Molecular BioSystems

REVIEW

View Article Online
View Journal

Cite this: DOI: 10.1039/c6mb00476h

Integrative analysis of human omics data using biomolecular networks

Jonathan L. Robinson[ab] and Jens Nielsen*[abc]

Received 16th June 2016,
Accepted 4th August 2016

DOI: 10.1039/c6mb00476h

www.rsc.org/molecularbiosystems

High-throughput '-omics' technologies have given rise to an increasing abundance of genome-scale data detailing human biology at the molecular level. Although these datasets have already made substantial contributions to a more comprehensive understanding of human physiology and diseases, their interpretation becomes increasingly cryptic and nontrivial as they continue to expand in size and complexity. Systems biology networks offer a scaffold upon which omics data can be integrated, facilitating the extraction of new and physiologically relevant information from the data. Two of the most prevalent networks that have been used for such integrative analyses of omics data are genome-scale metabolic models (GEMs) and protein–protein interaction (PPI) networks, both of which have demonstrated success among many different omics and sample types. This integrative approach seeks to unite 'top-down' omics data with 'bottom-up' biological networks in a synergistic fashion that draws on the strengths of both strategies. As the volume and resolution of high-throughput omics data continue to grow, integrative network-based analyses are expected to play an increasingly important role in their interpretation.

Introduction

The development and advancement of high-throughput transcriptomic, metabolomic, proteomic, and other genome-wide '-omic' profiling technologies have revolutionized the exploration of biological systems.[1,2] The highly parallelized nature of these methods enables the rapid quantification of molecular-level phenomena spanning the whole cell or tissue, thereby providing

[a] Department of Biology and Biological Engineering, Chalmers University of Technology, Kemivägen 10, SE412 96 Gothenburg, Sweden. E-mail: nielsenj@chalmers.se
[b] Wallenberg Centre for Protein Research, Chalmers University of Technology, Kemivägen 10, SE412 96 Gothenburg, Sweden
[c] Science for Life Laboratory, Royal Institute of Technology, SE171 21 Solna, Sweden

Jonathan L. Robinson

Jonathan Robinson received a BS in Chemistry and a BS in Chemical and Biological Engineering from Colorado State University, after which he earned a PhD in Chemical and Biological Engineering from Princeton University. He is currently a postdoctoral researcher in the division of Systems and Synthetic Biology at Chalmers University of Technology in Sweden, where he is working to develop a more quantitative understanding of human metabolism and protein secretion in the context of cancer and other diseases.

Jens Nielsen

Jens Nielsen has a PhD in Biochemical Engineering from the Danish Technical University, where he was appointed as full Professor in 1998, and founded the Center for Microbial Biotechnology. In 2008 he was recruited to Chalmers University of Technology, Sweden, where he was founding Head of the Department of Biology and Biological Engineering, and currently directs a research group of over 50 people. Jens Nielsen has published more than 550 papers that have received over 35 000 citations, and is a member of the National Academy of Engineering in the USA, the Royal Swedish Academy of Sciences, and the Royal Danish Academy of Science and Letters.

a holistic description of a system, in contrast to the reductionist view that is characteristic of more traditional molecular-scale measurements.[3,4] An increased breadth and depth in the information generated by these approaches is helping to elucidate the human genotype–phenotype relationship, an extensive understanding of which is critical to drive the paradigm shift in healthcare toward personalized medicine that has been promised with the advancement of omics technologies.[5–9] Interestingly, the time- and cost-efficiency of generating high-throughput omics datasets have progressed to such an extent that the subsequent processing and analysis of the data are more frequently becoming the rate-limiting step.[1,10,11] It is therefore critical to focus efforts on the development of computational methods to accurately and efficiently interpret increasingly large and complex omics datasets.[10,12]

An intuitive approach when faced with large datasets is the use of statistical methods, such as clustering, enrichment, or correlation analyses, as they can be useful in reducing the dimensionality of the data to identify patterns or anomalous behavior, which are often features of biological interest.[13,14] While many of these techniques have demonstrated success in the extraction of biologically relevant information from omics datasets, as well as the integration of multiple data types, they are inherently naive to the underlying biology of the system from which the data originated. This tends to yield a higher frequency of false-positives since the conclusions are drawn in the absence of known biological constraints, and they do not provide a mechanistic description (*i.e.*, an explanation of how the individual components and processes of a system interact to give rise to biological phenomena) of the results obtained.[15–18] In order to link the data to the underlying biology and exploit the existing knowledgebase, omics data can advantageously be integrated with biological networks.

Systems biology networks represent the physical and/or chemical relationships between the individual components that comprise a biological system.[19,20] These networks can be represented in a mathematical form and provide a scaffold upon which genome-wide omics data can be mapped, enhancing the data with the connectivity information encoded within the network architecture (Fig. 1).[18,21]

Two of the most common and successful network types that have been used for analyzing human omics datasets are genome-scale metabolic models (GEMs) and protein–protein interaction (PPI) networks. Additional networks, such as signaling networks and gene–protein or gene–gene interaction networks, have also been shown to aid in the interpretation of omics data,[22,23] but are beyond the scope of this review. GEMs are a mathematical representation of all known biochemical reactions and their associated enzymes and encoding genes that comprise the metabolic functionality of a cell.[24,25] The combination of gene, enzyme, metabolite, and reaction information allows for the mapping and integration of diverse omics datatypes with GEMs.[5,17,21] Furthermore, the extensive knowledgebase upon which GEMs are constructed, in addition to their highly curated nature, gives them an advantage of improved accuracy and a stronger connection to the governing biological phenomena

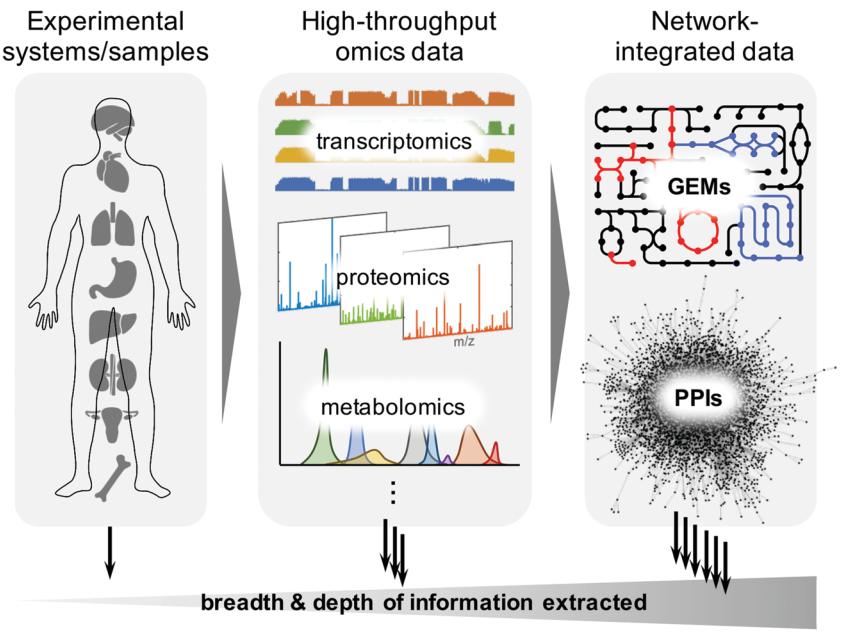

Fig. 1 The integration of omics data with biological networks enriches the information obtained. Moving from left to right, the panels represent an increase in the breadth and depth of information that can be extracted from a dataset. Left panel: Direct, targeted measurements of the human system typically provide low-dimensional data with very specific and narrow information content. Center panel: High-throughput omics profiling methods enable the acquisition of extensive quantities of molecular-level data in a rapid and increasingly cost-effective manner, but are inherently naive to the underlying biological mechanisms. Right panel: Systems biology networks such as GEMs and PPIs can be used to interpret omics datasets in the context of the governing metabolic processes and molecular interactions, enriching the information that is extracted, and helping to bridge the gap from genotype to phenotype.

compared to other network types.[17,26,27] PPI networks are experimentally generated maps detailing the interactions between different proteins in a system, and serve as an excellent complement to GEMs because they include many interactions that could not be inferred based on the existing knowledge such as protein class or function.[28,29] Although PPI networks can possess a number of false-positive interactions, and the biological meaning of an interaction can be ambiguous, these networks are valuable because they are experimentally generated, and are not limited to the scope of metabolism.[30,31] Furthermore, since experimental constraints often limit the coverage of some omics measurements (*e.g.*, proteomics), the interaction information encoded within PPI maps can be used to enable the analysis of proteins that were experimentally inaccessible.[32] We review here the latest advancements in human omics profiling, as well as the construction of two systems biology network types, GEMs and PPIs, and how they can be leveraged to perform integrative analyses of human omics data. The accumulating body of high-throughput omics data requires the use of a systematic framework to interpret the data in the context of biomolecular networks, thus enabling the extraction of novel and biologically relevant information.

Human omics profiling

The continued progression of high-throughput, genome-wide analytical techniques in the past decade has transformed the manner in which biological systems can be studied.[6,33] Advancements in techniques such as genome and mRNA sequencing,[34,35] protein mass spectrometry and immunohistochemistry,[36,37] and quantitative metabolomics[38] have improved the rate and quality at which biological data can be collected, and are available at a decreasing cost. The increased efficiency and accessibility of omics data are driving a paradigm shift in modern medicine from symptom-centric evaluation to personalized healthcare, where diagnoses and treatments are informed by the unique omics profile of each patient.[6,39] Before such a goal can be considered feasible, a more comprehensive understanding of the molecular components that comprise human biological systems, and how they behave as part of a highly interconnected and dynamic network, is required.

Since the completion of the human genome project in 2003, there have been an increasing number of concerted efforts to elucidate the genotype–phenotype connection, and quantify different layers of the involved biological networks using high-throughput omics approaches.[40–42] Although a great deal of information can be obtained from genomic sequences, they do not reveal which genes are expressed in a specific context (*e.g.*, tissue type, disease state, environment) or the dynamics of those expression changes. Over a decade ago, studies began to profile the transcriptomic response in different tissues,[43–45] and more recent efforts, such as the Human Protein Atlas (HPA),[46] Genotype-Tissue Expression (GTEx) project,[47] and functional annotation of the mammalian genome (FANTOM),[48] have collectively quantified the mRNA landscape in all major human tissues and organs.[49] The HPA database (proteinatlas.org) contains RNA sequencing (RNA-Seq) data spanning 32 different human tissues and organs, as well as 45 cell lines, in addition to their corresponding proteomic profiles.[46] The GTEx portal (gtexportal.org) offers RNA-Seq data for 8555 samples collected from 544 donors, covering 53 tissues.[47] Using cap analysis gene expression (CAGE),[50] transcriptomes were quantified for 573 human primary cell samples, 250 cancer cell lines, and 152 post-mortem tissue samples, and are available on the FANTOM online resource (fantom.gsc.riken.jp).[48] These and other human transcriptome profiling efforts have been reviewed in greater detail elsewhere.[49]

Quantification of the human proteome has also been an ongoing project, for which profiling techniques such as immunohistochemistry[46,51,52] and mass spectrometry[53,54] have been employed to characterize the tissue- and organ-specific proteome for most parts of the body. Many additional datatypes, such as metabolomics,[55,56] lipidomics,[57,58] and interactomics,[28] are rapidly advancing in their quality, depth, and coverage of human physiology. However, extracting useful information from these datasets and connecting it to the underlying molecular biology is not straightforward, as there are virtually infinite means by which the data can be analyzed and interpreted, but very few will yield novel biological insight.

The data-rich nature of the human omics field lends itself naturally to inferential statistical analyses, such as hierarchical clustering, principal component analysis (PCA), multiple factor analysis, machine learning, canonical correlation analysis, gene set analysis, and various combinations of these approaches.[12,14,59,60] These methods offer the advantage of requiring little or no prior knowledge of the biological system, which can in itself provide the benefit of an objective, unbiased approach.[61] For example, Fehrmann *et al.* conducted an extensive re-analysis of over 77 000 gene expression datasets spanning human, rat, and mouse, and across many tissue types, disease states, and therapeutic perturbations, upon which PCA was administered to identify a set of principal components (PCs) that captured the different expression-regulating factors.[62] In addition to enabling the inference of unknown gene functions based on similar patterns as known genes, the PCs were used to correct expression data for non-genetic differences, revealing a strong correlation of residual expression with copy number variation, perturbations of which have been implicated in a number of cancers.[63] Although statistical-based approaches constitute a powerful device for identifying patterns in the data, their disconnection from the underlying biochemistry can often lead to a relatively large frequency of false positives and overlooked phenomena (false negatives). To more efficiently extract information from these large omics datasets, the data must be reconnected with the biochemical network(s) governing the system upon which the measurement was conducted.

Genome-scale metabolic models

One type of biomolecular network that is commonly used to facilitate the interpretation and extraction of information from

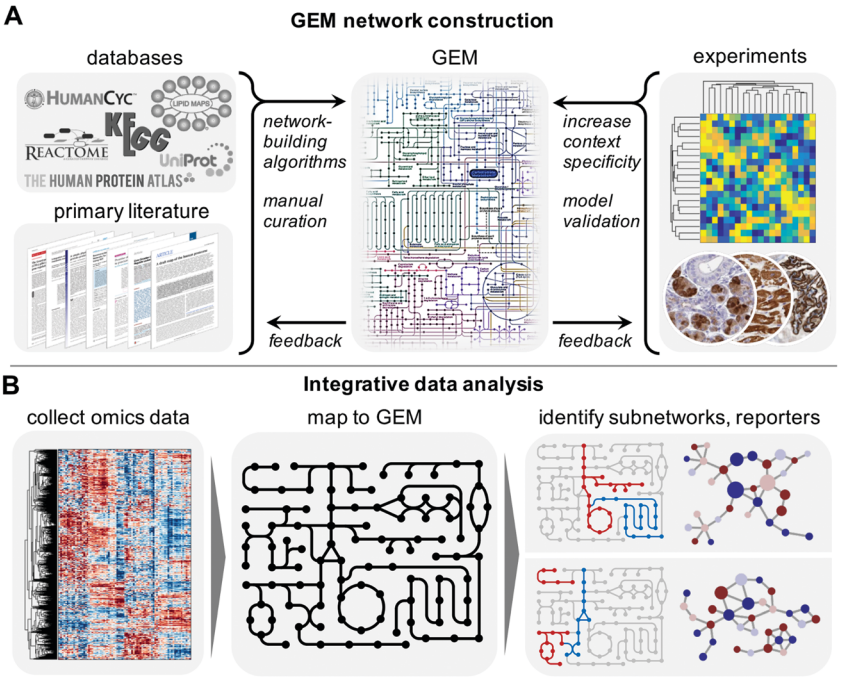

Fig. 2 Human GEM construction and integration with omics data. (A) GEMs are reconstructed from online databases and primary literature detailing the known or predicted reactions that comprise the metabolic network. Additional experimental data can be obtained to further refine and validate the model architecture. The construction is often an iterative process, where feedback from model analyses informs further experimentation and contributes to the existing knowledgebase. (B) Data obtained from omics methods, such as transcriptomic data, can be mapped onto a GEM *via* the encoded gene–reaction associations. This projection yields subnetworks or reporter metabolites/reactions that demonstrate significantly coordinated behavior of network components in the expression data.

omics datasets is the GEM. GEMs are mathematical representations of the complete network of biochemical reactions that can occur within a cell, and are assembled from the existing knowledgebase of the metabolites, reaction stoichiometry, compartmentalization, and gene–enzyme associations that define the metabolic network of that particular cell type or species (Fig. 2A).[17,64,65]

Since the first metabolic reconstruction in 1999 of *Haemophilus influenza*,[66] an increasingly large number of GEMs have been generated for a diverse range of species spanning bacteria, viruses, plants, and humans.[27] Human GEMs differ from those of unicellular organisms in that they are often constructed in a generic form to include the metabolites and reactions present among all organs and tissues, which cannot be represented by a single model type due to different gene expression profiles.[67] The initial generic human GEMs were Recon1[68] and the Edinburgh Human Metabolic Network (EHMN),[69] assembled from biological data collected over many decades from the primary literature, online databases, and high-quality genome annotations. Currently, the latest and most comprehensive generic human GEMs are the Human Metabolic Reaction 2.0 (HMR2)[26] and Recon2,[70] which are the result of updating previous versions with the current human biochemical knowledgebase (from the literature and databases[58,71–73]), and from the merger with many other, often more context-specific, models of human metabolism.[21,74–77]

The applications of GEMs are as diverse as the species for which they have been constructed, ranging from metabolic engineering for biofuel production to the identification of knowledge gaps in a particular reaction pathway.[17,78–80] Approaches such as constraint-based modeling, where mathematical bounds or relationships are enforced upon the network based on experimental data (such as nutrient uptake rates), are used to reduce the solution space to a more physiologically relevant subregion.[81] This technique can be used for simulation approaches such as flux balance analysis (FBA), where the depletion and generation of each metabolite are balanced to provide a set of algebraic equations constraining the fluxes. An optimization framework is then used to arrive at an even further constrained or unique solution of reaction fluxes through the network.[82]

Although simulations using GEMs have been beneficial in guiding metabolic engineering efforts of bacteria and yeast,[17,78,79] human GEMs possess substantial value in their network architecture alone. The construction of GEMs represents a 'bottom-up' approach in systems biology, where detailed knowledge of individual biological components and their interactions is assembled to form a larger, consolidated system.[5,15,83] This is in contrast to 'top-down' approaches, where a system is analyzed as a whole, often using high-throughput omics techniques to quantify how different conditions or perturbations impact different subsystems or individual components of the network.[5,15,83] An integrative approach seeks to bridge the top-down (omics datasets) and bottom-up (GEMs) strategies in a cooperative manner that exploits the strengths of both. In this way, the data are interpreted through a mechanistic framework that is firmly rooted to the underlying biochemistry.[21,84]

The structure of a GEM is comprised of metabolic, enzymatic, and genetic layers, as well as their connectivity, enabling integration with diverse omics datatypes.[17] A common integrative approach is the use of context-specific omics data (*i.e.*, originating from a specific tissue type or disease state) to specialize a generic human GEM, where reactions are removed or constrained to generate a context-specific GEM.[74,85–87] This procedure is highly flexible, where aspects such as the type of omics data and algorithm used to contextualize the GEM can be adapted according to the focus or applications of the study. A number of algorithms have been developed for this approach, including a transcriptome-based method by Åkesson *et al.*,[88] Gene Inactivity Moderated by Metabolism and Expression (GIMME),[85] Integrative Metabolic Analysis Tool (iMAT),[86] Metabolic Adjustment by Differential Expression (MADE),[89] Probabilistic Regulation of Metabolism (PROM),[90] and Integrative Network Inference for Tissues (INIT).[74] These and other approaches have been reviewed in greater detail elsewhere.[67,91–94]

Another means by which GEMs have been employed for integrative analysis, which is often done in combination with the context-specific reconstruction approach described above, is to map the omics data onto a GEM to enrich the data with the topological information encoded in the model architecture (*e.g.*, metabolites participating in the same reaction, or enzymes catalyzing reactions that share a common metabolite) (Fig. 2B).[5,21,95] In this way, patterns emerge that may otherwise be too diffuse or statistically unremarkable to detect in the raw (unmapped) omics data, yielding novel network insights such as transcriptional or metabolic regulatory circuits.[18,96] Furthermore, this integrative approach enables the use of incomplete datasets that do not exhibit an exhaustive coverage at the genome scale, where the connectivity information built into the GEM structure enables one to infer the behavior of neighboring components for which direct quantification was unavailable.[96]

An extensive body of work has demonstrated success in extracting new and valuable information from human omics data through a GEM-based integrative analysis.[21,97–104] For example, Folger *et al.* integrated microarray data from 59 cancer cell lines with the Recon1 human metabolic network reconstruction to generate a generic metabolic model of human cancer cells, and a growth reaction was added to enable the prediction of gene knockdowns that would perturb cancer cell proliferation.[105] Model-predicted genes were validated with experimental gene silencing data and with a set of human cancer genes for which mutations had been implicated in oncogenesis. The use of an integrated approach demonstrated improved performance over the independent use of the transcriptomic data or the Recon1 model alone. In another study, Shlomi *et al.* investigated the tissue specificity of human metabolism by mapping transcriptomic and proteomic data from ten different tissues onto Recon1, using an optimization framework that maximized the agreement between predicted flux distributions and the corresponding tissue-specific enzyme expression levels.[86] Their use of a biological network enabled the prediction of post-transcriptional regulatory effects based on differences between calculated reaction fluxes and enzyme expression levels, providing new information that could not be obtained from standard gene expression analysis approaches. Furthermore, the prediction of gene tissue-specificity using the network-based approach identified a larger number (three-fold greater) of gene–tissue associations, and with greater accuracy than the standard gene set enrichment analyses.[86]

Agren and colleagues integrated expression data from the HPA with the HMR human GEM using their INIT algorithm, generating context-specific GEMs for 69 different healthy cell types and 16 cancer types.[74] These GEMs were validated through comparison with measured protein abundances in different tissue types reported in the BRENDA database,[106] and through comparison of the liver-specific model with a previously-developed highly-curated model of human hepatocytes (HepatoNet1).[75] A comparison of different context-specific GEMs enabled the prediction of metabolic functions that were specific to cancer physiology, many of which were associated with enzymes that were previously known and used as drug targets. A later work extended and built upon this approach, generating personalized GEMs for 6 hepatocellular carcinoma patients from proteomic data,[107] and 11 cancer cell-line specific GEMs from RNA-Seq data.[108] These cancer-specific GEMs were then compared with 83 healthy cell-type GEMs to predict antimetabolites (drugs that are structurally similar to metabolites) that would inhibit cancer growth without perturbing normal cellular function. Many of the predicted antimetabolites overlapped with existing cancer treatments, and novel predictions were experimentally validated by demonstrating attenuated growth of cancer cell lines.[107,108] Since their antimetabolite prediction approach relied on the network connectivity information encoded within the GEM, it was not possible to extract this type of information using traditional statistical-based methods alone, further illustrating the added benefit of employing a network-based approach in the analysis of omics data.

Recently, Gatto *et al.*[104] performed an integrated analysis of 481 clear cell renal cell carcinoma (ccRCC) RNA-Seq datasets from The Cancer Genome Atlas (TCGA) with HMR2 to investigate the cancer-specific metabolic program that they previously identified[99] to be unique from that of many other cancer types. The investigation revealed a coordinated perturbation to glycosaminoglycan biosynthesis associated with metastatic ccRCC which was not only confirmed at the protein level using immunohistochemical staining of the ccRCC tissue, but was also sufficiently detectable in urine and plasma to constitute a promising new biomarker for non-invasive diagnosis of the disease.[104] These and many other studies, which have been reviewed at length elsewhere,[15,17,49,65,109–112] underscore the utility and flexibility of employing GEMs in integrative analyses of high-throughput human omics data.

Protein–protein interaction networks

Advancements in high-throughput genomic and transcriptomic technologies have greatly accelerated our understanding of biological systems at the gene-level, though it is ultimately the encoded proteins that execute many of the biological functions

that drive cellular behavior.[113,114] Furthermore, the interactions between proteins govern critical functions such as enzyme complex formation, cellular signaling, allosteric regulation, and post-translational processing of proteins (*e.g.*, folding, trafficking, degradation), and the disruption of such interactions has been implicated in diseases ranging from cancer to asthma.[31,40,115] Given that the transcript levels do not necessarily correlate with the abundance of their encoded proteins,[116] the direct quantification of the proteome and/or its interactions (interactome) serves as an important complement to gene-centric data in systems biology studies.

Many assays have been developed to quantify PPIs, and can differ in aspects such as the type of information they yield, how well they scale to a larger library size, and their cost. One of the most common techniques used to detect interactions between proteins is the yeast two-hybrid (Y2H) assay, where the two proteins are fused to different fragments of a transcription factor, and their interaction initiates the expression of a downstream reporter gene (Fig. 3A).[117] It offers the advantage of scaling to high-throughput methods, but is restricted to the subset of proteins that are able to enter the yeast nucleus.[118] The membrane Y2H (MYTH) assay[119] addresses the problem of restricted protein location by enabling the detection of membrane-bound protein interactions through fusion to split-ubiquitin,[120] though it prevents the detection of soluble proteins. These and other protein interaction assay techniques, such as mammalian protein–protein interaction trap (MAPPIT),[121] affinity purification-mass spectrometry (AP-MS),[122] and fluorescence resonance energy transfer (FRET),[123] are reviewed in greater detail elsewhere.[30,118]

There have been several high-throughput studies seeking to experimentally map the human PPI network, which was recently estimated to consist of approximately 130 000 binary interactions.[124] Mapping efforts began in 2005 when Stelzl *et al.*[125] and Rual *et al.*[126] used a Y2H-based approach to quantify approximately 3200 and 2800 binary PPIs, respectively. A later study by Ewing and colleagues utilized a mass spectrometry-based approach to conduct a similar analysis of human PPIs, resulting in ~6500 interactions between ~2200 proteins.[127] More recently, Rolland *et al.* sought to generate a high-quality binary interaction map of the human proteome using a systematic approach to enable more complete coverage of the possible interactome landscape.[28] The resulting human PPI map, which is the most comprehensive to date, included approximately 14 000 interactions, which were more broadly distributed among the proteome, spanning regions that had been sparsely covered by previous studies.

PPI networks serve as an excellent complement to GEMs because they are experimentally derived, and often possess novel interactions that cannot be ascertained from reaction–metabolite or gene–enzyme relationships alone. Generating the network from experimental data is also advantageous in that it is more objective, avoiding biases that could be introduced by the current knowledge of protein interactions.[28] Moreover,

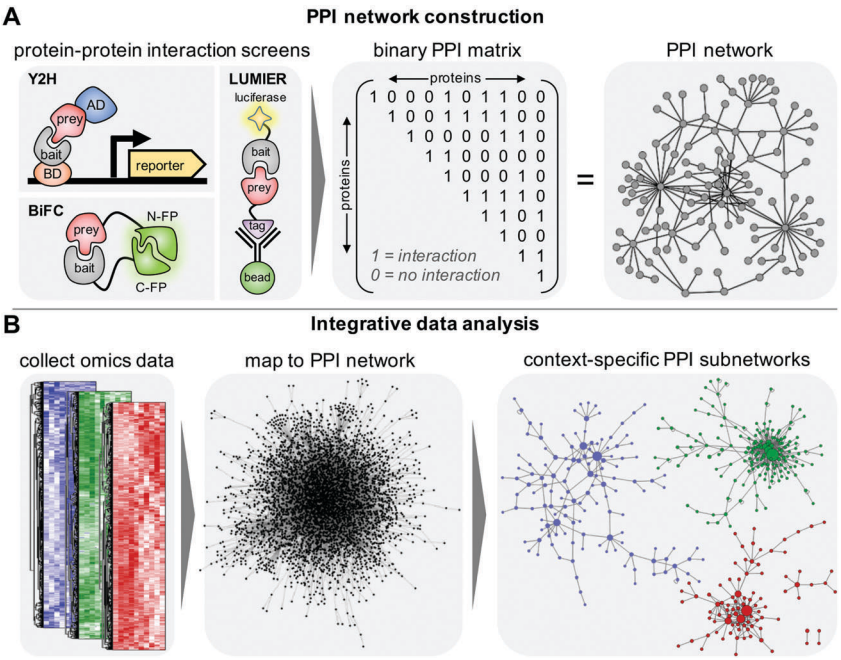

Fig. 3 Human PPI network construction and integration with omics data. (A) PPI networks are generated experimentally, where many different binary interaction screens can be used, such as Yeast Two-Hybrid (Y2H), Bimolecular Fluorescence Complementation (BiFC), and Luminescence-based Mammalian Interactome Mapping (LUMIER). BD is the DNA-binding domain, AD is the activation domain, and N-FP and C-FP are the N- and C-terminal fragments of a fluorescent protein. The results are arranged in a binary matrix indicating which pairs of proteins were found to interact, and can be represented graphically where proteins are nodes, and edges connect the pairs of interacting proteins. (B) High-throughput data, such as quantitative proteomics, can be integrated with PPIs to generate context-specific subnetworks that highlight groups of proteins or interactions that are significantly associated with the particular condition or perturbation explored.

PPI networks can, in principle, encompass all proteins comprising a biological system, whereas GEMs are generally limited to those involved in metabolic functions. Much like GEMs, PPI maps can serve as a scaffold upon which high-throughput omics data can be mapped for an integrative analysis, thus enriching the information that can be extracted from the data (Fig. 3B).[32,128–132] In an effort to determine mechanisms contributing to breast cancer progression and metastasis, Chuang *et al.*[128] mapped gene expression profiles obtained from primary breast tumors to a human PPI network (containing PPI information from both Stelzl *et al.*[125] and Rual *et al.*[126]), where transcriptomes were classified according to whether metastasis was detected in a follow-up evaluation of each patient. Subnetworks exhibiting large metastatic–nonmetastatic differences were identified as markers, and found to be enriched for functions such as apoptosis, tissue remodeling, and cell proliferation—processes implicated in cancer progression. Subnetwork markers exhibited greater predictive accuracy and reproducibility across different datasets compared to individual gene markers, demonstrating the value added by using a network-based approach. Moreover, many well-established breast cancer markers that were not found to be significantly differentially expressed when using traditional gene expression analyses were identified as members of significant subnetwork markers.

A study by Nibbe *et al.* involved the integration of proteomic data obtained from normal and colorectal cancer (CRC) tumor tissue biopsies with a human PPI network, where significantly involved proteins were used to "seed" the network and predict important sub-networks based on their connectivity and proximity to neighboring proteins.[32] The authors went further to map transcriptomic CRC data onto the proteome-informed PPI subnetworks to identify candidates that exhibited significant and synergistic differences between normal and tumor samples. Several of the candidates were experimentally validated by Western blot from tissue biopsies not included in the original screening. Recently, Greene *et al.* generated 144 tissue- and cell type-specific functional maps by integrating a massive collection of human omics data (including, for example, 980 gene expression datasets) with PPI networks using a Bayesian approach, to facilitate the understanding and prediction of tissue- and disease-specific gene functions.[129] To validate the approach, the authors experimentally assessed their predicted connection of 20 blood vessel cell network neighbors with *IL1B* by measuring gene expression in smooth muscle cells following IL-1β treatment, where 18 (90%) were confirmed to be among the top 500 genes exhibiting increased expression.

Balbin *et al.* used a PPI network in an integrative analysis of transcriptomic, proteomic, and phosphoproteomic data obtained from non-small cell lung cancer (NSCLC) cell lines.[130] Through the combination of a network pathway enrichment approach (Signaling Pathway Impact Analysis algorithm; SPIA[133]), with PPI data obtained from the STRING database,[134] the authors were able to investigate the changes in protein and transcript abundance specifically connected to KRAS. Three of the four proteins that they experimentally validated had not been explored previously in KRAS-dependent lung cancer.[130] In contrast, a naive analysis of the same data using traditional expression analyses missed 30 of the 115 candidate proteins identified using the integrated approach, and offered no insight into their specific role in lung cancer or their connection to known oncogenes.[130] Collectively, these studies demonstrate the versatility and advantages of using PPI networks in the integrative analyses of human omics data, enabling the extraction of new and biologically-relevant information that is otherwise inaccessible in the raw, unmapped datasets.

Conclusions

The quality, coverage, and rate at which human omics datasets are being generated are continuing to increase at a striking pace, and offer great potential in advancing the prevention, diagnosis, treatment, and understanding of human diseases.[1,8,15] Integrative analysis of these high-throughput datasets with biomolecular networks constitutes a promising approach that facilitates more efficient extraction of physiologically relevant information.[5,15,84,111] Systems biology networks such as GEMs and PPIs offer unique and complementary advantages in interpreting human omics data in the context of the underlying biology. Although the scope of using human GEMs is generally limited to the metabolic network they represent, they are highly curated due to the wealth of information available on metabolism, and provide a link between enzymes, metabolites, reactions, and genes.[16,111] PPI networks are prone to contain false-positives and can be ambiguous regarding the mechanics or meaning of an interaction, but they are experimentally generated, require no prior knowledge of the system, and can potentially span the entire proteome.[28,30,135]

The capacity for biomolecular networks to systematically process and interpret large datasets is critical to the realization of personalized medicine.[8,15] Given the heterogeneity and complexity with which diseases perturb human physiology, traditional expression analyses of omics data are insufficient to decipher the underlying cause(s) of system malfunction.[15,40] Network-based approaches enable a holistic analysis that accounts for the interactions among system components that comprise (or obscure) the mechanism of disease progression.[111,136] Furthermore, the connectivity information within GEMs and PPIs facilitates the prediction of potential targets for novel treatment options, or that of biomarkers for early disease detection.[21,136]

Beyond GEMs and PPIs, the construction and application of other networks, such as disorder–gene,[137] gene–gene,[23] and protein–DNA[138] networks have demonstrated success in extracting mechanistic information from biological systems, and can each offer a unique context to an omics dataset. Furthermore, the integration of high-throughput data with a combination of multiple different biological network types has the potential to increase the dimensionality and breadth of information that can be extracted, though the approach is nontrivial and will require further development. Although single omics datasets possess a wealth of information about a system, it is becoming increasingly clear that multi-omics approaches, wherein multiple types

of high-throughput omics assays are performed on the same system, are necessary to sufficiently capture the high dimensionality and complexity of biological systems.[4,130,139] Biological networks can serve as a core scaffold to facilitate the integration of diverse omics data types, providing a central structure upon which the different data are layered.[10,130,140] It is expected that GEMs, PPIs, and other network types will play an increasingly central role in the interpretation of high-throughput data as human biological profiling studies continue to shift toward more multi-omics-based strategies.

Acknowledgements

We would like to acknowledge funding from the Knut and Alice Wallenberg Foundation.

References

1 A. Alyass, M. Turcotte and D. Meyre, *BMC Med. Genomics*, 2015, **8**, 33.
2 S. H. Payne, *Trends Biochem. Sci.*, 2015, **40**, 1–3.
3 F. C. Fang and A. Casadevall, *Infect. Immun.*, 2011, **79**, 1401–1404.
4 W. W. Zhang, F. Li and L. Nie, *Microbiology*, 2010, **156**, 287–301.
5 M. L. Mo and B. Ø. Palsson, *Trends Biotechnol.*, 2009, **27**, 37–44.
6 R. Chen and M. Snyder, *Wiley Interdiscip. Rev.: Syst. Biol. Med.*, 2013, **5**, 73–82.
7 L. Hood, J. R. Heath, M. E. Phelps and B. Y. Lin, *Science*, 2004, **306**, 640–643.
8 Q. Tian, N. D. Price and L. Hood, *J. Intern. Med.*, 2012, **271**, 111–121.
9 C. Sheridan, *Nat. Biotechnol.*, 2015, **33**, 887–889.
10 B. Ø. Palsson and K. Zengler, *Nat. Chem. Biol.*, 2010, **6**, 787–789.
11 E. R. Mardis, *Genome Med.*, 2010, **2**, 84.
12 B. Berger, J. Peng and M. Singh, *Nat. Rev. Genet.*, 2013, **14**, 333–346.
13 C. Meng, O. A. Zeleznik, G. G. Thallinger, B. Kuster, A. M. Gholami and A. C. Culhane, *Brief. Bioinform.*, 2016, **17**, 628–641.
14 S. Pineda, F. X. Real, M. Kogevinas, A. Carrato, S. J. Chanock, N. Malats and K. Van Steen, *PLoS Genet.*, 2015, **11**, e1005689.
15 A. Mardinoglu and J. Nielsen, *J. Intern. Med.*, 2012, **271**, 142–154.
16 M. L. Mo, N. Jamshidi and B. Ø. Palsson, *Mol. BioSyst.*, 2007, **3**, 598–603.
17 E. J. O'Brien, J. M. Monk and B. Ø. Palsson, *Cell*, 2015, **161**, 971–987.
18 K. R. Patil and J. Nielsen, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 2685–2689.
19 A. L. Barabasi and Z. N. Oltvai, *Nat. Rev. Genet.*, 2004, **5**, U101–U115.
20 X. W. Zhu, M. Gerstein and M. Snyder, *Genes Dev.*, 2007, **21**, 1010–1024.
21 A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, I. Nookaew, P. Jacobson, A. J. Walley, P. Froguel, L. M. Carlsson, M. Uhlen and J. Nielsen, *Mol. Syst. Biol.*, 2013, **9**, 649.
22 K. Yugi, H. Kubota, Y. Toyoshima, R. Noguchi, K. Kawata, Y. Komori, S. Uda, K. Kunida, Y. Tomizawa, Y. Funato, H. Miki, M. Matsumoto, K. I. Nakayama, K. Kashikura, K. Endo, K. Ikeda, T. Soga and S. Kuroda, *Cell Rep.*, 2014, **8**, 1171–1183.
23 M. Costanzo, A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, C. S. Sevier, H. Ding, J. L. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R. P. St Onge, B. VanderSluis, T. Makhnevych, F. J. Vizeacoumar, S. Alizadeh, S. Bahr, R. L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z. Y. Lin, W. Liang, M. Marback, J. Paw, B. J. San Luis, E. Shuteriqi, A. H. Tong, N. van Dyk, I. M. Wallace, J. A. Whitney, M. T. Weirauch, G. Zhong, H. Zhu, W. A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pal, F. P. Roth, G. Giaever, C. Nislow, O. G. Troyanskaya, H. Bussey, G. D. Bader, A. C. Gingras, Q. D. Morris, P. M. Kim, C. A. Kaiser, C. L. Myers, B. J. Andrews and C. Boone, *Science*, 2010, **327**, 425–431.
24 M. W. Covert, C. H. Schilling, I. Famili, J. S. Edwards, I. I. Goryanin, E. Selkov and B. Ø. Palsson, *Trends Biochem. Sci.*, 2001, **26**, 179–186.
25 J. Förster, I. Famili, P. Fu, B. Ø. Palsson and J. Nielsen, *Genome Res.*, 2003, **13**, 244–253.
26 A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, M. Uhlen and J. Nielsen, *Nat. Commun.*, 2014, **5**, 3083.
27 J. Monk, J. Nogales and B. Ø. Palsson, *Nat. Biotechnol.*, 2014, **32**, 447–452.
28 T. Rolland, M. Tasan, B. Charloteaux, S. J. Pevzner, Q. Zhong, N. Sahni, S. Yi, I. Lemmens, C. Fontanillo, R. Mosca, A. Kamburov, S. D. Ghiassian, X. P. Yang, L. Ghamsari, D. Balcha, B. E. Begg, P. Braun, M. Brehme, M. P. Broly, A. R. Carvunis, D. Convery-Zupan, R. Corominas, J. Coulombe-Huntington, E. Dann, M. Dreze, A. Dricot, C. Y. Fan, E. Franzosa, F. Gebreab, B. J. Gutierrez, M. F. Hardy, M. Jin, S. L. Kang, R. Kiros, G. N. Lin, K. Luck, A. MacWilliams, J. R. Menche, R. R. Murray, A. Palagi, M. M. Poulin, X. Rambout, J. Rasla, P. Reichert, V. Romero, E. Ruyssinck, J. M. Sahalie, A. Scholz, A. A. Shah, A. Sharma, Y. Shen, K. Spirohn, S. Tam, A. O. Tejeda, S. A. Trigg, J. C. Twizere, K. Vega, J. Walsh, M. E. Cusick, Y. Xia, A. L. Barabasi, L. M. Iakoucheva, P. Aloy, J. De Las Rivas, J. Tavernier, M. A. Calderwood, D. E. Hill, T. Hao, F. P. Roth and M. Vidal, *Cell*, 2014, **159**, 1212–1226.
29 M. Vidal, M. E. Cusick and A. L. Barabasi, *Cell*, 2011, **144**, 986–998.
30 J. Snider, M. Kotlyar, P. Saraon, Z. Yao, I. Jurisica and I. Stagljar, *Mol. Syst. Biol.*, 2015, **11**, 848.
31 M. Vidal and S. Fields, *Nat. Methods*, 2014, **11**, 1203–1206.
32 R. K. Nibbe, M. Koyuturk and M. R. Chance, *PLoS Comput. Biol.*, 2010, **6**, e1000639.
33 A. Bordbar and B. Ø. Palsson, *J. Intern. Med.*, 2012, **271**, 131–141.

34 W. W. Soon, M. Hariharan and M. P. Snyder, *Mol. Syst. Biol.*, 2013, **9**, 640.
35 J. Shendure and H. Ji, *Nat. Biotechnol.*, 2008, **26**, 1135–1145.
36 T. Nilsson, M. Mann, R. Aebersold, J. R. Yates, 3rd, A. Bairoch and J. J. Bergeron, *Nat. Methods*, 2010, **7**, 681–685.
37 M. Uhlen, P. Oksvold, C. Algenas, C. Hamsten, L. Fagerberg, D. Klevebring, E. Lundberg, J. Odeberg, F. Ponten, T. Kondo and A. Sivertsson, *Mol. Cell. Proteomics*, 2012, **11**, M111.013458.
38 B. D. Bennett, J. Yuan, E. H. Kimball and J. D. Rabinowitz, *Nat. Protoc.*, 2008, **3**, 1299–1311.
39 J. Nielsen, *J. Intern. Med.*, 2012, **271**, 108–110.
40 A. L. Barabasi, N. Gulbahce and J. Loscalzo, *Nat. Rev. Genet.*, 2011, **12**, 56–68.
41 C. Genomes Project, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean and G. R. Abecasis, *Nature*, 2015, **526**, 68–74.
42 P. Legrain, R. Aebersold, A. Archakov, A. Bairoch, K. Bala, L. Beretta, J. Bergeron, C. H. Borchers, G. L. Corthals, C. E. Costello, E. W. Deutsch, B. Domon, W. Hancock, F. C. He, D. Hochstrasser, G. Marko-Varga, G. H. Salekdeh, S. Sechi, M. Snyder, S. Srivastava, M. Uhlen, C. H. Wu, T. Yamamoto, Y. K. Paik and G. S. Omenn, *Mol. Cell. Proteomics*, 2011, **10**, M111.009993.
43 A. I. Su, T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M. P. Cooke, J. R. Walker and J. B. Hogenesch, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 6062–6067.
44 C. V. Jongeneel, M. Delorenzi, C. Iseli, D. Zhou, C. D. Haudenschild, I. Khrebtukova, D. Kuznetsov, B. J. Stevenson, R. L. Strausberg, A. J. Simpson and T. J. Vasicek, *Genome Res.*, 2005, **15**, 1007–1014.
45 L. D. Hillier, G. Lennon, M. Becker, M. F. Bonaldo, B. Chiapelli, S. Chissoe, N. Dietrich, T. DuBuque, A. Favello, W. Gish, M. Hawkins, M. Hultman, T. Kucaba, M. Lacy, M. Le, N. Le, E. Mardis, B. Moore, M. Morris, J. Parsons, C. Prange, L. Rifkin, T. Rohlfing, K. Schellenberg, M. Bento Soares, F. Tan, J. Thierry-Meg, E. Trevaskis, K. Underwood, P. Wohldman, R. Waterston, R. Wilson and M. Marra, *Genome Res.*, 1996, **6**, 807–828.
46 M. Uhlen, L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen and F. Ponten, *Science*, 2015, **347**, 1260419.
47 G. T. Consortium, *Nat. Genet.*, 2013, **45**, 580–585.
48 A. R. R. Forrest, H. Kawaji, M. Rehli, J. K. Baillie, M. J. L. de Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jorgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. A. Semple, Y. Ishizu, R. S. Young, M. Francescatto, I. Alam, D. Albanese, G. M. Altschuler, T. Arakawa, J. A. C. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse, S. Pradhan-Bhatt, J. A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. M. Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H. C. Clevers, E. Dalla, C. A. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drablos, A. S. B. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furuno, J. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. H. Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. P. Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. J. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-sim, R. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. A. D. Morais, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. Van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, T. Nozaki, S. Ogishima, N. Ohkura, H. Ohmiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. D. Prendergast, O. J. L. Rackham, J. A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, H. Satoh, S. Savvi, A. Saxena, C. Schneider, E. A. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. J. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R. K. Swoboda, P. A. C. 't Hoen, M. Tagami, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyoda, T. Toyoda, E. Valen, M. van de Wetering, L. M. van den Berg, R. Verardo, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S. Watanabe, C. A. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S. E. Zabierowski, P. G. Zhang, X. B. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci,

Y. Hayashizaki, F. Consortium and R. P. C. DGT, *Nature*, 2014, **507**, 462–470.
49 M. Uhlen, B. M. Hallstrom, C. Lindskog, A. Mardinoglu, F. Ponten and J. Nielsen, *Mol. Syst. Biol.*, 2016, **12**, 862.
50 M. Kanamori-Katayama, M. Itoh, H. Kawaji, T. Lassmann, S. Katayama, M. Kojima, N. Bertin, A. Kaiho, N. Ninomiya, C. O. Daub, P. Carninci, A. R. Forrest and Y. Hayashizaki, *Genome Res.*, 2011, **21**, 1150–1159.
51 F. Ponten, M. Gry, L. Fagerberg, E. Lundberg, A. Asplund, L. Berglund, P. Oksvold, E. Bjorling, S. Hober, C. Kampf, S. Navani, P. Nilsson, J. Ottosson, A. Persson, H. Wernerus, K. Wester and M. Uhlen, *Mol. Syst. Biol.*, 2009, **5**, 337.
52 L. Fagerberg, B. M. Hallstrom, P. Oksvold, C. Kampf, D. Djureinovic, J. Odeberg, M. Habuka, S. Tahmasebpoor, A. Danielsson, K. Edlund, A. Asplund, E. Sjostedt, E. Lundberg, C. A. Szigyarto, M. Skogs, J. O. Takanen, H. Berling, H. Tegel, J. Mulder, P. Nilsson, J. M. Schwenk, C. Lindskog, F. Danielsson, A. Mardinoglu, A. Sivertsson, K. von Feilitzen, M. Forsberg, M. Zwahlen, I. Olsson, S. Navani, M. Huss, J. Nielsen, F. Ponten and M. Uhlen, *Mol. Cell. Proteomics*, 2014, **13**, 397–406.
53 M. S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T. C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda and A. Pandey, *Nature*, 2014, **509**, 575–581.
54 M. Wilhelm, J. Schlegl, H. Hahne, A. Moghaddas Gholami, M. Lieberenz, M. M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeer, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J. H. Boese, M. Bantscheff, A. Gerstmair, F. Faerber and B. Kuster, *Nature*, 2014, **509**, 582–587.
55 D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. F. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. G. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *Nucleic Acids Res.*, 2013, **41**, D801–D807.
56 S. Bijlsma, L. Bobeldijk, E. R. Verheij, R. Ramaker, S. Kochhar, I. A. Macdonald, B. van Ommen and A. K. Smilde, *Anal. Chem.*, 2006, **78**, 567–574.
57 O. Quehenberger, A. M. Armando, A. H. Brown, S. B. Milne, D. S. Myers, A. H. Merrill, S. Bandyopadhyay, K. N. Jones, S. Kelly, R. L. Shaner, C. M. Sullards, E. Wang, R. C. Murphy, R. M. Barkley, T. J. Leiker, C. R. H. Raetz, Z. Q. Guan, G. M. Laird, D. A. Six, D. W. Russell, J. G. McDonald, S. Subramaniam, E. Fahy and E. A. Dennis, *J. Lipid Res.*, 2010, **51**, 3299–3305.
58 E. Fahy, S. Subramaniam, R. C. Murphy, M. Nishijima, C. R. Raetz, T. Shimizu, F. Spener, G. van Meer, M. J. Wakelam and E. A. Dennis, *J. Lipid Res.*, 2009, **50**(suppl), S9–S14.
59 D. Rajasundaram and J. Selbig, *Curr. Opin. Plant Biol.*, 2016, **30**, 57–61.
60 L. Varemo, J. Nielsen and I. Nookaew, *Nucleic Acids Res.*, 2013, **41**, 4378–4391.
61 K. Dolinski and O. G. Troyanskaya, *Mol. Biol. Cell*, 2015, **26**, 2575–2578.
62 R. S. Fehrmann, J. M. Karjalainen, M. Krajewska, H. J. Westra, D. Maloney, A. Simeonov, T. H. Pers, J. N. Hirschhorn, R. C. Jansen, E. A. Schultes, H. H. van Haagen, E. G. de Vries, G. J. te Meerman, C. Wijmenga, M. A. van Vugt and L. Franke, *Nat. Genet.*, 2015, **47**, 115–125.
63 G. Ciriello, M. L. Miller, B. A. Aksoy, Y. Senbabaoglu, N. Schultz and C. Sander, *Nat. Genet.*, 2013, **45**, U1127–U1247.
64 J. Nielsen, *FEBS Lett.*, 2009, **583**, 3905–3913.
65 R. Saha, A. Chowdhury and C. D. Maranas, *Curr. Opin. Biotechnol.*, 2014, **29**, 39–45.
66 J. S. Edwards and B. Ø. Palsson, *J. Biol. Chem.*, 1999, **274**, 17410–17416.
67 J. Y. Ryu, H. U. Kim and S. Y. Lee, *Integr. Biol.*, 2015, **7**, 859–868.
68 N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas and B. Ø. Palsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 1777–1782.
69 H. Ma, A. Sorokin, A. Mazein, A. Selkov, E. Selkov, O. Demin and I. Goryanin, *Mol. Syst. Biol.*, 2007, **3**, 135.
70 I. Thiele, N. Swainston, R. M. T. Fleming, A. Hoppe, S. Sahoo, M. K. Aurich, H. Haraldsdottir, M. L. Mo, O. Rolfsson, M. D. Stobbe, S. G. Thorleifsson, R. Agren, C. Bolling, S. Bordel, A. K. Chavali, P. Dobson, W. B. Dunn, L. Endler, D. Hala, M. Hucka, D. Hull, D. Jameson, N. Jamshidi, J. J. Jonsson, N. Juty, S. Keating, I. Nookaew, N. Le Novere, N. Malys, A. Mazein, J. A. Papin, N. D. Price, E. Selkov, M. I. Sigurdsson, E. Simeonidis, N. Sonnenschein, K. Smallbone, A. Sorokin, J. H. G. M. van Beek, D. Weichart, I. Goryanin, J. Nielsen, H. V. Westerhoff, D. B. Kell, P. Mendes and B. Ø. Palsson, *Nat. Biotechnol.*, 2013, **31**, 419–425.
71 D. Croft, *Methods Mol. Biol.*, 2013, **1021**, 273–283.
72 M. Kanehisa, S. Goto, Y. Sato, M. Furumichi and M. Tanabe, *Nucleic Acids Res.*, 2012, **40**, D109–D114.
73 P. Romero, J. Wagg, M. L. Green, D. Kaiser, M. Krummenacker and P. D. Karp, *Genome Biol.*, 2005, **6**, R2.
74 R. Agren, S. Bordel, A. Mardinoglu, N. Pornputtapong, I. Nookaew and J. Nielsen, *PLoS Comput. Biol.*, 2012, **8**, e1002518.
75 C. Gille, C. Bolling, A. Hoppe, S. Bulik, S. Hoffmann, K. Hubner, A. Karlstadt, R. Ganeshan, M. Konig,

K. Rother, M. Weidlich, J. Behre and H. G. Holzhutter, *Mol. Syst. Biol.*, 2010, **6**, 411.
76 S. Sahoo, L. Franzson, J. J. Jonsson and I. Thiele, *Mol. BioSyst.*, 2012, **8**, 2545–2558.
77 S. Sahoo and I. Thiele, *Hum. Mol. Genet.*, 2013, **22**, 2705–2722.
78 M. Garcia-Albornoz and J. Nielsen, *Ind. Biotechnol.*, 2013, **9**, 203–214.
79 S. Y. Lee and H. U. Kim, *Nat. Biotechnol.*, 2015, **33**, 1061–1072.
80 M. A. Oberhardt, B. Ø. Palsson and J. A. Papin, *Mol. Syst. Biol.*, 2009, **5**, 320.
81 N. D. Price, J. A. Papin, C. H. Schilling and B. Ø. Palsson, *Trends Biotechnol.*, 2003, **21**, 162–169.
82 A. Varma and B. Ø. Palsson, *Nat. Biotechnol.*, 1994, **12**, 994–998.
83 D. Petranovic and J. Nielsen, *Trends Biotechnol.*, 2008, **26**, 584–590.
84 D. R. Hyduke, N. E. Lewis and B. Ø. Palsson, *Mol. BioSyst.*, 2013, **9**, 167–174.
85 S. A. Becker and B. Ø. Palsson, *PLoS Comput. Biol.*, 2008, **4**, e1000082.
86 T. Shlomi, M. N. Cabili, M. J. Herrgard, B. Ø. Palsson and E. Ruppin, *Nat. Biotechnol.*, 2008, **26**, 1003–1010.
87 L. Jerby, T. Shlomi and E. Ruppin, *Mol. Syst. Biol.*, 2010, **6**, 401.
88 M. Åkesson, J. Forster and J. Nielsen, *Metab. Eng.*, 2004, **6**, 285–293.
89 P. A. Jensen and J. A. Papin, *Bioinformatics*, 2011, **27**, 541–547.
90 S. Chandrasekaran and N. D. Price, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 17845–17850.
91 A. S. Blazier and J. A. Papin, *Front. Physiol.*, 2012, **3**, 299.
92 D. Machado and M. Herrgard, *PLoS Comput. Biol.*, 2014, **10**, e1003580.
93 O. Resendis-Antonio, C. Gonzalez-Torres, G. Jaime-Munoz, C. E. Hernandez-Patino and C. F. Salgado-Munoz, *Semin. Cancer Biol.*, 2015, **30**, 79–87.
94 M. K. Kim and D. S. Lun, *Comput. Struct. Biotechnol. J.*, 2014, **11**, 59–65.
95 A. P. Oliveira, K. R. Patil and J. Nielsen, *BMC Syst. Biol.*, 2008, **2**, 17.
96 T. Cakir, K. R. Patil, Z. Onsan, K. O. Ulgen, B. Kirdar and J. Nielsen, *Mol. Syst. Biol.*, 2006, **2**, 50.
97 C. Frezza, L. Zheng, O. Folger, K. N. Rajagopalan, E. D. MacKenzie, L. Jerby, M. Micaroni, B. Chaneton, J. Adam, A. Hedley, G. Kalna, I. P. Tomlinson, P. J. Pollard, D. G. Watson, R. J. Deberardinis, T. Shlomi, E. Ruppin and E. Gottlieb, *Nature*, 2011, **477**, 225–228.
98 F. Gatto, H. Miess, A. Schulze and J. Nielsen, *Sci. Rep.*, 2015, **5**, 10738.
99 F. Gatto, I. Nookaew and J. Nielsen, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E866–E875.
100 J. Hu, J. W. Locasale, J. H. Bielas, J. O'Sullivan, K. Sheahan, L. C. Cantley, M. G. Vander Heiden and D. Vitkup, *Nat. Biotechnol.*, 2013, **31**, 522–529.
101 A. Mardinoglu, C. Kampf, A. Asplund, L. Fagerberg, B. M. Hallstrom, K. Edlund, M. Bluher, F. Ponten, M. Uhlen and J. Nielsen, *J. Proteome Res.*, 2014, **13**, 5106–5119.
102 H. Nam, M. Campodonico, A. Bordbar, D. R. Hyduke, S. Kim, D. C. Zielinski and B. Ø. Palsson, *PLoS Comput. Biol.*, 2014, **10**, e1003837.
103 L. Varemo, C. Scheele, C. Broholm, A. Mardinoglu, C. Kampf, A. Asplund, I. Nookaew, M. Uhlen, B. K. Pedersen and J. Nielsen, *Cell Rep.*, 2015, **11**, 921–933.
104 F. Gatto, N. Volpi, H. Nilsson, I. Nookaew, M. Maruzzo, A. Roma, M. E. Johansson, U. Stierner, S. Lundstam, U. Basso and J. Nielsen, *Cell Rep.*, 2016, **15**, 1822–1836.
105 O. Folger, L. Jerby, C. Frezza, E. Gottlieb, E. Ruppin and T. Shlomi, *Mol. Syst. Biol.*, 2011, **7**, 501.
106 M. Scheer, A. Grote, A. Chang, I. Schomburg, C. Munaretto, M. Rother, C. Sohngen, M. Stelzer, J. Thiele and D. Schomburg, *Nucleic Acids Res.*, 2011, **39**, D670–D676.
107 R. Agren, A. Mardinoglu, A. Asplund, C. Kampf, M. Uhlen and J. Nielsen, *Mol. Syst. Biol.*, 2014, **10**, 721.
108 P. Ghaffari, A. Mardinoglu, A. Asplund, S. Shoaie, C. Kampf, M. Uhlen and J. Nielsen, *Sci. Rep.*, 2015, **5**, 8183.
109 P. Ghaffari, A. Mardinoglu and J. Nielsen, *Front. Physiol.*, 2015, **6**, 382.
110 L. Varemo, I. Nookaew and J. Nielsen, *Front. Physiol.*, 2013, **4**, 92.
111 A. Mardinoglu and J. Nielsen, *Curr. Opin. Biotechnol.*, 2015, **34**, 91–97.
112 A. Mardinoglu, F. Gatto and J. Nielsen, *Biotechnol. J.*, 2013, **8**, 985–996.
113 A. Bachi and T. Bonaldi, *J. Proteomics*, 2008, **71**, 357–367.
114 H. A. Ebhardt, A. Root, C. Sander and R. Aebersold, *Proteomics*, 2015, **15**, 3193–3208.
115 A. G. N. Wetie, I. Sokolowska, A. G. Woods, U. Roy, K. Deinhardt and C. C. Darie, *Cell. Mol. Life Sci.*, 2014, **71**, 205–228.
116 Y. Liu, A. Beyer and R. Aebersold, *Cell*, 2016, **165**, 535–550.
117 S. Fields and O. Song, *Nature*, 1989, **340**, 245–246.
118 B. Stynen, H. Tournu, J. Tavernier and P. Van Dijck, *Microbiol. Mol. Biol. Rev.*, 2012, **76**, 331–382.
119 I. Stagljar, C. Korostensky, N. Johnsson and S. te Heesen, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 5187–5192.
120 N. Johnsson and A. Varshavsky, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 10340–10344.
121 P. Ulrichts, I. Lemmens, D. Lavens, R. Beyaert and J. Tavernier, *Methods Mol. Biol.*, 2009, **517**, 133–144.
122 W. H. Dunham, M. Mullin and A. C. Gingras, *Proteomics*, 2012, **12**, 1576–1590.
123 S. Kumar, D. Alibhai, A. Margineanu, R. Laine, G. Kennedy, J. McGinty, S. Warren, D. Kelly, Y. Alexandrov, I. Munro, C. Talbot, D. W. Stuckey, C. Kimberly, B. Viellerobe, F. Lacombe, E. W. Lam, H. Taylor, M. J. Dallman, G. Stamp, E. J. Murray, F. Stuhmeier, A. Sardini, M. Katan, D. S. Elson, M. A. Neil, C. Dunsby and P. M. French, *ChemPhysChem*, 2011, **12**, 609–626.
124 K. Venkatesan, J. F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, M. Zenkner, X. Xin, K. I. Goh,

M. A. Yildirim, N. Simonis, K. Heinzmann, F. Gebreab, J. M. Sahalie, S. Cevik, C. Simon, A. S. de Smet, E. Dann, A. Smolyar, A. Vinayagam, H. Yu, D. Szeto, H. Borick, A. Dricot, N. Klitgord, R. R. Murray, C. Lin, M. Lalowski, J. Timm, K. Rau, C. Boone, P. Braun, M. E. Cusick, F. P. Roth, D. E. Hill, J. Tavernier, E. E. Wanker, A. L. Barabasi and M. Vidal, *Nat. Methods*, 2009, **6**, 83–90.

125 U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksoz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach and E. E. Wanker, *Cell*, 2005, **122**, 957–968.

126 J. F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth and M. Vidal, *Nature*, 2005, **437**, 1173–1178.

127 R. M. Ewing, P. Chu, F. Elisma, H. Li, P. Taylor, S. Climie, L. McBroom-Cerajewski, M. D. Robinson, L. O'Connor, M. Li, R. Taylor, M. Dharsee, Y. Ho, A. Heilbut, L. Moore, S. Zhang, O. Ornatsky, Y. V. Bukhman, M. Ethier, Y. Sheng, J. Vasilescu, M. Abu-Farha, J. P. Lambert, H. S. Duewel, I. I. Stewart, B. Kuehl, K. Hogue, K. Colwill, K. Gladwish, B. Muskat, R. Kinach, S. L. Adams, M. F. Moran, G. B. Morin, T. Topaloglou and D. Figeys, *Mol. Syst. Biol.*, 2007, **3**, 89.

128 H. Y. Chuang, E. Lee, Y. T. Liu, D. Lee and T. Ideker, *Mol. Syst. Biol.*, 2007, **3**, 140.

129 C. S. Greene, A. Krishnan, A. K. Wong, E. Ricciotti, R. A. Zelaya, D. S. Himmelstein, R. Zhang, B. M. Hartmann, E. Zaslavsky, S. C. Sealfon, D. I. Chasman, G. A. FitzGerald, K. Dolinski, T. Grosser and O. G. Troyanskaya, *Nat. Genet.*, 2015, **47**, 569–576.

130 O. A. Balbin, J. R. Prensner, A. Sahu, A. Yocum, S. Shankar, R. Malik, D. Fermin, S. M. Dhanasekaran, B. Chandler, D. Thomas, D. G. Beer, X. Cao, A. I. Nesvizhskii and A. M. Chinnaiyan, *Nat. Commun.*, 2013, **4**, 2617.

131 R. Barshir, O. Shwartz, I. Y. Smoly and E. Yeger-Lotem, *PLoS Comput. Biol.*, 2014, **10**, e1003632.

132 M. H. Schaefer, T. J. Lopes, N. Mah, J. E. Shoemaker, Y. Matsuoka, J. F. Fontaine, C. Louis-Jeune, A. J. Eisfeld, G. Neumann, C. Perez-Iratxeta, Y. Kawaoka, H. Kitano and M. A. Andrade-Navarro, *PLoS Comput. Biol.*, 2013, **9**, e1002860.

133 A. L. Tarca, S. Draghici, P. Khatri, S. S. Hassan, P. Mittal, J. S. Kim, C. J. Kim, J. P. Kusanovic and R. Romero, *Bioinformatics*, 2009, **25**, 75–82.

134 L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork and C. von Mering, *Nucleic Acids Res.*, 2009, **37**, D412–D416.

135 H. Ge, A. J. M. Walhout and M. Vidal, *Trends Genet.*, 2003, **19**, 551–560.

136 J. Loscalzo and A. L. Barabasi, *Wiley Interdiscip. Rev.: Syst. Biol. Med.*, 2011, **3**, 619–627.

137 K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal and A. L. Barabasi, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 8685–8690.

138 S. H. Hu, Z. Xie, A. Onishi, X. P. Yu, L. Z. Jiang, J. Lin, H. S. Rho, C. Woodard, H. Wang, J. S. Jeong, S. Y. Long, X. F. He, H. Wade, S. Blackshaw, J. Qian and H. Zhu, *Cell*, 2009, **139**, 610–622.

139 K. Yugi, H. Kubota, A. Hatano and S. Kuroda, *Trends Biotechnol.*, 2016, **34**, 276–290.

140 C. Zhang and Q. Hua, *Front. Physiol.*, 2016, **6**, 413.

