From systems biology to P4 medicine: applications in respiratory medicine

Guillaume Noell[1,2], Rosa Faner[1,2] and Alvar Agustí[1,2,3]

Number 4 in the Series "Personalised medicine in respiratory diseases"
Edited by Renaud Louis and Nicolas Roche

Affiliations: [1]Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. [2]CIBER Enfermedades Respiratorias (CIBERES), Barcelona, Spain. [3]Respiratory Institute, Hospital Clinic, Universitat de Barcelona, Barcelona, Spain.

Correspondence: Alvar Agustí, Respiratory Institute, Hospital Clínic, Villarroel 170, 08036 Barcelona, Spain. E-mail: AAGUSTI@clinic.cat

@ERSpublications
Systems biology and network medicine have the potential to transform medical research and practice
http://ow.ly/r3jR30hf35x

Cite this article as: Noell G, Faner R, Agustí A. From systems biology to P4 medicine: applications in respiratory medicine. *Eur Respir Rev* 2018; 27: 170110 [https://doi.org/10.1183/16000617.0110-2017].

ABSTRACT Human health and disease are emergent properties of a complex, nonlinear, dynamic multilevel biological system: the human body. Systems biology is a comprehensive research strategy that has the potential to understand these emergent properties holistically. It stems from advancements in medical diagnostics, "omics" data and bioinformatic computing power. It paves the way forward towards "P4 medicine" (predictive, preventive, personalised and participatory), which seeks to better intervene preventively to preserve health or therapeutically to cure diseases. In this review, we: 1) discuss the principles of systems biology; 2) elaborate on how P4 medicine has the potential to shift healthcare from reactive medicine (treatment of illness) to predict and prevent illness, in a revolution that will be personalised in nature, probabilistic in essence and participatory driven; 3) review the current state of the art of network (systems) medicine in three prevalent respiratory diseases (chronic obstructive pulmonary disease, asthma and lung cancer); and 4) outline current challenges and future goals in the field.

Introduction

Human health and disease are emergent properties of a complex, multilevel biological system that spans from the molecular domain to the microbiome, exposome and social levels (figure 1) [1, 2]. Ideally,

Previous articles in this series: No. 1: Chung KF. Personalised medicine in asthma: time for action. *Eur Respir Rev* 2017; 26: 170064. **No. 2:** Bonsignore MR, Suarez Giron MC, Marrone O, *et al*. Personalised medicine in sleep respiratory disorders: focus on obstructive sleep apnoea diagnosis and treatment. *Eur Respir Rev* 2017; 26: 170069. **No. 3:** Mascaux C, Tomasini P, Greillier L, *et al*. Personalised medicine for nonsmall cell lung cancer. *Eur Respir Rev* 2017; 26: 170066.

Received: Sept 28 2017 | Accepted after revision: Nov 30 2017

Support statement: This work was supported, in part, by Instituto de Salud Carlos III (PI12/01117, PI15/00799, CP16/00039), Recercaixa-2012 (AA084096), SEPAR (PI065/2013, PI192/2012, PI68/2015), AGAUR FI-DGR 2016 PhD grant and AstraZeneca foundation COPD Young Researcher Award. CERCA Programme/Generalitat de Catalunya. This work was developed at the building Centre de Recerca Biomèdica Cellex, Barcelona, Spain. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: None declared.

Provenance: Commissioned article, peer reviewed.

therefore, if we want to intervene prophylactically to preserve health or therapeutically to cure disease, in a safe and effective way, we should understand these dynamic gene–environment interactions in greater detail. Certainly, this will not be an easy task, but the alliance of new high-throughput "omic" methodologies, novel imaging techniques and current (and future) computational power can project us forward in this endeavour and eventually facilitate the development of novel therapeutic strategies (and the repurposing of old ones) [3]. However, as wisely highlighted by one of the anonymous reviewers of this paper, to whom we are grateful: "… full understanding of complex nonlinear systems in physics and biology might not be ever possible and, fortunately, might not be even required because probabilistic decisions are (and will become) more powerful than decisions based on precise mechanistic understanding. This is a real revolution already happening in society (Google and Amazon can predict your behaviour without knowing (less understanding) you). Similarly, Artificial Intelligence (AI) will be able soon to predict the clinical course and responsiveness to intervention based on probabilities rather than on deep understanding of the system …". We think that both concepts are actually synergistic since a more comprehensive and precise understanding of human biology (figure 1) will, no doubt, feed back to any AI platform, which will in turn provide new hypotheses to test iteratively. In any case, embracing a holistic scientific approach (as opposed to the reductionist research strategy used traditionally) for the understanding of human health and disease is a unique (and mandatory) opportunity to really move medical practice forward in the 21st century.

In this review, we: 1) discuss the principles of systems biology, a relatively recent research strategy that leverages from omics and bioinformatics to gain a holistic understanding of complex biological systems; 2) elaborate on how this can pave the way towards the effective deployment of the so-called "P4 medicine" (predictive, preventive, personalised and participatory) [4], which can shift healthcare from treatment of illness to prediction and prevention of illness, in a revolution that will be personalised in nature, probabilistic in essence and participatory driven; 3) review the state of the art of network (systems) medicine in three prevalent respiratory diseases (chronic obstructive pulmonary disease (COPD), asthma and lung cancer); and 4) outline current challenges and future goals in the field.

Systems biology

System approaches and emergent properties

System approaches stem from the premise that separate analysis of information gathered from different elements, compartments or levels of a dynamic system (figure 1) cannot yield appropriate understanding/

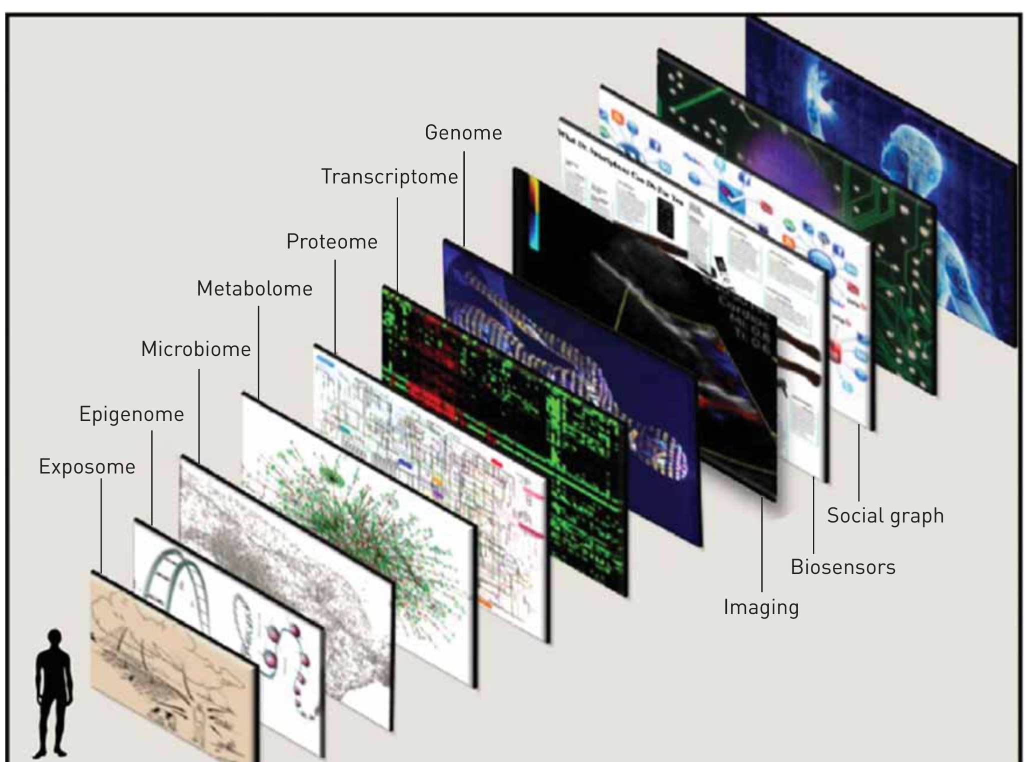

FIGURE 1 Multilevel layers of biological, environmental and social information ideally integrated in systems biomedicine approaches. For further explanations, see text. Reproduced and modified from [2] with permission.

prediction of the global behaviour of the system (so-called emergent properties, which are implicit in nonlinear systems) nor allow to fix it if found globally away from an homeokinetic state (*e.g.* disease *versus* health), with alterations that may spread throughout various levels or compartments of the system [5]. As Macklem [5] pointed out, emergent properties arise spontaneously as self-organised order from the nonlinear interactions of the different biological components and thus the overall emergent behaviour transcends the behaviour from each part in isolation. It follows that a more holistic approach, integrating information of the interacting parts and subentities into a single mathematical representation or model, can potentially offer better clues as to the causal chain of events that leads to the apparent phenotypic manifestations and how to remedy the situation [6]. Therefore, systems biology departs from the reductionist approach followed by traditional biomedical research by integrating (rather than taking apart) different biological levels (genes, molecules, cells, organs and the environment) and mechanisms, and shares a very similar goal with integrative physiology: to better understand holistically the systemic dynamic state of individuals [7, 8]. In this context, systems biology (and systems or network medicine) is nothing more than physiology, which has always meant to be multiscale and integrative [7, 8]. The difference is that today's availability of new tools, high-throughput technologies and computing power allows, for the first time, real physiology to be performed. In essence, it is all about perspective [9]. Before "perspective" (*i.e.* three-dimensional) painting was "invented", classical painting considered only two dimensions. Systems biology includes many different biological levels (dimensions) as well as the element of time dynamics. Hence, it has the potential to provide a much better definition for "the eye of the beholder" [9].

Biology as an informational science

In recent decades, faced with the biological complexity of human diseases, biomedical scientists have increasingly turned their efforts to apply high-throughput methodologies that embrace the Cartesian view that the human body is a system of formally interacting parts and that biology is an informational science. A nonexhaustive list of information sources (table 1) includes "omics" data ((epi-)genomic, transcriptomic, proteomic, metabolomic and microbiomic), single-cell analyses, phenotypic assays, extensive medical records and an endless list of environmental factors ("exposome"), such as smoking, exercise, diet and pollution, among others (figure 1). Common respiratory-specific levels of information are lung function and imaging.

System representation: networks

A network (or graph) is a practical graphical representation of complex data in the context of systems approaches (figure 2), where nodes are the elements of the system under study (*e.g.* genes, proteins, biochemical or physiological measures, individuals or patients, among many others) and edges (or links) connect nodes that interact somehow (causality, correlation). The network(s) constructions are hypothesis driven, *i.e.* there is not a single, fixed, network "template"; on the contrary, they can be "custom-made". Networks are used to make inferences regarding the emergent dynamic (spatial and temporal) behaviour of the system in response to perturbations of putative critical network elements (nodes and/or edges).

Diseases as network perturbations

Any disease can be viewed as a system in an abnormal state (a perturbed network) far from homeokinetic operating conditions [5], either with: 1) associated nonemergent (*i.e.* subclinical) alterations, or 2) observable phenotypic abnormalities (*i.e.* clinical symptoms) progressively departing from functional equilibrium towards partial system collapse (*i.e.* organ failure, *etc.*) or complete collapse (death). In opposition, perfect health, or wellness, can be viewed as the optimal and quantifiable state of a system in dynamic equilibrium (*i.e.* homeokinesis [5]).

Biological network properties

Several aspects of biomedical networks are due to their particular biological nature and must always be considered in a research setting [16]. In terms of "topology" (*i.e.* their spatial distribution) they are generally scale-free (as opposed to random networks). In this setting, "scale-free" means that this type of network contains many nodes with few connections and a few nodes with many links (hubs) (figure 2). This topology makes networks more robust against random perturbations [17] because of their higher modularity [18]. They are composed of loosely connected subparts (modules), which are groups of nodes highly connected internally but little to outsiders. Modules are usually coupled with specialised biological subtasks. Additionally, not all nodes are equal relative to the network structure. Central elements that are much more connected than the average are denominated "hub" nodes [19], while linkers between modules are termed "bottleneck" nodes (figure 2) [20]. Perturbations of these elements (hubs and bottlenecks) often alter the system behaviour drastically, whereas the impact of more peripheral nodes on systems behaviour (emergent properties) is often marginal. Other influential network properties with regard to the

TABLE 1 Common omics data types

	Assay	Platform	Main advantages and disadvantages	Standard bioinformatics pipelines
Genomics	Identify nucleotide variants (SNPs) in the whole genome associated with clinical traits (GWAS)	Genotyping arrays, whole-exome sequencing	SNP variability is stable during life; provides limited information in complex diseases due to several loci implicated	GWAS protocol review [10]
Transcriptomics	Quantify expression levels of cellular transcripts (*e.g.* mRNA)	Expression arrays, RNA sequencing	Widely used due to its high information content on cell status; differences in mRNA expression do not imply differences in proteins; does not take into account post-transcriptional modifications	RNA sequencing pipelines review [11]
Proteomics	Characterise protein expression levels of cells/samples	MS-based approaches	Expected to be closer to the phenotype; not widely used, expensive and more cumbersome analysis	Next-generation proteomics review [12]
Metabolomics	Characterise abundance profile of metabolites and their relative ratios	MS-based approaches	Representative of the cellular status; applicable to many biological fluids (*i.e.* breath, blood, urine, *etc.*); not widely used	Review of analytical methods for metabolomics [13]
Epigenomics	Determine modifications in DNA and small RNA that interfere with gene expression	DNA methylation analysis with arrays (Infinium MethylationEPIC 850K; Illumina, San Diego, CA, USA), next-generation sequencing, small RNA sequencing, arrays, *etc.*	Provides additional information to transcriptomics; related to exposures; more expensive than transcriptomics; sequencing-based approaches have computational tools in active development	Bioinformatics aspect of DNA methylation studies [14]
Microbiomics	Characterise bacterial (and viral) composition of a sample	Targeted sequencing of 16S rRNA gene, shotgun metagenomics sequencing	Provides information of external factors likely to be associated with disease; 16S sequencing does not differentiate between the presence of live/dead bacteria	Bioinformatics analysis for the characterisation of the human microbiome [15]

SNP: single nucleotide polymorphism; GWAS: genome-wide association study; MS: mass spectrometry.

robustness of the system include "redundancy" and "degeneracy" [21]. Finally, nodes and edges may be characterised qualitatively (*e.g.* fold-change sign for nodes that represent gene products) or quantitatively (*e.g.* chemical binding constant for edges that connect drug ligands to their target molecules) (figure 2).

Medical uses

Although systems biology is best suited for experimental models of disease, it can also provide actionable and useful insights in clinical medicine [22–24]. Systems (network) medicine can lead to the identification of disease biomarkers or drug targets, both defined as key nodes whose perturbation transits the state of the biological system from health to disease or *vice versa*. A paradigmatic example comes from the field of cancer and the observation that the sequential use of anticancer drugs enhances cell death by rewiring apoptotic signalling networks [25].

P4 medicine

The holistic approach of systems biology discussed earlier has enabled the emergence of a new comprehensive paradigm in medicine, called P4 medicine, for predictive, preventive, personalised and participatory [4, 26–28].

From treatment to prediction and prevention

Current western medicine mostly focuses on treating diseases and symptoms when they appear. Thus, the current healthcare system organisation (and its major stakeholders, *i.e.* hospitals and primary care centres, pharmaceutical industry, insurance companies, policy makers, providers (*e.g.* physicians) and patients) is based on the provision of medication and related health products to individuals once they are sick and

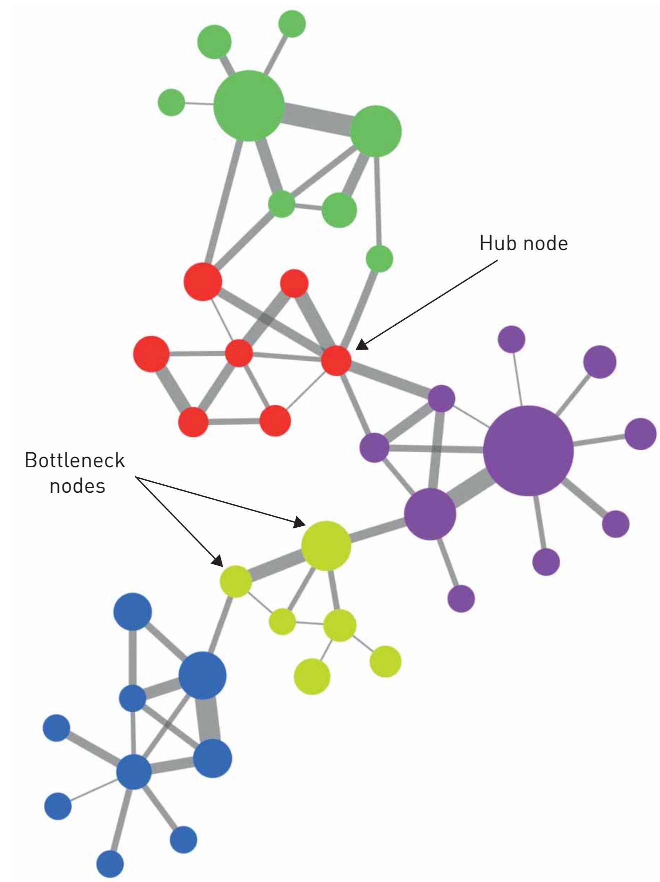

FIGURE 2 Network topology. Nodes are linked by edges. Node size represents a quantifiable node property (*e.g.* fold-change in two different experimental situations; this allows the inclusion of a dynamic component (*i.e.* time change) in the graphical representation of the network). Edge width represents the connection strength (*e.g.* correlation coefficient). Node colours identify different network modules. For further explanations, see text.

seeking treatment. In most countries, only a small fraction of public funds is currently devoted to prevention. Meanwhile, the burden and cost of chronic complex conditions (e.g. asthma, COPD, heart disease, stroke, cancer, type 2 diabetes, obesity and arthritis) is rising at an alarming rate [29, 30]. Hard or impossible to treat, these conditions may, however, be preventable to some (large?) extent. The hope to reach this goal (*e.g.* maintaining wellness) is fuelled by the growing understanding of risk factors and pathobiology of these diseases, gained (in part) as a result of the implementation of systems biomedicine approaches into research studies [31].

Personalised in nature

Since the dynamic life-long (from pre-womb to tomb [2]) interaction of genetic, environmental and social factors is what drives the physical state of individuals, and because no two individuals are biologically alike [2], the ideal preventive strategy should be tailored to each individual. In this setting, access to the individualised clouds of data, including omics data, digital medical records, and information related to the exposome, behaviour and social exposures, will be needed (and available).

Blood as a window for health assessment (liquid biopsies)

Blood stands out on the list of all the personalised data sources as it conveniently harbours dynamic critical information from many organs in the form of circulating organ-specific proteins, immune or signalling small molecules and cells (liquid biopsy). Moreover, cheap and mainstream nanosensor technology to measure these analytes longitudinally is on its way [32]. In the near future, these technologies may well serve to alert individuals in real-time to any high-risk alteration from healthy baseline measurements in order to prevent clinical complications such as organ failure, heart attacks [33], prion disease [34], liver injury [35], cancer recurrence [36], diabetes [37] or asthma attacks [38]. This scheme is expected to be especially powerful when combined with personalised genomic data, as well as other

https://doi.org/10.1183/16000617.0110-2017

biosensors continuously tracking essential variables, such as exhaled breath [39], urine [40], imaging [41] and/or ambient pathogens or allergens [42–44].

Participatory driven

Finally, the benefits of this new P4 medicine will only be possible if patients and healthy subjects become active agents in the continuous assessment and preservation of their health. The role of health providers, both traditional (physicians, nurses, physiotherapists) and novel (genetic counsellors, behavioural coaches), will evolve to facilitate actionable information to individuals, which they can use to maintain their health [45]. Importantly, a new legal framework of rights, obligations and protections for individuals/patients and health professionals alike remains to be established and implemented. The emergence of personalised "big" data repositories raises unprecedented ethical, privacy, confidentiality, security and policy issues related to information ownership, access and management. Of note, the insurance company regulatory framework is markedly unprepared in most countries.

How to do it?

Research strategy

In principle, there are two different approaches to analyse data in this setting: "supervised" analysis based on *a priori* knowledge (*e.g.* clinical characteristics of patients) and "unsupervised" analysis (*i.e.* hypothesis-free). Both strategies have advantages and disadvantages, and in a sense they are complementary; their characteristics are further discussed in the Analytical complexity section.

Input data

Systems biology leverages from several omics data types. The most commonly used data types are genomics, transcriptomics, proteomics, metabolomics, epigenomics and microbiomics. Table 1 summarises their definitions, available experimental platforms, advantages/disadvantages and the bioinformatics tools needed. In each omic, data is curated, normalised and the differences between groups are usually computed using general linear models [46, 47]. We acknowledge that exposomics and imaging are missing in table 1; this is on purpose as both fields are currently developing very actively [48, 49].

Analytical complexity

Single-level analysis

A common research approach is to perform standard (supervised or unsupervised) single-level omic analysis (table 1) and then use further bioinformatics tools to facilitate the translational interpretation (table 2 and figure 3). For instance, from a list of genes/proteins of interest, in order to identify underlying biological mechanisms, functional enrichment can be performed against many databases that host annotated information on functional roles (figure 3d): Gene Ontologies of biological processes, cellular components or molecular functions [62], KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways [63], Reactome pathways [64] and gene set enrichment analysis (GSEA) [50]. Furthermore, the

TABLE 2 Widely used tools to facilitate biomedical interpretation from omics analysis

Analytical tool	Goal	Advantages and disadvantages	Pipelines
Functional enrichment	From lists of identifiers (commonly genes) computes the over-representation in a specific molecular function, Gene Ontology, pathway, biological process, cell localisation, *etc.*	Noise and dimension reduction, helps interpret gene sets; useful to aggregate the individual gene contribution to overall changes; results are dependent on database knowledge and thus may be biased	Gene set enrichment analysis (GSEA): http://software.broadinstitute.org/gsea/index.jsp [50]; gene set variation analysis (GSVA) [51]; Enrichr: http://amp.pharm.mssm.edu/Enrichr [52]; FunRich: http://funrich.org [53]; STRING: https://string-db.org [54]
Data clustering	Classifies samples/variables based on their similarity in order to obtain homogeneous groups	Unsupervised, data driven and probabilistic; requires medium/large data sets	*k*-means [55, 56]; hierarchical bottom-up [57]; hierarchical top-down (divisive analysis clustering (DIANA)) [58]
Coexpression networks	From the dataset builds a correlation network to identify groups of related genes (modules), which can be investigated for biological functions and/or related to clinical traits	Coexpression in order to reflect causative processes must be coupled with functional enrichment and validation; correlations are affected by sample size of the dataset; requires proper data normalisation	Weighted gene coexpression networks analysis (WGCNA) [59]; conventional coexpression measures (Pearson/Spearman/Kendall, mutual information [60]); miRNA (targets)–genes [61]

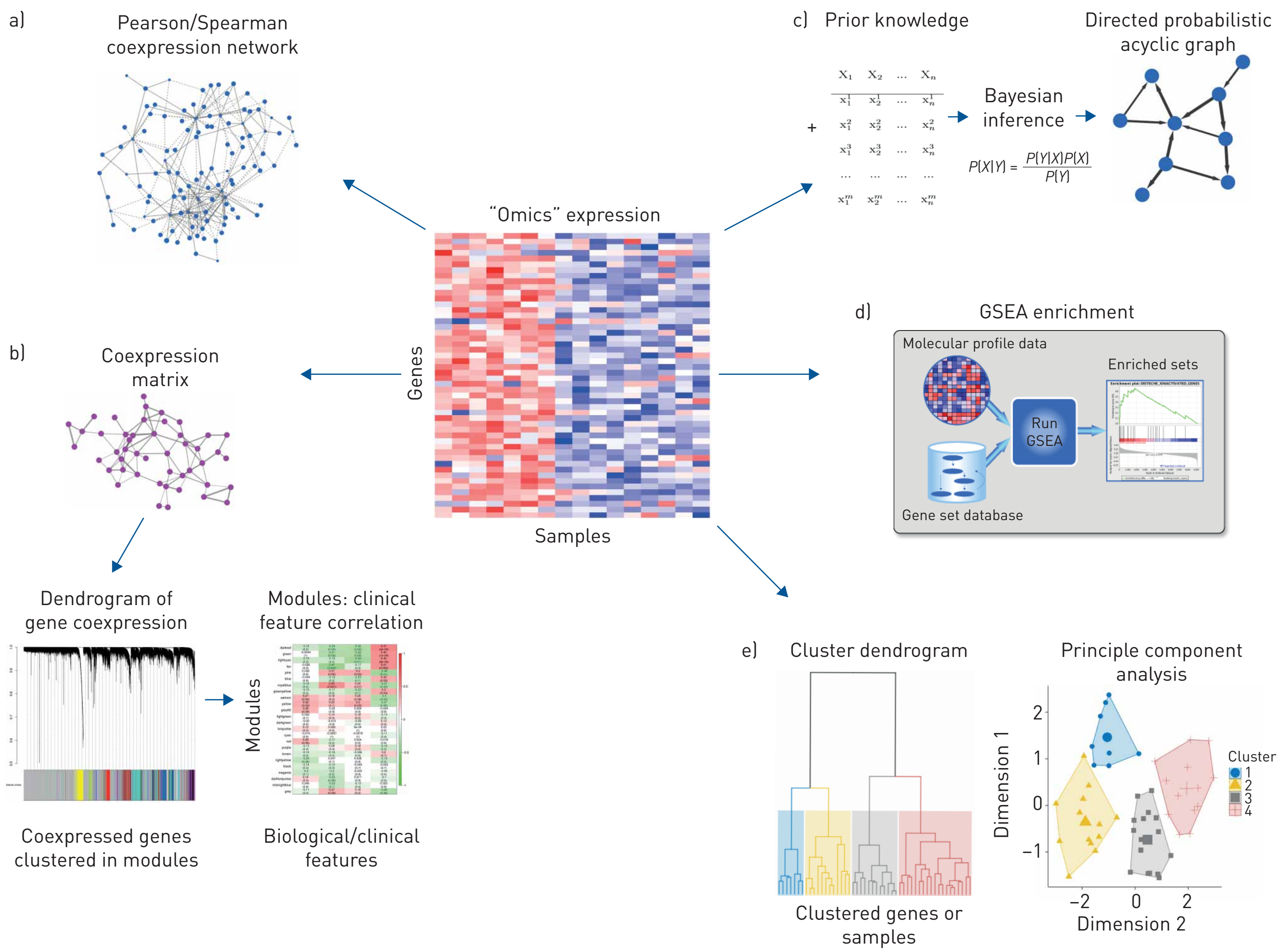

FIGURE 3 Summary of bioinformatic methodologies currently available. a) Correlation (*e.g.* Pearson/Spearman) network constructed from omics data. b) Weighted gene coexpression networks analysis (WGCNA) methodology. c) Bayesian networks approach. d) Gene set enrichment with gene set enrichment analysis (GSEA). e) *k*-means clustering. For further explanations, see text.

informational power can be augmented ("integrated") with additional molecular interactions available from public sources with the help of dedicated software (*e.g.* Cytoscape plugins [65] for the STRING protein–protein association database [54]). These exponentially growing public repositories host pathway interactions (*e.g.* gene regulatory, metabolic or signalling) of known relationships or putative associations that come from a wide range of experiments: theoretical, animal, experimentally modified (perturbed) biological conditions, *etc.* Additionally, in order to obtain homogeneous groups (of variables or samples), systems medicine uses a large variety of methods for clustering data (figure 3e) [66]. The methods vary in clustering criteria, computational efficiency/speed and cluster outputs. To stratify patients, relatively simple methods such as unsupervised learning algorithms (*e.g. k*-means [55, 56]) are commonly applied (table 2). However, more sophisticated methods can yield hierarchically organised clusters visualised as dendrograms: bottom-up agglomerative approaches (*e.g.* as described by WARD [57]) are preferred when clusters are expected to contain few observations, while top-down divisive approaches (*e.g.* DIANA (divisive analysis clustering) [58]) are more suited for estimating large numbers of clusters [67].

Finally, networks can be used to infer the structure of the data (table 2). Co-expression, co-occurrence or similarity networks can be built either across all study samples or separately for each medical condition that is to be compared (figure 3a). Many methods have been devised and are continuously under active development for building these networks. The simplest approach is to compute all gene–pair correlations with conventional coexpression measures (Pearson/Spearman/Kendall, mutual information [60]) (figure 3a), but more complex statistical procedures exist that cater for the specific features of biological systems. An extensively used procedure is termed weighted gene coexpression networks analysis (WGCNA) (figure 3b) [59, 68], which has the added value of clustering genes into nonoverlapping modules and correlating them with any imputed (usually clinical) characteristic. WGCNA can be complemented with functional enrichment analysis.

Multilevel analysis: the true revolution

Although some studies have and will continue to work successfully on a single omic level, recent decades have seen an ever-increasing body of work where several distinct omics datasets, including also other biological or clinical levels, are analysed conjointly using multiscale integrative methods such as SNF (similarity network fusion) [69]. This combination of levels has the potential to provide researchers with simultaneous information from several compartments of the biological system of interest, thus facilitating the modelling of the dynamic nonlinear relationships that characterise emergent properties (phenotypes) and complex diseases. Accordingly, this strategy would be able to provide more power to identify groups of patients affected with the same pathobiological mechanism or more power to probabilistically model (without understanding) the health *versus* disease states. The main multiscale analytical tools described to date are summarised in table 3. The "supervised" methods can be grouped mostly into either network-based, machine learning or multistep approaches [86], while the "unsupervised" can be further classified as based primarily on networks, Bayesian approaches or matrix factorisation (table 3).

Current applications of systems approaches in respiratory medicine

The pathogenesis of most common respiratory diseases is complex and largely undefined from a precise pathobiological point of view. Chronic respiratory conditions, such as asthma or COPD, are still diagnosed (and treated) based on respiratory symptoms and traditional lung function measures, but they are highly heterogeneous and often overlap. In fact, they are the end result of complex genetic and environmental interplays that are yet to be explicitly modelled. This poorly defined characterisation of the basic disease mechanisms results in nonspecific, mostly symptom-driven treatment options, or lack thereof, that may eventually be able to slow the progression of these diseases in fortunate, responsive patients.

Systems biology and network medicine approaches are being put forth in an effort to palliate this painful lack of knowledge and understanding by tackling two fundamental and interrelated matters: 1) as in other biomedical fields such as cancer, a novel classification (*i.e.* "taxonomy") of chronic airway diseases is needed, based not on clinical presentation (*i.e.* "phenotypes") but instead either on the underlying biological mechanisms (*i.e.* "endotypes") when characterised or resulting directly from data-driven probabilistic clustering of patients data; and 2) a more precise patient stratification that can be transferred to distinct and personalised preventive or therapeutic prognosis as well as improved prognosis (*i.e.* P4 medicine) is also needed, as recently highlighted in a review focused on biological therapies for airway diseases [87].

COPD

COPD is a heterogeneous disease with pulmonary and extrapulmonary manifestations [88], and variable response to pharmacological treatment [89], suggesting that the condition affects several distinct biological pathways. To characterise this heterogeneity at the molecular level, several studies have already used a number of different systems approaches. 1) WGCNA and GSEA showed that a molecular signature composed of gene modules related to B-cell activity, NK-cell activity or viral infection cellular markers might be detectable in peripheral blood months following COPD exacerbations [90]. 2) Xue *et al.* [91] used other network-centric procedures to reveal an unexpected loss of inflammatory signature in COPD patients, as well as an activation-independent core signature for human and murine macrophages. 3) Glass *et al.* [92] used the network inference analysis PANDA (Passing Attributes between Networks for Data Assimilation) [93], designed for improved integration of individual with public datasets, and discovered network rewiring of lymphocyte activation signalling circuits in a known gene variant implicated in COPD by genome-wide association studies. 4) Faner *et al.* [94] unravelled differences in the molecular pathogenesis of emphysema and bronchiolitis by performing correlation network analysis of lung transcriptomics on COPD patients. They found that B-cell-related genes were significantly enriched in emphysema (compared with COPD patients without emphysema), paving the way for differential therapeutic research on inflammatory pathways of the adaptive immune response. 5) Two COPD studies demonstrated the utility of unsupervised *k*-means clustering by identifying robust cluster associations with clinical characteristics and known COPD genetic variants [95, 96]. 6) Very recently, Ross *et al.* [97] introduced a new Bayesian method for COPD subtyping. They applied it to the COPDGene cohort and identified nine different patient subgroups with distinct disease progression trajectories. Of note, Ross *et al.* [97] prove that their sophisticated model has a better predictive capacity than multivariate ordinary least squares regression analysis.

TABLE 3 Analytical strategies for integrative multiomics analysis

Analytical tool	Strategy	Implementation features	Advantages and disadvantages	Pipelines
Unsupervised				
Network-based methods (variables)	Identification of subgroups of variables that span several data types	Coexpression measures are used to build a coexpression network across all data types, where variables are nodes and links represent correlations; clustering algorithms are latter applied to achieve groups of variables	Outperforms single data type network analysis; combines the network with known pathways; large multilayer network density complicates biological interpretation	PAthway Representation and Analysis by DIrect Reference on Graphical Models (PARADIGM) [70]; Lemon-Tree [71]
Network-based methods (patients)	Classifies patients on the basis of the similarity of their omic data	Constructs networks of samples for each available data type and efficiently fuses these into one comprehensive network	Identifies homogeneous individual clusters; goes beyond dichotomous patient classification by capturing continuous phenotypes; does not provide information on which biological features cluster individuals	Similarity network fusion (SNF) [69]
Bayesian methods	With the different layers of omics builds a Bayesian model of relations with the aim to create classifiers when related to clinical information	Computes probabilistic relationships among variables that can express mutual dependencies between them	Allows prior assumptions for each data type distribution and between data sets; good at modelling nonlinear relationships; high computational cost; difficulty of choosing prior distributions	Multiple dataset integration (MDI) [72]; patient-specific data fusion (PSDF) [73]; Bayesian consensus clustering (BCC) [74]; COpy Number and EXpression In Cancer (CONEXIC) [75]
Matrix factorisation methods	Projects variations among data sets onto dimension-reduced space	Algorithms aim to unravel a latent data matrix of reduced dimensionality that best explains observed variables' variations among all data sets	Noise, dimension and heterogeneity reduction; requires heavy computation time and large memory	Joint non-negative matrix factorisation (NMF) [76]; iCluster+ [77]; joint Bayes factor [78]
Supervised				
Network-based methods	Clinically meaningful groups are chosen as input for different network construction	Network construction relies on clinical characteristics of patients so as to allow analysis of underlying networks	Network models are tailored for prognosis or diagnosis; comparison of resulting networks is not straightforward	Analysis Tool for Heritable and Environmental Network Associations (ATHENA) [79]; jActiveModules [80]
Machine learning	Clinical covariates and omics data are included in a machine learning model for prediction or classification	Machine learning methods use various kernel-based frameworks for data transformation, integration and classifier training	The model obtained is better when large amounts of data are used for training the system; machine learning kernel parameterisation can be difficult	Semidefinite programming/support vector machine (SDP/SVM) [81]; feature selection multiple kernel learning (FSMKL) [82]
Multistep analysis	A step-by-step semiautomated data integration process	Works independently on the different layers prior to integrating them by identifying variables differentially expressed in several layers	User has more control flexibility over the workflow, filtering and selection of relevant biological/clinical information; relatively lower statistical or predictive power	Integrative Bayesian analysis of genomics data (iBAG) [83]; multiple concerted disruption (MCD) [84]; Anduril [85]

Asthma

Several international consortia have already applied systems biology and network medicine approaches to asthma research. 1) ADEPT (Airways Disease Endotyping for Personalised Therapeutics) and U-BIOPRED (Unbiased Biomarkers for the Prediction of Respiratory Disease Outcome Consortium) have, for instance, applied a clustering algorithm to two independent asthma cohorts on a small set of easily measurable clinical variables and successfully defined four longitudinally stable clusters of patients with distinct clinical and biomarker profiles (from blood, sputum and airway data) [98]. 2) Kuo *et al.* [99] recently reported three novel molecular phenotypes of asthma in the U-BIOPRED cohort by analysing sputum cell transcriptomics in asthmatic and nonasthmatic subjects. They applied hierarchical clustering of differentially expressed genes as well as gene set variation analysis, gene–protein coexpression and pathway enrichment analysis. 3) Sharma *et al.* [100] used network-based tools to analyse the predictive value of the asthma interactome, and characterised high-impact pathways central to the disease heterogeneity and drug response. 4) Qiu *et al.* [101] used PANDA on participants of the Childhood Asthma Management Program cohort to assess the differential connectivity between the gene regulatory network of good responders to inhaled corticosteroids *versus* that of poor responders. The method allowed them to integrate their dataset with public data interactions of genes, transcription factors and proteins, and eventually implicate several network hubs and transcription factors (as well as regulatory rewiring) in the heterogeneity of drug treatment effects. Specifically, the differential network topology of good responders *versus* that of poor responders revealed enriched corticosteroid-induced pro-apoptosis pathways in the former and anti-apoptosis pathways in the latter, as well as key regulatory transcription factors (hubs) that drove differential downstream gene expression in the two groups.

Lung cancer

Lung cancer is the leading cause of cancer death in the world. Lung cancer is highly heterogeneous genetically because of a high mutation rate, as well as extremely complex since it comprises a disparate subset of diseases with distinct and possibly overlapping pathobiologies that share a common phenotypic manifestation. Smoking is a core shared risk factor for COPD and lung cancer; up to 65–70% of lung cancer patients suffer both lung cancer and COPD [102, 103]. So far, no single satisfactory circulating (*i.e.* liquid biopsy) tumour marker has been properly validated, but recently a panel of six tumour markers showed a very high specificity and sensitivity in patients referred to a tertiary hospital because of the clinical suspicion of lung cancer [104, 105]. Given that inherited genetic variants play a significant role in lung cancer development [106], but contribute little to risk estimates of classical predictive statistical models [107–109], it is hoped that systems biology approaches will allow the comparison multilevel high-throughput omics data between tumour and normal tissue, and facilitate the identification of early diagnostic lung cancer biomarkers. WGCNA has already been used successfully in lung cancer research. 1) Tang *et al.* [110] related the gene expression profile of lung squamous cell carcinoma with five differentially expressed long noncoding RNAs that could help in prognosis evaluation. Their gene signature was statistically associated with overall survival in important clinical subsets (stage I, epidermal growth factor (EGFR) wild-type and EGFR mutant). 2) Tian *et al.* [111] analysed coexpression networks and protein–protein interactions of data available in public repositories (The Cancer Genome Atlas, KEGG and Gene Ontology).

What's next? Future challenges

For the successful development and implementation of systems biology and network medicine approaches in respiratory medicine, several challenges need to be faced and eventually solved.

Technical challenges

In any clinical study, only a fraction of the biological variability is captured (and therefore analysed) due to technical limitations (and cost) of the experimental tools available. The development of new experimental tools (*e.g.* high-throughput next-generation sequencing, mass spectrometry-based flow cytometry or real-time molecular imaging) will generate new information but, at the same time, massive amounts of (big) data that will have to be adequately handled, analysed and interpreted [112–114]. In this context, Riekeberg and Powers [115] recently reviewed the methodological advancements and successful applications of metabolomics, one the newest omic fields.

However, research would benefit not only from measuring "more" relevant variables, but also from estimating with better precision those variables already determined in the context of a more complete definition of reference and pathological ranges (that vary in time, across individuals and biological codeterminants) [116]. Of the variability supposedly present in experimental data, these currently unaccounted factors and batch effects should not be underrated since they can partly explain the general difficulty to replicate scientific findings in the biomedical field, of which respiratory biomedicine is not exempt.

Computational challenges

Computational methodologies and programming analytical tools are being constantly refined as they translate advancements from complementary areas such as AI and information science. However, challenges and difficulties remain. For instance, in differential expression (omics) analysis, one of the main difficulties is to accurately separate the biological signal from technical noise. As risk factors of complex diseases are polygenic (individualised genes have little independent effects), batch effects and technical heterogeneity are difficult to separate from protein, gene and epigenetic perturbations causing the disease. Computational analysis partly palliates this difficulty by integrating enormous amounts of data into models or network representations. However, there is not a single bioinformatic approach for the task and each has its own best-use cases, advantages and drawbacks. Regrettably, this lack of standard biostatistical and algorithmic procedures may create a lot of uncertainty as to the validity of the results and usually cannot fully guarantee reproducibility [117]. It is thus extremely important that bioinformatic researchers assess the sensitivity and optimal network thresholds of their implementations. Thus, replication and experimental validation of results must remain a research priority.

Drug discovery challenges

Network (poly) pharmacology differs from conventional drug discovery strategies by providing a powerful rationale for target identification that is based on the analysis of disease-specific biological networks. This novel paradigm consists of administering simultaneously multiple small molecules to target several biochemical network nodes in an attempt to "re-engineer" the network into its normal and healthy dynamic structure [118]. It has the potential to overcome the two major obstacles hindering the field: efficacy and toxicity. This new approach still requires considerable methodological developments. Proof-of-concept was obtained by combining methods as diverse as network analysis, text mining, molecular docking data and the STRING database [54] to integrate data from network pharmacology and metabolomics [119].

Healthcare system, educational and economic challenges

The milestones required for systems medicine to become a reality go far beyond mere scientific and technological progress. The structure of the healthcare system needs to substantially adapt to operate with multidisciplinary teams [113] of traditional (physicians, epidemiologists, computational biologists, IT specialists, statisticians) and new roles (genetic counsellor [45], behavioural coach, specialised educators), which cannot function without specific omics data storage facilities, diagnosis centres, standard analytical pipelines and managerial frameworks. Furthermore, systems biology and P4 medicine require specific education (*via* higher education degrees, complementary formations for hospital personnel or genomics training programmes [120]) since there is a growing gap between the amount of data generated by basic scientists and the clinical expertise available to analyse, interpret and translate it into clinical practice. Finally, because of the high cost of overcoming these challenges, there is a significant risk that individuals in developing countries will not benefit from this health revolution [121], unless guided by the appropriate political efforts from the international community and an informed public opinion [122–124]. Even within the richest nations, equitable access to P4 medicine is not guaranteed, and low-income individuals may not be able to afford the unsubsidised cost of omics data integration, diagnostic and clinical care [125, 126].

Conclusions

Human health and disease are emergent properties of a complex, multilevel and dynamic system: the human body. Systems biology and network medicine are comprehensive research strategies that have the potential to understand the emergent properties of the system holistically. By doing so, they are paving the way for a radical shift in medical practice that is evolving from a reactive proposition to a predictive, preventive, personalised and participatory (P4) approach. Respiratory medicine is in fact already contributing significantly to this change by leading the field of data-driven management [98, 99] as well as by applying multilevel network analysis to a variety of clinical conditions [127].

Acknowledgements

The authors thank the two anonymous reviewers of our manuscript for their very helpful and constructive criticisms and suggestions.

References

1 Agusti A, Celli B, Faner R. What does endotyping mean for treatment in chronic obstructive pulmonary disease? *Lancet* 2017; 390: 980–987.
2 Topol EJ. Individualized medicine from prewomb to tomb. *Cell* 2014; 157: 241–253.
3 Auffray C, Sagner M, Abdelhak S, *et al.* Viva Europa, a land of excellence in research and innovation for health and wellbeing. *Prog Prev Med* 2017; 2: e006.
4 Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. *Genome Med* 2009; 1: 2.
5 Macklem PT. Emergent phenomena and the secrets of life. *J Appl Physiol* 2008; 104: 1844–1846.
6 Loos RJF, Schadt EE. This I believe: gaining new insights through integrating "old" data. *Front Genet* 2012; 3: 137.

7 Greenhaff PL, Hargreaves M. 'Systems biology' in human exercise physiology: is it something different from integrative physiology? *J Physiol* 2011; 589: 1031–1036.
8 Grocott MP. Integrative physiology and systems biology: reductionism, emergence and causality. *Extrem Physiol Med* 2013; 2: 9.
9 Snyder LJ. Eye of the Beholder: Johannes Vermeer, Antoni van Leeuwenhoek, and the Reinvention of Seeing. New York, Norton, 2015.
10 Hamberg M, Backes C, Fehlmann T, *et al.* MiRTargetLink – miRNAs, genes and interaction networks. *Int J Mol Sci* 2016; 17: 564.
11 Teng M, Love MI, Davis CA, *et al.* A benchmark for RNA-seq quantification pipelines. *Genome Biol* 2016; 17: 74.
12 Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 2013; 14: 35–48.
13 Nobakht M Gh BF, Aliannejad R, Rezaei-Tavirani M, *et al.* The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis. *Biomarkers* 2015; 20: 5–16.
14 Adusumalli S, Mohd Omar MF, Soong R, *et al.* Methodological aspects of whole-genome bisulfite sequencing analysis. *Brief Bioinform* 2015; 16: 369–379.
15 Noecker C, McNally CP, Eng A, *et al.* High-resolution characterization of the human microbiome. *Transl Res* 2017; 179: 7–23.
16 Hu JX, Thomas CE, Brunak S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* 2016; 17: 615–629.
17 Barabasi AL, Albert R. Emergence of scaling in random networks. *Science* 1999; 286: 509–512.
18 Mitra K, Carvunis A-R, Ramesh SK, *et al.* Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet* 2013; 14: 719–732.
19 Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol* 2005; 22: 803–806.
20 Yu H, Kim PM, Sprecher E, *et al.* The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 2007; 3: e59.
21 Baffy G, Loscalzo J. Complexity and network dynamics in physiological adaptation: an integrated view. *Physiol Behav* 2014; 131: 49–56.
22 Schadt EE, Björkegren JLM. NEW: network-enabled wisdom in biology, medicine, and health care. *Sci Transl Med* 2012; 4: 115rv111.
23 Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature* 2009; 461: 218–223.
24 Silverman EK, Loscalzo J. Network medicine approaches to the genetics of complex diseases. *Discov Med* 2012; 14: 143–152.
25 Lee MJ, Ye AS, Gardino AK, *et al.* Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* 2012; 149: 780–794.
26 Hood L, Heath JR, Phelps ME, *et al.* Systems biology and new technologies enable predictive and preventative medicine. *Science* 2004; 306: 640–643.
27 Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J Proteome Res* 2004; 3: 179–196.
28 Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. *N Biotechnol* 2012; 29: 613–624.
29 Leadley RM, Armstrong N, Lee YC, *et al.* Chronic diseases in the European Union: the prevalence and health cost implications of chronic pain. *J Pain Palliat Care Pharmacother* 2012; 26: 310–325.
30 Stallard E. Estimates of the incidence, prevalence, duration, intensity, and cost of chronic disability among the U.S. elderly *N Am Actuar J* 2011; 15: 32–58.
31 Harrington RA, Liu ET. Quantitative biology and clinical trials: a perspective. *In:* Liu ET, Lauffenburger DA, eds. Systems Biomedicine: Concepts and Perspectives. San Diego, Academic Press, 2010; pp. 415–424.
32 Ferguson BS, Hoggarth DA, Maliniak D, *et al.* Real-time, aptamer-based tracking of circulating therapeutic agents in living animals. *Sci Transl Med* 2013; 5: 213ra165.
33 Damani S, Bacconi A, Libiger O, *et al.* Characterization of circulating endothelial cells in acute myocardial infarction. *Sci Transl Med* 2012; 4: 126ra133.
34 Hwang D, Lee IY, Yoo H, *et al.* A systems approach to prion disease. *Mol Syst Biol* 2009; 5: 252.
35 Wang K, Zhang S, Marzolf B, *et al.* Circulating microRNAs, potential biomarkers for drug-induced liver injury. *Proc Natl Acad Sci USA* 2009; 106: 4402–4407.
36 Iinuma H, Watanabe T, Mimori K, *et al.* Clinical significance of circulating tumor cells, including cancer stem-like cells, in peripheral blood for recurrence and prognosis in patients with Dukes' stage B and C colorectal cancer. *J Clin Oncol* 2011; 29: 1547–1555.
37 Sinnott M, Kinsley BT, Jackson AD, *et al.* Fasting plasma glucose as initial screening for diabetes and prediabetes in Irish adults: the Diabetes Mellitus and Vascular health initiative (DMVhi). *PLoS One* 2015; 10: e0122704.
38 Mosca T, Menezes MC, Dionigi PC, *et al.* C3 and C4 complement system components as biomarkers in the intermittent atopic asthma diagnosis. *J Pediatr* 2011; 87: 512–516.
39 Lawal O, Ahmed WM, Nijsen TME, *et al.* Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis. *Metabolomics* 2017; 13: 110.
40 Gonzalez-Guerrero AB, Maldonado J, Dante S, *et al.* Direct and label-free detection of the human growth hormone in urine by an ultrasensitive bimodal waveguide biosensor. *J Biophotonics* 2017; 10: 61–67.
41 Atukorale PU, Covarrubias G, Bauer L, *et al.* Vascular targeting of nanoparticles for molecular imaging of diseased endothelium. *Adv Drug Deliv Rev* 2017; 113: 141–156.
42 Chen R, Mias GI, Li-Pook-Than J, *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 2012; 148: 1293–1307.
43 Stanberry L, Mias GI, Haynes W, *et al.* Integrative analysis of longitudinal metabolomics data from a personal multi-omics profile. *Metabolites* 2013; 3: 741–760.
44 Sperisen P, Cominetti O, Martin FP. Longitudinal omics modeling and integration in clinical metabonomics research: challenges in childhood metabolic health research. *Front Mol Biosci* 2015; 2: 44.

45 Shelton CA, Whitcomb DC. Evolving roles for physicians and genetic counselors in managing complex genetic disorders. *Clin Transl Gastroenterol* 2015; 6: e124.
46 Ritchie ME, Phipson B, Wu D, *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.
47 Dai Z, Sheridan JM, Gearing LJ, *et al.* EdgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. *F1000Res* 2014; 3: 95.
48 Smith MT, de la Rosa R, Daniels SI. Using exposomics to assess cumulative risks and promote health. *Environ Mol Mutagen* 2015; 56: 715–723.
49 Verma G, Palombo A, Grigioni M, *et al.* Systems biology-driven hypotheses tested in vivo: the need to advancing molecular imaging tools. *Methods Mol Biol* 2018; 1702: 337–359.
50 Subramanian A, Tamayo P, Mootha VK, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; 102: 15545–15550.
51 Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013; 14: 7.
52 Chen EY, Tan CM, Kou Y, *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013; 14: 128.
53 Pathan M, Keerthikumar S, Ang CS, *et al.* FunRich: an open access standalone functional enrichment and interaction network analysis tool. *Proteomics* 2015; 15: 2597–2601.
54 Szklarczyk D, Franceschini A, Wyder S, *et al.* STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015; 43: D447–D452.
55 Lloyd S. Least squares quantization in PCM. *IEEE Trans Inf Theory* 1982; 28: 129–137.
56 Eisen MB, Spellman PT, Brown PO, *et al.* Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; 95: 14863–14868.
57 Ward JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 1963; 58: 236.
58 Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. New York, Wiley, 2009.
59 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.
60 Song L, Langfelder P, Horvath S. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics* 2012; 13: 328.
61 Fan Y, Siklenka K, Arora SK, *et al.* miRNet – dissecting miRNA–target interactions and functional associations through network-based visual analysis. *Nucleic Acids Res* 2016; 44: W135–W141.
62 Ashburner M, Ball CA, Blake JA, *et al.* Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25: 25–29.
63 Kanehisa M, Furumichi M, Tanabe M, *et al.* KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017; 45: D353–D361.
64 Fabregat A, Sidiropoulos K, Garapati P, *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res* 2016; 44: D481–D487.
65 Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13: 2498–2504.
66 Kerr G, Ruskin HJ, Crane M, *et al.* Techniques for clustering gene expression data. *Comput Biol Med* 2008; 38: 283–293.
67 Maimon,O, Rokach L. Data Mining and Knowledge Discovery Handbook. New York, Springer, 2005.
68 Botía JA, Vandrovcova J, Forabosco P, *et al.* An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst Biol* 2017; 11.
69 Wang B, Mezlini AM, Demir F, *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014; 11: 333–337.
70 Vaske CJ, Benz SC, Sanborn JZ, *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* 2010; 26: i237–i245.
71 Bonnet E, Calzone L, Michoel T. Integrative multi-omics module network inference with Lemon-Tree. *PLoS Comput Biol* 2015; 11: e1003983.
72 Kirk P, Griffin JE, Savage RS, *et al.* Bayesian correlated clustering to integrate multiple datasets. *Bioinformatics* 2012; 28: 3290–3297.
73 Yuan Y, Savage RS, Markowetz F. Patient-specific data fusion defines prognostic cancer subtypes. *PLoS Comput Biol* 2011; 7: e1002227.
74 Lock EF, Dunson DB. Bayesian consensus clustering. *Bioinformatics* 2013; 29: 2610–2616.
75 Akavia UD, Litvin O, Kim J, *et al.* An integrated approach to uncover drivers of cancer. *Cell* 2010; 143: 1005–1017.
76 Zhang S, Li Q, Liu J, *et al.* A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. *Bioinformatics* 2011; 27: i401–i409.
77 Mo Q, Wang S, Seshan VE, *et al.* Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc Natl Acad Sci USA* 2013; 110: 4245–4250.
78 Ray P, Zheng L, Lucas J, *et al.* Bayesian joint analysis of heterogeneous genomics data. *Bioinformatics* 2014; 30: 1370–1376.
79 Kim D, Li R, Dudek SM, *et al.* ATHENA: identifying interactions between different levels of genomic data associated with cancer clinical outcomes using grammatical evolution neural network. *BioData Min* 2013; 6: 23.
80 Ideker T, Ozier O, Schwikowski B, *et al.* Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 2002; 18: Suppl. 1, S233–S240.
81 Lanckriet GR, De Bie T, Cristianini N, *et al.* A statistical framework for genomic data fusion. *Bioinformatics* 2004; 20: 2626–2635.
82 Seoane JA, Day IN, Gaunt TR, *et al.* A pathway-based data integration framework for prediction of disease progression. *Bioinformatics* 2014; 30: 838–845.
83 Jennings EM, Morris JS, Carroll RJ, *et al.* Bayesian methods for expression-based integration of various types of genomics data. *EURASIP J Bioinform Syst Biol* 2013; 2013: 13.

84 Chari R, Coe BP, Vucic EA, *et al.* An integrative multi-dimensional genetic and epigenetic strategy to identify aberrant genes and pathways in cancer. *BMC Syst Biol* 2010; 4: 67.
85 Ovaska K, Laakso M, Haapa-Paananen S, *et al.* Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme. *Genome Med* 2010; 2: 65.
86 Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. *Front Genet* 2017; 8: 84.
87 Tan H-TT, Sugita K, Akdis CA. Novel biologicals for the treatment of allergic diseases and asthma. *Curr Allergy Asthma Rep* 2016; 16: 70.
88 Agusti A, Soriano JB. COPD as a systemic disease. *COPD* 2008; 5: 133–138.
89 Hanania NA. The impact of inhaled corticosteroid and long-acting beta-agonist combination therapy on outcomes in COPD. *Pulm Pharmacol Ther* 2008; 21: 540–550.
90 Morrow JD, Qiu W, Chhabra D, *et al.* Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. *BMC Med Genomics* 2015; 8: 1.
91 Xue J, Schmidt SV, Sander J, *et al.* Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity* 2014; 40: 274–288.
92 Glass K, Huttenhower C, Quackenbush J, *et al.* Passing messages between biological networks to refine predicted interactions. *PLoS One* 2013; 8: e64832.
93 Lao T, Glass K, Qiu W, *et al.* Haploinsufficiency of Hedgehog interacting protein causes increased emphysema induced by cigarette smoke through network rewiring. *Genome Med* 2015; 7: 12.
94 Faner R, Cruz T, Casserras T, *et al.* Network analysis of lung transcriptomics reveals a distinct B-cell signature in emphysema. *Am J Respir Crit Care Med* 2016; 193: 1242–1253.
95 Cho MH, Washko GR, Hoffmann TJ, *et al.* Cluster analysis in severe emphysema subjects using phenotype and genotype data: an exploratory investigation. *Respir Res* 2010; 11: 30.
96 Castaldi PJ, Dy J, Ross J, *et al.* Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax* 2014; 69: 415–422.
97 Ross JC, Castaldi PJ, Cho MH, *et al.* A Bayesian nonparametric model for disease subtyping: application to emphysema phenotypes. *IEEE Trans Med Imaging* 2017; 36: 343–354.
98 Loza MJ, Adcock I, Auffray C, *et al.* Longitudinally stable, clinically defined clusters of patients with asthma independently identified in the ADEPT and U-BIOPRED asthma studies. *Ann Am Thorac Soc* 2016; 13: Suppl. 1, S102–S103.
99 Kuo CS, Pavlidis S, Loza M, *et al.* T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur Respir J* 2017; 49: 1602135.
100 Sharma A, Menche J, Huang CC, *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: 3005–3020.
101 Qiu W, Guo F, Glass K, *et al.* Differential connectivity of gene regulatory networks distinguishes corticosteroid response in asthma. *J Allergy Clin Immunol* 2017; in press [http://doi.org/10.1016/j.jaci.2017.05.052].
102 Young RP, Hopkins RJ, Christmas T, *et al.* COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. *Eur Respir J* 2009; 34: 380–386.
103 de Torres JP, Bastarrika G, Wisnivesky JP, *et al.* Assessing the relationship between lung cancer risk and emphysema detected on low-dose CT of the chest. *Chest* 2007; 132: 1932–1938.
104 Molina R, Marrades RM, Auge JM, *et al.* Assessment of a combined panel of six serum tumor markers for lung cancer. *Am J Respir Crit Care Med* 2016; 193: 427–437.
105 Holdenrieder S, Pagliaro L, Morgenstern D, *et al.* Clinically meaningful use of blood tumor markers in oncology. *Biomed Res Int* 2016; 2016: 9795269.
106 Timofeeva MN, Hung RJ, Rafnar T, *et al.* Influence of common genetic variation on lung cancer risk: meta-analysis of 14900 cases and 29485 controls. *Hum Mol Genet* 2012; 21: 4980–4995.
107 Raji OY, Agbaje OF, Duffy SW, *et al.* Incorporation of a genetic factor into an epidemiologic model for prediction of individual risk of lung cancer: the Liverpool Lung Project. *Cancer Prev Res* 2010; 3: 664–669.
108 Spitz MR, Etzel CJ, Dong Q, *et al.* An expanded risk prediction model for lung cancer. *Cancer Prev Res* 2008; 1: 250–254.
109 Young RP, Hopkins RJ, Hay BA, *et al.* Lung cancer susceptibility model based on age, family history and genetic variants. *PLoS One* 2009; 4: e5302.
110 Tang R-X, Chen W-J, He R-Q, *et al.* Identification of a RNA-Seq based prognostic signature with five lncRNAs for lung squamous cell carcinoma. *Oncotarget* 2017; 8: 50761–50773.
111 Tian F, Zhao J, Fan X, *et al.* Weighted gene co-expression network analysis in identification of metastasis-related genes of lung squamous cell carcinoma based on the Cancer Genome Atlas database. *J Thorac Dis* 2017; 9: 42–53.
112 Merelli I, Perez-Sanchez H, Gesing S, *et al.* Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives. *Biomed Res Int* 2014; 2014: 134023.
113 Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities. *BMC Med Genomics* 2015; 8: 33.
114 Gligorijevic V, Malod-Dognin N, Przulj N. Integrative methods for analyzing big data in precision medicine. *Proteomics* 2016; 16: 741–758.
115 Riekeberg E, Powers R. New frontiers in metabolomics: from measurement to insight. *F1000Res* 2017; 6: 1148.
116 Ozarda Y. Reference intervals: current status, recent developments and future considerations. *Biochem Med* 2016; 26: 5–16.
117 Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011; 12: 56–68.
118 Haanstra JR, Bakker BM. Drug target identification through systems biology. *Drug Discov Today Technol* 2015; 15: 17–22.
119 Chen S, Jiang H, Cao Y, *et al.* Drug target identification using network analysis: taking active components in *Sini* decoction as an example. *Sci Rep* 2016; 6: 24245.
120 Cesario A, Auffray C, Russo P, *et al.* P4 medicine needs P4 education. *Curr Pharm Des* 2014; 20: 6071–6072.

121 Hardy B-J, Séguin B, Goodsaid F, *et al.* The next steps for genomic medicine: challenges and opportunities for the developing world. *Nat Rev Genet* 2008; 9: Suppl. 1, S23–S27.
122 Li A, Meyre D. Jumping on the train of personalized medicine: a primer for non-geneticist clinicians: part 1. Fundamental concepts in molecular genetics. *Curr Psychiatry Rev* 2014; 10: 91–100.
123 Li A, Meyre D. Jumping on the train of personalized medicine: a primer for non-geneticist clinicians: part 2. Fundamental concepts in genetic epidemiology. *Curr Psychiatry Rev* 2014; 10: 101–117.
124 Li A, Meyre D. Jumping on the train of personalized medicine: a primer for non-geneticist clinicians: part 3. Clinical applications in the personalized medicine area. *Curr Psychiatry Rev* 2014; 10: 118–132.
125 Mardis ER. The $1,000 genome, the $100,000 analysis? *Genome Med* 2010; 2: 84.
126 Phillips KA, Sakowski JA, Trosman J, *et al.* The economic value of personalized medicine tests: what we know and what we need to know. *Genet Med* 2014; 16: 251–257.
127 Noell G, Cosío BG, Faner R, *et al.* Multi-level differential network analysis of COPD exacerbations. *Eur Respir J* 2017; 50: 1700075.