be driven by increased protein synthesis and possibly increased catabolism of such amino acids. Furthermore, they also displayed significantly lower aspartate and carnosine levels. Carnosine is a naturally occurring dipeptide of β-alanine and histidine with putative roles as a pH buffer, neurotransmitter, and antioxidant agent which has been reported to inhibit cancer cell proliferation *in vivo* and *in vitro* (Bao et al.,

2018; Cheng et al., 2019; Ding et al., 2018b; Fouad et al., 2017; Iovine et al., 2014, 2016; Lee et al., 2018; Shen et al., 2014).

Additionally, acyl-carnitines and free carnitine were detected at significantly higher concentrations in SW480 than in the metastatic cell lines. As such metabolites are associated with β-oxidation(Longo et al., 2016), this confirms the dependency of SW480 on the β-oxidation determined in the fatty acid mitochondrial fuel assay (Figure 2K).

Remarkably, there were few biologically significant differences between SW620, LiM2, and LiM1. Indeed, when samples were clustered based on their metabolomics profile (Figure 3), replicates from LiM1 and LiM2 could not be separated.

Putative metabolic targets

The cell line-specific GSMMs were used to identify essential or synthetic lethal gene pairs that could selectively inhibit growth in the metastatic cell lines SW620 and LiM2. Overall, 26 single gene KO and 324 gene combinations were found to be able to impair the proliferation of SW620 and LiM2. Excluding the targets predicted to significantly impair the proliferation of normal colon epithelial cells (i.e., NCM460), 6 single putative gene targets and 9 putative synthetic lethal pairs were found to be selective for colorectal cancer cells (Table 1).

Most of the identified single targets were the direct result from the MCS analysis suggesting that such step significantly contributed to identifying the differential metabolic dependencies of the metastatic and non-metastatic cell lines. The exception is Pyruvate Kinase M1/2(PKM) which emerged as a target both because of the strong Warburg effect and the high expression of such isoenzyme in the cancer cell lines. Indeed, the key role of the PKM isoform 2 in cancer is widely recognized and its inhibition has emerged as a promising therapeutic strategy(Vander Heiden et al., 2010; Mazurek, 2011; Ning et al., 2017; Ye et al., 2012). The dependency of the cancer cell lines of study on PKM activity was validated with the Project DRIVE Database, which indicated that the silencing of such gene severely impaired the viability of both SW620 and SW480(McDonald et al., 2017). Such dependency was not integrated as part of MCS analysis as the predicted MCSs did not meet the required criteria (i.e. they required to inactivate more than the maximum allowed number of reactions, see Methods) and hence it emerged from the integration of the other layers of data. In NCM460, the model prediction is that they would be less susceptible to PKM inhibition because they had a lower ratio of PKM to Pyruvate Kinase L/R.

Concerning synthetic lethal pairs, we noted that many of them were associated with cystine transporters. Cystine (L-dicysteine) is the oxidized form of cysteine and the predominant form

in blood and culture media (Chawla et al., 1984; Crawhall et al., 1968). Cystine can be transported inside the cell through the Cystine/Glutamate antiporter system Xc- (coded by the genes SLC7A11 and SLC3A2)(Bridges et al., 2012) and through the cystine/neutral amino antiport acid system $b^{0,+}$ (coded by the genes SLC7A9 and SLC3A1)(Chillarón et al., 2010).

The remaining putative synthetic lethal genes were combinations with the glycolytic enzyme Glucose-6-Phosphate Isomerase (GPI) and most likely associated with the increased Warburg effect observed in the metastatic cell lines.

Metastatic cell lines are dependent on cystine uptake from the media

Close analysis revealed that both the dependency on Glutathione-Disulfide Reductase (GSR) and cystine transporters were connected. In the cytoplasm, cystine is reduced to cysteine primarily by reacting with glutathione(Bannai and Kitamura, 1980; States and Segal, 1973; Tietze et al., 1972), which is recycled by GSR. However, it has been reported that cystine can also be reduced by the thioredoxin reductase system(Holmgren, 1977; Mandal et al., 2010). Additionally, cysteine can also be produced from methionine, the other sulfur-containing amino acid, through the transsulfuration pathway(Sbodio et al., 2019) (Figure 4A). Furthermore, cysteine, in addition to being a proteogenic amino acid, is generally assumed to be the limiting substrate of glutathione synthesis(Estrela et al., 2016).

From the Project DRIVE gene screening(McDonald et al., 2017), it was predicted that silencing GSR had a significantly larger effect on the viability of SW620 than SW480. From MCS analysis it emerged that the most likely cause of such dependence was due to differences in cysteine metabolism. Namely, because of insufficient activity in both thioredoxin-dependent cystine reduction and cystathionine β-synthase in SW620, the glutathione reductase activity was essential to produce the cysteine needed to sustain both protein and glutathione synthesis (Figure 4A). Furthermore, based on the lack of significant differences in the gene expression levels of such enzymes between SW620 and LiM2, we hypothesized that LiM2 might share the same dependency. From this analysis, it emerged that both SW620 and LiM2 were largely cystine/cysteine auxotrophs. For such reason, the model simulation predicted as synthetic lethal combinations gene pairs that inhibited both the Xc- and $b^{0,+}$ systems.

To validate this hypothesis, we incubated SW480, SW620, and LiM2 in a culture medium without cystine. Proliferation was reduced significantly more in the metastatic cell lines, confirming that they were more dependent on cystine uptake from the media (Figure 4B). As expected, such effect could be rescued through the addition of N-acetyl cysteine (NAC) which can be deacylated to form cysteine(Whillier et al., 2009).

The metastatic cell lines are vulnerable to the inhibition of the Xc- system

Having established the dependence of the metastatic cell lines on cystine uptake, we proceeded to evaluate the therapeutic potential of inhibiting cystine transporters. With the simulated flux maps showing significantly more flux through the Xc- system across all analyzed cell lines (Figure 4C) and the evidences that its inhibition is sufficient to inhibit cell proliferation and tumorigenic capacity in several cancer types (Gout et al., 1997; Lo et al., 2008; Ma et al., 2015; Savaskan et al., 2008; Shin et al., 2018) we decided to focus on such system.

Hence, we evaluated the growth inhibitory effects of erastin, an irreversible inhibitor the Xc- system(Dixon et al., 2012, 2014; Sato et al., 2018), in SW480, SW620, LiM2 and the non-tumoral colon cell line NCM460. Remarkably erastin had been originally identified as a drug selective for cell lines with mutant RAS making it an attractive drug choice(Dixon et al., 2012). As expected, erastin had lower IC_{50} values for the metastatic cells than for the primary and non-tumoral cell lines (Figure 4D, Table 2).

Additionally, erastin had been reported to significantly reduce the viability of cancer cells by depleting glutathione and leading to an iron-dependent oxidative death termed ferroptosis (Dixon et al., 2012, 2014). In this regard, the effect of erastin on cell viability was evaluated in SW620, LiM2, SW480, and NCM460 and it was found to selectively and significantly increase apoptosis of the metastatic cell lines (Figure 4E).

The metastatic cell lines are dependent on MTHFD1 activity

Folate metabolism consists of two branches compartmentalized in the mitochondrial matrix and the cytosol. Both branches are interconnected through intermediaries and products that can be transported between the cytosol and the mitochondrial matrix, namely tetrahydrofolate, serine, glycine, ^{10}N-formyl tetrahydrofolate (CHO-THF) and formate (Tibbetts and Appling, 2010). Because of such redundancy, in most cancer types, inhibition of an enzymatic activity in any of the branches can be compensated by activities in the other branch(Ducker et al., 2016). However, this is not always the case, and some cancer types can be vulnerable to inhibition of activities specific to either of the branches(Ding et al., 2018a; Nishimura et al., 2019; Paone et al., 2014; Pikman et al., 2016).

In this regard, in the project DRIVE database, it was reported that SW620 but not SW480 were highly dependent on the expression of “Methenyltetrahydrofolate Cyclohydrolase, Formyltetrahydrofolate Synthetase” (MTHFD1) which codes for an enzyme which catalyzes three steps of the cytosolic branch of folate metabolism (Figure 5A).

To confirm that such dependency is related to its enzymatic activity, NCM460, SW480, SW620 and LiM2 were incubated with LY345899. LY345899 is an inhibitor of both MTHDF1 and MTHFD2 but with a significantly lower K_i for the former(Gustafsson et al., 2017; Schmidt et al., 2000). Such inhibitor was shown to selectively inhibit the proliferation of SW620 and LiM2, confirming the dependency of the metastatic cell lines on the MTHDF1 enzymatic activity (Figure 5B, Table 2).

MCS analysis predicted that such dependency emerged because the CHO-THF generated into the mitochondria could not be transported to the cytosol to compensate for MTHDF1 deficiency (Figure 5A). CHO-THF is the substrate of Phosphoribosylglycinamide Formyltransferase (GART) activity, which is necessary for de novo synthesis of purine and consequently is predicted to be required for proliferation. This suggests that in the metastatic cell lines, folate metabolism is, to some extent, uncoupled between the cytosol and the mitochondrial matrix.

The metastatic cell lines are vulnerable to SHMT inhibition and antifolates

Both metastatic cell lines displayed significantly higher fluxes through the cytosolic branch of folate metabolism than SW480 (Figure 5C). Indeed, when both the cytosolic and mitochondrial branches of folate metabolism were considered, SW620 and LiM2 also displayed significantly higher total activities than SW480 making folate metabolism an attractive therapeutic target.

In this regard SHIN2 (Ducker et al., 2017), a chemical inhibitor of both SHMT1/2, had similar growth inhibitory effect on all metastatic and non-metastatic cell lines suggesting that both the primary and the metastatic cell lines were dependent on SHMT activity for proliferation (Figure 5D, Table 2). However, SHIN2 was a selective drug for colorectal cancer cells as it had less growth inhibitory effect on the immortalized colon epithelial cell line NCM460. Even more, SHIN2 induced significantly more apoptosis in the metastatic cell lines than SW480 or NMC460 (Figure 5E).

Additionally, the therapeutic effectivity of the antifolates methotrexate, pemetrexed and lometrexol were also evaluated. All antifolates were shown to have greater growth inhibitory effects in the metastatic cell lines than in the primary cell line or healthy colon epithelial cell line (Figure S1A-C, Table 2).

High PDH and citrate synthase activities protect SW620 and LiM2 from carnosine accumulation

Due to the reported antineoplastic role of carnosine, it can be of therapeutic interest to identify the drivers of the synthesis of such metabolite in the cell models of study. The limiting substrate

for carnosine synthesis is generally assumed to be β-alanine due to both its low concentration and the low affinity of carnosine synthase for such metabolite(Sale et al., 2010). β-alanine can be synthesized from aspartate decarboxylation, spermine oxidation, uracil catabolism and from propionyl-CoA(Kanehisa and Goto, 2000).

Because in the SW480 cell line both carnosine and aspartate concentration were significantly higher than in the metastatic cell lines, it became apparent that aspartate decarboxylation was the predominant pathway fueling the synthesis of β-alanine, and, consequently, carnosine in such cell line. Aspartate decarboxylation is reported to be catalyzed by GAD1(glutamate decarboxylase 1) (Kanehisa and Goto, 2000; Porter and Martin, 1986) although recently it has been proposed that it can also be catalyzed by GADL1 (glutamate decarboxylase like 1)(Liu et al., 2012). Although we lacked gene expression measurements of GADL1, GAD1 was highly expressed in SW480, SW620, LiM1, and LiM2 and lowly expressed in NCM460(Provenzani et al., 2006; Sheffer et al., 2009; Urosevic et al., 2014).

In order to identify the drivers of carnosine synthesis, the condition-specific GSMMs were used to systematically analyze the effect of gene KOs on carnosine synthesis. According to such simulations, the most efficient intervention to increase carnosine synthesis was the inhibition of either Citrate Synthase (CS) or PDH. The model predicted that inhibiting either PDH or CS would lead to an accumulation of aspartate which would be redirected towards β-alanine synthesis fueling carnosine production (Figure 6). Indeed, the flux through CS and PDH was significantly higher in SW620 and LiM2 than SW480 (Figure 2G), suggesting that reduced flux through such reactions is likely to be the cause of the increased aspartate and carnosine content in SW480. However, it is worth noting that such targets would not affect NCM460's carnosine synthesis as their transcriptomics profile indicates they have low GAD1 expression(Sheffer et al., 2009).

Discussion

As part of our effort to identify the metabolic vulnerabilities in the colorectal metastatic populations, we performed a thorough characterization of the same-patient derived cell lines SW480, SW620 and LiM2. The lymph node and liver metastatic cell lines SW620 and LiM2, respectively, were shown to be endowed with enhanced proliferative and spheroid formation capacity compared to SW480, which were derived from the primary tumor site. Conversely, SW480 had higher motility and had less E-cadherin expression compared to SW620 and LiM2. In this regard, E-cadherin is beginning to emerge as a driver of metastatic spread(Gunasinghe et al., 2012; Hugo et al., 2017; Padmanaban et al., 2019; Reichert et al., 2018; Zheng et al., 2015). However, we note that LiM2, and to a lesser extent SW620, also had a strong expression of

mesenchymal markers such as N-cadherin, Fibronectin and Vimentin. Indeed, it has been suggested that the metastatic phenotype is often supported by an intermediate phenotype where the expression of both mesenchymal an epithelial markers coexist (Beerling et al., 2016; Lo and Zhang, 2018; Saitoh, 2018; Shibue and Weinberg, 2017; Thomson et al., 2019) and, indeed, this seems to be the case with the SW620 and LiM2 metastatic populations.

At the metabolic level, we determined that the metastatic cell lines (i.e., SW620 and LiM2) displayed a significantly more active Warburg effect than the primary cell line (i.e., SW480). In this regard, the Warburg effect is widely accepted to support fast proliferation rates by virtue of allowing both fast ATP production and redirection of glycolytic intermediates towards the synthesis of biomass components (Vander Heiden et al., 2009; Locasale and Cantley, 2011; Shlomi et al., 2011). Additionally, because the Warburg effect can produce ATP independent of the ROS-generating mitochondrial respiration, it can contribute to protecting cancer cells from oxidative stress during anoikis(Lu, 2019). Even more, acidification of the tumor microenvironment as a result of lactate secretion can also promote a gene expression program associated with invasiveness and metastasis in cancer cells(Rohani et al., 2019; Rozhin et al., 2013). Hence, the overactivated Warburg effect is poised to play a key role in enabling metastatic spread and targeting it has proven to be an effective strategy to reduce the formation of metastasis in several cancer models(Du et al., 2016; Kolesnik et al., 2015; Sheng et al., 2012; Sottnik et al., 2011; Sun et al., 2010). In this regard, genome-scale metabolic modeling identified several putative metabolic targets associated with the Warburg effect such as PKM and several synthetic lethal combinations with GPI.

Furthermore, the metastatic cell lines also presented a higher glutamine uptake and glutaminolysis than SW480. In this context, glutamine can support proliferation being both a carbon and nitrogen donor (Yang et al., 2017) and has been reported to enhance invasion and colonization capacities (Dornier et al., 2017; Jin et al., 2018; Rodrigues et al., 2016; Shelton et al., 2010; Yang et al., 2014). Additionally, the metastatic cell lines were also shown to be endowed with increased mitochondrial function. Even more, the mitochondrial fuel test determined that such function was highly flexible and could easily adapt to variable substrate availability. Hence, increased mitochondrial function might endow the metastatic cell lines with enhanced capacity to maintain a constant supply of ATP under variable, and potentially hostile, metabolic environments (Cannino et al., 2018). Finally, increased Warburg effect, glutaminolysis and respiration capacity are consistent with the known effects of MYC (Bhutia et al., 2015; Dang et al., 2006; Gao et al., 2009) suggesting that this oncogene might play a key role in the metabolic reprogramming underlying the metastatic phenotype.

Although the Warburg effect can contribute to reducing ROS production, ROS scavenging is largely dependent on glutathione(Bansal and Simon, 2018; Traverso et al., 2013). Indeed, glutathione levels tend to be higher in cancer cells than in healthy tissue and glutathione synthesis has been found to support metastasis in several cancer types (Andreassen et al., 2002; Estrela et al., 2002, 2016; Gal et al., 2015; Ortega et al., 2003). Glutathione is a tripeptide synthesized from glutamate, cysteine and glycine. In this context, increased glutathione synthesis is likely supported by the increased concentrations of glutamate and glycine found in the metastatic cell lines(Boysen et al., 2019; Sappington et al., 2016) as a result of the increased glutaminase and serine hydroxymethyltransferase activities. However, the limited substrate for glutathione synthesis is generally assumed to be cysteine(Estrela et al., 2016).

In this regard, we identified cysteine metabolism as a vulnerability in the metastatic cell lines. More in detail, the cell line specific GSMMs predicted that the metastatic cell lines were dependent on cystine uptake from the extracellular media as they were predicted to have insufficient capacity to produce enough cysteine through the transsulfuration pathway. Although cysteine uptake can be mediated by both the Xc- and $b^{0,+}$ transports, the Xc- system has been found to be overexpressed in several cancer types (Gout et al., 1997; Lo et al., 2008; Savaskan et al., 2008; Shin et al., 2018) including colorectal cancer (Ma et al., 2015). Additionally, its expression has been shown to be consistently higher in cell lines with mutant KRAS (Lim et al., 2019), such as SW480, SW620 and LiM2, and its activity and stability enhanced through the interaction with CD44v6 (Ishimoto et al., 2011) which is a marker of metastatic potential and CSC phenotype in colorectal cancer(Todaro et al., 2014). Here we determined that inhibition of Xc- with erastin can reduce both cell proliferation and cell viability in the metastatic cell lines (SW620 and LiM2). Even more, with the colorectal cancer cell line SW480 and the healthy colon epithelial cell line NMC460, higher drug doses were required to achieve growth inhibition and no decrease in viability was observed. Hence, erastin is highly specific for the metastatic cell lines and it is a potentially effective therapeutic strategy against metastasis in colorectal cancer. Indeed, such treatment would synergize well with electrophilic chemotherapeutic agents (e.g. cisplatin) (Ma et al., 2015; Sato et al., 2018) or radiotherapy (Cobler et al., 2018; Pan et al., 2019) as resistance to such therapeutic interventions is mostly glutathione-dependent.

Additionally, from genome-scale simulations, we also identified folate metabolism, particularly the cytosolic branch, as a pathway strongly upregulated in the metastatic cell lines compared to the primary colorectal cell line SW480. Indeed, because of its role supporting *de novo* nucleotide synthesis, epigenetic regulation and energy and redox balance, folate metabolism is often overactivated as part of cancer progression (Fan et al., 2014; Kim, 2005; Locasale, 2013; Meiser

et al., 2016; Newman and Maddocks, 2017; Tedeschi et al., 2013). In this regard, the antifolates methotrexate, pemetrexed and lometrexol had stronger growth inhibitory effect on the metastatic cell lines than on SW480 and NCM460 while the SHMT inhibitor SHIN2 was selective for the three colorectal cancer cell lines. However, LY345899, an inhibitor of the putative metabolic target MTHFD1, was the most selective inhibitor as it selectively inhibited proliferation in the metastatic cell lines (i.e., SW620 and LiM2) with little or no effect on the proliferation of both the primary colorectal cancer (SW480) and the healthy colon epithelial cell line (NCM460).

Such dependence is not the norm because MTHFD1, which catalyzes three steps of the cytosolic branch of folate metabolism, is generally made redundant by the activities of the mitochondrial branch (Ducker et al., 2016; Tibbetts and Appling, 2010). Indeed, according to the model predictions, MTHD1 is essential in the metastatic cell lines because of the reduced coupling between the cytosolic and mitochondrial branches of folate metabolism in the metastatic cell lines. On this subject, Serine Hydroxymethyltransferase 1 (SHMT1), Thymidylate Synthase and MTHFD1 have been reported to translocate to the nucleus during the S phase of the cell cycle in order to promote thymidylate synthesis (Anderson and Stover, 2009; Anderson et al., 2012; MacFarlane et al., 2011). Similarly, although purine synthesis is assumed to occur primarily in the cytoplasm(An et al., 2008), recent evidence suggests that several key enzymes of the pathway, such as GART, are partially localized in the nucleus where the pathway can be active in an MTHFD1 dependent manner(Sdelci et al., 2019). Hence, the dependency of MTHFD1, and the reduced coupling between the cytosolic and mitochondrial branches of the folate cycle could possibly be attributed to increased nuclear localization of the folate pathway into the nucleus in the metastatic cell lines in order to support increased DNA replication and RNA transcription. In this regard, because of their nuclear localization, in some cellular models, inhibition of SHMT1 or MTHFD1 has been reported to induce cell cycle arrest and compromise cell viability by reducing thymidylate synthesis and thus leading to uracil incorporation into DNA (Field et al., 2015; Kamynina et al., 2017; Macfarlane et al., 2011; Paone et al., 2014).

Finally, we determined that both SW480 and NCM460 had increased intracellular concentration of carnosine, a dipeptide of histidine and β-alanine, compared to the metastatic cell lines. Remarkably, carnosine has been reported to inhibit cell proliferation in several cancer models(Bao et al., 2018; Cheng et al., 2019; Ding et al., 2018b; Fouad et al., 2017; Shen et al., 2014), including colorectal cancer (Iovine et al., 2014, 2016; Lee et al., 2018) where it was also shown to enhance response to chemotherapy(Iovine et al., 2016). Additionally, carnosine has also been reported to have metabolic effects such as reducing the glycolytic, TCA cycle and

oxidative phosphorylation fluxes(Bao et al., 2018; Cheng et al., 2019; Shen et al., 2014). Although the mechanism of action of carnosine is not fully elucidated, it has been reported to inhibit HIFα (Iovine et al., 2014, 2016) and NF-κB (Fouad et al., 2017; Lee et al., 2018) which are key players in tumor progression, and indeed NF-κB levels were decreased in SW480 compared to SW620 and LiM2. Hence, we hypothesize that decreased carnosine content might contribute to the more aggressive phenotype of SW620 and LiM2.

In this regard, using genome-scale simulations, we identified that carnosine synthesis is low in both SW620 and LiM2 because their high PDH and CS activity contributes to maintaining a reduced concentration of aspartate, which can be decarboxylated to β-alanine in a GAD1-dependent manner. In this regard, a therapeutic intervention aimed at partially inhibiting CS could prove to be an effective therapeutic strategy against cancer cells expressing GAD1. Such inhibition could potentially be achieved with the drug Suramin, reported to inhibit CS (Salvarrey and Cazzulo, 1982; Xu et al., 2010). Indeed, Suramin, originally designed as an anti-parasitic drug, has shown therapeutic potential in cancer therapy due to its ability to antagonize growth factors (Bhargava et al., 2007; Stein et al., 1989; Waltenberger et al., 1996) and such effect might synergize with its putative capacity to induce carnosine synthesis.

Overall, the work here presented provides a greater understanding of the metabolic reprogramming supporting the metastatic phenotype and the metabolic dependencies that emerge from it in the same-patient derived SW480, SW620 and LiM2 cellular models. We observe significant metabolic differences between the metastatic and non-metastatic derived cell lines suggesting that the acquisition of the metastatic capacity is accompanied by a significant reprogramming of metabolism. Interestingly, there are only subtle differences between the metabolic phenotype of SW620 and LiM2 even though the latter are derived through *in vivo* selection of the SW620 clones with the most metastatic potential. This suggests that the metabolic phenotype of SW620 is already largely optimized for metastasis and that the increased metastatic potential in LiM2 is largely attributed to non-metabolic changes in line with previous findings (Urosevic et al., 2014). Indeed, this similarity can be exploited therapeutically as both populations were shown to share the same metabolic dependency on cystine uptake and folate metabolism and hence can be targeted by the same drugs. Targeting such dependencies paves the way for therapeutic interventions selective against the metastatic populations in colorectal cancer.

Methods

Cell lines and culture

SW480 cell line was obtained from ATCC. SW620 and its metastatic derivatives SW620-LiM1 and SW620-LiM2 were obtained from Dr. Gomis at IRB Barcelona. NCM460 cell line was obtained through a Material Transfer Agreement with INCELL. All cells were grown in DMEM with 12,5 mM glucose, 4 mM glutamine, 5% Fetal Bovine Serum and 1% Streptomycin/penicillin at 37ºC in a 5% CO_2 atmosphere.

Cell proliferation and viability assays

Cell proliferation was determined by flow cytometry using flow-count fluorospheres (Beckman Coulter) addition to resuspended cells in culture media. When determining IC_{50} using various concentrations of a drug, cell proliferation was assessed by HO33342 staining. At the end of incubation, cells were washed with PBS, lysed with 0.01% SDS and frozen at -20ºC, thawed at 37ºC and incubated with 4 µg/mL of HO33342 in 1M NaCl, 1 mM EDTA, 10 mM Tris-HCl pH 7.4 for 1 hour at 37ºC in the darkness. Fluorescence was measured at 460 nm after excitation with 337 nm in a FLUOstar OPTIMA Microplate Reader (BMG LABTECH GmbH, Ortenberg, Germany). Apoptotic cells were determined by imaging using Incucyte®, after incubation in the presence of a caspase-3/7 green reagent (Sartorius 4440) at 1:200 dilution. Apoptotic cells were also determined by flow citometry using Annexin V coupled with fluorescent isothiocyanate (FICT) addition to resuspended cells in 10 mM Hepes pH 7.4, 140 mM NaCl, 2.5 mM $CaCl_2$ buffer. Incubation with Annexin V-FICT was performed in the darkness during 30 minutes at room temperature and propidium iodide was added 1 minute before flow cytometry.

Spheroid assays

Cell lines were seeded on 24-well (10^4 cells/well) low attachment plates in medium containing EGF, BFGF, heparin, B27, insulin and hydrocortisone and incubating for one week. Spheroids were analyzed by phase-contrast microscopy and stained incubating them with 0.5 mg/mL MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) for 4h.

Wound healing assay

Cell lines were seeded on 24-well (6 x 10^5 cells/well) in 24-plates. The media was replaced after 24h by another media containing 0.5% of mytomicin and no Fetal Bovine Serum. After 1h incubation, an articficial wound was performed by scratching the monolayer using a pipette tip. The wound's width was measured at 0, 3, 7, 24 and 48h using phase-contrast microscope images.

Western blotting

Protein extracts were obtained from the cultured cells incubating for 20 minutes at 4ºC with RIPA buffer (50 mM Tris pH 8.0, 150 mM sodium chloride, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulphate, 1% protease inhibitor cocktail and 1% phosphatase inhibitor cocktail from Thermo Fisher Scientific Inc.) and scrapping, sonicating and centrifuging at 12.000 g for 15 minutes. Equal amounts of protein extracts were separated by SDS-PAGE gel (10%) and transferred to polyvinylidene fluoride transfer membranes. Membranes were blocked with 5% non-fat milk in PBS-0.1% Tween 20 and incubated with an specific primary antibody followed by an incubation with the appropiate Horse radish peroxidase (HRP)-labelled secondary antibody. HRP activity was assessed with Immobilon ECL Western Blotting Detection Kit Reagent and detected by exposition with photographic film. The primary antibodies used included rabbit anti-E-cadherin (ab1416, Abcam) diluted at 1:1000, mouse anti-N-cadherin (610920 BDtransductionsLab) diluted at 1:1000, mouse anti-γ-catenin (sc-514130, Santa Cruz Biotechnology) diluted at 1:10000, mouse anti-vimentin (MS-129, Thermo Fisher Scientific) diluted at 1:1000, rabbit anti-fibronectin1 (F3648, Sigma-Aldrich) diluted at 1:2000, rabbit anti-NF-κB (sc-103, Santa Cruz Biotechnology) diluted at 1:2000, rabbit anti-AKT (9272S, Cell Signalling Technology) diluted at 1:500, rabbit anti-P-AKT (9271, Cell Signaling Technology) diluted at 1:1000, rabbit anti-cMyc (Y69) (ab32072, Abcam) diluted at 1:10000, mouse anti-PDH (sub E1) (ab110330, Abcam) diluted at 1:1000, rabbit anti-P-PDH (sub E1-A) (S293) (ABS204, Merck Millipore) diluted at 1:10000, rabbit anti-GLS (ab93434, Abcam) diluted at 1:1000, rabbit anti-GLUD1 (ab166618, Abcam) diluted at 1:1000 and rabbit anti-TATA (TBS) (ab63766, Abcam) diluted at 1:1000. The last TATA antibody was used as a load control. Secondari antibodies included Anti-Mouse (GR304350-1, Abcam) diluted at 1:20000 and Anti-Rabbit (GR297013-4, Abcam) diluted at 1:20000.

Spectrophotometric measurements

Glucose, lactate, glutamine and glutamate uptake and production rates were measured from cell culture media concentrations at initial and final incubation time using a COBAS Mira Plus spectrophotometer (Horiba ABX Japan). The rates were normalized by number of cells and cell volume, assuming a constant rate during incubation time, which was verified in previous experiments. Determination of glucose was performed using Hexokinase and D-glucose-6-phosphate Dehydrogenase reactions (ABX Pentra Glucose HK CP, HORIBA ABX, Montpellier, France) and NADPH release was measured at 340 nm. Lactate was measured using Lactate Dehydrogenase (Roche) at 87.7 U/mL, 1.55 mg/mL NAD^+ in 0.2 M hydrazine, 12 mM EDTA pH 9 buffer and NADH release was measured at 340nm. Glutamate was quantified by the Glutamate

Dehydrogenase at 39 U/mL, 2.41 mM ADP, 3.9 mM NAD^+ in 0.5 M hydrazine, 0.5 M glycine pH 9 buffer and NADH release was measured at 340 nm. Glutamine was measured indirectly by first transforming it to glutamate by Glutaminase in 125 mM acetate pH 5 buffer for 30 minutes at 37ºC and then performing the same reaction as for glutamate concentration determination.

OCR measurements, Mito Stress and Mito Fuel Assays

Oxygen consumption rates were measured using a Seahorse XF24 Flux Analyzer (Seahorse Bioscience, North Billerica, MA, USA). Cells were seeded at 7.5 x 10^4 cells/well density for SW480 and 10^5 cell/well density for SW620, LiM1 and LiM2 in 24-well plate pre-coated with collagen (Advanced Biomatrix). Plating technique involved 100 µL seeding of cell suspension and 100 µL extra addition of medium 3 hours later once cells had attached to the surface. After overnight growth, medium was replaced with Seahorse medium (buffer-free DMEM, Sigma-Aldrich supplemented with glucose, glutamine and antibiotics). The plates were equilibrated in a 37ºC-incubator without CO_2 for 60 minutes. The cartridge with the sensors was hydrated with calibration solution (Seahorse Bioscience) overnight at 37ºC and loaded into the Seahorse Analyser at least 30 minutes before starting the experiment to calibrate the sensors.

For the Mito-Stress Assay, under baseline conditions (12.5 mM glucose, 4 mM glutamine, 1% streptomycin and penicillin) injections of 2.5 µM oligomycin (ATP synthetase inhibitor), 500 nM FCCP (Carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone, uncoupling agent) and 2 µM rotenone (inhibitor of complex I) together with 2 µM antimycin A (inhibitor of complex III) were made in order to calculate de different respiratory parameters (basal respiration, ATP production, proton leak, maximal respiration, spare capacity and non-mitochondrial respiration).

For glucose, glutamine and fatty acids Mito-Fuel Assay, baseline conditions were used (12.5 mM glucose, 4 mM glutamine and 1% streptomycin and penicillin). The capacity, dependency and flexibility for each substrate were assessed with the convenient injections of 2 µM UK5099 (inhibitor of pyruvate carrier) 3 µM BPTES (inhibitor of glutaminase) and 4 µM etomoxir (inhibitor of CPT1A).

Targeted metabolomics

Intracellular metabolite profiling and determination of uptake and secretion rates was performed using the Absolute IDQ™ p180 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria). For the quantification of intracellular metabolites, 5 x 10^6 cells in 100 mm plates (NCM460, SW480, SW620, LiM1 and LiM2 cell lines) were trypsinized and centrifuged at 500g for 5 minutes and metabolites were extracted from cell pellets. Pellet was resuspended in 70 µL of 85:15

EtOH:PBS buffer and sonicated 3 times for 5 seconds each, then submerged in liquid nitrogen for 30 seconds and thawed at 95ºC. Then, it was centrifuged 20.000 g, 5 minutes at 4ºC, supernatant was collected, and protein content was measured. For the determination of metabolites uptake and production, metabolites were extracted from cell media taken from the beginning and the end of a 24 h incubation in exponential growth and high confluence conditions and cell number was determined. Both extracts from pellets and media were plated in the Biocrates plate together with the calibration standards, and derivatized to be ready for UHPLC-MS reading (for amino acids and biogenic amines) and FIA-MS/MS (for lipids, sugars and acylcarnitines) according to the manufacturer's instructions. A total of 21 amino acids, 19 biogenic amines, 90 glycerophospholipids, 15 sphingolipids, 40 acylcarnitines and hexose sugars were analysed.

The data was normalized by protein and transformed into a Log2 scale. Clustering and heatmap and statistical analysis were performed using the Metaboanlyst web server(Chong et al., 2019).

Stable Isotope-Resolved Metabolomics

2.5 x 10^6 NCM460, SW480, SW620 and LiM2 cells were seeded in 100 mm plates and after 24 hours the media was changed for either glucose 50% enriched in [1,2-^{13}C]gluose, glutamine 50% enriched in [U-^{13}C]glutamine, or unlabelled substrates. Cells and media were obtained for metabolite extractions at 6 and 24 hours after the labelled substrates were added.

For polar intracellular metabolites analysis, the cells were washed with ice-cold PBS and scrapped with 1:1 metanol:water (adding first 1 mL of methanol, waiting for 1 minute and adding 1 mL of milliQ water afterwards). Then, the samples were sonicated (3 cycles of 5 seconds) and 2 mL of cold chloroform was added. After gentle shaking (30 minutes at 4ºC) the samples were centrifuged (20000 g, 15 minutes at 4ºC) and supernatand was completely dried under air flow. The extracted metabolites were derivatized by adding 50 µL of 2% methoxamine hydrochloride in pyridine for 90 minutes at 37ºC and (N-methyl-N-tert-butyldimethylsilyl) trifluoroacetamide + 1% tertbutyldimethylchlorosilate) for 60 minutes at 55ºC before GC-MS analysis (Agilent 7890A gas chromatograph coupled to a Agilent 7890A mass spectrometer, Agilent Technologies, Santa Clara, CA, USA) using electron impact mode.

For intracellular ribose analysis, RNA was isolated from cell pellets using Trizol reagent, mixing it with chloroform. The aqueous phase was obtained and cold isopropanol was added and centrifuged 12000 g, 15 minutes at 4ºC. The samples were washed several times using cold 75% ethanol and isolated RNA was quantified using a Nanodrop spectrophotometer (ND 1000 V3.1.0, Thermo Fisher Scientific). The samples of purified RNA were hydrolyzed in 2 mL of 2 M HCl at

100ºC for 2 hours, dried under air flow and derivatized using 100 µL of 2% hydroxylamine hydrochloride in pyridine at 100ºC for 30 minutes and 75 µL of acetic anhydride at 100ºC for 1 hour. Then, samples were dried under N_2 flow and resuspended in ethyl acetate before GC-MS analysis using chemical ionization mode.

For extracellular glucose analysis, glucose from cell culture media was isolated using Dowex-1X8/Dowex-50WX8 ion-exchange columns and samples were dried under air flow. Purified glucose was derivatized by incubation with 100 µL of 2% hydroxylamine hydrochloride in pyridine at 100ºC for 30 minutes and 75 µL of acetic anhydride at 100ºC for 1 hour. Then, samples were dried under N_2 flow and resuspended in ethyl acetate before GC-MS analysis using chemical ionization mode.

For extracellular lactate analysis, lactate from cell culture media was isolated adding HCl and 1 mL of ethyl acetate previous to air flow drying. Derivatization was performed by incubation with 200 µL of 2,2-dimethoxypropane and 50 µL of 0.5 N methanolic HCl at 75ºC for 1 hour and then adding 60 µL of n-propylamine at 100ºC for 1 hour more. After drying under N_2 flow, samples were filtered using glass wool through a Pasteur pipette and dried again under N_2 flow. Then, samples were resuspended and incubated with 200 µL of dichloromethane and 15 µL of heptafluorobutyric anhydride at room temperature for 10 minutes, dried under N_2 and resuspended under ethyl acetate before GC-MS analysis under chemical ionization mode.

Isotopologue fractions for each metabolite were measured from raw data using MSD5975C Data Analysis (Agilent Technologies), integrating peak areas and then correcting for natural heavy isotope enrichment using the Midcor software package(Selivanov et al., 2017).

Transcriptomics

Transcriptomics data were obtained from the gene expression omnibus repository(Barrett et al., 2013). For the cancer models (SW480, SW620, LIM1, LIM2), data was taken from GSE1323 (SW480-SW620)(Provenzani et al., 2006) and GSE33350 (SW620-LIM1-LIM2)(Urosevic et al., 2014). Because both sets shared the SW620 cell line, batch effect was corrected using the sva package for R(Leek et al., 2012). Transcriptomics for the NCM460 cell line were obtained from GSE41258(Sheffer et al., 2009) and could not be corrected for batch effects due to having only one replicate.

Transcriptomics were mapped to reactions using the gene protein reaction rules (GPR) defined in Recon2.2(Swainston et al., 2016). In detail, first OR operators were replaced by “MAX()” operators and AND operators by “MIN()” in the GPR expressions. Then, GPR expressions for each

reaction were evaluated, replacing gene IDs by their respective transcript abundances. Under such system, a reaction catalyzed by multiple isoenzymes will be mapped to the maximum gene expression values of all isoenzymes while reactions catalyzed by protein complexes will be mapped to the minimum transcript abundance of the complex's components.

Integrating extracellular flux measurements

Measures of extracellular metabolite concentrations (measured either through spectrophotometric methods or targeted metabolomics) were used to compute the rate of metabolite uptake or secretion normalized by cellular volume using the following equation(Tarrado-Castellarnau et al., 2017):

$$v_{M_{ex}} = \frac{M_1 - M_0}{N_1 - N_0} \cdot \mu \cdot \frac{1}{vol}$$

Where,

$v_{M_{ex}}$ is the estimated rate at which the metabolite M is produced/consumed per cell volume (µmol·h^{-1}·(µl cell volume)$^{-1}$)

M_1 and M_0 are metabolite abundances (µmol) measured at time points t_1 and t_0, respectively;

μ is the growth rate (h^{-1})

N_1 and N_0 are the cell numbers measured at time points t_1 and t_0, respectively.

vol is the cellular volume per cell (µl·cell^{-1}).

Exchange reactions in the cell line specific metabolic models were constrained to the 99.5% confidence intervals for $v_{M_{ex}}$ of the corresponding metabolites.

Integrating OCR measurements, Mito Stress and Mito Fuel Assays

For the SW620, SW480 and LiM2 cell lines, Seahorse measurements were integrated to constraint the oxidative metabolism. Firstly, OCR measurements were used to constrain the rate of oxygen consumption in the cell line specific models. Next, the percentage of OCR associated to ATP synthase, measured in the Mito Stress Assay, was used to constrain the flux through ATP synthase. Finally, from the Mito Fuel Assay, the dependency and capacity for fatty acid oxidation were integrated as the lower and upper bound, respectively, for transport of palmitate into the mitochondria.

Integrating intracellular amino acid and biogenic amines measurements

Metabolomics measured in the cellular pellet can be integrated with the proliferation rate to account for the dilution associated with proliferation. In the framework of constraint-based modeling, this can be represented by adding a sink reaction to measured metabolites the represents the requirements of metabolite synthesis to maintain the concentrations of such metabolite in steady state(Reimers and Reimers, 2016).

$$v_{M_{sink}} = [M] \cdot p \cdot \mu$$

Where:

$[M]$ is the concentration of metabolite M (µmol/mg prot)

p is the protein per cellular volume (mg prot/µl cell volume)

μ: is the proliferation rate in h^{-1}

$v_{M_{sink}}$ is the estimated flux through the sink reaction

To account for uncertainty, the 99.5% confidence intervals for $v_{M_{sink}}$ were added as flux boundaries in the model.

Using lipidomic measurements to personalize the biomass function

Across the analysed conditions, the most abundant phospholipids were the phosphatidylcholines (PC aa) PC aa C34:1, PC aa C34:2, PC aa C36:1 and PC aa C36:2 (Figure S2.A). Such species represent isomeric phosphatidylcholines C x:y where *x* is the total carbon number of both chains and *y* is the total number of unsaturations. The relative abundance of such species in each cell line was used to customize the biomass function in each cell line specific model.

Due to the large number of potential fatty acid chains combinations, Recon2.2 does not simulate individual phospholipid species. Instead, it simulates the fatty acid chains in phospholipids through an artificial "Rtotalcoa" metabolite that is synthesized from a combination of acyls-CoA with stoichiometric coefficients representing the relative abundance of each fatty acid chain in phospholipids.

The most abundant fatty acids in mammals are reported to be oleic (C18:1), palmitate (C16:0), stearic acid (18:0) and palmitoleic (16:1), in that order(Sheikh et al., 2005). Hence, we assumed that the side chains of the phosphatidylcholines primarily consisted of such fatty acids (Figure

S2.B) and their relative abundance was used as coefficients for "Rtotalcoa" synthesis (Figure S2.C).

It is worth noting, that even if other combinations of acyl-CoA can give rise to the measured isobaric phosphatidylcholines, the metabolic cost (NADPH, ATP, Acetyl-CoA) of producing any given phosphatidylcholine will depend primarily on the total length of the fatty acid chains and the number of unsaturations. Hence, the assumption that oleic, palmitate, stearic acid and palmitoleic, are the components of "Rtotalcoa" it's a valid approximation.

Integrating growth rates

The proliferation rates for each cell line were integrated to reflect the different proliferation rates of the cell lines of study. First maximum biomass production in SW620($v_{biomass}^{SW620}$) was computed using FBA(Orth et al., 2010). Then the biomass production of the remaining cell lines was set as follows:

$$v_{biomass}^{cell\ line} = v_{biomass}^{SW620} \cdot \frac{\mu^{cell\ line}}{\mu^{SW620}}$$

$v_{biomass}^{cell\ line}$ is then set as the upper bound for the biomass reaction allowing the models to accurately reflect the different growth rates of the different cell lines.

^{13}C MFA and GSMM integration

The central metabolism flux map consistent with the measured ^{13}C propagation and the measured rates of uptake and secretion for glucose, lactate and amino acids, and respiration data were computed in the framework of ^{13}C MFA (Antoniewicz, 2018).

^{13}C MFA was performed in a metabolic network of central carbon metabolism built from Recon2.2. The network comprised 347 reactions including, glycolysis, TCA cycle, pentose phosphate pathway, energy and redox metabolism and the main pathways of amino acid metabolism and synthesis for biomass components.

Using INCA (isotopomer network compartmental analysis) (Young, 2014), 95% confidence intervals for flux values were computed for all reactions in the network. Such confidence intervals were added as flux boundaries in the cell line specific GSMMs. While most reactions could be directly mapped, some reactions that were defined as single reactions in the ^{13}C MFA network were defined as multiple reactions in Recon2.2 (i.e. a reaction that can occur in multiple compartments or a reaction that can take either NAD or NAPD as cofactor). In such instances, the ^{13}C MFA confidence interval were used to constraint the summation of the flux through the equivalent reactions in Recon2.2.

Minimal cut set analysis

From the project DRIVE database, the list of essential metabolic genes in SW480 and SW620 was obtained(McDonald et al., 2017). A gene was considered essential if its Redundant siRNA Activity (RSA) score was equal or lower than -2. Conversely, a gene was defined as dispensable (i.e. not essential) if it had an RSA score larger or equal than -1.

Minimal cut set (MCS) analysis(Apaolaza et al., 2017, 2018) was performed to identify the MCSs containing the essential genes identified in Project DRIVE. MCSs are minimal sets of genes or reactions whose simultaneous removal directly blocks a metabolic task, in this case biomass production. MCS analysis was run in Recon2.2 with extracellular metabolite availability in the media defined based on the composition of DMEM. Furthermore, blocked reactions were removed and reactions that could only be part of linear pathways were grouped. MCS analysis was set to seek 8 MCS containing each Project DRIVE essential gene and all reactions were evaluated as possible gene set candidates. As reported by Apaolaza *et a*l(Apaolaza et al., 2017), in some instances the MCS algorithm can fail to reach optimality and provide a gene set, that although it contains MCS, is not minimal. When this occurred, we used FASTL(Pratapa et al., 2015) to identify the MCS within the returned set.

It can be assumed that if a gene metabolic function is essential in a given cell line, then this gene will be part of a MCS where the other MCS members have low activity. To integrate this information for SW480 and SW620 the following workflow was used for each cell line:

- *Rank essential genes based on their RSA score (low to high).*
- *For each essential gene in the ranked list:*
    - *Rank MCS containing the gene of interest based on the gene expression evidence of the reactions in the set, excluding reactions associated with the gene of interest. This is achieved by mapping transcriptomics to reactions using the GPR defined in Recon2.2. MCS with more than 8 reactions, not counting reactions associated with the gene of interest, are excluded.*
    - *For each MCS in the ranked list*
        - *Implement the MCS. Force the reactions in the set to be inactive, excluding the reactions associated with the essential gene. Following the definition of MCS the gene of interest is now an essential gene.*
        - *Use FVA*(Gudmundsson and Thiele, 2010) *to evaluate if the MCS is consistent with:*

- *Intracellular metabolomics (i.e., all detected metabolites can be produced).*
- *All measured rates of uptake and secretion and OCR measurements.*
- *^{13}C MFA flux intervals.*

- *Systematically simulate the effect of gene the KO for all metabolic gene defined as dispensable in Project Drive using FBA. Compute the number of false positives (dispensable genes that are predicted as essential by the model)*
- *If implementing the MCS a) increases the number of False positives OR b) is inconsistent with metabolomics, flux measurements or ^{13}C MFA:*
  - *Revert the MCS implementation.*
  - *Continue with the next MCS in the ranked list.*
- *Else:*
  - *Continue to the next essential gene in the ranked list.*

For LiM2, we implemented all the MCSs shared between SW480 and SW620 and the MCS in SW620 where none of the genes associated with the reactions in the set was significantly overexpressed.

In total, 8 ,7 and 6 MCSs were implemented into SW480, SW620 and LiM2, respectively.

Gene Inactivation Moderated by Metabolism, and Expression (GIMME)

GIMME was used to integrate transcriptomics together with the aforementioned data sets (e.g., extracellular fluxes, metabolomics, ^{13}C MFA, MCSs) to build cell line specific genome-scale flux maps. GIMME optimizes biomass production and then performs a second optimization where fluxes through reactions are minimized with a weight that is a function of the gene expression value mapped to each reaction(Schmidt et al., 2013). In our analysis, the minimization weight (w_i) of each reaction was defined as follows:

$$w_i = 1 + \max(Th - ge_i, 0)$$

where,

ge_i is the gene expression value mapped to reaction *i following the* GPR rules defined in Recon2.2.

Th is the gene expression threshold bellow which reactions are given additional minimization weight. In this analysis, it was set to maximum gene expression value for metabolic genes.

For each cell line, GIMME was run in a condition-specific GSMM obtained by implementing MCS, the personalized biomass function, metabolomics, respiration parameters, rates of metabolite uptake and secretion and ^{13}C MFA flux intervals in Recon2.2 as detailed in the previous sections. From the optimal GIMME solution, any inactive reaction with a mapped gene expression value under the 25th percentile of metabolic gene expression was removed. This allows pruning reactions catalyzed by lowly expressed enzymes from the network. Next, each flux was maximized and minimized to identify the ranges of feasible fluxes within the optimal GIMME solution with a tolerance of 99.9%(Schmidt et al., 2013). This space of solutions represents the space of most likely flux distributions in the conditions of study.

Finally, GIMME was also run with a 90% tolerance in a model integrating cell line specific MCS and biomass function but no other cell lines specific measurements such as ^{13}C MFA or metabolomics. Rather than representing the flux map under the conditions of study, the purpose of this model is to represent the metabolic potential of each cell line. As such, in such model, the flux boundaries for each reaction are modified to always include 0 (i.e., no reaction is forced to be active). These models, that we termed base models, serve as a framework to simulate gene KOs with MOMA(Segre et al., 2002).

Flux sampling and reference flux distribution selection

From the GIMME, a space of solutions is identified for each cell line. However, such space is still relatively wide and a strategy to select the most representative and accurate flux distributions from such space must be applied. For the SW480, SW620, LiM2 the following approach was used to select the most representative flux distributions:

1. Compute 1000 flux samples from within the GIMME solution space. Flux samples were computed using the Artificially Centered hit-and-run (ACHR) algorithm implemented into COBRApy (Ebrahim et al., 2013; Heirendt et al., 2019). ACHR was run with a thinning factor of 10000. The thinning factor defines the number of iterations between each returned sample and a large thinning factor reduces the correlation between samples resulting on a more representative set of samples.
2. Using each flux sample as a wild type flux distribution, systematically simulate the KO with MOMA(Segre et al., 2002) of all metabolic genes analysed in Project DRIVE in the cell line specific base model.
3. Use the following equation to give a discrepancy score to each flux sample based on how well they encapsulate gene essentiality/dispensability data

$$S_j = \sum_{i \in DRIVE} max(2 + RSA_i, 0) \cdot \left(1 - \frac{BiomassKO_i}{Biomass_{WT}}\right)$$

Where:

S_j is the discrepancy score for flux sample *j*

RSA_i is the RSA score for gene *i* (RSA scores have negative values and more negative values indicate more dependency on gene function). For LiM2, the DRIVE RSA measurements of for SW620 are used as it is the most closely related cell line.

$BiomassKO_i$ is the flux through the biomass reaction when the KO of gene *i* is simulated with MOMA(Segre et al., 2002) using flux sample *j* as input.

$Biomass_{WT}$ is the flux through the biomass reaction in the wild type (no reaction inactivated).

4. Select the top 100 flux samples with the least discrepancy score. The average of such flux samples will be used as reference flux distribution for the cell line of study.

Such an approach primarily serves to reduce the number of false positives (i.e., genes predicted as essential by the model and described as dispensable in DRIVE) that might emerge from unrepresentative flux distributions. Sensitivity for essential genes did not increase in this step as essential genes were already integrated as part of MCS.

For NCM460, lacking gene essentiality data from project DRIVE, 100 flux samples were computed with a thinning factor of 100000, and the reference flux distribution was defined as the average of such samples.

Identifying putative metabolic targets

To identify metabolic targets against colon cancer, gene KOs were systematically simulated for metabolic genes, single or in pairs. The reactions to be blocked by each gene KO(s) were determined by combining gene expression data with gene protein reaction rules of Recon2.2. A reaction was considered to be inactive if when a gene is inactivated (i.e., its expression set to 0) the mapped gene expression value decreased at least 16th fold. Then the effect of reaction KOs was simulated using the reference flux distribution computed for each cell line as input for running MOMA(Segre et al., 2002) in the framework of cell line specific base models.

Single gene KOs were systematically performed for all cell lines under study. Conversely, for SL pairs, due to the larger number of combinations to test (175000>), all potential SL pairs were only evaluated in SW620. The SL gene combinations that resulted in a biomass production below 15% of wild type in SW620 and displayed synergy were evaluated on the remaining cell lines. A

gene pair was considered to have synergy if the fraction of biomass production under the double KO was less than the product of the fraction of biomass production under the individual KOs.

A gene or gene pair was considered a selective target if it reduced the biomass production to 15% or less of the wild type in both SW620 and LiM2 while allowing a biomass production of 30% or more in the NCM460 cell line.

To predict the effect of gene KOs on carnosine and aspartate, the same approach was used but the flux through the sink reactions of carnosine and aspartate production were evaluated.

Declaration of Interests

The authors declare no competing interests.

Figure Legends

Figure 1

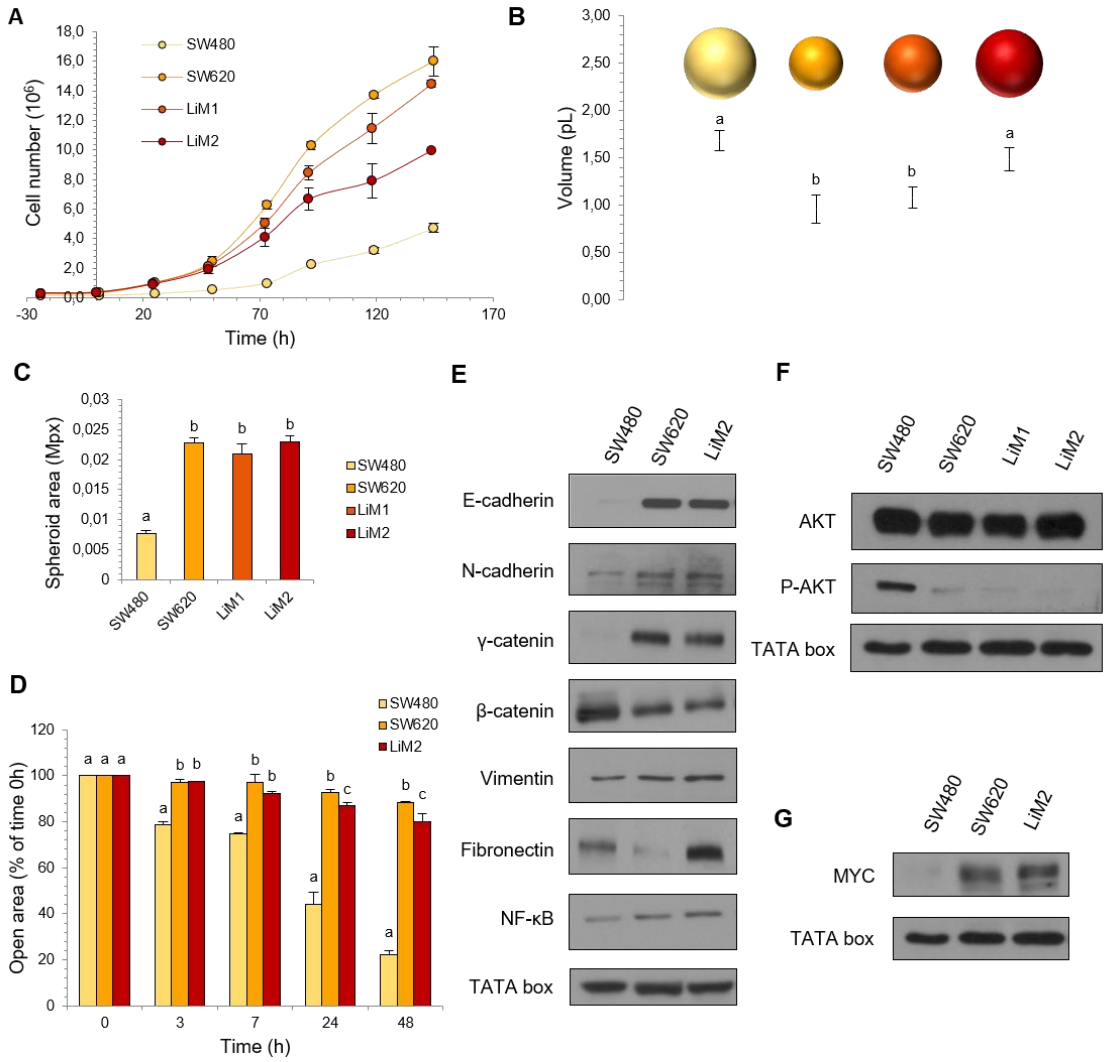


Figure 1. Characterization of the metastatic phenotype (A) Growth curve from 0-150h (changing the medium and measured each 24h). Medium DMEM 12,5 mM glucose and 4 mM glutamine, 5% FBS and 1% S/P. (B) Cell volume measured by Scepter® at various incubation times. (C) Spheroid formation assay done in low attachment plates and medium containing EGF, Bfgf, Heparine, B27, Insuline and Hydrocortisone. Quantification from images after a week incubation. (D) Migration area quantification from a wound healing assay for 0-48h of mitomycin incubation after the wound was made. (E) Epithelial and mesenchymal markers tested by western blotting. (F) and (G) AKT, P-AKT and c-MYC protein expression levels tested by western blotting. In all cases, a one-way ANOVA was performed for the factor "cell line", and Scheffe's test was used for multiple comparisons. Groups sharing the same letter do not show a significant difference with α=0.05.

Figure 2

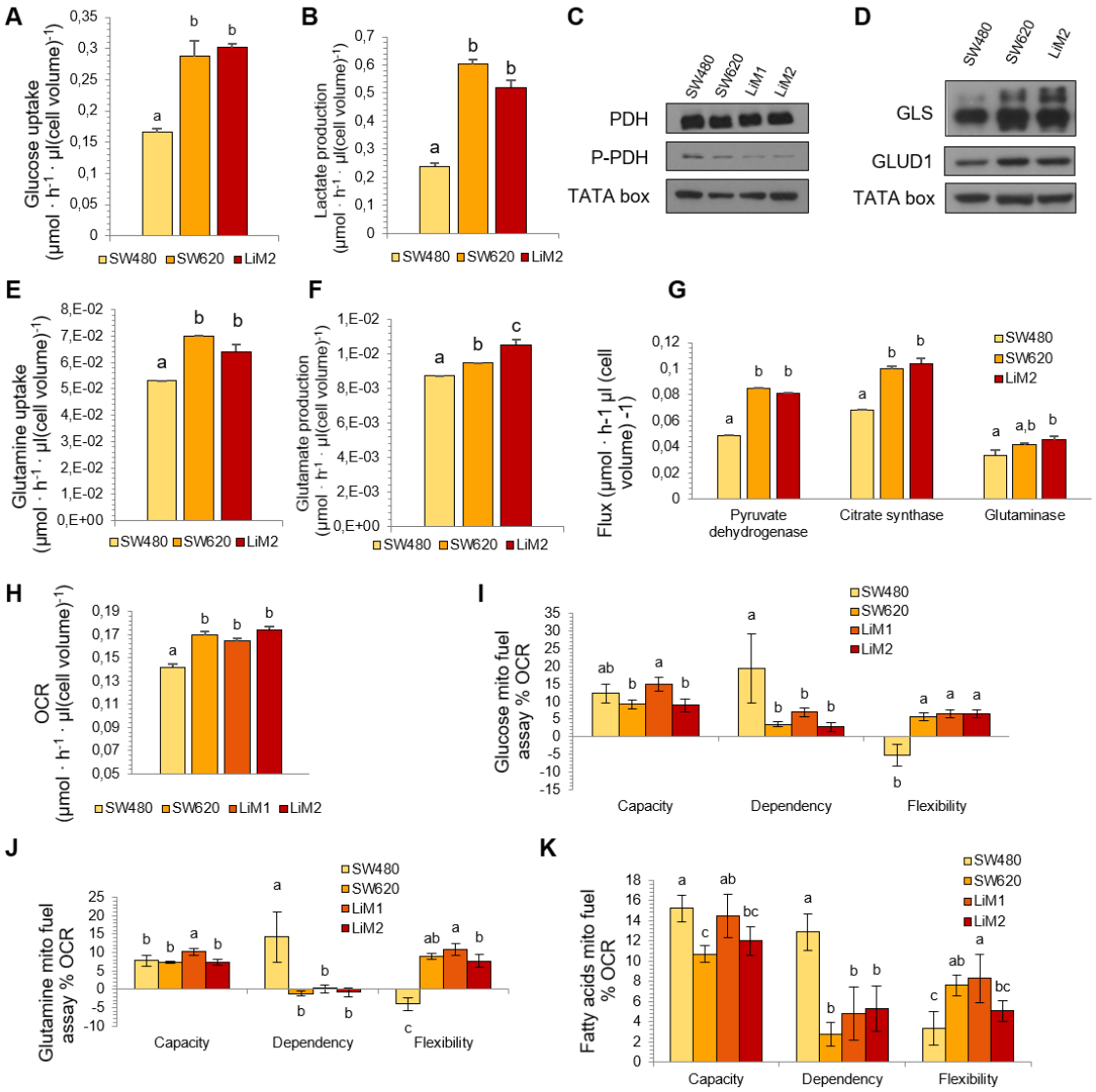


Figure 2. Metabolic changes upon metastatic progression (A) and (B) Glucose uptake and lactate production rates measured by spectrophotometry after 24h of cell culture, normalized by cell volume. (C) and (D) PDH (Pyruvate Dehydrogenase), P-PDH (Phosphorylated Pyruvate Dehydrogenase at S293), GLS (Glutaminase) and GLUD1 (Glutamate Dehydrogenase isoform 1) protein expression levels tested by western blotting.(E) and (F) Glutamine uptake and glutamate production rates measured by spectrophotometry from 0 to 24h of cell culture, normalized by cell volume. (G) Predicted Fluxes for pyruvate dehydrogenase, citrate synthase and glutaminase. Cell lines with a distinct letter (a,b,c) had no overlap in the 100 flux samples obtained from the multiomics integration workflow. (H) Oxygen consumption rate (OCR) values normalized by cell volume at a baseline condition of 12,5 mM glucose and 4 mM glutamine. (I), (J) and (K) Mito fuel assay for glucose, glutamine and fatty acids for each cell line. Percentage of oxygen consumption

rate (OCR) in respect to basal OCR in H. All parameters were assessed with the convenient injections of 2 µM UK5099, 3 µM BPTES and 4 µM etomoxir. In A,B,E,F H,I,J and K, a one-way ANOVA was performed for the factor “cell line”, and Scheffe’s test was used for multiple comparisons. Groups sharing the same letter do not show a significant difference with $\alpha=0.05$.

Figure 3

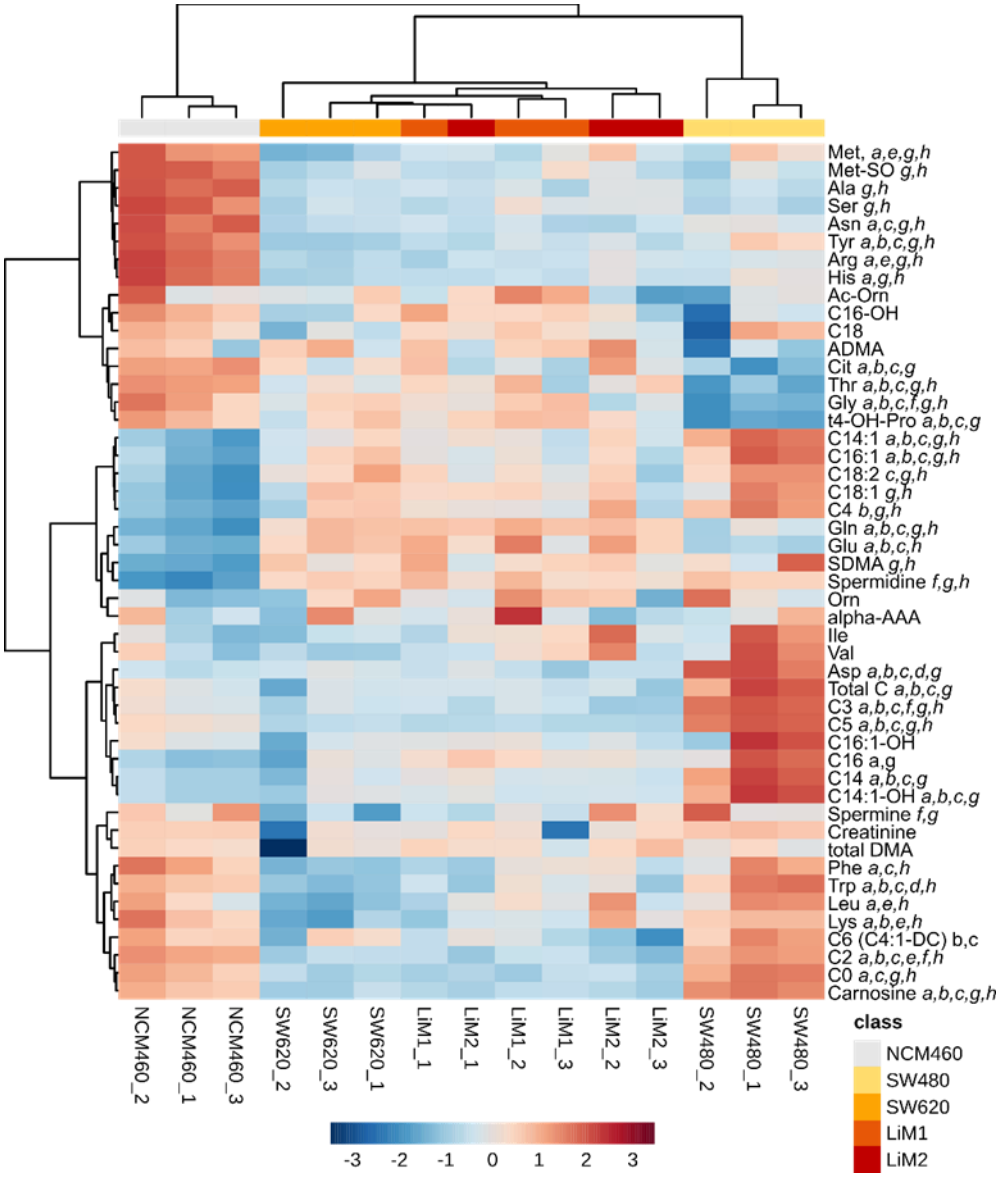


Figure 3. Targeted Metabolomics Heatmap for intracellular content of amino acid, biogenic amines, and acyl-carnitines. Metabolite concentrations were normalized by protein concentration, mean-centered and divided by the standard deviation of each metabolite. Hierarchical clustering for samples and features was performed using the Ward algorithm and Euclidian distance. ANOVA was used to determine statistically significant features, and Fisher's least significant difference method was used to evaluate statistically significant differences (a: SW480-SW620, b: SW480-LiM1 , c: SW480-LiM2, d: SW620-LiM1, e: SW620- LiM2, f:LiM1 - LiM2, g: NCM460-SW480, h: NCM460-SW620). C*x:y* denotes an acylcarnitine carrying fatty acids with *x* carbons and *y* unsaturations.

Figure 4

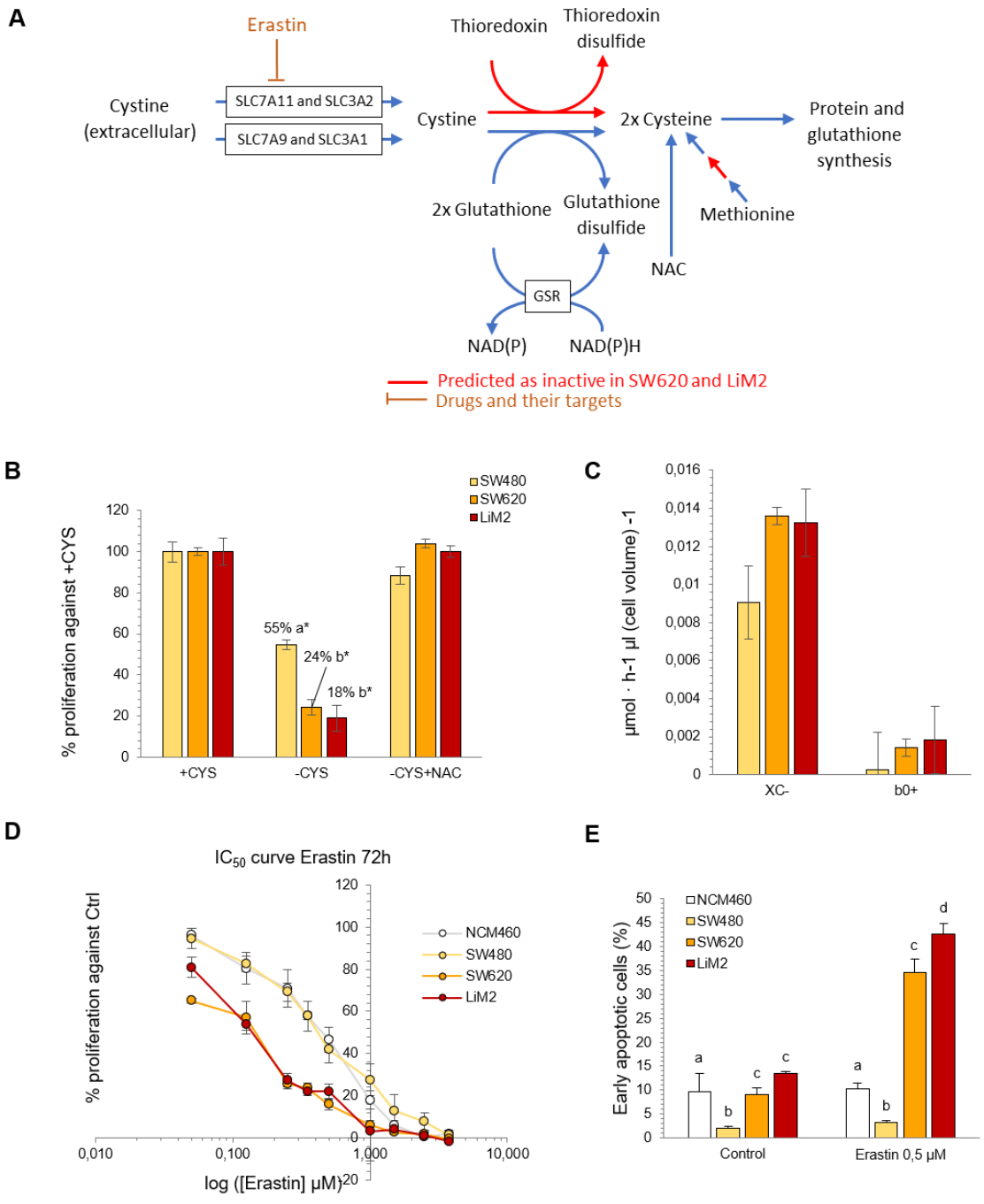


Figure 4. Metastatic cells are dependent on cystine uptake and vulnerable to system Xc- inhibition (A) Graphical representation of cysteine and glutathione metabolism. GSR: Glutathione-Disulfide Reductase. SLC3A1: B(0,+)-Type Amino Acid Transport Protein. SLC3A2: Solute Carrier Family 3 (Activators Of Dibasic And Neutral Amino Acid Transport), Member 2.

SLC7A11: Amino Acid Transport System Xc. SLC7A9: B(0,+)-Type Amino Acid Transporter 1. (B) Cell proliferation measured by DNA content using HO33342 under control conditions (+CYS) (12,5 mM glucose and 4 mM glutamine), under cysteine deprivation (-CYS) and adding N-acetylcysteine to cysteine deprivation (-CYS+NAC). * *t* Student test for –CYS or -CYS+NAC vs Control conditions, $p<0.05$. (C) Precited fluxes through the Xc- and $b^{0,+}$ system. (D) IC_{50} curve of the system Xc- inhibitor erastin assessed by cell proliferation measured by DNA content using HO33342 under various erastin concentrations after 72h incubation. (E) Percentage of early apoptotic cells measured by flow cytometry using Annexin V-PI under a concentration of 0.5 uM of erastin or control conditions after 72h incubation. In all cases, a one-way ANOVA was performed for the factor “cell line”, and Scheffe’s test was used for multiple comparisons. Groups sharing the same letter do not show a significant difference with $\alpha=0.05$.

Figure 5

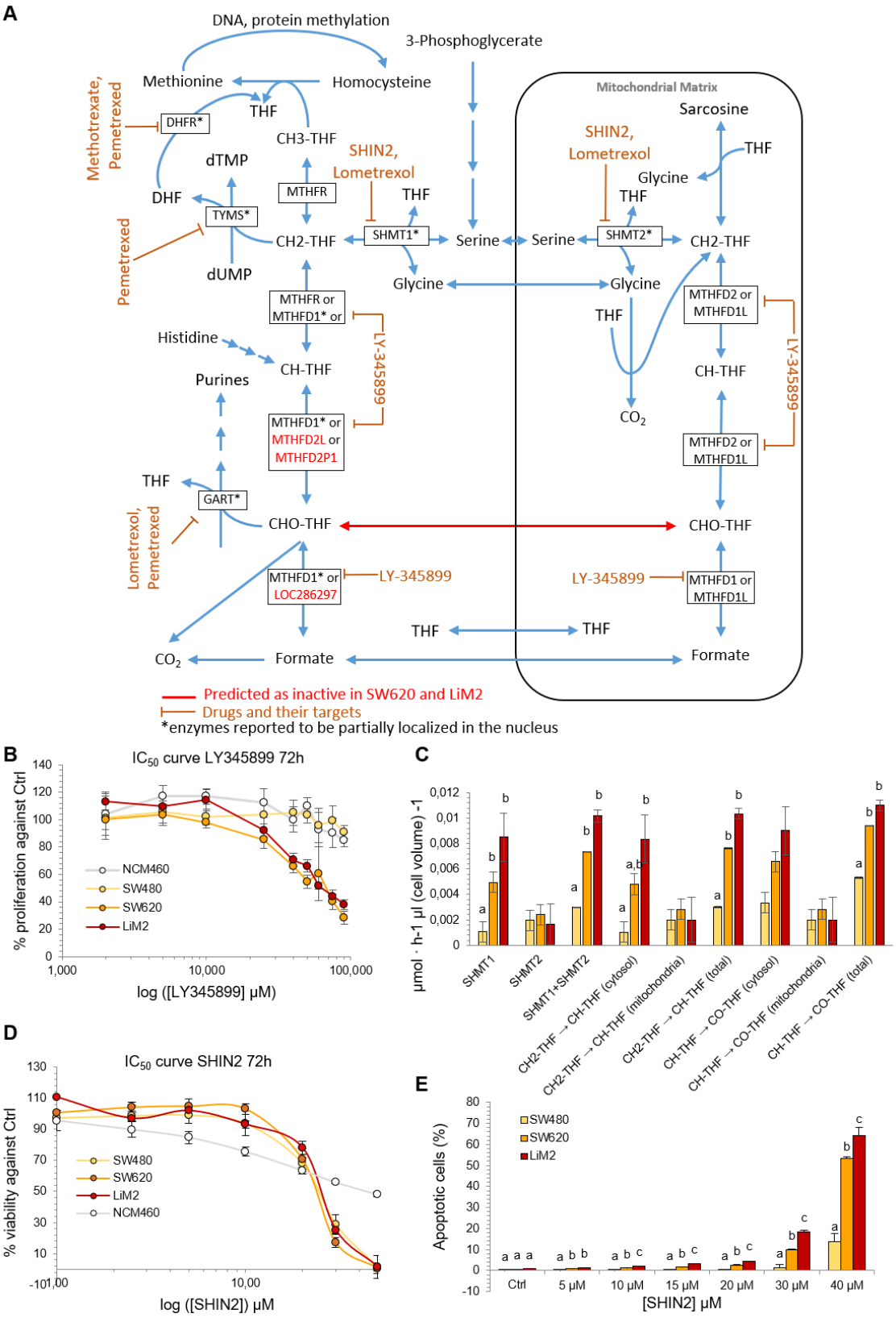

A
DNA, protein methylation
3-Phosphoglycerate
Methionine
Homocysteine
THF
CH3-THF
DHFR*
Methotrexate, Pemetrexed
dTMP
DHF
TYMS*
Pemetrexed
dUMP
MTHFR
SHIN2, Lometrexol
THF
SHMT1*
CH2-THF
Serine
Glycine
MTHFR or MTHFD1* or
Histidine
CH-THF
Purines
MTHFD1* or MTHFD2L or MTHFD2P1
LY-345899
THF
GART*
Lometrexol, Pemetrexed
CHO-THF
MTHFD1* or LOC286297
LY-345899
CO2
Formate
THF
Mitochondrial Matrix
Sarcosine
SHIN2, Lometrexol
THF
Glycine
THF
Serine
SHMT2*
CH2-THF
Glycine
THF
MTHFD2 or MTHFD1L
CH-THF
LY-345899
CO2
MTHFD2 or MTHFD1L
CHO-THF
LY-345899
MTHFD1 or MTHFD1L
THF
Formate
Predicted as inactive in SW620 and LiM2
Drugs and their targets
*enzymes reported to be partially localized in the nucleus
B
IC50 curve LY345899 72h
% proliferation against Ctrl
NCM460
SW480
SW620
LiM2
log ([LY345899] µM)
C
µmol · h-1 µl (cell volume) -1
SHMT1
SHMT2
SHMT1+SHMT2
CH2-THF → CH-THF (cytosol)
CH2-THF → CH-THF (mitochondria)
CH2-THF → CH-THF (total)
CH-THF → CO-THF (cytosol)
CH-THF → CO-THF (mitochondria)
CH-THF → CO-THF (total)
D
IC50 curve SHIN2 72h
% viability against Ctrl
SW480
SW620
LiM2
NCM460
log ([SHIN2]) µM
E
Apoptotic cells (%)
SW480
SW620
LiM2
Ctrl
5 µM
10 µM
15 µM
20 µM
30 µM
40 µM
[SHIN2] µM

Figure 5. Metastatic cells are vulnerable to MTHFD1 and SHMT inhibition (A) Graphical representation of folate metabolism. CH2-THF: N^5,N^{10} methylene-THF. CH3-THF: N^5-methyl tetrahydrofolate. CHO-THF: N^{10}-formyl tetrahydrofolate. CH-THF: N^5,N^{10}-methenyl tetrahydrofolate. THF: Tetrahydrofolate. DHFR: Dihydrofolate Reductase. GART: Trifunctional Purine Biosynthetic Protein Adenosine-3. LOC286297: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 1 Like Pseudogene. MTHFD1: Methylenetetrahydrofolate Dehydrogenase, Cyclohydrolase And Formyltetrahydrofolate Synthetase1. MTHFD1L: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent)1 Like. MTHFD2: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 2, Methenyltetrahydrofolate Cyclohydrolase. MTHFD2L: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent)2 Like. MTHFD2P1: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent)2, Methenyltetrahydrofolate Cyclohydrolase Pseudogene 1. MTHFR: Methylenetetrahydrofolate Reductase. SHMT1/2: Serine Hydroxymethyltransferase1/2. TYMS: Thymidylate Synthetase. (B) IC_{50} curve of the MTHFD1 inhibitor LY345899 determined by cell proliferation measured by DNA content using HO33342 under various LY345899 concentrations after 72h incubation.(C) Predicted flux values for different steps of cytosolic, mitochondrial folate and metabolism. Cell lines with a distinct letter (a,b,c) had no overlap in the 100 flux samples selected through multiomics integration workflow. (D) IC_{50} curve of the SHMT1 and 2 inhibitor SHIN2 assessed by cell proliferation measured by DNA content using HO33342 under various SHIN2 concentrations after 72h incubation. (E) Percentage of apoptotic cells measured by imaging with Incucyte® adding a caspase-3/7 green reagent with various SHIN2 concentrations after 72h incubation.A one-way ANOVA was performed for the factor “cell line”, and Scheffe’s test was used for multiple comparisons. Groups sharing the same letter do not show a significant difference with α=0.05.

Figure 6

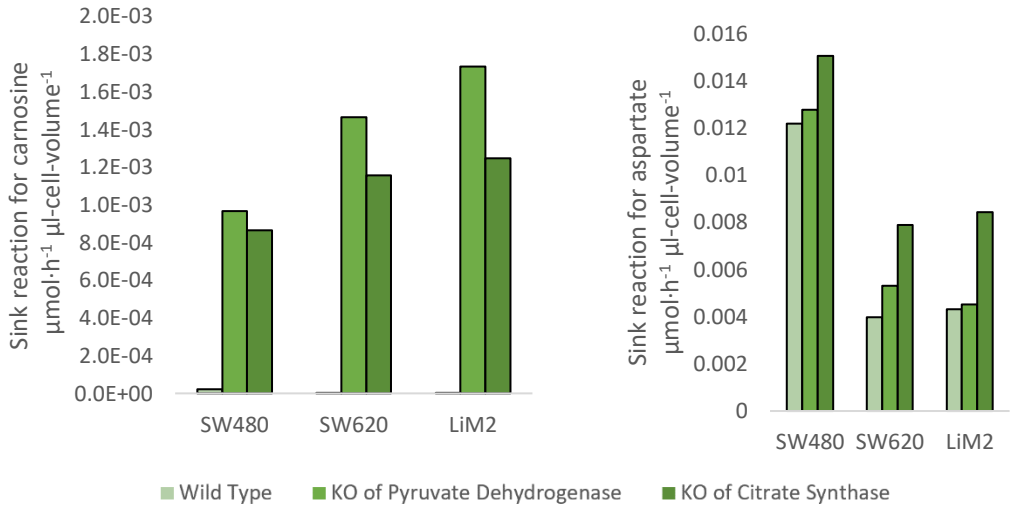


Figure 6. High PDH and citrate synthase activities protect SW620 and LiM2 from carnosine accumulation Simulated flux through the sink reactions for carnosine and aspartate in the wild type or under KO of citrate synthase or pyruvate dehydrogenase. Sink reactions are added to the model to simulate the required production of metabolites to compensate for their dilution due to cell proliferation. The flux through a sink reaction of a given metabolite is proportional to its intracellular concentration and the proliferation rate of the cell line.

Tables

Table 1 Putative metabolic targets. List of the individual or synthetic lethal pairs identified that can selectively impair SW620 and LiM2 proliferation while having a lesser effect on NCM460. * Indicates essential genes described as such in the Project DRIVE database and integrated as part of MCS analysis. **MTHFD1:** Methylenetetrahydrofolate Dehydrogenase, Cyclohydrolase And Formyltetrahydrofolate Synthetase 1. **GSR**: Glutathione-Disulfide Reductase. **PKM:** Pyruvate Kinase M1/2. **RRM1:** Ribonucleotide Reductase Catalytic Subunit M1. **PRODH**: Proline Dehydrogenase 1 **GUK1**: Guanylate Kinase 1. **SLC7A9**: Solute Carrier Family 7 Member 9, **SLC7A11**: Solute Carrier Family 7 Member 11, **SLC3A2**: Solute Carrier Family 3 Member 2, ***SLC3A1*** Solute Carrier Family 3 Member 1, ***GPI:*** Glucose-6-Phosphate Isomerase **CTH:** Cystathionine Gamma-Lyase. **DLST:** Dihydrolipoamide S-Succinyltransferase. **OGDH**: Oxoglutarate Dehydrogenase. **FASN**: Fatty Acid Synthase, **PPA2:** Pyrophosphatase (Inorganic) 2.

	Gene(s) symbol	**The predicted fraction of growth compared to wild type**			
		NCM460	**SW480**	**SW620**	**LiM2**
Single targets	MTHFD1*	100%	100%	0%	0%
	GSR*	99%	99%	0%	0%
	PKM	100%	0%	0%	0%
	RRM1*	99%	0%	0%	0%
	PRODH*	100%	0%	0%	0%
	GUK1*	100%	0%	0%	0%
Synthetic Lethal pairs	SLC7A9, SLC7A11	84%	85%	0%	0%
	SLC7A9, SLC3A2	84%	85%	0%	0%
	SLC3A1, SLC7A11	85%	85%	0%	0%
	SLC3A1, SLC3A2	85%	85%	0%	0%
	GPI, CTH	48%	42%	4%	7%
	GPI, DLST	38%	40%	5%	7%
	GPI, OGDH	38%	40%	5%	7%
	GPI, FASN	37%	28%	0%	0%
	GPI, PPA2	47%	18%	0%	0%

Table 2. IC_{50} values for the tested inhibitors. List of tested inhibitors and its IC_{50} values for NCM460, SW480, SW620 and LiM2 cell lines. The IC_{50} curves were performed assessing cell proliferation by staining DNA with HO33342 under various concentrations of the inhibitors. **Erastin** is an specific inhibitor of system Xc-, **LY345899** is an specific inhibitor of MTHFD1/2 (Methylenetetrahydrofolate Dehydrogenase, Cyclohydrolase And Formyltetrahydrofolate Synthetase 1/2), **SHIN2** is an specific inhibitor of SHMT1/2 (Serine Hydroxymethyltransferase 1/2), **methotrexate** is an inhibitor of DHFR (Dihydrofolate Reductase), **pemetrexed** is an inhibitor of DHFR, TYMS (Thymidylate Synthetase) and GART (Trifunctional Purine Biosynthetic Protein Adenosine-3), **lometrexol** is an inhibitor of GARFT and SHMT1/2.

Inhibitor	**IC_{50} values**			
	NCM460	**SW480**	**SW620**	**LiM2**
Erastin (µM)	0.65 ± 0.21	0.59 ± 0.15	0.11 ± 0.05	0.10 ± 0.06
LY345899 (µM)	191.1 ± 9.9	178.3 ± 45.3	57.5 ± 9.2	58.1 ± 12.7
SHIN2 (µM)	55.9 ± 5.7	19.8 ± 3.5	19.7 ± 3.3	20.8 ± 4.2
Methotrexate (nM)	72.8 ± 5.2	14.97 ± 2.5	10.9 ± 3.4	9.7 ± 2.3
Pemetrexed (nM)	68.1 ± 4.3	29.9 ± 4.2	15.8 ± 3.2	16.3 ± 2.5
Lometrexol (µM)	171.6 ± 20.4	4.2 ± 6.0	0.2 ± 0.2	0.5 ± 0.4

Supplemental Information

Figure S1

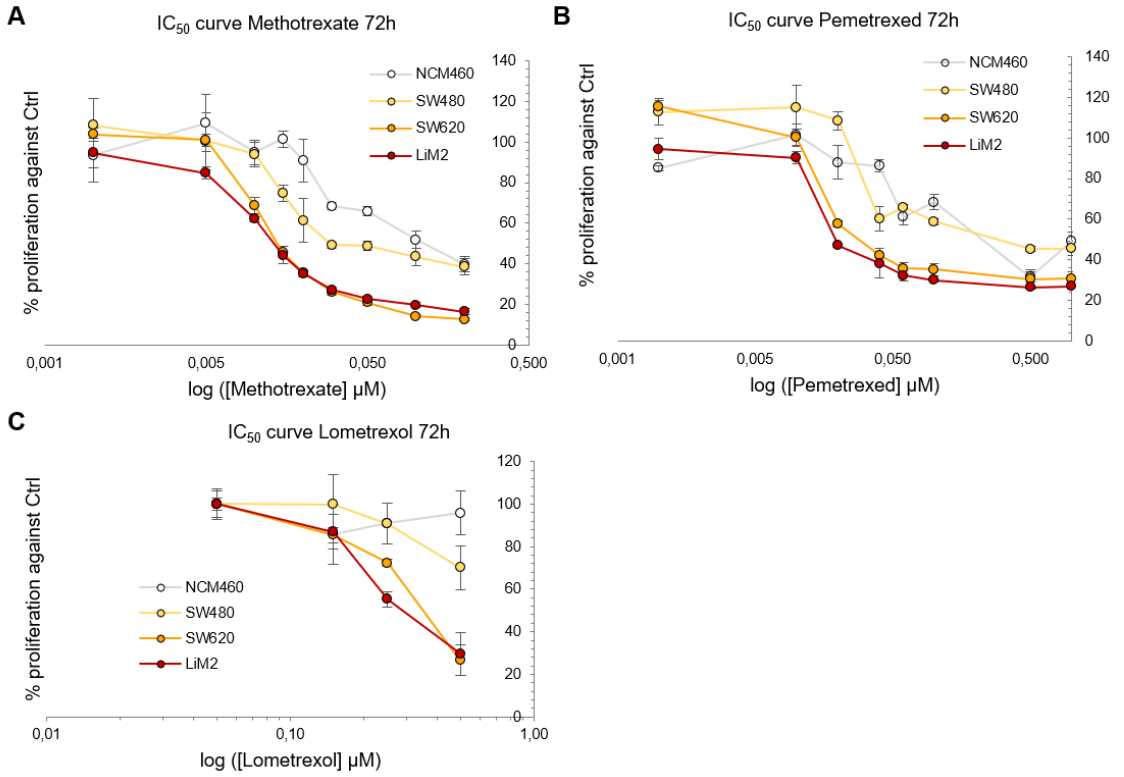
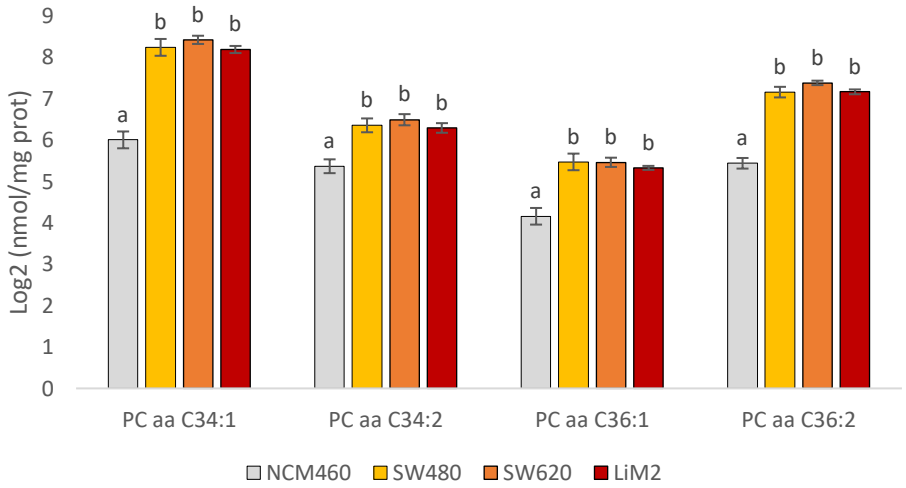


Figure S1. Metastatic cells are vulnerable to folate metabolism inhibitors (A) IC_{50} curve of methotrexate determined by cell proliferation measured by DNA content using HO33342 under various erastin concentrations after 72h incubation. (B) IC_{50} curve of pemetrexed determined by cell proliferation measured by DNA content using HO33342 under various pemetrexed concentrations after 72h incubation. (C) IC_{50} curve of lometrexol determinedd by cell proliferation measured by DNA content using HO33342 under various lometrexol concentrations after 72h incubation.

A: Concenration of the most abundant PC aa



B: Most probable acyl chains for the most abundant phosphatidylcholines

Isobaric species	Most probable acyl chains
PC aa C34:1	C16:0 + C18:1
PC aa C34:2	C16:1 + C18:1
PC aa C36:1	C:18 + C18:1
PC aa C36:2	C18:1+C18:1

C: Predicted relative abundance of each acyl chain each cell line

Acyl chain	NCM460	SW480	SW620	LiM2
C16:0	25.5%	30.5%	30.3%	29.9%
C16:1	19.9%	14.3%	13.8%	13.6%
C18:0	4.3%	3.0%	2.9%	3.1%
C18:1	50.4%	52.2%	53.0%	53.4%

Figure S2: A Levels of the of the most abundant PC AA. ANOVA was used to determine statistically significant features, and Fisher's least significant difference method was used to evaluate statistically significant differences. a,b indicates groups without statistically significant difference. B: Most probable acyl chains for each of the most abundant phosphatidylcholines C: Predicted relative abundance of each acyl chain each cell line.

References

Aktary, Z., Alaee, M., and Pasdar, M. (2017). Beyond cell-cell adhesion: Plakoglobin and the regulation of tumorigenesis and metastasis. Oncotarget *8*, 32270–32291.

An, S., Kumar, R., Sheets, E.D., and Benkovic, S.J. (2008). Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. Science *320*, 103–106.

Anderson, D.D., and Stover, P.J. (2009). SHMT1 and SHMT2 are functionally redundant in nuclear de novo thymidylate biosynthesis. PLoS One *4*, e5839.

Anderson, D.D., Woeller, C.F., Chiang, E.P., Shane, B., and Stover, P.J. (2012). Serine hydroxymethyltransferase anchors de Novo thymidylate synthesis pathway to nuclear lamina for DNA synthesis. J. Biol. Chem. *287*, 7051–7062.

Andreassen, K., Mortensen, B., Winberg, J.-O., and Huseby, N.-E. (2002). Increased resistance towards oxidative stress accompanies enhancement of metastatic potential obtained by repeated in vivo passage of colon carcinoma cells in syngeneic rats. Clin. Exp. Metastasis *19*, 623–629.

Antoniewicz, M.R. (2018). A guide to 13C metabolic flux analysis for the cancer biologist. Exp. Mol. Med. *50*, 19.

Apaolaza, I., San José-Eneriz, E., Tobalina, L., Miranda, E., Garate, L., Agirre, X., Prósper, F., and Planes, F.J. (2017). An in-silico approach to predict and exploit synthetic lethality in cancer metabolism. Nat. Commun. *8*, 1–9.

Apaolaza, I., Valcarcel, L.V., and Planes, F.J. (2018). gMCS: fast computation of genetic minimal cut sets in large networks. Bioinformatics 1–3.

Bannai, S., and Kitamura, E. (1980). Transport interaction of L-cystine and L-glutamate in human diploid fibroblasts in culture. J. Biol. Chem. *255*, 2372–2376.

Bansal, A., and Simon, M.C. (2018). Glutathione metabolism in cancer progression and treatment resistance. J. Cell Biol. *217*, 2291–2298.

Bao, Y., Ding, S., Cheng, J., Liu, Y., Wang, B., Xu, H., Shen, Y., and Lyu, J. (2018). Carnosine Inhibits the Proliferation of Human Cervical Gland Carcinoma Cells ThrCarnosine Inhibits the Proliferation of Human Cervical Gland Carcinoma Cells Through Inhibiting Both Mitochondrial Bioenergetics and Glycolysis Pathways and Retarding Cell Cyc. Integr. Cancer Ther. *17*, 80–91.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. *41*, D991-5.

Becker, S.A., and Palsson, B.O. (2008). Context-Specific Metabolic Networks Are Consistent with Experiments. PLoS Comput. Biol. *4*, e1000082.

Beerling, E., Seinstra, D., de Wit, E., Kester, L., van der Velden, D., Maynard, C., Schäfer, R., van Diest, P., Voest, E., van Oudenaarden, A., et al. (2016). Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. Cell Rep. *14*, 2281–2288.

Bhargava, S., Hotz, B., Hines, O.J., Reber, H.A., Buhr, H.J., and Hotz, H.G. (2007). Suramin Inhibits Not Only Tumor Growth and Metastasis but Also Angiogenesis in Experimental Pancreatic Cancer. J. Gastrointest. Surg. *11*, 171–178.

Bhutia, Y.D., Babu, E., Ramachandran, S., and Ganapathy, V. (2015). Amino Acid transporters in

cancer and their relevance to “glutamine addiction”: novel targets for the design of a new class of anticancer drugs. Cancer Res. *75*, 1782–1788.

Boysen, G., Jamshidi-Parsian, A., Davis, M.A., Siegel, E.R., Simecka, C.M., Kore, R.A., Dings, R.P.M., and Griffin, R.J. (2019). Glutaminase inhibitor CB-839 increases radiation sensitivity of lung tumor cells and human lung tumor xenografts in mice. Int. J. Radiat. Biol. *95*, 436–442.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. *68*, 394–424.

Bridges, R.J., Natale, N.R., and Patel, S.A. (2012). System xc^- cystine/glutamate antiporter: an update on molecular pharmacology and roles within the CNS. Br. J. Pharmacol. *165*, 20–34.

Cannino, G., Ciscato, F., Masgras, I., Sánchez-Martín, C., and Rasola, A. (2018). Metabolic Plasticity of Tumor Cell Mitochondria. Front. Oncol. *8*, 333.

Chae, Y.C., Vaira, V., Caino, M.C., Tang, H.Y., Seo, J.H., Kossenkov, A. V., Ottobrini, L., Martelli, C., Lucignani, G., Bertolini, I., et al. (2016). Mitochondrial Akt Regulation of Hypoxic Tumor Reprogramming. Cancer Cell *30*, 257–272.

Chawla, R.K., Lewis, F.W., Kutner, M.H., Bate, D.M., Roy, R.G.B., and Rudman, D. (1984). Plasma cysteine, cystine, and glutathione in cirrhosis. Gastroenterology *87*, 770–776.

Cheng, J.-Y., Yang, J.-B., Liu, Y., Xu, M., Huang, Y.-Y., Zhang, J.-J., Cao, P., Lyu, J.-X., and Shen, Y. (2019). Profiling and targeting of cellular mitochondrial bioenergetics: inhibition of human gastric cancer cell growth by carnosine. Acta Pharmacol. Sin. *40*, 938–948.

Chillarón, J., Font-Llitjós, M., Fort, J., Zorzano, A., Goldfarb, D.S., Nunes, V., and Palacín, M. (2010). Pathophysiology and treatment of cystinuria. Nat. Rev. Nephrol. *6*, 424–434.

Cho, K. Bin, Cho, M.K., Lee, W.Y., and Kang, K.W. (2010). Overexpression of c-myc induces epithelial mesenchymal transition in mammary epithelial cells. Cancer Lett. *293*, 230–239.

Chong, J., Yamamoto, M., and Xia, J. (2019). MetaboAnalystR 2.0: From Raw Spectra to Biological Insights. Metabolites *9*, 57.

Cobler, L., Zhang, H., Suri, P., Park, C., and Timmerman, L.A. (2018). xCT inhibition sensitizes tumors to γ-radiation via glutathione reduction. Oncotarget *9*, 32280–32297.

Crawhall, J.C., Lietman, P.S., Schneider, J.A., and Seegmiller, J.E. (1968). Cystinosis. Plasma cystine and cysteine concentrations and the effect of D-penicillamine and dietary treatment. Am. J. Med. *44*, 330–339.

Dang, C. V., O’Donnell, K.A., Zeller, K.I., Nguyen, T., Osthus, R.C., and Li, F. (2006). The c-Myc target gene network. Semin. Cancer Biol. *16*, 253–264.

Ding, K., Jiang, J., Chen, L., and Xu, X. (2018a). Methylenetetrahydrofolate Dehydrogenase 1 Silencing Expedites the Apoptosis of Non-Small Cell Lung Cancer Cells via Modulating DNA Methylation. Med. Sci. Monit. *24*, 7499–7507.

Ding, M., Jiao, G., Shi, H., and Chen, Y. (2018b). Investigations on in vitro anti-carcinogenic potential of L-carnosine in liver cancer cells. Cytotechnology *70*, 163–167.

Dixon, S.J., Lemberg, K.M., Lamprecht, M.R., Skouta, R., Zaitsev, E.M., Gleason, C.E., Patel, D.N., Bauer, A.J., Cantley, A.M., Yang, W.S., et al. (2012). Ferroptosis: an iron-dependent form of nonapoptotic cell death. Cell *149*, 1060–1072.

Dixon, S.J., Patel, D.N., Welsch, M., Skouta, R., Lee, E.D., Hayano, M., Thomas, A.G., Gleason,

C.E., Tatonetti, N.P., Slusher, B.S., et al. (2014). Pharmacological inhibition of cystine–glutamate exchange induces endoplasmic reticulum stress and ferroptosis. Elife *3*.

Dolfi, S.C., Chan, L.L.-Y., Qiu, J., Tedeschi, P.M., Bertino, J.R., Hirshfield, K.M., Oltvai, Z.N., and Vazquez, A. (2013). The metabolic demands of cancer cells are coupled to their size and protein synthesis rates. Cancer Metab. *1*, 20.

Dornier, E., Rabas, N., Mitchell, L., Novo, D., Dhayade, S., Marco, S., Mackay, G., Sumpton, D., Pallares, M., Nixon, C., et al. (2017). Glutaminolysis drives membrane trafficking to promote invasiveness of breast cancer cells. Nat. Commun. *8*, 2255.

Du, J., Yang, M., Chen, S., Li, D., Chang, Z., and Dong, Z. (2016). PDK1 promotes tumor growth and metastasis in a spontaneous breast cancer model. Oncogene *35*, 3314–3323.

Ducker, G.S., Chen, L., Morscher, R.J., Ghergurovich, J.M., Esposito, M., Teng, X., Kang, Y., and Rabinowitz, J.D. (2016). Reversal of Cytosolic One-Carbon Flux Compensates for Loss of the Mitochondrial Folate Pathway. Cell Metab. *23*, 1140–1153.

Ducker, G.S., Ghergurovich, J.M., Mainolfi, N., Suri, V., Jeong, S.K., Hsin-Jung Li, S., Friedman, A., Manfredi, M.G., Gitai, Z., Kim, H., et al. (2017). Human SHMT inhibitors reveal defective glycine import as a targetable metabolic vulnerability of diffuse large B-cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. *114*, 11404–11409.

Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst. Biol. *7*, 74.

Elia, I., Doglioni, G., and Fendt, S.M. (2018). Metabolic Hallmarks of Metastasis Formation. Trends Cell Biol. *28*, 673–684.

Estrela, J.M., Pellicer, J.A., Medina, I., Ortega, A., Carretero, J., Obrador, E., and Rodilla, V. (2002). γ-glutamyl transpeptidase overexpression increases metastatic growth of B16 melanoma cells in the mouse liver. Hepatology *35*, 74–81.

Estrela, J.M., Ortega, A., Mena, S., Sirerol, J.A., and Obrador, E. (2016). Glutathione in metastases: From mechanisms to clinical applications. Crit. Rev. Clin. Lab. Sci. *53*, 253–267.

Fan, J., Ye, J., Kamphorst, J.J., Shlomi, T., Thompson, C.B., and Rabinowitz, J.D. (2014). Quantitative flux analysis reveals folate-dependent NADPH production. Nature *510*, 298–302.

Field, M.S., Kamynina, E., Watkins, D., Rosenblatt, D.S., and Stover, P.J. (2015). Human mutations in methylenetetrahydrofolate dehydrogenase 1 impair nuclear de novo thymidylate biosynthesis. Proc. Natl. Acad. Sci. U. S. A. *112*, 400–405.

Fouad, A.A., Qutub, H.O., Al Rashed, A.S., and Al-Melhim, W.N. (2017). Therapeutic effect of carnosine in rat model of experimental liver carcinogenesis. Environ. Toxicol. Pharmacol. *56*, 10–14.

Gal, K. Le, Ibrahim, M.X., Wiel, C., Sayin, V.I., Akula, M.K., Karlsson, C., Dalin, M.G., Akyürek, L.M., Lindahl, P., Nilsson, J., et al. (2015). Antioxidants can increase melanoma metastasis in mice. Sci. Transl. Med. *7*, 1–8.

Gao, P., Tchernyshyov, I., Chang, T.-C., Lee, Y.-S., Kita, K., Ochi, T., Zeller, K.I., De Marzo, A.M., Van Eyk, J.E., Mendell, J.T., et al. (2009). c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. Nature *458*, 762–765.

Gout, P.W., Kang, Y.J., Buckley, D.J., Bruchovsky, N., and Buckley, A.R. (1997). Increased cystine uptake capability associated with malignant progression of Nb2 lymphoma cells. Leukemia *11*, 1329–1337.

Grille, S.J., Bellacosa, A., Upson, J., Klein-Szanto, A.J., van Roy, F., Lee-Kwon, W., Donowitz, M., Tsichlis, P.N., and Larue, L. (2003). The protein kinase Akt induces epithelial mesenchymal transition and promotes enhanced motility and invasiveness of squamous cell carcinoma lines. Cancer Res. *63*, 2172–2178.

Gudmundsson, S., and Thiele, I. (2010). Computationally efficient flux variability analysis. BMC Bioinformatics *11*, 489.

Gunasinghe, N.P.A.D., Wells, A., Thompson, E.W., and Hugo, H.J. (2012). Mesenchymal-epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. Cancer Metastasis Rev. *31*, 469–478.

Gustafsson, R., Jemth, A.-S., Gustafsson, N.M.S., Färnegårdh, K., Loseva, O., Wiita, E., Bonagas, N., Dahllund, L., Llona-Minguez, S., Häggblad, M., et al. (2017). Crystal Structure of the Emerging Cancer Target MTHFD2 in Complex with a Substrate-Based Inhibitor. Cancer Res. *77*, 937–948.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029–1033.

Vander Heiden, M.G., Christofk, H.R., Schuman, E., Subtelny, A.O., Sharfi, H., Harlow, E.E., Xian, J., and Cantley, L.C. (2010). Identification of small molecule inhibitors of pyruvate kinase M2. Biochem. Pharmacol. *79*, 1118–1124.

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat. Protoc. *14*, 639–702.

Hewitt, R.E., McMarlin, A., Kleiner, D., Wersto, R., Martin, P., Tsoskas, M., Stamp, G.W.H., and Stetler-Stevenson, W.G. (2000). Validation of a model of colon cancer progression. J. Pathol. *192*, 446–454.

Holmgren, A. (1977). Bovine thioredoxin system. Purification of thioredoxin reductase from calf liver and thymus and studies of its function in disulfide reduction. J. Biol. Chem. *252*, 4600–4606.

Hugo, H.J., Gunasinghe, N.P.A.D., Hollier, B.G., Tanaka, T., Blick, T., Toh, A., Hill, P., Gilles, C., Waltham, M., and Thompson, E.W. (2017). Epithelial requirement for in vitro proliferation and xenograft growth and metastasis of MDA-MB-468 human breast cancer cells: oncogenic rather than tumor-suppressive role of E-cadherin. Breast Cancer Res. *19*, 86.

Iovine, B., Oliviero, G., Garofalo, M., Orefice, M., Nocella, F., Borbone, N., Piccialli, V., Centore, R., Mazzone, M., Piccialli, G., et al. (2014). The anti-proliferative effect of L-carnosine correlates with a decreased expression of hypoxia inducible factor 1 alpha in human colon cancer cells. PLoS One *9*, e96755.

Iovine, B., Guardia, F., Irace, C., and Bevilacqua, M.A. (2016). L-carnosine dipeptide overcomes acquired resistance to 5-fluorouracil in HT29 human colon cancer cells via downregulation of HIF1-alpha and induction of apoptosis. Biochimie *127*, 196–204.

Ishimoto, T., Nagano, O., Yae, T., Tamada, M., Motohara, T., Oshima, H., Oshima, M., Ikeda, T., Asaba, R., Yagi, H., et al. (2011). CD44 Variant Regulates Redox Status in Cancer Cells by Stabilizing the xCT Subunit of System xc- and Thereby Promotes Tumor Growth. Cancer Cell *19*, 387–400.

Jana, A., Krett, N.L., Guzman, G., Khalid, A., Ozden, O., Staudacher, J.J., Bauer, J., Baik, S.H., Carroll, T., Yazici, C., et al. (2017). NFkB is essential for activin-induced colorectal cancer

migration via upregulation of PI3K-MDM2 pathway. Oncotarget *8*, 37377–37393.

Jin, L., Chun, J., Pan, C., Kumar, A., Zhang, G., Ha, Y., Li, D., Alesi, G.N., Kang, Y., Zhou, L., et al. (2018). The PLAG1-GDH1 Axis Promotes Anoikis Resistance and Tumor Metastasis through CamKK2-AMPK Signaling in LKB1-Deficient Lung Cancer. Mol. Cell *69*, 87-99.e7.

Kamarajugadda, S., Cai, Q., Chen, H., Nayak, S., Zhu, J., He, M., Jin, Y., Zhang, Y., Ai, L., Martin, S.S., et al. (2013). Manganese superoxide dismutase promotes anoikis resistance and tumor metastasis. Cell Death Dis. *4*, e504.

Kamynina, E.A., Lachenauer, E.R., DiRisio, A.C., Liebenthal, R.P., Field, M.S., and Stover, P.J. (2017). Arsenic trioxide targets MTHFD1 and SUMO-dependent nuclear de novo thymidylate biosynthesis. Proc. Natl. Acad. Sci. U. S. A. *114*, E2319–E2326.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Kim, Y.-I. (2005). Nutritional epigenetics: impact of folate deficiency on DNA methylation and colon cancer susceptibility. J. Nutr. *135*, 2703–2709.

Kolesnik, D.L., Pyaskovskaya, O.N., Boychuk, I. V, Dasyukevich, O.I., Melnikov, O.R., Tarasov, A.S., and Solyanik, G.I. (2015). Effect of dichloroacetate on Lewis lung carcinoma growth and metastasis. Exp. Oncol. *37*, 126–129.

Kuipers, E.J., Grady, W.M., Lieberman, D., Seufferlein, T., Sung, J.J., Boelens, P.G., van de Velde, C.J.H., and Watanabe, T. (2015). Colorectal cancer. Nat. Rev. Dis. Prim. *1*, 15065.

Lee, J., Park, J.R., Lee, H., Jang, S., Ryu, S.M., Kim, H., Kim, D., Jang, A., and Yang, S.R. (2018). l-carnosine induces apoptosis/cell cycle arrest via suppression of NF-κB/STAT1 pathway in HCT116 colorectal cancer cells. Vitr. Cell. Dev. Biol. - Anim. *54*, 505–512.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883.

Lehuédé, C., Dupuy, F., Rabinovitch, R., Jones, R.G., and Siegel, P.M. (2016). Metabolic Plasticity as a Determinant of Tumor Growth and Metastasis. Cancer Res. *76*, 5201–5208.

Li, A.E., Ito, H., Rovira, I.I., Kim, K.-S.S., Takeda, K., Yu, Z.-Y.Y., Ferrans, V.J., and Finkel, T. (1999). A role for reactive oxygen species in endothelial cell anoikis. Circ. Res. *85*, 304–310.

Lim, J.K.M., Delaidelli, A., Minaker, S.W., Zhang, H.F., Colovic, M., Yang, H., Negri, G.L., von Karstedt, S., Lockwood, W.W., Schaffer, P., et al. (2019). Cystine/glutamate antiporter xCT (SLC7A11) facilitates oncogenic RAS transformation by preserving intracellular redox balance. Proc. Natl. Acad. Sci. U. S. A. *116*, 9433–9442.

Liu, P., Ge, X., Ding, H., Jiang, H., Christensen, B.M., and Li, J. (2012). Role of glutamate decarboxylase-like protein 1 (GADL1) in taurine biosynthesis. J. Biol. Chem. *287*, 40898–40906.

Lo, H.C., and Zhang, X.H.-F. (2018). EMT in Metastasis: Finding the Right Balance. Dev. Cell *45*, 663–665.

Lo, M., Ling, V., Wang, Y.Z., and Gout, P.W. (2008). The xc- cystine/glutamate antiporter: a mediator of pancreatic cancer growth with a role in drug resistance. Br. J. Cancer *99*, 464–472.

Locasale, J.W. (2013). Serine, glycine and the one-carbon cycle: cancer metabolism is full circle. Nat. Rev. Cancer *13*, 572–583.

Locasale, J.W., and Cantley, L.C. (2011). Metabolic flux and the regulation of mammalian cell

growth. Cell Metab. *14*, 443–451.

Longo, N., Frigeni, M., and Pasquali, M. (2016). Carnitine transport and fatty acid oxidation. Biochim. Biophys. Acta *1863*, 2422–2435.

Lu, J. (2019). The Warburg metabolism fuels tumor metastasis. Cancer Metastasis Rev. *38*, 157–164.

Ma, M.-Z., Chen, G., Wang, P., Lu, W., Zhu, C., Song, M., Yang, J., Wen, S., Xu, R.-H., Hu, Y., et al. (2015). Xc- inhibitor sulfasalazine sensitizes colorectal cancer to cisplatin by a GSH-dependent mechanism. Cancer Lett. *368*, 88–96.

Macfarlane, A.J., Perry, C.A., McEntee, M.F., Lin, D.M., and Stover, P.J. (2011). Shmt1 heterozygosity impairs folate-dependent thymidylate synthesis capacity and modifies risk of Apc(min)-mediated intestinal cancer risk. Cancer Res. *71*, 2098–2107.

MacFarlane, A.J., Anderson, D.D., Flodby, P., Perry, C.A., Allen, R.H., Stabler, S.P., and Stover, P.J. (2011). Nuclear localization of de novo thymidylate biosynthesis pathway is required to prevent uracil accumulation in DNA. J. Biol. Chem. *286*, 44015–44022.

Mandal, P.K., Seiler, A., Perisic, T., Kölle, P., Canak, A.B., Förster, H., Weiss, N., Kremmer, E., Lieberman, M.W., Bannai, S., et al. (2010). System xc- and thioredoxin reductase 1 cooperatively rescue glutathione deficiency. J. Biol. Chem. *285*, 22244–22253.

Mazurek, S. (2011). Pyruvate kinase type M2: A key regulator of the metabolic budget system in tumor cells. Int. J. Biochem. Cell Biol. *43*, 969–980.

McDonald, E.R., de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. Cell *170*, 577-592.e10.

Meiser, J., Tumanov, S., Maddocks, O., Labuschagne, C.F., Athineos, D., Van Den Broek, N., Mackay, G.M., Gottlieb, E., Blyth, K., Vousden, K., et al. (2016). Serine one-carbon catabolism with formate overflow. Sci. Adv. *2*, 1–11.

Mittal, V. (2018). Epithelial Mesenchymal Transition in Tumor Metastasis. Annu. Rev. Pathol. *13*, 395–412.

Moyer, M.P., Manzano, L.A., Merriman, R.L., Stauffer, J.S., and Tanzer, L.R. (1996). NCM460, a normal human colon mucosal epithelial cell line. In Vitro Cell. Dev. Biol. Anim. *32*, 315–317.

Newman, A.C., and Maddocks, O.D.K. (2017). One-carbon metabolism in cancer. Br. J. Cancer *116*, 1499–1504.

Ning, X., Qi, H., Li, R., Li, Y., Jin, Y., McNutt, M.A., Liu, J., and Yin, Y. (2017). Discovery of novel naphthoquinone derivatives as inhibitors of the tumor cell specific M2 isoform of pyruvate kinase. Eur. J. Med. Chem. *138*, 343–352.

Nishimura, T., Nakata, A., Chen, X., Nishi, K., Meguro-Horike, M., Sasaki, S., Kita, K., Horike, S. ichi, Saitoh, K., Kato, K., et al. (2019). Cancer stem-like properties and gefitinib resistance are dependent on purine synthetic metabolism mediated by the mitochondrial enzyme MTHFD2. Oncogene *38*, 2464–2481.

Ortega, A.L., Carretero, J., Obrador, E., Gambini, J., Asensi, M., Rodilla, V., and Estrela, J.M. (2003). Tumor cytotoxicity by endothelial cells: Impairment of the mitochondrial system for glutathione uptake in mouse B16 melanoma cells that survive after in vitro interaction with the hepatic sinusoidal endothelium. J. Biol. Chem. *278*, 13888–13897.

Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010). What is flux balance analysis? Nat. Biotechnol. *28*, 245–248.

Oskarsson, T., Batlle, E., and Massagué, J. (2014). Metastatic stem cells: sources, niches, and vital pathways. Cell Stem Cell *14*, 306–321.

Padmanaban, V., Krol, I., Suhail, Y., Szczerba, B.M., Aceto, N., Bader, J.S., and Ewald, A.J. (2019). E-cadherin is required for metastasis in multiple models of breast cancer. Nature *324*, 297–314.

Pan, X., Lin, Z., Jiang, D., Yu, Y., Yang, D., Zhou, H., Zhan, D., Liu, S., Peng, G., Chen, Z., et al. (2019). Erastin decreases radioresistance of NSCLC cells partially by inducing GPX4-mediated ferroptosis. Oncol. Lett. *17*, 3001–3008.

Paone, A., Marani, M., Fiascarelli, A., Rinaldo, S., Giardina, G., Contestabile, R., Paiardini, A., and Cutruzzolà, F. (2014). SHMT1 knockdown induces apoptosis in lung cancer cells by causing uracil misincorporation. Cell Death Dis. *5*, e1525.

Pascual, G., Domínguez, D., and Benitah, S.A. (2018). The contributions of cancer cell metabolism to metastasis. Dis. Model. Mech. *11*.

Pikman, Y., Puissant, A., Alexe, G., Furman, A., Chen, L.M., Frumm, S.M., Ross, L., Fenouille, N., Bassil, C.F., Lewis, C.A., et al. (2016). Targeting MTHFD2 in acute myeloid leukemia. J. Exp. Med. *213*, 1285–1306.

Porter, T.G., and Martin, D.L. (1986). Non-steady-state kinetics of brain glutamate decarboxylase resulting from interconversion of the apo- and holoenzyme. Biochim. Biophys. Acta *874*, 235–244.

Pratapa, A., Balachandran, S., and Raman, K. (2015). Fast-SL: An efficient algorithm to identify synthetic lethal sets in metabolic networks. Bioinformatics *31*, 3299–3305.

Provenzani, A., Fronza, R., Loreni, F., Pascale, A., Amadio, M., and Quattrone, A. (2006). Global alterations in mRNA polysomal recruitment in a cell model of colorectal cancer progression to metastasis. Carcinogenesis *27*, 1323–1333.

Reichert, M., Bakir, B., Moreira, L., Pitarresi, J.R., Feldmann, K., Simon, L., Suzuki, K., Maddipati, R., Rhim, A.D., Schlitter, A.M., et al. (2018). Regulation of Epithelial Plasticity Determines Metastatic Organotropism in Pancreatic Cancer. Dev. Cell *45*, 696-711.e8.

Reimers, A.M., and Reimers, A.C. (2016). The steady-state assumption in oscillating and growing systems. J. Theor. Biol. *406*, 176–186.

Rodrigues, M.F., Obre, E., de Melo, F.H.M., Santos, G.C., Galina, A., Jasiulionis, M.G., Rossignol, R., Rumjanek, F.D., and Amoêdo, N.D. (2016). Enhanced OXPHOS, glutaminolysis and β-oxidation constitute the metastatic phenotype of melanoma cells. Biochem. J. *473*, 703–715.

Rohani, N., Hao, L., Alexis, M.S., Joughin, B.A., Krismer, K., Moufarrej, M.N., Soltis, A.R., Lauffenburger, D.A., Yaffe, M.B., Burge, C.B., et al. (2019). Acidification of Tumor at Stromal Boundaries Drives Transcriptome Alterations Associated with Aggressive Phenotypes. Cancer Res. *79*, 1952–1966.

Rozhin, J., Sameni, M., Ziegler, G., Sloane, B., Cornnell, H.H., Ibrahim-Hashim, A., Bailey, K., Balagurunathan, Y., Rothberg, J.M., Sloane, B.F., et al. (2013). Pericellular pH Affects Distribution and Secretion of Cathepsin B in Malignant Cells. Cancer Res. *54*, 6517–6525.

Saitoh, M. (2018). Involvement of partial EMT in cancer progression. J. Biochem. *164*, 257–264.

Sale, C., Saunders, B., and Harris, R.C. (2010). Effect of beta-alanine supplementation on muscle carnosine concentrations and exercise performance. Amino Acids *39*, 321–333.

Salvarrey, M.S., and Cazzulo, J.J. (1982). Citrate synthase from Crithidia fasciculata: Inhibition by adenine nucleotides and suramin. Comp. Biochem. Physiol. Part B Comp. Biochem. *72*, 165–168.

Sappington, D.R., Siegel, E.R., Hiatt, G., Desai, A., Penney, R.B., Jamshidi-Parsian, A., Griffin, R.J., and Boysen, G. (2016). Glutamine drives glutathione synthesis and contributes to radiation sensitivity of A549 and H460 lung cancer cell lines. Biochim. Biophys. Acta *1860*, 836–843.

Sato, M., Kusumi, R., Hamashima, S., Kobayashi, S., Sasaki, S., Komiyama, Y., Izumikawa, T., Conrad, M., Bannai, S., and Sato, H. (2018). The ferroptosis inducer erastin irreversibly inhibits system xc- and synergizes with cisplatin to increase cisplatin's cytotoxicity in cancer cells. Sci. Rep. *8*, 1–9.

Savaskan, N.E., Heckel, A., Hahnen, E., Engelhorn, T., Doerfler, A., Ganslandt, O., Nimsky, C., Buchfelder, M., and Eyüpoglu, I.Y. (2008). Small interfering RNA–mediated xCT silencing in gliomas inhibits neurodegeneration and alleviates brain edema. Nat. Med. *14*, 629–632.

Sbodio, J.I., Snyder, S.H., and Paul, B.D. (2019). Regulators of the transsulfuration pathway. Br. J. Pharmacol. *176*, 583–593.

Schmidt, A., Wu, H., MacKenzie, R.E., Chen, V.J., Bewly, J.R., Ray, J.E., Toth, J.E., and Cygler, M. (2000). Structures of three inhibitor complexes provide insight into the reaction mechanism of the human methylenetetrahydrofolate dehydrogenase/cyclohydrolase. Biochemistry *39*, 6325–6335.

Schmidt, B.J., Ebrahim, A., Metz, T.O., Adkins, J.N., Palsson, B.Ø., and Hyduke, D.R. (2013). GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics *29*, 2900–2908.

Schnoor, M. (2015). E-cadherin Is Important for the Maintenance of Intestinal Epithelial Homeostasis Under Basal and Inflammatory Conditions. Dig. Dis. Sci. *60*, 816–818.

Sdelci, S., Rendeiro, A.F., Rathert, P., You, W., Lin, J.-M.G., Ringler, A., Hofstätter, G., Moll, H.P., Gürtl, B., Farlik, M., et al. (2019). MTHFD1 interaction with BRD4 links folate metabolism to transcriptional regulation. Nat. Genet. *51*, 990–998.

Segre, D., Vitkup, D., and Church, G.M. (2002). Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. *99*, 15112–15117.

Selivanov, V.A., Benito, A.A., Miranda, A., Aguilar, E., Polat, I.H., Centelles, J.J., Jayaraman, A., Lee, P.W.N.N., Marin, S., and Cascante, M. (2017). MIDcor, an R-program for deciphering mass interferences in mass spectra of metabolites enriched in stable isotopes. BMC Bioinformatics *18*, 88.

Sheffer, M., Bacolod, M.D., Zuk, O., Giardina, S.F., Pincas, H., Barany, F., Paty, P.B., Gerald, W.L., Notterman, D.A., and Domany, E. (2009). Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proc. Natl. Acad. Sci. U. S. A. *106*, 7131–7136.

Sheikh, K., Förster, J., and Nielsen, L.K. (2005). Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus. Biotechnol. Prog. *21*, 112–121.

Shelton, L.M., Huysentruyt, L.C., and Seyfried, T.N. (2010). Glutamine targeting inhibits

systemic metastasis in the VM-M3 murine tumor model. Int. J. Cancer *127*, 2478–2485.

Shen, Y., Yang, J., Li, J., Shi, X., Ouyang, L., Tian, Y., and Lu, J. (2014). Carnosine inhibits the proliferation of human gastric cancer SGC-7901 cells through both of the mitochondrial respiration and glycolysis pathways. PLoS One *9*.

Sheng, S.L., Liu, J.J., Dai, Y.H., Sun, X.G., Xiong, X.P., and Huang, G. (2012). Knockdown of lactate dehydrogenase A suppresses tumor growth and metastasis of human hepatocellular carcinoma. FEBS J. *279*, 3898–3910.

Shibue, T., and Weinberg, R.A. (2017). EMT, CSCs, and drug resistance: The mechanistic link and clinical implications. Nat. Rev. Clin. Oncol. *14*, 611–629.

Shin, S.-S., Jeong, B.-S., Wall, B.A., Li, J., Shan, N.L., Wen, Y., Goydos, J.S., and Chen, S. (2018). Participation of xCT in melanoma cell proliferation in vitro and tumorigenesis in vivo. Oncogenesis *7*, 86.

Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R., and Ruppin, E. (2011). Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect. PLoS Comput. Biol. *7*, 1–8.

Sottnik, J.L., Lori, J.C., Rose, B.J., and Thamm, D.H. (2011). Glycolysis inhibition by 2-deoxy-d-glucose reverts the metastatic phenotype in vitro and in vivo. Clin. Exp. Metastasis *28*, 865–875.

States, B., and Segal, S. (1973). Interrelationship of glutathione–cystine transhydrogenase and glutathione reductase in developing rat intestine. Biochem. J. *132*, 623–631.

Stein, C.A., LaRocca, R. V, Thomas, R., McAtee, N., and Myers, C.E. (1989). Suramin: an anticancer drug with a unique mechanism of action. J. Clin. Oncol. *7*, 499–508.

Sun, R.C., Fadia, M., Dahlstrom, J.E., Parish, C.R., Board, P.G., and Blackburn, A.C. (2010). Reversal of the glycolytic phenotype by dichloroacetate inhibits metastatic breast cancer cell growth in vitro and in vivo. Breast Cancer Res. Treat. *120*, 253–260.

Swainston, N., Smallbone, K., Hefzi, H., Dobson, P.D., Brewer, J., Hanscho, M., Zielinski, D.C., Ang, K.S., Gardiner, N.J., Gutierrez, J.M., et al. (2016). Recon 2.2: from reconstruction to model of human metabolism. Metabolomics *12*.

Tarrado-Castellarnau, M., de Atauri, P., and Cascante, M. (2016). Oncogenic regulation of tumor metabolic reprogramming. Oncotarget *7*, 62726–62753.

Tarrado-Castellarnau, M., de Atauri, P., Tarragó-Celada, J., Perarnau, J., Yuneva, M., Thomson, T.M., and Cascante, M. (2017). De novo MYC addiction as an adaptive response of cancer cells to CDK4/6 inhibition. Mol. Syst. Biol. *13*, 940.

Tedeschi, P.M., Markert, E.K., Gounder, M., Lin, H., Dvorzhinski, D., Dolfi, S.C., Chan, L.L.Y., Qiu, J., DiPaola, R.S., Hirshfield, K.M., et al. (2013). Contribution of serine, folate and glycine metabolism to the ATP, NADPH and purine requirements of cancer cells. Cell Death Dis. *4*, e877-12.

Thomson, T.M.M., Balcells, C., Cascante, M., Thomson, T.M.M., Balcells, C., and Cascante, M. (2019). Metabolic Plasticity and Epithelial-Mesenchymal Transition. J. Clin. Med. *8*, 967.

Tibbetts, A.S., and Appling, D.R. (2010). Compartmentalization of Mammalian Folate-Mediated One-Carbon Metabolism. Annu. Rev. Nutr. *30*, 57–81.

Tietze, F., Bradley, K.H., and Schulman, J.D. (1972). Enzymic reduction of cystine by subcellular

fractions of cultured and peripheral leukocytes from normal and cystinotic individuals. Pediatr. Res. *6*, 649–658.

Todaro, M., Gaggianesi, M., Catalano, V., Benfante, A., Iovino, F., Biffoni, M., Apuzzo, T., Sperduti, I., Volpe, S., Cocorullo, G., et al. (2014). CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. Cell Stem Cell *14*, 342–356.

Traverso, N., Ricciarelli, R., Nitti, M., Marengo, B., Furfaro, A.L., Pronzato, M.A., Marinari, U.M., and Domenicotti, C. (2013). Role of glutathione in cancer progression and chemoresistance. Oxid. Med. Cell. Longev. *2013*, 972913.

Urosevic, J., Garcia-Albéniz, X., Planet, E., Real, S., Céspedes, M.V., Guiu, M., Fernandez, E., Bellmunt, A., Gawrzak, S., Pavlovic, M., et al. (2014). Colon cancer cells colonize the lung from established liver metastases through p38 MAPK signalling and PTHLH. Nat. Cell Biol. *16*, 685–694.

Waltenberger, J., Mayr, U., Frank, H., and Hombach, V. (1996). Suramin is a Potent Inhibitor of Vascular Endothelial Growth Factor. A Contribution to the Molecular Basis of its Antiangiogenic Action. J. Mol. Cell. Cardiol. *28*, 1523–1529.

Whillier, S., Raftos, J.E., Chapman, B., and Kuchel, P.W. (2009). Role of N-acetylcysteine and cystine in glutathione synthesis in human erythrocytes. Redox Rep. *14*, 115–124.

Xu, M., You, J., Hou, N., Zhang, H., Chen, G., and Yang, Z. (2010). Mitochondrial enzymes and citrate transporter contribute to the aluminium-induced citrate secretion from soybean (Glycine max) roots. Funct. Plant Biol. *37*, 285.

Yamamoto, H., Murata, K., Fukunaga, M., Ohnishi, T., Noura, S., Miyake, Y., Kato, T., Ohtsuka, M., Nakamura, Y., Takemasa, I., et al. (2016). Micrometastasis Volume in Lymph Nodes Determines Disease Recurrence Rate of Stage II Colorectal Cancer: A Prospective Multicenter Trial. Clin. Cancer Res. *22*, 3201–3208.

Yan, M., Xu, Q., Zhang, P., Zhou, X., Zhang, Z., and Chen, W. (2010). Correlation of NF-kappaB signal pathway with tumor metastasis of human head and neck squamous cell carcinoma. BMC Cancer *10*, 437.

Yang, J., Nie, J., Ma, X., Wei, Y., Peng, Y., and Wei, X. (2019). Targeting PI3K in cancer: mechanisms and advances in clinical trials. Mol. Cancer *18*, 26.

Yang, L., Moss, T., Mangala, L.S., Marini, J., Zhao, H., Wahlig, S., Armaiz-Pena, G., Jiang, D., Achreja, A., Win, J., et al. (2014). Metabolic shifts toward glutamine regulate tumor growth, invasion and bioenergetics in ovarian cancer. Mol. Syst. Biol. *10*, 728.

Yang, L., Venneti, S., and Nagrath, D. (2017). Glutaminolysis: A Hallmark of Cancer Metabolism. Annu. Rev. Biomed. Eng. *19*, 163–194.

Ye, J., Mancuso, A., Tong, X., Ward, P.S., Fan, J., Rabinowitz, J.D., and Thompson, C.B. (2012). Pyruvate kinase M2 promotes de novo serine synthesis to sustain mTORC1 activity and cell proliferation. Proc. Natl. Acad. Sci. *109*, 6904–6909.

Yeh, J.J., Routh, E.D., Rubinas, T., Peacock, J., Martin, T.D., Shen, X.J., Sandler, R.S., Kim, H.J., Keku, T.O., and Der, C.J. (2009). KRAS/BRAF mutation status and ERK1/2 activation as biomarkers for MEK1/2 inhibitor therapy in colorectal cancer. Mol. Cancer Ther. *8*, 834–843.

Yin, S., Cheryan, V.T., Xu, L., Rishi, A.K., and Reddy, K.B. (2017). Myc mediates cancer stem-like cells and EMT changes in triple negative breast cancers cells. PLoS One *12*, e0183578.

Young, J.D. (2014). INCA: a computational platform for isotopically non-stationary metabolic

flux analysis. Bioinformatics *30*, 1333–1335.

Zheng, X., Carstens, J.L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., Wu, C.-C., LeBleu, V.S., and Kalluri, R. (2015). Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature *527*, 525–530.

9. Appendix

9 Appendix

Cancer cell metabolism as new targets for novel designed therapies

Igor Mar n De Mas, Esther Aguilar, Anusha ayaraman, Ibrahim olat, Alfonso Mart n- ernab , Rohit harat, Carles oguet, Enric Mil , al zs app, osep Centelles, Marta Cascante.

The importance of post-translational modifications in systems biology approaches to identify therapeutic targets in cancer metabolism

Alfonso Mart n- ernab , Cristina alcells, osep Tarrag -Celada, Carles oguet, Sandrine ourgoin-Voillard, Michel Seve, Marta Cascante.

Tracing metabolic fluxes using mass spectrometry: stable isotope-resolved metabolomics in health and disease

Cristina alcells, Carles oguet, osep Tarrag -Celada, edro de Atauri, Silvia Marin, Marta Cascante.

Metabolomics in Systems Medicine: an overview of methods and applications

Effrosyni Kara itsou, Carles oguet, edro de Atauri, Kim Kultima, ayam Emami Khoonsari, Vitor A Martins dos Santos, Edoardo Saccenti, Antonio Rosato, Marta Cascante.

Review
Mini-Focus Issue: Systems biology

For reprint orders, please contact reprints@future-science.com

Future Medicinal Chemistry

Cancer cell metabolism as new targets for novel designed therapies

Metabolic processes are altered in cancer cells, which obtain advantages from this metabolic reprogramming in terms of energy production and synthesis of biomolecules that sustain their uncontrolled proliferation. Due to the conceptual progresses in the last decade, metabolic reprogramming was recently included as one of the new hallmarks of cancer. The advent of high-throughput technologies to amass an abundance of omic data, together with the development of new computational methods that allow the integration and analysis of omic data by using genome-scale reconstructions of human metabolism, have increased and accelerated the discovery and development of anticancer drugs and tumor-specific metabolic biomarkers. Here we review and discuss the latest advances in the context of metabolic reprogramming and the future in cancer research.

Igor Marín de Mas[‡,1,2], Esther Aguilar[‡,1], Anusha Jayaraman[‡,1], Ibrahim H Polat[1], Alfonso Martín-Bernabé[1], Rohit Bharat[1], Carles Foguet[1], Enric Milà[1], Balázs Papp[2], Josep J Centelles[1] & Marta Cascante*[,1]

[1]Department of Biochemistry & Molecular Biology, Faculty of Biology, IBUB, Universitat de Barcelona & Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Unit Associated with CSIC, Diagonal 643, E-08028-Barcelona, Spain

[2]Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, Temesvári krt. 62, H-6726 Szeged, Hungary

*Author for correspondence:
Tel.: +34 934021593
Fax: +34 934021559
martacascante@ub.edu

‡Authors contributed equally

Cancer is still one of the major causes of death worldwide and the statistics are devastating. According to the WHO the global burden of cancer has risen to 14.1 million new cases and 8.2 million cancer deaths in 2012 and the estimates predict that it could increase in its global incidence [1].

It was proposed 15 years ago by Hanahan and Weinberg that cancer development relies on the following basic biological capabilities, known as the 'hallmarks of cancer' that are acquired during the multistep process of tumor development: the capability to sustain proliferative signaling, resistance to cell death, evasion of growth suppression, ability of replicative immortality, tumor-promoting inflammation, genome instability and mutation, induction of angiogenesis and activation of invasion and metastasis. Owing to conceptual progress in the last decade, two new hallmarks, metabolic reprogramming and evasion of immune destruction, have been identified (Figure 1) [2].

Nowadays, it is widely recognized that metabolic reprogramming is essential to sustain tumor progression. Several metabolic adaptations described in cancer cells, such as the metabolization of glucose to lactate in the presence of oxygen (Warburg effect), are quite common among different cancer types. These changes are promoted by genetic and epigenetic alterations producing mutations or alterations in the expression of key metabolic enzymes that modify flux distributions in metabolic networks, providing advantages to cancer cells in terms of energy production and synthesis of biomolecules [3,4].

Understanding the mechanisms that trigger metabolic reprogramming in cancer cells and its role in tumoral progression is crucial, not only from a biological but also from a clinical stance, since this can be the basis towards improving existing cancer therapies or developing new ones.

In this review, we discuss the role of: the crosstalk between oncogenic signaling pathways and metabolism; the influence of non-genetic factors, such as tumor microenvironment, on metabolic reprogramming of cancer and stromal cells; the changes in isoenzymes patterns as potential therapeutic targets; and the new computational tools used by a systems biology approach in drug-target and biomarker discovery based on genome-scale metabolic models (GSMMs). Finally, we also discuss the future challenges in

FUTURE SCIENCE part of fsg

123 Duggal R, Minev B, Geissinger U *et al.* Biotherapeutic approaches to target cancer stem cells. *J. Stem Cells* 8(3–4), 135–149 (2013).

124 Philips MR, Cox AD. Geranylgeranyltransferase I as a target for anti-cancer drugs. *J. Clin. Invest.* 117(5), 1223–1225 (2007).

125 Dudakovic A, Tong H, Hohl R. Geranylgeranyl diphosphate depletion inhibits breast cancer cell migration. *Invest. New Drugs* 29(5), 912–920 (2011).

126 Hebar A, Valent P, Selzer E. The impact of molecular targets in cancer drug development: major hurdles and future strategies. *Expert Rev. Clin. Pharmacol.* 6(1), 23–34 (2013).

127 Conde-Pueyo N, Munteanu A, Solé RV, Rodríguez-Caso C. Human synthetic lethal inference as potential anti-cancer target gene detection. *BMC Syst. Biol.* 3(1), 116 (2009).

128 Barbash DA, Lorigan JG. Lethality in *Drosophilamelanogast er*/*Drosophilasimulans* species hybrids is not associated with substantial transcriptional misregulation. *J. Exp. Zool. Part B Mol. Dev. Evol.* 308(1), 74–84 (2007).

129 Ren S, Zeng B, Qian X. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. *BMC Bioinformatics* 14(Suppl. 2), S17 (2013).

130 Frezza C, Zheng L, Folger O *et al.* Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* 477(7363), 225–228 (2011).

131 Zhao Y, Butler EB, Tan M. Targeting cellular metabolism to improve cancer therapeutics. *Cell Death Dis.* 4, e532 (2013).

132 Vaupel P. Tumor microenvironmental physiology and its implications for radiation oncology. *Sem. Radiat. Oncol.* 14(3), 198–206 (2004).

133 Tannock IF, Lee CM, Tunggal JK, Cowan DS, Egorin MJ. Limited penetration of anticancer drugs through tumor tissue a potential cause of resistance of solid tumors to chemotherapy. *Clin. Cancer Res.* 8(3), 878–884 (2002).

134 Olive KP, Jacobetz MA, Davidson CJ *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* 324(5933), 1457–1461 (2009).

135 Hulit J, Howell A, Gandara R, Sartini M, Arafat H, Bevilacqua G. Creating a tumor-resistant microenvironment. *Cell Cycle* 12(3), 480–490 (2013).

136 Ye C, Zou W, Xu N, Liu L. Metabolic model reconstruction and analysis of an artificial microbial ecosystem for vitamin C production. *J. Biotech.* 182, 61–67 (2014).

137 Peinado H, Olmeda D, Cano A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat. Rev. Cancer* 7(6), 415–428 (2007).

138 Celià-Terrassa T, Meca-Cortés Ó, Mateo F *et al.* Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *J. Clin. Invest.* 122(5), 1849 (2012).

139 Korpal M, Ell BJ, Buffa FM *et al.* Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat. Med.* 17(9), 1101–1108 (2011).

140 Ocaña OH, Córcoles R, Fabra Á *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* 22(6), 709–724 (2012).

141 Rundqvist H, Johnson RS. Hypoxia and metastasis in breast cancer. In: *Diverse Effects of Hypoxia on Tumor Progression.* Simon MC (Ed.) Springer, Berlin/Heidelberg, Germany, 121–139 (2010).

142 Nagy JA, Chang S-H, Shih S-C, Dvorak AM, Dvorak HF. Heterogeneity of the tumor vasculature. *Semin. Thromb. Hemost.* 36(3), 321–331 (2010).

143 Mazzone M, Dettori D, Leite De Oliveira R *et al.* Heterozygous deficiency of PHD2 restores tumor oxygenation and inhibits metastasis via endothelial normalization. *Cell* 136(5), 839–851 (2009).

144 Rolny C, Mazzone M, Tugues S *et al.* HRG inhibits tumor growth and metastasis by inducing macrophage polarization and vessel normalization through downregulation of PlGF. *Cancer Cell* 19(1), 31–44 (2011).

145 Yizhak K, Gabay O, Cohen H, Ruppin E. Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat. Commun.* 4, 2632 (2013).

146 De Mas IM, Selivanov VA, Marin S *et al.* Compartmentation of glycogen metabolism revealed from 13C isotopologue distributions. *BMC Syst. Biol.* 5(1), 175 (2011).

147 Hassan S, Buchanan M, Jahan K *et al.* CXCR4 peptide antagonist inhibits primary breast tumor growth, metastasis and enhances the efficacy of anti-VEGF treatment or docetaxel in a transgenic mouse model. *Int. J. Cancer* 129(1), 225–232 (2011).

148 Coenegrachts L, Maes C, Torrekens S *et al.* Anti-placental growth factor reduces bone metastasis by blocking tumor cell engraftment and osteoclast differentiation. *Cancer Res.* 70(16), 6537–6547 (2010).

149 Hiratsuka S, Duda DG, Huang Y *et al.* CXC receptor type 4 promotes metastasis by activating p38 mitogen-activated protein kinase in myeloid differentiation antigen (Gr-1)-positive cells. *Proc. Natl Acad. Sci. USA* 108(1), 302–307 (2011).

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

The importance of post-translational modifications in systems biology approaches to identify therapeutic targets in cancer metabolism

Alfonso Martín-Bernabé[1,2,3,4,5], Cristina Balcells[1,5], Josep Tarragó-Celada[1,5], Carles Foguet[1,5], Sandrine Bourgoin-Voillard[2,3,4], Michel Seve[2,3,4] and Marta Cascante[1]

Abstract

Cancer metabolism is reprogrammed to fulfill the needs of proliferation and migration, which is accomplished through different levels of regulation. In recent years, new advances in protein post-translational modifications (PTMs) research have revealed a complex layer of regulatory mechanisms through which PTMs control cell signaling and metabolic pathways, contributing to the diverse metabolic phenotypes found in cancer. Despite the efficacy of current modeling approaches to study cancer metabolism they still lack the capacity to integrate PTMs in their predictions. Here we will review the importance of PTMs in cancer metabolic reprogramming and suggest ways in which computational predictions could be enhanced through the integration of PTMs.

Addresses

[1] Department of Biochemistry and Molecular Biomedicine, Universitat de Barcelona and Institute of Biomedicine of Universitat de Barcelona (IBUB), Barcelona, Spain

[2] Univ. Grenoble Alpes, LBFA et BEeSy, PROMETHEE Proteomic Platform, Grenoble, France

[3] Inserm, U1055, PROMETHEE Proteomic Platform, Grenoble, France

[4] CHU de Grenoble, Institut de Biologie et de Pathologie, PROMETHEE Proteomic Platform, Grenoble, France

Corresponding author: Cascante, Marta (martacascante@ub.edu)

Email addresses: alf.martin@gmail.com (A. Martín-Bernabé), crisgatsu@gmail.com (C. Balcells), jtarragocelada@ub.edu (J. Tarragó-Celada), cfoguet@ub.edu (C. Foguet), sandrine.bourgoin@univ-grenoble-alpes.fr (S. Bourgoin-Voillard), michel.seve@univ-grenoble-alpes.fr (M. Seve)

[5] These authors contributed equally to this work.

Current Opinion in Systems Biology 2017, **3**:161–169

This review comes from a themed issue on **Clinical and translational systems biology (2017)**

Edited by **Jesper Tegnér** and **David Gomez-Cabrero**

For a complete overview see the Issue and the Editorial

Available online 19 May 2017

http://dx.doi.org/10.1016/j.coisb.2017.05.011

2452-3100/© 2017 Elsevier Ltd. All rights reserved.

Introduction

Cancer is a complex and heterogeneous disease, which involves alteration of multiple biological processes. Some of these alterations are geared towards metabolic reprogramming, which provides advantages to cancer cells in terms of energy production and synthesis of biomolecules and is essential for tumor progression. Understanding metabolic reprogramming in heterogeneous tumor cell populations is key to identify metabolic vulnerabilities in cancer that can be exploited in therapy [1].

Metabolism is a complex network of biochemical reactions that requires a systemic view. Computational models have emerged as platforms to integrate multi-omics data (genomics, transcriptomics, proteomics, metabolomics, etc.), and therefore to understand the underlying metabolic phenotype. However current modeling approaches have a limited capacity to integrate modifications that occur at a post-translational level hence limiting their usefulness at analyzing cancer metabolic reprogramming.

Here, we review recent studies regarding the role of post-translational modifications (PTMs) in cancer metabolic reprogramming, mainly focusing on central carbon metabolism. In this context, we assess how computational predictions through genome-scale metabolic models (GSMMs) and kinetic models could be enhanced through the integration of PTMs.

Metabolic reprogramming in cancer

The most common metabolic adaptation in cancer is the Warburg effect, consisting of high glucose uptake, glycolytic activity, and lactate production even under aerobic conditions [2]. However, metabolic reprogramming is not limited to the Warburg effect and usually involves the enhancement of other essential metabolic pathways including pentose phosphate pathway (PPP), glutaminolysis, and amino acid and lipid metabolisms [3].

Tumor cells display metabolic flexibility, which allows them to undergo metabolic switches and use different energy and carbon sources depending on their

42. Nguyen TL, Durán RV: **Prolyl hydroxylase domain enzymes and their role in cell signaling and cancer metabolism**. *Int J Biochem Cell Biol* 2016, **80**:71–80.

43. Luo W, Hu H, Chang R, Zhong J, Knabel M, O'Meally R, Cole RN, Pandey A, Semenza GL: **Pyruvate kinase M2 is a PHD3-stimulated coactivator for hypoxia-inducible factor 1**. *Cell* 2011, **145**:732–744.

44. Kikuchi D, Minamishima YA, Nakayama K: **Prolyl-hydroxylase PHD3 interacts with pyruvate dehydrogenase (PDH)-E1β and regulates the cellular PDH activity**. *Biochem Biophys Res Commun* 2014, **451**:288–294.

45. German NJ, Yoon H, Yusuf RZ, Scadden DT, Kaelin WG, Haigis MC: **PHD3 loss in cancer enables metabolic reliance on fatty acid oxidation via deactivation of ACC2**. *Mol Cell* 2016, **63**:1006–1020.

46. Theurillat JP, Udeshi ND, Errington WJ, Svinkina T, Baca SC, Pop M, Wild PJ, Blattner M, Groner AC, Rubin MA, Moch H, Privé GG, Carr SA, Garraway LA: **Prostate cancer. Ubiquitylome analysis identifies dysregulation of effector substrates in SPOP-mutant prostate cancer**. *Science* 2014 Oct 3, **346**:85–89.

47. Xu YM, Wang HJ, Chen F, Guo WH, Wang YY, Li HY, Tang JH, Ding Y, Shen YC, Li M, *et al.*: **HRD1 suppresses the growth and metastasis of breast cancer cells by promoting IGF-1R degradation**. *Oncotarget* 2015, **6**:42854–42867.

48. Han C, Yang L, Choi HH, Baddour J, Achreja A, Liu Y, Li Y, Li J, Wan G, Huang C, *et al.*: **Amplification of USP13 drives ovarian cancer metabolism**. *Nat Commun* 2016, **7**:13525.

49. • Liu K, Li F, Han H, Chen Y, Mao Z, Luo J, Zhao Y, Zheng B, Gu W, Zhao W: **Parkin regulates the activity of pyruvate kinase M2**. *J Biol Chem* 2016, **291**:10307–10317.

This work elucidated the role of Parkin in tumor cell metabolism by ubiquitination of PKM2.

50. Bian X, Chen H, Yang P, Li Y, Zhang F, Zhang J, Wang W, Zhao W, Zhang S, Chen Q, *et al.*: **Nur77 suppresses hepatocellular carcinoma via switching glucose metabolism toward gluconeogenesis through attenuating phosphoenolpyruvate carboxykinase sumoylation**. *Nat Commun* 2017, **8**:1–14.

51. Tan M, Peng C, Anderson KA, Chhoy P, Xie Z, Dai L, Park J, Chen Y, Huang H, Zhang Y, *et al.*: **Lysine glutarylation is a protein posttranslational modification regulated by SIRT5**. *Cell Metab* 2014, **19**:605–617.

52. Hirschey MD, Zhao Y: **Metabolic regulation by lysine malonylation, succinylation, and glutarylation**. *Mol Cell Proteomics* 2015, **14**:2308–2315.

53. Du Y, Cai T, Li T, Xue P, Zhou B, He X, Wei P, Liu P, Yang F, Wei T: **Lysine malonylation is elevated in type 2 diabetic mouse models and enriched in metabolic associated proteins**. *Mol Cell Proteomics* 2015, **14**:227–236.

54. Xu H, Chen X, Xu X, Shi R, Suo S, Cheng K, Zheng Z, Wang M, Wang L, Zhao Y, *et al.*: **Lysine acetylation and succinylation in HeLa cells and their essential roles in response to UV-induced stress**. *Sci Rep* 2016, **6**:30212.

55. Zhou L, Wang F, Sun R, Chen X, Zhang M, Xu Q, Wang Y, Wang S, Xiong Y, Guan K: **SIRT5 promotes IDH2 desuccinylation and G6PD deglutarylation to enhance cellular antioxidant defense**. *EMBO Rep* 2016, **36**:1–12.

56. Marín de Mas I, Aguilar E, Jayaraman A, Polat IH, Martín-Bernabé A, Bharat R, Foguet C, Milà E, Papp B, Centelles JC, *et al.*: **Cancer cell metabolism as new targets for novel designed therapies**. *Future Med Chem* 2014, **6**:1791–1810.

57. •• Megchelenbrink W, Katzir R, Lu X, Ruppin E, Notebaart R a: **Synthetic dosage lethality in the human metabolic network is highly predictive of tumor growth and cancer patient survival**. *Proc Natl Acad Sci U S A* 2015, **112**:12217–12222.

This study presents a method that can sistematically identify metabolic vulnerabilities specific for cancer cells overexpressing a given metabolic gene.

58. Yizhak K, Chaneton B, Gottlieb E, Ruppin E: **Modeling cancer metabolism on a genome scale**. *Mol Syst Biol* 2015, **11**. 817–817.

59. König M, Bulik S, Holzhütter H-G: **Quantifying the contribution of the liver to glucose homeostasis: a detailed kinetic model of human hepatic glucose metabolism**. *PLoS Comput Biol* 2012, **8**:e1002577.

60. Stanford NJ, Lubitz T, Smallbone K, Klipp E, Mendes P, Liebermeister W: **Systematic construction of kinetic models from genome-scale metabolic networks**. *PLoS One* 2013, **8**.

61. Khodayari A, Maranas CD: **A genome-scale Escherichia coli kinetic metabolic model satisfying flux data for multiple mutant strains**. *Nat Commun* 2016, **7**:1–12.

Trends in Analytical Chemistry xxx (xxxx) xxx

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Tracing metabolic fluxes using mass spectrometry: Stable isotope-resolved metabolomics in health and disease

Cristina Balcells [a, b, 1], Carles Foguet [a, b, c, 1], Josep Tarragó-Celada [a, b, 1], Pedro de Atauri [a, b, c], Silvia Marin [a, b, c, **], Marta Cascante [a, b, c, *]

[a] Department of Biochemistry and Molecular Biomedicine, Faculty of Biology, Universitat de Barcelona, Diagonal, 643, 08028 Barcelona, Spain
[b] Institute of Biomedicine of Universitat de Barcelona (IBUB), Barcelona, Spain
[c] Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Instituto de Salud Carlos III (ISCIII), Madrid, Spain

ARTICLE INFO

Article history:
Available online xxx

Keywords:
SIRM
MFA
MS
Metabolic fluxes
Metabolic regulation
Cancer metabolism
Metabolic diseases

ABSTRACT

Tracing metabolic fluxes, defined as the reaction and transport rates in living cells, is essential to characterize metabolic phenotypes. One of the most informative methods to predict fluxes is stable isotope-resolved metabolomics (SIRM). In SIRM, a biological system is fed with substrates labeled with stable heavy isotopes. This isotopic label propagates along metabolic pathways and is incorporated into metabolites. After incubation, metabolites are extracted, and the incorporation of the isotopic label is quantified with isotope-sensitive analytical techniques, either mass spectrometry (MS) or nuclear magnetic resonance (NMR). Here we review the most suitable and widely-used MS platforms and methodologies for SIRM. We also provide an overview of state of the art in the analysis of SIRM data to trace metabolic fluxes, covering both local flux predictions and network-wide flux analysis. Finally, we highlight the role of SIRM in shaping our current understanding of metabolism in both health and pathological conditions.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Metabolic fluxes, defined as the reaction and transport rates in living cells, are a close reflection of the metabolic phenotype [1]. Hence, tracing metabolic fluxes is key to understanding the mechanisms of metabolic regulation in both health and disease. Extracellular fluxes (rates of uptake and secretion of metabolites) can be quantified by measuring changes in metabolite concentrations in the extracellular media [2]. However, intracellular fluxes cannot be directly estimated from measurements of intracellular metabolite concentrations as they are not informative of the fluxes leading to and originating from any given metabolite [3–5]. This limitation has led to the development of stable isotope-resolved metabolomics (SIRM).

In SIRM, a biological system is fed with one or more metabolic substrates labeled with stable heavy isotopes (^{13}C, ^{15}N, ^{2}H, ^{18}O, etc.). These labeled molecules, usually referred to as tracers, are metabolized by the system of interest through different metabolic pathways. An isotopic label propagates in a time, flux and pathway-dependent manner, generating characteristic labeling patterns in metabolites, which can provide information about the fluxes through pathways leading to such metabolites [1,2].

The propagation of a label from tracers to metabolites is quantified by isotope-sensitive analytical techniques, namely nuclear magnetic resonance (NMR) and mass spectrometry (MS). NMR measures the resonance of nuclei with a net spin under a magnetic field. NMR, which is commonly used as a structural determination technique, can be used to quantify the abundance of positional isotopomers (isomers with heavy isotopes substitutions in specific positions) with isotopes that possess a net nuclear spin (e.g., ^{13}C and ^{15}N) [6,7]. By contrast, MS separates and quantifies ionized molecules based on their mass-to-charge ratio (m/z). Therefore, as the mass of an ionized molecule will increase upon heavy isotope incorporation, MS can be used to quantify the relative abundance of

* Corresponding author. Department of Biochemistry and Molecular Biomedicine, Faculty of Biology, Universitat de Barcelona, Av. Diagonal, 643, 08028 Barcelona, Spain.
** Corresponding author. Department of Biochemistry and Molecular Biomedicine, Faculty of Biology, Universitat de Barcelona, Av. Diagonal, 643, 08028 Barcelona, Spain.
E-mail addresses: silviamarin@ub.edu (S. Marin), martacascante@ub.edu (M. Cascante).
[1] These authors contributed equally to this work, and they are listed in alphabetic order.

https://doi.org/10.1016/j.trac.2018.12.025
0165-9936/© 2019 Elsevier B.V. All rights reserved.

Please cite this article as: C. Balcells et al., Tracing metabolic fluxes using mass spectrometry: Stable isotope-resolved metabolomics in health and disease, Trends in Analytical Chemistry, https://doi.org/10.1016/j.trac.2018.12.025

line, J. Biol. Chem. 283 (2008) 20621–20627. https://doi.org/10.1074/jbc.M706494200.

[98] C.L. Perez, M.R. Van Gilst, A ^{13}C isotope labeling strategy reveals the influence of insulin signaling on lipogenesis in C. elegans, Cell Metab. 8 (2008) 266–274. https://doi.org/10.1016/j.cmet.2008.08.007.

[99] S.B. Crown, N. Marze, M.R. Antoniewicz, Catabolism of branched chain amino acids contributes significantly to synthesis of odd-chain and even-chain fatty acids in 3T3-L1 adipocytes, PLoS One 10 (2015) e0145850. https://doi.org/10.1371/journal.pone.0145850.

[100] J.S. Kirkwood, C.L. Miranda, G. Bobe, C.S. Maier, J.F. Stevens, 18O-Tracer metabolomics reveals protein turnover and CDP-choline cycle activity in differentiating 3T3-L1 pre-adipocytes, PLoS One 11 (2016) 1–18. https://doi.org/10.1371/journal.pone.0157118.

[101] G. Aubert, O.J. Martin, J.L. Horton, L. Lai, R.B. Vega, T.C. Leone, T. Koves, S.J. Gardell, M. Krüger, C.L. Hoppel, E.D. Lewandowski, P.A. Crawford, D.M. Muoio, D.P. Kelly, The failing heart relies on ketone bodies as a fuel, Circulation 133 (2016) 698–705. https://doi.org/10.1161/CIRCULATIONAHA.115.017355.

[102] P. Mirtschink, J. Krishnan, F. Grimm, A. Sarre, M. Hörl, M. Kayikci, N. Fankhauser, Y. Christinat, C. Cortijo, O. Feehan, A. Vukolic, S. Sossalla, S.N. Stehr, J. Ule, N. Zamboni, T. Pedrazzini, W. Krek, HIF-driven SF3B1 induces KHK-C to enforce fructolysis and heart disease, Nature 522 (2015) 444–449. https://doi.org/10.1038/nature14508.

[103] L. Gu, G.-F. Zhang, R.S. Kombu, F. Allen, G. Kutz, W.-U. Brewer, C.R. Roe, H. Brunengraber, Parenteral and enteral metabolism of anaplerotic triheptanoin in normal rats. II. Effects on lipolysis, glucose production, and liver acyl-CoA profile, Am. J. Physiol. Endocrinol. Metab. 298 (2010) E362–E371. https://doi.org/10.1152/ajpendo.00384.2009.

[104] N. Carinhas, A. Koshkin, D.A.M. Pais, P.M. Alves, A.P. Teixeira, 13 C-metabolic flux analysis of human adenovirus infection: implications for viral vector production, Biotechnol. Bioeng. 114 (2017) 195–207. https://doi.org/10.1002/bit.26063.

[105] S. Watanabe, M. Zimmermann, M.B. Goodwin, U. Sauer, C.E. Barry, H.I. Boshoff, Fumarate reductase activity maintains an energized membrane in anaerobic Mycobacterium tuberculosis, PLoS Pathog. 7 (2011) e1002287. https://doi.org/10.1371/journal.ppat.1002287.

[106] E.H. Ma, G. Bantug, T. Griss, S. Condotta, R.M. Johnson, B. Samborska, N. Mainolfi, V. Suri, H. Guak, M.L. Balmer, M.J. Verway, T.C. Raissi, H. Tsui, G. Boukhaled, S. Henriques da Costa, C. Frezza, C.M. Krawczyk, A. Friedman, M. Manfredi, M.J. Richer, C. Hess, R.G. Jones, Serine is an essential metabolite for effector T cell expansion, Cell Metab. 25 (2017) 345–357. https://doi.org/10.1016/j.cmet.2016.12.011.

[107] P. Sadiku, J.A. Willson, R.S. Dickinson, F. Murphy, A.J. Harris, A. Lewis, D. Sammut, A.S. Mirchandani, E. Ryan, E.R. Watts, A.A.R. Thompson, H.M. Marriott, D.H. Dockrell, C.T. Taylor, M. Schneider, P.H. Maxwell, E.R. Chilvers, M. Mazzone, V. Moral, C.W. Pugh, P.J. Ratcliffe, C.J. Schofield, B. Ghesquiere, P. Carmeliet, M.K.B. Whyte, S.R. Walmsley, Prolyl hydroxylase 2 inactivation enhances glycogen storage and promotes excessive neutrophilic responses, J. Clin. Invest. 127 (2017) 3407–3420. https://doi.org/10.1172/JCI90848.

[108] D. Drago, V. Basso, E. Gaude, G. Volpe, L. Peruzzotti-Jametti, A. Bachi, G. Musco, A. Andolfo, C. Frezza, A. Mondino, S. Pluchino, Metabolic determinants of the immune modulatory function of neural stem cells, J. Neuroinflammation 13 (2016) 1–18. https://doi.org/10.1186/s12974-016-0667-7.

[109] N.J. Matheson, J. Sumner, K. Wals, R. Rapiteanu, M.P. Weekes, R. Vigan, J. Weinelt, M. Schindler, R. Antrobus, A.S.H. Costa, C. Frezza, C.B. Clish, S.J.D. Neil, P.J. Lehner, Cell surface proteomic map of HIV infection reveals antagonism of amino acid metabolism by Vpu and Nef, Cell Host Microbe 18 (2015) 409–423. https://doi.org/10.1016/j.chom.2015.09.003.

[110] C. Des Rosiers, C.A. Fernandez, F. David, H. Brunengraber, Reversibility of the mitochondrial isocitrate dehydrogenase reaction in the perfused rat liver. Evidence from isotopomer analysis of citric acid cycle intermediates, J. Biol. Chem. 269 (1994) 27179–27182. http://www.ncbi.nlm.nih.gov/pubmed/7961626.

[111] G.-F. Zhang, S. Sadhukhan, R.A. Ibarra, S.M. Lauden, C.-Y. Chuang, S. Sushailo, P. Chatterjee, V.E. Anderson, G.P. Tochtrop, H. Brunengraber, Metabolism of γ-hydroxybutyrate in perfused rat livers, Biochem. J. 444 (2012) 333–341. https://doi.org/10.1042/BJ20112046.

[112] S.Y. Lunt, V. Muralidhar, A.M. Hosios, W.J. Israelsen, D.Y. Gui, L. Newhouse, M. Ogrodzinski, V. Hecht, K. Xu, P.N.M. Acevedo, D.P. Hollern, G. Bellinger, T.L. Dayton, S. Christen, I. Elia, A.T. Dinh, G. Stephanopoulos, S.R. Manalis, M.B. Yaffe, E.R. Andrechek, S.M. Fendt, M.G. Vander Heiden, Pyruvate kinase isoform expression alters nucleotide synthesis to impact cell proliferation, Mol. Cell. 57 (2015) 95–107. https://doi.org/10.1016/j.molcel.2014.10.027.

[113] A.R. Grassian, S.J. Parker, S.M. Davidson, A.S. Divakaruni, C.R. Green, X. Zhang, K.L. Slocum, M. Pu, F. Lin, C. Vickers, C. Joud-Caldwell, F. Chung, H. Yin, E.D. Handly, C. Straub, J.D. Growney, M.G. Vander Heiden, A.N. Murphy, R. Pagliarini, C.M. Metallo, IDH1 mutations alter citric acid cycle metabolism and increase dependence on oxidative mitochondrial metabolism, Cancer Res. 74 (2014) 3317–3331. https://doi.org/10.1158/0008-5472.CAN-14-0772-T.

[114] S. Christen, D. Lorendeau, R. Schmieder, D. Broekaert, K. Metzger, K. Veys, I. Elia, J.M. Buescher, M.F. Orth, S.M. Davidson, T.G.P. Grünewald, K. De Bock, S.-M. Fendt, Breast cancer-derived lung metastases show increased pyruvate carboxylase-dependent anaplerosis, Cell Rep. 17 (2016) 837–848. https://doi.org/10.1016/j.celrep.2016.09.042.

[115] I. Marín de Mas, E. Aguilar, E. Zodda, C. Balcells, S. Marin, G. Dallmann, T.M. Thomson, B. Papp, M. Cascante, Model-driven discovery of long-chain fatty acid metabolic reprogramming in heterogeneous prostate cancer cells, PLoS Comput. Biol. 14 (2018) e1005914. https://doi.org/10.1371/journal.pcbi.1005914.

[116] M. Tarrado-Castellarnau, P. de Atauri, J. Tarragó-Celada, J. Perarnau, M. Yuneva, T.M. Thomson, M. Cascante, De novo MYC addiction as an adaptive response of cancer cells to CDK4/6 inhibition, Mol. Syst. Biol. 13 (2017) 940. https://doi.org/10.15252/msb.20167321.

[117] J.R. Krycer, K. Yugi, A. Hirayama, D.J. Fazakerley, L.E. Quek, R. Scalzo, S. Ohno, M.P. Hodson, S. Ikeda, F. Shoji, K. Suzuki, W. Domanova, B.L. Parker, M.E. Nelson, S.J. Humphrey, N. Turner, K.L. Hoehn, G.J. Cooney, T. Soga, S. Kuroda, D.E. James, Dynamic metabolomics reveals that insulin primes the adipocyte for glucose metabolism, Cell Rep. 21 (2017) 3536–3547. https://doi.org/10.1016/j.celrep.2017.11.085.

[118] E. Brunk, S. Sahoo, D.C. Zielinski, A. Altunkaya, A. Dräger, N. Mih, F. Gatto, A. Nilsson, G.A. Preciat Gonzalez, M.K. Aurich, A. Prlic, A. Sastry, A.D. Danielsdottir, A. Heinken, A. Noronha, P.W. Rose, S.K. Burley, R.M.T. Fleming, J. Nielsen, I. Thiele, B.O. Palsson, Recon3D enables a three-dimensional view of gene variation in human metabolism, Nat. Biotechnol. 36 (2018) 272–281. https://doi.org/10.1038/nbt.4072.

Please cite this article as: C. Balcells et al., Tracing metabolic fluxes using mass spectrometry: Stable isotope-resolved metabolomics in health and disease, Trends in Analytical Chemistry, https://doi.org/10.1016/j.trac.2018.12.025

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Metabolomics in systems medicine: an overview of methods and applications

Effrosyni Karakitsou[1,2,†], Carles Foguet[1,3,†], Pedro de Atauri[1,3], Kim Kultima[4], Payam Emami Khoonsari[4], Vitor A. P. Martins dos Santos[5,6], Edoardo Saccenti[5], Antonio Rosato[7] and Marta Cascante[1,3]

Abstract

Patient-derived metabolomics offers valuable insights into the metabolic phenotype underlying diseases with a strong metabolic component. Thus, these data sets will be pivotal to the implementation of personalized medicine strategies in health and disease. However, to take full advantage of such data sets, they must be integrated with other omics within a coherent pathophysiological framework to enable improved diagnostics, to identify therapeutic interventions, and to accurately stratify patients. Herein, we provide an overview of the state-of-the-art data analysis and modeling approaches applicable to metabolomics data and of their potential for systems medicine.

Addresses

[1] Department of Biochemistry and Molecular Biomedicine and Institute of Biomedicine (IBUB), Faculty of Biology, Universitat de Barcelona (UB), Barcelona, Spain

[2] Institute of Cancer and Genomic Sciences and Centre for Computational Biology, University of Birmingham, B15 2TT, Birmingham, UK

[3] Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Metabolomics Node at INB-Bioinformatics Platform, Instituto de Salud Carlos III (ISCIII), Madrid, Spain

[4] Department of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden

[5] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708WE, Wageningen, the Netherlands

[6] LifeGlimmer GmbH, Markelstraße 38, 12163 Berlin, Germany

[7] Department of Chemistry and Magnetic Resonance Center, University of Florence, Italy

Corresponding authors: Saccenti, Edoardo (edoardo.saccenti@wur.nl); Rosato, Antonio (rosato@cerm.unifi.it); Cascante, Marta (martacascante@ub.edu)

[†] These authors contributed equally to this work.

Current Opinion in Systems Biology 2019, **15**:91–99

This reviews comes from a themed issue on **Gene regulation**

Edited by **Mariko Okada** and **Shinya Kuroda**

For a complete overview see the Issue and the Editorial

Available online 29 March 2019

https://doi.org/10.1016/j.coisb.2019.03.009

2452-3100/© 2019 Published by Elsevier Ltd.

Keywords

Systems medicine, Metabolomics, Multiomics, Kinetic modellig, Constraint-based modelling, Personalized medicine.

The brave new world of systems medicine

Systems biology treats biological systems as ensembles of networks at multiple levels, starting from the molecular level and from there gradually addressing more complex systems such as cells, tissues, organs, whole organisms, and finally analyzing population dynamics. Systems biology aims to describe and predict the behavior of groups of interacting components. To do so, it uses mathematical and computational tools to analyze measurements collected by systematic high-throughput technologies such as (post)genomics, metabolomics, or proteomics among others. The goal of systems approaches is to boost our understanding of biology by overcoming the limitations of reductive science, which addresses individual genes, proteins, metabolites, pathways, or cells, and thus does not account for the properties emerging from their interactions [1,2].

Current medical science is mostly conducted using the reductionist approach [3,4]. This limits our ability to grasp how multiple variables interact with one another to create emergent effects [3] and hampers our understanding of diseases, as well as our capability of delivering better treatments. Systems medicine can be regarded as the application of systems biology to human physiology in a clinical context [5,6]. It addresses the aforementioned issues by applying iterative and reciprocal feedback between clinical research and practice through computational, statistical, and mathematical multiscale analysis. This includes modeling of disease progression and remission, treatment responses, and adverse events both at the epidemiological and patient level. This new paradigm of systems science and medicine strongly complements the traditional reductionist approach (Figure 1).

the journal of biological databases and curation 2015, **2015**: bav068.

40. Brunk E, *et al.*: **Recon3D enables a three-dimensional view of gene variation in human metabolism**. *Nat Biotechnol* 2018, **36**: 272.

41. Foguet C, *et al.*: **HepatoDyn: a dynamic model of hepatocyte metabolism that integrates 13C isotopomer data**. *PLoS Comput Biol* 2016, **12**, e1004899.

42. Puigjaner J, *et al.*: **Comparison of control analysis data using different approaches: modelling and experiments with muscle extract**. *FEBS (Fed Eur Biochem Soc) Lett* 1997, **418**: 47–52.

43. Steuer R, *et al.*: **Structural kinetic modeling of metabolic networks**. *Proc Natl Acad Sci Unit States Am* 2006, **103**:11868.

44. Grimbs S, *et al.*: **The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks**. *Mol Syst Biol* 2007, **3**.

45. Jamshidi N, Palsson BØ: **Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models**. *Biophys J* 2010, **98**:175–185.

46. ** Nilsson A, Nielsen J: **Genome scale metabolic modeling of cancer**. *Metab Eng* 2017, **43**:103–112.
A comprehensive review of GSMMs focused on their application to the study of Cancer metabolism.

47. Zur H, Ruppin E, Shlomi T: **iMAT: an integrative metabolic analysis tool**. *Bioinformatics* 2010, **26**:3140–3142.

48. Jensen PA, Papin JA: **Functional integration of a metabolic network model and expression data without arbitrary thresholding**. *Bioinformatics* 2011, **27**:541–547.

49. Agren R, *et al.*: **Reconstruction of genome-scale Active metabolic networks for 69 human cell types and 16 cancer types using INIT**. *PLoS Comput Biol* 2012, **8**, e1002518.

50. Schmidt BJ, *et al.*: **GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data**. *Bioinformatics* 2013, **29**:2900–2908.

51. Yizhak K, *et al.*: **Model-based identification of drug targets that revert disrupted metabolism and its application to ageing**. *Nat Commun* 2013, **4**:2632.

52. Galhardo M, *et al.*: **Integrated analysis of transcript-level regulation of metabolism reveals disease-relevant nodes of the human metabolic network**. *Nucleic Acids Res* 2014, **42**: 1474–1496.

53. Reimers A-M, Reimers AC: **The steady-state assumption in oscillating and growing systems**. *J Theor Biol* 2016, **406**: 176–186.

54. Orth JD, Thiele I, Palsson BØ: **What is flux balance analysis?** *Nat Biotechnol* 2010, **28**:245–248.

55. Ramakrishna R, *et al.*: **Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints**. *Am J Physiol Regul Integr Comp Physiol* 2001, **280**:R695–R704.

56. Gille C, *et al.*: **HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology**. *Mol Syst Biol* 2010, **6**:411.

57. Bordbar A, *et al.*: **Constraint-based models predict metabolic and associated cellular functions**. *Nat Rev Genet* 2014, **15**: 107.

58. Rienksma RA, *et al.*: **Modeling the metabolic state of Mycobacterium tuberculosis upon infection**. *Front Cell Infect Microbiol* 2018, **8**.

59. Wang T, *et al.*: **Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic ras**. *Cell* 2017, **168**:890–903. e15.

60. Suhre K, *et al.*: **Human metabolic individuality in biomedical and pharmaceutical research**. *Nature* 2011, **477**:54–60.

61. * Zaghlool SB, *et al.*: **Deep molecular phenotypes link complex disorders and physiological insult to CpG methylation**. *Hum Mol Genet* 2018, **27**:1106–1121.
combined Epigenome-Wide ASsociation data (EWAS) with cytosine-guanine dinucleotide (CpG) methylation and other multi-omics datasets and revealed a causal effect of metabolite levels on methylation of obesity-associated CpG sites.

62. * Haanstra JR, *et al.*: **Targeting pathogen metabolism without collateral damage to the host**. *Sci Rep* 2017, **7**:40406.
This study highlights how kinetic modelling can assist drug design by identifying targets against pathogen metabolism with have minimal side effects on the host.

63. * Berndt N, *et al.*: **HEPATOKIN1 is a biochemistry-based model of liver metabolism for applications in medicine and pharmacology**. *Nat Commun* 2018, **9**:2386.
In this study, the authors build a large-scale kinetic model of hepatocyte metabolism and integrated proteomics to highlight the metabolic differences in the metabolic phenotype hepatocytes compared to adenoma and hepatocellular carcinoma cells.

64. ** Bordbar A, *et al.*: **Personalized whole-cell kinetic models of metabolism for discovery in genomics and pharmacodynamics**. *Cell Systems* 2015, **1**:283–292.
constructed personalized whole cell kinetic models of erythrocyte and showed that personalized kinetic rate constants are the best representation of the genotype. They were also able to identify individuals at risk for a drug side effect.

65. Mardinoglu A, *et al.*: **Integration of clinical data with a genome-scale metabolic model of the human adipocyte**. *Mol Syst Biol* 2013, **9**:649.

66. von Kamp A, Klamt S: **Enumeration of smallest intervention strategies in genome-scale metabolic networks**. *PLoS Comput Biol* 2014, **10**, e1003378.

67. Pratapa A, Balachandran S, Raman K: **Fast-SL: an efficient algorithm to identify synthetic lethal sets in metabolic networks**. *Bioinformatics* 2015, **31**:3299–3305.

68. Zhan T, Boutros M: **Towards a compendium of essential genes – from model organisms to synthetic lethality in cancer cells**. *Crit Rev Biochem Mol Biol* 2016, **51**:74–85.

69. Folger O, *et al.*: **Predicting selective drug targets in cancer through metabolic networks**. *Mol Syst Biol* 2011, **7**:501.

70. Agren R, *et al.*: **Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling**. *Mol Syst Biol* 2014, **10**:721.

71. Mardinoglu A, *et al.*: **Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease**. *Nat Commun* 2014, **5**:3083.

72. Väremo L, *et al.*: **Proteome- and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes**. *Cell Rep* 2015, **11**:921–933.

73. Kuiper JS, *et al.*: **Social relationships and risk of dementia: a systematic review and meta-analysis of longitudinal cohort studies**. *Ageing Res Rev* 2015, **22**:39–57.

74. Hamburg MA, Collins FS: **The path to personalized medicine**. *N Engl J Med* 2010, **363**:301–304.

75. Liu X, *et al.*: **Identifying disease genes and module biomarkers by differential interactions**. *J Am Med Inform Assoc : JAMIA* 2012, **19**:241–248.

76. Liu R, *et al.*: **Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers**. *Med Res Rev* 2014, **34**:455–478.

77. Zhang W, *et al.*: **Diagnosing phenotypes of single-sample individuals by edge biomarkers**. *J Mol Cell Biol* 2015, **7**:231–241.

78. Chen R, Snyder M: **Systems biology: personalized medicine for the future?** *Curr Opin Pharmacol* 2012, **12**:623–628.

79. Flores M, *et al.*: **P4 medicine: how systems medicine will transform the healthcare sector and society**. *Pers Med* 2013, **10**:565–576.