



Application of 'omics technologies to biomarker discovery in inflammatory lung diseases

Craig E. Wheelock^{1,2}, Victoria M. Goss³, David Balgoma^{1,2}, Ben Nicholas³, Joost Brandsma³, Paul J. Skipp⁴, Stuart Snowden^{1,2}, Dominic Burg^{3,4}, Arnaldo D'Amico⁵, Ildiko Horvath⁶, Amphun Chaiboonchoe⁷, Hassan Ahmed⁷, Stéphane Ballereau⁷, Christos Rossios⁸, Kian Fan Chung⁸, Paolo Montuschi⁹, Stephen J. Fowler¹⁰, Ian M. Adcock⁸, Anthony D. Postle³, Sven-Erik Dahlén^{2,11}, Anthony Rowe¹², Peter J. Sterk¹³, Charles Auffray⁷, Ratko Djukanović³ and the U-BIOPRED Study Group

Affiliations: ¹Dept of Medical Biochemistry and Biophysics, Division of Physiological Chemistry II, Karolinska Institutet, Stockholm, ²Centre for Allergy Research, Karolinska Institutet, Stockholm, and ¹¹Unit for Experimental Asthma and Allergy Research, Karolinska Institutet, Stockholm, Sweden. ³Clinical and Experimental Sciences, University of Southampton Faculty of Medicine and NIHR Southampton Respiratory Biomedical Research Unit, Southampton, ⁴Centre for Biological Sciences, University of Southampton Faculty of Natural and Environmental Sciences and Centre for Proteomic Research, Southampton, ⁵Airways Disease, National Heart and Lung Institute, Imperial College London, London, ¹⁰University of Manchester and Lancashire Teaching Hospitals NHS Foundation Trust, Respiratory Research Group, Education and Research Centre, Wythenshawe Hospital, Manchester, and ¹²Janssen Research and Development Ltd, High Wycombe, UK. ⁵Dept of Electronic Engineering, University of Rome "Tor Vergata", Rome, and ⁹Dept of Pharmacology, Faculty of Medicine, Catholic University of the Sacred Heart, Rome, Italy. ⁶Dept of Pulmonology, Semmelweis University, Budapest, Hungary. ⁷CNRS Institute of Biological Sciences, European Institute for Systems Biology and Medicine, Claude Bernard University, Lyon, France. ¹³Dept of Respiratory Medicine, Academic Medical Center University of Amsterdam, Amsterdam, The Netherlands.

Correspondence: R. Djukanović, Clinical and Experimental Sciences, University of Southampton, and Southampton Respiratory Biomedical Research Unit, Mailpoint 810, Sir Henry Wellcome Laboratories, Level F, Southampton, University Hospitals Southampton NHS Foundation Trust, Southampton, SO16 6YD, UK. E-mail: R.Djukanovic@soton.ac.uk

ABSTRACT Inflammatory lung diseases are highly complex in respect of pathogenesis and relationships between inflammation, clinical disease and response to treatment. Sophisticated large-scale analytical methods to quantify gene expression (transcriptomics), proteins (proteomics), lipids (lipidomics) and metabolites (metabolomics) in the lungs, blood and urine are now available to identify biomarkers that define disease in terms of combined clinical, physiological and patho-biological abnormalities. The aspiration is that these approaches will improve diagnosis, *i.e.* define pathological phenotypes, and facilitate the monitoring of disease and therapy, and also, unravel underlying molecular pathways. Biomarker studies can either select predefined biomarker(s) measured by specific methods or apply an "unbiased" approach involving detection platforms that are indiscriminate in focus. This article reviews the technologies presently available to study biomarkers of lung disease within the 'omics field. The contributions of the individual 'omics analytical platforms to the field of respiratory diseases are summarised, with the goal of providing background on their respective abilities to contribute to systems medicine-based studies of lung disease.



@ERSpublications

Summary of the application of 'omics-based analytical platforms for biomarker discovery in inflammatory lung diseases <http://ow.ly/mjGGc>

Received: May 16 2012 | Accepted after revision: Dec 14 2012 | First published online: Feb 08 2013

Support statement: This work was supported by the U-BIOPRED consortium (Unbiased Biomarkers for the PREDiction of respiratory disease outcomes, Grant Agreement IMI No.115010). S. Ballereau and C. Auffray were also supported by the FP7-MeDALL Consortium (Mechanisms of the Development of Allergy, Grant Agreement FP7 No.264357).

Conflict of interest: Disclosures can be found alongside the online version of this article at www.erj.ersjournals.com

Copyright ©ERS 2013

Introduction

Inflammatory lung diseases are highly complex in respect of pathogenesis and relationships between inflammation, clinical disease and response to treatment. While interstitial lung diseases have long been viewed as a spectrum of distinct pathological conditions with different clinical outcomes [1], asthma and chronic obstructive pulmonary disease (COPD) have only recently been recognised as syndromes consisting of several disease entities [2–6]. Sophisticated, high-throughput, large-scale analytical methods to quantify gene expression, proteins and lipids as well as other metabolites in the lungs, blood and urine are now available. These methods offer the potential to identify biomarkers that define airway obstructive diseases in terms of combined clinical, physiological and patho-biological abnormalities. The aspiration is that these approaches will improve diagnosis, *i.e.* define disease phenotypes and facilitate the monitoring of disease activity and therapy. In research terms, this information will also help unravel the complex molecular pathways underpinning disease.

In broad terms, biomarker studies can either select predefined biomarker(s) measured by specific methods or apply an “unbiased” approach involving use of indiscriminate detection platforms. This article reviews the technologies currently available to study biomarkers of lung disease within the so-called ‘omics field, a term that was first used to define the studies of genomes (genomics) and gene expression (transcriptomics) of cells, tissues, organs and organisms and has subsequently been adopted for studies of proteins (proteomics), lipids (lipidomics) and metabolites (metabolomics). More recently, measurement of volatile organic compounds (VOCs) in exhaled breath condensate has been termed “breathomics”. The use of the ‘omics term reflects an experimental paradigm based upon the acquisition of large-scale datasets from a single sample with the aim of identifying biomarkers of disease and/or elucidating novel functional or pathological mechanisms (fig. 1). An ‘omics experimental design often involves a hypothesis-generating component in which a broad encompassing dataset is acquired to provide insight into novel processes in disease, rather than focusing on reductionist “molecular medicine-based” targeted methodologies. ‘Omics approaches are resource-intensive, analytically demanding and require the use of sophisticated statistical and modelling approaches to analyse datasets consisting of hundreds to thousands of variables in order to minimise false positives (Type I error) and false negatives (Type II error). The collection of ‘omics-based datasets is often an integral component of systems biology studies, which seek to integrate data and thereby understand key fluctuations in the homeodynamics of the experimental system in question (*i.e.* disease, phenotype, therapeutic intervention). Regardless of the technique chosen, the diagnostic accuracy of potential identified biomarkers has to be examined and validated according to international recommendations based on STARD-guidelines [8, 9].

The principal biological matrices available for biomarker discovery in respiratory diseases are: whole lung tissue and cells isolated from lung parenchyma; bronchoalveolar lavage fluid (BALF); spontaneous or induced sputum; exhaled air; exhaled breath condensate (EBC); blood (cells, serum and plasma); and urine. Their advantages and disadvantages for application in the various ‘omics platforms are summarised in table 1, with a glossary of terms provided in table 2. For example, lung tissue contains combined transcriptomes, proteomes and metabolomes of multiple cell types, which complicates analysis and data interpretation. By comparison, EBC may have a proteomic profile that is too simple to empower meaningful biomarker studies [10]. Accordingly, it is important that the potential constraints of a given matrix are considered when evaluating method development for any of the ‘omics approaches discussed in this review.

A key consideration for ‘omics methods is the wide dynamic range of analytes in biological samples [12, 13]. This can result in ‘crowding out’ of less abundant analytes by more abundant ones, and as a consequence requires methodologies to enrich low abundance components, but at the price of adverse effects on assay reproducibility. This range in concentrations can also have repercussions for statistical modelling (*e.g.* use of univariate scaling *versus* no scaling of data, which both make very different assumptions regarding the biological significance of the data). Analysis is further complicated by the physico-chemical diversity of some analytes, *e.g.* proteins due to alternative splicing, RNA editing, subunit oligomerisation and post-translational modifications [14]. Lipids have a wide diversity of structural and physical properties, ranging from neutral molecules, such as triacylglycerols and sterols, through polar glycerophospholipids, to signalling molecules such as eicosanoids and other oxylipins, which comprise numerous isomers. In addition, developed methodologies rarely consider stereochemistry in compound analysis and identification, which can have profound effects on observed biological parameters. This structural and physico-chemical diversity, in combination with dynamic ranges of several orders of magnitude, results in an abundance of analytical challenges associated with each of the different ‘omics platforms that need to be considered in method development on a matrix-specific basis. In addition, the variety in the acquired data

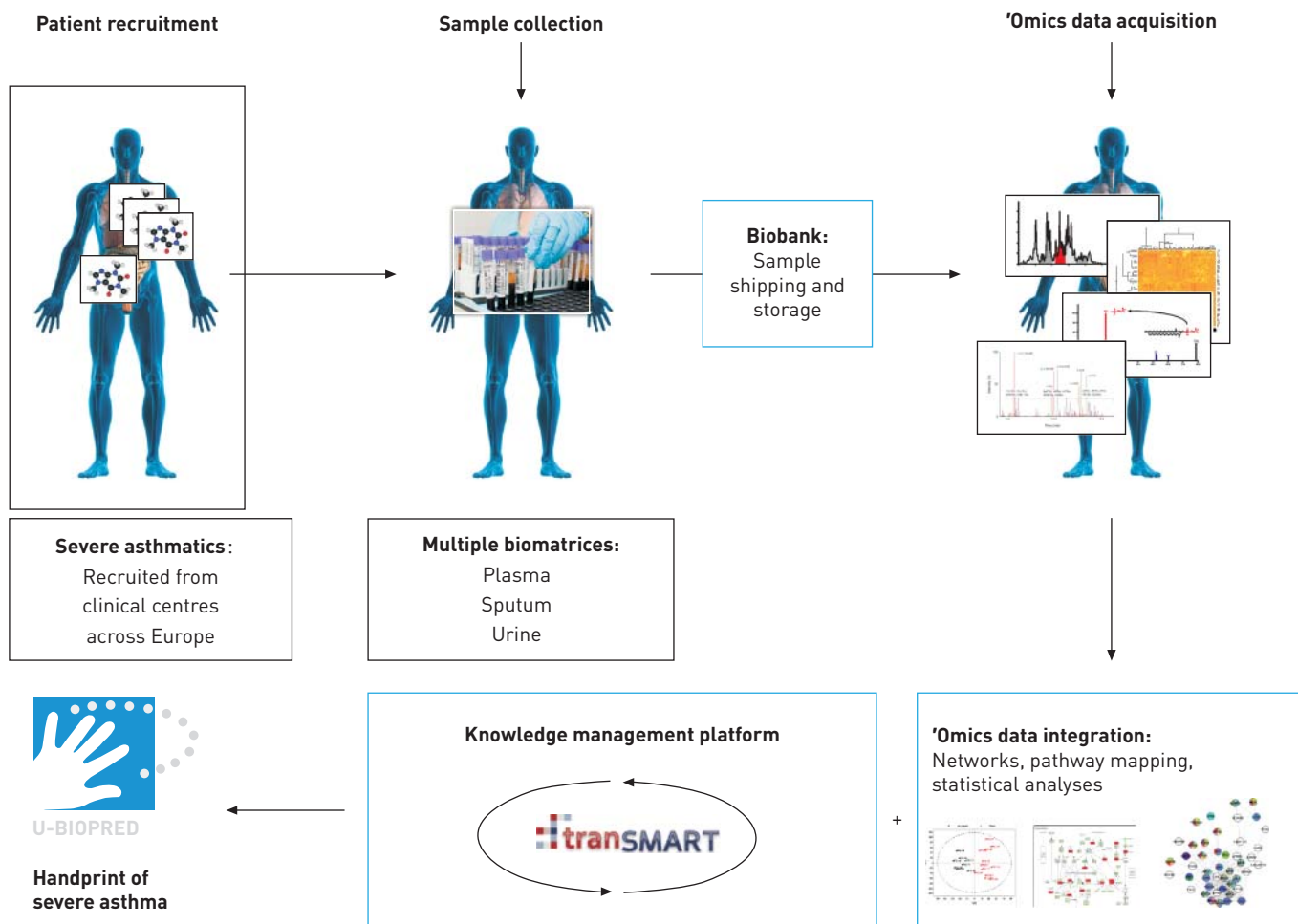


FIGURE 1 Flowchart of the 'omics-based workflow for unbiased clinical biomarker discovery employed in the U-BIOPRED project that employs transSMART as its knowledge management platform [7].

structures raises multiple statistical issues when integrating 'omics data from disparate analytical platforms and biological matrices [15, 16].

Mass spectrometry overview

The complexity, and hence power, of unbiased biomarker studies is growing rapidly due to an increasing portfolio of available mass spectrometry methodologies. Mass spectrometers determine the molecular mass of molecules using mass analysers (fig. 2), which broadly exist in five distinct formats of increasing mass resolution and accuracy: quadrupole, ion trap, time of flight (ToF), Orbitrap and Fourier Transform Ion Cyclotron Resonance (FT-ICR). The major advance that enabled mass spectrometry to address a wide range of biological questions was the development of electrospray ionisation (ESI) as a means of ionising analytes; a critical step since mass spectrometers detect charged ions [17, 18]. ESI is a soft ionisation technique that charges analytes by nebulising a liquid flow from a capillary held at a high potential. Ions generated by ESI are stable as opposed to other soft ionisation methods such as matrix-assisted laser desorption ionisation (MALDI), which produces ions in the excited state that decay rapidly. ESI also results in less fragmentation and hence facilitates detection of the molecular ion. This advance has made the detection of proteins, peptides, less volatile lipids and metabolites relatively straightforward, and ESI is currently the most commonly used technique for introducing analytes into a mass spectrometer.

Mass spectrometry methods vary in respect of throughput (*i.e.* time of analysis), sensitivity and selectivity, as well as robustness, ease of use and cost. For example, surface enhanced laser desorption ionisation mass spectrometry (SELDI-MS) separates and captures protein subsets on a surface based on specific biophysical properties, such as hydrophobicity, net anionic or cationic charge, prior to analysis by MALDI-ToF mass spectrometry. However, lack of resolution and mass accuracy hinders unambiguous identification of protein peaks [19, 20]. Accordingly, identification of candidate biomarkers of pulmonary disease has had

TABLE 1 Advantages and disadvantages of different clinical matrices for biomarker discovery studies in respiratory disease

| Type of sample | Advantages | Disadvantages |
|-------------------------------------|---|---|
| Bronchoalveolar lavage fluid | Collected from specific area of the lung Collected under clinical and reproducible conditions Pre-existing proof of principle in this sample type Reasonably complex analytical matrix | High salt content Dilute analytes; often requires concentration before analysis, leading to variability Difficult to normalise for sample dilution Possible bias towards leakage of tissue or inflammatory cell components in diseased patients Collection procedure may cause tissue damage Not well tolerated as a procedure, particularly in diseased patients Cannot perform longitudinal sampling |
| Induced sputum | “Natural” biofluid Internationally recognised SOP for collection Choice of sample type (i.e. with/without mucolytic agent) Reasonably complex analytical matrix | High salt content Difficult to normalise for sample dilution (amylase sometimes used) Source in the lungs poorly defined Use of reducing agents can affect analyte characteristics Presence of cellular and/or bacterial debris Potential high content of plasma infiltration components, especially in asthma Salivary contamination during collection High quantities of mucins affect proteomics readouts |
| Blood (serum or plasma) | Minimally invasive Easy to collect with standardised protocols Widely used in proteomics and lipidomics studies, composition relatively well documented Relatively consistent and easy to define protocol in multicentre studies Contains a large number of potential targets | Large dynamic range of protein content (~10 orders of magnitude) Lipid composition dominated by lipoproteins, possibly masking minor components Distant from the tissue of interest, so potential bias towards systemic changes in disease |
| Exhaled breath condensate | Noninvasive Suitable for analysis of non-volatile components Suitable for longitudinal study Safe Feasible in children Metabolomic analysis by NMR spectroscopy validated | Very dilute, often requires concentration steps leading to variability Potential variability due to differences in droplet dilution Samples whole airway, difficult to localise changes Difficult to normalise protein/chemical composition for total content High variability in sample quality Relatively few proteins present |
| Exhaled air volatiles | Noninvasive (Pseudo) real-time Inexpensive and portable equipment available Diagnostic accuracies confirmed between independent centres | Data not exchangeable between devices Samples airways, lungs and systemic volatiles (not specific) Requires standardisation of sampling and analysis |
| Urine | Noninvasive Easy to collect with standardised protocols Conducive to longitudinal sampling regimens Ideal for eicosanoids | Distant from the tissue of interest, so biases towards systemic changes in disease Excretion of metabolites may be changed by kidney or liver diseases |
| Lung tissues | Collected from specific area of lung Reflects local changes Essential for transcriptomics analysis Can be combined with immunohistochemistry Results can be compared to genomic and cytology data Complex proteome | Difficult to obtain, either <i>post mortem</i> or using bronchoscopy Contains multiple cellular types, difficult to identify origin of biomarkers Homogenisation of tissue can be difficult Limited quantity of tissue |

SOP: standard operating procedure; NMR: nuclear magnetic resonance.

only limited success [21–23]. The majority of ‘omics efforts have therefore moved towards high-resolution instruments that provide increased mass accuracy to facilitate molecular species identification. For example, FT-ICR mass spectrometers currently provide the highest mass accuracy and resolution, often sufficient for calculation of elemental formulae, but they are relatively low-throughput and costly. The Orbitrap technology-based systems interfaced with a linear ion trap (LTQ Orbitrap) have high mass accuracy and

TABLE 2 Glossary of terms related to 'omics-based studies of respiratory diseases

| Term | Definition |
|---------------------------------------|--|
| Transcriptomics | The study of the set of all RNA molecules, including messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA) and other non-coding RNA (ncRNA) such as microRNAs produced in a cell type or in a population of cells/tissues. |
| Non-coding RNAs (ncRNA) | A functional RNA molecule that is not translated into a protein. |
| Exon | A nucleic acid sequence that is represented in the mature form of an RNA molecule after portions of precursor RNA (introns) have either been removed by splicing or two or more sequences joined by splicing. |
| RNA sequencing (RNA-seq) | A recently developed approach to transcriptome profiling that uses deep-sequencing technologies (high-throughput next generation sequencing) to sequence cDNA with deep coverage and base-level resolution. This provides a far more precise measurement of levels of transcripts and, in addition, information on differently spliced transcripts and post-transcriptional editing. |
| Microarray | A DNA microarray is a collection of microscopic DNA spots attached to a solid surface. Each DNA spot contains a specific DNA sequence or probes that can hybridise a cDNA or cRNA target under high-stringency conditions. DNA microarrays are used to measure the expression levels of large numbers of genes simultaneously. |
| Proteomics | The study of the entire protein complement of an organism or system. |
| Electrophoresis | The movement of particles relative to a solution under the influence of a uniform electric field. |
| Lipid | Broadly defined as a hydrophobic or amphipathic small molecule that may originate entirely or in part by carbanion-based condensations of thioesters and/or by carbocation-based condensations of isoprene units. |
| Lipidomics | The large-scale study or acquisition of the lipid composition of a biological system. |
| Structural lipid | The amphipathic components of the cell membrane including: glycerophospholipids, sphingolipids and sterols (also known as molecular or membrane lipids). |
| Eicosanoid | Biologically active products of arachidonic acid including prostaglandins, thromboxanes, leukotrienes and other oxygenated derivatives (also known as signalling lipids or lipid mediators). |
| Oxylipin | Oxygenated compounds that are formed from fatty acids by pathways involving at least one step of dioxygen dependent oxidation (include the eicosanoids as well as analogous compounds derived from alternative fatty acid substrates). |
| Phospholipid | Lipids built on a glycerol backbone that contain a phosphate group, a head group and two fatty acyl/alkyl chains; commonly structural components of cell membranes as they can form bilayers. |
| Sphingosine-1-phosphate | A signalling phospholipid built on a sphingosine backbone, rather than glycerol. Sphingosine can be released from ceramide and subsequently phosphorylated to become bioactive, for instance, as a ligand for cell surface receptors. |
| Lysophosphatidic acid | A potent lipid mediator known to be involved in cell proliferation. Can be formed via a number of different biosynthetic pathways by the sequential action of phospholipase enzymes. |
| Lung surfactant | A complex mixture of lipids and proteins whose primary function is to reduce alveolar surface tension and allow lung compliance. |
| Metabolomics | The analysis of the whole of the metabolome under a given set of conditions. |
| Metabolome | Complete complement of all of the low molecular weight molecules (<1500 amu) present in the biological compartment in a particular physiological state under a given set of environmental conditions. |
| Metabolite profiling | Targeted quantification of a predefined subset of metabolite components of the metabolome that usually are of related chemical structure and/or biological activity. |
| Breathomics | The large-scale study of exhaled metabolites [11]. |
| Breathcloud | Online internet database of eNose breathprints matching various diseases, which can be used for direct calculation of diagnostic probability all around the world. |
| Breathprint | Composite signature of an exhaled molecular mixture, as obtained by 'omics analysis. |
| eNose | Electronic nose; composed of a nano-sensor array capturing mixtures of volatile organic compounds. |
| Exhaled breath condensate | The fluid captured by cooling exhaled air, containing non-volatile metabolites. |
| VOCs | Volatile organic compounds; in exhaled air these represent metabolites from respiratory or systemic origin. |
| Multivariate statistics | A form of statistics encompassing the simultaneous observation and analysis of more than one statistical variable. These methods reduce the dimensionality of datasets, thereby improving the interpretation and visualisation of data. A common application is to separate the signal from the noise in data with many variables. |
| Pattern recognition algorithms | Mathematical recognition of consistency between complex datasets, e.g. between composite molecular mixtures from validation and training samples. This can be used for probabilistic diagnostic assessment (positive and negative predictive value for disease). |
| Systems biology | Research approach to understand and model the system behaviour as a whole, linking the highly complex interactions between its components to its emergent properties, based on perturbation of the system and measures of its response with large-scale experimental data (primarily obtained via 'omics techniques). |
| Biomarker | Trait or substance such as an exogenous chemical molecule, endogenous antibody, protein, transcript or even cells objectively detectable or measurable in the body, blood or tissue, that indicates a biological state, and in biomedical research that correlates with the risk or stage of a disease, with response to treatment or intervention, or with exposure to environmental factors. |
| Biomarker profile | Also known as molecular signature or fingerprint; expression profile comprised of clusters of differentially expressed genes, proteins and/or metabolites indicative of a biological process. |
| Molecular network | Graph representation of interactions (edges) between molecules (nodes) within a cell for molecular systems such as transcriptional regulation, protein-protein interactions, or metabolites and reactions, whose topology analysis for example enables identification of important elements. |
| Machine learning | Diverse set of computational and statistical tools used where human expertise does not exist or cannot be formulated to design algorithms for the identification of patterns in empirical datasets that can be generalised to improve prediction of features in previously unseen datasets. |

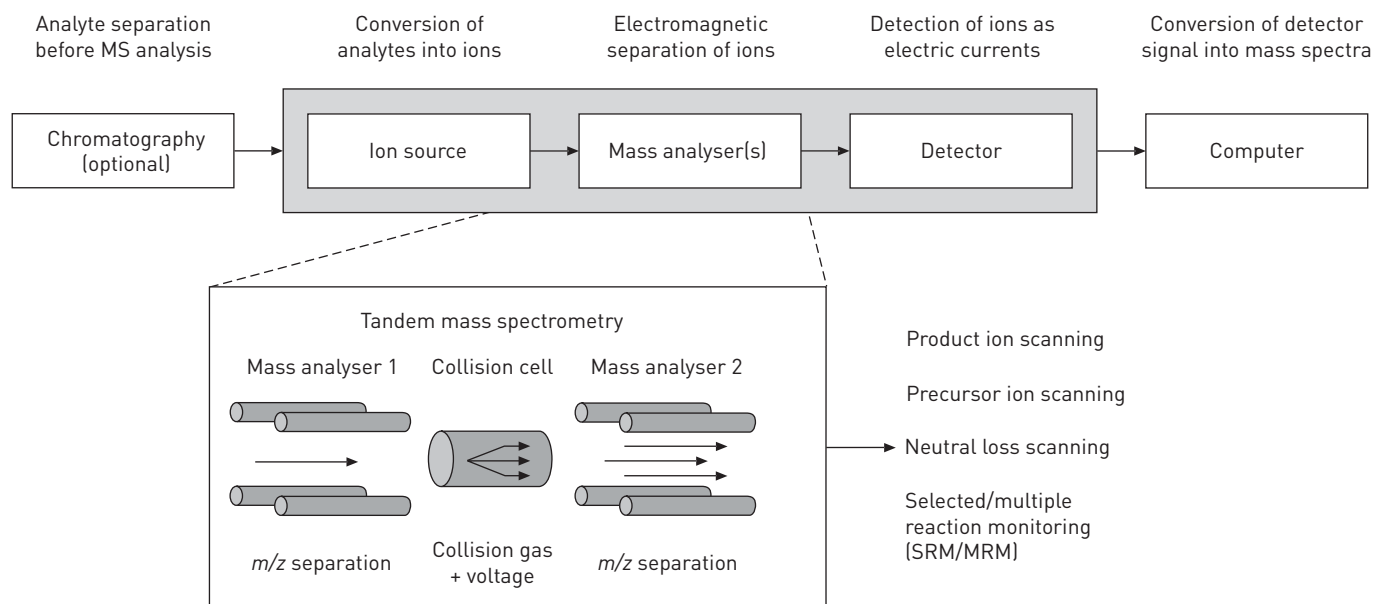


FIGURE 2 Mass spectrometry (MS) is an analytical technique that can be used to determine the mass, elemental composition or chemical structure of molecules. In a typical configuration the target analytes are first converted into charged particles (ionisation) before introduction into the mass spectrometer. Once inside, the ions are separated by a mass analyser according to their m/z ratio using electromagnetic fields under vacuum. The separated ions are then recorded by a detector, and the detector signal is converted into a mass spectrum that can be stored and manipulated on a computer. Mass spectrometry is often preceded by some form of chromatography (gas, liquid or thin layer chromatography) to separate analytes of interest before analysis. It is also common for mass spectrometers to have multiple mass analysers for targeted manipulation of the ions within the instrument. A widely used example is triple quadrupole mass spectrometry, in which the first analyser is followed by a collision cell for fragmentation of the ions, and a second analyser. This configuration allows for a number of scan experiments that can be used to elucidate the structure of ions of interest, or increase the sensitivity of the instrument to a specific (set of) ion(s).

resolving power and are extensively employed in 'omics-based applications [24–27]. While they may have a lower mass resolution than the FT-ICR systems [28], Orbitrap technology-based instruments have greater throughput, are more robust and are significantly less expensive. The re-emergence of ion-mobility separation, which separates charged molecules based upon their shape and conformation, offers an orthogonal dimension of separation. Recently, ion-mobility separation has been combined with quadrupole-ToF analysis [29], as well as MALDI interfaces to increase resolution for mass spectrometry imaging [30]. Triple quadrupole systems (MS/MS) employing multiple/selected reaction monitoring (MRM/SRM) are the workhorses of bioanalytical chemistry and are extensively used in quantifying proteins/peptides, lipids and metabolites, partly due to their robustness and wide dynamic range [31]. MRM provides maximum sensitivity for analytes separated by high-performance liquid chromatography (HPLC), albeit at the expense of limited spectral data and loss of mass resolution. Mass spectrometer versatility and, therefore, utility as vehicles for biomarker discovery derives from combining mass analysers into hybrid analytical platforms, in order to achieve a wide range of specificities and sensitivities.

Methods that directly infuse highly soluble samples by ESI into a mass spectrometer increase reproducibility and greatly facilitate high-throughput analysis, albeit at the expense of simplification of complex mixtures of analytes. Direct infusion has been widely used for shotgun lipidomics, which employs a combination of diagnostic precursor and neutral loss scans to characterise the molecular species compositions of individual classes of lipids [32]. Because all analytes are measured under identical ionisation conditions, they are readily quantifiable using appropriate internal recovery standards. The advent of ESI has provided a simple approach to introducing the eluate from a liquid chromatography (LC) column directly into any of the mass spectrometry analysers, in turn providing an additional analytical dimension by harnessing the extensive range of HPLC column technologies to separate analytes prior to introduction into the mass spectrometer. The choice of chromatography employed for the separation of a sample is equally important for data generation. For clinical biomarker discovery, the combination of reverse phase HPLC with ESI is the most common configuration. Recent advances include the shift to higher-pressure LC systems such as ultra-performance liquid chromatography (UPLC) and ultra-high performance liquid chromatography, which offer increased resolution, speed and sensitivity for 'omics-based approaches. In addition, new approaches using alternative mobile phases such as CO_2 have recently been developed (e.g. UltraPerformance Convergence Chromatography; UPC²). There are multiple stationary phases available that are compatible with mass spectrometry, ranging from hydrophobic reverse phase (C18) columns to

traditional normal phase hydrophilic interaction liquid chromatography (HILIC) and ion exchange systems. Capillary electrophoresis has also been successfully coupled to mass spectrometry for 'omics-based applications [33]. In addition, gas chromatography-based systems (GC) can be used for the separation of volatile, thermally stable compounds of relatively low polarity. These systems have proven useful for quantification of many of the small molecules involved in basic metabolism, *e.g.* in breathomics approaches [34, 35].

The combination of LC-MS/MS analyses can provide detailed compositional and structural analyses tailored to the LC elution profile, with enhanced sensitivity for low abundance components. Single- and multi-dimensional LC-MS/MS have been used for so-called "shotgun" proteomics. One approach that has been successfully applied in many proteomic laboratories is multidimensional protein identification technology (MudPIT) [36]. Although attention in recent years has focused on using orthogonal separations to reduce the complexity of a biological sample, the high pressure capabilities of UPLC and nano-UPLC have enabled the implementation of long columns (*i.e.* 50 cm) for efficient separation of biological samples [36] and have also been used quantitatively [35]. The eventual choice of separation technique and column depends upon the target compounds for biomarker discovery because no current methodology is truly global or comprehensive in its ability to capture a full proteome, lipidome or metabolome.

Transcriptomics

The ability to determine the differential expression of RNA transcripts (transcriptomics) over time and/or between cells and disease has transformed our understanding of cellular function [37]. Transcriptomics analysis aims to describe and quantify RNA species such as mRNAs, non-coding RNAs and small RNAs, and their variations in response to external stimuli or disease. Expression profiling by microarrays has been very successful and widely used, with, for example, >40 000 citations currently in PubMed [38]. Using microarrays it is possible to detect variations in expression of many, but not all, transcribed genes under both normal and perturbed conditions. Direct sequencing offers the potential for the detection of more transcripts and their variants, but relies on a less mature technology.

The improvements in microarray analysis and interpretation have been due to concerted efforts by many groups across the world to introduce quality control standards and guidelines for complete microarray workflows [39–42]. These advances took 10 years to establish for microarrays, but should be addressed more rapidly for newly emerging technologies [39, 40]. RNA sequencing (RNA-seq) offers several advantages over microarrays and has generated important results across diverse species [43–45]. It is considered completely unbiased, because it does not rely on a set of predefined probes selected for the array chip and covers the whole transcriptome, enabling the discovery of novel exons, isoforms and even previously undetected transcripts [45]. In addition, RNA-seq methods have low background noise, a large dynamic range, are highly accurate and reproducible, and produce data comparable to that of microarrays [45, 46]. However, some of the specific protocols used may introduce bias due to amplification, fragmentation and ligation processes having some sequence preferences [37, 40, 47]. A limitation of both these methods is the need to validate expression values using RT-qPCR. Emerging technologies that use miniaturised high-throughput RT-qPCR approaches or multiplex direct visualisation and counting of RNA molecules have been developed, but these approaches must be standardised and applied across platforms [37, 40]. The current advantages of microarrays include their relatively low running cost compared with sequencing as well as the maturity of the analysis strategies and experimental designs for dealing with the known biases inherent in microarray data [38].

The application of transcriptomics to lung diseases is transforming our views on the molecular classification of chronic lung diseases, as well as opening novel avenues for biomarker discovery using disease tissue or surrogate cells and monitoring drug responses [48, 49]. For example, the use of microarrays has confirmed the presence of distinct subsets of mild/moderate asthmatic patients on the basis of their expression of Th2 cytokines and has shown that gene expression profiles in airway epithelial cells can predict drug responsiveness [50]. The patient population expressing a high level of Th2 cytokines, the so-called Th2-high phenotype, expresses distinct features of airways inflammation over a variable continuum, correlating significantly with local and systemic measures of allergy and eosinophilia [51], responds better to corticosteroids than the Th2-low phenotype [50] and is also linked to markers of airway remodelling [50]. High periostin levels seen in the Th2-high phenotype have also been shown to distinguish patients with severe asthma who respond to anti-interleukin-13 antibody therapy [52].

Microarrays have also been used to distinguish mRNA expression profiles in peripheral blood CD4⁺ T-cells in children with frequent and infrequent wheeze due to viral exposure [53] and to demonstrate distinct microRNA profiles in CD8⁺ T-cells isolated from patients with severe and non-severe asthma [54]. In animal models of asthma, microarrays have furthermore demonstrated profound effects of combined

allergen and viral challenge on murine lung gene expression, thus emphasising a key role for Toll-like receptors, novel serine protease inhibitors as well as chemokines and cytokines that recruit inflammatory cells into the airway [55]. In a similar manner, TILLEY *et al.* [56] have analysed the effect of cigarette smoke on transcriptome patterns in small airway epithelial cells [56] and alveolar macrophages [57], and demonstrated differences in expression profiles in some “healthy” smokers that may be relevant to the pathogenesis of COPD [57, 58]. The key role of oxidative stress effects on airway epithelial cells in driving COPD was also demonstrated using gene microarrays [59].

A direct comparison between microarrays and RNA-seq has been performed on bronchial epithelial cells from never-smokers and smokers with and without lung cancer. The results showed a significant correlation between the two techniques although RNA-seq detected more smoking- and cancer-related transcripts than the microarrays [60]. The same research group was able to demonstrate a correlation between transcriptomic readouts analysed by microarrays and proteomics [61], which highlighted the presence of altered protein expression in the absence of differential transcription. The pharmaceutical industry has used microarrays to analyse disease and drug effects on numerous cells and tissues using standard procedures and platforms (*e.g.* Affymetrix microarrays). Their databases are capable of integrating these data with clinical data from patients, which represents a major driver for the use of microarrays in drug discovery [37, 40]. However, there is a need for further analysis of samples from the site of disease to validate the relevance of the detected differences [37, 40].

Summary

The current state-of-the art with regard to drug discovery remains expression profiling by microarrays due to: 1) the maturity of the available tools/platforms; 2) existing data available for comparison purposes; and 3) relatively lower cost. However, ongoing research toward standardising deep-sequencing approaches and comparison of results from sequencing and microarray analysis of the same samples will lead to analysis programmes that enable direct comparison between the two technologies [38]. Detecting genes with low expression will remain a problem for both approaches, but there are some applications, such as transcript discovery and isoform identification, where RNA-seq is the preferred method [38]. As the cost of sequencing decreases and the availability of new generation sequencing platforms increases, a switch in approaches will occur progressively over the next few years. However, given the substantial agreement between the two methods, it is unlikely that the microarray data currently being generated or in existing databases will become obsolete. Rather, this information will likely be complemented and extended in depth and coverage by the sequence-based data, providing deeper insight into respiratory physiology and disease. In addition, expression levels of panels of mRNAs will provide biomarkers for disease (sub)types and efficacy of novel drugs.

Proteomics

Quantification of proteins has been the basis of numerous studies of lung diseases, but it is only more recently that unbiased proteomics approaches, combining mass spectrometry with either gel- or non-gel-based methods for protein/peptide separation, have been used [62, 63]. The breadth of proteins identified has largely depended on the method applied, since each detects proteins with varying selectivity and sensitivity. Some methodologies have been primarily descriptive, whilst others have been more quantitative. For example, the qualitative technique GeLC-MS/MS was used to define the proteome of human sputum [64] and has more recently been applied to the analysis of BALF in mouse and non-primate asthma models, where disease specific biomarkers relating to response to corticosteroid treatment were identified [65].

One key aim of any proteomics study is to quantify potential biomarkers associated with disease. SELDI-ToF-MS has been extensively used in several respiratory studies. For example, serum amyloid protein (SAA) was identified by this method as a novel blood biomarker of acute exacerbations of COPD, which was confirmed by ELISA [66]. Because SAA levels correlated with infection, the study suggested a role in infection or exacerbations rather than just COPD. Other SELDI-MS analyses of BALF have identified CCSP10, neutrophil defensins 1 and 2, and calgranulins 1 and 2 (S100A8 and S100A9) as being altered in smokers with COPD when compared to asymptomatic smokers [67]. However, despite improvements in the technology and relative affordability of the equipment, the SELDI-ToF-MS platform does not yet provide sufficient resolution or reproducibility for use in clinical diagnostics. This is primarily due to the lack of mass accuracy and accompanying drift over the data collection period, making it difficult to identify protein peaks unambiguously [19, 20].

Other quantitative studies have applied different means of protein separation prior to mass spectrometry analysis. NICHOLAS *et al.* [68] used two-dimensional gel electrophoresis to separate proteins in induced sputum samples from patients with COPD and healthy smokers prior to MS/MS analysis. They identified 44 differentially expressed protein spots in COPD, two of which were further validated by Western blot analysis

and ELISA: lipocalin-1 and apolipoprotein A1. Interestingly, the majority of differentially expressed proteins were reduced in COPD and many of them could be functionally associated with innate immunity. Improvements to the two-dimensional gel electrophoresis in the form of differential protein labelling with multiple dyes has recently emphasised the utility of the two-dimensional gel technique for finding biomarkers of lung diseases in challenging biofluids such as sputum [69]. KOHLER *et al.* [70] recently utilised multiplexed differential gel electrophoresis to identify a female-dominated subphenotype of COPD. In this study, a subset of 19 proteins in alveolar macrophages primarily originating from the lysosomal activity and oxidative phosphorylation pathways were found to provide classification of female COPD patients with 78% predictive power. The enduring appeal of two-dimensional gel methods lies in the ease of quantitation and identification of biomarkers, as demonstrated by the identification of the poly immunoglobulin receptor as a biomarker of COPD.

Although two-dimensional gel electrophoresis offers excellent resolving power, the approach has several limitations [71], including a limited ability to resolve proteins with extremes of molecular weight and isoelectric point and the challenge of accurately aligning individual proteins spots on different gels for comparison between patient populations. Consequently, gel-free approaches, such as the stable isotope labelling approach (iTRAQ, isobaric tags for relative and absolute quantitation), have been employed, an example being a study to identify biomarkers of smoking in human plasma samples [72]. Here, the authors increased proteome coverage through depletion of the top 14 most abundant proteins, identifying 113 low abundance proteins, 16 of which were differentially expressed between smokers and nonsmokers. The iTRAQ approach, coupled with nano-LC-LTQ-Orbitrap, has also been used to compare the proteomes of bronchial biopsy samples from healthy and asthmatic subjects and to identify differences in response to glucocorticoid treatment [73]. An extensive number of proteins were identified despite limited material for analysis and lack of abundant protein depletion. Seven proteins were differentially expressed in asthma when compared to controls and seven were modified in response to budesonide treatment.

Advances in the fields of mass spectrometry instrumentation (increased sensitivity, resolution and mass accuracy), LC separation and bioinformatics have led to label-free approaches becoming the method of choice for proteomics analyses (fig. 3). This is because the approach offers a rapid, simple and low cost measurement of protein expression levels in complex biological samples. GHARIB *et al.* [75] used a label-free approach with quantitation based upon spectral counting to assign 17 differentially expressed proteins in induced sputum. This subset was enriched for proteins associated with processes involved in protease inhibitory activity, defence response, immunity and inflammation; the method robustly classified asthmatic and control subjects. A further study using FT-ICR mass spectrometry to identify isoforms of SP-A in BALF from patients with cystic fibrosis, chronic bronchitis and pulmonary alveolar proteinosis observed qualitative differences in SP-A isoforms in patients with pulmonary alveolar proteinosis when compared to other diseases examined [76]. However, it is now accepted that a single protein biomarker of complex disease is unlikely to be sufficient for disease classification and diagnosis, and that a successful strategy would consist of developing a panel of biomarkers [77]. Impressively, researchers studying idiopathic

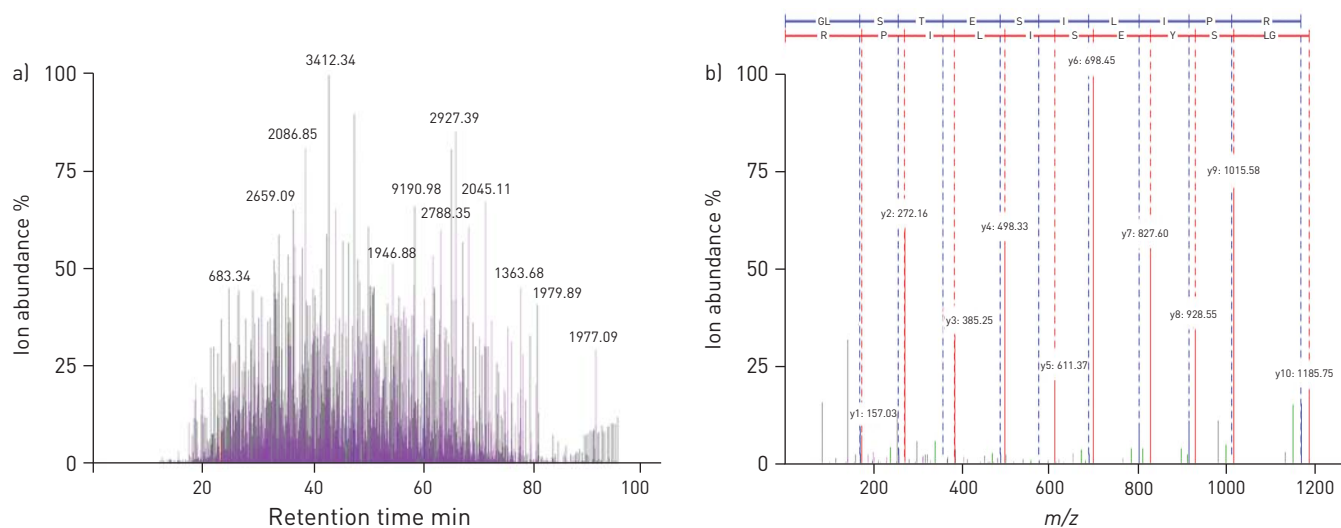


FIGURE 3 Analysis of human induced sputum analysed using the label-free approach, LC-MS^E, SILVA *et al.* [74]. a) The obtained LC-ion chromatogram of peptides measured in sputum and b) an MS/MS spectrum of a peptide ($[M+H]^{2+}$ $m/z=1185.68$) identified from lipocalin-1.

pneumonia syndrome identified a set of 81 disease-associated protein biomarkers using a label-free approach and were able to stratify patients likely to respond to cytokine neutralisation therapy [78].

Summary

The current state-of-the-art in proteomics analysis of respiratory samples uses quantitative label-free LC-MS approaches that offer: 1) improved (faster and more sensitive) detection of proteins in a range of biological sample types; 2) relative and absolute quantitation ($\text{ng}\cdot\text{mL}^{-1}$); and 3) unlike multiplexing approaches such as iTRAQ, provide independent data collection of samples, in theory allowing the comparison of a theoretically unlimited number of clinical samples. Technical advances in mass spectrometry instrumentation and informatics offer an exciting prospect for the future of diagnostic and prognostic marker discovery in respiratory disease. Such advances include the application of unbiased data-independent LC-MS acquisition strategies [74] that allow the information obtained from a biological sample to be maximised by acquiring accurate mass-to-charge (m/z) ratio values for all precursor ions and their corresponding fragment ions within a single analytical run. These almost complete datasets can subsequently be interrogated post-analysis, not only for peptides and proteins, but also for other target molecules such as lipids and metabolites, facilitating “one-stop” multidimensional biomarker discovery. However, despite the recent advances in mass spectrometry instrumentation, the speed of analysis is currently limiting and represents a key challenge for the future of clinical proteomics. Advances in the miniaturisation of front-end separations through nanospray microfluidics [79] may offer a solution to address this challenge, providing further increased sensitivity and sample throughput, an essential prerequisite for the advent of individualised protein biomarker discovery.

Lipidomics

Lipids make up ~90% of lung surfactant, which is vital for maintaining small airways and alveolar potency and they play a significant role in lung disease. A discussion of lipids in the lung can be divided into 1) high-abundance structural or molecular lipids (e.g. phospholipids) and 2) low-abundance signalling lipids (e.g. eicosanoids). Lipids may have a primary role, in which alterations in lipid composition, biosynthesis or downstream metabolism impact directly on exacerbation and disease severity. Alternatively, lipid composition may be altered as a result of the disease process and thereby provide biomarkers to stratify the condition, monitor the effect of a drug, or predict the likelihood of exacerbation. Furthermore, the biological importance of lipid metabolic products formed from cell membrane-associated arachidonic acid has been demonstrated in numerous studies [80], although the biological role of structurally analogous compounds remains unclear.

Phosphatidylcholine

The glycerophospholipid phosphatidylcholine (PC) is the main lipid class in lung surfactant, comprising up to half of the total lipid content. This surface active lipid is reduced in sputum, but not BALF [81] of asthmatic patients, suggesting that compositional alterations to BALF PC are related to plasma infiltration in the airways rather than altered surfactant metabolism in the alveolus [82]. Accordingly, treatment of asthmatics with exogenous surfactant that has a “normal” PC content may be beneficial [83, 84]. Lysophosphatidylcholine (LPC), generated by PLA_2 activity on PC, contributes to the pathogenesis of lung disorders, including acute respiratory distress syndrome [85]. Antigen challenge in an animal model of asthma increases the concentration of an alveolar type II cell-specific PLA_2 isoform and decreases surfactant phospholipid, with treatment with specific PLA_2 inhibitors showing therapeutic benefit [86]. Interestingly, surfactant protein A inhibited this type II cell-specific PLA_2 enzyme.

Phosphatidylglycerol

Phosphatidylglycerol is a minor component of typical mammalian cell membranes, but it is the second most abundant surfactant phospholipid where it promotes adsorption of PC to the air:liquid interface. Secretory PLA_2 preferentially binds and hydrolyses acidic phospholipids like phosphatidylglycerol, and local allergen challenge in asthmatic subjects led to decreased amounts of phosphatidylglycerol and an increased ratio of PC to phosphatidylglycerol, which correlated with poor surface tension function [87]. An additional role for phosphatidylglycerol binding to Toll-like receptor 4 has been proposed in virally-induced asthma exacerbations; treatment of bronchial epithelial cells with phosphatidylglycerol reduced their inflammatory response to respiratory syncytial virus (RSV), and lung installation of phosphatidylglycerol in mice significantly reduced their susceptibility to infection with RSV [88]. Recently, decreased phosphatidylglycerol in exhaled particles has been described for asthmatic subjects compared with control volunteers [89].

Sphingosine-1-phosphate and lysophosphatidic acid

Sphingosine-1-phosphate (S1P), which is synthesised by sphingosine kinase-mediated phosphorylation of sphingosine [90], is elevated in BALF of asthmatic patients following antigen challenge [91] and, among other effects, stimulates contraction of airway smooth muscle [92]. The “Orm” family of proteins are regulators of sphingolipid synthesis [93] and single nucleotide polymorphisms within the ORMDL3 locus have been associated with severe childhood asthma [94], with possible effects on the production of S1P and the development of asthma. Lysophosphatidic acid (LPA) is generated by the action of autotaxin on LPC and can enhance airway smooth muscle contractility [95, 96]. BALF LPA levels have been shown to increase in a sensitised asthma mouse model, in which a direct link was observed between the expression of the LPA2 receptor and lung inflammation [97].

Eicosanoids

Eicosanoids comprise a large group of biologically active signalling molecules produced by enzymatic and auto-oxidative processes from arachidonic acid and other membrane-bound polyunsaturated fatty acids. The term oxylipin was introduced as an encompassing label for oxygenated compounds that are formed from fatty acids by reaction(s) involving at least one step of mono- or dioxygenase-catalysed oxygenation. Accordingly, this term includes the well-known eicosanoids synthesised from arachidonic acid, as well as related compounds formed by oxygenation of polyunsaturated fatty acids of longer and shorter chain length. There are thousands of potential analogues synthesised from different fatty acid precursors (e.g. DHA and EPA), most of which have as of yet undetermined biological roles. The use of LC with low particle size and MRM mass spectrometry has recently made it possible to quantify hundreds of these molecules simultaneously [98, 99], providing new insights into their role in respiratory disease [80]. A particular advantage of analysing eicosanoids is that measurement of indicative urinary metabolites is usually a sensitive approach to monitoring pulmonary biosynthesis. In particular, urinary eicosanoid profiles can reflect asthma exacerbations or induction of bronchoconstriction by, for example, allergen challenge. This is possible because the resting levels of eicosanoids and their downstream metabolites are very low, whereas there is a massive increase in their release into the circulation following induction of *de novo* biosynthesis, as reviewed by KUPCZYK *et al.* [100].

Prostaglandins are produced following oxidation of arachidonic acid by cyclooxygenases (COX-1 or COX-2) and specific prostaglandin synthases into the five primary COX products: PGE₂, PGD₂, PGF_{2α}, PGI₂ and thromboxane A₂ (TXA₂). Arguably the best-studied prostaglandin is PGE₂, which has a prominent, but complex, role in lung pathology [101]. PGE₂ has a bronchodilator effect and inhibits responses to allergens and other triggers of bronchoconstriction, presumably by an anti-inflammatory effect on mast cells [102]. In contrast, PGD₂, together with its early appearing metabolite 9α,11β-PGF₂, causes bronchoconstriction in subjects with asthma [103–105]. The 9α,11β-metabolite and tetranor-metabolites can be measured in blood and urine, and serve as an index of endogenous PGD₂, which is biosynthesised by mast cells. In asthmatics, urinary concentrations of 9α,11β-PGF₂ increase in response to allergen exposure and other trigger factors of airway obstruction. Asthmatics have higher urinary levels of the tetranor metabolites of PGD₂ than non-asthmatic control subjects, whereas levels of PGE₂ are comparable [106]. TXA₂ is a potent bronchoconstrictor that has been considered as a target for asthma therapy. The levels of the enzymatically formed product 11-dehydro-TXB₂ are, however, more reliable as indicators of endogenous TXA₂ biosynthesis, and 11-dehydro-TXB₂ is increased in the urine of atopic asthmatics following allergen-induced bronchoconstriction [107, 108].

Leukotrienes (LT) are formed by 5-lipoxygenase (5-LOX)-catalysed conversion of arachidonic acid to LTA₄ [109], which is subsequently converted to either LTB₄ *via* LTA₄ hydrolase or to cysteinyl leukotrienes (CysLTs) *via* LTC₄ synthase. Analogous pathways exist *via* 15-lipoxygenase activity, leading to the synthesis of lipoxins and eoxins as well as the associated hydroxyeicosatetraenoic acids (HETEs). CysLTs are potent contractile agonists of human airway and vascular smooth muscle [110–112]. LTE₄ is to a large extent excreted in urine without additional metabolism and increased urinary LTE₄ levels are used as a biomarker of disease severity (e.g. asthma exacerbations) [100]. There is extensive evidence that monitoring of urinary LTE₄ provides valuable information about mechanisms of inflammation in asthma and other airway diseases [80]. Lipoxins (LX) are short-lived eicosanoids that can support the resolution of inflammation [113]. Studies measuring lipoxins have suggested a protective role for LXA₄ and 15-epi-LXA₄ in asthma [114, 115]. HETEs are monohydroxy fatty acids primarily produced *via* LOX metabolism (5-LOX and 12/15-LOX in humans [116]), although they can also be generated non-enzymatically. The lipid mediator 15-HETE is the major arachidonic acid metabolite in human bronchi [117] and several studies have suggested that high 15-HETE levels are indicative of pro-inflammatory responses in asthma [118, 119]. In reactions analogous to the biosynthesis of CysLTs, 14,15-LTA₄ can be transformed further to the eoxins 14,15-LTC₄, 14,15-LTD₄ and 14,15-LTE₄ [120]. The biological function of eoxins remains unclear, but significant EXC₄

levels were observed in BALF of patients with a range of diseases, including eosinophilic pneumonia and asthma [121]. Eoxin levels were elevated in EBC from asthmatic relative to healthy children, with results suggesting a relationship between asthma severity and eoxin levels [122].

Isoprostanes, derived from arachidonic acid *via* auto-oxidation, have been primarily studied as markers of oxidative stress in lung diseases [123, 124]. Although they are not enzymatic products, they have distinct biological activities. Patients with pulmonary hypertension have increased isoprostane levels relative to healthy controls, and the response to inhaled NO has been correlated to basal levels of these compounds [125]. Among oxidative markers of lung diseases, 8-iso-PGF_{2 α} is a good candidate to study the influence of oxidative stress, because it shows strong constriction properties in smooth muscle *in vitro* through activation of the FP receptor [126]. Indeed, high urinary levels of 8-iso-PGF_{2 α} have been documented in extrinsic allergic alveolitis patients [127, 128].

Summary

The relevance of lipids in respiratory diseases has been well established, and there is a strong argument for their inclusion in systems biology-based efforts to identify biomarkers and explore disease mechanisms. The current state-of-the-art in lipidomics involves a combination of fast scanning tandem MS/MS and high-resolution mass spectrometers (*e.g.* Orbitrap technology-based systems) to identify individual lipid species (fig. 4). These platforms offer multiple advantages including: 1) increased accuracy; 2) flexibility in performing structural confirmation experiments; and 3) formatting for relatively high throughput analyses. Recent advances include the development of lipid-based informatics resources, such as the LIPID MAPS Lipidomics Gateway and specific lipid-based software (*e.g.* LipidView, SimLipid, LipidXplorer) designed to aid in the identification of multiple lipid species in a single analysis, and the ability to determine the position of unsaturated bonds in the fatty acid moieties of molecular lipids [129]. A limitation of the current LC-MS methodologies is their inability to perform exact quantification due to a paucity of authentic analytical standards, which moreover often co-elute with the large number of species acquired in a lipidomics profiling approach. Combined with the scarcity of databases or spectral libraries for compound identification, this makes routine lipidomics challenging. The ability to identify and quantify individual lipid species remains the key obstacle in most lipidomics studies; however, it is expected that future technical advances will significantly increase our ability to quantify such complex lipidomics profiles.

Metabolomics

Metabolomics is defined as “the analysis of the whole metabolome under a given set of physiological, environmental and/or clinical conditions” [130]. The exact definition of the metabolome varies, but can generally be considered to be the “quantitative complement of all of the low molecular weight molecules present in a particular physiological or developmental state” (*e.g.* metabolome of metabolic processes, cells,

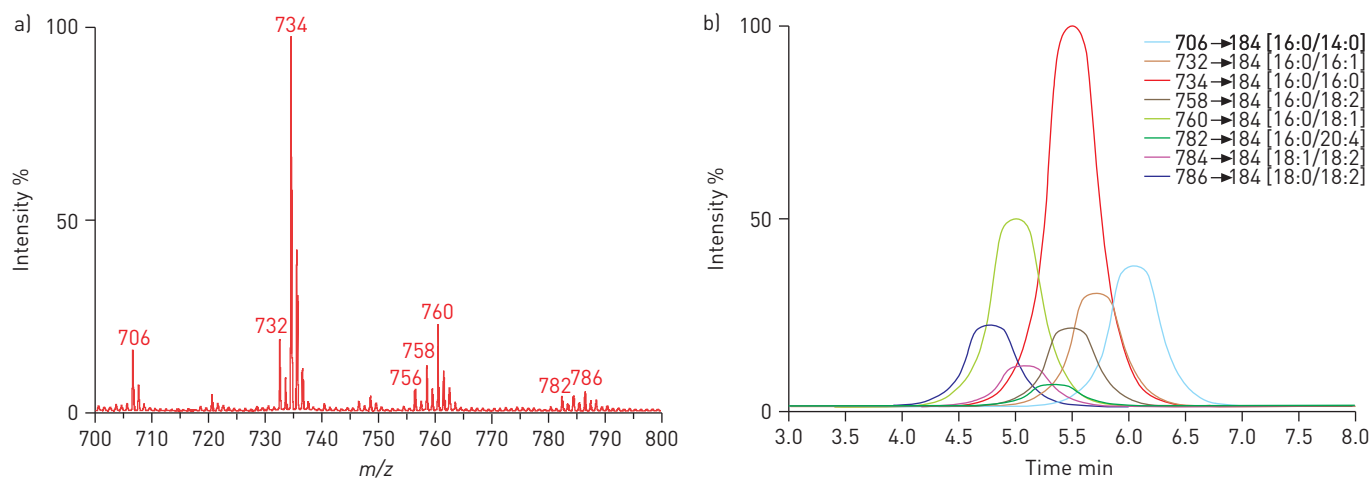


FIGURE 4 a) Direct infusion full scan and b) multiple reaction monitoring (MRM) mass spectra of human bronchoalveolar lavage lipid extracts. In the direct infusion example, the sample is introduced directly into the mass spectrometer without prior separation and a complete mass spectrum of all ions (full scan) is obtained without fragmentation. This is a rapid method for screening the lipid composition of a sample, but does not yield any structural information beyond the molecular mass, and may lack the sensitivity to detect analytes that occur in low abundance in the sample. In the MRM example the analytes are separated by liquid chromatography immediately prior to analysis, which enhances sensitivity. Furthermore, both mass analysers are set to only detect a specific mass; the first analyser selectively measures one precursor ion, which is then fragmented in a collision cell, and the second analyser selectively measures one of its fragments (the product ion). This allows for highly specific and sensitive screening of target analytes, using known transitions that are characteristic for a molecule's fragmentation pattern, albeit at the cost of ignoring the remaining composition of the sample.

tissues, organs or organisms) [131]. The application of metabolomics in the study of respiratory diseases is in its infancy, lagging behind other diseases (e.g. cancer and cardiovascular disease). It offers the ability to 1) classify specific respiratory diseases and sub-phenotypes (e.g. mild *versus* severe asthma) and 2) identify a “quantitative disease phenotype” (i.e. specific profiles/concentrations of metabolites diagnostic or prognostic for disease) [132]. Many of the initial metabolomics studies in the respiratory field were conducted with nuclear magnetic resonance (NMR) spectroscopy due to its ease of application and non-destructive nature, but mass spectrometry is increasingly used because of improved sensitivity and specificity.

Initial applications of metabolomics approaches in asthma studies have proven promising. A NMR-based metabolomics study identified 70 urinary metabolites that were collectively discriminant for a model of stable asthmatics as well as for a model of exacerbated asthmatics *versus* stable asthmatics, both with 94% accuracy [133]. Robust multivariate modelling (partial least squares-discriminant analysis) identified 23 metabolites as being altered, with TCA cycle metabolites significantly increased in both classification models (succinate, fumarate, oxaloacetate, 2-oxoglutarate and cis-aconitate). MATTARUCHI *et al.* [134] successfully classified a range of atopic asthma states using multivariate models (orthogonal projections to latent structures-discriminant analysis) generated from non-targeted LC-MS profiling of urine. The first model differentiated asthmatics and healthy controls with 98% accuracy, the second distinguished between medicated and non-medicated asthmatics with an accuracy of 96%, and the third separated well- and poorly controlled asthmatics with an accuracy of 100%. Focused investigation in asthma revealed reduced excretion of urocanic and methyl-imidazoleacetic acid as well as a metabolite resembling an Ile-Pro fragment. CARRARO *et al.* [35] used NMR profiles of EBC to classify asthma with an accuracy of 86%, which was a slight improvement over the 81% accuracy based on exhaled NO and forced expiratory volume in 1 s (FEV₁).

NMR analysis of serum differentiated moderate (Global Initiative for Chronic Obstructive Lung Disease (GOLD) III) from severe COPD (GOLD IV) with an accuracy of 82% [135]. The discrimination of patients from healthy controls was due to decreased levels of the branched-chain amino acids (BCAA) valine and isoleucine, possibly the result of weight loss due to proteolysis in patients with cachexia because BCAAs have been shown to correlate with body mass index [135]. LC-MS analysis of metabolites in plasma successfully classified emphysematous COPD and non-emphysematous patients with an accuracy of 64.3% using hierarchical clustering [136]. However, multivariate modelling (linear discriminant analysis) of the top seven biomarkers (whose structures were not identified) improved classification accuracy to 96.5% [136]. NMR profiling of EBC successfully modelled stable cystic fibrosis patients with an accuracy of 96% (91% accuracy and 96% specificity). A second model differentiated unstable cystic fibrosis with an accuracy of 95% (86% accuracy and 94% specificity) [137].

Summary

High resolution LC-MS-based platforms currently represent the forefront of metabolomics techniques due to: 1) their ability to simultaneously measure large numbers of unrelated metabolites; 2) capability of analysing metabolites with a wide range of chemical properties; and 3) relatively straight forward sample preparation compared to other metabolomics techniques. One of the most significant recent advances in metabolomics is the development of kit-based technologies such as that sold by BIOCRATES [138]. These technologies are straightforward and illustrate the potential for metabolomics analysis to become a routine component of the 'omics toolbox. However, metabolomics approaches have a number of limitations; first, no platform can currently measure all the metabolites present within a sample and secondly, there is a lack of tools for data annotation, meaning that only a fraction of the information present within a data set can be properly interpreted. Thus, the most significant advances in the field will require developing tools or methods to deal with these limitations. To date, the majority of metabolomics studies in respiratory disease have focused on developing and validating the approach; however, in the future metabolomics will play an important role in identifying biomarkers and elucidating mechanisms of disease. The high classification accuracy of the models generated from material collected noninvasively (e.g. urine) suggests that metabolomics can play a central role in discrimination of disease and quantitative phenotyping. A long-term goal is the ability to identify prognostic and/or diagnostic patterns of metabolites in relation to disease. In addition, the identification of individual metabolites responsible for differentiating between patients with respiratory disease and healthy controls will provide valuable information on the metabolic mechanisms of disease. Accordingly, it is expected that the application of metabolomics approaches (both NMR- and mass spectrometry-based) will increase significantly.

Breathomics

Exhaled air contains a complex mixture of VOCs in the gas phase [139] and non-volatile compounds derived from condensed water vapour and aerosol particles (so-called EBC) [140]. The origin of exhaled

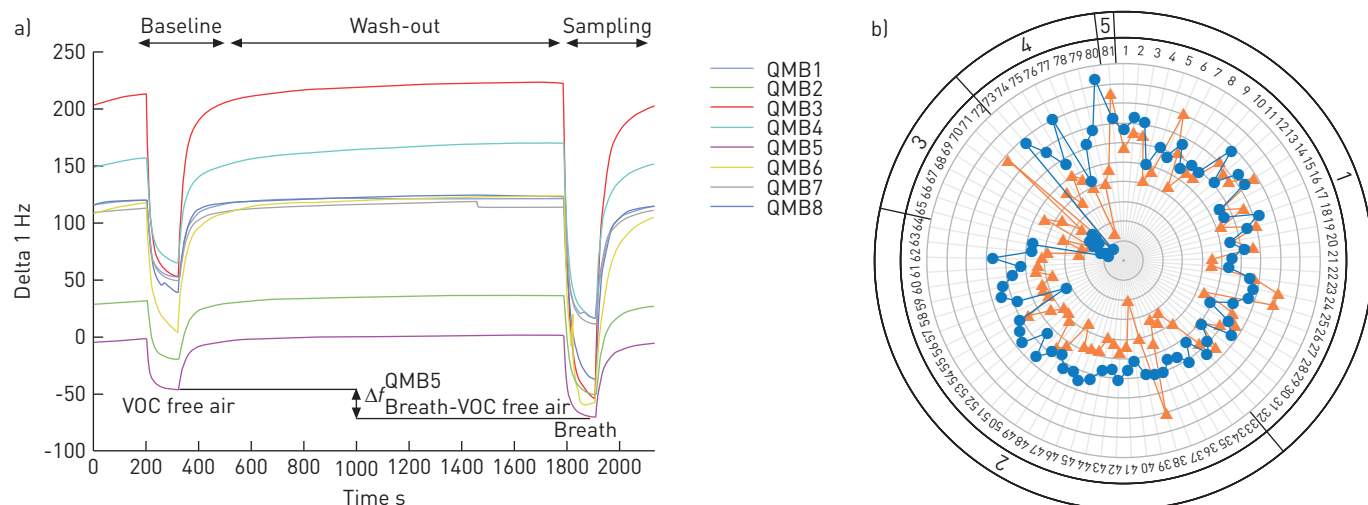


FIGURE 5 Breathprints of eNoses. a) A typical breathprint from a patient with asthma obtained with a non-commercial eNose consisting of eight quartz microbalance (QMB) gas sensors coated by molecular films of metalloporphyrins. Deflections represent the response (change in sensor frequency) to volatile organic compound (VOC)-free air used as a baseline and the patient breath sample. The actual response (Δf) for individual sensors is calculated by subtracting baseline response to VOC-free air from patient breath sample response (sampling). A wash-out phase is performed between baseline and sampling. QMB5: sensor 5. b) Breathprints of exhaled air as obtained in two different patients with severe asthma: one nonsmoking patient (circles) and one smoking patient (triangles). The breathprints are generated by an eNose platform, combining five different eNoses (four different brands, one duplicated brand) adding up to an array of 81 sensors in total. The 81 sensors from the five eNoses (1–5 are listed in a circular display, comprising: carbon-black polymer composite, quartz crystal microbalance metalloporphyrin, metal oxide semiconductor sensors and a field asymmetric ion mobility spectrometer). The signals from all sensors have been normalised towards an arbitrary unit at a scale between 0 (centre) and 100 (outer circle). The sensor array exhibits differential signals between the two patients, demonstrating two different signatures or “breathprints”. When using adequate training and validation sets eNose signals can be examined for their diagnostic accuracy for (subphenotypes) of disease [144, 145].

metabolites varies because they result from both systemic and local metabolic, inflammatory and oxidative activities. The advantage of exhaled air analysis is its noninvasive nature. Metabolomics approaches in exhaled air are currently referred to as “breathomics”. The standard for detecting individual molecular compounds in VOCs is GC-MS [141], but other sophisticated analytical equipment, such as proton transfer reaction mass spectrometry (PTR-MS), ion mobility spectrometry, and selected ion flow tube mass spectrometry (SIFT-MS), can also be used. The combined molecular composition of gas mixtures can be assessed by electronic noses (eNoses) [142, 143] based on arrays of nano-sensors that do not identify individual chemical constituents but patterns of interactions. The output of eNoses is a signature of the VOC mixture, which can be regarded as a fingerprint of a complex gas mixture (fig. 5). There are several principles underpinning eNose sensors, including conducting polymers, metal oxide, metal oxide field effect transistors, surface or bulk acoustic waves, optical sensors, colorimetric sensors, ion mobility spectrometry, infrared spectroscopy, gold nanoparticles, and also GC-MS [142, 143]. Analysis of eNose data involves pattern recognition algorithms to develop signatures of complex, exhaled air mixtures (breathprints) [146]. Breathprints can be discriminated by trained dogs and this observation has facilitated research on olfactory sensation and signal pathways to develop new analytic techniques [147]. EBC can be analysed by almost any (bio)assay, in which the limits of detection represent the most common problem. Metabolomics of EBC by NMR spectroscopy is currently the most promising approach (see below) [35, 137, 148].

GC-MS analysis of exhaled breath

Apart from lung cancer and other diseases [149], GC-MS has been employed in the study of inflammatory lung diseases, including asthma [150, 151], cystic fibrosis [152] and COPD [153]. When analysing 945 compounds in exhaled breath, a total of eight compounds allowed for a 92% correct classification of asthmatic and non-asthmatic children [150]. Notably, the profile of exhaled VOCs was found to be associated with either predominantly eosinophilic or neutrophilic inflammation in asthma and COPD [154, 155]. This suggests that breathomics is suitable for noninvasive subphenotyping of inflammatory airway diseases and, possibly, for disease monitoring.

eNose

Studies conducted with different eNose sensor systems suggest that asthmatics can be discriminated from healthy controls with accuracies between 80–100% [151, 156]. Interestingly, COPD patients can also be discriminated from asthmatics [157], a finding confirmed by external validation according to STARD guidelines [158]. Such confirmation in newly recruited patients from different hospitals is essential for

limiting false discovery rates [144] and validating diagnostic accuracy [9]. It should be emphasised that the observed differences in eNose breathprints between asthma and COPD are not due to differences in smoking habits because comparison of asthmatics and nonsmoking COPD patients showed the same differences [157, 158]. Similarly, the level of airways obstruction *per se* does not seem to affect the eNose signal because the breathprints are stable before and after acute bronchoconstriction and bronchodilation in asthma [159]. eNose systems have also been applied successfully in studies of other diseases, such as lung cancer [160]; however, full validation of these results is pending. The critical unresolved issue in current eNose research is the need for mapping (*i.e.* performing quantitative comparison between devices) [161] since sensor signals are not identical between different devices.

Exhaled breath condensate

Determination of pH, adenosine and eicosanoids in EBC has provided useful information on the pathophysiological processes in asthma and COPD [140]. Recently, protein multiplex assays have been applied to EBC, showing sufficient signal in asthmatics when using a breath-recycling condenser method [162]. The most promising development is high-throughput, metabolomics analysis of EBC by NMR spectroscopy, which is a robust analytical technique [163]. Several independent laboratories have recently demonstrated profiles of metabolites in EBC, providing discriminatory signals for asthma [35], COPD [148] and cystic fibrosis [137]. These studies suggest that both volatile and non-volatile exhaled breath metabolomics are ready for stringent validation of diagnostic accuracies.

Summary

The current status of breathomics is that: 1) analytical instruments for the individual identification of gaseous metabolites in exhaled air are in place (GC-MS) and are used for pathophysiological research; 2) the validation of on site, portable eNoses in the clinical diagnosis and monitoring of respiratory disease is ongoing, but has not yet been finalised; and 3) NMR spectroscopy is the most powerful method for analysis of fluid-phase EBC. The recent advances in breathomics research include the development of tailored nano-sensor arrays for specific disease entities. In addition, the first studies demonstrating external validation of diagnostic accuracy by eNoses have recently been published, whilst longitudinal monitoring studies in asthma and COPD are underway. The largest limitation of using eNoses is poor between-device comparability of numerical data. In addition, currently, only one single brand of eNose is commercially available (Cyranose 320). This limits rapid progress in multicentre studies, but can be overcome by centralised analysis on a multi-eNose platform of shipped breath samples in desorption tubes. The future prospects of eNose analysis in medicine are widely recognised. Cheap, on site, real-time breath analysis with online analysis against existing databases on a “Breathcloud” is a feasible prospect for diagnostic assessment in low-income countries (*e.g.* tuberculosis) as well as in high-income countries (*e.g.* lung cancer, and asthma in infants). The positioning of eNoses in these fields will require a maximal negative-predictive value (high sensitivity), allowing reduced and selective usage of more expensive, invasive and/or hazardous diagnostic procedures.

Integrating clinical and functional genomics data into fingerprints and phenotype handprints

Systems biology

Properties of a biological system are not only defined by the simple addition of elementary functions, but also emerge from the interactions between its elements at each level of biological organisation (molecules, organelles, cells, tissues and organs) [164]. Inferring interactions between these constituents (*e.g.* genes, proteins and ligands) and unravelling their regulatory mechanisms is key to defining the emergent properties. Systems biology approaches aim to understand the system behaviour as a whole. They produce a convincing mathematical and computational model linking the highly complex interactions between system components to emergent properties [165, 166]. The primary challenges encountered in implementing systems biology approaches are: 1) the complexity of biological systems; 2) the multi-scale nature of the range of biological information encoded in DNA, RNA, proteins, metabolites and interaction networks at different levels of biological organisation (*e.g.* in cells, tissues, organs and the entire organism), which occur over various timescales; 3) the vast amount of data generated by omics technologies; and 4) the scatter of heterogeneous knowledge. Accordingly, integrative systems biology approaches combine experimental methods with mathematical and computational methods to model and simulate molecular, subcellular, cellular, and organ-level structures and processes [164]. This approach offers the ability to gain a deeper understanding of functional and regulatory pathways that play central roles in the behaviour of complex biological systems. A typical workflow includes: 1) data processing, 2) inference of networks representing relationships between the molecular entities surveyed, 3) deep curation of available data and knowledge, 4) simulation of the system’s behaviour and 5) model analysis.

Processing and preliminary analysis

The typical analysis of an 'omics data set (following standard quality control) proceeds in four steps: 1) detection of raw signals (microarray hybridisations, mass spectra, eNose patterns, *etc.*), 2) preprocessing (subtraction of background noise, smoothing, peak detection, calculation of levels of expression), 3) normalisation of data and 4) identification of differentially expressed genes, peptides, metabolites or lipids for further data analysis, including feature selection, clustering, classification and pathway/network analyses. For example, in the U-BIOPRED (Unbiased BIOMarkers in PREDiction of respiratory disease outcomes) project, a project funded by the Innovative Medicines Initiative and focusing on severe asthma, an automated data analysis pipeline has been developed initially for lipidomics and proteomics. Within the general workflow each type of 'omics data requires the use of specific bioinformatics tools. Transcriptomics is considered to have the most developed, well established and robust data analysis pipeline, followed by proteomics and lipidomics, for which specific databases have been built by several consortia (*e.g.* LIPIDMAPS [39, 167, 168]). Integration of 'omics data requires normalisation of data from different platforms, data formats and identifiers. This is typically performed by transforming values to obtain zero mean and unit standard deviation. The analysis of each type of 'omics data requires specific selection from a wide range of statistical, data mining and machine learning techniques adapted for unbiased/biased, unsupervised/supervised and uni-/multi-variate analyses. Networks linking the individual data readouts are well-suited tools to represent interactions between entities as they depict the wide range of relationships (edges) observed between very large numbers of elements (nodes) [169]. Furthermore, powerful statistical and computational techniques, such as those borrowed from graph theory, are applied to the analysis of biological networks, *e.g.* to identify key proteins or master regulators, *i.e.* nodes interacting with a very large number of immediate neighbours. Other major methods in a typical workflow include power and sample size calculation, feature selection (*e.g.* bootstrapping, wrapper), principal component analysis, clustering (*e.g.* hierarchical, k-means, ward, biclustering), and classification (*e.g.* support vector machine and Bayesian networks). A detailed description of the advantages and limitations of the methods and their combinations is beyond the scope of this article, but these aspects have been recently reviewed elsewhere [169, 170].

Multi-omics integration

Identifying similar patterns of 'omics data can be performed using clustering methods (conventional and biclustering) and further functional analyses using network and pathway inference, representation and analysis software tools (*e.g.* ingenuity pathway analysis [171] and Cytoscape [172]). Causal relationships between entities measured with 'omics technologies under different conditions and/or at different time points can be modelled in probabilistic causal networks, using the Bayesian paradigm to estimate the probability of relationships based on prior knowledge, or using mutual information (a measure of dependence or reciprocal informativeness between two variables). These methods were first developed to analyse data of a single type (*e.g.* transcriptomics gene expression profiles), but have now been extended to integrate information on genome-wide genetic variation, DNA-binding and protein-protein interactions [170]. A useful example is the ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) algorithm, which was specifically designed to scale-up the complexity of regulatory networks [173, 174]. Although a large amount of data is generated by 'omics approaches, the data are still generally too scarce compared to the high number of possible interactions tested. One consequence is that method accuracy is often tested on simulated datasets, which do not reflect true biological complexity, rather than on widely accepted benchmarks [170]. A major drawback of this approach is that it does not provide access to mechanisms and causality, which have to be addressed by other methods. Deep curation relies not only on 'omics data but also on its integration with the vast amount of knowledge available in the literature and pathway databases after curation by experts (*e.g.* KEGG [175]), and can therefore include mechanisms and causal relationships, implemented using standards such as Systems Biology Markup Language (SBML) or Systems Biology Graphical Notation (SBGN) [176], and widely used visualisation and modelling tools such as CellDesigner [177] or Cytoscape [178].

The major drawbacks of pathways used in deep curation are their inaccuracy, incompleteness, lack of documentation of the context (*e.g.* tissue) and the huge amount of time and expertise required for proper curation and regular updating. These issues are being addressed by automatic text mining [179, 180] and community-based efforts (*e.g.* WikiPathways [181]). The importance of dynamics of interactions in time and space is not captured by static models generated by data-driven probabilistic networks or by deep curation. Dynamic models mainly use ordinary and partial differential equations, or Boolean networks (*i.e.* based on logical data type) for gene regulation, while calibration relies mostly on measures obtained with *in vitro* assays.

Data, information and knowledge management

The ultimate value of systems biology and medicine is in the integration of 'omics data across platforms and cellular levels. This requires effective knowledge management tools and computational platforms to collect,

manage, analyse and share clinical and experimental data, and integrate them with prior knowledge stored in public databases (e.g. PubMed [182], BIND [183], Reactome [184], KEGG [185]). Software platforms aim to render data and knowledge available at any step of the workflow, provide high interoperability, avoid errors in handling and analysis of data or models, and thereby improve and accelerate the full analysis. As biological datasets are described using highly heterogeneous formats, nomenclatures and data schema, the development of standards is essential to enable their integrative analysis. Standards for data management address issues of minimum information (e.g. Minimum Information About a Microarray Experiment [MIAME] [41]), file format (i.e. how the information should be stored, usually XML-based) and ontologies (e.g. Gene Ontology [GO] [186] and Systems Biology Ontology [SBO] [187]). A functional interface is required and it enables users to browse, query and retrieve information for genes, proteins, lipids, metabolites or pathways and networks of interest. Several software platforms have been developed for this purpose: 1) spreadsheet tab-delimited template-based files used *via* specific interfaces to analysis software; 2) online wiki-based secure data and analysis tools; 3) laboratory information management systems (LIMS); 4) workflow management systems, such as Galaxy [188] for genomics and Konstanz Information Miner (KNIME [189]); and 5) Ensembl [190] and UCSC [191] genome browsers. More recent efforts include: 1) integration of transcriptomics and protein–protein interactions such as the Sage Bionetworks initiative; 2) open source software integration such as the Garuda Alliance and the tranSMART platform [192]; and 3) commercial proprietary systems from IDBS (ClinicalSense), Oracle (Translational Research Solution) and BioMax Informatics (BioXM).

The tranSMART platform [7, 193] was originally developed by the pharmaceutical company Janssen Research and Development to effectively manage knowledge associated to its own internal biomarker research. In parallel it was made available to external research groups as an enabling platform for translational research collaborations. tranSMART enables research teams to manage both analysis results and the patient level clinical, omics and genetics data of biomarker studies. It enables researchers to explore the different types of data produced in a biomarker study, generate and test a novel hypothesis within a study and explore the relationships between studies. As an open platform, tranSMART also leverages other open-source tools such as those of the academic i2b2 consortium [194] or the R project for Statistical Computing [195]. tranSMART is now being used by several consortia (U-BIOPRED, OncoTrack, SAFE-T, PreDiCT-TB, BTCure, eTRIKS, EMIF) supported by the Innovative Medicines Initiative (fig. 1). Alternatively, BioXM, developed by BioMax Informatics, has been used for knowledge management of the BioBridge project that studied COPD [196], as well as a number of other projects.

Replication

These experimental and analytical tools have enabled the identification of many molecular fingerprints [145]. However, uncertainties in their reported accuracy have resulted in overly optimistic expectations on the predictive value of molecular profiles. Indeed, only a fraction of the reported signatures have been validated and proven to be useful. The integration of omics datasets for multiple biological levels and data types remains a challenge. Indeed, such complex datasets suffer from biological and technical biases, noise and errors that may lead to false positive and false negative discoveries. To overcome these limitations, best practices and guidelines for the development of omics-based molecular profiles continue to evolve [145]. The vast amount and diversity of information obtained with omics technologies cannot serve as a surrogate for appropriate experimental design. First, an efficient design relies on hypothesis formulation, phenotype definition, power and sample size calculation, multiple testing correction, and plans for replication and experimental validation [197–204]. Secondly, efficient implementation requires standardised experimental protocols and quality control procedures, data annotation, representation and modelling with novel algorithms and data integration tools. These measures help to reduce, albeit not totally eliminate, potential errors and to strike a suitable balance between sensitivity and specificity, thereby improving the accuracy of prognostic and diagnostic biomarkers. Defining profiles with multiple types of omics data may greatly improve their usefulness. This strategy is being applied, for example, in attempts to integrate transcriptomics with protein–protein interaction networks and/or metabolomics [170, 205], while the first integrated personal omics profile for a single subject over time has recently been reported [206], supporting the current general trend towards personalised medicine.

Predictive power of systems biology

Statistical analysis of omics datasets pose problems because of the large number of features that they measure and sheer volume of data that they generate. Standard statistical methods are not directly applicable without correction for multiple testing or assessment of false discovery rates, which are suitable for identification of individual and independent biomarkers. The strength of systems biology over individual biomarkers is two-fold: its integration of independent, single omics datasets to define their intersection and its focus on networks of interrelated elements that are collectively changing in relation to

disease or external stimuli. This reduces the numbers of patients needed to demonstrate differences between clinical phenotypes and effect of treatment, and is the strategy now being successfully implemented in several clinical studies, e.g. in respiratory [174, 207], cardiovascular [208], infectious [209] and neurological [210] diseases, as well as cancer [211, 212] and nephrology [213].

Towards systems medicine of respiratory diseases

Systems biology approaches have been successfully applied to respiratory diseases and have, for example, suggested that skeletal muscle degeneration in COPD may be caused by cell hypoxia due to abnormal expression of histone modifiers linked to poor coordination between remodelling of several tissues and energy sources [207]. Several large-scale multicentre collaborative projects have now started to develop such methods to decipher the development of respiratory diseases. Their common goal is to identify novel, complex biomarker profiles that combine diverse clinical, biological and functional genomics data types into molecular fingerprints and disease phenotype handprints. These novel diagnosis and prognosis tools aim to improve disease prevention and help identify new drug targets for better, personalised therapy [214]. Such “systems medicine” projects rely on the joint efforts of multidisciplinary experts from academic research institutes, hospital centres, small companies and the pharmaceutical industry and will help advance translational medicine [214]. A major challenge encountered in these projects is to define the optimal range, combination and depth of experimental methods necessary to improve understanding of disease and its treatment (e.g. whole or targeted transcriptomics and/or proteomics/metabolomics/lipidomics). Financial and time constraints are important factors and even more so in the context of clinical applications and public health. The U-BIOPRED [215], AirPROM (Airway Disease PRedicting Outcomes through patient-specific computational Modelling) [216] and MeDALL (Mechanisms of the Development of ALLergy) [217] consortia are implementing this research strategy in a coordinated manner to overcome hurdles in understanding and treating severe asthma (U-BIOPRED), COPD (AirPROM) and allergic diseases (MeDALL) [37, 218, 219]. Unbiased approaches necessitate comprehensive genome-wide initial analyses, which may then be adapted to specific objectives and available biological resources. Another project, Synergy-COPD [220], aims to produce a computer model of the mechanisms of COPD built using epidemiological data, clinical trials and physician interviews, translated into patient-based models that will contribute to replicating human physiology. Thus, iterative perturbation of a biological system of interest *ex vivo* and/or *in vivo*, and *in silico* in large-scale experiments to generate and then refine integrative phenotype handprints holds the promise for deeper understanding, diagnosis and treatment of respiratory and other complex chronic diseases [214].

Concluding remarks

The use of ‘omics approaches to elucidate mechanisms of disease has grown exponentially in recent years, driven by marked improvements in analytical platforms, with increasing resolution and sensitivity as well as increased throughput and reduced cost. Paramount to making good use of the vast amount of data generated by the ‘omics methods is the creation of appropriate knowledge management and data handling platforms, together with judicious application of bioinformatics, statistics and modelling tools. It is also important to acknowledge that ‘omics-based tests, including both prognostic and diagnostic tools based upon shifts in patterns of variables, are highly prone to errors and require rigorous statistical handling. It has been recommended by the Institute of Medicine that “all information needed to verify the test discovery process be disclosed through publication or patent application” and that “the computational procedures must be “locked down” (recorded and no longer changed) and then confirmed with a new set of samples not used in the initial discovery” [221]. The use of systems biology to analyse ‘omics data comes with significant challenges, and will have to comply with similar rules that properly describe how studies with such large datasets can be designed with adequate power, taking advantage of the dimensionality reduction introduced by the identification of network modules in the biomarker discovery process.

Translation of all these capabilities into stratified medicine has yet to take place and will require large collaborative efforts delivered with the help of public–industrial partnership schemes that not only fund the programmes, but also bring together the considerable expertise that exists in both academia and the pharmaceutical industry. Whilst there are challenges to these new operational models, not least their complexities, it is hoped that such global systemic approaches to disease, combined with well-proven reductionist/focused approaches that have been central to the development of modern science, will lead to paradigm shifts in disease characterisation. It should finally be emphasised that the quality of the data generated by the ‘omics approach is highly dependent upon the quality of the clinical phenotyping of the subjects, further stressing the importance of rigorous phenotyping. Ultimately, the real test will be to demonstrate that such integrated approaches provide novel insight into disease mechanisms, speed up the drug discovery process and enable early disease detection. Accordingly, the nascent field of systems

medicine needs to prove its ability in both the clinic and the laboratory to deliver on its promise of shifting the paradigm of clinical study towards a large-scale biology discovery approach to detecting, understanding, treating and, ultimately, curing and preventing disease.

References

- 1 American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. *Am J Respir Crit Care Med* 2002; 165: 277–304.
- 2 Moore WC, Meyers DA, Wenzel SE, *et al.* Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010; 181: 315–323.
- 3 Han MK, Agusti A, Calverley PM, *et al.* Chronic obstructive pulmonary disease phenotypes: the future of COPD. *Am J Respir Crit Care Med* 2010; 182: 598–604.
- 4 Kitaguchi Y, Fujimoto K, Kubo K, *et al.* Characteristics of COPD phenotypes classified according to the findings of HRCT. *Respir Med* 2006; 100: 1742–1752.
- 5 The Global Initiative for Asthma. www.ginasthma.org/
- 6 The Global Initiative for Chronic Obstructive Lung Disease. www.goldcopd.org/
- 7 Szalma S, Koka V, Khasanova T, *et al.* Effective knowledge management in translational medicine. *J Transl Med* 2010; 8: 68.
- 8 Standards for the Reporting of Diagnostic Accuracy Studies. www.stard-statement.org Date last updated: April 22, 2008. Date last accessed: June 27, 2013.
- 9 Bossuyt PM, Reitsma J, Bruns DE, *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138: W1–W12.
- 10 Griese M, Noss J, von Bredow C. Protein pattern of exhaled breath condensate and saliva. *Proteomics* 2002; 2: 690–696.
- 11 Breathomics Platform Netherlands. www.breathomics.nl Date last updated: June 20, 2013. Date last accessed: June 27, 2013.
- 12 Anderson LN, Anderson N. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Prot* 2002; 1: 854–867.
- 13 Krug S, Kastenmüller G, Stücker F, *et al.* The dynamic range of the human metabolome revealed by challenges. *FASEB J* 2012; 26: 2607–2619.
- 14 Bray D. Molecular prodigality. *Science* 2003; 299: 1189–1190.
- 15 Joyce AR, Palsson BO. The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* 2006; 7: 198–210.
- 16 Wheelock CE, Wheelock AM, Kawashima S, *et al.* Systems biology approaches and pathway tools for investigating cardiovascular disease. *Mol Biosyst* 2009; 5: 588–602.
- 17 Fenn J, Mann M, Meng C, *et al.* Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989; 246: 64–71.
- 18 Wilm M. Principles of electrospray ionization. *Mol Cell Proteomics* 2011; 10: M111.009407.
- 19 Albrethsen J. Reproducibility in protein profiling by MALDI-TOF mass spectrometry. *Clin Chem* 2007; 53: 852–858.
- 20 McLerran D, Grizzle WE, Feng Z, *et al.* SELDI-TOF MS whole serum proteomic profiling with IMAC surface does not reliably detect prostate cancer. *Clin Chem* 2008; 54: 53–60.
- 21 Bowler RP, Canham ME, Ellison MC. Surface enhanced laser desorption/ionization (SELDI) time-of-flight mass spectrometry to identify patients with chronic obstructive pulmonary disease. *COPD* 2006; 3: 41–50.
- 22 Calvo FQ, Fillet M, de Seny D, *et al.* Biomarker discovery in asthma-related inflammation and remodeling. *Proteomics* 2009; 9: 2163–2170.
- 23 Gomes-Alves P, Imrie M, Gray RD, *et al.* SELDI-TOF biomarker signatures for cystic fibrosis, asthma and chronic obstructive pulmonary disease. *Clin Biochem* 2010; 43: 168–177.
- 24 Makarov A, Scigelova M. Coupling liquid chromatography to Orbitrap mass spectrometry. *J Chromatogr A* 2010; 1217: 3938–3945.
- 25 Schuhmann K, Herzog R, Schwudke D, *et al.* Bottom-up shotgun lipidomics by higher energy collisional dissociation on LTQ Orbitrap mass spectrometers. *Anal Chem* 2011; 83: 5480–5487.
- 26 Masoodi M, Eiden M, Koulman A, *et al.* Comprehensive lipidomics analysis of bioactive lipids in complex regulatory networks. *Anal Chem* 2010; 82: 8176–8185.
- 27 Stoop MP, Coulter L, Rosenling T, *et al.* Quantitative proteomics and metabolomics analysis of normal human cerebrospinal fluid samples. *Mol Cell Proteomics* 2010; 9: 2063–2075.
- 28 Weber RJM, Southam AD, Sommer U, *et al.* Characterization of isotopic abundance measurements in high resolution FT-ICR and Orbitrap mass spectra for improved confidence of metabolite identification. *Anal Chem* 2011; 83: 3737–3743.
- 29 Broadbelt KG, Rivera KD, Paterson DS, *et al.* Brainstem deficiency of the 14-3-3 regulator of serotonin synthesis: a proteomics analysis in the sudden infant death syndrome. *Mol Cell Proteomics* 2012; 11: M111.009530.
- 30 Hart P, Francese S, Claude E, *et al.* MALDI-MS imaging of lipids in *ex vivo* human skin. *Anal Bioanal Chem* 2011; 401: 115–125.
- 31 Gallien S, Duriez E, Domon B. Selected reaction monitoring applied to proteomics. *J Mass Spectrom* 2011; 46: 298–312.
- 32 Han X, Gross RW. Shotgun lipidomics: multidimensional MS analysis of cellular lipidomes. *Expert Rev Proteomics* 2005; 2: 253–264.
- 33 Soga T, Ohashi Y, Ueno Y, *et al.* Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res* 2003; 2: 488–494.
- 34 Cáp P, Chládek J, Pehal F, *et al.* Gas chromatography/mass spectrometry analysis of exhaled leukotrienes in asthmatic patients. *Thorax* 2004; 59: 465–470.
- 35 Carraro S, Rezzi S, Reniero F, *et al.* Metabolomics applied to exhaled breath condensate in childhood asthma. *Am J Respir Crit Care Med* 2007; 175: 986–990.

- 36 Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001; 19: 242–247.
- 37 Auffray C, Adcock IM, Chung KF, *et al.* An integrative systems biology approach to understanding pulmonary diseases. *Chest* 2010; 137: 1410–1416.
- 38 Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol* 2011; 9: 34.
- 39 Imbeaud S, Auffray C. ‘The 39 steps’ in gene expression profiling: critical issues and proposed best practices for microarray experiments. *Drug Discov Today* 2005; 10: 1175–1182.
- 40 Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. *Genome Med* 2009; 1: 2.
- 41 Brazma A, Hingamp P, Quackenbush J, *et al.* Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat Genet* 2001; 29: 365–371.
- 42 Brazma A. Minimum Information About a Microarray Experiment (MIAME) – successes, failures, challenges. *Scientific World Journal* 2009; 9: 420–423.
- 43 Hurd PJ, Nelson CJ. Advantages of next-generation sequencing *versus* the microarray in epigenetic research. *Brief Funct Genomic Proteomic* 2009; 8: 174–183.
- 44 Pickrell JK, Marioni JC, Pai AA, *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 2010; 464: 768–772.
- 45 Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009; 10: 57–63.
- 46 Mortazavi A, Williams BA, McCue K, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008; 5: 621–628.
- 47 Raz T, Kapranov P, Lipson D, *et al.* Protocol dependence of sequencing-based gene expression measurements. *PLoS One* 2011; 6: e19287.
- 48 Ask K, Eickelberg O, Gauldie J, *et al.* Have advanced research technologies made real impact on respiratory medicine? *Respirology* 2010; 15: 876–880.
- 49 Woodruff PG, Boushey HA, Dolganov GM, *et al.* Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. *Proc Natl Acad Sci USA* 2007; 104: 15858–15863.
- 50 Woodruff PG, Modrek B, Choy DF, *et al.* T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med* 2009; 180: 388–395.
- 51 Choy DF, Modrek B, Abbas AR, *et al.* Gene expression patterns of Th2 inflammation and intercellular communication in asthmatic airways. *J Immunol* 2011; 186: 1861–1869.
- 52 Corren J, Lemanske RF, Hanania NA, *et al.* Lebrikizumab treatment in adults with asthma. *N Engl J Med* 2011; 365: 1088–1098.
- 53 Kapitein B, Hoekstra MO, Nijhuis EH, *et al.* Gene expression in CD4⁺ T-cells reflects heterogeneity in infant wheezing phenotypes. *Eur Respir J* 2008; 32: 1203–1212.
- 54 Tsitsiou E, Williams AE, Moschos SA, *et al.* Transcriptome analysis shows activation of circulating CD8⁺ T cells in patients with severe asthma. *J Allergy Clin Immunol* 2012; 129: 95–103.
- 55 Al-Garawi A, Husain M, Ilieva D, *et al.* Shifting of immune responsiveness to house dust mite by influenza A infection: genomic insights. *J Immunol* 2012; 188: 832–843.
- 56 Tilley AE, O’Connor TP, Hackett NR, *et al.* Biologic phenotyping of the human small airway epithelial response to cigarette smoking. *PLoS One* 2011; 6: e22798.
- 57 Heguy A, O’Connor T, Luettich K, *et al.* Gene expression profiling of human alveolar macrophages of phenotypically normal smokers and nonsmokers reveals a previously unrecognized subset of genes modulated by cigarette smoking. *J Mol Med* 2006; 84: 318–328.
- 58 Harvey BG, Heguy A, Leopold P, *et al.* Modification of gene expression of the small airway epithelium in response to cigarette smoking. *J Mol Med* 2007; 85: 39–53.
- 59 Pierrou S, Broberg P, O’Donnell RA, *et al.* Expression of genes involved in oxidative stress responses in airway epithelial cells of smokers with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2007; 175: 577–586.
- 60 Beane J, Vick J, Schembri F, *et al.* Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prev Res (Phila)* 2011; 4: 803–817.
- 61 Steiling K, Kadar AY, Bergerat A, *et al.* Comparison of proteomic and transcriptomic profiles in the bronchial airway epithelium of current and never smokers. *PLoS One* 2009; 4: e5043.
- 62 Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* 2009; 11: 49–79.
- 63 Palmblad M, Tiss A, Cramer R. Mass spectrometry in clinical proteomics – from the present to the future. *Proteomics Clin Appl* 2009; 3: 6–17.
- 64 Nicholas B, Skipp P, Mould R, *et al.* Shotgun proteomic analysis of human-induced sputum. *Proteomics* 2006; 6: 4390–4401.
- 65 Louten J, Mattson JD, Malinao MC, *et al.* Biomarkers of disease and treatment in murine and cynomolgus models of chronic asthma. *Biomark Insights* 2012; 7: 87–104.
- 66 Bozinovski S, Hutchinson A, Thompson M, *et al.* Serum amyloid A is a biomarker of acute exacerbations of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2008; 177: 269–278.
- 67 Merkel D, Rist W, Seither P, *et al.* Proteomic study of human bronchoalveolar lavage fluids from smokers with chronic obstructive pulmonary disease by combining surface-enhanced laser desorption/ionization-mass spectrometry profiling with mass spectrometric protein identification. *Proteomics* 2005; 5: 2972–2980.
- 68 Nicholas BL, Skipp P, Barton S, *et al.* Identification of lipocalin and apolipoprotein A1 as biomarkers of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010; 181: 1049–1060.
- 69 Ohlmeier S, Mazur W, Linja-Aho A, *et al.* Sputum proteomics identifies elevated PIGR levels in smokers and mild-to-moderate COPD. *J Proteome Res* 2012; 11: 599–608.
- 70 Kohler M, Sandberg A, Kjellqvist S, *et al.* Gender differences in the bronchoalveolar lavage cell proteome of patients with COPD. *J Allergy Clin Immunol* 2012 [In press DOI: 10.106/j.jaci.2012.09.024].
- 71 Haynes PA, Yates JR. Proteome profiling – pitfalls and progress. *Yeast* 2000; 17: 81–87.

- 72 Bortner JD, Richie JP, Das A, *et al.* Proteomic profiling of human plasma by iTRAQ reveals down-regulation of ITI-HC3 and VDBP by cigarette smoking. *J Proteome Res* 2010; 10: 1151–1159.
- 73 O’Neil SE, Lundbäck B, Lötvall J. Proteomics in asthma and COPD phenotypes and endotypes for biomarker discovery and improved understanding of disease entities. *J Proteomics* 2011; 75: 192–201.
- 74 Silva JC, Gorenstein MV, Li GZ, *et al.* Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* 2006; 5: 144–156.
- 75 Gharib SA, Nguyen EV, Lai Y, *et al.* Induced sputum proteome in healthy subjects and asthmatic patients. *J Allergy Clin Immunol* 2011; 128: 1176–1184.
- 76 Bai Y, Galetskiy D, Damoc E, *et al.* High resolution mass spectrometric alveolar proteomics: identification of surfactant protein SP-A and SP-D modifications in proteinosis and cystic fibrosis patients. *Proteomics* 2004; 4: 2300–2309.
- 77 Fu Q, Schoenhoff FS, Savage WJ, *et al.* Multiplex assays for biomarker research and clinical application: Translational science coming of age. *Proteomics Clin Appl* 2010; 4: 271–284.
- 78 Schlatter DM, Dazard JE, Ewing RM, *et al.* Human biomarker discovery and predictive models for disease progression for idiopathic pneumonia syndrome following allogeneic stem cell transplantation. *Mol Cell Proteomics* 2012; 11: M111.015479.
- 79 Freire SLS, Yang H, Wheeler AR. A practical interface for microfluidics and nanoelectrospray mass spectrometry. *Electrophoresis* 2008; 29: 1836–1843.
- 80 Lundström SL, Balgoma D, Wheelock ÅM, *et al.* Lipid mediator profiling in pulmonary disease. *Curr Pharm Biotechnol* 2011; 12: 1026–1052.
- 81 Wright SM, Hockey PM, Enhorning G, *et al.* Altered airway surfactant phospholipid composition and reduced lung function in asthma. *J Appl Physiol* 2000; 89: 1283–1292.
- 82 Heeley EL, Hohlfeld JM, Krug N, *et al.* Phospholipid molecular species of bronchoalveolar lavage fluid after local allergen challenge in asthma. *Am J Physiol Lung Cell Mol Physiol* 2000; 278: L305–L311.
- 83 Erpenbeck VJ, Krug N, Hohlfeld JM. Therapeutic use of surfactant components in allergic asthma. *Naunyn Schmiedebergs Arch Pharmacol* 2009; 379: 217–224.
- 84 Babu KS, Woodcock DA, Smith SE, *et al.* Inhaled synthetic surfactant abolishes the early allergen-induced response in asthma. *Eur Respir J* 2003; 21: 1046–1049.
- 85 Arbibe L, Koumanov K, Vial D, *et al.* Generation of lyso-phospholipids from surfactant in acute lung injury is mediated by type-II phospholipase A2 and inhibited by a direct surfactant protein A-phospholipase A2 protein interaction. *J Clin Invest* 1998; 102: 1152–1160.
- 86 Furue S, Mikawa K, Nishina K, *et al.* Therapeutic time-window of a group IIA phospholipase A2 inhibitor in rabbit acute lung injury: correlation with lung surfactant protection. *Crit Care Med* 2001; 29: 719–727.
- 87 Hite RD, Seeds MC, Bowton DL, *et al.* Surfactant phospholipid changes after antigen challenge: a role for phosphatidylglycerol in dysfunction. *Am J Physiol Lung Cell Mol Physiol* 2005; 288: L610–L617.
- 88 Numata M, Chu HW, Dakhama A, *et al.* Pulmonary surfactant phosphatidylglycerol inhibits respiratory syncytial virus-induced inflammation and infection. *Proc Natl Acad Sci USA* 2010; 107: 320–325.
- 89 Almstrand AC, Josefson M, Bredberg A, *et al.* TOF-SIMS analysis of exhaled particles from patients with asthma and healthy controls. *Eur Respir J* 2012; 39: 59–66.
- 90 Kim RH, Takabe K, Milstien S, *et al.* Export and functions of sphingosine-1-phosphate. *Biochim Biophys Acta* 2009; 1791: 692–696.
- 91 Ammit AJ, Hastie AT, Edsall LC, *et al.* Sphingosine 1-phosphate modulates human airway smooth muscle cell functions that promote inflammation and airway remodeling in asthma. *FASEB J* 2001; 15: 1212–1214.
- 92 Ryan JJ, Spiegel S. The role of sphingosine-1-phosphate and its receptors in asthma. *Drug News Perspect* 2008; 21: 89–96.
- 93 Breslow DK, Collins SR, Bodenmiller B, *et al.* Orm family proteins mediate sphingolipid homeostasis. *Nature* 2010; 463: 1048–1053.
- 94 Moffatt MF, Kabesch M, Liang L, *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; 448: 470–473.
- 95 Clarke DL, Dakshinamurti S, Larsson AK, *et al.* Lipid metabolites as regulators of airway smooth muscle function. *Pulm Pharmacol Ther* 2009; 22: 426–435.
- 96 Toews ML, Ediger TL, Romberger DJ, *et al.* Lysophosphatidic acid in airway function and disease. *Biochim Biophys Acta* 2002; 1582: 240–250.
- 97 Zhao Y, Tong J, He D, *et al.* Role of lysophosphatidic acid receptor LPA2 in the development of allergic airway inflammation in a murine model of asthma. *Respir Res* 2009; 10: 114.
- 98 Lundström SL, Levänen B, Nording M, *et al.* Asthmatics exhibit altered oxylipin profiles compared to healthy individuals after subway air exposure. *PLoS One* 2011; 6: e23864.
- 99 Lundström SL, Yang J, Källberg HJ, *et al.* Allergic asthmatics show divergent lipid mediator profiles from healthy controls both at baseline and following birch pollen provocation. *PLoS One* 2012; 7: e33780.
- 100 Kupczyk M, Lundstrom S, Dahlen B, *et al.* Lipid mediators in severe asthma. In: Chung KF, Bel EH, Wenzel SE, eds. *Difficult-to-treat Severe Asthma*. *Eur Respir Monogr* 2011; 51: 218–235.
- 101 Chung KF. Evaluation of selective prostaglandin E₂ (PGE₂) receptor agonists as therapeutic agents for the treatment of asthma. *Sci STKE* 2005; 2005: pe47.
- 102 Raud J, Dahlén SE, Sydbom A, *et al.* Enhancement of acute allergic inflammation by indomethacin is reversed by prostaglandin E₂: apparent correlation with *in vivo* modulation of mediator release. *Proc Natl Acad Sci USA* 1988; 85: 2315–2319.
- 103 Dahlén SE, Kumlin M. Monitoring mast cell activation by prostaglandin D₂ *in vivo*. *Thorax* 2004; 59: 453–455.
- 104 Bochenek G, Nagraba K, Nizankowska E, *et al.* A controlled study of 9 α ,11 β -PGF₂ (a prostaglandin D₂ metabolite) in plasma and urine of patients with bronchial asthma and healthy controls after aspirin challenge. *J Allergy Clin Immunol* 2003; 111: 743–749.
- 105 Bochenek G, Nizankowska E, Gielicz A, *et al.* Plasma 9 α ,11 β -PGF₂, a PGD₂ metabolite, as a sensitive marker of mast cell activation by allergen in bronchial asthma. *Thorax* 2004; 59: 459–464.
- 106 Daham K, Song WL, Lawson JA, *et al.* Effects of celecoxib on major prostaglandins in asthma. *Clin Exp Allergy* 2011; 41: 36–45.

- 107 Sladek K, Dworski R, Fitzgerald GA, *et al.* Allergen-stimulated release of thromboxane A₂ and leukotriene E₄ in humans: effect of indomethacin. *Am Rev Respir Dis* 1990; 141: 1441–1445.
- 108 Kumlin M, Dahlén B, Björck T, *et al.* Urinary excretion of leukotriene E₄ and 11-dehydro-thromboxane B₂ in response to bronchial provocations with allergen, aspirin, leukotriene D₄, and histamine in asthmatics. *Am Rev Respir Dis* 1992; 146: 96–103.
- 109 Dixon RAF, Diehl RE, Opas E, *et al.* Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis. *Nature* 1990; 343: 282–284.
- 110 Dahlen SE, Hedqvist P, Hammarstrom S, *et al.* Leukotrienes are potent constrictors of human bronchi. *Nature* 1980; 288: 484–486.
- 111 Hanna CJ, Bach MK, Pare PD, *et al.* Slow-reacting substances (leukotrienes) contract human airway and pulmonary vascular smooth muscle *in vitro*. *Nature* 1981; 290: 343–344.
- 112 Weiss J, Drazen J, Coles N, *et al.* Bronchoconstrictor effects of leukotriene C in humans. *Science* 1982; 216: 196–198.
- 113 Serhan CN, Samuelsson B. Lipoxins: a new series of eicosanoids (biosynthesis, stereochemistry, and biological activities). *Adv Expl Med Biol* 1988; 229: 1–14.
- 114 Levy BD, De Sanctis GT, Devchand PR, *et al.* Multi-pronged inhibition of airway hyper-responsiveness and inflammation by lipoxin A₄. *Nat Med* 2002; 8: 1018–1023.
- 115 Bonnans C, Vachier I, Chavis C, *et al.* Lipoxins are potential endogenous antiinflammatory mediators in asthma. *Am J Respir Crit Care Med* 2002; 165: 1531–1535.
- 116 Haeggström JZ, Funk CD. Lipoxygenase and leukotriene pathways: biochemistry, biology, and roles in disease. *Chem Rev* 2011; 111: 5866–5898.
- 117 Kumlin M, Hamberg M, Granström E, *et al.* 15(S)-Hydroxyeicosatetraenoic acid is the major arachidonic acid metabolite in human bronchi: association with airway epithelium. *Arch Biochem Biophys* 1990; 282: 254–262.
- 118 Chu HW, Balzar S, Westcott JY, *et al.* Expression and activation of 15-lipoxygenase pathway in severe asthma: relationship to eosinophilic phenotype and collagen deposition. *Clin Exp Allergy* 2002; 32: 1558–1565.
- 119 Liu C, Xu D, Liu L, *et al.* 15-Lipoxygenase-1 induces expression and release of chemokines in cultured human lung epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 2009; 297: L196–L203.
- 120 Feltenmark S, Gautam N, Brunnström Å, *et al.* Eoxins are proinflammatory arachidonic acid metabolites produced via the 15-lipoxygenase-1 pathway in human eosinophils and mast cells. *Proc Natl Acad Sci USA* 2008; 105: 680–685.
- 121 Ono E, Mita H, Taniguchi M, *et al.* Concentration of 14,15-leukotriene C₄ (eoxin C₄) in bronchoalveolar lavage fluid. *Clin Exp Allergy* 2009; 39: 1348–1352.
- 122 Sachs-Olsen C, Sanak M, Lang AM, *et al.* Eoxins: a new inflammatory pathway in childhood asthma. *J Allergy Clin Immunol* 2010; 126: 859–867.
- 123 Montuschi P, Currò D, Ragazzoni E, *et al.* Anaphylaxis increases 8-iso-prostaglandin F_{2α} release from guinea-pig lung *in vitro*. *Eur J Pharmacol* 1999; 365: 59–64.
- 124 Dworski R, Murray JJ, Roberts LJ 2nd, *et al.* Allergen-induced synthesis of F₂-isoprostanes in atopic asthmatics. *Am J Respir Crit Care Med* 1999; 160: 1947–1951.
- 125 Cracowski JL, Cracowski C, Bessard G, *et al.* Increased lipid peroxidation in patients with pulmonary hypertension. *Am J Respir Crit Care Med* 2001; 164: 1038–1042.
- 126 Wood LG, Gibson PG, Garg ML. Biomarkers of lipid peroxidation, airway inflammation and asthma. *Euro Respir J* 2003; 21: 177–186.
- 127 Morrow JD, Hill KE, Burk RF, *et al.* A series of prostaglandin F₂-like compounds are produced *in vivo* in humans by a non-cyclooxygenase, free radical-catalyzed mechanism. *Proc Natl Acad Sci USA* 1990; 87: 9383–9387.
- 128 Higashi A, Higashi N, Tsuburai T, *et al.* Involvement of eicosanoids and surfactant protein D in extrinsic allergic alveolitis. *Eur Respir J* 2005; 26: 1069–1073.
- 129 Thomas MC, Mitchell TW, Harman DG, *et al.* Ozone-induced dissociation: elucidation of double bond position within mass-selected lipid ions. *Anal Chem* 2007; 80: 303–311.
- 130 Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genom* 2001; 2: 155–168.
- 131 Oliver SG, Winson MK, Kell DB, *et al.* Systematic functional analysis of the yeast genome. *Trends Biotechnol* 1998; 16: 373–378.
- 132 Snowden S, Dahlén SE, Wheelock CE. Application of metabolomics approaches to the study of respiratory diseases. *Bioanalysis* 2012; 4: 2265–2290.
- 133 Saude EJ, Skappak CD, Regush S, *et al.* Metabolomic profiling of asthma: diagnostic utility of urine nuclear magnetic resonance spectroscopy. *J Allergy Clin Immunol* 2011; 127: 757–764.
- 134 Mattarucchi E, Baraldi E, Guillou C. Metabolomics applied to urine samples in childhood asthma; differentiation between asthma phenotypes and identification of relevant metabolites. *Biomed Chromatogr* 2012; 26: 89–94.
- 135 Ubhi BK, Riley JH, Shaw PA, *et al.* Metabolic profiling detects biomarkers of protein degradation in COPD patients. *Eur Respir J* 2012; 40: 345–355.
- 136 Paige M, Burdick MD, Kim S, *et al.* Pilot analysis of the plasma metabolite profiles associated with emphysematous chronic obstructive pulmonary disease phenotype. *Biochem Biophys Res Commun* 2011; 413: 588–593.
- 137 Montuschi P, Paris D, Melck D, *et al.* NMR spectroscopy metabolomic profiling of exhaled breath condensate in patients with stable and unstable cystic fibrosis. *Thorax* 2012; 67: 222–228.
- 138 Sonntag D, Scandurra F, Friedrich T, *et al.* Targeted metabolomics for bioprocessing. *BMC Proc* 2011; 5: Suppl. 8, P27.
- 139 Pauling L, Robinson A, Teranishi RR, *et al.* Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc Natl Acad Sci USA* 1971; 68: 2374–2376.
- 140 Horvath I, Hunt J, Barnes PJ, *et al.* Exhaled breath condensate: methodological recommendations and unresolved questions. *Eur Respir J* 2005; 26: 523–548.
- 141 Phillips M. Method for the collection and assay of volatile organic compounds in breath. *Anal Biochem* 1997; 247: 272–278.
- 142 D'Amico A, Di Natale C, Paolesse R, *et al.* Olfactory systems for medical applications. *Sensor Actuat B: Chemical* 2008; 130: 458–465.
- 143 Röck F, Barsan N, Weimar U. Electronic nose: current status and future trends. *Chem Rev* 2008; 108: 705–713.

- 144 Broadhurst D, Kell D. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* 2006; 2: 171–196.
- 145 Sung J, Wang Y, Chandrasekaran S, *et al.* Molecular signatures from omics data: from chaos to consensus. *Biotechnol J* 2012; 7: 946–957.
- 146 Scott SM, James D, Ali Z. Data analysis for electronic nose systems. *Microchim Acta* 2006; 156: 183–207.
- 147 Horváth I, Lázár Z, Gyulai N, *et al.* Exhaled biomarkers in lung cancer. *Eur Respir J* 2009; 34: 261–275.
- 148 de Laurentiis G, Paris D, Melck D, *et al.* Metabonomic analysis of exhaled breath condensate in adults by nuclear magnetic resonance spectroscopy. *Eur Respir J* 2008; 32: 1175–1183.
- 149 Phillips M, Altorki N, Austin JH, *et al.* Detection of lung cancer using weighted digital analysis of breath biomarkers. *Clin Chim Acta* 2008; 393: 76–84.
- 150 Dallinga JW, Robroeks CM, van Berkel JJ, *et al.* Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children. *Clin Exp Allergy* 2010; 40: 68–76.
- 151 Montuschi P, Santonico M, Mondino C, *et al.* Diagnostic performance of an electronic nose, fractional exhaled nitric oxide, and lung function testing in asthma. *Chest* 2010; 137: 790–796.
- 152 Bennett L, Ciaffoni L, Denzer W, *et al.* A chemometric study on human breath mass spectra for biomarker identification in cystic fibrosis. *J Breath Res* 2009; 3: 046002.
- 153 Van Berkel JJ, Dallinga JW, Möller GM, *et al.* A profile of volatile organic compounds in breath discriminates COPD patients from controls. *Respir Med* 2010; 104: 557–563.
- 154 Ibrahim B, Basanta M, Cadden P, *et al.* Non-invasive phenotyping using exhaled volatile organic compounds in asthma. *Thorax* 2011; 66: 804–809.
- 155 Fens N, de Nijs SB, Peters S, *et al.* Exhaled air molecular profiling in relation to inflammatory subtype and activity in COPD. *Eur Respir J* 2011; 38: 1301–1309.
- 156 Dragonieri S, Schot R, Mertens BJ, *et al.* An electronic nose in the discrimination of patients with asthma and controls. *J Allergy Clin Immunol* 2007; 120: 856–862.
- 157 Fens N, Zwinderman AH, van der Schee MP, *et al.* Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma. *Am J Respir Crit Care Med* 2009; 180: 1076–1082.
- 158 Fens N, Roldaan AC, van der Schee MP, *et al.* External validation of exhaled breath profiling using an electronic nose in the discrimination of asthma with fixed airways obstruction and chronic obstructive pulmonary disease. *Clin Exp Allergy* 2011; 41: 1371–1378.
- 159 Lazar Z, Fens N, van der Maten J, *et al.* Electronic nose breathprints are independent of acute changes in airway caliber in asthma. *Sensors (Basel)* 2010; 10: 9127–9138.
- 160 Peng G, Hakim M, Broza YY, *et al.* Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *Br J Cancer* 2010; 103: 542–551.
- 161 Shaham O, Carmel L, Harel D. On mappings between electronic noses. *Sensor Actuat B: Chemical* 2005; 106: 76–82.
- 162 Robroeks CM, Rijkers GT, Jöbsis Q, *et al.* Increased cytokines, chemokines and soluble adhesion molecules in exhaled breath condensate of asthmatic children. *Clin Exp Allergy* 2010; 40: 77–84.
- 163 Motta A, Paris D, Melck D, *et al.* Nuclear magnetic resonance-based metabolomics of exhaled breath condensate: methodological aspects. *Eur Respir J* 2012; 39: 498–500.
- 164 Auffray C, Imbeaud S, Roux-Rouquié M, *et al.* From functional genomics to systems biology: concepts and practices. *C R Biol* 2003; 326: 879–892.
- 165 Kohl P, Noble D. Systems biology and the virtual physiological human. *Mol Syst Biol* 2009; 5: 292.
- 166 Westerhoff H, Kolodkin A, Conradie R, *et al.* Systems biology towards life in silico: mathematics of the control of living cells. *J Math Biol* 2009; 58: 7–34.
- 167 Schmelzer K, Fahy E, Subramaniam S, *et al.* The lipid maps initiative in lipidomics. *Method Enzymol* 2007; 432: 171–183.
- 168 Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 2007; 99: 147–157.
- 169 Diez D, Wheelock AM, Goto S, *et al.* The use of network analyses for elucidating mechanisms in cardiovascular disease. *Mol Biosyst* 2010; 6: 289–304.
- 170 Ghosh S, Matsuoka Y, Asai Y, *et al.* Software for systems biology: from tools to integrated platforms. *Nat Rev Genet* 2011; 12: 821–832.
- 171 IPA. www.ingenuity.com/products/ipa
- 172 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.
- 173 Margolin AA, Nemenman I, Basso K, *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006; 7: Suppl. 1, S7.
- 174 Diez D, Goto S, Fahy JV, *et al.* Network analysis identifies a putative role for the PPAR and type 1 interferon pathways in glucocorticoid actions in asthmatics. *BMC Med Genomics* 2012; 5: 27.
- 175 Kanehisa M, Goto S, Furumichi M, *et al.* KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010; 38: D355–D360.
- 176 Novere NL, Hucka M, Mi H, *et al.* The systems biology graphical notation. *Nat Biotech* 2009; 27: 735–741.
- 177 Funahashi A, Matsuoka Y, Jouraku A, *et al.* CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Proc IEEE* 2008; 96: 1254–1265.
- 178 Smoot ME, Ono K, Ruscheinski J, *et al.* Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011; 27: 431–432.
- 179 Krallinger M, Valencia A, Hirschman L. Linking genes to literature: text mining, information extraction, and retrieval applications for biology. *Genome Biol* 2008; 9: Suppl. 2, S8.
- 180 Tsuruoka Y, Miwa M, Hamamoto K, *et al.* Discovering and visualizing indirect associations between biomedical concepts. *Bioinformatics* 2011; 27: i111–i119.
- 181 Kelder T, van Iersel MP, Hanspers K, *et al.* WikiPathways: building research communities on biological pathways. *Nucleic Acids Res* 2012; 40: D1301–D1307.
- 182 PubMed. National Center for Biotechnology Information. www.ncbi.nlm.nih.gov/pubmed
- 183 Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res* 2003; 31: 248–50.

- 184 Matthews L, Gopinath G, Gillespie M, *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* 2009; 37: D619–D622.
- 185 KEGG: Kyoto Encyclopedia of Genes and Genomes. www.genome.jp/kegg/
- 186 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.
- 187 Courtot M, Juty N, Knüpfer C, *et al.* Controlled vocabularies and semantics in systems biology. *Mol Syst Biol* 2011; 7: 543.
- 188 Galaxy. <http://galaxyproject.org> Date last updated: November 6, 2012. Date last accessed: July 1, 2013.
- 189 Berthold MR, Cebron N, Dill F, *et al.* KNIME – the Konstanz information miner: version 2.0 and beyond. *SIGKDD Explorations* 2009; 11: 26–31.
- 190 Flicek P, Amode MR, Barrell D, *et al.* Ensembl 2012. *Nucleic Acids Res* 2012; 40: D84–D90.
- 191 Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Briefing Bioinform* 2012 [In press DOI: 10.1093/bib/bbs038].
- 192 transSMART. www.transmartfoundation.org Date last updated: June 27, 2013. Date last accessed: June 27, 2013.
- 193 Perakslis ED, Van Dam J, Szalma S. How informatics can potentiate precompetitive open-source collaboration to jump-start drug discovery and development. *Clin Pharmacol Ther* 2010; 87: 614–616.
- 194 Informatics for Integrating Biology and the Bedside. A National Center for Biomedical Computing. www.i2b2.org/ Date last accessed: June 27, 2013.
- 195 The R Project for Statistical Computing. www.r-project.org Date last updated: May 16, 2013. Date last accessed: June 27, 2013.
- 196 Maier D, Kalus W, Wolff M, *et al.* Knowledge management for systems biology a general and visually driven framework applied to translational medicine. *BMC Syst Biol* 2011; 5: 38.
- 197 Figueroa RL, Zeng-Treitler Q, Kandula S, *et al.* Predicting sample size required for classification performance. *BMC Med Inform Decis Mak* 2012; 12: 8.
- 198 Burton PR, Hansell AL, Fortier I, *et al.* Size matters: just how big is BIG? Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol* 2009; 38: 263–273.
- 199 Jung SH, Young SS. Power and sample size calculation for microarray studies. *J Biopharm Stat* 2012; 22: 30–42.
- 200 Liu P, Hwang JT. Quick calculation for sample size while controlling false discovery rate with application to microarray analysis. *Bioinformatics* 2007; 23: 739–746.
- 201 Matsui S, Noma H. Estimating effect sizes of differentially expressed genes for power and sample-size assessments in microarray experiments. *Biometrics* 2011; 67: 1225–1235.
- 202 Page GP, Edwards JW, Gadbury GL, *et al.* The PowerAtlas: a power and sample size atlas for microarray experimental design and research. *BMC Bioinformatics* 2006; 7: 84.
- 203 Kato BS, Nicholson G, Neiman M, *et al.* Variance decomposition of protein profiles from antibody arrays using a longitudinal twin model. *Proteome Sci* 2011; 9: 73.
- 204 Nicholson G, Rantalainen M, Maher AD, *et al.* Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol Syst Biol* 2011; 7: 525.
- 205 Gehlenborg N, O'Donoghue S, Baliga NS, *et al.* Visualization of omics data for systems biology. *Nat Methods* 2010; 7: Suppl. 3, S56–S68.
- 206 Chen R, Mias George I, Li-Pook-Than J, *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 2012; 148: 1293–307.
- 207 Turan N, Kalko S, Stincone A, *et al.* A systems biology approach identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease. *PLoS Comput Biol* 2011; 7: e1002129.
- 208 Lusis AJ. A thematic review series: systems biology approaches to metabolic and cardiovascular disorders. *J Lipid Res* 2006; 47: 1887–1890.
- 209 Young D, Stark J, Kirschner D. Systems biology of persistent infection: tuberculosis as a case study. *Nat Rev Micro* 2008; 6: 520–528.
- 210 Hwang D, Lee IY, Yoo H, *et al.* A systems approach to prion disease. *Mol Syst Biol* 2009; 5: 252.
- 211 Casado-Vela J, Cebrián A, Gómez del Pulgar MT, *et al.* Approaches for the study of cancer: towards the integration of genomics, proteomics and metabolomics. *Clin Transl Oncol* 2011; 13: 617–628.
- 212 Chuang HY, Lee E, Liu YT, *et al.* Network-based classification of breast cancer metastasis. *Mol Syst Biol* 2007; 3: 140.
- 213 He JC, Chuang PY, Ma'Ayan A, *et al.* Systems biology of kidney diseases. *Kidney Int* 2012; 81: 22–39.
- 214 Hood L, Balling R, Auffray C. Revolutionizing medicine in the 21st century through systems approaches. *Biotechnol J* 2012; 7: 992–1001.
- 215 U-BIOPRED. www.ubiopred.european-lung-foundation.org Date last updated: June 14, 2013. Date last accessed: June 27, 2013.
- 216 AirPROM. www.airprom.european-lung-foundation.org Date last updated: May 8, 2013. Date last accessed: June 27, 2013.
- 217 MeDALL. Mechanisms of the Development of Allergy. <http://medall-fp7.eu/> Last updated: April 2013. Date last accessed: June 27, 2013.
- 218 Bousquet J, Anto J, Auffray C, *et al.* MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. *Allergy* 2011; 66: 596–604.
- 219 Brightling CE, Gupta S, Gonen S, *et al.* Lung damage and airway remodelling in severe asthma. *Clin Exp Allergy* 2012; 42: 638–649.
- 220 Synergy-COPD project. www.synergy-copd.eu Date last updated: March 12, 2013. Date last accessed: June 27, 2013.
- 221 Institute of Medicine. Evolution of Translational Omics: Lessons Learned and the Path Forward. Washington, The National Academies Press, 2012.