UNIVERSITAT DE BARCELONA

A Systems Medicine approach to multimorbidity

Towards personalised care for patients with Chronic Obstructive Pulmonary Disease

Ákos Tényi

UNIVERSITAT DE BARCELONA

DOCTORAL THESIS

A Systems Medicine approach to multimorbidity

Towards personalised care for patients with Chronic Obstructive Pulmonary Disease

Ákos Tényi

2018

DOCTORAL PROGRAM IN MEDICINE AND TRANSLATIONAL RESEARCH

Supervised by

Dr. Isaac Cano Franco and Prof. Josep Roca Torrent

"Research is to see what everybody else has seen,

and to think what nobody else has thought."

ALBERT SZENT-GYÖRGYI

(Nobel Prize in Physiology-Medicine, 1937)

Memoria de la tesis doctoral presentada por

Ákos Tényi

para optar al grado de doctor en Medicina e Investigación Traslacional

Trabajo realizado bajo la dirección de

Dr. Isaac Cano Franco
Prof. Josep Roca Torrent

Tesis inscrita en el programa de Doctorado de Medicina e Investigación Traslacional
Facultad de Medicina,
Universidad de Barcelona

Barcelona 2018

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my advisor Prof. Josep Roca for the continuous support, feedback and intellectual guidance that helped me in all the time of the research and writing of this PhD thesis. Without his immense knowledge and patience, this thesis would not have been possible.

I would like to thank to my co-director Isaac Cano and close collaborators David Gomez-Cabrero and Marta Cascante for their support, insightful comments and their scientific expertise that always helped to get through major obstacles during the writing of the thesis. I give my gratitude to Dieter Maier for his professional guidance during my first steps in the scientific world.

This thesis could not have been done without all the fun and adventures with my friends that always helped me to reignite inner creativity and draw motivation to move forward. I am very grateful for my parents for all the values and qualities that they have instilled in me and for their constant support and encouragement throughout the entire thesis. I would like to express my deepest gratitude to my Sylvia for her whole-hearted emotional and professional support and for being my main inspirations in the last push of the work.

GRANTS

This PhD thesis was partly supported by:

- Synergy-COPD (Modeling and Simulation Environment for Systems Medicine: Chronic Obstructive Pulmonary Disease (COPD) as a use case): grant FP7-ICT-2011-270086.

- FIS PI15/00576 - Despliegue de servicios innovadores para la atención personalizada de enfermedades crónicas en un marco de salud digital - PITES-TliSS ISCIII 2016-18.

- NEXTCARE (Innovation in Integrated Care Services for Chronic Patients): grant COMRDI15-1-0016- 2016).

- CONNECARE (Personalised Connected Care for Complex Chronic Patients): grant H2020, GA nº 689802

CONTENTS

LIST OF FIGURES

LIST OF ACRONYMS

AHA	American Heart Association
CAT	COPD Assessment Test
CDSS	Clinical Decision Support System
COPD	Chronic Obstructive Pulmonary Disease
COPDkb	COPD Knowledge Base
CPM	Clinical Predictive Modelling
CVD	Cardiovascular Disorders
CRG	Clinical Risk Groups
DSS	Decision Support System
EHR	Electronic Health Records
FAIR	Findable, Accessible, Interoperable, Reusable
FEV1	Forced Expiratory Volume in One Second
FVC	Forced Vital Capacity
GDPR	European General Data Protection Regulation
GMA	Adjusted Morbidity Groups
GOLD	Global Initiative for Obstructive Lung Disease
GRN	Gene Regulatory Network
HPC	High Performance Computing
ICD	International Classification of Disease
ICT	Information and Communication Technologies
LHS	Learning Healthcare System
mMRC	modified Medical Research Council
NCD	Non-Communicable Diseases
PDSS	Patient Decision Support System
PPI	Protein-Protein Interaction
T2DM	Type 2 Diabetes Mellitus

ABSTRACT

Multimorbidity (i.e. the presence of more than one chronic disease in the same patient) and comorbidity (i.e. the presence of more than one chronic disease in the presence of an index disease) are main sources of healthcare dysfunction in chronic patients and avoidable costs in conventional health systems worldwide. By affecting the majority of the population in Western societies, multimorbidity prompts the need for revisiting the single disease approach followed by contemporary clinical practice and elaborate treatments that target shared mechanisms of associated diseases with the potential of decelerating or even halting multimorbid disease progression. However, our current understanding on disease interactions is rather limited; and, although many disorders have been associated based on their shared molecular traits and their observed co-occurrence in different populations, no comprehensive approach has been outlined to translate this knowledge into clinical practice

This PhD thesis aims to explore multimorbidity from a systems medicine perspective on the specific use-case of chronic obstructive pulmonary disease (COPD), with the outlook of generalising these methods to a broader set of chronic respiratory diseases, and other non-communicable diseases. COPD is a major cause of morbidity and mortality worldwide, and its disease manifestations often involves non-pulmonary effects, including highly prevalent comorbidities, such as type 2 diabetes and cardiovascular diseases, and systemic effects.

The thesis investigates molecular disturbances in the skeletal muscle of patients with COPD and their body systems level interactions to identify signature biological pathways that potentially play key role in systemic effects and comorbidities. Furthermore, the thesis analyses population level patterns of COPD comorbidities and investigates their role in the health risk of patients with COPD, indicating its major negative impact on highly relevant clinical events, use of healthcare resources and prognosis. Finally, the thesis identifies personalized health risk prediction and service selection as potential tools for the integration and transfer of scientific evidence on multimorbidity to daily clinical practice and explores real-world implementation strategies in cloud-based environments.

The thesis outcomes indicate the need for a novel, systems perspective on patients with COPD that considers non-pulmonary manifestations both at the staging and the management of the disease. Moreover, the thesis provides specific actionable insights for the development of innovative interventions targeting the prevention of non-pulmonary manifestations. The application of the outcomes of the thesis has a credible potential to contribute to personalized care for chronic patients.

INTRODUCTION

1. THE EVOLVING SCENARIO

Multimorbidity (i.e. the presence of more than one chronic disease in the same patient) and comorbidity (i.e. the presence of additional chronic diseases in the presence of an index disease) are main sources of dysfunctions and avoidable costs in conventional health systems worldwide [1, 2].

The demonstrated population-wide patterns of disease co-occurrence [3–5], pathophysiological linkage of certain diseases [4, 6, 7] and the sheer size of the problem , i.e. multimorbidity affects the majority of elderly worldwide [8], prompts the need for revisiting the single disease approach of contemporary clinical practice [8], which presumes independence among comorbid conditions. Research and development of novel treatments that target the source of interactions among associated diseases are needed with the potential of decelerating or even halting multimorbid disease progression. In this context, several questions arise and need to be answered for the successful clinical application of multimorbidity principles: Which conditions are associated? What is the nature of their relation (e.g. causal, common environmental factor)? What is the molecular cause of the relation? How to translate this information into daily clinical practice?

Since the early 2000s, two key phenomena are prompting substantial changes in both biomedical research and clinical management of comorbidity (**Figure 1**). Firstly, systems biology methodologies (i.e. 'omics' technologies, use of computational modelling, etc.) are being progressively embedded into medical practice shaping the practicalities of systems medicine [9–13]. This new field promises a novel approach to disease, shifting its definition from phenotypical signs and symptoms towards molecular subtypes (i.e. endotypes) of diseases, which is indispensable for precise characterization of disease relations and for the evaluation of shared mechanisms [14]. Simultaneously, digital health initiatives and wearable devices are facilitating access to an enormous amount of patient-related information from whole populations to personal levels, and state-of-the art computational models and machine learning tools demonstrate high potential for health prediction [15, 16, 25, 17–24]. Given the extremely long and expensive bench to clinics cycles of the biomedical sector, these technologies promise a fast-track approach where scientific evidence can support clinical care while simultaneously collected insights from daily clinical practice promote new scientific discoveries and optimize healthcare optimization [26].

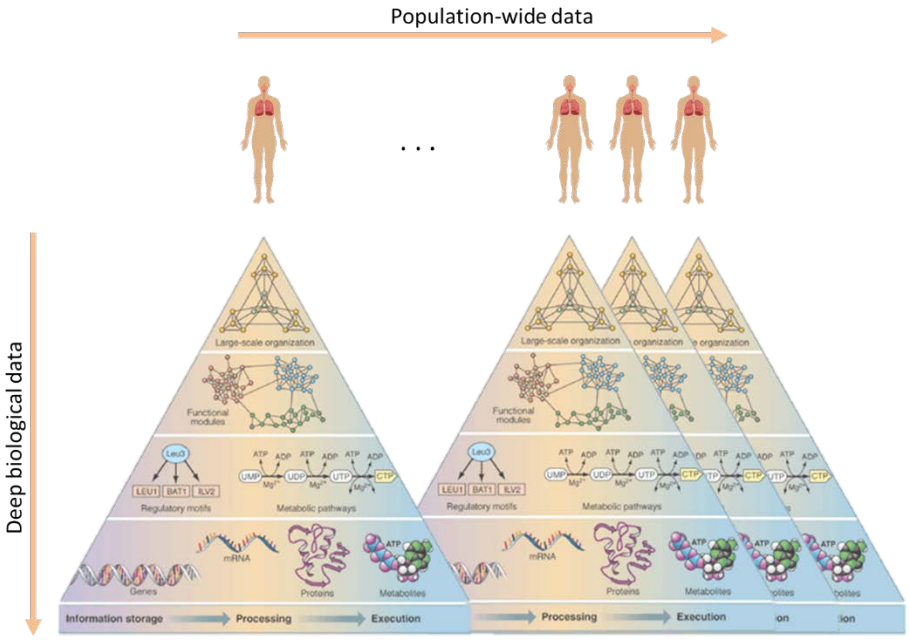


Figure 1. Data trends in systems medicine. Molecular measurement technologies and related systems biology methodologies are facilitating a deeper understanding of biological mechanisms. Meanwhile, digital health initiatives and wearable devices are facilitating access to enormous amount of patient-related information from whole populations (x-axis) and from different organs in the same patient (y-axis). Parts of the figure were adapted from [27].

This PhD thesis research work aims to explore multimorbidity from a systems medicine perspective on the concrete and practical use case of chronic obstructive pulmonary disease (COPD). COPD constitutes an ideal use case due to several factors, including: i) its high impact on healthcare and its ever-increasing burden; ii) its heterogeneous disease manifestations, and progress, often involving extra-pulmonary effects, including highly prevalent comorbidities (e.g. type 2 diabetes mellitus (T2DM), cardiovascular disorders (CVD), anxiety-depression and lung cancer); and, iii) its well described systemic effects with evidence suggesting associations with comorbidities in terms of underlying mechanisms.

The PhD thesis applies systems biology tools to analyse the underlying molecular mechanisms of COPD systemic effects that might partly explain disease heterogeneity. Furthermore, the PhD thesis aims to improve knowledge on the impact of comorbidities on healthcare and their potential role in strategies for health risk assessment and personalised medicine. The proposed approach should lead to generation of novel biomedical knowledge for the enhancement of patient stratification while exploring applicable strategies for enhanced patient management.

1.1 CONCEPTUALIZATION OF MULTIMORBIDITY

Burden and challenges

As alluded to above, multimorbidity generates an ever-increasing burden on healthcare systems world-wide. It affects the majority of individuals older than 65 years, such that in certain western populations the number of people living with multiple condition exceeds those living with only one [8] (**Figure 2**). Risk populations include females, people with lower socioeconomic status and those living with mental health problems [8, 28], indicating the multifactorial nature of the phenomenon. Multimorbidity has a major effect on patients' lives, most often manifested in the form of polypharmacy (i.e. the use of multiple drugs with potential unexpected adverse effects) [29, 30] and patients' frailty [31]. It also shows positive correlations with number of outpatient visits [32, 33] and consequently healthcare costs, such that some studies showed near exponential association between the number of co-existing conditions and healthcare use and costs [34, 35].

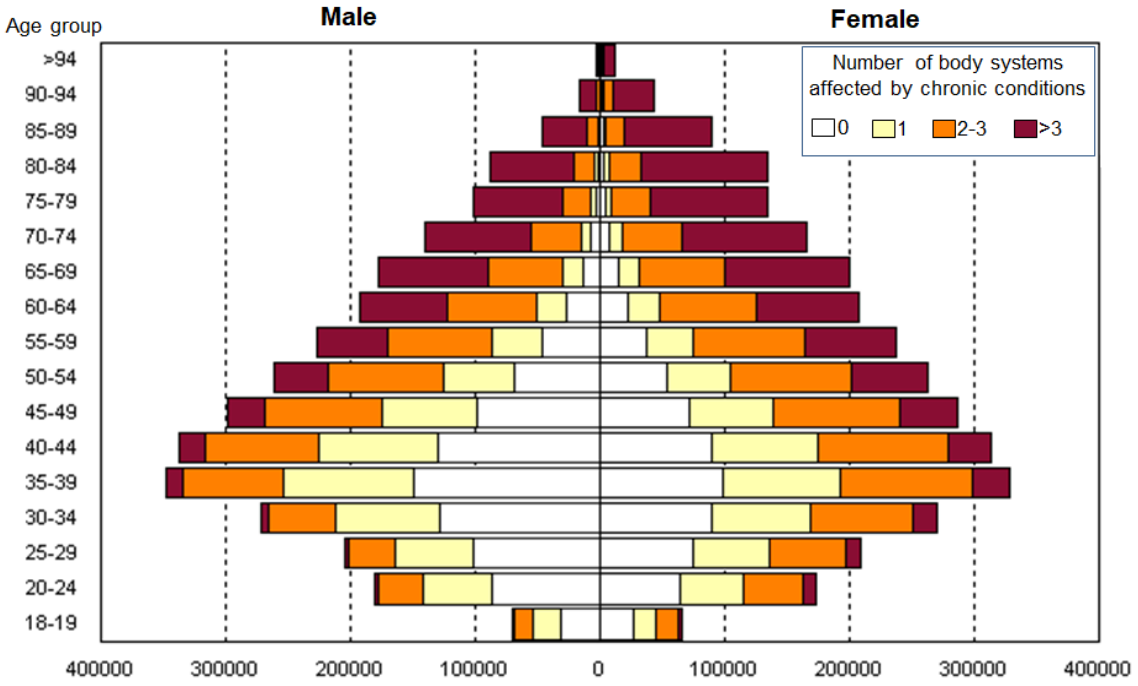


Figure 2. Multimorbidity prevalence in the Catalan region (Spain) (7.5M citizens). Number of males (left) and females (right) living with chronic conditions by age group. Bars are coloured by the number of body systems affected by chronic conditions. Data was retrieved form the Catalan Health Surveillance System in 2014.

Therefore, multimorbidity emerges as one of the main sources of dysfunctions and avoidable costs in conventional health systems worldwide [1, 2] and highlights a complex physiological phenomenon that is influenced by the lifelong interplay of genetic and environmental factors. The understanding of these complex interactions

needs novel conceptual frames and approaches that are able to accurately capture high biological complexity. However, there are several barriers in the current medical practice that are hindering the fine-grained understanding of the multimorbidity phenomenon.

One of the main barriers lies in the contemporary disease classification system [14]. The roots of this system date back to the 19th century and defines disease according to the observational correlation of pathological signs and clinical symptoms, often with an organ-centred approach. While two hundred years ago this system provided an acceptable base for patient management, our current understanding of disease shows that conditions with similar symptoms can arise from molecularly distinct mechanisms (e.g. spectrum of obstructive pulmonary diseases), leading to obsolete disease definitions. Furthermore, in the conventional approach, disease prevention is conceptually difficult, as disease can only be diagnosed when physical manifestations are already developed.

Another barrier lies in the traditional medical definition of comorbidity and multimorbidity, which refers to the terms rather as independently co-occurring conditions. This assumption is one of the main reasons behind the use of a single disease approach (i.e. independent treatment of each condition) in contemporary clinical practice. Whereas diseases can co-occur by chance, several recent studies have indicated the complex pairwise interaction of comorbidities, including non-random, population-wide co-occurrence patterns, as well as genetic and metabolic interactions [3, 4, 6, 7, 36]. Therefore, more recent models of multimorbidity already acknowledge that independent disease co-occurrence is only one type of relationship that can occur between two diseases [37]. Diseases can also arise interdependently owing to environmental exposures (e.g. smoking, diet, etc.) causing damage to multiple organs (e.g. COPD, lung cancer, CVD) [38–40] and also due to causative interactions among co-occurring conditions (e.g. skeletal muscle dysfunction, a systemic effect of CODP [41]) (**Figure 3**).

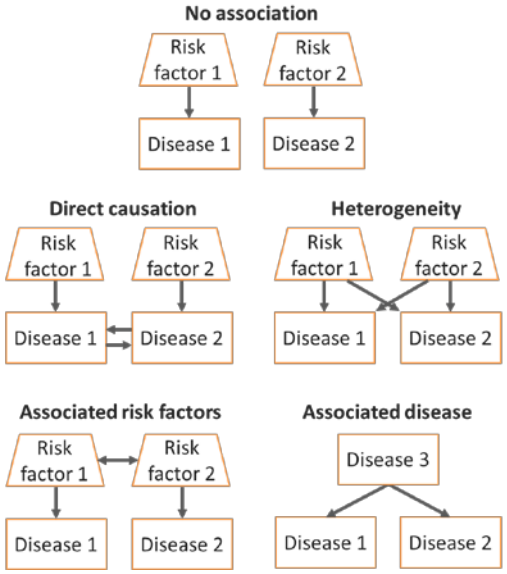


Figure 3. Recent models of multimorbidity. Valderas et al. describe five etiological models of comorbid diseases that relies on the interaction between diseases and risk factors [37]: i) no association: two disease co-occur by chance in the same patient; ii) direct causation: one disease may cause the other, e.g. COPD → depression; iii) heterogeneity: the risk factors for each disease are not correlated but each one of them can cause either disease, e.g. smoking, age and ischemic heart disease, lung cancer; iv) associated risk factors: The risk factors for each disease are correlated, e.g. smoking, alcohol and COPD, liver cirrhosis; v) associated disease: the presence of the diagnostic features of each diagnosis is due to a third distinct disease, e.g. hypertension (D1), tension headache (D2) and pheochromocytoma (D3). Figure was adapted from [37].

In summary, a novel approach to classify human diseases and to define disease interactions, based on the systemic understanding of the molecular underpinning of diseases, arise as a major unmet challenge in modern medicine [36] and an indispensable step for the accurate characterisation of the nature of comorbidity relations.

Systems concept of disease and comorbidity

The advent of novel measurement technologies (e.g. omics) has led to the emergence of novel research fields concentrating on the holistic understanding of biology (i.e. systems biology) and medicine (i.e. systems medicine) and recently the substantially reshaping of the view of disease and disease interactions. Systems biology is based on the observation that biological systems are inherently complex and often irreducible to the elementary properties of their individual components, thus the understanding of these concepts need a holistic approach, studying an organism as a whole living system. Systems medicine is the application of systems biology to medical research

and practice with the additional goal to integrate a variety of data at all relevant levels of cellular organization with clinical and patient-reported disease markers [9].

It is widely accepted that clinical phenotypes are rarely straight-forward consequences of an abnormality in a single gene, but rather reflect the interplay of multiple molecular processes. Thus, the identification of the core set of molecular elements and processes that underlie a disease (i.e. disease modules) have become the main objective of contemporary network medicine approaches [4, 36]. This shift of attention from clinical phenotypes to endotypes, i.e. molecular subtypes of the disease (pathophenotypes), is a major potential of network medicine that promises a more fine-grained disease classification as well as a new way to conceptualize the relation of comorbid conditions.

Figure 4 shows the different levels of abstraction in the systems model of disease. In this model, clinical expression of a disease can arise from several endotypes that are identified by disease modules, the molecular handprints of a disease [42]. Molecular elements of the modules (e.g. genes, proteins, metabolites) are part of a multi-level biological network that connects actively interacting molecular elements on the cellular level. The genotype gives the base topology of this network that due to the lifelong exposure to different environmental factors (e.g. smoking, diet), can be perturbed and rewired, i.e. it can change the active interactions. These changes are expressed on different biological levels (i.e. genetic, epigenetic, transcriptomic, post-transcriptional regulation, etc.) and result in altered cellular functions. Disease modules emerge when a system is not robust enough to balance perturbations and abnormal processes lead to phenotypic abnormalities [43].

In this systems model of disease, comorbidities can arise when specific genes, proteins, or metabolites participate in several disease modules, i.e. through overlapping disease modules. Thus, endotypes define the potential space for the interaction of diseases, where perturbations caused by a disease can potentially provoke other disorders. Whereas this is mainly conceptual model of comorbidity, a recent study demonstrated that in fact disease with overlapping disease modules display significantly more similar symptoms and co-occur more often than the ones that do not overlap [4], which also indicates the potential of the approach.

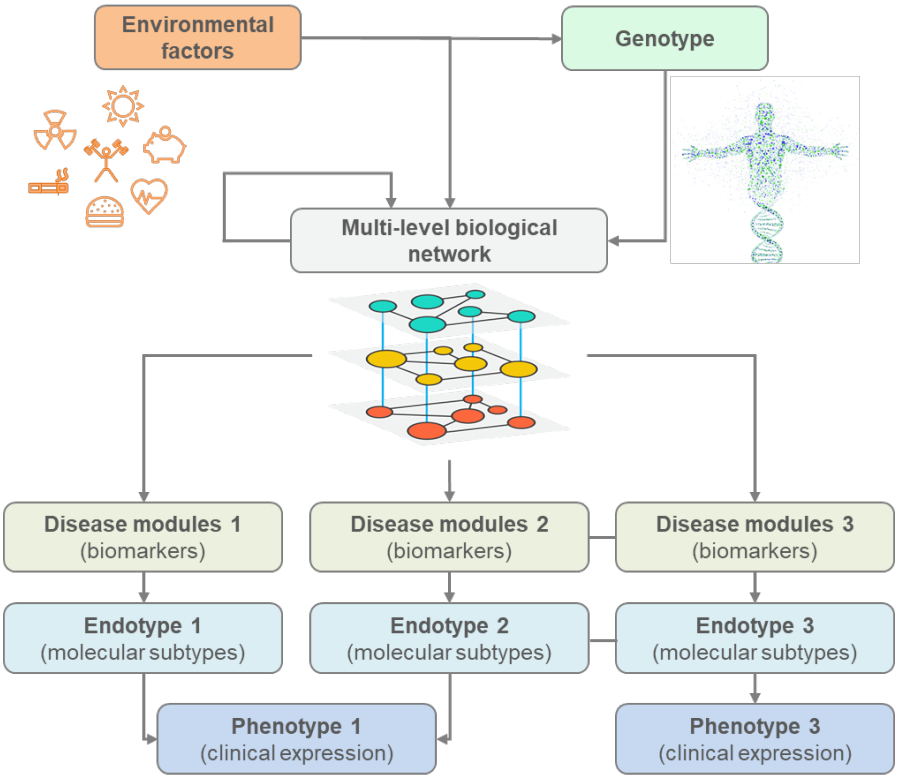


Figure 4. Systems model of disease. Different levels of biological conceptualization of disease, showing the associations between genes, environment, biomarkers, endotypes and phenotypes. The personal genotype defines the base interactions of one's interactome, i.e. the multi-level biological network that connects all actively interacting molecular elements on the cellular level. Continuous exposure to environmental factors shape both the genotype (e.g. mutations) and the interactome (e.g. rewiring) potentially leading to abnormal cellular regulatory processes, represented as disease modules in the interactome. Clinical expression of diseases (phenotypes) can arise from different molecular subtypes (endotypes), which in turn can be identified by a set of measurable biomarkers, i.e. representative elements of the disease modules. Comorbid conditions can arise from molecular interactions amongst different disease modules, which can be identified at the endotype level but not at the phenotype level.

1.2 COPD: A MODEL OF HETEROGENEOUS CHRONIC DISEASE

COPD is a major public health problem and it is one of the five major priorities of the non-communicable diseases (NCDs) policy of the World Health Organization, together with CVD, cancer, T2DM and mental disorders [44]. In 2010, COPD was responsible for three million deaths (6% of all deaths), which makes COPD the third leading cause of death worldwide, climbing two ranks higher from the fifth place in 1990 [45]. This trend is especially alarming, since COPD is one of the only disease with worsening death rates in Western societies (**Figure 5**) [1]. Projections on COPD prevalence and costs over the next fifteen years indicate a continuous escalating burden, mainly due to population

ageing and impact of comorbidities, on both health and social support systems [46, 47], such that in 2020 it is projected to rank fifth worldwide in terms of burden of disease [48].

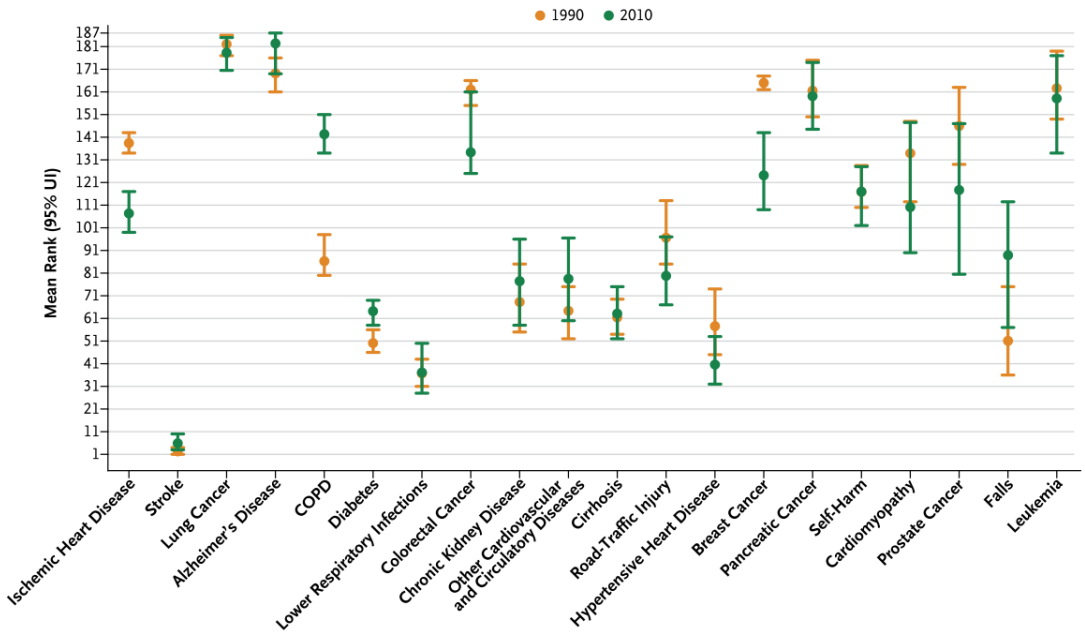


Figure 5. Age-standardized relative rate of death in the United States in 1990 and 2010, as ranked among 187 other countries. Ranks (y-axis) for the 20 leading cause of deaths in the US (x-axis) are shown. Rank 1 is the lowest age-standardized mortality across the 187 countries, and rank 187 is the highest. Bars indicate 95% uncertainty intervals (UIs). COPD denotes chronic obstructive pulmonary disease. This figure was adopted from [1].

COPD is a preventable and partially treatable disorder caused by inhalation of irritants, mainly tobacco smoking. However, only 15-20% of all tobacco smokers are prone to develop the disease. Early pulmonary-related manifestations of the disease are characterized by inflammatory phenomena of peripheral airways progressing to destruction of lung parenchyma (i.e. emphysema), abnormalities in pulmonary airways, inflammatory changes and remodelling in pulmonary circulation. The most characteristic functional finding in the clinical arena is expiratory flow limitation, due to reduced pulmonary elastance and increased airways resistance, assessed by a reduced Forced Expiratory Volume in one second (FEV_1) / Forced Vital Capacity (FVC)) ratio in forced spirometry testing [48].

Highly prevalent chronic conditions such as CVD, T2DM – metabolic syndrome and lung cancer are often co-occurring in patients with COPD. These comorbidities influence not only the severity of the symptoms and the quality of life of individual patients, but also the risk of hospitalization and eventually death [40, 48, 49]. Co-occurrence of several comorbid conditions at the same time in patients with COPD is a long standing

observation [50, 51], such that this comorbidity clustering has shown to be the major characteristic to be taken into account for management of patients with COPD [52]. There is evidence indicating that comorbidity clustering is only partly explained by shared risk factors (i.e. tobacco smoking, unhealthy diet and sedentarism) [50, 53, 54], highlighting the potential role of shared underlying biological mechanisms in the development of comorbidities.

Briefly, several concurrent facts are determining the interest of COPD as a use case representative of the challenges associated to NCDs; that is: i) the high prevalence of COPD, approximately 9% of the adult population above 40 years of age; ii) the elevated, and still steadily increasing, burden of patients with COPD on healthcare and social support services mainly due to ageing and high prevalence of co-morbid conditions; and, iii) the poor understanding of factors modulating heterogeneities both in terms of clinical manifestations and COPD progression due to insufficient knowledge on underlying mechanisms of both pulmonary and non-pulmonary manifestations of the disease.

The latter has important consequences in two relevant areas: i) generating potential problems in terms of COPD taxonomy overlaps with other pulmonary diseases presenting airflow limitation [55, 56]; and ii) poor knowledge on underlying mechanisms of patients with COPD limiting mechanism-oriented therapeutic strategies.

Current approaches to COPD heterogeneity

The Global initiative for Obstructive Lung Disease (GOLD) has played a key role in raising awareness of COPD and has been an important initiative towards the formulation of evidence-based medicine for care of patients with COPD. GOLD recommendations aim to provide expert advice on diagnosis and management of the disease worldwide, often through implementation of its recommendations into national guidelines [57].

Classification of patients with COPD into severity groups looking for predictive value has been one of the relevant aspects of GOLD recommendations. In the first two reports (2001, 2006), COPD staging was solely based on the alteration of FEV_1 in patients with low FEV_1/FVC ratio [58, 59]. The next two updates of GOLD recommendations (2011, 2017) supplemented the degree of airflow limitation (FEV_1) with assessment of symptoms (mMRC dyspnoea scale or the COPD Assessment Test (CAT)) and the risk of frequent exacerbations [48, 60].

The current GOLD recommendations [60] report disease severity based on FEV_1 impairment, that is: i) GOLD I – FEV_1 > 80% reference values – light disease; ii) GOLD II – FEV_1 between 50% and 80% - moderate disease; iii) GOLD III – FEV_1 between 30% and 50% FEV1 – severe disease; and, iv) GOLD IV – FEV_1 < 30% reference values - respiratory failure. Meanwhile, recommendations for pharmacological interventions are based on the intensity of symptoms and risk for frequent exacerbations of the pulmonary disease.

Non-pulmonary phenomena; that is, systemic effects and co-morbidities are mentioned in the GOLD recommendations as factors associated to a negative impact on patient prognosis deserving therapeutic interventions. There is mounting evidence indicating that impairment of the central organ (lung) in patients with COPD only partly explains disease prognosis [54]. The negative impact of comorbid conditions on prognosis is acknowledged since the 2006 GOLD report and it gained increasing attention in the recent reporting due to emerging evidence on its role in COPD heterogeneity [61]. Despite all this evidence, current guidelines still suggest that *"presence of comorbidities should not alter COPD treatment, and comorbidities should be treated per usual standards regardless of the presence of COPD"*. It is of note, however, that there is still confusion between systemic effects (conditions that are suspected to be in causal relationship with COPD) and comorbidities of the disease because of the descriptive nature of the classifications that are poorly based on deep knowledge of underlying mechanisms.

All in all, GOLD recommendations have represented a progress in standard of care of the pulmonary disorder, but it shows well accepted limitations for a comprehensive assessment of patients with COPD [62–66]. Several alternative approaches have emerged aiming at contributing to the prognostic assessment of these patients using alternative indices. Broadly used measures include prognostic indices combining several patient characteristics, such as BODE index (including body-mass index, airflow obstruction, dyspnoea, and exercise capacity) [67], DOSE index (including dyspnoea, obstruction, smoking, exacerbations) and the simplified ADO index (including age, dyspnoea, and airflow obstruction) [68, 69].

It is of note that some of the above indices specifically address prediction of mortality rather than staging of patients for management purposes. It is of note that general multimorbidity indices also have broad applicability in the field, showing a broad spectrum of possibilities. This includes simple diseases count, Charlson Index [70] or other, more complex indices like Clinical Risk Groups (CRG) [71] or Adjusted Morbidity

Groups (GMA) [72, 73], which evaluates prognosis based on age and weightings for specific comorbid conditions.

Challenges of COPD care

In summary, there are several factors that challenge our current understanding of COPD and that are needed to be overcome for a more personalized care of patients with COPD. One of the main challenges is the lack of our understanding of the underlying biological mechanisms of heterogeneous disease manifestations. There are several factors that may lead to disease heterogeneity and that need further attention and actions in the future.

Firstly, the simplistic diagnostic criteria of COPD (i.e. presence of respiratory symptoms like dyspnoea and cough, history of inhalation of irritants, and measurable airflow limitation) is one of the main sources of heterogeneity, often leading to overlapping diagnoses with different chronic obstructive airways diseases (e.g. asthma, bronchiectasis, bronchiolitis) [55] or to the exclusion of important emphysema-related disorders with similar treatment needs [74, 75]. This highlights the need for a better understanding of the key pathobiological mechanisms (endotypes) that drive the disease or its subtypes and that can function as potential therapeutic targets [76]. Their identification should lead to novel disease taxonomies and will likely result in the re-definition of COPD [77].

Secondly, contemporary lung-centric view of the disease often fails to explain the observed heterogeneity and other COPD specific phenomena, such as comorbidity clustering. This suggests a non-pulmonary component of the disease that needs a better understanding of the interplay between pulmonary and non-pulmonary manifestations. To date, the most supported hypothesis was that systemic inflammation induced by lung inflammatory processes can cause non-pulmonary effects seen in these patients [78–80]. Despite the potential relevance of systemic inflammation, many questions about its origin, mechanisms, and effects remain unanswered [81]. In fact, persistent systemic inflammation, associated with negative physiological effects, was shown to be present in only 16% of patient with COPD [82], which, as pointed out by Mohan and colleagues [83], is much less than the prevalence of certain co-occurring diseases. Furthermore, several recent studies discarded the link between low-grade systemic inflammation and cardiovascular or muscle manifestations in COPD [50, 83, 84], indicating that inflammation is unlikely to be the only factor relating these manifestations. Therefore, investigating underlying mechanisms and pathways shared by COPD, comorbidities and systemic effects [85] is

needed and should lead to better characterization of the disease and likely will have relevant implications on patient management [86].

Another major challenge is the descriptive nature of current classification of non-pulmonary manifestations in COPD. There is a marked confusion among systemic effects, complications and comorbidities. For example, is anxiety-depression a systemic effect, a COPD complication or a co-morbid condition? We lack knowledge on how comorbidities develop and interact with COPD and with each other, concerning the patients' wellbeing and survival. This indicates the need for a better understanding of these factors including gaining insight into the causal links between COPD and its comorbidities, which would require both a population level view of disease co-occurrence and molecular level evidences on the shared molecular mechanisms.

Finally, current tools for predicting disease severity have not yet been fully validated, while others have shown relatively low performance, suggesting the need for novel strategies for risk assessment and service selection including cost-effective strategies to prevent and stop spread of comorbidities in patients with COPD [73].

Overall, there is a strong rationale for a systems approach in COPD research, i.e. understanding the disease with all its components rather than concentrating only on the pulmonary axis. A better understanding of COPD heterogeneity [87] should permit the development and implementation of personalised therapeutic strategies that are specific to subgroups of patients, as well as the development of novel therapies [88], leading to a significant decrease of disease burden.

In summary, current COPD care is in great need of innovative approaches that help to overcome the long-running stagnation of this field. Better characterization of disease severity, novel tools for patient classification and prediction of survival and other health related outcomes, as well as novel therapies, are necessary to enhance COPD care and to progress towards a subject-specific risk prediction and stratification for personalized management of patients [65, 89].

1.3 SYNERGY-COPD: A SYSTEMS MEDICINE APPROACH TO COPD HETEROGENEITY

Synergy-COPD (2011-2014) [90] was a European Union project within the Virtual Physiological Human call of the 7th Framework programme (FP7-ICT-270086) tackling the main challenges of current COPD care. The unsuccessful attempts of traditional clinical approaches to answer the major questions regarding COPD heterogeneity were major motivations for Synergy-COPD to consider a holistic approach of the disease, focusing on non-pulmonary effects and their underlying mechanisms. The central

biomedical hypothesis of the project was that COPD heterogeneities cannot be solely explained by the activity of pulmonary disease and that abnormalities in co-regulation of core metabolic pathways (bioenergetics, inflammation, tissue remodelling, oxidative stress) at systemic level might also play a key role on both systemic effects of COPD and comorbidity clustering. The main biological objective of the project was to characterize two specific aspects of COPD heterogeneity, namely i) understand the mechanisms of a specific systemic effect of COPD, i.e. skeletal muscle dysfunction that is widely accepted to be provoked by COPD; and ii) analyse comorbidity clustering observed in patients with COPD.

The systems approach of the project aimed to use computer-based modelling techniques and to create an integrated environment that enables the seamless communication of novel biological knowledge into clinical practice, as well as the dynamic collection and update of data for biological research (**Figure 6**). The design of Synergy-COPD was based on the interaction among four main components: i) the COPD knowledge base (COPDkb); ii) a simulation environment for computational modelling; iii) clinical decision support systems (CDSS); and iv) an adaptive case management system for integrated care of complex chronic patients.

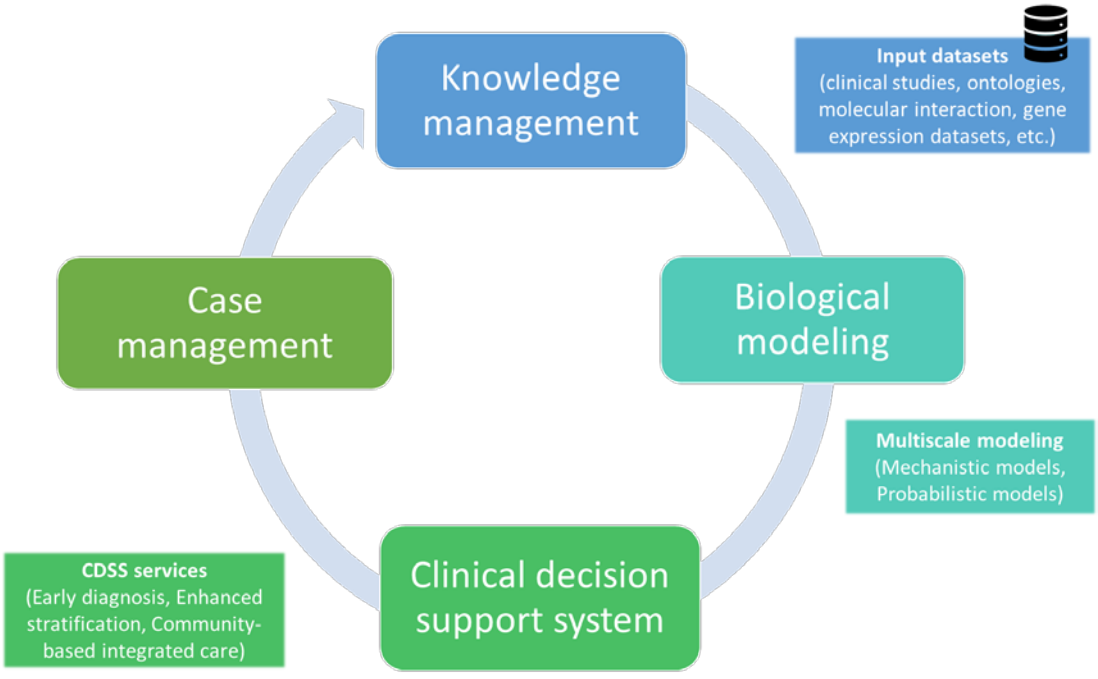


Figure 6. Schematic diagram of the Synergy-COPD project design. The design of Synergy-COPD was based on the interaction among four main components: i) a knowledge management system filled with data from clinical studies on COPD and related conditions, and publicly available data sources; ii) a simulation environment for computational/in-silico modelling; iii) rules generated using the computational models that can drive clinical decision systems (CDSS); and iv) an adaptive case management approach for integrated care of complex chronic patients.

The project aimed to contribute to each of these components by: i) developing and filling the COPDkb using clinical studies and publicly available data sources (e.g. genetic databases, pathways databases, etc.) as well as to enrich its functionalities with tools that can help to create biological models; ii) develop in-silico models to explore the main biomedical questions of the project; and, ultimately, iii) use in-silico validated computational models to feed CDSS for enhanced individual health risk assessment and stratification, leading to innovative patient management strategies. Synergy-COPD aimed to explore a wide spectrum of modelling techniques, including mechanistic and probabilistic modelling, as well as hybrid models combining these two approaches (discussed in detail in the "Current modelling approaches in systems medicine" section of the thesis).
Core aspects of the project design and specific parts of it have been reported in the Synergy monograph [61, 91–94].

From outcomes and challenges to new research directions

During the lifetime of the project (2011-2014), significant outcomes were generated, as well as challenges and barriers of the project were identified. Worth to mention that the contribution of the PhD candidate to the Synergy-COPD project started with a research protocol carried out in the frame of the BioHealth Computing Erasmus Mundus master course in 2013. At this time the project had been already running for 2 years and thus several challenges were identified that provided the core research directions for this PhD thesis. Some of these main challenges are summarised below.
First, early modelling efforts of the project indicated the need for further research on the molecular mechanisms of skeletal muscle dysfunction. Some of the challenges were faced when trying to integrate models at different scales for more accurate predictions. Other modelling challenges were related to newly available measurement data that required an analysis pipeline that can integrate data coming from different compartments (i.e. blood, muscle) and of different measurement type (i.e. transcriptomics, various metabolic measures, clinical measures, etc.). Second, registry data based analysis of comorbidity clustering indicated need for the validation of the results on an independent population. During the project, efforts in this direction were mainly hindered by data harmonization issues due to differences between disease coding versions of the International Classification of Diseases (ICD) used in different healthcare systems (i.e. USA; Sweden and Spain). The validation of the results using the same medical coding was identified as central need to prove the high impact of the findings

2. MULTIMORBIDITY RESEARCH ROADMAP

The evolving scenario of multimorbidity outlines a general roadmap that should facilitate comprehensive characterization of disease relations (**Figure 7**). In this context, population-health and systems medicine approaches are key for the proper understanding of the common mechanisms of comorbid conditions.

Population-health approaches show exciting potential to explore population-wide patterns of disease associations and temporal disease progression, however, their true potential is currently limited by investigating diseases on the phenotype level. Disease classification relying on disease endotypes should lead to efficient retrieval of accurate disease comorbidity maps and help to identify causative effects amongst comorbid diseases and risk factors. In turn these maps should guide molecular research to explore the underlying cause of the interactions and finally to target the shared mechanisms that should allow for: i) better early case identification; ii) definition of better preventive strategies; and iii) to explore novel therapeutic approaches.

This multimorbidity research roadmap highlights three main fields that are indispensable for its successful implementation, namely: i) assessment of the multimorbidity dynamics, ii) systems medicine tools that can streamline biomedical research results into clinical knowledge; and iii) health risk assessment and stratification tools to support clinical decisions. The upcoming three sections summarize the state of the art and the main challenges of these fields that support the understanding of the work done in the PhD thesis.

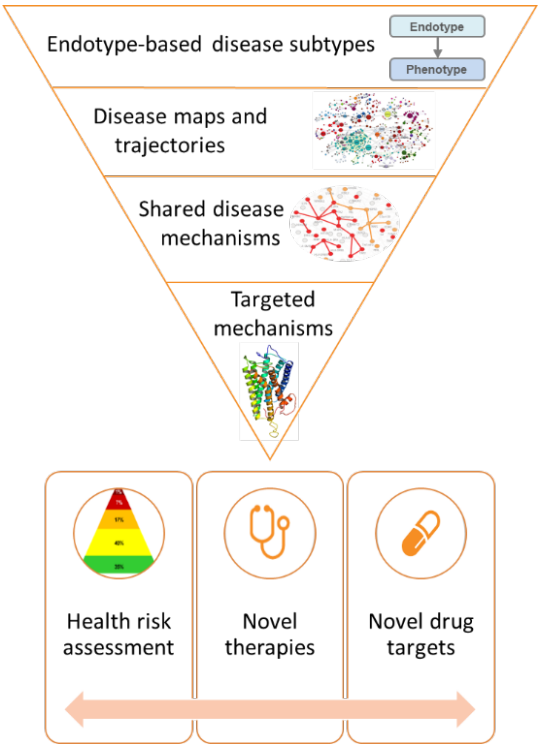


Figure 7. Multimorbidity research roadmap. Main steps that facilitate the comprehensive characterization of disease relations and their transfer to healthcare. Endotype-based disease subtypes should enable an accurate global view of disease relations, guiding molecular research to uncover mechanisms that can be targeted by novel therapies and pharmacological interventions and could lead to personalized health risk assessment. All of the stages of the pyramid are currently co-existing approaches, however their combined use shall contribute to the acceleration of multimorbidity research and knowledge transfer to healthcare.

2.1 ASSESSING DYNAMIC COMORBIDITY RELATIONS

Some of the main challenges regarding the assessment of comorbidity relations lie in the lack of clear boundaries between diseases (e.g. spectrum of obstructive pulmonary diseases) and their diverse molecular background, indicating that analysis of their relationships requires further studies taking into consideration several dimensions. For example, from a molecular perspective, a pair of diseases can be related because they both have been associated with the same genetic or metabolic problem; whereas, from an epidemiological perspective, diseases can be related when they affect the same individuals substantially more often than expected by chance alone. State of the art research on disease co-occurrence is aiming to quantify and identify comorbidities with two main strategies: i) mechanism-based analysis of disease co-occurrence; and, ii) phenotype-based analysis of disease co-occurrence (Cross-sectional disease maps and Temporal disease trajectories).

Mechanism-based analysis of disease co-occurrence

The ultimate aim of biomedical research is to decipher the molecular mechanisms that are driving diseases and can help to find points of interventions, potentially driving the system back to its normal behaviour or slow disease progression. When looking at comorbid diseases, there is further potential to find mechanisms that drive the emergence of several comorbidities. Recently, the disease module hypothesis opened new avenues in comorbidity research and several works have assessed comorbidity relation based on shared molecular traits of diseases. For example, Goh et al. created a network of Mendelian gene-disease associations by connecting diseases that have been associated with mutation in the same genes [6], whereas Lee et al. constructed a network in which two diseases are linked if mutated enzymes associated with them catalyse adjacent metabolic reactions [7]. Lately, Menche et. al. also showed that the distance of disease modules in the interactome has a strong influence on comorbidity relations, such that diseases with modules situated closer in the interactome have higher risk of showing phenotype-based comorbid relation than the ones situated further away, as well closer diseases showed higher similarity in terms of symptoms, expressed genes and associated biological function [4]. Uncovering mechanism-based comorbidity has huge potential and can lead to the discovery of novel drug targets, development of new therapies and enhance clinical decision making.

Phenotype-based analysis of disease co-occurrence

Access to well annotated national registry data sources recently opened new avenues to analyse disease occurrence and disease progression on a population level, permitting unbiased study designs for the analysis of the relation of all diseases enlisted in registries. Based on the observation that diseases with shared molecular mechanisms have direct epidemiological consequences [4, 36], disease phenotype-based methods are powerful tools to define the comorbidity relation of disease pairs. The phenotype-based approaches define the existence of comorbidity relationship if two diseases co-occur more often than it would be expected only by chance.

Cross-sectional disease maps have emerged as an early use of these approaches, such as the work of Hidalgo and colleagues [3], that represented pair-wise comorbidities, measured by relative risk and binary correlation, as a disease map. This representation, in fact, allows for the identification of specific comorbidity patterns that occur in the population. The strength of co-occurrence-based comorbidity relies in highlighting diseases that affect large populations and that potentially share common molecular mechanisms.

From the **temporal point of view**, comorbidities can have different manifestations, which need to be taken into account when assigning treatment or assessing the risks of a patient. One thing to consider is that comorbidities can appear one after each other or occur within a given period of time without being simultaneously present at any given time point [37]. The sequence in which the comorbidities appear may also have special importance, e.g. the treatment may be very different for a patient with depression who is newly diagnosed with COPD and a patient being diagnose with COPD first and then with depression. Earlier approaches to research such time-critical disease associations followed hypothesis-driven designs, where a selected index-disease is compared against other comorbid conditions. Access to well annotated national registry data sources recently opened new avenues to population level analysis of disease progression.

Earlier data-driven studies using network-approach showed potential for exploring temporal and non-temporal patterns of comorbidities in elderly patients [3] and their change over age groups [95]. However, these studies considered relatively short periods of time (1-3 years), reducing the possibility of observing the development of novel diagnoses. Recent studies aimed to address this problem, for example Jensen et al. analysed 14.9 years of registry data to identify **temporal disease trajectories**, defined as ordered series of diagnoses observed in a patient [5]. Five main trajectory patterns were identified, from which the one centred on COPD is shown in **Figure 8**. Although the introduced methodology cannot be used to conclude causative effects due to possible confounding factors, it can inform on possible directions of progression of the disease, which may be used to predict future disease outcomes. Similar methodology showed potential in predicting mortality in sepsis patients and indicated the potential to incorporate trajectory based risk into clinical mortality risk scores [96]. All in all, temporal diagnosis trajectories reassure the existence of patterns in the sequence of developing comorbid conditions.

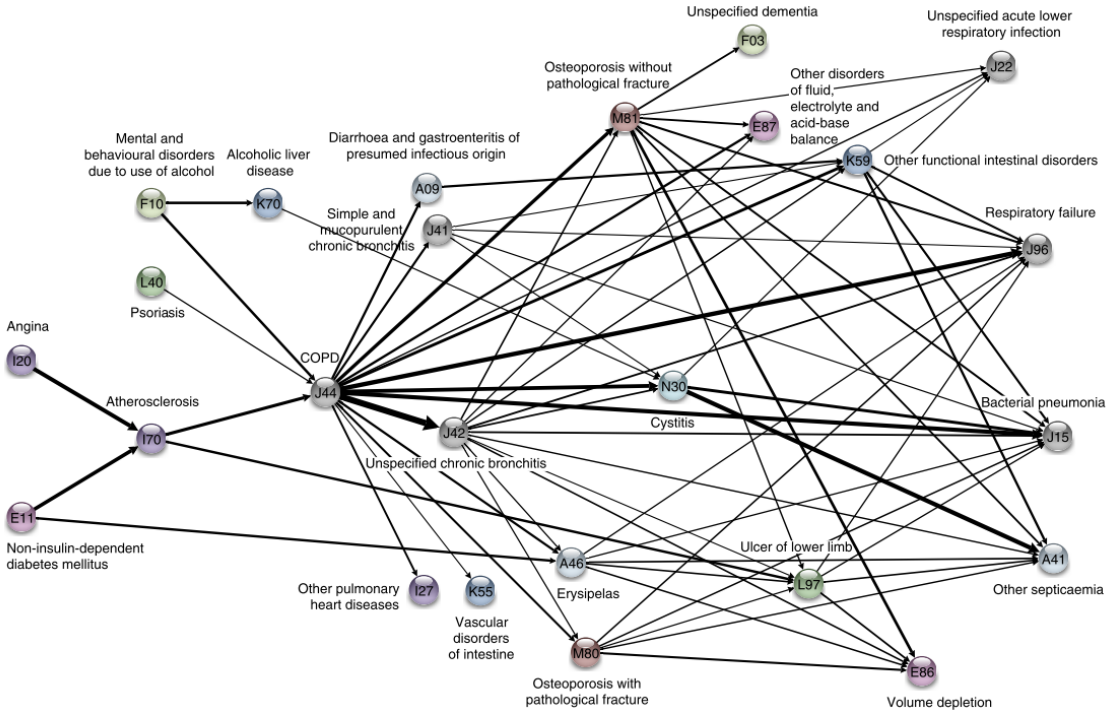


Figure 8. COPD disease trajectory. The result of the analysis of Jensen et al. [5], showing the COPD cluster with five preceding diagnoses leading to COPD and some of the possible following diagnoses.

To summarize, the current state of the art indicates that diseases progress and co-occur according to the dynamics of underlying molecular mechanisms. Further research however is needed to elaborate on specific disease relations and point out actionable factors and specific uses of these approaches.

2.2 CURRENT MODELLING APPROACHES IN SYSTEMS MEDICINE

Due to the complexity of the interactions of biological systems, computational and mathematical modelling has become part of the common toolbox of systems biology research. Over the years, several modelling approaches have emerged but in general it can be summarized by two main strategies: mechanistic, knowledge-based modelling approaches and probabilistic modelling and network analysis approaches.

Mechanistic models are aiming to simulate the specific biological functions, often based on ordinary differential equations. Used in conjunction with quantitative experimental data, such models are powerful tools for understanding systems dynamics. Their simulations provide quantitative and temporal predictions, which can be crucial for understanding biological processes. Wide variety of tools and mature simulation environments are available to create such models and simulate their behaviour [97]. A limiting factor in the application of such models is that they often

need deep understanding of the phenomena that they model. Any given model only works with few, but very specific, parameters, whose quality strongly influence the results of the simulation. These factors and the advancement of high throughput 'omics' measurements, producing thousands of data points per experiment (e.g. Affymetrix Human U133 Plus2 Gene Chip can measure over 47,000 transcripts per sample), led to high demand for tools that can handle large quantity of data and filter out biologically important details.

Probabilistic modelling and network analysis tools concentrates on quantifying the interaction of biological entities and in the interaction network detect communities that play key role in the studied biological function or disease. Functional analysis of these entities can be used to characterize processes involved in diseases and reveal molecular abnormalities. Besides, they can be used as feature extraction tools to be further processed with machine learning algorithms and predict disease related risks. Due to their importance and wide use, this approach is further discussed in the next sections.

The **combination of different modelling approaches** is also a promising field of research. In theory, using the top-down approach of network analysis models to retrieve specific functional blocs in a manner that it can be subjected to mechanistic modelling can lead to systems of models simulating the behaviour of a cell, tissue, organ or even a human being. Despite the complexity of this approach, there are several methodological attempts [97], as well as several projects aiming to address this need, such as the ones under the Virtual Physiological Human topic of the 7th Framework Programme of the European Commission [98].

Probabilistic modelling and network analysis: finding biological function in molecular data

To study biological function, molecular biology for long relies on reverse engineering cellular processes by perturbing biological systems, i.e. removing or changing molecular entities one-by-one and examining their effect on the system. Traditionally, deciphered molecular functions are organized into **canonical pathways**: a series of interactions among molecules in a cell that leads to a certain product or a change in a cell, depicted in network diagrams. A wide variety of resources are available to browse pathways (e.g. KEGG [99], Reactome [100], WikiPathways [101]), which are most often used in conjunction with gene set enrichment tools to analyse lists of genes derived from experimental setups (e.g. Enrichr [102, 103], DAVID [104]) (see **Figure 9**).

Early gene expression analysis pipelines consisted of two main steps: i) derive differentially expressed genes from the comparison of case-control profiles and then (ii) use enrichment tools to find pathways, in which the differentially expressed genes represented in greater number than it would be expected by chance. These gene expression analyses however showed quite discouraging results in terms of reproducibility and comprehensibility. While enrichment tools are still broadly used to derive functional insight to gene lists, main limitations have been identified, namely: i) the use of canonical pathway for functional annotation, ii) the raw use of differentially expressed profiles, and iii) noisy experimental techniques and pipelines that recent techniques are trying to address.

Main **limitation of canonical pathway resources** is owing to two main factors: i) they are mostly derived from hypothesis-driven experiments collected in exceedingly diverse contexts, encompassing a large variety of experimental conditions (e.g. different species, cell types/tissues, diseases) and/or in-vitro models, therefore they represent generalized pathways, i.e. they lack specificity to any tissues, diseases, etc. [105, 106]; and, ii) they represent cell functions as separate entities that are shown to be much more interrelated than this traditional representation [107, 108]. Two main approaches emerged to address these limitations. One of them concentrates on the enhanced representation of available knowledge and, the other focuses on data-driven recreation of pathways.

The **knowledge-driven approach** aims to integrate various molecular resources and create high-quality and expert community-driven conceptual representation of mechanisms specific to a disease, in a machine-readable manner. Following such an approach, the Disease Map community constructed maps for cancer signalling [109], Alzheimer's [110, 111], Parkinson's [12] and influenza [112] and showed possible use cases for these enhanced pathways. In fact, some knowledge-driven approaches could be viewed as the next step of data driven investigations that is needed for the consolidation of their results.

The **data driven approach** concentrates on recreating pathways from single measurements that ensures their specificity to the experimental context, i.e. disease or tissue. Thanks to recent breakthroughs in high-throughput experimental methods (e.g. microarray, RNA-seq for transcriptional measurements, yeast 2-hybrid methods for measurement of protein-protein interactions), there is an increasing interest in these approaches. Due to the strongly specific biological processes (i.e. tissue specific processes, disease specific processes, environment- and genotype specific processes), dynamic pathway reconstruction comes with a great promise to facilitate

personalized medicine [113]. In recent years, two main approaches seem to dominate this field [114] (**Figure 9**):

- ✓ **Interaction network-based module identification methods**. These methods concentrate on finding neighbourhoods of interacting molecules that represent functional and disease related mechanisms. Experimental data (i.e. expression profile of genes in a cell/tissue) is often used in conjunction with a network of physical interaction of gene products to identify modules that show different expression profile compared to a control condition [115].
- ✓ **Network inference methods**, based on solely experimental data, aim to infer gene regulatory network (GRN) using correlation and other information theory methods [97, 116, 117]. Consequently, they can infer logical relations of non-interacting genes, which can be especially important to identify the regulatory relationships and/or interactions between biological components.

These methodologies, while representing different approaches, can be used in a complementary manner as they study different aspects of cell mechanisms, as well as one can reassure the findings of the other.

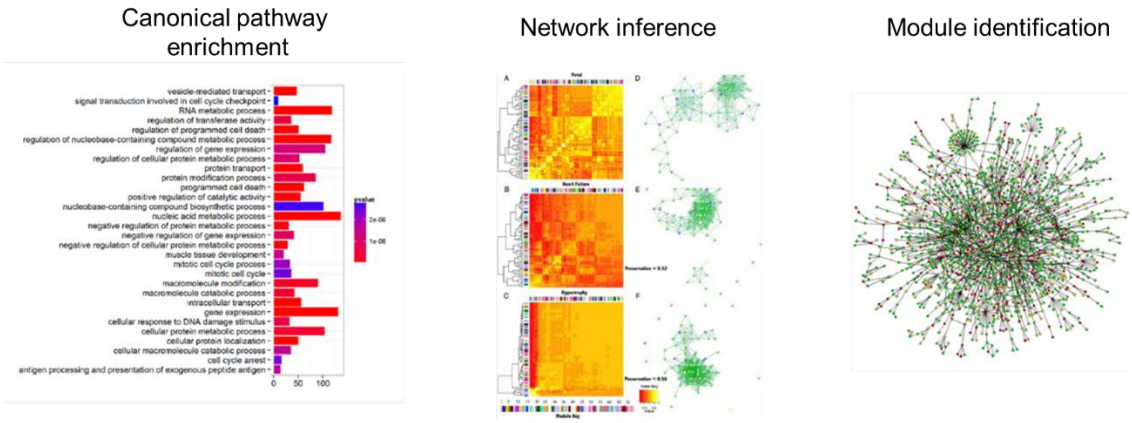


Figure 9. Main approaches to analyse genomic data. Canonical pathway enrichment analysis tests the statistical enrichment of pathways elements in gene lists. Network inference tools aim to recreate gene regulatory networks based on the correlation of measured expression levels. Module identification methods use interaction networks (e.g. PPI, metabolic, etc.) and measured expression levels to identify active communities.

Whereas in the frame of the Synergy-COPD project there were intentions to integrate all the above modelling techniques (mechanistic, network-based and GRN), the current PhD thesis mainly concentrates on the use of interaction networks. This decision was mainly based on the robust properties of this approach, i.e. it uses a fixed, functionally organized map of biological interactions, which works well for filtering random noise [118] that is especially important when using noisy biological

measurements, such as microarray data, as well as when working with low sample sizes that usually characterize clinical studies.

Major criticisms of this approach are based on the incompleteness of currently available interaction networks, which is a possible source of false discoveries and could lead to omitting valuable information. Whereas these points are arguably true, recent results points out that these effects are minor compared to the potentials of these methods and proving their potential for systemic studies of disease mechanisms [119]. For these reasons interaction networks are described in detail in the next section.

Module identification for personalized disease profiles

Cellular organization is thought to be fundamentally modular [120, 121]. At the molecular level, modules are defined as groups of genes, gene products or metabolites that are functionally coordinated, physically interacting and/or co-regulated [36, 120–122]. They are described as drivers of common biological processes, and as the functional building blocks of the cell [36, 120–122]. To create a complete map of biological modules, large networks of intermolecular interactions are being measured systematically for humans and many model species. Most attention is focused on **protein-protein interaction** (PPI) networks, whose nodes are proteins linked to each other via physical (binding) interactions [123, 124]. Such networks can also include protein–DNA physical interactions, metabolic pathways or even functional associations. These networks all together are often referred as the interactome [118].

When integrating biological interactions, some very interesting **properties** emerge from the network structure. One property is the scale-free nature of biological networks, highlighting hub proteins that interact with more proteins than most of the nodes in the network and that have special biological role, i.e. they tend to be encoded by essential genes, they are more conserved proteins, they are more prone to have greater phenotypic effect than other proteins in the network, and they tend not to be tissue specific and disease related proteins [6, 36].

Another important property of biological networks is their modular structure. A module is a network substructure with densely connected nodes and sparser connections to other, non-module nodes. Such modules are often specialized for certain subtasks. Three types of modules are distinguished when looking at their context specificity. *Topological modules* represent densely connected neighbourhood in the network, based on purely topological measures. A *functional module* represents a community in the interactome that plays specific role in a cellular process, whereas

disease modules represent altered molecular processes that lead to the emergence of phenotypic signs of diseases. While topological models are hardcoded in the network, the latter two are dynamic structures in the network and depend largely on the interplay of the interactome with environmental factors.

These concepts are connected by the local hypothesis, i.e. molecular elements involved in the same function or dysregulated by the same disease show high propensity to interact with each other [6, 36]. In this context, cellular components of a topological module have closely related functions, thus it approximates a functional module; and a disease is a result of alterations in a particular functional module, i.e. a disease module is a disease specific manifestation of the functional module. A recent study has also demonstrated the validity of the local hypothesis on relations between disease modules. Menche and colleagues have shown that disease pairs with overlapping disease modules display significant molecular similarity, elevated co-expression of their associated genes, and similar disease symptoms and high comorbidity, whereas non-overlapping ones lacked any detectable pathobiological relationships [4].

Numerous approaches have been developed based on these properties to mine such networks for identifying biological modules. Especially PPI networks are gaining increasing attention in the biomedical research field due to their ability to highlight complex cellular mechanisms [115, 125, 126]. Earlier methods concentrated on clustering proteins based on topological features of the network such as degree and betweenness centrality [127]. However integrative approaches are continuously gaining larger interest. One of the most successful integrative approaches has been to overlay networks with molecular profiles to identify '***active modules***' [115]. Molecular profiles (i.e. transcriptomics, genomics, proteomics, epigenetics, etc.) capture dynamic and process-specific information that is correlated with cellular functions or disease states, complementing static interaction data, derived under a single experimental condition and representing a generalized state of a cell. Active modules or network hotspots are network regions showing marked changes in molecular activity (e.g. transcriptomic expression) or phenotypic signatures (e.g. mutational abundance) that are associated with a given cellular response [115, 128, 129]. By mapping differential network changes across conditions, these methods can inform on cellular rewiring happening between the studied conditions [43].

Many computational techniques have been developed that automate the large-scale identification of active modules in an unbiased manner. An exciting new technique in this field, often referred as **diffusion-flow** and **network-propagation methods**, aims

to simulate the spread of influence of protein activity (often modelled by gene differentially expression) to physical interaction partners [115]. For example, the HotNet2 algorithm is based on a modified heat diffusion model, where proteins function as heat sources and physical interactions amongst them are links where heat can diffuse, i.e. simulating the spread of the influence of each protein to its interaction partners. This algorithm has been shown to retrieve biologically more meaningful modules and has been successfully applied to characterize mutational profiles in various cancer types [130–132].

In summary, the proliferation of 'omics' measurements, both in biological research and recently in the clinics, creates high demand for methodologies that can retrieve and analyse biologically important features from large amount of measurement data. In this context, network analysis is a promising framework with high potential for generalization and broad applicability. Further research, however is still needed to elaborate the potentials of these tools in generating knowledge that can be easily transferred to healthcare, for instance, for enhanced health risk assessment of patient with multimorbidity.

2.3 HEALTH RISK ASSESSMENT AND SERVICE SELECTION

Challenges of clinical decision in multimorbidity care

Decision making in the clinical setting traditionally faces two types of challenges. On the one hand, current healthcare systems are reactive (i.e. aims to solve specific disease events, such as pneumonia or appendicitis on the backdrop of chronic diseases); as well as disease-centred (i.e. it targets the management of each health condition of a patient independently), leading to suboptimal treatment for patients with multiple, related chronic disease.

On the other hand, the design and selection of appropriate therapeutic plans and services also constitutes a substantial challenge for multimorbid patients, for whom currently available interventions are rather limited in terms of actionable factors and show only limited improvement in clinical outcomes and health service use [35, 133].

Clinical decisions are traditionally based on knowledge, general and field-specific, of the health professional, previous experience, as well as intuition. Lately, rule-based decision making, relying on robust medical evidence, often generated through randomized controlled trials, has also became an essential component of clinical decision making process.

These factors highlight the potential for a more patient-oriented healthcare, where patient care is integrated amongst healthcare tiers and that provide personalized and cost-effective preventive interventions to modulate disease progression. In recent years, health risk assessment and stratification strategies emerged as potential facilitator of these efforts.

Health risk assessment and stratification for personalized medicine

Patient-based health risk assessment – In the clinical management domain, risk prediction of well-defined medical problems aims to support health professionals in the decision making process for a given patient. The definition is valid for example for the process solving a clinical episode of pneumonia or exacerbation of a chronic disease, but also to define mid- or long-term action plans for chronic patients aiming at developing an optimal care plan.

When considering one disease in a given patient, prognosis is essentially based on two main parameters: i) severity, defined as the degree of alteration of the organ/systems caused by the disease, which have an impact on the functional reserve; and, ii) activity, defined by the rate of progression of the disease. Appropriate markers of these two phenomena contribute to define both risk and prognosis of the patient which, in turn, facilitates his/her classification into risk strata; that is, patient stratification. Moreover, the combination of patient stratification and the identification of the disease end-points, or target outcomes, are the two key elements to define specific therapeutic strategies or action plans for the patient. The ultimate aim is to classify the patient in the appropriate health tier and identify the type of service that would allow optimization of healthcare provision.

Population-based health risk assessment – Risk assessment in the health services domain is differentiated from individual clinical risk prediction in the sense that it examines patient risk in the context of the entire population of a geographical region. These tools are broadly used by policy makers and/or payers for service commissioning or other uses like risk adjustment analyses or actuarial approaches. However, these tools also show great potential to be applied in the healthcare arena. Population-based health risk assessment allows identification of subsets of citizens with similar healthcare requirements and thus facilitates both case finding and screening. The former, case finding, identifies highly vulnerable patients, allocated at the tip of the risk pyramid who are prone to major deleterious health events such as unplanned hospital admissions/re-admissions, fast functional decline and/or death [134, 135]. Likewise, performing screening for discovery of cases with non-manifest illnesses may

benefit from early diagnosis and cost-effective preventive interventions [136]. Comprehensive descriptions of the characteristics of health risk predictive modelling and the logistics required for deployment are reported elsewhere [137–141].

Current status and factors limiting evolution – However, real world healthcare settings face high levels of complexity that are imposing huge challenges, specifically on risk assessment and patient stratification for adequate service selection. Main determinants of such complexity are: i) patient heterogeneity with lack of appropriate biomarkers and/or insufficiently defined end-points of the disease; ii) co-existence of one main disease and several accompanying disorders (or co-morbidities), in some cases showing shared mechanisms that may explain comorbidity clustering [142]; iii) poor control on factors determining health status beyond the clinical scenario (socio-demographics, biological and lifestyle related data); iv) patient health risk is a dynamic phenomenon with sometimes unexpected events that requires high levels of flexibility in terms of event-handling by the case manager in charge of the patient; and, v) fragmentation of healthcare services.

While adoption of rule-based clinical decision support has made substantial progresses over the last years, the use of computational models for patient-based health risk predictive modelling to enhance clinical decision support is still in its infancy. Limitations of patient-based risk prediction revolve around three main factors. Firstly, there is a need for the enrichment of predictive modelling based on clinical information with other sources of multidisciplinary health data (i.e. informal care, population-health, biological data, etc.) obtained with a multilevel approach (**Figure 10**). A second factor is to ensure general applicability and transferability of current predictive tools, addressing specific clinical issues with a high predictive power, to other populations outside the source study groups. Last, but not least, there is a clear need to overcome limitations of use of risk factors as prognostic factors, as described in [143].

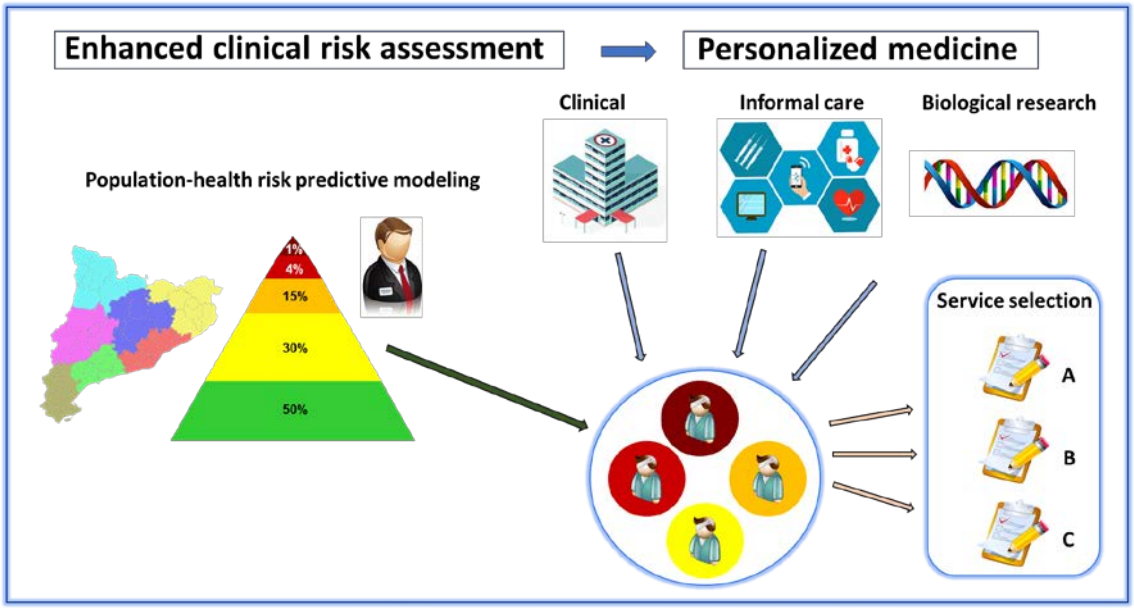


Figure 10. Patient-based health risk assessment is used as synonymous with enhanced clinical risk assessment, which adopts of a holistic approach that fosters inclusion of covariates from multilevel data sources, namely: i) clinical, ii) informal care; iii) biological research; and, iv) outcomes from population-health risk predictive modelling, resulting in enhanced patient-based stratification and optimization of service selection. This approach paves the way towards personalized medicine. Population-health risk predictive modelling includes all the citizens in a given geographical area

Population-based health risk assessment relies on elaborated tools that are broadly implemented in some healthcare systems (e.g., CRGs, GMAs), however recent projects [73, 144, 145] identified several limiting factors of these tools, such as difficulties in the comparability and transferability of risk prediction tools amongst regions, as well as they identified evolving requirements for them such as: i) integration between healthcare and social services; and, ii) implementation of synergies between population-health and clinically oriented risk predictive modelling, as described in [73].

Learning healthcare systems

The implementation of the setting described above implies realization of a new health paradigm, i.e. the learning healthcare systems, recently described by the American Heart Association (AHA) [26]. In 2013, the Institute of Medicine reported Best Care and Lower Cost: The Path to Continuously Learning Health Care in America [26] wherein the concept of Learning Healthcare Systems (LHS) was formulated as a strategy to improve the quality and efficiency of healthcare. A recent document generated by the AHA [26] further develops the concept of LHS and proposes specific steps to make it operational and evaluate its implementation, see Table 2 in [26].

Briefly, LHS uses health information technology and the health data infrastructure to apply scientific evidence at the point of clinical care while simultaneously collecting insights from that care to promote innovation in optimal healthcare delivery and to fuel new scientific discovery [26]. Such a system creates an iterative learning process where evidence informs practice and practice informs evidence (**Figure 11**).

The main goal of LHS is to facilitate an optimal care decision and delivery by reducing the complexity of the massive amount of clinical data that's being produced every day and to improve efficiency of health outcomes both in terms of well-being and expenditures. The LHS relies on the availability of health-related data and tools that process it, such as predictive modelling and clinical decision support contributing to the acceleration of evidence diffusion to practice, help to identify gaps in care and to target interventions to appropriate population.

Main technical building blocks of such a system are data availability, predictive modelling, service selection and clinical decision support systems. Data is key because it provides the continuous feedback from the practice, while predictive modelling processes and extracts important information from the produced data. Computed patient risk then can be used to stratify patients to intervention groups that help in the optimal service selection for the patient.

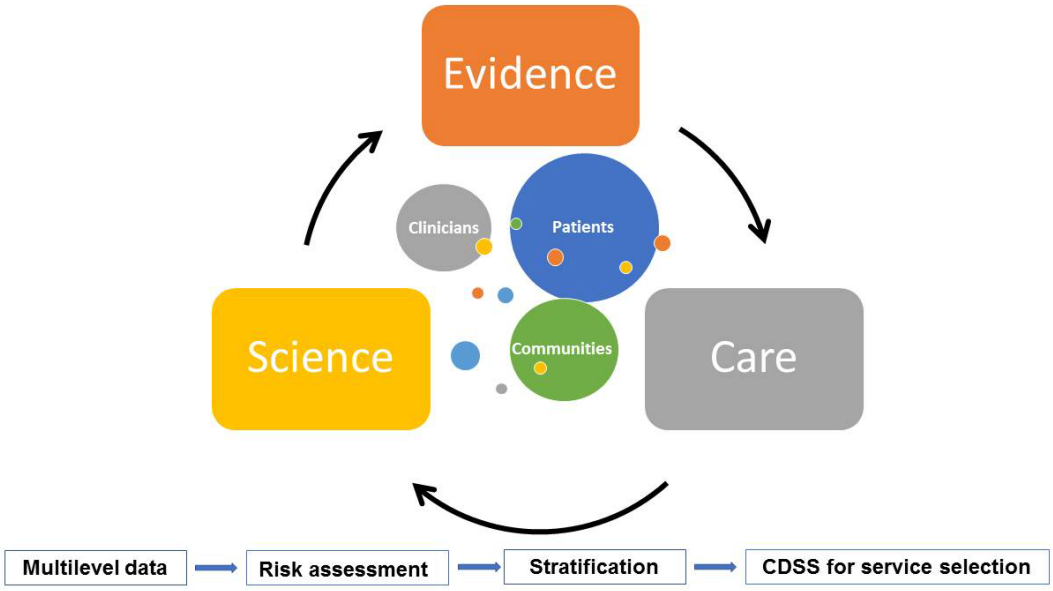


Figure 11. Basic traits of a Learning Healthcare System (LHS). It constitutes an organizational concept technologically supported by the Digital Health Framework. The LHS fosters generation of scientifically-based evidence and speed-up its applicability into healthcare. Regarding health risk predictive modelling, it shall allow inclusion of covariates from multilevel data sources which should enhance model robustness and eventually transferability feeding clinical decision support systems (CDSS) for appropriate integrated care service selection.

3. SUMMARY OF THE INTRODUCTION

The epidemics of non-communicable diseases and the need for cost-containment in healthcare are clearly indicating that the current approach for care delivery needs to be profoundly reshaped. On the one hand, personalized medicine is essential for achieving better health outcomes and optimal resource allocation. On the other hand, the burden of multimorbid conditions on healthcare and patient lives strongly questions the single disease approach followed by contemporary clinical practice and suggests the need for novel strategies concentrating on multimorbidity prevention and treatment.

The emergence of novel biomedical technologies and related methods in systems medicine outline novel strategies for mechanism-based approach to diseases. In this context, identifying specific biological mechanisms underlying a disease and that potentially lead to the emergence of other diseases constitute major actions to achieve current healthcare goals, alluded to above.

Furthermore, access to population wide registries and other patient related multidisciplinary data sources opens new avenues in data driven analysis of diseases. Population-wide patterns of disease co-occurrences can facilitate the characterisation of disease interactions for the better understanding of the comorbidity challenge. Furthermore, these resources show a key role in enhancing health risk assessment and patient stratification facilitating decision in the clinical practice in general and personalised service selection in particular.

Synergy-COPD, based on the current challenges of COPD care, envisioned an integrated environment to facilitate the systemic study of chronic diseases and the seamless communication of results into clinical practice. The work of the current PhD thesis was based on the foundations of this project, as well as develops on its vision answering challenges identified during the project lifetime.

HYPOTHESIS

The central hypothesis of the PhD thesis builds on the emerging biological evidence that clustering of comorbid conditions, a phenomenon seen in complex chronic patients, could be due to shared abnormalities in relevant biological pathways (i.e. bioenergetics, inflammation and tissue remodelling). It is assumed that a systems understanding of the patient conditions may help to uncover the molecular mechanisms and lead to the design of preventive and targeted therapeutic strategies aiming at modulating patient prognosis.

The PhD thesis focuses on non-pulmonary phenomena of COPD (i.e., systemic effects and comorbidities) often observed in patients with COPD, as a paradigm of complex chronic disease.

MAIN OBJECTIVES

The general objective of the PhD thesis is threefold: i) to investigate molecular disturbances at body systems level that may lead to a better understanding of characteristic systemic effects and comorbidities of patients with COPD; ii) to analyse population level patterns of COPD comorbidities and investigate their role in the health risk of patients with COPD; and, iii) to explore technological strategies and tools that facilitate the transfer of the collected knowledge on comorbidity into clinical practice.

More specifically, the objectives of this work are:

OBJECTIVE 1 – ANALYSE THE MECHANISMS OF SKELETAL MUSCLE DYSFUNCTION IN PATIENTS WITH COPD

Rationale: Skeletal muscle dysfunction is a systemic effect of COPD with prominent negative impact on prognosis. Due to the multifactorial nature of the condition, its underlying mechanisms are still unclear, hindering the development of novel preventive and therapeutic strategies. Systemic approaches with the aim of uncovering underlying network dynamics show potential to point out relevant biological pathways that drives abnormal conditions.

Objective: The thesis aims: i) to develop a modelling tool to facilitate the integration of existing biological models for enhanced prediction of oxidative stress in the skeletal

muscle of patients with COPD (**Manuscript 1**); and, ii) to explore the underlying mechanisms of skeletal muscle dysfunction in patients with COPD, before and after exercise training, by assessing transcriptionally active network modules (**Manuscript 2**).

Manuscript 1

Tényi Á, de Atauri P, Gomez-Cabrero D, Cano I, Clarke K, Falciani F, Cascante M, Roca J, Maier D. ChainRank, a chain prioritisation method for contextualisation of biological networks. *BMC Bioinformatics* 2016; 17: 17.

Manuscript 2

Tényi Á, Cano I, Marabita F, Kiani N, Kalko SG, Barreiro E, de Atauri P, Cascante M, Gomez-Cabrero D, Roca J. Network modules uncover mechanisms of skeletal muscle dysfunction in COPD patients. *J. Transl. Med.* BioMed Central; 2018; 16: 34.

OBJECTIVE 2 – ANALYSE MULTIMORBIDITY RISK IN PATIENTS WITH COPD

Rationale: Recent publication, using the Medicare (US) registries, showed increased risk in COPD to develop comorbid conditions compared to non-COPD population. These findings, in accordance with common comorbidity patterns, highlight the potential role of multimorbidity as a health risk factor with potential use in personalised health risk assessment.

Objective: The thesis aims to elucidate the role of comorbid conditions in the health risk of patient with COPD in Catalonia (ES), with a two-fold objective: i) to assess the comorbidity risk of patients with COPD in Catalonia (**Manuscript 3**); and, ii) to analyse the burden of multimorbidity and its effect on adverse hospitalization related events in patients with COPD with a population-health analysis (**Manuscript 4**).

Manuscript 3

Tényi Á, Vela E, Cano I, Cleries M, Monterde D, Gomez-Cabrero D and Roca J. Risk and temporal order of disease diagnosis of comorbidities in patients with COPD: a population health perspective. *BMJ Open Respiratory Research* British Medical Journal Publishing Group; 2018; 0: e000302.

Manuscript 4

Vela E, **Tényi Á,** Cano I, Monterde D, Cleries M, Garcia-Altes A, Hernandez C, Escarrabill J, Roca J. Population-based analysis of patients with COPD in Catalonia: a cohort study with implications for clinical management. *BMJ Open* British Medical Journal Publishing Group; 2018; 8: e017283.

OBJECTIVE 3 – ESTABLISH A NEW PERSPECTIVE ON COPD NON-PULMONARY EFFECTS AND FACILITATE KNOWLEDGE TRANSFER TO HEALTHCARE

Rationale: Current standard of care recommendations for COPD concentrates mostly on pulmonary events of the disease, while non-pulmonary effects are given suboptimal consideration in the assessment and management of the disease. Meanwhile, the increasing amount of health-related data being generated in different tiers of healthcare, i.e. formal care, informal care and biomedical research, bears with great potential to revolutionize healthcare and facilitate personalized health risk prediction and stratification for better assessment and management of diseases.

Objective: The thesis aims to summarize consolidated outcomes on the role of non-pulmonary effects in COPD heterogeneity, as addressed in the three main biomedical dimensions of the Synergy-COPD project (**Manuscript 5**): i) Skeletal muscle dysfunction as a systemic effect of COPD; ii) Comorbidities; and, iii) Proposals for enhanced transfer of knowledge into clinical practice. Furthermore, the thesis aims to revise the main determinants of knowledge transfer into healthcare and opportunities of Big Data analytics in the health area (**Manuscript 6**).

Manuscript 5

Cano I, Gomez-Cabrero D, **Tényi Á**, Jesper Tegner, Wagner P, Maier D, Miralles F, Cascante M, and Roca J. Non-pulmonary manifestations of Chronic Obstructive Pulmonary Disease: mechanisms, risk assessment and clinical management. Submitted to Respiratory Research – April 2018.

Manuscript 6

Cano I, **Tényi Á**, Vela E, Miralles F, Roca J. Perspectives on Big Data applications of health information. *Curr. Opin. Syst. Biol.* 2017; 3: 36–42.

RESULTS

MANUSCRIPT 1: CHAINRANK, A CHAIN PRIORITISATION METHOD FOR CONTEXTUALISATION OF BIOLOGICAL NETWORKS.

Tényi Á, de Atauri P, Gomez-Cabrero D, Cano I, Clarke K, Falciani F, Cascante M, Roca J, Maier D. ChainRank, a chain prioritisation method for contextualisation of biological networks. *BMC Bioinformatics* 2016; 17: 17.

- Journal: *BMC Bioinformatics*
- Impact factor: *2.448*
- Quartile: *1st*
- Status: *Published*

Tényi et al. BMC Bioinformatics (2016) 17:17
DOI 10.1186/s12859-015-0864-x

BMC Bioinformatics

METHODOLOGY ARTICLE Open Access

ChainRank, a chain prioritisation method for contextualisation of biological networks

Ákos Tényi[1,4*], Pedro de Atauri[1,2], David Gomez-Cabrero[3], Isaac Cano[1,4], Kim Clarke[5], Francesco Falciani[5], Marta Cascante[1,2*], Josep Roca[1,4] and Dieter Maier[6]

Abstract

Background: Advances in high throughput technologies and growth of biomedical knowledge have contributed to an exponential increase in associative data. These data can be represented in the form of complex networks of biological associations, which are suitable for systems analyses. However, these networks usually lack both, context specificity in time and space as well as the distinctive borders, which are usually assigned in the classical pathway view of molecular events (e.g. signal transduction). This complexity and high interconnectedness call for automated techniques that can identify smaller targeted subnetworks specific to a given research context (e.g. a disease scenario).

Results: Our method, named ChainRank, finds relevant subnetworks by identifying and scoring chains of interactions that link specific network components. Scores can be generated from integrating multiple general and context specific measures (e.g. experimental molecular data from expression to proteomics and metabolomics, literature evidence, network topology). The performance of the novel ChainRank method was evaluated on recreating selected signalling pathways from a human protein interaction network. Specifically, we recreated skeletal muscle specific signaling networks in healthy and chronic obstructive pulmonary disease (COPD) contexts. The analysis showed that ChainRank can identify main mediators of context specific molecular signalling. An improvement of up to factor 2.5 was shown in the precision of finding proteins of the recreated pathways compared to random simulation.

Conclusions: ChainRank provides a framework, which can integrate several user-defined scores and evaluate their combined effect on ranking interaction chains linking input data sets. It can be used to contextualise networks, identify signaling and regulatory path amongst targeted genes or to analyse synthetic lethality in the context of anticancer therapy. ChainRank is implemented in R programming language and freely available at https://github.com/atenyi/ChainRank.

Keywords: Biological networks, Protein-protein interaction, Data integration, Filtering, Computational biology, Bioinformatics, Systems biology, COPD

Background

Canonical pathways are widely used tools to represent signal transduction and molecular networks. They generally rely on literature-based information, mostly derived from hypothesis-driven experiments collected in exceedingly diverse contexts, encompassing a large variety of experimental conditions (e.g. different species, cell-types/tissues, diseases) and/or in-vitro models. Multiple layers of information (e.g. direction of a signalling event, type of interactions or cartoon graphics) make literature-based pathways a highly accepted and convenient source of information in biological research. However, the emergence of high-throughput technologies has shown several limitations of the approach.

By incorporating non-hypothesis based interactions, high-throughput methods have revealed many previously unrecognised pathway components [1–3]. Moreover, different studies have shown high interconnectedness of signalling pathways indicating larger complexity than the conventional separate representation of molecular events [4, 5]. Furthermore an increasing amount of evidence suggests the dependence of biological, cellular and disease outcomes on the complex of interactions between genes,

* Correspondence: tenyi.akos@ub.edu; martacascante@ub.edu
[1]Hospital Clínic-Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Research Institute, Universitat de Barcelona, C/Villarroel 170, 08036 Barcelona, Spain
Full list of author information is available at the end of the article

proteins and other molecules [6] which is rarely addressed in pathway databases. Consequently, it is currently apparent that the classical pathway approach is too simplistic to properly describe complex cellular events [7–9].

With advances in high-throughput technologies an increasing number of genome scale association data became available. This scenario facilitates the construction of data-driven biological networks, integrating experimental data, e.g. on protein-protein interactions (PPI), gene regulation and metabolic interactions, offering a systems approach to model molecular events [10]. However, these networks are too large for human interpretation and their context specific origin is often unaccounted in databases. Therefore, filtering these networks and identifying subnetworks that are important in a certain context (e.g. disease/health, tissue/cell) are major challenges that make up an active field of research.

An appealing approach for relevant subnetwork identification is to model the flow of biological information (e.g. cell signalling) using chains of interactions. In the case of protein signalling this means that every protein in a chain can modify the consequent protein, transmitting a biological signal (the alternative term "path" is avoided here to prevent confusion with signalling pathways). Multiple alternative chains which allow to traverse from a start to an endpoint may exist within a network. Following this logic Scott et al. [11] successfully developed an algorithm to identify protein signalling cascades in a protein network for pathway discovery purposes. They used interaction reliability and functional enrichment based scoring to calculate the significance of the chains. They showed that this technique has a potential in recovering known pathways in yeast, however, their algorithm lack context specificity and is not publicly available. Other methodologies use gene expression data to get more context specific results. Teku et al. [12] developed a filtering method to identify a core T cell network using the immunome interactome. They used a co-expression based weighting of the interaction network to compute the significance of the links. However, expression based specificity is not the only factor defining the importance of a protein in an added context. Functional module identification methods based on topological structures of unweighted PPI networks are another active area of research. For example lately, Liekens et al. [13] introduced a solely network based methodology for gene prioritisation using an integrated interaction network. According to the assessment of the authors, this method, despite its exclusively topology based search algorithm, was reported to outperform earlier gene prioritisation algorithms based on data fusion of heterogeneous data sources [13]. Recent reviews on pathway discovery approaches provide further examples for the interested readers [14, 15].

Here, we present ChainRank, an enhanced search and prioritisation tool that allows combining multiple biological evidences (e.g. topology, experimental molecular data from expression to proteomics and metabolomics, literature evidence, meta-analysis results, phenotype association) as scores. Similarly to the work of Scott et al. [11], our method uses a chain based network search algorithm to retrieve chains linking user defined start and end nodes, e.g. biomarkers associated with a disease state. In this work, we show that combining different context specific and topological scores together with a chain based search approach that simulates real interaction mechanisms – instead of focusing on individual biological elements or their associations – can improve the prediction of underlying pathway mechanisms. We introduce a framework over the search algorithm that can incorporate multiple user defined scores and thus is able to contextualize search results to e.g. disease states or tissues. Furthermore, we show that this framework can evaluate the combined effect of these scores to simulate complex phenotypes, e.g. tissue specific effects of a certain disease. According to our knowledge this is the first method relying on a chain based approach that is able to incorporate various scores and combine them and this is the first study showing the effect of combining different scores.

To assess ChainRank, we evaluated three scores (topological, tissue specific and disease state specific) to prioritise chains within a PPI network and evaluate them against known gold standard signalling pathways. We focused our analysis on muscle dysfunctions in chronic obstructive pulmonary disease (COPD) because of its specificity to a distinctive tissue, and also because of its clinical relevance. We introduce two complex, biologically motivated scores that we created integrating multiple differential expression studies as well as expression, protein and metabolite data to describe tissue- and disease wise importance of the network proteins. We also present a score describing topological importance and show the combined effects of the developed scores. Evaluating the precision and recall of finding gold standard (GS) proteins in our top scoring results, we show a considerable increase in precision with comparably good recall rate, compared to a simulated random scoring. Furthermore we show that combining different scores can further improve the performance of the prioritisation. The results demonstrate that our method can effectively identify pathway elements in a context specific manner. Potential use cases are the identification of disease specific networks, assessment of pathway interactions, simulation of the spread of perturbing effects amongst networks (mode-of-actions) and the elucidation of mechanistic relations between biomarkers.

Our method is implemented in the popular R framework and freely available at https://github.com/atenyi/ChainRank.

Methods

The ChainRank method consists of two main steps. The first step searches for all chains connecting start and end nodes in a network (Fig. 1c-d). For example given a start node S which interacts with node C1 which interacts with proteins C2 and E1 (Fig. 1c), as such we define two chains between S and E, namely S-C1-E1 and S-C1-C2-E (Fig. 1f). The next step involves annotating the network nodes with scores and computing the chain scores and p-values to provide a ranking and selection (Fig. 1e-f).

Chain search

The chain search step is used directly to evaluate all potential chains connecting start and end nodes within the initial network. This task translates to the "all simple paths" graph theoretical problem [16] that seeks to find all simple (non-cyclical) paths between two vertices. A graph of n vertices contains n! simple paths which makes a brute force search an NP hard problem. However, for signalling and gene regulatory networks the biological relevance of connections between two entities diminishes with increasing relative distance, i.e. the given distance relative to the shortest distance [17, 18]. Therefore, the problem can be addressed by introducing a depth limit for the search that is greater or equal to the distance of the shortest path linking the start and end nodes. This problem can be optimally solved by a depth limited depth first search (DFS) algorithm. The basic DFS algorithm traverses the network from the starting node and explores a branch of the network before backtracking (Fig. 1c). Using a depth limit the search is halted if a chain would exceed a specified k maximal length (depth limit) which is defined as the number of nodes a chain contains (Fig. 1b). This algorithm has O (b^k) time complexity, where b is the branching factor of the graph and due to its exhaustive nature it finds an optimal solution within the depth limit k [19]. We implemented a recursive version of this algorithm and extended it to be able to search simple paths amongst multiple start and end nodes. Chains connecting start and end nodes are stored and serve as the output of the algorithm. The method was implemented in R programming language. The pseudo code of the algorithm is detailed in the Additional file 1: Text S1.

Scoring and prioritisation using p-values

In order to create a general prioritisation framework, we introduced the concept of element scores. Such scores are mapped to network nodes and describe a specific

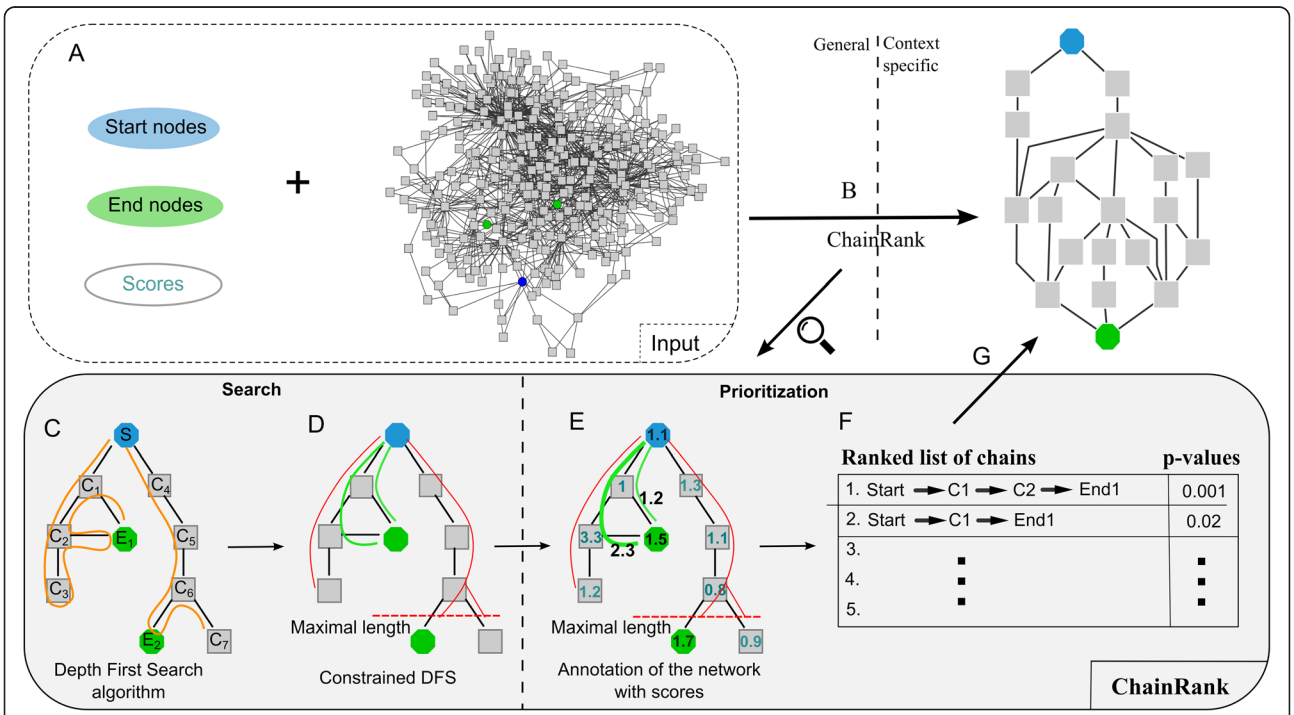


Fig. 1 Schematic overview of the ChainRank method (**a-b**) and its workflow (**c-f**). **a** The input parameters of the algorithm are the investigated research targets (start and end nodes), a network and the defined scores. **b** The ChainRank method produces a context specific subnetwork specific to the research targets. **c** The method is based on a depth-first search (DFS) algorithm that traverses the network from the starting node and explores a branch as far as possible before backtracking. **d** DFS is constrained to search only chains that has a maximal length or smaller. **e** Network nodes are annotated with scores. Chain scores then calculated by the sum of the scores of chain elements, normalized by the length of the chain. **f** The significance of the chains are calculated using the chain scores. **g** Most significant chains are selected and used for the construction of a context and target specific subnetwork

property of a biological entity that the node represents. This score can include both topological and biological characteristics (e.g. the connectivity of a node or tissue specific expression of the protein/gene that the node represents or experimental support for a protein-protein interaction) (Fig. 1) and a node can hold one or more separate scores. We used these measures to characterize the interaction chains. Our aim was to maximize the overall score of the nodes in a chain, therefore we used the sum of their element scores to calculate the chain scores. Furthermore, to exclude length based biases we normalized this score by the length of the chain to get the final chain score, thus $S = \sum_i^l s_i/l$ where S denotes the chain score, l is the length of the chain and s_i is the score of the i^{th} element of the chain.

Certain research situations involve several biological contexts, e.g. disease effects on specific tissue. To address such needs, we introduced the concept of combined scores. We introduced three different strategies to combine the scores: (i) Combined scores are calculated as the weighted product of the normalized element scores mapped to a node, using the formula $c_k = \sum_j^n w_j s_{kj}$, where c_k is the combined score of the k^{th} node, n is the number of scores, s_j is the j^{th} element score normalized to the range [0,1] and w_j is the weight corresponding to the j^{th} score, (ii) the filtering strategy pre-filters the chains using a threshold for the score s_1, and then it re-ranks the filtered chains with score s_2 and (iii) the intersection strategy keeps only those chains that are under a specified threshold for all the selected scores.

To evaluate the chain scores, we calculate the significance of the chains. We simulated random networks, constructed by shuffling the weights and edges of the initial network, while preserving the vertex degrees. For a given chain with score s, its score p-value is defined as the percentage of top-scoring chains in random networks that have score s or higher [11].

We also use the score p-value to generate the list of prioritised chains. Depending on the application a score p-value cut-off can be utilized to select the most significant chains or alternatively the top scoring n chains can be selected. Assembling the filtered chains allows for the reconstruction of a subnetwork that is specific to the start and end nodes and to the context the score defines.

Evaluation and performance

To evaluate a computational method one can either apply a measure of stability by cross-validating multiple runs or, ideally, derive precision and sensitivity information from comparison against a standard of truth. As described in the introduction there is a lack of context aware pathways which could be used as standard of truth. In order to evaluate the results of the ChainRank we therefore validate our method on two levels. First, the significance of the chain scores is evaluated. Second, a reference pathway is used as a validation set and the enrichment of its members in the top results or the ranked chains is assessed for the evaluation. This validation set is referred to as the gold standard (GS). To judge the stability of the method we compute the precision and recall of the top n chains or alternatively use a p-value cut-off. For the validation, positives (P) are defined as the validation elements represented in the input network but not included in the start and end proteins. To determine the precision, the occurrence of the validation set elements are counted in the top chains (excluding start and end proteins), i.e. the true positives (TP), while non-validation set elements represents the false positives (FP). Thus, $Precision = TP/(TP + FP)$ and $Recall = TP/P$. Due to the lack of well-defined GS, reaching high precision values is a highly challenging task. Therefore to represent our results in a more informative way we defined the metric of improvement. To compute the improvement of a ranking we simulate a random score, i.e. we perform a random sampling from the chains to select the top results. Then, we compute *ovement* = (*Precision of ranking*)/(*Precision of random ranking*).

Results

In order to assess the performance of our method we studied its applicability in protein interaction network based pathway reconstruction. We specified the domain of interest to muscle dysfunctions in chronic obstructive pulmonary disease (COPD) because of its specificity to a distinctive tissue, its clinical relevance as well as the wealth of literature mining and experimental data available for our analysis [20]. We designed two application cases, each with a specific GS pathway (Table 1.). First, we aimed to recreate a subnetwork of the IGF-Akt pathway [21] describing regulation of protein synthesis, an important aspect of muscle remodelling (Fig. 2a). In the second case, our goal was to represent the disease specific involvement of parts of a canonical signalling pathway. We used disease specific varieties of the canonical MAPK pathway: the EGF-PI3K and ROS-TGFa-EGFR pathways (Additional file 1: Table S1, Fig. S7), that are based on literature mining for COPD related signal transduction events [20]. We note that evidence for the involvement of these specific parts of the GS pathways is not excluding potential involvement of additional parts. For the evaluation we selected specific chains from these pathways defined by start and end proteins that we refer as gold standards (Table 1).

PPI network

For the investigations we utilized the complete human PPI network as the input network. At the time of the analysis it contained 1.6 million protein interactions that

Table 1 Overview of the networks used in the evaluation process and the gold standards. Gold standard representation is shown in the original PPI network and in the selected networks. *Edges* signify the number of edges connecting GS *Nodes* in the network

Application case	Network properties		Gold standard (GS)	Start protein	End protein	GS representation	
	Edges	Nodes				Nodes	Edges
Human PPI network	61872	10167	IGF-Akt pathway	-	-	13	20
			COPD specific MAPK	-	-	21	34
Muscle specific case	847	308	IGF-Akt pathway	IGF1	RPS6KB1	9	10
COPD related case	544	152	COPD specific MAPK	EGFR	SRF, CREBBP, ELK1, MYC	11	8

were collected and merged from different publicly available databases and integrated into the COPD knowledge base [20]. We quality filtered this network by including only those interactions that were supported by at least one piece of experimental evidence (in contrast to purely computationally predicted ones). This resulted in a PPI network of around 10,000 nodes and 62,000 interactions (Table 1).

Subnetwork selection and performance

Within this general PPI network we are only interested in the specific subnetwork that potentially connects our start and target set, here determined by the endpoints of our selected gold standard pathways. In order to retrieve this subnetwork as starting point for the ChainRank method, we applied the BioXM knowledge management environment network search tool [22]. This tool is based on a heuristic breadth-first search algorithm, allows nodes to be preferred or penalised based on their connectivity and it retrieves those nodes in the input network that have the potential to link targeted nodes within a *k* maximal distance. Consequently, with this step we omit those unnecessary nodes and edges that does not lead to any targeted endpoints in a *k* maximal length chain. Therefore, we decided to set the *k* distance cut-off for the breadth first search centered on the distance between the start and the target in our reference GS path. Furthermore, Baudot et al. [18] showed that canonical signal transduction pathways are enriched for highly connected protein hubs; therefore, we set the algorithm to encourage the integration of canonical interactors preferring highly connected proteins. We generated two subnetworks (IGF-Akt proximity and MAPK proximity subnetworks, Table 1.). Because heuristic subnetwork generation methods introduce an element of variability, we evaluated its effect by creating further networks with different parameterisation and analysed them in terms of their overall influence on the ChainRank results which was not significant (Additional file 1: Table S2 and S4).

As an alternative to the heuristic network selection step the ChainRank method could be used to evaluate all potential chains of a given maximal length within the overall network. However, the corresponding computational requirements quickly become prohibitive as longer

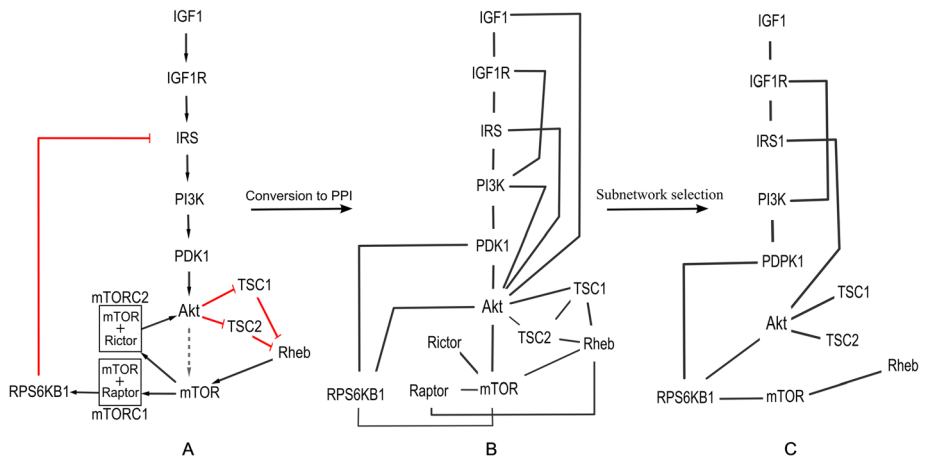


Fig. 2 Changes in the representation of IGF-Akt pathway. **a** shows the protein synthesis regulation related part of IGF-Akt pathway [21]. In this pathway representation relations of proteins are well annotated with their directionality, type of interaction, etc. **b** shows that during manual conversion of the canonical pathways to PPI network representation (by retrieving all PPI interactions between the fixed set of e nodes) this kind of information is lost and only undirected edges represent physical binding between proteins. Moreover, the structure of the pathway is altered by extra edges and some connections are missing. **c** shows the effect of the network selection which removes redundant edges but retains most of the nodes

chains are explored in dense networks (see Chain search). Runtime of the chain search for the muscle-specific network (314 nodes, 865 edges) with a maximal length of 8 is 14.5 min on a 2.4 GHz processor, finding more than 9000 chains. In addition, we note that the size of the network that the ChainRank method can process in realistic time depends strongly on the network complexity (more runtime data on different networks is available in Additional file 1: Table S3).

Evaluation of the input network

In order to set a realistic gold standard (GS) for the evaluation we analysed the changes in the canonical GS during its manual conversion to a PPI representation and then the effect of the network selection (Fig. 2). In canonical pathways relations of proteins are manually selected and well annotated with their directionality, type of interaction, etc. During the conversion of these pathways to a PPI network representation the annotation is lost and only physical interaction without pre-selection are depicted. Therefore edges appear/disappear during the conversion and protein complexes become individual, interacting nodes. These findings show the high complexity of searching in PPI networks and demonstrate that the exact recreation of a canonical pathway cannot be the ultimate metric of the evaluation process but rather the relative improvement between unranked and ranked searches.

Scores

As mentioned in the introduction there are several methods that use gene expression data to investigate domain specific traits. While ChainRank is able to incorporate gene expression scores, here we focus on more complex scores to represent localisation or disease relevance. We also introduced a topology based score.

1. Localisation score: To show the capabilities of the method in tissue-specific filtering we created a muscle specificity score. Using this prioritisation with the ChainRank method would result in those interaction chains that contain mostly muscle specific proteins being highly ranked. To create this score we collected publicly available gene expression measurements from Gene Expression Omnibus (GEO) [23], studying a large amount of different conditions in different tissues. We compared the mean variability of the genes' expression value in muscle to their mean variability in the rest of the body. Genes with highly variable expression levels under different conditions in muscle but lower variability in other parts of the body receive higher scores while genes that are not typically variable in muscle or are variable throughout all tissues receive lower score. The corresponding proteins were mapped to genes to be applicable for PPI network based analyses. Details on the included data sets and the exact methodology can be found in the Additional file 1: Text S2.
2. Relevance score: This score describes the relevance of a protein in a specific biological process — in this case a disease. To generate a disease specific score we used studies that investigated the effect of COPD on skeletal muscle and other mechanisms that related to this disease. The selected studies incorporated diverse experimental paradigms such as proteomics, metabolomics and gene expression. From these studies we extracted all genes or proteins (depending on the type of analysis) that were shown to be significantly changed in the disease context. Then we computed the score by counting how many times a gene/protein occurs with high significance in any of these study results. The first study we utilized investigated the training effect on the muscle of COPD patients [24] integrating measurements of gene expression, metabolism and protein carbonylation [25–27]. In addition, as part of this research study, the effect of angiogenesis on gene expression in young (<30 year) and elderly (>60) persons was examined (detailed in the Additional file 1: Text S3). Finally, an analysis on inactivity-induced wasting in mouse glycolytic muscle was used to construct the score [28] (detailed in Additional file 1: Text S3). We used HomoloGene [29] to find homologous human genes for the mouse genes and we mapped the genes to the related proteins in all the studies.
3. Connectivity score: We used a topology based score to characterize the degree centrality of the proteins in the network. We reversed the degree centrality to compute the score, thus $Connectivity\ score(v) = |dc(v) - \max(dc(V))| + 1$, where $dc(v)$ is the degree centrality of $v \in V$ vertice. This score is a good measure to distinguish between general hub like proteins with high degree centrality (and thus with low scores) and specific proteins with lower degree centrality (and thus high scores).

To test the sensitivity of our algorithm to different scores that explain similar biological phenomena, we introduced two additional scores from external data sources. As an alternative to Localisation score, we retrieved the Tissue Specificity (TS) score from the Human Protein Atlas [30], which corresponds to the score calculated as the fold change to the second highest tissue (for further information see Additional file 1: Text S4). As an alternative to Relevance score we created the Fold change (Fc) score, which we retrieved from a recent publication that reported RNA-seq data for 98 COPD subjects and 91 controls [31]. Score was computed as $Fc = \log2(COPD/control)$, where COPD and control is the gene expression value of the signed group.

Evaluation of the scores: distribution, correlations and the length of the chains

In order to check for the independence of our selected scores we examined their correlation and their relation to the length of the chains. We used the IGF-Akt proximity subnetwork, with maximal length 8 for this analysis. Figure 3a shows that the expression and relevance scores show a slight correlation which can be explained by the fact that in this case the relevance score (among other aspects, such as protein carbonylation and metabolites) includes data on gene expression in muscle tissue. Therefore, although the relevance rank is based on experiments with specific environmental factors, the expression data is expected to show some correlation with the general muscle expression measurements. The other variables are uncorrelated, therefore we can assume that the different ranks explain different properties of the chains. We found that normalisation of the chain scores by the number of chain nodes removes most of the length dependency (Fig. 3b). We note that different topological properties of the networks might have effect on the connectivity scores' length dependence. Furthermore, we showed that the distribution of scores in the generated subnetworks (Subnetwork selection and performance, Table 1.) represents well the distribution of scores over the whole PPI network (Additional file 1: Figure S1).

Evaluation of the performance of the ChainRank method

Having prepared the networks, we applied the ChainRank method on them. To determine the maximal length parameter for the analysis we took into consideration the distance of the start and end proteins in the GS. For the muscle specific application the canonical distance would be 9, however, due to the differences of the PPI representation of complexes (see in Evaluation of the input network, Fig. 2b) we rationalized using a maximal length 8. For the COPD specific application we used 7 for maximal length, following similar reasoning. In the evaluation process we assessed the improvement of the different scores in finding GS proteins in the top ranked results compared to random prioritisation. We evaluated the performance both by using only individual scores to rank and also by combining the scores. Figure 4 details the dependence of performance on different p-value cut-offs.

For the muscle specific case we ran the ChainRank using the IGF-Akt proximity subnetwork and maximal length 8, retrieved 9351 chains. For the COPD application case (MAPK proximity subnetwork) we computed the chains with maximal length 7, finding 71838 chains. In this case Relevance scores showed high discrepancy from normal distribution therefore the introduced p-value calculation can be misleading for this score. Instead, we show our results by the number of top chains in this scenario.

In the muscle specific scenario, results show that the Connectivity score has the highest improvement of the scores (Fig. 4a). Detailed analysis reveals that this score show especially high improvement with very low p-values however, with growing p-values this improvement quickly decreases to an average of factor 1.8-2 for significant chains. Furthermore, in the top 5 chains Connectivity already finds one of the shortest GS path represented in the input network (Fig. 2c), i.e. IGF1-Akt-mTOR-RPS6KB1. Localisation also introduces an improvement of factor 1.5 amongst the significant chains and maximizes the Recall under 0.001 p-value (Additional file 1: Figure S2). In the MAPK scenario the Relevance score outperformed the other scores showing consistent improvement in top chains (Fig. 4b, Additional file 1: Figure S3). We analyzed the robustness of the algorithm by comparing the performance of Localisation score to TS score in the Muscle specific case and Relevance to Fc score in the COPD specific case. Results showed that the method produces similar improvement for the scores in these scenarios (Additional file 1: Figure S6) and thus it is robust to changes of the scores.

We also investigated the performance of the defined combined strategies. We computed the Combined score as the equal weighted sum of the three normalized scores and evaluated its improvement. With these settings this score could not improve over the best individual scores and therefore we do not report further results. Furthermore, we applied the filtering strategy for both scenarios. For the IGF-Akt case we used a Connectivity filter before evaluating the chains by the Localisation score. We applied a threshold of 0.05 for the filtering. This method introduces a strong and stable increase in improvement (Fig. 4a) which shows good applicability in arbitrary sized subnetwork retrieval. For the MAPK application we investigated the effects of COPD on muscle, therefore we used Localisation filtering and evaluated Relevance on the reduced list of chains. We used the top quartile of the ranked chains to set a filtering threshold. Together with the intersection strategy, in which we applied the same parameters, filtering introduced comparable improvement to Relevance score. To conclude we showed that combining different scores can improve the prediction power of the algorithm and they are capable to mimic complex biological contexts.

We evaluated the receiver operating characteristic (ROC) curve and the area under curve (AUC) (Fig. 5) which shows the significant improvement over random scoring. Next, we investigated the effect of the maximal length parameter on the improvement of the chain scores. We found that length does not have a significant effect on the ranking performance (Additional file 1: Figure S4).

Finally, we identified relevant thresholds that can be used to construct significant subnetworks and recreate the target pathways. Taking into account the improvement-

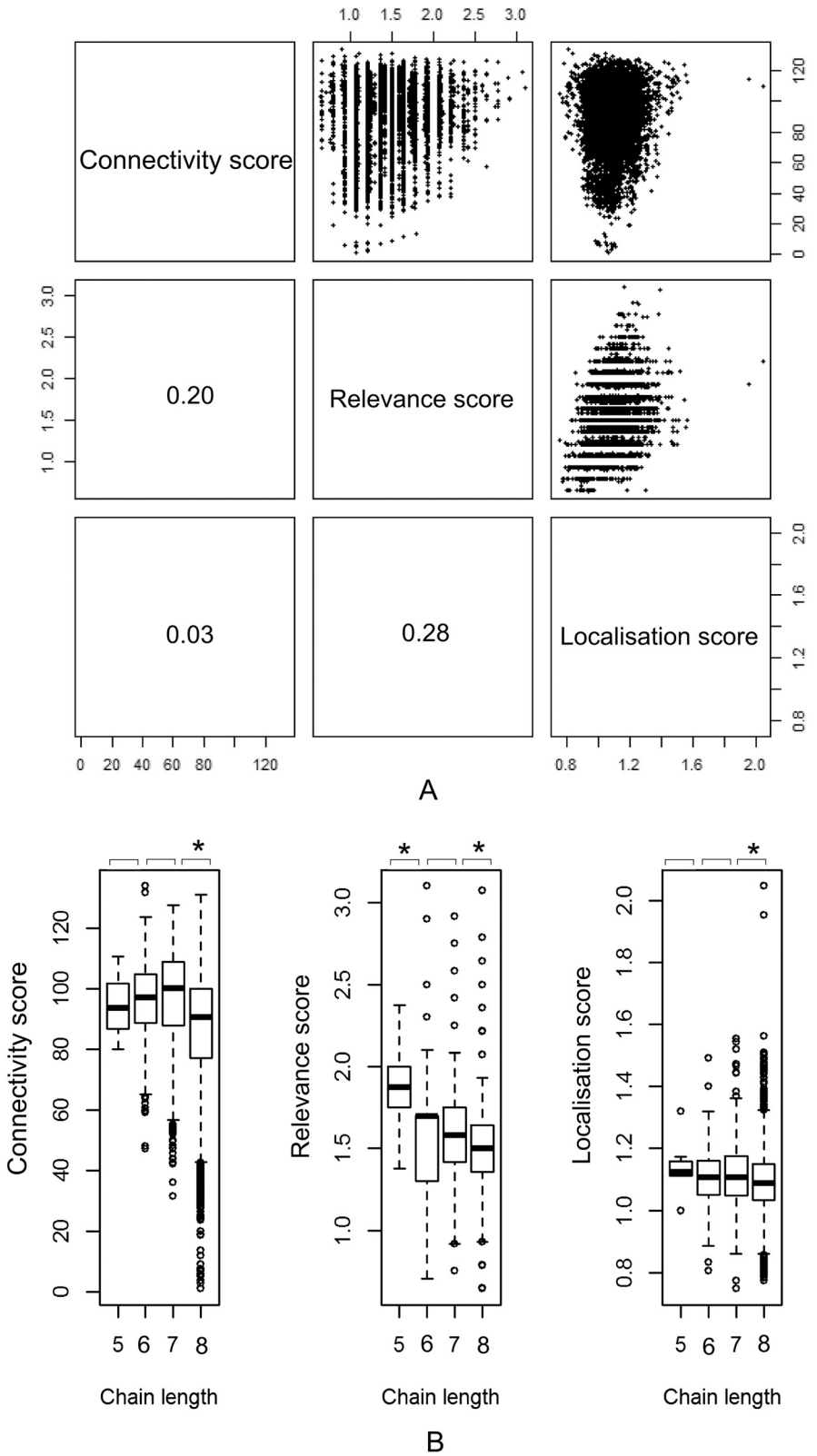


Fig. 3 Statistical evaluation of the scores. **a** shows the correlation between the chain ranks, correlation values are indicated in the lower triangle. **b** shows the relation of the length of the chains to the chain scores. Statistical significance between the different length chains' scores is indicated (*p ≤0.05)

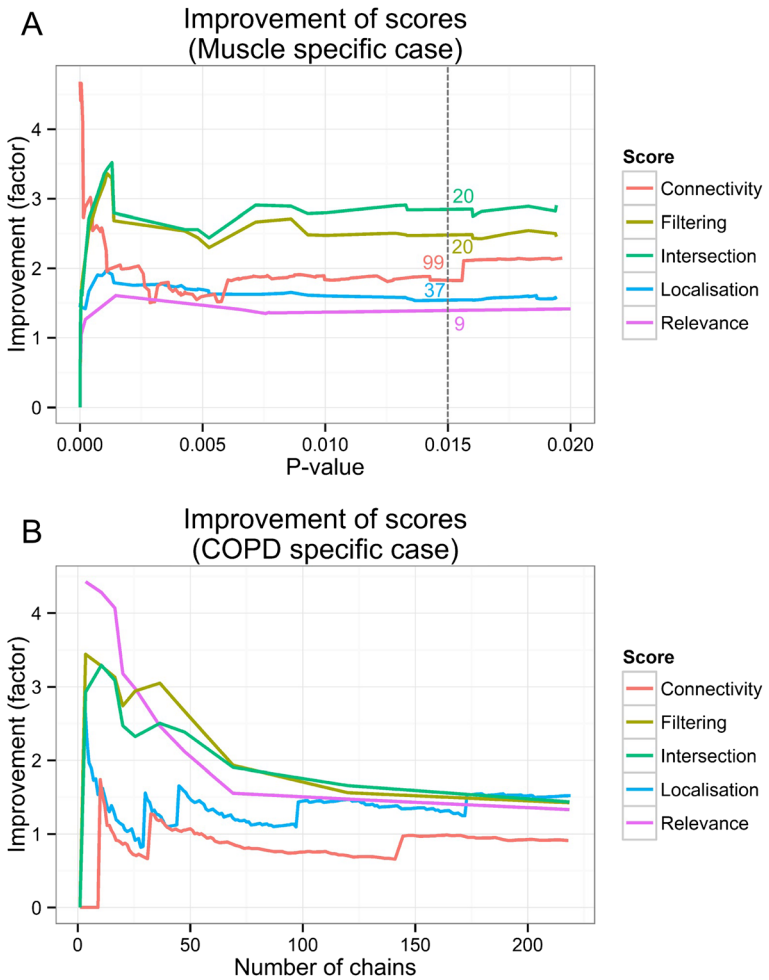


Fig. 4 Improvement of the different scores. **a** Muscle specific case: Intersection is defined as the common chains that have both a p-values ≤0.05 with Connectivity and Localisation score. These chains are shown with their Localisation score p-values. For Filtering chains with Connectivity score p-values ≤0.05 were selected, re-ranked and evaluated by Localisation score. Number of chains are indicated at p =0.015 for each score. **b** COPD specific case, Intersection here is defined as the common chains in the top quartile of chains ranked by Localisation score and Relevance score. These chains are shown with the number of top chains ranked by Relevance. For Filtering the top quartile of chains ranked by Localisation score were selected, re-ranked and evaluated by Relevance

recall trade-off we found a p-value of 0.015 or the number of chains 50 as a good cut-off value. With these thresholds we show high improvement over random in finding targeted GS proteins (Fig. 6). Assembly of chains under the cut-off value shows that the algorithm finds the main chains connecting the targeted start and end proteins and identifies relevant alternative chains with a recall of 67 % and a precision of 30 % (Figs. 6, 7 and Additional file 1: Figures S2, S8). As a further evaluation of the approach, we show that the distribution of scores in the recreated pathways are different from the original network. In the recreated pathways, the distribution means of the simple scores are shifted to higher values. The combined scores can further alter this effect, producing a score distribution that resembles more to the GS (Additional file 1: Figure S5) indicating that the scores indeed capture biological context.

Discussion and conclusion

In recent decades huge amounts of data have been accumulated in biological research but up to now these valuable data sources remain underutilized in terms of applications for integrative analysis and data mining. Systemic use of biological data could help to create more personalized and contextualized information and overcome the current rigid and generally simplistic representations of mechanisms involved in biological processes and their regulation. This calls for bioinformatics tools that can facilitate data analysis and help in the interpretation of these huge datasets. Biological networks could play an important role in this procedure as they have already shown their utility in many applications. Current high-throughput methods, however, are prone to errors e.g. in yeast two-hybrid systems high false positive rates and platform-specific biases [32] still

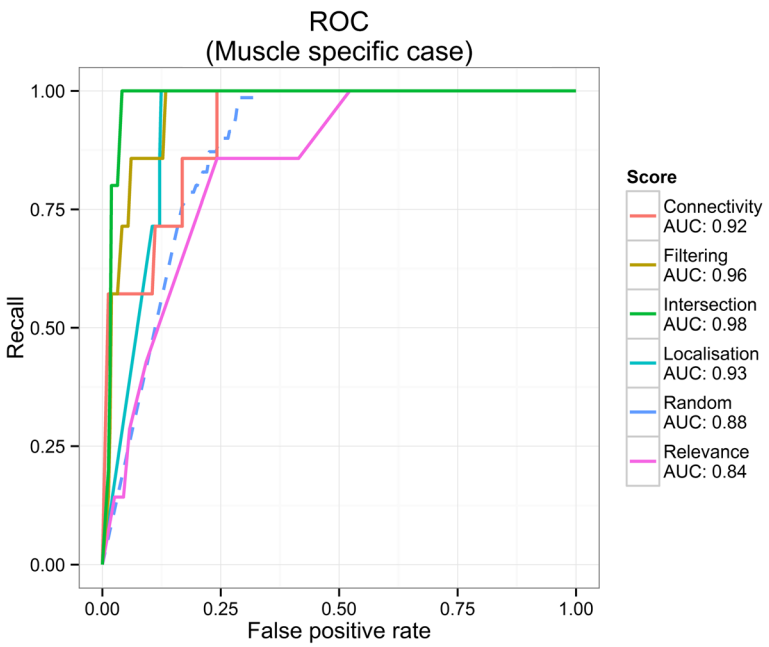


Fig. 5 ROC curve and AUC of the muscle specific case. Localisation and its combined scores with Connectivity shows the highest AUC. Random score (*dashed line*) already has an increased performance over completely random guess (*diagonal line*, not shown), which can be accounted for by the constraints introduced by the underlying network topology. AUC values appear below the names of the scores

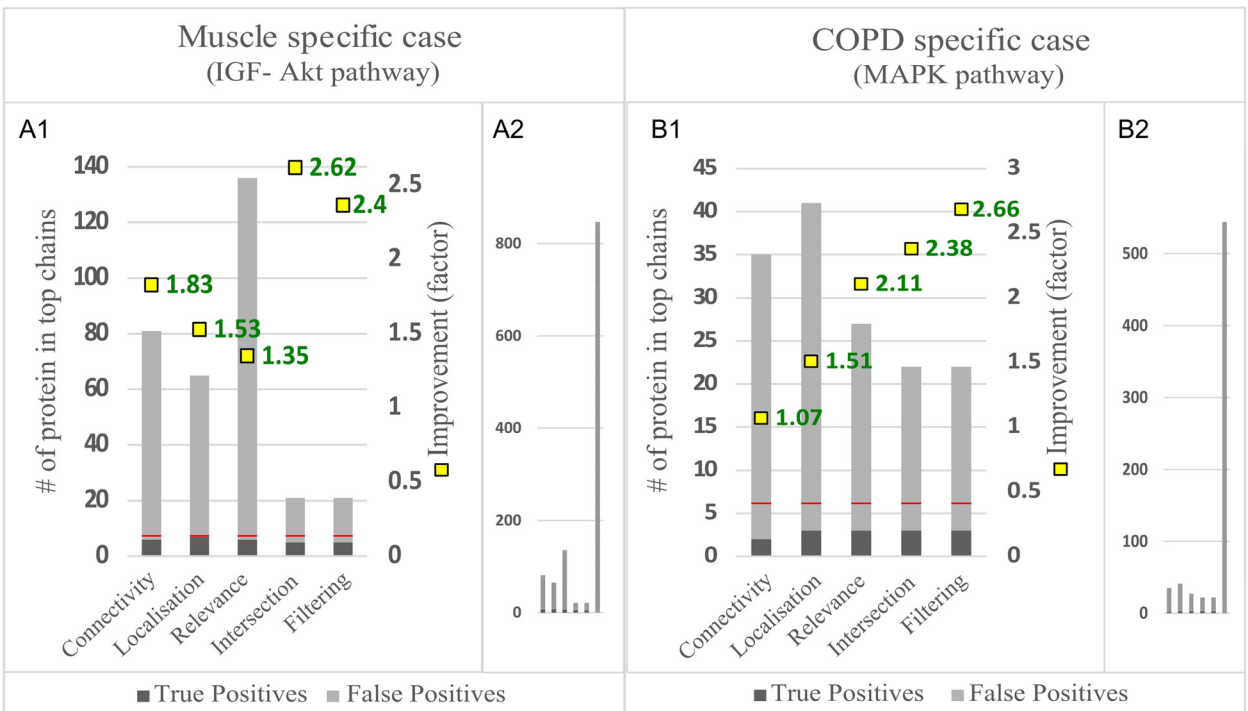


Fig. 6 Results of the evaluation after cut-off. (A1) For the muscle specific case 0.015 was used as cut-off threshold. (B1) shows the performance of COPD specific case evaluating the top 50 chains. Red lines show the number of GS proteins (positives), second axis shows the improvement. To represent the enrichment capabilities of the method we compare the recreated pathways to the input network (A2, B2)

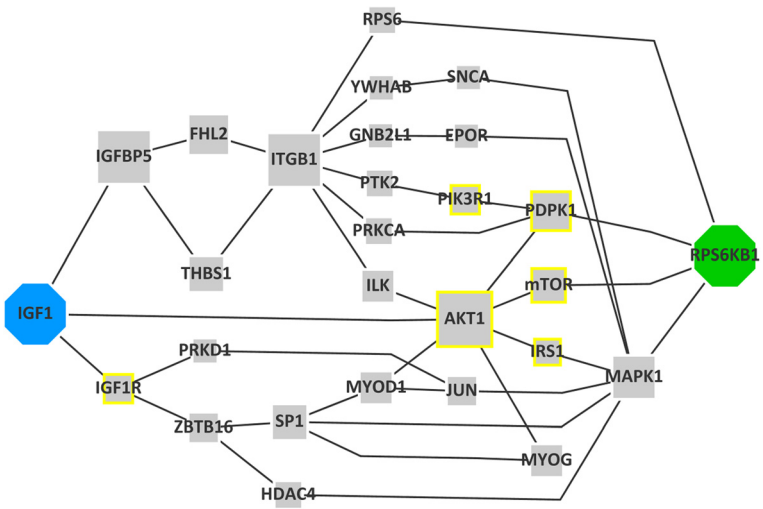


Fig. 7 The recreated IGF-Akt pathway. The results of ChainRank were filtered by taking the intersection of the chains that has lower Connectivity and Localisation score p-values lower than 0.05, then the ones with p-values ≤0.015 were assembled into a network (Fig. 6 (a) *Intersection*). The size of the nodes represents the occurrence of a protein in the top chains. Octagons indicate the start and end proteins, nodes with yellow border shows the gold standard proteins

remain problematic. As a result, inconsistencies could be present in the PPI networks that create alteration in the representation of signalling pathways [4, 33] which our results also confirmed (Evaluation of the input network).

The ChainRank method introduces a data-driven biological search tool that can be applied in widespread research situations. Our goal was to create a tool that can retrieve context specific subnetworks by using different evidences (e.g. expression profile, literature mining). Evaluating a specific application case is a complex task, which we addressed by recreating selected gold standard pathways.

Overall, our evaluation results showed that the generated scores can create domain specific effects. We showed that filtering the chains by scores and intersecting top scoring chains can create improvements in precision and can be applied to simulate complex biological contexts. Although this evaluation is limited only to a few contexts (muscle and COPD) we believe that it gives a representative result to show the general applicability of the method and encourage its usage. Using the three developed ranks, we showed a 50 % improvement (factor 1.5), on average, in the precision of finding gold standard proteins in our top ranked chains. We also showed that combining ranks, for example by pre-filtering with one score before ranking by another, can improve the precision by up to a factor of 2.5. We achieved as high as 11 % improvement in the area under the receiver operating curve (AUC) (Fig. 5) which compares favourably with Bader's results [34] who reports a similar improvement but with a less generic framework and using protein complexes as a gold standard. Our results are comparable to [12] and [11] who use signal transduction pathways in yeast and human respectively as gold standards and report recall of 50–85 %, and precision of 18–42 %. Therefore, our method generalises the achievements introduced by Scott et al. [11] and Teku et al. [12] by introducing additional, non-expression based evidences and allowing to tune for multiple contexts such as tissue specificity or disease association. We were able to replicate our results with different pathways (IGF-Akt, COPD specific MAPK sub-pathways) and different initial conditions (different input networks). Overall the evaluation showed strong evidence that the method provides improved specificity to generate context-specific networks and therefore supports the viability of the concept.

Although we only showed the applicability of our methodology using PPI networks and in two different contexts (muscle and COPD), it is a generic tool that is applicable for various network types, like metabolic networks or disease networks. Integrated networks incorporating several interactome layers, like proteomics, metabolomics, diseases, etc. can also be used with the method. In addition, scoring criteria can be easily created using various private and public data sources. Although, the new criteria would have to be validated, the accumulation of different context profiles could pave the way for an integrated analysis framework. The differences in performance of individual scores in different biological context (Fig. 4) underscore the importance of appropriate selection of scores depending on the scientific question.

The method can be utilized to analyse many research questions, for example: a) given a set of data-driven associations, e.g. oxidative stress and proteolysis, what is the most likely causal, mechanistic connection in a given

context? b) what are the common mechanisms driving different diseases, e.g. systemic effects of COPD and diabetes mellitus type 2? c) can computational modelling be supported by reducing the number of interactions to the biologically most relevant ones and thereby generate manageable complexity [35]? Another promising application field could be the analysis of synthetic lethality in the context of anticancer therapy. By providing evidence-supported alternatives to classical consensus pathways ChainRank could open up new avenues of investigation. A possible avenue is the improvement of the search algorithm for example to use "information propagation" methods [36] to include information from the neighbourhood of a chain into the ranking and thereby see whether biological modularity can be used to further enhance the context specificity of the results. Another interesting aspect would be to implement and compare the current exhaustive search with a heuristic search algorithm that is possibly usable on the full multi-million node and association network that makes up our current biological knowledge.

Availability

Project home page: https://github.com/atenyi/ChainRank
Programming language: R

Additional file

Additional file 1: This file contains supplementary tables, figures and further information on the scores and the search algorithm. (PDF 1.43 MB)

Abbreviations

AUC: area under curve; COPD: Chronic Obstructive Pulmunary Disease; FP: false positive; GS: gold standard; P: positive; PPI: protein-protein interaction; ROC: receiver operating characteristic; TP: true positive.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AT has written the software code, executed the evaluation and prepared the manuscript together with DM who also supervised the software development and co-analysed the evaluation results. DM, DGC, FF made substantial contributions to conception and design of the method. DGC prepared the Localisation score. KC prepared the study measuring the effect of angiogenesis on gene expression in young (<30 year) and elderly (>60) persons that was used for the Relevance score. PA, IC, MC and JR participated in designing the evaluation, drafting the manuscript and coordinating the study. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Peter Davidson who provided the analysis results on the study of inactivity-induced wasting in mouse glycolytic muscle. This research was supported by the Spanish Government and FEDER Funds SAF2011-25726, the AGAUR (2009SGR911 and 2009SGR-1308), the European Commission (FP7) grants SYNERGY-COPD (no. FP7-ICT-2009-270086) and METAFLUX (PITN-GA-2010-264780) and the Icrea Academia award (granted to M. Cascante).

Author details

[1]Hospital Clínic-Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Research Institute, Universitat de Barcelona, C/Villarroel 170, 08036 Barcelona, Spain. [2]Departament de Bioquimica i Biologia Molecular, Facultat de Biologia-IBUB, Universitat de Barcelona, 08028 Barcelona, Spain. [3]Unit of computational Medicine, Center for Molecular Medicine, Department of Medicine, Karolinska Institute and Karolinska University Hospital, SE-171 76 Stockholm, Sweden. [4]Centro de Investigación en Red de Enfermedades Respiratorias (CibeRes), 07110 Palma de Mallorca, Spain. [5]Integrative Systems Biology, University of Liverpool, L69 3BX Liverpool, UK. [6]Biomax Informatics AG, D-82152 Planegg, Germany.

Received: 18 June 2015 Accepted: 17 December 2015
Published online: 05 January 2016

References

1. De Chassey B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaugué S, et al. Hepatitis C virus infection protein network. Mol Syst Biol. 2008;4:230.
2. Huang PH, Mukasa A, Bonavia R, Flynn RA, Brewer ZE, Cavenee WK, et al. Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. Proc Natl Acad Sci U S A. 2007;104:12867–72.
3. Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, et al. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science. 2000;287:873–80.
4. Kirouac DC, Saez-Rodriguez J, Swantek J, Burke JM, Lauffenburger DA, Sorger PK: Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks. BMC Syst Biol 2012, 6:29.
5. Natarajan M, Lin K-M, Hsueh RC, Sternweis PC, Ranganathan R. A global analysis of cross-talk in a mammalian cellular signalling network. Nat Cell Biol. 2006;8:571–80.
6. Ben-Hamo R, Gidoni M, Efroni S. PhenoNet: identification of key networks associated with disease phenotype. Bioinformatics. 2014;30:2399–405.
7. Li Y, Agarwal P, Rajagopalan D. A global pathway crosstalk network. Bioinformatics. 2008;24:1442–7.
8. Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, et al. High-throughput mapping of a dynamic signaling network in mammalian cells. Science. 2005;307:1621–5.
9. Bandyopadhyay S, Chiang C, Srivastava J, Gersten M, White S, Bell R, et al. A human MAP kinase interactome. Nat Methods. 2010;7:801–5.
10. Vidal M, Cusick ME, Barabási A-L. Interactome networks and human disease. Cell. 2011;144:986–98.
11. Scott J, Ideker T, Karp RM, Sharan R. Efficient algorithms for detecting signaling pathways in protein interaction networks. J Comput Biol. 2006;13:133–44.
12. Teku GN, Ortutay C, Vihinen M. Identification of core T cell network based on immunome interactome. BMC Syst Biol. 2014;8:17.
13. Liekens AM, De Knijf J, Daelemans W, Goethals B, De Rijk P, Del-Favero J: BioGraph: unsupervised biomedical knowledge discovery via automated hypothesis generation. Genome Biol 2011, 12:R57.
14. Chen B, Fan W, Liu J, Wu F-X. Identifying protein complexes and functional modules–from static PPI networks to dynamic PPI networks. Brief Bioinform. 2014;15:177–94.
15. Mitra K, Carvunis A-R, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. Nat Rev Genet. 2013;14:719–32.
16. Black PE, Pieterse V. "all simple paths." In: Dictionary of Algorithms and Data Structures [online]. 2008.
17. Poyatos JF, Hurst LD. How biologically relevant are interaction-based modules in protein networks? Genome Biol. 2004;5:R93.
18. Baudot A, Angelelli J-B, Guénoche A, Jacq B, Brun C. Defining a modular signalling network from the fly interactome. BMC Syst Biol. 2008;2:45.
19. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 3rd ed. 2009.
20. Maier D, Kalus W, Wolff M, Kalko SG, Roca J, Marin de Mas I, et al. Knowledge management for systems biology a general and visually driven framework applied to translational medicine. BMC Syst Biol. 2011;5:38.
21. Schiaffino S, Mammucari C. Regulation of skeletal muscle growth by the IGF1-Akt/PKB pathway: insights from genetic models. Skelet Muscle. 2011;1:4.
22. Cano I, Tényi Á, Schueller C, Wolff M, Huertas Migueláñez MM, Gomez-Cabrero D, et al. The COPD Knowledge Base: enabling data analysis and computational simulation in translational COPD research. J Transl Med. 2014;12 Suppl 2:S6.
23. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207–10.
24. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, et al. A systems biology approach identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease. PLoS Comput Biol. 2011;7, e1002129.

25. Barreiro E, Rabinovich R, Marin-Corral J, Barbera JA, Gea J, Roca J. Chronic endurance exercise induces quadriceps nitrosative stress in patients with severe COPD. Thorax. 2009;64:13–9.
26. Rabinovich RA, Bastos R, Ardite E, Llinas L, Orozco-Levi M, Gea J, et al. Mitochondrial dysfunction in COPD patients with low body mass index. Eur Respir J. 2007;29:643–50.
27. Radom-Aizik S, Kaminski N, Hayek S, Halkin H, Cooper DM, Ben-Dov I. Effects of exercise training on quadriceps muscle gene expression in chronic obstructive pulmonary disease. J Appl Physiol (Bethesda, Md 1985). 2007;102:1976–84.
28. Bialek P, Morris C, Parkington J, St Andre M, Owens J, Yaworsky P, et al. Distinct protein degradation profiles are induced by different disuse models of skeletal muscle atrophy. Physiol Genomics. 2011;43:1075–86.
29. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L: Database resources of the national center for biotechnology. Nucleic Acids Res 2003, 31:28–33.
30. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science (80-). 2015;347:1260419.
31. Kim WJ, Lim JH, Lee JS, Lee S-D, Kim JH, Oh Y-M. Comprehensive Analysis of Transcriptome Sequencing Data in the Lung Tissues of COPD Subjects. Int J Genomics. 2015;2015:206937.
32. Huang H, Bader JS. Precision and recall estimates for two-hybrid screens. Bioinformatics. 2009;25:372–8.
33. Ramírez F, Schlicker A, Assenov Y, Lengauer T, Albrecht M. Computational analysis of human protein interaction networks. Proteomics. 2007;7:2541–52.
34. Bader JS. Greedily building protein networks with confidence. Bioinformatics. 2003;19:1869–74.
35. Gomez-Cabrero D, Menche J, Cano I, Abugessaisa I, Huertas-Migueláñez M, Tenyi A, et al. Systems Medicine: from molecular features and models to the clinic in COPD. J Transl Med. 2014;12 Suppl 2 Suppl 2:S4.
36. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. Nat Methods. 2013;10:1108–15.

MANUSCRIPT 2: NETWORK MODULES UNCOVER MECHANISMS OF SKELETAL MUSCLE DYSFUNCTION IN COPD PATIENTS

Tényi Á, Cano I, Marabita F, Kiani N, Kalko SG, Barreiro E, de Atauri P, Cascante M, Gomez-Cabrero D, Roca J. Network modules uncover mechanisms of skeletal muscle dysfunction in COPD patients. *J. Transl. Med.* BioMed Central; 2018; 16: 34.

- Journal: *Journal of Translational Medicine*
- Impact factor: *3.786*
- Quartile: *1st*
- Status: *Published*

Tényi *et al. J Transl Med (2018) 16:34*
https://doi.org/10.1186/s12967-018-1405-y

Journal of Translational Medicine

RESEARCH

Open Access

Network modules uncover mechanisms of skeletal muscle dysfunction in COPD patients

Ákos Tényi[1,2*], Isaac Cano[1,2], Francesco Marabita[3,4], Narsis Kiani[3,4], Susana G. Kalko[1,2,5], Esther Barreiro[2,6], Pedro de Atauri[7], Marta Cascante[7†], David Gomez-Cabrero[3,4,8†] and Josep Roca[1,2*†]

Abstract

Background: Chronic obstructive pulmonary disease (COPD) patients often show skeletal muscle dysfunction that has a prominent negative impact on prognosis. The study aims to further explore underlying mechanisms of skeletal muscle dysfunction as a characteristic systemic effect of COPD, potentially modifiable with preventive interventions (i.e. muscle training). The research analyzes network module associated pathways and evaluates the findings using independent measurements.

Methods: We characterized the transcriptionally active network modules of interacting proteins in the vastus lateralis of COPD patients (n = 15, FEV_1 46 ± 12% pred, age 68 ± 7 years) and healthy sedentary controls (n = 12, age 65 ± 9 years), at rest and after an 8-week endurance training program. Network modules were functionally evaluated using experimental data derived from the same study groups.

Results: At baseline, we identified four COPD specific network modules indicating abnormalities in creatinine metabolism, calcium homeostasis, oxidative stress and inflammatory responses, showing statistically significant associations with exercise capacity (VO_2 peak, Watts peak, BODE index and blood lactate levels) ($P < 0.05$ each), but not with lung function (FEV_1). Training-induced network modules displayed marked differences between COPD and controls. Healthy subjects specific training adaptations were significantly associated with cell bioenergetics ($P < 0.05$) which, in turn, showed strong relationships with training-induced plasma metabolomic changes; whereas, effects of training in COPD were constrained to muscle remodeling.

Conclusion: In summary, altered muscle bioenergetics appears as the most striking finding, potentially driving other abnormal skeletal muscle responses.

Trial registration The study was based on a retrospectively registered trial (May 2017), ClinicalTrials.gov identifier: NCT03169270

Keywords: Gene modules, Chronic obstructive pulmonary disease, Exercise training, Systems medicine, Muscular weakness

*Correspondence: tenyi@clinic.cat; jroca@clinic.cat
†Marta Cascante, David Gomez-Cabrero and Josep Roca are senior authors
[1] Hospital Clinic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain
Full list of author information is available at the end of the article

Background

Patients with chronic obstructive pulmonary disease (COPD) show marked individual variability of both clinical manifestations and disease progression with relevant implications on prognosis and management [1].

The 2017 GOLD update [2] recommends lung function measurements (FEV_1) to assess COPD severity; whereas both symptoms intensity and history of COPD exacerbations are recommended indexes for the modulation of pharmacological therapy. However, these patients can also show systemic effects [3] and co-morbid conditions [4–6] that are independently associated with poor prognosis [1]. Enhanced knowledge of the underlying mechanisms of these two phenomena constitutes a key step toward a better understanding of COPD heterogeneity and its implications in patient management [7].

The current study focuses on the analysis of skeletal muscle dysfunction as a characteristic systemic effect of COPD, potentially modifiable with preventive interventions, i.e. exercise training [3, 4, 8–10]. Several studies addressed the question of training adaptation of COPD muscle, ranging from studies investigating expression of specific proteins [11–13] to modeling mitochondrial mechanisms [10] and systemic exploration of canonical pathways' using gene expression [14, 15]. However, a comprehensive view of the disease mechanisms, highlighting potential biomarkers and pathway dynamics with functional implications is still missing. In our study, we applied a robust systems biology approach assuming that proteins associated to biological functions or diseases interact with each other conforming distinct neighborhoods, or network modules, in the human interactome [16, 17]. In other words, the network modules consist of clusters of active proteins (approximated in the study by transcriptionally active genes), showing high probability of functional interactions. We hypothesize that the identification, functional characterization and independent functional evaluation of such network modules can help to determine how their disturbance may lead to disease and how therapy may affect the molecular machinery [18].

To further explore the underlying mechanisms of skeletal muscle dysfunction, we compared healthy persons and COPD patients before and after exercise training. In the pre-training analysis (Fig. 1), we described transcriptionally active network modules that are specific to the skeletal muscle of COPD patients. Likewise, in the assessment of adaptive mechanisms of endurance training, we compared the differences between COPD and healthy muscle adaptation. Functional implications were initially explored through the analysis of network module associated pathways and representative differentially expressed genes. In a subsequent step, we evaluated the functional interpretation of the network modules, and relevant genes, with previous experimental data obtained in the same study groups [19, 20].

To the best of our knowledge the current research provides an innovative approach by retrieving disease specific pathway mechanisms and performing an integrative analysis of the relationships of transcriptomics with metabolic, redox, inflammatory and clinical measurements to investigate COPD muscle dysfunction and training-induced adaptive changes in these patients. We believe that the study sheds novel light on underlying mechanisms of the disease with potential implications for the design of innovative preventive strategies.

Methods

Study dataset

The current study is based on a dataset of microarray gene expression measurements (Human U133 Plus2 Gene Chips) performed on open biopsies from the limb muscle vastus lateralis, reported in [15]. In all participants, these were obtained at rest, before and after an 8-week high intensity endurance training program (Fig. 1a). The study groups (Table 1) included fifteen COPD patients and twelve healthy but sedentary age-matched controls. The training program is explained in details in Additional file 1: Section 1 and in the related studies [15, 19, 20].

Analysis strategy

Briefly, network modules were identified for each differential condition with the HotNet2 algorithm [22] (Fig. 1b), using the genes' adjusted differential expression profile and selected protein–protein interaction (PPI) networks [17, 23]. Thereafter (Fig. 1c), each module was functionally characterized using gene ontology (GO) [24] term enrichment analysis and literature mining. Finally, (Fig. 1d), the validity of the matched module functions was evaluated using previous experimental data (Tables 1, 2).

Statistical analysis

Differential gene expression

To evaluate the baseline (pre-training) effects, we computed the differential gene expression between COPD and healthy individuals, referred as COPD disease effects (COPD-DE) (Fig. 1a). To evaluate the training induced changes in the molecular mechanisms (training effects, TE), we investigated the post and pre-training differential gene expression in COPD (COPD-TE) and healthy (Healthy-TE) separately (Fig. 1a). The non-parametric rank product method [25] was used to compute the significance and false discovery rate (FDR) of differential

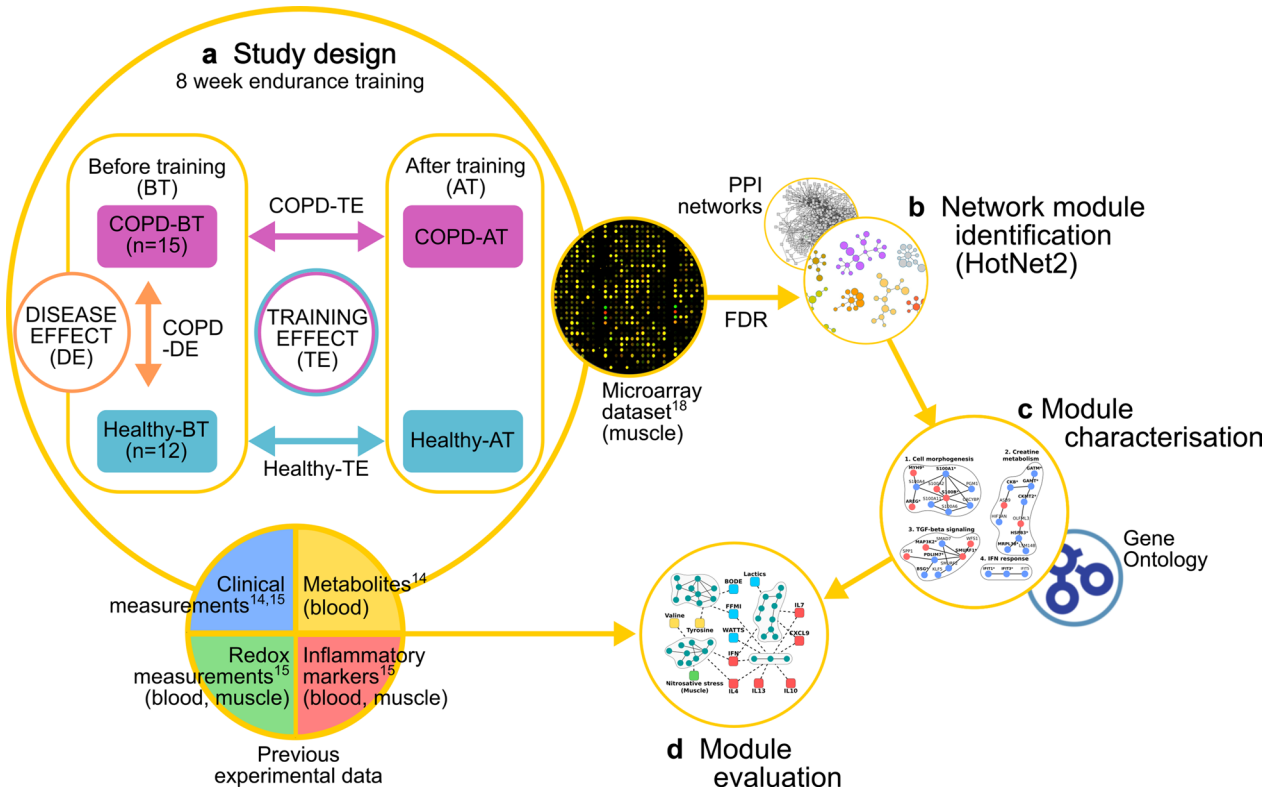


Fig. 1 Schematic diagram of the workflow of the study. (a) Study design of the used datasets. COPD patients (n = 15) and healthy controls (n = 12) were studied before (BT) and after (AT) an 8-week endurance training program. Measurements of skeletal gene expression [15] were used for network modules identification. Differential conditions of COPD disease effects (COPD-DE) and training-induced effects in COPD (COPD-TE) and in healthy muscles (Healthy-TE) were analyzed in the study. (b) Network modules were identified for each differential condition with the HotNet2 algorithm [22], using the gene's false discovery rate (FDR) adjusted differential expression P values and selected protein–protein interaction (PPI) networks [17, 23] as explained in details in Additional file 1: Section 1. Thereafter (c), each module was functionally characterized using gene ontology (GO) term enrichment analysis. (d) Correlation of network modules with independent multilevel measurements was analyzed for evaluation purposes. Specifically, independent measurements were sampled both pre- and post-training and consisted of physiological parameters measured with a constant-work rate exercise at 75% of pre-training maximum peak exercise, inflammatory and redox biomarkers measured in plasma and in skeletal muscle [20], as well as plasma metabolomics measured at rest and after exercise [19]

gene expression, due to its reliable and consistent performance with noisy, low sample size measurements.

Network module identification

For each condition, we used the HotNet2 algorithm [22] to identify network modules (Fig. 1c), taking into account: (i) the FDR of differential gene expression and, (ii) publicly-available high quality protein–protein interaction (PPI) networks [17, 23] (see Fig. 1b). A statistical test included in the HotNet2 algorithm was used to determine the significance of the number and size of the network modules. The HotNet2 algorithm was selected due to its specific feature of the use of a heat diffusion/random walk model to simulate the spread of influence of protein activity to their physical interaction partners. This feature makes this approach less reliant on the significance test and enables the identification of key proteins with less significant changes but with high biological meaning (i.e. due to topology: hub proteins, high betweenness centrality proteins, etc.). For more details see extended methods in Additional file 1: Section 1.

Functional characterization

We conducted Gene Set Analysis to investigate the enrichment of GO terms in modules (Fig. 1c) using the clusterProfiler R library [26]. Network modules were considered functionally significant if it had at least one associated GO term that: (i) had Benjamin–Hochberg corrected P value < 0.05; and, (ii) were related to at least two module genes.

Evaluation of the module functions with experimental data

To evaluate the identified functions, modules were compared with experimental data obtained in the same study group (Fig. 1a, Tables 1, 2), firstly the

Table 1 Characteristics of the study groups

	Healthy	COPD
Sex (M/F)	10/2	15/0
Age, years	65 ± 9	69 ± 7
FFMI, kg/m^2	21 ± 2	19 ± 3
FEV_1, L (mean % pred)	3.46 ± 0.69 (107)	1.34 ± 0.37 (46)*
FEV_1/FVC	0.75 ± 0.04	0.43 ± 0.08*
VO_2 peak, L/min (mean VO_2 peak/kg)	1.70 ± 0.5 (22)	0.91 ± 0.3 (14)*
[La]a peak, mEq/L	10.60 ± 2.7	6.8 ± 2.3*
VO_2 peak training diff (post–pre), L/min	0.25 ± 0.11†	0.14 ± 0.18†
[La]a training diff (post–pre), mEq/L	− 4.60 ± 0.6†	− 1.5 ± 2†

Results are expressed as mean ± SD

In the post-training study, lactate measurements during constant-work rate exercise were done at the same workload and duration than the pre-training exercise protocol

FFMI fat free mass index, *FEV_1* forced expiratory volume in the first second, *FEV1/FVC* FEV_1 to forced vital capacity ratio, *VO_2 peak* peak oxygen uptake difference post minus pre-training, *[La]a* arterial lactate concentration difference

Unpaired t test was used to compare controls and COPD, * P < 0.05. Paired t test was used to compare post-training and baseline time points in both healthy controls and COPD patients, †P < 0.05. Low FFMI was defined as < 17.05 kg/m^2 for men [21]. It is of note that three COPD patients were discarded from the analysis because they did not pass the Agilent analysis

transcriptional activity of the network modules were summarized using their first three principal components, i.e. their first three eigengenes [27, 28], which on average explained 83% of the modules' overall variability. Then, associations of the principal components with the previous experimental data [19, 20] were identified using non-parametrical Kendall correlation and selecting those associations with absolute value of *rho* ($|R|$) $\geq$ 0.4 and P value (P) < 0.05. Significant differentially expressed genes within functionally significant network modules were also considered for comparison with previous experimental data.

Results

Study workflow

In the pre-training analysis, we describe transcriptionally active network modules that are specific to COPD patients (Fig. 1a–c). Likewise, in the analysis of the training-induced effects, we separately analyze network modules that changed in response to training in COPD and healthy and compare the differences between them. Functional implications of the network modules were initially determined through the analysis of pathways associated to module genes and then specific mechanisms were deduced from the gene functions and interactions. In a subsequent step, the network modules and representative genes of specific pathways are compared with previous experimental data obtained in the two study groups both showing clear training-induced physiological responses, as described in Fig. 1d and in Tables 1 and 2.

Alterations in skeletal muscle of COPD patients at rest

The pre-training study identified four significant COPD specific network modules, that were functionally characterized, on the basis of significantly enriched GO terms in the modules (see Additional file 2: Table S5), as: creatine metabolism, Ca^{2+} dependent binding, TGF-β signaling and Interferon response (Fig. 2a).

Defective skeletal muscle energy metabolism in COPD was indicated by the creatine metabolism module. The module presented four out of the nine genes of the creatine metabolism pathway significantly down-regulated, two related to creatine synthesis (GAMT, GATM) and two creatine kinase (CK) genes (CKB, CKMT2). Overall, down-regulation of creatine metabolism suggests impairment of muscle energy production, which is consistent with studies showing low baseline creatine kinase and ATP concentrations [29, 30]; and low post-exercise recovery rate in COPD skeletal muscle [31–33].

Table 2 Summary of experimental data obtained from the same study groups

Measurements	COPD versus health Summary of results
Plasma metabolomics [19]	The two groups showed differences in metabolomic profiles at rest (P < 0.05). Levels of valine, alanine and isoleucine were associated with FFMI (P < 0.01 each)
Plasma metabolomics training diff [19]	In Healthy, training generated marked changes in amino acids, creatine, succinate, pyruvate, glucose and lactate (P < 0.05 each). But, COPD patients only showed lactate decrease (P < 0.05)
Inflammatory cytokines [20]	COPD patients showed high levels of circulating cytokines (P < 0.05), not seen in healthy
Inflammatory cytokines training diff [20]	No training-induced changes were observed in circulating cytokines levels
Redox status [20]	COPD patients showed blood and muscle oxidative stress at baseline. Muscle and blood protein carbonylation levels were correlated (P < 0.05)
Redox status training diff [20]	In COPD patients, protein nitration levels decreased after training

Summary description of the results of previous experimental measurements on plasma metabolomics [19], as well as on both muscle and blood inflammatory cytokines and redox status [20], carried out at rest before training and after endurance training. The term training diff refers to training-induced adaptive changes. For comprehensive list of measured variables see Additional file 2: Tables S2, S7 for the differentials

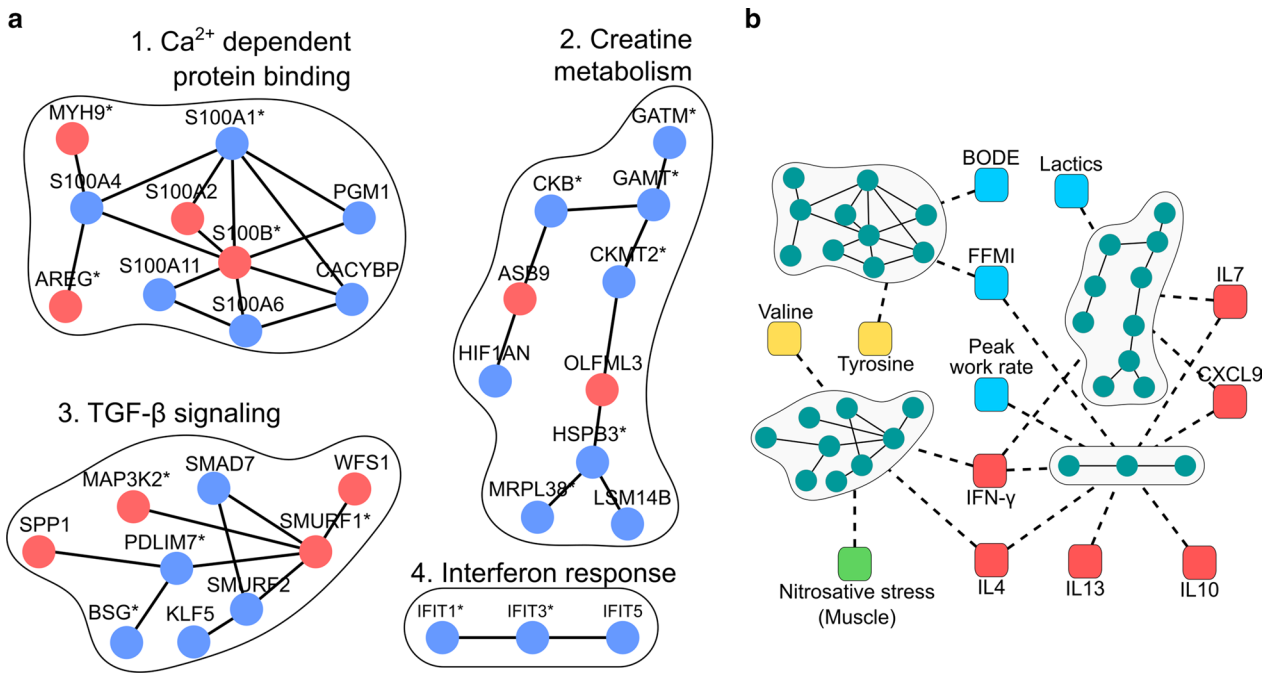


Fig. 2 Disease effects (COPD-DE) network modules. **a** The four network modules associated to COPD disease effects and their composing genes. Genes are colored according to their differential regulation, namely: up regulation—red nodes; and down regulation—blue nodes. Significantly differentially expressed genes are indicated by * (FDR ≤ 0.05) (for detailed information see Additional file 2: Table S6). **b** The significant correlations of independent measurements with any of the network modules' first three principal components. Blue squares depict exercise related independent variables [19]; red squares show cytokines measured in blood [20]; yellow squares correspond to amino acids measured in serum [19]; and, green squares represents redox biomarkers [20]

It is of note that impaired creatine metabolism would primarily affect work performance and Ca^{2+} homeostasis, especially in the presence of oxidative stress [34, 35]. In line with this, we found clustering of S100 family calcium-dependent protein binding genes in the Ca^{2+} dependent protein binding module. The potential deleterious effect of the module is well represented by the down-regulation of S100A1 gene, which could lead to abnormal sarcoplasmic reticulum Ca^{2+} content and fluxes, deteriorating muscle contractility and work performance [36, 37]. Furthermore, several module genes (S100B, S100A4, S100A6, MYH9) are related to cell morphogenic processes.

The TGF-β signaling module displayed an interplay of genes related to muscle remodeling (SMURF1, SMURF2, SMAD7) and cellular stress response (MAP3K2, SPP1). Abnormal TGF-β signaling was suggested by up-regulation of its inhibitor SMURF1 and further strengthened by the observed down-regulation of SMURF1's binding competitor (PDLIM7) [38] potentially leading to increased protein degradation by ubiquitination [39]. The associated gene functions suggest an interplay between TGF-β signaling and oxidative stress, which has been reported in the literature highlighting the specific role of SMURF1 in these processes [40, 41]. Furthermore, overexpression of SMURF1 may attenuate IFN-γ-mediated immune responses of the Jak-STAT pathway, by inhibiting STAT1 [42, 43], positive regulator of IFIT gene expression [44], which could explain systematic down-regulation of these genes in the interferon response module.

Evaluation of alterations in COPD patients at rest

In order to evaluate the functions of the COPD specific network modules, their association with previous experimental data was analyzed (see Fig. 2b and for details Additional file 2: Table S8).

The Creatine metabolism module showed statistically significant associations with systemic inflammatory markers, namely IFN-γ ($|R| = 0.42$, $P = 0.041$), IL7 ($|R| = 0.5$, $P = 0.016$) and CXCL9 ($|R| = 0.58$, $P = 0.003$) as well as with pre-training blood lactate levels at a constant-work rate exercise at 75% VO_2 peak ($|R| = 0.49$, $P = 0.013$) suggesting relationships between altered cell bioenergetics and abnormal inflammatory processes.

The association of the Ca^{2+} dependent protein binding module with muscle mass (FFMI) ($|R| = 0.45$, $P = 0.026$)

and with exercise capacity, expressed by the composite BODE index [45] ($|R| = 0.47$, $P = 0.033$) confirms the physiological impact of defective Ca^{2+} homeostasis. Such an association at module level is further strengthened by the correlations of the S100A1 gene expression with both VO_2 peak ($R = 0.52$, $P = 0.006$) (Fig. 3a) and peak work rate (Watts peak) ($R = 0.53$, $P = 0.005$).

Consistent with the functional analysis, TGF-β signaling module showed significant correlations with increased skeletal muscle nitrosative stress in COPD patients ($|R| = 0.49$, $P = 0.031$), as well as with abnormally low levels of blood valine amino acids ($|R| = 0.41$, $P = 0.047$). Likewise, blood cytokines IFN ($|R| = 0.44$, $P = 0.03$) and IL4 ($|R| = 0.47$, $P = 0.024$) also showed significant associations with the module. At gene level, statistically significant negative correlations were observed between SMURF1 and nitrosative stress levels in skeletal muscle of COPD patients ($R = -0.66$, $P = 0.017$), not seen in healthy subjects (Fig. 3b).

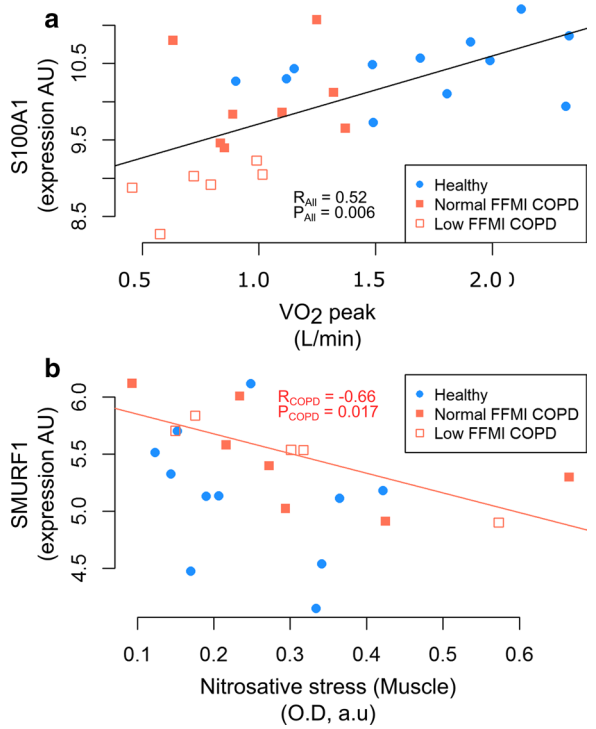


Fig. 3 Relationships between genes from COPD specific modules (disease effects) and previous experimental data. **a** The relationships between S100A1, from the Ca^{2+} dependent binding module, and VO_2max. The two groups, healthy subjects (blue circles) and COPD patients (low and normal FFMI, empty and filled squares, respectively) fell on the same regression line ($R = 0.52$, $P = 0.006$, FDR = 0.026). **b** The relationships between SMURF1 from the TGF-β signaling module and skeletal muscle nitrosative stress. A statistical significant correlation was seen in the COPD group, both normal and low FFMI ($R = -0.67$, $P = 0.018$ and FDR = 0.07), but not in healthy subjects ($R = -0.2$, $P = 0.55$)

As expected, interferon response module showed significant correlations with IFN ($|R| = 0.63$, $P = 0.015$) and several other cytokines, which presented elevated blood levels in COPD patients (Table 2). Furthermore, the module also presented significant relationships with FFMI ($|R| = 0.47$, $P = 0.019$) and peak work rate ($|R| = 0.40$, $P = 0.047$) in COPD patients.

Inefficient training-induced responses in COPD patients

In the analysis of the training-induced effects (TE), we identified and evaluated network modules separately for COPD patients (COPD-TE) and for healthy sedentary subjects (Healthy-TE). The research identified a total of six functionally enriched network modules (Fig. 4a).

It is of note that Hippo signaling was the only COPD-TE specific module; whereas, some genes of the Interferon response were observed in both COPD-TE and Healthy-TE. Likewise, Oxidative phosphorylation, Amino acid biosynthesis, Epigenetic regulation of metabolic processes and Intracellular transport functional modules were only observed in Healthy-TE and were named after significantly enriched GO terms in the modules.

In COPD-TE, the Hippo pathway module suggests abnormal training-induced activation of skeletal muscle remodeling, as reported in detail in the extended results section in Additional file 1: Section 1.

Endurance training induced inflammatory responses in skeletal muscle, as indicated by the Interferon response module that showed a consistent increase in gene expression levels in both COPD-TE and Healthy-TE. The module could signal the local inflammatory response to muscle damage caused by exercise, which reportedly coincides with muscle repair, regeneration, and growth [46].

It is of note that the four Healthy-TE network modules indicated strong associations of training responses with bioenergetics changes and their joint regulation with other molecular functions (see extended results in Additional file 1: Section 2).

Evaluation of training-induced responses in COPD patients

The analysis of associations between TE network modules and previous experimental data was carried out in COPD-TE and Healthy-TE separately, as displayed in Fig. 4b (for details see Additional file 2: Table S8). We observed a significant association between training-induced increase in peak work rate (Watts) and the interferon response module in the two groups ($|R|_{COPD} = 0.48$, $P_{COPD} = 0.019$; $|R|_{Healthy} = 0.53$, $P_{Healthy} = 0.018$), suggesting training-induced increase of inflammatory responses both in healthy subjects and in COPD patients. However, the most relevant findings were the strong relationships between Healthy-TE network modules

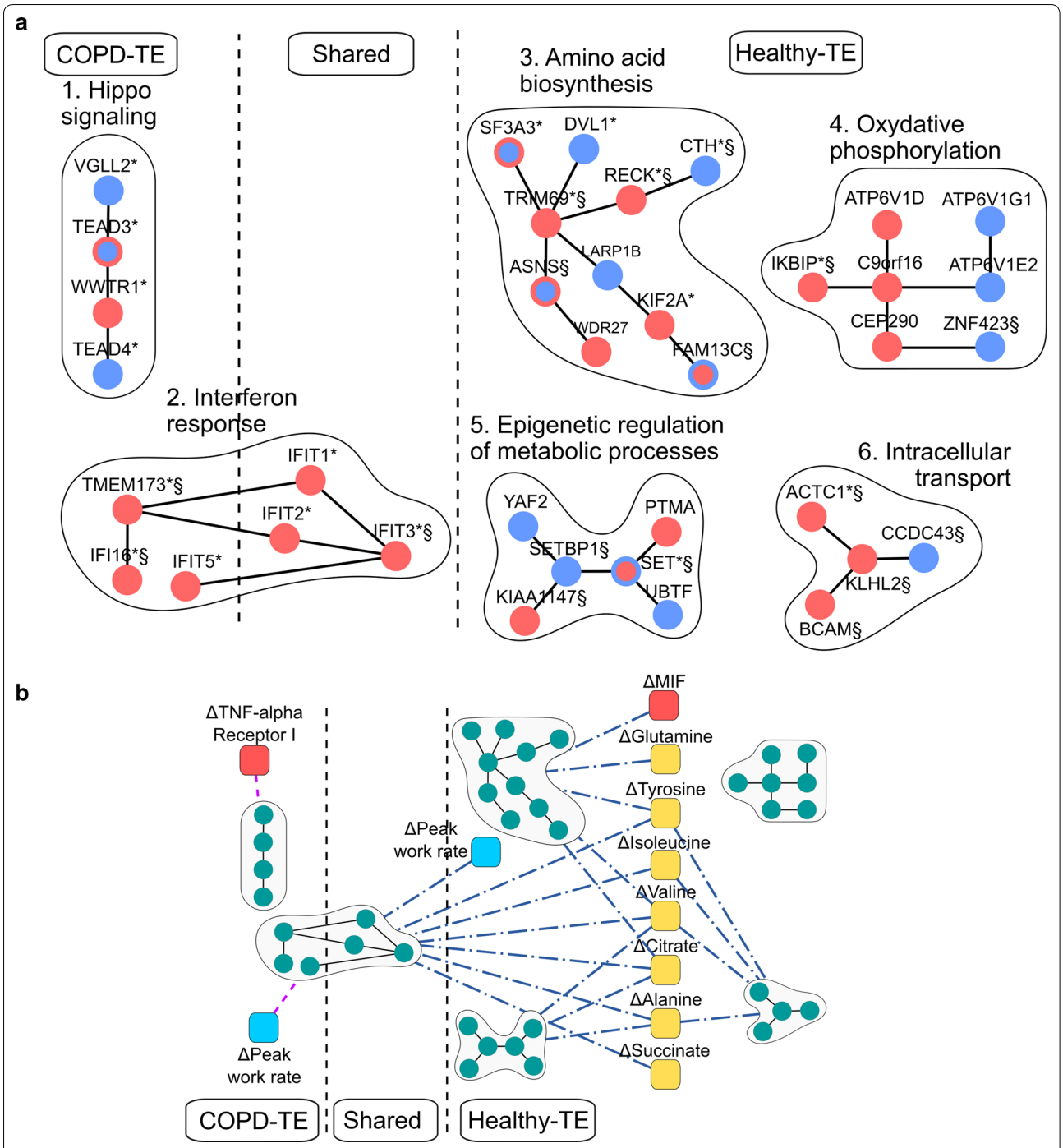


Fig. 4 Training effects (TE) network modules. **a** Active network modules identified in case of COPD-TE, Healthy-TE and in both (shared). Genes are colored according to their differential regulation in COPD-TE (inner color of the nodes) and in Healthy-TE (border color of the nodes): up regulation with training (red circles), down regulation with training (blue circles). Modules are named after significantly enriched GO terms. Training differential expression significance is signed by * for COPD-TE, and § for Healthy-TE (FDR < 0.05) (for detailed information see Additional file 2: Table S6). **b** The significant correlations of the independent measurements with any of the significantly-changed training modules' first three principal components in COPD, depicted as purple dashed lines, and in healthy subjects, depicted as blue dotted-dashed lines. Blue squares depict exercise related independent variables; red squares show cytokines measured in blood; and yellow squares correspond to amino acids measured in serum

associated to different aspects of muscle bioenergetics and metabolomics training-induced changes, not seen in COPD patients. Likewise, statistically significant associations were observed between training-induced transcriptional changes at gene level and plasma metabolomics responses in healthy subjects, but not in COPD patients, as shown in Fig. 5 wherein the relationships between training-induced changes the splicing factor SF3A3 (ΔSF3A3) and Δglutamine are depicted for healthy subjects ($R_{Healthy} = 0.7$, $P_{Healthy} = 0.001$) and for COPD patients ($R_{COPD} = -0.14$, $P_{COPD} = 0.518$).

Discussion

The approach adopted in the current study contributed to uncover novel interactions among biological pathways of skeletal muscle dysfunction in COPD patients, as well as suggest biomarkers, while reinforcing previous results on the mechanisms related to the disease. The applied methodological framework also shows high potential to explore relations between clinical and omics platforms, facilitating interpretation of biological measurements.

The four network modules identified in the pre-training analysis (Fig. 2a) correspond to COPD specific mechanisms related to abnormal energy production and contractility, as well as to alterations in both inflammatory and oxidative stress pathways. Moreover, they showed significant associations with previous measurements carried out in the same study group (Figs. 2b, 3). To be noted that lung function (FEV_1) only presented a weak negative relationship with the interferon module that did not meet the inclusion criteria of the analysis. In contrast, several COPD specific modules, and genes (Fig. 2b) consistently showed associations with different indices reflecting exercise capacity, namely: BODE, VO_2 peak, Watts peak and lactate levels.

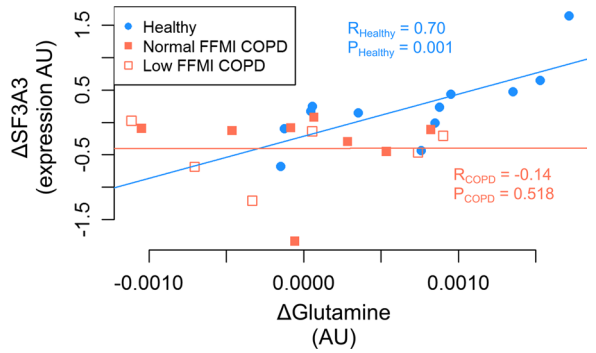


Fig. 5 Relationships between genes from Healthy-TE specific modules and previous experimental plasma metabolomics data. The figure depicts the relationships between training-induced changes in both SF3A3, from the Amino acid biosynthesis module, and glutamine. A strong correlation was seen in healthy subjects (blue circles) ($R = 0.70$, $P = 0.001$), but not in COPD patients (low and normal FFMI, empty and filled red squares, respectively) ($R = -0.14$, $P = 0.518$)

The two study groups showed significant physiological training effects as displayed in Table 1. The differences in the training-induced responses between COPD and healthy (Fig. 4a) further contributed to shed novel light on the underlying mechanisms of skeletal muscle dysfunction in these patients. The most striking finding was that the physiological bioenergetics responses, strongly correlated with plasma metabolomics (Fig. 4b), were not observed in the patients. Instead, in the COPD group, the training-induced changes were mostly related with skeletal muscle remodeling (Hippo signaling pathway), without significant adaptive changes in oxidative phosphorylation and related bioenergetics pathways. It is of note that in a post hoc analysis, we explored the impact of FFMI on the modules, which consistently indicated that training adaptation seen in COPD patients with normal FFMI were more similar to the ones of healthy subjects than those observed in COPD patients with low FFMI (see Additional file 1: Section 2). Regarding the training-induced inflammatory responses, the healthy and COPD groups only shared part of the genes of the network module that indicates increased inflammatory changes induced by training in COPD. It is of note that significant associations of peak work rate with the inflammatory network modules were observed in the disease effects (Fig. 2a, b) and in the training-induced effects (Fig. 4a, b).

As acknowledged below, the current study cannot inform on causality and temporal sequence of the skeletal muscle abnormalities observed in the COPD group. The marked differences between COPD patients and healthy subjects regarding training adaptations of skeletal muscle bioenergetics (Fig. 4a) seem to suggest that the abnormal energy production, already depicted in the pre-training analysis (Fig. 2a), is the most visible and likely the primary phenomenon of skeletal muscle dysfunction in COPD. It is of note that a recent report using data from the same study group [10], but focusing on the analysis of gene regulatory networks, highlighted the existence of significant COPD abnormalities at mitochondrial level with impact on skeletal muscle inflammatory responses, and explored potential therapeutic strategies.

Abnormal bioenergetics may likely trigger changes in skeletal muscle Ca^{2+} homeostasis, which ultimately may lead to impairment of the contractile mechanisms and alterations in muscle morphogenesis, as suggested by the Ca^{2+} dependent protein binding module (Fig. 2a) and the Hippo signaling pathway module (Fig. 4a). These mechanisms might be related to generation of abnormal muscle fiber type distribution with increased glycolytic

(Type II and IIX) to oxidative (Type I) fiber ratio in these patients [47, 48]. In the study, physiological inflammatory response pathways at baseline showed to be inhibited, potentially by SMURF1, and most likely be modulated by oxidative stress, which might indicate counter-regulatory processes related with low-grade systemic inflammation. The partly abnormal training-induced inflammatory responses observed in the study might also constitute a secondary phenomenon modulated by nitroso-redox disequilibrium reported in these patients [3, 8, 20].

The network biology techniques used in the current study to identify and characterize skeletal muscle network modules are gaining increasing attention in the biomedical research field due to their ability to highlight complex cellular disease mechanisms [49–51]. An added potential of PPI based methods is the constraint that the interaction network represents, whose topology already encodes basic biological functions [16, 52] and provides high performance in predicting biologically meaningful pathways [53]. Additionally, the model used in HotNet2, simulating the spread of influence of protein activity, enables the identification of key proteins with less significant changes but with high biological meaning due to surrounding expression patterns as well as due to topology (e.g. hub proteins, proteins with high betweenness centrality, etc.), which complement standard differential expression measures with deeper biological insights. We believe that the approach adopted in the current study facilitates a comprehensive analysis and understanding of complex cellular mechanisms overcoming limitations of traditional research only addressing analysis of target biological pathways. Furthermore, the applied methodology has high potential for creating a standardized analysis pipeline for the integrative analysis of multi-level data.

Study limitations

We acknowledge, however, that further longitudinal studies are needed to support the above statements, as well as to properly clarify the relationships between skeletal muscle dysfunction and pulmonary impairment provoked by the disease. We also acknowledge that the microarray dataset used in the study is lacking standard qPCR validation of specific biomarkers, which we aimed to overcome by showing the high concordance of specific markers with qPCR validation of two earlier studies on skeletal muscle of patients with COPD (Additional file 2: Table S9). However, we believe that given our system-based approach, the validation of a few genes is less relevant compared to the functional evaluation of the modules with independent measurements that was conducted in the study. Furthermore, the completeness and/or bias of the publicly available PPI networks [17, 23] are intrinsic limitations of the methodological approach which, additionally, does not provide information on causality. The rather small sample size constituted a problem such that a type II error limiting our interpretations of the results cannot be excluded. Furthermore, the limitation of sample size has been addressed using robust statistical approaches at each step of the analysis. In particular, when choosing the HotNet2 algorithm and its application, as explained in detail in the extended methods and, in general, when considering protein–protein interactions (PPI) network based methods, which offer a more robust performance in small sample size environments [54] compared to other systems medicine approaches [15]. Moreover, the identification of both statistically and biologically significant relationships of the resulting functional modules (and genes) with previous experimental multilevel data obtained in the same study groups [19, 20] provided additional robustness to the evaluation and functional characterization of the core findings of the study. Summing up, different factors emerging from the study design, such as sample size, noisy clinical environment and factors originating from the modeling technique in use, such as (i) current constraints of available PPI networks, (ii) modeling proteins levels with gene expression, (iii) relying on arbitrary significance thresholds, and (iv) comparing of measurements of different body compartments (blood, muscle) may lead to confounding results, which prompts for future validation of the study. The consistency of the results, however highlights the potential of biological modeling as a preliminary step for future discoveries. The above mentioned factors may also explain that the study did not identify specific pathways that are known to play a significant role in skeletal muscle dysfunction in COPD, such as the FoxO signaling pathway [3, 14, 55].

We acknowledge that differences in training intensity between healthy subjects and COPD patients (Table 1) should be considered in the interpretation of the results. However, the findings of the study are supported by the following factors: (i) pre-training COPD specific findings; and, (ii) qualitative nature of the training-induced differences between healthy and COPD unlikely explained only by differences in training intensity. A final methodological consideration is that the COPD group includes only males, which constitute an over-representation of this gender (Table 1), as compared to current COPD prevalence in men. However, no reports on gender specificity of the findings have been found neither in the literature nor in our dataset (Additional file 1: Figure S5).

Future work

We believe that the current study significantly contributed to enhance our understanding of skeletal muscle

dysfunction in patients with COPD. Further research addressing the molecular mechanisms of impaired muscle energy production in these patients should shed light on remaining challenges such as, causality, lung-muscle interactions and design of cost-effective strategies aiming at preventing non-pulmonary effects in COPD patients. The central role of impaired bioenergetics seems to endorse that promotion of daily physical activity at early disease stages may have a role preventing skeletal muscle dysfunction in these patients. We believe that future longitudinal studies using the current methodological approach will generate further evidence supporting our interpretations of the current study findings.

A better knowledge on underlying mechanisms of non-pulmonary effects of COPD should necessarily lead to enhanced patient risk assessment and better health service selection. Moreover, continuous progresses in our understanding of mechanisms of COPD heterogeneity might prompt the need for revisiting the taxonomies of obstructive airways diseases.

Conclusions

The research provides a comprehensive view of the core mechanisms involved in skeletal muscle dysfunction as a systemic effect of COPD. The results indicate that COPD patients show impaired training-induced responses in skeletal muscle bioenergetics, with abnormal inflammatory changes and altered tissue remodeling, as compared to healthy sedentary subjects. The current network medicine approach shows high potential for future longitudinal analyses exploring preventive strategies addressing non-pulmonary effects of COPD.

Additional files

Additional file 1: Section 1. Expanded methods: TRAINING program, HotNet2 and parameters, Interactome construction. **Section 2.** Expanded results: Training induced adaptation, Impact of FFMI on disease effect modules, Impact of FFMI on training effect modules. Training effect differences between COPD and healthy. Gender effect on the transcriptomics profile. **Figure S1.** Overlap between nodes and edges in the interaction networks. **Figure S2.** Disease effect modules of the COPD subgroups. **Figure S3.** Training effect modules of the COPD subgroups. **Figure S4.** Difference between COPD and healthy training effects. **Figure S5.** Gender effect on the transcriptomics profile.

Additional file 2: Table S1. Number of differentially expressed genes in the different groups and conditions. **Table S2.** Previous measurements: List of variables measured. **Table S3.** Network modules: HotNet2 results. **Table S4.** Network modules: HotNet2 consensus. **Table S5.** Network modules: Functional characterization. **Table S6.** Network modules: Gene differential expression. **Table S7.** Previous measurements: Differentials. **Table S8.** Previous measurements: Association with network modules. **Table S9.** Comparison of microarray results with qPCR validation of external datasets.

Abbreviations

COPD: chronic obstructive pulmonary disease; COPD-DE: COPD disease effects; COPD-TE: COPD training effects; Healthy-TE: healthy training effects; COPDN: normal FFMI COPD; COPDL: low FFMI COPD; FDR: false discovery rate; PPI: protein–protein interaction; GO: gene ontology; FFMI: fat free mass index; FEV1: forced expiratory volume in the first second; FEV1/FVC: FEV1 to forced vital capacity ratio; VO_2 peak: peak oxygen uptake difference post minus pre-training; [La]a: arterial lactate concentration difference; WATTS: peak work rate.

Authors' contributions

Study conception and design, AT, DG-C, MC, JR. Data acquisition, AT, FM and NK. Data analysis, AT, DG-C, MC, JR. Manuscript preparation, AT, IC, DG-C and JR. Manuscript revision, AT, IC, NK, FM, SK, PA, MC, EB, DG-C and JR. All authors read and approved the final manuscript.

Author details

[1] Hospital Clinic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain. [2] Center for Biomedical Network Research in Respiratory Diseases (CIBERES), Madrid, Spain. [3] Unit of Computational Medicine, Department of Medicine, Karolinska Institute, 171 77 Stockholm, Sweden. [4] Center for Molecular Medicine, Karolinska Institutet, 171 77 Stockholm, Sweden. [5] Bioinformatics Core Facility, IDIBAPS-CEK, Hospital Clínic, University de Barcelona, Barcelona, Spain. [6] Pulmonology Dept, Muscle and Respiratory System Research Unit, IMIM-Hospital del Mar, Universitat Pompeu Fabra, PRBB, Barcelona, Spain. [7] Departament de Bioquimica i Biologia Molecular, Facultat de Biologia-IBUB, Universitat de Barcelona, 08028 Barcelona, Spain. [8] Mucosal and Salivary Biology Division, King's College London Dental Institute, London SE1 9RT, UK.

Acknowledgements

We want to acknowledge the support of the European Commission (FP7) grants SYNERGY-COPD (no. FP7-ICT-2009-270086), the AGAUR (2009SGR911 and 2014SGR1017), CERCA Programme / Generalitat de Catalunya (2014SGR661) and The 'ICREA Academia' prize for excellence in research, ICREA foundation-Generalitat de Catalunya (to M.C.).

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The microarray dataset(s) supporting the conclusions of this article is(are) available in the GEO repository, GSE27536, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE27536.

Consent for publication

Not applicable.

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations and were approved by the Ethical Committee at Hospital Clinic and written informed consent was signed by each participant.

Funding

This research was supported by the European Commission (H2020) grant CONNECARE (no. PHC-25-2015-689802).

Publisher's Note

Received: 31 May 2017 Accepted: 12 February 2018
Published online: 20 February 2018

References

1. Celli BR, Decramer M, Wedzicha JA, Wilson KC, Agustí A, Criner GJ, et al. An official American Thoracic Society/European Respiratory Society statement: research questions in COPD. Eur Respir J. 2015;45:879–905.

2. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. Am J Respir Crit Care Med. 2017;195:557–82.
3. Maltais F, Decramer M, Casaburi R, Barreiro E, Burelle Y, Debigaré R, et al. An official American Thoracic Society/European Respiratory Society statement: update on limb muscle dysfunction in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2014;189:e15–62.
4. Barnes PJ. Mechanisms of development of multimorbidity in the elderly. Eur Respir J. 2015;45:790–806.
5. Vanfleteren LEGW, Spruit MA, Groenen M, Gaffron S, van Empel VPM, Bruijnzeel PLB, et al. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2013;187:728–35.
6. Divo MJ, Casanova C, Marin JM, Pinto-Plata VM, De-Torres JP, Zulueta JJ, et al. COPD comorbidities network. Eur Respir J. 2015;46:640–50.
7. Roca J, Vargas C, Cano I, Selivanov V, Barreiro E, Maier D, et al. Chronic obstructive pulmonary disease heterogeneity: challenges for health risk assessment, stratification and management. J Transl Med. 2014;12:S3.
8. Rabinovich RA, Ardite E, Troosters T, Carbó N, Alonso J, Gonzalez de Suso JM, et al. Reduced muscle redox capacity after endurance training in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2001;164:1114–8.
9. Rabinovich RA, Bastos R, Ardite E, Llinas L, Orozco-Levi M, Gea J, et al. Mitochondrial dysfunction in COPD patients with low body mass index. Eur Respir J. 2007;29:643–50.
10. Marín de Mas I, Fanchon E, Papp B, Kalko S, Roca J, Cascante M. Molecular mechanisms underlying COPD-muscle dysfunction unveiled through a systems medicine approach. Bioinformatics. 2017;33:95–103.
11. Puig-Vilanova E, Rodriguez DA, Lloreta J, Ausin P, Pascual-Guardia S, Broquetas J, et al. Oxidative stress, redox signaling pathways, and autophagy in cachectic muscles of male patients with advanced COPD and lung cancer. Free Radic Biol Med. 2015;79:91–108.
12. Fermoselle C, Rabinovich R, Ausin P, Puig-Vilanova E, Coronell C, Sanchez F, et al. Does oxidative stress modulate limb muscle atrophy in severe COPD patients? Eur Respir J. 2012;40:851–62.
13. Barreiro E, Rabinovich R, Marin-Corral J, Barbera JA, Gea J, Roca J. Chronic endurance exercise induces quadriceps nitrosative stress in patients with severe COPD. Thorax. 2009;64:13–9.
14. Rabinovich RA, Drost E, Manning JR, Dunbar DR, Díaz-Ramos M, Lahkdar R, et al. Genome-wide mRNA expression profiling in vastus lateralis of COPD patients with low and normal fat free mass index and healthy controls. Respir Res. 2015;16:1.
15. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, et al. A systems biology approach identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease. PLoS Comput Biol. 2011;7:e1002129.
16. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12:56–68.
17. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, et al. Disease networks. Uncovering disease–disease relationships through the incomplete interactome. Science. 2015;347:1257601.
18. Dittrich MT, Klau GW, Rosenwald A, Dandekar T, Müller T. Identifying functional modules in protein–protein interaction networks: an integrated exact approach. Bioinformatics. 2008;24:i223–31.
19. Rodríguez DA, Alcarraz-Vizán G, Díaz-Moralli S, Reed M, Gómez FP, Falciani F, et al. Plasma metabolic profile in COPD patients: effects of exercise and endurance training. Metabolomics. 2011;8:508–16.
20. Rodriguez DA, Kalko S, Puig-Vilanova E, Perez-Olabarría M, Falciani F, Gea J, et al. Muscle and blood redox status after exercise training in severe COPD patients. Free Radic Biol Med. 2012;52:88–94.
21. Vestbo J, Prescott E, Almdal T, Dahl M, Nordestgaard BG, Andersen T, et al. Body mass, fat-free body mass, and prognosis in patients with chronic obstructive pulmonary disease from a random population sample: findings from the Copenhagen City Heart Study. Am J Respir Crit Care Med. 2006;173:79–83.
22. Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge JV, Thomas JL, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat Genet. 2014;47:106–14.
23. Rolland T, Taşan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, et al. A proteome-scale map of the human interactome network. Cell. 2014;159:1212–26.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.
25. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. 2004;573:83–92.
26. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.
27. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol. 2007;1:54.
28. Davidsen PK, Herbert JM, Antczak P, Clarke K, Ferrer E, Peinado VI, et al. A systems biology approach reveals a link between systemic cytokines and skeletal muscle energy metabolism in a rodent smoking model and human COPD. Genome Med. 2014;6:59.
29. Barreiro E, Gea J, Matar G, Hussain SNA. Expression and carbonylation of creatine kinase in the quadriceps femoris muscles of patients with chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol. 2005;33:636–42.
30. Barreiro E, Rabinovich R, Marin-Corral J, Barbera JA, Gea J, Roca J. Chronic endurance exercise induces quadriceps nitrosative stress in patients with severe COPD. Thorax. 2008;64:13–9.
31. Wuyam B, Payen JF, Levy P, Bensaidane H, Reutenauer H, Le Bas JF, et al. Metabolism and aerobic capacity of skeletal muscle in chronic respiratory failure related to chronic obstructive pulmonary disease. Eur Respir J. 1992;5:157–62.
32. Tada H, Kato H, Misawa T, Sasaki F, Hayashi S, Takahashi H, et al. 31P-Nuclear magnetic resonance evidence of abnormal skeletal muscle metabolism in patients with chronic lung disease and congestive heart failure. Eur Respir J. 1992;5:163–9.
33. Kutsuzawa T, Shioya S, Kurita D, Haida M, Ohta Y, Yamabayashi H. Muscle energy metabolism and nutritional status in patients with chronic obstructive pulmonary disease. A 31P magnetic resonance study. Am J Respir Crit Care Med. 1995;152:647–52.
34. Steeghs K, Benders A, Oerlemans F, de Haan A, Heerschap A, Ruitenbeek W, et al. Altered Ca^{2+} responses in muscles with combined mitochondrial and cytosolic creatine kinase deficiencies. Cell. 1997;89:93–103.
35. Schlattner U, Tokarska-Schlattner M, Wallimann T. Mitochondrial creatine kinase in human health and disease. Biochim Biophys Acta Mol Basis Dis. 2006;1762:164–80.
36. Prosser BL, Wright NT, Hernãndez-Ochoa EO, Varney KM, Liu Y, Olojo RO, et al. S100A1 binds to the calmodulin-binding site of ryanodine receptor and modulates skeletal muscle excitation–contraction coupling. J Biol Chem. 2008;283:5046–57.
37. Donato R, Cannon BR, Sorci G, Riuzzi F, Hsu K, Weber DJ, et al. Functions of S100 proteins. Curr Mol Med. 2013;13:24–57.
38. Sangadala S, Boden SD, Viggeswarapu M, Liu Y, Titus L. LIM mineralization protein-1 potentiates bone morphogenetic protein responsiveness via a novel interaction with Smurf1 resulting in decreased ubiquitination of smads. J Biol Chem. 2006;281:17212–9.
39. Goodman CA, Hornberger TA. New roles for Smad signaling and phosphatidic acid in the regulation of skeletal muscle mass. F1000Prime Rep. 2014;6:20.
40. Liu R-M, Gaston Pravia KA. Oxidative stress and glutathione in TGF-beta-mediated fibrogenesis. Free Radic Biol Med. 2010;48:1–15.
41. Krstić J, Trivanović D, Mojsilović S, Santibanez JF. Transforming growth factor-beta and oxidative stress interplay: implications in tumorigenesis and cancer progression. Oxid Med Cell Longev. 2015;2015:1–15.
42. Yuan C, Qi J, Zhao X, Gao C. Smurf1 protein negatively regulates interferon-γ signaling through promoting STAT1 protein ubiquitination and degradation. J Biol Chem. 2012;287:17006–15.
43. Doles JD, Olwin BB. The impact of JAK-STAT signaling on muscle regeneration. Nat Med. 2014;20:1094–5.
44. Zhou X, Michal JJ, Zhang L, Ding B, Lunney JK, Liu B, et al. Interferon induced IFIT family genes in host antiviral defense. Int J Biol Sci. 2013;9:200–8.
45. Celli BR, Cote CG, Marin JM, Casanova C, Montes de Oca M, Mendez RA, et al. The body-mass index, airflow obstruction, dyspnea, and exercise

capacity index in chronic obstructive pulmonary disease. N Engl J Med. 2004;350:1005–12.
46. Cheng M, Nguyen M-H, Fantuzzi G, Koh TJ. Endogenous interferon-γ is required for efficient skeletal muscle regeneration. AJP Cell Physiol. 2008;294:C1183–91.
47. Richardson RS, Leek BT, Gavin TP, Haseler LJ, Mudaliar SRD, Henry R, et al. Reduced mechanical efficiency in chronic obstructive pulmonary disease but normal peak VO_2 with small muscle mass exercise. Am J Respir Crit Care Med. 2004;169:89–96.
48. Tsika RW, Schramm C, Simmer G, Fitzsimons DP, Moss RL, Ji J. Overexpression of TEAD-1 in transgenic mouse striated muscles produces a slower skeletal muscle contractile phenotype. J Biol Chem. 2008;283:36154–67.
49. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics. 2002;18(Suppl 1):S233–40.
50. Mitra K, Carvunis A-R, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. Nat Rev Genet. 2013;14:719–32.
51. Sharma A, Menche J, Huang CC, Ort T, Zhou X, Kitsak M, et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. Hum Mol Genet. 2015;24:3005–20.
52. Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, et al. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature. 2004;430:88–93.
53. Tényi Á, de Atauri P, Gomez-Cabrero D, Cano I, Clarke K, Falciani F, et al. ChainRank, a chain prioritisation method for contextualisation of biological networks. BMC Bioinform. 2016;17:17.
54. Diez D, Agustí A, Wheelock CE. Network analysis in the investigation of chronic respiratory diseases. From basics to application. Am J Respir Crit Care Med. 2014;190:981–8.
55. Langen RCJ, Gosker HR, Remels AHV, Schols AMWJ. Triggers and mechanisms of skeletal muscle wasting in chronic obstructive pulmonary disease. Int J Biochem Cell Biol. 2013;45:2245–56.

MANUSCRIPT 3: RISK AND TEMPORAL ORDER OF DISEASE DIAGNOSIS OF COMORBIDITIES IN PATIENTS WITH COPD: A POPULATION HEALTH PERSPECTIVE

Tényi Á, Vela E, Cano I, Cleries M, Monterde D, Gomez-Cabrero D and Roca J. Risk and temporal order of disease diagnosis of comorbidities in patients with COPD: a population health perspective. *BMJ Open Respiratory Research* British Medical Journal Publishing Group; 2018; 0: e017283.

- Journal: *BMJ Open Respiratory Research*
- Impact factor: -
- Quartile: *3rd*
- Status: *Published*

BMJ Open Respiratory Research

Risk and temporal order of disease diagnosis of comorbidities in patients with COPD: a population health perspective

Ákos Tényi,[1,2] Emili Vela,[3] Isaac Cano,[1,2] Montserrat Cleries,[3] David Monterde,[4] David Gomez-Cabrero,[5,6,7] Josep Roca[1,2]

To cite: Tényi Á, Vela E, Cano I, *et al*. Risk and temporal order of disease diagnosis of comorbidities in patients with COPD: a population health perspective. *BMJ Open Resp Res* Published Online First: [please include Day Month Year]. doi:10.1136/bmjresp-2018-000302

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmjresp-2018-000302).

DG-C and JR are senior authors

Received 27 March 2018
Revised 22 May 2018

For numbered affiliations see end of article.

Correspondence to
Dr Josep Roca;
jroca@clinic.cat

ABSTRACT

Introduction Comorbidities in patients with chronic obstructive pulmonary disease (COPD) generate a major burden on healthcare. Identification of cost-effective strategies aiming at preventing and enhancing management of comorbid conditions in patients with COPD requires deeper knowledge on epidemiological patterns and on shared biological pathways explaining co-occurrence of diseases.

Methods The study assesses the co-occurrence of several chronic conditions in patients with COPD using two different datasets: Catalan Healthcare Surveillance System (CHSS) (ES, 1.4 million registries) and Medicare (USA, 13 million registries). Temporal order of disease diagnosis was analysed in the CHSS dataset.

Results The results demonstrate higher prevalence of most of the diseases, as comorbid conditions, in elderly (>65) patients with COPD compared with non-COPD subjects, an effect observed in both CHSS and Medicare datasets. Analysis of temporal order of disease diagnosis showed that comorbid conditions in elderly patients with COPD tend to appear after the diagnosis of the obstructive disease, rather than before it.

Conclusion The results provide a population health perspective of the comorbidity challenge in patients with COPD, indicating the increased risk of developing comorbid conditions in these patients. The research reinforces the need for novel approaches in the prevention and management of comorbidities in patients with COPD to effectively reduce the overall burden of the disease on these patients.

INTRODUCTION

Projections on healthcare impact of chronic obstructive pulmonary disease (COPD) over the next 15 years indicate a rapidly escalating health and societal burden mainly due to population ageing and comorbidities.[1 2] It is well-known that highly prevalent chronic conditions such as cardiovascular disorders, type 2 diabetes mellitus—metabolic syndrome and/or anxiety–depression often occur as comorbid conditions in patients with COPD.[3]

Whereas the current standards on COPD management[4] acknowledge the adverse effects of comorbidities on COPD prognosis, they suggest that 'presence of comorbidities should not alter COPD treatment, and comorbidities should be treated per usual standards regardless of the presence of COPD'. However, recent evidence prompts the need for novel approaches in the prevention and management of comorbidities in patients with COPD to effectively reduce the overall burden of the disease.[5 6]

Identification of such cost-effective strategies aiming at preventing and enhancing management of comorbid conditions in patients with COPD requires deeper knowledge on epidemiological patterns and shared biological pathways explaining co-occurrence of diseases.[7] Recently, Gomez-Cabrero *et al*[8] reported the higher risk of developing certain comorbidities in patients with COPD, as compared with patients without COPD. The study used a data-driven analysis of Medicare registries from 13 million hospitalised patients over 65 years. The authors also proposed underlying biological mechanisms that may explain the identified comorbidities. Another direction of comorbidity research aims to uncover temporal disease co-occurrence patterns, showing great potential to explain the dynamics of disease co-occurrence and to highlight characteristic disease sequences potentially caused by underlying mechanisms and common risk factors. As an example, a recent study identified COPD as a central disease with rapid progression to many other conditions, stressing the importance of its early diagnosis.[9]

In order to gain deeper knowledge on epidemiological patterns explaining co-occurrence of diseases,[7] the primary aim of the current study is to reinforce previous

evidence on the higher risk of comorbidities in patients with COPD.[8] To this end, we conducted a similar analysis to the recent work by Gomez-Cabrero *et al*[8] on an independent dataset retrieved from the Catalan Healthcare Surveillance System (CHSS) in Spain,[10] which accounts for 1.4 million patients over 65 years with chronic conditions recruited across all healthcare tiers. The research also explored the temporal order of disease diagnosis of COPD and comorbidities at a population level which might help to further understand the dynamics of comorbidity clustering often seen in patients with COPD.[3 11]

METHODS

Dataset and study population

The study is based on registry data of over 7.6 million inhabitants retrieved from the CHSS, including: primary care consultations, hospital-related events (hospitalisations, emergency room consultations and specialised outpatient visits), pharmacy, mental health events, socio-sanitary services and other items, such as home-based respiratory therapies, dialysis, outpatient rehabilitation and non-urgent healthcare transportation.[12 13] The study used a cross-sectional analysis, incorporating all patients over 65 years and registered in the CHSS who were active and alive during 2016 (n=1 433 376).

The research considered only chronic conditions of the patients, expressed with the Chronic Condition Indicator for International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) coding (Agency for Healthcare Research and Quality, USA).[14] Diagnosis of COPD was based on the ICD-9 coding (online supplementary table S1) declared by the patient's responsible physician, either a primary care professional or a specialist. The study did not take into account background clinical information nor forced spirometry data. The study identified patient with COPD (n=211 418) and their concomitant diseases, which were aggregated in 27 disease groups (DGs) (see online supplementary table S2) representing clinically significant traits.[8] DGs with <1% prevalence in the study population were excluded from the study.

Heterogeneities between the Medicare dataset[8] and the current study were identified (table 1) and were taken into account in the analysis of the results. Briefly, the Medicare dataset included registries from hospitalised patients over 65 years, considering both acute and chronic conditions, from 1990 to 1993. In contrast, the current study considered chronic diagnosis of patients over 65 years during 2016, obtained from a broader healthcare scenario, including all hospital visits since 2005 as well as diagnoses made in the primary care centres since the first visit of the patient.

Statistical analysis

For each DG, period prevalence was computed as the proportion of existing DG cases between 1 January 2016 and 31 December 2016 compared with the total population in the dataset. Age-associated prevalence was computed for each DG in patients with COPD and without COPD for 5-year age windows between ages 65 and 90 (eg, 75 denotes the prevalence between 73 and 77 years, both included).

Comorbidity association between COPD and DGs was measured using relative risk (RR) and phi correlation coefficient (Φ) (for detailed definition, see online supplementary methods).[8 15] Significance of these measures were assessed at the stringent threshold $p<0.0001$, associated with a Bonferroni corrected p-value<0.01. Healthcare system-related differences in comorbidity associations were compared using two-sided t-tests of RR measures, where $p<0.0001$ was considered significant.

Temporal order of disease diagnosis

Diagnosis history of the patients included in the CHSS registries was used to study the temporal order of disease diagnosis (COPD ↔ disease). Date of diagnosis for a given DG was defined by the first diagnosis of any corresponding disease in the DG. Patient DG diagnoses, made before and after the COPD diagnosis, were counted to characterise their temporal order. Cases in which COPD was diagnosed simultaneously with other DGs for the first time in the same visit were disregarded from the temporal analysis (see online supplementary figure S1).

For significant comorbid conditions, the directionality of temporal order of disease diagnosis was tested. *Preferred* direction (the one that appears more often) was assigned to those COPD–DG pairs where significantly more patients were diagnosed with the DG before COPD or the other way around, using a binomial test for each direction with a probability of success equals to 0.5 and sample size $N_{DG}+N_{simultaneous}+N_{COPD}$, where $N_{simultaneous}$ indicates the number of patients with simultaneous first diagnosis of COPD and the DG.[9] To estimate the strength of the

Table 1 Description of the datasets and methodological considerations of the current study and the previous study of Gomez-Cabrero and colleagues[8]

	Study population	Study period	Scope of data	Diseases considered
Current study	1.4 million (CHSS)	2016 +diagnosis history	Primary care, hospital claims, social care, others	Chronic
Gomez-Cabrero *et al*[8]	13 million (Medicare)	1990–1993	Hospital claims	Chronic, acute

CHSS, Catalan Healthcare Surveillance System.

directional associations, the causal information fraction (CIF) was used, which, in contrast to RR and Φ, considers the order of occurrence of the disease pairs and emphasises possible causative effects.[16] CIF is defined between a pair of diseases i and j, as

$$f_{i\to j} = \frac{n_{i\to j}}{n_i} - \frac{n_j}{2N}$$

where n_i and n_j denote the prevalence of the diseases, $n_{i\to j}$ is the number of individuals diagnosed with disease i followed by disease j and N is the population size. In the current analysis, CIF was used to compute COPD–DG associations.

RESULTS

Comorbidity risk in patients with COPD in different healthcare systems

Nine DGs were discarded from the comparative analysis due to containing solely acute diseases (two DGs) and because of showing <1% prevalence (seven DGs) (online supplementary table S3).

The prevalence of 18 DGs included in the analysis is indicated in figure 1 by the size of the bars. Comparison of prevalence results within the CHSS dataset (red bars) indicates that patients with COPD (dark colour) have higher risk of developing most of the DGs compared with a general patient of the healthcare system (light colour). Identical prevalence patterns are observed in the Medicare dataset (cyan bars). These results are also consistent with the analysis of COPD comorbidity risk based on comorbidity measures (ie, RR, Φ-correlation), indicating significant (p<0.0001) disease association between COPD and all the DGs in both healthcare systems (online supplementary table S3). The comparative analysis between the two datasets (figure 1) shows significant differences in the RR for several DGs that are fully explainable by the heterogeneities in the data sources described in the section Methods. For example, acute diseases, not considered in the current analysis (CHSS), accounted for more than 90% of Medicare cases in *endocrine disorders* and *skin alterations* (online supplementary table S4) leading to the visible differences in prevalence.

Figure 2 compares the age-associated prevalence of *heart diseases*, *circulatory disorders* and *digestive alterations*[8] between patients with COPD and without COPD in the two datasets. The figure indicates that in the two datasets the prevalence of the three DGs is consistently higher in patients with COPD (red lines), and that similar age-associated comorbidity patterns are observed. Interestingly, the prevalence of *heart diseases* for the two groups, COPD and non-COPD, is higher in the Medicare dataset than in the current study, representative of a Mediterranean population with mostly non-hospitalised patients. However, prevalence of *heart diseases* increases more steeply with age in the CHSS dataset. Similar age-associated prevalence of the remaining DGs in the CHSS dataset is displayed in online supplementary figure S2.

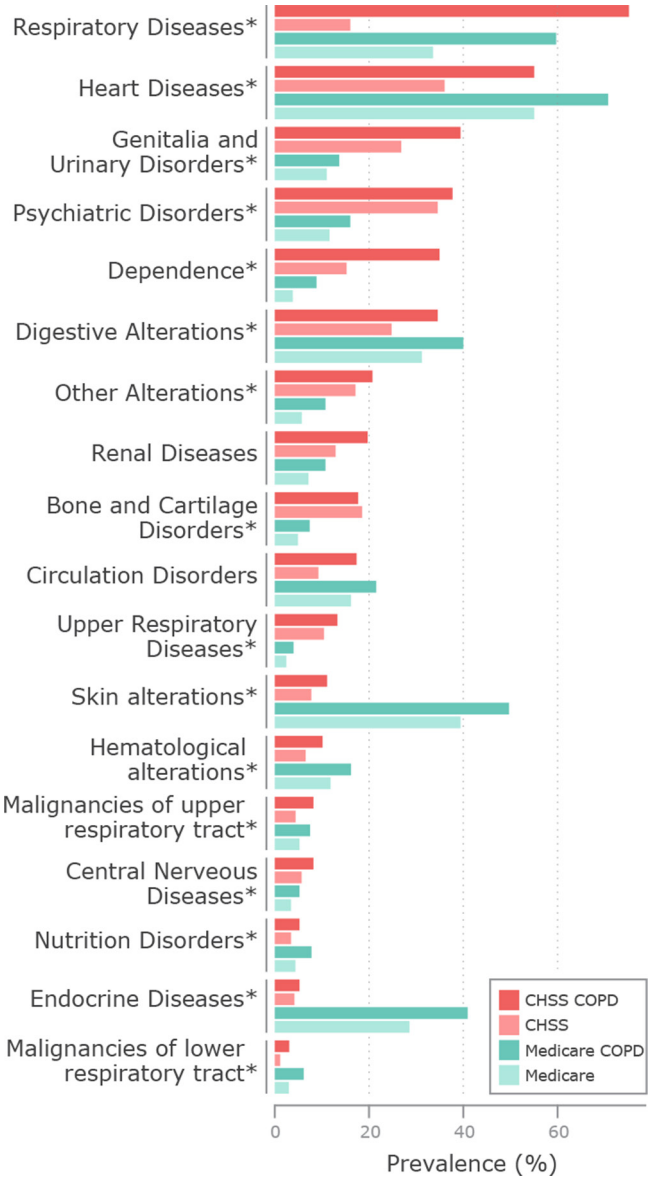


Figure 1 Prevalence (x axis) of disease groups (DGs) (y axis) in the population of Medicare (light cyan) and Catalan Healthcare Surveillance System (CHSS) (light red), and in patients with chronic obstructive pulmonary disease (COPD) in Medicare (dark cyan) and in CHSS (dark red). The comparative analysis within datasets shows that the prevalence of most of the DGs is higher in patients with COPD (dark colour) than in the entire population (light colour). Differences in the prevalence between datasets are fully explainable by methodological heterogeneities, detailed in the main text. Healthcare system-related differences in comorbidity associations were compared using two-sided t-tests of relative risk measures (*), p<0.0001.

Temporal order of disease diagnosis

The analysis of temporal order of disease diagnosis with respect to COPD is shown in figure 3A. Red bars indicate the patients in whom the first diagnosis of a disease from a given DG was done before the diagnosis of COPD. The number of patients in whom the corresponding DG was

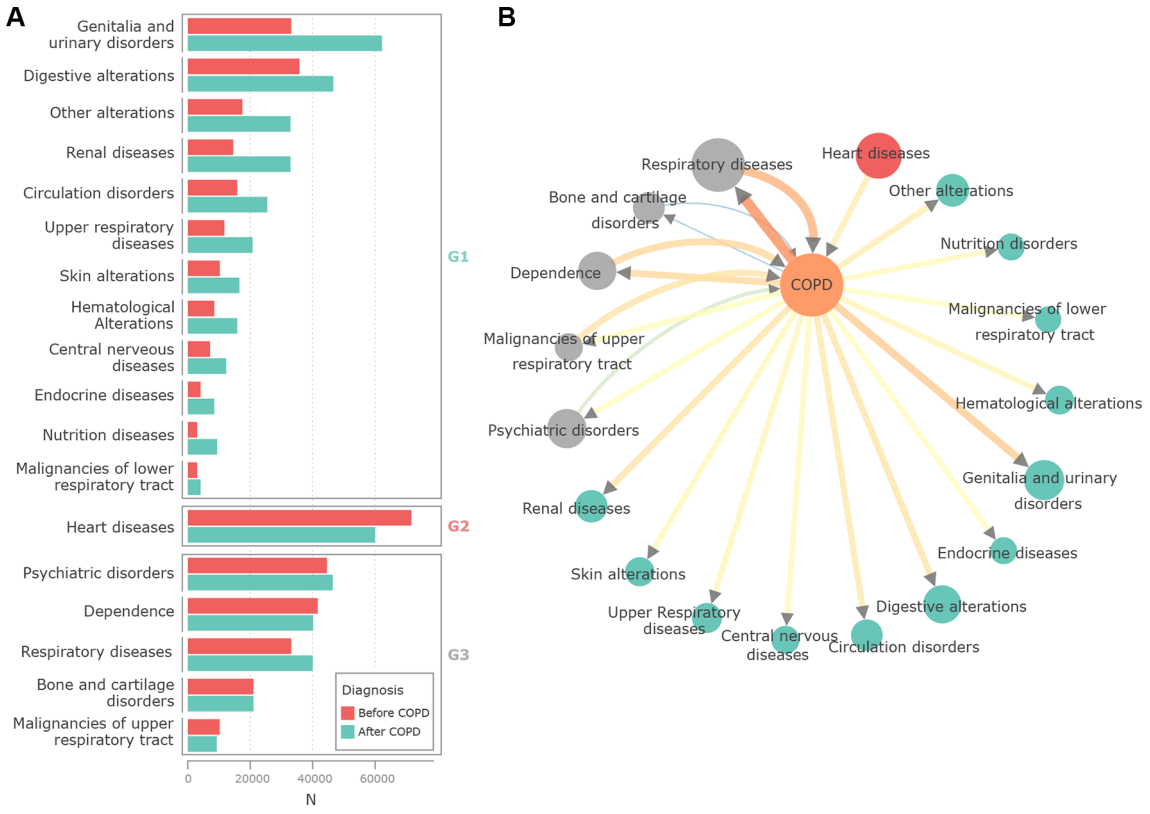


Figure 3 (A) Temporal order of pairwise diagnoses in patients with chronic obstructive pulmonary disease (COPD). Red bars show the number of patients whose first diagnosis of a disease from the corresponding disease group (DG) happened before COPD, whereas cyan bars show the cases when such diagnoses were done after COPD. DGs are grouped into preferred directions: (1) G1, DG diagnosis after COPD, (2) G2, disease diagnosis before COPD and (3) G3, no significant directionality. (B) Elderly comorbidity network. Network nodes represent COPD and the different DGs. DGs are coloured by their directionality grouping: cyan for G1, red for G2 and grey for G3. The size of the nodes is proportional to the number of cases affected by both COPD and the DG, colour and thickness of edges are proportional to the strength of the directional association based on the causal information fraction measure. It is of note that simultaneous diagnoses were excluded from the analysis or accounted for when computing binomial directionality. This mainly influenced the data shown on *respiratory diseases* (>45% of COPD diagnoses were made simultaneously, online supplementary figure S1).

identified after the diagnosis of COPD are indicated by the cyan bars. It is of note, that for the majority of DGs, COPD was diagnosed first (G1). Interestingly, only *heart diseases* (G2) were more often diagnosed before COPD than after COPD. A third group of DGs (G3) showed no preferred direction. Figure 3B translates these interac-tions to a network representation, with directional edges based on the grouping. Directional strengths of the asso-ciation (ie, CIF measure).

DISCUSSION

The current research confirms in the CHSS dataset that patients with COPD are in higher risk of developing certain comorbidities than in patients without COPD, along with the results reported in Gomez-Cabrero *et al*.[8] Despite marked methodological heterogeneities, similar age-related prevalence patterns were also observed between the current study and the Medicare dataset in elderly patients, as displayed in figure 2. These results provide a population health perspective of the comorbidity challenge in patients with COPD. It is of note that they are in line with a recent independent report carried out using a similar population-based analysis.[17]

It is known that clinical prevalence of certain comorbidities is higher in patients with COPD than in those without COPD.[18 19] In this study, we show that this effect is also observable using registry data, independently of the population (ES, USA) and the specificities of the healthcare system (figure 1). This relation also persists if studying both acute and chronic (Medicare, USA) or only chronic diseases (CHSS, ES), suggesting the validity of disease interactions on the functional trait level represented by the DGs. Interestingly, age-related patterns of elevated comorbidity risk are similarly observable in the different healthcare systems reinforcing that this effect is persistent in the elderly population (>65 years).

The results also showed that comorbid conditions in elderly patients with COPD tend to appear after the diagnosis of the obstructive disease, which seems to be in line with other studies on disease trajectories.[9 20] Different reports[17 21] have suggested age-dependent comorbidity patterns in patients with COPD, which indicate an

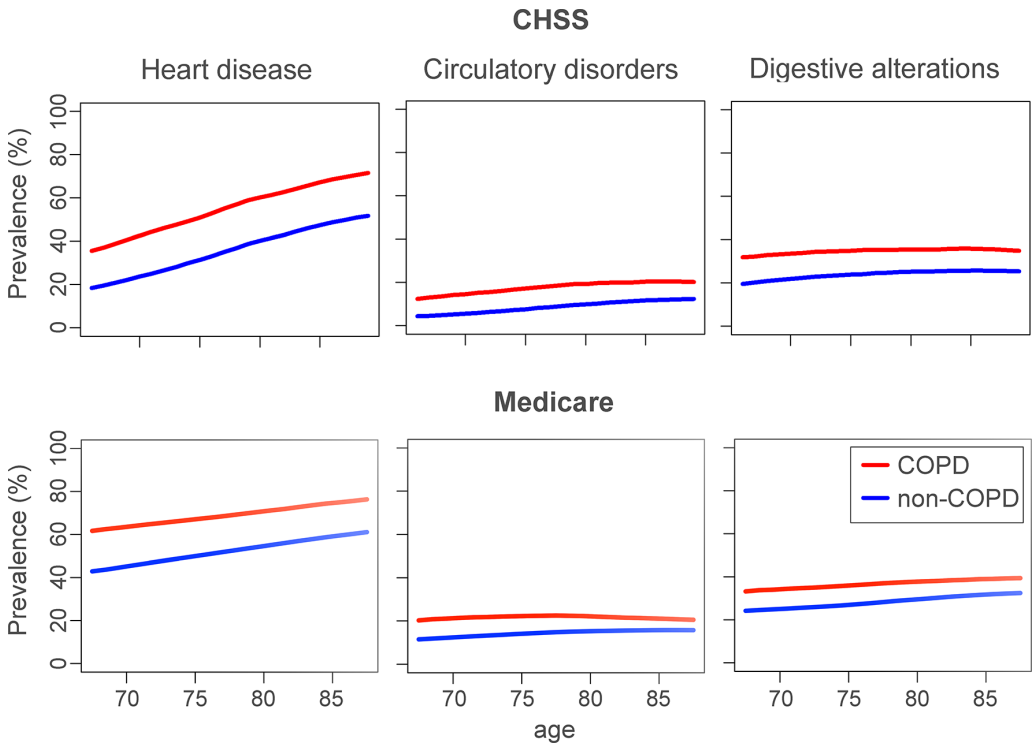


Figure 2 Comparison of the age-associated prevalence (y axis) in the Catalan Healthcare Surveillance System (CHSS) and Medicare datasets of selected disease groups in patients with chronic obstructive pulmonary disease (COPD) (red) and non-COPD (blue) individuals over windows of 5 years (x axis). This figure shows that patients with COPD in both datasets showed a higher risk for *heart disease*, *circulatory disorders* and *digestive alterations*.

interesting direction for future analysis of temporal order of disease diagnoses. Furthermore, the results showing the distinguished role of cardiovascular health in COPD are also in line with earlier studies on comorbidity clustering[21] and with its consistent relation to systemic effects of the disease.[22 23] This indicate potential synergies between the management of pulmonary and cardiovascular health and the promotion of physical activity from early stages of the disease with the potential to modulate prognosis in these patients.

The increased risk of developing comorbid conditions in patients with COPD, as well deleterious interactions among concurrent diseases,[24 25] indicates the need for refining current strategies aiming at reducing the burden of COPD on healthcare systems. In this context, prevention and appropriate treatment of comorbidities arise as central goals in the management of patients with COPD. These considerations are especially relevant in light of reports indicating that the majority of hospital admissions (and increased patients costs) are associated with comorbidities instead of the pulmonary events.[6]

Considering the high predictive potential of comorbidity groupers on these events,[6] further evaluation of other modalities of disease interactions, such as temporal order of appearance,[9 26] concomitant clinical characteristics,[21] lifestyle and genetic risk,[27] constitute as interesting next step towards high accuracy health risk prediction, as proposed in Dueñas-Espín *et al*.[10] Furthermore, a better knowledge of the underlying molecular mechanisms that modulate susceptibility for developing comorbidities in patients with COPD also emerges as major priority. It constitutes an initial step toward elaboration of cost-efficient strategies to prevent comorbid conditions.[28]

However, a large-scale combined approach is indispensable in order to define efficient strategies coping with comorbidity clustering in patients with COPD. It involves meaningful integration of registry data with other information sources reflecting a broader health status, such as electronic health records, environmental and occupational exposures and genetic risks. Furthermore, the evolution of population-based analyses towards personalised approaches, such as identifying personal disease progression[9 26] or their transcriptional patterns[29] and comparing it with similar patients profiles,[30–32] is needed to address disease heterogeneity often seen in COPD and to progress towards personalised health risk assessment and service selection.[6 10]

Finally, it is acknowledged that risk of developing comorbidities can be modulated by confounding risk factors not included in the current analysis, such as degree of airflow limitation and smoking history. It is important to note, however, that risk factors alone cannot explain the observed effects,[17 33] which reinforces the need for a large-scale, combined approaches incorporating patient information at population and patient-specific level.

It is acknowledged that registry information alone reflects underdiagnosis of COPD and the lack of forced spirometry data constitutes a significant limitation for accurate

diagnosis of COPD. These barriers also indicate the need for speeding-up efforts to facilitate integration between clinical and registry data.

CONCLUSION

The current research confirms that patients with COPD are in higher risk of developing certain comorbidities than patients without COPD. The study results, as well as ongoing research on time-related analyses of disease trajectories,[7 9] strengthen the need for further investigations on underlying mechanisms of non-pulmonary phenomena observed in patients with COPD with focus on altered regulation of biological pathways likely shared by different comorbid conditions. Furthermore, the study suggests the need for exploring novel modalities for health risk assessment and patient management aiming at consolidating cost-effective strategies to prevent comorbidities.[6 10] In this context, current standard of care recommendations[4] should necessarily evolve from the current organ-centred orientation to a systems approach.

Author affiliations
[1]Hospital Clinic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain
[2]Center for Biomedical Network Research in Respiratory Diseases (CIBERES), Madrid, Spain
[3]Unitat d'Informació i Coneixement, Servei Catala de la Salut de la Generalitat de Catalunya, Barcelona, Catalunya, Spain
[4]Serveis Centrals, Institut Català de la Salut, Barcelona, Spain
[5]Unit of Computational Medicine, Department of Medicine, Karolinska Institute, Stockholm, Sweden
[6]Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden
[7]Mucosal and Salivary Biology Division, King's College London Dental Institute, London, UK

Acknowledgements We want to acknowledge the support of NEXTCARE team (COMRDI15-1-0016), AGAUR research groups (2009SGR911 and 2014SGR661) and CERCA Programme/Generalitat de Catalunya.

Contributors Study conception and design: AT, IC, DG-C and JR. Data acquisition: EV, DM and MC. Data analysis: AT, EV, DM and MC. Manuscript preparation: AT, EV, IC, DG-C and JR. Manuscript revision: all authors.

Funding This work was supported by the European Commission grant CONNECARE (H2020-689802).

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Khakban A, Sin DD, FitzGerald JM, *et al*. The projected epidemic of COPD hospitalizations over the next 15 years: a population based perspective. *Am J Respir Crit Care Med* 2016.
2. McLean S, Hoogendoorn M, Hoogenveen RT, *et al*. Projecting the COPD population and costs in England and Scotland: 2011 to 2030. *Sci Rep* 2016;6:31893.
3. Vanfleteren LE, Spruit MA, Groenen M, *et al*. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013;187:728–35.
4. Vogelmeier CF, Criner GJ, Martinez FJ, *et al*. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. *Am J Respir Crit Care Med* 2017;195:557–82.
5. Chen W, FitzGerald JM, Sin DD, *et al*. Excess economic burden of comorbidities in COPD: a 15-year population-based study. *Eur Respir J* 2017;50:1700393.
6. Vela E, Tényi Á, Cano I, *et al*. Population-based analysis of patients with COPD in Catalonia: a cohort study with implications for clinical management. *BMJ Open* 2018;8:e017283.
7. Hu JX, Thomas CE, Brunak S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* 2016;17:615–29.
8. Gomez-Cabrero D, Menche J, Vargas C, *et al*. From comorbidities of chronic obstructive pulmonary disease to identification of shared molecular mechanisms by data integration. *BMC Bioinformatics* 2016;17:23–35.
9. Jensen AB, Moseley PL, Oprea TI, *et al*. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat Commun* 2014;5.
10. Dueñas-Espín I, Vela E, Pauws S, *et al*. Proposals for enhanced health risk assessment and stratification in an integrated care scenario. *BMJ Open* 2016;6:e010301.
11. Garcia-Aymerich J, Gómez FP, Benet M, *et al*. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. *Thorax* 2011;66:430–7.
12. Farré N, Vela E, Clèries M, *et al*. Real world heart failure epidemiology and outcome: A population-based analysis of 88,195 patients. *PLoS One* 2017;12:e0172745.
13. Department of Health C. Programa públic d'analítica de dades per a la recerca i la innovació en salut (PADRIS). 2017. http://salutweb.gencat.cat/web/.content/home/ambits_tematics/linies_dactuacio/recerca/enllacos/Programa_analitica_dades_PADRIS_aquas2017_publica.pdf
14. Cherkin DC, Deyo RA, Volinn E, *et al*. Use of the International Classification of Diseases (ICD-9-CM) to identify hospitalizations for mechanical low back problems in administrative databases. *Spine* 1992;17:817–25.
15. Hidalgo CA, Blumm N, Barabási AL, *et al*. A dynamic network approach for the study of human phenotypes. *PLoS Comput Biol* 2009;5:e1000353.
16. Kannan V, Swartz F, Kiani NA, *et al*. Conditional disease development extracted from longitudinal health care cohort data using layered network construction. *Sci Rep* 2016;6:26170.
17. Divo MJ, Celli BR, Poblador-Plou B, *et al*. Chronic Obstructive Pulmonary Disease (COPD) as a disease of early aging: Evidence from the EpiChron Cohort. *PLoS One* 2018;13:e0193143.
18. Cazzola M, Bettoncelli G, Sessa E, *et al*. Prevalence of comorbidities in patients with chronic obstructive pulmonary disease. *Respiration* 2010;80:112–9.
19. Mapel DW, Hurley JS, Frost FJ, *et al*. Health care utilization in chronic obstructive pulmonary disease. A case-control study in a health maintenance organization. *Arch Intern Med* 2000;160:2653–8.
20. Giannoula A, Gutierrez-Sacristán A, Bravo Á, *et al*. Identifying temporal patterns in patient disease trajectories using dynamic time warping: A population-based study. *Sci Rep* 2018;8:4216.
21. Divo MJ, Casanova C, Marin JM, *et al*. COPD comorbidities network. *Eur Respir J* 2015;46:640–50.
22. Cano I, Selivanov V, Gomez-Cabrero D, *et al*. Oxygen pathway modeling estimates high reactive oxygen species production above the highest permanent human habitation. *PLoS One* 2014;9:e111068.
23. Tényi Á, Cano I, Marabita F, *et al*. Network modules uncover mechanisms of skeletal muscle dysfunction in COPD patients. *J Transl Med* 2018;16:34.
24. Almagro P, De la Sierra A, Acosta E, *et al*. Spirometrically confirmed chronic obstructive pulmonary disease worsens long-term prognosis after percutaneous coronary intervention. *Am J Respir Crit Care Med* 2018;197:824–6.
25. Roversi S, Fabbri LM, Sin DD, *et al*. Chronic obstructive pulmonary disease and cardiac diseases. an urgent need for integrated care. *Am J Respir Crit Care Med* 2016;194:1319–36.

26. Beck MK, Jensen AB, Nielsen AB, *et al*. Diagnosis trajectories of prior multi-morbidity predict sepsis mortality. *Sci Rep* 2016;6:36624.
27. Muse ED, Wineinger NE, Spencer EG, *et al*. Validation of a genetic risk score for atrial fibrillation: A prospective multicenter cohort study. *PLoS Med* 2018;15:e1002525.
28. Divo M, Cabrera C. Which are the most relevant comorbidities In COPD? *Barcelona Respiratory Network* 2016;2:215–28.
29. Menche J, Guney E, Sharma A, *et al*. Integrating personalized gene expression profiles into predictive disease-associated gene pools. *NPJ Syst Biol Appl* 2017;3:10.
30. Sharafoddini A, Dubin JA, Lee J. Patient similarity in prediction models based on health data: a scoping review. *JMIR Med Inform* 2017;5:e7.
31. Brown SA. Patient similarity: emerging concepts in systems and precision medicine. *Front Physiol* 2016;7:561.
32. Torkamani A, Andersen KG, Steinhubl SR, *et al*. High-definition medicine. *Cell* 2017;170:828–43.
33. McGarvey LP, John M, Anderson JA, *et al*. Ascertainment of cause-specific mortality in COPD: operations of the TORCH Clinical Endpoint Committee. *Thorax* 2007;62:411–5.

MANUSCRIPT 4: POPULATION-BASED ANALYSIS OF PATIENTS WITH COPD IN CATALONIA: A COHORT STUDY WITH IMPLICATIONS FOR CLINICAL MANAGEMENT

Vela E, **Tényi Á**, Cano I, Monterde D, Cleries M, Garcia-Altes A, Hernandez C, Escarrabill J, Roca J. Population-based analysis of patients with COPD in Catalonia: a cohort study with implications for clinical management. *BMJ Open* British Medical Journal Publishing Group; 2018; 8: e017283.

- Journal: *BMJ Open*
- Impact factor: *2.369*
- Quartile: *1st*
- Status: *Published*

BMJ Open

Population-based analysis of patients with COPD in Catalonia: a cohort study with implications for clinical management

Emili Vela,[1] Ákos Tényi,[2,3] Isaac Cano,[2,3] David Monterde,[4] Montserrat Cleries,[1] Anna Garcia-Altes,[5] Carme Hernandez,[2,3] Joan Escarrabill,[2,6] Josep Roca[2,3]

To cite: Vela E, Tényi Á, Cano I, *et al*. Population-based analysis of patients with COPD in Catalonia: a cohort study with implications for clinical management. *BMJ Open* 2018;**8**:e017283. doi:10.1136/bmjopen-2017-017283

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-017283).

Received 13 April 2017
Revised 13 December 2017
Accepted 19 December 2017

[1]Area d'Atenció Sanitària, Servei Català de la Salut, Barcelona, Catalonia, Spain
[2]Hospital Clinic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain
[3]Center for Biomedical Network Research in Respiratory Diseases (CIBERES), Madrid, Spain
[4]Institut Català de la Salut, Serveis Centrals, Catalunya, Spain
[5]Agencia de Qualitat i Avaluació Sanitaries de Catalunya (AQuAS), Catalunya, Spain
[6]Master Plan for Respiratory Diseases (PDMAR), Ministry of Health (Catalonia) REDISSEC, Health Services Research on Chronic Patients Network, Instituto de Salud Carlos III, Barcelona, Spain

Correspondence to
Professor Josep Roca;
jroca@clinic.cat

ABSTRACT

Background Clinical management of patients with chronic obstructive pulmonary disease (COPD) shows potential for improvement provided that patients' heterogeneities are better understood. The study addresses the impact of comorbidities and its role in health risk assessment.

Objective To explore the potential of health registry information to enhance clinical risk assessment and stratification.

Design Fixed cohort study including all registered patients with COPD in Catalonia (Spain) (7.5 million citizens) at 31 December 2014 with 1-year (2015) follow-up.

Methods A total of 264 830 patients with COPD diagnosis, based on the International Classification of Diseases (Ninth Revision) coding, were assessed. Performance of multiple logistic regression models for the six main dependent variables of the study: mortality, hospitalisations (patients with one or more admissions; all cases and COPD-related), multiple hospitalisations (patients with at least two admissions; all causes and COPD-related) and users with high healthcare costs. Neither clinical nor forced spirometry data were available.

Results Multimorbidity, assessed with the adjusted morbidity grouper, was the covariate with the highest impact in the predictive models, which in turn showed high performance measured by the C-statistics: (1) mortality (0.83), (2 and 3) hospitalisations (all causes: 0.77; COPD-related: 0.81), (4 and 5) multiple hospitalisations (all causes: 0.80; COPD-related: 0.87) and (6) users with high healthcare costs (0.76). Fifteen per cent of individuals with highest healthcare costs to year ratio represented 59% of the overall costs of patients with COPD.

Conclusions The results stress the impact of assessing multimorbidity with the adjusted morbidity grouper on considered health indicators, which has implications for enhanced COPD staging and clinical management.

Trial registration number NCT02956395.

Strengths and limitations of this study

- The main strength of the study is that it contributes to risk prediction of relevant clinical events in patients with chronic obstructive pulmonary disease (COPD).
- The study shows high potential to assess health risk factors at population level indicating the high impact of comorbidities. It can contribute to define innovative strategies aiming at reducing the healthcare impact of patients with COPD.
- Full potential of the approach should be proven by integrating registry information and electronic medical records.
- Lack of clinical information, spirometric data and history of tobacco smoking reduces the potential for standardised risk characterisation of patients with COPD.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is one of the major disorders included in the WHO programme addressing non-communicable diseases.[1] It is estimated that COPD will become the third leading cause of death by 2020.[2] Moreover, projections on COPD prevalence and costs over the next 15 years indicate a rapidly escalating burden, mainly due to population ageing, on both health and social support systems.[3 4]

While acknowledging the progress made in terms of standard of care recommendations,[5] it is accepted that a better understanding of patients' heterogeneities constitutes a key challenge to further enhance both prevention and management of patients with COPD aiming at healthcare value generation.[6 7] Recent studies indicate a high impact of comorbidities on use of healthcare resources in patients with COPD prompting the need for assessing novel integrated care strategies with a patient-oriented approach.[8 9] It is well accepted that several prevalent chronic conditions often occur as clusters of comorbidities in patients with COPD[10–12] and potential explanatory mechanisms for the phenomenon have been proposed.[13 14]

The current study addresses comorbidities in patients with COPD based on the hypothesis that assessment of all patients with COPD living in a given geographical area, a population-based analysis of the patients with the disease, can provide valuable insights

to better understand COPD heterogeneity. Specifically, the research objective was to explore the potential of the health registry information contained in the Health Surveillance System of the region of Catalonia (Spain) (7.5 million inhabitants) to enhance health risk assessment and stratification in the clinical arena.[15 16] To this end, we analysed a total of 264 830 patients with COPD from all healthcare layers, registered in 31 December 2014 and followed up in 2015, to elaborate predictive models for six key health indicators: (1) mortality, (2 and 3) hospitalisations (all causes and COPD-related), (4 and 5) multiple hospitalisations (all causes and COPD-related) and (6) users with high healthcare costs. Ultimately, this study aimed to consider the weight of the different covariates included in the predictive models to provide useful information to enhance clinical management.

The study is a relevant component of the programme for collaborative management of complex chronic patients being deployed in the region during the period 2016–2020.[17] An ancillary aim of the research was to assess factors determining the economic impact of the patients with the disease, as well as to identify areas of action to increase healthcare efficiencies in the management of these patients.

METHODS

Population-based risk assessment: adjusted morbidity grouper

The Catalan Health Surveillance System (CHSS) includes updated registries of the region of Catalonia (Spain) (7.5 million inhabitants) from Primary Care, Hospital-related events (hospitalisations, emergency room consultations and specialised outpatient visits), Pharmacy, Mental Health, Socio-sanitary services and other items (home-based respiratory therapies, dialysis, outpatient rehabilitation and non-urgent healthcare transportation) since 2011.[18 19] It allows analyses on use of healthcare resources, pharmacy consumption, prevalence of key disorders and population-based health risk assessment.[15 16] It is of note that although integration of CHSS registry data with electronic medical records is not yet in place, it constitutes the main goal of the PADRIS programme,[20] officially launched on January 2017.

The regional population-based health risk assessment tool, named GMA (Adjusted Morbidity Groups), is used to elaborate the health risk strata pyramid of the general population of Catalonia (figure 1, left triangle).[15 16] The GMA tool predicts individual patient risk, periodically updated on a 6-month basis, based on multimorbidity information gathered from CHSS registry data. The rationale behind the use of GMA, against alternative health risk assessment tools, is that it complies with four main recommended criteria,[15] that is, (1) a population health approach (uses the entire population of 7.5 million inhabitants of the region), (2) publicly owned without licensing constraints, (3) open source computational algorithms and (4) the adjusted morbidity grouper relies mostly on statistical criteria, as opposed to other tools that include expert-based coefficients, thus facilitating quick transferability to other territories. Detailed descriptions of the GMA, as well as its evaluation, have been reported elsewhere.[15 16] Methodological details of the GMA algorithm are described in the online supplementary figure 1S.

Study dataset and design

The current study design is a fixed cohort analysis of the entire population of 264 830 patients, alive and above 39 years of age, included in the CHSS on 31 December 2014 with the clinical diagnosis of COPD. New patients registered during 2015 were not included in the study. For the study purposes, the diagnosis of COPD was based on the International Classification of Diseases (Ninth Edition) coding[21] (online supplementary table 1S) declared by the patient's responsible physician, either a

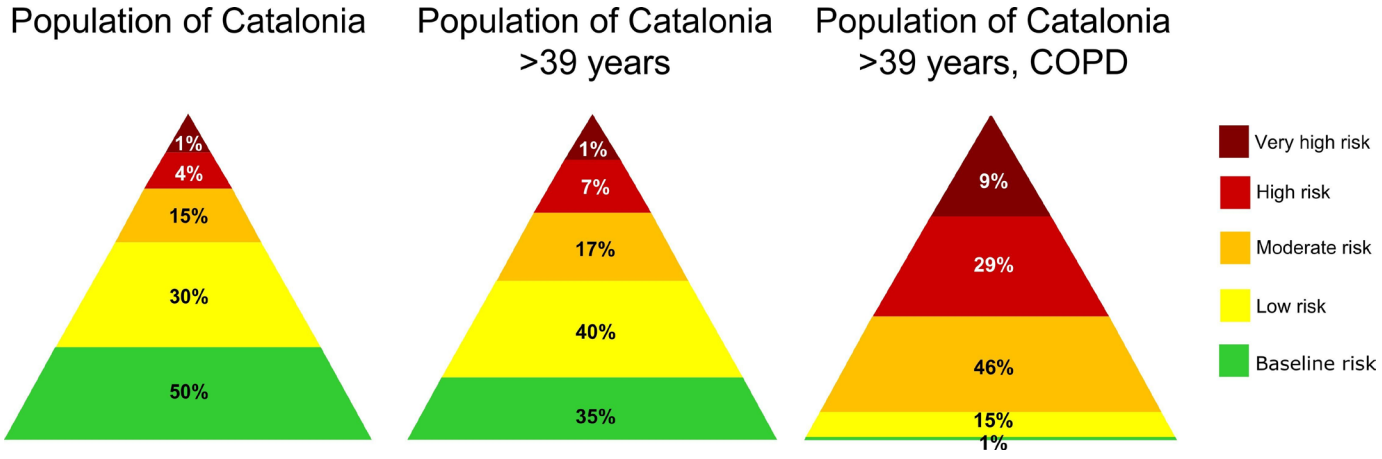


Figure 1 GMA health risk grades. The left triangle depicts the distribution of all Catalan citizens, expressed as percentage, in five arbitrary health risk layers defined using the GMA (Adjusted Morbidity Groups)[15 16] as a population-based health risk stratification tool. The central triangle indicates the distribution in these five health risk layers of the subset of Catalan citizens older than 39 years. The right triangle displays the distribution of the study group of patients with chronic obstructive pulmonary disease (COPD) (264 830 patients) across the GMA health risk grades: baseline, low, moderate, high and very high.

primary care professional or a specialist. The study did not take into account background clinical information, nor forced spirometry data.

The analyses were carried out for the entire population of patients with COPD, but the study considered different subsets of patients based on the GMA health risk grading: (1) baseline, (2) low, (3) moderate, (4) high and (5) very-high-risk patients. The thresholds defining these subsets of patients correspond to the percentiles 50, 80, 95 and 99 of the GMA grading for the general population (figure 1, left triangle). For the purposes of the study, patients with COPD falling into the baseline health risk group (1%) were merged with the low-risk group such that only four GMA grades were considered in the analysis (figure 1, right triangle).

Health risk predictive modelling was elaborated with registry data from year 2015 for the six main dependent variables: (1) mortality, (2 and 3) hospitalisation (all causes and COPD-related), (4 and 5) multiple hospitalisations (all causes and COPD-related) and (6) users with high healthcare costs, as described below. In the study, dependent variables were defined as binary variables. Mortality was defined as true if a patient died during the period from 1 January to 31 December 2015, regardless of whether it occurred in the hospital, at the patient's home or in other settings such as skilled nursing facilities. The category all causes hospitalisations was true if a patient had one or more hospital admissions due to any cause during 2015. COPD-related hospitalisations refer only to events triggered by acute exacerbations of COPD. Multiple hospitalisations were defined as two or more hospital admissions. Users with high healthcare costs refer to subjects above percentile 85 (PCT85) in terms of yearly healthcare costs during 2015.

The study used retrospective deidentified data from administrative databases (CHSS). Therefore, neither informed consent nor ethical committee approval was required according to the current legislation in Catalonia. The analyses were developed under the umbrella of the Nextcare project (http://www.nextcarecat.cat/) (https://clinicaltrials.gov/: NCT02956395).

Assessment of economic burden

Allocation of healthcare expenditure to each patient, including pharmacy, was done through the Personal Health Identification Number since it allows each billing invoice to be attributed to a given patient. The healthcare expenditure includes hospitalisation, primary care, pharmacy, health transport, respiratory home care therapies, outpatient visits and skilled nursing facilities. The key outcome variable of the analysis was 1-year healthcare resource use and expenditure by patient.[19] Calculation of individual healthcare costs was done by addition of costs for each item included in the CHSS registries alluded to above. This methodology allows calculation of the total healthcare expenditure in patients with COPD. Consequently, it allows to perform a holistic analysis of healthcare expenditure.

Statistical analysis

The study outcomes are described for the entire population of patients with COPD. Comparisons among the four subgroups defined by GMA health risk grades were done. The results are summarised in the main manuscript and complementary information is reported in detail in the online supplementary material. In this population-based analysis (table 1), age and number of chronic comorbidities are summarised as mean and SD, the proportion of women and the morbidities are expressed as percentages, while mortality rate, hospitalisation rate and COPD-related hospitalisation rate are expressed per 100 patients with COPD among patient groups. Comparisons among groups were done using analysis of variance for continuous variables, and χ^2 test for binary and nominal variables.

Statistical analyses were performed using SPSS software V.18.0. All statistical tests and confidence intervals were constructed with a type I error (alpha) level of 5%, and P values lower than 0.05 were considered statistically significant.

Predictive modelling

Multiple logistic regression analyses were used to generate health risk predictive models for the six main outcome variables of the study, namely, (1) mortality, (2 and 3) hospitalisation (all causes and COPD-related), (4 and 5) multiple hospitalisations (all causes and COPD-related) and (6) users with high healthcare costs. The following model independent variables were considered: age, sex, GMA health risk grades, previous history of hospital admissions, emergency room consultations and use of social support services. Each independent variable was included in the model as a categorical variable to allow for possible non-linearity in the relationship between variables and the relevant outcomes. For each measure, we collapsed the uppermost categories to ensure there were enough individuals in each cell to allow estimation of the parameters. All covariates were entered in the model one by one and retained when they showed a significant contribution to the predictive accuracy (P<0.10). The predictive role of each covariate into the model was assessed with a log-likelihood ratio test. To evaluate the performance of the resulting predictive models, we calculated the C-statistics (ie, the area under the receiver operating characteristic curve).[22 23]

RESULTS

Population-based analysis

The distribution of the entire population of the region in the health risk strata pyramid is depicted in figure 1, left triangle, wherein citizens are distributed in five health risk grades defined by the GMA percentiles, namely, (1) 50% of the population with baseline health risk (green), (2) 30% of individuals with low health risk (yellow), (3) 15% with moderate health risk (orange), (4) four percent with high risk and (5) 1% with very high health risk (brown).

Table 1 Main characteristics of the study group by GMA health risk grades

	Low risk	Moderate risk	High risk	Very high risk	Total
Patients (n)	41 809	121 918	76 237	24 866	264 830
Age (years)*	60.8±11.8	69.4±11.8	75.1±10.9	77.4±10.1	70.5±12.5
Women (%)*	36.0	36.2	35.9	35.3	36.0
Morbidity					
No of chronic comorbidities*	2.9±1.1	5.2±1.4	7.0±1.6	8.4±1.7	5.6±2.1
Diabetes (%)*	5.2	22.7	39.4	52.6	27.5
Heart failure (%)*	0.2	5.2	30.7	63.7	17.2
Hypertension (%)*	24.0	60.4	81.1	88.9	63.3
Renal failure (%)*	0.3	5.8	23.4	47.8	13.9
Dementia (%)*	0.3	2.6	7.8	13.6	4.7
Cirrhosis (%)	0.5	1.6	3.0	4.8	2.1
Depression (%)	8.8	20.1	28.6	34.2	22.1
Stroke (%)*	0.5	5.9	19.2	31.8	11.3
Ischaemic coronary disease (%)*	0.9	9.0	28.0	42.7	16.4
Malignancy (%)*	4.2	17.0	32.1	40.9	21.6
Locomotor system (%)*	29.3	56.3	67.9	72.0	56.8
Osteoporosis (%)*	3.9	10.0	14.3	17.9	11.0
Arthrosis (%)*	9.8	32.8	47.6	51.0	35.1
2015 events	n (%)	n (%)	n (%)	n (%)	n (%)
Mortality*	391 (0.9)	3213 (2.6)	6792 (8.9)	5955 (23.9)	16 351 (6.2)
Patients with hospitalisations (all causes)*	1785 (4.3)	12 451 (10.2)	19 417 (25.5)	12 496 (50.3)	46 149 (17.4)
Patients with hospitalisations (COPD related)*	404 (1.0)	2769 (2.3)	4697 (6.2)	3600 (14.5)	11 470 (4.3)
Patients with multiple hospitalisations (all causes)*	366 (0.9)	3026 (2.5)	6687 (8.8)	6035 (24. 0)	16 114 (6.1)
Patients with multiple hospitalisations (COPD related)*	71 (0.2)	524 (0.4)	1169 (1.5)	1210 (4.9)	2974 (1.1)
Users with high healthcare costs*	1119 (2.7)	10 064 (8.3)	17 333 (22.7)	11 609 (46.7)	40 125 (15.2)

Age and number of chronic comorbidities expressed as mean±SD; gender and morbidities are expressed as percentages; comparisons among risk grades were done using analysis of variance for continuous variables and χ^2 test for binary and nominal variables.
*P<0.01.
COPD, chronic obstructive pulmonary disease; GMA, Adjusted Morbidity Groups.

These risk strata show strong associations with mortality, hospital admissions, use of healthcare resources and expenditures, as reported in detail elsewhere.[15 16]

The central triangle (figure 1) indicates the distribution of the general population older than 39 years in the five health risk grades, described above. It constitutes the reference risk strata pyramid to be compared with the population-based COPD analysis. Finally, the distribution of the study group of patients with COPD is depicted in the right triangle that indicates the effects of the chronic pulmonary disease on health risk stratification. As expected, most of the patients with COPD (84%) were distributed between moderate (46%), high (29%) and very high (9%) health risk layers. Only 15% of the patients fell in the low risk level (yellow) and 1% were allocated in the baseline health risk level (green).

The study showed a COPD prevalence of 6.6% of all subjects above 39 years of age. The mean age was 70 (SD 12.5) years. Women represented 36% of the study group. On average, these patients presented 5.6 (SD 2.1) comorbid conditions. The mortality rate was 6.2%. The hospitalisation rate, all causes, was 17.4% (n=46 149), whereas the rate of COPD-related hospitalisations was 4.3% (n=11 470), which represents 24.9% of the total number of hospitalisations registered in these patients. As expected, the rate of comorbidities showed a consistent increase with GMA grading, which was the most apparent for cardiovascular disorders, type 2 diabetes mellitus–metabolic syndrome and/or anxiety–depression, as displayed in table 1.

Predictive modelling

The summary information of the predictive modelling for each of the six outcome variables—mortality, hospitalisations (all causes and COPD-related hospitalisations), multiple hospitalisations (all causes and COPD-related

Table 2 Summary description of the six predictive models

	Mortality	Hospitalisations		Multiple hospitalisations		Users with high healthcare costs (PCT85)
		All causes	COPD related	All causes	COPD related	
C-statistics (AUC)	0.829	0.766	0.807	0.803	0.865	0.763
Covariates	OR					
Sex						
Male	1	1	1	1	1	1
Female	0.701	0.825	0.885	0.818	0.823	0.84
Age (years)						
40–54	1	1	1	1	1	1
55–64	1.693	1.254	1.435	1.308	1.428	0.884
65–74	2.103	1.506	1.675	1.576	1.7	0.91
75–84	3.749	2.054	2.017	1.983	1.634	0.898
>84	9.911	2.827	2.049	2.335	1.272	0.681
Admissions						
Group A	1	1	1	1	1	1
Group B	1.394	1.903	4.699	1.952	6.952	1.59
Group C	1.368	2.039	8.592	2.266	15.214	1.577
GMA grade						
Low	1	1	1	1	1	1
Moderate	1.826	1.896	1.644	2.128	1.649	3.125
High	4.316	3.998	2.646	5.419	2.852	8.858
Very high	8.919	7.851	3.548	11.042	3.891	20.285
Emergency room visits						
0	1	1	1	1	1	1
1–2	1.106	1.405	1.37	1.526	1.505	1.353
3–5	1.276	1.865	1.684	2.179	2.013	1.771
>5	1.643	2.615	2.235	3.354	3.032	2.428
Social support						
No	1	1	1	1	1	1
Yes	2.289	1.008	0.892	0.917	0.922	0.877

The intensities of grey background colour reflect the magnitude of ORs, being white when value is one and stronger grey when they are closer to 0 or have a higher positive value.
Admissions:
Group A corresponds to patients with no registries of hospital admissions within the period 2011–2014; Group B includes patients with history of admissions before 2014, but without admissions in that year; Group C includes patients with hospital admissions during 2014. Graphical representation and details of the six predictive models with the corresponding 95% CIs of the ORs are shown in the online supplementary figures 2S–7S and tables 2S–7S.
AUC, area under the receiver operating characteristic curve; COPD, chronic obstructive pulmonary disease; GMA, Adjusted Morbidity Groups.

repeated admissions) and users with high healthcare costs—are depicted in table 2, wherein significant covariates, the corresponding ORs and the C-statistics are indicated for each predictive model.

It is of note that age, closely followed by GMA grading, showed the two highest independent associations in the mortality model. Likewise, GMA grading depicted the highest predictive role in three out of the six models: hospitalisations (all causes), multiple hospitalisations (all causes) and users with high healthcare costs. Interestingly, for COPD-related events (hospitalisations and multiple hospitalisations), the covariate with highest predictive role was history of hospitalisations, whereas the ORs of the covariate GMA grades were markedly lower than in the other predictive models.

Economic impact of patients with COPD

Figure 2 compares the costs of the different items, expressed as percentages, for the general population of Catalonia (outer circle) and those generated by the study

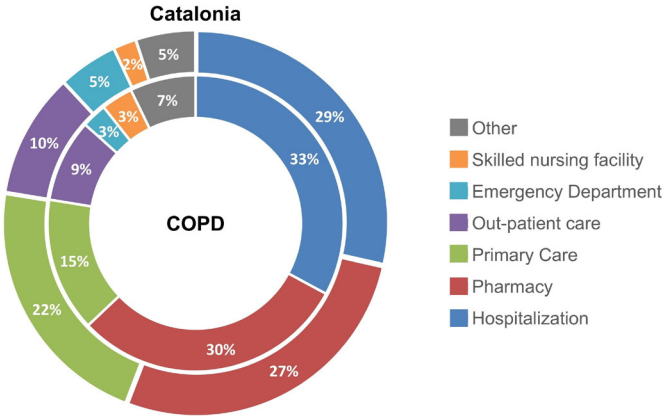


Figure 2 Indicates the distribution of costs of main items, expressed as percentages. The outer circle corresponds to overall cost for the Catalan Health System, whereas the inner circle indicates the corresponding relative costs ascribed to patients with chronic obstructive pulmonary disease (COPD). The absolute values are (1) Hospitalisation (€2291.8 million and €356.6 million, respectively), (2) Pharmacy (€2193.4 million and €325.8 million), (3) Primary care (€1745.0 million and €158.9 million), (4) Outpatient specialised care (€842.9 million and €98.1 million), (5) Emergency department (€401.5 million and €29.7 million), (6) Skilled nursing facility (Catalonia €155.1 million; COPD €37.5 million), and (7) Other (€404.0 million and €78.0 million). The last item, Others, includes home-based respiratory therapies, dialysis, outpatient rehabilitation and non-urgent healthcare transportation.

group of patients with COPD (inner circle). Briefly, COPD hospitalisations, pharmacy, skilled nursing care and other costs (respiratory therapies) are above the mean cost of the corresponding items in the general population of the region.

The average healthcare costs of patients with COPD per year was €4238 as compared with a mean of €987 a year per citizen in the region. It is of note that the cost generated by patients with COPD represents 13.5% of the overall healthcare costs in the region.

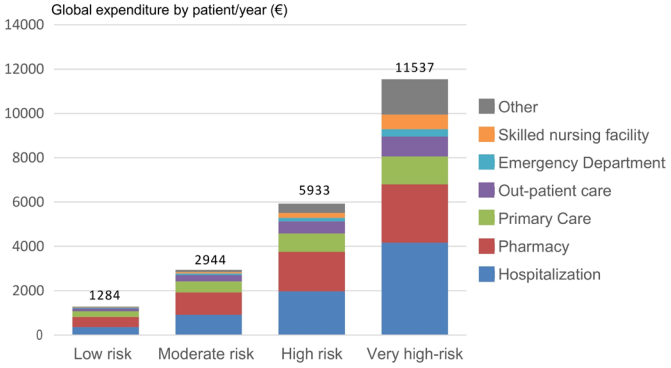


Figure 3 Average patient cost per year and relative contribution of the seven items (see text and figure 2 legend) for the four subgroups of patients with chronic obstructive pulmonary disease classified according to the GMA (Adjusted Morbidity Groups) scoring from low health risk (left column) to very high health risk (right column).

The cost analysis by GMA grading (figure 3) clearly showed a steady increase of costs per patient/year with considered GMA risk grades: (1) low, €1284; (2) moderate, €2944; (3) high, €5933; and (4) very high, €11 537. Hospitalisation and pharmacy are the two items with highest impact on costs associated to GMA grading. It is worth to mention that we observed huge differences (7.41 times) between the users with high healthcare costs per year (PCT85: €16 131), which represents 59% of the overall costs of patients with COPD, and patients with COPD below PCT85 (€2177). Online supplementary table 8S indicates use of healthcare resources by GMA category.

DISCUSSION

Interpretation of the main findings

The main finding of the current study was the identification of comorbidities, expressed as GMA grades, as the covariate with highest discriminative impact on target events (mortality, hospitalisations) (table 2) and on yearly healthcare costs per patient (figure 3). These results can be relevant for commissioning of innovative healthcare services aiming at preventing materialisation of recent predictions on increases of healthcare impact of patients with COPD over the next 15 years.[3 4] However, they may also foster enhanced clinical management of individual cases with COPD.

It is of note that the performance of these models (table 2 and online supplementary figure S2–S7) shows quite acceptable goodness of fit (ie, C-statistics), indicating the potential of exploring synergies between population-health risk assessment (ie, GMA grading system) and clinical information to enhance health risk assessment and stratification in the clinical arena, as reported in Dueñas-Espín *et al*.[15] Ultimately, integration between registry data and electronic medical records of healthcare providers emerge as a high priority goal to properly pave the way towards personalised medicine for patients with chronic disorders.[24]

As expected, the covariate GMA showed a higher contribution in all causes hospital-related events than in only COPD-related admissions (table 2). It is of note that the latter represented only 22% of all admissions (table 1). While current GOLD recommendations address pulmonary events explaining frequent COPD exacerbations, the results of the current study suggest the need for further analyses aiming at identifying specific management needs for complex patients with COPD and comorbid conditions.

Two complementary healthcare strategies

The distribution of patients with COPD in the regional health risk stratification pyramid (figure 1, right triangle) allows identification of two different scenarios with well-defined associated challenges.

Patients with high health risk

The distribution of yearly healthcare costs per patient (figure 3) provides a strong rationale for targeting

patients close to the tip of the risk stratification pyramid as candidates for large-scale deployment of innovative services based on care coordination. However, two main limiting factors should be overcome in order to achieve proper designs of integrated care services.[25] First, poor comparability among interventions assessing effects on integrated care management for patients with COPD indicates an urgent need for standardisation of service workflows. Second is the limited healthcare impact of standard interventions addressed to patients at the tip of the pyramid,[26] that is, the low ratio between magnitude of the interventional effects and the resources devoted to achieve them, which may imply little healthcare value generation.[6 7] These two factors strengthen the need for further evidence on cost-effectiveness of well-defined integrated care interventions for complex chronic patients.[27]

Patients with low and moderate health risk

A better understanding of underlying mechanisms of comorbidity clustering in these patients emerges as a central need to effectively slow down patients' progress towards the tip of the pyramid. A natural consequence should be the development of efficient preventive interventions[28] aiming at delaying patients' worsening in terms of health risk scoring, which is recognised as a central unmet need for enhanced COPD management.

Strengths and limitations of the study

The uniqueness of the current study is that it was carried out using registry data that allow population-based analyses of all patients with a given condition(s) in the region. It is of note that complexities involved in implementation and optimisation of large-scale health information technology systems often impede health assessment of the entire population in a real-world setting.[29]

We acknowledge, however, that the use of registry information alone reflects underdiagnosis of COPD and constitutes a significant limitation that may explain a rather low figure for COPD prevalence in the region.[5 30] The lack of clinical information, spirometric data and history of tobacco smoking reduces the potential for a proper characterisation of patients with COPD. As indicated above, the barriers associated to the existence of health information silos further strengthens the need for speeding up the current efforts to achieve real integration between clinical and registry data[5] that will open new avenues to enhance both medical knowledge and clinical practice.

Episodes of inpatient and outpatient care carried out in private hospitals were not available for analysis because private providers do not use the Personal Health Identification Number. Nevertheless, the vast majority (approximately 97.5% in 2015) of COPD hospitalisations are done in public hospitals.

Clinical impact and perspectives generated by the study

The current study confirms that prevalent chronic conditions such as cardiovascular disorders, type 2 diabetes mellitus–metabolic syndrome and/or anxiety-depression often occur in high-risk patients with COPD.[10–12] We believe that the study provides a strong rationale for further research on subject-specific health risk prediction and stratification aiming at early identification of patients with low to moderate health risk who are prone to develop comorbid conditions in order to enhance preventive management in a cost-effective manner.[31 32] The current research may contribute to foster future developments of GOLD recommendations[5] addressing non-pulmonary manifestations of COPD that should have a positive impact on both staging and management of complex chronic patients.

Moreover, the study reinforces the ongoing strategies aiming at speeding up the evolution of the current health surveillance system in Catalonia towards a Digital Health Framework conceptually formulated in Cano *et al*[33 34] with potential to articulate three categories of data: (1) outcomes from population-based health-risk predictive models; (2) healthcare and biomedical research knowledge resulting from integration of clinical, physiological and biological/molecular information; and (3) informal care information from in-place personal health folders[15 17] encompassing information on lifestyle, adherence profile, socioeconomic status, social support and environmental factors. It is envisaged that inclusion of all these covariates influencing patient health should markedly increase the predictive accuracy and facilitate clinical decision-making based on sound estimates of the prognosis of an individual.

CONCLUSIONS

The current study provides a population-based analysis of 264 830 patients with COPD based on administrative health registries in Catalonia. The results illustrate the high impact of comorbidities on undesirable clinical events. We believe that the results highly encourage further developments fostering interoperability between health registries and electronic medical records to enhance clinical risk prediction.

Acknowledgements We want to acknowledge the support of NEXTCARE team (COMRDI15-1-0016), AGAUR research groups (2009SGR911 and 2014SGR661) and CERCA Programme/Generalitat de Catalunya.

Contributors Study conception and design: EV, AT, IC and JR. Data acquisition: EV, DM and MC. Data analysis: EV, DM and MC. Manuscript preparation: EV, AT, IC, CH, JE, AGA and JR. Manuscript revision: all authors.

Funding This work was supported by the European Commission grants CONNECARE (H2020-689802).

Competing interests None declared.

Patient consent Obtained.

Ethics approval Hospital Clinic of Barcelona.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. WHO. *Innovative care for chronic conditions: building blocks for action*. Geneva: World Health Organization (WHO/MNC/CCH/02.01), 2002. http://www.who.int/chp/knowledge/publications/iccglobalreport.pdf (accessed 9 Mar 2017).
2. Murray CJ, Lopez AD. Measuring the global burden of disease. *N Engl J Med* 2013;369:448–57.
3. Khakban A, Sin DD, FitzGerald JM, *et al*. The projected epidemic of COPD hospitalizations over the next 15 years: a population based perspective. *Am J Respir Crit Care Med* 2016:rccm.201606-1162PP.
4. McLean S, Hoogendoorn M, Hoogenveen RT, *et al*. Projecting the COPD population and costs in England and Scotland: 2011 to 2030. *Sci Rep* 2016;6:31893.
5. Vogelmeier CF, Criner GJ, Martinez FJ, *et al*. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am J Respir Crit Care Med* 2017;195:557–82.
6. Porter ME. What is value in health care? *N Engl J Med* 2010;363:2477–81.
7. Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med* 2016;374:504–6.
8. van Boven JFM. Costly comorbidities of COPD: the ignored side of the coin? *Eur Respir J* 2017;50:1700917.
9. Chen W, FitzGerald JM, Sin DD, *et al*. Excess economic burden of comorbidities in COPD: a 15-year population-based study. *Eur Respir J* 2017;50:1700393.
10. Vanfleteren LE, Spruit MA, Groenen M, *et al*. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013;187:728–35.
11. Divo MJ, Casanova C, Marin JM, *et al*. COPD comorbidities network. *Eur Respir J* 2015;46:640–50.
12. Vanfleteren LE, Spruit MA, Franssen FM. Tailoring the approach to multimorbidity in adults with respiratory disease: the NICE guideline. *Eur Respir J* 2017;49:1601696.
13. Barnes PJ. Mechanisms of development of multimorbidity in the elderly. *Eur Respir J* 2015;45:790–806.
14. Menche J, Sharma A, Kitsak M, *et al*. Disease networks. Uncovering disease–disease relationships through the incomplete interactome. *Science* 2015;347:1257601.
15. Dueñas-Espín I, Vela E, Pauws S, *et al*. Proposals for enhanced health risk assessment and stratification in an integrated care scenario. *BMJ Open* 2016;6:e010301.
16. Monterde D, Vela E, Clèries M. Los grupos de morbilidad ajustados: nuevo agrupador de morbilidad poblacional de utilidad en el ámbito de la atención primaria. *Atención Primaria* 2016;48:674–82.
17. Department of Health C. Catalonia health plan for 2016–2020. 2016. http://salutweb.gencat.cat/web/.content/home/el_departament/Pla_salut/pla_salut_2016_2020/Documents/Pla_salut_Catalunya_2016_2020.pdf
18. Farré N, Vela E, Clèries M, *et al*. Real world heart failure epidemiology and outcome: a population-based analysis of 88 195 patients. *PLoS One* 2017;12:e0172745.
19. Farré N, Vela E, Clèries M, *et al*. Medical resource use and expenditure in patients with chronic heart failure: a population-based analysis of 88 195 patients. *Eur J Heart Fail* 2016;18:1132–40.
20. Department of Health C. Programa públic d'analítica de dades per a la recerca i la innovació en salut (PADRIS). 2017. http://salutweb.gencat.cat/web/.content/home/ambits_tematics/linies_dactuacio/recerca/enllacos/Programa_analitica_dades_PADRIS_aquas2017_publica.pdf
21. Cherkin DC, Deyo RA, Volinn E, *et al*. Use of the International Classification of Diseases (ICD-9-CM) to identify hospitalizations for mechanical low back problems in administrative databases. *Spine* 1992;17:817–25.
22. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 2007;115:654–7.
23. Ware JH. The limitations of risk factors as prognostic tools. *N Engl J Med* 2006;355:2615–7.
24. Cano I, Tenyi A, Vela E, *et al*. Perspectives on big data applications of health information. *Curr Opin Syst Biol* 2017;3:36–42.
25. McWilliams JM, Schwartz AL. Focusing on high-cost patients—the key to addressing high costs? *N Engl J Med* 2017;376:807–9.
26. Lewis GH. 'Impactibility models': identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs. *Milbank Q* 2010;88:240–55.
27. Cano I, Dueñas-Espín I, Hernandez C, *et al*. Protocol for regional implementation of community-based collaborative management of complex chronic patients. *NPJ Prim Care Respir Med* 2017;27:44.
28. Escarrabill J, Torrente E, Esquinas C, *et al*. Auditoría clínica de los pacientes que ingresan en el hospital por agudización de EPOC. Estudio MAG-1. *Arch Bronconeumol* 2015;51:483–9.
29. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful optimization of large-scale health information technology. *J Am Med Inform Assoc* 2017;24:182–7.
30. Miravitlles M, Soriano JB, García-Río F, *et al*. Prevalence of COPD in Spain: impact of undiagnosed COPD on quality of life and daily life activities. *Thorax* 2009;64:863–8.
31. Gomez-Cabrero D, Menche J, Vargas C, *et al*. From comorbidities of chronic obstructive pulmonary disease to identification of shared molecular mechanisms by data integration. *BMC Bioinformatics* 2016;17:23–35.
32. Roca J, Vargas C, Cano I, *et al*. Chronic obstructive pulmonary disease heterogeneity: challenges for health risk assessment, stratification and management. *J Transl Med* 2014;12 (Suppl 2):S3.
33. Cano I, Lluch-Ariet M, Gomez-Cabrero D, *et al*. Biomedical research in a digital health framework. *J Transl Med* 2014;12 (Suppl 2):S10.
34. Cano I, Alonso A, Hernandez C, *et al*. An adaptive case management system to support integrated care services: lessons learned from the NEXES project. *J Biomed Inform* 2015;55:11–22.

MANUSCRIPT 5: Non-pulmonary manifestations of Chronic Obstructive Pulmonary Disease: mechanisms, risk assessment and clinical management

Cano I, Gomez-Cabrero D, **Tényi Á**, Jesper Tegner, Wagner P, Maier D, Miralles F, Cascante M, and Roca J. Non-pulmonary manifestations of Chronic Obstructive Pulmonary Disease: mechanisms, risk assessment and clinical management. *Submitted to Respiratory Research – April 2018*.

- Journal: *Respiratory Research*
- Impact factor: *3.841*
- Quartile: *1st*
- Status: *Submitted to Respiratory Research – April 2018*.

A systems approach to non-pulmonary manifestations of COPD

Isaac Cano[1,2], David Gomez-Cabrero[3]*, Akos Tenyi[1,2], Jesper Tegner[3,4], Peter Wagner[5], Dieter Maier[6], Felip Miralles[7], Marta Cascante[8], and Josep Roca[1,2]*

*The two authors contributed equally

[1] Hospital Clínic. IDIBAPS, Facultat de Medicina. Universitat de Barcelona, 08036, Barcelona, Catalunya, Spain.

[2] Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Spain.

[3] Unit of Computational Medicine, Department of Medicine, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

[4] Biological and Environmental Sciences and Engineering Division (BESE), Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955–6900, Kingdom of Saudi Arabia

[5] Division of Physiology, Pulmonary and Critical Care Medicine, University of California, San Diego, La Jolla, California.

[6] Biomax Informatics AG, Robert-Koch-Str. 2, Planegg, Germany.

[7] Technology Centre of Catalonia (Eurecat). Barcelona, Catalonia. Spain.

[8] Departament de Bioquimica i Biologia Molecular i IBUB, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Catalunya, Spain

Correspondence

Josep Roca (jroca@clinic.cat). Hospital Clínic, IDIBAPS, Universitat de Barcelona, C/Villarroel 170, 08036, Barcelona, Spain. Phone: +34-932275747, Fax: +34-932275455.

Article type: Review

Abstract

Background: The Synergy-COPD project was conceived as a systems medicine approach to study underlying biological mechanisms of skeletal muscle dysfunction and the phenomenon of co-morbidity clustering observed in patients suffering from chronic obstructive pulmonary disease (COPD). The overarching hypothesis was that non-pulmonary manifestations cannot solely be explained by the activity of the pulmonary disease. This paper summarizes the biomedical outcomes of the project.

Main body: Synergy-COPD identified abnormalities in co-regulation of bioenergetics, inflammation and tissue remodelling processes, operating as central players in non-pulmonary manifestations, and a relevant role for oxidative stress as a key characteristic mechanism in these patients. The findings showed significant associations with aerobic capacity, but not with lung function. In addition, a data-driven analysis of the Medicare dataset indicated higher risk for co-morbidities in patients with COPD. Moreover, a population-health risk assessment of COPD cases in Catalonia (Spain) suggested a high predictive role of co-morbidities in terms of mortality, hospitalizations, multiple hospital admissions, and high healthcare costs.

Conclusions: These findings on mechanisms of non-pulmonary phenomena and co-morbidities, indicate the need for novel risk assessment strategies. Synergy-COPD outcomes strongly points out that current standards for clinical management should be complemented by a patient-oriented approach considering enhanced comorbidity prevention and management.

Key words: Bioenergetics; Multimorbidity; Predictive Medicine; Redox disequilibrium; Systems Medicine.

Background

Patients with chronic obstructive pulmonary disease (COPD) produce a huge health and societal burden worldwide [1–4], which unfortunately is expected to increase in the coming years mainly due to population ageing [5, 6].

While acknowledging the progress achieved in terms of standard of care recommendations addressing COPD [2], a deeper knowledge of non-pulmonary manifestations [7–9], as well as their implications for clinical care, are well-recognized unmet needs.

The EU project Synergy-COPD [10] was formulated on the basis of the hypothesis that a better insight into the biological mechanisms involved in non-pulmonary manifestations [11–13] may contribute to improved care of patients suffering from COPD. The project (**Figure 1**) was therefore conceived as a systems medicine analysis of unknown aspects of two specific non-pulmonary manifestations: skeletal muscle dysfunction/wasting [7] and the phenomenon of co-morbidity clustering [13]. It is of note that the latter seems to be only partly explained by shared risk factors among concurrent diseases [8, 13], such as tobacco smoking, nutritional disbalance and sedentarism.

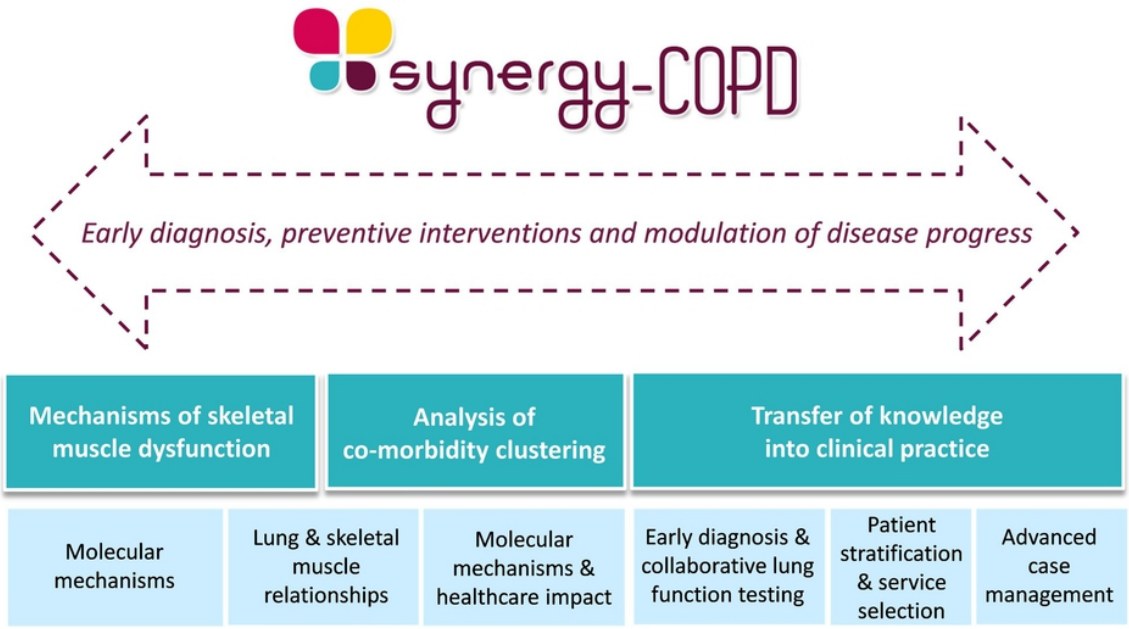


Figure 1 – Main biomedical areas of the EU project Synergy-COPD. The project analysed experimental, "omics", physiological, clinical and epidemiological information addressing underlying mechanisms of two major non-pulmonary phenomena often seen in patients with COPD: Skeletal muscle dysfunction and Co-morbidities. A third area of the project was to design and evaluate strategies for transferring novel knowledge into

clinical practice with a patient-oriented approach. The ultimate aim was to explore strategies leading to modulation of disease progress and cost-effective management of these patients.

Synergy-COPD was designed as an iterative research process wherein data from several sources, encompassing animal experimentation [14], human studies [15–17], epidemiological research [18] and registry information [19, 20], were analysed using several, and in some cases complementary, computational modelling techniques.

Besides its core biomedical dimension, the project had two additional objectives: i) technological developments facilitating generation and integration of knowledge; and, ii) transfer of acquired knowledge into clinical settings.

Whereas the details of the project design, methodological issues and initial results of Synergy-COPD have been previously described in a dedicated monograph [21], the current paper summarizes consolidated outcomes for each of the three biomedical areas of the project (**Figure 1**): i) Skeletal muscle dysfunction; ii) Co-morbidities; and, iii) Proposals for enhanced transfer of knowledge into clinical practice.

The report also presents limitations, and facilitators, encountered during the project lifespan with impact on future developments and action plans. It is of note, that the current manuscript does not specifically address the technological aspects of Synergy-COPD.

Skeletal muscle dysfunction

Skeletal muscle dysfunction is a well-accepted systemic effect of COPD associated with poor prognosis and high use of healthcare resources, independently of the degree of lung function impairment (FEV_1) [2, 7]. The phenomenon has a multifactorial basis and it may affect up to 30% of patients [22] with a wide spectrum of manifestations, from subclinical findings to overt muscle wasting, defined as a low Fat Free Mass Index (FFMI<16kg/m^2 in men and FFMI<15kg/m^2 in women) [23]. Synergy-COPD performed six studies [14, 24–28] addressing two specific aspects of skeletal muscle dysfunction, namely: i) Molecular mechanisms [14, 24, 28]; and, ii) Lung-limb muscle relationships associated to oxygen transport [25–27].

Molecular mechanisms – The transcriptionally active network modules of interacting proteins in the vastus lateralis of patients with COPD (n= 15, FEV1 46 ± 12 % predicted, age 68 ± 7 yrs.) and healthy sedentary controls (n= 12, age 65 ±9 yrs.) [28] were characterized for each differential condition, at rest and after an 8-week endurance training program, using the HotNet2 algorithm [29]. Results were evaluated

with previous experimental multilevel data derived from the same study groups [15–17], not used in the analysis. At baseline, the study identified four COPD specific network modules indicating abnormalities in creatinine metabolism, calcium homeostasis, oxidative stress and inflammatory responses, showing statistically significant ($p<0.05$) associations with exercise capacity (VO_2 peak, BODE index and blood lactate levels), but not with lung function (FEV_1). It is of note that endurance training-induced effects, assessed through changes in the network modules, displayed marked differences between COPD patients and controls (**Figure 2**). In healthy subjects, skeletal muscle training adaptations were significantly associated with changes in cell bioenergetics ($p<0.05$) which, in turn, showed strong relationships with plasma metabolomics adaptations; whereas, effects of training in COPD were confined to muscle remodelling with abnormal inflammatory changes.

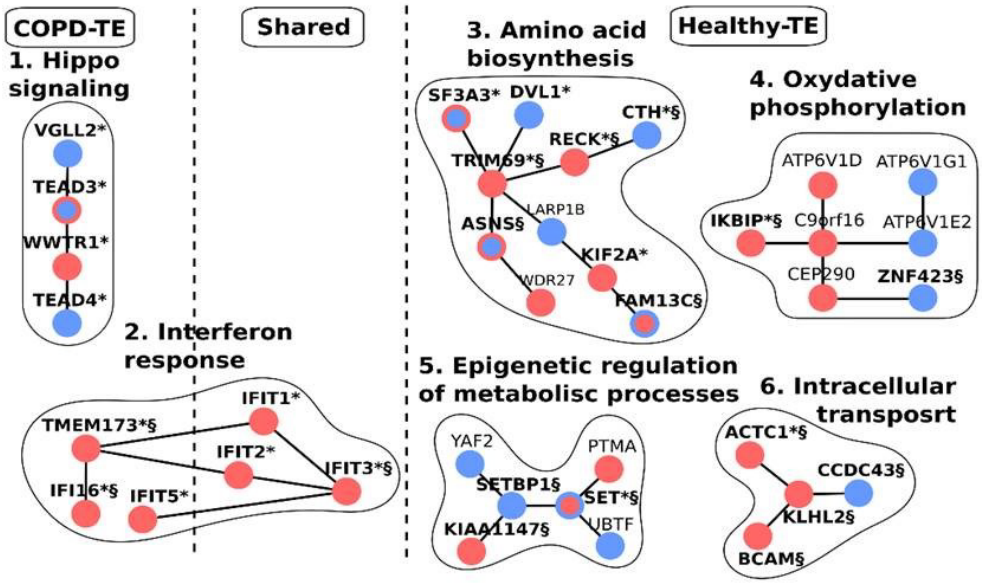


Figure 2 – Main Training effects (TE) on network modules. The figure depicts network modules that were found to be modified by the 8-week training program. The figure depicts on the left, network modules identified in patients with COPD (COPD-TE); on the right, network modules identified in healthy controls (Healthy-TE); in the middle, (sub) modules that were identified in both groups (Shared). Genes are colored according to their differential regulation in COPD-TE (inner colour of the nodes) and in Healthy-TE (border colour of the nodes): up regulation with training (red circles), down regulation with training (blue circles). Modules are named after significantly enriched Gene Ontology terms. Training differential expression significance is signed by * for COPD-TE, and § for Healthy-TE (False Discovery Rate < 0.05) [28].

The defective muscle adaptation to training in patients with COPD [24] was also assessed using a novel computational approach combining the benefits of probabilistic and classical model-driven methods (i.e., Thomas networks) [30] to address the complex dynamics of energy-metabolism-associated gene regulatory networks. Briefly, alterations in tricarboxylic acid (TCA) cycle, electron transport chain, creatine kinase, cytokines regulation and insulin receptor factor were unveiled as key players in the abnormal training-induced responses in patients with COPD. Moreover, the study [24] suggested new avenues to explore drug repurposing in subsets of patients with COPD.

Synergy-COPD also showed that guinea pigs exposed to long-term cigarette smoking accurately reflected most of the transcriptional changes observed in dysfunctional limb muscle of patients with COPD when compared to matched controls [14]. Using network inference techniques, the study indicated that the expression profile in whole lung of genes encoding for soluble inflammatory mediators is informative of the molecular state of skeletal muscles in the guinea pig smoking model. These findings suggested abnormal interactions between lung and skeletal muscle that deserve to be further explored. It was shown that two cytokines [CXCL10 and CXCL9, Chemokine (C-X-C motif) ligand 9 and ligand 10, respectively] are promising candidate inflammatory signals linked to the regulation of central metabolism genes in skeletal muscles. These two cytokines had been reported as biomarkers of abnormal cardiac remodelling [31, 32].

Dissociation between lung function and limb muscle bioenergetics – Synergy-COPD developed one integrated mechanistic mathematical model to predict intracellular oxygen levels [25, 26] which was linked to mitochondrial metabolism [33], to estimate mitochondrial reactive oxygen species (ROS) production [27]. A sensitivity analysis of the relationships between cell oxygenation and mitochondrial ROS production was carried out [34] in a group of 21 patients with mild to severe COPD (GOLD: II 33%; III 43%, and, IV 24%) [35] exercising at peak aerobic capacity (VO_2 peak). Two central messages emerged from this analysis: i) Both estimated tissue oxygenation levels and VO_2 peak were unrelated to measured FEV_1, and ii) Low intracellular oxygen levels, stimulating abnormally high mitochondrial ROS production, were predicted to occur at peak exercise in these patients [27, 36]. The latter is consistent with the deleterious effects of high intensity training on skeletal muscle performance observed in severe COPD patients [37].

Lessons learned and clinical perspectives – Anomalous gene regulatory network dynamics in skeletal muscle of patients with COPD constituted the most striking findings of the above studies. Skeletal muscle abnormalities leading to less efficient

energy metabolism, and potentially driving other abnormal skeletal muscle responses were common characteristics. In our view, the consistent relationship between these results and aerobic capacity makes it a priority to characterize each patient's exercise performance and physical activity in the clinic. The results of Synergy-COPD endorse the need for optimizing early cardiopulmonary rehabilitation strategies, and promote physicial activity, as a way to modulate prognosis in these patients [38]. Moreover, the results foster the need for novel longitudinal studies with multilevel measurements to further refine operational strategies for prevention of skeletal muscle dysfunction and for a better understanding of the abnormal interplay between lung and skeletal muscle in patients with COPD.

Co-morbidities

The high impact of co-morbidities on healthcare burden prompts the need for revisiting current strategies for management of multimorbidity in these patients [9, 39, 40]. Whereas the current recommendations suggest to deal with each co-morbidity as a separate condition [2], different studies have identified some recurring co-morbidity patterns in patients with COPD [8, 41]. However, there are several key unanswered questions, namely: i) Is co-morbidity prevalence in patients with COPD fully explained by well identified risk factors, especially tobacco smoking and air pollution?; ii) Can abnormalities in gene-regulatory pathway dynamics be shared by co-morbid conditions explaining the common patterns of co-morbidities?; iii) Can susceptibility for development of co-morbidities be early identified and prevented?; and, iv) Should the importance of co-morbidities in patients with COPD trigger novel interventional approaches both at population and at individual levels?. Synergy-COPD partly addressed these issues through two different data driven analyses [19, 20], as briefly described below.

Co-morbidity risk and shared molecular mechanisms – A data driven analysis of 13 million people from the Medicare dataset [19] identified higher risk for co-morbidity clustering in patients with COPD, for all age windows explored, when compared to patients without COPD (**Figure 3**). It is of note that a similar analysis carried out with the Catalan health surveillance dataset [20] confirm these findings. Gomez-Cabrero et. al. [19] also identified a dysregulated set of genes shared by common co-morbid conditions (e.g., IL5, TNF, JUP and some genes of the HLA family), as previously reported in [8, 42].

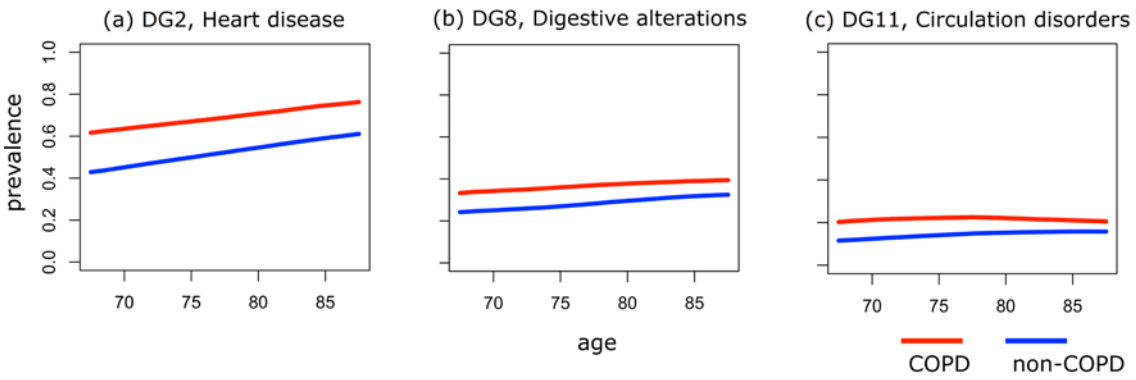


Figure 3 – Prevalence of selected disease groups (DG), based on ICD9 coding, over age for COPD and non-COPD individuals from the Medicare dataset, as previously published in [19]. Each panel shows, for a given DG, the prevalence of the DG in non-COPD (blue) and in COPD (red) individuals using a 0 to 1 scale. DGs 2 (heart diseases), 8 (digestive alterations) and 11 (circulation disorders) are shown in panels (a), (b) and (c), respectively. In all cases the prevalence was higher in COPD patients. However, the effect of age was different among groups [19].

Population-health analyses of co-morbidities – The project also explored the impact of co-morbidities in patients with COPD [20] through a population-health analysis, which accounted for all registered cases with diagnosis of COPD (n=264,830) in Catalonia (7.5 million inhabitants) using a fixed cohort design (**Figure 4**). Main study findings were the identification of co-morbidities, aggregated with an adjusted morbidity grouper (*AMG - Adjusted Morbidity Groups*) [44], as the covariate with highest discriminative impact on target events such as mortality, hospitalizations (all causes and COPD-related), multiple admissions (related and not related to COPD) and on annual healthcare costs per patient. The results may suggest the need for designing novel strategies for management of COPD patients aiming at containing the increasing healthcare burden of these patients, a well-identified unmet need [5, 6, 9, 39]. A highly relevant information of the study was that less than 25% of unplanned hospitalizations in patients with COPD were registered as associated with pulmonary events which may have important implications in the management of these patients. In the study, AMG-based predictive modelling derived from the same study dataset showed an acceptable predictive capacity [area under the Receiver Operating Characteristic (ROC) curve of 0.83 (mortality), 0.77 (all causes of hospitalizations), 0.81 (COPD-related hospitalizations), 0.80 (all causes of multiple hospitalizations), 0.87 (COPD-related multiple hospitalizations) and 0.76 (users with high healthcare costs)]. These results generate two relevant messages. Firstly, the high impact of co-morbidities (AMG grading) on clinical events in patients with COPD which strengthen the need for novel

strategies for both prevention and enhanced management of co-morbid conditions in these patients. Secondly, the results indicate the potential for exploring synergies among population-health analyses, clinical information and information on underlying biological mechanisms of COPD to enhance health risk assessment and stratification in the clinical arena (**Figure 4**) [43].

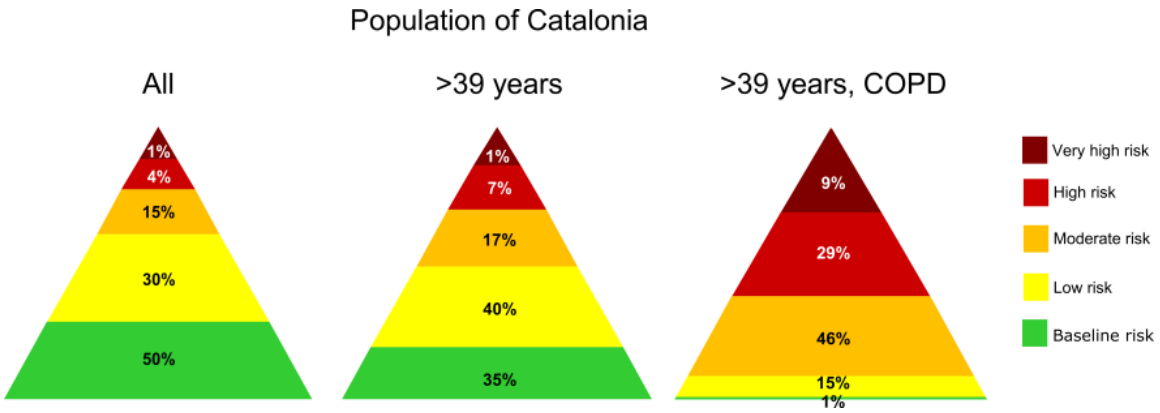


Figure 4 – Population-based health risk stratification pyramid in Catalonia previously published in [20]. The left triangle depicts the distribution of all citizens in the region (7.5M people), expressed as percentage, in five arbitrary health risk layers defined using the AMG (Adjusted Morbidity Groups) [20, 43, 44] as a population-based health risk stratification tool. The thresholds defining the five health risk layers in all three triangles correspond to the percentiles 50, 80, 95 and 99 of the AMG grading calculated for the general population (left triangle).The central triangle indicates the distribution in these five health risk layers of the subset of citizens older than 39 years. The right triangle displays the distribution of the study group of all cases with COPD in the region (264,830 patients) across the AMG health risk grades [20].

Lessons learned and clinical perspectives – Synergy-COPD results indicate that longitudinal analyses of co-morbidity clustering should constitute a priority area for biomedical research with important implications on patient management. It should ultimately aim for better understanding COPD heterogeneities to enable early identification of patients susceptible for the development of abnormal dynamics of key gene regulatory networks. To this end, recent analyses of disease trajectories [45–47] are prompting a highly attractive research scenario leading to a better understanding of co-morbidity clustering observed both in cross-sectional analyses [19, 20] and in longitudinal studies [45]. The rationale is that known risk factors would generate target co-morbid conditions especially in those patients with COPD that may show early indicators of abnormal gene regulatory network dynamics [14, 24]. Synergy-COPD seems to provide appropriate grounds for further research aimed at identifying such early indicators (**Figure 5**). Moreover, the holistic approach used in the project may

shed light on related issues such as identification of shared mechanisms to explain increased rate of lung cancer in these patients, and possible links between skeletal muscle dysfunction and co-morbidity clustering. It is of note that recent advances in targeted plasma metabolomics analyses may provide an operational approach for early identification of candidate patients for preventive interventions [48, 49].

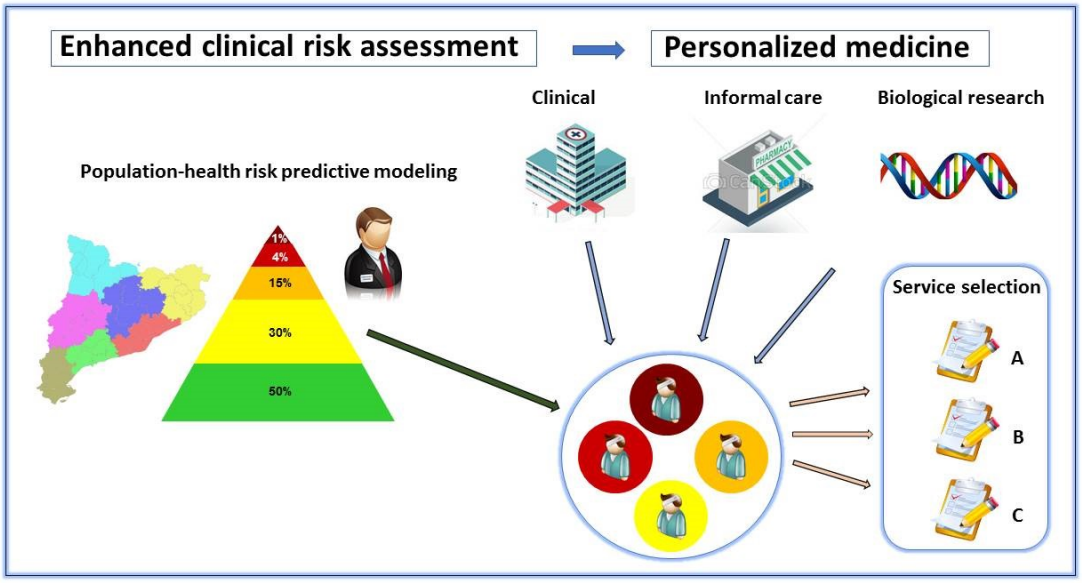


Figure 5 – A holistic approach of health risk assessment involves inclusion of covariates from multilevel domains, namely: i) Clinical, ii) Informal care; iii) Biological research; and, iv) Outcomes from population-health risk predictive modelling, which includes all the citizens in a given geographical area and results in enhanced patient-based stratification and optimization of service selection. Left-hand-side of the figure displays the Catalan risk stratification-pyramid in 2014 [20, 43, 44]. Synergy-COPD hypothesizes synergies between population-health predictive modelling and clinical risk assessment that should pave the way towards personalized medicine for patients with chronic diseases, as explained in [43].

Transfer into clinical practice

As alluded to above, a core component of Synergy-COPD was the transfer of the acquired knowledge on non-pulmonary phenomena seen in patients with COPD into the clinical arena with a twofold purpose. Firstly, to enhance individual health-risk assessment and service selection, leading to innovative management strategies for these patients. A second general aim was to explore and identify novel organizational and technological settings to facilitate generation of medical evidence through enhanced interplay between healthcare and biomedical research [50, 51].

Accordingly, the project promoted convergence of Synergy-COPD outcomes with specific ongoing initiatives contributing to large scale deployment of care coordination for chronic patients in Catalonia [52], namely: i) Early diagnosis and collaborative lung function testing; ii) Enhanced health risk assessment and service selection; and, iii) Management of cases with advanced disease.

Early diagnosis and collaborative lung function testing – Previous studies have indicated that collaborative forced spirometry (FS) testing between specialized and primary care may generate significant healthcare efficiencies and provide valuable information on longitudinal changes of lung function either spontaneously or due to therapeutic interventions. The project hypothesized that the approach may overcome historical limitations for extensive use of forced spirometry in primary care, due to suboptimal quality of testing while generating cost-effectiveness [53].

All the above elements were supported by Synergy-COPD outcomes and substantiated the pivotal components of the collaborative FS program [54] across healthcare tiers being currently deployed in Catalonia. The three principal expected outcomes of the program are: i) Early diagnosis; ii) Enhanced quality and accessibility of FS testing across the health system; and, iii) Availability of longitudinal FS testing information for individual risk health risk assessment.

Enhanced patient stratification and service selection – Synergy-COPD outcomes on non-pulmonary phenomena are clearly opening exciting new avenues mainly addressing secondary prevention in patients in low to high layers of the health-risk pyramid (**Figure 4**), in whom there is still potential for modulating prognosis. It is of note that the statement refers to health risk assessment of patients with COPD, not to severity, and/or activity, of the pulmonary disease. This is because the latter is only one of the elements influencing patient-based health risk assessment.

Patients with COPD showing low to high health-risk require management of the pulmonary disease following standards of care recommendations [2]. However, the results of the project clearly indicate the need for preventive interventions on actionable factors determining non-pulmonary phenomena in order to 1) effectively modulate patients' natural history and improve prognosis, which are recognized as unmet needs for enhanced COPD management; and 2) reduce health care costs.

Specifically, the project points out on two main directions in terms of early diagnosis, monitoring and action. Firstly, detection and prevention of abnormalities in skeletal muscle bioenergetics. In this regard, promotion of daily physical activity appears as a relevant intervention in these patients [55–57]. A second major area for action is early

identification of susceptibility to develop common co-morbid conditions, as well as enhanced management of patients with co-morbidities through patient-oriented strategies rather than to current disease-oriented approaches. However, full development and validation of suitable non-pharmacological and pharmacological interventions on non-pulmonary phenomena in these patients will require further longitudinal research on shared underlying mechanisms.

Synergy-COPD makes a strong argument for the high potential of population-health risk assessment (**Figures 4** and **5**) in contributing to enhance clincal health risk assessment for decision making, as proposed in [43].

Advanced case management – The project did not focus on management of patients close to the tip of the health-risk stratification pyramid (**Figure 4**). However, a logical priority is the achievement of cost-effective management of high and very high health-risk cases, which imposes a strong need for care coordination across healthcare tiers (i.e., integrated care) towards: i) Prevention of severe exacerbations leading to hospital admissions, both related and not related to COPD; ii) Enhanced resilience and health-related quality of life; and, iii) Increased survival.

We acknowledge, however, two main limiting factors that should be solved in order to achieve adoption of integrated care services in patients with COPD [58]. Firstly, poor comparability among interventions assessing effects of integrated care on patients with COPD indicates an urgent need for standardization of service workflows able to be adapted to the evolving conditions of patients. It is compulsory in order to generate evidence, as well as to facilitate transferability of outcomes among sites [2]. To this end, assessment methodologies based on implementation science are clearly needed [59–61]. A second challenge is to overcome the limited healthcare impact of usual interventions addressed to patients that are high consumers of healthcare resources [62]. That is, low ratio between the magnitude of intervention benefits and the resources devoted to achieve them, which may imply little healthcare value generation [63, 64]. These two factors strengthen the need for further evidence on cost-effectiveness of well-defined integrated care interventions for complex chronic patients [65].

It is acknowledged that the messages raised from the population-health risk assessment analysis were originated only from registry information. However, because of its clinical relevance, they clearly deserve further research and actions in this direction.

Conclusions

The Synergy-COPD project showed the potential of systems medicine to generate knowledge on underlying mechanisms of non-pulmonary effects in patients with COPD with impact on clinical practice. The project results indicate the need for novel therapeutic and care coordination strategies that should be validated in future longitudinal studies carried out in real-world settings.

The biomedical results from Synergy-COPD are opening new avenues to better understand the interplay of factors modulating non-pulmonary manifestations in patients with COPD. Abnormalities in co-regulation of core biological pathways (i.e., bioenergetics, inflammation and tissue remodelling) at systemic level seem to play a central role on both skeletal muscle dysfunction and co-morbidity clustering [66], with evidence of the relevant role of oxidative stress as a characteristic mechanism in these patients [13]. All in all, Synergy-COPD strongly points out the need for a broader vision in the care and management of COPD by adopting a patient-oriented approach that addresses much more than just the pulmonary manifestations of the disease.

Further research is still needed to identify potential causal factors of non-pulmonary manifestations, namely: i) in-born genetic susceptibility; ii) epigenetic changes associated with unhealthy lifestyles and/or to activity/severity of pulmonary disease; and, iii) unknown interactions with gut microbiome, among others. Accordingly, a better understanding of the interplay between pulmonary disease and systemic alterations in these patients also constitutes an unmet need. Likewise, identification of metabolomics patterns facilitating early identification of subsets of patients with COPD that are candidates for secondary prevention of non-pulmonary manifestations would also be a major achievement. It is hoped that plasma samples would be sufficient for this [67].

We propose that both the design and evaluation of novel non-pharmacological and pharmacological preventive interventions in COPD patients will require well-designed longitudinal studies using an integrated, multidisciplinary, systems approach. Ultimately, multilevel integrative analyses of registry data, biomedical research information, electronic medical records and informal care data seems to constitute a high priority to properly pave the way toward enhanced clinical management and personalized medicine for patients with chronic disorders (**Figure 5**) [68–70].

One of the major strengths of the Synergy-COPD design was the combination of well-defined biomedical goals with parallel technological developments beyond the current state of the art in terms of novel modelling approaches, knowledge generation tools and information and communication technologies supporting care coordination. As

described in [21, 71], conceptualization of a digital health framework and formulation of the roadmap for its ongoing deployment, together with technologies supporting systematic collection of different types of data over time, for the benefit of clinical and biomedical research are urgently needed [72–74]. Overall, the results indicate that convergence between a systems approach to COPD and care coordination may conform an optimal scenario to foster cross-fertilization between biomedical research and clinical practice [50, 75].

List of abbreviations

Adjusted Morbidity Groups (AMG)

chronic obstructive pulmonary disease (COPD)

Forced expiratory volume during the first second (FEV1)

Fat Free Mass Index (FFMI)

forced spirometry (FS)

Receiver Operating Characteristic (ROC)

reactive oxygen species (ROS)

tricarboxylic acid (TCA)

Declarations

Authors' contributions

Conception and design: IC, DGC, AT,JT, PW, DM, FM, MC, JR; Analysis and interpretation: IC, DGC, AT, PW, MC, JR; Drafting the work or revising it critically for final approval: IC, DGC, AT,JT, PW, DM, FM, MC, JR;

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations and were approved by the Ethical Committee at Hospital Clinic and written informed consent was signed by each participant.

Consent for publication

Not applicable

Availability of data and materials

The microarray dataset(s) supporting the conclusions of this article is(are) available in the GEO repository, GSE27536,

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE27536.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the grants CONNECARE (H2020-689802).

Acknowledgments

We want to acknowledge the support of Synergy-COPD (FP7-ICT-2009-270086), NEXTCARE (COMRDI15-1-0016), PITES-TliSS (PI15/00576), AGAUR research groups (2009SGR911 and 2014SGR661), and CERCA Programme / Generalitat de Catalunya.

References

1. Murray CJL, Lopez AD. Measuring the Global Burden of Disease. *N Engl J Med* 2013; 369: 448–457.

2. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, Chen R, Decramer M, Fabbri LM, Frith P, Halpin DMG, López Varela MV, Nishimura M, Roche N, Rodriguez-Roisin R, Sin DD, Singh D, Stockley R, Vestbo J, Wedzicha JA, Agustí A. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am J Respir Crit Care Med* American Thoracic Society; 2017; 195: 557–582.

3. WHO. 2008-2013 Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases. World Health Organization 2008. http://www.who.int/nmh/Actionplan-PC-NCD-2008.pdf. Date last updated: Apr 18 2000. Date last accessed: Nov 14 2017. .

4. WHO. Global Status Report on Noncommunicable Diseases 2014. World Health Organization 2014. http://apps.who.int/medicinedocs/es/m/abstract/Js21756en/. Date last updated: Jan 19 2015. Date last accessed: Nov 14 2017. .

5. Khakban A, Sin DD, FitzGerald JM, McManus BM, Ng R, Hollander Z, Sadatsafavi M. The Projected Epidemic of Chronic Obstructive Pulmonary Disease Hospitalizations over the Next 15 Years. A Population-based Perspective. *Am J Respir Crit Care Med* United States: American Thoracic Society; 2017; 195: 287–291.

6. McLean S, Hoogendoorn M, Hoogenveen RT, Feenstra TL, Wild S, Simpson CR, Mölken MR, Sheikh A. Projecting the COPD population and costs in England and Scotland: 2011 to 2030. *Sci Rep* 2016; 6: 31893.

7. Maltais F, Decramer M, Casaburi R, Barreiro E, Burelle Y, Debigare R, Dekhuijzen PN, Franssen F, Gayan-Ramirez G, Gea J, Gosker HR, Gosselink R, Hayot M, Hussain SN, Janssens W, Polkey MI, Roca J, Saey D, Schols AM, Spruit MA, Steiner M, Taivassalo T, Troosters T, Vogiatzis I, Wagner PD. An official american thoracic society/european respiratory society statement: update on limb muscle dysfunction in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2014; 189: e15-62.

8. Vanfleteren LE, Spruit MA, Groenen M, Gaffron S, van Empel VP, Bruijnzeel PL, Rutten EP, Op 't Roodt J, Wouters EF, Franssen FM. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013/02/09. 2013; 187: 728–735.

9. Vanfleteren LE, Spruit MA, Franssen FM. Tailoring the approach to multimorbidity in adults with respiratory disease: the NICE guideline. *Eur Respir J* 2017; 49: 1601696.

10. Synergy-COPD. Modelling and simulation environment for systems medicine: Chronic obstructive pulmonary disease (COPD) as a use case, FP7-ICT-270086. 2010. http://www.synergy-copd.eu/. Date last updated: May 23 2013. Date last accessed: Nov 14 2017. .

11. Dustin ML. Signaling at neuro/immune synapses. *J Clin Invest* 2012; 122: 1149–1155.

12. Poole DC, Hirai DM, Copp SW, Musch TI. Muscle oxygen transport and utilization in heart failure: implications for exercise (in)tolerance. *Am J Physiol Hear Circ Physiol* 2012; 302: H1050-63.

13. Barnes PJ. Mechanisms of development of multimorbidity in the elderly. *Eur Respir J* 2015; 45: 790–806.

14. Davidsen PK, Herbert JM, Antczak P, Clarke K, Ferrer E, Peinado VI, Gonzalez C, Roca J, Egginton S, Barberá JA, Falciani F. A systems biology approach reveals a link between systemic cytokines and skeletal muscle energy metabolism in a rodent smoking model and human COPD. *Genome Med* 2014; 6: 59.

15. Rodríguez D, Alcarraz-Vizán G, Díaz-Moralli S, Reed M, Gómez F, Falciani F, Günther U, Roca J, Cascante M. Plasma metabolic profile in COPD patients: effects of exercise and

endurance training. *Metabolomics* Springer US; 2012; 8: 508–516.

16. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, Curnow SJ, Rodriguez DA, Cascante M, O'Neill L, Egginton S, Roca J, Falciani F. A Systems Biology Approach Identifies Molecular Networks Defining Skeletal Muscle Abnormalities in Chronic Obstructive Pulmonary Disease. *PLoS Comput Biol* Public Library of Science; 2011; 7: e1002129.

17. Rodriguez DA, Kalko S, Puig-Vilanova E, Perez-Olabarría M, Falciani F, Gea J, Cascante M, Barreiro E, Roca J. Muscle and blood redox status after exercise training in severe COPD patients. *Free Radic Biol Med* BioMed Central; 2012; 52: 88–94.

18. Garcia-Aymerich J, Gomez FP, Benet M, Farrero E, Basagana X, Gayete A, Pare C, Freixa X, Ferrer J, Ferrer A, Roca J, Galdiz JB, Sauleda J, Monso E, Gea J, Barbera JA, Agusti A, Anto JM. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. *Thorax* 2011; 66: 430–437.

19. Gomez-Cabrero D, Menche J, Vargas C, Cano I, Maier D, Barabási A-L, Tegnér J, Roca J. From comorbidities of chronic obstructive pulmonary disease to identification of shared molecular mechanisms by data integration. *BMC Bioinformatics* BioMed Central; 2016; 17: 441.

20. Vela E, Tényi Á, Cano I, Monterde D, Clèries M, Garcia-Altes A, Hernandez C, Escarrabill J, Roca J, Cleries M, Garcia-Altes A, Hernandez C, Escarrabill J, Roca J. Population-based analysis of COPD patients in Catalonia: implications for case management. *BMJ Open* British Medical Journal Publishing Group; 2018; 8: e017283.

21. Gomez-Cabrero D, Lluch-Ariet M, Tegnér J, Cascante M, Miralles F, Roca J, Synergy-COPD consortium. Synergy-COPD: a systems approach for understanding and managing chronic diseases. *J Transl Med* 2014; 12: S2.

22. Gea J, Pascual S, Casadevall C, Orozco-Levi M, Barreiro E. Muscle dysfunction in chronic obstructive pulmonary disease: update on causes and biological findings. *J Thorac Dis* 2015; 7: E418–E438.

23. Vestbo J, Prescott E, Almdal T, Dahl M, Nordestgaard BG, Andersen T, Sørensen TIA, Lange P. Body Mass, Fat-Free Body Mass, and Prognosis in Patients with Chronic Obstructive Pulmonary Disease from a Random Population Sample. *Am J Respir Crit Care Med* 2006; 173: 79–83.

24. Marín de Mas I, Fanchon E, Papp B, Kalko S, Roca J, Cascante M. Molecular mechanisms underlying COPD-muscle dysfunction unveiled through a systems medicine approach. *Bioinformatics* 2017; 33: 95–103.

25. Cano I, Mickael M, Gomez-Cabrero D, Tegnér J, Roca J, Wagner PD. Importance of mitochondrial PO2 in maximal O2 transport and utilization: A theoretical analysis. *Respir Physiol Neurobiol* 2013; 189: 477–483.

26. Cano I, Roca J, Wagner PD. Effects of lung ventilation-perfusion and muscle metabolism-perfusion heterogeneities on maximal O2 transport and utilization. *J Physiol* 2015; 593: 1841–1856.

27. Cano I, Selivanov V, Gomez-Cabrero D, Tegnér J, Roca J, Wagner PD, Cascante M. Oxygen Pathway Modeling Estimates High Reactive Oxygen Species Production above the Highest Permanent Human Habitation. West J, editor. *PLoS One* Public Library of Science; 2014; 9: e111068.

28. Tényi Á, Cano I, Marabita F, Kiani N, Kalko SG, Barreiro E, De Atauri P, Cascante M, Gomez-Cabrero D, Roca J, Tényi A, Cano I, Marabita F, Kiani N, Kalko SG, Barreiro E, De Atauri P, Cascante M, Gomez-Cabrero D, Roca D. Network modules uncover mechanisms of skeletal muscle dysfunction in COPD patients. *J Trans Med* 2018; 16: 34.

29. Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge J V, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, Lawrence MS, Gonzalez-Perez A, Tamborero D, Cheng Y, Ryslik GA, Lopez-Bigas N, Getz G, Ding L, Raphael BJ. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet* 2015; 47: 106–114.

30. Thomas R, Kaufman M. Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multistationarity and other nontrivial behavior. *Chaos* 2001; 11: 170–

179.

31. Altara R, Gu Y-M, Struijker-Boudier HAJ, Thijs L, Staessen JA, Blankesteijn WM. Left Ventricular Dysfunction and CXCR3 Ligands in Hypertension: From Animal Experiments to a Population-Based Pilot Study. Monleon D, editor. *PLoS One* 2015; 10: e0141394.

32. Altara R, Manca M, Hessel MH, Gu Y, van Vark LC, Akkerhuis KM, Staessen JA, Struijker-Boudier HAJ, Booz GW, Blankesteijn WM. CXCL10 Is a Circulating Inflammatory Marker in Patients with Advanced Heart Failure: a Pilot Study. *J Cardiovasc Transl Res* 2016; 9: 302–314.

33. Selivanov VA, Votyakova T V, Pivtoraiko VN, Zeak J, Sukhomlin T, Trucco M, Roca J, Cascante M. Reactive Oxygen Species Production by Forward and Reverse Electron Fluxes in the Mitochondrial Respiratory Chain. *PLoS Comput Biol* Public Library of Science; 2011; 7: e1001115.

34. Cano I. Predictive Medicine for Chronic Patients in an Integrated Care Scenario. Chronic Obstructive Pulmonary Disease as Use Case. 2014. http://hdl.handle.net/10803/283315. Date last updated: Sept 03 2014. Date last accessed: Nov 14 2017. Universitat de Barcelona;

35. Blanco I, Gimeno E, Munoz PA, Pizarro S, Gistau C, Rodriguez-Roisin R, Roca J, Barberà JA. Hemodynamic and Gas Exchange Effects of Sildenafil in Patients with Chronic Obstructive Pulmonary Disease and Pulmonary Hypertension. *Am J Respir Crit Care Med* 2010; 181: 270–278.

36. Roca J, Vargas C, Cano I, Selivanov V, Barreiro E, Maier D, Falciani F, Wagner PD, Cascante M, Garcia-Aymerich J, Kalko S, Marin I, Tegner J, Escarrabill J, Agustí A, Gomez-Cabrero D, Synergy-COPD consortium. Chronic Obstructive Pulmonary Disease Heterogeneity. Challenges for Health Risk Assessment, Stratification and Management. *BMC J Transl Med* 2014; 12: S3.

37. Barreiro E, Rabinovich R, Marin-Corral J, Barberà J a, Gea J, Roca J. Chronic endurance exercise induces quadriceps nitrosative stress in patients with severe COPD. *Thorax* 2009; 64: 13–19.

38. Spruit MA, Augustin IML, Vanfleteren LE, Janssen DJA, Gaffron S, Pennings H-J, Smeenk F, Pieters W, van den Bergh JJAM, Michels A-J, Groenen MTJ, Rutten EPA, Wouters EFM, Franssen FME, CIRO+ Rehabilitation Network. Differential response to pulmonary rehabilitation in COPD: multidimensional profiling. *Eur Respir J* 2015; 46: 1625–1635.

39. van Boven JFM. Costly comorbidities of COPD: the ignored side of the coin? *Eur Respir J* 2017; 50: 1700917.

40. Chen W, FitzGerald JM, Sin DD, Sadatsafavi M, Canadian Respiratory Research Network. Excess economic burden of comorbidities in COPD: a 15-year population-based study. *Eur Respir J* 2017; 50: 1700393.

41. Divo MJ, Casanova C, Marin JM, Pinto-Plata VM, de-Torres JP, Zulueta JJ, Cabrera C, Zagaceta J, Sanchez-Salcedo P, Berto J, Davila RB, Alcaide AB, Cote C, Celli BR, BODE Collaborative Group. COPD comorbidities network. *Eur Respir J* 2015; 46: 640–650.

42. Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, Zulueta J, Cabrera C, Zagaceta J, Hunninghake G, Celli B. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012; 186: 155–161.

43. Dueñas-Espín I, Vela E, Pauws S, Bescos C, Cano I, Cleries M, Contel JC, de Manuel Keenoy E, Garcia-Aymerich J, Gomez-Cabrero D, Kaye R, Lahr MMH, Lluch-Ariet M, Moharra M, Monterde D, Mora J, Nalin M, Pavlickova A, Piera J, Ponce S, Santaeugenia S, Schonenberg H, Störk S, Tegner J, Velickovski F, Westerteicher C, Roca J. Proposals for enhanced health risk assessment and stratification in an integrated care scenario. *BMJ Open* 2016; 6: e010301.

44. Monterde D, Vela E, Clèries M. Los grupos de morbilidad ajustados: nuevo agrupador de morbilidad poblacional de utilidad en el ámbito de la atención primaria. *Atención Primaria* 2016; 48: 674–682.

45. Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, Jensen PB,

Jensen LJ, Brunak S. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat Commun* Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014; 5: 4022.

46. Hu JX, Thomas CE, Brunak S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2016; 17: 615–629.

47. Kannan V, Swartz F, Kiani NA, Silberberg G, Tsipras G, Gomez-Cabrero D, Alexanderson K, Tegnèr J. Conditional Disease Development extracted from Longitudinal Health Care Cohort Data using Layered Network Construction. *Sci Rep* Nature Publishing Group; 2016; 6: 26170.

48. Trivedi DK, Hollywood KA, Goodacre R, Wardy JL, Kaddurah-Daouk NW, Kristal R, London BS, Mendes J, Morrison P, Nikolau N, Robertson B, Sumner D, Taylor LW, Werf C van der, Ommen M, O BVF. Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horizons Transl Med* Academic Press, Oxford; 2017; 3: 294–305.

49. Rankin NJ, Preiss D, Welsh P, Sattar N. Applying metabolomics to cardiometabolic intervention studies and trials: past experiences and a roadmap for the future. *Int J Epidemiol* Oxford University Press; 2016; 45: 1351–1371.

50. Maddox TM, Albert NM, Borden WB, Curtis LH, Ferguson TB, Kao DP, Marcus GM, Peterson ED, Redberg R, Rumsfeld JS, Shah ND, Tcheng JE, American Heart Association Council on Quality of Care and Outcomes Research; Council on Cardiovascular Disease in the Young; Council on Clinical Cardiology; Council on Functional Genomics and Translational Biology; and Stroke Council. The Learning Healthcare System and Cardiovascular Care: A Scientific Statement From the American Heart Association. *Circulation* 2017; 135: e826–e857.

51. Medicine I, America CLHCS, McGinnis JM, Stuckhardt L, Saunders R, Smith M. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. National Academies Press; 2013.

52. Nextcare. Innovation in Integrated Care Services for Chronic Patients, COMRDI15-1-0016. 2016. http://www.nextcarecat.cat/.Date last updated: April 01 2017. Date last accessed: Nov 14 2017. .

53. Burgos F. Impact of Information and Communication Technologies on Remote Testing. Forced Spirometry as a Use Case. 2013. http://hdl.handle.net/10803/284587. Date last updated: Sept 12 2013. Date last accessed: Nov 15 2017. Universitat de Barcelona;

54. Vargas C, Burgos F, Cano I, Blanco I, Caminal P, Escarrabill J, Gallego C, Llauger MA, Miralles F, Solans O, Vallverdú M, Velickovski F, Roca J. Protocol for regional implementation of collaborative lung function testing. *NPJ Prim care Respir Med* Nature Publishing Group; 2016; 26: 16024.

55. Maltais F. Exercise and COPD: Therapeutic Responses, Disease-Related Outcomes, and Activity-Promotion Strategies. *Phys Sportsmed* 2013; 41: 66–80.

56. Mendoza L, Horta P, Espinoza J, Aguilera M, Balmaceda N, Castro A, Ruiz M, Díaz O, Hopkinson NS. Pedometers to enhance physical activity in COPD: a randomised controlled trial. *Eur Respir J* European Respiratory Society; 2015; 45: 347–354.

57. Demeyer H, Louvaris Z, Frei A, Rabinovich RA, de Jong C, Gimeno-Santos E, Loeckx M, Buttery SC, Rubio N, Van der Molen T, Hopkinson NS, Vogiatzis I, Puhan MA, Garcia-Aymerich J, Polkey MI, Troosters T, Mr Papp PROactive study group and the PROactive consortium. Physical activity is increased by a 12-week semiautomated telecoaching programme in patients with COPD: a multicentre randomised controlled trial. *Thorax* 2017; 72: 415–423.

58. Hernandez C, Alonso A, Garcia-Aymerich J, Grimsmo A, Vontetsianos T, Cuyàs FG, Altes AG, Vogiatzis I, Garåsen H, Pellise L, Wienhofen L, Cano I, Meya M, Martinez JI, Escarrabill J, Roca J. Integrated care services: lessons learned from the deployment of the NEXES project. *Int J Integr Care* 2015; 15: e006.

59. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. *Br Med J* 2013; 347: 2–7.

60. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, Rycroft-Malone J, Meissner P, Murray E, Patel A, Sheikh A, Taylor SJC, StaRI Group. Standards for Reporting Implementation Studies (StaRI): explanation and elaboration document. *BMJ Open* BMJ Publishing Group; 2017; 7: e013318.

61. Weiss CH, Krishnan JA, Au DH, Bender BG, Carson SS, Cattamanchi A, Cloutier MM, Cooke CR, Erickson K, George M, Gerald JK, Gerald LB, Goss CH, Gould MK, Hyzy R, Kahn JM, Mittman BS, Mosesón EM, Mularski RA, Parthasarathy S, Patel SR, Rand CS, Redeker NS, Reiss TF, Riekert KA, Rubenfeld GD, Tate JA, Wilson KC, Thomson CC. An Official American Thoracic Society Research Statement: Implementation Science in Pulmonary, Critical Care, and Sleep Medicine. *Am J Respir Crit Care Med* 2016; 194: 1015–1025.

62. Lewis GH. "Impactibility models": identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs. *Milbank Q* Milbank Memorial Fund; 2010; 88: 240–255.

63. Porter M. What is value in health care? *N Engl J Med* 2010; : 2477–2481.

64. Porter ME, Ph D, Larsson S, Ph D, Lee TH. Standardizing Patient Outcomes Measurement. *N Engl J Med* 2016; 374: 10–12.

65. Cano I, Dueñas-Espín I, Hernandez C, de Batlle J, Benavent J, Contel JC, Baltaxe E, Escarrabill J, Fernández JM, Garcia-Aymerich J, Mas MÀ, Miralles F, Moharra M, Piera J, Salas T, Santaeugènia S, Soler N, Torres G, Vargiu E, Vela E, Roca J. Protocol for regional implementation of community-based collaborative management of complex chronic patients. *npj Prim Care Respir Med* 2017; 27: 44.

66. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabasi AL. Uncovering disease-disease relationships through the incomplete interactome. *Science* 2015; 347: 1257601–1257601.

67. Celli BR, Locantore N, Tal-Singer R, Riley J, Miller B, Vestbo J, Yates JC, Silverman EK, Owen CA, Divo M, Pinto-Plata V, Wouters EFM, Faner R, Agusti A, ECLIPSE Study Investigators. Emphysema and extrapulmonary tissue loss in COPD: a multi-organ loss of tissue phenotype. *Eur Respir J* 2018; 51: 1702146.

68. Cano I, Tenyi A, Vela E, Miralles F, Roca J. Perspectives on Big Data applications of health information. *Curr Opin Syst Biol* 2017; 3: 36–42.

69. Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, McDonald DT, Kusebauch U, Moss CL, Zhou Y, Qin S, Moritz RL, Brogaard K, Omenn GS, Lovejoy JC, Hood L. A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat Biotechnol* 2017; 35: 747–756.

70. Torkamani A, Andersen KG, Steinhubl SR, Topol EJ. High-Definition Medicine. *Cell* 2017; 170: 828–843.

71. Cano I, Lluch-Ariet M, Gomez-Cabrero D, Maier D, Kalko S, Cascante M, Tegner J, Miralles F, Herrera D, Roca J, Synergy-COPD consortium. Biomedical research in a Digital Health Framework. *J Transl Med* 2014; 12: S10.

72. Abugessaisa I, Gomez-Cabrero D, Snir O, Lindblad S, Klareskog L, Malmstrom V, Tegner J. Implementation of the CDC translational informatics platform - from genetic variants to the national Swedish Rheumatology Quality Register. *J Transl Med* 2013; 11: 85.

73. Abugessaisa I, Saevarsdottir S, Tsipras G, Lindblad S, Sandin C, Nikamo P, Ståhle M, Malmström V, Klareskog L, Tegnér J. Accelerating Translational Research by Clinically Driven Development of an Informatics Platform–A Case Study. *PLoS One* Public Library of Science; 2014; 9: e104382.

74. Lippi G, Jansen-Duerr P, Viña J, Durrance-Bagale A, Abugessaisa I, Gomez-Cabrero D, Tegnér J, Grillari J, Erusalimsky J, Sinclair A, Rodriguez-Manãs L, FRAILOMIC consortium. Laboratory biomarkers and frailty: presentation of the FRAILOMIC initiative. *Clin Chem Lab Med* 2015; 53: e253-5.

75. Roca J, Cano I, Gomez-Cabrero D, Tegner J. From Systems Understanding to Personalized Medicine: Lessons and Recommendations Based on a Multidisciplinary and Translational Analysis of COPD. *Methods Mol Biol* United States; 2016; 1386: 283–303.

MANUSCRIPT 6: Perspectives on Big Data applications of health information.

Cano I, **Tenyi A**, Vela E, Miralles F, Roca J. Perspectives on Big Data applications of health information. *Curr. Opin. Syst. Biol.* 2017; 3: 36–42.

- Journal: *Current Opinion in Systems Biology*
- Impact factor: -
- Quartile: -
- Status: *Published*

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Perspectives on Big Data applications of health information

Isaac Cano[a], Akos Tenyi[a], Emili Vela[b], Felip Miralles[c] and Josep Roca[a]

Abstract

Recent advances on prospective monitoring and retrospective analysis of health information at national or regional level are generating high expectations for the application of Big Data technologies that aim to analyze at real time high-volumes and/or complex of data from healthcare delivery (e.g., electronic health records, laboratory and radiology information, electronic prescriptions, etc.) and citizens' lifestyles (e.g., personal health records, personal monitoring devices, social networks, etc.). Along these same lines, advances in the field of genomics are revolutionizing biomedical research, both in terms of data volume and prospects, as well as in terms of the social impact it entails. The potential of Big Data applications that consider all of the above levels of health information lies in the possibility of combining and integrating de-identified health information to allow secondary uses of data. This is the use and re-use of various sources of health information for purposes in addition to the direct clinical care of specific patients or the direct investigation of specific biomedical research hypotheses. Current applications include: epidemiological and pharmacovigilance studies, facilitating recruitment to randomized controlled trials, carrying out audits and benchmarking studies, financial and service planning, and ultimately supporting the generation of novel biomedical research outcomes.

Addresses

[a] Hospital Clinic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), CIBERES, Universitat de Barcelona, Barcelona, Catalonia, Spain

[b] CatSalut, Servei Català de la Salut, Barcelona, Catalonia, Spain

[c] Technology Centre of Catalonia (Eurecat), Barcelona, Catalonia, Spain

Corresponding author: Cano, Isaac, Hospital Clínic de Barcelona, IDIBAPS, Villarroel 170, 08036 Barcelona, Spain. Fax: +34 932 275 455. (iscano@clinic.cat)

Current Opinion in Systems Biology 2017, **3**:36–42

This review comes from a themed issue on **Clinical and translational systems biology (2017)**

Edited by **Jesper Tegnér** and **David Gomez-Cabrero**

For a complete overview see the Issue and the Editorial

Available online 27 April 2017

http://dx.doi.org/10.1016/j.coisb.2017.04.012

Keywords

Digital health, Secondary use of data, Health analytics, Predictive modeling, Health forecasting.

Introduction

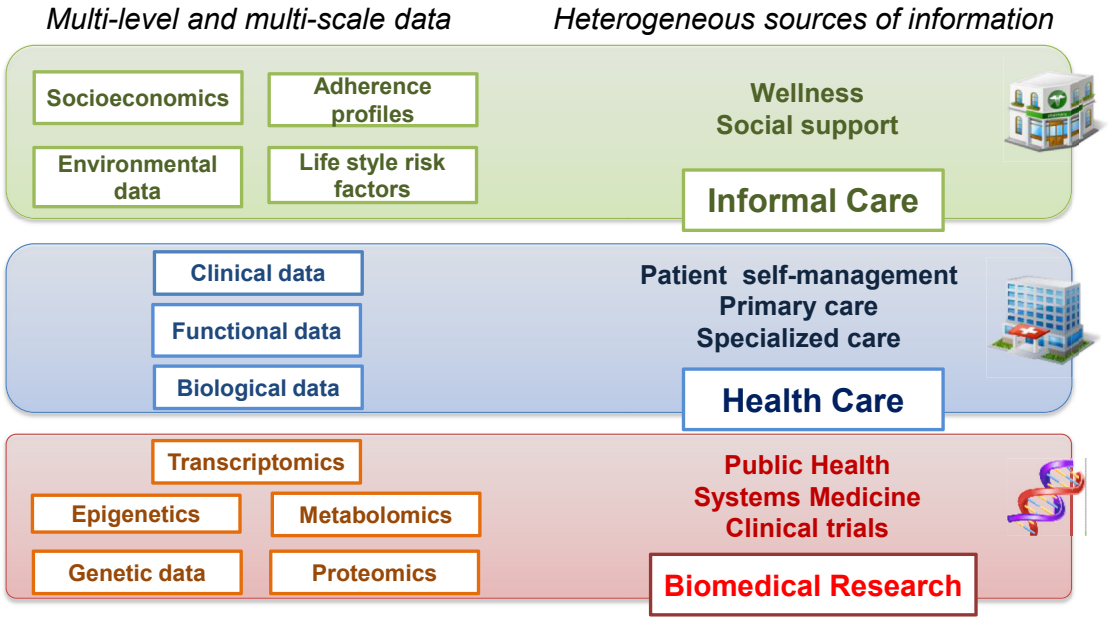
The high prevalence of chronic patients with one or more associated disorders, known as multi-morbidity, is the main source of dysfunctions and avoidable costs in conventional health systems worldwide [1,2]. In this scenario, health risk assessment and stratification are widely accepted tools facilitating large-scale adoption of integrated care of chronic patients [3,4] while generating efficient healthcare and supporting the vision of personalized medicine. However, only a small proportion of the huge potential of risk predictive modeling is being applied [5] for health forecasting of chronic patients due to the lack of in-place procedures for accessing and mining health information from daily clinical practice.

Applying holistic strategies for subject-specific risk prediction and stratification, that consider multilevel covariates influencing patient health, would increase the predictive accuracy and facilitate clinical decision-making based on sound estimates of individual prognosis [6]. For instance, on a daily clinical setting early identification of patient susceptibility to multimorbidity might enable cost-effective preventive strategies (pharmacological and non-pharmacological) and enhance management of chronic patients [7].

Such strategies require dealing with highly complex data and creating new biomedical knowledge, which opens entirely new translational medicine scenarios and requires interplay between clinical practice and biomedical research. This holistic approach generates novel requirements to be adopted by the field. Firstly, the need for multilevel integration of heterogeneous patient information, namely: socio-economical, life-style, behavioral, clinical, physiological, cellular and "omics" data [8], and their use for the study of disease mechanisms. Secondly, the need to extend current trends on open data from the biomedical community [9] to the clinical practice and the whole society, by engaging citizens and solving privacy and regulatory constraints.

A core element for addressing current unmet needs in any given healthcare setting is the deployment of a Digital Health Framework (DHF), as displayed in **Figure 1** and extensively described in [10]. A DHF aims at fostering communication among health registries

Figure 1



Key dimensions of a digital health framework for enhancing communication among informal care, health care and biomedical research, as first described at [10].

containing health information from various sources, namely: (i) healthcare; (ii) informal care, with special emphasis on environmental and self-management information potentially gathered via personal health folders; and (iii) biomedical research.

Driven by recent advances on big data applications [11–17] and the potentials of the vast amount of accumulated patient data, the scope of this manuscript is to firstly report on the key dimensions of a DHF to provide unified access to health registries with all information about the patient's health determinants. Then, we report on current and future potential applications to gain new understanding about the patient's health through big data and modeling tools [15]. It shall ultimately contribute to enhance dynamic health risk assessment and patient stratification, as well as a technological facilitator to support collaborative case management [10].

Health information within a digital health framework

Worldwide, most national and/or regional health services have positioned themselves to allow secondary uses of digitalized real-world data for quality and safety of care, financial management and most recently for research purposes [18]. Examples include, the Clinical Practice Research Datalink health registry [19,20] in UK, the National population-based registries in the Nordic countries [21] or the Medicare registries from the Centers for Medicare & Medicaid Services (www.cms.gov) in US. However, most of these health registries are being adapted and expanded in order to include necessary information not being captured by formal care providers, namely: i) informal care, and, ii) biomedical research. Briefly, informal care includes any aspect with impact on health (e.g. life style, environmental and behavioral aspects, etc.) occurring in the community, whereas biomedical research refers to all research levels from bench to clinical and to public health. This requires development of policies and software solutions [11,12,15,16] that enable smooth data collection and storage as well as data linkage in order to facilitate the extraction of relevant data, its analysis and the communication of findings to relevant parties [10].

Formal care information

Apart from administrative and reimbursement needs, formal care registries aim to focus on health care of the patient, so they contain information from all clinicians involved in the patient's care. This information is mainly captured through Electronic Health Records (EHRs). EHRs refer to the electronic systems that health care professionals use to manage, store, share, and increasingly to analyze heterogeneous health information from emergency room visits and hospitalizations, primary care visits, mental health centers, socio-sanitary centers, drug prescriptions, etc. EHRs allow doctors to better keep track of patients' health information in structured (e.g., ICD [22], HL7-CDA [23], etc.) or non-structured (e.g., free text, pdf, etc.) formats and make it easier to ensure privacy and security of patients' health

information by recording and tracking who has accessed what information and encrypting the information.

Standards like HL7 (Health Level 7 [24]) have contributed to normalize the adoption of EHRs, but current approaches are mostly generating information silos. Hence, further adoption is needed to effectively enable communication across healthcare providers though the exchange of health information of various sources need to be linked in order to provide a comprehensive picture of patients' entire care pathway and care history. To this end, focus should also be set on integrating formal care registries with contextual, real-world data, directly gathered from patients.

Informal care information

Patients are increasingly being considered as potential sources of health information to be linked with provider's EHRs through the use of Personal Health Records (PHRs). PHRs [25] can be used to keep track of information on an ongoing treatment or active monitoring as well as for the management of health conditions partnering with care professionals. The PHR can also contain data on the patient's health priorities toward self-management, such as tracking food, daily activity, blood pressure, etc. Nowadays, a lot of products and mobile health apps [26], aiming at giving patients greater and better control over their health conditions, are available on the market, mostly as standalone applications.

Nevertheless, current PHRs do not reach the scale to have an impact at population level, yet. In fact, they are mostly disconnected from any formal healthcare support (i.e., supervision of the data generated by a healthcare professional, integration with EHRs, etc.). These limitations should be overcome by making sure that health apps support evidence-based health services and that are not purely designed in the interest of market entities, and by establishing reference policies and platforms to support dialog between health apps and PHRs/EHRs [27].

Biomedical research information

To enhance scientific analysis, it is crucial to complement and expand previously discussed sources of health information (e.g. data repositories of formal and informal care settings) with necessary information from biomedical research, not yet linked. This requires stepwise implementation strategies wherein the following key aspects clearly emerge as main short-term priorities.

Firstly, standardization (ISA-Tab [28], MIAME, etc.) is a prerequisite for proper management of biomedical data and metadata that has been already implemented in some public repositories (e.g. Ensembl, Uniprot and KEGG). A second element is the convergence of ongoing developments in the area of knowledge management giving particular priority to the assessment of multilevel interactions that should foster the bridging between omics-generated knowledge and the clinical arena (e.g., disease maps [29–31]). Last but not least, developments aiming at generating user-friendly portals for clinicians devoted to translational research are required. Good examples are the COPD-KB [32], eTRIKS [33], EHR4CR [34] and BioMart [35] platforms.

Potential of health information within a digital health framework

Incorporation of non-clinical information into patient management bears with a major potential to improve formal healthcare. Current advancements in the development of innovative digital health devices and applications [36] are outlining a unique environment where patient-reported outcomes could be used to personalize care [37]. Such informal resources of health information on the one hand potentially enable patients to be more active players in their own health [38,39]. O n the other hand, informal care data can greatly enrich the clinical insight to patient's life. Over the already proven applicability in telemedicine programs [36], self-tracked data such as physical activity has high potential to be used in identifying groups of patients potentially addressed by different interventions [40].

Moreover, articulation of this scenario with systems-oriented biomedical research would provide continuous cross-fertilization between research and patient care [41]. Tools for the integration, management and exploration of high-throughput molecular analyses in the context of clinical care have flourished in the recent years [33,41–43], however current applications mostly include separate cohort studies. Integrating such systems with working EMR solutions in hospitals would enable the development of dynamic predictive modeling approaches, which by taking into account broader biological background of a patient, should facilitate the way toward truly personalized medicine [44]. This will open entirely new and fascinating sce-narios for the interplay between clinical practice and biomedical research professionals.

In this setting, user-profiled business intelligence functionalities and clinical decision support systems can facilitate the use of the same information in different medical services. For example, primary care professionals, specialized care and social care workers could access patient information on clinically-oriented interfaces; whereas, a more detailed view of the patient data could facilitate the work of translational research scientists and clinicians interested in biomedical research. To achieve this scenario huge volume of

information in various format have to be processed and made available for clinical use, which calls for the use of big data tools. An emerging idea in this field is personalized predictive analytics based on patient similarity. At the arrival of a new patient this approach aims to identify similar patients from the historical data and derive insights from their records to provide personalized predictions [45–47]. Examples of current applications of this approach varies from prediction of heart failure from telemonitoring data [48], risk factor Identification of similar patients [49] and personalized treatment and drug recommendation systems [50,51], which list could be further broadened in the DHF scenario.

Overall, the unique potential of information from health data within a digital health framework is the enhanced extraction/generation of novel impactful knowledge through the integration of multiple information sources [48–51]. Development of new models for patient stratification based on this foundation would help to define the most appropriate action plan for the patients, supporting the vision of personalized healthcare. A proper implementation strategy, tackling privacy and regulatory constraints, would highly contribute to enhance healthcare outcomes and patient experience of care while reducing costs and improving the health of populations.

Barriers and opportunities

Barriers and opportunities to enable the previously described potential Big Data applications in Health [15] have been identified in a European Union study on Big Data in Public Health, Telemedicine and Healthcare [52]. As a result of the systematic review recommendations were identified for ten relevant fields, which has been taken into account to structure the following list of potential areas of improvement:

Standards and protocols

Health data is not always available in a digitized form. Its transformation into structured formats (e.g., HL7 CDA [23]) and to move health registries out of current silos in formal care, informal care and biomedical research might be costly. Moreover, current developments focus on standards to guarantee data standardization and interoperability (e.g., ICD [22], DICOM [53], SNOMED [54], HL7 [24], ISATab [28], etc.), but do not consider data quality and how to manage patient identity across data sources (e.g., unique patient identifiers).

Technological developments

New technological and software developments can improve the utility and security of health registries and enable data analysis in real-time settings. However, in order to run preprocessing routines and machine learning algorithms to build predictive models and perform integrative multi-scale simulations [55], it arises the need to allocate clusters of computers working in a collaborative way [56] and supporting novel stacks of privacy-preserving software frameworks and tools [57] which require expert Big Data scientists and engineers.

Data analytics

High awareness and understanding of the added-value of Big Data applications with Health information can promote the development of success stories. Considering that real-time, menu-driven, user-friendly and transparent data analytics tools might not be fully developed yet [58], entrepreneurs [59] and early-adopters [60] might foster the use of innovative Big Data analytics in health.

Privacy and data protection

Balancing the priorities of maintaining and promoting public health and R&D with information from health registries against the privacy of personal data might be a challenge. A shift of collective mind-set toward open data and data sharing scenarios [61] that encourage data authorship as an incentive to data sharing [62] and that are compliant with well-designed and aligned privacy policies might enable a more transparent and comprehensive data value chain.

Legal aspects

Although the General Data Protection Regulation (EU) 2016/679 [63] provides more precise definitions of health data, consent and scientific research, most rules relevant for health (such as the eventual requirement of informed consent, the potential use of professional secrecy as an obstacle to share health information, and the many references to member state laws) might hinder gathering and sharing personal health data. Therefore, there is an urgent EU need for aligning existing fragmented national legislations [64] on collection, storage, analysis, use and dissemination of health data toward the foundation of global legal frameworks to support development and assessment of digital health services [65].

Stakeholders

There is an increasing need for the coordination of interests and responsibilities among different stakeholders (e.g., payers, healthcare providers, academia, clinicians, patients and patients associations, etc.). Involving opinion leaders in different public and private stakeholders groups in public consultations [66] might reduce risk while increasing acceptance and the probability of successful applications.

Business models

Huge potential health and economic benefits can be envisaged in terms of accelerating cross-fertilization between knowledge generation (biomedical research)

and both health and informal care data. Progress in this direction will be strongly associated to innovative business models, such as bundle payment for care improvement [67–69], providing sustainability of platforms beyond specific projects that triggered the initial settings.

Conclusions

In the current paper, latent sources of information about the patient's health determinants are presented and potential strategies are proposed within a digital health framework that aims to support emerging requirements of applied systems medicine. Development of patient health risk assessment and stratification models, integrating health information from informal care, formal healthcare and biomedical research, while creating new knowledge on disease mechanisms, are foreseen as the most promising strategies to support healthcare professionals. The current manuscript also covered current barriers and opportunities concerning technological, legal and economic aspects for an effective improvement of patient experience of care and the health of populations.

Acknowledgments

This manuscript was supported by CONNECARE H2020-689802, NEXTCARE COMRDI15-1-0016, FIS-PITES (PI15/00576) and Generalitat de Catalunya (2014SGR661).

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Murray CJL, Lopez AD: **Measuring the global burden of disease**. *N Engl J Med* 2013, **369**:448–457, http://dx.doi.org/10.1056/NEJMra1201534.

2. Blumenthal D, Chernof B, Fulmer T, Lumpkin J, Selberg J: **Caring for high-need, high-cost patients — an urgent priority**. *N Engl J Med* 2016, **375**:909–911, http://dx.doi.org/10.1056/NEJMp1608511.

3. Cano I, Alonso A, Hernandez C, Burgos F, Barberan-Garcia A, Roldan J, *et al.*: **An adaptive case management system to support integrated care services: lessons learned from the NEXES project**. *J Biomed Inf* 2015, **55**:11–22, http://dx.doi.org/10.1016/j.jbi.2015.02.011.

4. Dueñas-Espín I, Vela E, Pauws S, Bescos C, Cano I, Cleries M, *et al.*: **Proposals for enhanced health risk assessment and stratification in an integrated care scenario**. *BMJ Open* 2016:6, http://dx.doi.org/10.1136/bmjopen-2015-010301.

5. Doos L, Packer C, Ward D, Simpson S, Stevens A: **Past speculations of the future: a review of the methods used for forecasting emerging health technologies**. *BMJ Open* 2016, **6**, http://dx.doi.org/10.1136/bmjopen-2015-010479. e010479.

6. Vanfleteren LEGW, Spruit MA, Groenen M, Gaffron S, van Empel VPM, Bruijnzeel PLB, *et al.*: **Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease**. *Am J Respir Crit Care Med* 2013, **187**:728–735, http://dx.doi.org/10.1164/rccm.201209-1665OC.

7. Vanfleteren LEGW, Spruit MA, Franssen FME: **Tailoring the approach to multimorbidity in adults with respiratory disease: the NICE guideline**. *Eur Respir J* 2017, **49**:1601696, http://dx.doi.org/10.1183/13993003.01696-2016.

8. •• Bousquet J, Anto JM, Sterk PJ, Adcock IM, Chung KF, Roca J, *et al.*: **Systems medicine and integrated care to combat chronic noncommunicable diseases**. *Genome Med* 2011, **3**:43, http://dx.doi.org/10.1186/gm259.
One of the reference papers of Systems Medicine field that propose an innovative, integrated, cost-effective health system to combat major non-communicable diseases based on a robust and extensive knowledge management infrastructure that contains individual patient information.

9. E.C. Global genomic data-sharing effort kicks off. Nature News. 06 March 2014. URL http://www.nature.com/news/global-Genomic-Data-Sharing-Effort-Kicks-off-114826 [accessed 18.03.14] (Archived by WebCite® http://www.webcitation.org/6OAELb42a) 2014.

10. • Cano I, Lluch-Ariet M, Gomez-Cabrero D, Maier D, Kalko S, Cascante M, *et al.*: **Biomedical research in a digital health framework**. *J Transl Med* 2014, **12**:S10, http://dx.doi.org/10.1186/1479-5876-12-S2-S10.
This paper conveys conceptual approaches to embrace the emerging requirements – data and tools – of applying systems medicine into healthcare with a three-tier strategy articulating formal healthcare, informal care and biomedical research.

11. Viceconti M, Hunter P, Hose R: **Big data, big knowledge: big data for personalized healthcare**. *IEEE J Biomed Heal Inf* 2015, **19**:1209–1215, http://dx.doi.org/10.1109/JBHI.2015.2406883.

12. Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ: **Big data for health**. *IEEE J Biomed Heal Inf* 2015, **19**:1193–1208, http://dx.doi.org/10.1109/JBHI.2015.2450362.

13. Luo J, Wu M, Gopukumar D, Zhao Y: **Big data application in biomedical research and health care: a literature review**. *Biomed Inf Insights* 2016, **8**:1–10, http://dx.doi.org/10.4137/BII.S31559.

14. Hamad R, Modrek S, Kubo J, Goldstein BA, Cullen MR: **Using "big data" to capture overall health status: properties and predictive value of a claims-based health risk score**. *PLoS One* 2015, **10**, http://dx.doi.org/10.1371/journal.pone.0126054. e0126054.

15. Murdoch TB, Detsky AS: **The inevitable application of big data to health care**. *JAMA* 2013, **309**:1351–1352, http://dx.doi.org/10.1001/jama.2013.393.

16. Belle A, Thiagarajan R, Soroushmehr SMR, Navidi F, Beard DA, Najarian K: **Big data analytics in healthcare**. *Biomed Res Int* 2015, **2015**:370194, http://dx.doi.org/10.1155/2015/370194.

17. Gerstein M: **Genomics: ENCODE leads the way on big data**. *Nature* 2012, **489**:208.

18. Robertson ARR, Nurmatov U, Sood HS, Cresswell K, Smith P, Sheikh A: **A systematic scoping review of the domains and innovations in secondary uses of digitised health-related data**. *J Innov Heal Inf* 2016, **23**:611, http://dx.doi.org/10.14236/jhi.v23i3.841.

19. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, *et al.*: **Data resource profile: Clinical Practice Research Datalink (CPRD)**. *Int J Epidemiol* 2015, **44**:827–836, http://dx.doi.org/10.1093/ije/dyv098.

20. McDonald L, Lambrelli D, Wasiak R, Ramagopalan SV: **Real-world data in the United Kingdom: opportunities and challenges**. *BMC Med* 2016, **14**:97, http://dx.doi.org/10.1186/s12916-016-0647-x.

21. Ludvigsson J, Nørgaard M, Weiderpass E, Håberg S, LaFolie P, Sarkkola C, *et al.*: **Ethical aspects of registry-based research in the Nordic countries**. *Clin Epidemiol* 2015, **23**:491, http://dx.doi.org/10.2147/CLEP.S90589.

22. Chabra S: **International Classification of Diseases, 10th Revision, coding for prematurity: need for standardized nomenclature**. *Health Care Manag (Frederick)* 2015, **34**:123–127, http://dx.doi.org/10.1097/HCM.0000000000000053.

23. Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, *et al.*: **HL7 clinical document architecture, release 2**. *J Am Med Inf Assoc* 2006, **13**:30–39, http://dx.doi.org/10.1197/jamia.M1888.

24. Kabachinski J: **What is health level 7?** *Biomed Instrum Technol* 2006, **40**:375–379.

25. •• Irizarry T, DeVito Dabbs A, Curran CR: **Patient portals and patient engagement: a state of the science review**. *J Med Internet Res* 2015, **17**:e148, http://dx.doi.org/10.2196/jmir.4255.
This state of the science review presents the definition, background, and how current literature addresses the encouragement and support of patient engagement through the patient portal, and provides a summary of future directions for patient portal research and development to meaningfully impact patient engagement.

26. Fiordelli M, Diviani N, Schulz PJ: **Mapping mHealth research: a decade of evolution**. *J Med Internet Res* 2013, **15**:e95, http://dx.doi.org/10.2196/jmir.2430.

27. • eHealth Action Plan 2012–2020. Green Paper on mobile health ("mHealth"). URL: https://ec.europa.eu/digital-agenda/en/news/green-paper-mobile-health-mhealth. [accessed 22.06.14] (Archived by WebCite® at http://www.webcitation.org/6QWamz0Y4) 2014.
This state of the art review present the definition, background, and how current literature addresses the encouragement and support of patient engagement through the patient portal, and provide a summary of future directions for patient portal research and development to meaningfully impact patient engagement.

28. Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, *et al.*: **Toward interoperable bioscience data**. *Nat Genet* 2012, **44**:121–126, http://dx.doi.org/10.1038/ng.1054.

29. Barabasi AL: **Network medicine–from obesity to the "diseasome."** *N Engl J Med* 2007, **357**:404–407, http://dx.doi.org/10.1056/NEJMe078114.

30. Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, *et al.*: **Integrating pathways of Parkinson's disease in a molecular interaction map**. *Mol Neurobiol* 2014, **49**:88–102, http://dx.doi.org/10.1007/s12035-013-8489-4.

31. Bhinder B, Elemento O: **Towards a better cancer precision medicine: systems biology meets immunotherapy**. *Curr Opin Syst Biol* 2017, **2**:66–72, http://dx.doi.org/10.1016/j.coisb.2017.01.006.

32. Cano I, Tényi Á, Schueller C, Wolff M, Huertas Migueláñez M, Gomez-Cabrero D, *et al.*: **The COPD knowledge base: enabling data analysis and computational simulation in translational COPD research**. *BMC J Transl Med* Sept 2014. http://www.translational-Medicine.com; Sept 2014.

33. • Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, *et al.*: **Making sense of big data in health research: towards an EU action plan**. *Genome Med* 2016, **8**:71, http://dx.doi.org/10.1186/s13073-016-0323-y.
This opinion paper describe the key challenges that result from making sense out of big data in health and using these data for the benefit of the patient and the healthcare system. The authors also highlight key technical, legal, and ethical issues that we face to develop evidence-based personalized medicine, and put forward five recommendations for the European Union (EU) and member states' policy makers.

34. Daniel C, Ouagne D, Sadou E, Forsberg K, Gilchrist MM, Zapletal E, *et al.*: **Cross border semantic interoperability for clinical research: the EHR4CR semantic resources and services**. *AMIA Summits Transl Sci Proc* 2016, **2016**:51–59.

35. Kasprzyk A: **BioMart: driving a paradigm change in biological data management**. *Database (Oxford)* 2011, **2011**, http://dx.doi.org/10.1093/database/bar049. bar049.

36. Bhavnani SP, Narula J, Sengupta PP: **Mobile technology and the digitization of healthcare**. *Eur Heart J* 2016, **37**:1428–1438, http://dx.doi.org/10.1093/eurheartj/ehv770.

37. Weiler A: **mHealth and big data will bring meaning and value to patient-reported outcomes**. *mHealth* 2016, **2**:2, http://dx.doi.org/10.3978/j.issn.2306-9740.2016.01.02.

38. Guarneri MR, Perego P: *Games and gamification for healthy behaviours: the experience of PEGASO fit 4 future*. Cham: Springer; 2017:100–109, http://dx.doi.org/10.1007/978-3-319-49655-9_14.

39. Guarneri MR, Brocca MD, Piras L: *Patient's empowerment and behaviour change: complementary approaches in EU projects PALANTE and PEGASO*. Cham: Springer; 2017:359–369, http://dx.doi.org/10.1007/978-3-319-49655-9_43.

40. Mesquita R, Spina G, Pitta F, Donaire-Gonzalez D, Deering BM, Patel MS, *et al.*: **Physical activity patterns and clusters in 1001 patients with COPD**. *Chron Respir Dis* 2017, http://dx.doi.org/10.1177/1479972316687207.

41. Canuel V, Rance B, Avillach P, Degoulet P, Burgun A: **Translational research platforms integrating clinical and omics data: a review of publicly available solutions**. *Brief Bioinform* 2015, **16**:280–290, http://dx.doi.org/10.1093/bib/bbu006.

42. Szalma S, Koka V, Khasanova T, Perakslis ED: **Effective knowledge management in translational medicine**. *J Transl Med* 2010, **8**:68, http://dx.doi.org/10.1186/1479-5876-8-68.

43. Losko S, Heumann K: **Semantic data integration and knowledge management to represent biological network associations**. *Methods Mol Biol* 2009, **563**:241–258, http://dx.doi.org/10.1007/978-1-60761-175-2_13.

44. Auffray C, Charron D, Hood L: **Predictive, preventive, personalized and participatory medicine: back to the future**. *Genome Med* 2010, **2**:57.

45. Gallego B, Walter SR, Day RO, Dunn AG, Sivaraman V, Shah N, *et al.*: **Bringing cohort studies to the bedside: framework for a "green button" to support clinical decision-making**. *J Comp Eff Res* 2015, **4**:191–197, http://dx.doi.org/10.2217/cer.15.12.

46. Brown S-A: **Patient similarity: emerging concepts in systems and precision medicine**. *Front Physiol* 2016, **7**:561, http://dx.doi.org/10.3389/fphys.2016.00561.

47. Sharafoddini A, Dubin JA, Lee J: **Patient similarity in prediction models based on health data: a scoping review**. *JMIR Med Inf* 2017, **5**:e7, http://dx.doi.org/10.2196/medinform.6730.

48. Henriques J, Carvalho P, Paredes S, Rocha T, Habetha J, Antunes M, *et al.*: **Prediction of heart failure decompensation events by trend analysis of telemonitoring data**. *IEEE J Biomed Heal Inf* 2015, **19**:1757–1769, http://dx.doi.org/10.1109/JBHI.2014.2358715.

49. Ng K, Sun J, Hu J, Wang F: **Personalized predictive modeling and risk factor identification using patient similarity**. *AMIA Jt Summits Transl Sci Proc* 2015, **2015**:132–136.

50. Wang Y, Tian Y, Tian L-L, Qian Y-M, Li J-S: **An electronic medical record system with treatment recommendations based on patient similarity**. *J Med Syst* 2015, **39**:55, http://dx.doi.org/10.1007/s10916-015-0237-z.

51. Zhang P, Wang F, Hu J, Sorrentino R: **Towards personalized medicine: leveraging patient similarity and drug similarity analytics**. *AMIA Jt Summits Transl Sci Proc* 2014, **2014**: 132–136.

52. •• Renner A-T, Bobek J, Habl C, Anja L: **Study on big data in public health**. *Telemedicine Healthc* 2016, http://dx.doi.org/10.4172/2375-4273.C1.026.
This study on Big Data in Public Health, Telemedicine and Healthcare selects ten priority examples of use cases of big data in health and identifies potential policy actions for the implementation of big data in health. The recommendations outlined in this study aim to feed into the European Commission's actions to develop a big data value chain in the context of the European Digital Single Market strategy.

53. Digital Imaging and Communications in Medicine (DICOM) Standard, National Electrical Manufacturers Association, Rosslyn, VA, USA (available free at http://medical.nema.org/) n.d. http://dx.doi.org/NEMA PS3/ISO 12052.

54. Stearns MQ, Price C, Spackman KA, Wang AY: **SNOMED clinical terms: overview of the development process and project status**. *Proc AMIA Symp* 2001:662–666.

55. Tegnér J, Zenil H, Kiani NA, Ball G, Gomez-Cabrero D: **A perspective on bridging scales and design of models using low-dimensional manifolds and data-driven model inference**. *Philos Trans R Soc Lond A Math Phys Eng Sci* 2016:374.

56. Stein LD: **Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges**. *Nat Rev Genet* 2008, **9**:678–688, http://dx.doi.org/10.1038/nrg2414.

57. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, *et al.*: **DataSHIELD: taking the analysis to the data, not the data to the analysis**. *Int J Epidemiol* 2014, **43**:1929–1944, http://dx.doi.org/10.1093/ije/dyu188.

58. Raghupathi W, Raghupathi V: **Big data analytics in healthcare: promise and potential**. *Heal Inf Sci Syst* 2014, **2**:3, http://dx.doi.org/10.1186/2047-2501-2-3.

59. Atomian: *Atomian medical records*. 2016. http://www.atomian.com/atomian-medical-records/.

60. McCowan C, Thomson E, Szmigielski CA, Kalra D, Sullivan FM, Prokosch H-U, *et al.*: **Using electronic health records to support clinical trials: a report on stakeholder engagement for EHR4CR**. *Biomed Res Int* 2015, **2015**, http://dx.doi.org/10.1155/2015/707891. 707891.

61. Longo DL, Drazen JM: **Data sharing**. *N Engl J Med* 2016, **374**:276–277, http://dx.doi.org/10.1056/NEJMe1516564.

62. Bierer BE, Crosas M, Pierce HH: **Data authorship as an incentive to data sharing**. *N Engl J Med* 2017, http://dx.doi.org/10.1056/NEJMsb1616595.

63. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Da. Off J Eur Union 2016. http://eur-lex.europa.eu/eli/reg/2016/679/oj.

64. SMART. Study on the Legal Framework for Interoperable eHealth in Europe. URL http://www.ehealthnews.eu/images/stories/pdf/ehealth-Legal-Fmwk-Final-Report.pdf [accessed 20.08.13] (Archived by WebCite® http//www.webcitation.org/6J0kRIL0R) 2009.

65. Garell C, Svedberg P, Nygren JM: **A legal framework to support development and assessment of digital health services**. *JMIR Med Inf* 2016, **4**:e17, http://dx.doi.org/10.2196/medinform.5401.

66. European Commission: *Public consultation on building the European data economy*. 2017. https://ec.europa.eu/digital-single-market/en/news/public-consultation-building-european-data-economy.

67. Tol J, Swinkels ICS, Struijs JN, Veenhof C, de Bakker DH: *Integrating care by implementation of bundled payments: results from a national survey on the experience of Dutch dietitians*. 2013: 2013.

68. Hirsch JA, Leslie-Mazwi TM, Barr RM, McGinty G, Nicola GN, Silva E, *et al.*: **The bundled payments for care improvement initiative**. *J Neurointerv Surg* 2016, **8**:547–548, http://dx.doi.org/10.1136/neurintsurg-2015-011746.

69. Shih T, Chen LM, Nallamothu BK: **Will bundled payments change health care? Examining the evidence thus far in cardiovascular care**. *Circulation* 2015, **131**:2151–2158, http://dx.doi.org/10.1161/CIRCULATIONAHA.114.010393.

DISCUSSION

The work presented in this PhD thesis has introduced a systems medicine approach to study specific aspects of multimorbidity in chronic diseases, including the development of a novel system biology tool, an analysis of molecular disease mechanisms, an analysis of population-wide comorbidity patterns, and a review and a perspectives paper. The thesis generated outcomes on the concrete and practical use case of chronic obstructive pulmonary disease (COPD) heterogeneity and comorbidity clustering, with the outlook of generalising these methods to non-communicable diseases. In this chapter the main findings, clinical implications, limitations and opportunities of the work are summarised.

MAIN FINDINGS

This PhD thesis has achieved relevant outcomes that support the initial hypothesis and fully achieved main points of the general objectives, namely:

- ✓ To investigate molecular disturbances at body systems level that may lead to a better understanding of characteristic systemic effects and comorbidities of patients with COPD (**Manuscript 1**, **Manuscript 2**, **Manuscript 5**).
- ✓ To analyse population level patterns of COPD comorbidities and investigate their role in the health risk of patients with COPD (**Manuscript 3**, **Manuscript 4**, **Manuscript 5**).
- ✓ To explore technological strategies and tools that facilitate the transfer of the collected knowledge on comorbidity into clinical practice (**Manuscript 1**, **Manuscript 5**, **Manuscript 6**).

Firstly, the PhD thesis introduced a novel knowledge management tool (**Manuscript 1**) for targeted molecular analysis of underlying **disease mechanisms of skeletal muscle dysfunction in patients with COPD**. The presented ChainRank approach facilitated the integration of Synergy-COPD models at different scales for more accurate predictions of COPD molecular events, as well as it indicated high potential as a general search tool for exploring interactions between relevant biological pathways. It is of note that the methodological approach of ChainRank was integrated as part of the knowledge management platform of Biomax Informatics AG.

Next, a systems biology analysis was outlined in **Manuscript 2** to further study the mechanisms of skeletal muscle dysfunction in patients with COPD and the causes of their abnormal adaptation to exercise training. This research work, together with three

other studies [146–148], also aimed to reveal the general underlying causes of comorbidity clustering in COPD using different modelling approaches (i.e. mechanistic, probabilistic and hybrid modelling approaches and animal experiments) (**Manuscript 5**). Overarching outcome of these studies indicated complex co-regulation of core biological pathways (i.e. bioenergetics, inflammation, oxidative stress and tissue remodelling) both on muscle and body systems level (blood, lung). While these results need further validation in independent datasets and comparisons with molecular causes of other comorbid condition of COPD, a major conclusive outcome of these studies was the strong relation of muscle related dysregulations and training adaptation to cardiopulmonary factors, aerobic capacity, whereas the pulmonary severity of COPD did not show significant relation to muscular abnormalities. Further research of the PhD thesis also highlighted the distinctive role of cardiovascular diseases in COPD comorbidities (**Manuscript 3, Manuscript 4**). These findings have far reaching potential in COPD care, starting from defining the need for better characterization of exercise performance in the clinic practice and the promotion of physical activity from early stages of the disease.

The PhD thesis also generated outcomes with respect to the **risk of multimorbidity** in patients with COPD using a population-health approach. The adverse effects of multimorbidity on patients' health is largely acknowledged since 2006 in COPD treatment guidelines [39], however changes in treatment regimens of the patients are yet discouraged in the latest 2017 report [60], indicating that further evidence is needed on actionable factors for clinical application. In **Manuscript 3**, the thesis validated that patients with COPD are in increased risk to co-occur with other diseases compared to non-COPD patients, regardless of the population and healthcare system specificities of different regions (i.e. Catalonia, US). These findings indicated the potential role of multimorbidity as a health risk factor in patients with COPD that was evaluated in **Manuscript 4** by constructing health risk assessment models to predict unexpected medical events in patients with COPD. The promising performance of the models and the prominent role of multimorbidity in these models presented a powerful argument for its role in clinical staging of the disease and their potential in clinical decision support. Furthermore, the outcomes on temporal disease diagnosis patterns, presented in **Manuscript 3**, could further improve the performance of the predictive models used for health risk assessment.

Finally, the thesis made substantial effort to consolidate and summarize current knowledge on COPD comorbidities and contributed to establish **a new perspective of non-pulmonary effects in COPD,** involving strategies for transferring biomedical

research results to healthcare (**Manuscript 5**). Big Data analytics, as the most promising translational field, was further analysed and main implementation challenges were assessed, including strategies to overcome current distributed and inaccessible storage of health data, and other technological and legal elements of this process (**Manuscript 6**).

Clinical implications

The multilevel systems medicine approach adopted by the thesis is of particular importance in clarifying COPD heterogeneity and comorbidity clustering due to several factors that hindered earlier progress in this field. First, lung-centric view of the disease often failed to explain the observed heterogeneity and other COPD specific phenomena, such as clustering of comorbid conditions. Second, hypotheses on the interplay between pulmonary and non-pulmonary manifestations of COPD was based on simplistic assumptions that could not explain the observed comorbidity related phenomena. Finally, the descriptive nature of current classification of non-pulmonary manifestations in COPD leads to confusion in the field and indicated lack of knowledge on how comorbidities develop and interact with COPD and with each other, concerning the patients' wellbeing and survival.

The holistic analysis of health factors, applied in this PhD thesis, enabled to produce several outcomes with relevant implications on clinical management of COPD heterogeneity and comorbidity clustering. First, the revealed complex endotype of skeletal muscle dysfunction and its interplay with multiple systemic factors should guide further investigations on the causes of comorbidity clustering in COPD. These results also indicate that management of patients with COPD must be increasingly based on a better understanding of underlying disease mechanisms. Such systems medicine approach should also help to overcome current problems of disease taxonomy, accurately identify patients in risk and should open new avenues for preventive and personalised management of these patients, including drug repurposing.

Second, the proven impact of multimorbidity in patients with COPD strongly points out that current staging of the disease should move from assessing patient risk and interventional needs based solely on pulmonary manifestations, towards a patient-oriented approach that addresses patient as a whole, including non-pulmonary manifestations of the disease. In this context, management of patients with COPD should also move towards integrative approaches targeting prevention of non-pulmonary manifestations in patients with COPD. A specific example pointed out by

the thesis is cardiovascular health that was shown to be a risk factor for both COPD and skeletal muscle dysfunction, indicating synergies between the management of pulmonary and cardiovascular health, including the promotion of physical activity of patients with COPD, which could potentially slow down the progression of other comorbid diseases.

Finally, the thesis provides strong rationale for the application of health risk assessment in daily clinical decision making. These tools show high potential to enhance healthcare outcomes and patient experience of care while reducing costs and improving the health of populations. For instance, in a clinical setting, early identification of patient susceptibility to multimorbidity might enable cost-effective preventive strategies (pharmacological and non-pharmacological) and enhance management of chronic patients reducing healthcare burden. In COPD care, health risk assessment should address 3 priority areas: i) lung abnormalities (early COPD progression); ii) interplay between pulmonary and systemic effects; and, iii) comorbidities, that should allow for preventive and predictive approaches,

Future developments should consider implementation of appropriate decision support systems in real world settings, gaining inputs from both evidence-based clinical rules and enhanced clinical predictive modelling, that contribute to foster GOLD recommendations [60] by addressing non-pulmonary manifestations of COPD and generate a positive impact on staging and management of COPD.

Strengths and limitations

Several principal limitations in the studies presented in the thesis need to be acknowledged. Methodological limitations, originating from the applied modelling techniques in **Manuscript 1** and in **Manuscript 2**, are acknowledged, such as i) current constraints of available PPI networks [119, 123, 124, 149], ii) modelling proteins levels with gene expression, and iii) comparing body compartments (blood, muscle), may lead to confounding results. It is of note, that these limitations represent current challenges of state-of-the-art systems biology research and utility of similar tools have been broadly demonstrated [36, 109, 114, 119, 132, 150–152].

Moreover, the molecular analysis described in **Manuscript 2** faced several challenges originated from the study design, such as the rather small sample size of the study combined with noisy, heterogeneous in-vivo measurements. These effects were mainly addressed using robust statistical approaches at each step of the analysis, and at project level, through comparing the results of different analysis strategies using the same dataset [147, 153], as well as comparing outcomes to an analysis based on animal

experimentation [148] (**Manuscript 5**). However, it is acknowledged that further longitudinal studies with multilevel measurements are needed to support the main biomedical outcomes and to refine operational strategies.

We acknowledge that molecular heterogeneity of COPD was not specifically addressed in the current PhD thesis. Main reason for this is that disease heterogeneity is rarely taken into account in contemporary network analysis methodologies, which mainly rely on phenotype based grouping of patients, assuming molecular homogeneity legitimated by group based statistics. To avoid such potentially false assumptions, unsupervised approaches for defining sub-groups with similar molecular [154, 155] or disease trajectory background [5, 156] are key for better characterisation of heterogeneity.

Finally, registry information for the identification of COPD cases in **Manuscript 3** and in **Manuscript 4** reflects under-diagnosis of COPD and constitutes a significant limitation. The lack of clinical information forced spirometry data, history of tobacco smoking and information on other risk common factors reduces the potential for exhaustive characterization of patients with COPD. The full spectrum of this information, however, is hardly accessible and their integration raises several technical, ethical and privacy questions that are currently poorly solved in the biomedical sector. Addressing these issues has immense potentials to bridge current gaps between biomedical research and clinical care, as discussed in the upcoming section.

CHALLENGES AND OPPORTUNITIES

The current PhD thesis indicated the potential of systems medicine approaches of health risk assessment for personalized clinical decision making and for large-scale adoption of integrated chronic care. However, their application currently faces major limitations when it comes to accessing and mining health data, stored in distributed silos of information. In this context, integrating and analysing highly complex data would open new avenues for digital health in the clinical arena. Incorporating multi-level determinants of health into risk models would substantially increase the predictive accuracy and facilitate clinical decision-making [157]. Such strategies, however generate several requirements to be adopted by the field. Firstly, there is a need for multilevel integration of heterogeneous patient information, namely: socio-economical, lifestyle, behavioural, clinical, physiological, cellular and "omics" data [157]; and its better exploitation for the study of disease mechanisms. Notwithstanding that already available knowledge is not used enough for bringing conclusions and

hypothesis to clinical practice. Secondly, the open data trends in biomedical research showed high innovative power in this field [158] and thus should be similarly extended to clinical practice by solving privacy and regulatory constraints. Thirdly, the incorporation of non-clinical information into patient management bears with a major potential to improve formal healthcare.

Current advancements in the development of innovative digital health devices and applications [159] are outlining a unique environment where patient-reported outcomes could be used to personalise healthcare delivery [160]. Furthermore, such informal resources of health information potentially enable patients to be more active players in their own health [161, 162] and can greatly enrich the clinical insight to patients' life. Over the already proven applicability in telemedicine programs [159], self-tracked data such as physical activity, as well as patient reported outcomes, have high potential to be used for the identification of patient groups with similar therapeutic needs that can be addressed by more precise interventions [163].

Systems for the integration, management and exploration of high-throughput molecular analyses in the context of clinical care have flourished in the recent years [164–167], however current applications are mainly restricted to separate cohort studies. The integration of such systems with in-place electronic health records (EHR) in hospitals and in primary care centres, would enable the development of dynamic predictive modelling approaches, opening up entirely new and fascinating scenarios for the interplay between clinical practice and biomedical research [168]. In this setting, user-profiled business intelligence functionalities and decision support systems (DSS) would facilitate the use of the same information in different medical services. For example, primary care professionals, specialized care and social care workers could access patient information on clinically-oriented interfaces; whereas, a more detailed view of the patient data could facilitate the work of translational research scientists and clinicians interested in biomedical research. Articulation of this scenario with the systems-oriented biomedical research approach of this PhD thesis would provide continuous cross-fertilization between research and patient care [167].

An ideal digital health and care setting (**Figure 12**) should facilitate an optimal support to care decisions and delivery by reducing the complexity of the massive amount of clinical and multi-disciplinary data being produced every day and to improve efficiency of health outcomes both in terms of well-being and expenditures. Such a health system relies on the availability of health-related data, tools that process it, such as clinical predictive modelling (CPM), and personalised diagnostic and treatment tools, such as

clinical and patient decision support systems (CDSS/PDSS), contributing to the acceleration of evidence diffusion to practice, helping to identify gaps in care and to target interventions to the most appropriate sub-population of patients.

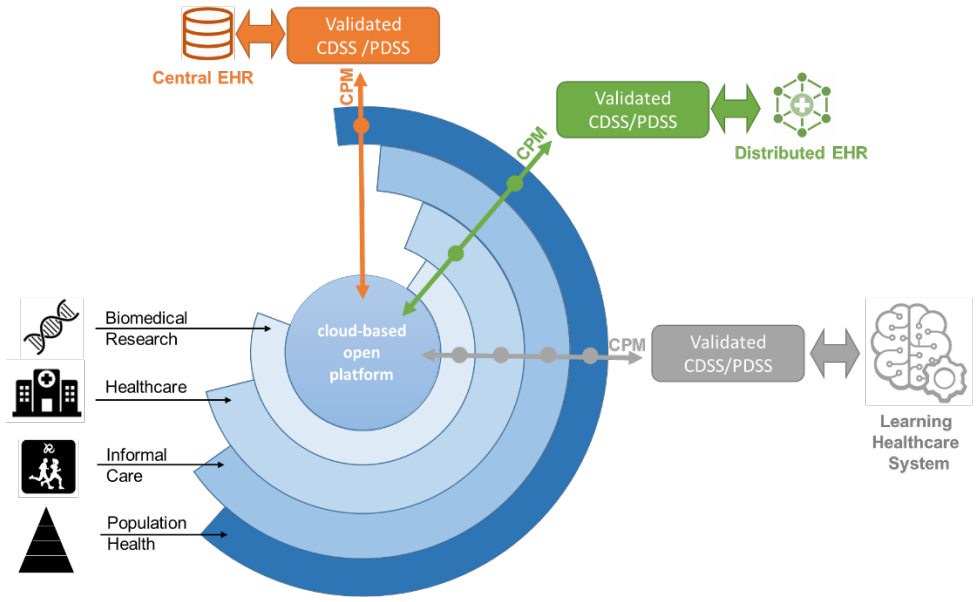


Figure 12. Ideal digital health and care concept for dynamic enhancement of clinical predictive modelling (CPM) feeding clinical DSS (CDSS) and/or patient DSS (PDSS) in cloud-based environments. Development of enhanced CPM to feed DSS will require consideration, and eventual integration, of computational modelling of four different dimensions: i) Underlying biological mechanisms; ii) Current evidence-based clinical knowledge; iii) Patients' self-tracked data, including sensors, behavioural, environmental and social information; and, iv) Population-based health risk assessment data. CDSS/PDSS should be designed to interoperate with existing centralised or distributed hospital information systems, and ultimately with learning health systems (i.e. systems in which science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral by-product of the delivery experience).

This PhD thesis has identified four main interrelated enablers of this scenario [169, 170]: i) **Cloud-based tools and services**: allowing secure analysis of patient-centric distributed and multi-disciplinary health-related information; ii) **In-silico modelling**: Systems Medicine approaches to generate CPM that feed CDSS/PDSS; iii) **Implementation and evaluation**: strategies for real-world implementation and assessment of cloud-based services, and, iv) **Governance, regulatory aspects and service adoption**.

Cloud-based tools and services

End-to-end exploitation of cloud infrastructures for large-scale data analytics has been so far held back by the lack of a well-integrated set of reliable and flexible services, and user-friendly interfaces. This problem is particularly true for medical research and clinical applications where the additional complexity of handling personal information requires particular care. In this context, a cloud-based data analytics platform shall unlock the full potential of CPM by solving issues around integration, harmonization and privacy of data coming from different sources in an integrated manner, and by providing a general interface for developing and deploying predictive models.

Added benefit of such solution includes rapid local prototyping, cost-effective parameter sweeping and validation scale-out to high-performance, large-scale modelling. It should also enable the deployment of the same services across different infrastructures, institutes, laboratories or projects according to local policies, while maintaining interoperability and consistency at the platform layer. Specific implementation of such platform should identify and deploy cloud-based services for private and public uses with a design that is versatile, scalable, trusted and abstract enough to support a wide range of CPM approaches and application in broad operational contexts for digital health and daily clinical practice.

A high-level description of a proposed cloud-based data analytics platform is displayed in **Figure 13** indicating the 4 types of data sources that are considered for multi-disciplinary computational modelling: i) healthcare data from electronic health records; ii) patient self-tracked information, including: sensor-based data, patient reported outcomes and other social/environmental information; iii) population-based information from health registries; and, iv) biomedical research information. Computational modelling in combining these 4 dimensions should provide the basis for enhanced CPM and elaboration of cloud-based CDSS/PDSS embedded into clinical processes.

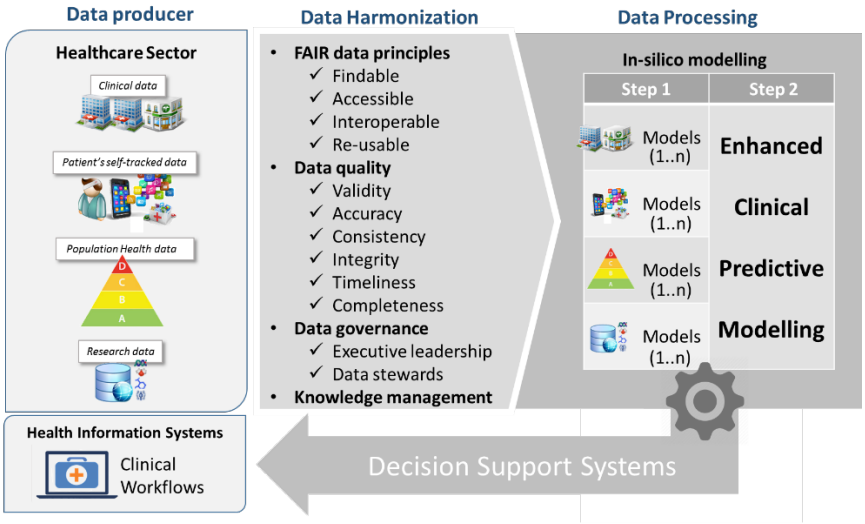


Figure 13. Conceptual view of the proposed cloud-based data analytics platform. Properly harmonised multi-disciplinary data sources provide the basis for enhanced clinical predictive models that feed real-time decision support systems able to guide health professionals in the clinical decision making process.

In the implementation process of such platform, specific requirements include i) development of data interoperability engine for standardised data exchange and semantic interoperability among multi-disciplinary data sources; ii) deployment of technologies to handle personal data in compliance with agreed legal, policy and standardization requirements; and iii) defining standard interfaces in compliance with user needs.

Data Interoperability Engine – A critical aspect of the proposed platform is the development of an efficient data interoperability engine compliant with European regulations and FAIR data principles: i) Findable: data must be easy to find by both humans and computer systems; ii) Accessible: data must be put in long-term storage in such a way that either the data itself or its metadata can be accessed easily; iii) Interoperable: datasets can be combined by humans as well as computer systems, in which the use of shared vocabularies and/or ontologies is of special importance; and, iv) Re-usable: data can be used for future research and to be processed further by computer programs. To this end, several initiatives are available for consideration, e.g. the ELIXIR (www.elixir-europe.org) interoperability platform or commercial medical data federation tools such as FedEHR (www.fedehr.com).

Data Security and Data Protection – Given the need to handle sensitive data, particular attention is needed to be paid to security and data protection aspects. Future data architectures should assume that sensitive data cannot be moved around, but rather code and models have to be moved to the data. When considering such architectures three levels of data should be considered: i) public data; ii) anonymized

data; iii) pseudo-anonymized data. Public data includes public datasets that can be freely shared and access from well-known, community curated data banks and repositories. In this context, creation of public "data lakes" shared across the community and used for example to train the machine learning services and build reference models is needed. Anonymized data need to be defined in compliance with existing and future legal definitions and need to be accessed and contributed based on agreed policies of use within specific projects or communities. Such data, although anonymized, is richer than completely public data as it might convey a specific context of use, thus it is of high interest for research purposes. Finally, pseudo-anonymized data should be used to generate models for the CDSS/PDSS that can then be applied to patients in clinically relevant scenarios. In light of the new European General Data Protection Regulation (GDPR) new technologies are also needed to be considered to manage and audit access to data and to move in the direction of implementing the GDPR norms of data ownership and usage. Data transactions and "smart" user consents could be stored and managed using blockchain technologies to enable the implementation of transparent, end-to-end policies for data governance.

User-Profiled Interfaces: One of the main barriers for translational researcher and clinicians in accessing cloud-based services is currently the lack of intuitive, easy-to-use user interfaces. The main objective of the technical implementation should be to provide an intuitive way for healthcare researchers and practitioners to exploit the capabilities of distributed infrastructures, i.e. accessing wide range of data sources and services without having to understand any of the complexity of the system. At least two levels of interfaces should be provided in a desired platform, which should be co-designed with the communities of use of reference:

- ✓ **Data analysis interfaces for translational researchers**: this type of interfaces provide an intuitive, scriptable, portable access to the platform services. Technologies like Jupyter Notebooks, CERN SWAN (Service for Web based ANalysis) and other similar approaches provide good examples of such environments. The possibility of supporting different types of scripting engines including familiar systems such as the R analysis framework is critical to ease the transition from local computers to cloud-based services.
- ✓ **Decision support interfaces for healthcare professionals and patients**: this type of interfaces should provide clean, web-based access to specialized services and applications, hiding from the end-user (i.e. healthcare professional and patients) the complexity of the underlying infrastructure are service layers.

In-silico modelling

To comply with the vision of the cloud-based data analytics platform, methodologies used in biological and clinical modelling should converge towards standard operations and tools that can be integrated into general pipelines and implemented in analysis platforms. These pipelines shall be ready to analyse data independent of their source or type, i.e. molecular, clinical or wearable measurements, and integrate them in an operational manner. This assumes pre-processed and well formatted data input, as well as standardised outputs (**Figure 13**). When considering subject-specific health risk prediction and stratification as the desired output of such system, the framework of machine learning defines these input and output needs as well as the identified challenges. In this regard, main factors to consider should be the dimensionality of the data sources, i.e. number of features that are used for health prediction, their sample size and differences in sample sizes when considering the integration of multiple datasets. Registry data, EHR data and wearable technologies come with the great promise to bring biomedical research to the Big Data era with population/subpopulation size data, whereas molecular data have great potential to gain biological insight into disease mechanisms, however for these data sources population-wide availability is yet awaited. The integration of such data sources should enable mining health related patterns from data with state-of-the art technologies, such as deep learning that show exciting potential for identifying non-linear patterns from large amount of raw biomedical data [171–173]. Major potential of this technology is that it promises a universal approximator for many learning and prediction tasks that could substitute several processes that are currently done separately in biomedical and machine learning fields. A fascinating way of using deep learning could help to select biologically important features, organise them into higher abstraction level biological assemblies (e.g. pathways, disease modules), highlight their role in the disease and also to predict disease risk using them [171]. A major obstacle, however, is that they are often associated with the need for large longitudinal sample sizes, which is a barrier especially in molecular data sources. Furthermore, more research is needed in this field to abolish its current stigmatization as a "black box" approach, which is often seen as a barrier for clinical application.

Moreover, integration of data from different sources and with different formats is a major challenge of in-silico modelling pipelines. Current practice shows that in different fields different models evolve, such as the disease maps, disease trajectories, mechanistic models, other multiscale hybrid modelling already combining some of the previous approaches; or, data-driven approaches using machine learning.

In order not to waste the field specific knowledge encapsulated in these models, integrative approaches are needed. In this context, patient similarity framework shows immense potential to be applied in the clinical field, mainly due its high abstraction level leading to broad applicability, its patient centred approach and its transparent methodology, which is especially important for acceptability from the clinical side [174, 175]. Patient similarity enables the separate comparison of patients on different biological organisational levels, e.g. using molecular profile (transcriptome, genome, epigenome, etc.), clinical traits, comorbidities, and allows to retrieve groups of similar patients, or the most successful treatments based on similar cases, as well as to predict health risk on an unsupervised manner [176–179].

Implementation and evaluation

For successful adoption in real world settings, interoperability of the proposed cloud-based data analytics platform (**Figure 14**) with healthcare information systems is indispensable. On the one hand, cloud-based services should be integrated at site level with the required structured and unstructured data sources.

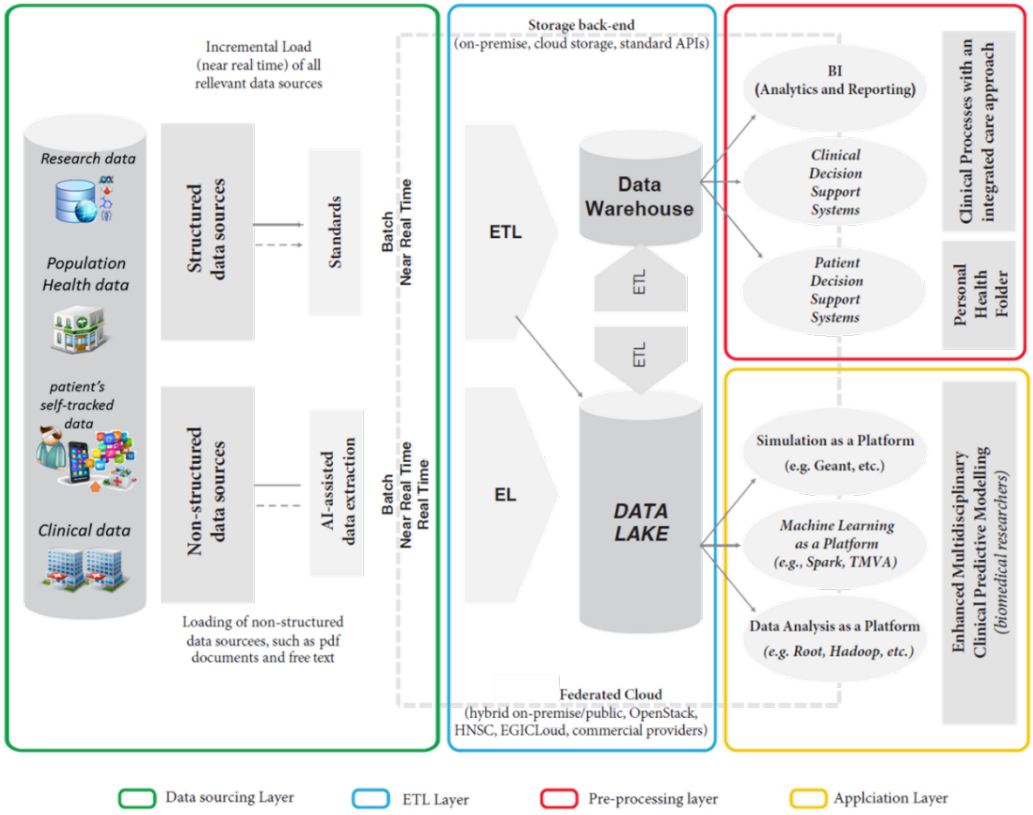


Figure 14. Proposed interoperability architecture. The data sourcing layer is responsible for integrating cloud-based services to required structured and unstructured data sources at site-level. The ETL (Extract-Transform-Load) layer is where data is extracted from homogeneous or heterogeneous data sources and transformed for storing in the proper format or structure for the purposes of querying

and analysis the target data lake or data warehouse. Finally, the application layer provides access to common platform services for data analytics whereas the pre-processing layer is responsible for integrating DSS with site-specific clinical workstations and patient gateways (e.g. Cat@Salut La Meva Salut in Catalonia, ES).

On the other hand, information systems departments of clinical sites should take into account in-place health information exchange infrastructures, where standard terminology (e.g. SNOMED-CT, SERAM, SEMN, LOINC, etc.), message encoding (e.g. HL7 2.x / 3.x, MLHIM, openEHR, ISO 13606, etc.), message routing and security (e.g. IPSec, Audit trail, Node authentication, etc.) are of special importance. Where available, existing controlled vocabularies such as the ICD-10 or the FMA human anatomy, standards for data and metadata format (e.g. ISA-TAB) and content (e.g. MSI or MIAME) should be used. Where standards are currently not yet broadly accepted, agreements should be generated to deploy site-level interoperability middleware based on an HL7 FHIR standard specification (e.g. HAPI FHIR, hapifhir.io).
Moreover, successful adoption in real world settings will likely require specific evaluation designs ranging from standard randomized controlled trials to different implementation science designs [180, 181] depending upon both the Technology Readiness Level of the proposed cloud-based data analytics platform and the clinical setting wherein it should be implemented.

Governance, regulatory aspects and service adoption

Governance of a cloud-based system based on FAIR data principles generates several major challenges in terms of: i) data/services administration accessibility; ii) continuous control of quality assurance programs; iii) compliance with ethical and regulatory issues; as well as, iv) sustainability of the approach over time. These management and governance challenges are expected to be overcome by adopting block chain technology that allow complete traceability of transactions, and most importantly finely granular enforcement of rules in observance of regulation at data origins, and of consent design. All actors of cloud-based systems should be able to verify that sharing, analysis and other handling and use of the information is in accordance of the individual's intentions and applicable laws, regulations and processes.
Regulatory aspects – Several ethical and regulatory issues currently fall into grey areas in terms of regulation (i.e. computational modelling and DSS assessment for medical use). In this regard, generation of recommendations on grey areas should be addressed in future works, with the specific objectives to be covered:

- ✓ To ensure the protection of privacy and compliance with the GDPR (EU) 2016/679, Directive 95/46/EC and ISO27001 conformance for secure storage of pseudonymised multi-disciplinary data, which will improve trust in health research and therefore will facilitate adoption of innovative digital health services;
- ✓ To advise on applicable legislation and regulations to which innovative mathematical models, and components using them, need to comply before they can be deployed and used in healthcare settings;
- ✓ To ensure the ethical use of innovative models within patient and professional decision-making.

Service adoption – Adoption and both organizational (e.g. data processing agreements, liability/responsibility aspects, etc.) and financial sustainability (e.g. entrepreneurial actions) of cloud-based services constitutes a major challenge, wherein service models such as those developed in projects like MyHealthMyData (ICT-18-2016-732907) could be considered. The development of a roadmap for large scale deployment and adoption of cloud services at national and EU level should be of main interest for future initiatives pursuing adoption of novel cloud-based services.

Summary

The proposed cloud-based data analytics platform has been conceived to successfully address the implicated potentials of health risk assessment and stratification and to facilitate large-scale adoption of integrated care of chronic patients [73, 182], contributing to enhance healthcare outcomes and patient experience of care while reducing costs and improving the health of populations. Applying holistic strategies for subject-specific risk prediction and stratification, that consider multilevel covariates influencing patient health, would increase the predictive accuracy and facilitate clinical decision-making based on sound estimates of individual prognosis [157]. Future developments and evaluation of novel DSS fed by enhanced CPM tackling the multimorbidity phenomena constitutes an efficient manner to bring the achievements of the PhD thesis into the real clinical scenario.

CONCLUSIONS

The PhD thesis achieved main points of the general objectives, namely: i) to perform a systems analysis of patients with COPD by investigating molecular disturbances at body systems level leading to a better understanding of characteristic systemic effects and comorbidities of patient with COPD; ii) to analyse population level patterns of COPD comorbidities and investigate their role in the health risk of patients with COPD; and iii) to explore technological strategies and tools that facilitate the transfer of the collected knowledge on comorbidity into clinical practice. Accordingly, the following conclusions arise:

1. Non-pulmonary manifestations in patients with Chronic Obstructive Pulmonary Disease (COPD) have a major negative impact on: highly relevant clinical events, use of healthcare resources and prognosis. Accordingly, the following indications were made:
    a. Actionable insights on non-pulmonary phenomena should be included in the clinical staging of these patients in an operational manner.
    b. Management of patients with COPD should be revisited to incorporate an integrative approach to non-pulmonary phenomena.
    c. Innovative cost-effective interventions, and pharmacological and non-pharmacological treatments targeting prevention of non-pulmonary manifestations in patients with COPD should be developed, and properly assessed.
2. Abnormal co-regulation of core biological pathways (i.e. bioenergetics, inflammation, tissue remodelling and oxidative stress), both in skeletal muscle and at body systems level, are common characteristics of patients with COPD, which potentially play a major role in comorbidity clustering.
3. Consistent relationships between cardiovascular health, skeletal muscle dysfunction and clinical outcomes in patients with COPD was identified, which makes it a priority to characterize patient exercise performance and physical activity in the clinic, and to adopt early cardiopulmonary rehabilitation strategies to modulate prognosis and prevent comorbidity clustering in these patients.
4. Multimorbidity is a strong predictor of unplanned medical events in patients with COPD and shows high potential to be used for personalized health risk prediction and service workflow selection.
5. Personalized health risk prediction was identified as a high potential tool for the integration and transfer of scientific evidence on multimorbidity to daily clinical

practice. Limiting factors of its present applicability were explored and implementation strategies based on cloud computing solutions were proposed.

SUMMARY IN ENGLISH

Background

Multimorbidity (i.e. the presence of more than one chronic disease in the same patient) and comorbidity (i.e. the presence of more than one chronic disease in the presence of an index disease) are main sources of dysfunction in chronic patients and avoidable costs in conventional health systems worldwide. By affecting a majority of elderly population worldwide, multimorbidity prompts the need for revisiting the single disease approach followed by contemporary clinical practice and elaborate strategies that target shared mechanisms of associated diseases with the potential of preventing, decelerating or even halting multimorbid disease progression. However, our current understanding on disease interactions is rather limited, and although many disorders have been associated based on their shared molecular traits and their observed co-occurrence in different populations, no comprehensive approach has been outlined to translate this knowledge into clinical practice.

The advent of novel measurement technologies (e.g. omics) and recent initiatives on digital health (e.g. registries, electronic health records) are facilitating access to an enormous amount of patient-related information from whole populations to molecular levels. State-of-the art computational models and machine learning tools demonstrate high potential for health prediction and together with systems biology are shaping the practicalities of systems medicine. Given the extremely long and expensive bench to clinics cycles of the biomedical sector, systems medicine promises a fast track approach where scientific evidence support clinical care, while simultaneously collected insights from daily clinical practice promote new scientific discoveries and optimize healthcare.

The PhD thesis aims to explore multimorbidity from a systems medicine perspective on the concrete and practical use case of chronic obstructive pulmonary disease (COPD). COPD constitutes an ideal use case due to several factors, including: i) its high impact on healthcare and its ever-increasing burden; ii) its heterogeneous disease manifestations, and progress, often involving extra-pulmonary effects, including highly prevalent comorbidities (e.g. type 2 diabetes mellitus, cardiovascular disorders, anxiety-depression and lung cancer); and, iii) its well described systemic effects that are suggested associations with comorbidities in terms of underlying mechanisms.

Hypothesis

The central hypothesis of the PhD thesis builds on the emerging biological evidence that clustering of comorbid conditions, a phenomenon seen in complex chronic patients, could be due to shared abnormalities in relevant biological pathways (i.e. bioenergetics, inflammation and tissue remodelling). It is assumed that a systems understanding of the patient conditions may help to uncover the molecular mechanisms and lead to the design of preventive and targeted therapeutic strategies aiming at modulating patient prognosis.
The PhD thesis focuses on non-pulmonary phenomena of COPD; that is, systemic effects and comorbidities, often observed in patients with COPD as a paradigm of complex chronic disease.

Objectives

The general objective of the PhD thesis is threefold: i) to investigate molecular disturbances at body systems level that may lead to a better understanding of characteristic systemic effects and comorbidities of patients with COPD; ii) to analyse population level patterns of COPD comorbidities and investigate their role in the health risk of patients with COPD; and, iii) to explore technological strategies and tools that facilitate the transfer of the collected knowledge on comorbidity into clinical practice.

Main findings

Firstly, the PhD thesis introduced a novel knowledge management tool for targeted molecular analysis of underlying **disease mechanisms of skeletal muscle dysfunction in patients with COPD**. Second, a network analysis approach was outlined to further study this systemic effect, as well as the causes of abnormal adaptation of COPD muscle to exercise training. Furthermore, this work together with three other studies also aimed to reveal the general underlying causes of comorbidity clustering in COPD, using different modelling approaches. Overarching outcome of these studies indicates abnormalities in the complex co-regulation of core biological pathways (i.e. bioenergetics, inflammation, oxidative stress and tissue remodelling) both on muscle and body systems level (blood, lung), which paves the way for the development of novel pharmacological and non-pharmacological preventive interventions on non-pulmonary phenomena in patients with COPD. Furthermore, results indicated strong relation of muscle related dysregulations to aerobic capacity, in opposed to pulmonary

severity of COPD. These findings have far reaching potential in COPD care, starting from defining the need for better characterization of exercise performance in the clinic practice and the promotion of physical activity from early stages of the disease. This PhD thesis also generated outcomes with respect to the **risk of multimorbidity** in patients with COPD using a population health approach. The thesis validated that patients with COPD are in increased risk to co-occur with other diseases compared to non-COPD patients, regardless of the population and healthcare system specificities of different regions (i.e. Catalonia, US). These findings indicated the potential role of multimorbidity as a risk factor for COPD, that was evaluated in the PhD thesis by constructing health risk assessment models to predict unexpected medical events in patients with COPD. The promising performance of the models and the prominent role of multimorbidity in these models presented a powerful argument for its role in clinical staging of the disease and their potential in clinical decision support.

Conclusions

The PhD thesis achieved main points of the general objectives, namely: i) to perform a systems analysis of patients with COPD by investigating molecular disturbances at body systems level leading to a better understanding of characteristic systemic effects and comorbidities of patient with COPD; ii) to analyse population level patterns of COPD comorbidities and investigate their role in the health risk of patients with COPD; and iii) to explore technological strategies and tools that facilitate the transfer of the collected knowledge on comorbidity into clinical practice. Accordingly, the following conclusions arise:

1. Non-pulmonary manifestations in patients with Chronic Obstructive Pulmonary Disease (COPD) have a major negative impact on: highly relevant clinical events, use of healthcare resources and prognosis. Accordingly, the following indications were made:
    a. Actionable insights on non-pulmonary phenomena should be included in the clinical staging of these patients in an operational manner.
    b. Management of patients with COPD should be revisited to incorporate an integrative approach to non-pulmonary phenomena.
    c. Innovative cost-effective interventions, and pharmacological and non-pharmacological treatments targeting prevention of non-pulmonary manifestations in patients with COPD should be developed, and properly assessed.

2. Abnormal co-regulation of core biological pathways (i.e. bioenergetics, inflammation, tissue remodelling and oxidative stress), both in skeletal muscle and at body systems level, are common characteristics of patients with COPD, which potentially play a major role in comorbidity clustering.
3. Consistent relationships between cardiovascular health, skeletal muscle dysfunction and clinical outcomes in patients with COPD was identified, which makes it a priority to characterize patient exercise performance and physical activity in the clinic, and to adopt early cardiopulmonary rehabilitation strategies to modulate prognosis and prevent comorbidity clustering in these patients.
4. Multimorbidity is a strong predictor of unplanned medical events in patients with COPD and shows high potential to be used for personalized health risk prediction and service workflow selection.
5. Personalized health risk prediction was identified as a high potential tool for the integration and transfer of scientific evidence on multimorbidity to daily clinical practice. Limiting factors of its present applicability were explored and implementation strategies based on cloud computing solutions were proposed.

RESUM EN CATALÀ

Introducció

Tant la multimorbiditat (la presència de més d'una malaltia crònica en el mateix pacient), com la comorbiditat (la presència de més d'una malaltia crònica quan hi ha una malaltia de referència) són una font important de disfuncions en l'atenció sanitària dels pacients crònics i generen importants despeses evitables en sistemes de salut arreu del món. La multimorbiditat/comorbiditat afecta la majoria de població de més de 65 anys. El seu gran impacte sanitari i social fa necessària la revisió d'aspectes essencials de la pràctica mèdica convencional, molt enfocada al tractament de cada malaltia de forma aïllada. En aquest sentit, cal elaborar estratègies que considerin els mecanismes biològics comuns entre patologies, per tal de prevenir, retardar o fins i tot aturar la progressió del fenomen. Malauradament, el poc coneixement dels mecanismes biològics que modulen les interaccions entre malalties és un factor limitant important. Hi ha estudis sobre els mecanismes moleculars comuns entre malalties i s'han realitzat anàlisis poblacionals de la multimorbiditat, però no existeix encara una aproximació holística per tal de traduir aquest coneixement a la pràctica clínica.

L'aparició de noves tecnologies òmiques, així com iniciatives recents en l'àmbit de la salut digital, han facilitat l'accés a una quantitat enorme d'informació dels pacients, tant a nivell poblacional com a nivell molecular. A més, les eines computacionals i d'aprenentatge automàtic existents estan demostrant un gran potencial predictiu que, conjuntament amb les metodologies de la biologia de sistemes, estan conformant els aspectes pràctics del desplegament de la medicina de sistemes. De forma progressiva, aquesta última esdevé una via efectiva per accelerar el rol de l'evidència científica com a suport a la atenció clínica. De forma recíproca, la digitalització sistemàtica de la pràctica clínica diària, permet la generació de noves descobertes científiques i la optimització de l'assistència sanitària.

Aquesta tesis doctoral pretén explorar la multimorbiditat des d'una perspectiva de medicina de sistemes, considerant com a cas d'ús concret i pràctic la malaltia pulmonar obstructiva crònica (MPOC). La MPOC constitueix un cas d'ús ideal a causa de diversos factors: i) el seu alt impacte a nivell sanitari; ii) la heterogeneïtat en quant a manifestacions i progrés, sovint amb efectes extra-pulmonars, incloent de forma freqüent comorbiditats com la diabetis mellitus tipus 2, trastorns cardiovasculars, l'ansietat-depressió i el càncer de pulmó; i, iii) els efectes sistèmics de la malaltia

pulmonar, que podrien presentar mecanismes biològics comuns a algunes comorbiditats.

Hipòtesis

La hipòtesi central d'aquesta tesis doctoral considera que la multimorbiditat podria explicar-se per alteracions en les xarxes de regulació de mecanismes biològics rellevants com la bioenergètica, inflamació i remodelació de teixits. En aquest sentit, l'anàlisi holística del problema podria millorar la comprensió dels mecanismes moleculars que modulen les associacions entre malalties i, per tant, facilitar el disseny d'estratègies terapèutiques preventives i dirigides a modular el pronòstic dels pacients.

Aquesta tesis doctoral estudia els fenòmens extra-pulmonars de la MPOC; és a dir, efectes sistèmics (disfunció del múscul esquelètic) i comorbiditats, com a paradigma de malalties cròniques complexes.

Objectius

L'objectiu general d'aquesta tesis doctoral és triple: i) l'anàlisi holístic de pacients amb MPOC amb focus en la disfunció muscular i les comorbiditats; ii) avaluar el paper de les comorbiditats en el risc de salut dels pacients amb MPOC, tant a nivell poblacional com individual; i, iii) explorar estratègies tecnològiques i eines de salut digital que facilitin la transferència de coneixement a la pràctica clínica diària.

Resultats

El primer manuscrit de la tesi descriu una nova eina de gestió del coneixement per l'anàlisi molecular dels mecanismes de disfunció del múscul esquelètic en pacients amb MPOC. També dins el primer objectiu de la tesi, s'efectua un anàlisi de xarxes orientat a la identificació de mòduls biològics explicatius de la disfunció muscular i de l'adaptació anòmala d'aquests malalts a l'entrenament físic, tal com es descriu en el segon manuscrit. Els tres articles següents exploren, des de diferents perspectives, l'impacte i mecanismes de les comorbiditats en els pacients amb MPOC. Els principals resultats d'aquests estudis indiquen una complexa i anormal regulació de vies biològiques principals, com es el cas de la bioenergètica, inflamació, estrès oxidatiu i remodelació de teixits, tant a nivell del múscul com a nivell sistèmic (sang, pulmó). Aquests resultats obren noves vies per a intervencions preventives, tant farmacològiques com no farmacològiques, sobre els fenòmens no pulmonars que presenten els pacients amb MPOC. Els resultats indiquen una associació de les

alteracions musculars amb la capacitat aeròbica, i no pas amb la gravetat de la malaltia pulmonar. Aquestes troballes tenen un gran potencial en la millora de la gestió dels pacients amb MPOC, començant per la necessitat d'una millor caracterització de la capacitat aeròbica en la pràctica clínica i la promoció d'activitat física des de les primeres etapes de la malaltia.

La tesi també ha generat resultats d'interès en relació amb el risc de multimorbiditat en pacients amb MPOC, mitjançant un enfocament de salut poblacional. Els resultats evidencien que els pacients amb MPOC presenten un risc mes elevat de comorbiditat que els pacients sense MPOC, independentment de les especificitats de la població i del sistema sanitari de les àrees analitzades (Catalunya, EUA). La tesi també demostra el paper de la multimorbiditat com a factor modulador del risc clínic dels pacients amb MPOC. Aquests resultats indiquen l'interès de l'ús de la multimobiditat en l'estadiatge dels pacients amb MPOC i en l'elaboració d'eines de suport al procés de decisió clínica.

Conclusions

Aquesta tesi doctoral ha assolit els objectius generals plantejats i proposa les següents conclusions:

1. Les manifestacions no pulmonars en els pacients amb malaltia pulmonar obstructiva crònica (MPOC) tenen un impacte negatiu respecte a esdeveniments de gran rellevància clínica, ús de recursos sanitaris i pronòstic. En conseqüència, es fan les següents recomanacions:
    a. Els fenòmens no pulmonars de la MPOC s'haurien d'incloure de manera operativa en l'estadiatge d'aquests pacients.
    b. S'hauria de redefinir la gestió clínica dels pacients amb MPOC tot incorporant un enfocament holístic dels fenòmens no pulmonars.
    c. S'haurien de desenvolupar i avaluar correctament noves intervencions, farmacològiques i no farmacològiques, per a la prevenció de les manifestacions no pulmonars en pacients amb MPOC.
2. Les alteracions de la regulació de vies biològiques rellevants com la bioenergètica, inflamació, estrès oxidatiu i la remodelació de teixits a nivell del múscul esquelètic, i també a nivell sistèmic, s'observa en els pacients amb MPOC i pot tenir un paper important en les co-morbiditats.
3. Les relacions entre alteracions cardiovasculars, disfunció del múscul esquelètic i altres aspectes clínics dels pacients amb MPOC, indiquen la necessitat de caracteritzar la capacitat aeròbica i els nivells d'activitat física en la pràctica

clínica, així com la implementació d'estratègies de rehabilitació cardiopulmonar en les primeres etapes de la malaltia, per tal de modular la prognosis dels malalts i prevenir l'aparició de comorbiditats.

4. La multimorbiditat és un bon predictor d'esdeveniments clínics rellevants en pacients amb MPOC i mostra un gran potencial per a personalitzar l'estimació de risc i la selecció de serveis.
5. La predicció de risc de forma personalitzada s'ha identificat com una eina amb molt potencial per a la gestió de la multimorbiditat en la pràctica clínica diària. S'han explorat els factors limitants de la seva aplicabilitat i s'han proposat estratègies d'implementació d'eines predictives adients, basades en solucions de computació en el núvol.

RESUMEN EN CASTELLANO

Introducción

Tanto la multimorbilidad (la presencia de más de una enfermedad crónica en un mismo paciente) como la comorbilidad (la presencia de más de una enfermedad crónica en presencia de una enfermedad de referencia) son una fuente importante de disfunciones en la atención sanitaria de los pacientes crónicos y generan importantes costes evitables en los sistemas de salud de todo el mundo. La multimorbilidad/comorbilidad afecta a la mayoría de la población de más de 65 años. Debido a su gran impacto sanitario y social, resulta necesaria la revisión de aspectos esenciales de la práctica médica convencional, muy enfocada en el tratamiento de cada enfermedad de forma aislada. En este sentido, es necesario elaborar estrategias que consideren mecanismos biológicos comunes entre patologías, con el fin de prevenir, retrasar o incluso detener la progresión del fenómeno. Desgraciadamente, el escaso conocimiento de los mecanismos biológicos que modulan las interacciones entre enfermedades es un factor limitante importante. Existen estudios sobre los mecanismos moleculares comunes entre enfermedades y se han realizados análisis poblaciones de la multimorbilidad, pero no existe aún una aproximación holística que permita traducir este conocimiento a la práctica clínica.

La aparición de nuevas tecnologías ómicas, así como recientes iniciativas en el ámbito de la salud digital, han facilitado el acceso a una cantidad enorme de información sobre los pacientes, tanto a nivel poblacional como a nivel molecular. Además, las herramientas computacionales y de aprendizaje automático existentes demuestran un gran potencial predictivo que, conjuntamente con las metodologías de biología de sistemas, están conformando los aspectos prácticos de la medicina de sistemas. De manera progresiva esta última se está convirtiendo en una vía efectiva para acelerar el papel de la evidencia científica como soporte a la atención clínica. De forma recíproca, la digitalización sistemática de la práctica clínica diaria permite la generación de nuevos descubrimientos científicos y la optimización de la asistencia sanitaria.

Esta tesis doctoral pretende explorar la multimorbilidad desde una perspectiva de medicina de sistemas, considerando como caso de uso concreto y práctico la enfermedad pulmonar obstructiva crónica (EPOC). La EPOC constituye un caso de uso ideal debido a diversos factores: i) su alto impacto a nivel sanitario; ii) la heterogeneidad en cuanto a manifestaciones y progreso, a menudo con efectos extra pulmonares, incluyendo de forma frecuente comorbilidades como la diabetes mellitus tipo 2, trastornos cardiovasculares, la ansiedad-depresión y el cáncer de pulmón; y,

iii) los efectos sistémicos de la enfermedad pulmonar, que podrían presentar mecanismos biológicos comunes a algunas comorbilidades.

Hipótesis

La hipótesis central de esta tesis doctoral considera que la multimorbilidad podría explicarse por alteraciones en las redes de regulación de mecanismos biológicos relevantes como la bioenergética, inflamación y remodelación de tejidos. En este sentido, el análisis holístico del problema podría mejorar la comprensión de los mecanismos moleculares que modulan las asociaciones entre enfermedades y, por tanto, facilitar el diseño de estrategias terapéuticas preventivas y dirigidas a modular el pronóstico de los pacientes.

Esta tesis doctoral estudia los fenómenos extra pulmonares de la EPOC; es decir, efectos sistémicos (disfunción del músculo esquelético) y comorbilidades, como paradigma de enfermedades crónicas complejas.

Objetivos

El objetivo general de esta tesis doctoral es triple: i) el análisis holístico de pacientes con EPOC focalizando en la disfunción muscular y la comorbilidades; ii) evaluar el papel de las comorbilidades en el riesgo de salud de los pacientes con EPOC, tanto a nivel poblacional como individual; y, iii) explorar estrategias tecnológicas y herramientas de salud digital que faciliten la transferencia de conocimiento a la práctica clínica diaria.

Resultados

El primer manuscrito de la tesis describe una nueva herramienta de gestión del conocimiento para el análisis molecular de los mecanismos de disfunción del músculo esquelético en pacientes con EPOC. Incluido en el primer objetivo de la tesis, se efectúa un análisis de redes orientado a la identificación de módulos biológicos que explican la disfunción muscular y la adaptación anómala de estos pacientes al entrenamiento físico, tal y cómo se describe en el segundo manuscrito. Los tres artículos siguientes exploran, desde perspectivas diferentes, el impacto y mecanismos de las comorbilidades en los pacientes con EPOC. Los principales resultados de estos estudios indican una compleja y anormal regulación de vías biológicas principales, como es el caso de la bioenergética, inflamación, estrés oxidativo y remodelación de tejidos, tanto a nivel del músculo como a nivel sistémico (sangre, pulmón). Estos resultados abren nuevas vías para intervenciones preventivas, tanto farmacológicas

como no farmacológicas, sobre los fenómenos no pulmonares que presentan los pacientes con EPOC. Los resultados indican una asociación de las alteraciones musculares con la capacidad aeróbica, y no con la gravedad de la enfermedad pulmonar. Estos hallazgos tienen un gran potencial en la mejora de la gestión de los pacientes con EPOC, empezando por la necesidad de una mejor caracterización de la capacidad aeróbica en la práctica clínica y la promoción de actividad física desde etapas tempranas de la enfermedad.

La tesis también ha generado resultados de interés en relación con el riesgo de multimorbilidad en pacientes con EPOC, mediante un enfoque de salud poblacional. Los resultados evidencian que los pacientes con EPOC presentan un mayor riesgo de comorbilidad que los pacientes sin EPOC, independientemente de las especificidades de la población y del sistema sanitario de las áreas analizadas (Cataluña, EUA). La tesis demuestra también el papel de la multimorbilidad como factor modulador del riesgo clínico de los pacientes con EPOC. Estos resultados indican la conveniencia del uso de la multimorbilidad en el estadiaje de los pacientes con EPOC y en la elaboración de herramientas de soporte al proceso de decisión clínica.

Conclusiones

Esta tesis doctoral ha conseguido los objetivos generales planteados y propone las siguientes conclusiones:

1. Las manifestaciones no pulmonares en los pacientes con enfermedad pulmonar obstructiva crónica (EPOC) tienen un impacto negativo respecto a eventos de gran relevancia clínica, uso de recursos sanitarios y pronóstico. En consecuencia, se formulan las siguientes recomendaciones:
    a) Los fenómenos no pulmonares de la EPOC deberían incluirse de manera operativa en el estadiaje de estos pacientes.
    b) Se debería redefinir la gestión clínica de los pacientes con EPOC incorporando un enfoque holístico de los fenómenos no pulmonares.
    c) Se deberían desarrollar y evaluar correctamente nuevas intervenciones, farmacológicas y no farmacológicas, para la prevención de las manifestaciones no pulmonares en pacientes con EPOC.
2. Las alteraciones de la regulación de vías biológicas relevantes como la bioenergética, inflamación, estrés oxidativo y la remodelación de tejidos a nivel del músculo esquelético y también a nivel sistémico, se observa en pacientes con EPOC y puede tener un papel importante en las comorbilidades.

3. Las relaciones entre alteraciones cardiovasculares, disfunción del músculo esquelético y otros aspectos clínicos de los pacientes con EPOC, indican la necesidad de caracterizar la capacidad aeróbica y los niveles de actividad física en la práctica clínica, así como la implementación de estrategias de rehabilitación cardiopulmonar en las primeras etapas de la enfermedad, con el fin de modular el pronóstico de los pacientes y prevenir la aparición de comorbilidades.
4. La multimorbilidad es un buen predictor de eventos clínicos relevantes en pacientes con EPOC y muestra un gran potencial para personalizar la estimación de riesgo y la selección de servicios.
5. La predicción del riesgo de forma personalizada se ha identificado como una herramienta con alto potencial para la gestión de la multimorbilidad en la práctica clínica diaria. Se han explorado los factores limitantes de su aplicabilidad y se han propuesto estrategias de implementación de herramientas predictivas adecuadas, basadas en soluciones de computación en la nube.

REFERENCES

1. Murray CJL, Lopez AD. Measuring the Global Burden of Disease. *N. Engl. J. Med.* Massachusetts Medical Society; 2013; 369: 448–457.

2. Blumenthal D, Chernof B, Fulmer T, Lumpkin J, Selberg J. Caring for High-Need, High-Cost Patients — An Urgent Priority. *N. Engl. J. Med.* Massachusetts Medical Society; 2016; 375: 909–911.

3. Hidalgo CA, Blumm N, Barabási AL, Christakis NA. A Dynamic Network Approach for the Study of Human Phenotypes. *PLoS Comput. Biol.* Public Library of Science; 2009; 5: e1000353.

4. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabasi A-L. Uncovering disease-disease relationships through the incomplete interactome. *Science (80-.).* 2015; 347: 1257601–1257601.

5. Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, Jensen PB, Jensen LJ, Brunak S. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat. Commun.* Nature Publishing Group; 2014; 5: 4022-.

6. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* National Academy of Sciences; 2007; 104: 8685–8690.

7. Lee D-S, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabási A-L. The implications of human metabolic network topology for disease comorbidity. *Proc. Natl. Acad. Sci. U. S. A.* National Academy of Sciences; 2008; 105: 9880–9885.

8. Barnett K, Mercer SWS, Norbury M, Watt GG, Wyke S, Guthrie B, WHO, Wolff J, Starfield B, Anderson G, Starfield B, Shi L, Macinko J, Schoen C, Osborn R, Akker M van den, Buntinx F, Metsemakers J, Roos S, Knottnerus J, Walker A, Salisbury C, Johnson C, Purdy S, Valderas J, Montgomery A, Gijsen R, Hoeymans N, Schellevis F, Ruwaard D, et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet (London, England)* World Health Organization, Geneva, Switzerland; 2012; 380: 37–43.

9. Bousquet J, Anto JM, Sterk PJ, Adcock IM, Chung KF, Roca J, Agusti A, Brightling C, Cambon-Thomsen A, Cesario A, Abdelhak S, Antonarakis SE, Avignon A, Ballabio A, Baraldi E, Baranov A, Bieber T, Bockaert J, Brahmachari S, Brambilla C, Bringer J, Dauzat M, Ernberg I, Fabbri L, Froguel P, Galas D, Gojobori T, Hunter P, Jorgensen C, Kauffmann F, et al. Systems medicine and integrated care to combat chronic noncommunicable diseases. *Genome Med.* BioMed Central; 2011; 3: 43.

10. Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, Fang H, Neumann S, Tong W, Amaral-Zettler L, Begley K, Booth T, Bougueleret L, Burns G, Chapman B, Clark T, Coleman LA, Copeland J, Das S, de Daruvar A, de Matos P, Dix I, Edmunds S, Evelo CT, Forster MJ, Gaudet P, Gilbert J, Goble C, Griffin JL, Jacob D, et al. Toward interoperable bioscience data. *Nat Genet* 2012/01/28. 2012; 44: 121–126.

11. Barabási A-L. Network Medicine — From Obesity to the "Diseasome." N. Engl. J. Med. 2007. p. 404–407.

12. Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, Crespo I, Perumal TM, Jurkowski W, Antony PMA, Diederich N, Buttini M, Kodama A, Satagopam VP, Eifes S, del Sol A, Schneider R, Kitano H, Balling R. Integrating Pathways of Parkinson's Disease in a Molecular Interaction Map. *Mol. Neurobiol.* 2014; 49: 88–102.

13. Bhinder B, Elemento O. Towards a better cancer precision medicine: Systems biology meets immunotherapy. *Curr. Opin. Syst. Biol.* 2017; 2: 67–73.

14. Loscalzo J, Kohane I, Barabasi A-L. Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. *Mol. Syst. Biol.* 2007; 3: 124.

15. Cano I, Lluch-Ariet M, Gomez-Cabrero D, Maier D, Kalko S, Cascante M, Tegnér J, Miralles F, Herrera D, Roca J. Biomedical research in a Digital Health Framework. *J. Transl. Med.* BioMed Central; 2014; 12: S10.

16. Viceconti M, Hunter P, Hose R. Big data, big knowledge: big data for personalized healthcare. *IEEE J. Biomed. Heal. informatics* United States; 2015; 19: 1209–1215.

17. Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ. Big Data for Health. IEEE J. Biomed. Heal. Informatics 2015. p. 1193–1208.

18. Hamad R, Modrek S, Kubo J, Goldstein BA, Cullen MR. Using "big data" to capture overall health status: properties and predictive value of a claims-based health risk score. *PLoS One* 2015; 10: e0126054.

19. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* United States; 2013; 309: 1351–1352.

20. Belle A, Thiagarajan R, Soroushmehr SMR, Navidi F, Beard DA, Najarian K. Big Data Analytics in Healthcare. *Biomed Res. Int.* United States; 2015; 2015: 370194.

21. Luo J, Wu M, Gopukumar D, Zhao Y. Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed. Inform. Insights* SAGE Publications; 2016; 8: 1–10.

22. Gerstein M. Genomics: ENCODE leads the way on big data. *Nature* 2012; 489: 208–208.

23. Robertson ARR, Nurmatov U, Sood HS, Cresswell K, Smith P, Sheikh A. A systematic scoping review of the domains and innovations in secondary uses of digitised health-related data. *J. Innov. Heal. informatics* 2016; 23: 611–619.

24. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int. J. Epidemiol.* 2015; 44: 827–836.

25. McDonald L, Lambrelli D, Wasiak R, Ramagopalan S V. Real-world data in the United Kingdom: opportunities and challenges. *BMC Med.* BioMed Central; 2016; 14: 97.

26. Maddox TM, Albert NM, Borden WB, Curtis LH, Ferguson TB, Kao DP, Marcus GM, Peterson ED, Redberg R, Rumsfeld JS, Shah ND, Tcheng JE. The Learning Healthcare System and Cardiovascular Care: A Scientific Statement From the American Heart Association. *Circulation* 2017; .

27. Oltvai ZN, Barabási A-L. Systems biology. Life's complexity pyramid. *Science* American Association for the Advancement of Science; 2002; 298: 763–764.

28. Violan C, Foguet-Boreu Q, Flores-Mateo G, Salisbury C, Blom J, Freitag M, Glynn L, Muth C, Valderas JM. Prevalence, Determinants and Patterns of Multimorbidity in Primary Care: A Systematic Review of Observational Studies. Scuteri A, editor. *PLoS One* 2014; 9: e102149.

29. Halling A, Fridh G, Ovhed I. Validating the Johns Hopkins ACG Case-Mix System of the elderly in Swedish primary health care. *BMC Public Health* 2006; 6: 171.

30. Cashion W, McClellan W, Howard G, Goyal A, Kleinbaum D, Goodman M, Prince V, Muntner P, McClure LA, McClellan A, Judd S. Geographic region and racial variations in polypharmacy in the United States. *Ann. Epidemiol.* 2015; 25: 433–438.e1.

31. Orueta JF, Nuño-Solinis R, Mateos M, Vergara I, Grandes G, Esnaola S. Predictive risk modelling in the Spanish population: a cross-sectional study. *BMC Health Serv. Res.* 2013; 13: 269.

32. Glynn LG, Valderas JM, Healy P, Burke E, Newell J, Gillespie P, Murphy AW. The prevalence of multimorbidity in primary care and its effect on health care utilization and cost. *Fam. Pract.* 2011; 28: 516–523.

33. Pymont C, Butterworth P. Longitudinal cohort study describing persistent frequent attenders in Australian primary healthcare. *BMJ Open* British Medical Journal Publishing

Group; 2015; 5: e008975.

34. Lehnert T, Heider D, Leicht H, Heinrich S, Corrieri S, Luppa M, Riedel-Heller S, König H-H. Review: Health Care Utilization and Costs of Elderly Persons With Multiple Chronic Conditions. *Med. Care Res. Rev.* 2011; 68: 387–420.

35. McPhail SM. Multimorbidity in chronic disease: Impact on health care resources and costs. Risk Manag. Healthc. Policy Dove Press; 2016. p. 143–156.

36. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 2011; 12: 56–68.

37. Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. *Ann. Fam. Med.* American Academy of Family Physicians; 2009; 7: 357–363.

38. Thomas MC, Brownlee M, Susztak K, Sharma K, Jandeleit-Dahm KAM, Zoungas S, Rossing P, Groop P-H, Cooper ME. Diabetic kidney disease. *Nat. Rev. Dis. Prim.* 2015; 1: 15018.

39. Mannino DM, Doherty DE, Sonia Buist A. Global Initiative on Obstructive Lung Disease (GOLD) classification of lung disease and mortality: findings from the Atherosclerosis Risk in Communities (ARIC) study. *Respir. Med.* Elsevier; 2006; 100: 115–122.

40. Agustí A, Vestbo J. Current Controversies and Future Perspectives in Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* 2011; 184: 507–513.

41. Choudhury G, Rabinovich R, MacNee W. Comorbidities and Systemic Effects of Chronic Obstructive Pulmonary Disease. *Clin. Chest Med.* 2014; 35: 101–130.

42. Wheelock CE, Goss VM, Balgoma D, Nicholas B, Brandsma J, Skipp PJ, Snowden S, Burg D, D'Amico A, Horvath I, Chaiboonchoe A, Ahmed H, Ballereau S, Rossios C, Chung KF, Montuschi P, Fowler SJ, Adcock IM, Postle AD, Dahlén S-E, Rowe A, Sterk PJ, Auffray C, Djukanovic R. Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. *Eur. Respir. J.* 2013; 42: 802–825.

43. Hu JX, Thomas CE, Brunak S. Network biology concepts in complex disease comorbidities. *Nat. Rev. Genet.* 2016; 17: 615–629.

44. World Health Organization. Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases The six objectives of the 2008-2013. *Geneva World Heal. Organ.* 2008; : 48.

45. Wang H, Dwyer-Lindgren L, Lofgren KT, Rajaratnam JK, Marcus JR, Levin-Rector A, Levitz CE, Lopez AD, Murray CJ. Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380: 2071–2094.

46. Khakban A, Sin DD, FitzGerald JM, McManus B, Ng R, Hollander Z, Sadatsafavi M. The Projected Epidemic of COPD Hospitalizations Over the Next 15 Years: A Population Based Perspective. *Am. J. Respir. Crit. Care Med.* American Thoracic Society; 2016; : rccm.201606-1162PP.

47. McLean S, Hoogendoorn M, Hoogenveen RT, Feenstra TL, Wild S, Simpson CR, Mölken MR, Sheikh A. Projecting the COPD population and costs in England and Scotland: 2011 to 2030. *Sci. Rep.* 2016; 6: 31893.

48. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. Am. J. Respir. Crit. Care Med. American Thoracic Society; 2013. p. 347–365.

49. Fabbri LM, Luppi F, Beghé B, Rabe KF. Complex chronic comorbidities of COPD. *Eur. Respir. J.* European Respiratory Society; 2008; 31: 204–212.

50. Vanfleteren LEGW, Spruit MA, Groenen M, Gaffron S, van Empel VPM, Bruijnzeel PLB, Rutten EPA, Op 't Roodt J, Wouters EFM, Franssen FME. Clusters of comorbidities based

on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 2013; 187: 728–735.

51. Divo M, Cote C, De Torres JP, Casanova C, Marin JM, Pinto-Plata V, Zulueta J, Cabrera C, Zagaceta J, Hunninghake G, Celli B. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 2012/05/09. 2012; 186: 155–161.

52. Garcia-Aymerich J, Gómez FP, Benet M, Farrero E, Basagaña X, Gayete A, Paré C, Freixa X, Ferrer J, Ferrer A, Roca J, Gáldiz JB, Sauleda J, Monsó E, Gea J, Barberà J a, Agustí A, Antó JM. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. *Thorax* 2010; 66: 430–437.

53. Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, Zulueta J, Cabrera C, Zagaceta J, Hunninghake G, Celli B. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012; 186: 155–161.

54. Fabbri LM, Beghé B, Agustí A. COPD and the Solar System. *Am. J. Respir. Crit. Care Med.* 2012; 186: 117–119.

55. Pavord ID, Wardlaw AJ. The A to E of airway disease. *Clin. Exp. Allergy* 2009; 40: 62–67.

56. Celli BR, Locantore N, Tal-Singer R, Riley J, Miller B, Vestbo J, Yates JC, Silverman EK, Owen CA, Divo M, Pinto-Plata V, Wouters EFM, Faner R, Agusti A, ECLIPSE Study Investigators. Emphysema and extrapulmonary tissue loss in COPD: a multi-organ loss of tissue phenotype. *Eur. Respir. J.* 2018; 51: 1702146.

57. Rodriguez-Roisin R, Rabe KF, Vestbo J, Vogelmeier C, Agustí A. Global Initiative for Chronic Obstructive Lung Disease (GOLD) 20th Anniversary: a brief history of time. *Eur. Respir. J.* 2017; 50.

58. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS, GOLD Scientific Committee. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am. J. Respir. Crit. Care Med.* 2001; 163: 1256–1276.

59. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, van Weel C, Zielinski J, Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* 2007; 176: 532–555.

60. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, Chen R, Decramer M, Fabbri LM, Frith P, Halpin DMG, López Varela MV, Nishimura M, Roche N, Rodriguez-Roisin R, Sin DD, Singh D, Stockley R, Vestbo J, Wedzicha JA, Agustí A. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am. J. Respir. Crit. Care Med.* American Thoracic Society; 2017; 195: 557–582.

61. Roca J, Vargas C, Cano I, Selivanov V, Barreiro E, Maier D, Falciani F, Wagner P, Cascante M, Garcia-Aymerich J, Kalko S, De Mas I, Tegnér J, Escarrabill J, Agustí A, Gomez-Cabrero D. Chronic Obstructive Pulmonary Disease heterogeneity: challenges for health risk assessment, stratification and management. *J. Transl. Med.* BioMed Central; 2014; 12: S3.

62. Singh D, Roche N, Halpin D, Agusti A, Wedzicha JA, Martinez FJ. Current Controversies in the Pharmacological Treatment of Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* 2016; 194: 541–549.

63. Vestbo J, Agusti A, Wouters EF, Bakke P, Calverley PM, Celli B, Coxson H, Crim C, Edwards LD, Locantore N, Lomas DA, MacNee W, Miller B, Rennard SI, Silverman EK, Yates JC, Tal-Singer R. Should we view chronic obstructive pulmonary disease differently after ECLIPSE? A clinical perspective from the study team. *Am J Respir Crit Care Med* 2014/02/21. 2014; 189: 1022–1030.

64. Singh D, Barnes PJ, Stockley R, Lopez Valera MV, Vogelmeier C, Agusti A. Pharmacological treatment of COPD: the devil is always in the detail. *Eur. Respir. J.* European Respiratory Society; 2018; 51: 1800263.

65. Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *Eur. Respir. Rev.* European Respiratory Society; 2017; 26: 160061.

66. Kahnert K, Alter P, Young D, Lucke T, Heinrich J, Huber RM, Behr J, Wacker M, Biertz F, Watz H, Bals R, Welte T, Wirtz H, Herth F, Vestbo J, Wouters EF, Vogelmeier CF, Jörres RA. The revised GOLD 2017 COPD categorization in relation to comorbidities. *Respir. Med.* 2018; 134: 79–85.

67. Celli BR, Cote CG, Marin JM, Casanova C, Montes de Oca M, Mendez RA, Pinto Plata V, Cabral HJ. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 2004; 350: 1005–1012.

68. Puhan MA, Garcia-Aymerich J, Frey M, ter Riet G, Antó JM, Agustí AG, Gómez FP, Rodríguez-Roisín R, Moons KGM, Kessels AG, Held U. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet (London, England)* Elsevier; 2009; 374: 704–711.

69. Puhan MA, Hansel NN, Sobradillo P, Enright P, Lange P, Hickson D, Menezes AM, ter Riet G, Held U, Domingo-Salvany A, Mosenifar Z, Antó JM, Moons KGM, Kessels A, Garcia-Aymerich J, International COPD Cohorts Collaboration Working Group for the ICCCW. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open* British Medical Journal Publishing Group; 2012; 2: e002152.

70. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.* 1987; 40: 373–383.

71. Hughes JS, Averill RF, Eisenhandler J, Goldfield NI, Muldoon J, Neff JM, Gay JC. Clinical Risk Groups (CRGs). *Med. Care* 2004; 42: 81–90.

72. Monterde D, Vela E, Clèries M. Los grupos de morbilidad ajustados: nuevo agrupador de morbilidad poblacional de utilidad en el ámbito de la atención primaria. *Atención Primaria* Elsevier; 2016; 48: 674–682.

73. Dueñas-Espín I, Vela E, Pauws S, Bescos C, Cano I, Cleries M, Contel JC, de Manuel Keenoy E, Garcia-Aymerich J, Gomez-Cabrero D, Kaye R, Lahr MMH, Lluch-Ariet M, Moharra M, Monterde D, Mora J, Nalin M, Pavlickova A, Piera J, Ponce S, Santaeugenia S, Schonenberg H, Störk S, Tegner J, Velickovski F, Westerteicher C, Roca J. Proposals for enhanced health risk assessment and stratification in an integrated care scenario. *BMJ Open* British Medical Journal Publishing Group; 2016; 6: e010301.

74. Regan EA, Lynch DA, Curran-Everett D, Curtis JL, Austin JHM, Grenier PA, Kauczor H-U, Bailey WC, DeMeo DL, Casaburi RH, Friedman P, Van Beek EJR, Hokanson JE, Bowler RP, Beaty TH, Washko GR, Han MK, Kim V, Kim SS, Yagihashi K, Washington L, McEvoy CE, Tanner C, Mannino DM, Make BJ, Silverman EK, Crapo JD, Genetic Epidemiology of COPD (COPDGene) Investigators. Clinical and Radiologic Disease in Smokers With Normal Spirometry. *JAMA Intern. Med.* 2015; 175: 1539.

75. Fabbri LM. Smoking, Not COPD, as the Disease. *N. Engl. J. Med.* Massachusetts Medical Society; 2016; 374: 1885–1886.

76. Lococo F, Cesario A, Del Bufalo A, Ciarrocchi A, Prinzi G, Mina M, Bonassi S, Russo P. Novel therapeutic strategy in the management of COPD: a systems medicine approach. *Curr. Med. Chem.* 2015; 22: 3655–3675.

77. Wardlaw AJ, Silverman M, Siva R, Pavord ID, Green R. Multi-dimensional phenotyping: Towards a new taxonomy for airway disease. Clin. Exp. Allergy 2005. p. 1254–1262.

78. Agusti A. Systemic Effects of Chronic Obstructive Pulmonary Disease: What We Know and What We Don't Know (but Should). *Proc. Am. Thorac. Soc.* 2007; 4: 522–525.

79. Gea J, Barreiro E, Orozco-Levi M. Systemic Inflammation in COPD. Clin. Pulm. Med. 2009. p. 233–242.

80. Sevenoaks MJ, Stockley RA. Chronic Obstructive Pulmonary Disease, inflammation and co-morbidity – a common inflammatory phenotype? *Respir. Res.* BioMed Central; 2006; 7: 70.

81. Van Eeden S, Leipsic J, Paul Man SF, Sin DD. The Relationship between Lung Inflammation and Cardiovascular Disease. *Am. J. Respir. Crit. Care Med.* 2012; 186: 11–16.

82. Agusti A, Calverley PM, Celli B, Coxson HO, Edwards LD, Lomas DA, MacNee W, Miller BE, Rennard S, Silverman EK, Tal-Singer R, Wouters E, Yates JC, Vestbo J, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) investigators. Characterisation of COPD heterogeneity in the ECLIPSE cohort. *Respir. Res.* 2010; 11: 122.

83. Mohan D, Forman JR, Allinder M, McEniery CM, Bolton CE, Cockcroft JR, MacNee W, Fuld J, Marchong M, Gale NS, Fisk M, Nagarajan S, Cheriyan J, Lomas DA, Calverley PMA, Miller BE, Tal-Singer R, Wilkinson IB, Polkey MI. Fibrinogen does not relate to cardiovascular or muscle manifestations in COPD: cross-sectional data from the ERICA study. *Thorax* BMJ Publishing Group Ltd; 2018; : thoraxjnl-2018-211556.

84. Vanfleteren LEGW, Spruit MA, Groenen MTJ, Bruijnzeel PLB, Taib Z, Rutten EPA, Op 't Roodt J, Akkermans MA, Wouters EFM, Franssen FME. Arterial stiffness in patients with COPD: the role of systemic inflammation and the effects of pulmonary rehabilitation. *Eur. Respir. J.* 2014; 43: 1306–1315.

85. Barnes PJ. Mechanisms of development of multimorbidity in the elderly. *Eur. Respir. J.* 2015; 45: 790–806.

86. Poole DC, Hirai DM, Copp SW, Musch TI. Muscle oxygen transport and utilization in heart failure: implications for exercise (in)tolerance. *Am J Physiol Hear. Circ Physiol* 2011/11/22. 2012; 302: H1050-63.

87. Rennard SI. COPD Heterogeneity: What This Will Mean in Practice. *Respir. Care* 2011; 56: 1181–1187.

88. Chen X, Xu X, Xiao F. Heterogeneity of chronic obstructive pulmonary disease: from phenotype to genotype. *Front. Med.* 2013; 7: 425–432.

89. Motegi T, Jones RC, Ishii T, Hattori K, Kusunoki Y, Yamada R, Gemma K, Kida K, Furutate R. A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations. *Int. J. Chron. Obstruct. Pulmon. Dis.* 2013; 8: 259.

90. Synergy-COPD consortium. Modelling and simulation environment for systems medicine: Chronic obstructive pulmonary disease (COPD) as a use case, FP7-ICT-270086. 2010. Available from: http://www.synergy-copd.eu/. 2010; .

91. Miralles F, Gomez-Cabrero D, Lluch-Ariet M, Tegnér J, Cascante M, Roca J, Synergy-COPD consortium S-C. Predictive medicine: outcomes, challenges and opportunities in the Synergy-COPD project. *J. Transl. Med.* BioMed Central; 2014; 12 Suppl 2: S12.

92. Roca J, Cano I, Gomez-Cabrero D, Tegnér J. From systems understanding to personalized medicine: Lessons and recommendations based on a multidisciplinary and translational analysis of COPD. *Methods Mol. Biol.* 2016. p. 283–303.

93. Gomez-Cabrero D, Lluch-Ariet M, Tegnér J, Cascante M, Miralles F, Roca J, Synergy-COPD consortium the S-C. Synergy-COPD: a systems approach for understanding and managing chronic diseases. *J. Transl. Med.* BioMed Central; 2014; 12 Suppl 2: S2.

94. Cano I, Tényi Á, Schueller C, Wolff M, Huertas Migueláñez MM, Gomez-Cabrero D, Antczak P, Roca J, Cascante M, Falciani F, Maier D. The COPD Knowledge Base: enabling data analysis and computational simulation in translational COPD research. *J. Transl. Med.* 2014; 12 Suppl 2: S6.

95. Chmiel A, Klimek P, Thurner S. Spreading of diseases through comorbidity networks

across life and gender. *New J. Phys.* 2014; 16: 115013.

96. Beck MK, Jensen AB, Nielsen AB, Perner A, Moseley PL, Brunak S. Diagnosis trajectories of prior multi-morbidity predict sepsis mortality. *Sci. Rep.* Nature Publishing Group; 2016; 6: 36624.

97. Le Novère N. Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* Nature Publishing Group; 2015; 16: 146–158.

98. Viceconti M, Hunter P. The Virtual Physiological Human: Ten Years After. *Annu. Rev. Biomed. Eng.* 2016; 18: 103–123.

99. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012; 40: D109–D114.

100. Vastrik I, D'Eustachio P, Schmidt E, Joshi-Tope G, Gopinath G, Croft D, de Bono B, Gillespie M, Jassal B, Lewis S, Matthews L, Wu G, Birney E, Stein L. Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.* 2007; 8: R39.

101. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, Mélius J, Cirillo E, Coort SL, Digles D, Ehrhart F, Giesbertz P, Kalafati M, Martens M, Miller R, Nishida K, Rieswijk L, Waagmeester A, Eijssen LMT, Evelo CT, Pico AR, Willighagen EL. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* 2018; 46: D661–D667.

102. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013; 14: 128.

103. Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016; 44: W90–W97.

104. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37: 1–13.

105. Kitsak M, Sharma A, Menche J, Guney E, Ghiassian SD, Loscalzo J, Barabási A-L. Tissue Specificity of Human Disease Module. *Sci. Rep.* Nature Publishing Group; 2016; 6: 35241.

106. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, Zhang R, Hartmann BM, Zaslavsky E, Sealfon SC, Chasman DI, FitzGerald GA, Dolinski K, Grosser T, Troyanskaya OG. Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* 2015; 47: 569–576.

107. Kirouac DC, Saez-Rodriguez J, Swantek J, Burke JM, Lauffenburger DA, Sorger PK. Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks. *BMC Syst. Biol.* 2012; 6: 29.

108. Natarajan M, Lin K-M, Hsueh RC, Sternweis PC, Ranganathan R. A global analysis of cross-talk in a mammalian cellular signalling network. *Nat. Cell Biol.* 2006; 8: 571–580.

109. Kuperstein I, Bonnet E, Nguyen H-A, Cohen D, Viara E, Grieco L, Fourquet S, Calzone L, Russo C, Kondratova M, Dutreix M, Barillot E, Zinovyev A. Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis* Nature Publishing Group; 2015; 4: e160–e160.

110. Mizuno S, Iijima R, Ogishima S, Kikuchi M, Matsuoka Y, Ghosh S, Miyamoto T, Miyashita A, Kuwano R, Tanaka H. AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. *BMC Syst. Biol.* 2012; 6: 52.

111. Ogishima S, Mizuno S, Kikuchi M, Miyashita A, Kuwano R, Tanaka H, Nakaya J. AlzPathway, an Updated Map of Curated Signaling Pathways: Towards Deciphering Alzheimer's Disease Pathogenesis. *Methods Mol. Biol.* 2016. p. 423–432.

112. Matsuoka Y, Matsumae H, Katoh M, Eisfeld AJ, Neumann G, Hase T, Ghosh S, Shoemaker JE, Lopes TJS, Watanabe T, Watanabe S, Fukuyama S, Kitano H, Kawaoka Y. A comprehensive map of the influenza A virus replication cycle. *BMC Syst. Biol.* BioMed Central; 2013; 7: 97.

113. Loscalzo J, Barabasi A-L. Systems biology and the future of medicine. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2011; 3: 619–627.

114. Diez D, Agustí A, Wheelock CE. Network Analysis in the Investigation of Chronic Respiratory Diseases. From Basics to Application. *Am. J. Respir. Crit. Care Med.* 2014; 190: 981–988.

115. Mitra K, Carvunis A-R, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. *Nat. Rev. Genet.* 2013; 14: 719–732.

116. Carter H, Hofree M, Ideker T. Genotype to phenotype via network analysis. *Curr. Opin. Genet. Dev.* 2013; 23: 611–621.

117. Yu D, Kim M, Xiao G, Hwang TH. Review of Biological Network Data and Its Applications. *Genomics Inform.* 2013; 11: 200–210.

118. Vidal M, Cusick ME, Barabási A-L, Shi Z, Berger MF, Newburger DE, Saulrieta K, Smith Z, Shah MV, Radhakrishnan M, Al. E. Interactome networks and human disease. *Cell* Oxford University Press, New York; 2011; 144: 986–998.

119. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási A-L. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 2015; 347: 1257601.

120. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature* 1999; 402: C47–C52.

121. Alon U. Biological networks: the tinkerer as an engineer. *Science* 2003; 301: 1866–1867.

122. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 2004; 5: 101–113.

123. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang L V, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005; 437: 1173–1178.

124. Rolland T, Taşan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly MP, Carvunis A-R, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan C, Franzosa E, Gebreab F, Gutierrez BJ, et al. A Proteome-Scale Map of the Human Interactome Network. *Cell* 2014; 159: 1212–1226.

125. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 2002; 18: S233–S240.

126. Sharma A, Menche J, Huang CC, Ort T, Zhou X, Kitsak M, Sahni N, Thibault D, Voung L, Guo F, Ghiassian SD, Gulbahce N, Baribaud F, Tocker J, Dobrin R, Barnathan E, Liu H, Panettieri RA, Tantisira KG, Qiu W, Raby BA, Silverman EK, Vidal M, Weiss ST, Barabási A-L. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum. Mol. Genet.* 2015; 24: 3005–3020.

127. Chen B, Fan W, Liu J, Wu F-X. Identifying protein complexes and functional modules--from static PPI networks to dynamic PPI networks. *Brief. Bioinform.* 2014; 15: 177–194.

128. Segal E, Wang H, Koller D. Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics* 2003; 19 Suppl 1: i264–71.

129. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HYK, Chen R, Miriami E, Karczewski KJ,

Hariharan M, Dewey FE, Cheng Y, Clark MJ, Im H, Habegger L, Balasubramanian S, O'Huallachain M, Dudley JT, Hillenmeyer S, Haraksingh R, Sharon D, Euskirchen G, Lacroute P, Bettinger K, Boyle AP, Kasowski M, Grubert F, Seki S, Garcia M, Whirl-Carrillo M, Gallardo M, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* NIH Public Access; 2012; 148: 1293–1307.

130. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013; 499: 43–49.

131. Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge J V, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, Lawrence MS, Gonzalez-Perez A, Tamborero D, Cheng Y, Ryslik GA, Lopez-Bigas N, Getz G, Ding L, Raphael BJ. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* 2014; 47: 106–114.

132. Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, Davis C, Wheeler DA, Murray BA, Schmidt L, Vocke CD, Peto M, Al Mamun AAM, Shinbrot E, Sethi A, Brooks S, Rathmell WK, Brooks AN, Hoadley KA, Robertson AG, Brooks D, Bowlby R, Sadeghi S, Shen H, Weisenberger DJ, Bootwalla M, Baylin SB, Laird PW, Cherniack AD, Saksena G, Haake S, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N. Engl. J. Med.* Massachusetts Medical Society; 2015; .

133. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. In: Smith SM, editor. *Cochrane Database Syst. Rev.* Chichester, UK: John Wiley & Sons, Ltd; 2016. p. CD006560.

134. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S. Risk prediction models for hospital readmission: a systematic review. *JAMA* NIH Public Access; 2011; 306: 1688–1698.

135. Roland M, Abel G. Reducing emergency admissions: are we on the right track? *BMJ* 2012; 345.

136. Johnson TL, Rinehart DJ, Durfee J, Brewer D, Batal H, Blum J, Oronce CI, Melinkovich P, Gabow P. For Many Patients Who Use Large Amounts Of Health Care Services, The Need Is Intense Yet Temporary. *Health Aff.* 2015; 34: 1312–1319.

137. Lewis G, Curry N, Bardsley M. Choosing a predictive risk model : a guide for commissioners in England. 2011; : 20.

138. NIGB National Information Governance Board. Advice on Risk Prediction and Stratification Activities. 2012 p. 1–10.

139. Lewis GH. "Impactibility models": Identifying the subgroup of high-risk patients most amenable to hospital-avoidance programs. *Milbank Q.* 2010; 88: 240–255.

140. Orkin FK. Risk stratification, risk adjustment, and other risks. *Anesthesiology* 2010; 113: 1001–1003.

141. JW T. Risk Adjustment for Measuring Health Care Outcomes, 3rd edition. *Int. J. Qual. Heal. Care* 2004; 16: 181.

142. Gomez-Cabrero D, Menche J, Vargas C, Cano I, Maier D, Barabási A-L, Tegnér J, Roca J. From comorbidities of chronic obstructive pulmonary disease to identification of shared molecular mechanisms by data integration. *BMC Bioinformatics* BioMed Central; 2016; 17: 23–35.

143. Ware JH. The limitations of risk factors as prognostic tools. *N. Engl. J. Med.* 2006; 355: 2615–2617.

144. Mora J, Iturralde MD, Prieto L, Domingo C, Gagnon M-P, Martínez-Carazo C, March AG, De Massari D, Martí T, Nalin M, Avolio F, Bousquet J, Keenoy E de M, all ASSEHS group. Key aspects related to implementation of risk stratification in health care systems-the ASSEHS study. *BMC Health Serv. Res.* 2017; 17: 331.

145. Keenoy E de M, Nalin M, Alhambra T, Avolio F, Baroni I, Bedbrook A, Barbara Branchini, Segura JCC, Iturralde MD, Massari D De, Rico CD, Erreguerena I, Sukia SE, Garces J, March AG, López F, Lepore V, Linkeviciute A, Tino M, Carazo CM, Ortuondo EM, Mora J, Mendia JFO, Pauws S, Prieto L, Robusto F, Ródenas F, Romano M, Segui L, Bousquet J. White paper on Deployment of Stratification Methods. The ASSEHS project. 2016.

146. Cano I, Selivanov V, Gomez-Cabrero D, Tegnér J, Roca J, Wagner PD, Cascante M. Oxygen pathway modeling estimates high reactive oxygen species production above the highest permanent human habitation. *PLoS One* Public Library of Science; 2014; 9: e111068.

147. Marín de Mas I, Fanchon E, Papp B, Kalko S, Roca J, Cascante M. Molecular mechanisms underlying COPD-muscle dysfunction unveiled through a systems medicine approach. *Bioinformatics* Oxford University Press; 2017; 33: 95–103.

148. Davidsen PK, Herbert JM, Antczak P, Clarke K, Ferrer E, Peinado VI, Gonzalez C, Roca J, Egginton S, Barberá JA, Falciani F. A systems biology approach reveals a link between systemic cytokines and skeletal muscle energy metabolism in a rodent smoking model and human COPD. *Genome Med*. 2014; 6: 59.

149. Gillis J, Ballouz S, Pavlidis P. Bias tradeoffs in the creation and analysis of protein-protein interaction networks. *J. Proteomics* 2014; 100: 44–54.

150. Ben-Hamo R, Gidoni M, Efroni S. PhenoNet: identification of key networks associated with disease phenotype. *Bioinformatics* 2014; 30: 2399–2405.

151. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. *Nat. Methods* Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013; 10: 1108–1115.

152. Langfelder P, Horvath S, Fisher R, Zhou X, Kao M, Wong W, Steffen M, Petti A, Aach J, D'haeseleer P, Church G, Stuart J, Segal E, Koller D, Kim S, Zhang B, Horvath S, Carey V, Gentry J, Whalen E, Gentleman R, Schaefer J, Strimmer K, Chuang C, Jen C, Chen C, Shieh G, Cokus S, Rose S, Haynor D, et al. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* BioMed Central; 2008; 9: 559.

153. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, Curnow SJ, Rodriguez D a, Cascante M, O'Neill L, Egginton S, Roca J, Falciani F. A systems biology approach identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease. *PLoS Comput. Biol.* 2011; 7: e1002129.

154. Menche J, Guney E, Sharma A, Branigan PJ, Loza MJ, Baribaud F, Dobrin R, Barabási A-L. Integrating personalized gene expression profiles into predictive disease-associated gene pools. *npj Syst. Biol. Appl.* Nature Publishing Group; 2017; 3: 10.

155. Li C-X, Wheelock CE, Sköld CM, Wheelock ÅM. Integration of multi-omics datasets enables molecular classification of COPD. *Eur. Respir. J.* European Respiratory Society; 2018; : 1701930.

156. Giannoula A, Gutierrez-Sacristán A, Bravo Á, Sanz F, Furlong LI. Identifying temporal patterns in patient disease trajectories using dynamic time warping: A population-based study. *Sci. Rep.* 2018; 8: 4216.

157. Vanfleteren LE, Spruit MA, Groenen M, Gaffron S, van Empel VP, Bruijnzeel PL, Rutten EP, Op 't Roodt J, Wouters EF, Franssen FM. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013/02/09. 2013; 187: 728–735.

158. E. C. Global genomic data-sharing effort kicks off. Nature News. 06 March 2014. *URLhttp//www.nature.com/news/global-genomic-data-sharing-effort-kicks-off-1.14826. Accessed 2014-03-18. (Archived by WebCite® http//www.webcitation.org/6OAELb42a)* 2014; .

159. Bhavnani SP, Narula J, Sengupta PP. Mobile technology and the digitization of healthcare. *Eur. Heart J.* Oxford University Press; 2016; 37: 1428–1438.

160. Weiler A. mHealth and big data will bring meaning and value to patient-reported outcomes. *mHealth* AME Publications; 2016; 2: 2.

161. Guarneri MR, Perego P. Games and Gamification for Healthy Behaviours: The Experience of PEGASO Fit 4 Future. Springer, Cham; 2017. p. 100–109.

162. Guarneri MR, Brocca MD, Piras L. Patient's Empowerment and Behaviour Change: Complementary Approaches in EU Projects PALANTE and PEGASO. Springer, Cham; 2017. p. 359–369.

163. Mesquita R, Spina G, Pitta F, Donaire-Gonzalez D, Deering BM, Patel MS, Mitchell KE, Alison J, van Gestel AJ, Zogg S, Gagnon P, Singh SJ, Hopkinson NS, Miedinger D, Benzo RP, Maltais F, Paggiaro P, McKeough ZJ, Polkey MI, Hill K, D-C Man W, Clarenbach CF, Hernandes NA, Savi D, Wootton S, Furlanetto KC, Cindy Ng LW, Vaes AW, Jenkins C, Eastwood PR, et al. Physical activity patterns and clusters in 1001 patients with COPD. *Chron. Respir. Dis.* SAGE PublicationsSage UK: London, England; 2017; : 147997231668720.

164. Szalma S, Koka V, Khasanova T, Perakslis ED. Effective knowledge management in translational medicine. *J. Transl. Med.* BioMed Central; 2010; 8: 68.

165. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, Bernal-Delgado E, Blomberg N, Bock C, Conesa A, Del Signore S, Delogne C, Devilee P, Di Meglio A, Eijkemans M, Flicek P, Graf N, Grimm V, Guchelaar H-J, Guo Y-K, Gut IG, Hanbury A, Hanif S, Hilgers R-D, Honrado Á, Hose DR, Houwing-Duistermaat J, Hubbard T, Janacek SH, Karanikas H, et al. Making sense of big data in health research: Towards an EU action plan. *Genome Med.* BioMed Central; 2016; 8: 71.

166. Losko S, Heumann K. Semantic Data Integration and Knowledge Management to Represent Biological Network Associations. *Methods Mol. Biol.* 2009. p. 241–258.

167. Canuel V, Rance B, Avillach P, Degoulet P, Burgun A. Translational research platforms integrating clinical and omics data: a review of publicly available solutions. *Brief. Bioinform.* Oxford University Press; 2015; 16: 280–290.

168. Auffray C, Charron D, Hood L. Predictive, preventive, personalized and participatory medicine: back to the future. *Genome Med* 2010; 2: 57.

169. Cano I, Tenyi A, Vela E, Miralles F, Roca J. Perspectives on Big Data applications of health information. *Curr. Opin. Syst. Biol.* 2017; 3: 36–42.

170. Renner A-T, Bobek J, Habl C, Anja L. Study on Big Data in Public Health, Telemedicine and Healthcare. 2016.

171. Alkawaa FM, Chaudhary K, Garmire LX. Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. *J. Proteome Res.* American Chemical Society; 2017; : acs.jproteome.7b00595.

172. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Sci. Rep.* Nature Publishing Group; 2016; 6: 26094.

173. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, Liu PJ, Liu X, Marcus J, Sun M, Sundberg P, Yee H, Zhang K, Zhang Y, Flores G, Duggan GE, Irvine J, Le Q, Litsch K, Mossin A, Tansuwan J, Wang D, Wexler J, Wilson J, Ludwig D, Volchenboum SL, Chou K, Pearson M, Madabushi S, Shah NH, et al. Scalable and accurate deep learning with electronic health records. *npj Digit. Med.* Nature Publishing Group; 2018; 1: 18.

174. Brown S-A. Patient Similarity: Emerging Concepts in Systems and Precision Medicine. *Front. Physiol.* Frontiers Media SA; 2016; 7: 561.

175. Gallego B, Walter SR, Day RO, Dunn AG, Sivaraman V, Shah N, Longhurst CA, Coiera E. Bringing cohort studies to the bedside: framework for a 'green button' to support clinical decision-making. *J. Comp. Eff. Res.* Future Medicine Ltd London, UK; 2015; 4: 191–197.

176. Lee J, Maslove DM, Dubin JA, Jonge E de, Keizer N de, Bruining H. Personalized Mortality Prediction Driven by Electronic Medical Data and a Patient Similarity Metric. Emmert-Streib F, editor. *PLoS One* ACM Press; 2015; 10: e0127428.

177. Ng K, Sun J, Hu J, Wang F. Personalized Predictive Modeling and Risk Factor Identification using Patient Similarity. *AMIA Jt. Summits Transl. Sci. proceedings. AMIA Jt. Summits Transl. Sci.* American Medical Informatics Association; 2015; 2015: 132–136.

178. Panahiazar M, Taslimitehrani V, Pereira NL, Pathak J. Using EHRs for Heart Failure Therapy Recommendation Using Multidimensional Patient Similarity Analytics. *Stud. Health Technol. Inform.* 2015; 210: 369–373.

179. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A. Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* Nature Research; 2014; 11: 333–337.

180. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, Rycroft-Malone J, Meissner P, Murray E, Patel A, Sheikh A, Taylor SJC, StaRI Group. Standards for Reporting Implementation Studies (StaRI): explanation and elaboration document. *BMJ Open* BMJ Publishing Group; 2017; 7: e013318.

181. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. *Br. Med. J.* 2013; 347: 2–7.

182. Cano I, Alonso A, Hernandez C, Burgos F, Barberan-Garcia A, Roldan J, Roca J. An adaptive case management system to support integrated care services: Lessons learned from the NEXES project. *J. Biomed. Inform.* 2015; 55: 11–22.