TESI DOCTORAL UPF / 2022

Department of Medicine and Life Sciences

An in silico Clinical Trial platform based on Systems Biology approaches

Guillem Jorba Argemí

Supervisor: Dr. Jose Manuel Mas

Co-Supervisor: Dr. Baldomero Oliva

ANAXOMICS
Systems biology solutions

upf. Universitat Pompeu Fabra Barcelona

The research leading to present thesis has received support from a Marie Curie Scholarship, and has been part of the DRIVE project (https://drive-autophagy.eu/), which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765912.

Resum

El procés de descobriment i desenvolupament de fàrmacs (DDD) és costós tant econòmic com temporalment, i el qual resulta en nivells d'èxit molt reduïts. Una gran quantitat de compostos son analitzats i molts descartats en cada estadi mentre que, ja en fases d'assaigs clínics, només un 10% d'aquests acaben aprovats per les agències reguladores. Tot i que existeix un gran ventall d'eines bioinformàtiques per les fases inicials del DDD, poques opcions estan disponibles en fases posteriors, on s'hi inverteix una gran quantitat de temps i diners. Aquest fet és bàsicament degut a la complexitat existent en la parametrització d'assaigs clínics, on hi intervenen humans i fàrmacs, i on múltiples eines s'haurien de fusionar per simular la realitat. En aquesta tesi, descric una nova metodologia per realitzar assaigs clínics in silico (ISCTs), basats en biologia de sistemes, amb l'objectiu d'assistir en el procés de desenvolupament de fàrmacs. Aquesta eina usa pacients virtuals, pel que ajudarà també a moure's cap a una medicina personalitzada o individualitzada.

Paraules clau: assaig clínic in silico; farmacologia de sistemes quantitativa; biologia de sistemes; pacients virtuals.

Abstract

Drug discovery and development (DDD) is a time consuming and expensive process with low rates of success. A big amount of drug candidates are tested and discarded in each phase and, in clinical trial phases, only about 10% of those are finally approved by regulatory agencies. While many bioinformatic tools are available at early stages of DDD, few options have been approved for posterior phases, where most investment and time is required. This is mainly because of the complexity inherent in clinical trials setting, where human cohorts are being tested with a drug, meaning that multiple tools must be merged in order to fully match reality, or have enough predictability. In this thesis, I describe a novel methodology to undergo in silico Clinial Trials (ISCTs), based on a Systems Biology approach, with the aim of assisting drug development process. It uses virtual patient populations, which will also help shifting towards individualized, personalized medicine.

Keywords: in silico Clinical Trial (ISCT); quantitative systems pharmacology (QSP); systems biology; virtual patients (VPs).

Contents

Part I

INTRODUCTION

Chapter 1

General Introduction

1.1 Drug discovery/development and Motivation

Drug discovery and development (DDD) is a time consuming and expensive, but necessary, process that allows to bring secure and effective drugs to market. Its starting point is mainly a medical condition with no available treatment, or which is not satisfactorily tackled using available treatment procedures. Commonly referred as an unmet medical need, it triggers and demands the search for new drugs/interventions which need be either targeting a non-tackled disease, offering additional advantages over existing treatments, presenting reduced adverse effects, or having fewer drug-drug interactions. Hence, resulting in an overall quality improvement in a patient's life [360]. The current process pipeline consists of a series of defined phases and steps (figure 1). As the name already suggests, two major stages can be defined, Drug Discovery and Drug Development (figure 1).

While several reliable computational tools are available at early stages of DDD, few options have been approved for posterior phases, during

drug development stage, where most investment and time is required and withdrawals at this point will lead to substantial financial losses [277]. This is mainly due to the inherent complexity of modeling an in-vivo setting, where many factors must be taken into account. Another barrier is that simulations at this stage should be highly reliable, if animals or human individuals are to be replaced by virtual subjects.

That being said, and due to the enhancement of computer models and technologies, better and more reliable models are being developed on a yearly basis. Patient-specific computational modeling is receiving much attention and funding due to its potential in reliably modeling animal and human experimentation, reducing their need in preclinical and clinical phases [404]. This is very appealing to pharmaceutical industries as it will save them money and time, while avoiding some ethical issues associated with animal and human research.

1.2 Drug Discovery

Although other strategies are used, drug discovery process usually initiates with a disease target identification via basic research; that is, a biomolecule related to a disease that can be targeted and modulated to ameliorate or treat a condition. A key point at this stage is to identify a target that is 'druggable', meaning that it can be accessible to a drug compound [178]. Next comes the target validation step, which can make use of several *in vitro*, *vivo* and/or *silico* techniques with the aim of demonstrating and confirming that the modulation of the identified target molecule in a disease environment does indeed correspond to a measurable benefit. Widely used approaches include the identification of the structure activity relationship (SAR), the knockdown or over expression mutants for the target, or the monitoring of

known signaling pathways downstream the potential target [183].

Once defined and validated, the Lead Identification step proceeds, which is the process carried out in order to find a potential drug candidate for that specific target. The initial goal at this phase is to identify 'hit' molecules or compounds which interact with the biological target, and that show a measurable, reproducible response [168]. Hits will become 'lead' compounds if they comply with a series of properties and if are able to fulfill a list of requirements, like initial toxicity or specificity studies. Dose-response curves will already be generated at this step, in order to determine the biological activity of the compound. Finally, an a priori save, effective and potent hit will go on as lead molecule [187]. Due to the low success rates at this and subsequent stages, many compounds are usually used and tested in parallel screening assays until one or several lead candidates are found [48].

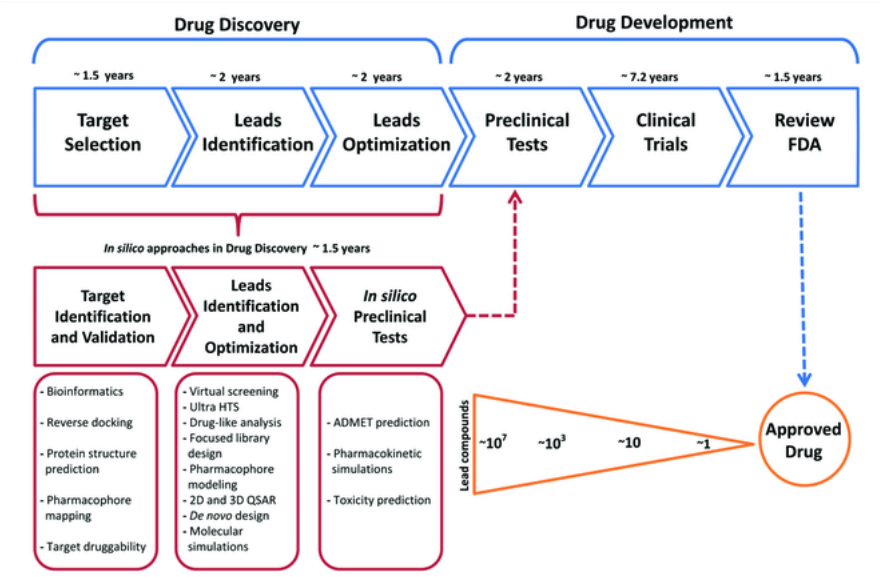

Figure 1: DDD process scheme and the main available computational tools. Obtained from: Rodenhizer et al., 2018 [335].

When a lead molecule is identified, an optimization process is undergone by chemically modifying the selected compound to improve its physicochemical properties. While the focus is put onto efficacy or potency, much effort is also taken at improving absorption, metabolization, distribution and excretion (AMDE) properties, especially drug's solubility, absorption or clearance, as they are often molecule drop-off limitations. Both lead identification and optimization phases are often intertwined by testing multiple lead molecules and variations at the same time, until a proper, robust compound is selected [168]. Moreover, further analyses are carried out to ensure the initial benefits of the compound and its compliance with de standards defined. Those include a series of pharmacokinetic, toxicity, efficacy, stability and bioavailability requirements, which are restricting limitations before entering preclinical phase in the Drug Development stage [177, 35].

Other strategies for drug discovery widely being used include: (i) drug repurposing, which uses compounds already available in the market and tests its efficacy over other diseases or pathological conditions; (ii) drug repositioning, similarly to repurposing, is based on the usage of unapproved drugs directed to other illnesses; and (iii) drug rescue, following the same idea, focus on giving a second chance to chemical and biological entities that were previously investigated but not further developed [283, 299].

1.3 Drug Development

Next step after a promising compound has been identified and tested, is the scaling up to animal models and human individuals, in preclinical and clinical trials (CT) phases I, II, III and IV. In order for a new molecule to be accepted into CT phases, a permission must be requested

by the submission of an Investigational New Drug (IND) application [170]. This document, requested by the regulatory agencies like the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA), must contain information on pharmacology and toxicology studies in animal models, chemical manufacturing information (including formulation, stability studies, and quality control measures), and detailed clinical protocols describing how the clinical compound shall be used and studied in subsequent human population CTs. After approval, the new drug can move towards human clinical trials, where it will be tested for safety and efficacy. Compound testing in humans is generally conducted in four phases. Each one of them is carried out as a separate, independent trial, where the researchers must submit the resulting output data gathered to the corresponding regulatory agency for approval, before moving to next phase [177, 35].

1.3.1 Preclinical Phase

Preclinical studies aim to ensure the new, promising compound will be valid for subsequent human testing and usage, by proving the efficacy and safety of the compound, as well as monitoring for possible side effects, in animal models. Pharmacokinetic data is also gathered at this point to evaluate the distribution of the lead molecule, and to calculate main parameters like half-life or absorption. At this stage, and because several animal models are usually being used, it is of upmost importance to carefully design the experiments in order to ensure safety and ethics are respected in all steps [95, 168]. Prior knowledge and early experimentation during drug discovery stage is put into practice, especially in the pharmacokinetics and pharmacodynamics fields, as a proportion of compounds fail to proceed because of high toxicity of their by-products [187]. The aims at this point

are to reduce to a minimum the risk to the healthy volunteers and the patients to whom the drug will be given in clinical trials [168].

Rats are habitually the first animal model choice for preclinical studies because of their small size, easy managing, low cost and wide understanding [187, 168]. Increasing dosages are commonly applied orally or intravenously, to evaluate the compound's pharmacokinetic properties like absorption and clearance, as well as monitoring toxicology and side effects. Efficacy is also tested in diseased animal models. Finally, if both efficacy and safety parameters lay within acceptable range, the IND is requested and the initial doses for human trials is defined. Compound manufacture (drug supply), dosing method, and formulation are also requirements for a successful IND report [377].

1.3.2 Clinical Phase I

Also named First-in-Human (FIH) trials, clinical phase I goal is to determine the safety of the new compound in human individuals. They comprise between 20 and 100 healthy subjects, which are given ascending dose profiles. The maximum tolerated dose, dose limiting toxicities, as well as drug's pharmacokinetic profile will be identified in these initial human trials, and the gathered information will be used to select the appropriate dosing profiles in subsequent phases [35]. Careful and thorough monitoring is carried out by following each individual's response to given dosages, in order to identify possible signs of toxicity or side effects.

1.3.3 Clinical Phase II

When a compound has proven safe in initial FIH studies, clinical phase II starts, focusing on effectiveness and response, while still monitoring

safety and tolerance. Here, few hundred patients are included to test the dose-response profiles regarding the disease or condition. After determining the appropriate dosing schedule, the effectiveness is tested and the trials can move on [35, 168]. In this phase, the treatment response is first evaluated in diseased human patients. Many drugs fail to demonstrate enough effectivity, or show high toxicity, and consequently they do not continue to further steps [364].

1.3.4 Clinical Phase III

Clinical phase III trials aim to confirm the effectiveness found in previous phase, as well as to keep ensuring safety, recruiting a broader number of patients of many thousands. These usually include double-blinded and randomized trials to confirm previous evidences, still monitoring for adverse events, as they might appear in larger populations. Due to time and complex designs, this is the costliest phase of the whole DDD process. Here, the safety/effectivity trade-off is measured, and different patient cohorts evaluated. When completed, a New Drug Application (NDA) is submitted to the agencies to request market approval [35, 168].

1.3.5 Clinical Phase IV

Once a drug is approved for commercial use, post-marketing monitoring stage begins. During that period, the compound has to be monitored by the sponsor (typically the manufacturer) in the so-called Clinical Phase IV or post-market studies. This is comprised by a pharmacovigilance group of experts who will monitor the new medication users in order to detect rare or late occurring adverse effect not earlier detected, or even to support efficacy measures, in a wider population group. It also serves as means to identify special cohorts of patients, not included or highly misrepresented

in previous phases.

1.4 Drug development struggles

As described above, DDD is a time consuming and expensive process only bearable by big pharmaceutical industries, and where successful rate is very low. Many thousands of molecules and its derivates are tested before entering clinical stages to bring a potential candidate molecule into human trials, although only about 10% of those finally end up approved by regulatory agencies and reach market [35, 160]. For that, a lot of emphasis has been put to increase early identification of most promising compounds, allowing to discard poor molecules as soon as possible in order to avoid investing money in further development, lowering the risk of withdrawal. Taking into account all these complexities and low success ratios, the consequence is a high cost to bring a new drug into the market, which is estimated in US$ 2.6 billion in average [97]. Time is also one of the key factors in DDD, as it takes around 10 full years since a new target is found, until the candidate compound is finally accepted by the regulatory agencies [360].

1.5 Computer modeling in drug development

Several approaches have been developed in order to model or simulate the effect and behavior of drugs. Their main objective is to increase the pharmaceutical research and development productivity, by assisting in the decision making during clinical stages, hence reducing the usage of both animal and human experimentation [259, 340, 338]. Although many approaches exist, focus is given here to the following fields: (i) drug simulation using pharmacokinetics, able to model drug behaviour inside a

body or compartments system; and (ii) systems biology modeling, widely used for interpreting complex systems interactions.

1.5.1 Pharmacokinetics/pharmacodynamics

Pharmacokinetics (PK) and Pharmacodynamics (PD) fields are often described as "what body does to the drug/chemical" and "what the drug/chemical does to the body" respectively. These terms are hence used to refer to the study of behavior and effect of drugs or chemicals in the body. PK aims at describing the four elements of absorption, distribution, metabolism and elimination (ADME) that describes the fate of a compound inside the body. On the other hand, PD describes the interactions of drugs and metabolites with biological targets and molecules, and their observed effects.

There are two major types of PK models, Non-compartment Analysis (NCA) and compartment physiological analysis [200]. For PD, models are focused on empirically fitting tissue dose and response, and are also majorly categorized into two types. The first group include direct effect models that assume chemical effects are directly proportional to receptor occupancy (i.e. linear transduction), while the other is the indirect effect model in which response is due to chemicals indirect effect to the synthesis or degradation of a response variable [201].

Both types of drug modeling, PK and PD, can be linked together which is often referred as PBPK/PD models, or PBPK (Physiologically-based pharmacokinetic models) [388, 125].

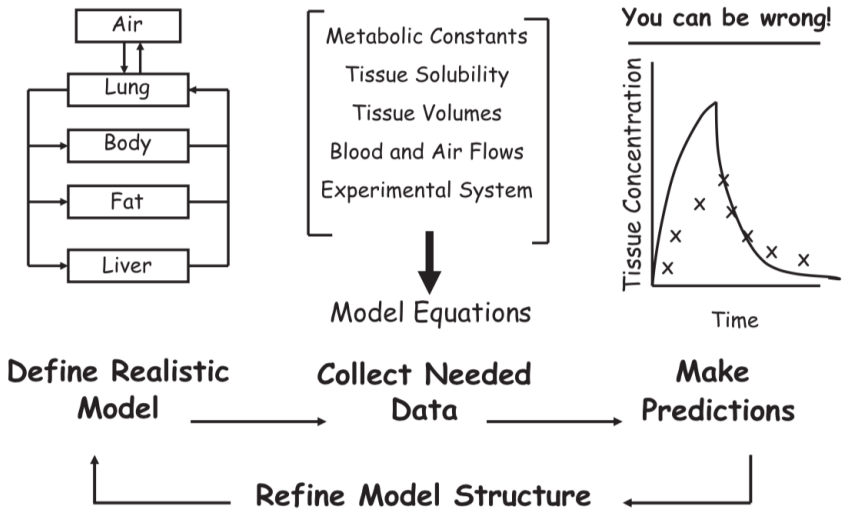

Figure 2: Steps for the PBPK model development. Obtained from: Reddy et al., 2005 [327].

Physiologically-based pharmacokinetic (PBPK) models

PBPK models (figure 2) use mathematical and computer modeling in order to integrate physiological and compounds' properties information to predict or simulate drug and/or molecules behavior inside a whole-body system [107]. They arouse during the 1980s in the need of improved modeling systems capable of providing help in assessing the risk or toxicity of drugs [118, 312]. However, it was not until 2010 that its implementation was widely used, mainly in order to model ADME properties of compounds [384, 231]. Thanks to the vast amount of data available and higher-performance computing power, PBPK models have been improved and are used nowadays within a wide variety of goals by predicting and modeling how drugs and/or molecules will behave and distribute inside the body.

While classical PK models use data from in vivo PK studies or scientific literature information on ADME characteristics to build "virtual" single

or multi compartment systems, PBPK adds the individuals' anatomical physiology data and biological processes to improve those models [118]. PBPK are based on a set of mathematical representations of the body composed of many parameters describing the physiology of the individual (eg. tissue/organ volumes, blood flows), and differential equations that will describe the fate of the compound inside the system. Similar to some PK models, they are composed of multiple compartments, representing all or the main organs/tissues of the body. Those are usually interconnected by blood flows, although alternative or more complex models also include lymphatic or even respiratory system [432, 247, 197]. While simpler models are preferred, complexity can increase until the user demands, by adding additional information like drug metabolism, cytochrome clearance, inter-organ compound dissemination, etc. [140, 386, 212, 191]. The resulting models are finally solved by using complex software algorithms.

PBPK building and parametrization can be achieved by bottom-up or top-down approaches. In the first one, drug or molecule characteristics are used to predict the ADME properties in the model. In contrast, top-down approaches use real observed experimental data to fit the missing parameters.

As for the PD information of the models, referring to the molecular events leading from the interaction of the drug/metabolite with a receptor to a pharmacodynamic or pharmacologic response, the following basic assumption can be applied:
$[Drug - Receptorcomplex] \rightarrow Response/Effect$

By applying the classical theory of drug-receptor interaction and the well-known modified Hill equations, the effect can thus be calculated as:
$E = Eo + Emax * C(t)/EC50 + C(t)$

Where, $Emax$ is the maximum response, $EC50$ is the concentration at which 50% of $Emax$ occurs and Eo is the baseline response. $C(t)$ is the effective chemical concentration i.e. concentration at the target site. Again, more complex models can include metabolites fate, as well as alternative response models as sigmoidal or indirect response models.

PBPK Applications

PBPK models, aside for simulation purposes, are mainly applied for risk assessment including: (i) in-vitro-to in-vivo extrapolation, or IVIVE [268], used in prediction in-vivo toxicity as well as adverse effects based on in-vitro dose response data [34, 39]; (ii) cross-species extrapolation, for risk assessment when translating from animals to humans trials, offering enhanced results compared to classical PK/PD approaches; (iii) cross-route extrapolation, for comparing drug exposure between different intake routes; (iv) dose extrapolation, especially useful for non-lineal response/toxicity drugs due to metabolism saturation, can be simulated in PBPK models bye, for example, including cytochrome interaction information; (v) time extrapolation, for better predicting differences between single and multiple dosages [78, 80]. Additionally, PBPK models allow the integration of other useful information about different patient cohorts for specific modeling, as pregnancy related characteristics or child individuals.

1.5.2 Systems Biology

Biological entities are mainly organized in complex systems of interconnected parts or units, which can be studied at different scales or levels like population, individual, organ, cellular or molecular. Systems biology aims to study those systems by means of mathematical and computer models, using a holistic approach [385, 213]. Two major pipelines

can be followed in this field (figure 3) : (i) bottom-up, and (ii) top-down. Bottom-up approaches aims at understanding and using information of the system constitutive parts (bottom), in order to integrate the new data into a bigger model and generate predictions on the system's behavior (top). They rely on detailed mechanistic studies or low-throughput proteomic experiments in order to extract new relationship data into the corresponding models, like yeast two-hybrid or tandem affinity purification yeast-2-hybrid assays [55, 406], co-immunoprecipitation [251], co-expression of genes [42], sequence homology [357], and others [333]. Nowadays, all this data is integrated into big databases where it can be easily extracted and included in the models. On the other hand, top-down methodologies start from a wider view of the system, using data to fit the model, and then extracting new insights from a molecular, bottom perspective. This approach has been lately the predominant one in the molecular scale, mainly because of the recent improvements in the -omics field. Top-down approaches mainly focus on big high-throughput data analyses, with the aim of characterizing the intrinsic or closer to the bottom pathways and/or interactions, responsible of the results variance [56, 100].

Networks are usually used in the systems biology field in order to represent and understand the behavior of the system as a whole, its components, and subprocesses, as a subfield called Network Biology. Given the mentioned interrelated nature of biological processes, network analysis is particularly suited for the purpose. In Network Biology, the nodes of the networks are being used for representing proteins, genes, or even diseases in higher level schemes. Network edges, then, depict the relationships between those biological entities. There exist four main types of network representations for studying the human organism: (i) protein interaction networks, where

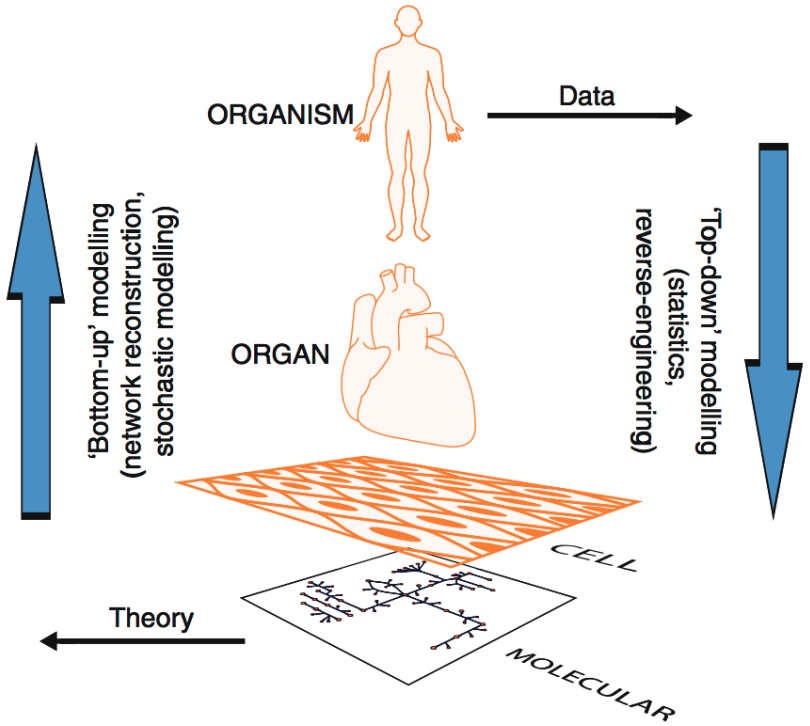

Figure 3: 'Top-down' vs. 'bottom-up' approaches for modeling biological systems. Obtained from: Edwards et. al 2013 [100].

nodes are proteins and edges are protein-protein physical interactions; (ii) metabolic networks, where nodes are metabolites and proteins, and edges are metabolic reactions; (iii) gene networks, where nodes are transcription factors and genes, and edges are regulatory interactions; and (iv) disease networks, where nodes are diseases, and edges represent different types of relationships such as shared genes [314].

From the molecular perspective, networks have helped integrating and englobing several types of information like protein-protein interactions (PPIs), gene expression patterns or enzymatic catalysis, into one single model, which can then be perturbed in order to extract new insights on pathway understanding. However, initially, most models focused on

describing and simulating rather small, specific networks, e.g., molecular pathways or a disease interaction context, in an effort to use a simplistic approach. Big effort has been put into trying to include all available connections and processes describing the functioning of the entire human cell, namely interactome (Box 1) [405] .

Network Medicine

Network medicine appeared as a sub-discipline of systems biology, with the aim of using biological networks for understanding diseases [27]. Basically, the idea of these model networks is that understanding complexity of gene regulation, metabolic reactions, and protein-protein interactions as complex networks, will shed light on the causes and mechanisms of diseases [145]. Cell systems have shown to have a highly complex interconnectivity of its components, implying that a single abnormality or miss-regulation can have an impact on a greater scale. For that, the understanding of diseases usually complies the analysis of many parts of the network [146, 351]. Finally, the ultimate goal of network medicine research is to develop a more general understanding of how perturbations propagate in a system by identifying the pathways, sub-types of disease states, and key components in the networks that can be targeted, analyzed or modeled in clinical interventions [350].

There are two major steps at generating network medicine models: (i) the construction of the network; and (ii) the analysis of the generated system. As in other systems biology approaches, different networks here can be built depending on the type of data used: (i) protein networks, using PPIs; (ii) gene co-expression and regulatory networks, which make use of phenotype-specific gene expression data mainly from high-throughput analyses; (iii) metabolic networks, representing a set

Box 1 - The Human Interactome

As more and more information become available, human protein networks or human interactome arouses as a global network system englobing all information related to human or human cell internal interactions. Actually, most of the different level biological networks are also interconnected between them, in a form of 'network of networks' [28]. As such, interactomes are generated from an ensemble of PPIs, gene expression relations, pathway activations/inhibitions and/or complex formation. By creating these model networks, a novel way to analyze the whole cell provides multiple insights for investigators.

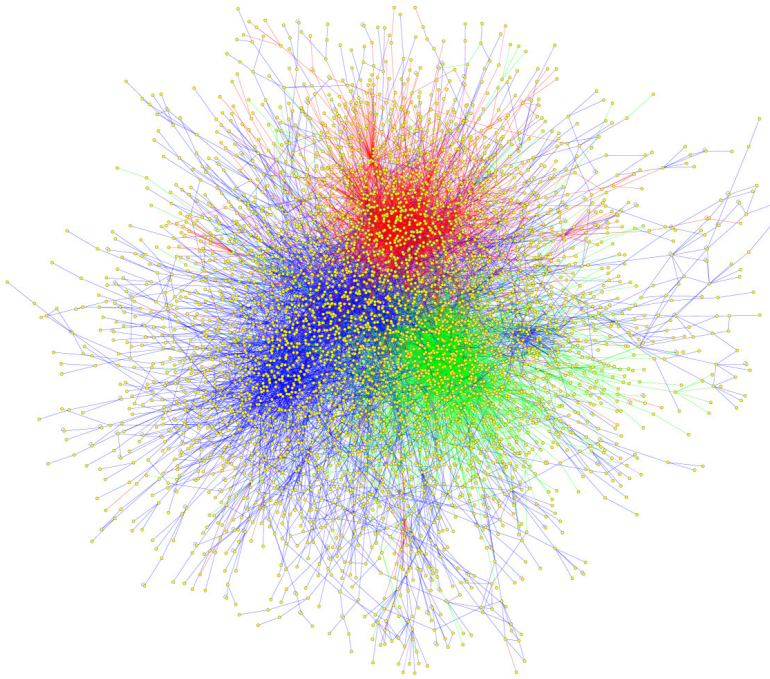

Figure 4: Interactome networks can be viewed as a merge of several protein/gene link data like PPIs (blue), gene regulation (red), and/or enzymatic reactions (gren). Obtained from: Lander et. al 2010 [235].

of biochemical interactions between metabolites and enzymes; among others; or (iv) interactome, which uses a merge of different information like PPIs, co-expression or metabolic reactions. On the other hand, the analysis of these systems requires the understanding their structural or topological properties. Because of the complexity of the constructed networks, the resulting models can be viewed in a global scale, searching for properties like average path length, degree distribution, diameter, clustering coefficients, or controllability (Liu et al., 2011). Alternatively, local analyses are used by focusing on sub-network motifs including a fixed number of connecting nodes (typically 3 or 4) to identify recurrent patterns [278].

Different approaches have been developed in order to build and study these complex networks and interactions between them, e.g.: (i) Passing Attributes between Networks for Data Assimilation (PANDA) algorithm [143]; (ii) INtegrated DiffErential Expression and Differential network analysis (INDEED) [451] and DICER [16], focused on differential network analyses for comparing control vs disease networks resulting from differential expression analyses; (iii) ARACNe [264] and CLR [109], which are gene-gene interaction network methods that attempt to understand regulatory associations by accounting for connections within a shared neighborhood of genes; or (iv) Therapeutic Mapping System (TPMS) [18], which generates mathematical models to elucidate the mechanism of action of drug-pathophysiology relations (Box 2).

1.5.3 Quantitative Systems Pharmacology

Quantitative Systems Pharmacology (QSP) is a novel discipline that emerged as means to integrate systems biology approaches with

Box 2 - Therapeutic Performance Mapping System (TPMS)
TPMS is a top-down systems biology approach that integrates pharmacological knowledge to create mathematical models that simulate human patho-physiology in silico [339, 164, 148, 182]. That is, it uses protein network models to extract the possible pathways, or mechanisms of action (MoAs), connecting drug targets and a set of pathologically related proteins. Two core elements of the TPMS technology are the human protein network (HPN) or interactome, combined with a Sampling Method algorithm used for generating the mathematical models. The HPN consists of a protein interaction network built using information from a compendium of databases like: (i) KEGG [205]; (ii) BioGRID [68]; (iii) IntAct [300]; (iv) REACTOME [108]; (v) TRRUST [157]; (vi) and HPRD [319]. Thus, it integrates several types of protein relations like PPIs, complexes formation, gene co-regulation or reaction information.
Sampling Methods are used to solve the missing parameters of the HPN, which is transformed into mathematical models capable of both reproducing existing knowledge and predicting new data. A collection of known drug-pathology/AEs relationships are considered as prior information and integrated into a table ("truth table" [19]) as training data, which ensure the models will reproduce the observed behavior of a human cell. The truth table is constructed by using an array of databases that accumulates biological and clinical data and provides biological and pharmacological action-response relationships (such as drug-indication pairs). In order to model both drugs and pathologies in the HPN, drugs are translated, characterized, into protein targets and pathologies into 'effectors', or proteins displaying a measurable effect produced by the pathology. TPMS method then generates mathematical models similar to Multilayer Perceptons where neurons are proteins and edges are the links between proteins in the HPN, and signal is transduced from input (drug target) to output (pathology effectors) [198].

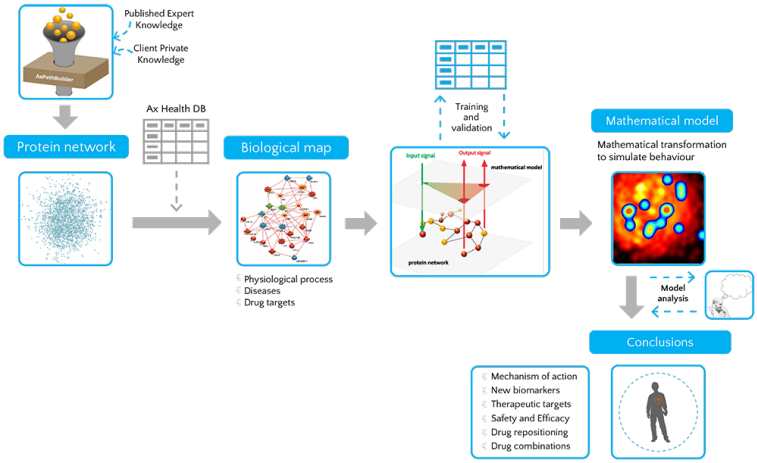

Figure 5: TPMS modelling scheme. Obtained from: https://www.anaxomics.com/tpms.php#tpms [18]

pharmacodynamic (PD), pharmacokinetic (PK) or physiologically-based pharmacokinetic (PBPK) models. Together, they are able to include the physiological characteristics and molecular/cellular/organ systems, leading to the identification and design of safer and more effective drug therapies [185, 243]. QSP concept is gaining more attention since the release in 2011 of a National Institutes of Health white paper based on a series of workshops and discussions with field investigators and industry [1]. More recently, according to an expert consensus, QSP was defined as "a quantitative or computational framework to support translational drug discovery and development by integrating knowledge on biochemical, biological, physiological, pharmacological, and clinical systems" [290]. While systems biology or networking methodologies aims at generating a general interaction model for identifying MoAs, biomarkers, or adverse events at a mechanistic level, QSP focus on overcoming the gap between pre-clinical and clinical science by integrating pre-clinical knowledge and

methods from Systems Biology with clinical pharmacology. It does so by understanding the clinically relevant mechanisms of action of drugs, and using this knowledge to optimize therapy in a way that achieves maximum effect and minimal toxicity in a given individual [12, 410].

Several approaches for QSP have been tackled, including statistical (Bayesian), Boolean, temporal (ordinary differential equations), spatio-temporal (partial differential equations), agent-based, integrative, empirical curve fitting, and machine learning that enable integrating molecular pathways with clinical results and pharmacology [139]. However, many already published QSP models are constructed as multi-compartment nonlinear systems of ordinary differential equations (ODE) [70]. Table 1 summarizes some examples of QSP models approaches developed so far.

Table 1: List of published QSP models and their company/institution affiliations.

Title	**Institutions**	**Reference**
A Strategy for Developing New Treatment Paradigms for Neuropsychiatric and Neurocognitive Symptoms in Alzheimer's Disease	In Silico Biosciences/ University of Pennsylvania/ Oregon Health & Science University	[138]
A Computer-Based Quantitative Systems Pharmacology Model of Negative Symptoms in Schizophrenia: Exploring Glycine Modulation of Excitation-Inhibition Balance	In Silico Biosciences/ Oregon Health & Science University/ University of Pennsylvania	[374]

table continues

Title	**Institutions**	**Reference**
Systems pharmacology of the nerve growth factor pathway: use of a systems biology model for the identification of key drug targets using sensitivity analysis and the integration of physiology and pharmacology	Xenologiq/ Astellas/ Pfizer	[37]
Development and Application of a Quantitative Systems Pharmacology (QSP) Model of Complement Pathway to Evaluate Treatments for Autoimmune Diseases	GlaxoSmithKline	[26]
A Humanized Clinically Calibrated Quantitative Systems Pharmacology Model for Hypokinetic Motor Symptoms in Parkinson's Disease	In Silico Biosciences/ Washington State University/ University of Pennsylvania	[334]
Systems Pharmacology Modeling in Neuroscience: Prediction and Outcome of PF-04995274, a 5-HT4 Partial Agonist, in a Clinical Scopolamine Impairment Trial	Pfizer	[289]
A Dynamic Quantitative Systems Pharmacology Model of Inflammatory Bowel Disease: Part 1 - Model Framework	Pfizer/ Janssen/ Takeda	[337]

table continues

Title	**Institutions**	**Reference**
Benefits and Challenges of a QSP Approach Through Case Study: Evaluation of a Hypothetical GLP-1/GIP Dual Agonist Therapy	Pfizer	[332]
Effects of IL-1β-Blocking Therapies in Type 2 Diabetes Mellitus: A Quantitative Systems Pharmacology Modeling Approach to Explore Underlying Mechanisms	AstraZeneca/ MedImmune	[302]
Radiation and PD-(L)1 Treatment Combinations: Immune Response and Dose Optimization via a Predictive Systems Model	AstraZeneca	[228]
Clinical Responses to ERK Inhibition in BRAF V600E-Mutant Colorectal Cancer Predicted Using a Computation Model	Genentech	[222]
Computational Modeling of ERBB2-Amplified Breast Cancer Identifies Combined ErbB2/3 Blockade as Superior to the Combination of MEK and AKT Inhibitors	Merrimack	[221]

table continues

Title	Institutions	Reference
Systems Pharmacology-Based Approach for Dissecting the Active Ingredients and Potential Targets of the Chinese Herbal BJF for the Treatment of COPD	Henan University of Traditional Chinese Medicine	[247]
QSP Toolbox: Computational Implementation of Integrated Workflow Components for Deploying Multi-Scale Mechanistic Models	Bristol-Myers Squibb	[70]
Characterization and Prediction of Cardiovascular Effects of Fingolimod and Siponimod Using a Systems Pharmacology Modeling Approach	Novartis/ Leiden Academic Centre for Drug Research	[365]
The role of quantitative systems pharmacology modeling in the prediction and explanation of idiosyncratic drug-induced liver injury	DILIsym Services	[430]

1.5.4 Precision medicine

Precision medicine, often referred as personalize medicine, is a rather new field focused on using patient's specific characteristics, like genotype or physiological data, in order to adapt or choose the best treatment scheduling approach. The National Research Council's Toward Precision Medicine agreed on using the following definition of precision medicine:

"The tailoring of medical treatment to the individual characteristics of each patient... to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment. Preventative or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not" [293]. Basically, its goal is to identify cohorts or sub-groups of patients for which a specific treatment has been proven beneficial or an unfavorable and use this knowledge for future similar individuals. This field has gained interest in last few years due to the enhancement of technologies allowing data obtention wider and faster for both investigators and medical care doctors [267].

In several modeling fields like systems biology, but especially in PBPK or the newly QSP modelling, in order to reflect world's variability, personalized or precision medicine are gaining importance with the aim of accounting for population variabilities. In an effort of moving from the standard/global-patient models to grasping different cohort/patient characteristic responses, and thanks to current computer capabilities, there has been an increase interest in the usage of Virtual Patients (VPs) and Virtual Populations (VPop). VPs and VPops allow the generation of patient distributions with varying characteristics, which can then be used for modeling and analyzing different plausible settings [11].

However, VPop generation techniques must prove reliable and able to compute real-like, robust population distribution values. Several algorithms have been proposed and used in that sense [11, 71].

1.5.5 In silico clinical trials (ISCT)

While the concept of clinical trial simulation (CTS) had already emerged in the early 2000s [147, 79, 142], in silico clinical trials appeared a few years later [17, 411, 248], and was only recently during 2010s when the term was properly defined and accepted thanks to the foundation of organizations like VPH Institute and the Avicenna consortium [412, 22, 304]. The latter took a first step in designing and preparing a first draft englobing a series of standards towards this new line of software development and provided a consensus definition for ISCT: "the use of individualised computer simulation in the development or regulatory evaluation of a medicinal product or medical device/medical intervention" [403].

Although the perception is changing, there are still some barriers in the usage of ISCT approaches: (i) cultural resistance from the researchers -mainly biologists, pharmacologists, and medics with limited computer modeling background; (ii) regulators, albeit some are now indeed promoting their use, they historically did not accept in silico obtained evidences; and (iii) the already mentioned complexities associated with the construction and predictability for quantitative modeling of living organisms [304]. However, both regulatory agencies and industry have already started promoting and encouraging the usage of computer models. In 2016, the FDA issued the draft guidance “Reporting of Computational Modeling Studies in Medical Device Submissions", aimed at FDA staff and industry [113]. They also participated in the formation of committees and consortiums focused on advancing in computational modeling of medical devices [21, 353]. Governments are also pushing forward in silico approaches, as US approved a Congress bill stating: “...urges FDA to engage with device and drug sponsors to explore greater use, where

appropriate, of In Silico trials for advancing new devices and drug therapy applications" [396]. In parallel, the European Parliament suggested the European Medicine Agency (EMA) to "...develop a framework for the regulatory acceptance of alternative models and shall take into consideration the opportunities presented by these new concepts which aim at providing for more predictive medicines. These concepts may be based on human relevant computer or cellular models, pathways of toxicity, or adverse outcome pathways" [103].

The aim of ISCTs is to move toward a more patient-specific modeling approaches, while using reliable modeling techniques. Different types of methodologies, similar to those of QSP, have been proposed for that purpose: (i) agent-based modeling (ABM), used in precision medicine to find optimal, personalized dosage and timing for drug administration or vaccine development [59, 216]; (ii) differential equations, widely used in pharmacokinetic modeling with good results, but also to model tumor and pathogen dynamics [348, 328, 152]; (iii) machine learning, and systems biology, which uses artificial intelligence technology to investigate drug mechanism of action and drug effects [65].

Despite the initial doubts, carefully implemented ISCT offer many advantages for DDD either in combination or supplementing real CTs, and especially for drug repurposing cases were many information is already available: (i) lower cost and implementation time; (ii) the possibility of generating and testing thousands of virtual patients; (iii) the possibility of testing extreme, rare cases; and (4) the possibility of testing several treatment variations [402]. An example of the promising benefits are the few groups that have already started developing their ISCT tools with encouraging results [306, 65, 361, 206].

Chapter 2

Objectives

The main goal of this thesis is to define the general strategy and to develop the computational tools to perform in silico Clinical Trials (ISCTs). They will work over a powerful systems biology approach and allow the individual patient evaluation of different scenarios where a treatment can be applied. All this can be divided in the following working objectives:

- Usage of a Systems Biology approach, namely Therapeutic Mapping System (TPMS). It is focused in extracting the Mechanisms of Action (MoA) of drugs by using sampling methods over a Human Protein Network (HPN).
  - Use the TPMS methodology in a head-to-head study in NSCLC to explore the efficacies of two treatments. Additionally, undergo a thorough analysis in the MoA models generated to extract information on possible resistances, metastasis prevention, and harmful drug-drug interactions.
  - As TPMS models contain not one but an ensemble of

MoA solutions, extend the approach application by analyzing generated models as prototype-patients in a Heart Failure context, in a first approach to moving towards personalized medicine. The working hypothesis is that prototype-patient models can be grouped and compared to explore the molecular differences between best and worst treatment responders.

- Development and implementation of an ISCT platform, a software dedicated to the simulation of Clinical Trials in silico by computing and evaluating treatment responses in virtually generated patients.
  - Design of a general physiologically-based pharmacokinetic (PBPK) model method able to simulate any, or most, drugs' behavior inside a human body system. For that, use a data-driven approach which models drug using their information on compound administration, body distribution and excretion properties.
  - Develop a tool to generate Virtual Patients (VPs) containing both PBPK and SB descriptors in order to move towards personalized medicine, by providing demographic and molecular patient variability.
  - Combine PBPK with the TPMS approach to generate Quantitative Systems Pharmacology (QSP) models. For that, use the PBPK output compound concentrations as input restrictions for the TPMS-SB approach.
  - Apply the ISCT platform in an ADHD head-to-head case-study. Then, analyze the generated virtual patients for demographic and molecular differences influencing treatment effect or predisposition.

Part II

SYSTEMS BIOLOGY APPROACH (TPMS)

Chapter 3

Using TPMS models to compare two first-line treatments for ALK+ NSCLC

3.1 Introduction

As explained in Box2, TPMS consists of a Systems Biology technology focused on understanding the Mechanisms of Action (MoA) of stimulus (e.g., drugs)-response (e.g., disease) pairs. The resulting models, which contains an ensemble of solutions, are able to explain the study region of the Human Interactome and can be used, among other possibilities, to understand the protein pathways linking stimulus and response, look for possible biomarkers, or predict probable adverse reactions [309, 164, 339].

In the following article, TPMS was used in the context of ALK+ non scamous cell lung carcinoma (NSCLC) to evaluate two first-line treatments, alectinib and brigatinib. The TPMS models were computed to generate

their respective MoAs with the aim of predicting their mechanistic effect, possible resistance mechanisms bypass, and the impact of potential resistances and other treatment interferences. In this article, I participated in the supervision of the models generated, as well as generating part of the data analysis.

3.2 Head to head evaluation of second generation ALK inhibitors brigatinib and alectinib as first-line treatment for ALK+ NSCLC using an in silico systems biology-based approach

Oncotarget. 2021 Feb 16;12(4):316-332.
PMID: 33659043
PMCID: PMC7899557
DOI: 10.18632/oncotarget.27875

Enric Carcereny[1], Alonso Fernández-Nistal[2], Araceli López[2], Carmen Montoto[2], Andrea Naves[2], Cristina Segú-Vergés[3], Mireia Coma[3], Guillem Jorba[3,4], Baldomero Oliva[4], Jose Manuel Mas[4]

[1]Catalan Institute of Oncology B-ARGO Group, Hospital Germans Trias i Pujol, Badalona, Spain.
[2]Takeda Farmacéutica España, Madrid, Spain.
[3]Anaxomics Biotech, Barcelona, Spain.
[4]Structural Bioinformatics (GRIB-IMIM), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain.

3.2.1 Abstract

Around 3-7% of patients with non-small cell lung cancer (NSCLC), which represent 85% of diagnosed lung cancers, have a rearrangement in the ALK gene that produces an abnormal activity of the ALK protein cell signaling pathway. The developed ALK tyrosine kinase inhibitors (TKIs), such as crizotinib, ceritinib, alectinib, brigatinib and lorlatinb present good performance treating ALK+ NSCLC, although all patients invariably develop resistance due to ALK secondary mutations or bypass mechanisms. In the present study, we compare the potential differences between brigatinib and alectinib's mechanisms of action as first-line treatment for ALK+ NSCLC in a systems biology-based in silico setting. Therapeutic performance mapping system (TPMS) technology was used to characterize the mechanisms of action of brigatinib and alectinib and the impact of potential resistances and drug interferences with concomitant treatments. The analyses indicate that brigatinib and alectinib affect cell growth, apoptosis and immune evasion through ALK inhibition. However, brigatinib seems to achieve a more diverse downstream effect due to a broader cancer-related kinase target spectrum. Brigatinib also shows a robust effect over invasiveness and central nervous system metastasis-related mechanisms, whereas alectinib seems to have a greater impact on the immune evasion mechanism. Based on this in silico head to head study, we conclude that brigatinib shows a predicted efficacy similar to alectinib and could be a good candidate in a first-line setting against ALK+ NSCLC. Future investigation involving clinical studies will be needed to confirm these findings. These in silico systems biology-based models could be applied for exploring other unanswered questions.

3.2.2 Introduction

Lung cancer (LC) remains the leading cause of death worldwide, with an estimated 1.6 million deaths each year [389, 429]. Despite significant therapeutic advances over the last decade, over half of patients diagnosed with LC die within one year of diagnosis and the five-year survival is around 18% [173]. About 85% of LCs are diagnosed as the subtype non-small cell lung cancer (NSCLC), adenocarcinoma being one of the most common histological subtypes. In adenocarcinoma, several driver mutations have been identified, including mutations/alterations of the epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase (ALK), and ROS1, among others; most of them are therapeutically targetable [150]. Around 3–7% of NSCLC cases present active ALK rearrangement (ALK+ NSCLC) that produces an abnormal activity of the ALK protein cell signaling pathway and causes the cancer cells to grow and metastasize [343, 281]. Central nervous system (CNS) metastasis is a common finding in NSCLCs, occurring in 10% of patients, and even more frequent in ALK+ NSCLCs, were the frequency of CNS metastasis is around 20–30% at the time of diagnosis [149, 195]. CNS is also the most common site of relapse [439].

Thus far, three generations of ALK tyrosine kinase inhibitors (TKIs) have been developed. Some of the drugs that target the abnormal ALK protein are crizotinib (first generation), ceritinib, alectinib, brigatinib, ensartinib (second generation) and lorlatinib (third generation) [378]. However, despite their effectiveness in ALK+ NSCLC cases, all patients invariably develop treatment resistance at some point. Consequently, it is of the upmost importance to adequately use the currently available treatments in the correct order to maximise the life span of NSCLC patients. The most common progression mechanisms for all ALKi are: 1) ALK secondary

mutations, which affect not only crizotinib-treated patients (around 20–30% of patients [434]), but also second and third generation ALKi [245, 250]; and 2) bypass mechanisms (i.e., activation of other parallel pro-proliferative signaling pathways [132]). Crizotinib has been used as first-line since its approval in 2011 in the United States [435] and in 2015 in Europe[434]. The results of its phase III trial (PROFILE 1014) demonstrated that crizotinib was superior to standard chemotherapy [367]. Second and third generation ALK TKIs are effective in treating numerous crizotinib-resistant ALK mutations and are used after crizotinib, and some of them have even replaced crizotinib as first-line option among patients with ALK-rearranged NSCLC [10, 397, 445]. Second generation ALKi ceritinib, alectinib and brigatinib have been approved for the treatment of ALK+ NSCLC patients after treatment with crizotinib (ceritinib in 2014 [124] and in 2015 [452]; alectinib in 2015 [114] and in 2017 [9], while brigatinib in 2017 [14] and 2018 [13] for the United States and Europe, respectively) and as first-line TKI treatments (in 2017, ceritinib [122, 452] and alectinib[9, 121] and, in 2020, brigatinib [13, 123], for the United States and Europe, respectively).

At the time of the design and performance of the study, clinical trials had provided promising results for both brigatinib and alectinib as first-line TKIs in TKI-naïve ALK+ NSCLC patients, compared to crizotinib. The ALEX phase III study (https://clinicaltrials.gov/ number, NCT02075840), showed that alectinib had a superior investigator-assessed PFS versus crizotinib (HR, 0.47; $P < 0.001$ [311]). At the second interim analysis of the ALTA-1L phase III trial (https://clinicaltrials.gov/ number, NCT02737501) the blinded independent review committee (BIRC)-assessed HR of PFS was 0.49 (log rank $P < 0.0001$) [63]. Moreover, both drugs present relevant intracranial efficacy: alectinib demonstrated superior efficacy versus crizotinib

regardless of baseline CNS metastases [60] and brigatinib significantly delayed both CNS progression (without prior systemic progression) and systemic progression (without prior intracranial progression) compared with crizotinib [62]. Regarding ceritinib, direct comparison in first line has only been performed with chemotherapy [116, 371], although indirect comparison showed better results for ceritinib than crizotinib [246]. At the time of the beginning of the current study, two other ALK inhibitors were being tested in first line in comparison to crizotinib, although no results were available (ensartinib in eXalt3, NCT02767804, or lorlatinib in CROWN, NCT03052608).

Although no direct comparison between alectinib, brigatinib and ceritinib has been performed in a first-line setting, there are indirect comparisons in second line from which hypotheses can be drawn. Ceritinib, alectinib and brigatinib are effective in crizotinib-refractory ALK+ NSCLC patients [356, 291, 176], but no direct comparison between these drugs after crizotinib is available. An ongoing trial, ALTA-3 (https://clinicaltrials.gov/ number, NCT03596866), compares the efficacy of alectinib versus brigatinib in ALK+ NSCLC patients who had progressed on crizotinib; besides, and according to the current lack of direct comparisons, indirect analyses using available data have been performed to compare them. In fact, a matching-adjusted indirect comparison (MAIC) [326] between these drugs in crizotinib-refractory ALK+ NSCLC patients (using clinical data from the ALTA trial – date February 21, 2017 –, ASCEND-1 [215], ASCEND-2 [280] NP28761 [61] and NP28673 [30]) suggested that brigatinib may have prolonged PFS and OS versus ceritinib and prolonged PFS versus alectinib in patients after progression with crizotinib.

From a safety perspective, all ALKi are considered to be safe and tolerable

in a similar fashion [52], although they show adverse events (AE), some of them common and others drug-specific. A systematic review [208] concluded that crizotinib was associated with more gastrointestinal and visual events, alectinib tended to have more hepatic and musculoskeletal AEs, ceritinib presented the highest incidence of clinically significant gastrointestinal AEs and laboratory abnormalities and brigatinib had a unique profile of increased early onset pulmonary AEs and hypertension associated with the 180 mg dose; these pulmonary AEs were found to be reduced when using the recommended initial dose of 90 mg [347]. This systematic review also suggested ceritinib to be less preferred by clinicians due to its safety profile. Regardless of their differences, most of the safety concerns associated with the mentioned ALKi can be minimized reducing administration dose [208] .

According to first-line results with brigatinib and alectinib and indirect results of ceritinib, this last drug seems to have a lower efficacy both at systemic and cerebral levels when compared to brigatinib and alectinib [326]. As tolerability of all these ALKi is similar and alectinib has become the standard of care, a head to head clinical trial comparing brigatinib and alectinib as first-line therapy would be very interesting. However, since this head to head is not planned, results obtained with these second generation ALKi in the ALTA-2 study in second line, as well as in the MAIC analysis, will help to elucidate and refine the first-line therapy outline. Besides, in silico investigational approaches may be an alternative to compare the potential benefits of both drugs.

Concerning the mechanism of action of alectinib and brigatinib, both share ALK as a protein target, but they display completely different target profiles that could be determinant to define each drug mechanism. Beside

ALK, brigatinib targets other tyrosine-protein kinases receptors such as EGFR [15, 189, 394], receptor-type tyrosine-protein kinase FLT3 [15, 175], tyrosine-protein kinase FER [375], ROS1 [175, 325], and insulin-like growth factor 1 receptor (IGF1R) [15, 189, 175, 276]. On the other hand, alectinib inhibits RET with comparable potency to ALK [225].

In silico tools are useful resources for predicting several (bio)chemical and (patho)physiological characteristics of likewise potential drugs [431]. These methods are used to improve in vivo and in vitro models and refine experimental programs of clinical and general biomedical studies involving lab work [390], and, in the long run, can reduce lab work and effectively succeed in 3R (reduce, reuse, recycle) [193]. Overall, these systems can be employed for the exploration of anticancer drug mechanisms of action and their efficacy in specific patient profiles.

In the present study, we created in silico systems biology-based mechanistic models of two first-line approved second generation ALKi, brigatinib and alectinib, in order to explore the potential differences between them with the aim of providing information or raising hypotheses towards the identification of strengths and weaknesses of the mechanisms of action of both drugs as first-line treatment for ALK+ NSCLC patients.

3.2.3 Results

The main pathophysiological processes (namely "motives") described to be involved in ALK+ NSCLC were: (1) Cell growth and proliferation, (2) Sustained angiogenesis, (3) Evading apoptosis, (4) Tissue invasion and metastasis, (5) Immune evasion (Table 4). Subsequently, each pathophysiological process was functionally characterized at protein level to determine its molecular effectors and used for focusing the analysis

towards ALK+ NSCLC in a human biological network context (Figure 6 and Supplementary Table 2). Brigatinib and alectinib protein target profiles were also carefully characterized and used in the posterior analyses (Figure 6 and Table 5). Mechanistic systems biology models of brigatinib and alectinib obtained with TPMS technology were constructed with accuracy values of 94% to evaluate their mechanism of action and potential treatment efficacy in ALK+ NSCLC. Two distinct modelling approaches were used for that purpose: Artificial neural networks (ANN) [339], with the aim of detecting biological relationships; and sampling-based methods [198], in order to explain those relationships. A Sobol sensibility analysis was applied to brigatinib and alectinib mechanistic models in order to evaluate their robustness. The results of this analysis are available in the Supplementary Methods.

Effect of brigatinib and alectinib on cell growth, apoptosis and immune evasion through ALK and non-ALK inhibition

The relationships of each drug target with ALK+ NSCLC main pathophysiological motives were evaluated by the ANN and the results are shown in Table 1. The ANN analysis showed that, in general, alectinib presented a slightly lower correlation with ALK+ NSCLC pathophysiology than brigatinib (around 80% of the score obtained by brigatinib).

Evaluation of the relations between individual pathophysiological motives and drug targets suggested that both drugs affect cell growth and proliferation, apoptosis evasion and immune evasion through ALK inhibition. Regarding alectinib, its inhibition of RET might occur through modulation of the tumour immune response. On the other side, brigatinib non-ALK targets might affect the pathophysiological motives already

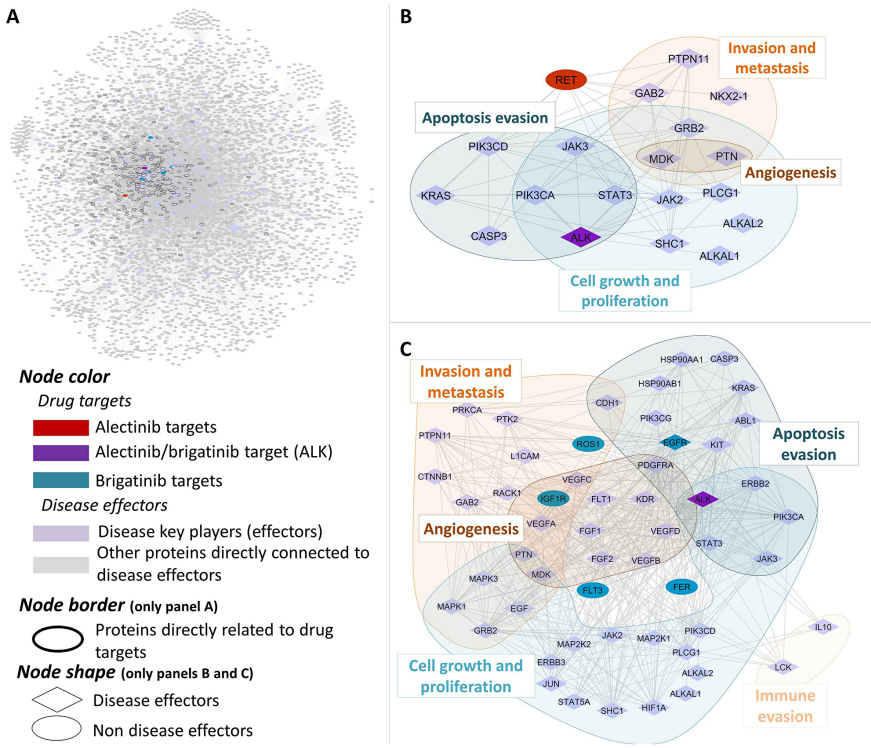

Figure 6: Human protein networks around ALK+ NSCLC molecular pathophysiology.
General overview (A) and centered on main disease players (indicating their pathophysiological motives) and their relationship to alectinib (B) and brigatinib (C) drug targets.

Table 2: Effect of brigatinib's and alectinib's drug targets in ALK+ NSCLC.
Percentage of effectors reversed indicate the proportion of proteins in each motive with a significant difference (FDR < 0.05) and stronger modulation considering the total number of effectors affected in the mechanism of action (MoA) for each drug.

Pathophysiological processes (motives)	Brigatinib's targets						Alectinib's targets	
	ALK	*FLT3*	*FER*	*ROS1*	*IGF1R*	*EGFR*	*ALK*	*RET*
NSCLC ALK$^+$	++	+	+	+	+	+++	++	+
Cell growth and proliferation	+++	++	++	+	+++	+++	+++	+
Evading apoptosis	+++	++	+	+	++	++++	+++	+
Sustained angiogenesis	+	+	++	+	+++	+	+	+
Tissue invasion and metastasis	+	+	++	+	++	++	+	+
Immune evasion	+++	+++	+	+	+	++	+++	++++

affected by ALK inhibition (FLT3, IGF1R, and especially EGFR), as well as angiogenesis and invasiveness through FER and IGF1R inhibition.

Brigatinib and alectinib non-ALK targets affect differently cancer-related processes, including proliferation, apoptosis evasion, invasiveness and immune evasion

The comparison of the predicted mechanisms of action obtained by the mechanistic systems biology modeling using TPMS technology (Figure 7 and Supplementary Table 4) shows that both drugs act through ALK and some overlapping intracellular mechanisms (involving SHC1, GRB2, RASK). However, brigatinib seems to achieve a more diverse downstream effect, through PI3K, ERK and JAK/STAT.

A further evaluation of the impact of each drug on the activity of each protein present in the mechanisms of action, and on the pathophysiological motives previously defined was carried out. This analysis showed that brigatinib, compared to alectinib, has a stronger effect (TSignal) on most of the proteins and all the motives defining ALK+ NSCLC (Figure 8) except in immune evasion (Table 2), for which alectinib presents a greater effect.

Effect of brigatinib and alectinib on invasiveness and central nervous system metastasis

As shown in Table 2, brigatinib was predicted to have a potential stronger effect on metastasis effectors, which are related to invasiveness promotion and metastasis-site characteristics. In order to assess the possible role of each drug on brain metastasis, eight protein/gene effectors known to have a more important role in brain metastasis than in primary tumours were considered: FGFR1 [288, 318, 320]; Ki-67 [418, 270], ROBO1 [391];

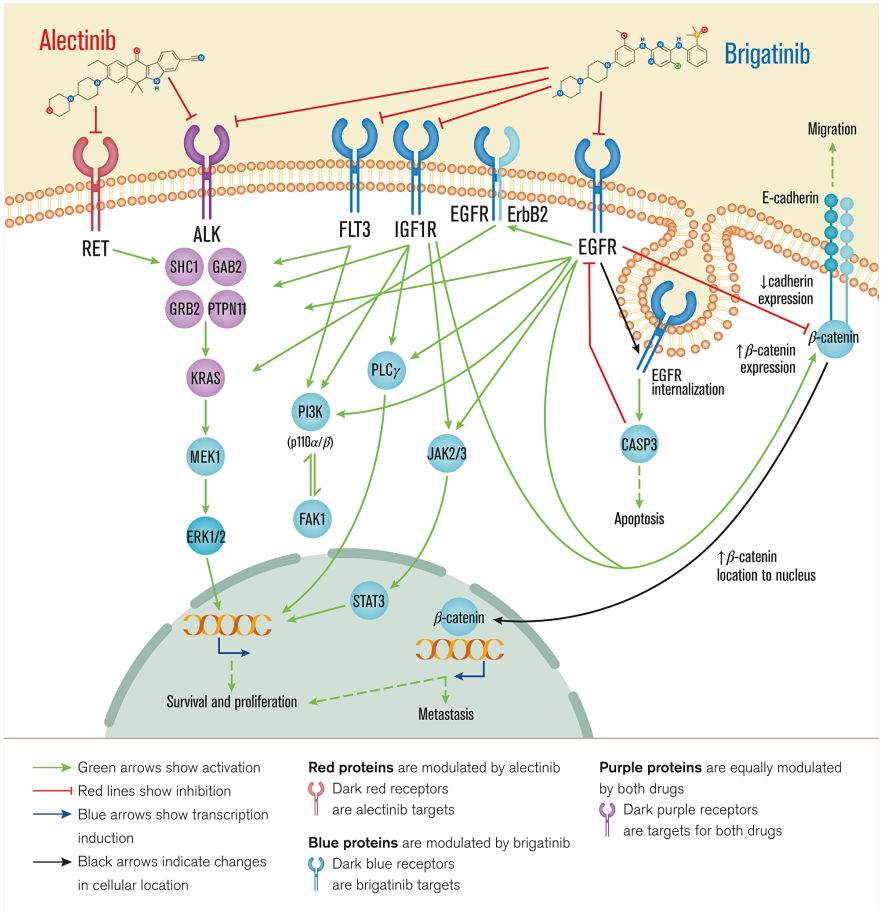

Figure 7: Overview of brigatinib's and alectinib's mechanisms of action.
Receptor targets of each drug are depicted through the cell membrane and the following pathways and pathophysiological motives affected are depicted from the cell surface to the nucleus. Alectinib acts through ALK and RET, involved mainly in survival and proliferation, while brigatinib acts also through ALK and FLT3, IGF1R, and EGFR, signaling through overlapping intracellular mechanisms affecting cell survival and proliferation, metastasis, apoptosis and migration. Bibliographical validation information of interactions on the predicted mechanisms of action are shown in Supplementary data (Supplementary Table 1).

S100A7 [174, 427]; S100B [33, 69]; SIRT1 [159]; SLIT2 [392]; and VEGFA [288, 174]. Out of these, six (Ki-67, ROBO1, S100A7, S100B, SLIT2, VEGFA) were found to be significantly more inhibited by brigatinib than alectinib (FDR < 0.05 and a change in TSignal > 20%). The current

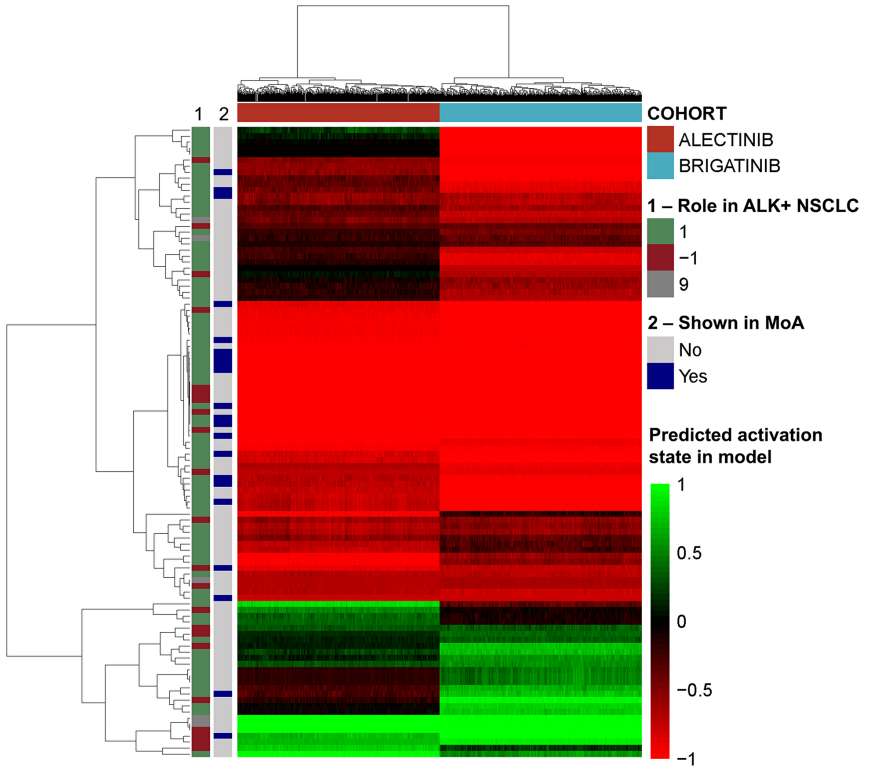

Figure 8: Heatmap of the effect induced by brigatinib and alectinib in each model solution over the effectors of the pathology.
The vertical bars indicate the pathological effect of the effectors (1 if activated in the pathology, -1 if inhibited in the pathology, 9 if complex role) and whether the proteins appear in Figure 7 (“MoA”).

analysis also showed an association between brigatinib and prevention of brain metastasis, mainly through EGFR and IGF1R.

Susceptibility of brigatinib and alectinib to bypass resistance mechanisms

The impact of resistance mechanisms, via protein mutations, on brigatinib’s and alectinib’s mechanisms of action TPMS models was evaluated over a total of 935 proteins and 2805 modifications (activation, inhibition, deletion) (Supplementary Table 5). Of those, 55 different modifications

Table 3: Effect of brigatinib and alectinib over each pathophysiological motive measured by Tsignal.
Percentage of effectors reversed indicate the proportion of proteins in each motive with a significant difference (FDR < 0.05) and stronger modulation considering the total number of effectors affected in the mechanism of action (MoA) for each drug.

Motive	% Effectors more reversed		Drug with highest Tsignal	FDR *T*-Test
	Brigatinib MoA	Alectinib MoA		
Cell growth and proliferation	87%	10%	Brigatinib	< 0.05
Evading apoptosis	69%	27%	Brigatinib	< 0.05
Sustained angiogenesis	65%	29%	Brigatinib	< 0.05
Tissue invasion and metastasis	63%	37%	Brigatinib	< 0.05
Immune evasion	25%	75%	Alectinib	< 0.05

were identified as potential treatment resistances for brigatinib and 93 for alectinib. Among them, 37 were shared between the two drugs. The potential resistance mechanisms that affected alectinib to a greater extent than brigatinib were, mostly, related to the alternative pro-proliferative signaling mechanisms by which NSCLC cells could continue to proliferate. These mechanisms included proteins like MET, ERBB, FGFR, NTRK1 or PDGFR, among others. On the other hand, the potential resistance mechanisms that affected either both drugs or brigatinib to a greater extent than alectinib were mainly potential downstream mediators such as SHC1, KRAS, PI3K or ERK.

Effect of concomitant treatments on the mechanism of action of brigatinib and alectinib

The potential interference with brigatinib or alectinib mechanisms of action was evaluated using a total of 654 drugs (Supplementary Table 6). The drugs that may impact the brigatinib mechanism may include angiotensin-receptor blockers, barbiturates and bisphosphonates. This can be mitigated by adjusting the brigatinib dose. For alectinib, drugs such as non-peptide inhibitors of the antidiuretic hormone can also interfere with

it. However, concomitant use of brigatinib with strong/moderate CYP3A inhibitors/inducers can be managed [392].

3.2.4 Discussion

The second generation ALKi brigatinib and alectinib have demonstrated efficacy in second line treatment in crizotinib-refractory ALK+ NSCLC patients, and in the first-line setting in ALEX (alectinib) and ALTA-1L (brigatinib) clinical trials. In the absence of a head to head trial between brigatinib and alectinib in the first-line setting, and beside the efficacy data and toxicity profile information obtained in independent trials, information from clinical trials in a second-line setting and indirect approaches may help to elucidate the best therapy against ALK+ NSCLC. In the present study, we applied in silico systems biology approaches to compare brigatinib and alectinib as first-line treatment for ALK+ NSCLC at a mechanistic level and thus highlighting the strengths and weaknesses of each ALK inhibitor. The present study indicates that both brigatinib and alectinib could be reasonable choices for first-line treatment, as also previously suggested by other authors [347]. The results obtained by our in silico model allow differentiating between the mechanisms of action of each drug, suggesting that both drugs may have similar efficacy as first-line treatment, and brigatinib may have higher impact in most studied pathways than alectinib. Other specific characteristics were highlighted for each drug.

According to previous publications [10, 347], brigatinib acts as a multi-kinase inhibitor with a broad-spectrum activity against ALK, FLT3, FER, ROS1, IGF1R, and EGFR targets, while alectinib acts on ALK and the proto-oncogene RET [225]. The analyses performed in this study to further determine the differences between brigatinib and alectinib's mechanisms

of action point towards a potentially relevant role of RET, EFGR, IGF1R and FLT3 (besides ALK) in treating NSCLC. All these targets had been previously related to a greater or lesser extent to NSCLC development [382, 40, 449, 186, 249].

As predicted by TPMS analyses (Figure 7), brigatinib targets appear to show a more diverse range of effects compared to alectinib, mediated by ALK inhibition on NSCLC, such as: cell growth and proliferation (as for example STAT3 [330], PI3K [266], K-Ras [330, 282] or erbB2 [75] signaling); evading apoptosis (through EGFR-CASP3 interplay [369]); acting over sustained angiogenesis (IGF1R signaling [163, 294]); and tissue invasion and metastasis processes (modulating the E-cadherin-β-catenin axis [253, 420, 421, 192]). These predicted results and the observed broad range of different effects of brigatinib could be explained by a wider cancer-related target profile of brigatinib. Moreover, it could also be associated with the relatively longer PFS observed with brigatinib in the crizotinib-refractory setting as compared to alectinib [133]. On the other hand, the analyses also suggest that alectinib might have a greater effect on immune evasion regulation through RET inhibition.

Central nervous system (CNS) is one of the most common sites of first progression in ALK+ NSCLC [133]. Even while receiving crizotinib (in around 25–50% of cases), efficacy end points are lower in relation to the CNS than overall [92, 324, 443]. In our study, brigatinib was predicted to have a potentially more robust impact on brain metastasis effectors than alectinib. Inhibition of EGFR might prevent CADH1 reduction mediated by PI3K/FAK1 and thus inhibit tissue invasion. Blocking EGFR and IGF1R pathways might also prevent β-catenin (CTNB1) upregulation, accumulation in the nucleus and transcription factor function. Intracranial

responses to TKIs have also been observed in previous studies. In the phase III trial ALTA-1L (https://clinicaltrials.gov/: NCT02737501) brigatinib was associated with a higher intracranial objective response rate (iORR) (78%) in individuals with ALK TKI-naive ALK+ NSCLC with baseline brain metastases compared to crizotinib (26%) [347]. Alectinib also showed superior intracranial activity versus crizotinib (81% and 50%, respectively) in the ALEX clinical trial, although progression in the brain with both agents has also been observed [131]. This intracranial efficacy is clearly explained by brain bioavailability in the case of alectinib, which shows a very good blood-brain barrier penetration [131], without being affected by MDR1/p-gp modulation [9, 4]. Brigatinib might be susceptible to MDR1/p-gp modulation [244], although no major concerns were raised by the regulatory bodies [13]. Thus, some other factors might explain brigatinib's activity in the brain. The mechanistic study of the current analysis suggests that brigatinib might be able to reverse the activation of a greater percentage of metastasis effectors and, specifically, brain-related metastasis effectors, compared to alectinib. These effectors include the well-known proliferation marker Ki-67 [383] overexpressed in brain metastasis when compared to primary tumours [418, 270]; the ROBO1/SLIT2 axis, increased in brain metastasis [391] and involved in cell migration [227]; the pro-angiogenic VEGFA, related to increased brain metastatic potential [449, 199, 380]; and damage signal proteins (DAMPS) S100 proteins, involved in increased proliferation, anti-apoptotic, and migration capabilities [303, 194]], which are increased in serum of brain metastatic patients and brain metastasis models [33, 69, 449, 303, 194]. Enhanced mechanistic impact over these - and other non brain-specific - metastasis effectors by brigatinib might explain its activity in the brain despite its lower blood-brain barrier penetration. The ongoing ALTA-3 trial

will provide valuable information including intracranial progression after brigatinib versus alectinib in crizotinib-refractory patients that might help better understand their anti-metastatic mechanisms.

The acquisition of resistance to TKI therapy still seems inevitable. However, next generation TKIs are able to more strongly inhibit ALK – both in its wild type form and presenting secondary mutations – suggesting a better control over these progression mechanisms. In fact, brigatinib presents a high selectivity for ALK and low propensity for pharmacological failure [10, 445], showing higher potency than alectinib towards ALK-rearrangement fusions [132, 120, 66]. Brigatinib selectivity over ALK has been also proven in patients with ALK fusion proteins with and without secondary mutations [32, 209]. Little is known about the capacity of ALK TKIs to prevent bypass resistance mechanisms. The evaluation of the impact of developing non-ALK-related resistances on the efficacy of the drugs performed in the current study suggests that alectinib might be more susceptible to bypass resistance mechanisms. The results of our in silico analysis also suggest that brigatinib might block or prevent the development of upstream bypass resistance mechanisms more effectively than alectinib, which could translate into resistance-free treatment for a longer period of time. This would probably occur due to a mechanism of action that reaches a larger number of intracellular effectors involved in ALK-independent resistance mechanisms, including JAK/STAT, MEK/ERK, PI3K or PLCγ [209, 342]. According to our in silico results, brigatinib is predicted to modulate these pathways that are involved in different NSCLC-related pathophysiological processes, more strongly than alectinib. Thus, given the broader impact of brigatinib on ALK secondary mutations compared to other ALK TKIs [10, 445] and the results of the current analysis regarding

bypass mechanisms, it could be hypothesized that brigatinib would prevent the generation of a wider spectrum of resistance mechanisms compared to alectinib. This low resistance predisposition of brigatinib could be related to the efficacy results in terms of PFS observed in the indirect comparison (MAIC) between brigatinib and alectinib/ceritinib by Reckamp [326]. However, further pre-clinical and clinical studies are needed to validate these hypotheses, and ALTA-3, comparing brigatinib to alectinib in ALK+ NSCLC patients who had progressed on crizotinib, might provide interesting conclusions in this regard.

There are two dimensions in which drugs can affect each other: through metabolic and mechanistic interactions. According to the recommendations of the technical specifications [9, 13], whereas both drugs interact with CYP3A – among other enzymes and transporters–, only brigatinib has strict interactions with the usage of inductors, inhibitors and substrates of CYP3A family cytochromes [13, 383]. The current study evaluated the mechanistic interaction between drugs commonly used in cancer patients, regarding the interference of the signal induced by the targets of co-treatments.

According to the current knowledge and the data herein presented, brigatinib might be more prone to present relevant metabolic and mechanistic interactions with other drugs than alectinib, which might be a safer option in poly-treated patients. Use of more than one drug (e.g., to treat cancer or treatment-derived complications, or pre-existing conditions) is common in cancer patients, and polypharmacy (5 or more concomitant drugs) has been shown to occur at a higher frequency in cancer survivors than in non-cancer age- and sex-matched controls [285]. Polypharmacy is especially common among the elderly or in end-of-life settings [238]. Thus, drug interactions must be carefully taken into account when considering different treatment

options. However, as NSCLC adenocarcinoma patients tend to be younger and tend to be non-smokers compared to other cancer patients [120, 336], potential drug interference due to polypharmacy might not represent a determinant factor for treatment selection in clinical practice.

As previously stated, ALKi activity is affected by several factors, including tumour intrinsic characteristics (e.g., ALK fusion gene variants or presence of other primary gene co-mutations) and extrinsic factors (e.g., impact of prior treatments such as presence of ALK secondary mutations, or development of by-pass resistances), and also drug-dependent characteristics (e.g., blood-brain barrier crossing).

The current study aimed to explore mechanistic differences between brigatinib and alectinib that could affect efficacy of both drugs in an in silico approach. However, beside efficacy data, drug toxicity profile is an important determinant of treatment selection. According to previous publications, we considered that although all ALKi present common and specific adverse events, alectinib and brigatinib are similarly well tolerated and can be managed by reducing dose or interrupting treatment [208].

In order to better contextualize the hypotheses raised from the mechanistic analyses, other parameters need to be considered and have been herein discussed (ALK secondary mutations, safety concerns), and must be taken into account in the clinical practice. Besides, in silico modelling approaches can be used as predictive tools and hypothesis generators, limited by the information about diseases and drugs. For example, unknown targets or not yet described pathophysiological processes might have a role in the mechanisms of action of the evaluated drugs. Nevertheless, the models were built by considering the whole human protein network and a wide range of drug-pathology relationships (Supplementary Table 7) [198],

not only limited to NSCLC or oncologic indications, and they present cross-validation accuracies above 80% in the case of ANN models and above 90% in sampling methods-based models. Thus, even if modelling approaches based on systems biology are limited by the amount of available information and some assumptions have to be made, in silico techniques are helpful for understanding fundamental processes in cancer [136, 233]. These approaches allow us to explore investigational or marketed drugs with reduced experimental cost and in different settings. This proves to be especially important if clinical investigations are not going to be done soon or are complex to be conducted, as in the case of the brigatinib versus alectinib head to head study in a first-line setting. Similarly, a comparison to other second and third generation ALKi that have recently shown benefit with respect to crizotinib in the first line setting (ensartinib in eXalt3, NCT02767804, or lorlatinib in CROWN, NCT03052608) could provide further insights into the mechanisms behind ALK+NSCLC treatment. Thus, systems biology and artificial intelligence approaches can contribute to exploring unanswered questions and this may guide the development of ALK TKIs and the identification of the optimal treatment sequence in ALK+ NSCLC patients. Further in silico studies with the aim of identifying the best treatment sequence after brigatinib are ongoing.

3.2.5 Materials and methods

Molecular characterization of ALK+ NSCLC pathophysiology and drugs

To carefully characterize the pathophysiology of ALK+ NSCLC, we conducted an extensive and detailed full-length review of relevant review articles over the last 5 years in the PubMed database (from December

3rd 2013 to December 3rd 2018) using the following search string: ((“ALK+-positive” [TITLE] or “ALK+” [TITLE]) and (“Non-Small Cell Lung Cancer” [TITLE] or “NSCLC” [TITLE]) AND (“MOLECULAR” [TITLE/ABSTRACT] or “PATHOGENESIS” [TITLE/ABSTRACT] or “PATHOPHYSIOLOGY” [TITLE/ABSTRACT]) and Review[ptyp]) and ((“Non-Small Cell Lung Cancer” [TITLE] or “NSCLC” [TITLE]) AND (“MOLECULAR” [TITLE/ABSTRACT] or “PATHOGENESIS” [TITLE/ABSTRACT] or “PATHOPHYSIOLOGY” [TITLE/ABSTRACT]) and Review[ptyp]). The search was also expanded using article reference lists. The main pathophysiological processes (motives) described to be involved in ALK+ NSCLC were identified (Table 2). Subsequently, each motive was further functionally characterized at protein level to determine its molecular effectors. A total of 174 proteins were identified (Supplementary Table 2).

For drug protein target profile definition (brigatinib and alectinib), a dedicated review of databases (DrugBank [425], STITCH [381], SuperTarget [161]) and of scientific literature was performed (Supplementary Table 3).

TPMS technology: systems biology-based model creation

Therapeutic Performance Mapping System (TPMS) (Anaxomics Biotech, Barcelona, Spain) is a top-down systems biology approach based on artificial intelligence and pattern recognition models. This methodology integrates available biological, pharmacological and medical information to generate mathematical models that simulate the mechanisms of action of drugs in a pathophysiological human context (Figure 9). TPMS models are trained using a compendium of biological and clinical data characteristics

of the human physiology (Table 6).

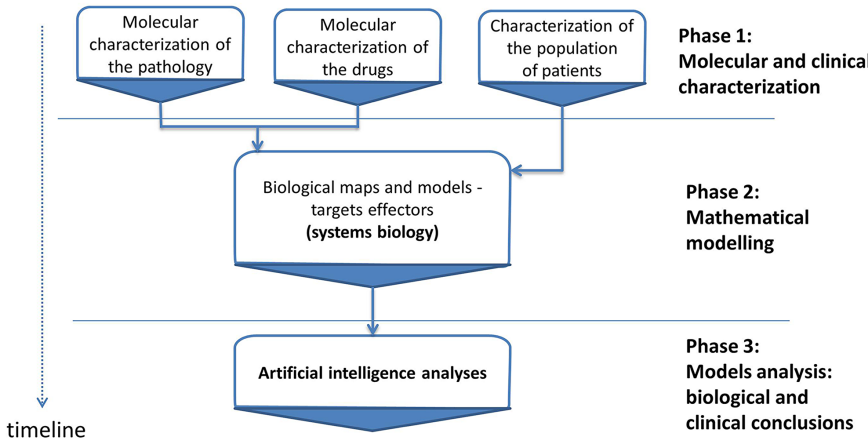

Figure 9: Study workflow.
Overview of the in silico study approach showing the main phases employed to simulate the mechanisms of action (MoA) of brigatinib and alectinib with respect to ALK+NSCLC molecular characterization. TPMS is a validated top-down systems biology approach that integrates all available biological, pharmacological and medical knowledge (protein network, truth table and specific data compilation) by means of pattern recognition models and artificial intelligence to create mathematical models that simulate in silico the behavior of human physiology.

Mechanism of action models

In order to obtain the mechanism of action (MoA) of brigatinib and alectinib, drug-ALK+ NSCLC mathematical models were generated following the same methodology as described in Jorba [198] and applied in previous studies [320, 257, 181]. As input, TPMS takes the activation (+1) and inactivation (–1) of the drug target proteins (Supplementary Table 3), and as output the protein states of the pathology of interest (Supplementary Table 2). It then optimizes the paths between both protein sets and computes the activation and inactivation values of the full human interactome. The resulting subnetwork of proteins with non-null outputs and their values will define the MoA of the drug. The impact of each drug over the activity of the proteins effectors of the pathophysiological disease was quantified using the

Tsignal (i.e., the average signal values of the protein effectors), as described in Jorba [198]. More detailed information on the modelling methodology can be found in Supplementary Methods.

Sobol sensibility analysis

In order to analyze the impact of the noise in the final MoAs affecting the biological conclusions reached, a Sobol sensibility analysis was performed over the constructed TPMS mathematical models [446]. Detailed of the analysis implementation can be found in Supplementary Methods.

Drug-(patho)physiology motive relation finding

Artificial neural networks (ANN) were used to identify relations between proteins (e.g., drug targets) and clinical elements of the network [47], which is an approach previously used and validated by several publications [325, 119, 164, 165]. This strategy was used to perform an efficacy evaluation of brigatinib and alectinib from each of their targets towards ALK+ NSCLC pathophysiological motives and its corresponding proteins. Detailed information on the modelling methodology can be found in Supplementary Methods.

Evaluation of the impact of potential resistances over the mechanisms of action

In order to identify possible cancer resistances, the TPMS models were evaluated for possible mutations to identify the key nodes or proteins with higher impact on the effector proteins TSignal. Because both brigatinib and alectinib mechanisms of action had a vast amount of proteins or nodes (> 5000), the universe of possible key nodes was reduced to the list of proteins around ALK+ NSCLC effectors and around the drugs' target proteins.

To evaluate the addition of a mutation in the system, the impact of a protein activation, inhibition and deletion over the mechanisms of action of brigatinib and alectinib was tested. To do so, the resulting TSignal of the altered models was computed and compared to the original one. Finally, the p-value of the difference between the Tsignals over ALK+ NSCLC with and without the mutation was calculated, and the ones with p-values ≤ 0.022 were selected (Supplementary Table 5).

Evaluation of drug interferences over the mechanisms of action

To identify possible co-treatment interferences, a list of pharmacological treatments potentially co-administered with brigatinib or alectinib was created and evaluated. To do so, we generated a list of all treatments for common conditions (either in the general population and in the ALK+ NSCLC population) and treatments for brigatinib/alectinib-associated adverse drug reactions, according to DrugBank database [425]. After that, the mechanisms of action of brigatinib and alectinib was perturbed by activating the co-treatment protein targets, each drug one by one, and the TSignal was computed. Finally, the differences in the TSignal between the original and the perturbed system and the corresponding p-values were calculated, and the ones with p-values < 0.1 and 0.05 were selected (Supplementary Table 6).

3.2.6 Conclusions

An in silico head to head based on the mechanism of action evaluation between brigatinib and alectinib has been performed highlighting the advantages of using one before the other from an efficacy point of view. Brigatinib appears to have a wider mechanism of action, presenting targets that potentially act more strongly in most of the ALK+ NSCLC

pathophysiological pathways, including invasiveness to the CNS. On the other side, alectinib-induced RET inhibition might contribute to reducing the tumour immune evasion mechanisms. In general, both drugs are known to be well-tolerated and, although shown and predicted to have a similar efficacy for the treatment of ALK+ NSCLC in a first-line setting, the differences in their target profiles might allow for identification, in subsequent studies, of different patient profiles that might benefit from either of them, beside considering potential safety concerns in specific patient subpopulations. Future clinical studies will be needed to confirm these findings. The used approach can be applied for the evaluation of other next-generation ALKi, even if not yet approved, or exploring other questions, such as optimal treatment sequence.

3.2.7 Author contributions

AFN, AL and EC contributed to setting up fundamental questions regarding ALK+ NSCLC treatment. EC, CSV and MC contributed to the study design. AFN, AL, CSV, CM and AN contributed to data acquisition. CSV, MC, GJ, JMM and BO contributed to data analysis. EC contributed to clinical interpretation of data. AFN, AL, CM, AN and CSV contributed to writing the manuscript. EC and MC contributed to the critical revision of the manuscript.

3.2.8 Acknowledgements

The authors thank Cristina Lorca-Oró for her assistance in writing and editing the manuscript and Pedro Filipe for his assistance in the sensitivity analysis. Medical writing was funded by Takeda Farmacéutica España.

3.2.9 Conflicts of interest

EC: Speaker or advisory board: Takeda, Astra Zeneca, Roche, MSD, Novartis, Boheringer Ingelheim, Pfizer, BMS; Travel expenses: Roche, Pfizer, Takeda, BMS. AL, CM, AN and AFN are full time employees of Takeda Farmacéutica España S.A. CSV, GJ, MC and JMM are full time employees of Anaxomics Biotech. BO declares no competing interests.

3.2.10 Funding

The study was funded by Takeda Farmacéutica Spain. GJ has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement (ref: 765912).

3.2.11 Supplementary files

www.ncbi.nlm.nih.gov/pmc/articles/PMC7899557/bin/oncotarget-12-316-s001.pdf

www.ncbi.nlm.nih.gov/pmc/articles/PMC7899557/bin/oncotarget-12-316-s002.xlsx

3.2.12 Supplementary materials

TPMS technology: systems biology-based model creation and analysis for ALK+ NSCLC

Systems biology-based models were created using the Therapeutic Performance Mapping System (TPMS) to investigate the molecular Mechanisms of Action (MoA) of brigatinib and alectinib towards the modulation of ALK+ NSCLC. TPMS is a validated top-down systems

biology approach that integrates all available biological, pharmacological and medical knowledge by means of pattern recognition models and artificial intelligence to create mathematical models that simulate in silico the behavior of human physiology. The methodology employed and detailed herein has been previously described [198] and applied elsewhere [339, 257, 181, 165]. Biological maps were transformed into a mathematical model capable of both reproducing existing knowledge and predicting new data. TPMS technology uses a set of artificial intelligence algorithms to generate the human physiology over the human biological network [339, 220, 87, 346].

Human protein network (HPN) and truth table construction

A human protein network (HPN) was created in order to obtain the MoAs by including information from many public and private databases (KEGG [204, 205], REACTOME [93], INTACT [300], BIOGRID [349], HPRD [319], and TRRUST [158] and information extracted from scientific literature.

For the construction of the truth table, a selected collection of known input-output physiological signals considered the "truths" were collated into a table (Supplementary Table 7) and was used for training the models [19]. The truth table was based on a compendium of different databases that contain biological and clinical data [426, 232] and provides biological and pharmacological input-output relationships (such as drug indication pairs). Information relating biological processes (adverse drug reactions, indications, diseases and molecular pathways) to their molecular effectors, i.e., each one of the proteins involved in the physiological process, was extracted from the biological effectors database (BED) (Anaxomics Biotech SL, [198]). The biological or pathological conditions under study were also

included in the truth table and molecularly characterized through specific scientific literature search and hand-curated assignment of proteins to the conditions (Supplementary Table 2). The obtained final models had to be able to reproduce every rule contained in the truth table, and we defined the error of a model as the percentage of all the rules with which the model does not comply, while the accuracy was defined as the percentage of all the rules complied with.

Modelling strategies

Two complementary modelling strategies were used, (a) TPMS Artificial Neural Networks (ANNs) [339] and (b) TPMS Sampling-based Methods [198], to compare the efficacy of the drugs (defined as their targets, see Supplementary Table 3) and to compute the MoA models.

(a)ANNs are supervised algorithms that identify relations between proteins (e.g., drug targets) and clinical elements of a protein network [339, 47, 164, 269, 321] by inferring the probability of the existence of a specific relationship between two or more protein sets, based on the validation of the predictive capacity of the model towards the truth table. The learning methodology used consisted in an architecture of stratified ensembles of neural networks as a model, trained with a gradient descent algorithm to approximate the values of the given truth table. The neural network model used consisted in a Multilayer Perceptron (MLP) neural network classifier. MLP gradient descent training depends on randomization initialization and to avoid random errors 1000 MLPs are trained with the training subset and the best 100 MLPs are used. In order to correctly predict the effect of a drug independently of the number of targets, a different ensemble of neural networks are trained for a different subset of drugs according to their number

of targets (drugs with 1 target, 2 targets, 3 targets). Then, the predictions for a query drug are calculated by all the ensembles, and pondered according to the number of targets of the query drug (the difference between the number of targets of the query and the number of targets of the drugs used to calculate each ensemble is used to ponder the result of each ensemble). A cross-validation with the truth table information showed that the accuracy of the described ANNs to reproduce the indications compiled in DrugBank [426, 425] is 81.7% for those drugs with all targets in the human biological network.

(b)Sampling-based methods generate models similar to a MLP over the previously constructed HPN, where neurons are the proteins and the edges of the network are used to transfer the information (Figure 10) [198]. This methodology was used for describing with high capability all plausible relationships between an input (or stimulus) and an output (or response). Sampling-based methods use optimization algorithms [87] to solve each parameter of the equation, i.e. the weights associated to the links between the nodes in the human protein network. In this approach, the network is limited by considering only interactions that connect drug targets with protein effectors in a maximum of three steps. The values of activation (+1) and inactivation (–1) of the protein targets of the drugs in the truth table were considered as input signals whereas the output is defined as the values of activation and inactivation of the proteins describing the phenotype (as retrieved from the BED). Each node of the protein network receives as input the output of the connected nodes in the direction flow from targets to effectors, weighted by each link weight (Figure 10). The sum of inputs is transformed by a hyperbolic tangent function to generate the score of the node (neuron), which becomes the 'output signal' of the

current node towards the nodes. The weight parameters are obtained by Stochastic Optimization Method based on Simulated Annealing [87], which uses probabilistic measures derived from the biological evidence to adjust network interaction types and strengths. Since the number of entries in the truth table is always smaller than the number of parameters (link weights) required by the algorithm, any process modelled by TPMS considers a population of different solutions.

Mechanisms of action elucidation

The MoAs obtained with the TPMS simulates potential interactions between drug targets and protein effectors associated to prototype-ALK+ NSCLC patients. In order to validate this approach, the intensity of the model's response, divided in TSignal and number of protein effectors activated, was used to understand the relationships between all potential mechanisms and compare sets of MoAs from different views (Figure 10) [198].

Intensity of the response

We defined the "intensity" of the response as follows: 1) the quantity of protein effectors (#) that reach an expected signal sign; and 2) the strength or amount of the output signal reaching the effectors (i.e., a global measure of the output signal, named TSignal). Given a protein effector "i", which reaches a signal value yi , and vi being the effector sign according to the BED (active or inactive) and n is the total number of effectors described for a phenotype, it was determined:

Number of effectors achieving the expected sign

Assuming that a drug may be able to activate/ inactivate protein effectors reverting a disease/indication model phenotype. Using Dirac's δ (i.e.

$\delta(0) = 1$, and zero otherwise), the equation to calculate number of effectors achieving the expected sign for drug indications was defined as:

$\#_{indication} = sum_n^{i=1}\delta\left(v_i + \frac{y_i}{|y_i|}\right)$

TSignal

The average output values of the protein effectors. For each effector, it was counted as positive signal if the sign is correct, and negative otherwise. When a drug affects a disease phenotype, vi and yi have opposite sign and it is necessary to change the sign in the corresponding equation:

$TSignal_{indication} = -\frac{1}{n}\sum_{i=1}^{n} v_i y_i$

Sobol sensitivity analysis

An adapted methodology for ensembles of high dimensional algorithms was applied following the definition of Sobol Sensitive Analysis [446]. According to the Sobol terminology, TPMS models can be redefined as follows:

$TSignal = TPMS(X) for X = \{X_1, X_2 \dots X_n\}$

Where xi is each of the parameters used in the TPMS models. Then, the variation of TSignal for each xi parameter can be expressed as:

$\frac{dTSignal}{dX_i} = \frac{dTPMS(X)}{dX_i}$

Consequently, the variation of the simultaneous parameters xi and xj can be estimated as:

$\frac{dTSignal}{dX_i dX_j} = \frac{dTPMS(X)}{dX_i dX_j}$

Using the previous equation descriptions, we measured the impact of varying random parameters over output TSignal in two different approaches, those being local analysis and global analysis [446].

Local sensitivity analysis

Local sensitivity analysis evaluates changes in the model outputs (TSignal)

Table 4: Pathophysiological processes (motives) and number of seeds proteins of ALK+ NSCLC

MOTIVE #	Motive name	# proteins
1	CELL GROWTH AND PROLIFERATION	46
2	SUSTAINED ANGIOGENESIS	20
3	EVADING APOPTOSIS	50
4	TISSUE INVASION AND METASTASIS	73
5	IMMUNE EVASION	16

Table 5: Brigatinib and alectinib characterized protein targets

Drug name	Uniprot ID	Gene Name	Effect	DrugBank candidate	Stitch candidate	Supertarget candidate	References
Alectinib	Q9UM73	ALK	-1	yes	yes	no	PMID: 28455243
Alectinib	P07949	RET	-1	no	yes	no	PMID: 25349307
Brigatinib	Q9UM73	ALK	-1	yes	yes	no	PMID: 2714483; 29075144
Brigatinib	P00533	EGFR*	-1	yes	no	no	PMID: 29451020; 29075144; 28287083
Brigatinib	P36888	FLT3	-1	yes	no	no	PMID: 27144831; 29451020
Brigatinib	P08922	ROS1	-1	no	no	no	PMID: 28680831; 29451020
Brigatinib	P16591	FER	-1	no	no	no	PMID: 29540831; FDA Multi-discipline review
Brigatinib	P08069	IGF1R	-1	yes	no	no	PMID: 27144831; 29451020; 29075144; 29403310

*Highest affinity for EGFR (L858R) mutated form.

with respect to variations in a single model parameter. This effect was measured in the TPMS-models for both alectinib and brigatinib MoA models (Figure 11).

Global sensitivity analysis

In the global sensitivity analysis, all parameters are varied simultaneously over the entire parameter space to measure the effects of their interactions on the model output. Given the high dimensionality of the TPMS ensemble models, this measure has been estimated by a MonteCarlo experiment to introduce random values (noise) in sets of 1200 candidate parameters. These final TSignal effects were measured by altering combinations of parameters in subsets of 1, 2, 3, 4, 5, 10, 15, 20 and 30 parameters simultaneously, from the candidate parameters list (Figure 12).

Table 6: Summary of data used for model construction (Human Protein Network (HPN) and truth table)

Entry type		# Entries
In-house databases information		
	Considered Interactions	437.071
	Considered Proteins	16.961
	Characterized Drugs	5.414
	Drug Targets	2,690
	Characterized Clinical Conditions	253
	Clinical Conditions Key Proteins Characterized	4.076
Truth table information		
	Curated drug-indications restrictions	180,264 (1,731 positive)
	Drug-ADRs restrictions	30,096 (2,460 positive)
	Drug-indications/ADRs protein correlations	2.175

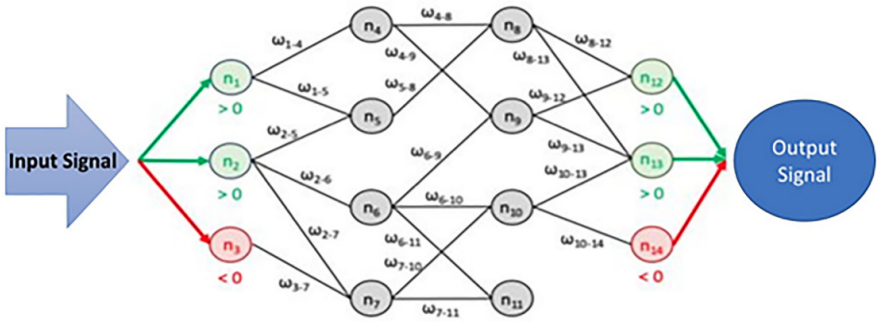

Figure 10: TPMS schematic representation of the input/output signals information over the Human Protein Network (HPN) using a Multilayer Perceptron-like and sampling method to predict the Mechanisms of Action (MoAs) of a drug

Sensitivity results

Although TPMS-models have about 5000 parameters, only a small percentage of them showed a real impact on the output, which was less notorious in brigatinib than alectinib (Figures 11 and 12). Nevertheless, the impact of some of the protein parameters are of great importance, meaning that TPMS models had to carefully adjust to all the restrictions defines in the truth table, while completing the drug-pathology model. We can see this as most protein parameters are actually part of the ALK+ NSCLC effectors (like P27361, P28482 and P414921, among others), which will definitely

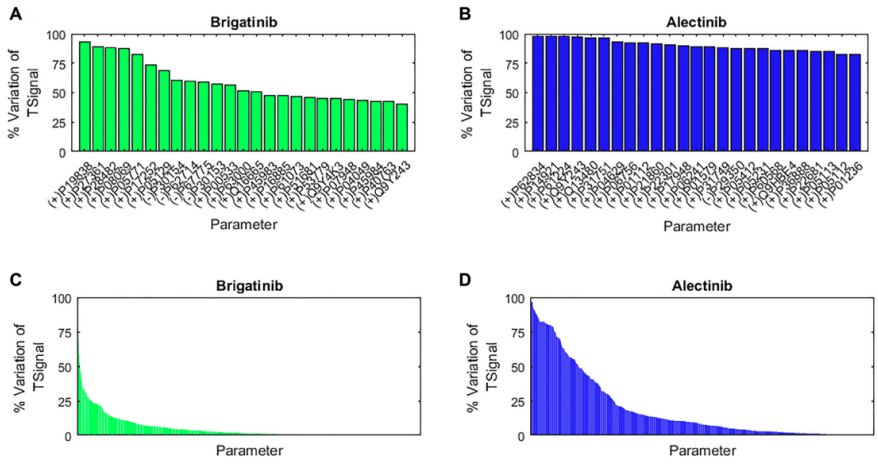

Figure 11: MoA models difference of TSignal measured for each individual parameters variation.
(A and B) show the % of variation of the output TSignal for the 25 most sensible parameters in brigatinib and alectinib MoA models, respectivelty. (C and D) show the % of variation of the output TSignal for all parameters in brigatinib and alectinib MoA models, respectively. Parameters are ordered from the ones affecting the most to the models, to the ones affecting the less. For the sake of visual simplicity, the parameter names of C and D are not displayed in the x axis.

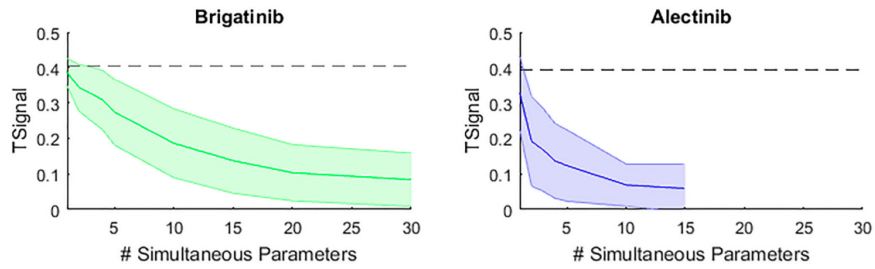

Figure 12: MoA models TSignal measured for individual and multiple parameters variation.
The output TSignal when variating 1 until 30 random parameters simultaneously in brigatinib and alectinib MoA models, respectively, is shown. The gray, dashed line represents the original TSignal values for each of the drugs MoA.

have a huge effect on the final TSignal according to its definition in Equation 2.

3.3 Discussion

Both alectinin and brigatinib, second generation ALK tyrosine kinase inhibitors (TKIs), are drugs that had already demonstrated good efficacy in second line treatments in patients with ALK+ non-squamous cell lung carcinoma (NSCLC) in previous clinical trials, ALEX and ALTA-1L respectively [60, 63]. However, no head-to-head study to compare the two drugs had been performed. Here, TPMS systems biology and ANN approaches were used to compute both mechanisms of action (MoA) and explore their mechanistic differences in silico.

3.3.1 TPMS models

Systems biology models were constructed with TPMS technology in order to mechanistically explain the relation between brigatinib and alectinib drugs with NSCLC. Those models were able to explain the MoA linking each drug with the study disease, in order for posterior comparison of highlighted pathways. The systems accuracy were above 94%, that is, the percentage of truth table or known biological relationships complied after model optimization.

A Sobol sensibility analysis was performed over the constructed models in order to analyze the impact of noise in the final MoAs and the weight of each parameter or protein. Although less notorious in brigatinib than alectinib, a portion of them showed a real impact on the output, highlighting their importance in the predicted pathways.

3.3.2 ANNs

Artificial neural network models were constructed over the same protein network as the MoA models, but with the aim of detecting relationship between an input protein set, stimulus (e.g., drugs), and a response (e.g., disease).

These networks are build using topological and functional measures of the protein nodes, and after being trained with known protein relationship data, can be used to predict protein sets relationships. The resulting models were able to reproduce more than 80% of the training set data.

3.3.3 Response effect

Using the ANN response profiles, both drugs appear to show similar intensity effect towards ALK+ NSCLC, with brigatinib presenting a slightly higher relation to the disease effectors. However, when focusing on the specific activated pathways, britgatinib showed a wider interplay of functions compared to alectinib, such as: (i) cell growth and proliferation; (ii) apoptosis evasion; (iii) sustained angiogenesis; and (iv) tissue invasion and metastasis processes. Those functions were linked to its more diverse cancer-related target profile, which correlate with previous findings associating brigatinib to relatively longer progression free survival times compared to alectinib [133].

3.3.4 Brain metastasis

Central nervous system progression in ALK+ NSCLC patients is rather common [134]. In previous studies, both drugs had already showed higher intracranial objective response rate (iORR) than another NSCLC drug, crizotinib, in individuals with ALK TKI-naive ALK+ NSCLC [347, 131].

Here, when analyzing both drug models, brigatinib was predicted to be less prone to brain metastasis, which could help explain its prevention levels despite having low BBB penetration. Specifically, we found that brigatinib targets are able to reach brain-related metastasis proteins, including: (i) the well-known proliferation marker Ki-67 [383]; (ii) the ROBO1/SLIT2 axis, increased in brain metastasis [391] and involved in cell migration [227]; (iii) the pro-angiogenic VEGFA, related to increased brain metastatic potential [449, 199, 380]; and (iv) damage signal proteins (DAMPS) S100 proteins, involved in increased proliferation, anti-apoptotic, and migration capabilities [303, 194], which are increased in serum of brain metastatic patients and brain metastasis models [33, 69, 449, 303, 194]. In contrast, alectinib intracranial efficacy can be easily explained by its elevated blood-brain barrier (BBB) penetration [131].

3.3.5 Drug resistances

Although these two next generation TKIs are able to strongly inhibit ALK, both in its wild type form and with secondary mutations, resistance to TKI therapy is still an issue [32, 209]. In that sense, little is known about the capacity of ALK TKIs to prevent bypass resistance mechanisms. For that, we use both TPMS models to test the effect of random single or multiple pinpoint mutations, by activating or inhibiting proteins from the network systems. The in silico evaluation of the robustness towards non-ALK-related resistances suggests that brigatinib might be able to block or prevent the development of upstream bypass resistance mechanisms more effectively than alectinib. This could translate into resistance-free treatment for a longer period of time. That effect was predicted to occur due to a mechanism of action that reaches a larger number of intracellular effectors involved in ALK-independent resistance mechanisms, including

JAK/STAT, MEK/ERK, PI3K or PLCγ [209, 342]. The effect was also seen on the Sobol analyses, were alectinib presented more sensitive proteins. So, brigatinib would modulate these pathways that are involved in different NSCLC-related pathophysiological processes more robustly than alectinib.

3.3.6 Drug-drug interactions

The usage of more than one drug (e.g., to treat cancer or treatment-derived complications, or pre-existing conditions) is common in cancer patients, and polypharmacy (5 or more concomitant drugs) has been shown to occur at a higher frequency in cancer survivors than in non-cancer age- and sex-matched controls [285]. Whereas both drugs interact with CYP3A, among other enzymes and transporters, only brigatinib has been described to interact with other inductors, inhibitors and substrates of CYP3A family cytochromes [13, 383].

In this study, the generated models were evaluated with an array of over 600 drugs. The results indicated that brigatinib might be more prone to drug interactions, including angiotensin-receptor blockers, barbiturates and bisphosphonate; hence alectinib, which still shoed interaction with non-peptide inhibitors of the antidiuretic hormone, may be a safer option in poly-treated patients.

3.4 Concluding remarks

TPMS systems biology models allowed a fast in silico evaluation and comparison of alectinib and brigatinib as treatments for ALK+ NSCLC. The results were centered on the MoAs of both drugs but not their physiological properties which, in the case of brain metastasis pharmacological action, helped understand the similar effect of both drugs although presenting

different BBB penetration ratios.

Moreover, these MoA models are resourceful in the sense that not only the drug's highlighted pathways could be investigated, but also drug resistances and interactions could be evaluated. So, these types of models come handy when investigating with complex diseases like cancer were many signaling pathways take part in the pathogenesis [323].

Chapter 4

Systems biology network approaches to simulate and analyze prototype-patients

4.1 Introduction

In collaboration with Joaquim Aguirre-Plans, the other first co-author of the following article and investigating at GRIB (Research Programme on Biomedical Informatics), we applied the systems biology network tools of TPMS [18] and GUILDify v2.0 [6] to model the response of the drug combination sacubitril+valsartan towards the phenotypes of heart failure (HF) and macular degeneration (MD), using the theoretical models as "prototype-patients".

Briefly, TPMS generates an ensemble of solutions that can include clusters of solutions defining an array of MoAs for a particular stimulus-response relationship. Because TPMS searches for all MoA pathway possibilities,

resulting in not one but several MoA pathways evaluated as a whole, that could be extrapolated to the response of different patients, here named prototype-patients. The sampling system allows for the finding of pathways that may not appear on a one-solution approach used in other models and comes handy for relating to distinct patient responses; specially in heart related diseases, where much variability is observed [254]. It is also a step forward precision or personalized medicine, were individual patients are modeled. So, we assumed that MoAs associated to good or bad patient responders to the treatment can be identified by their intensity of model's outcome measure, namely tSignal. We stratified prototype-patients in two categories, best and worst, belonging to the highest 25% solutions or prototype-patients, and worst 25%. Then, we identified several biomarker proteins that allowed to differentiate such prototype-patients in best vs worst responders. Posteriorly, we applied GUILDify v2.0, which uses diffusion-based algorithms to identify the modules associated to a disease or side effect, to detect the disease components of heart failure and macular degeneration, assess how the treatment combination target proteins' signal were overlapping in both phenotypes, and search the biomarker proteins identified by TPMS in this context.

4.2 In-silico simulated prototype-patients using TPMS technology to study a potential adverse effect of sacubitril and valsartan

PLoS One. 2020 Feb 13;15(2):e0228926.

PMID: 32053711

PMCID: PMC7018085

DOI: 10.1371/journal.pone.0228926

Guillem Jorba[1,2,#], Joaquim Aguirre-Plans[2,#], Valentin Junet[1,3], Cristina Segú-Vergés[1], José Luis Ruiz[1], Albert Pujol[1], Narcis Fernandez-Fuentes[4], José Manuel Mas[1,*], Baldo Oliva[2,*]

[1]Anaxomics Biotech SL, Barcelona 08008, Catalonia, Spain

[2]Structural Bioinformatics Group, Research Programme on Biomedical Informatics, Department of Experimental and Health Science, Universitat Pompeu Fabra, Barcelona 08003, Catalonia, Spain

[3]Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193, Catalonia, Spain

[4]Department of Biosciences, U Science Tech, Universitat de Vic-Universitat Central de Catalunya, Vic 08500, Catalonia, Spain

[#]These authors contributed equally to this work and are considered to be co-first authors

[*]Corresponding authors

4.2.1 Abstract

Unveiling the mechanism of action of a drug is key to understand the benefits and adverse reactions of a medication in an organism. However, in complex diseases such as heart diseases there is not a unique mechanism of action but a wide range of different responses depending on the patient. Exploring this collection of mechanisms is one of the clues for a future personalized medicine. The Therapeutic Performance Mapping System (TPMS) is a Systems Biology approach that generates multiple models of the mechanism of action of a drug. Each molecular mechanism generated could be associated to particular individuals, here defined as prototype-patients, hence the generation of models using TPMS technology may be used for detecting adverse effects to specific patients.

TPMS operates by (1) modelling the responses in humans with an accurate description of a protein network and (2) applying a Multilayer Perceptron-like and sampling strategy to find all plausible solutions. In the present study, TPMS is applied to explore the diversity of mechanisms of action of the drug combination sacubitril/valsartan. We use TPMS to generate a wide range of models explaining the relationship between sacubitril/valsartan and heart failure (the indication), as well as evaluating their association with macular degeneration (a potential adverse effect). Among the models generated, we identify a set of mechanisms of action associated to a better response in terms of heart failure treatment, which could also be associated to macular degeneration development. Finally, a set of 30 potential biomarkers are proposed to identify mechanisms (or prototype-patients) more prone of suffering macular degeneration when presenting good heart failure response. All prototype-patients models generated are completely theoretical and therefore they do not necessarily involve clinical effects in real patients. Data and accession to software are available at http://sbi.upf.edu/data/tpms/.

4.2.2 Introduction

Systems biology methods are an increasingly recurring strategy to understand the molecular effects of a drug in complex clinical settings [304]. Some of these methods apply computer science techniques and mathematical approaches to simulate the responses of a drug. In 2005, the Virtual Physiological Human initiative was founded with the objective of developing computational models of patients [401]. Later, they defined the concept of In Silico Clinical Trials as “the use of individualized computer simulation in the development or regulatory evaluation of a medicinal product, medical device, or medical intervention” [403]. Since then, In

Silico Clinical Trials have been adopted in several occasions in preclinical and clinical trials [304].

However, current methodologies do not consider the inter-patient variability intrinsic to pharmacological treatments, missing relevant information that should be incorporated into the models. Indeed, there are many parameters influencing the Mechanisms of Action (MoA) in such therapies, including demographic data of the patient, co-treatments or clinical history. Thus, by modelling all molecular mechanisms affected by the drug, the diversity of responses observed in patients during or after the treatment could be explained.

The Therapeutic Performance Mapping System (TPMS) [18] is a method used to elucidate all the possible MoAs that could exist between an input drug and a pathology or adverse effect. It is a systems biology approach based on the simulation of patient-specific protein-protein interaction networks. TPMS incorporates data from different resources and uses the information from the drugs and diseases under study to generate multiple models of potential MoAs. In the last years, TPMS has been broadly used in different clinical areas and with different objectives [321, 164, 148, 309, 181, 339, 257, 182], in some cases being validated in the posterior experiments [164, 257, 182]. Our working hypothesis is that a set of MoAs can represent the different responses to a drug in cells and that a real population of patients is the result of a myriad of cell responses. Thus, we define a prototype-patient as an abstract case with all cells responding to a single MoA.

Here, we propose the application of TPMS and protein-network approaches in the specific case study of the drug combination sacubitril/valsartan, used for the treatment of Heart Failure (HF). HF is becoming a major

health problem in the western world due to its increasing hospitalization rates [101], with a prevalence being influenced by many factors like age, nutritional habits, lifestyles or genetics. This complicates the development of treatments and the identification of universal biomarkers to stratify the population. To facilitate this segmentation, it is necessary to understand the molecular details of the treatment and the pathology. Sacubitril/valsartan (marketed by Novartis as Entresto®) is a drug combination that shows better results than conventional treatments by reducing cardiovascular deaths and heart failure (HF) readmissions [274]. In pharmacological terms, it is an angiotensin receptor-neprilysin inhibitor. Consequently, it triggers the natriuretic peptide system by inhibiting neprilysin (NEP) and inhibits renin-angiotensin-aldosterone system by blocking the type-1 angiotensin II receptor (AT1R) [359]. In a previous work, TPMS was already applied to unveil the MoA of sacubitril/valsartan synergy, revealing its effect against two molecular processes [181]: the left ventricular extracellular matrix remodeling, mediated by proteins like gap junction alpha-1 protein or matrix metalloproteinase-9; and the cardiomyocyte apoptosis, through modulation of glycogen synthase kinase-3 beta. However, several publications warned about the potential long-term negative implications of using a neprilysin inhibitor like sacubitril [359, 317, 115, 331, 448]. Neprilysin plays a critical role at maintaining the amyloid-β homeostasis in the brain, and the alteration of amyloid-β levels has been linked to a potential long-term development of Alzheimer's disease or Macular Degeneration (MD) [359, 115, 448, 29, 297]. During the clinical trials PARADIGM-HF and PARAGON-HF with sacubitril/valsartan no serious effects were detected [274, 368]. Still, their patient follow-up was relatively short and not specialized in finding neurodegenerative specific symptoms. For this reason, in a forthcoming PERSPECTIVE trial (NCT02884206) a battery of

cognitive tests was taken [331]. In line with this, the application of systems biology methods may shed light to the potential relationship between the treatment and the adverse effect.

In this study, we used TPMS and GUILDify v2.0 to analyze the relationship between sacubitril/valsartan, HF and MD in entirely theoretical models. Because these are theoretical models it is important to note that they are not associated with clinical effects in real patients, they only point on potential mechanisms to explain potential adverse effects. We analyzed a population of MoAs that describe the possible protein links from a sacubitril/valsartan treatment to HF and MD phenotypes. We clustered the MoAs in groups according to their response intensity and labelled them as high or low efficacy of treating HF and possibility of causing MD. We then compared these sets of MoAs and proposed a list of biomarkers to identify potential cases of MD when using sacubitril/valsartan. Simultaneously, we used GUILDify v2.0 web server [6] as an alternative approach to compare the biomarkers proposed by TPMS and reinforce the results.

4.2.3 Materials and methods

Biological Effectors Database (BED) to molecularly describe specific clinical conditions

Biological Effectors Database (BED) [321, 363] describes more than 300 clinical conditions as sets of genes and proteins (effectors) that can be "active", "inactive" or "neutral". For example, in a metabolic protein-like network, an enzyme will become "active" in the presence of a catalyst, or become inactivated when interacting with an inhibitor (see further details in supplementary material).

TPMS modelling

The Therapeutic Performance Mapping System (TPMS) is a tool that creates mathematical models of the protein pathways underlying a drug/pathology to explain a clinical outcome or phenotype [18, 321, 164, 148, 309, 181, 339]. These models find MoAs that explain how a Stimulus (i.e. proteins activated or inhibited by a drug) produces a Response (i.e. proteins active or inhibited in a phenotype). In the present case study, we applied TPMS to the drug-indication pair sacubitril/valsartan and HF. Regarding the drug, we retrieved the sacubitril/valsartan targets from DrugBank [425], PubChem [217], STITCH [381], SuperTarget [161] and hand curated literature revision. As for the indication, we retrieved the proteins associated with the phenotype from the BED [321, 363].

Building the Human Protein Network (HPN)

To apply the TPMS approach and create the mathematical models of MoAs, a Human Protein Network (HPN) is needed beforehand. In this study, we used a protein-protein interactions network created from the integration of public and private databases: KEGG [205], BioGRID [68], IntAct [300], REACTOME [108], TRRUST [157], and HPRD [319]. In addition, information extracted from scientific literature, which was manually curated, was also included and used for trimming the network. The resulting HPN considers interactions corresponding to different tissues to take into account the effect of the Stimulus in the whole body.

Defining active/inactive nodes

We define the state of human proteins as active or inactive for a particular phenotype, including its expression (as active) or repression (as inactive) extracted from the GSE57345 gene expression dataset [254] as in

Iborra-Egea et al [181] (see further details in supplementary material).

Description of the mathematical models

The algorithm of TPMS takes as input signals the activation (+1) and inactivation (-1) of the drug target proteins, and as output the BED protein states of the pathology. It then optimizes the paths between both protein sets and computes the activation and inactivation values of all proteins in the HPN. Each node of the protein network receives as input the output of the incoming connected nodes and every link is given a weight (ω_l). The sum of inputs is transformed by a hyperbolic tangent function that generates a score for every node, which becomes the "output signal" towards the outgoing connected nodes. The ω_l parameters are obtained by optimization, using a Stochastic Optimization Method based on Simulated Annealing [88]. The models are then trained by using the general restrictions (i.e. defined as edges and nodes with the property of being active or inactive) and the specific conditions set by the user. Details of the approach are shown in Fig 1 and supplementary material.

Measures to compare sets of MoAs

To understand the relationships between all potential mechanisms we defined some measures of comparison between different sets of solutions. We expect that a drug will revert the conditions of a disease phenotype; subsequently, a drug should inactivate the active protein effectors of a pathology-phenotype and activate the inactive ones. In this section we describe the measures used in the present study to analyze and compare sets of MoAs from different views (see further details in supplementary material).

TSignal

To quantify the intensity of the response of a MoA, we defined TSignal as the average signal arriving at the protein effectors (equation in supplementary material).

Distance between two sets of MoAs

We used the modified Hausdorff distance (MHD) introduced by Dubuisson and Jain [99] as the distance between two or more sets of MoAs in order to determine their similarity. Details of the equations are explained in the supplementary material.

Potential biomarkers extracted from MoAs

In order to extract potential biomarkers when comparing sets of MoAs, we first defined the best-classifier proteins. These are proteins inside the HPN that allow to better classify between groups of models and are identified following a Data-Science strategy (see supplementary material). Best-classifier proteins are usually strongly related to the intensity of a response and are proteins with values differently distributed between the groups of MoAs analyzed. For this study, and for the sake of simplicity, we focused only on the 200 proteins (or pair of proteins) showing the higher classification accuracy. Assuming the hypothesis that the selected MoAs are representative of individual prototype-patients, these proteins could be used as biomarkers to classify a cohort of patients.

Then, we applied the Mann-Whitney U test to compare the distributions of the best-classifier proteins values between the groups and selected those proteins with significant difference (p-value$<$ 0.01). We also restricted the list to proteins having an average value with opposite sign among groups (i.e. positive vs. negative or vice versa) and named them as differential best-classifier proteins. By following this strategy, we can identify two groups of differential best-classifier proteins: those active in the first group

(positive output signal in average) and inactive in the other (negative output signal in average), and the opposite.

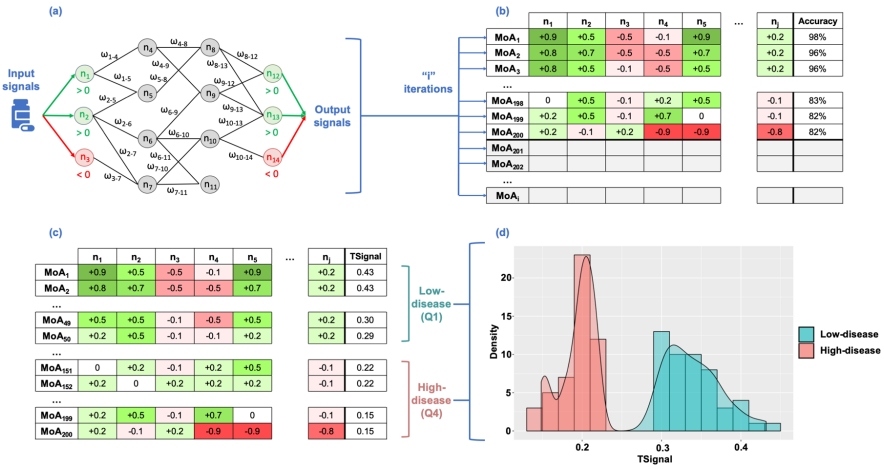

Figure 13: Scheme of how to apply TPMS to find the Mechanisms of Action (MoA) of a drug.
(a) Scheme of the method, transmitting information over the Human Protein Network (HPN) using a Multilayer Perceptron-like and sampling. **(b)** After a given number of iterations, we obtain a collection of Mechanisms of Actions (MoA). Rows represent the MoAs and columns the output signal values of the proteins (nodes of the network). The final column shows the accuracy of the model as a percentage of the number restrictions accomplished. **(c)** 200 MoAs are selected (coloured in the slide) and sorted by TSignal. The first quartile is defined as the Low-disease group, and the fourth quartile as High-disease group. The distribution of the output signals of the two groups of MoA are shown in **(d)** (High-disease in red and Low-disease is in blue).

4.2.4 Results and discussion

We applied TPMS to the HPN using as input signals the drug targets of sacubitril/valsartan (NEP / AT1R) and as output signals the proteins associated with HF extracted from the BED. Out of all MoAs found by TPMS, we selected the 200 satisfying the largest number of restrictions (and at least 80% of them) to perform further analysis.

Note that TPMS was only executed once, optimizing the results to satisfy the restrictions on HF data. The values of MD are obtained by measuring

the signal arriving at the MD effectors, which are part of the HPN and also receive signal. This procedure was chosen because we defined HF as the indication of the drug (sacubitril/valsartan), while MD is a potential adverse effect.

Stratification of MoAs

In order to compare models related to a good or bad response to the treatment, or those more prone to lead towards potential MD adverse effect, we stratified the MoAs. For HF, or treatment response, MoAs were ranked by their TSignal and then split in four quartiles. The first quartile (top 25%) contains MoAs with higher intensity of the response, which in turn corresponds to lower values of the effectors associated with HF phenotype (we named them as "Low"-disease MoAs). On the contrary, the fourth quartile (bottom 25%) collects MoAs with lower intensity of response (thus, we named as "High"-disease MoAs) (S1 File). On the other hand, for MD, the first quartile (top 25%) contains MoAs with higher intensity, which as an adverse event, correspond to models with high values of the effectors associated to MD (we named them as High-adverseEvent MoAs). The fourth quartile (bottom 25%) collects MoAs with lower intensity of response (thus, we named as Low- adverseEvent MoAs) (S1 File). Note that, in the following steps and because HF and MD groups were extracted from the same 200 set of models, common MoAs between different HF and MD-defined sets could be expected.

Comparison of MoAs with high/low TSignal associated to HF or MD

We calculated the modified Hausdorff distance between the groups of MoAs (High-MD, Low-MD, High-HF and Low-HF) to elucidate their similarity values (S1 File). In this sense, the higher the distance between the groups

is, the more different they are. We used these distances to calculate a dendrogram tree (see S1 File) showing that MoAs associated with a bad response to sacubitril/valsartan for HF (high-HF) are more similar (i.e. closer) to MoAs linked to a stronger MD adverse effect (high-MD). It is remarkable that the distances between Low- and High-HF and between Low- and High-MD are larger than the cross distances between HF and MD. However, by the definition of distance (equation 3 in supplementary material), it cannot account for the dispersion among the MoAs within and between each group. Therefore, for each set we calculated the mean Euclidean distance between all the points and its center, defined by the average of all points (see S1 File). As a result, all groups showed very similar dispersion values.

In order to have a global and graphical view of the distance between the individual MoAs, we generated a multidimensional scaling (MDS) plot calculated using MATLAB (see Fig 2). MDS plots display the pairwise distances in two dimensions while preserving the clustering characteristics (i.e. close MoAs are also close in the 2D-plot and far MoAs are also far in 2D). Focusing on the Low-HF group depicted in blue circles, we observe that there is no clear tendency to cluster with any of the MD groups. There are few cases of Low-HF MoAs coinciding in the space with Low- or High-MD MoAs. This implies that a good response to sacubitril/valsartan of HF patients would not be usually linked to the development of MD. Moreover, no clear distinction is found when plotting only the MD MoAs within the Low-HF group (see S1 File). However, regarding the set of High-HF MoAs, we can differentiate two clusters of MoAs: one related to the High-MD group (green crosses); and the other close to MoAs of the Low-MD group (black crosses) (see S1 File).

Assuming the hypothesis that different MoAs correspond to distinct prototype-patients, we conclude that for the specific set of patients for which sacubitril/valsartan works best reducing HF, it would be more difficult to differentiate between those presenting MD and those who do not. Instead, for the High-HF group, patients having MD could indeed be easily distinguished from those not presenting MD as side effect. However, because Low-HF group has more relevance to the clinics, specific functional analyses were performed in this specific group, as seen in following sections. Finally, we highlight that, as these distinct groups of prototype-patients are theoretical simulations, they don't reflect the clinical effects of real patients.

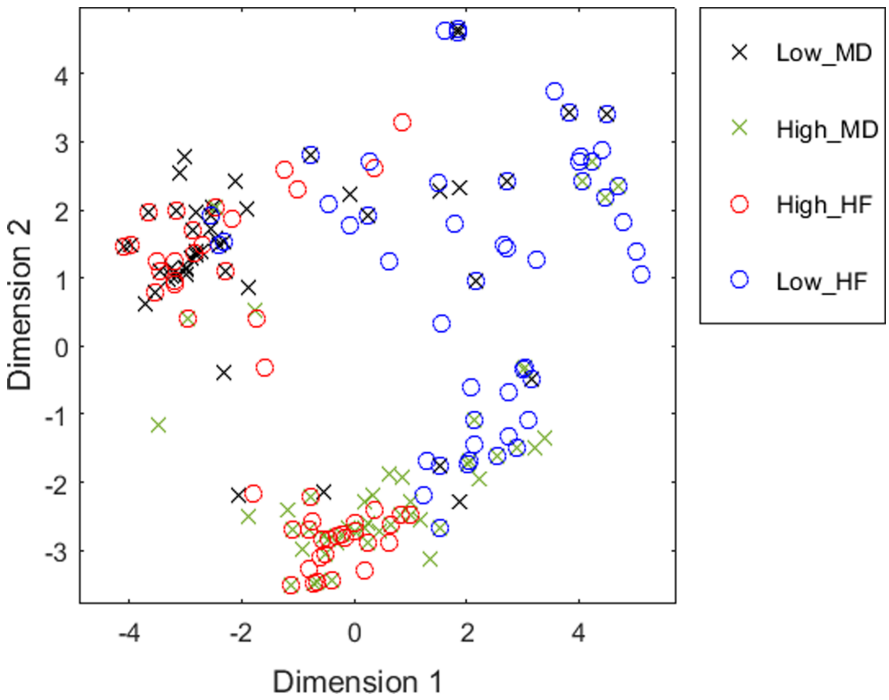

Figure 14: Multidimensional scaling plot of the distances between the Mechanisms of Action (MoA) of the four groups defined.
Each point represents a MoA. Axes are defined by the most representative dimensions.

Identification and functional analysis of potential biomarkers

For this section, we identified the nodes (i.e. proteins) significantly differentiating two groups of models (using a Mann-Whitney U test) for which the average of output signals have opposite signs (see methods in refsect33). After that, the function of the identified proteins was extracted from Gene Ontology (GO).

Identification of best-classifier proteins differentiating HF responses

After comparing High- vs Low- HF groups, we found a total of 45 differential best-classifier proteins associated with the treatment response (6 Low-HF-active/High-HF-inactive and 39 Low-HF-inactive/High-HF-active) (see Fig 3A and S1 File). To pinpoint the biological role of these proteins, we first identified the GO enriched functions (see S1 File) and then searched in the literature for evidences linking them with HF. As a result, we found that the differential best-classifier proteins Low-HF-active/High-HF-inactive point towards an important role for actin nucleation and polymerization mechanisms in drug response (reflected by the functions regulation of actin nucleation, regulation of Arp2/3 complex-mediated actin nucleation, SCAR complex, filopodium tip, or dendrite extension). In fact, the alteration of actin nucleation and polymerization mechanisms has been reported in heart failure [307, 207, 72]. Interestingly, a role for the activation of another differential best-classifier candidate, ATGR2, has been proposed to mediate some of the beneficial effects of angiotensin II receptor type 1 antagonists, such as valsartan [255, 355]. On the other hand, the results of the differential best-classifier proteins Low-HF-inactive/High-HF-active are linked to phosphatidylinositol kinase mediated pathways (phosphatidylinositol-3,4-bisphosphate 5-kinase

activity) and MAP kinase mediated pathways (MAP kinase kinase activity, best classifier proteins MAPK1, MAPK3, MAPK11, MAPK12 or MAPK13). In this case, both signaling pathways have been associated to cardiac hypertrophy and subsequent heart failure [20, 104]. These outcomes clearly lead towards the idea that High-HF models are a representation of prototype-patients with a worst response to the treatment, while Low-HF models are related to more beneficial response to the medication. A more detailed explanation can be found in the supplementary material.

Identification of best-classifier proteins differentiating MD responses

We identified 57 differential best-classifier proteins of MD (28 Low-MD-active/High-MD-inactive and 29 Low-MD-inactive/High-MD-active) (see Fig 3B and S1 File). Again, we searched for relationships between these proteins and MD by identifying the GO enriched functions (see S1 File) and searching for links in the literature. Some of the proteins and functions highlighted in the current analysis had been related to MD in previous works. The presence of dendritic spine development and dorsal/ventral axon guidance related proteins emphasizes the role of sacubitril/valsartan in dendritic and synaptic plasticity mechanisms, which had been previously linked to MD [379]. Furthermore, valsartan treatment has been reported to promote dendritic spine development in other related neurodegenerative diseases, such as Alzheimer's disease [366]. Other enriched functions are implicated in growth factor related pathways, which are known to be involved in wet MD pathogenesis [127]. Moreover, neovascularization in the wet variant of MD has been linked to the signaling of some of the growth factors detected as sacubitril/valsartan-associated MD classifiers in this study, including FGF1 [127] and PDGF [144, 151]. A more detailed explanation can be found in

the supplementary material.

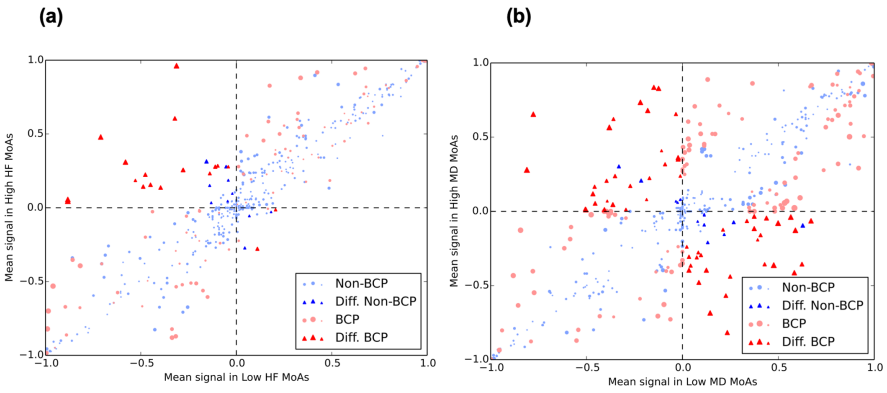

Figure 15: Scatter plot of the mean signal values of Low and High-"disease" Mechanisms of Action (MoA).
Scatter plot of the mean signal values of Low-"disease" and High-"disease" MoAs for each protein using as disease Heart Failure (HF) in **(a)** and Macular Degeneration (MD) in **(b)** . The average of the output signal of each protein in High-group is presented versus its value in Low-group. Differential signals (Diff., shown as triangles) are defined as those with opposite sign when comparing High versus Low average, and a p-value < 0.01 when calculating the Mann-Whitney U test between the two distributions of signals. Best-classifier proteins (BCP) are colored in red, otherwise they are blue. Sizes of markers are proportional to p-values of the Mann-Whitney U test.

Identification of potential biomarkers differentiating MD responses in Low-HF

Because of its clinical relevance, we decided to focus on analyzing the special case of prototype-patients in which the treatment reduces HF (Low-HF) but produces MD adverse effect (High-HF). In order to find these prototype-patients, we: (i) identified 13 Low-HF $\cap$ Low-MD MoAs and 12 Low-HF $\cap$ High-MD MoAs; and (ii) compared the protein signal of the two groups and proposed 30 potential biomarkers (Table 1). Among the proposed biomarkers, we found 16 proteins active in Low-HF $\cap$ Low-MD MoAs but inactive in Low-HF $\cap$ High-MD (15 of them shared with MD best-classifier proteins). On the other hand, 14 proteins were identified as inactive in Low-HF $\cap$ Low-MD and active

in Low-HF ∩ High-MD MoAs (12 of them were MD best-classifier proteins). We calculated the GO enriched functions of these two groups and observed that "phosphatidylinositol bisphosphate kinase activity" is enriched among proteins that are active in Low-HF ∩ Low-MD MoAs. Instead, "fibrinolysis" was found to be enriched among proteins active in Low-HF ∩ High-MD MoAs (Table 2). With this, we conclude that among the group of prototype-patients for which sacubitril/valsartan improves HF treatment response, the modulation of fibrinolysis could play a role at inducing the MD adverse effect. Moreover, we propose 12 best-classifier proteins that may be considered as biomarkers for good prognosis of the side effect.

In fact, since neovascular MD development is characterized by subretinal extravasations of novel vessels derived from the choroid (CNV) and the subsequent hemorrhage into the photoreceptor cell layer in the macula region [292], it might be reasonable to think that the modulation of fibrinolysis and blood coagulation pathways could play a role. The reported implication of some fibrinolysis related classifiers, such as FGB, SERPINE1 (PAI-1), and SERPING1, in neovascular MD development seems to support this hypothesis [440, 239, 167]. Besides, valsartan might be implicated in this mechanism, since it has been reported to modulate PAI-1 levels and promote fibrinolysis in different animal and human models [279, 301]. In addition, the presence of several other MD related classifiers in this list, such as IRS2 [8], PTGS2 [419], DCN [419] and FGF1 [362], further supports the interest of the classifiers as biomarkers of MD development in sacubitril/valsartan good responders. Still, we would like to highlight that the biomarkers have been proposed using a theoretical approach, and that the clinical effects studied may not be present in real patients.

Table 7: Potential biomarker proteins, with opposite signal in Low-HF ∩ Low-MD and Low-HF ∩ High-MD MoAs.

	Uniprot ID	Gene symbol	Gene name	⟨LMD⟩	⟨HMD⟩	$\sqrt{\lvert LMDxHMD \rvert}$	Adjusted P-value	BCP
1	**P02675**	**FGB**	**Fibrinogen beta chain**	-0.576	0.814	0.685	1.297E-03	MD
2	O43639	NCK2	Cytoplasmic protein NCK2	0.620	-0.697	0.657	1.656E-04	MD
3	P54762	EPHB1	Ephrin type-B receptor 1	0.317	-0.677	0.464	3.669E-04	HF&MD
4	Q9Y4H2	IRS2	Insulin receptor substrate 2	0.417	-0.465	0.440	8.181E-04	MD
5	O60674	JAK2	Tyrosine-protein kinase JAK2	-0.747	0.249	0.431	1.656E-04	MD
6	P06241	FYN	Tyrosine-protein kinase Fyn	0.591	-0.236	0.373	2.466E-04	HF&MD
7	P30530	AXL	Tyrosine-protein kinase receptor UFO	0.392	-0.330	0.360	2.111E-04	MD
8	**Q02297**	**NRG1**	**Pro-neuregulin-1, membrane-bound isoform**	0.672	-0.188	0.355	2.111E-04	MD
9	P32004	L1CAM	Neural cell adhesion molecule L1	-0.373	0.309	0.339	1.297E-03	HF&MD
10	Q05586	GRIN1	Glutamate receptor ionotropic, NMDA 1	-0.174	0.620	0.329	1.955E-04	MD
11	**P05230**	**FGF1**	**Fibroblast growth factor 1**	-0.152	0.688	0.323	8.181E-04	HF&MD
12	**P18084**	**ITGB5**	**Integrin beta-5**	0.436	-0.236	0.321	2.111E-04	MD
13	**P01583**	**IL1A**	**Interleukin-1 alpha**	0.174	-0.472	0.287	1.955E-04	MD
14	P10275	AR	Androgen receptor	0.349	-0.201	0.265	8.008E-04	MD
15	P15941	MUC1	Mucin-1 subunit alpha	0.099	-0.652	0.254	6.905E-04	HF&MD
16	O14757	CHEK1	Serine/threonine-protein kinase Chk1	0.436	-0.142	0.248	1.549E-03	MD
17	P15391	CD19	B-lymphocyte antigen CD19	-0.131	0.357	0.216	8.160E-03	MD
18	**P61981**	**YWHAG**	**14-3-3 protein gamma, N-terminally processed**	0.174	-0.236	0.203	2.783E-03	-
19	Q9Y478	PRKAB1	5'-AMP-activated protein kinase subunit beta-1	0.261	-0.142	0.192	5.682E-03	MD
20	P62158	CALM1; CALM2; CALM3	Calmodulin-1 {ECO:0000312|HGNC:HGNC:1442}	-0.282	0.107	0.174	9.405E-03	MD
21	**P06748**	**NPM1**	**Nucleophosmin**	0.261	-0.107	0.167	3.618E-03	MD
22	O15357	INPPL1	Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2	-0.261	0.094	0.157	3.618E-03	MD
23	P17081	RHOQ	Rho-related GTP-binding protein RhoQ	-0.218	0.094	0.143	9.794E-03	MD
24	P35354	PTGS2	Prostaglandin G/H synthase 2	0.044	-0.472	0.143	3.669E-04	MD
25	P42684	ABL2	Abelson tyrosine-protein kinase 2	-0.218	0.094	0.143	9.794E-03	MD
26	**Q15109**	**AGER**	**Advanced glycosylation end product-specific receptor**	-0.267	0.063	0.130	8.160E-03	-
27	P07585	DCN	Decorin	-0.044	0.236	0.101	5.682E-03	MD
28	**P05155**	**SERPING1**	**Plasma protease C1 inhibitor**	-0.044	0.236	0.101	5.682E-03	MD
29	**P05121**	**SERPINE1**	**Plasminogen activator inhibitor 1**	-0.044	0.236	0.101	5.682E-03	-
30	P14770	GP9	Platelet glycoprotein IX	0.044	-0.236	0.101	5.682E-03	MD

Highlighted cells correspond to proteins that are part of the Top-HF ∪ Top-MD ∪ Top-Drug set, the top-scoring proteins according to GUILDify. Columns show: the protein name (as UniprotID, gene-symbol and gene-name), the average of the signal in in Low-MD (<LMD>) and High-MD (<HMD>) in the selected sets of MoAs and a measure of the strength of the signal in both distributions (calculated as $\sqrt{\lvert LMDxHMD \rvert}$), the significance (adjusted P-value) ensuring that both distributions of signals are different, and whether the protein has been considered best-classifier in MD of HF (BCP).

https://doi.org/10.1371/journal.pone.0228926.t001

Analysis of proposed biomarkers with GUILDify

In the previous section, we proposed 30 proteins that could potentially help to identify HF patients at risk of developing MD. To corroborate these biomarkers, we tested how many of them are found using a different approach also based on the use of functional networks. For this purpose, we used GUILDify v2.0 [6], a web server that extends the information of disease-gene associations through the protein-protein interactions network. GUILDify scores proteins according to their proximity with the genes associated with a disease (seeds). Using this web server, we identify a list

Table 8: Top 10 Gene Ontology functions enriched from proteins with opposite signal in Low-HF ∩ Low-MD and Low-HF ∩ High-MD MoAs.

	Low-HF ∩ LMD+ HMD-			Low-HF ∩ HMD+ LMD-			Overlapped functions		
	GO name	LOD	P-val.	GO name	LOD	P-val.	GO name	LOD	P-val.
1	phosphatidylinositol-4,5-bisphosphate 3-kinase activity	1.89	0.03600	fibrinolysis	2.51	0.00050	response to stimulus	1.19	<0.00050
2	cellular response to UV	1.87	0.04200	negative regulation of wound healing	2.13	0.00050	positive regulation of transport	1.24	<0.00050
3	phosphatidylinositol bisphosphate kinase activity	1.87	0.04200	negative regulation of blood coagulation	2.12	0.00850	positive regulation of biological process	1.13	0.00051
4	vascular endothelial growth factor receptor signaling pathway	1.86	0.04200	negative regulation of hemostasis	2.12	0.00850	positive regulation of developmental process	1.18	<0.00050
5	positive regulation of protein kinase B signaling	1.70	0.01050	negative regulation of coagulation	2.10	0.01050	positive regulation of cellular process	1.04	0.00294
6	negative regulation of apoptotic signaling pathway	1.68	0.00050	platelet alpha granule lumen	1.96	0.02300	positive regulation of response to stimulus	1.04	0.00417
7	peptidyl-tyrosine phosphorylation	1.63	0.01400	regulation of epithelial cell apoptotic process	1.96	0.02300	-	-	-
8	regulation of apoptotic signaling pathway	1.63	<0.00050	regulation of blood coagulation	1.91	0.02800	-	-	-
9	peptidyl-tyrosine modification	1.62	0.01400	regulation of hemostasis	1.91	0.02800	-	-	-
10	protein tyrosine kinase activity	1.61	0.01850	regulation of coagulation	1.89	0.03450	-	-	-

Functional enrichment analysis from FuncAssociate [50].

https://doi.org/10.1371/journal.pone.0228926.t002

of top-scoring proteins that are critical on transmitting the perturbation of disease genes through the network. The network used by GUILDify is completely independent from the HPN used in the TPMS, becoming an ideal, independent context to test the potential biomarkers.

Thus, we used GUILDify to indicate which of the potential biomarkers identified by TPMS may have a relevant role in the molecular mechanism of the drug. We ran GUILDify using the two targets of sacubitril/valsartan (NEP, AT1R) as seeds, and selected the top 2% scored nodes (defined as the "top-drug" set). We did the same with the phenotypes of HF and MD, using as seeds the 124 effectors of HF and 163 effectors of MD from the BED database. We merged the top scored sets of HF, MD and top-drug ("top-drug ∪ top-HF ∪ top-MD") and studied the overlap with the set of 30 biomarkers proposed in the previous section. 10 of the candidate biomarkers are found in the merged set "top-drug ∪ top-HF ∪ top-MD" and are consequently significant (see S1 File).

Some of these candidates can be functionally linked to both diseases and the

drug under study. For example, among these 10 classifiers, AGER has been implicated in both HF [162], through extracellular matrix remodeling, and MD development [25], through inflammation, oxidative stress, and basal laminar deposit formation between retinal pigment epithelium cells and the basal membrane; furthermore, this receptor is known to be modulated by AT1R [313], valsartan target. Similarly, FGF1 has been proposed to improve cardiac function after HF [137], as well as to promote choroid neovascularization leading to MD [127]. Moreover, FGF1 is regulated by angiotensin II through ATGR2 [234], another protein suggested as classifier in the current analysis that is known to mediate some of the effects of AT1R antagonists, such as valsartan [255, 355]. Another candidate, NRG1, has been linked to myocardial regeneration after HF [135] and is known to lessen the development of neurodegenerative diseases such as Alzheimer's disease [437], which shares similar pathological features with MD [202]. NRG1 is also linked to the expression of neprilysin [437], sacubitril target. ITGB5 has been identified as risk locus for HF [400] and its modulation has been linked to lipofucsin accumulation in MD [203]. Interestingly, ATGR1 inhibitors have been reported to modulate ITGB5 expression in animal models [211]. Finally, IL1A has been proposed as an essential mediator of HF pathogenesis [57, 393] through inflammation modulations, and serum levels of this protein have been found increased in MD patients [286]. In addition, as described in previous sections, classifiers FGB, SERPINE1, and SERPING1 have been linked to MD [440, 239, 167] and are also known to play a role in HF development [442, 442, 275, 67]. According to these findings, the 10 potential biomarkers proposed by TPMS and identified with GUILDify might be prioritized when studying good responder HF patients at risk of MD development.

4.2.5 Limitations

Although TPMS returns the amount of signal from the drug arriving to the rest of the proteins in the HPN, this signal is only a qualitative measure. We are not using data about the dosage of the drug or the quantity of expression of the proteins. However, we are already working to make TPMS move towards the growing tendency of Quantitative Systems Pharmacology. The quantification of the availability of drugs in the target tissue for each patient opens the opportunity to have an accurate patient simulation to do in silico clinical trials.

4.2.6 Conclusions

It exists an increasing need for new tools to get closer to real life clinical problems and the Systems Biology-based computational methods could be the solution needed. The specific case of sacubitril/valsartan stands out because of the amount of resources invested in the safety of the drug and the concern on the possible risk of inducing amyloid accumulation-associated conditions, such as macular degeneration (MD), in the long term. In this study, we applied TPMS technology to uncover different Mechanisms of Action (MoAs) of sacubitril/valsartan over heart failure (HF) and reveal its molecular relationship with MD. For this approach, we hypothesize that each MoA would correspond to a prototype-patient. The method is then used to generate a wide battery of MoAs by performing an in silico trial of the drug and pathology under study. TPMS computes the models by using a hand curated Human Protein Network and applying a Multilayer Perceptron-like and sampling method strategy to find all plausible solutions. After analyzing the models generated, we found different sets of proteins able to classify the models according to HF

treatment efficacy or MD treatment relationship. The sets include functions such as PI3K and MAPK kinase signaling pathways, involved in HF-related cardiac hypertrophy, or fibrinolysis and coagulation processes (e.g. FGB, SERPINE1 or SERPING1) and growth factors (e.g. FGF1 or PDGF) related to MD induction. Furthermore, we propose 30 biomarker candidates to identify patients potentially developing MD under a successful treatment with sacubitril/valsartan. Out of this 30, 10 biomarkers were also found in the alternative, independent molecular context proposed by GUILDify, including some HF and MD effectors such as AGER, NRG1, ITGB5 or IL1A. Further studies might prospectively validate the herein raised hypothesis.

We notice that the models generated with TPMS are completely theoretical and thus, they are not associated with clinical effects of real patients. Consequently, the biomarkers proposed on the basis of these models are also theoretical and would require an experimental validation. Still, TPMS represents a huge improvement for studying the hypothetical relationship between a drug and an adverse effect. Until now, there were not enough tools that allow to perform an exhaustive study on the MoAs of an adverse effect. Now, with the MoAs and biomarkers proposed by TPMS, we provide the tools for this type of research.

4.2.7 Acknowledgements

Public funders provided support for authors salaries: JAP, NFF and BO received support from the Spanish Ministry of Economy (MINECO) [BIO2017-85329-R] [RYC-2015-17519]; “Unidad de Excelencia María de Maeztu”, funded by the Spanish Ministry of Economy [ref: MDM-2014-0370]. The Research Programme on Biomedical Informatics

(GRIB) is a member of the Spanish National Bioinformatics Institute (INB), PRB2-ISCIII and is supported by grant PT13/0001/0023, of the PE I+D+i 2013-2016, funded by ISCIII and FEDER. GJ has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 765912. VJ is part of a project (COSMIC; www.cosmic-h2020.eu) that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 765158.

4.2.8 Supporting information

S1 File. Extended version of materials and methods; S1-S5 Figs; S1-S13 Tables. https://doi.org/10.1371/journal.pone.0228926.s001 (DOCX)

4.2.9 Extended version of materials and methods

Biological Effectors Database (BED) to molecularly describe specific clinical conditions

Patient-like characteristics are modelled using clinical data and/or experimental molecular data. There are many databases providing clinical data of patients, adverse drug reactions, diseases or indications (e.g. ClinicalTrials.gov, SIDER, ChEMBL, PubChem, DrugBank...). Many other databases provide molecular data defining the existing human genes and/or proteins and describing the relationships between them (IntAct, BioGRID, REACTOME...). Combining both, clinical and molecular information available, the BED describes more than 300 clinical phenotypes as sets of genes and proteins (effectors) that can be "active", "inactive" or "neutral" [363, 321]. For example, in a metabolic protein-like network, an enzyme will become "active" in the presence of a catalyst, or become

inactivated when interacting with an inhibitor. Alternatively, in a genetic network, genes are active when they are expressed (experimentally detected as over-expression) and inactive when they are repressed (experimentally detected as under-expression). Additionally, in protein-protein interaction (PPI) networks, some proteins carry out their interactions only when they are phosphorylated, thus becoming active, and vice versa by dephosphorylation. By default, neutral proteins remain unaffected, neither active nor inactive, for a particular phenotype.

The methodology used for assigning the protein effectors to each pathology starts by defining the pathophysiological processes (functions) according to the general definitions used by the scientists studying the disease. Then, a review of the most recent, relevant and accepted information in the field is performed through PubMed queries, starting from general pathophysiology reviews. An expansion of the effector candidate's identification is done through reading the relevant original papers from the references or adding searches of important concepts that are not covered enough (molecularly wise) within the reviews read. The final goal of the characterization is to select proteins with an accepted functional role within the disease, and specifically within the functions that define the disease to center the analysis.

HF effectors

Regarding the molecular basis of HF BED proteins, they were characterized as described above and in Iborra-Egea et al. (2017) [181]. The definition used of heart failure in the current study has been performed according to the indication of Entresto and to the EMA Assessment report [105]. Thus, it is centered in processes associated to long term changes related to cardiac remodeling (as discussed in the paper were the models were initially presented [181]), that can be cause and consequence of heart failure, not

necessarily caused by ischemic causes. The identified functions are detailed in Supplementary Table 12.

MD effectors

MD pathophysiology is tightly related to protein accumulation [242, 180, 73]. However, the characterization used for the current study not only included this function, but also other processes associated to MD pathophysiology, including neovascularization, characteristics of wet Age-Related MD and changes associated to geographic atrophy (late stage dry Age-Related MD) [292]. The functions are detailed in Supplementary Table 13.

TPMS modelling

The Therapeutic Performance Mapping System (TPMS) is a tool that creates mathematical models of a drug/pathology protein pathways to explain a clinical outcome or phenotype [321, 181, 18, 164, 148, 309, 339]. These models find MoAs that explain how a Stimulus (i.e. proteins activated or inhibited by a drug) produces a Response (i.e. proteins active or inhibited in a phenotype). As an example of usage, here we applied TPMS to the drug-indication pair sacubitril/valsartan and HF. Regarding the drug, we retrieved the sacubitril/valsartan targets from DrugBank [425], PubChem [217], STITCH [381], SuperTarget [161] and hand curated literature revision. As for the indication, we retrieved the proteins whose modulations had been associated with HF from the BED [363, 321]. Finally, after applying the TPMS methodology, we obtained a set of connected proteins (subnetworks) with associated activities, each subnetwork with a potential explanation of the molecular mechanism of the drug in agreement with what had been previously described (i.e. a potential MoA).

Building the Human protein network (HPN)

To apply the TPMS approach and create the mathematical models of MoAs, an HPN is needed beforehand. In this study, we used a PPI network created from the integration of public and private databases: KEGG [205], BioGRID [68], IntAct [300], REACTOME [108], TRRUST [157], and HPRD [319]. In addition, information extracted from scientific literature, which was manually curated, was also included and used for trimming the network. The resulting HPN considers interactions corresponding to different tissues to take into account the effect of the Stimulus in the whole body.

Defining model restrictions

A collection of restrictions, defined as the true set of edges and nodes with the property of being active or inactive, are used for validating the models obtained with TPMS. We define two types of restrictions depending on its specificity. The general or global restrictions are those used in all approaches and describe a wide expanse of knowledge about protein interactions and relations. This information is obtained from HPRD [319], DIP [349], TRRUST [157], INTACT [300], REACTOME [108], BIOGRID [68], SIDER [232] and DrugBank [425]. These set of restrictions help indicate what proteins are active or inactive, and their interactions, in a general human being. Additionally, specific restrictions regarding the phenotype under study can also be used, usually derived from high throughput data or additional protein knowledge.

For this study, we added specific information to our models concerning the changes of gene expression induced by sacubitril/valsartan on HF patients. Specifically, we used the GSE57345 gene expression dataset [254] , extracted from GEO database, as in Iborra-Egea et al. (2017) [181]. We

calculated the expression fold change of genes associated with the HPN and mapped them as activated or inhibited proteins (active if they corresponded to over-expressed genes and inactive -inhibited- for under-expressed).

Description of the mathematical models

The algorithm of TPMS for generating the models is similar to a Multilayer Perceptron of an Artificial Neural Network over the HPN (where neurons are the proteins and the edges of the network are used to transfer the information). It takes as input signals the activation (+1) and inactivation (-1) of the drug target proteins and as output the BED protein states of the pathology phenotype. The network is limited to only interactions that connect the drug targets with any other protein in the HPN in a maximum of three steps to avoid signal noisiness. Once set, the algorithm optimizes the paths between both input and output protein sets and computes the activation and inactivation values of the all proteins in the HPN. The parameters to solve are the weights associated to the links between every node pair (ω_i). Each node of the protein network receives as input the output of the incoming connected nodes, which are weighted by each link weight. The sum of inputs is transformed by a hyperbolic tangent function to generate the score of the node (neuron), which become the "output signal" of the current node towards outgoing nodes. Details of the approach are shown in Fig 1a, where n_5 is linked to n_1 and n_2. The output signal of n_5 is $n_5 = tanh(n_1 * \omega_{1-5} + n_2 * \omega_{2-5})$. The ω_i parameters are obtained by optimization, using a Stochastic Optimization Method based on Simulated Annealing [87], such that the values of the effector nodes are the closest to their expected values, and always adjusting to the maximum of the restrictions mentioned above. The iterative process of optimization usually requires between 106 and 109 iterations, until satisfying at least the 80%

of the restrictions and the values of the effectors. However, the number of ω_i parameters can be very high (between 100,000 and 400,000 depending on the size of the subnetwork) and the size of the collection of restrictions (approximately 107) is usually not enough to find a unique solution. For that, many final models can be obtained and manual curation can be applied to select and modify the network and reduce the space of exploration.

Measures to compare sets of MoAs

TPMS returns a set of MoAs describing potential relationships between the targets of a drug and the biological protein effectors of a disease. We hypothesize that TPMS solutions represent MoAs in different prototype-patients. Therefore, we needed to define some comparison measures in order to understand the relationships between all potential mechanisms and compare sets of MoAs from different views.

Intensity of the response

We defined the "intensity" of the response as a pair: 1) the number of protein effectors (#) achieving an expected signal sign; and 2) a measure of the strength of the output signal of the effectors (i.e. a global measure of the output signal, named TSignal). For the present study, however, only the TSignal was used.

Assuming y_i as the value achieved by a protein effector "i", while v_i is the effector sign according to the BED (active or inactive) and n is the total number of effectors described for a phenotype, we define:

- Number of effectors achieving the expected sign: We expect that a drug will revert the conditions of a disease phenotype, while it may reach the effectors of an adverse event. Consequently, a drug should

inactivate the active protein effectors of a pathology-phenotype and activate the inactive ones, but it could activate/inhibit other adverse event effectors with the same sign as described in the BED. Using Dirac's δ (i.e. $\delta(0) = 1$, and zero otherwise), for drug indications the formula is defined as following:

$\#_{indication} = sum_n^{i=1} \delta \left(v_i + \frac{y_i}{|y_i|} \right)$

Therefore, in the case of the disease effectors we only count the effectors with a BED value of opposite sign to the signal arriving from the drug.

However, for adverse events, the formula changes because we count the effectors that are affected by the drug, such that the signal arriving from the drug has the same sign as in the BED:

$\#_{adverse_event} = sum_n^{i=1} \delta \left(v_i - \frac{y_i}{|y_i|} \right)$

- TSignal: The average of the output values of the protein effectors such that the proteins with correct sign are considered as positive signal, and the ones with the incorrect sign considered as negative signal. For a drug affecting the phenotype of a disease, this implies that v_i and y_i have opposite sign and we need to change the sign:

  $TSignal_{indication} = -\frac{1}{n} \sum_{i=1}^{n} v_i y_i$

  On the contrary, to test if a drug induces an adverse event, we check if the output signal has the same sign as the effectors of the desired phenotype, and therefore TSignal is defined as:

  $TSignal_{adverse_event} = \frac{1}{n} \sum_{i=1}^{n} v_i y_i$

Distance between two sets of MoAs

We used the modified Hausdorff distance (MHD) introduced by Dubuisson and Jain [99] as the distance between two or more sets of MoAs in order

to determine their similarity. We used the distance measures between two (finite) point sets A and B as following: For $a \in A$, $d(a,b) = min_{(b \in B)} d(a,b)$

and

$d_A(B) = \frac{1}{|A|} \sum_{a \in A} d(a,b)$

Where $|A|$ is the number of elements in A, $d(\cdot,\cdot)$ is the Euclidean distance and "a" and "b" are n-tuples of the activities (output signals) of the nodes of two MoAs (a in A and b in B). Then, we defined the MHD as:
$d_{MHD}(A,B) = max(d_A(B), d_B(A))$

Note that the MHD is a semimetric and not a metric, since the triangular inequality does not hold.

Potential biomarkers extracted from MoAs

Identification of Best-Classifier Proteins

In order to extract potential biomarkers from comparing sets of MoAs, we first defined the best-classifier proteins, specific proteins helping us to infer biological associations and distinguish the responses of drugs on a population (i.e. potential biomarkers). Best-classifier proteins (single or pairs) are the proteins inside the HPN that allow to better classify samples between groups of MoAs. These classifiers are determined by a Data-Science strategy, which is based on a set of Feature Selection algorithms combined with several Base Classifiers. The feature selection used for single proteins was brute force [153], so analyzing one feature or protein at a time, while for protein pairs the following selection methods were used: elastic net [450]; entropy and correlation [308]; LASSO [387]; random forest [169]; GLM random sets [260]; ReliefF [219];

Ridge regression [438]; simple regression [129]; Wilcoxon test [74]; and Wilcoxon test with correlation [74]. Several base classifiers were applied to distinguish the two groups using the selected features: optimal threshold; linear regression [129]; Multilayer Perceptron Network [230]; Generalized Linear Model [260]; elastic net [298]; and optimal quadratic threshold [358]. Finally, after a k-fold cross-validation (k=10) [226] was applied, the proteins were sorted by the balanced accuracy [46] of the classification. For this study, only the 200 proteins (or pair of proteins) with highest balanced accuracy were selected as best-classifier proteins. Assuming the hypothesis that the selected MoAs are representative of individual prototype-patients, these proteins could then be used as biomarkers to classify a cohort of patients by the activity or absence of activity of the proteins.

Identification of differential Best-Classifier Proteins

Each best-classifier protein has a specific distribution of signal values corresponding to each group of MoAs. We applied the Mann-Whitney U test to compare the two distributions and selected those proteins having a significantly different distribution (p-value$<$ 0.01). We also restricted the list to proteins having an average value with opposite sign among groups (i.e. positive vs. negative or vice versa), and named them as differential best-classifier proteins. By following this strategy, we can identify two groups of differential best-classifier proteins: those active in the first group (positive output signal in average) and inactive in the other (negative output signal in average), and the opposite.

Types of proteins not considered

- Non-differential Best-Classifier Proteins: Those are proteins in

which, even if the mean signal in both groups is very similar, the machine learning algorithms are still able to differentiate High- and Low- MoAs based on their distribution values. For example, in the upper right corner of Figure 2a we find the protein P29353, the 181st best protein to classify High- and Low- HF models (cross-validation AUC = 0.67, P-value = 1.12·10-4). P29353 has a Low-HF mean signal of 0.99999999948 and a High-HF mean signal of 0.9999999985. As showed in Supplementary Fig 4a, the High- and Low- HF signals values are both very close to each other. However, if we explore the distribution of signals considering all the decimals given by TPMS (Supplementary Fig 4b), we can observe a slight difference between the two distributions. This fact allowed the machine learning algorithms to include the protein as a best-classifier protein, but was then rejected as a differential best-classifier protein after applying the Mann-Whitney U test.

- Differential non-Best-Classifier Proteins: Those are proteins that, when comparing the signals between groups, they have significantly opposite sign. However, they are not considered Best-Classifier Proteins because they are not among the top 200 proteins selected by the machine learning algorithms. For example, the protein P40763 is the 241st best feature on distinguishing High- and Low- Heart Failure Mechanisms of Action (cross-validation AUC = 0.66, P-value = 1.22·10-3). The distribution of High- and Low signals are represented in Supplementary Fig 5. In the figure we can appreciate how the distributions of High- and Low- signals are overlapped, complicating their differentiation. Still, the p-value of the cross-validation is below 0.05, reflecting the potential of this feature to differentiate the distinct types of Mechanisms of Action.

4.2.10 Extended version of results and discussion

We applied TPMS to the HPN using as input signals the drug targets of sacubitril/valsartan (NEP / AT1R) and as output signals the proteins associated with HF extracted from the BED. Out of all MoAs found by TPMS, we selected the 200 satisfying the largest number of restrictions (and at least 80% of them) to perform further analysis.

Note that TPMS was only executed once, optimizing the results to satisfy the restrictions on HF data. The values of MD are obtained by measuring the signal arriving at the MD effectors, which are part of the HPN and also receive signal. This procedure was chosen because we defined HF as the indication of the drug (sacubitril/valsartan), while MD is a potential adverse effect.

Stratification of MoAs

In order to compare models related to a good or bad response to the treatment, or those more prone to lead towards potential MD adverse effect, we stratified the MoAs. For HF, or treatment response, MoAs were ranked by their TSignal and then split in four quartiles. The first quartile (top 25%) contains MoAs with higher intensity of the response, which in turn corresponds to lower values of the effectors associated with HF phenotype (we named them as "Low"-disease MoAs). On the contrary, the fourth quartile (bottom 25%) collects MoAs with lower intensity of response (thus, we named as "High"-disease MoAs) (Supplementary Fig 1a). On the other hand, for MD, the first quartile (top 25%) contains MoAs with higher intensity, which as an adverse event, correspond to models with high values of the effectors associated to MD (we named them as High-adverseEvent MoAs). The fourth quartile (bottom 25%)

collects MoAs with lower intensity of response (thus, we named as Low-adverseEvent MoAs) (Supplementary Fig 1b). Note that, in the following steps and because HF and MD groups were extracted from the same 200 set of models, common MoAs between different HF and MD-defined sets could be expected.

Comparison of MoAs with high/low TSignal associated to HF or MD

We calculated the modified Hausdorff distance between the groups of MoAs (High-MD, Low-MD, High-HF and Low-HF) to elucidate their similarity values (Supplementary Table 5). In this sense, the higher distance between the groups, the more different they are. We used these distances to calculate a dendrogram tree (see Supplementary Fig 2) showing that MoAs associated with a bad response to sacubitril/valsartan for HF (high-HF) are more similar (i.e. closer) to MoAs linked to a stronger MD adverse effect (high-MD). It is remarkable that the distances between Low-HF and High-HF and between Low-MD and High-MD are larger than the cross distances between HF and MD. However, by the definition of distance (equation 3 in supplementary material), we cannot account for the dispersion among the MoAs within and between each group. Therefore, for each set we calculated the mean Euclidean distance between all the points and its center, defined by the average of all points (see Supplementary Table 6). As a result, all groups showed very similar dispersion values.

In order to have a global and graphical view of the distance between the individual MoAs, we generated a multidimensional scaling (MDS) plot calculated using MATLAB (see Fig 2). MDS plots display the pairwise distances in two dimensions while preserving the clustering characteristics (i.e. close MoAs are also close in the 2D-plot and far MoAs are also far in

2D). Focusing on the Low-HF group depicted in blue circles, we observe that there is no clear tendency to cluster with any of the MD groups. There are few cases of Low-HF MoAs coinciding in the space with Low- or High-MD MoAs. This implies that a good response to sacubitril/valsartan of HF patients would not be usually linked to the development of MD. Moreover, no clear distinction is found when plotting only the MD MoAs within the Low-HF group (see Supplementary Fig 3a). However, regarding the set of High-HF MoAs, we can differentiate two clusters of MoAs: one related to the High-MD group (green crosses); and the other close to MoAs of the Low-MD group (black crosses) (see Supplementary Fig 3b).

Assuming the hypothesis that different MoAs correspond to distinct prototype-patients, we conclude that for the specific set of patients for which sacubitril/valsartan works best reducing HF, it would be more difficult to differentiate between those presenting MD and those who do not. Instead, for the High-HF group, patients having MD could indeed be easily distinguished from those not presenting MD as side effect. However, because Low-HF group has more relevance to the clinics, specific functional analyses were performed in this specific group, as seen in following sections.

Identification and functional analysis of potential biomarkers

For this section, we identified the nodes (i.e. proteins) significantly differentiating two groups of models (using a Mann-Whitney U test) for which the average of output signals have opposite signs (see methods in 3.3). After that, the function of the identified proteins was extracted from Gene Ontology (GO).

Identification of best-classifier proteins differentiating HF responses

After the model stratification regarding the HF groups, we selected the 200 best-classifier proteins to differentiate the two groups of MoAs. Among these proteins, we identified the differential best-classifier proteins as explained in the methodology, and ended up with two groups: those active in Low-HF (the average of output signals in Low-HF MoAs is positive) and inactive in High-HF (the average of output signals in High-HF MoAs is negative); and those active in High-HF but inactive in Low-HF. Out of the starting 200 best-classifier proteins, we found a total of 45 differential best-classifier proteins associated with the treatment response (6 in the first group and 39 in the second) (see Supplementary Table 1). Fig 3a displays all the proteins average signal values for the MoAs of Low-HF vs High-HF. Most of the proteins with opposite signs between the two cohorts were also selected as differential best-classifier proteins.

To pinpoint the biological role of these proteins, we first identified the GO enriched functions (see Supplementary Table 2) and then searched in the literature for evidences linking them with HF. The enrichment used for this proceeding was calculated using the software FuncAssociate [38]. Among the enriched functions, we found processes associated with the SCAR complex, the positive regulation of actin nucleation, the regulation of neurotrophin TRK receptor and dendrite extension. We used the same procedure to extract the GO functions associated to the differential best-classifier proteins that are inactive in Low-HF but active in High-HF. We detected functions such as phosphatidylinositol kinase activity, MAP kinase activity, DNA damage induced protein phosphorylation and superoxide anion generation. Although some enriched functions are shared by both sets, such as Fc gamma receptor signaling, the majority of functions identified are different (see Supplementary Table 2).

Some of the proteins and functions highlighted in the current analysis have been related to myocardial function. On the one hand, our findings show that differential best-classifier proteins Low-HF-active/High-HF-inactive point towards an important role for actin nucleation and polymerization mechanisms in drug response (reflected by the functions regulation of actin nucleation, regulation of Arp2/3 complex-mediated actin nucleation, SCAR complex, filopodium tip, or dendrite extension). In fact, the alteration of actin nucleation and polymerization mechanisms has been reported in heart failure [307, 207, 72]. Interestingly, a role for the activation of another differential best-classifier candidate, ATGR2, has been proposed to mediate some of the beneficial effects of angiotensin II receptor type 1 antagonists, such as valsartan [255, 355].

On the other hand, the results of the differential best-classifier proteins Low-HF-inactive/High-HF-active are linked to phosphatidylinositol kinase mediated pathways (phosphatidylinositol-3,4-bisphosphate 5-kinase activity) and MAP kinase mediated pathways (MAP kinase kinase activity, best classifier proteins MAPK1, MAPK3, MAPK11, MAPK12 or MAPK13). In this case, both signaling pathways have been associated to cardiac hypertrophy and subsequent heart failure [20, 104]. These outcomes clearly leads towards the idea that High-HF models are a representation of prototype-patients with a worst response to the treatment, while Low-HF models are related to more beneficial response to the medication.

Identification of best-classifier proteins differentiating MD responses

We similarly classified MoAs in High-MD and Low-MD identified the differential best-classifier proteins active in Low-MD but inactive in High-MD, and vice versa. As before, we compared the distributions of Low-MD and High-MD output signals of the

best-classifier proteins and calculate the average of the signal in all MoAs in Low- and High- MD. Out of 200 best-classifier proteins, we identified 28 Low-MD-active/High-MD-inactive and 29 Low-MD-inactive/High-MD-active (see Supplementary Table 3). Fig 3b shows the plot for all proteins classified by their average output signal in Low-MD and High-MD models.

Again, we calculated the GO enriched functions for these groups of proteins (see Supplementary Table 4). For the first group (Low-MD-active/High-MD-inactive) we obtained unique functions such as dendritic spine development, positive regulation of vascular endothelial growth factor production and phosphotyrosine binding. For the second group (Low-MD-inactive/High-MD-active), we found functions such as dorsal/ventral axon guidance, fibroblast growth factor receptor binding and response to toxic substance. However, phosphatidylinositol bisphosphate kinase activity showed up as enriched function in both groups.

Some of the proteins and functions underlined in the current analysis had previously been related to MD. The presence of dendritic spine development and dorsal/ventral axon guidance related proteins among the differential best-classifiers points towards a role for sacubitril/valsartan-associated MD in dendritic and synaptic plasticity mechanisms, which had been previously linked to the condition [379]. Furthermore, valsartan treatment has been reported to promote dendritic spine development in other related neurodegenerative diseases, such as Alzheimer's disease [366]. Other functions enriched within the differential best-classifier proteins (Low-MD-inactive/High-MD-active) are implicated in growth factor related pathways, which are known to be involved in wet MD pathogenesis [127]. Moreover, neovascularization in the wet variant of MD has

been linked to the signaling of some of the growth factors detected as sacubitril/valsartan-associated MD classifiers in this study, including FGF1 [127] and PDGF [144, 151].

Identification of potential biomarkers differentiating MD responses in Low-HF

We previously mentioned that some MoAs could be shared between the different groups of HF and MD (Supplementary Table 7). Knowing that, we focused on the shared MoAs between Low-HF and High-MD to analyze the special case comprising prototype-patients in which the treatment best reduces HF disease but increases MD adverse effect. In order to identify these patients, we compared the Low-HF ∩ Low-MD with Low-HF ∩ High-MD MoAs; Table 1 shows the 30 biomarkers identified. On the one hand, we found 16 proteins active in Low-HF ∩ Low-MD MoAs but inactive in Low-HF ∩ High-MD (15 of them shared with MD best-classifier proteins). On the other hand, 14 proteins were identified as inactive in Low-HF ∩ Low-MD and active in Low-HF ∩ High-MD MoAs (12 of them were MD best-classifier proteins). We calculated the GO enriched functions of these two groups and observed that "phosphatidylinositol bisphosphate kinase activity" is enriched among proteins that are active in Low-HF ∩ Low-MD MoAs. Instead, "fibrinolysis" was found to be enriched among proteins active in Low-HF ∩ High-MD MoAs (Table 2). With this, we conclude that among the group of prototype-patients for which sacubitril/valsartan improves HF treatment response, the modulation of fibrinolysis could play a role at inducing the MD adverse effect. Moreover, we propose 12 best-classifier proteins that may be considered as biomarkers for good prognosis of the side effect.

In fact, since neovascular MD development is characterized by subretinal

extravasations of novel vessels derived from the choroid (CNV) and the subsequent hemorrhage into the photoreceptor cell layer in the macula region [292], it might be reasonable to think that the modulation of fibrinolysis and blood coagulation pathways could play a role. The reported implication of some fibrinolysis related classifiers, such as FGB, SERPINE1 (PAI-1), and SERPING1, in neovascular MD development seems to support this hypothesis [440, 239, 167]. Besides, valsartan might be implicated in this mechanism, since it has been reported to modulate PAI-1 levels and promote fibrinolysis in different animal and human models [279, 301].

In addition, the presence of several other MD related classifiers in this list, such as IRS2 [8], PTGS2 [444], DCN [419] and FGF1 [362], further supports the interest of the classifiers as biomarkers of MD development in sacubitril/valsartan good responders.

Analysis of proposed biomarkers with GUILDify

In the previous section, we proposed 30 proteins that could potentially help to identify HF patients at risk of developing MD. To corroborate these biomarkers, we tested how many of them are found using a different approach also based on the use of functional networks. For this purpose, we used GUILDify v2.0 [6], a web server that extends the information of disease-gene associations through the protein-protein interactions network. GUILDify scores proteins according to their proximity with the genes associated with a disease (seeds). Using this web server, we identify a list of top-scoring proteins that are critical on transmitting the perturbation of disease genes through the network. The network used by GUILDify is completely independent from the HPN used in the TPMS, becoming an ideal, independent context to test the potential biomarkers.

Thus, we used GUILDify to indicate which of the potential biomarkers identified by TPMS may have a relevant role in the molecular mechanism of the drug. We ran GUILDify using the two targets of sacubitril/valsartan (NEP, AT1R) as seeds, and selected the top 2% scored nodes (defined as the "top-drug" set). We did the same with the phenotypes of HF and MD, using as seeds the 124 effectors of HF and 163 effectors of MD from the BED database. We merged the top scored sets of HF, MD and top-drug ("top-drug ∪ top-HF ∪ top-MD") and studied the overlap with the set of differential best-classifier proteins associated with MD and HF. Supplementary Table 8 shows the result of this analysis, with a significant representation of best-classifier proteins in most of the sets, especially on MD best-classifier proteins. Supplementary Table 9 shows the list of 13 proteins involved in this overlap. We have also checked the overlap with the 30 biomarkers proposed in the previous section, of which 10 are found in the merged set "top-drug ∪ top-HF ∪ top-MD" and are consequently significant (see Supplementary Tables 10 and 11).

Some of these candidates can be functionally linked to both diseases and the drug under study. For example, among these 10 classifiers, AGER has been implicated in both HF [162], through extracellular matrix remodeling, and MD development [25], through inflammation, oxidative stress, and basal laminar deposit formation between retinal pigment epithelium cells and the basal membrane; furthermore, this receptor is known to be modulated by AT1R [313], valsartan target. Similarly, FGF1 has been proposed to improve cardiac function after HF [137], as well as to promote choroid neovascularization leading to MD [127]. Moreover, FGF1 is regulated by angiotensin II through ATGR2 [234], another protein suggested as classifier in the current analysis that is known to mediate some of the effects of AT1R

antagonists, such as valsartan [255, 355]. Another candidate, NRG1, has been linked to myocardial regeneration after HF [135] and is known to lessen the development of neurodegenerative diseases such as Alzheimer's disease [437], which shares similar pathological features with MD [202]. NRG1 is also linked to the expression of neprilysin [437], sacubitril target. ITGB5 has been identified as risk locus for HF [400] and its modulation has been linked to lipofucsin accumulation in MD [203]. Interestingly, ATGR1 inhibitors have been reported to modulate ITGB5 expression in animal models [211]. Finally, IL1A has been proposed as an essential mediator of HF pathogenesis [57, 393] through inflammation modulations, and serum levels of this protein have been found increased in MD patients [286]. In addition, as described in previous sections, classifiers FGB, SERPINE1, and SERPING1 have been linked to MD [440, 239, 167] and are also known to play a role in HF development [447, 442, 275, 67]. According to these findings, the 10 potential biomarkers proposed by TPMS and identified with GUILDify might be prioritized when studying good responder HF patients at risk of MD development.

4.3 Discussion

Heart failure (HF) related hospitalizations are rising, especially in the developed countries [101]. Its prevalence is being influenced by different factors like age, nutritional habits, lifestyles or genetic, which complicates the development of treatments and the identification of universal biomarkers. Sacubitril/valsartan (marketed by Novartis as Entresto) is a drug combination that shows better results than conventional treatments, reducing both cardiovascular deaths and HF readmissions [274]. Many resources have been invested in the safety of this treatment,

although recent concern was risen on the possible risk of inducing amyloid accumulation-associated conditions in the long term, such as macular degeneration (MD). In this study, TPMS technology [18] was applied to uncover the different Mechanisms of Action (MoAs) of sacubitril/valsartan over HF and reveal its molecular relationship with MD. For this approach, we hypothesize that each generated MoA could correspond to a prototype-patient, and we stratified them in best vs worst responders to extract possible biomarkers. Additionally, we used GUILDify v2.0 [6] to compare the biomarkers proposed by TPMS and reinforce the results.

4.3.1 From TPMS models toward prototype-patients

TPMS was used to generate a model for understanding the pathways linking Sacubitril/valsartan treatment towards HF. As an ensemble model, a total of 200 MoA solutions satisfying the largest number of restrictions of the training set (and at least 80% of them) were computed. We then used the model intensity signal parameter, tSignal, to define four groups each bearing 50 prototype-patients, or MoA solutions. On one hand, we selected the models showing the highest (first quartile) or lowest (forth quartile) signal values for HF, and on the other the highest and lowest for MD. The cohorts were then analyzed to extract different sets of proteins able to classify the models according to treatment efficacy (as tSignal) toward HF or MD adverse event relationship, proposed as potential biomarkers. The GO enriched functions of these sets were then identified in order to pinpoint their biological role.

4.3.2 Identification and functional analysis of potential biomarkers

For the HF group comparison, best MoA-prototype-patient responders vs worst, we identified 6 proteins differentially active in best treatment responders, but inactive in bad responders. GO functions related to this protein set were found to be associated to processes like: (i) the SCAR complex; (ii) the positive regulation of actin nucleation; (iii) the regulation of neurotrophin TRK receptor; and (iv) to dendrite extension. These results coincide with previous findings in HF, where alteration of actin nucleation and polymerization mechanisms was observed [307, 207, 72]. Next, 45 additional proteins were found to be inactive in best responders, but active in the other group. Those were linked to: (i) phosphatidylinositol kinase mediated pathways; and (ii) MAP kinase mediated pathways. In this case, both signaling pathways had been associated to cardiac hypertrophy and subsequent HF [38, 292].

In case of MD groups, we identified 28 proteins differentially active in the group of lower probability of displaying MD, but inactive in the high group. The GO functions related to this set were: (i) dendritic spine development; (ii) positive regulation of vascular endothelial growth factor production; and (iii) phosphotyrosine binding. The presence of dendritic spine development and dorsal/ventral axon guidance had been previously linked as a role for sacubitril/valsartan-associated MD [440]. Additionally, valsartan had been reported to promote dendritic spine development in other related neurodegenerative diseases, such as Alzheimer's disease [239]. Regarding the proteins inactive in low MD group but active in high MD group comparison, 29 proteins were identified. The set was involved in functions such as: (i) dorsal/ventral axon guidance; (ii) fibroblast growth factor

receptor binding; and (iii) response to toxic substance. Those are functions implicated in growth factor related pathways, which are known to be involved in wet MD pathogenesis [167]. Interestingly, phosphatidylinositol bisphosphate kinase activity showed up as enriched function in both groups.

4.4 Concluding remarks

The differences found in the functional analyses support the initial hypothesis that the MoA models could represent possible different patients, where high response signals are related to a good response, and low response signals to bad or lower response. It also corroborates the previous usage of TPMS, by linking tSginal to the goodness of a response, as applied elsewhere [182, 181, 148, 257].

Finally, it favors the interpretation of individual or prototype models to be used in the finding of stereotype patients in order to develop or apply more accurate treatments.

Part III

ISCT PLATFORM

Chapter 5

An ISCT platform using TPMS models

5.1 Introduction

In order to move towards a more real-like or clinical application of the TPMS MoA models, I designed and applied a series of algorithms which enabled the generation of in silico clinical trials. The main idea was to exponentiate the generation of MoA models into multiple models representing multiple patients. Moreover, quantitative restrictions were included, by means of a Physiologically-based Pharmacokinetic (PBPK) modeling, in order to be able to simulate and differentiate distinct dosing and treatment schemes.

Because of the complexity of the resulting approach, involving different modeling types and procedures, a platform was defined as a semi-supervised stepwise protocol, mimicking real Clinical Trial's (CTs) pipeline. The resulting platform is divided into three main parts or phases: (i) initiation,

(ii) modeling, and (iii) analysis (Figure 16).

5.2 Initiation Phase

The first phase is the study definition which, asides from a data recollection step, includes the molecular characterization of the study pathology, adverse events and drug targets, as well as any other principal or secondary variable linked to the CT. This step is the same followed by any previous TPMS model analysis in order to map diseases and effects in the human protein network models, and which is also described in both Publication 1 (section 3.2) and 2 (section 4.2) of the present thesis.

Another key aspect of the platform is the definition of a Virtual Population (VPop), intended to represent and assign particular parameters to each of the models to be generated, now as Virtual Patients (VPs). Because a quantitative approach was added in the modeling procedure, I designed an algorithm able to assign both qualitative (e.g., molecular) and quantitative (e.g., physiological) parameters to the VPs. In this first phase, the characterization of the population details is defined.

Finally, a sample size algorithm was included in order to provide information on the statistical power attributed to the resulting models, and guide in choosing the appropriate number of patients.

5.3 Modeling Phase

Next, the modeling phase follows with the actual generation of the VPs. Here, the VPop algorithm is used, which is divided in two steps. The first is focused on assigning the physiological values (e.g., age, gender, weight) to the VPs for generating specific PBPK models. The second

one consists in using the population disease/molecular information ratios to assign pathological tags to the VPs, which will be later translated into molecular restrictions by the TPMS models.

After VPs are defined, a PBPK model is generated for each of patient. For that, I implemented a PBPK model strategy based on the SimBiology tool of MATLAB [184], consisting of a general whole-body model, which can then be parametrized for each VP depending on their physiological and drug treatment characteristics.

Subsequently, using the Interactome maps and PBPK outcome data, a QSP model using the TPMS is generated for each VP, including both molecular and pharmacological restrictions.

5.4 Analysis Phase

Finally, a set of tools were set available in order to analyze the generated models and data by using the Data Science strategy implemented in Anaxomics [198].

In the following publication, I described the details regarding the methodology used for carrying out an ISCT, in the case study of ADHD.

5.5 Methods to develop an in silico clinical trial: Computational head-to-head comparison of lisdexamfetamine and methylphenidate

Frontiers in Psychiatry. 2021 Nov 3;12:741170.
PMID: 34803764
PMCID: PMC8595241

DOI: 10.3389/fpsyt.2021.741170

José Ramón Gutiérrez Casares[1,*], Javier Quintero[2], Guillem Jorba[3,4,*], Valentin Junet[4,5], Vicente Martínez[6], Tamara Pozo-Rubio[6], Baldomero Oliva[3], Xavier Daura[5,7], José Manuel Mas[4] and Carmen Montoto[6]

[1]Unidad Ambulatoria de Psiquiatría y Salud Mental de la Infancia, Niñez y Adolescencia. Hospital Perpetuo Socorro., Spain

[2]Servicio de Psiquiatría. Hospital Universitario Infanta Leonor. Universidad Complutense., Spain

[3]Structural Bioinformatics (GRIB-IMIM). Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra,, Spain

[4]Anaxomics Biotech SL, Spain

[5]Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Spain

[6]Medical Department. Takeda Farmacéutica España, Spain

[7]Catalan Institution for Research and Advanced Studies (ICREA), Spain

[*]Corresponding authors

5.5.1 Abstract

Regulatory agencies encourage computer modelling and simulation to reduce the time and cost of clinical trials. Although still not classified in formal guidelines, system biology-based models represent a powerful tool for generating hypotheses with great molecular detail. Herein, we have applied a mechanistic head-to-head in silico clinical trial (ISCT) between two treatments for attention-deficit/hyperactivity disorder, to wit lisdexamfetamine (LDX) and methylphenidate (MPH). The ISCT was generated through three phases comprising (i) the molecular characterization of drugs and pathologies, (ii) the generation of adult

and children virtual populations (vPOPs) totaling 2,600 individuals and the creation of physiologically based pharmacokinetic (PBPK) and quantitative systems pharmacology (QSP) models, and (iii) data analysis with artificial intelligence methods. The characteristics of our vPOPs were in close agreement with real reference populations extracted from clinical trials, as did our PBPK models with in vivo parameters. The mechanisms of action of LDX and MPH were obtained from QSP models combining PBPK modelling of dosing schemes and systems biology-based modelling technology, i.e. therapeutic performance mapping system. The step-by-step process described here to undertake a head-to-head ISCT would allow obtaining mechanistic conclusions that could be extrapolated or used for predictions to a certain extent at the clinical level. Altogether, these computational techniques are proven an excellent tool for hypothesis-generation and would help reach a personalized medicine.

Keywords: attention-deficit/hyperactivity disorder, lisdexamfetamine, methylphenidate, mathematical modeling, in silico clinical trial

5.5.2 Introduction

To reduce clinical trials time and cost and to improve their outcomes' conclusiveness, regulatory agencies encourage the use of computer modeling and simulation (CM&S) approaches to optimize randomized clinical trials [171]. CM&S approaches are based on the analysis of existing data and experience, including real-world data studies, pharmacometrics modeling or, more recently, in silico clinical trials (ISCT). Although the concept emerged in the early 2000s [147, 79, 142], the term and proper definition of ISCT was widely established and accepted during the 2010 decade with the foundation of specific organizations to promote

the implementation of these approaches, such as the VPH Institute in 2011 or the Avicenna Alliance, founded by the European Commission, to create the research roadmap for ISCT [403]. In addition to its economic advantages, ISCT allow the exploration of drugs and diseases in many settings, thus, reducing risks for patients and the use of animal models to test hypotheses. CM&S and artificial intelligence-based approaches are crucial to achieving personalized, preventive, predictive, participative, and precise—the so-called 5P—medicine and healthcare [49].

Systems Biology and MID3 Guidelines

One of the most promising computational tools encompassing these concepts is systems biology or systems medicine [403, 51, 428, 417]. During the last 20 years, the US and European medicines agencies (FDA and EMA), in collaboration with the pharmaceutical industry, have been developing the guidelines and good practices to which these computational approaches should adhere. One of these guidelines is MID3, which describes the quantitative framework for predicting and extrapolating models' conclusions [409, 263]. Establishing three categories based on the relevance of the conclusions, MID3 is meant to guide industry decision-making [265] or regulatory assessment [102]. Accordingly, models can be classified as (i) "LOW" impact, when information obtained from them cannot be directly used to make clinical or commercial decisions [e.g., physiologically based pharmacokinetic (PBPK)] models; (ii) "MEDIUM" impact, for models providing helpful information for strategic conditioning of future trial data [e.g., studies to determine optimal dosing, target population, sample size, design of future trials, or study of mechanisms of action (MoA) of compounds]; and (iii) "HIGH" impact, for cases where conclusions support decision-making without the need for additional

experimental or trial studies (e.g., simulations replacing direct clinical trial data in children or oncologic patients that provide evidence on efficacy and safety to uphold regulatory submission package and labeling). While pharmacometric models are under evaluation for acceptance as HIGH impact models, systems biology-based models are still in debate [287]. However, they possess an undeniable great potential in providing molecular detail, generating hypotheses, and suggesting specific molecular solutions to complex pathophysiological problems.

Proof-of-Concept: ADHD

Attention-deficit/hyperactivity disorder (ADHD) is a complex ailment with a prevalence in children ranging from 6 to 10% [5]. Besides, ADHD exhibits an important long-term persistence [5], affecting $\tilde{5}$% of adults [214, 316, 436]. Around 30–50% of children with ADHD continue to manifest symptoms, inattention in particular, in adulthood [3, 424]. Comorbid psychiatric disorders are present in up to 67% of ADHD pediatric-adolescent patients [236] and almost 80% of adults [210]. These comorbidities can complicate ADHD diagnosis and treatment [273, 296] and include depression, anxiety, bipolar disorder, binge eating, tics, conduct disorder, personality disorder and non-alcoholic substance abuse, among others [424, 262]. Recent findings suggest a direct relationship between ADHD and the development of these comorbidities [296, 172, 43], likely involving a genetic connection [224], although results on this subject remain controversial.

ADHD management comprises pharmacologic and non-pharmacologic treatments. Medications include stimulant [amphetamines and methylphenidate (MPH)] and non-stimulant drugs (atomoxetine,

extended-release clonidine, and guanfacine), with the former being recommended as first-line treatment [91]. Several modifications to improve the characteristics of amphetamines have been performed, among which the design of the prodrug lisdexamfetamine (LDX, Vyvanse©in the US and Elvanse ©in Europe) and the development of extended-release formulations [such as the osmotic release oral system (OROS) of MPH, Concerta©or Medikinet©retard]. Although a pediatric clinical trial analyzing LDX and MPH is currently ongoing [84, 85], there are no explicitly designed head-to-head trials comparing these treatments, neither on the pediatric nor adult population.

We present here the methods of the Therapeutic Performance Mapping System (TPMS) technology, which allow the generation of virtual patients and PBPK and systems biology-based models with the purpose of performing ISCTs. To demonstrate the applicability of the method, we used as case-study a mechanistic head-to-head ISCT between LDX and MPH (Elvanse©vs. Concerta©in the pediatric-adolescent population and Elvanse©vs. Medikinet©retard in the adult population) using a crossover–like design. The objective of this ISCT was to model the efficacies of the two drugs and compare them in a virtual head-to-head setting. Additionally, we describe an approach to measure and compare the output results in terms of efficacy of the two medications, the molecular mechanisms triggered, and the response to ADHD management in a diverse population of virtual patients, including patients with the most common psychiatric comorbidities.

5.5.3 Methods

This methods study details the steps and modeling approaches to carry out the ISCT (Figure 16). Before the study trial (phase I), drugs and pathological conditions were molecularly characterized and reference populations defined. In the modeling stage (phase II), a series of virtual populations and PBPK and quantitative systems pharmacology (QSP) models were generated and embedded in the ISCT as a means of virtual patient recruitment. At this step, the models were optimized to reproduce known clinical efficacy findings according to the primary outcome of the study, i.e., the model-based clinical efficacy-related measure herein proposed, based on modeled protein activity over ADHD molecular definition. Finally, in the analysis phase (phase III), the molecular variability among patients was explored by analyzing all ADHD models, patient by patient.

Population Definition—Virtual Patients

Two types of virtual populations (vPOPs) were generated: adult (>18 years old) and pediatric-adolescent (6–17 years old) vPOPs. As reference demographic and comorbidity parameters to generate the VPOPs, the following studies were used: NCT00730249 [329] (MPH) and NCT00337285 [141] (LDX) for adults; and NCT00763971 study [84] (LDX and MPH) for the pediatric-adolescent population. These clinical trials presented standard inclusion and exclusion criteria for ADHD evaluation, which were appropriate for the case-study herein proposed and showed homogeneous demographic values when compared to other clinical trials with equivalent inclusion and exclusion criteria.

Additionally, standard population distribution data was used to fill

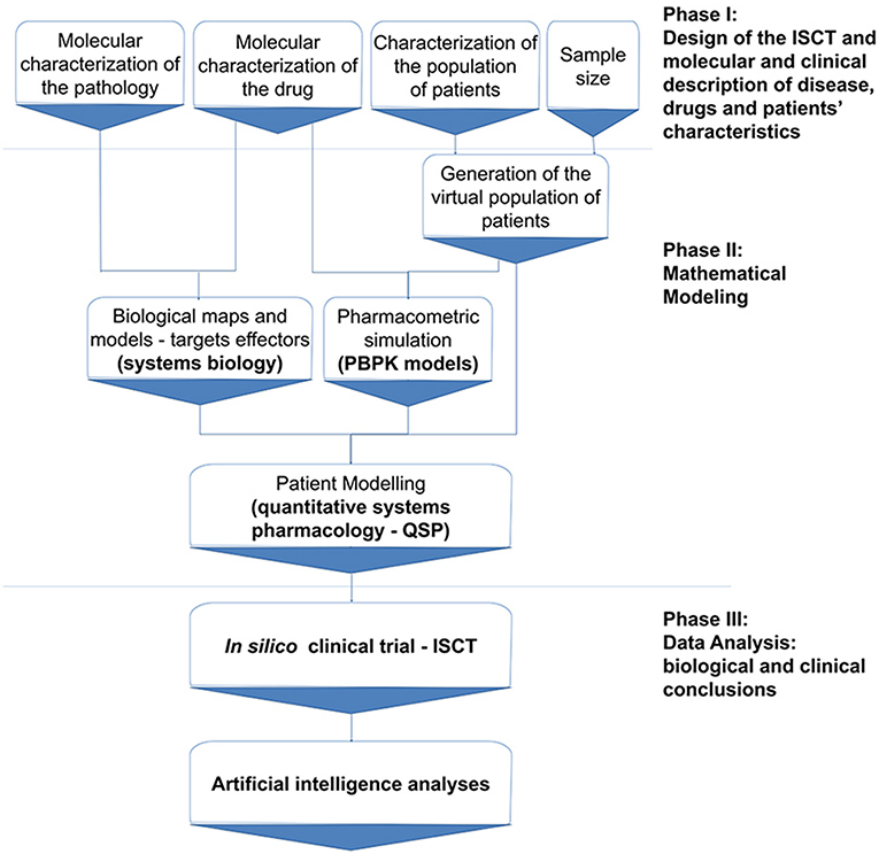

Figure 16: In silico clinical trial protocol overview.
The protocol is divided into three main stages: Phase I, including trial design and information compilation; Phase II, comprising mathematical modeling; and Phase III, consisting of data analysis according to the trial design. ISCT, in silico clinical trial; PBPK, Physiologically based pharmacokinetic; QSP, Quantitative systems biology.

incomplete demographic parameters. For adults, ESS Round 7 [106] was used, while data from the World Health Organization (WHO) growth information [258] was retrieved for the pediatric-adolescent population.

All virtual patients created had ADHD, and specific branches for the different comorbidities were also generated, as previously described [198]. ADHD and comorbidities definitions were obtained by thorough literature review of current molecular knowledge on each condition (see

Supplementary Methods in Supplementary Material 1; Supplementary Tables A, B in Supplementary Material 2).

Sample Size Calculation

Since data on treated and non-treated patients is not available, we considered that a number of patients large enough to discriminate among ADHD patients and healthy individuals would also be large enough to detect efficacy-associated changes for each drug. Therefore, to generate enough patients and ensure having sufficient statistical power when performing data analyses, the sample size approach described below was carried out. Because TPMS' drug efficacy outcomes are based on predicted protein activity (i.e., tSignal, Equation 1—defined in section Systems Biology Maps and Models), this methodology was based on experimental measures that can relate to protein activity variability, particularly gene expression.

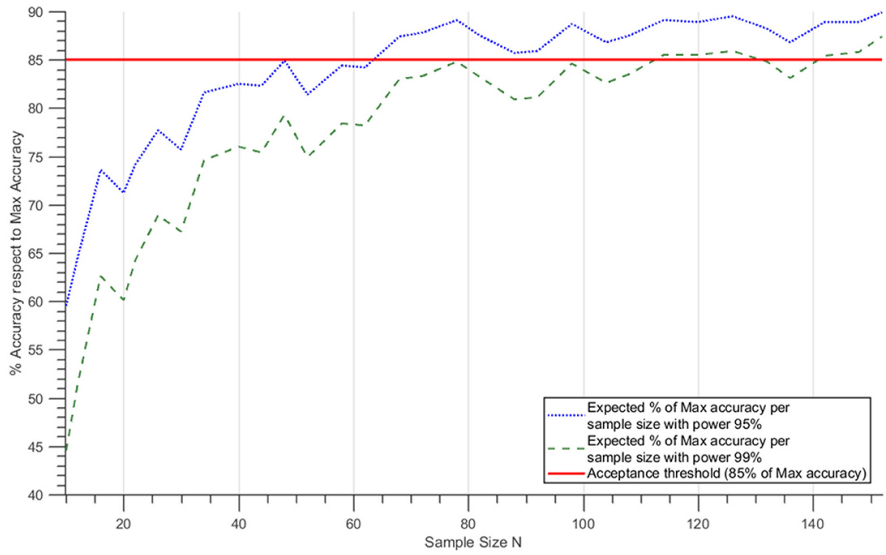

Figure 17: Expected percentage of best accuracy as a function of sample size. Dotted blue and discontinuous green lines correspond to the mean % best accuracy reached for each sample size at statistical power 95 and 99%, respectively, assuming a normal distribution of the accuracy variation and estimating the means and the standard deviation for each sample size. The red line shows the 85% Max accuracy level.

First, gene expression data groups identified as control- "healthy" and case- "disease" were retrieved from Gene Expression Omnibus (GEO) experiments [31] and then treated and normalized using R packages, parameters, and steps defined by Law et al. [237]. Afterwards, a protocol based on the method introduced by Mukherjee et al. [284] and Figueroa et al. [117] was followed to explore the variation in accuracy and statistical power induced by changes in the sample size. To that end, GEO patient-normalized gene expression datasets are submitted to sampling-without-replacement combined with a linear regression classification method [129]. The latter allows the identification of the best classifiers (proteins) to separate control-healthy from case-disease patients, and these classifiers are used to compute the highest possible accuracy ("Max accuracy"). Progressive sampling is then applied to obtain subsets of balanced samples from both cohorts (case-disease vs. control) in a 1:1 ratio. These subsets are tested for sample sizes ranging from eight to the number of the smallest cohort performing 100 repetitions per sample size. Each subset is used to train a linear classifier based on two features extracted by feature selection procedures previously described [198]. The accuracy achieved for each classifier is estimated using k-fold cross-validation (k = 10) [226]. Finally, taking as reference the Max accuracy, the percentage of max accuracy reached for each subset of samples and total samples is calculated using the classifiers obtained for that subset.

For the present ADHD study-case, RNAseq records from the entry GSE159104 [272] were selected, where two cohorts of patients were already identified and labeled as control (healthy) and ADHD (case-disease). The variability within the genes or proteins involved in the ADHD molecular definition (see Supplementary Methods in Supplementary Material 1;

Supplementary Tables A, B in Supplementary Material 2) was evaluated for the 154 samples (78 control, 76 ADHD) included in the GEO experiment. After finalizing the abovementioned procedure, statistical powers of 95 and 99% were used, based on classification errors [284], and a value of 85% of Max accuracy was set as minimum valid threshold (Figure 17). Considering a statistical power of 95%, we deemed 68 samples (34 control and 34 ADHD) to be enough to achieve the objectives of the analysis in our simulation. Under these premises, 142 samples (71 control and 71 ADHD) were adequate to reach the target accuracy with 99% power (although more RNAseq samples would be required to ensure curve stabilization). Accordingly, at least 100 virtual patients were built per each patient group (minimum sample size of 200 samples per analysis).

Patient Distribution

The two populations, adult and pediatric-adolescent, were segmented into nine arms each (a total of 18 arms) to facilitate the simulation and the analysis. One arm accounted for ADHD without any comorbidity, while the eight additional arms contained patients with ADHD and one, or a combination, of comorbid psychiatric conditions.

Each of the arms accounting for comorbidities had 100 patients, while arms related to ADHD alone consisted of 500 patients, with the aim of maximizing the number of patients with different demographical characteristics. Consequently, a total of 2,600 patients were included in the simulation: 1,300 adults (Figure 18) and 1,300 children-adolescents (Figure 19). All of them were treated sequentially with LDX and MPH using the adequate dosing scheme. According to the study's in silico nature, the files containing the models of each virtual patient could be cloned; thus, no wash-out period was needed.

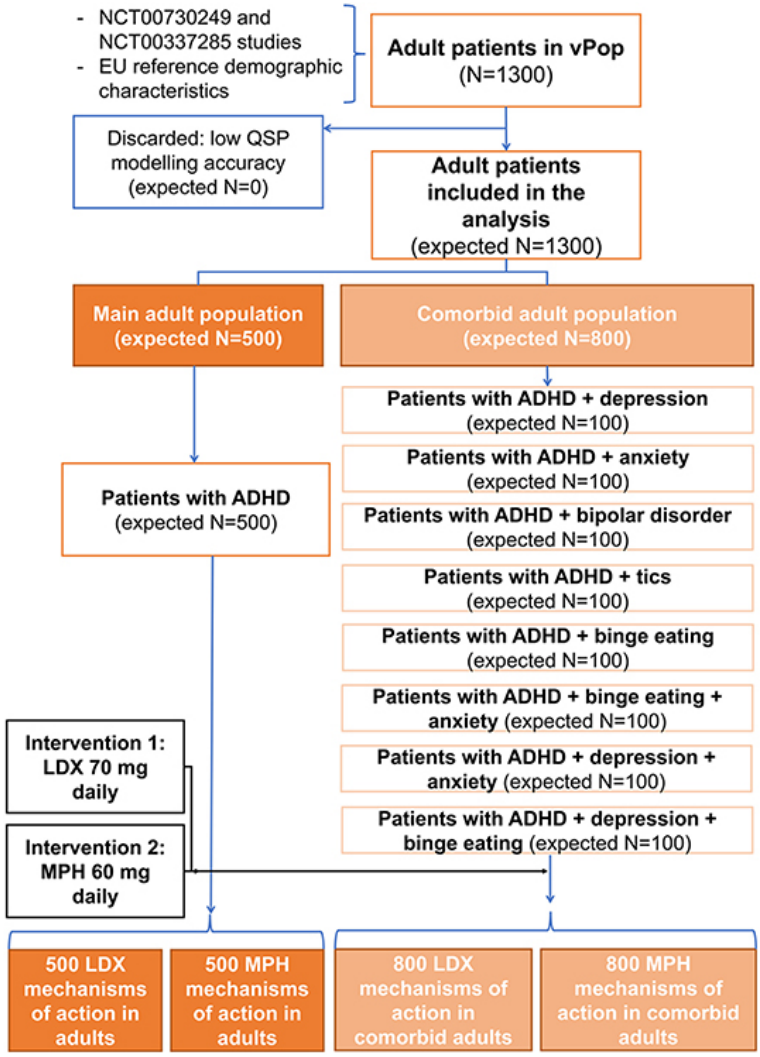

Figure 18: Comorbidities distribution and treatment allocation in the adult virtual population.
ADHD, Attention-deficit/hyperactivity disorder; LDX, Lisdexamfetamine; MPH, Methylphenidate; QSP, Quantitative systems biology; vPOP, Virtual population.

Intervention Definition

According to their population group, the patients included in the ISCT were treated with different formulations and doses of LDX and MPH in a two-period crossover-like study design. For all patients, the same initial

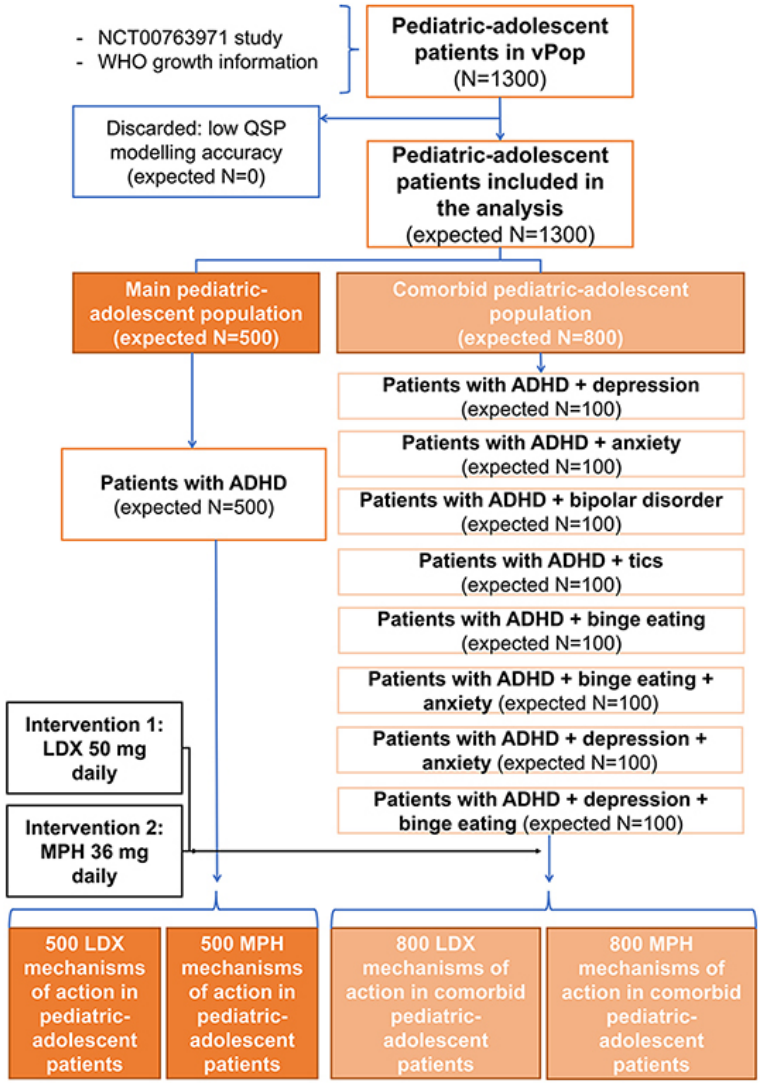

Figure 19: Comorbidities distribution and treatment allocation in the pediatric-adolescent virtual population.
ADHD, Attention-deficit/hyperactivity disorder; LDX, Lisdexamfetamine; MPH, Methylphenidate; QSP, Quantitative systems biology; vPOP, Virtual population.

state was used at each period, hence carryover effect was assumed zero. Dosage, molecular target profile, and pharmacokinetic information were needed for the QSP modeling herein proposed.

Table 9: Identified protein targets for lisdexamfetamine and methylphenidate.

Gene name	Protein name	Effect*	Reference of LDX target	Reference of MPH target
TAAR1	Trace amine-associated receptor 1	1	(49)	–
SLC18A2	Synaptic vesicular amine transporter (VMAT2)	−1	(49, 50)	–
SLC6A3	Sodium-dependent dopamine transporter (DAT)	−1	(50–52)	(53–55)
SLC6A2	Sodium-dependent noradrenaline transporter (NET)	−1	(50, 52)	(53–55)
SLC6A4	Sodium-dependent serotonin transporter (SERT)	−1	(50)	–
MAOA	Amine oxidase (flavin-containing) A	−1	(52, 56)	–
MAOB	Amine oxidase (flavin-containing) B	−1	(52, 56)	–
HTR1A	5-hydroxytryptamine receptor 1A	1	–	(57, 58)

*Effect refers to the drug's action on the protein, 1 denotes activation of protein function, −1 inhibition of protein function.
LDX, lisdexamfetamine; MPH, methylphenidate.

Dosage

Dosage schemes were simulated differently in the pediatric-adolescent and adult populations according to usual clinical practices. Adults were treated with LDX (Elvanse©) 70 mg and children with LDX 50 mg. Different doses and types of modified release systems for MPH were considered in the simulation, corresponding to different commercial formulations: (i) for adults, Medikinet

copyright 60 mg with modified-release (also known as Medikinet©XL or Medikinet©Retard), based in multiarticular beads that combine 50% immediate and 50% extended-release [89]; and (ii) for the pediatric-adolescent population, Concerta©36 mg, an osmotic release system (OROS technology) with a 22% of the total amount available for immediate release (the remaining 78% corresponding to the osmotically controlled extended-release) [98].

Molecular Target Profile The molecular target profile identification was performed through a review of official regulatory sources [European Medicines Agency—EMA, European Public Assessment Report (EPAR)—and Food and Drug Administration—FDA, Multidisciplinary and Chemistry reviews and Label], drug-target–dedicated databases [DrugBank [425], STITCH [381], SuperTarget [161]] and the scientific literature (the

Table 10: Summary of pharmacometrics information used for PBPK modeling.

Drug	% Bioavailability (Ref.)	Main clearance organ
Elvanse®	96.4 (59)	Kidney (60)
Medikinet® with modified release	30 (61, 62)	Kidney (63)
Concerta®	32 (64)	Kidney (63, 65)

specific searches performed can be found in Supplementary Methods in Supplementary Material 1). This information was integrated into the TPMS technology-based MoA models for each drug. Table 9 contains the proteins defining the target profile of LDX and MPH.

Pharmacokinetics Information

Bioavailability and drug's information on main clearance organ were retrieved from published studies and set for the corresponding PBPK models (Table 10). Moreover, previous PK studies were used to fit the generated PBPK models, to parameterize absorption and drugs' clearance ratios, and to validate the models. The reference studies used were Krishnan and Zhang [229] for LDX in adults, Boellner et al. [50] for LDX in children, the EPAR [41] for Medikinet©with modified-release, and Maldonado [261] for Concerta©. All three drugs were administered orally and crossed the blood-brain barrier.

Modeling Methodology

TPMS ISCT is divided into three types of modeling approaches (Figure 16). First, virtual patients are generated containing demographic information and disease tags. Afterwards, PBPK models are constructed using each patient's demographic variables, which are then used to infer inter-patient

specific drug concentration-related knowledge. Finally, the patient-specific drug concentration and disease-related data, and protein mapping according to pathophysiological information, are used for generating patient-specific MoA-QSP models of the drugs under study, here MPH and LDX.

Virtual Population Modeling

For the construction or recruitment of vPOPs, randomized populational demographic characteristics are generated using two types of data sources: (i) original or reference population with demographic characteristics to be mimicked [age, weight, height, and/or body mass index (BMI)]; and (ii) standard population distributions, retrieved from populational studies. For the present ADHD study-case, the recruitment of each vPOP was based on the demographical parametric descriptors defined in section Population Definition—Virtual Patients' [reference clinical trials (30, 32, 33), European standard population [106], and WHO growth information [258]].

For adult population, an adapted version of the algorithm proposed by Allen et al. [11] was used to generate the population of individuals virtually recruited in the trial. As a first step, this algorithm generates a multivariate normal distribution (MVND) with the demographic means and standard deviations from the original population. The standard population distribution values are used to fill in the potential missing demographic information. A simulated annealing strategy is then used to minimize a cost function by using the patients generated in the MVND as starting points (see Supplementary Methods in Supplementary Material 1).

In the pediatric-adolescent population, a modification of the protocol used for adult population was applied to adjust better the dependence of morphometric measures for ages 0–17 years. First, the standard population distribution, taken from the growth information published by

the WHO [258], was used to create a reference MVND. Then, a sampling strategy based on a Metropolis-Hastings method [71] was applied to reach the original population distributions (see Supplementary Methods in Supplementary Material 1).

The final distribution values for adult and pediatric-adolescent populations were statistically compared (one sample z-test) to the original means and standard deviations; only populations not significantly different from the original population (p-value > 0.05) were accepted and kept for posterior modeling steps. For both population types, corresponding comorbidity-related tags were assigned to the patients allocated to each of the 18 ISCT arms (Figure 19).

Demographic parameters were used to obtain accurate and individualized PBPK models of the drugs, while comorbidity data, once translated into molecular information, influenced the patients' corresponding QSP models.

Systems Biology Maps and Models

TPMS technology [198] generates mathematical models that use known biological, medical, and pharmacological information as training data (see Supplementary Table C in Supplementary Material 2) to simulate the behavior of drugs and the pathophysiology of diseases in terms of changes in protein activity. This methodology uses supervised machine learning methods based on a human protein functional network to infer information at the clinical and protein levels. Here, TPMS was used to build the mathematical models to simulate the behavior of LDX and MPH over ADHD by modeling the changes in proteins' activity defining the disease. While generating TPMS models, molecular information relating to psychiatric comorbidities was added to denote the different neurophysiological ADHD patient types.

The resulting models allowed the extraction of several protein activity measures. Therefore, the model-derived parameter tSignal (Equation 1) [198], which ranges between 1 and -1, applied to the molecular definition of clinical conditions (in this case, ADHD molecular definition, as detailed in Supplementary Table B in Supplementary Material 2) permitted access to clinically relevant information at a model-patient level.

$$TSignal = \frac{1}{n} \sum_{i=1}^{n} v_i y_i$$

Where n is the number of proteins defining the protein set; vi are the protein signs (active or inactive) according to each disease/comorbidity definitions; and yi are the resulting modeled signal values achieved by each protein "i" after stimulating the model with the corresponding drug.

Physiologically Based Pharmacokinetic Models

A PBPK model per virtual patient was built to describe the relationship between drug doses and drug concentration in different organs within the human body. The PBPK model structure used consists of 14 predefined compartments representing the human body's main organs and tissues, a simplified version of a previously reported model [310] (Figure 20). Blood acts as the central compartment by interconnecting the rest of the system through blood flows, and the whole system can be disturbed by administering a drug dose in any of the following organs or compartments: gut (oral drugs), blood (intravenous drugs), or skin (subcutaneous drugs). Similarly, clearance of drugs and compounds is restricted to three compartments: gut, liver, and kidneys. The equations associated with blood flow rates and organ/tissue volumes are taken from Brochot and Quindroit [53]. These variables depend on cardiac frequency, age, BMI, and gender and yield individualized models as described elsewhere [77]. Here, blood volume was readjusted to fit the volume of distribution of each compound

for optimized modeling.

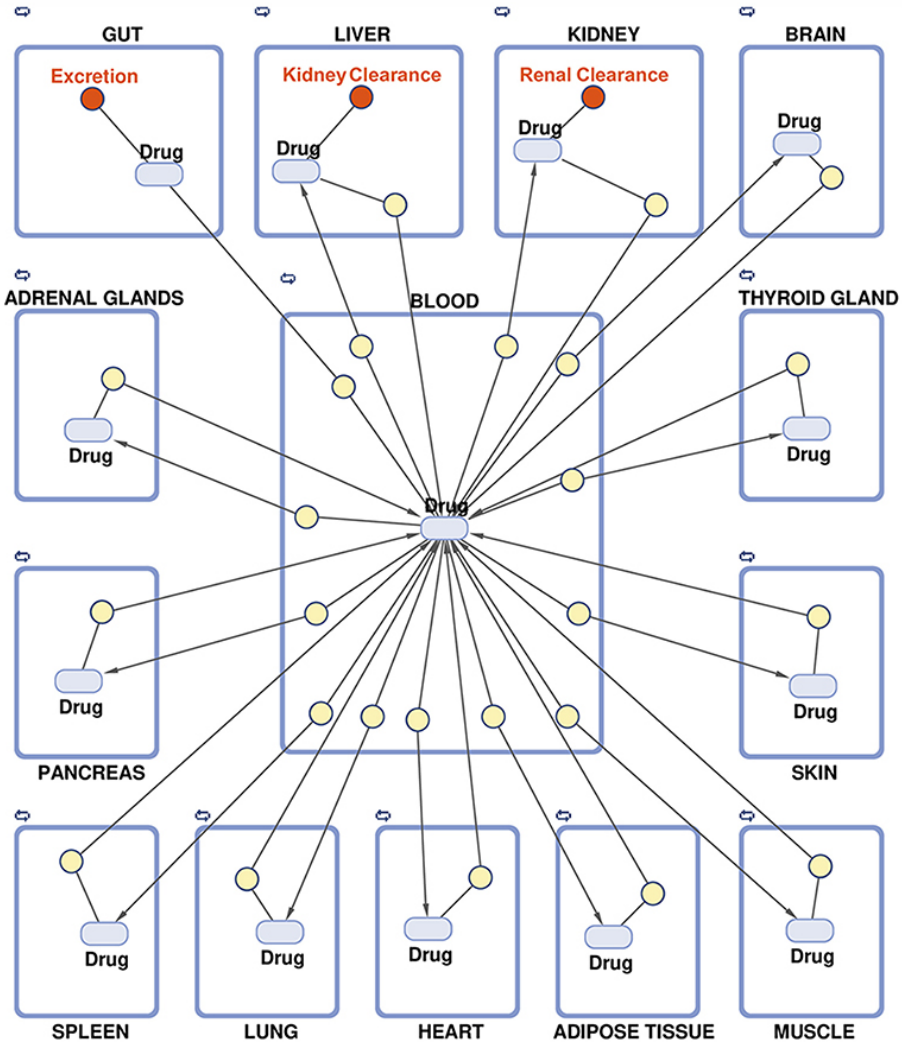

Figure 20: Schematic representation of the multi-compartment model for physiologically based pharmacokinetic modeling.

Parameters related to the anatomy and physiology of each specific patient's human body were used to mathematically describe drugs' internal flow, i.e., drugs' absorption, distribution, metabolism, and excretion (ADME) processes [341]. The drug's absorption and clearance constant parameters were calculated by fitting the general model to existing

real pharmacokinetics data points for d-amphetamine (d-Amph, active compound for LDX) and MPH (Table 10) [229, 50, 41, 261]. For other ADHD drugs (see Supplementary Methods in Supplementary Material 1), pharmacokinetics data used can be found in Supplementary Table D in Supplementary Material 2. Regarding MPH, as extended-release capsules are not easily simulated, an approximation using repeated administration of lower doses was used. This strategy had already been described for the two MPH extended-release formulations used here, and resulted in similar concentration dynamics: (i) for Medikinet©with modified-release, considered to have an equivalent MPH bioavailability to Ritalin©[96], a twice-a-day administration was simulated with half the dosage for each simulated administration, and (ii) for Concerta©, three administrations were simulated with one-third of the original dose for each administration [154].

The whole PBPK compartments model is implemented in MATLAB^{TM}, and differential equations describing the kinetics of the compounds and the fitting procedures are integrated by using SimBiology Toolkit [184].

Quantitative Systems Pharmacology Models—Quantitative Mechanism of Action

A QSP model enclosing PBPK model outputs and TPMS model maps was generated for each patient of the vPOPs. QSP models are generated following the TPMS methodology previously described [198] but incorporating drug concentration data at different timepoints in addition to molecular inputs, which add patient-specific quantitative data. To this end, a set of drug concentration timepoints in the target tissue—brain in this study—can be associated with the modulation of the drug's target proteins. Additionally, by applying the EC50 equation definition and using

clinical efficacy observations, the drug's effect on the disease-characterized proteins in the target tissue can also be calculated (see Supplementary Methods in Supplementary Material 1). Accordingly, the resulting MPH and LDX drug's target modulation-efficacy relationships were used as extra parameters in the TPMS training set, resulting in the final QSP models. The latter had the same output format as the systems biology MoA models previously described [198], but included quantitative information related to drug concentration. Hence, these models were used to answer additional questions related to individual differences among patients or treatment comparisons. At least 50 mathematical solutions per patient were computed during the QSP modeling to account for intra-patient variability, with accuracies >85% with respect to the TPMS training set [198].

Efficacy Outcomes and Measures Definition and Optimization

Molecular Measures

Due to the systems-biology–based nature of the virtual patients' resulting models, all measures were centered on protein activity. As previously described [198], after modeling a drug MoA on each patient, a protein activity value in the range (-1, 1) was obtained. These values can be either analyzed individually or combined in protein functional groups to evaluate biological concepts, such as diseases or comorbidities.

Efficacy Outcome

As for any clinical trial, in which the primary outcome is usually related to the drug's efficacy, our primary case-study goal was to identify and compare both drug's efficacies. Accordingly, a selection and conversion methodology were defined to select the protein set within the ADHD definition that best explained a chosen efficacy metric, and we transformed

the protein activities of that set into a model-derived measure that correlated with an actual clinical measure. The clinical variable used here was the ADHD Rating Scale IV (ADHD-RS IV, change from baseline). Three steps were followed to convert TPMS-model protein activities into ADHD-RS IV values: (i) select a model-derived activity measure (i.e., tSignal) that could be used as a proxy for efficacy; (ii) carry out ADHD molecular characterization, which consisted on a curated review of the scientific literature available in the PubMed database to identify proteins functionally involved in ADHD (see Supplementary Methods in Supplementary Material 1; and Supplementary Tables A, B in Supplementary Material 2; and (iii) optimize by trimming the ADHD molecular definition using real clinical trial efficacy observations (using ADHD-RS IV). In the third step, a series of eligible ADHD clinical trials meeting our inclusion criteria and measuring ADHD-RS IV in relevant drugs (Table 11; Supplementary Methods in Supplementary Material 1) were compiled. The reported ADHD-RS IV values were then used for ADHD molecular definition refinement through Pearson's correlation (Supplementary Methods in Supplementary Material 1); the final ADHD definitions used for outcome measurement are displayed in Supplementary Table E in Supplementary Material 2.

Model-derived ADHD outcome measures were optimized separately for adults' and pediatric-adolescent's clinical trials to reduce noise on the molecular definition.

Data Analysis

For the analysis of the population demographic and PBPK parameters, descriptive statistics were used (mean and standard deviation, frequency tables, or pie charts), and appropriate parametric and non-parametric tests

Table 11: List of clinical trials used for attention-deficit/hyperactivity disorder model-derived efficacy measure optimization.

Clinical trial number/PMID	Title	References
Adult clinical trials		
PMID: 17137560	Efficacy and safety of dexmethylphenidate extended-release capsules in adults with attention-deficit/hyperactivity disorder	(78)
PMID: 20576091	Randomized, double-blind, placebo-controlled, crossover study of the efficacy and safety of lisdexamfetamine dimesylate in adults with attention-deficit/hyperactivity disorder: novel findings using a simulated adult workplace environment design	(79)
NCT00337285	A long-term, open-label, and single-arm study of NRP104 30, 50, or 70 mg per day in adults with attention deficit hyperactivity disorder (ADHD)	(33)
NCT01270555	Efficacy of bupropion SR for attention deficit hyperactivity disorder (ADHD) in adults with recent past or current substance use disorders	(80)
NCT01259492	A 40-week, randomized, double-blind, placebo-controlled, multicenter efficacy and safety study of methylphenidate HCl extended release in the treatment of adult patients with childhood-onset ADHD	(81)
NCT02141113	Double-blind, randomized, placebo-controlled, single- center, dose optimization study evaluating efficacy and safety of guanfacine hydrochloride in combination with Psychostimulants in adults aged 18–65 years with a diagnosis of ADHD	(82)
NCT02604407	A phase 3, randomized, double-blind, multicenter, placebo-controlled, forced-dose titration, safety and efficacy study of SHP465 in adults aged 18–55 years with attention-deficit/hyperactivity disorder (ADHD)	(83)
Pediatric-adolescent clinical trials		
NCT00507065	A phase III, randomized, multicenter, double-blind, parallel-group, placebo-controlled safety and efficacy study of ADDERALL XR with an open label extension, in the treatment of adolescents aged 13–17 with ADHD	(84)
PMID: 17577466	Efficacy and tolerability of lisdexamfetamine dimesylate (NRP-104) in children with attention-deficit/hyperactivity disorder: a phase III, multicenter, randomized, double-blind, forced-dose, parallel-group study	(85)
NCT00447278	A study comparing the effect of atomoxetine vs. other standard care therapy on the long term functioning in attention-deficit/hyperactivity disorder (ADHD) children and adolescents (ADHD LIFE)	(86)
NCT00393042	Sleep and tolerability of extended release dexmethylphenidate vs. mixed amphetamine salts: a double blind, placebo controlled study (SAT STUDY)	(87)
PMID: 21241954	Clonidine extended-release tablets for pediatric patients with attention-deficit/hyperactivity disorder	(88)
NCT00763971	A phase III, randomized, double-blind, multicentre, parallel-group, placebo- and active-controlled, dose-optimization safety and efficacy study of lisdexamfetamine dimesylate (LDX) in children and adolescents aged 6–17 with attention-deficit/hyperactivity disorder (ADHD)	(30)
NCT01244490	A phase 3, randomized, double-blind, multicentre, parallel-group, placebo- and active-reference, dose-optimization efficacy and safety study of extended-release guanfacine hydrochloride in children and adolescents aged 6–17 years with attention-deficit/hyperactivity disorder	(89)
NCT01328756	A phase 4, open-label, multicentre, safety study of lisdexamfetamine dimesylate in children and adolescents with attention-deficit/hyperactivity disorder (ADHD)	(90)

applied. The p-value was taken as a measure of the significance of the fitting to the reference population.

The data was analyzed employing MATLAB^{TM} functions and Python or R packages to compare means and/or standard deviation between data distributions. Analyses with <30 samples were treated with non-parametric tests, while comparisons involving more than 30 samples were performed assuming a normal distribution and treated with parametric tests; in all cases, the applied test was reported. The statistical significance level was set at $p < 0.05$. False discovery rate (FDR) was used to control type I errors by applying the Benjamini-Hochberg [36] multi-test correction method, whenever relevant. All analyses were performed according to the described analytical strategy.

The accuracies of systems biology and QSP models were calculated for each solution within each individual model and expressed as the percentage of compliance of all drug-pathophysiology relationships included in the training set [198].

To evaluate the sensitivity of systems biology models, a local sensitivity analysis based in the SOBOL methodology [446] was performed to explore whether the variation in the protein activity (-1, 1) of the proteins in the models influenced the MoA models response of the two drugs (ADHD, as defined in Supplementary Tables A, B in Supplementary Material 2). According to the SOBOL terminology, TPMS models could be redefined as:
$tSignal(ADHD) = TPMS(X)$ for $X = X_1, X_2, X_3, \ldots, X_n$

Where Xi corresponds to each one of the parameters (here protein nodes activity) used in the models. Then, the variation of response model tSignal for each Xi parameter variation can be expressed as:
$dTPMSd(X_i) = d(tSignal)d(X_i)$

The tSignal difference compared to the original model was computed for all values in the range tested, and the mean for each protein was calculated and evaluated as a percentage with respect to the maximal possible tSignal variation, set as 2 [(-1, 1) difference] minus the original tSignal.

An unsupervised clustering strategy was applied to obtain groups of two to seven clusters of MoAs to evaluate the molecular variability of the generated models. The two (adults and children-adolescents) complete sets of 1,000 QSP ADHD patient mechanistic models (500 for LDX and 500 for MPH) were evaluated separately, taking into account the final activation values of the ADHD protein effectors modulated by both drugs.

Clusters were obtained using K means algorithm [130]. The clustering analysis was performed using all features (effector proteins) and principal component analysis (PCA) dimensionality reduction with five dimensions [196]. Four quality indicators were used to select the optimal number of clusters: Hopkins statistics [24] to measure the cluster tendency of a data set; Silhouette index [344] to weigh the cohesion of the clusters and Jaccard Bootstrap Index [76] to gauge the similarity and diversity of sample sets. Clusters were also filtered by heavily unbalanced groups, according to the Silhouette index ratio [344]. Classification analysis, as described elsewhere [198], were applied to molecularly describe the identified clusters.

Ethics

Only aggregated patient data from published clinical trials were used in the current project [84, 329, 141, 373, 423, 82, 179, 58, 422, 372, 44, 81, 376, 190, 166, 86, 83]. Aggregated patient data prevents individual patients' identification and, thus, avoids the need for approval from an ethics committee or institutional review board.

Computational Availability

All simulations described in this project were executed in the Anaxomics' cloud computing, which integrates more than 800 computational threads in machines with 64 Gigabytes of RAM. Software, databases, and tools are the property of Anaxomics Biotech.

Table 12: Demographic characteristics of the adult virtual population and the reference population.

	Virtual population ($N = 500$)	Reference population ($N = 511$)	p-value[a]
Sex (% females)	41.6	41.6	NA
Age (years)	36.98 ± 10.31	36.58 ± 10.10	0.37
Height (cm)	172.03 ± 9.98	171.7 ± 9.4	0.43
Weight (kg)	77.74 ± 16.95	78.75 ± 17.20	0.19
BMI (kg/m^2)	26.51 ± 6.55	NA[b]	NA[b]

Figures are mean ± standard deviation unless otherwise stated.
BMI, Body mass index; NA, not applicable.
[a]*Calculated with the unpaired two-tailed T Student's test.*
[b]*For demographic data not provided in the reference clinical trial, European mean values were used.*

5.5.4 Results

Demographic Characteristics

The characteristics of adult and pediatric-adolescent vPOPs of ADHD patients were generated from the proportions of demographic characteristics reported in the corresponding clinical trials. Additionally, eight subpopulations with different comorbidities (depression, anxiety, bipolar disorder, tics, and binge eating disorder) were created using the same method for both populations to evaluate the impact of comorbidities on the drugs' efficacy. The characteristics of our modeled vPOPs can be found in Figure 21A for adults and Figure 21B for the pediatric-adolescent population. The characteristics of the adult vPOP showed no significant differences with real reference populations extracted from clinical trials (Table 12). The same was true for the pediatric-adolescent vPOP (Table 13).

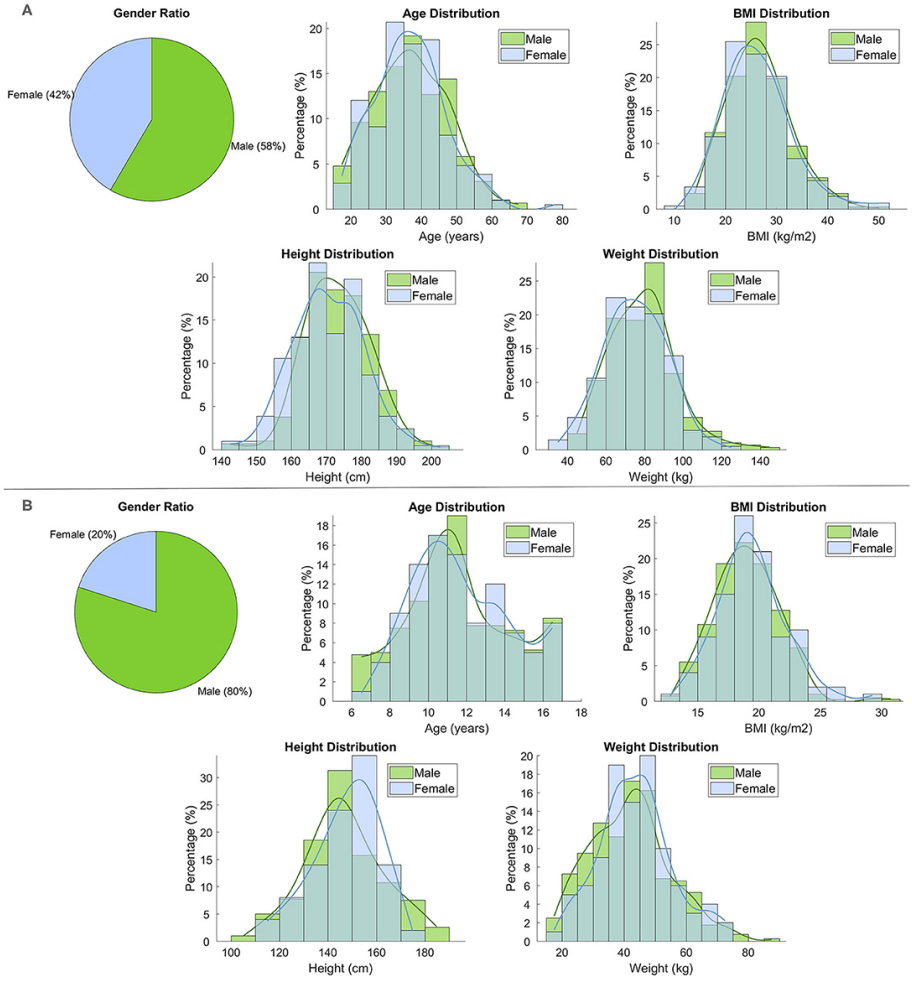

Figure 21: Demographic characteristics (sex, age, BMI, height, and weight) of (A) the adult virtual population (N = 500) and (B) the pediatric-adolescent virtual population (N = 500).

Local Sensitivity Analysis of Systems Biology-Based Models

TPMS-derived MoA models were subjected to sensitivity SOBOL analysis to evaluate whether the variation of molecular parameters would affect the models' response and to identify key molecules. The sensitivity evaluation was carried out for a range of values (-1, 1) for each protein. Although these models have about 5000 parameters, less than a third of them showed a real

Table 13: Demographic characteristics of the pediatric-adolescent virtual population and the reference population.

	Virtual population ($N = 500$)	Reference population ($N = 111$)	p-value[a]
Sex (% females)	20.0	20.0	NA
Age (years)	11.11 ± 2.73	10.90 ± 2.80	0.09
Height (cm)	147 ± 15.92	NA[b]	NA[b]
Weight (kg)	42.41 ± 12.70	43.60 ± 15.10	0.08
BMI (kg/m^2)	19.1 ± 2.7	19.1 ± 3.4	0.81

Figures are mean ± standard deviation unless otherwise stated.
BMI, Body mass index; NA, not applicable.
[a]*Calculated with the unpaired two-tailed T Student's test.*
[b]*For demographic data not provided in the reference clinical trial, European mean values were used.*

impact (difference >15%) on the output, which was less notorious in MPH (max difference Ĩ7%) than in LDX (max difference ̃32%) (Supplementary Table F in Supplementary Material 2). Interestingly, from the 30 most sensitive proteins, some were shared between both mechanisms (namely, NFKB1, PRKCA, PRKCZ, TRAF6, and PRKCB).

Physiologically Based Pharmacokinetic Models

PBPK models simulating the available drug concentration in blood over time were obtained for LDX and MPH and for the two studied populations. Drug concentration models were fitted to real data resulting in similar blood drug concentration levels for a standard adult (male, 40 years old, 175 cm, 70 kg) and child (male, 8 years old, 30 kg, 130 cm) (Figure 22). PBPK model simulations complied with the observed in vivo curves, even for the case of MPH in children and adults, where approximating repeated administration of lower doses was required to model the modified-release

formulations.

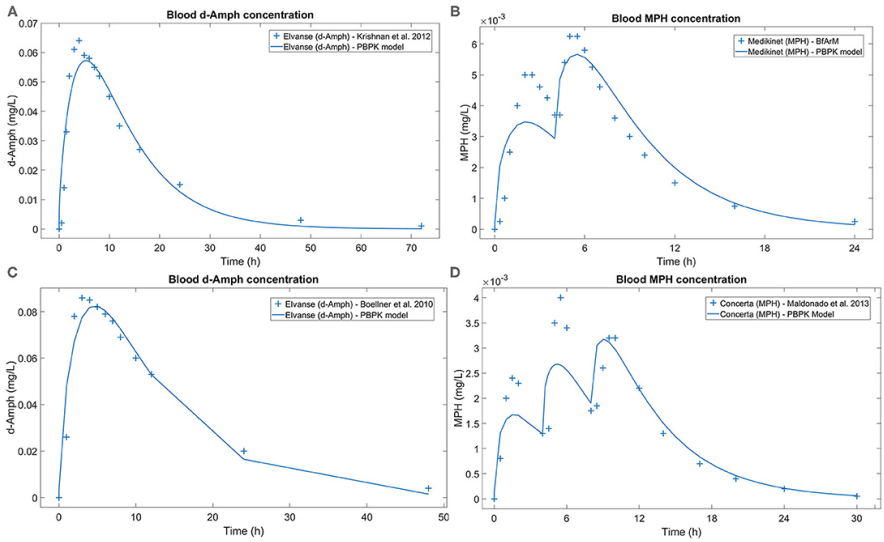

Figure 22: Blood d-Amph and MPH concentration comparison between real datapoints and the curve resulting from the PBPK model.
(A) Generated for a standard adult patient after a single 70 mg dose of LDX, real datapoints obtained from Krishnan et al. [229]; (B) generated for a standard adult patient after two 10 mg doses of MPH every 4 h, real datapoints obtained from BfArM [41]; (C) generated for a standard pediatric patient after a single 50 mg dose of LDX, real datapoints obtained from Boellner et al. [50]; and (D) model generated for a standard pediatric patient after three 5 mg doses of MPH every 4 h, real datapoints obtained from Maldonado et al. [261]. d-Amph, d-Amphetamine; LDX, Lisdexamfetamine; MPH, Methylphenidate; PBPK, Physiologically based pharmacokinetic.

Efficacy Outcomes and Measures Definition and Optimization

After the process of optimizing by trimming, 83 proteins (out of 86) were included in the pediatric-adolescent ADHD definition (ρ = -0.81) and 66 proteins (out of 86) in the adult ADHD definition (ρ = -0.79). The resulting molecular definitions found after optimizing the model-derived efficacy measures for each conditions' clinical efficacy can be found in Supplementary Table G in Supplementary Material 2. The subsequent regression lines, as well as the different study points used, are represented

in Figures 24 and 25.

Quantitative Systems Pharmacology Models in the Virtual Populations

The MoA of LDX and MPH in our populations of interest, inferred from QSP models, were obtained by combining PBPK models of the dosing schemes of these drugs and TPMS technology, which modeled the MoA of both drugs in ADHD. The simulation analyzed the whole available data on pathologies, drugs, and the population. The mean accuracy values obtained in mechanistic models for ADHD virtual patients were: 91.63% (adults treated with LDX), 91.71% (adults treated with MPH), 91.68% (children-adolescents treated with LDX), and 91.69% (children-adolescents treated with MPH). Thus, for each patient, activation/inhibition patterns of all proteins associated with the MoA of LDX and MPH were obtained. Drugs' efficacy on ADHD measured over each virtual patient was exclusively estimated using the above mentioned tSignal formula (Equation 1), which summed up the activity values of ADHD effector proteins. The tSignal formula was applied to the list of ADHD effector proteins optimized to fit clinical observations and provided high accuracy QSP models for the whole set of 1,300 patients comprising adults and children.

The ADHD population was subjected to clustering analysis to explore molecular variability within the LDX and MPH mechanistic models. The optimal number of clusters for adults was four different clusters, whereas three main clusters were identified for children, according to Hopkins statistics (0.82 and 0.89, respectively), Silhouette index (0.31 and 0.33, respectively), and Jaccard Bootstrap index (0.52 and 0.57, respectively). These results reflected drug-independent patient intrinsic variability since they clustered in a non–drug-dependent manner (Table 14). Clusters were

Table 14: Distribution of LDX and MPH mechanistic models in the generated clusters.

Drug	Cluster 1 (Red)	Cluster 2 (Green)	Cluster 3 (Blue)	Cluster 4 (Purple)
Adult models clustering				
LDX	122	123	150	105
MPH	98	154	106	142
Drug	**Cluster 1 (Red)**	**Cluster 2 (Green)**	**Cluster 3 (Blue)**	
Pediatric-adolescent models clustering				
LDX	230	179	91	
MPH	215	118	167	

LDX, Lisdexamfetamine; MPH, Methylphenidate.

represented using the two main components of PCA (Figure 23), which explained 66.7 and 12.4% of the observed variability in adults, and 61.4 and 19.4% in children-adolescents, respectively. The five most relevant proteins in the PC1 (eigenvector 1) of each population were – IL4, AKT3, NTRK2, IL5, and NTF3 for adults and – CRY1, AKT3, CRY2, AKT1, and AKT2 for children-adolescents.

We also found that clustering was associated to differences in treatment efficacy in adults (ANOVA p-value = 2.515e-08) and children-adolescents (ANOVA p-value = 1.194e-09). In adults, cluster 4 showed the highest mean tSignal (p-value = 2.263e-09), while cluster 2 was the one presenting the lowest (p-value = 6.835e-04). In children-adolescents, the tSignal of cluster 1 was significantly higher (p-value = 2.752e-05) and that of cluster 3 was significantly lower (p-value = 7.397e-04) than the rest (Student's T-test). To further characterize the clusters, we performed an ANOVA analysis to

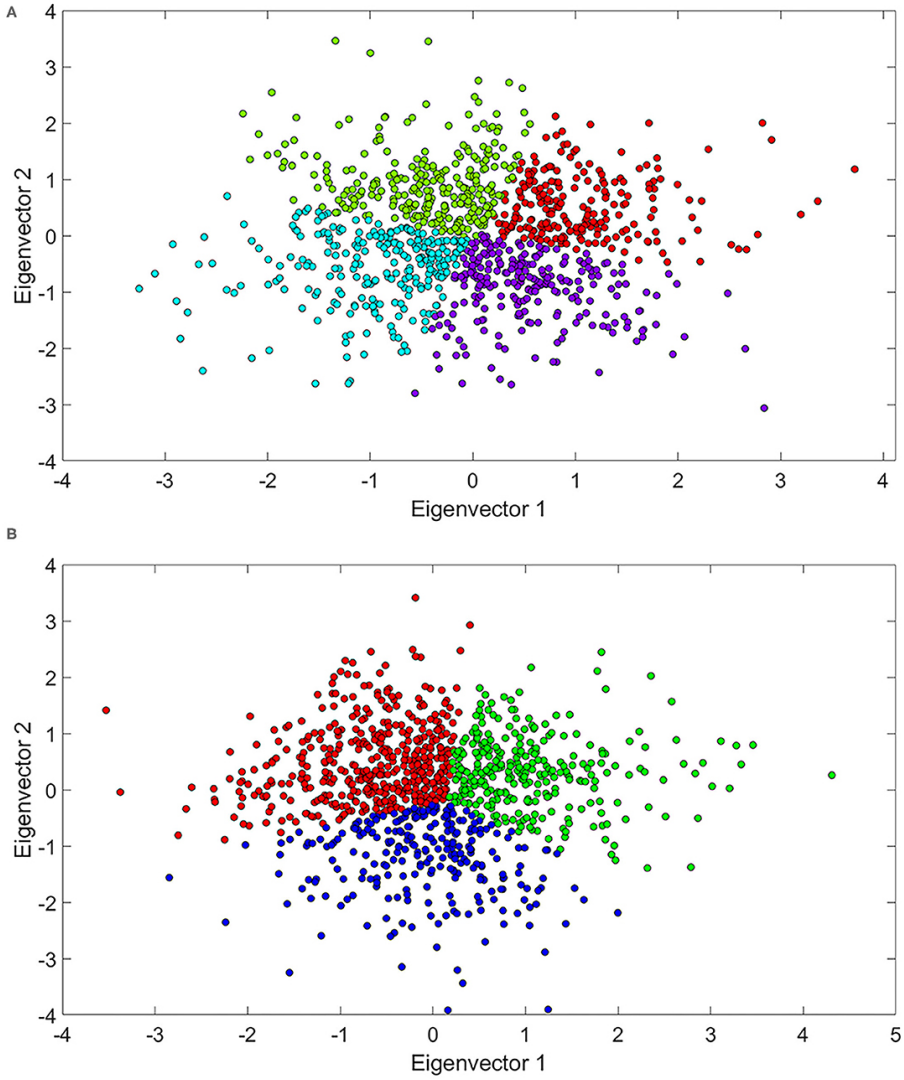

Figure 23: PCA representation, based on the modulation of ADHD effectors, of LDX and MPH mechanisms of action in (A) ADHD adult patients and (B) ADHD children-adolescent patients.
ADHD, Attention-deficit/hyperactivity disorder; LDX, Lisdexamfetamine; MPH, Methylphenidate; PCA, Principal component analysis.

identify potential differences on the demographic characteristics within the clusters. In the overall analysis, only weight was significantly different (p-value < 0.05) in adults (ANOVA p-value = 0.019), and no differences were found in children-adolescents. When comparing each cluster against

Table 15: Results of the comparison analysis between demographic characteristics within the clusters.

ID	Age, years	Height, cm	Weight, kg	BMI, kg/m^2	Gender, M:F ratio
Adults					
1	36.49 ± 9.96 (0.424)	172.19 ± 10.62 (0.797)	78.79 ± 17.67 (0.297)	26.85 ± 6.82 (0.387)	0.63 (0.141)
2	37.16 ± 10.73 (0.739)	172.02 ± 9.94 (0.979)	79.58 ± 16.95 (**0.034**)	27.02 ± 6.04 (0.131)	0.60 (0.454)
3	36.20 ± 9.75 (0.158)	171.80 ± 9.92 (0.659)	77.33 ± 15.73 (0.651)	26.53 ± 6.67 (0.953)	0.59 (0.826)
4	38.04 ± 10.73 (0.064)	172.16 ± 10.01 (0.820)	75.17 ± 17.23 (**0.006**)	25.62 ± 6.67 (**0.014**)	0.52 (**0.016**)
Children-adolescents					
1	10.95 ± 2.82 (0.095)	145.96 ± 16.13 (0.064)	41.42 ± 12.60 (**0.026**)	18.93 ± 2.69 (**0.031**)	0.81 (0.634)
2	11.22 ± 2.57 (0.431)	147.44 ± 14.53 (0.568)	42.70 ± 12.00 (0.646)	19.25 ± 2.71 (0.400)	0.76 (0.067)
3	11.27 ± 2.75 (0.283)	148.29 ± 16.96 (0.131)	43.80 ± 13.51 (**0.041**)	19.36 ± 2.68 (0.118)	0.83 (0.170)

Figures are mean ± standard deviation (and p-values).*
**T-student-test: each cluster vs. rest of clusters.*
ANOVA, Analysis of variance; BMI, Body mass index; ID, Cluster ID; M:F ratio, Male to female ratio.
p-values in bold are those considered significant ($p < 0.05$).

the rest in adults, we found that BMI, weight, and gender ratio were significantly lower in cluster 4, while weight was slightly higher in cluster 2. In children- adolescents, BMI and weight were significantly lower in cluster 1, while weight was slightly higher in cluster 3 (Table 15).

5.5.5 Discussion

Herein, the technology to create populations of virtual patients and the subsequent ISCT is described in the case-study of LDX and MPH head-to-head comparison in the context of ADHD treatment. Adult and pediatric-adolescent vPOPs were obtained, and PBPK and QSP models were generated successfully to provide the basis for identifying mechanistic differences between the two drugs, patient cohort differences, inter- and intra-patient response variability.

Preliminary evaluation of the models revealed some insights on the factors affecting MoA-related treatment efficacy. The sensitivity analysis of systems biology MoA models provided a list of common proteins that

might affect both drugs' efficacy: proteins involved in the NF-κB signaling pathway (NFKB1 and TRAF6) and PKC (alpha, beta, and zeta types). The pleiotropic nature of these proteins and their involvement in several signaling processes could explain their potential impact in the sensitivity of mechanistic models. However, a detailed evaluation of each drug mechanistic model should provide further knowledge on the key proteins involved.

QSP model clustering analysis indicated the presence of several response patterns, not clearly defined by drug treatment. The protein activity-based unsupervised clustering was somehow associated to response level, and PCA analysis revealed some relevant proteins that could be exerting this effect, including: dopamine signaling-related AKT proteins [223], neurotrophins related to neural viability and dopamine regulation in ADHD [252, 23, 45, 305], circadian rhythm proteins related to ADHD and comorbidities–associated sleep disturbances [2, 295, 90], and cytokines related to neuroinflammation and Th2 response and ADHD [241, 398, 399, 188]. Regarding clinical characteristics, a possible correlation was found between lower weight, female gender, and lower BMI, with a higher tSignal or better efficacy in adult population. Similar results were found in the children-adolescent population, were higher tSignals were found in the group with lower weight and BMI. In this sense, previous reports had already suggested a relationship between drug efficacy and BMI [112].

Related Work

Virtual populations have been generated in the past to assist in solving complex medical issues. The FDA has accepted a type 1 diabetes simulator to replace animal testing in pre-clinical trials [407]. Besides,

in silico cloning of data from individual type 1 diabetes patients to improve algorithms for closed-loop insulin delivery systems has been reported in 12 and 47 virtual patients in studies that aimed to tackle the challenging problem of inter- and intra-subject variability [155, 408]. Likewise, a virtual population of 50 individuals has been generated to test in silico drug cardiotoxicity and account for inter-subject variability in clinical studies with toxicological endpoints [315]. Such approaches were warranted considering the high variability of the evaluated pharmacokinetic parameters in a short time. However, they were limited in the number of virtual patients that could be generated accurately. Therefore, considering that ADHD is not as varying in brief periods and that such pharmacokinetic detail was unnecessary, a higher number of patients could be generated in our study. On the other hand, a multi-compartment model with a large virtual population size has been published on trauma-induced critical illness that showed how the molecular and cellular events taken as a whole could manifest heterogeneously on individuals [54]. These results were in agreement with ours, which showed different clusters of patients that could correspond to different response profiles to a certain point, independent from drug treatment.

Virtual populations combined with PBPK modeling have been used successfully to predict the pharmacokinetic profile of a drug and evaluate potential drug-drug interactions for a specific ethnicity [218]. In addition, a PBPK model combined with systems-biology techniques has been reported and validated as an efficient tool for assessing risk exposure to certain volatile organic compounds [345]. Furthermore, multi-compartment QSP has been used to model immunotherapies in breast cancer [415]. When associated with pharmacokinetics and pharmacodynamics data, it has been

reported in an in silico virtual clinical trial to analyze predictive biomarkers in certain breast cancers [416]. Hence, PBPK and QSP models have been established as powerful computational tools for in silico simulations.

Finally, only a few in silico head-to-head trials have been published. A recent study compared two insulin therapies for type 1 diabetes treatment by using the abovementioned FDA-approved simulator and pharmacokinetics models to compare two designs, crossover and parallel [352]. The parallel design was justified because it would likely be preferred in a real setting for practical reasons, which is not necessarily true in the case of our study on ADHD. Another head-to-head mechanistic study comparing two lung cancer treatments has been reported by our group, whereby a similar approach to the one here described was undertaken [64]. However, our previous study did not require generating virtual populations nor used PBPK or QSP models to reach its conclusions.

These examples of application of in silico modeling approaches in different therapeutic areas bear witness to an increasing tendency to use newly available high performance computing technologies in the field of biomedicine. The use of these technologies will help advancing toward the implementation of precision medicine pipelines and personalizing the healthcare provided to patients.

Strengths and Limitations

TPMS models are constructed considering the whole human protein network and a wide range of drug-pathology relationships, not only limited to ADHD or psychiatric ailments, which, in part, attenuates the potential bias on information regarding drugs or disease of interest. As defined by Jorba et al. [198], only MoA models with accuracies above 85% against

the training set were used to ensure good quality and general extrapolation of results. This systems biology-based methodology has been reported to be successful, with results validated by in vitro and/or in vivo models [339, 182, 257].

Only limited by computational power, ISCT allows enrolling a large number of patients with several neurophysiological ADHD subtypes, which can be difficult, costly, and even not feasible in a conventional clinical trial setting. Virtual patients generated in our study were defined by the drugs' molecular mechanisms, allowing the exploration of the complete clinical and molecular landscape of each patient. Furthermore, our ISCT design had a large enough sample size and considered pools of mathematical solutions for each patient—instead of a single mathematical model per patient—which ensured that the simulations were robust and appropriate for data analysis.

However, our study presented some limitations. Firstly, our models depended on the current knowledge of human physiology, particularly on the drugs and disease under investigation, as well as protein interactions and pathways described and involved in the MoA. Therefore, our models could have been susceptible to missing data, errors, and bias, and some aspects could have been overlooked. For instance, unknown targets or yet undescribed pathophysiological ADHD processes might play a role in the MoA of the evaluated drugs. ADHD and its associated comorbid psychiatric disorders present a high genetic and signaling overlap [111, 110], which could act as confounding factors at the clinical and molecular levels. Accordingly, the molecular characterizations used for modeling could be biased; prospective data could expand the knowledge on these diseases and, therefore, improve our model-derived conclusions.

Secondly, our approach considered only the impact of demographic characteristics on the PBPK modeling (i.e., drugs' absorption, distribution, metabolism, and excretion). However, other consequences of these characteristics at the ADHD pathophysiology level were not considered, because of the absence of (i) clear molecular information to include in the ADHD definitions for each patient profile, and (ii) reliable sources of information to properly model these characteristics at the molecular level. This limitation could prevent the modeling and detection of relevant results regarding these characteristics, such as age-related neurodevelopment [370, 94], differences between children, adolescent and adults [441, 128] or the potential role of sex-dependent differences [322, 256, 126]. Future data on large sets of patients, or specific research on the impact of those characteristics on ADHD, might allow to improve our models and derived conclusions.

Thirdly, all mathematical models are subjected to the limitation of not being able to fit 100% the training data information. In our approach, while we obtained a pediatric virtual population with demographic values non-significantly different from the reference clinical trial population, the obtained p-values for age and weight were close to the significance threshold. These parameters proved to be more difficult to fit in pediatric than in adult virtual populations. While clinical trials only report average weight and age, general pediatric population weight distributions obtained from growth information [258] are age-dependent. Accordingly, setting a higher threshold of significance ($p > 0.1$ or even $p > 0.2$) during the randomization procedure might ensure obtaining a fitter population, especially regarding the pediatric case. In this specific scenario, as the case-study objective for the generation of the ISCT was a head-to-head

between LDX and MPH using the exact same population, this bias was not expected to significantly affect the results. TPMS-based models are not an exception either [198]. Each virtual patient was constructed with at least 50 solutions, and a population sample size larger than the minimum calculated was used to dampen this effect. TPMS models present an inherent variability, rendering them useful to explore molecular variability within human physiology [198]; through an adequate management of the model's variability and considering an appropriate sample size, the best solutions could be obtained.

Finally, our study's primary outcome was generated with information from literature on the drugs used for ADHD treatment and their measured clinical effect. The values used for the training process were the average values reported in those publications (Table 11). However, a great dispersion was observed. For instance, while the mean ADHD-RS IV value associated with amphetamine was -18.1, the authors report a range of response between -4.68 and -31.52, representing a 74% deviation from the mean, clearly higher than the dispersion values generated with our models (Figure 24). The dispersion identified in clinical trials was probably due to demographic and metabolic differences between patient cohorts and how the principal variable was measured. This effect appeared in all analyzed drugs, and we estimated an average dispersion of 57% for all of them. The dispersion in the efficacy measured from the clinical trials cannot be mathematically treated without accessing patient data, which is not available; at this point, we had to resort to a naïve pooled approach, risking its associated limitations. In such cases, the best approach to obtain a drug efficacy value is to compute the mean of the values reported by different authors. Selection bias can also induce errors when using external data. To attenuate its

effect, clinical trials assessing a wide range of drugs were used in our study. On the other hand, another limitation associated to the outcome measure used would be of clinical nature; ADHD symptom scales are based on questionnaires to the patient or the physician, that comprise several aspects of a complex psychiatric disease. These clinical measures might not be as directly associated to molecular or biologically measurable factors (such as blood pressure when studying hypertension). To minimize both technical and clinical limitations of the outcome measure used, we selected the ADHD-RS IV scale as this was the scale with the largest amount of clinical trial information for different mechanisms of action, so the model efficacy measures could be properly optimized to fit clinical data. Our approach tried to compile the largest amount of available information around patients, disease, and treatments at the molecular and clinical level and provided benchmarks to validate the different steps of the study. Nonetheless, a corroboration of the herein described procedure to infer new actual clinical results with independent (existent or new) experiments is called for.

In our ISCT, most of these sources of error could translate into an error in the estimation of the principal variable, evaluated by the Pearson correlation coefficient (Figures 24 and 25). Interestingly, the Pearson correlation coefficients obtained after lineal regression adjustment for the adult and pediatric-adolescent populations were high given the large dispersion shown in values from clinical practice for the same drugs (Supplementary Table E in Supplementary Material 2).

5.5.6 Conclusions

The methods here illustrated described the step-by-step process for creating a virtual population of patients treated with two drugs for

ADHD management, LDX and MPH, with the aim of designing an ISCT for their comparison head-to-head. Our study provided adult and pediatric-adolescent vPOPs and generated QSP models to infer, after analysis, the MoA of these two drugs. This theoretical model, and its use for a head-to-head analysis, would allow obtaining conclusions classified as MEDIUM impact according to MID3 guidelines. Although experimental and clinical assays are warranted to validate or refute these potential results before translation into clinical practice, the mechanistic-driven modeling techniques used here should be accepted as hypothesis-generation solid tools with a remarkable ability to provide molecular detail. Besides, from a scientific evidence point of view, complementing meta-analyses with theoretical models, such as the ones here presented, can palliate the lack of costly, though necessary, head-to-head clinical trials. Altogether, in silico techniques can contribute to advancing the understanding of diseases' pathophysiology and the molecular MoA of available therapies, with the ultimate goal of reaching personalized medicine.

5.5.7 Data Availability Statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

5.5.8 Author Contributions

JRG-C, JQ, VM, TP-R, CM, and JM conceived the study. JRG-C, JQ, GJ, and VJ performed the investigation and drafted the first version of the manuscript. GJ, VJ, BO, XD, and JM developed the methodology of the study. GJ, VJ, and JM undertook software related tasks. VM, TP-R, and CM managed the project and JRG-C, JQ, BO, XD, and JM supervised and

validated it. GJ and VJ aided in visualization tasks. All authors reviewed and edited its final version.

5.5.9 Funding

This study was funded by Takeda. Public funders provided support for some of the authors' salaries: GJ has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 765912. VJ is part of a project (COSMIC; www.cosmic-h2020.eu) that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 765158.

5.5.10 Conflict of Interest

JRG-C has served as speaker for Takeda and Shire and has received research funding from Shire. JQ has served as speaker and/or on scientific advisory boards for Takeda, Janssen, and Rubio. GJ, VJ, and JM are full-time employees at Anaxomics Biotech. VM, TP-R, and CM are full-time employees at Takeda. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

5.5.11 Publisher's Note

5.5.12 Acknowledgments

The authors would like to thank Juan Manuel García Illarramendi (Anaxomics Biotech, Barcelona) for his assistance on the RNAseq data to protein transformation for sample size calculation; Mireia Coma, Cristina Segú-Vergés, and Helena Bartra (Anaxomics Biotech, Barcelona) for compiling data; Judith Farrés (Anaxomics Biotech, Barcelona) for supervising the methodology; and Matías Rey-Carrizo (BCN Medical Writing) for providing editorial support.

5.5.13 Supplementary Material

The Supplementary Material for this article can be found online at:
https://www.frontiersin.org/articles/10.3389/fpsyt.2021.741170/full#supplementary-material

5.5.14 Supplementary Methods

Bibliography-based characterization and searches

Bibliography-based drug characterization

Aside from a review of official regulatory documentation and drug-target dedicated databases, a review of the currently available bibliography regarding known targets of the drugs was performed in PubMed on April 27, 2020. The specific searches performed were the following:

- ("Elvanse" [Title/abstract] OR "Vyvanse" [Title/abstract] OR "Lisdexamfetamine" [Title/Abstract]) AND ("molecular" [Title/Abstract] OR "mechanism" [Title/Abstract] OR "pathophysiology" [Title/Abstract] OR "pathogenesis" [Title/Abstract] OR "mode" [Title/Abstract] OR "action"

[Title/Abstract] OR "signaling" [Title/Abstract] OR "signalling "[Title/Abstract] OR "expression" [Title/Abstract] OR "activation" [Title/Abstract] OR "inhibition" [Title/Abstract] OR "activity" [Title/Abstract])

- ("Methylphenidate" [Title] OR "Medikinet" [title/abstract] OR "Concerta" [title/abstract] OR “Medikinet”[title/abstract]) AND ("molecular" [Title] OR "mechanism" [Title] OR "pathophysiology" [Title] OR "pathogenesis" [Title] OR "action" [Title] OR "signaling" [Title] OR "signalling " [Title])

All articles were analyzed at the title and abstract level. The presence of molecular information was reviewed in depth to identify protein/gene candidates to be considered drug target candidates.

Bibliography-based conditions characterization

For disease characterization, we initiated an extensive and careful full-length review of relevant articles in the PubMed database (up to January 21, 2020) that included the following search strings:

- ADHD: ("Attention deficit hyperactivity disorder" [Title] OR "ADHD" [Title] OR "Attention-Deficit/Hyperactivity Disorder" [Title]) AND (“pathogenesis” [Title/Abstract] OR “pathophysiology” [Title/Abstract] OR “molecular” [Title/Abstract]) AND Review [ptyp].

- Depression: ("Depression" [Title] OR "Major Depressive Disorder" [Title]) AND ("Molecular" [Title/Abstract] AND ("Pathophysiology" [Title/Abstract]) OR ("Pathogenesis" [Title/Abstract]) AND (Review[ptyp] AND "2015/01/28" [PDat] : "2020/01/28" [PDat])

- Anxiety: (“Anxiety disorders” [Title] OR "Anxiety" [Title]) AND ("Pathogenesis" [Title] OR "Pathophysiology" [Title] OR "Molecular"[Title])
- Bipolar Disorder: ("Manic-Depressive" [Title] OR "Bipolar" [Title] OR "Manic Depressive" [Title] OR "Manic Disorder" [Title] OR "Manic Depression" [Title]) AND (“Pathogenesis” [Title/Abstract] OR “Pathophysiology” [Title/Abstract] OR “Molecular” [Title/Abstract]) AND Review[ptyp]
- Tics Disorder: (“TICS” [Title] OR "Tic Disorder" [Title] OR "Tic Disorders" [Title]) AND ("Pathogenesis" [Title/Abstract] OR "Pathophysiology" [Title/Abstract] OR "Molecular" [Title/Abstract])
- Binge eating: ("Binge Eating Disorder" [Title] OR "Binge Eating" [Title]) AND ("Pathogenesis" [Title/Abstract] OR "Pathophysiology" [Title/Abstract] OR "Molecular" [Title/Abstract])

The list of publications identified in the specific searches was retrieved and assessed at the title and abstract level. If molecular information describing pathophysiology conditions was found, the full texts were thoroughly reviewed to identify the main pathophysiological processes known to be involved in the diseases (Table A in the S2 File). Subsequently, each pathophysiological process was further characterized at the protein level by using the retrieved publications. Accordingly, proteins whose activity (or lack thereof) are functionally associated with the development of the condition were identified (Table B in the S2 File).

Clinical trial information compilation

To accurately obtain a model-derived efficacy value to fit clinical efficacy values, clinical trials that assess the efficacy of drugs currently approved and

commonly used in clinical practice were retrieved from clinicaltrials.gov and PubMed. Only phase III clinical trials that were interventional (i.e. included at least an arm treated with the drug of interest), completed, and had published results were considered. Once listed, these clinical trials were evaluated to select the efficacy scale most frequently used and the best representative of drug variability (i.e. that included data from a vast number of drugs). Accordingly, the ADHD Rating Scale IV (ADHDRS IV) and the following list of drugs were considered: Amphetamine, Atomoxetine, Bupropion, Clonidine, Dexmethylphenidate, Guanfacine, Lisdexamfetamine, and Methylphenidate. The selected drugs were characterized at the molecular target and pharmacokinetic levels (Table D in the S2 File).

Modelling methodology and algorithms

Virtual population modelling

Adult virtual populations (vPOPs) are created by assigning demographic variables (age, weight, height, and body mass index [BMI]) to virtually generated patients by using a modification of the algorithm proposed by Allen et al. [11]. Accordingly, we define as original population the one whose characteristics we would like to mimic and as standard population the reference use to complete missing demographic characteristics. Firstly, a multivariate normal distribution (MVND) with the given means and standard deviations for each variable from the original population is created. When data about a specific parameter is not available, the information is taken from the standard population distribution (in this study, for the adult population, a standard European distribution [106] was used). Since BMI, weight, and height are related through equation 1, only age and one pair of the morphometric parameters (weight and BMI, weight and height, or BMI

and height) are generated.

$$BMI(kg/m^2) = weight(kg)/height(m)^2$$

Secondly, a cost function based on the original population demographic parameters is used with the objective of being minimized until the generated population resembles the available information on the original population. A simulated annealing strategy is used to minimize the cost function by using as starting points the patients generated according to MVND values.

Let n be the number of patients to generate; μ_i, σ_i, m_i, M_i the mean, standard deviations, minimums, and maximums of the original population's age, height, weight, and BMI (i); X_i the n dimensional vector containing the variable i's values of the generated population; and $X_{i,j}$ the $nx2$ matrix containing the generated population's values of the variables i and j so that it represents the concatenation of X_i with X_j. For any data vector X generated with the multivariate distribution, let mean(X), std(X), min(X) and max(X) be the mean, standard deviation, minimum, and maximum of X. Two different cost functions are defined: the first one contains only age as a single item, and the second equation is based on equation 1, relating two of the morphometric parameters (BMI, weight, and height).

For age (a):

$$f_a(X_a) = (mean(X_a) - \mu_a)^6 + (std(X_a) - \sigma_a)^4 + \sum_{x_a \in X_a} max\left(\left(x_a - \frac{m_a+M_a}{2}\right)^2 - \left(\frac{m_a+M_a}{2}\right)^2, 0\right)$$

For a pair of the morphometric parameters (BMI, weight, and height) (i,j):

$$f1_{i,j}(X_{i,j}) = (mean(X_i) - \mu_i)^6 + (std(X_i) - \sigma_i)^4 + (min(X_i) - m_i)^2 + (max(X_i) - M_i)^2 + (mean(X_j) - \mu_j)^6 + (std(X_j) - \sigma_j)^4 + (min(X_j) - m_j)^2 + (max(X_j) - M_j)^2$$

Since i and j are generated independently from the remaining demographic variable (either BMI, weight, or height), and to ensure the latter stays in a plausible range, the cost function was extended by using equation 3, which also depends on k:

i = weight; j = BMI; k = height

$$f2.1_{i,j,k}(X_{i,j}) = \left(mean(\sqrt{\frac{X_i}{X_j}}) - \mu_k\right)^6 + \left(std(\sqrt{\frac{X_i}{Xj}}) - \sigma_k\right)^4 + \sum_{(x_i,x_j)\in X_{i,j}} \left(max\left(\left(\sqrt{\frac{x_i}{x_j}} - \frac{m_k+M_k}{2}\right)2 - \left(\frac{M_k+m_k}{2}\right)^2, 0\right)\right)$$

i = height; j = BMI; k = weight

$$f2.2_{i,j,k}(X_{i,j}) = \left(mean(X_i^2 \cdot X_j) - \mu_k\right)^6 + \left(std(X_i^2 \cdot X_j) - \sigma_k\right)^4 + \sum_{(x_i,x_j)\in X_{i,j}} \left(max\left(\left(x_i^2 \cdot x_j - \frac{m_k+M_k}{2}\right)2 - \left(\frac{M_k+m_k}{2}\right)^2, 0\right)\right)$$

i = height; j = weight; k = BMI

$$f2.1_{i,j,k}(X_{i,j}) = \left(mean(\frac{X_i}{X_j^2}) - \mu_k\right)^6 + \left(std(\frac{X_i}{X_j^2}) - \sigma_k\right)^4 + \sum_{(x_i,x_j)\in X_{i,j}} \left(max\left(\left(fracx_ix_j^2 - \frac{m_k+M_k}{2}\right)2 - \left(\frac{M_k+m_k}{2}\right)^2, 0\right)\right)$$

Where $\frac{X_i}{X_j}$ corresponds to the pointwise division, i.e. a vector where the lth element is the lth element of X_i divided by the lth element of X_j. Similarly, $X_i \cdot X_j$ and $X^2 = X_i \cdot X_i$ are the pointwise multiplications. The final cost function equation for BMI/weight/height results as follows:

$$f_{i,j,k}(X_{i,j}) = f1_{i,j}(X_{i,j}) + f2_{i,j,k}(X_{i,j})$$

A one-sample z-test (alpha= 0.05) is also used to ensure the new population preserves the original data distributions; otherwise, the population is recalculated.

In pediatric-adolescent virtual populations, and because morphometric measures drastically depend on age, a slightly different approach was undertaken. The first step was generating a sample population using the percentiles information reported by the World Health Organization

in pediatric-adolescent populations [258] and by randomly generating plausible values setting the percentiles as probability density function points [156]. Subsequently, a resampling strategy based on the Metropolis-Hastings method [71] was used to select a sub-population fulfilling the means and standard deviations of the original population. An initial sub-population set was chosen randomly. Then, an iterative process was carried out by continuously replacing patients from the sub-sample with the ones in the generated sample population. If the newly replaced patient resulted in a sub-population closer to the original target population, that population was chosen. Otherwise, the old patient remained in the population. This process continued until a p-value > 0.05 was obtained, according to a one sample z-test between distribution values of the subpopulation and the original population.

Quantitative Systems Pharmacology models – Physiologically-based pharmacokinetic data integration in Therapeutic Performance Mapping System models

The training data used by Therapeutic Performance Mapping System (TPMS) models (Table C in S2 File) consists of physiologically known stimulus-response relationships (e.g. drugindication) that must be achieved [198]. Therefore, to integrate the patient-specific physiologicallybased pharmacokinetic (PBPK) concentration data with TPMS resulting in the final Quantitative Systems Pharmacology (QSP) models, the individuals' drug concentration variation data need to (i) be transformed into a protein activity-like measure (stimulus); and (ii) be associated to a molecular effect (response). To that end, drug concentration is translated into drug target protein's grade of activation/inhibition and, in parallel, is related to the pathology-ADHD in this case - inhibition or reduction by using a set

of equations based on the half-maximal response concentration (EC50) definition and clinical efficacy data. To calculate the stimulus, the following procedure and equations are used. PBPK models describe the variation of drug concentration in the different compartments over time. Given a specific compartment, the drug concentration (C) variation over time can be expressed as a vector of drug concentration values for i timepoints (C_i). The drug concentration is then related to its protein targets (Table 9). According to the TPMS definition, the activity of a protein P can be treated as a normalized vector with values in the range [-1,1], where 1 represents the maximal functional capability of the component to develop its activation functional role, -1 corresponds to the maximum inhibition capacity, and 0 represents the null capability of developing tasks. For each setting, and assuming that the maximal absolute value of P will be obtained when drug concentration is maximal (max(C)), we define Sig as the protein sign (+1 when drug activates P, -1 for inhibition), and a vector of protein target activity over time i as:

$P_i = C_i/max(C) * Sig$

Response values are also calculated by using drug concentration over time (C_i) but refer to the drug effect over the disease. The transformation into pathology response values can be obtained applying the concept of $EC50$, according to equation 9 [414], being Eff the effect of the drug over the pathology, which will range from 0 to 1:

$Eff = \frac{C}{EC50+C}$

As real drug's $EC50$ were not available, $EC50'$ was here defined as a model-derived proxy related to (clinical) efficacy or drug effect. Also, the resulting models' tSignals were used as a modelderived measure of the drug's impact on the pathology (Eff' i.e. the model equivalent to the

parameter Eff). Then, to pre-calculate the $EC50'$ and be able to obtain the Eff', the theoretical mechanism of action model between the drugs under study and the molecular descriptors of the pathology were built, as described by Jorba et al. [198], and the resulting tSignals extracted. The latter was assumed as a drug's maximal effect and, by applying equation 9, $EC50'$ could be defined as a function of the Eff' when the maximal drug concentration was achieved (max(C)):

$$EC50' = \left(\frac{max(C)}{Eff'} - max(C)\right)$$

To render $EC50'$ an estimate of clinical efficacy, it was weighted taking into account the real clinical efficacies ($clEff$) of the whole set of drugs to be considered in the study; thus, $EC50'$ was a parameter relative to the set of drug efficacies included in the analysis. For each drug, let $clEff$ be the clinical efficacy value; max($clEff$) the maximum clinical efficacy found from the whole set of drugs considered; and Eff' the tSignal extracted from the TPMS model. Then:

$$EC50' = \frac{\left(\frac{max(C)}{Eff'} - max(C)\right)}{\frac{clEff}{max(clEff)}}$$

In this study, the set of drugs used for $EC50'$ calculation was the same set of drugs that were used for the intervention outcome optimization and are summarized in Table D in the S2 File. Finally, the response values of each drug could be computed by using the corresponding $EC50'$, and rewriting equation 9 as shown in equation 12.

$$Eff'_i = \frac{C_i}{EC50' + C_i}$$

As a result, a set of stimulus(P_i)-response(Eff'_i) vector pairs could be computed for each of the drugs, one pair per each drug's protein target, and were added to the training set to construct the patient-specific QSP models.

Efficacy outcome – Clinical efficacy measure

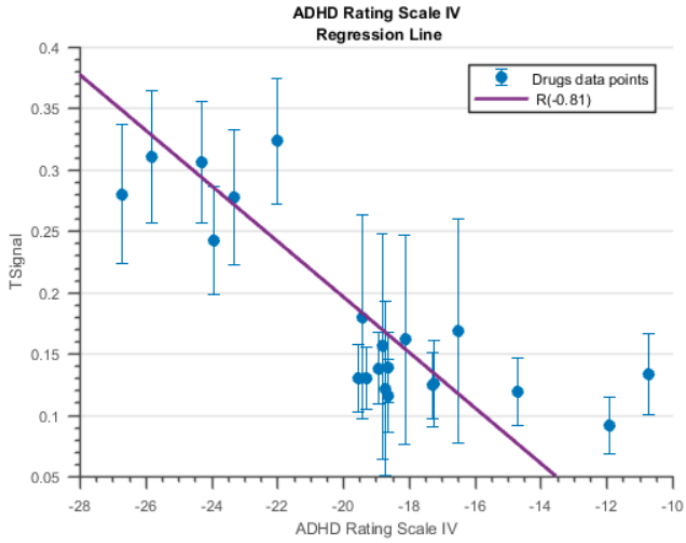

Figure 24: Regression line between optimized ADHD tSignal in the pediatric-adolescent population in relation to ADHD-RS IV, change from baseline values (R = -0.81)

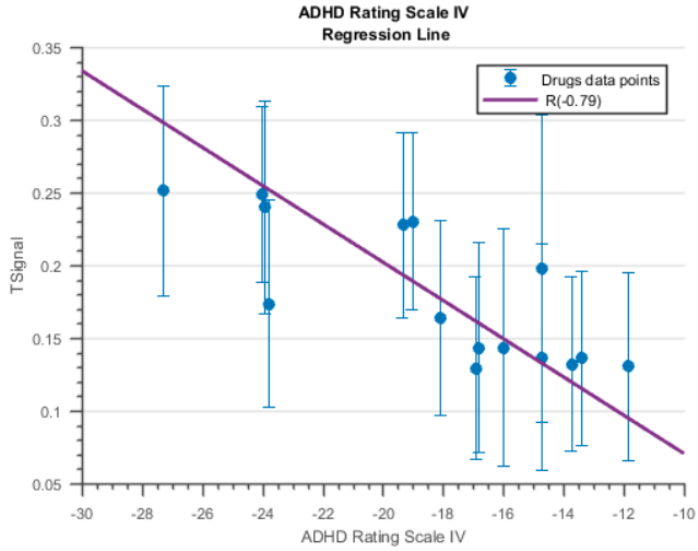

Figure 25: Regression line between optimized ADHD tSignal in adult patients in relation to ADHD-RS IV, change from baseline values (R = -0.79)

ADHD was characterized and the tSginal of the subsequent protein set was chosen as modelderived efficacy measure. Then, the TPMS-based MoA

models of the selected drug's clinical trials were built (summary in Table E in the S2 File), and the ADHD-tSignals were computed and used to measure optimization. In order to link the clinical efficacy measure, ADHD-RS IV, with the model-derived value, ADHD-tSignal, linear regression analysis between both variables was performed to parameterize the following equation:

$$Clinicalefficacymeasure = A * model - derivedefficacy + B$$

The ADHD-tSignal strongly depends on the disease's molecular characterization. However, the initial bibliography-based definition could lead to the inclusion of proteins not related to the drugs under study or that might not have a clear role in the clinical manifestations affecting the clinical scale. Accordingly, the optimization process was centered by determining the molecular definition (ADHD protein subset) of the pathological condition whose tSignal would best correlate to clinically observed efficacies (ADHD-RS IV). This process was designed to maximize the absolute value of the Pearson correlation coefficient (|ρ|) between clinical and tSignal values, maintaining molecular information from the bibliography-based characterization. Thus, to identify the best A and B parameters in equation 13 that linked clinical efficacy measures with tSignals, proteins within the ADHD molecular definition were discarded iteratively until a molecular definition provided a strong correlation between both variables [7].

5.6 Discussion

In this methods publication, the methodology to create populations of virtual patients, with their QSP models, and the implementation of the subsequent ISCT is described. The article is centered on

the case-study of lisdexamfetamine (LDX) and methylphenidate (MPH) head-to-head comparison in the context of attention-deficit/hyperactivity disorder (ADHD) treatment. The models and platform herein proposed could be used as a hypothesis-generation tool with a remarkable ability to provide molecular detail, although experimental and clinical assays will be needed to validate the potential results before translation into clinical practice. Nevertheless, from a scientific evidence point of view, complementing meta-analyses with theoretical models, such as the platform here presented, can palliate the lack of costly, though necessary, head-to-head clinical trials. Moreover, they can provide critical information on sub-cohort responses or adverse event prediction.

5.6.1 Virtual populations

Virtual populations (VPops) have been already used to assist in solving complex medical issues, while allowing the analysis of patient variability, and replacing animal testing in pre-clinical trials [407, 155, 408]. Most of these approaches took into account the high variability of the evaluated pharmacokinetic parameters in a short time. However, they were limited in the number of virtual patients that could be generated accurately.

The algorithm herein proposed uses multivariate normal distributions (MVND) and, because of the intrinsic different characteristics and distributions of child and adult populations, two solving approaches are described based on sub-sampling and simulated annealing respectively. They both allow the rapid generation of VPops with real-like physiological characteristic distributions.

5.6.2 PBPK modelling

Most PBPK models are drug- and/or disease-specific, which limit their scalability to other areas. However, here a general PBPK model is described representing a whole-body system and able to simulate the behavior of any given drug. Although it does not include metabolism or interactions with other body components, it can model the basic behavior of a drug in all main body compartments thanks to its data-drive approach. Moreover, as in the case-study example, late-release compound could be also modeled resulting in real-like concentration curves.

5.6.3 QSP modeling

Virtual populations combined with PBPK modeling have been used to successfully predict the pharmacokinetic profile of a drug and evaluate potential drug-drug interactions for a specific ethnicity [218]. In addition, a PBPK model combined with systems-biology techniques has been reported and validated as an efficient tool for assessing risk exposure to certain volatile organic compounds [345]. Furthermore, multi-compartment QSP approach has been used to model immunotherapies in breast cancer [415]. When associated with pharmacokinetics and pharmacodynamics data, it has been reported in an in silico virtual clinical trial to analyze predictive biomarkers in certain breast cancers [416]. Hence, PBPK and QSP models have been established as powerful computational tools for in silico simulations.

In this ISCT platform PBPK models are connected to TPMS, enabling the generation of QSP models partly dependent on treatment schedules and body characteristics. In order to link both model types, one focused on compound concentration and the other on protein interactions, a system

is proposed were the drug concentrations are linked to target activation and disease treatment efficacy which, thanks to the protein characterization of both, can then be translated into protein activations/inhibitions. This information is later taken by the TPMS-systems biology models as input data.

5.6.4 Application of the ISCT platform

Only a few in silico head-to-head trials have been published. A recent study compared two insulin therapies for type 1 diabetes treatment by using the above mentioned FDA-approved simulator and pharmacokinetics models to compare two designs, crossover and parallel [352]. In publication 1 (section 3.2), another head-to-head mechanistic study was done using TPMS models, comparing two lung cancer treatments. Although using similar models, our previous study did not require the generation of virtual populations nor used PBPK or QSP models to reach its conclusions.

Here, a more extensive application of TPMS models was performed. As a result of the study, we found different clusters of patients that could correspond to different response profiles to a certain point, independent from drug treatment. These results were in agreement with previous findings, where a multi-compartment model with a large virtual population size was published on trauma-induced critical illness, and which showed that molecular and cellular events taken as a whole could be manifested heterogeneously on individuals [54].

5.7 Concluding remarks

Recent application of in silico modeling approaches in different therapeutic areas bear witness to an increasing tendency to use newly available high

performance computing technologies in the field of biomedicine. The use of these technologies will help in drug development decision making, while advancing toward the implementation of precision medicine pipelines and personalizing the healthcare provided to patients.

The herein generated theoretical models inside a ISCT platform, and its use for a head-to-head analysis, would allow obtaining conclusions classified as MEDIUM impact according to MID3 guidelines [263, 409].

Part IV

CONCLUSIONS

Chapter 6

General Discussion

In this thesis, a novel methodology for the implementation and execution of ISCTs, based on a powerful systems biology approach, is being described. ISCT is still a new and expanding field with low implementation consensus, despite recent efforts [304, 403]. While some models and approaches have been proposed through the usage of Bayesian models, ML (machine learning) algorithms or PBPK [304], here I propose the usage of a patient specific individualized QSP approach by merging systems biology with PBPK modeling, as earlier mentioned in Avicenna Roadmap [403].

As discussed in section 5.5, I have built an ISCT platform consisting on a series of semi-supervised, sequential methodologies that allow the generation of many VPops and treatment branches, which can later be analyzed and compared. After defining the treatment strategy/ies and target population (protocol), the process starts by the generation of a population, by means of defining or mirroring a reference population demographic distributions values, as well as assigning molecular/disease tags. Next, a QSP (PBPK+SysBio) model is built for each patient, taking into account

their specific physiological and molecular characteristics (age, sex, bmi, comorbidities, co-treatments) and the system is perturbed with the treatment schedule. Finally, the population is analyzed as a whole, in cohorts or individually, in order to extract new insights.

6.1 Systems biology

Systems biology approaches have been previously developed in order to elucidate the MoA of drugs [1, 302, 309, 380, 314, 181]. Here, the TPMS methodology has been used, which is based on a HPN that uses and combines data from several databases, as described in Box2 and section 3.2 and section 4.2, and previously used elsewhere [181, 182, 164, 269, 309, 339]. Aside from PPI information, the network also incorporates information regarding protein complexes, metabolism and gene-coregulation. TPMS goal is to elucidate the MoA or pathways of drugs from their target proteins until a patho-physiological related protein set, in order to extract information of their protein based effect and possible adverse events.

This type of models, which can be perturbed in order to follow its effects over the HPN, have been proven useful in answering wide range hypotheses, like: understanding pathologies, predicting adverse events, predicting drug-drug interactions, finding possible biomarkers [181, 182, 164, 269, 309, 339].

6.2 Physiologically-based pharmacokinetics

Many complex, usually disease-specific, PBPK models have been generated and some already accepted by regulatory agencies to be part of the DDD process to avoid animal testing [413, 395, 433]. Here, a rather fundamental

but generic PBPK model has been built using SimBiology Toolbox (Release 5.7) from MATLAB^{TM}[271] in order to simulate the absorption, distribution and excretion of drugs and compounds in a whole-body human system.

As discussed in section 5.5, this PBPK model is data-driven and relies on previously generated information, either actual PK concentration-time curves or drug's PK parameters (eg. Tmax, Thalf), to parametrize and simulate the drug's behavior. This comes handy when analyzing known compounds, although would not serve for newly generated drugs with unknown PK data. However, the goal of this approach is not to predict drug properties, which other algorithms can already do, but to simulate their behavior inside the human body organs and tissues, in order to extract a distributions and concentration curve patterns that will be used afterwards.

6.3 QSP model

By combining the drug concentration profile from the PBPK model and the TPMS technology, a novel QSP model has been generated and described in section 5.5. Although the model largely relies on the systems biology network properties, the addition of PK information as extra training data allows the differentiation between distinct dosing profiles, as well as comparison of different PBPK systems, corresponding to different individuals.

It is important to notice that a specific strategy has been followed in order to integrate both models' data. As explained in the article, previous CT results are used in order to transform the concentration values into efficacies, which are then used as signal values. This results in models that that can be used to extract information on specific patient cohorts, as well as comparing

different drug profiles. However, as a data-driven approach, it relies on previous PK and CTs data in order to function optimally, and leaves few margins for novel drug data prediction.

6.4 Virtual populations

In order to move towards an individualized, personalized medicine, which is only feasible thanks to recent computer capabilities, the usage of virtual patients (VPs) and populations (VPops) is increasing [11, 354, 240]. Following this premise, the first and important step in the configuration of the present ISCT platform is the generation of a user defined reference VPop, for assigning both demographic and molecular characteristics to virtually generated patients; as realistic populations will enable better cohort understanding in posterior analyses.

Two distinct methodologies for generating virtual populations, one for adults and another for child-adolescents, have been implemented based on previous works [11, 71]. Both have proven valid for creating Vpops that mimic reference population demographic distributions, with no statistical difference. Because of the core of the algorithms used, no extreme patients are generated, with eg. 150 years old or 250 cm tall. The whole demographic distribution range and initial patient molecular characteristics assignation can be simulated, which allow the generation of specific, rare cohorts not or little represented in actual CTs.

In order to choose an appropriate population sample size, a tool has also been developed to compute the statistical power of generating different cohort sizes, as described in subsection 5.5.3.

6.5 ISCT platform

As depicted in section 5.5, a platform for performing ISCT has been generated which includes the algorithms from sample size determination and generation of VPs, until the QSP modeling and posterior analysis. Although it needs data form previous studies, this tool can be used to compare a different treatments of a drug and could prove paricularly useful for drug repurposing studies. So, this tool it is intended to help in decision making when moving through CT stages in Drug Development processes, as it has been exemplified in the paper of section 5.5.

6.6 Advantages and limitations

The resulting QSP models rely on the TPMS technology, which has been proven to generate robust models. Those can be used not only for MoA extraction, but also for AEs prediction, analysis of drug-drug interaction or biomarker prediction.

As a MEDIUM impact according to MID3 guidelines, this methodology could be used to justify key decision making points during DDD phases.

As a data-driven approach, it needs previous PBPK experiments and CTs information in order to build the models and make optimal further predictions. This would be an inconvenient for newly developed drugs, but not for repurposing trials, where lot of information has already been generated.

Although tissue/organ is taken into account at modeling drug concentration in PBPK models, drug-tissue/organ interactions are not taken into account and could result in some bias.

Because the HPN used depends on the current knowledge of human physiology, as well as protein interactions and pathways described and involved in the MoA, the resulting models could be susceptible to missing data, errors, and bias.

Chapter 7

Conclusions

In the present thesis, I have used both systems biology and a novel in silico clinical trial platform in order to gain insights in the mechanisms of actions of drugs, as well as extracting relevant information in the several disease contexts. Along with this thesis, I have come to the following conclusions:

- The Therapeutic Mapping System is a powerful and robust systems biology modeling approach which enables the computation of the mechanism of action of drug-pathology relations (chapter 3 and chapter 4).
- The computational models, using the TPMS, of britgatinib and alectinib towards ALK+ NSCLC, display similar effectivity. However, a broader response was identified for brigatinib which might infer better benefits, especially towards preventing resistance mechanisms (chapter 3).
- Artificial neural networks can be used to predict drug-drug interactions and resistances and, jointly with TPMS, revealed that

while alectinib might be more susceptible towards developing resistances, it would also present less predicted interactions with other treatments (chapter 3).

- Because TPMS models use a sampling method approach, generating several possible mechanism of action solutions, those can be interpreted as different patients and analyzed as such to extract cohort insights. This permitted the identification of 30 promising biomarkers related to heart failure diseases and macular degeneration (chapter 4).

- TPMS technology can be coupled to a physiologically-based modeling approach in order to generate personalized quantitative systems pharmacology (QSP) models (chapter 5).

- All-together, the developed ISCT platform allow the simulation of a clinical trial. As so, it can help finding insights in drug effectiveness, drug comparisons, biomarkers and adverse event identification, and extraction of patient-cohort differences (chapter 5).

List of Figures

List of Tables

Bibliography

[1] Abernethy, D. R., Woodcock, J., and Lesko, L. J. (2011). Pharmacological mechanism-based drug safety assessment and prediction. *Clinical Pharmacology & Therapeutics*, 89(6):793–797.

[2] Accardo, J., Marcus, C., Leonard, M., Shults, J., Meltzer, L., and Elia, J. (2012). Associations between psychiatric comorbidities and sleep disturbances in children with attention-deficit/hyperactivity disorder. *J Dev Behav Pediatr*, 33:97–105.

[3] Achenbach, T., Howell, C., McConaughy, S., and Stanger, C. (1998). Six-year predictors of problems in a national sample: Iv. young adult signs of disturbance. *J Am Acad Child Adolesc Psychiatry*, 37:718–27.

[4] Addeo, A., Tabbò, F., Robinson, T., Buffoni, L., and Novello, S. (2018). Precision medicine in alk rearranged nsclc: A rapidly evolving scenario. *Crit Rev Oncol Hematol*, 122:150–56.

[5] Adler, L., Faraone, S., Spencer, T., Berglund, P., Alperin, S., and Kessler, R. (2017). The structure of adult adhd. *Int J Methods Psychiatr Res*, page 26 1555.

[6] Aguirre-Plans, J., Piñero, J., Sanz, F., Furlong, L. I., Fernandez-Fuentes, N., Oliva, B., and Guney, E. (2019). GUILDify v2.0: A tool to identify molecular networks underlying human diseases, their comorbidities and their druggable targets. *Journal of Molecular Biology*, 431(13):2477–2484.

[7] Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18(3):91–93.

[8] Albert-Fort, M., Hombrebueno, JR, P.-V., S, S.-G., S, D.-L., M, P.-D., and M.D. (2014). Retinal neurodegenerative changes in the adult insulin receptor substrate-2 deficient mouse. *Exp Eye Res*, 124:1–10. pmid: 24792588.

[9] Alecensa (2019). European medicines agency (EMA). Internet. https://www.ema.europa.eu/en/medicines/human/EPAR/alecensa.

[10] Ali, R., Arshad, J., Palacio, S., and Mudad, R. (2019). Brigatinib for *ALK*-positive metastatic non-small-cell lung cancer: design, development and place in therapy. *Drug Design, Development and Therapy*, Volume 13:569–580.

[11] Allen, R., Rieger, T., and Musante, C. (2016). Efficient generation and selection of virtual populations in quantitative systems pharmacology models. *CPT: Pharmacometrics & Systems Pharmacology*, 5(3):140–146.

[12] Allerheiligen, S. R. B., Abernethy, D. R., Altman, R. B., Brouwer, K. L. R., Califano, A., David, Z., D'argenio, Iyengar, R., Jusko, W. J., Lalonde, R. L., Lauffenburger, D. A., Shoichet, B. K., Stevens, J. L., Subramaniam, S., Graaf, P. H. V. D., Vicini, P., and Ward, R. J. (2011). Quantitative and systems pharmacology in the post-genomic era : New approaches to discovering drugs and understanding therapeutic mechanisms. In Ward, R., editor, *An NIH White Paper by the QSP Workshop Group*.

[13] Alunbrig (2019). European medicines agency (EMA). Internet. https://www.ema.europa.eu/en/medicines/human/EPAR/alunbrig.

[14] Alunbrig (n.d.). https://www.alunbrig.com.

[15] Amanam, I., Gupta, R., Mambetsariev, I., and Salgia, R. (2018). The brigatinib experience: a new generation of therapy for alk-positive non-small-cell lung cancer. *Future Oncol*, 14:1897–908.

[16] Amar, D., Safer, H., and Shamir, R. (2013). Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Computational Biology*, 9(3):e1002955.

[17] An, G., Faeder, J., and Vodovotz, Y. (2008). Translational systems biology: Introduction of an engineering approach to the pathophysiology of the burn patient. *Journal of Burn Care & Research*, 29(2):277–285.

[18] Anaxomics (n.d.). Anaxomics biotech sl. tpms technology. Internet. http://www.anaxomics.com/tpms.php.

[19] Anellis, I. H. (2012). Peirce's truth-functional analysis and the origin of the truth table. *History and Philosophy of Logic*, 33(1):87–97.

[20] Aoyagi, T. and Matsui, T. (2011). Phosphoinositide-3 kinase signaling in cardiac hypertrophy and heart failure. *Curr Pharm Des*, 17:1818–24. pmid: 21631421.

[21] ASME (n.d.). Asme standardisation committee v&v-40: Verification and validation in computational modeling of medical devices. https://cstools.asme.org/csconnect/committeepages.cfm?committee=100003367.

[22] Avicienna (n.d.). Avicienna institute. http://avicenna-isct.org.

[23] Banaschewski, T., Becker, K., Scherag, S., Franke, B., and Coghill, D. (2010). Molecular genetics of attention-deficit/hyperactivity disorder: an overview. *Eur Child Adolesc Psychiatry*, 19:237–57.

[24] Banerjee, A. and Davë, R. (2004). *Validating clusters using the Hopkins statistic*. IEEE, Budapest.

[25] Banevicius, M., Vilkeviciute, A., Kriauciuniene, L., Liutkeviciene, R., and Deltuva, V. (2018). The association between variants of receptor for advanced glycation end products (rage) gene polymorphisms and age-related macular degeneration. *Med Sci Monit*, 24:190–199. pmid: 29317590.

[26] Bansal, L., Neisen, J., Nichols, E.-M., and Damian, V. (2017). Development and application of a quantitative systems pharmacology (qsp) model of complement pathway to evaluate treatments for autoimmune diseases. *J. Pharmacokinet Pharmacodyn*, 44:S91–S92.

[27] Barabási, A.-L., Gulbahce, N., and Loscalzo, J. (2010). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68.

[28] Barabási, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113.

[29] Baranello, R., Bharani, K., Padmaraju, V., Chopra, N., Lahiri, D., and Greig, N. (2015). Amyloid-beta protein clearance and degradation (abcd) pathways and their role in alzheimer's disease. *Curr Alzheimer Res*, 12:32–46. pmid: 25523424.

[30] Barlesi, F., Dingemans, A., Yang, J., Ou, S., Ahn, J., Petris, L., Hughes, B., Lena, H., Bordogna, W., Golding, S., Morcos, P., Balas, B., Zeaiter, A., and Kim, D. (2016). Updated efficacy and safety from

the global phase ii np28673 study of alectinib in patients (pts) with previously treated alk þ non-small-cell lung cancer (nsclc. *Ann Oncol*, 27:1263.

[31] Barrett, T., Wilhite, S., Ledoux, P., Evangelista, C., Kim, I., and Tomashevsky, M. (2013). Ncbi geo: archive for functional genomics data sets–update. *Nucleic Acids Res*, 41:991–5.

[32] Bazhenova, L., Hodgson, J., Langer, C., Simon, G., Gettinger, S., Ou, S., Reckamp, K., West, H., Chiappori, A., Koh, H., Molina, JR, S., AT, P., and J.D. (2017). Activity of brigatinib (brg) in crizotinib (crz)-resistant alk+ nsclc patients (pts) according to alk plasma mutation status. *J Clin Oncol*, 35:9065.

[33] Bechmann, T., Madsen, J., Brandslund, I., Lund, E., Ormstrup, T., Jakobsen, E., Jylling, A., Steffensen, K., and Jakobsen, A. (2013). Predicting brain metastases of breast cancer based on serum s100b and serum her2. *Oncol Lett*, 6:1265–70.

[34] Bell, S. M., Chang, X., Wambaugh, J. F., Allen, D. G., Bartels, M., Brouwer, K. L., Casey, W. M., Choksi, N., Ferguson, S. S., Fraczkiewicz, G., Jarabek, A. M., Ke, A., Lumen, A., Lynn, S. G., Paini, A., Price, P. S., Ring, C., Simon, T. W., Sipes, N. S., Sprankle, C. S., Strickland, J., Troutman, J., Wetmore, B. A., and Kleinstreuer, N. C. (2018). In vitro to in vivo extrapolation for high throughput prioritization and decision making. *Toxicology in Vitro*, 47:213–227.

[35] Benjamin, E. B. (2015). *Basic Principles of Drug Discovery and Development*. Academic Press.

[36] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B*, 57:289–300.

[37] Benson, N., Matsuura, T., Smirnov, S., Demin, O., Jones, H. M., Dua, P., and van der Graaf, P. H. (2013). Systems pharmacology of the nerve growth factor pathway: use of a systems biology model for the identification of key drug targets using sensitivity analysis and the integration of physiology and pharmacology. *Interface Focus*, 3(2):20120071.

[38] Berriz, G., Beaver, J., Cenik, C., Tasan, M., and Roth, F. (2009). Next generation software for functional trend analysis. *Bioinformatics*, 25:3043–3044. pmid: 19717575.

[39] Bessems, J., Coecke, S., Gouliarmou, V., Whelan, M., and Worth, A. (2015). Eurl ecvam strategy for achieving 3rs impact in the assessment of toxicokinetics and systemic toxicity. Scientific analysis or review LB-NA-27315-EN-N, Publications Office of the European Union, Luxembourg (Luxembourg).

[40] Bethune, G., Bethune, D., Ridgway, N., and Xu, Z. (2010). Epidermal growth factor receptor (egfr) in lung cancer: an overview and update. *J Thorac Dis*, 2:48–51.

[41] BfArM (2020). Decentralised procedure. public assessment report. medikinet 5, 10, 20 mg. medikinet retard 5, 10, 20, 30, 40 mg. methylphenidate hydrochloride.

[42] Bhardwaj, N. and Lu, H. (2005). Correlation between gene expression profiles and protein-protein interactions within and across genomes. *Bioinformatics*, 21(11):2730–2738.

[43] Biederman, J. (2004). Impact of comorbidity in adults with attention-deficit/hyperactivity disorder. *J Clin Psychiatry*, 65:3–7.

[44] Biederman, J., Krishnan, S., Zhang, Y., McGough, J. J., and Findling, R. L. (2007). Efficacy and tolerability of lisdexamfetamine dimesylate (NRP-104) in children with attention-deficit/hyperactivity disorder: A phase III, multicenter, randomized, double-blind, forced-dose, parallel-group study. *Clinical Therapeutics*, 29(3):450–463.

[45] Bilgiç, A., Toker, A., Işik, Ü., and Kilinç, I. (2017). Serum brain-derived neurotrophic factor, glial-derived neurotrophic factor, nerve growth factor, and neurotrophin-3 levels in children with attention-deficit/hyperactivity disorder. *Eur Child Adolesc Psychiatry*, 26:355–63.

[46] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML (2008). Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100:2008.

[47] Bishop, C. (2006). Pattern recognition and machine learning. *Springer Science and Business Media*.

[48] Bleicher, K. H., Böhm, H.-J., Müller, K., and Alanine, A. I. (2003). Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery*, 2(5):369–378.

[49] Blobel, B., Ruotsalainen, P., Brochhausen, M., Oemig, F., and Uribe, G. (2020). Autonomous systems and artificial intelligence in healthcare transformation to 5p medicine - ethical challenges. *Stud Heal Technol Inf.*, pages 270 1089–93.

[50] Boellner, S., Stark, J., Krishnan, S., and Zhang, Y. (2010). Pharmacokinetics of lisdexamfetamine dimesylate and its active metabolite, d-amphetamine, with increasing oral doses of lisdexamfetamine dimesylate in children with attention-deficit/hyperactivity disorder: a single-dose, randomized, open-label, crossover. clin ther.

[51] Bousquet, J., Jorgensen, C., Dauzat, M., Cesario, A., Camuzat, T., and Bourret, R. (2014). Systems medicine approaches for the definition of complex phenotypes in chronic diseases and ageing. from concept to implementation and policies. *Curr Pharm Des.*, 20:5928–44.

[52] Breadner, D., Blanchette, P., Shanmuganathan, S., Boldt, R., and Raphael, J. (2020). Efficacy and safety of alk inhibitors in alk-rearranged non-small cell lung cancer: A systematic review and meta-analysis. *Lung Cancer*, 144:57–63.

[53] Brochot, C. and Quindroit, P. (2018). *Modelling the Fate of Chemicals in Humans Using a Lifetime Physiologically Based Pharmacokinetic (PBPK) Model in MERLIN Expo*. Springer International Publishing, Cham.

[54] Brown, D., Namas, R., Almahmoud, K., Zaaqoq, A., Sarkar, J., and Barclay, D. (2015). Trauma in silico: individual-specific mathematical models and virtual clinical populations. *Sci Transl Med*, 7:285 61.

[55] Brückner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. *International Journal of Molecular Sciences*, 10(6):2763–2788.

[56] Bruggeman, F. J., Hornberg, J. J., Boogerd, F. C., and Westerhoff, H. V. (2007). Introduction to systems biology. *Plant Systems Biology*, 97:1–19.

[57] Bujak, M. and Frangogiannis, N. (2009). The role of il-1 in the pathogenesis of heart disease. arch immunol ther exp (warsz. pmid: 19479203.

[58] Butterfield, M., Saal, J., Young, B., and Young, J. (2016). Supplementary guanfacine hydrochloride as a treatment of

attention deficit hyperactivity disorder in adults: a double blind, placebo-controlled study. *Psychiatry Res*, 236:136–41.

[59] Calonaci, C., Chiacchio, F., and Pappalardo, F. (2012). Optimal vaccination schedule search using genetic algorithm over MPI technology. *BMC Medical Informatics and Decision Making*, 12(1).

[60] Camidge, D., Dziadziuszko, R., Peters, S., Mok, T., Noe, J., Nowicka, M., Gadgeel, S., Cheema, P., Pavlakis, N., Marinis, F., Cho, B., Zhang, L., and Moro-Sibilot, D. (2019a). Updated efficacy and safety data and impact of the eml4-alk fusion variant on the efficacy of alectinib in untreated alk-positive advanced non-small cell lung cancer in the global phase iii alex study. *J Thorac Oncol*, 14:1233–43.

[61] Camidge, D., Gadgeel, S., Ou, S., Gandhi, L., Riely, G., Cetnar, J., West, H., Socinski, M., Chiappori, A., Mekhail, T., Chao, B., Borghaei, H., and Gold, K. (2017). Ma07.02 updated efficacy and safety data from the phase 2 np28761 study of alectinib in alk-positive non-small-cell lung cancer. *J Thorac Oncol*, 12:378.

[62] Camidge, D., Kim, D., Tiseo, M., Langer, C., Ahn, M., Shaw, A., Huber, R., Hochmair, M., Lee, D., Bazhenova, L., Gold, K., Ou, S., and West, H. (2018). Exploratory analysis of brigatinib activity in patients with anaplastic lymphoma kinase-positive non–small-cell lung cancer and brain metastases in two clinical trials. *J Clin Oncol*, 36:2693–701.

[63] Camidge, R., Kim, H., Ahn, M.-J., Yang, J.-H., Han, J.-Y., Hochmair, M., Lee, K., Delmonte, A., Campelo, M. G., Kim, D.-W., Griesinger, F., Felip, E., Califano, R., Spira, A., Gettinger, S., Tiseo, M., Ni, Q., Zhang, P., and Popat, S. (2019b). Brigatinib vs crizotinib in patients with ALK inhibitor-naive advanced ALK+ NSCLC: Updated results from the phase III ALTA-1l trial. *Annals of Oncology*, 30:ix195–ix196.

[64] Carcereny, E., Fernández-Nistal, A., López, A., Montoto, C., Naves, A., Segú-Vergés, C., Coma, M., Jorba, G., Oliva, B., and Mas, J. M. (2021). Head to head evaluation of second generation alk inhibitors brigatinib and alectinib as first-line treatment for alk+ nsclc using an in silico systems biology-based approach. *Oncotarget*, 12(4):316–332.

[65] Carlier, A., Vasilevich, A., Marechal, M., de Boer, J., and Geris, L. (2018). In silico clinical trials for pediatric orphan diseases. *Scientific Reports*, 8(1).

[66] Ceccon, M., Mologni, L., Giudici, G., Piazza, R., Pirola, A., Fontana, D., and Gambacorti-Passerini, C. (2015). Treatment efficacy

and resistance mechanisms using the second-generation alk inhibitor ap26113 in human npm-alk-positive anaplastic large cell lymphoma. *Mol Cancer Res*, 13:775–83 10 1158 1541–7786 –14–0157.

[67] Chakravarthy, U., Wong, T., Fletcher, A., Piault, E., Evans, C., and Zlateva, G. (2010). Clinical risk factors for age-related macular degeneration: A systematic review and meta-analysis. *BMC Ophthalmol*, 10:31.

[68] Chatr-aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N. K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., Stark, C., Breitkreutz, B.-J., Dolinski, K., and Tyers, M. (2016). The BioGRID interaction database: 2017 update. *Nucleic Acids Research*, 45(D1):D369–D379.

[69] Chen, L., Hu, X., Wu, H., Jia, Y., Liu, J., Mu, X., Wu, H., and Zhao, Y. (2019). Over-expression of s100b protein as a serum marker of brain metastasis in non-small cell lung cancer and its prognostic value. *Pathol Res Pract*, 215:427–32.

[70] Cheng, Y., Thalhauser, C. J., Smithline, S., Pagidala, J., Miladinov, M., Vezina, H. E., Gupta, M., Leil, T. A., and Schmidt, B. J. (2017). QSP toolbox: Computational implementation of integrated workflow components for deploying multi-scale mechanistic models. *The AAPS Journal*, 19(4):1002–1016.

[71] Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorith. *The American Statistician*, 49:327–335.

[72] Childers, R., Sunyecz, I., West, T., Cismowski, M., Lucchesi, P., and Gooch, K. (2018). Role of the cytoskeleton in the development of a hypofibrotic cardiac fibroblast phenotype in volume overload heart failure. *Am J Physiol Heart Circ Physiol*, 316:596– 608. pmid: 30575422.

[73] Chiras, D., Kitsos, G., Petersen, M. B., Skalidakis, I., and Kroupis, C. (2014). Oxidative stress in dry age-related macular degeneration and exfoliation syndrome. *Critical Reviews in Clinical Laboratory Sciences*, 52(1):12–27.

[74] Christin, C., Hoefsloot, H. C., Smilde, A. K., Hoekman, B., Suits, F., Bischoff, R., and Horvatovich, P. (2013). A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. *Molecular & Cellular Proteomics*, 12(1):263–276.

[75] Chuang, J., Stehr, H., Liang, Y., Das, M., Huang, J., Diehn, M., Wakelee, H., and Neal, J. (2017). Erbb2-mutated metastatic non-small cell lung cancer: Response and resistance to targeted therapies. *J Thorac Oncol*, 12:833–42.

[76] Chung, N., Miasojedow, B., Startek, M., and Gambin, A. (2019). Jaccard/tanimoto similarity test and estimation methods for biological presence-absence data. *BMC Bioinformatics.*, 20:644.

[77] Ciffroy, P., Alfonso, B., Altenpohl, A., Banjac, Z., Bierkens, J., Brochot, C., Critto, A., Wilde, T. D., Fait, G., Fierens, T., Garratt, J., Giubilato, E., Grange, E., Johansson, E., Radomyski, A., Reschwann, K., Suciu, N., Tanaka, T., Tediosi, A., Holderbeke, M. V., and Verdonck, F. (2016). Modelling the exposure to chemicals for risk assessment: a comprehensive library of multimedia and PBPK models for integration, prediction, uncertainty and sensitivity analysis – the MERLIN-expo tool. *Science of The Total Environment*, 568:770–784.

[78] Clark, R. W., Sutfin, T. A., Ruggeri, R. B., Willauer, A. T., Sugarman, E. D., Magnus-Aryitey, G., Cosgrove, P. G., Sand, T. M., Wester, R. T., Williams, J. A., Perlman, M. E., and Bamberger, M. J. (2004). Raising high-density lipoprotein in humans through inhibition of cholesteryl ester transfer protein. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 24(3):490–497.

[79] Clermont, G., Bartels, J., Kumar, R., Constantine, G., Vodovotz, Y., and Chow, C. (2004). In silico design of clinical trials: A method coming of age. *Critical Care Medicine*, 32(10):2061–2070.

[80] Clewell, H. J., Teeguarden, J., McDonald, T., Sarangapani, R., Lawrence, G., Covington, T., Gentry, R., and Shipp, A. (2002). Review and evaluation of the potential impact of age- and gender-specific pharmacokinetic differences on tissue dosimetry. *Critical Reviews in Toxicology*, 32(5):329–389.

[81] ClinicalTrials.gov (2010). Identifier: Nct00447278. a study comparing the effect of atomoxetine versus other standard care therapy on the long term functioning in attention-deficit/hyperactivity disorder (adhd) children and adolescents (adhd life).
https://clinicaltrials.gov/ct2/show/nct00447278.

[82] ClinicalTrialsgov (2013). Efficacy of bupropion for attention deficit hyperactivity disorder (adhd) in adults (nct01270555.

[83] Coghill, D., Banaschewski, T., Bliss, C., Robertson, B., and Zuddas, A. (2018). Cognitive function of children and adolescents with attention-deficit/hyperactivity disorder in a 2-year open-label study of lisdexamfetamine dimesylate. *CNS Drugs.*, 32:85–95.

[84] Coghill, D., Banaschewski, T., Lecendreux, M., Soutullo, C., Johnson, M., and Zuddas, A. (2013). European, randomized, phase 3 study of lisdexamfetamine dimesylate in children and adolescents with attention-deficit/hyperactivity disorder. *Eur Neuropsychopharmacol.*, 23:1208–18.

[85] Coghill, D., Banaschewski, T., Lecendreux, M., Zuddas, A., Dittmann, R., and Otero, I. (2014). Efficacy of lisdexamfetamine dimesylate throughout the day in children and adolescents with attention-deficit/hyperactivity disorder: results from a randomized, controlled trial. *Eur Child Adolesc Psychiatry*, 23:61–8.

[86] Coghill, D., Banaschewski, T., Nagy, P., Otero, I., Soutullo, C., and Yan, B. (2017). Long-term safety and efficacy of lisdexamfetamine dimesylate in children and adolescents with adhd: a phase iv, 2-year, open-label study in europe. *CNS Drugs.*, 31:625–38.

[87] Collet, P. and Rennard, J.-P. (2008). *"Stochastic Optimization Algorithms." Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications*. Number 1121-1137. IGI Global.

[88] Collet, P. and Rennard, J.-P. (2011). Stochastic optimization algorithms. *Intell Inf Technol*, pages 1121–1137.

[89] Connor, D. and Steingard, R. (2004). New formulations of stimulants for attention-deficit hyperactivity disorder: therapeutic potential. *CNS Drugs*, 18:1011–30.

[90] Coogan, A., Schenk, M., Palm, D., Uzoni, A., Grube, J., and Tsang, A. (2019). Impact of adult attention deficit hyperactivity disorder and medication status on sleep/wake behavior and molecular circadian rhythms. *Neuropsychopharmacology*, 44:1198–206.

[91] Cortese, S. (2020). Pharmacologic treatment of attention deficit-hyperactivity disorder. *N Engl J Med*, 383:1050–6.

[92] Costa, D., Kobayashi, S., Pandya, S., Yeo, W., Shen, Z., Tan, W., and Wilner, K. (2011). Csf concentration of the anaplastic lymphoma kinase inhibitor crizotinib. *J Clin Oncol*, 29:e443–45.

[93] Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M. R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., Birney, E., Hermjakob, H., Stein, L., and D'Eustachio, P. (2013). The reactome pathway knowledgebase. *Nucleic Acids Research*, 42(D1):D472–D477.

[94] Dark, C., Homman-Ludiye, J., and Bryson-Richardson, R. (2018). The role of adhd associated genes in neurodevelopment. *Dev Biol.*, 438:69–83.

[95] Deore, A. B., Dhumane, J. R., Wagh, R., and Sonawane, R. (2019). The stages of drug discovery and development process. *Asian Journal of Pharmaceutical Research and Development*, 7(6):62–67.

[96] Dew, R. and Kollins, S. (2010). Lisdexamfetamine dimesylate: a new option in stimulant treatment for adhd. expert opin pharmacother.

[97] DiMasi, J. A., Grabowski, H. G., and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of r&d costs. *Journal of Health Economics*, 47:20–33.

[98] Döpfner, M., Ose, C., Fischer, R., Ammer, R., and Scherag, A. (2011). Comparison of the efficacy of two different modified release methylphenidate preparations for children and adolescents with attention-deficit/hyperactivity disorder in a natural setting: comparison of the efficacy of medikinet® retard and concerta®-a rand. *J Child Adolesc Psychopharmacol*, 21:445–54.

[99] Dubuisson, M.-P. and Jain, A. (1994). A modified hausdorff distance for object matching.

[100] Edwards, L. M. and Thiele, I. (2013). Applying systems biology methods to the study of human physiology in extreme environments. *Extreme Physiology & Medicine*, 2(1).

[101] EES, R., AW, H., KP, W., A, L., MAJ, L., and FH, R. (2016). Epidemiology of heart failure: The prevalence of heart failure and ventricular dysfunction in older adults over time. a systematic review. *Eur J Heart Fail*, 18:242–52. pmid: 26727047.

[102] EMA (2011). Internet. https://www.ema.europa.eu/en/documents/presentation/presentation-role-modelling-simulation-regulatory-decision-making-europe_en.pdf.

[103] EMA (n.d.). European parliament. authorisation and supervision of veterinary medicinal products. https://www.europarl.europa.eu/doceo/document/ta-8-2016-0088_en.html?redirect.

[104] Ennis, I., Aiello, E., Cingolani, H., and Perez, N. (2013). The autocrine/paracrine loop after myocardial stretch: Mineralocorticoid receptor activation. *Curr Cardiol Rev*, 9:230–40. pmid: 23909633.

[105] Entresto (2015). European medicines agency (EMA). Internet. https://www.ema.europa.eu/en/documents/assessment-report/entresto-epar-public-assessment-report_en.pdf.

[106] ESS7 (2014). The european social survey. ess round 7: European social survey round 7 data. https://www.europeansocialsurvey.org/data/download.html?r=7 (accessed June, 2020).

[107] Fàbrega, F., Nadal, M., Schuhmacher, M., Domingo, J. L., and Kumar, V. (2016). Influence of the uncertainty in the validation of PBPK models: A case-study for PFOS and PFOA. *Regulatory Toxicology and Pharmacology*, 77:230–239.

[108] Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., and D'Eustachio, P. (2017). The reactome pathway knowledgebase. *Nucleic Acids Research*, 46(D1):D649–D655.

[109] Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J. J., and Gardner, T. S. (2007). Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biology*, 5(1):e8.

[110] Faraone, S. and Larsson, H. (2019). Genetics of attention deficit hyperactivity disorder. mol psychiatry.

[111] Faraone, S. and Mick, E. (2010). Molecular genetics of attention deficit hyperactivity disorder. psychiatr clin north am.

[112] Faraone, S., Spencer, T., Kollins, S., Glatt, S., and Goodman, D. (2012). Dose response effects of lisdexamfetamine dimesylate treatment in adults with adhd: an exploratory study. *J Atten Disord*, 16:118–27.

[113] FDA (2016). Reporting of computational modeling studies in medical device submissions; guidance for industry and food and drug administration staff; availability.

[114] FDA (2019). Us food and drug administration (fda). highlights of prescribing information: alecensa. Internet. http://www.fda.gov/CompanionDiagnostics.

[115] Feldman, A., Haller, J., and DeKosky, S. (2016). Valsartan/sacubitril for heart failure: Reconciling disparities between preclinical and clinical investigations. *JAMA—J Am Med Assoc*, 315:25–26. pmid: 26641736.

[116] Felip, E., Orlov, S., Park, K., Yu, C., Tsai, C., Nishio, M., Cobo Dols, M., McKeage, M., Su, W., Mok, T., Scagliotti, G., Spigel, D., and Branle, F. (2015). Ascend-3: A single-arm, open-label, multicenter phase ii study of ceritinib in alki-naïve adult patients (pts) with alk-rearranged (alk+) non-small cell lung cancer (nsclc. *J Clin Oncol*, 33:8060.

[117] Figueroa, R., Zeng-Treitler, Q., Kandula, S., and Ngo, L. (2012). Predicting sample size required for classification performance. *BMC Med Inf Decis Mak.*, page 12 8.

[118] Fisher, J., Gearhart, J., and Lin, Z., editors (2020). *Physiologically Based Pharmacokinetic (PBPK) Modeling*. Academic Press.

[119] Fiuza-Luces, C., Santos-Lozano, A., Llavero, F., Campo, R., Nogales-Gadea, G., Díez-Bermejo, J., Baladrón, C., González-Murillo, Á., Arenas, J., Martín, M., Andreu, A., Pinós, T., and Gálvez, B. (2018). Muscle molecular adaptations to endurance exercise training are conditioned by glycogen availability: a proteomics-based analysis in the mcardle mouse model. *J Physiol*, 596:1035–61.

[120] Fontana, D., Ceccon, M., Gambacorti-Passerini, C., and Mologni, L. (2015). Activity of second-generation alk inhibitors against crizotinib-resistant mutants in an npm-alk model compared to eml4-alk. *Cancer Med*, 4:953–65.

[121] Food and Administration, D. (2017a). Alectinib approved for (alk) positive metastatic non-small cell lung cancer (nsclc.

[122] Food and Administration, D. (2017b). Fda broadens ceritinib indication to previously untreated alk-positive metastatic nsclc.

[123] Food and Administration, D. (n.d.). Fda approves brigatinib for alk-positive metastatic nsclc.

https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-brigatinib-alk-positive-metastatic-nsclc.

[124] Food, U. and Administration, D. (2014). Fda approves zykadia for late-stage lung cancer: breakthrough therapy drug approved four months ahead of review completion goal date. Scholar.

[125] Foxenberg, R. J., Ellison, C. A., Knaak, J. B., Ma, C., and Olson, J. R. (2011). Cytochrome p450-specific human PBPK/PD models for the organophosphorus pesticides: Chlorpyrifos and parathion. *Toxicology*, 285(1-2):57–66.

[126] Franceschini, A. and Fattore, L. (2021). Gender-specific approach in psychiatric diseases: because sex matters. *Eur J Pharmacol*, 896:173895.

[127] Frank, R. (1997). Growth factors in age-related macular degeneration: Pathogenic and therapeutic implications. *Ophthalmic Res*, 29:341–53. pmid: 9323725.

[128] Franke, B., Michelini, G., Asherson, P., Banaschewski, T., Bilbow, A., and Buitelaar, J. (2018). Live fast, die young? a review on the developmental trajectories of adhd across the lifespan. *Eur Neuropsychopharmacol.*, 28:1059–88.

[129] Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Elsevier.

[130] Gabow, H., editor (2007). *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, USA. Society for Industrial and Applied Mathematics.

[131] Gadgeel, S. M., Shaw, A. T., Govindan, R., Gandhi, L., Socinski, M. A., Camidge, D. R., Petris, L. D., Kim, D.-W., Chiappori, A., Moro-Sibilot, D. L., Duruisseaux, M., Crino, L., Pas, T. D., Dansin, E., Tessmer, A., Yang, J. C.-H., Han, J.-Y., Bordogna, W., Golding, S., Zeaiter, A., and Ou, S.-H. I. (2016). Pooled analysis of CNS response to alectinib in two studies of pretreated patients with ALK-positive non–small-cell lung cancer. *Journal of Clinical Oncology*, 34(34):4079–4085.

[132] Gainor, J., Dardaei, L., Yoda, S., Friboulet, L., Leshchiner, I., Katayama, R., Dagogo-Jack, I., Gadgeel, S., Schultz, K., Singh, M., Chin, E., Parks, M., and Lee, D. (2016). Molecular mechanisms of resistance to first- and second-generation alk inhibitors in alk-rearranged lung cancer. *Cancer Discov*, 6:1118–33.

[133] Gainor, J., Ou, S., Logan, J., Borges, L., and Shaw, A. (2013). The central nervous system as a sanctuary site in alk-positive non-small-cell lung cancer. *J Thorac Oncol*, 8:1570–73.

[134] Gainor, J. F. and Shaw, A. T. (2013). Emerging paradigms in the development of resistance to tyrosine kinase inhibitors in lung cancer. *Journal of Clinical Oncology*, 31(31):3987–3996.

[135] Galindo, C., Ryzhov, S., and Sawyer, D. (2014). Neuregulin as a heart failure therapy and mediator of reverse remodeling. *Curr Heart Fail Rep.*, 11:40–9. pmid: 24234399.

[136] Gao, T. and Qian, J. (2019). Eagle: an algorithm that utilizes a small number of genomic features to predict tissue/cell type-specific enhancer-gene interactions. *PLoS Comput Biol*, 15:e1007436.

[137] Garbayo, E., Gavira, J., Yebenes, M., Pelacho, B., Abizanda, G., and Lana, H. (2016). Catheter-based intramyocardial injection of fgf1 or nrg1-loaded mps improves cardiac function in a preclinical model of ischemia-reperfusion. *Sci Rep*, 6:25932 27184924.

[138] Geerts, H., Roberts, P., Spiros, A., and Carr, R. (2013). A strategy for developing new treatment paradigms for neuropsychiatric and neurocognitive symptoms in alzheimer's disease. *Frontiers in Pharmacology*, 4.

[139] Geerts, H., Wikswo, J., van der Graaf, P. H., Bai, J. P., Gaiteri, C., Bennett, D., Swalley, S. E., Schuck, E., Kaddurah-Daouk, R., Tsaioun, K., and Pelleymounter, M. (2020). Quantitative systems pharmacology for neuroscience drug discovery and development: Current status, opportunities, and challenges. *CPT: Pharmacometrics & Systems Pharmacology*, 9(1):5–20.

[140] Gérard, C., Bleyzac, N., Girard, P., Freyer, G., Bertrand, Y., and Tod, M. (2010). Influence of dosing schedule on organ exposure to cyclosporin in pediatric hematopoietic stem cell transplantation: Analysis with a PBPK model. *Pharmaceutical Research*, 27(12):2602–2613.

[141] Ginsberg, L., Katic, A., Adeyi, B., Dirks, B., Babcock, T., and Lasser, R. (2011). Long-term treatment outcomes with lisdexamfetamine dimesylate for adults with attention-deficit/hyperactivity disorder stratified by baseline severity. curr med res opin.

[142] Girard, P., Cucherat, M., Guez, D., Boissel, J.-P., Cucherat, M., Durrleman, S., Girard, P., Guez, D., Koen, R., Laveille, C., Mathiex-Fortunet, H., Micallef, J., Missoum, N., Paintaud, G., Perault, M.-C., Tansey, M., Thomas, J.-L., Treluyer, J.-M., Variol, P., and Waegemans, T. (2004). Simulation des essais cliniques dans le développement des médicaments. *Therapies*, 59(3):287–295. XIXèmes Rencontres Nationales de Pharmacologie Clinique, Giens 28-30 septembre 2003.

[143] Glass, K., Huttenhower, C., Quackenbush, J., and Yuan, G.-C. (2013). Passing messages between biological networks to refine predicted interactions. *PLoS ONE*, 8(5):e64832.

[144] Glenn, J. and Stitt, A. (2009). The role of advanced glycation end products in retinal ageing and disease. *Biochim Biophys Acta—Gen Subj*, 1790:1109–16. pmid: 19409449.

[145] Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabási, A.-L. (2007). The human disease network. *Proceedings of the National Academy of Sciences*, 104(21):8685–8690.

[146] Goldstein, D. B. (2009). Common genetic variation and human traits. *New England Journal of Medicine*, 360(17):1696–1698.

[147] Gomeni, R., D'Angeli, C., and Bye, A. (2002). In silico prediction of optimal in vivo delivery properties using convolution-based model and clinical trial simulation. *Pharmaceutical Research*, 19(1):99–103.

[148] Gómez-Serrano, M., Camafeita, E., García-Santos, E., López, J. A., Rubio, M. A., Sánchez-Pernaute, A., Torres, A., Vázquez, J., and Peral, B. (2016). Proteome-wide alterations on adipose tissue from obese patients as age-, diabetes- and gender-specific hallmarks. *Scientific Reports*, 6(1).

[149] Griesinger, F., Roeper, J., Pöttgen, C., Willborn, K., and Eberhardt, W. (2018). Brain metastases in alk-positive nsclc - time to adjust current treatment algorithms. *Oncotarget*, 9:35181–94. 10.18632/oncotarget.26073.

[150] Griffin, R. and Ramirez, R. (2017). Molecular targets in non–small cell lung cancer. *Ochsner J*, 17:388–92 10 1043 –17–0033.

[151] Grossniklaus, H. and Green, W. (2004). Choroidal neovascularization. *Am J Ophthalmol*, 137:496–503. pmid: 15013874.

[152] Gullo, F., van der Garde, M., Russo, G., Pennisi, M., Motta, S., Pappalardo, F., and Watt, S. (2015). Computational modeling of the expansion of human cord blood cd133 hematopoietic stem/progenitor cells with different cytokine combinations. *Bioinformatics*, 31(15):2514–2522.

[153] Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182.

[154] Guze, S. (1995). Diagnostic and statistical manual of mental disorders. *DSM-IV). Am J Psychiatry*, page 152.

[155] Haidar, A., Wilinska, M., Graveston, J., and Hovorka, R. (2013). Stochastic virtual population of subjects with type 1 diabetes for the assessment of closed-loop glucose controllers. *IEEE Trans Biomed Eng*, 60:3524–33.

[156] Halleyhit (2022). Generate random numbers according to pdf or cdf (https://www.mathworks.com/matlabcentral/fileexchange/68492-generate-random-numbers-according-to-pdf-or-cdf), matlab central file exchange. recuperado december 5, 2022.

[157] Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C. Y., Lee, M., Kim, E., Lee, S., Kang, B., Jeong, D., Kim, Y., Jeon, H.-N., Jung, H., Nam, S., Chung, M., Kim, J.-H., and Lee, I. (2017). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research*, 46(D1):D380–D386.

[158] Han, H., Shim, H., Shin, D., Shim, J. E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., Kim, H., Kim, K., Yang, S., Bae, D., Yun, A., Kim, S., Kim, C. Y., Cho, H. J., Kang, B., Shin, S., and Lee, I. (2015). TRRUST: a reference database of human transcriptional regulatory interactions. *Scientific Reports*, 5(1).

[159] Han, L., Liang, X., Chen, L., Bao, S., and Yan, Z. (2013). Sirt1 is highly expressed in brain metastasis tissues of non-small cell lung cancer (nsclc) and in positive regulation of nsclc cell migration. *Int J Clin Exp Pathol*, 6:2357–65.

[160] Hay, M., Thomas, D. W., Craighead, J. L., Economides, C., and Rosenthal, J. (2014). Clinical development success rates for investigational drugs. *Nature Biotechnology*, 32(1):40–51.

[161] Hecker, N., Ahmed, J., Eichborn, J., Dunkel, M., Macha, K., and Eckert, A. (2011). Supertarget goes quantitative: update on drug-target interactions. *Nucleic Acids Res*, 40:1113– 1117 22067455.

[162] Hegab, Z., Gibbons, S., Neyses, L., and Mamas, M. (2012). Role of advanced glycation end products in cardiovascular disease. *World J Cardiol*, 4:90–102. pmid: 22558488.

[163] Heidegger, I., Kern, J., Ofer, P., Klocker, H., and Massoner, P. (2014). Oncogenic functions of igf1r and insr in prostate cancer include enhanced tumor growth, cell migration and angiogenesis. *Oncotarget*, 5:2723–35.

[164] Herrando-Grabulosa, M., Mulet, R., Pujol, A., Mas, J. M., Navarro, X., Aloy, P., Coma, M., and Casas, C. (2016). Novel neuroprotective multicomponent therapy for amyotrophic lateral sclerosis designed by networked systems. *PLOS ONE*, 11(1):e0147626.

[165] Herreros-Villanueva, M., Perez-Palacios, R., Castillo, S., Segu, C., Sardon, T., Mas, J., Martin, A., and Arroyo, R. (2018). Biological relationships between mirnas used for colorectal cancer screening. *J Mol Biomark Diagn*, 09:10 4172 2155–9929 1000398.

[166] Hervas, A., Huss, M., Johnson, M., McNicholas, F., Stralen, J., and Sreckovic, S. (2014). Efficacy and safety of extended-release guanfacine hydrochloride in children and adolescents with attention-deficit/hyperactivity disorder: a randomized, controlled, phase iii trial. *Eur Neuropsychopharmacol.*, 24:1861–72.

[167] Higgins, P. (2015). Balancing ahr-dependent pro-oxidant and nrf2-responsive anti-oxidant pathways in age-related retinopathy: Is serpine1 expression a therapeutic target in disease onset and progression? *J Mol Genet Med*, 8:101 25237384.

[168] Hill, R. G. and Rang, H. P., editors (2012). *Drug Discovery and Development*. Churchill Livingstone.

[169] Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1.

[170] Holbein, M. E. B. (2009). Understanding FDA regulatory requirements for investigational new drug applications for sponsor-investigators. *Journal of Investigative Medicine*, 57(6):688–694.

[171] Holford, N., Ma, S., and Ploeger, B. (2010). Clinical trial simulation: a review. clin pharmacol ther.

[172] Howard, A., Kennedy, T., Mitchell, J., Sibley, M., Hinshaw, S., and Arnold, L. (2020). Early substance use in the pathway from childhood attention-deficit/hyperactivity disorder (adhd) to young adult substance use: evidence of statistical mediation and substance specificity. *Psychol Addict Behav.*, 34:281–92.

[173] Howlader, N., Noone, A., Krapcho, M., Miller, D., Brest, A., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D., Chen, H., Feuer, E., and Cronin (eds), K. (2019). Seer cancer statistics review, 1975-2016, national cancer institute. bethesda, md, https://seer.cancer.gov/csr/1975_2016/, based on november 2018 seer data submission, posted to the seer web site. *Cancer Statistics Review*.

[174] Hu, L., Zhang, J., Zhu, H., Min, J., Feng, Y., and Zhang, H. (2010). Biological characteristics of a specific brain metastatic cell line derived from human lung adenocarcinoma. *Med Oncol*, 27:708–14.

[175] Huang, W., Liu, S., Zou, D., Thomas, M., Wang, Y., Zhou, T., Romero, J., Kohlmann, A., Li, F., Qi, J., Cai, L., Dwight, T., and Xu, Y. (2016). Discovery of brigatinib (ap26113), a phosphine oxide-containing, potent, orally active inhibitor of anaplastic lymphoma kinase. *J Med Chem*, 59:4948–64.

[176] Huber, R., Hansen, K., Paz-Ares Rodríguez, L., West, H., Reckamp, K., Leighl, N., Tiseo, M., Smit, E., Kim, D., Gettinger, S., Hochmair, M., Kim, S., and Langer, C. (2020). Brigatinib in crizotinib-refractory alk+ nsclc: 2-year follow-up on systemic and intracranial outcomes in the phase 2 alta trial. *J Thorac Oncol*, 15:404–15.

[177] Huber, W. (2005). A new strategy for improved secondary screening and lead optimization using high-resolution spr characterization of compound-target interactions. *Journal of Molecular Recognition*, 18(4):273–281.

[178] Hughes, J., Rees, S., Kalindjian, S., and Philpott, K. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, 162(6):1239–1249.

[179] Huss, M., Ginsberg, Y., Tvedten, T., Arngrim, T., Philipsen, A., and Carter, K. (2014). Methylphenidate hydrochloride modified-release in adults with attention deficit hyperactivity disorder: a randomized double-blind placebo-controlled trial. adv ther.

[180] Hyttinen, J. M., Amadio, M., Viiri, J., Pascale, A., Salminen, A., and Kaarniranta, K. (2014). Clearance of misfolded and aggregated proteins

by aggrephagy and implications for aggregation diseases. *Ageing Research Reviews*, 18:16–28.

[181] Iborra-Egea, O., Gálvez-Montón, C., Roura, S., Perea-Gil, I., Prat-Vidal, C., and Soler-Botija, C. (2017). Mechanisms of action of sacubitril/valsartan on cardiac remodeling: a systems biology approach. *npj Syst Biol Appl*, 3:1–8.

[182] Iborra-Egea, O., Santiago-Vacas, E., Yurista, S. R., Lupón, J., Packer, M., Heymans, S., Zannad, F., Butler, J., Pascual-Figal, D., Lax, A., Núñez, J., de Boer, R. A., and Bayés-Genís, A. (2019). Unraveling the molecular mechanism of action of empagliflozin in heart failure with reduced ejection fraction with or without diabetes. *JACC: Basic to Translational Science*, 4(7):831–840.

[183] Imming, P., Sinning, C., and Meyer, A. (2006). Drugs, their targets and the nature and number of drug targets. *Nature Reviews Drug Discovery*, 5(10):821–834.

[184] Inc, T. M. (2017). *MATLAB and SimBiology Toolbox Release 5.7*. The MathWorks Inc, Natick, MA.

[185] Iyengar, R., Zhao, S., Chung, S.-W., Mager, D. E., and Gallo, J. M. (2012). Merging systems biology with pharmacodynamics. *Science Translational Medicine*, 4(126).

[186] J, V., S, O., RM, S., G, S., PM, B. E. H., and RC, R. (2009). Crosstalk between epidermal growth factor receptor- and insulin-like growth factor-1 receptor signaling: implications for cancer therapy. *Curr Cancer Drug Targets*, 9:748–60 10 2174 156800909789271495.

[187] J. O'Donnell III, J., Somberg, J., Idemyor, V., and O'Donnell, J. T., editors (2019). *Drug Discovery and Development*. CRC Press.

[188] Jackson-Cowan, L., Cole, E., Arbiser, J., Silverberg, J., and Lawley, L. (2021). Th2 sensitization in the skin-gut-brain axis: how early-life th2-mediated inflammation may negatively perpetuate developmental and psychologic abnormalities. *Pediatr Dermatol.* Epub ahead of print].

[189] Jain, R. and Chen, H. (2017). Spotlight on brigatinib and its potential in the treatment of patients with metastatic alk-positive non-small cell lung cancer who are resistant or intolerant to crizotinib. *Lung Cancer (Auckl*, 8:169–77.

[190] Jain, R., Segal, S., Kollins, S., and Khayrallah, M. (2011). Clonidine extended-release tablets for pediatric patients with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*, 50:171–9.

[191] Jamei, M., Bajot, F., Neuhoff, S., Barter, Z., Yang, J., Rostami-Hodjegan, A., and Rowland-Yeo, K. (2013). A mechanistic framework for in vitro–in vivo extrapolation of liver membrane transporters: Prediction of drug–drug interaction between rosuvastatin and cyclosporine. *Clinical Pharmacokinetics*, 53(1):73–87.

[192] Jamwal, G., Singh, G., Dar, M. S., Singh, P., Bano, N., Syed, S. H., Sandhu, P., Akhter, Y., Monga, S. P., and Dar, M. J. (2018). Identification of a unique loss-of-function mutation in IGF1r and a crosstalk between IGF1r and wnt/β-catenin signaling pathways. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1865(6):920–931.

[193] Jean-Quartier, C., Jeanquartier, F., Jurisica, I., and Holzinger, A. (2018). In silico cancer research towards 3r. *BMC Cancer*, 18:408 10 1186 12885–018–4302–0.

[194] Jiang, W., Jia, Q., Liu, L., Zhao, X., Tan, A., Ma, N., and Zhang, H. (2011). S100b promotes the proliferation, migration and invasion of specific brain metastatic lung adenocarcinoma cell line. *Cell Biochem Funct*, 29:582–88.

[195] Johung, K., Yeh, N., Desai, N., Williams, T., Lautenschlaeger, T., Arvold, N., Ning, M., Attia, A., Lovly, C., Goldberg, S., Beal, K., Yu, J., and Kavanagh, B. (2016). Extended survival and prognostic factors for patients with alk-rearranged non-small-cell lung cancer and brain metastasis. *J Clin Oncol*, 34:123–29.

[196] Jolliffe, I. (2002). *Principal Component Analysis*. Springer, Berlin.

[197] Jones, H. M., Dickins, M., Youdim, K., Gosset, J. R., Attkins, N. J., Hay, T. L., Gurrell, I. K., Logan, Y. R., Bungay, P. J., Jones, B. C., and Gardner, I. B. (2011). Application of PBPK modelling in drug discovery and development at pfizer. *Xenobiotica*, 42(1):94–106.

[198] Jorba, G., Aguirre-Plans, J., Junet, V., Segú-Vergés, C., Ruiz, J. L., Pujol, A., Fernández-Fuentes, N., Mas, J. M., and Oliva, B. (2020). In-silico simulated prototype-patients using tpms technology to study a potential adverse effect of sacubitril and valsartan. *PLOS ONE*, 15(2):e0228926.

[199] juan Chen, L., ya Li, X., qiu Zhao, Y., jing Liu, W., juan Wu, H., Liu, J., qian Mu, X., and bo Wu, H. (2017). Down-regulated microRNA-375 expression as a predictive biomarker in non-small cell lung cancer brain metastasis and its prognostic significance. *Pathology - Research and Practice*, 213(8):882–888.

[200] Jusko, W. J. (2013). Moving from basic toward systems pharmacodynamic models. *Journal of Pharmaceutical Sciences*, 102(9):2930–2940.

[201] Jusko, W. J. and Ko, H. C. (1994). Physiologic indirect response models characterize diverse types of pharmacodynamic effects. *Clinical Pharmacology and Therapeutics*, 56(4):406–419.

[202] Kaarniranta, K., Salminen, A., Haapasalo, A., Soininen, H., and Hiltunen, M. (2011). Age-related macular degeneration (amd): Alzheimer's disease in the eye? *J Alzheimer's Dis*, 24:615–31. pmid: 21297256.

[203] Kaarniranta, K., Sinha, D., Blasiak, J., Kauppinen, A., Veréb, Z., and Salminen, A. (2013). Autophagy and heterophagy dysregulation leads to retinal pigment epithelium dysfunction and development of age-related macular degeneration. *Autophagy*, 9:973–84. pmid: 23590900.

[204] Kanehisa, M. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30.

[205] Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361.

[206] Karanasiou, G. S., Tsobou, P. I., Tachos, N. S., Antonini, L., Petrini, L., Pennati, G., Gijsen, F., Nezami, F. R., Tzafiri, R., Vaughan, T., and et al. (2020). Design and implementation of in silico clinical trial for bioresorbable vascular scaffolds. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*.

[207] Karsanov, N., Pirtskhalaishvili, M., Semerikova, V., and Losaberidze, N. (1986). Thin myofilament proteins in norm and heart failure i. polymerizability of myocardial straub actin in acute and chronic heart failure. *Basic Res Cardiol*, 81:199–212. pmid: 3741358.

[208] Kassem, L., Shohdy, K., Lasheen, S., Abdel-Rahman, O., Ali, A., and Abdel-Malek, R. (2019). Safety issues with the alk inhibitors in

the treatment of nsclc: A systematic review. *Crit Rev Oncol Hematol*, 134:56–64.

[209] Katayama, R. (2018). Drug resistance in anaplastic lymphoma kinase-rearranged lung cancer. *Cancer Sci*, 109:572–80.

[210] Katzman, M., Bilkey, T., Chokka, P., Fallu, A., and Klassen, L. (2017). Adult adhd and comorbid disorders: clinical implications of a dimensional approach. *BMC Psychiatry*, 17:302.

[211] Kawano, H., Cody, R. J., Graf, K., Goetze, S., Kawano, Y., Schnee, J., Law, R. E., and Hsueh, W. A. (2000). Angiotensin II enhances integrin and α-actinin expression in adult rat cardiac fibroblasts. *Hypertension*, 35(1):273–279.

[212] Ke, A. B., Nallani, S. C., Zhao, P., Rostami-Hodjegan, A., Isoherranen, N., and Unadkat, J. D. (2013). A physiologically based pharmacokinetic model to predict disposition of CYP2d6 and CYP1a2 metabolized drugs in pregnant women. *Drug Metabolism and Disposition*, 41(4):801–813.

[213] Kell, D. B. and Goodacre, R. (2014). Metabolomics and systems pharmacology: why and how to model the human metabolic network for drug discovery. *Drug Discovery Today*, 19(2):171–182.

[214] Kessler, R., Adler, L., Barkley, R., Biederman, J., Conners, C., and Demler, O. (2006). The prevalence and correlates of adult adhd in the united states: results from the national comorbidity survey replication. *Am J Psychiatry*, 163:716–23.

[215] Kim, D., Mehra, R., Tan, D., Felip, E., Chow, L., Camidge, D., Vansteenkiste, J., Sharma, S., Pas, T., Riely, G., Solomon, B., Wolf, J., and Thomas, M. (2016a). Activity and safety of ceritinib in patients with alk-rearranged non-small-cell lung cancer (ascend-1): updated results from the multicentre, open-label, phase 1 trial. *Lancet Oncol*, 17:452–63 10 1016 1470–2045 15 00614–2.

[216] Kim, P. S. and Lee, P. P. (2012). Modeling protective anti-tumor immunity via preventative cancer vaccines using a hybrid agent-based and delay differential equation approach. *PLoS Computational Biology*, 8(10):e1002742.

[217] Kim, S., Thiessen, P., Bolton, E., Chen, J., Fu, G., and Gindulyte, A. (2016b). Pubchem substance and compound databases. *Nucleic Acids Res*, 44:1202– 1213 26400175.

[218] Kim, Y., Hatley, O., Rhee, S., Yi, S., Lee, H., and Yoon, S. (2019). Development of a korean-specific virtual population for physiologically based pharmacokinetic modelling and simulation. *Biopharm Drug Dispos.*, 40:135–50.

[219] Kira, K. and Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, AAAI'92, pages 129–134. AAAI Press.

[220] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.

[221] Kirouac, D. C., Du, J. Y., Lahdenranta, J., Overland, R., Yarar, D., Paragas, V., Pace, E., McDonagh, C. F., Nielsen, U. B., and Onsum, M. D. (2013). Computational modeling of $erbb_2$-amplified breast cancer identifies combined ErbB2/3 blockade as superior to the combination of MEK and AKT inhibitors. *Science Signaling*, 6(288).

[222] Kirouac, D. C., Schaefer, G., Chan, J., Merchant, M., Orr, C., Huang, S.-M. A., Moffat, J., Liu, L., Gadkar, K., and Ramanujan, S. (2017). Clinical responses to ERK inhibition in BRAF v600e-mutant colorectal cancer predicted using a computational model. *npj Systems Biology and Applications*, 3(1).

[223] Kitagishi, Y., Minami, A., Nakanishi, A., Ogura, Y., and Matsuda, S. (2015). Neuron membrane trafficking and protein kinases involved in autism and adhd. *Int J Mol Sci*, 16:3095–115.

[224] Kittel-Schneider, S. and Reif, A. (2020). Adulte aufmerksamkeitsdefizit-/hyperaktivitätsstörung und komorbidität: neue befunde zu epidemiologischen und genetischen faktoren. *Der Nervenarzt*, 91(7):575–582.

[225] Kodama, T., Tsukaguchi, T., Satoh, Y., Yoshida, M., Watanabe, Y., Kondoh, O., and Sakamoto, H. (2014). Alectinib shows potent antitumor activity against ret-rearranged non-small cell lung cancer. *Mol Cancer Ther*, 13:2910–18 10 1158 1535–7163 –14–0274.

[226] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, 2:1137–1143.

[227] Kong, R., Yi, F., Wen, P., Liu, J., Chen, X., Ren, J., Li, X., Shang, Y., Nie, Y., Wu, K., Fan, D., Zhu, L., Feng, W., and Wu, J. Y. (2015).

Myo9b is a key player in SLIT/ROBO-mediated lung tumor suppression. *Journal of Clinical Investigation*, 125(12):4407–4420.

[228] Kosinsky, Y., Dovedi, S. J., Peskov, K., Voronova, V., Chu, L., Tomkinson, H., Al-Huniti, N., Stanski, D. R., and Helmlinger, G. (2018). Radiation and PD-(l)1 treatment combinations: immune response and dose optimization via a predictive systems model. *Journal for ImmunoTherapy of Cancer*, 6(1).

[229] Krishnan, S. and Zhang, Y. (2008). Relative bioavailability of lisdexamfetamine 70-mg capsules in fasted and fed healthy adult volunteers and in solution: a single-dose, crossover pharmacokinetic study. *J Clin Pharmacol*, 48:293–302.

[230] Kubat, M. (1999). Neural networks: a comprehensive foundation by simon haykin, macmillan, 1994, isbn 0-02-352781-7. *The Knowledge Engineering Review*, 13(4):409–412.

[231] Kuepfer, L., Niederalt, C., Wendl, T., Schlender, J.-F., Block, M., Eissing, T., and Teutonico, D. (2016). PBPK modelling of intracellular drug delivery through active and passive transport processes. In *Intracellular Delivery III*, pages 363–374. Springer International Publishing.

[232] Kuhn, M., Letunic, I., Jensen, L. J., and Bork, P. (2015). The SIDER database of drugs and side effects. *Nucleic Acids Research*, 44(D1):D1075–D1079.

[233] Kuperstein, I., Bonnet, E., Nguyen, H., Cohen, D., Viara, E., Grieco, L., Fourquet, S., Calzone, L., Russo, C., Kondratova, M., Dutreix, M., Barillot, E., and Zinovyev, A. (2015). Atlas of cancer signalling network: a systems biology resource for integrative analysis of cancer data with google maps. *Oncogenesis*, 4:e160–14.

[234] Lakó-Futó, Z., Szokodi, I., Sármán, B., Földes, G., Tokola, H., and Ilves, M. (2003). Evidence for a functional role of angiotensin ii type 2 receptor in the cardiac hypertrophic process in vivo in the rat heart. *Circulation*, 108:2414–22. pmid: 14568903.

[235] Lander, A. D. (2010). The edges of understanding. *BMC Biology*, 8(1).

[236] Larson, K., Russ, S., Kahn, R., and Halfon, N. (2011). Patterns of comorbidity, functioning, and service use for us children with adhd, 2007. *Pediatrics.*, 127:462–70.

[237] Law, C., Alhamdoosh, M., Su, S., Dong, X., Tian, L., and Smyth, G. (2018). Rna-seq analysis is easy as 1-2-3 with limma, glimma and edger. *F1000Research.*, 5:1408.

[238] LeBlanc, T., McNeil, M., Kamal, A., Currow, D., and Abernethy, A. (2015). Polypharmacy in patients with advanced cancer and the role of medication discontinuation. *Lancet Oncol*, 16:e333–41:10 1016 1470–2045 15 00080–7.

[239] Lee, A., Kulkarni, M., Fang, A., Edelstein, S., Osborn, M., and Brantley, M. (2010). The effect of genetic variants in serping1 on the risk of neovascular age-related macular degeneration. *Br J Ophthalmol*, 94:915–7. pmid: 20606025.

[240] Lee, J., Kim, H., Kim, K. H., Jung, D., Jowsey, T., and Webster, C. S. (2020). Effective virtual patient simulators for medical communication training: A systematic review. *Medical Education*, 54(9):786–795.

[241] Leffa, D. and Torres, I. (2018). Rohde la a review on the role of inflammation in attention-deficit/hyperactivity disorder. *Neuroimmunomodulation.*, 25:328–33.

[242] Leger, F., Fernagut, P.-O., Canron, M.-H., Léoni, S., Vital, C., Tison, F., Bezard, E., and Vital, A. (2011). Protein aggregation in the aging retina. *Journal of Neuropathology & Experimental Neurology*, 70(1):63–68.

[243] Leil, T. A. and Ermakov, S. (2015). Editorial: The emerging discipline of quantitative systems pharmacology. *Frontiers in Pharmacology*, 6.

[244] Li, D., Yang, W., Arthur, C., Liu, J., Cruz-Niera, C., and Yang, M. (2018a). Systems biology analysis reveals new insights into invasive lung cancer. *BMC Syst Biol*, 12:117 10 1186 12918–018–0637–.

[245] Li, J., Huang, Y., Wu, M., Wu, C., Li, X., and Bao, J. (2018b). Structure and energy based quantitative missense variant effect analysis provides insights into drug resistance mechanisms of anaplastic lymphoma kinase mutations. *Sci Rep*, 8:10664 10 1038 41598–018–28752–9.

[246] Li, J., Knoll, S., Bocharova, I., Tang, W., and Signorovitch, J. (2019). Comparative efficacy of first-line ceritinib and crizotinib in advanced or metastatic anaplastic lymphoma kinase-positive non-small cell lung cancer: an adjusted indirect comparison with external controls. *Curr Med Res Opin*, 35:105–11.

[247] Li, J., Zhao, P., wang, Y., zheng, C., Li, Y., and Tian, Y. (2015). Systems pharmacology-based approach for dissecting the active ingredients and potential targets of the chinese herbal bufei jianpi formula for the treatment of COPD. *International Journal of Chronic Obstructive Pulmonary Disease*, page 2633.

[248] Li, N. Y. K., Verdolini, K., Clermont, G., Mi, Q., Rubinstein, E. N., Hebda, P. A., and Vodovotz, Y. (2008). A patient-specific in silico model of inflammation and healing tested in acute vocal fold injury. *PLoS ONE*, 3(7):e2789.

[249] Lim, S., Kim, S., Kim, K., Jang, H., Ahn, S., Kim, K., Kim, N., Park, W., Lee, S., Kim, S., Park, S., Park, J., and Park, Y. (2017). The implication of flt3 amplification for flt targeted therapeutics in solid tumors. *Oncotarget*, 8:3237–45.

[250] Lin, J. J., Zhu, V. W., Yoda, S., Yeap, B. Y., Schrock, A. B., Dagogo-Jack, I., Jessop, N. A., Jiang, G. Y., Le, L. P., Gowen, K., Stephens, P. J., Ross, J. S., Ali, S. M., Miller, V. A., Johnson, M. L., Lovly, C. M., Hata, A. N., Gainor, J. F., Iafrate, A. J., Shaw, A. T., and Ou, S.-H. I. (2018). Impact of eml_4-ALK variant on resistance mechanisms and clinical outcomes in ALK-positive lung cancer. *Journal of Clinical Oncology*, 36(12):1199–1206.

[251] Lin, J.-S. and Lai, E.-M. (2017). Protein–protein interactions: Co-immunoprecipitation. In *Methods in Molecular Biology*, pages 211–219. Springer New York.

[252] Liu, D., Shen, X., Yuan, F., Guo, O., Zhong, Y., and Chen, J. (2015a). The physiology of bdnf and its relationship with adhd. *Mol Neurobiol.*, 52:1467–76.

[253] Liu, X., Su, L., and Liu, X. (2013). Loss of cdh1 up-regulates epidermal growth factor receptor via phosphorylation of ybx1 in non-small cell lung cancer cells. *FEBS Lett*, 587:3995–4000.

[254] Liu, Y., Morley, M., Brandimarto, J., Hannenhalli, S., Hu, Y., and Ashley, E. (2015b). Rna-seq identifies novel myocardial gene expression signatures of heart failure. *Genomics*, 105:83–9. pmid: 25528681.

[255] Liu, Y., Yang, X., Sharov, V., Nass, O., Sabbah, H., and Peterson, E. (1997). Effects of angiotensin-converting enzyme inhibitors and angiotensin ii type 1 receptor antagonists in rats with heart failure: Role of kinins and angiotensin ii type 2 receptors. *J Clin Invest*, 99:1926–35. pmid: 9109437.

[256] Loke, H., Harley, V., and Lee, J. (2015). Biological factors underlying sex differences in neurological disorders: focus on sry. *Int J Biochem Cell Biol*, 65:139–50.

[257] Lorén, V., Garcia-Jaraquemada, A., Naves, J., Carmona, X., Mañosa, M., and Aransay, A. (2019). Anp32e, a protein involved in steroid-refractoriness in ulcerative colitis, identified by a systems biology approach. *J Crohn's Colitis*, 13:351–361. pmid: 30329026.

[258] M, O., AW, O., E, B., A, S., C, N., and J, S. (2007). Development of a who growth reference for school-aged children and adolescents. *Bull World Health Organ.*, 85:660–7.

[259] Madden, J. C., Enoch, S. J., Paini, A., and Cronin, M. T. (2020). A review of in silico tools as alternatives to animal testing: Principles, resources and applications. *SAGE Publications*, 48(4):146–172.

[260] Madsen, H. and Thyregod, P. (2011). A generalized linear model with binomial distribution and probit link function has been used as classifier. In *Introduction to General and Generalized Linear Models.* Chapman & Hall/CRC.

[261] Maldonado, R. (2013). Comparison of the pharmacokinetics and clinical efficacy of new extended-release formulations of methylphenidate. *Expert Opin Drug Metab Toxicol.*, 9:1001–14.

[262] Mannuzza, S., Klein, R., Bessler, A., Malloy, P., and LaPadula, M. (1998). Adult psychiatric status of hyperactive boys grown up. *Am J Psychiatry*, 155:493–8.

[263] Manolis, E., Rohou, S., Hemmings, R., Salmonson, T., Karlsson, M., and Milligan, P. (2013). The role of modeling and simulation in development and registration of medicinal products: Output from the EFPIA/EMA modeling and simulation workshop. *CPT: Pharmacometrics & Systems Pharmacology*, 2(2):31.

[264] Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., and Califano, A. (2006). ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7(S1).

[265] Marshall, S., Hemmings, R., Josephson, F., Karlsson, M., Posch, M., and Steimer, J.-L. (2013). Modeling and simulation to optimize the design and analysis of confirmatory trials, characterize risk-benefit, and support label claims. *CPT Pharmacometrics Syst Pharmacol*, page 2 27.

[266] Marsit, C., Zheng, S., Aldape, K., Hinds, P., Nelson, H., Wiencke, J., and Kelsey, K. (2005). Pten expression in non-small-cell lung cancer: evaluating its relation to tumor characteristics, allelic loss, and epigenetic alteration. *Hum Pathol*, 36:768–76.

[267] Marson, F. A. L. (2017). Personalized or precision medicine? the example of cystic fibrosis. *Frontiers in Pharmacology*, 8.

[268] Martin, S. A., McLanahan, E. D., Bushnell, P. J., Hunter, E. S., and El-Masri, H. (2014). Species extrapolation of life-stage physiologically-based pharmacokinetic (PBPK) models to investigate the developmental toxicology of ethanol using in vitro to in vivo (IVIVE) methods. *Toxicological Sciences*, 143(2):512–535.

[269] Martín-Segura, A., Casadomé-Perales, Á., Fazzari, P., Mas, J. M., Artigas, L., Valls, R., Nebreda, A. R., and Dotti, C. G. (2019). Aging increases hippocampal DUSP2 by a membrane cholesterol loss-mediated RTK/p38mapk activation mechanism. *Frontiers in Neurology*, 10.

[270] Matheus, R., Bernardi, F., Gallo, C., Silva, A., Rodrigues, O., Capelozzi, M., Lopes, A., Fenezelian, S., Saldiva, P., and Capelozzi, V. (2004). Nuclear markers (star volume, mitotic index, agnor and ki-67) of the primary tumor and its metastasis in non-small cell lung carcinomas. *Pathol Res Pract*, 200:13–23.

[271] MATLAB (2017). *version 9.2.0.538062 (R2017a)*. The MathWorks Inc, Natick, Massachusetts.

[272] McCaffrey, T. A., Laurent, G. S., Shtokalo, D., Antonets, D., Vyatkin, Y., Jones, D., Battison, E., and Nigg, J. T. (2020). Biomarker discovery in attention deficit hyperactivity disorder: RNA sequencing of whole blood in discordant twin and case-controlled cohorts. *BMC Medical Genomics*, 13(1).

[273] McGough, J. and Barkley, R. (2004). Diagnostic controversies in adult attention deficit hyperactivity disorder. *Am J Psychiatry*, 161:1948–56.

[274] McMurray, J., Packer, M., Desai, A., Gong, J., Lefkowitz, M., and Rizkala, A. (2014). Angiotensin–neprilysin inhibition versus enalapril in heart failure. *N Engl J Med*, 371:993–1004. pmid: 25176015.

[275] Messaoudi, S., Azibani, F., Delcayre, C., and Jaisser, F. (2012). Aldosterone, mineralocorticoid receptor, and heart failure. *Mol Cell Endocrinol*, 350:266–72. pmid: 21784127.

[276] Mezquita, L. and Planchard, D. (2018). The role of brigatinib in crizotinib-resistant non-small cell lung cancer. *Cancer Manag Res*, 10:123–30.

[277] Milligan, P. A., Brown, M. J., Marchant, B., Martin, S. W., van der Graaf, P. H., Benson, N., Nucci, G., Nichols, D. J., Boyd, R. A., Mandema, J. W., Krishnaswami, S., Zwillich, S., Gruben, D., Anziano, R. J., Stock, T. C., and Lalonde, R. L. (2013). Model-based drug development: A rational approach to efficiently accelerate drug development. *Clinical Pharmacology & Therapeutics*, 93(6):502–514.

[278] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827.

[279] Miyata, M., Ikeda, Y., Nakamura, S., Sasaki, T., Abe, S., and Minagoe, S. (2012). Effects of valsartan on fibrinolysis in hypertensive patients with metabolic syndrome. *Circ J*, 76:843–51. pmid: 22451451.

[280] Mok, T., Spigel, D., Felip, E., Marinis, F., Ahn, M., Groen, H., Wakelee, H., Hida, T., Crino, L., Nishio, M., Scagliotti, G., Branle, F., and Emeremni, C. (2015). Ascend-2: A single-arm, open-label, multicenter phase ii study of ceritinib in adult patients (pts) with alk-rearranged (alk+) non-small cell lung cancer (nsclc) previously treated with chemotherapy and crizotinib (crz. *Journal of Clinical Oncology*, 33:8059 10 1200 2015 33 15 8059.

[281] Morris, T., Khoo, C., and Solomon, B. (2019). Targeting ros1 rearrangements in non-small cell lung cancer: Crizotinib and newer generation tyrosine kinase inhibitors. *Drugs*, 79:1277–86 10 1007 40265–019–01164–3.

[282] Mou, H., Moore, J., Malonia, S., Li, Y., Ozata, D., Hough, S., Song, C., Smith, J., Fischer, A., Weng, Z., Green, M., and Xue, W. (2017). Genetic disruption of oncogenic kras sensitizes lung cancer cells to fas receptor-mediated apoptosis. *Proc Natl Acad Sci U S A*, 114:3648–53.

[283] Mucke, H. A. and Mucke, E. (2015). Sources and targets for drug repurposing: Landscaping transitions in therapeutic space. *ASSAY and Drug Development Technologies*, 13(6):319–324.

[284] Mukherjee, S., Tamayo, P., Rogers, S., Rifkin, R., Engle, A., and Campbell, C. (2003). Estimating dataset size requirements for classifying dna microarray data. *J Comput Biol*, 10:119–42.

[285] Murphy, C., Fullington, H., Alvarez, C., Betts, A., Lee, S., Haggstrom, D., and Halm, E. (2018). Polypharmacy and patterns of prescription medication use among cancer survivors. *Cancer*, 124:2850–57.

[286] Nassar, K., Grisanti, S., Elfar, E., Lüke, J., Lüke, M., and Grisanti, S. (2015). Serum cytokines as biomarkers for age-related macular degeneration. *Graefe's Arch Clin Exp Ophthalmol*, 253:699–704. pmid: 25056526.

[287] Nguyen, T., Mouksassi, M., Holford, N., Al-Huniti, N., Freedman, I., and Hooker, A. (2017). Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT Pharmacometrics Syst Pharmacol*, 6:87–109.

[288] Ni, W., Chen, W., and Lu, Y. (2018). Emerging findings into molecular mechanism of brain metastasis. *Cancer Med*, 7:3820–33.

[289] Nicholas, T., Duvvuri, S., Leurent, C., Raunig, D., Rapp, T., Iredale, P., Rowinski, C., Carr, R., Roberts, P., Spiros, A., and Geerts, H. (2013). Systems pharmacology modeling in neuroscience: Prediction and outcome of PF-04995274, a 5-HT4 partial agonist, in a clinical scopolamine impairment trial. *Advances in Alzheimer's Disease*, 02(03):83–98.

[290] Nijsen, M. J., Wu, F., Bansal, L., Bradshaw-Pierce, E., Chan, J. R., Liederer, B. M., Mettetal, J. T., Schroeder, P., Schuck, E., Tsai, A., Xu, C., Chimalakonda, A., Le, K., Penney, M., Topp, B., Yamada, A., and Spilker, M. E. (2018). Preclinical QSP modeling in the pharmaceutical industry: An IQ consortium survey examining the current landscape. *CPT: Pharmacometrics & Systems Pharmacology*, 7(3):135–146.

[291] Novello, S., Mazières, J., Oh, I., Castro, J., Migliorino, M., Helland, Å., Dziadziuszko, R., Griesinger, F., Kotb, A., Zeaiter, A., Cardona, A., Balas, B., and Johannsdottir, H. (2018). Alectinib versus chemotherapy in crizotinib-pretreated anaplastic lymphoma kinase (alk)-positive non-small-cell lung cancer: results from the phase iii alur study. *Ann Oncol*, 29:1409–16.

[292] Nowak, J. (2014). Amd-the retinal disease with an unprecised etiopathogenesis: In search of effective therapeutics. *Acta Pol Pharm—Drug Res*, 71:900–16.

[293] NRC (2011). *Toward Precision Medicine. National Research Council*. National Academies Press.

[294] Nurwidya, F., Andarini, S., Takahashi, F., Syahruddin, E., and Takahashi, K. (2016). Implications of insulin-like growth factor 1 receptor activation in lung cancer. *Malays J Med Sci*, 23:9–21.

[295] O, E. O., M, E. K., S, G., K, B., Y, W., and A, Ö. (2020). Human cry1 variants associate with attention deficit/hyperactivity disorder. *J Clin Invest*, 130:3885–900.

[296] Ohnishi, T., Kobayashi, H., Yajima, T., Koyama, T., and Noguchi, K. (2019). Psychiatric comorbidities in adult attention-deficit/hyperactivity disorder: prevalence and patterns in the routine clinical setting. *Innov Clin Neurosci.*, 16:11–6.

[297] Ohno-Matsui, K. (2011). Parallel findings in age-related macular degeneration and alzheimer's disease. *Prog Retin Eye Res*, 30:217–238. pmid: 21440663.

[298] Olivas, E. S., Guerrero, J. D. M., Sober, M. M., Benedito, J. R. M., and López, A. J. S. (2009). *Handbook Of Research On Machine Learning Applications and Trends: Algorithms, Methods and Techniques (2 Volumes)*. Information Science Reference - Imprint of: IGI Publishing, 701 E. Chocolate Avenue, Suite 200, Hershey, PA.

[299] Oprea, T. I. and Mestres, J. (2012). Drug repurposing: Far beyond new targets for old drugs. *The AAPS Journal*, 14(4):759–763.

[300] Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N. H., Chavali, G., Chen, C., del Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R. C., Meldal, B., Melidoni, A. N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G., and Hermjakob, H. (2013). The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, 42(D1):D358–D363.

[301] Oubiña, M., Heras, N., Vázquez-Pérez, S., Cediel, E., Sanz-Rosa, D., and Ruilope, L. (2002). Valsartan improves fibrinolytic balance in atherosclerotic rabbits. *J Hypertens*, 20:303–10. pmid: 11821716.

[302] Palmér, R., Nyman, E., Penney, M., Marley, A., Cedersund, G., and Agoram, B. (2014). Effects of IL-1β-blocking therapies in type 2 diabetes mellitus: A quantitative systems pharmacology modeling

approach to explore underlying mechanisms. *CPT: Pharmacometrics & Systems Pharmacology*, 3(6):118.

[303] Pang, X., Min, J., Liu, L., Liu, Y., Ma, N., and Zhang, H. (2012). S100b protein as a possible participant in the brain metastasis of nsclc. *Med Oncol*, 29:2626–32.

[304] Pappalardo, F., Russo, G., Tshinanu, F. M., and Viceconti, M. (2018). In silico clinical trials: concepts and early adoptions. *Briefings in Bioinformatics*, 20(5):1699–1708.

[305] Park, S., Kim, B., Kim, J., Shin, M., Cho, S., and Kim, J. (2014). Neurotrophin 3 genotype and emotional adverse effects of osmotic-release oral system methylphenidate (oros-mph) in children with attention-deficit/ hyperactivity disorder. *J Psychopharmacol*, 28:220–6.

[306] Passini, E., Britton, O. J., Lu, H. R., Rohrbacher, J., Hermans, A. N., Gallacher, D. J., Greig, R. J. H., Bueno-Orovio, A., and Rodriguez, B. (2017). Human in silico drug trials demonstrate higher accuracy than animal models in predicting clinical pro-arrhythmic cardiotoxicity. *Frontiers in Physiology*, 8.

[307] Patel, V., Wang, Z., Fan, D., Zhabyeyev, P., Basu, R., and Das, S. (2013). Loss of p47phox subunit enhances susceptibility to biomechanical stress and heart failure because of dysregulation of cortactin and actin filaments. *Circ Res*, 112:1542–56. pmid: 23553616.

[308] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2012). Scikit-learn: Machine learning in python. *CoRR*, abs/1201.0490.

[309] Perera, S., Artigas, L., Mulet, R., Mas, J., and Sardón, T. (2014). Systems biology applied to non-alcoholic fatty liver disease (nafld): treatment selection based on the mechanism of action of nutraceuticals. *Nutrafoods*, 13:61–68.

[310] Peters, S. (2008). Evaluation of a generic physiologically based pharmacokinetic model for lineshape analysis. *Clin Pharmacokinet*, 47:261–75.

[311] Peters, S., Camidge, D., Shaw, A., Gadgeel, S., Ahn, J., Kim, D., Ou, S., Pérol, M., Dziadziuszko, R., Rosell, R., Zeaiter, A., Mitry, E., and Golding, S. (2017). Alectinib versus crizotinib in untreated alk-positive non–small-cell lung cancer. *N Engl J Med*, 377:829–38.

[312] Peters, S. A. (2012). *Physiologically-Based Pharmacokinetic (PBPK) Modeling and Simulations: Principles, Methods, and Applications in the Pharmaceutical Industry*. John Wiley & Sons, Inc.

[313] Pickering, R., Tikellis, C., Rosado, C., Tsorotes, D., Dimitropoulos, A., and Smith, M. (2019). Transactivation of rage mediates angiotensin-induced inflammation and atherogenesis. *J Clin Invest*, 129:406–421. pmid: 30530993.

[314] Piñero, J., Berenstein, A., Gonzalez-Perez, A., Chernomoretz, A., and Furlong, L. I. (2016). Uncovering disease mechanisms through network biology in the era of next generation sequencing. *Scientific Reports*, 6(1).

[315] Polak, S., Fijorek, K., Glinka, A., Wisniowska, B., and Mendyk, A. (2012). Virtual population generator for human cardiomyocytes parameters: in silico drug cardiotoxicity assessment. toxicol mech methods.

[316] Polanczyk, G., Willcutt, E., Salum, G., Kieling, C., and Rohde, L. (2014). Adhd. prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *Int J Epidemiol*, 43:434–42.

[317] Ponikowski, P., Voors, A., Anker, S., Bueno, H., Cleland, J., and Coats, A. (2016). 2016 esc guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*, 37:2129–2200 27206819.

[318] Popper, H. (2016). Progression and metastasis of lung cancer. *Cancer Metastasis Rev*, 35:75–91 10 1007 10555–016–9618–0.

[319] Prasad, T. S. K., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D. S., Sebastian, A., Rani, S., Ray, S., Kishore, C. J. H., Kanth, S., Ahmed, M., Kashyap, M. K., Mohmood, R., Ramachandra, Y. L., Krishna, V., Rahiman, B. A., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R., and Pandey, A. (2009). Human protein reference database–2009 update. *Nucleic Acids Research*, 37(Database):D767–D772.

[320] Preusser, M., Berghoff, A., Berger, W., Ilhan-Mutlu, A., Dinhof, C., Widhalm, G., Dieckmann, K., Wöhrer, A., Hackl, M., Deimling, A., Streubel, B., and Birner, P. (2014). High rate of fgfr1 amplifications

in brain metastases of squamous and non-squamous lung cancer. *Lung Cancer*, 83:83–89.

[321] Pujol, A., Mosca, R., Farrés, J., and Aloy, P. (2010). Unveiling the role of network and systems biology in drug discovery. *Trends in Pharmacological Sciences*, 31(3):115–123.

[322] Quinn, P. (2005). Treating adolescent girls and women with adhd: gender-specific issues. *J Clin Psychol*, 61:579–87.

[323] Rai, S., Raj, U., and Varadwaj, P. K. (2018). Systems biology: A powerful tool for drug development. *Current Topics in Medicinal Chemistry*, 18(20):1745–1754.

[324] Rangachari, D., Yamaguchi, N., VanderLaan, P., Folch, E., Mahadevan, A., Floyd, SR, U., EJ, W., ET, D., SE, H., MS, C., and D.B. (2015). Brain metastases in patients with egfr-mutated or alk-rearranged non-small-cell lung cancers. *Lung Cancer*, 88:108–11.

[325] Rashdan, S., Gerber, D., Jain, R., and Chen, H. (2017). A crowded, but still varied, space: brigatinib in anaplastic lymphoma kinase-rearranged non-small cell lung cancer. *Transl Cancer Res*, 6:78–82.

[326] Reckamp, K., Lin, H., Huang, J., Proskorovsky, I., Reichmann, W., Krotneva, S., Kerstein, D., Huang, H., and Lee, J. (2019). Comparative efficacy of brigatinib versus ceritinib and alectinib in patients with crizotinib-refractory anaplastic lymphoma kinase-positive non-small cell lung cancer. *Curr Med Res Opin*, 35:569–76.

[327] Reddy, M., Yang, R., Andersen, M., and Clewell, H. (2005). *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. Wiley.

[328] Rejniak, K. A. and McCawley, L. J. (2010). Current trends in mathematical modeling of tumor–microenvironment interactions: a survey of tools and applications. *Experimental Biology and Medicine*, 235(4):411–423.

[329] Retz, W., Rösler, M., Ose, C., Scherag, A., Alm, B., and Philipsen, A. (2012). Multiscale assessment of treatment efficacy in adults with adhd: a randomized placebo-controlled, multi-centre study with extended-release methylphenidate. *World J Biol Psychyatry*, 13:48–59.

[330] Ricciuti, B., Giglio, A., Mecca, C., Arcuri, C., Marini, S., Metro, G., Baglivo, S., Sidoni, A., Bellezza, G., Crinò, L., and Chiari, R. (2018). Precision medicine against alk-positive non-small cell lung cancer: beyond crizotinib. *Med Oncol*, 35:72 10 1007 12032–018–1133–4.

[331] Riddell, E. and Vader, J. (2017). Potential expanded indications for neprilysin inhibitors. current heart failure reports. pmid: 28281174.

[332] Rieger, T. R. and Musante, C. J. (2016). Benefits and challenges of a QSP approach through case study: Evaluation of a hypothetical GLP-1/GIP dual agonist therapy. *European Journal of Pharmaceutical Sciences*, 94:15–19.

[333] Rivas, J. D. L. and Fontanillo, C. (2010). Protein–protein interactions essentials: Key concepts to building and analyzing interactome networks. *PLoS Computational Biology*, 6(6):e1000807.

[334] Roberts, P., Spiros, A., and Geerts, H. (2016). A humanized clinically calibrated quantitative systems pharmacology model for hypokinetic motor symptoms in parkinson's disease. *Frontiers in Pharmacology*, 7.

[335] Rodenhizer, D., Dean, T., D'Arcangelo, E., and McGuigan, A. P. (2018). The current landscape of 3d in vitro tumor models: What cancer hallmarks are accessible for drug discovery? *Advanced Healthcare Materials*, 7(8):1701174.

[336] Rodig, S., Mino-Kenudson, M., Dacic, S., Yeap, B., Shaw, A., Barletta, J., Stubbs, H., Law, K., Lindeman, N., Mark, E., Janne, P., Lynch, T., and Johnson, B. (2009). Unique clinicopathologic features characterize alk-rearranged lung adenocarcinoma in the western population. *Clin Cancer Res*, 15:5216–23 10 1158 1078–0432 –09–0802.

[337] Rogers, K. V., Martin, S. W., Bhattacharya, I., Singh, R. S. P., and Nayak, S. (2020). A dynamic quantitative systems pharmacology model of inflammatory bowel disease: Part 1 – model framework. *Clinical and Translational Science*, 14(1):239–248.

[338] Rogers, M., Lyster, P., and Okita, R. (2013). NIH support for the emergence of quantitative and systems pharmacology. *CPT: Pharmacometrics & Systems Pharmacology*, 2(4):37.

[339] Romeo-Guitart, D., Forés, J., Herrando-Grabulosa, M., Valls, R., Leiva-Rodríguez, T., Galea, E., González-Pérez, F., Navarro, X., Petegnief, V., Bosch, A., Coma, M., Mas, J. M., and Casas, C.

(2018). Neuroprotective drug for nerve trauma revealed using artificial intelligence. *Scientific Reports*, 8(1).

[340] Roncaglioni, A., Toropov, A. A., Toropova, A. P., and Benfenati, E. (2013). In silico methods to predict drug toxicity. *Current Opinion in Pharmacology*, 13(5):802–806. Anti-infectives * New technologies.

[341] Rostami-Hodjegan, A. (2012). Physiologically based pharmacokinetics joined with in vitro-in vivo extrapolation of adme: a marriage under the arch of systems pharmacology. clin pharmacol ther.

[342] Rothenstein, J. and Chooback, N. (2018). Alk inhibitors, resistance development, clinical trials. *Curr Oncol*, 25:59–67.

[343] Rotow, J. and Bivona, T. (2017). Understanding and targeting resistance mechanisms in nsclc. *Nat Rev Cancer*, 17:637–58.

[344] Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*, 20:53–65.

[345] Ruiz, P., Emond, C., McLanahan, E., Joshi-Barr, S., and Mumtaz, M. (2020). Exploring mechanistic toxicity of mixtures using pbpk modeling and computational systems biology. *Toxicol Sci.*, 174:38–50.

[346] Russell, S. J. and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach (2nd Edition)*. Pearson Education.

[347] Sabari, J. K., Santini, F., Schram, A. M., Bergagnini, I., Chen, R., Mrad, C., Lai, W. V., Arbour, K. C., and Drilon, A. (2017). The activity, safety, and evolving role of brigatinib in patients with *ALK*-rearranged non-small cell lung cancers. *OncoTargets and Therapy*, Volume 10:1983–1992.

[348] Sager, J. E., Yu, J., Ragueneau-Majlessi, I., and Isoherranen, N. (2015). Physiologically based pharmacokinetic (PBPK) modeling and simulation approaches: A systematic review of published models, applications, and model verification. *Drug Metabolism and Disposition*, 43(11):1823–1837.

[349] Salwinski, L. (2004). The database of interacting proteins: 2004 update. *Nucleic Acids Research*, 32(90001):449D–451.

[350] Schadt, E. E. and Björkegren, J. L. M. (2012). NEW: Network-enabled wisdom in biology, medicine, and health care. *Science Translational Medicine*, 4(115).

[351] Schadt, E. E., Friend, S. H., and Shaywitz, D. A. (2009). A network view of disease and compound screening. *Nature Reviews Drug Discovery*, 8(4):286–295.

[352] Schiavon, M., Visentin, R., Giegerich, C., Sieber, J., Dalla Man, C., and Cobelli, C. (2020). In silico head-to-head comparison of insulin glargine 300 u/ml and insulin degludec 100 u/ml in type 1 diabetes. *Diabetes Technol Ther.*, 22:553–61.

[353] Schiestl, R., Bardot, D., and Myers, K. (2014). Computer modeling and simulation project update. mdic 2014 annual meeting. http://mdic.org/wp-content/uploads/2014/06/computer-modeling-simulation-cms-project-update.pdf.

[354] Schmidt, B. J., Casey, F. P., Paterson, T., and Chan, J. R. (2013). Alternate virtual populations elucidate the type i interferon signature predictive of the response to rituximab in rheumatoid arthritis. *BMC Bioinformatics*, 14(1).

[355] Schrier, R., Abdallah, J., Weinberger, H., and Abraham, W. (2000). Therapy of heart failure. *Kidney Int*, 57:1418–25. pmid: 10760077.

[356] Shaw, A., Kim, T., Crinò, L., Gridelli, C., Kiura, K., Liu, G., Novello, S., Bearz, A., Gautschi, O., Mok, T., Nishio, M., Scagliotti, G., and Spigel, D. (2017). Ceritinib versus chemotherapy in patients with alk-rearranged non-small-cell lung cancer previously given chemotherapy and crizotinib (ascend-5): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol*, 18:874–86 10 1016 1470–2045 17 30339–.

[357] Shen, Y. and Bax, A. (2007). Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *Journal of Biomolecular NMR*, 38(4):289–302.

[358] Shimizu, K., Short, D. A., and Kedem, B. (1993). Single- and double-threshold methods for estimating the variance of area rain rate. *Journal of the Meteorological Society of Japan. Ser. II*, 71(6):673–683.

[359] Singh, J., Burrell, L., Cherif, M., Squire, I., Clark, A., and Lang, C. (2017). Sacubitril/valsartan: Beyond natriuretic peptides. *Heart*, 103:1569–1577. pmid: 28689178.

[360] Sinha, S. and Vohora, D. (2018). Chapter 2 - drug discovery and development: An overview. In Vohora, D. and Singh, G., editors, *Pharmaceutical Medicine and Translational Clinical Research*, pages 19–32. Academic Press, Boston.

[361] Sinisi, S., Alimguzhin, V., Mancini, T., Tronci, E., and Leeners, B. (2020). Complete populations of virtual patients for in silico clinical trials. *Bioinformatics*, 36(22-23):5465–5472.

[362] Skeie, J., Zeng, S., Faidley, E., and Mullins, R. (2011). Angiogenin in age-related macular degeneration. *Mol Vis*, 17:576–82. pmid: 21364907.

[363] SL, A. B. (2018). Biological effectors database.

[364] Smietana, K., Siatkowski, M., and Møller, M. (2016). Trends in clinical success rates. *Nature Reviews Drug Discovery*, 15(6):379–380.

[365] Snelder, N., Ploeger, B. A., Luttringer, O., Rigel, D. F., Webb, R. L., Feldman, D., Fu, F., Beil, M., Jin, L., Stanski, D. R., and Danhof, M. (2016). Characterization and prediction of cardiovascular effects of fingolimod and siponimod using a systems pharmacology modeling approach. *Journal of Pharmacology and Experimental Therapeutics*, 360(2):356–367.

[366] Sohn, Y., Lee, N., Chung, A., Saavedra, J., Scott Turner, R., and Pak, D. (2013). Antihypertensive drug valsartan promotes dendritic spine density by altering ampa receptor trafficking. *Biochem Biophys Res Commun*, 439:464–70. pmid: 24012668.

[367] Solomon, B., Mok, T., Kim, D., Wu, Y., Nakagawa, K., Mekhail, T., Felip, E., Cappuzzo, F., Paolini, J., Usari, T., Iyer, S., Reisman, A., and Wilner, K. (2014). First-line crizotinib versus chemotherapy in alk-positive lung cancer. *N Engl J Med*, 371:2167–77.

[368] Solomon, S., Rizkala, A., Gong, J., Wang, W., Anand, I., and Ge, J. (2017). Angiotensin receptor neprilysin inhibition in heart failure with preserved ejection fraction: Rationale and design of the paragon-hf trial. *JACC Hear Fail*, 5:471–482. pmid: 28662936.

[369] Song, J., Lee, S., Hong, J., Chang, S., Choe, H., and Choi, J. (2009). Epidermal growth factor competes with egf receptor inhibitors to induce cell death in egfr-overexpressing tumor cells. *Cancer Lett*, 283:135–42.

[370] Sonuga-Barke, E. and Halperin, J. (2010). Developmental phenotypes and causal pathways in attention deficithyperactivity disorder: potential targets for early intervention? *J Child Psychol Psychiatry*, 51:368–89.

[371] Soria, J., Tan, D., Chiari, R., Wu, Y., Paz-Ares, L., Wolf, J., Geater, S., Orlov, S., Cortinovis, D., Yu, C., Hochmair, M., Cortot, A., and

Tsai, C. (2017). First-line ceritinib versus platinum-based chemotherapy in advanced alk-rearranged non-small-cell lung cancer (ascend-4): a randomised, open-label, phase 3 study. *Lancet*, 389:917–29 10 1016 0140–6736 17 30123–.

[372] Spencer, T. J., Adler, L. A., McGough, J. J., Muniz, R., Jiang, H., and Pestreich, L. (2007). Efficacy and safety of dexmethylphenidate extended-release capsules in adults with attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 61(12):1380–1387.

[373] Spencer, T. J., Wilens, T. E., Biederman, J., Weisler, R. H., Read, S. C., and Pratt, R. (2006). Efficacy and safety of mixed amphetamine salts extended release (adderall XR) in the management of attention-deficit/hyperactivity disorder in adolescent patients: A 4-week, randomized, double-blind, placebo-controlled, parallel-group study. *Clinical Therapeutics*, 28(2):266–279.

[374] Spiros, A., Roberts, P., and Geerts, H. (2014). A computer-based quantitative systems pharmacology model of negative symptoms in schizophrenia: exploring glycine modulation of excitation-inhibition balance. *Frontiers in Pharmacology*, 5.

[375] Stanicka, J., Rieger, L., O'Shea, S., Cox, O., Coleman, M., O'Flanagan, C., Addario, B., McCabe, N., Kennedy, R., and O'Connor, R. (2018). Fes-related tyrosine kinase activates the insulin-like growth factor-1 receptor at sites of cell adhesion. *Oncogene*, 37:3131–50 10 1038 41388–017–0113–.

[376] Stein, M., Waldman, I., Charney, E., Aryal, S., Sable, C., and Gruber, R. (2011). Dose effects and comparative effectiveness of extended release dexmethylphenidate and mixed amphetamine salts. *J Child Adolesc Psychopharmacol*, 21:581–8.

[377] Strovel, J., Sittampalam, S., Coussens, N., Hughes, M., Inglese, J., Kurtz, A., Andalibi, A., Patton, L., Austin, C., Baltezor, M., Beckloff, M., Weingarten, M., and Weir, S. (2012). *Early Drug Discovery and Development Guidelines: For Academic Researchers, Collaborators, and Start-up Companies*. Eli Lilly & Company and the National Center for Advancing Translational Sciences.

[378] Sullivan, I. and Planchard, D. (2016). Alk inhibitors in non-small cell lung cancer: the latest evidence and developments. *Ther Adv Med Oncol*, 8:32–47.

[379] Sullivan, R., WoldeMussie, E., and Pow, D. (2007). Dendritic and synaptic plasticity of neurons in the human age-related macular degeneration retina. *Investig Ophthalmol Vis Sci*, 48:2782–91. pmid: 17525213.

[380] Sun, D., Li, X., Ma, M., Liu, J., Xu, Y., Ye, L., Hou, H., Wang, C., Li, X., and Jiang, Y. (2015). The predictive value and potential mechanisms of miRNA-328 and miRNA-378 for brain metastases in operable and advanced non-small-cell lung cancer. *Japanese Journal of Clinical Oncology*, 45(5):464–473.

[381] Szklarczyk, D., Santos, A., Mering, C., Jensen, L., Bork, P., and Kuhn, M. (2016). Stitch 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res*, 44:380–84.

[382] Takeuchi, K., Soda, M., Togashi, Y., Suzuki, R., Sakata, S., Hatano, S., Asaka, R., Hamanaka, W., Ninomiya, H., Uehara, H., Lim Choi, Y., Satoh, Y., and Okumura, S. (2012). Ret, ros1 and alk fusions in lung cancer. *Nat Med*, 18:378–81.

[383] Tan, J., Li, M., Zhong, W., Hu, C., Gu, Q., and Xie, Y. (2017). Tyrosine kinase inhibitors show different anti-brain metastases efficacy in NSCLC: A direct comparative analysis of icotinib, gefitinib, and erlotinib in a nude mouse model. *Oncotarget*, 8(58):98771–98781.

[384] Tan, Y.-M., Worley, R. R., Leonard, J. A., and Fisher, J. W. (2018). Challenges associated with applying physiologically based pharmacokinetic modeling for public health decision-making. *Toxicological Sciences*, 162(2):341–348.

[385] Tavassoly, I., Goldfarb, J., and Iyengar, R. (2018). Systems biology primer: the basic methods and approaches. *Essays in Biochemistry*, 62(4):487–500.

[386] Teeguarden, J. G., Housand, C. J., Smith, J. N., Hinderliter, P. M., Gunawan, R., and Timchalk, C. A. (2013). A multi-route model of nicotine–cotinine pharmacokinetics, pharmacodynamics and brain nicotinic acetylcholine receptor binding in humans. *Regulatory Toxicology and Pharmacology*, 65(1):12–28.

[387] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

[388] Timchalk, C. (2002). A physiologically based pharmacokinetic and pharmacodynamic (PBPK/PD) model for the organophosphate insecticide chlorpyrifos in rats and humans. *Toxicological Sciences*, 66(1):34–53.

[389] Torre, L., Bray, F., Siegel, R., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA Cancer J Clin*, 65(3322/caac.21262):87–108.

[390] Trisilowati, M. D. (2012). In silico experimental modeling of cancer treatment. *ISRN Oncol*, page 828701.

[391] Tseng, R.-C., Lee, S.-H., Hsu, H.-S., Chen, B.-H., Tsai, W.-C., Tzao, C., and Wang, Y.-C. (2010). SLIT2 attenuation during lung cancer progression deregulates β-catenin and e-cadherin and associates with poor prognosis. *Cancer Research*, 70(2):543–551.

[392] Tugnait, M., Gupta, N., Hanley, M., Sonnichsen, D., Kerstein, D., Dorer, D., Venkatakrishnan, K., and Narasimhan, N. (2020). Effects of strong cyp2c8 or cyp3a inhibition and cyp3a induction on the pharmacokinetics of brigatinib, an oral anaplastic lymphoma kinase inhibitor, in healthy volunteers. *Clin Pharmacol Drug Dev*, 9:214–23.

[393] Turner, N. (2014). Effects of interleukin-1 on cardiac fibroblast function: Relevance to post-myocardial infarction remodelling. *Vascul Pharmacol*, 60:1–7. pmid: 23806284.

[394] Uchibori, K., Inase, N., Araki, M., Kamada, M., Sato, S., Okuno, Y., Fujita, N., and Katayama, R. (2017). Brigatinib combined with anti-egfr antibody overcomes osimertinib resistance in egfr-mutated non-small-cell lung cancer. *Nat Commun*, 8:14768.

[395] Umehara, K., Huth, F., Jin, Y., Schiller, H., Aslanis, V., Heimbach, T., and He, H. (2019). Drug-drug interaction (DDI) assessments of ruxolitinib, a dual substrate of CYP3a4 and CYP2c9, using a verified physiologically based pharmacokinetic (PBPK) model to support regulatory submissions. *Drug Metabolism and Personalized Therapy*, 34(2).

[396] US (n.d.). Agriculture, rural development, food and drug administration, and related agencies appropriations bill, 2016. https://www.congress.gov/114/crpt/srpt82/crpt-114srpt82.pdf.

[397] Vavalà, T. and Novello, S. (2018). Alectinib in the treatment of alk-positive non-small cell lung cancer: an update on its properties,

efficacy, safety and place in therapy. *Ther Adv Med Oncol*, 10:1758835918789364.

[398] Verlaet, A., Breynaert, A., Ceulemans, B., Bruyne, T., Fransen, E., and Pieters, L. (2019). Oxidative stress and immune aberrancies in attention-deficit/hyperactivity disorder (adhd): a case-control comparison. *Eur Child Adolesc Psychiatry*, 28:719–29.

[399] Verlaet, A., Noriega, D., Hermans, N., and Savelkoul, H. (2014). Nutrition, immunological mechanisms and dietary immunomodulation in adhd. eur child adolesc psychiatry.

[400] Verweij, N., Eppinga, R., Hagemeijer, Y., and Harst, P. (2017). Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. *Sci Rep*, 7:2761 28584231.

[401] Viceconti, M. and Clapworthy, G. (2006). The virtual physiological human: challenges and opportunities. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, pages 812–815.

[402] Viceconti, M., Cobelli, C., Haddad, T., Himes, A., Kovatchev, B., and Palmer, M. (2017). In silico assessment of biomedical products: The conundrum of rare but not so rare events in two case studies. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 231(5):455–466.

[403] Viceconti, M., Henney, A., and Morley-Fletcher, E. (2026). *in silico* clinical trials: How computer simulation will transform the biomedical industry. *Research and Technological Development Roadmap, Avicenna Consortium, Brussels*.

[404] Viceconti, M. and McCulloch, A. D. (2011). Policy needs and options for a common approach towards modelling and simulation of human physiology and diseases with a focus on the virtual physiological human. *Studies in Health Technology and Informatics*, 170:49–82.

[405] Vidal, M., Cusick, M. E., and Barabási, A.-L. (2011). Interactome networks and human disease. *Cell*, 144(6):986–998.

[406] Vidal, M. and Fields, S. (2014). The yeast two-hybrid assay: still finding connections after 25 years. *Nature Methods*, 11(12):1203–1206.

[407] Visentin, R., Dalla Man, C., Kovatchev, B., and Cobelli, C. (2014). The university of virginia/padova type 1 diabetes simulator matches the glucose traces of a clinical trial. diabetes technol ther.

[408] Visentin, R., Man, C., and Cobelli, C. (2016). One-day bayesian cloning of type 1 diabetes subjects: toward a single-day uva/padova type 1 diabetes simulator. *IEEE Trans Biomed Eng*, 63:2416–24.

[409] Visser, S., Norton, J., Marshall, S., and O'Kelly, M. (2018). Common best practice in modeling and simulation across quantitative disciplines: a comparison of independently emerging proposals. *Stat Biopharm Res*, 10:72–5.

[410] Visser, S. A. G., de Alwis, D. P., Kerbusch, T., Stone, J. A., and Allerheiligen, S. R. B. (2014). Implementation of quantitative and systems pharmacology in large pharma. *CPT: Pharmacometrics & Systems Pharmacology*, 3(10):142.

[411] Vodovotz, Y., Csete, M., Bartels, J., Chang, S., and An, G. (2008). Translational systems biology of inflammation. *PLoS Computational Biology*, 4(4):e1000014.

[412] VPH (accessed 2022). Virtual physiological human institute for integrative biomedical research. http://www.vph-institute.org/.

[413] Wagner, C., Pan, Y., Hsu, V., Grillo, J. A., Zhang, L., Reynolds, K. S., Sinha, V., and Zhao, P. (2014). Predicting the effect of cytochrome p450 inhibitors on substrate drugs: Analysis of physiologically based pharmacokinetic modeling submissions to the US food and drug administration. *Clinical Pharmacokinetics*, 54(1):117–127.

[414] Wagner, J. (1968). Kinetics of pharmacologic response i. proposed relationships between response and drug concentration in the intact animal and man. *Journal of Theoretical Biology*, 20(2):173–201.

[415] Wang, H., Milberg, O., Bartelink, I., Vicini, P., Wang, B., and Narwal, R. (2019). In silico simulation of a clinical trial with anti-ctla-4 and anti-pd-l1 immunotherapies in metastatic breast cancer using a systems pharmacology model. *R Soc Open Sci*, page 6.

[416] Wang, H., Sové, R., Jafarnejad, M., Rahmeh, S., Jaffee, E., and Stearns, V. (2020). Conducting a virtual clinical trial in her2-negative breast cancer using a quantitative systems pharmacology model with an epigenetic modulator and immune checkpoint inhibitors. *Front Bioeng Biotechnol.*, 8:141.

[417] Wang, R., Maron, B., and Loscalzo, J. (2015). Systems medicine: evolution of systems biology from bench to bedside. *Wiley Interdiscip Rev Syst Biol Med.*, 7:141–61.

[418] Wang, X., Jiao, W., Zhao, Y., Zhang, L., Yao, R., Wang, Y., Wang, M., Luo, Y., and Zhao, J. (2016). Cug-binding protein 1 (cugbp1) expression and prognosis of brain metastases from non-small cell lung cancer. *Thorac Cancer*, 7:32–38 10 1111 1759–7714 12268.

[419] Wang, X., Ma, W., Han, S., Meng, Z., Zhao, L., Yin, Y., Wang, Y., and Li, J. (2017). TGF-β participates choroid neovascularization through smad2/3-VEGF/TNF-α signaling in mice with laser-induced wet age-related macular degeneration. *Scientific Reports*, 7(1).

[420] Wang, X., Wang, H., Li, G., Song, Y., Wang, S., Zhu, F., Guo, C., Zhang, L., and Shi, Y. (2014). Activated macrophages down-regulate expression of e-cadherin in hepatocellular carcinoma cells via NF–κb/slug pathway. *Tumor Biology*, 35(9):8893–8901.

[421] Warsito, D., Sjöström, S., Andersson, S., Larsson, O., and Sehat, B. (2012). Nuclear igf1r is a transcriptional co-activator of lef1/tcf. *EMBO Rep*, 13:244–50.

[422] Weisler, R. H., Greenbaum, M., Arnold, V., Yu, M., Yan, B., Jaffee, M., and Robertson, B. (2017). Efficacy and safety of SHP465 mixed amphetamine salts in the treatment of attention-deficit/hyperactivity disorder in adults: Results of a randomized, double-blind, placebo-controlled, forced-dose clinical study. *CNS Drugs*, 31(8):685–697.

[423] Wigal, T., Brams, M., Gasior, M., Gao, J., Squires, L., and and, J. G. (2010). Randomized, double-blind, placebo-controlled, crossover study of the efficacy and safety of lisdexamfetamine dimesylate in adults with attention-deficit/hyperactivity disorder: novel findings using a simulated adult workplace environment design. *Behavioral and Brain Functions*, 6(1):34.

[424] Wilens, T. E. (2004). Attention-deficit/hyperactivity disorder in adults. *JAMA*, 292(5):619.

[425] Wishart, D., Feunang, Y., Guo, A., Lo, E., Marcu, A., Grant, JR, S., T, J., D, L., C, S., Z, A., N, I., I, L., and Y. (2018). Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res*, 46:1074–82.

[426] Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2007). DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research*, 36(suppl_1):D901–D906.

[427] Wolf, R., Howard, O., Dong, H., Voscopoulos, C., Boeshans, K., Winston, J., Divi, R., Gunsior, M., Goldsmith, P., Ahvazi, B., Chavakis, T., Oppenheim, J., and Yuspa, S. (2008). Chemotactic activity of s100a7 (psoriasin) is mediated by the receptor for advanced glycation end products and potentiates inflammation with highly homologous but functionally distinct s100a15. *J Immunol*, 181:1499–506.

[428] Wolkenhauer, O., Auffray, C., Brass, O., Clairambault, J., Deutsch, A., and Drasdo, D. (2014). Enabling multiscale modeling in systems medicine. *Genome Med*, 6:21.

[429] Wong, M., Lao, X., Ho, K., Goggins, W., and Tse, S. (2017). Incidence and mortality of lung cancer: global trends and association with socioeconomic status. *Sci Rep*, 7:14300 10 1038 41598–017–14513–7.

[430] Woodhead, J. L., Watkins, P. B., Howell, B. A., Siler, S. Q., and Shoda, L. K. (2017). The role of quantitative systems pharmacology modeling in the prediction and explanation of idiosyncratic drug-induced liver injury. *Drug Metabolism and Pharmacokinetics*, 32(1):40–45.

[431] Workman, P., Aboagye, E., Balkwill, F., Balmain, A., Bruder, G., Chaplin, D., Double, J., Everitt, J., Farningham, D., Glennie, M., Kelland, L., Robinson, V., and Stratford, I. (2010). Guidelines for the welfare and use of animals in cancer research. *Br J Cancer*, 102:1555–77.

[432] Wu, F., Gaohua, L., Zhao, P., Jamei, M., Huang, S.-M., Bashaw, E. D., and Lee, S.-C. (2014). Predicting nonlinear pharmacokinetics of omeprazole enantiomers and racemic drug using physiologically based pharmacokinetic modeling and simulation: Application to predict drug/genetic interactions. *Pharmaceutical Research*, 31(8):1919–1929.

[433] Wu, F., Krishna, G., and Surapaneni, S. (2020). Physiologically based pharmacokinetic modeling to assess metabolic drug–drug interaction risks and inform the drug label for fedratinib. *Cancer Chemotherapy and Pharmacology*, 86(4):461–473.

[434] Xalkori (2019). European medicines agency (ema). https://www.ema.europa.eu/en/medicines/human/epar/xalkori.

[435] Xalkori (n.d.). List of cleared or approved companion diagnostic devices (in vitro and imaging tools). https://www.fda.gov/medical-devices/in-vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-in-vitro-and-imaging-tools.

[436] Xu, G., Strathearn, L., Liu, B., Yang, B., and Bao, W. (2018). Twenty-year trends in diagnosed attention-deficit/hyperactivity disorder among us children and adolescents, 1997-2016. *JAMA Netw Open*, page 1 181471.

[437] Xu, J., Winter, F., Farrokhi, C., Rockenstein, E., Mante, M., and Adame, A. (2016). Neuregulin 1 improves cognitive deficits and neuropathology in an alzheimer's disease model. *Sci Rep*, 6:31692 27558862.

[438] Xuan, G., Zhu, X., Chai, P., Zhang, Z., Shi, Y. Q., and Fu, D. (2006). Feature selection based on the bhattacharyya distance. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 3, pages 1232–1235.

[439] Yoshida, T., Oya, Y., Tanaka, K., Shimizu, J., Horio, Y., Kuroda, H., Sakao, Y., Hida, T., and Yatabe, Y. (2016). Clinical impact of crizotinib on central nervous system progression in alk-positive non-small lung cancer. *Lung Cancer*, 97:43–47.

[440] Yuan, X., Gu, X., Crabb, J., Yue, X., Shadrach, K., and Hollyfield, J. (2010). Quantitative proteomics: Comparison of the macular bruch membrane/choroid complex from age-related macular degeneration and normal eyes. *Mol Cell Proteomics*, 9:1031–46. pmid: 20177130.

[441] Zalsman, G. and Shilton, T. (2016). Adult adhd: a new disease? *Int J Psychiatry Clin Pract*, 20:70–6.

[442] Zaman, A., French, C., Schneider, D., and Sobel, B. (2009). A profibrotic effect of plasminogen activator inhibitor type-1 (pai-1) in the heart. *Exp Biol Med*, 234:246–54. pmid: 19144865.

[443] Zhang, I., Zaorsky, N., Palmer, J., Mehra, R., and Lu, B. (2015a). Targeting brain metastases in alk-rearranged non-small-cell lung cancer. *Lancet Oncol*, 16:e510–21:10 1016 1470–2045 15 00013–3.

[444] Zhang, R., Liu, Z., Zhang, H., Zhang, Y., and Lin, D. (2016a). The cox-2-selective antagonist (ns-398) inhibits choroidal neovascularization and subretinal fibrosis. *PLoS One*, 11:0146808 26760305.

[445] Zhang, S., Anjum, R., Squillace, R., Nadworny, S., Zhou, T., Keats, J., Ning, Y., Wardwell, S., Miller, D., Song, Y., Eichinger, L., Moran, L., and Huang, W. (2016b). The potent alk inhibitor brigatinib (ap26113) overcomes mechanisms of resistance to first- and second-generation alk inhibitors in preclinical models. *Clin Cancer Res*, 22:5527–38 10 1158 1078–0432 –16–0569. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm395299.htm.

[446] Zhang, X., Trame, M., Lesko, L., and Schmidt, S. (2015b). Sobol sensitivity analysis: A tool to guide the development and evaluation of systems pharmacology models. *CPT Pharmacometrics Syst Pharmacol*, 4:69–79.

[447] Zhang, Y., Vernooij, F., Ibrahim, I., Ooi, S., Gijsberts, C., and Schoneveld, A. (2016c). Extracellular vesicle proteins associated with systemic vascular events correlate with heart failure: An observational study in a dyspnoea cohort. *PLoS One*, 11:0148073. pmid: 26820481.

[448] Zhang, Z. l., Li, R., Yang, F., and Xi, L. (2018). Natriuretic peptide family as diagnostic/prognostic biomarker and treatment modality in management of adult and geriatric patients with heart failure: Remaining issues and challenges. *J Geriatr Cardiol*, 15:540–546. pmid: 30344534.

[449] Zhao, J., Zou, M., Lv, J., Han, Y., Wang, G., and Wang, G. (2018). Effective treatment of pulmonary adenocarcinoma harboring triple EGFR mutations of l858r, t790m, and *cis*-c797s by osimertinib, bevacizumab, and brigatinib combination therapy: a case report. *OncoTargets and Therapy*, Volume 11:5545–5550.

[450] Zou, H. and Hastie, T. (2005). Addendum: Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):768–768.

[451] Zuo, Y., Cui, Y., Poto, C. D., Varghese, R. S., Yu, G., Li, R., and Ressom, H. W. (2016). INDEED: Integrated differential expression and differential network analysis of omic data for biomarker discovery. *Methods*, 111:12–20.

[452] Zykadia (2019). European medicines agency (EMA). Internet. https://www.ema.europa.eu/en/medicines/human/EPAR/zykadia.

Appendix A

Annex

Aside from the publications mentioned in the text, and as part of the DRIVE project, I also participated in the elaboration of the following publication:

Muhammed Kocak, Saba Ezazi Erdi, Guillem Jorba, Inés Maestro, Judith Farrés, Vladimir Kirkin, Ana Martinez & Ole Pless (2022) Targeting autophagy in disease: established and new strategies, Autophagy, 18:3, 473-495, DOI: 10.1080/15548627.2021.1936359

Abstract

Macroautophagy/autophagy is an evolutionarily conserved pathway responsible for clearing cytosolic aggregated proteins, damaged organelles or invading microorganisms. Dysfunctional autophagy leads to pathological accumulation of the cargo, which has been linked to a range of human diseases, including neurodegenerative diseases, infectious and autoimmune diseases and various forms of cancer. Cumulative work in animal models, application of genetic tools and pharmacologically active compounds, has suggested the potential therapeutic value of autophagy modulation in disease, as diverse as Huntington, Salmonella infection, or pancreatic cancer. Autophagy activation versus inhibition strategies are being explored, while the role of autophagy in pathophysiology is being studied in parallel. However, the progress of preclinical and clinical development of autophagy modulators has been greatly hampered by the paucity of selective pharmacological agents and biomarkers to dissect their precise impact on various forms of autophagy and cellular responses. Here, we summarize established and new strategies in autophagy-related drug discovery and indicate a path toward establishing a more efficient discovery

of autophagy-selective pharmacological agents. With this knowledge at hand, modern concepts for therapeutic exploitation of autophagy might become more plausible.Abbreviations: ALS: amyotrophic lateral sclerosis; AMPK: AMP-activated protein kinase; ATG: autophagy-related gene; AUTAC: autophagy-targeting chimera; CNS: central nervous system; CQ: chloroquine; GABARAP: gamma-aminobutyric acid type A receptor-associated protein; HCQ: hydroxychloroquine; LYTAC: lysosome targeting chimera; MAP1LC3/LC3: microtubule associated protein 1 light chain 3; MTOR: mechanistic target of rapamycin kinase; NDD: neurodegenerative disease; PDAC: pancreatic ductal adenocarcinoma; PE: phosphatidylethanolamine; PIK3C3/VPS34: phosphatidylinositol 3-kinase catalytic subunit type 3; PtdIns3K: class III phosphatidylinositol 3-kinase; PtdIns3P: phosphatidylinositol 3-phosphate; PROTAC: proteolysis-targeting chimera; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; SQSTM1/p62: sequestosome 1; ULK1: unc-51 like autophagy activating kinase 1.