REVIEW

Integrating transcriptome and proteome profiling: Strategies and applications

Dhirendra Kumar[1], Gourja Bansal[1], Ankita Narang[1], Trayambak Basak[1,2], Tahseen Abbas[1,2] and Debasis Dash[1,2]*

[1] G.N. Ramachandran Knowledge Center for Genome Informatics, CSIR-Institute of Genomics and Integrative Biology, South Campus, Sukhdev Vihar, New Delhi, INDIA
[2] Academy of Scientific & Innovative Research (AcSIR), CSIR-IGIB South Campus, New Delhi, India

Received: March 16, 2016
Revised: June 12, 2016
Accepted: June 23, 2016

Discovering the gene expression signature associated with a cellular state is one of the basic quests in majority of biological studies. For most of the clinical and cellular manifestations, these molecular differences may be exhibited across multiple layers of gene regulation like genomic variations, gene expression, protein translation and post-translational modifications. These system wide variations are dynamic in nature and their crosstalk is overwhelmingly complex, thus analyzing them separately may not be very informative. This necessitates the integrative analysis of such multiple layers of information to understand the interplay of the individual components of the biological system. Recent developments in high throughput RNA sequencing and mass spectrometric (MS) technologies to probe transcripts and proteins made these as preferred methods for understanding global gene regulation. Subsequently, improvements in "big-data" analysis techniques enable novel conclusions to be drawn from integrative transcriptomic-proteomic analysis. The unified analyses of both these data types have been rewarding for several biological objectives like improving genome annotation, predicting RNA-protein quantities, deciphering gene regulations, discovering disease markers and drug targets. There are different ways in which transcriptomics and proteomics data can be integrated; each aiming for different research objectives. Here, we review various studies, approaches and computational tools targeted for integrative analysis of these two high-throughput omics methods.

Keywords:
Bioinformatics / Network biology / Post translational modifications (PTM) / Proteogenomics / Ribosome profiling / RNA-seq

1 Introduction

Gene expression profiles, either in the form of transcriptome and/or proteome, provide means to explore and determine underlying molecular and cellular processes. The last decade has brought about several key advances in high-throughput technologies like RNA sequencing and shotgun proteomics that have now enabled probing transcript and protein expression at an unprecedented depth and coverage [1]. Due to the reduced time and cost, these methods have fast gained broad applicability in the study of various biological systems, ranging from pathogens to embryonic development and cancer [2]. Now, the power of high-throughput techniques can be leveraged to make inferences at the genome-wide scale rather than quantifying the expression of a few genes using the conventional experimental methods [3].

While the genome remains nearly static for an organism, its transcriptome and proteome rapidly change, albeit in a

Correspondence: Dr. Debasis Dash, G.N. Ramachandran Knowledge Center for Genome Informatics, CSIR-Institute of Genomics and Integrative Biology, South Campus, Sukhdev Vihar, Mathura Road, Delhi 110025, India
E-mail: ddash@igib.res.in, ddash@igib.in
Fax: +91-11-29879301

Abbreviations: ANN, artificial neural network; **PTM**, post-translational modifications; **RBS**, ribosomal binding site; **TIS**, translation initiation site; **TSS**, transcriptional start site

*Additional corresponding author: Dhirendra Kumar,
E-mail: dhirendra8kumar@gmail.com
Colour Online: See the article online to view Figs. 1–3 in colour.

tightly regulated manner in response to different environmental perturbations and growth conditions. This dynamicity of a cell or tissue can be estimated by measuring its transcriptome (expressed portion of the genome) and/or proteome (expressed protein set from genome). However, another order of dynamicity is added by post-transcriptional and post-translational regulations of gene expression. These include, but are not limited to, alternative splicing [4] and editing of expressed transcripts and post-translational addition of covalent modifications [5] to proteins. Thus, several isoforms or proteoforms with distinct structural and functional attributes may originate from a given gene. Further, the crosstalk between these different biological macro-molecules presents a stupendous number of possibilities in which molecular patterns can be reflected [6]. Deciphering these gene expression patterns specifically associated with a given biological state is fundamental to understanding cellular processes and diseases.

Sustained developments in the nucleotide sequencing technology, especially RNA-sequencing, has resulted in an explosive growth in the number and quality of transcriptome sequencing for various tissues and diseases [7]. It enables profiling of the slightest changes in gene expression between two conditions and thus, is rich in information. Besides the expression profiling of annotated genes, it may also reveal alternate transcriptional start site (TSS) usage, novel splice variants, intergenic transcripts, expression quantitative trait loci (eQTLs) [8] and fusion transcripts; all of which may have functional implications in the cellular context [9]. Transcriptome sequencing has been immensely beneficial in finding key markers of various human cancers. It has also been immensely useful in discovering potential drug targets when integrated with epigenetic marks like DNA methylation and histone modification. [10].

Recent advances in instrumentation and analytical methods associated with mass spectrometry (MS)-based proteome profiling makes it a powerful technique for probing gene expression changes at the protein level [11, 12]. The major goal of proteomics is to build the complete proteome map of a species which must include precise cellular localization of each protein, delineate signaling pathways and describe their regulatory PTMs [13]. Combined with liquid chromatography (LC), tandem mass spectrometry (MS/MS) also provides a sensitive method to capture protein quantity differences [14, 15]. This unique ability of proteomics to discover the post-translational modifications and their quantities may allow segregation of active proteoforms from the inactive ones and thus, may facilitate drawing accurate functional inferences from gene expression data.

Although, both transcriptome and proteome profiling methods are rich in biological information, these are limited in their abilities to provide a comprehensive perspective of the system when analyzed individually. Despite high sensitivity, estimation of transcript expression is not sufficient to provide a picture of the true biological state, as mRNA profiling does not capture regulatory processes or post-transcriptional modifications that might affect the amount of active protein [16]. Similarly, proteomics lacks the sensitivity to detect low abundant proteins and is limited in its ability to identify novel proteoforms resulting from alternative splicing or SNPs. However, integrative analysis of these "omics" datasets may present complementary information towards drawing more informative conclusions.

In recent years, researchers are integrating knowledge from transcriptomic and proteomic data to gain meaningful insights. By integrating expressed transcript information in proteomics data analysis, various discoveries like novel coding genes, alternate translation initiation sites (TIS), splice variants, single amino acid polymorphism, etc., can be made [17]. Similarly, integrative transcriptomic-proteomic analysis highlighted a poor correlation between the quantities of these two macromolecules indicative of a complex regulatory mechanism controlling expression both at the RNA and the protein levels. In the following sections, we have reviewed the notable studies and approaches that have used integrative transcriptomic and proteomic analyses. Broadly, we have summarized these approaches into the following categories:

(i) Transcriptome as a template for proteomics data analysis.
(ii) Integrative analyses to decipher gene expression and regulation.
(iii) Clinical applications of integrative transcriptome and proteome analyses.

2 Transcriptome as a template for proteomics data analysis

A regular shotgun proteomics experiment utilizes a mass spectrometer to acquire the mass to charge ratios of peptides and their fragments resulting in a unique mass spectrum for each peptide [15]. These tandem mass spectra (MS/MS) are then searched against a database of probable protein sequences so as to identify the peptides and proteins expressed in that sample [18]. Alternate approaches are of sequencing peptides de novo from the MS/MS spectra. Although there have been tremendous improvements in *de novo* peptide discovery software like PEAKS [19], etc., these still have a remarkably reduced sensitivity compared to database search method and are also suspected of high false positives [20]. Thus, database search approach is the preferred method of peptide discovery from spectral data. However, the drawback of this method is its dependency on the database. Peptide identification algorithms like MASCOT [21], MassWiz [22], etc., can only identify peptides present in the search database. This necessitates that proteomics search databases should be complete and accurate. On the contrary, most of the proteome databases used for spectral searches are the *in silico* annotated proteomes, which generally contain errors and lacks completeness [23, 24]. Additionally, tissue and individual-specific proteome variations attributable to SNPs and splice variants are not represented in the annotated reference proteome.

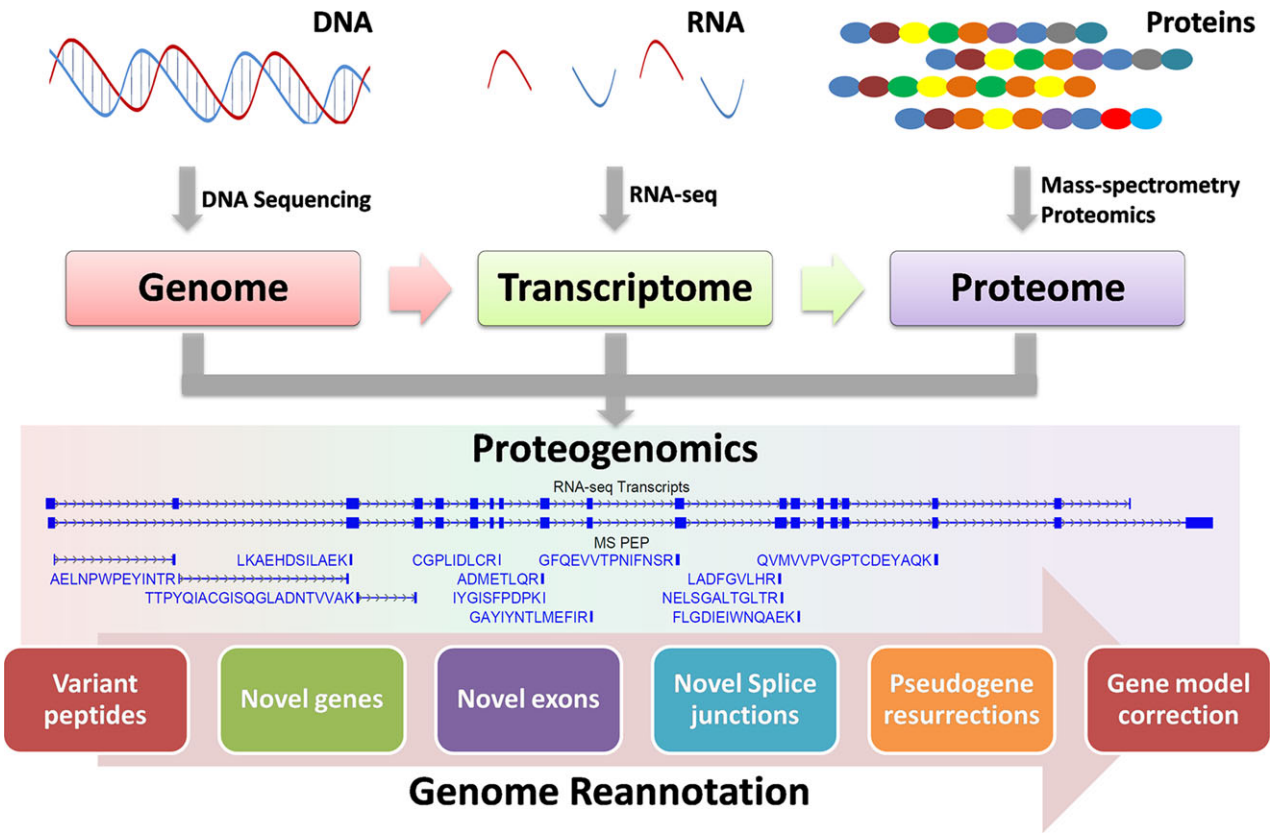

Figure 1. Integrating the mRNA sequencing and peptide sequencing for proteogenomic discovery and genome annotation.

Proteogenomics is a new direction of research where proteomics data is utilized to annotate genes in the genome [25, 26]. However, for this task, proteogenomic methods rely on custom genomic databases, other than the annotated reference proteome, to discover novel peptides indicative of novel translated regions [27]. While genomic databases such as six frame translated genome or ab initio predicted genes may include majority of the novel translation possibilities for a prokaryotic genome [28–30], it would have a limited representation of protein isoforms possible from a eukaryotic genome [31]. Proteomic diversity of a eukaryote is largely attributed to alternative mRNA splicing [4]. The number of possible splice variants and resulting protein isoforms, rise exponentially with the number of the exons in a gene. For complex organisms like human, the number of exons for a few genes is more than a hundred. Including all these possible exon combinations in the search database would increase the search space drastically and thus make peptide identification a time consuming and high false positive generating process [32]. Additionally, to capture intergenic or intronic novel coding regions, a six frame translated eukaryotic genome may increase the database search size thousand folds as only a minor fraction of most of the eukaryotic genome is believed to be coding. Thus, genomic databases-based proteogenomic analysis of eukaryotic organisms only promises a limited success. Rather, by integrating high-throughput transcriptome profiling, such as captured by RNA-seq, may provide a compact, condition specific search database for fast and better proteogenomic analysis [33].

In general, transcripts are assembled from RNA-seq reads and these are translated in three reading frames to create the proteomics search database. The mass spectra are searched against this custom transcriptomic database and also against annotated protein database. Peptides unique to the transcriptome database indicate novel protein isoforms not annotated for the organisms (Fig. 1). RNA-seq reads may also be utilized to capture non-synonymous nucleotide variations (both genomic and RNA editing) into the database which may allow detection of variant peptides potentially involved in various diseases including cancers [34]. RNA-seq analysis may also indicate fusion transcripts which may be checked for their coding potential by integrating it with the proteomics data [35]. Thus, integrating the transcripts, profiled from high-throughput RNA-seq methods, in proteogenomic analyses may lead to discovery of translated novel exons, splice variants, non-synonymous mutations, novel genes, correction of annotations of translation initiation site (TIS) and also the detection of fusion proteins [17]. In cases where the genome is yet to be sequenced, RNA-seq data can be *de novo* assembled

into potential transcripts which greatly enhances the peptide detection from such under-explored organisms [36].

Various studies have benefited from such an integrative proteogenomic analysis. Most notable of these are the recently reported draft human proteome maps which captured several new peptides and proteins previously missed [37, 38]. Although both these studies have been criticized for their lenient false discovery rate (FDR) measures [39, 40], they added several new proteins to the previously known human proteome. High error rates are common in proteogenomic studies primarily due to extra-large genomic or transcriptomic search database and large amounts of MS/MS data being searched. Both peptide and protein level FDR estimation and result filters are necessary to be implemented in proteogenomics studies in order to avoid any error to propagate in gene and protein annotations in reference proteomes. Recently, we could detect several novel proteoforms in the rat genome by integrating publicly available RNA-seq and proteomics data using a stringent analysis strategy [33]. Kelkar et al. utilized transcriptomic and proteomics profiling from various different organs of zebrafish to comprehensively re-annotate the genome of this model organism [24]. Violette et al. discovered several new toxins from cone snail venom by utilizing similar integrative approach [41]. Further, Dutertre et al. extended the discovery of conopeptide toxins from cone snails by integrating mRNA sequencing with high resolution MS [42]. In another study, new protein toxins could be identified in jellyfish, whose genome is yet to be sequenced, by integrating *de novo* assembled transcripts to create a proteomic search database [36]. Similar studies have been rewarding for high-throughput re-annotation of various other genomes [43]. Integrative proteogenomic approaches have also been employed to discover novel contributors in various human diseases primarily cancers. Zhang et al. have carried out an extensive proteogenomic characterization of colon cancer for candidate prioritization by integrating genomic variations, mRNA expression and protein discovery [34]. Similarly, Ruggles et al. have identified various single nucleotide and splice variants in patient-derived breast cancer xenografts [44].

Numerous software tools have been developed to facilitate this integrative proteogenomic analysis from RNA-seq and MS proteomics data. CustomProDB is an R package which creates proteomics search database by incorporating the genomic variations called from RNA-seq data [45]. Similarly, SpliceDB creates an MS data searchable compact proteome database of splice patterns from RNA-seq reads [46]. MSProGene constructs a sample specific proteome database from RNA-seq data and also stores the peptide shared-ness among database entries which may be utilized to resolve the expression of isoforms [47]. Enosi is a pipeline which provides a complete solution for proteogenomic re-annotation of genomes by utilizing the RNA-seq reads to identify peptides from MS data [48]. PGTools, a similar tool, facilitates the discovery of novel peptides from human disease samples by integrating the transcriptome data in the protein discovery process [35]. Integrated transcriptomic-proteomic pipeline (ITP) incorporates the reference-based transcriptome assembly to build the comprehensive database to search MS data and annotate eukaryotic genomes [33]. PPLine is a proteogenomic tool which performs integrated transcriptomic proteomic analysis to discover variant peptides accumulating amino-acid variations [49]. QUILTS enables the discovery of coding SNPs and splice variants from the proteogenomic survey of human disease samples by creating individual-specific search databases [44]. Galaxy-P is an extension of web base Galaxy framework and allows comprehensive yet flexible integrative analysis of transcriptomic and proteomics datasets [50]. It also extends the capabilities of the commercial software ProteinPilot towards proteogenomics. ProteinPilot encompasses the Paragon algorithm which discovers peptides without various restrictions of the search parameters and may be beneficial in proteogenomic studies [51].

There are only a few studies where above-mentioned tools are compared with each other for their performances. In a recent study, we observed that many of these tools actually provide complementary results in peptide discovery suggesting the use of multiple tools for a comprehensive proteogenomic analysis [33]. Enosi appears to be among the most comprehensive standalone tool for such integration. However, ITP utilizes RNA quantities to derive better conclusions for the expressed isoforms and may be useful in similar studies. While above mentioned tools have extended our ability towards integrating RNA-seq within proteogenomics studies, they still require considerable bioinformatics skills to enable their effective use. This attribute limits their reach to experimental scientists and improvements in this aspect are badly required. Although tools like GalaxyP and PGTools are relatively easier to use, proteogenomic software are far from being in routine analysis primarily due to difficulties in their implementation.

Ribosome profiling (RIBO-seq), an upcoming nucleotide sequencing technology, captures transcripts which are being actively translated by ribosomes [52]. It provides both qualitative and quantitative information about in vivo protein translation. It not only delineates ribosome occupied protein-coding transcripts from non-coding transcripts but also reflects transcriptome-wide protein translation efficiency and rates [53]. Thus, it may serve to bridge the gap between transcriptome and proteome [52]. However, it should be noted that RIBO-seq may not be a substitute for MS-based proteome profiling. While RIBO-seq provides much better depth and coverage of protein translation and can detect translation on transcripts which typically result in low abundant proteins, it does not reflect the post translational events like protein degradation rates and PTMs which affect active protein quantities within a cell. Utilizing the complementary nature of ribosome profiling, peptide discovery from MS/MS data can be further improved. Search databases can be created from RIBO-seq profiled transcripts providing a compact, precise and specific search space for MS data. Menschaert et al. have demonstrated that utilizing RIBO-seq transcripts in proteogenomics study not only increases

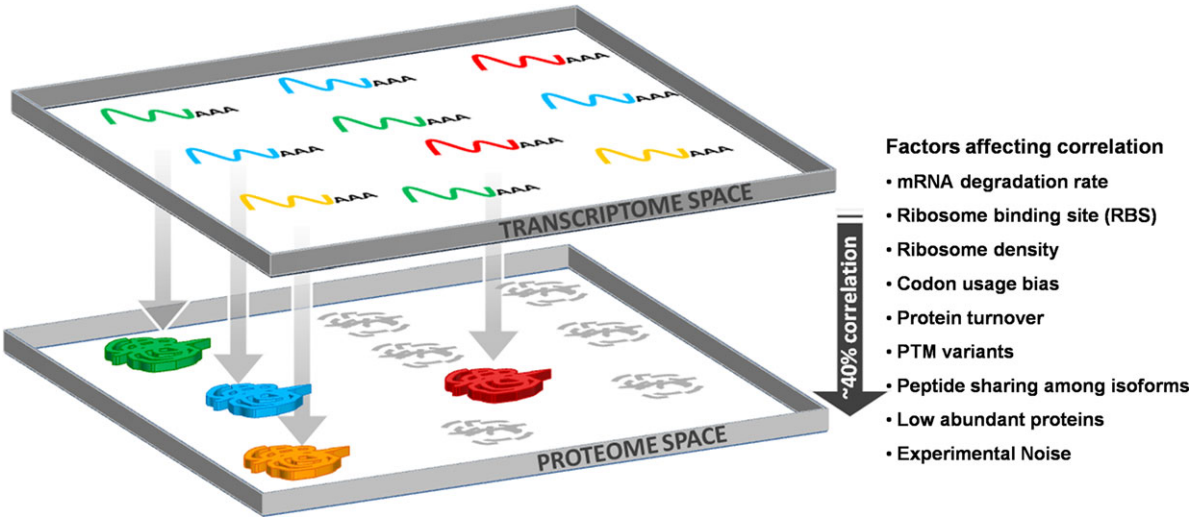

Figure 2. Factors influencing the correlation between mRNA–protein quantities.

peptide discovery but also helps in discovering alternate TISs [54]. Similarly, Koch et al. could detect various alternate TIS and new proteins from human cancer cell line by integrating the RIBO-seq with proteogenomics approach [55]. A recently developed software PROTEOFORMER enables creation of a MS friendly search database from RIBO-seq reads and thus allows seamless integration of proteomics with ribosome profiling [56].

The availability of analysis software and extensive transcriptomic, and proteomic datasets in public repositories like GEO [57] and PRIDE [58,59], allows comprehensive proteogenomic analyses leading to refinement of genome annotations for model organisms, discovery of variant peptides and isoforms, discovery of novel disease markers and fusion proteins, etc.

3 Integrative analysis to decipher gene expression and regulation

For decades, gene expression has primarily been studied at the transcript level with the assumption that transcript quantities are indicative of active protein quantities [57]. The explicit concept of central dogma which determines the functional aspects of genetic codes through gene expression (mRNA) and protein translation has been the hallmark of cellular functional entity. However, several studies have revealed that the measured quantities of mRNAs and proteins correlate only modestly [60,61]. Various biological factors, along with the inherent experimental noise in high-throughput technologies and inadequacy of statistical tools, can be attributed to the poor correlation coefficient observed in those studies [62]. A strong correlation between transcriptomic and proteomic data would allow predictability of protein quantities from highly sensitive transcriptomic methods. However, the lack of correlation indicates possibility of experimental error or biological uncoupling between the levels of mRNA and protein.

Correlation coefficients vary across different organisms ranging from 0.2 to 0.47 in bacteria, 0.34 to 0.87 in yeast and 0.09 to 0.46 in multi-cellular organisms [63]. Presence of weak ribosome binding site (Shine-Dalgarno for prokaryotes and Kozak for Eukaryotes), regulatory proteins, codon usage bias, half-life difference between protein and mRNA are some of the biological reasons which could be attributed towards weak correlation between measured RNA and proteins (Fig. 2). Translation efficiency is majorly affected by the number of ribosomes in a transcription unit which is generally known as the ribosome density. Experiments on yeast cells showed that mRNA species having more number of ribosomes attached have higher translation rates [64]. Schwanhäusser et al. reported that mRNAs are five times less stable and 900 times less abundant than proteins in mammalian cells [62]. Ubiquitination, phosphorylation and cellular localization are some of the post-translational regulations which can affect protein half-lives and thus, their detection [60]. Studies have reported that mRNA abundance can predict protein abundance only partially, for ~40% genes. Various post-transcriptional regulations, levels of PTMs, consideration of experiment noise, etc., need to be factored in the equation to accurately explain the remaining 60% variations [61].

Given that the correlation between transcriptome and proteome data is low, their joint analysis may allow gaining useful insights about the mRNA-protein expression dynamics. There are different approaches that have been utilized to better understand gene expression and its regulation by integrative analysis of transcriptomic and proteomic datasets. By creating a reference data set using the union of proteomic

and transcriptomic data from different samples, Nathanael et al. discovered a significant number of bacterial metabolism enzymes in *Bradyrhizobium japonicum* which were not identified from proteomic data alone [65]. Instead of a direct correlation analysis Purizian et al. have used topological network methods (over-connection analysis, hidden node analysis, rank aggregation and network analysis) leading to the identification of common regulatory machinery and signaling pathways that might contribute to the altered state of this regulatory network in psoriatic lesion [66].

Several approaches have been used to tackle the issue of missing data in proteomic datasets in order to provide an unbiased biological interpretation from integration studies done on temporal expression data. The estimation of missing values can be done using Nearest-neighbor and Bayesian principal component analysis (BPCA) methods, and by integrating the gene ontology (GO) information into the proteomic data imputation. Zero-inflated poisson (ZIP) linear regression model and a stochastic gradient boosted trees (GBT) nonlinear model uncover possible relationships between transcriptomic and proteomic data and improve the predictability of abundances of experimentally undetected proteins. Using non-linear optimization model by implementing GBT method, Garcia et al. estimated missing protein expression in sulfate reducing bacterium, *Desulfovibrio vulgaris* [67]. After estimating the non-linear relationship and missing protein expression values, they could validate the results using literature knowledge. In another study, Li et al. used artificial neural network (ANN) approach to predict the abundance of experimentally undetected proteins in *D. vulgaris* [68]. Authors also quantified the contribution of various sequence measurable factors like mRNA abundance, protein instability index, gene length, effective number of codons and codon adaptation index (CAI) in predicting missing proteomic values by using a multiple logistic regression analysis.

Rogers et al. proposed a coupled clustering approach which creates a certain number of transcriptomic and proteomic clusters and provides a conditional probability of a gene to be in a protein cluster given that it is in an mRNA cluster. Authors have used this approach on time course data for human mammary epithelial cell line (HMEC) stimulated with epidermal growth factor (EGF). Using this approach, authors revealed a complex relationship between transcriptome and proteome with most mRNA clusters linked to at least two protein clusters, and vice versa [69].

Further, to gain functional insights from the integrative analysis, several bio-informatics approaches have been undertaken to develop comprehensive tools like biomaRt [70], Cytoscape [71], VANTED [72], ChromeTracks [73], IPA (http://www.ingenuity.com/), etc. Several integrative tools are being developed based on the basic conceptual implementations that corroborate to both the data sets. Simple expressional statistical correlation among different states, extracting the common functional context [74], topology-based analysis such as a hidden-node-based network analysis [75], clustering based on abundance similarity [69], dynamic modeling (employing such as Bayesian network, Boolean network, etc.) [76, 77]) have remained the key approaches for such different analytical platforms. Recently, Kuo et al. have performed a comparative study for several of the mentioned integrative tools determining their advantages and disadvantages. Additionally, they developed a new tool to integrate transcriptomic, proteomic and metabolic data sets [78]. Despite these advances, there is a dire need to perform a benchmarking study with similar data sets to document the capabilities of these bioinformatics pipelines which would provide ease of use to the researchers.

Besides estimating the protein quantities, discovering the PTMs on these proteins is crucial to understand the underlying signal transduction and cellular responses. PTMs are specialized covalent modifications on specific amino acids of any particular protein in a given biological system [79]. PTMs may allow further segregation of protein quantities into active protein quantities. These modifications act as an important basis for a protein's structural as well as functional entity [80]. For instance, several cellular signaling pathways like MAP kinase, JAK-STAT, AKT, etc., are driven by specific phosphorylation events and have been implicated in many important diseases [81–83] like cancer, different metabolic diseases, etc. Further to exemplify, glycosylation of basement membrane proteins such as collagen IV have been shown to be important in providing the scaffold for tissue-structure maintenance [84]. Importantly, a significant portion of these covalent modifications are specific enzyme dependent. The level and activity of these enzymes are crucial for PTM establishment at a specific time and space. The advantages of next generation-based gene expression analytical platforms are to document subtle changes for almost all the transcripts in a given biological system. Further, the role of common transcription factors which can drive the occurrence of enzymatically catalyzed PTMs can also be inferred from the integrative analyses. Interestingly, the concept of individual somatic genetic variations and alternative splicing events could also contribute to the PTM variation which is limited by the use of standard proteomic database search strategy in any bottom-up proteomics experiment [85, 86]. Thus, the integration of both approaches could be used as an important handle to correlate the enzyme-dependent PTMs to address relevant biological questions. Towards this end, there is a huge scope for the development of newer bioinformatics pipelines to determine the level of PTMs and their accordance with the transcript levels of respective enzymes.

4 Clinical applications of integrative transcriptome and proteome analyses

Comparative transcriptome and/or proteome analyses have been applied to numerous disease studies to identify gene expression signatures specific to the disease state compared to non-diseased or healthy controls. Analysis of transcriptomes

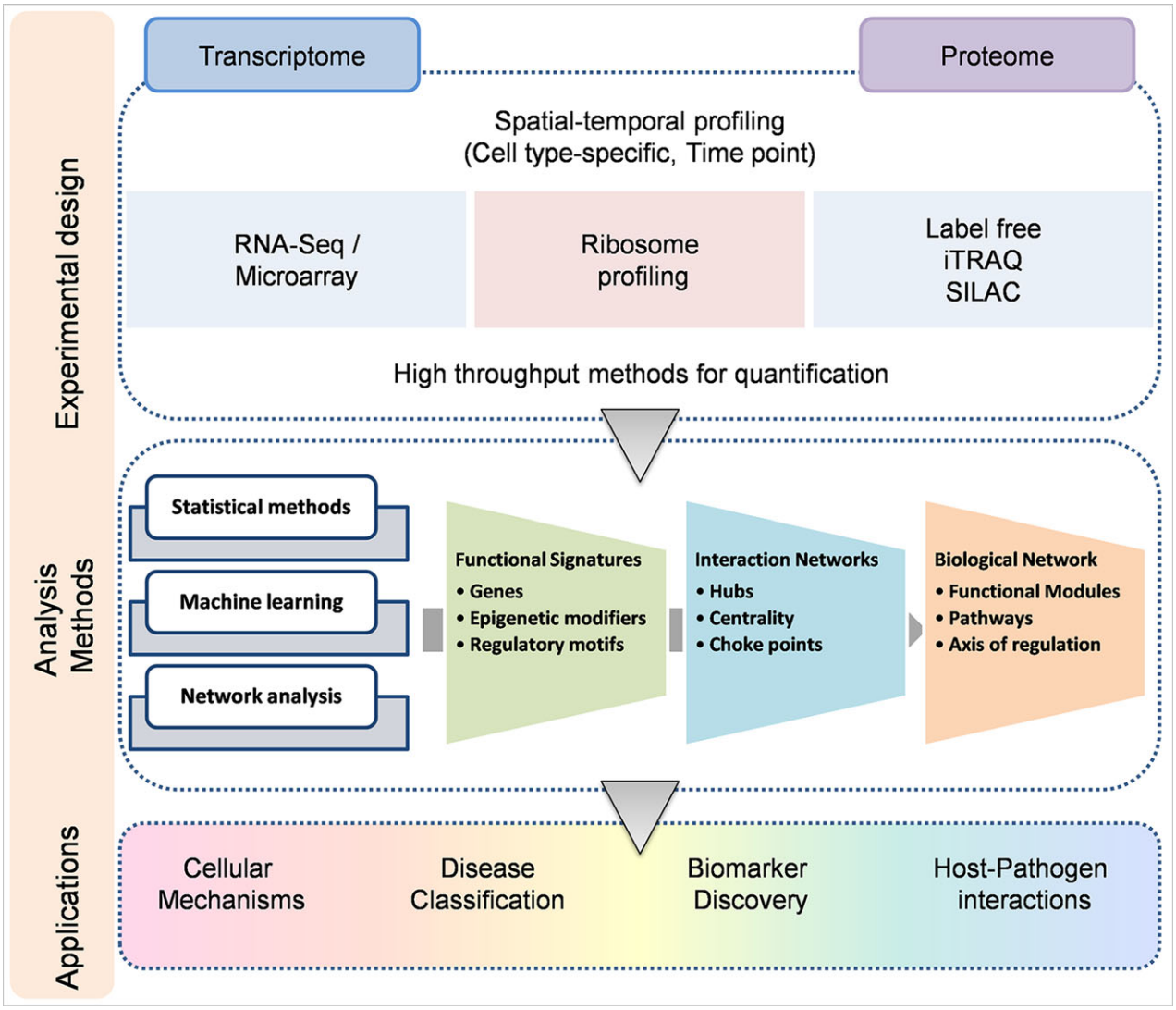

Figure 3. Integrating transcriptomics and proteomics to decipher disease biology. (a) Experimental design: different platforms are available for generation of high-throughput transcriptome and proteome data. Ribosome profiling which quantitates the translation rate can provide better estimates of correlation of transcriptome and proteome datasets. To understand the dynamics of a disease or specific phenotype, expression and translation can be measured for different time frames (temporal) and/or in specific spaces – cell/tissues (spatial). (b) Analysis design: at first stage, transcriptome and proteome are analyzed separately to find signature genes using various statistical or machine learning approaches. Interaction networks from signature genes are further analyzed via network biology approaches to find underlying specific biology processes or pathway that are impacted in disease or phenotype. (c) Applications: integrated omics have diverse applications: Signature genes are often used as biomarkers for disease prognosis and diagnosis or helps in classification of diseases. Inferences from biological processes or pathways may provide insights about disease mechanisms or therapeutics interventions.

and proteomes from diseased individuals may provide cues about the functional and molecular correlates (disease mechanism) of many complex diseases like diabetes [87–89], psoriasis [66], cardiovascular diseases [90] and cancer [91, 92]. Complex disorders are often an outcome of multiple epistatic gene interactions. Notwithstanding the concerted efforts put in by the genomic community to delineate the causal genes or variants, only a small fraction of heritability can be explained till date for most of the common complex disorders. This issue of "missing heritability" or "missing variance" actually reflects the complex etiology of such diseases and thus cannot be explained by addressing only one tier of regulation. Apart from the complex regulatory architecture, the phenotypic manifestation may well be tissue specific and thus targeted and specific experimental design in such cases becomes an important consideration. Infection biology has gained a lot from integrative omics. The detailed mechanisms of many host-pathogen interactions have been worked out. Pathogen surveillance and epidemiological data have specially provided new dimensions to public health. The establishment of an

infection and further sustenance of the pathogen requires a fine balance between activation and inactivation of complex biological processes in both the host and the pathogen. To understand the host tissue response in lymph node tuberculosis (an extra-pulmonary manifestation of tuberculosis), Maji et al. integrated transcriptomics and proteomics data from the host tissue to identify molecular signatures that provide mechanistic insights into the disease pathology or manifestation [93]. Likewise, Villar et al. demonstrated how bacterium ensures its survival in a tick's cell by modulating the endoplasmic reticulum stress towards protein degradation pathway rather than apoptosis [94]. Molecular signatures identified using such integrated approaches also provide a new paradigm for translational research (Fig. 3). Such signatures could be used as potential biomarkers for efficient treatment. Disease biomarkers identified using integrative analyses are better indicators of disease prognosis and diagnosis because chances of false positives are minimized. Shimwell et al. used combined transcriptome and proteome analyses as a non-invasive method towards identifying urinary biomarkers for urothelial carcinoma. Integrative frameworks can also be an alternative strategy towards identification of novel drug targets [91]. Tarun et al. made a remarkable effort in this direction by identifying unique pathways in malaria that can be used as potential drug targets to prevent infection [95]. To understand and/or determine the impact of therapeutic interventions, a clear understanding of the underlying molecular processes is required. Zheng et al. described a systems-level approach or a roadmap to integrate transcriptome and proteome study for understanding the complex biochemical mechanism of combination of therapies in case of promyelocytic leukemia [92].

To understand the phenomenon of organ-specific aging effects in rats, Ori et al. integrated transcriptomics, proteomics and ribosome profiling datasets to identify the cellular changes that manifests at different levels [96]. Age related transcriptional and translation outcomes were found to be different in the brain and the liver at different stages. Liver and brain are impacted more at the transcriptional and the translational levels, respectively. Protein localization and post-translational phospho-proteomic analysis also revealed altered outcomes in aged animals.

In the studies based on an integrated omics design, findings may be analyzed either by taking a consensus of different approaches or a single approach for generating testable hypothesis and other(s) can subsequently substantiate or validate those by focusing on a targeted set of prioritized candidates. In the following sections, we will discuss the different analytical frameworks and considerations for integrated experimental designs.

4.1 Sample classification and identification of biomarkers/features

Gerling et al. discussed the use of various statistical methods/tests like ANOVA, t-test, multiple test correction statistics, Principal component analysis (PCA) and K-means clustering for comparative analysis of high dimensional transcriptome and proteome datasets. Authors demonstrated the applicability of these methods in identifying the molecular markers for susceptibility to autoimmune diabetes [89]. Differential signatures (features) obtained from such analysis can also be used for classifying disease samples. Recently Swan et al. applied machine learning (ML)-based software named RGIFE (Rule-guided Iterative Feature Elimination) to identify biomarkers for osteoarthritis using both transcriptomics and proteomics datasets [97]. This algorithm was shown to be compatible with multiple data types and has better performance as it combines heuristic framework with accuracy from two different methods used in ML paradigm.

4.2 Integrating omics data through network biology approach

To gain a system-wide understanding of expression regulation, differentially regulated or deregulated genes from the transcriptome and the proteome datasets should be threaded via network biology approaches. Nodes (genes) and edges (connection between nodes) are the two basic components required to construct a gene expression network. Connection or relation between genes can be defined based on any kind of interaction data–Protein-protein interactions (PPIs), co-expression, genetic interactions, etc. Consequently, networks can be analysed from two different contexts–Network properties and network tools (based on knowledge derived annotations). Network properties or attributes (like size, connectedness between nodes, centrality and so on) can be used as the method to prioritize or rank in order to find common regulatory points [66]. However, network tools based on annotations from pathways, ontologies and other functional data (either literature curated or prediction methods) can find functional modules from data which is impacted in diseased phenotypes.

5 Concluding remarks

Integrative analyses of transcripts and proteins hold immense promise towards providing a better understanding of the gene regulation, genome annotation and the intricate biological processes that underlie any disease manifestations. Significant advancements in molecular profiling techniques as well as computational resources and data analytics have provided possibility to perform multivariate analysis for systems-level understanding from multi-dimensional data. While the dynamics of transcriptome and protein expression remains informative to gain insights into the biological processes, a system-wide understanding can be achieved by further integrating the other important components of biological systems. An ideal systems biology study therefore, would require an integration of all major forms of functional regulations, i.e., epigenome, genome, transcriptome, translatome, proteome and metabolome. Thus, integrating multiple layers of high-throughput omics data is an immediate necessity in

life sciences research. One application of such "multi-omics" analysis would be in the precision medicine aspects where a panel of markers identified from integrative analysis may provide a better predictability in the diagnosis or prognosis of a particular disease.

We thank Dr. Gajinder Pal Singh and Mr. Aniket Bhattacharya for insightful comments while proof-reading the manuscript. We acknowledge Dr. Pramod Gautam for his assistance in preparing the figures. Financial assistance from Council of Scientific and Industrial Research (CSIR) in the form of CSIR-Genesis project (BSC-0121) is duly acknowledged.

The authors have declared no conflict of interest.

6 References

[1] Nagalakshmi, U., Wang, Z., Waern, K., Shou, C. et al., The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 2008, *320*, 1344–1349.

[2] Wang, Z., Gerstein, M., Snyder, M., RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev.Genet.* 2009, *10*, 57–63.

[3] Ge, H., Walhout, A. J., Vidal, M., Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genet.* 2003, *19*, 551–560.

[4] Nilsen, T. W., Graveley, B. R., Expansion of the eukaryotic proteome by alternative splicing. *Nature* 2010, *463*, 457–463.

[5] Khoury, G. A., Baliban, R. C., Floudas, C. A., Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* 2011, *1*.

[6] Brivanlou, A. H., Darnell, J. E., Jr., Signal transduction and the control of gene expression. *Science* 2002, *295*, 813–818.

[7] Ozsolak, F., Milos, P. M., RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 2011, *12*, 87–98.

[8] The Genotype-Tissue Expression (GTEx), Project. *Nat. Genet.* 2013, *45*, 580–585.

[9] Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. et al., Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 2008, *5*, 621–628.

[10] Pathania, R., Ramachandran, S., Elangovan, S., Padia, R. et al., DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. *Nat. Commun.* 2015, *6*, 6910.

[11] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, *422*, 198–207.

[12] Steen, H., Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell. Biol.* 2004, *5*, 699–711.

[13] Kumar, D., Jain, A., Dash, D., Probing the missing human proteome: a computational perspective. *J. Proteome. Res.* 2015, 4949–4958.

[14] Yadav, A. K., Bhardwaj, G., Basak, T., Kumar, D. et al., A systematic analysis of eluted fraction of plasma post immunoaffinity depletion: implications in biomarker discovery. *PLoS One* 2011, *6*, e24442.

[15] Washburn, M. P., Wolters, D., Yates, J. R., III, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 2001, *19*, 242–247.

[16] Vogel, C., Marcotte, E. M., Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 2012, *13*, 227–232.

[17] Nesvizhskii, A. I., Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* 2014, *11*, 1114–1125.

[18] Eng, J. K., Searle, B. C., Clauser, K. R., Tabb, D. L., A face in the crowd: recognizing peptides through database search. *Mol. Cell. Proteomics* 2011, *10*, R111.009522.

[19] Ma, B., Zhang, K., Hendrie, C., Liang, C. et al., PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 2003, *17*, 2337–2342.

[20] Frank, A. M., Savitski, M. M., Nielsen, M. L., Zubarev, R. A. et al., De novo peptide sequencing and identification with precision mass spectrometry. *J. Proteome. Res.* 2007, *6*, 114–123.

[21] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551–3567.

[22] Yadav, A. K., Kumar, D., Dash, D., MassWiz: a novel scoring algorithm with target-decoy based analysis pipeline for tandem mass spectrometry. *J. Proteome. Res.* 2011, *10*, 2154–2160.

[23] Kumar, D., Yadav, A. K., Kadimi, P. K., Nagaraj, S. H. et al., Proteogenomic analysis of Bradyrhizobium japonicum USDA110 using GenoSuite, an automated multi-algorithmic pipeline. *Mol. Cell. Proteomics* 2013, *12*, 3388–3397.

[24] Kelkar, D. S., Provost, E., Chaerkady, R., Muthusamy, B. et al., Annotation of the zebrafish genome through an integrated transcriptomic and proteomic analysis. *Mol. Cell. Proteomics* 2014, *13*, 3184–3198.

[25] Yates, J. R., III, Eng, J. K., McCormack, A. L., Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal. Chem.* 1995, *67*, 3202–3210.

[26] Jaffe, J. D., Berg, H. C., Church, G. M., Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* 2004, *4*, 59–77.

[27] Renuse, S., Chaerkady, R., Pandey, A., Proteogenomics. *Proteomics* 2011, *11*, 620–630.

[28] Kelkar, D. S., Kumar, D., Kumar, P., Balakrishnan, L. et al., Proteogenomic analysis of Mycobacterium tuberculosis by high resolution mass spectrometry. *Mol.Cell. Proteomics* 2011, *10*, M111.011445.

[29] Kumar, D., Mondal, A. K., Kutum, R., Dash, D., Proteogenomics of rare taxonomic phyla: a prospective treasure trove of protein coding genes. *Proteomics* 2016, *16*, 226–240.

[30] Kumar, D., Mondal, A. K., Yadav, A. K., Dash, D., Discovery of rare protein-coding genes in model methylotroph Methylobacterium extorquens AM1. *Proteomics* 2014, *14*, 2790–2794.

[31] Risk, B. A., Spitzer, W. J., Giddings, M. C., Peppy: proteogenomic search software. *J. Proteome Res.* 2013, *12*, 3019–3025.

[32] Castellana, N., Bafna, V., Proteogenomics to discover the full coding content of genomes: a computational perspective. *J. Proteomics* 2010, *73*, 2124–2135.

[33] Kumar, D., Yadav, A. K., Jia, X., Mulvenna, J. et al., Integrated transcriptomic-proteomic analysis using a proteogenomic workflow refines rat genome annotation. *Mol. Cell. Proteomics* 2016, *15*, 329–339.

[34] Zhang, B., Wang, J., Wang, X., Zhu, J. et al., Proteogenomic characterization of human colon and rectal cancer. *Nature* 2014, *513*, 382–387.

[35] Nagaraj, S. H., Waddell, N., Madugundu, A. K., Wood, S. et al., PGTools: a software suite for proteogenomic data analysis and visualization. *J. Proteome. Res.* 2015, *14*, 2255–2266.

[36] Brinkman, D. L., Jia, X., Potriquet, J., Kumar, D. et al., Transcriptome and venom proteome of the box jellyfish Chironex fleckeri. *BMC Genomics* 2015, *16*, 407.

[37] Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S. et al., A draft map of the human proteome. *Nature* 2014, *509*, 575–581.

[38] Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas, Gholami A. et al., Mass-spectrometry-based draft of the human proteome. *Nature* 2014, *509*, 582–587.

[39] Serang, O., Kall, L., Solution to statistical challenges in proteomics is more statistics, not less. *J. Proteome Res.* 2015, *14*, 4099–4103.

[40] Ezkurdia, I., Vazquez, J., Valencia, A., Tress, M., Analyzing the first drafts of the human proteome. *J. Proteome Res.* 2014, *13*, 3854–3855.

[41] Violette, A., Biass, D., Dutertre, S., Koua, D. et al., Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach. *J. Proteomics* 2012, *75*, 5215–5225.

[42] Dutertre, S., Jin, A. H., Kaas, Q., Jones, A. et al., Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteomics* 2013, *12*, 312–329.

[43] Zhu, X., Xie, S., Armengaud, J., Xie, W. et al., Tissue-specific proteogenomic analysis of Plutella xylostella larval midgut using a multi-algorithm pipeline. *Mol. Cell. Proteomics* 2016, 1791–1807.

[44] Ruggles, K. V., Tang, Z., Wang, X., Grover, H. et al., An analysis of the sensitivity of proteogenomic mapping of somatic mutations and novel splicing events in cancer. *Mol. Cell. Proteomics* 2016, *15*, 1060–1071.

[45] Wang, X., Zhang, B., customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics* 2013, *29*, 3235–3237.

[46] Woo, S., Cha, S. W., Merrihew, G., He, Y. et al., Proteogenomic database construction driven from large scale RNA-seq data. *J. Proteome. Res.* 2014, *13*, 21–28.

[47] Zickmann, F., Renard, B. Y., MSProGene: integrative proteogenomics beyond six-frames and single nucleotide polymorphisms. *Bioinformatics* 2015, *31*, i106–i115.

[48] Castellana, N. E., Shen, Z., He, Y., Walley, J. W. et al., An automated proteogenomic method uses mass spectrometry to reveal novel genes in Zea mays. *Mol. Cell. Proteomics* 2014, *13*, 157–167.

[49] Krasnov, G. S., Dmitriev, A. A., Kudryavtseva, A. V., Shargunov, A. V. et al., PPLine: an automated pipeline for SNP, SAP, and splice variant detection in the context of proteogenomics. *J. Proteome Res.* 2015, *14*, 3729–3737.

[50] Sheynkman, G. M., Johnson, J. E., Jagtap, P. D., Shortreed, M. R. et al., Using galaxy-P to leverage RNA-Seq for the discovery of novel protein variations. *BMC Genomics* 2014, *15*, 703.

[51] Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A. et al., The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* 2007, *6*, 1638–1655.

[52] Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., Weissman, J. S., Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009, *324*, 218–223.

[53] Ingolia, N. T., Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* 2014, *15*, 205–213.

[54] Menschaert, G., Van, Criekinge W., Notelaers, T., Koch, A. et al., Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. *Mol. Cell. Proteomics* 2013, *12*, 1780–1790.

[55] Koch, A., Gawron, D., Steyaert, S., Ndah, E. et al., A proteogenomics approach integrating proteomics and ribosome profiling increases the efficiency of protein identification and enables the discovery of alternative translation start sites. *Proteomics* 2014, *14*, 2688–2698.

[56] Crappe, J., Ndah, E., Koch, A., Steyaert, S. et al., PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. *Nucleic Acids Res.* 2015, *43*, e29.

[57] Edgar, R., Domrachev, M., Lash, A. E., Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002, *30*, 207–210.

[58] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A. et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 2014, *32*, 223–226.

[59] Vizcaino, J. A., Cote, R., Reisinger, F., Barsnes, H. et al., The Proteomics Identifications database: 2010 update. *Nucleic Acids Res.* 2010, *38*, D736–D742.

[60] Maier, T., Guell, M., Serrano, L., Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* 2009, *583*, 3966–3973.

[61] Gygi, S. P., Rochon, Y., Franza, B. R., Aebersold, R., Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 1999, *19*, 1720–1730.

[62] Schwanhausser, B., Busse, D., Li, N., Dittmar, G. et al., Global quantification of mammalian gene expression control. *Nature* 2011, *473*, 337–342.

[63] de Sousa, Abreu R., Penalva, L. O., Marcotte, E. M., Vogel, C., Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* 2009, *5*, 1512–1526.

[64] Arava, Y., Wang, Y., Storey, J. D., Liu, C. L. et al., Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* 2003, *100*, 3889–3894.

[65] Delmotte, N., Ahrens, C. H., Knief, C., Qeli, E. et al., An integrated proteomics and transcriptomics reference data set provides new insights into the Bradyrhizobium japonicum bacteroid metabolism in soybean root nodules. *Proteomics* 2010, *10*, 1391–1400.

[66] Piruzian, E., Bruskin, S., Ishkin, A., Abdeev, R. et al., Integrated network analysis of transcriptomic and proteomic data in psoriasis. *BMC Syst. Biol.* 2010, *4*, 41.

[67] Torres-Garcia, W., Zhang, W., Runger, G. C., Johnson, R. H. et al., Integrative analysis of transcriptomic and proteomic data of Desulfovibrio vulgaris: a non-linear model to predict abundance of undetected proteins. *Bioinformatics* 2009, *25*, 1905–1914.

[68] Li, F., Nie, L., Wu, G., Qiao, J., Zhang, W., Prediction and characterization of missing proteomic data in Desulfovibrio vulgaris. *Comp. Funct. Genomics* 2011, *2011*, 780973.

[69] Rogers, S., Girolami, M., Kolch, W., Waters, K. M. et al., Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. *Bioinformatics* 2008, *24*, 2894–2900.

[70] Durinck, S., Spellman, P. T., Birney, E., Huber, W., Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 2009, *4*, 1184–1191.

[71] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S. et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003, *13*, 2498–2504.

[72] Junker, B. H., Klukas, C., Schreiber, F., VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* 2006, *7*, 109.

[73] Symons, S., Zipplies, C., Battke, F., Nieselt, K., Integrative systems biology visualization with MAYDAY. *J. Integr. Bioinform.* 2010, *7*.

[74] Perco, P., Muhlberger, I., Mayer, G., Oberbauer, R. et al., Linking transcriptomic and proteomic data on the level of protein interaction networks. *Electrophoresis* 2010, *31*, 1780–1789.

[75] Imielinski, M., Cha, S., Rejtar, T., Richardson, E. A. et al., Integrated proteomic, transcriptomic, and biological network analysis of breast carcinoma reveals molecular features of tumorigenesis and clinical relapse. *Mol. Cell. Proteomics* 2012, *11*, M111.

[76] Liang, S., Fuhrman, S., Somogyi, R., Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.* 1998, *3*, 18–29.

[77] Guthke, R., Moller, U., Hoffmann, M., Thies, F. et al., Dynamic network reconstruction from gene expression data applied to immune response during bacterial infection. *Bioinformatics* 2005, *21*, 1626–1634.

[78] Kuo, T. C., Tian, T. F., Tseng, Y. J., 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. *BMC. Syst. Biol.* 2013, *7*, 64.

[79] Chaube, R., Absolute quantitation of post-translational modifications. *Front Chem.* 2014, *2*, 58.

[80] Farley, A. R., Link, A. J., Identification and quantification of protein posttranslational modifications. *Methods Enzymol.* 2009, *463*, 725–763.

[81] Olow, A., Chen, Z., Niedner, R. H., Wolf, D. M. et al., An atlas of the human kinome reveals the mutational landscape underlying dysregulated phosphorylation cascades in cancer. *Cancer Res.* 2016, *76*, 1733–1745.

[82] Affolter, A., Muller, M. F., Sommer, K., Stenzinger, A. et al., Targeting irradiation-induced mitogen-activated protein kinase activation in vitro and in an ex vivo model for human head and neck cancer. *Head Neck* 2016, *38*, E2049–E2067.

[83] Su, M., Huang, J., Li, J., Qin, X. et al., ZLM-7 exhibits anti-angiogenic effects via impaired endothelial cell function and blockade of VEGF/VEGFR-2 signaling. *Oncotarget* 2016.

[84] Basak, T., Vega-Montoto, L., Zimmerman, L. J., Tabb, D. L. et al., Comprehensive characterization of glycosylation and hydroxylation of basement membrane collagen IV by high-resolution mass spectrometry. *J. Proteome Res.* 2016, *15*, 245–258.

[85] Cesnik, A. J., Shortreed, M. R., Sheynkman, G. M., Frey, B. L., Smith, L. M., Human proteomic variation revealed by combining RNA-seq proteogenomics and global post-translational modification (G-PTM) search strategy. *J. Proteome Res.* 2016, *15*, 800–808.

[86] Onwuli, D. O., Beltran-Alvarez, P., An update on transcriptional and post-translational regulation of brain voltage-gated sodium channels. *Amino Acids* 2016, *48*, 641–651.

[87] Varemo, L., Scheele, C., Broholm, C., Mardinoglu, A. et al., Proteome- and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes. *Cell Rep.* 2015, *11*, 921–933.

[88] Meierhofer, D., Weidner, C., Sauer, S., Integrative analysis of transcriptomics, proteomics, and metabolomics data of white adipose and liver tissue of high-fat diet and rosiglitazone-treated insulin-resistant mice identified pathway alterations and molecular hubs. *J. Proteome Res.* 2014, *13*, 5592–5602.

[89] Gerling, I. C., Singh, S., Lenchik, N. I., Marshall, D. R. et al., New data analysis and mining approaches identify unique proteome and transcriptome markers of susceptibility to autoimmune diabetes. *Mol. Cell. Proteomics* 2006, *5*, 293–305.

[90] Azimzadeh, O., Sievert, W., Sarioglu, H., Merl-Pham, J. et al., Integrative proteomics and targeted transcriptomics analyses in cardiac endothelial cells unravel mechanisms of long-term radiation-induced vascular dysfunction. *J. Proteome Res.* 2015, *14*, 1203–1219.

[91] Shimwell, N. J., Bryan, R. T., Wei, W., James, N. D. et al., Combined proteome and transcriptome analyses for the

discovery of urinary biomarkers for urothelial carcinoma. *Br. J. Cancer* 2013, *108*, 1854–1861.

[92] Zheng, P. Z., Wang, K. K., Zhang, Q. Y., Huang, Q. H. et al., Systems analysis of transcriptome and proteome in retinoic acid/arsenic trioxide-induced cell differentiation/apoptosis of promyelocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 2005, *102*, 7653–7658.

[93] Maji, A., Misra, R., Kumar, Mondal A., Kumar, D. et al., Expression profiling of lymph nodes in tuberculosis patients reveal inflammatory milieu at site of infection. *Sci. Rep.* 2015, *5*, 15214.

[94] Villar, M., Ayllon, N., Alberdi, P., Moreno, A. et al., Integrated metabolomics, transcriptomics and proteomics identifies metabolic pathways affected by anaplasma phagocytophilum infection in tick cells. *Mol. Cell. Proteomics* 2015, *14*, 3154–3172.

[95] Tarun, A. S., Peng, X., Dumpit, R. F., Ogata, Y. et al., A combined transcriptome and proteome survey of malaria parasite liver stages. *Proc. Natl. Acad. Sci. U. S. A.* 2008, *105*, 305–310.

[96] Ori A., Toyama B. H., Bock T., Iskar M. et al., Integrated transcriptome and proteome analyses reveal organ-specific proteome deterioration in old rats. *Cell Systems* 2016, *1*, 224–237.

[97] Swan, A. L., Stekel, D. J., Hodgman, C., Allaway, D. et al., A machine learning heuristic to identify biologically relevant and minimal biomarker panels from omics data. *BMC Genomics* 2015, 16 Suppl *1*, S2.

