

Northumbria Research Link

Citation: Gomez-Cabrero, David, Menche, Jörg, Cano, Isaac, Abugessaisa, Imad, Huertas-Migueláñez, Mercedes, Tenyi, Akos, de Mas, Igor, Kiani, Narsis A, Marabita, Francesco, Falciani, Francesco, Burrowes, Kelly, Maier, Dieter, Wagner, Peter, Selivanov, Vitaly, Cascante, Marta, Roca, Josep, Barabási, Albert-László and Tegnér, Jesper (2014) Systems Medicine: from molecular features and models to the clinic in COPD. *Journal of Translational Medicine*, 12 (Suppl2). S4. ISSN 1479-5876

Published by: BioMed Central

URL: <https://doi.org/10.1186/1479-5876-12-S2-S4> <<https://doi.org/10.1186/1479-5876-12-S2-S4>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/44683/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



RESEARCH

Open Access

Systems Medicine: from molecular features and models to the clinic in COPD

David Gomez-Cabrero^{1†}, Jörg Menche^{2,3†}, Isaac Cano⁴, Imad Abugessaisa¹, Mercedes Huertas-Migueláñez⁵, Akos Tenyi⁶, Igor Marin de Mas⁶, Narsis A. Kiani¹, Francesco Marabita¹, Francesco Falciani⁷, Kelly Burrowes⁸, Dieter Maier⁹, Peter Wagner¹⁰, Vitaly Selivanov⁶, Marta Cascante⁶, Josep Roca⁴, Albert-László Barabási^{2,3,11,12,13*†}, Jesper Tegnér^{1*†}

Abstract

Background and hypothesis: *Chronic Obstructive Pulmonary Disease* (COPD) patients are characterized by heterogeneous clinical manifestations and patterns of disease progression. Two major factors that can be used to identify COPD subtypes are muscle dysfunction/wasting and co-morbidity patterns. We hypothesized that COPD heterogeneity is in part the result of complex interactions between several genes and pathways. We explored the possibility of using a Systems Medicine approach to identify such pathways, as well as to generate predictive computational models that may be used in clinic practice.

Objective and method: Our overarching goal is to generate clinically applicable predictive models that characterize COPD heterogeneity through a Systems Medicine approach. To this end we have developed a general framework, consisting of three steps/objectives: (1) feature identification, (2) model generation and statistical validation, and (3) application and validation of the predictive models in the clinical scenario. We used muscle dysfunction and co-morbidity as test cases for this framework.

Results: In the study of muscle wasting we identified relevant features (genes) by a network analysis and generated predictive models that integrate mechanistic and probabilistic models. This allowed us to characterize muscle wasting as a general de-regulation of pathway interactions. In the co-morbidity analysis we identified relevant features (genes/pathways) by the integration of gene-disease and disease-disease associations. We further present a detailed characterization of co-morbidities in COPD patients that was implemented into a predictive model. In both use cases we were able to achieve predictive modeling but we also identified several key challenges, the most pressing being the validation and implementation into actual clinical practice.

Conclusions: The results confirm the potential of the Systems Medicine approach to study complex diseases and generate clinically relevant predictive models. Our study also highlights important obstacles and bottlenecks for such approaches (e.g. data availability and normalization of frameworks among others) and suggests specific proposals to overcome them.

* Correspondence: alb@neu.edu; jesper.tegner@ki.se

† Contributed equally

¹Unit of computational Medicine, Center for Molecular Medicine, Department of Medicine, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden

²Center for Complex Network Research, Northeastern University Physics Department, Boston, MA 02115, USA

Full list of author information is available at the end of the article

Introduction

Recent years have seen a paradigm shift in Life Sciences: “from a fragmented to a systems approach, linear to non-linear methodology and from genome to physiome based analysis” [1]. Systems Medicine, as an adaptation and extension of *Systems Biology*, embraces this paradigm and is becoming a cornerstone in the study of complex diseases. A general introduction to Systems Medicine is provided in [2]. In this article, part of a Supplement dedicated to the Synergy-COPD project [2], we review and assess the Synergy-COPD’s Systems Medicine approach to study Chronic Obstructive Pulmonary Disease (COPD), both a chronic and a complex disease.

While the characterization of COPD has been extensively investigated and there is a continuous refinement of guidelines (e.g. GOLD), there is yet no consensus on a phenotypic definition of the term “COPD patient”. For instance in [3] several sub-types of COPD patients were identified; see also [4] within this Supplement for further details on the heterogeneity in COPD. Briefly, within Synergy-COPD we aim to characterize two sources of heterogeneity: first we investigated the systemic effects associated with skeletal muscle dysfunction in COPD patients (*MusclDYS*). Second, we aimed to characterize co-morbidity patterns of COPD patients (*CoMorb*). Finally, we also investigated the interplay between the different heterogeneities, which may provide a novel description of COPD as being driven by the interaction between several factors.

Using COPD as a case-study we explore the notion that Systems Medicine provides tools to investigate and characterize disease heterogeneity. To this end our analyses follow a general three-step procedure: (1) first, we need to identify the relevant biomarkers (or more generally *features of interest, FoI*) for each case of heterogeneity; (2) in a second step, predictive models with the potential to be applied in the clinic are developed and validated statistically. Third and final, (3) the usefulness of the models in a clinical scenario has to be validated. To achieve this goal we integrated a wide variety of available resources, such as prior domain knowledge, relevant data-bases and existing probabilistic and mechanistic models.

The rest of this article is organized as follows: the next section details the Systems Medicine framework used in Synergy-COPD. The third and fourth sections describe the application of this framework in the characterization of *MusclDYS* and *CoMorb*. These sections include a brief description of the questions, the obtained results and the limitations of the proposed methodology. The final section provides the conclusions and summarizes the identified remaining challenges.

The Synergy-COPD’s systems medicine approach

Systems Medicine provides a comprehensive and general framework to investigate the complex interactions

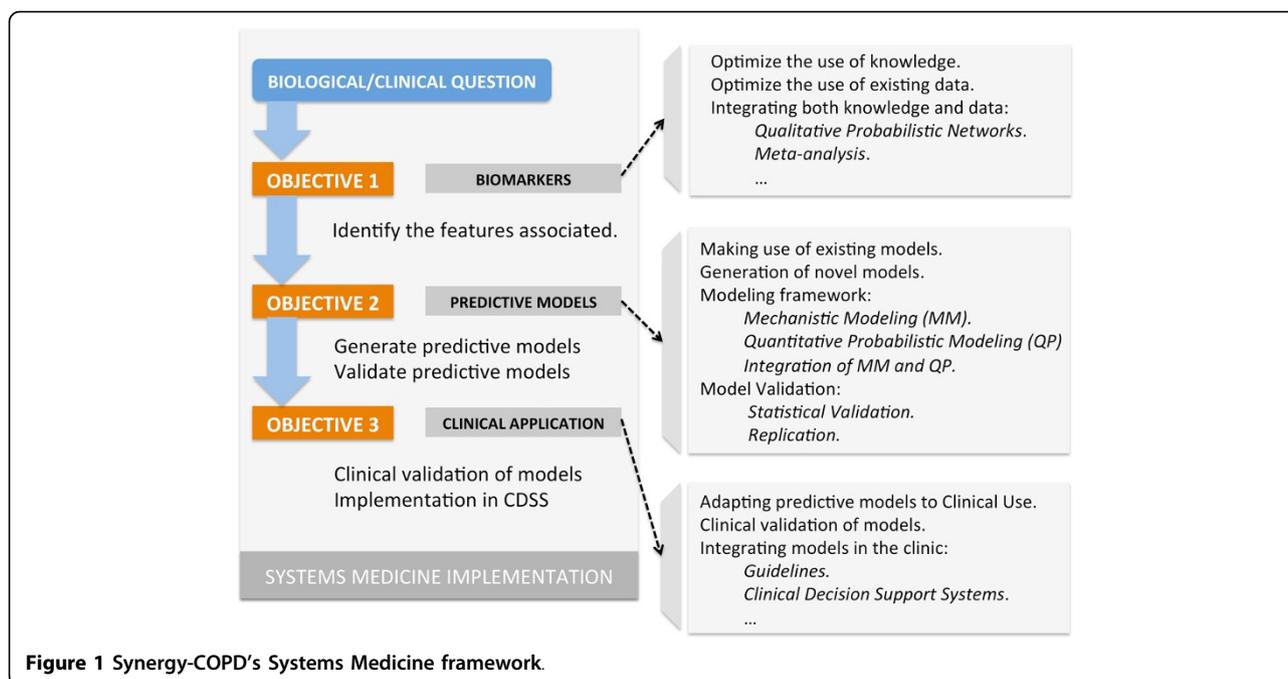
implicated in human disease in an integrated fashion. Consequently, there is no single defined set of methodologies associated with Systems Medicine. Instead, any methodology useful to investigate the question under study “as a system” can be considered as relevant to explore and validate. While the concrete focus of Synergy-COPD lay on COPD, we aimed at developing a more general framework that may also be applicable to other complex diseases. We therefore started by defining a generic three-step objective plan that sets the goals in our studies of *MusclDys* and *CoMorb* (Figure 1). The plan was then concretized and adapted to each question accordingly.

Our final goal was a characterization of the disease heterogeneities that can be transferred to clinical practice (Figure 1, Objective 3), in particular by implementing it into a Clinical Decision Support system (CDSS, [5]). The first step (Objective 1) towards this goal is to identify the relevant, *i.e.* most predictive, biological features among the large amount of available data, *e.g.* genes, metabolites and clinical variables, among many others. The second step (Objective 2) is to integrate these features into predictive models and to validate them. In the following we briefly review each objective for the two case studies *MusclDys* and *CoMorb* and introduce the respective different resources and methodologies that were used.

Objective 1, (Biomarker identification): having defined a question of interest (*e.g.* *MusclDys*) we first need to identify the relevant associated features. The core of this Objective is formed by publicly available data-sets and knowledge (*e.g.* Gene Omnibus [6]) that were integrated into a user-friendly knowledge-base [7]. The different methodologies for *MusclDys* and *CoMorb* are detailed in two separate sections.

Objective 2 (Predictive modeling): the identified features are now used in predictive models that may provide insights into the question of interest. For instance, in *MusclDys* we aim to predict the effects of muscle dysfunction in a given patient. In *CoMorb* we aim to compute the probability of developing a specific co-morbidity in COPD patients. Those quantitative models are question-specific and require both statistical (*e.g.* through cross-validation [8]) and biological validation.

Objective 3 (Clinical application): bridging the gap between a predictive model and its use in clinical practice constitutes an important and challenging task: Beyond the basic statistical and biological validation of a model, it also needs to be clinically relevant in the context of personalized medicine. In this objective, predictive models are reviewed for their possible uses in a CDSS. Once a model is considered useful in principle, both a thorough clinical validation and an optimal CDSS implementation are required.



Understanding COPD skeletal muscle dysfunction (*MusclDys*) through systems medicine

Objective 1: Biomarker identification

To identify the relevant features associated with muscle dysfunction/wasting we used existing data and knowledge through (existing and novel) network-based methodologies. We considered the Biobridge clinical study [2] as the core of the data and extended it through publicly available data-sets from GEO [6]. Among the most relevant data-sets are the gene expression profiling of sputum in COPD ex-smokers (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22148>) and Peripheral Blood Mononuclear Cell (PBMC) profiling of COPD patients by COPDgene [9]. For those data-sets used but not publicly available we had permission to access and analyze the data. Next we describe the interactome-based methodologies and results.

The Interactome

The etiology of COPD involves a multitude of intertwined molecular processes, many of which still remain unknown. These processes are embedded in the larger context of the *Interactome*, referring to a single comprehensive network integrating all molecular interactions, such as protein-protein interactions, regulatory protein-DNA interactions or metabolic interactions (see Figure 2). While ongoing efforts to systematically map the complete *Interactome* stand only at the beginning, currently available databases (Table 1) already include several hundreds of thousands of interactions. In order to explore such large networks, Systems Medicine has extensively adopted tools from network science [10-12].

Network approaches to human disease are based on the observation that the cellular components associated with a specific disease are not scattered randomly within the *Interactome*, but segregate in certain neighborhoods or *disease modules*. The identification of the specific disease modules is therefore an important step towards a holistic understanding of how molecular variations with small isolated effect sizes collectively give rise to a certain disease phenotype. The local agglomeration of disease-associated proteins within the *Interactome* can be used in this process, by extrapolating from the connectivity patterns of known disease-associated proteins to infer novel disease proteins [13,14]. Other applications of this principle include the identification of pathway members [15] or prioritization of weak GWAS loci [16]. In [17] a COPD specific protein interaction network was constructed around genes differentially expressed between healthy and COPD subjects. This network was then queried for potential drug targets that could reverse the expression changes. COPD specific gene expression is also the basis of a Systems Medicine method proposed in [18] that aims at identifying subgroups of COPD patients with different molecular signatures.

Typically, only direct *physical* (binding) interactions are considered in the *Interactome*. Another line of Systems Medicine network approaches uses *functional* networks, where links may also represent indirect associations, for example co-expression [19] or genetic interaction [20,21]. These networks are usually assembled from specific experimental data, rather than the more general

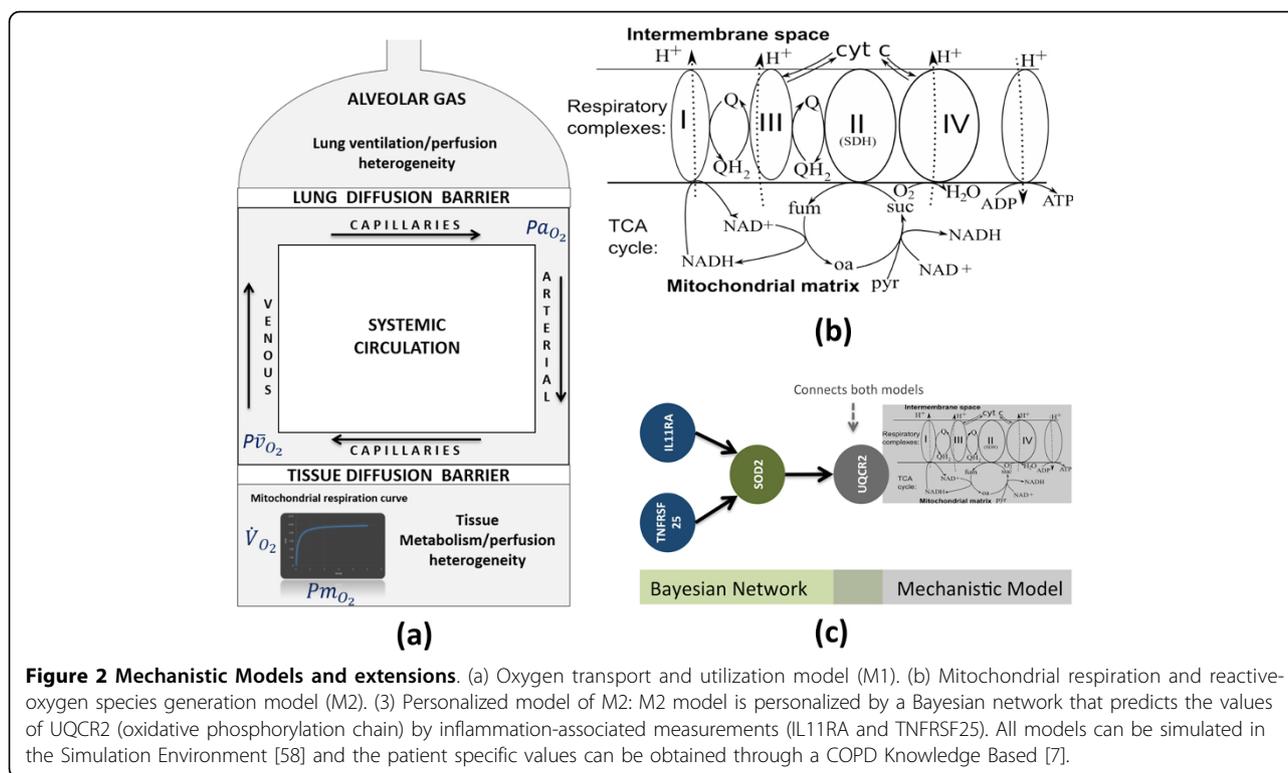


Figure 2 Mechanistic Models and extensions. (a) Oxygen transport and utilization model (M1). (b) Mitochondrial respiration and reactive-oxygen species generation model (M2). (c) Personalized model of M2: M2 model is personalized by a Bayesian network that predicts the values of UQC2 (oxidative phosphorylation chain) by inflammation-associated measurements (IL11RA and TNFRSF25). All models can be simulated in the Simulation Environment [58] and the patient specific values can be obtained through a COPD Knowledge Based [7].

interaction data in public databases. An example for the use of functional networks in COPD is given by the study of [22]. In order to explore the molecular basis of muscle degeneration in COPD, they developed a network model integrating several types of relevant measurements, such as blood cytokine levels and muscle gene expression. They were thereby able to identify several tissue remodeling and bioenergetics pathways that fail to coordinate in COPD diseased muscles.

Both physical and functional-based networks are useful tools that allow the identification of de-regulated functional elements, and to zoom-in into the interactions driving that de-regulation. We considered this information to be relevant in the generation of predictive models addressing the characterization of heterogeneity in COPD.

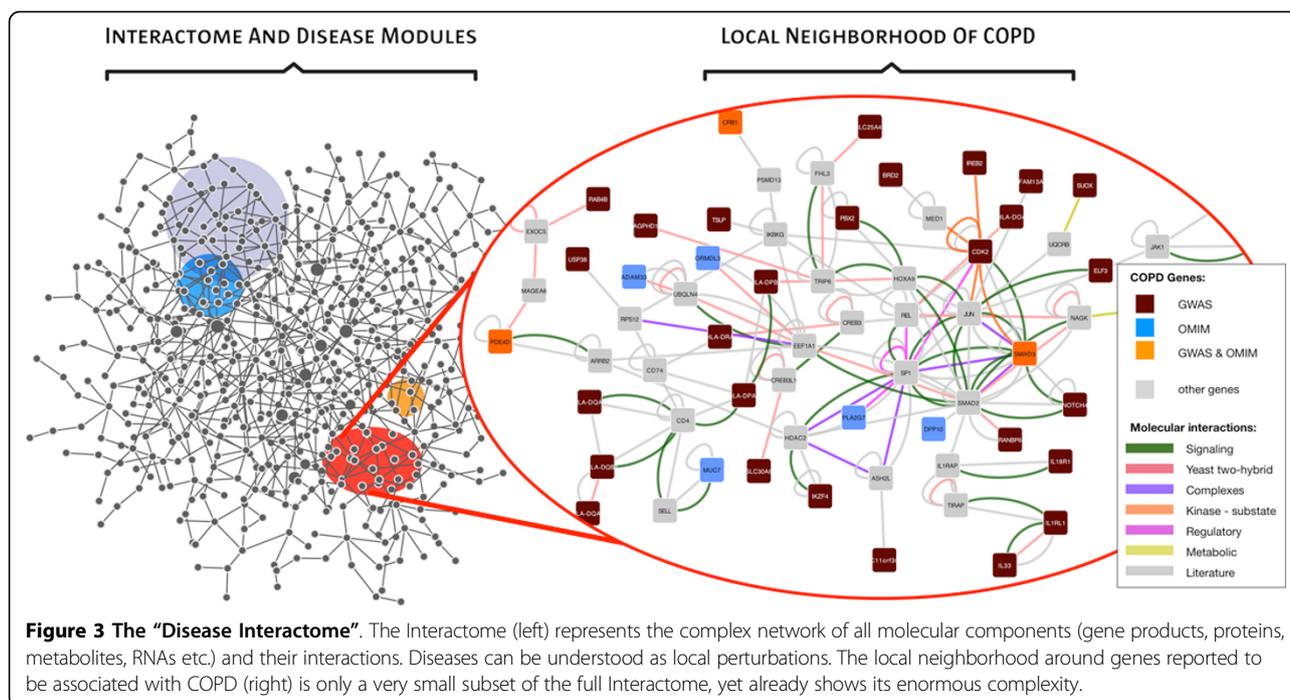
Methodologies and results

To characterize skeletal muscle dysfunction in COPD patients before and after training we evaluated two

hypotheses. In the first hypothesis, *MusclDys* is characterized by the de-regulated activity of a selected set of pathways; namely *transcription*, *proteolysis*, *immune activation* and/or *oxidative phosphorylation* (see [4] for more details). Using these pathways as a reference an initial list of associated genes was downloaded from the Synergy-COPD Knowledge-base SKB [7] and then filtered by clinical and biological experts. We followed a network approach similar to one described in the Bio-Bridge analysis [22] and extended it by including differential expressed genes, metabolites, cytokines, and clinical variables in addition to the filtered list of genes. We generated a network for each combination of healthy vs COPD and untrained vs trained individuals, obtaining two main results. First, we identified a module (*i.e.* a network-based cluster of genes, *Mod1*) that is present in all networks and shows the interaction between *mRNA-translation*, *Insulin* and *mTOR* pathways. Interestingly, this module was also associated to immune

Table 1 Major resources for protein interaction data:

Database / Interaction type	Reference
Databases integrating several sources	IntAct [59], MINT [60], BioGRID [61], HPRD [62], MIPS [63], STRING [64]
Protein complexes:	CORUM [65], [66]
Binary interactions (high-throughput)	CCSB-HI (CCSB), [67;68]
Regulatory interactions	TRANSFAC [69]
Kinase-substrate Interactions	PhosphositePlus [70]



markers such as *IL1B*. A second result is the general loss of co-regulation (including *Mod 1* loss) in COPD patients' muscle after training, independently confirming [22]. These promising results require further corroborating data in order to obtain a robust predictive model.

In a second hypothesis, we considered that given the relevance of bioenergetics and immune markers in COPD patients, we could use them to explain the level of *Reactive Oxidative Species* (ROS, major markers of oxidative stress) in COPD patients' muscle. In order to identify the sub-network(s) linking immune and bioenergetics genes to ROS-associated genes (from [23;24]) in the context of COPD we developed a chain-based methodology [25] named ChainRank. Briefly, the methodology identifies relevant sub-networks by identifying and scoring chains of interactions that link specific targets. The type of interactions to include in the chain search are selected by the user; we selected among those interactome-networks included in the Synergy-COPD Knowledge-Base [7]. Scores are generated from the integration of multiple general and context specific measures. Finally, the algorithm allows the identification of genes that are over-represented in highly ranked chains as relevant features. This list was then used for generating personalized predictive models in Objective 3.

Objective 2: Predictive models

The generation of predictive models in *MusclDys* involves three steps, beginning with the identification of existing mechanistic models that were then updated and

adapted to better estimate particular features of interest (FoI). In this process we make use of Objective 1's results to either identify FoI or to personalize the models by adding disease-related parameters.

The identification of mechanistic existing models

Many phenomena in physiology and biology are of essentially nonlinear nature and therefore require quantitative descriptions in addition to qualitative ones [26]. We considered three well-described physiological models of interest to the characterization of COPD: (i) *Oxygen transport and utilization* [27,28] (*M-OX*); (ii) *Cell Bioenergetics, mitochondrial respiration and reactive-oxygen-species generation* (ROS) [23,24], (*M-ROS*); and, (iii) *Spatial heterogeneities of lung ventilation and perfusion* [29] (*M-HET*). The first two models (*M-OX* and *M-ROS*) are relevant for the characterization of the systemic effects of the disease in skeletal muscle, as they provide mechanistic description of both the oxygen pathway and ROS generation. The third model (*M-HET*) is relevant for the study of pulmonary events in a sub-set of COPD patients [3] with low pulmonary density (high emphysema score) and mild airway remodeling resulting in mild to moderate airflow limitation.

Oxygen transport and utilization [27,28] (*M-OS*): The model details the determinants of oxygen transport from air to mitochondria and characterizes oxygen utilization at mitochondrial level during maximum exercise. In summary, it constitutes the most complete integrative approach of the interplay among factors modulating oxygen transport (lungs, hearth, blood and skeletal

muscle) and oxygen utilization at muscle level (see Figure 2 (a)). In COPD patients, the oxygen transfer capacity from the atmosphere to the cell, as well as its utilization at mitochondrial level, can be limited. It is therefore of interest to observe the effects of such a limitation at all levels of the transport chain [30] and, in particular, its effects on muscle's mitochondria.

Cell Bioenergetics, mitochondrial respiration and reactive-oxygen-species generation (M-ROS): The model details mitochondrial respiration and its relation with the production of Reactive Oxidative Species (ROS). The model integrates two sub-models, the Electron Chain model and the TCA cycle module [23,24]. ROS production in the mitochondrial respiratory chain is a signal of cellular adaptation to the environment, but a sharp increase is incompatible with cell survival; therefore the predicting ROS production is a relevant task. Moreover, in smoking-related COPD patients the antioxidant capacity is severely reduced and further decreases after smoking cessation due to endogenous production of ROS [31,32]. The integrated model (M-OX + M-ROS) generated within the Synergy-COPD project allows estimating quantitatively the relationships between determinants of cell oxygenation and mitochondrial ROS generation. The interplay between ROS levels and the antioxidant capacity of the redox system ultimately determines tissue oxidative and nitrosative stress with important implications on pathway regulation and cell damage.

Spatial heterogeneities of lung ventilation and perfusion [29] (M-HET): The anatomy-based multi-scale model of the human pulmonary circulation allows for the study of pre- and post-occlusion flow and embolus-generated flow redistribution, among other features. It combines four independent simulations of model geometry, tissue mechanics, ventilation and blood flow allowing for a local description of alveolar ventilation and pulmonary blood flow. The lung modeling approaches are described in this Supplement in detail in a separate paper as interactive work with the FP7 EU Project *AirPROM*. A major relevance of lung modeling in Synergy-COPD is the characterization of patients with reduced lung density but without classical COPD symptoms of airway obstruction [3]. The inclusion of this modeling approach in Synergy-COPD had two main goals. Firstly, the analysis of the impact of spatial heterogeneities of lung ventilation and perfusion on blood oxygenation and, secondly, the study of the subset of COPD patients showing dissociation between high emphysema score and low intensity of airway remodeling, as indicated above and described in [3].

Updating existing models

In order to characterize skeletal muscle dysfunction in COPD (*MusclDys*) we modified the oxygen transport

and utilization model (M-OX) and included the mitochondrial respiration [33]. The outcomes of this model were two fold: (1) it increased the physiological validity of the model by estimating the mitochondrial P_{O_2} , (2) it allowed for the integration with bioenergetics models (M-ROS). A second extension of the model was the modeling of lung ventilation/perfusion heterogeneities and [33]; these extensions (see Figure 2(a)) allows better estimation and better personalization of the model in COPD.

In addition, we integrated models M-OX and M-ROS (IM) to model the relation between oxygen transport and ROS generation, which constitutes a major marker of skeletal muscle dysfunction. Parameter models were investigated again in the case of [23,24], to provide better ROS estimations. A major outcome of the integrated model analysis [34] is that it permits to estimate the effect of various states of oxygen supply and demand of mitochondrial P_{O_2} on ROS production; this is relevant in COPD patients with airway obstruction symptoms [22].

Novel models

In Synergy-COPD we generated novel models (Objective 2) by integrating existing mechanistic models, developed for the healthy individual, with features of interest (FoI) associated to *MusclDys* (obtained in Objective 1). Having identified ROS as a major marker of muscle dysfunction, we aimed to predict ROS status by surrogate variables identified in Objective 1. Initially those surrogate variables were immune markers from protein measurements in the blood as shown in [22]. To integrate the effect of the surrogate variables (SV) with the integrated model of oxygen transport and ROS generation we proposed to use SV values (which are more commonly used in clinical diagnostics) to estimate a subset of the integrated model parameter values. As a technical solution we proposed the use of Bayesian networks (see Figure 2(c), [35]) to connect immune markers and selected model parameters. This connection was also possible through the use of methodologies and results obtained during Objective 1 such as the ChainRank methodology briefly described elsewhere in the manuscript. However, the generation of accurate linking-by-Bayesian networks is limited by the requirement of a large number of samples. Therefore, to increase accuracy we made use of public available muscle-related data-sets in GEO [6] and included estimates of relations from other sources such as text-mining [36]) into the Bayesian network generation. While we considered that the proposed approach to be technically valid, still we need to increase the sample size to generate useful models, therefore the requirement of follow-up studies. The model can be run in the Synergy-COPD Simulation Environment and the patient specific values can be obtained through a COPD Knowledge Base [2,7].

As a second approach, we integrated transcriptomic data from muscle biopsies and literature-based data into a mathematical discrete model [37]. By this model-driven approach we aimed to determine the processes that lead the abnormal adaptation to training in COPD patients and the role of ROS in this process. Since skeletal muscle mitochondrial dysfunction is a central actor in COPD [38] this approach was based on those genes associated to selected mitochondrial processes from Objective 1's candidate biomarkers obtained in [22]. The modeling was achieved by inferring the activity state of a gene regulatory network (GRN) in six different states: Control group, COPD with normal body mass index (BMI) and COPD with low BMI before and after undergoing 8 weeks of training program [22]. We carried out this task in two parts: 1) GRN reconstructions and 2) Integration of GRN into a discrete model.

As a first step in the GRN reconstruction we curated the list of candidate biomarkers to be included by the re-analysis of the transcriptomic data of the six different states used previously in [22]. For this aim, we used statistical methods such as rank product [39] to determine the gene candidates and Gene Ontology and Human Proteins Atlas databases [40,41] to filter those genes associated with mitochondrial processes in skeletal muscle. Next, to determine gene associations we used IPA software and DroID [42,43]. Finally, In order to correct incomplete or erroneous annotations and identify the direction and the sign of the interactions, we manually curated the GRN reconstruction using a large number of bibliographic data sources.

The GRN reconstruction was then converted into a mathematical discrete model based on the *Thomas formalism* [44] by mechanistically describing the interactions between those mitochondrial-associated genes that were differentially expressed between states. In order to refine the accuracy of our model predictions, we used public available muscle-related data-sets in GEO [6] to impose constraints to our model. We integrated these constraints in the form of inequalities based on probabilistic approaches: if we observed a strong Pearson correlation ($\rho > 0.9$) between two non-connected genes, their expression values were forced to evolve in the same direction. The rationale is just the opposite in the case of a strong anti-correlation. Then, summing up, we propose a method by which the interaction between genes are determined by performing a tissue and organelle specific GRN reconstruction and the constraints are defined using probabilistic approaches, finally both, the GRN reconstruction and the constraints are integrated into a discrete model in order to unveil the mechanisms governing the adaptation to training in the groups of study.

Together, both probabilistic approaches show a way forward to close the inherent under-determination gap

of deterministic, quantitative models by coupling data driven and knowledge driven approaches.

Objective 3: Clinical application and limitations in Synergy-COPD

The models proposed to address *MusclDys* are still far from the clinical practice. We consider that the methodologies to achieve such goal do exist, but the data publicly available is limited. Several lessons can be learnt:

- (1) While the use of mechanistic models is very valid to understand biological systems and diseases, they have serious limitations in the study of complex diseases. Complex diseases may be described as the combination of many factors, and mechanistic models will require too many parameters and consequently too much data to be *yet* clinically effective.
- (2) However statistical predictive models (e.g. linear models, Bayesian networks, etc.), not necessarily mechanistically accurate, may provide a valid technical solution. The implementation of such technical solution requires the use of large amount of data to ensure accuracy and statistical validation. For this data to be obtained we consider necessary (1) to strengthen the policies promoting data-sharing (especially in the clinical context) and (2) the generation of large data-sets with proper experimental designs and clinically-driven hypotheses.
- (3) Clinically driven research needs to be re-designed to align the different objectives described in Figure 1. We observed that the re-use of data is necessary but complex, and minimal modifications (such as extended questionnaires to patients providing samples) in existing bio-banks may be very useful.

Understanding COPD co-morbidities through systems medicine

COPD has been associated with several diseases such as lung cancer [45], metabolic syndrome and cardiovascular diseases [46]. However, not all COPD patients share the same diseases or exhibit the same degree of co-morbidity. We hypothesized that the particular co-morbidities in a given COPD patient can be understood from his/her particular set of de-regulated pathways and genes. Identifying genes and pathways that are shared between COPD associated diseases could therefore allow for a more detailed characterization of COPD and its co-morbidities. In the first subsection we briefly describe our method to identify such potential biomarkers (pathways and/or genes). We then introduce our initial predictive models in the second subsection and finally discuss the clinical applications. Some of the data sets supporting this article are available in HuDiNe repository (<http://barabasilab.neu.edu/projects/>)

hudine/resource/data/data.html); for the rest we had permission to access and analyze the data.

Biomarker and Co-morbidity identification

Using 13 million health records from U.S. Medicare [47], we identified 27 disease groups (DG) with significantly elevated risks to co-occur with COPD. These groups included both well-established associations like cardiovascular diseases or lung cancer, but also unexpected ones that could be interesting candidates for more focused follow-up investigations. In order to elucidate possible shared molecular origins between the disease groups and COPD, we considered their respective implicated pathways: for each disease group, we first constructed a comprehensive list of known associated genes from the literature (by pooling several sources of gene-disease associations such as OMIM, NIH Thesaurus and text-mining among others). We then performed a pathway enrichment analysis for each disease group. The results show that there are a number of pathways that are shared between different disease groups, suggesting that the observed co-morbidities are indeed rooted in shared molecular mechanisms. By further inspecting the genes within prevalent pathways we were able to identify a number of genes with the potential to characterize COPD co-morbidity. We are investigating if those markers may predict the level of co-morbidity. This could be of immediate relevance for the clinical practice, as co-morbidity has been associated to lower overall quality of life [48] and increased mortality [47,49]. To date, a number of interesting outcomes of this analysis remain to be validated in further studies. However, with currently available data we considered that the disease groups and co-factors such as age and gender could be used to generate predictive models.

Predictive models

We selected disease groups that are highly prevalent in COPD patients, such as heart and circulation associated diseases and digestive alterations. The observation that their prevalences vary with age prompted us to develop a first model (Objective 2) aiming to predict the probability for specific co-morbidities in COPD patients over different age strata; in this case we made use of the ranking of co-morbidities from Objective 1 to select those diseases of major interest in the generation of the predictive models. This model may be used as support information for clinicians in the daily practice (e.g. predictive medicine, or comparing observed symptoms with candidate co-morbidities). For a more robust clinical validation, however, follow-up studies in different cohorts will be required; for this reason we consider the comorbidity modeling as part of Objective 2, but closer to the Objective 3 than any other model presented.

Clinical application and limitations in Synergy-COPD

While co-morbidity is being generally accepted as a relevant clinical factor [47,50-55], co-morbid predictive models are rarely reaching the clinical practice. Our experiences gained throughout the Synergy-COPD project suggest several limitations:

- (1) Incompatibilities in the medical nomenclature. While there are several large health registries available to investigate co-morbidities (e.g. Medicare, Swedish Registry and others), there is yet to agree a common diagnostic standard even for simplified administrative coding. In Sweden, for instance, nowadays ICD10 is being used, yet the registry also includes information coded in ICD7 to ICD9. In comparison, Medicare (as used in [47]) is mainly using ICD9 codes. Maps between ICD coding do exist, but they are not accurate, and every new ICD coding system may represent a different conceptual approach. ICD11 will represent a new challenge.
- (2) There are many studies investigating specific co-morbidities, not only in COPD but in many other diseases. Yet, two major limitations inhibit the integration of these studies in larger meta-analyses: (1) diseases may be defined differently in each study and in many cases no official coding is followed; (2) the selection of diseases is biased towards well established and expected diseases. These two limitations reflect that most studies are developed to validate specific hypotheses. We believe that broader studies and normalized questionnaires will eventually facilitate meta-analyses and thereby increase the power of co-morbidity studies.
- (3) Finally, previous large-scale studies are often limited to one type of data: either -omics data were collected, but no co-morbidity information, or the other way round. We believe that in the future it will be crucial to combine these two approaches.

Conclusions

Despite the massive amounts of data collected in medical research throughout the last decades, our understanding of complex diseases still remains very limited. The fundamental shortcoming of our knowledge may be illustrated by a popular quote attributed to Ernest Rutherford: *all science is either physics or stamp collecting* [56]. Systems Medicine bears the promise of facilitating the transition from stamps to understanding. Indeed we are convinced that a systems perspective is necessary for the integration of all hitherto largely disconnected facts, thereby ultimately enabling clinically predictive tools. Synergy-COPD presents a large-scale case study of Systems Medicine applied to COPD, recognizing that both *Clinical Research*

and *Clinical Decision Systems* require the development of integrative quantitative models. Developing such models is a complex task which we addressed by adhering to a 3-step framework: (1) feature identification, (2) model generation and statistical validation, (3) clinical validation and implementation. We developed and used the framework targeting specifically the characterization of muscle-related systemic effects and co-morbidity as use-cases thus grounding the methodology in real-world applications. In both use-cases we were able to identify candidate biomarkers that may help characterizing COPD heterogeneity, and developed models with the potential to be considered in future Clinical Decision Support Systems (e.g. co-morbidity prevention and prognosis among other objectives). Throughout the project we identified several key factors that are currently limiting the clinical applicability of our approach: the most important ones were data availability, normalization of frameworks (e.g. ICD codes in co-morbidity) and the necessity of broader and optimized experimental designs (e.g. the inclusion of co-morbidity information in genomic studies).

In conclusion, we consider that the first steps to bridge the gap between basic research and clinical practice are built, however further steps are required to complete the path. To exploit the full potential of our results, future follow-ups are required for statistical and clinical validation, and once validated, predictive models (supported by longitudinal studies) will make a strong case for clinical applications. Further considerations on challenges and future are discussed in [57] on this Supplement.

We are at the juncture of a very exciting era, where Systems Medicine offers the possibility of a real connection between research and clinical applications. While Synergy-COPD may only represent a minor milestone along a long road, we are convinced it is a relevant and instructive case.

Competing interests

DM is part of Biomax Informatics AG. The rest of authors declare they have no competing interests.

Authors' contributions

DGC and JM defined an initial draft of the manuscript. DGC, JM, JR, JT, and LB reviewed and defined the final structure. DGC and JM wrote the manuscript. All authors first reviewed their specific sections in detail, then reviewed the full document, in both cases they proposed modifications; finally all authors agreed on the final version.

Acknowledgements

We would like to thank all the members of the Synergy-COPD consortium for their support.

Declaration

Publication of this article has been funded by the Synergy-COPD European project (FP7-ICT-270086). The opinions expressed in this manuscript are those of the authors and are not necessarily those of Synergy-COPD project's partners or the European Commission
This article has been published as part of *Journal of Translational Medicine* Volume 12 Supplement 2, 2014: Systems medicine in chronic diseases: COPD

as a use case. The full contents of the supplement are available online at <http://www.translational-medicine.com/supplements/12/S2>.

Authors' details

¹Unit of computational Medicine, Center for Molecular Medicine, Department of Medicine, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden. ²Center for Complex Network Research, Northeastern University Physics Department, Boston, MA 02115, USA. ³Department of Theoretical Physics, Budapest University of Technology and Economics, H-1111 Budafoki út. 8., Budapest, Hungary. ⁴Hospital Clinic, IDIBAPS, CIBERES, Universitat de Barcelona, Barcelona, Catalunya, Spain. ⁵Barcelona Digital Technology Centre Carrer Roc Boronat, 117 08018 Barcelona. ⁶Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. ⁷Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool, UK. ⁸Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford, OX1 3QD, UK. ⁹Biomax Informatics AG, Munich, Germany. ¹⁰School of Medicine, University of California, San Diego, San Diego, CA 92093-0623A, USA. ¹¹Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Smith Bldg., Rm. 858A, 450 Brookline Ave, Boston, MA 02215, USA. ¹²Center for Network Science, Central European University, Nadoru. 9, 1051 Budapest, Hungary. ¹³Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115 USA.

Published: 28 November 2014

References

1. Majumder D, Mukherjee A: **A passage through systems biology to systems medicine: adoption of middle-out rational approaches towards the understanding of therapeutic outcomes in cancer.** *The Analyst* 2011, **136**(4):663-78.
2. Gomez-Cabrero , David , Lluch-Ariet , Magi , Tegner Jesper, Cascante Marta, Miralles Felip, Roca Josep, the Synergy-COPD consortium: **Synergy-COPD: A systems Approach for understanding and managing Chronic Diseases.** *Journal of Translational Medicine* 2014, **12**(Suppl 2):S2.
3. Garcia-Aymerich J, Gómez FP, Benet M, Ferrero E, Basagaña X, Gayete A, Antó JM: **Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes.** *Thorax* 2011.
4. Josep Roca, Claudia Vargas, Isaac Cano, Vitaly Selivanov, Esther Barreiro, Dieter Maier, Francesco Falciani, Peter Wagner, Marta Cascante, Judith Garcia-Aymerich, Susana Kalko, Igor Marin, Jesper Tegner, Joan Escarrabill, Alvar Agustí, David Gomez-Cabrero, the Synergy-COPD consortium: **Chronic Obstructive Pulmonary Disease Heterogeneity: Challenges for Health Risk Assessment, Stratification and Management.** *BMC Journal of Translational Medicine, to appear* 2014, **12**(Suppl 2):S3.
5. Filip Velickovski, Luigi Ceccaroni, Josep Roca, Felip Burgos, Galdiz Juan B, Marina Nueria, Magi Lluch-Ariet: **Clinical Decision Support Systems (CDSS) for preventive management of COPD patients.** *Journal of Translational Medicine* 2014, **12**(Suppl 2):S9.
6. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A: **NCBI GEO: archive for functional genomics data sets—update.** *Nucleic Acids Res* 2013, **41**(Database):D991-5, Jan.
7. Isaac Cano, Ákos Tényi, Christine Schueller, Martin Wolff, Mercedes Huertas Migueláñez M, David Gomez-Cabrero, Philipp Antczak, Josep Roca, Marta Cascante, Francesco Falciani, Dieter Maier: **The COPD Knowledge Base: enabling data analysis and computational simulation in translational COPD research.** *Journal of Translational Medicine* 2014, **12**(Suppl 2):S6.
8. Picard , Richard , Cook , Dennis : **"Cross-Validation of Regression Models".** *Journal of the American Statistical Association* 1984, **79**(387):575-583.
9. Bahr TM, Hughes GJ, Armstrong M, Reisdorph R, et al: **Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease.** *Am J Respir Cell Mol Biol* 2013, **49**(2):316-23, Aug.
10. Vidal M, Cusick ME, Barabási AL: **Interactome networks and human disease.** *Cell* 2011, **144**(6):986-98.
11. Schadt EE: **Molecular networks as sensors and drivers of common human diseases.** *Nature* 2009, **461**:218-223.
12. Barabási AL, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease.** In *Nature reviews. Volume 12. Genetics*; 2011:(1):56-68.

13. Köhler S, et al: **Walking the interactome for prioritization of candidate disease genes.** *The American Journal of Human Genetics* 2008, **82**:949-958.
14. Oti M, Snel B, Huynen MA, Brunner HG: **Predicting disease genes using protein-protein interactions.** *Journal of medical genetics* 2006, **43**:691-698.
15. Navlakha S, Gitter A, Bar-Joseph Z: **A network-based approach for predicting missing pathway interactions.** *PLoS computational biology* 2012, **8**:e1002640.
16. Sharma A, Gulbahce N, Pezner SJ, Menche J, Ladenvall C, Folkersen L, Eriksson P, Orho-Melander M, Barabási AL: **Network-based Analysis of Genome Wide Association Data Provides Novel Candidate Genes for Lipid and Lipoprotein Traits.** *Molecular & Cellular Proteomics* **12.11**(2013):3398-3408.
17. Bao H, Wang J, Zhou D, Han Z, Su L, Zhang Y, Li Q: **Protein-Protein Interaction Network Analysis in Chronic Obstructive Pulmonary Disease.** *Lung (2800)* 2013.
18. Menche J, Sharma A, Cho MH, Mayer RJ, Rennard SI, Celli B, Barabási AL: **A divisive Shuffling Approach (ViStA) for gene expression analysis to identify subtypes in Chronic Obstructive Pulmonary Disease.** *BMC Systems Biology* 2014, **8**(Suppl 2):S8.
19. Stuart JM, Segal E, Kolker D, Kim SK: **A gene-coexpression network for global discovery of conserved genetic modules.** *Science* 2003, **302**:249-255.
20. Davierwala , Armaity P, et al: **The synthetic genetic interaction spectrum of essential genes.** *Nature genetics* 2005, **37**(10):1147-1152.
21. Franke L, van Bakel H, Fokkens L, de Jong ED, Egmont-Petersen M, Wijmenga C: **Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes.** *The American Journal of Human Genetics* 2006, **78**:1011-1025.
22. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, Falciani F: **A Systems Biology Approach Identifies Molecular Networks Defining Skeletal Muscle Abnormalities in Chronic Obstructive Pulmonary Disease.** In *PLoS Computational Biology* S. Miyano 2011, **7**(9):e1002129.
23. Selivanov Va, Votyakova TV, Pivtoraiko VN, Zeak J, Sukhomlin T, Trucco M, Cascante M: **Reactive oxygen species production by forward and reverse electron fluxes in the mitochondrial respiratory chain.** *PLoS Computational Biology* 2011, **7**(3):e1001115.
24. Selivanov Va, Votyakova TV, Zeak Ja, Trucco M, Roca J, Cascante M: **Bistability of mitochondrial respiration underlies paradoxical reactive oxygen species generation induced by anoxia.** *PLoS Computational Biology* 2009, **5**(12):e1000619.
25. Ákos Tényi, Pedro de Atauri, David Gomez-Cabrero, Isaac Cano Isaac, Francesco Falciani Francesco, Marta Cascante Marta, Josep Roca, Dieter Maier: **ChainRank, a chain search based method for prioritization and contextualisation of biological sub-networks.** 2014, Submitted.
26. Wolkenhauer O, Auffray C, Jaster R, Steinhoff G, Dammann O: **The road from systems biology to systems medicine.** *Pediatric research* 2013, **73**(4 Pt 2):502-7.
27. Wagner PD: **A theoretical analysis of factors determining VO2max at sea level and altitude.** *Resp Physiol* 1996a, **106**(3):329-43.
28. Wagner PD: **Determinants of maximal oxygen transport and utilization.** *Annual Review of Physiology* 1996b, **58**:21-50.
29. Burrowes KS, Clark aR, Tawhai MH: **Blood flow redistribution and ventilation-perfusion mismatch during embolic pulmonary arterial occlusion.** *Pulmonary Circulation* 2011, **1**(3):365-76.
30. Heffner JE: **The story of oxygen.** 2013, **58**(1):18-31.
31. Kirkham Pa, Barnes PJ: **Oxidative stress in COPD.** *Chest* 2013, **144**(1):266-73.
32. Barnes PJ: **Cellular and Molecular Mechanisms of Chronic Obstructive Pulmonary Disease.** *Clinics in Chest Medicine* 2014, **35**(1):71-86.
33. Cano I, Mickael M, Gomez-Cabrero D, Tegnér J, Roca J, Wagner PD: **Importance of Mitochondrial in Maximal O2 Transport and Utilization: a Theoretical Analysis.** *Respiratory Physiology & Neurobiology* 2013.
34. Selivanov Cano V, Gomez-Cabrero David, Tegnér Jesper, Roca J, Wagner PD, Cascante M: **Mathematical Modeling of Skeletal Muscle Reactive Oxygen Species Generation during Maximal Exercise as a function of Mitochondrial.** 2014, submitted.
35. Scutari M: **Learning Bayesian Networks with the bnlearn R Package.** 2010, **35**(3).
36. Haibe-Kains B, Olsen C, Djebbari A, Bontempi G, Correll M, Bouton C, Quackenbush J: **Predictive networks: a flexible, open source, web application for integration and analysis of human gene networks.** *Nucleic Acids Research* 2012, **40**(Database):D866-75.
37. de Mas Igor Marin, Fanchon E, Selivanov Vitaly A, Papp Balázs, Josep Roca, Marta Cascante: **A novel discrete model-driven approach unveils abnormal metabolic adaptation to training in COPD.**, Submitted.
38. Meyer A, Zoll J, Charles AL, Charloux A, de Blay F, Diemunsch P, Sibilia J, Piquard F, Geny B: **Skeletal muscle mitochondrial dysfunction during chronic obstructive pulmonary disease: central actor and therapeutic target.** *Exp Physiol* 2013, **98**(6):1063-78.
39. 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} Letters* 2004, **573**(1-3):83-92.
40. Thomas PD, Mi H, Lewis S: **Ontology annotation: mapping genomic regions to biological function.** *Curr Opin Chem Biol* 2007, **11**(1):4-11.
41. Uhlén M, Björling E, Agaton C, Szijarto CA, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, Berglund L, Bergström K, Brumer H, Cerjan D, Ekström M, Eloheid A, Eriksson C, Fagerberg L, Falk R, Fall J, Forsberg M, Björklund MG, Gumbel K, Halimi A, Hallin I, Hamsten C, Hansson M, Hedhammar M, Hercules G, Kampf C, Larsson K, Lindskog M, Lodewyckx W, Lund J, Lundeberg J, Magnusson K, Malm E, Nilsson P, Odling J, Oksvold P, Olsson I, Oster E, Ottosson J, Paavilainen L, Persson A, Rimini R, Rockberg J, Runeson G, Mumbel K, Halimi A, Sköllerömo A, Steen J, Stenvall M, Sterky F, Strömberg S, Sundberg M, Tegel H, Tourle S, Wahlund E, Waldén A, Wan J, Wernérus H, Westberg J, Wester K, Wrethagen U, Xu LL, Hober S, Pontén F: **A human protein atlas for normal and cancer tissues based on antibody proteomics.** *Mol Cell Proteomics* 2005, **4**(12):1920-32.
42. Murali T, Pacifico S, Yu J, Guest S, Roberts GG 3rd, Finley RL Jr: **DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila.** *Nucleic Acids Res* 2011, **39**(Database):D736-43.
43. Krämer A, Green J, Pollard J Jr, Tugendreich S: **Causal analysis approaches in Ingenuity Pathway Analysis.** *Bioinformatics* 15 2014, **30**(4):523-30.
44. Thomas R, Multistationarity Kaufman M: **the Basis of Cell Differentiation and Memory. II Logical Analysis of Regulatory Networks in Term of Feedback Circuits** *Chaos* 2001 **11**:180-195.
45. Houghton aM: **Mechanistic links between COPD and lung cancer.** *Nature Reviews Cancer* 2013, **13**(4):233-45.
46. Müllerova H, Agustí A, Erqou S, Mapel DW: **Cardiovascular comorbidity in COPD: systematic literature review.** *Chest* 2013, **144**(4):1163-78.
47. Hidalgo Ca, Blumm N, Barabási AL, Christakis Na: **A dynamic network approach for the study of human phenotypes.** *PLoS Computational Biology* 2009, **5**(4):e1000353.
48. Fabbri LM, Beghé B, Agustí A: **COPD and the solar system: introducing the chronic obstructive pulmonary disease comorbidome.** *American Journal of Respiratory and Critical Care Medicine* 2012, **186**(2):117-9.
49. Divo M, Cote C, de Torres JP, Casanova C, Marin JM, Pinto-Plata V, Celli B: **Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease.** *American Journal of Respiratory and Critical Care Medicine* 2012, **186**(2):155-61.
50. Areias V, Carreira S, Anciães M, Pinto P, Bárbara C: **Co-morbidities in patients with gold stage 4 chronic obstructive pulmonary disease.** *Revista Portuguesa de Pneumologia* 2013.
51. Decramer M, Janssens W, Miravittles M: **Chronic obstructive pulmonary disease.** *Lancet* 2012, **379**(9823):1341-51.
52. Hernandez C, Jansa M, Vidal M, Nuñez M, Bertran MJ, Garcia-Aymerich J, Roca J: **The burden of chronic disorders on hospital admissions prompts the need for new modalities of care: a cross-sectional analysis in a tertiary hospital.** *QJM: Monthly Journal of the Association of Physicians* 2009, **102**(3):193-202.
53. Inghammar M, Ekbohm A, Engström G, Ljungberg B, Romanus V, Löfdahl C-G, Eggesten A: **COPD and the risk of tuberculosis—a population-based cohort study.** *PLoS One* 2010, **5**(4):e10138.
54. Khdour MR, Hawwa AF, Kidney JC, Smyth BM, McElroy JC: **Potential risk factors for medication non-adherence in patients with chronic obstructive pulmonary disease (COPD).** *European Journal of Clinical Pharmacology* 2012, **68**(10):1365-73.
55. MacLay JD, MacNee W: **Cardiovascular disease in COPD: mechanisms.** *Chest* 2013, **143**(3):798-807.
56. Birks JB: **Rutherford at Manchester.** Heywood. London; 1962.
57. Felip Miralles, David Gomez-Cabrero, Magi Lluch-Ariet, Jesper Tegnér, Marta Cascante, Josep Roca, the Synergy-COPD consortium: **Predictive**

- Medicine: Outcomes, Challenges and Opportunities in the Synergy-COPD project.** *Journal of Translational Medicine* 2014, **12**(Suppl 2):S12.
58. Mercedes Huertas Migueláñez M, Daniel Mora, Isaac Cano, Dieter Maier, David Gomez-Cabrero, Magi Lluch-Ariet, Felip Miralles: **Simulation Environment and Graphical Visualization Environment: a COPD use-case.** *Journal of Translational Medicine* 2014, **12**(Suppl 2):S7.
 59. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J, Kerssemakers J, Leroy C, Menden M, Michaut M, Montecchi-Palazzi L, Neuhauser SN, Orchard S, Perreau V, Roechert B, van Eijk K, Hermjakob H: **The intact molecular interaction database in 2010.** *Nucleic acids research* 2010, **38**:D525-D531, 2010.
 60. Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G: **Mint, (2010) The molecular interaction database: 2009 update.** *Nucleic acids research* 2010, **38**:D532-D539.
 61. Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, Reguly T, Rust JM, Winter A, Dolinski K, Tyers M: **The biogrid interaction database: 2011 update.** *Nucleic acids research* 2011, **39**:D698-D704, 2011.
 62. Prasad TK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadrans S, Chaerkady R, Pandey A: **Human protein reference database2009 update.** *Nucleic acids research* 2009, **37**:D767-D772.
 63. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes HW, Ruepp A, Frishman D: **The MIPS mammalian protein-protein interaction database.** *Bioinformatics* 2005, **21**(6):832-834.
 64. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ: **STRING v9. 1: protein-protein interaction networks, with increased coverage and integration.** *Nucleic acids research* 41 **D1**(2013):D808-D815.
 65. Ruepp A, Waegle B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Mewes HW: **Corum: the comprehensive resource of mammalian protein complexes 2009.** *Nucleic acids research* 2010, **38**:D497-D501.
 66. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, Babu M, Craig SA, Hu P, Wan C, Vlasblom J, Dar VU, Bezginov A, Clark GW, Wu GC, Wodak SJ, Tillier ER, Paccanaro A, Marcotte EM, Emili A: **A census of human soluble protein complexes.** *Cell* 2012, **150**(5):1068-1081.
 67. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzläff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksöz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE: **A human protein-protein interaction network: a resource for annotating the proteome.** *Cell* 2005, **122**:957-968.
 68. Rual JF, 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 LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, *et al*: **Towards a proteome-scale map of the human protein-protein interaction network.** *Nature* 2005, **437**:1173-1178.
 69. Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E: **Transfac: transcriptional regulation, from patterns to profiles.** *Nucleic acids research* 2003, **31**:374-378.
 70. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M: **Phosphositeplus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.** *Nucleic acids research* 2012, **40**:D261-D270.

doi:10.1186/1479-5876-12-S2-S4

Cite this article as: Gomez-Cabrero *et al*: Systems Medicine: from molecular features and models to the clinic in COPD. *Journal of Translational Medicine* 2014 **12**(Suppl 2):S4.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

