

Exploiting Interdata Relationships in Next-generation Proteomics Analysis

Authors

Burcu Vitrinel, Hiromi W. L. Koh, Funda Mujgan Kar, Shuvadeep Maity, Justin Rendleman, Hyungwon Choi, and Christine Vogel

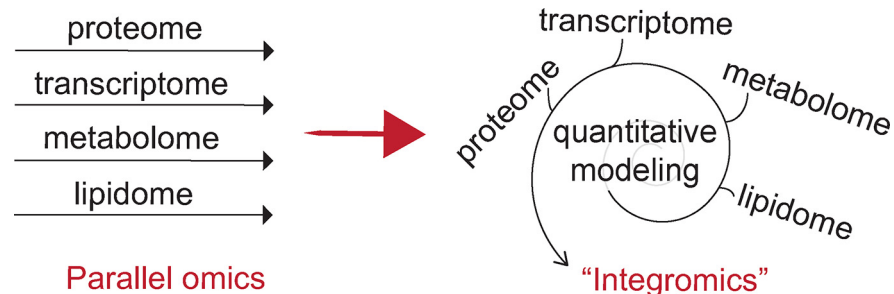
Correspondence

cvogel@nyu.edu

In Brief

Mass spectrometry-based proteomics and other technologies have matured to enable routine acquisition of system-wide data sets that describe concentrations, modifications, and interactions of proteins, mRNAs, and other molecules. Productive integrative studies differ from parallel data analysis by quantitative modeling of the relationships between data. We outline steps and considerations towards integromic studies to exploit the synergy between data sets.

Graphical Abstract



Highlights

- Technological advances produce many system-wide 'omics data sets.
- Integrating proteomics with other data needs quantitative modeling of inter-data relationships.
- Some tools have emerged for considering specific properties of 'omics data.
- Productive integration of proteomic and other data can provide exciting new insights.



Exploiting Interdata Relationships in Next-generation Proteomics Analysis*

✉ Burcu Vitrinel[‡]**, Hiromi W. L. Koh[§]¶**, Funda Mujgan Kar[‡], Shuvadeep Maity[‡], Justin Rendleman[‡], ✉ Hyungwon Choi[§]¶**, and ✉ Christine Vogel[‡]¶**

Mass spectrometry based proteomics and other technologies have matured to enable routine quantitative, system-wide analysis of concentrations, modifications, and interactions of proteins, mRNAs, and other molecules. These studies have allowed us to move toward a new field concerned with mining information from the combination of these orthogonal data sets, perhaps called “integromics.” We highlight examples of recent studies and tools that aim at relating proteomic information to mRNAs, genetic associations, and changes in small molecules and lipids. We argue that productive data integration differs from parallel acquisition and interpretation and should move toward quantitative modeling of the relationships between the data. These relationships might be expressed by temporal information retrieved from time series experiments, rate equations to model synthesis and degradation, or networks of causal, evolutionary, physical, and other interactions. We outline steps and considerations toward such integromic studies to exploit the synergy between data sets. *Molecular & Cellular Proteomics 18: S5–S14, 2019. DOI: 10.1074/mcp.MR118.001246.*

THE PAST AND THE PRESENT: DIRECTIONS IN PROTEOMIC DATA INTEGRATION

Recent large-scale analysis of protein concentrations, modifications, and interactions has seen tremendous advances, pushing us to consider the next steps in multiomics studies. Some of the new work lies “outside the box” of standard parallel mining of individual data sets and attempts to model the relationships between proteomic variations and other molecular changes to gain insights at their interface (Fig. 1). A new field might be born: “integromics.” Integromics studies have included information on the dynamics of mRNA and protein concentration changes, but also other molecules, such as lipids and metabolites, or completely orthogonal information on genomic variation across a population of samples. Because several excellent reviews discuss the relationship between protein and mRNAs, as well as proteogenomic approaches, e.g. (1–4), we will focus here on other new di-

rections that have emerged in the last few years, e.g. with respect to combination of proteomics with other technologies or other data types. We will also discuss components of such integrative analysis.

Integrating Proteomic and Transcriptomic Measurements

Correlating Protein and mRNA Concentrations—The first and long-debated question in integrative proteomic studies concerns the correlation between mRNA and protein concentrations in a steady-state system, i.e. in unperturbed cells that do not change over time. High correlation between mRNA and protein concentrations implies that transcription determines the cellular architecture; low correlation implies a dominant role for post-transcriptional regulation. For yeast and mammalian cells, estimates started to appear over ten years ago and differed considerably (5): although most studies agreed on substantial contribution of post-transcriptional regulation to the overall expression landscape (6, 7), some argued for a dominant role of transcription (8).

In 2012, we attempted to synthesize these findings into a common theme: transcription regulation might often act as an on-off switch, whereas translation and protein degradation fine-tune actual concentrations, like a rheostat (2). This two-step process attributes to the different response signals for different groups of genes (9). A 2015 study of bone-marrow derived dendritic cells exposed to lipopolysaccharide supported this view (10): indeed, most of the responses to the stimulus were initiated by RNA expression changes. In comparison, protein levels for housekeeping genes were also altered substantially when examining absolute molecule numbers. Because of the high protein concentrations, the resulting fold-changes remained comparatively small. We observed a similar trend in cancer cells responding to protein misfolding: protein concentration changes were much smaller in magnitude than mRNA expression changes (11). In addition, the transcriptome returned to pretreatment levels after ~12 h whereas proteins did not reach steady state at that time. Our

From the [‡]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY; [§]Department of Medicine, Yong Loo Lin School of Medicine, National University Singapore, Singapore; [¶]Institute of Molecular and Cell Biology, Agency for Science, Technology, and Research, Singapore

✂ Author's Choice—Final version open access under the terms of the Creative Commons CC-BY license.

Received November 28, 2018, and in revised form, May 1, 2019

Published, MCP Papers in Press, May 24, 2019, DOI 10.1074/mcp.MR118.001246

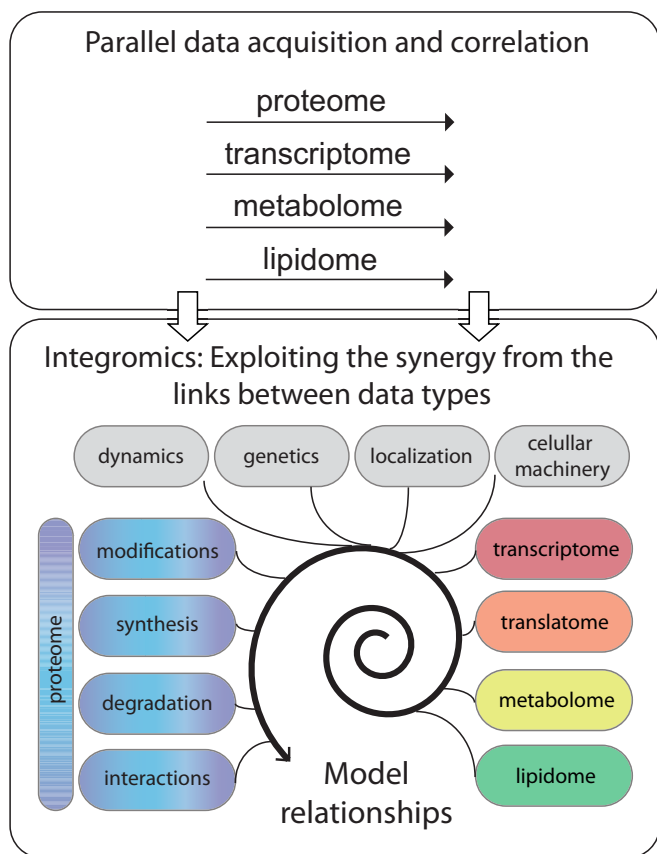


FIG. 1. Moving from multiomics studies that acquire and analyze data sets in parallel to modeling and exploiting the interactions between data.

most recent study suggests that the cell might implement such discordance observed between the transcriptome and the proteome in different ways, e.g. through gene specific increase in translation via short regulatory elements despite no transcription change or an increase in transcription but delayed translation (12).

Another possible explanation for the disparity between transcript and protein levels has been discussed for yeast undergoing meiosis (13). When comparing protein expression levels with those of transcripts and parallel ribosome footprinting data, the authors noticed impaired translation of several genes through new isoforms which they named “long undecoded transcript isoforms” (LUTIs)¹. They proposed that a single transcription factor can activate the canonical transcript for some genes and LUTI for others.

Finally, several theoretical studies highlighted the unexpected conservation of protein-RNA ratios across tissues (14, 15), and the fact that protein concentrations of orthologs

¹ The abbreviations used are: LUTI, long undecoded transcript isoforms; SILAC, stable isotope labeling of amino acids; QTL, quantitative tract loci; CETSA, cellular thermal shift assays; MKGI, meta-dimensional knowledge-driven genomic interactions; PECA, protein expression control analysis.

appear to be more conserved across organisms than mRNA concentrations (16–18). However, some of these observations are because of an effect like Simpson’s paradox (19): some relationships may become reversed or masked by the opposing effects of other data types. Therefore, orthologous protein concentrations might be correlated across genes, but not as much as some studies suggest (20, 21).

In addition to these insights into the relationship between protein and mRNA concentrations, the future might bring more integration of these paired data sets with additional information, such as the temporal changes in response to a stimulus, physical interactions between proteins, or measurements of synthesis and turnover rates. Examples of such new directions include the time-resolved studies described above, or recent work involving Down Syndrome patients (22). The study analyzed protein and mRNA concentrations in samples from identical twins where one twin is healthy and one has Down Syndrome. Thanks to careful integration of these data with protein stability measurements, the authors demonstrated the major role of degradation in maintaining stoichiometry in protein complexes, despite minor effects on overall mRNA or protein levels.

Proteoforms and Alternative Splice Variants—Another factor lowering RNA and protein correlations arises from alternative splicing and the production of protein isoforms, and many efforts in integrative proteomics concern the complete mapping of the resulting proteome diversity (13, 15, 23–25). Integrating data from mRNA sequencing and proteomics, Liu *et al.* attempted to map the entire human proteome with respect to its variants (26). The authors identified a significant contribution of alternative splicing to proteome composition and diversity, with respect to alternative translation initiation, alternative splicing, and post-translational modifications. However, these estimates are not uncontended - other studies suggest that the number of functional variants per protein might be very small (27–29). The reason for these discrepancies may lie in technical challenges to identify critical peptides that mark variants and isoforms (29) or in the fact that most proteins get expressed only one isoform per tissue (23). Therefore, identification of functional proteoforms and alternative splice variants remains a daunting task.

Combining Proteomics with other ‘Omics Data

Measurements of Translation—Several recent studies have moved beyond simple assessment of the relationships between concentrations and toward identification of the underlying processes that determine concentrations and concentration changes. One example is the twin study mentioned above that included examination of protein degradation through use of dynamic proteomics (22).

Other examples arise from the inclusion of sequencing data that identifies ribosome footprint positions along mRNAs, estimating translation efficiency and regulatory elements (30).

Several such studies exist and examined meiosis or the response to environmental stress (13, 15, 23–25, 31, 32). For example, when comparing genome-wide transcriptome, proteome, and ribosome profiles across diverse stresses, Ho *et al.* found that ribosomes appeared to dissociate from some transcripts, delaying their translation into the corresponding proteins (32). The authors suggested that this process frees ribosomes which could then be used toward the synthesis of stress response proteins.

However, importantly, the association of ribosomes with mRNAs may not always reflect actual translation output, as can be measured by proteomics methods such as pulsed SILAC (stable isotope labeling of amino acids) (33). Ribosomes might attach to mRNAs leading to reported footprints, but not actively translate and produce protein. Such discrepancy was observed for data from multiple myeloma cells (34). Under unperturbed conditions, ribosome footprinting and pulsed-SILAC translation measurements largely correlated across genes. However, when the cells were perturbed through inhibition of protein degradation, the correlation vanished: pulsed-SILAC was able to detect global alterations in translation rates across genes, whereas ribosome footprinting failed to do so. Therefore, albeit proteomics methods often pose technical challenges and result in smaller coverage than sequencing methods, some biological questions might demand the use of proteomics for translation measurements over assessment of ribosome footprints.

Genetic Association and Quantitative Trait Loci—An entirely orthogonal area of proteomic data integration lies in their use as molecular phenotypes that are then associated with specific genomic regions to discover Quantitative Trait Loci (QTL). Associations with mRNA expression phenotypes are typically called eQTLs, whereas associations with protein expression render pQTLs. The relationship between eQTLs and pQTLs is complex and still only incompletely understood (35, 36). For example, in yeast, the genomic position of eQTLs and pQTLs seems to overlap only little (37), and *cis* regulation is common for eQTLs but not at the level of the proteome (38). One important role of pQTLs appear to be maintenance of the stoichiometry of protein complexes and pathways (39).

More recent studies in blood plasma cells confirmed the complexity of the relationship between pQTLs and eQTLs. For example, they detected several pQTLs that affected protein levels in *trans*, illustrating how pQTLs can identify effects hidden at the mRNA level (40). Other studies found several pQTLs acting in *cis* and reported substantial overlap between pQTLs and *cis*-eQTLs—contrasting what had been observed before (41–43).

Post-translational Modifications—Protocols to measure post-translational modifications such as phosphorylation, ubiquitination, and SUMOylation are now readily available for routine use. Recent work integrated such measurements with other 'omics data, *i.e.* protein concentrations. For example, a study in mice showed substantial overlap between protein

abundance and phosphorylation levels but revealed differences in the temporal patterns upon induction of a high-fat diet (44). A similar observation was made in samples from breast cancer patients: the phosphoproteome grouped into clusters that were undetectable at the mRNA or protein level, illustrating the need for collecting multiple data types (45).

However, it is crucial to move beyond simple parallel analysis to models that attempt to reveal and exploit relationships between the data (46, 47). Such analyses are necessary to understand causal relationships, *e.g.* the role of multiple, often successive protein modifications in signal transduction cascades. A step toward such analyses arises from a study of phosphorylation, acetylation, and methylation events across 45 untreated and treated lung cancer cell lines (48). Using machine learning and a comprehensive protein-protein interaction network, the authors found many multimodification events that acted in a mutually exclusive pattern: the protein had either one type of modification or the other, but not both. Such pattern suggests that protein modifications are used to direct signaling pathways into different routes with an “exclusive OR” gate depending on the type of modification.

Integrating Proteomics with Other Technologies—Another type of integrative developments uses mass spectrometry in conjunction with other technologies in form of a new method. Such analyses might not be strictly “integrative” yet, as they typically acquire only one data type. However, they might inspire and encourage proteomicists to venture more frequently into new territory to map novel aspects of biology.

For example, recent work combined proteomics with polysome profiling, a technique that exploits differential sedimentation of ribosome-bound mRNAs and the unbound the small and large subunits in a sucrose density gradient (49). The analyses examined mammalian cells during mitosis to identify changes in ribosome composition and function (50), as well as phosphorylation changes (51). Although the authors could not confirm earlier findings on varying composition of the ribosome core (52), they found extensive differential protein phosphorylation across the polysome profile leading to identification of a new regulatory phosphorylation event on a ribosome subunit.

Proteomics has also been combined with cellular thermal shift assays (CETSA) to monitor the thermal stability of the proteome (53). The method exploits the fact that, depending on their structural properties, proteins “melt” at higher temperatures and collapse (54). Integrating such stability measurements with estimates of protein abundance and solubility, the study found that many intrinsically disordered and mitotically phosphorylated proteins were stabilized and more solubilized during mitosis, suggesting a fundamental remodeling of the biophysical environment during the cell cycle. A more recent study using CETSA examined a similar system and found very little proteome remodeling during mitosis, but substantial changes in protein-protein interactions (55)—empha-

sizing the importance to look beyond simple concentration measurements for new discoveries.

A third expansion of traditional proteomics employs limited proteolytic digest prior to mass spectrometry, providing an indirect readout of a protein's structural stability through identification of peptides that are accessible to the protease (56, 57). Leuenerger *et al.* used this technique with samples processed at different temperatures to estimate thermal melting points of proteins from *Escherichia coli*, *Saccharomyces cerevisiae*, *Thermus thermophilus*, and human cells (58). Their results confirmed the complexity of the relationship between protein concentrations and stability and the need for a multifaceted analysis: although highly expressed proteins were particularly stable, a specific subset of proteins was destabilized by higher temperatures leading to cellular collapse.

Metabolomics and Lipid Measurements—Besides nucleic acids, the cell contains many other molecules with various functions—and proteomic data integration has made big strides toward understanding them. Such studies have, for example, combined limited proteolysis-based proteomics with screening of small molecules for their effect on protein folding (59). The authors found ~140 new proteins that appeared to bind ATP, and many cases in which ATP or other metabolites affected the structure and protease accessibility of the protein. The role of ATP beyond being an energy source was also confirmed by use of CETSA in combination with proteomics: ATP-binding membrane proteins shifted in their sensitivity to detergents, suggesting that the molecule stabilizes the protein structure and increases solubility (60).

Other examples support this relationship between metabolism and the proteome. In *Drosophila* cells, measurements and calculations of phase differences revealed a link between oscillations of protein concentrations and their downstream metabolites (61). In other studies, integration of proteome, metabolome, and lipidome measurements revealed new functions of genes in mitochondrial coenzyme Q biosynthesis and regulation (62, 63). Finally, combining such measurements with genomic data defined lipid-QTLs like those discussed for mRNAs, protein, and translation discussed above. Indeed, lipid-QTLs differed substantially from protein QTLs as discovered by analysis of a BXD mouse population (64, 65). These findings were enabled through combination of multiple 'omics measurements, but analysis is currently restricted to mostly correlative observations. The next and exciting challenge lies in identifying causal relationships behind molecules and pathways.

New Tools and Techniques for Integrative Analysis

General Tools—The growing demand for effective data integration has spurred the development of numerous computational tools and approaches. Several recent reviews comprehensively summarized these advances (66, 67). In addition, integrative tools have begun to emerge to model multiomics

data including proteomics data (68). From a biologist's point of view, the major challenge in integrating multiple 'omics data may lie not in lacking availability, but rather in the selection of appropriate tools for a given research question. As the principal aims of the tools are diverse, ranging from the discovery of a sparse set of data features associated with a phenotype of interest to the integrated clustering of samples, finding the right tool can be demanding (Fig. 2). Therefore, no single computational tool can provide solutions for every problem, and successful application depends on the user's proper understanding of the biological question and the functionalities of each tool.

Visualization—An important starting point for successful integration is the visualization of integrated data to provide a holistic view of merged data. In the past, multiomics analyses typically used a series of heatmaps. However, heatmaps lose their merit with increasing data size, and they fail to show the connections between different types of molecules. Therefore, several tools have been designed to show networks of inter-feature relationships. One example is the 3Omics tool in which biological networks are created based on a coexpression analysis of multiomics data and the resulting networks are displayed (69). More recently, OmicsNet has been developed to simultaneously map multiple 'omics data onto a network consisting of protein-protein interactions, metabolic relationships and co-expression information (70). The tool creates a global view of the network ensemble with flexible options for customization. Its visualization is unique as it not only preserves a modular network layout within each 'omics data but also shows the connections between different layers.

Data Interpretation—Another key to successful data integration is finding tools that can effectively harmonize the biological information across heterogeneous platforms and facilitate biological interpretation of merged findings. Many statistical methods already exist that can integrate multiple data sets to perform sample clustering or classification (71). In many cases, however, these tools treat different 'omics data as equally contributing features and ignore biological relationships between different types of molecules. For example, a recently published MultiOmics Factor Analysis tool (MOFA) is an unsupervised method for inference of a set of latent factors which can best capture sources of variabilities across the different 'omics data sets using a probabilistic Bayesian framework (72). However, as there may be no biological relationship between the features from different data types with high loading scores, the latent factors identified may not always necessarily be interpretable.

In contrast, computational methods that directly derive molecular interactions or communities of molecules from the correlation structures often yield more biological interpretable results. For example, a tool called Meta-dimensional Knowledge-driven Genomic Interactions (MKGI) maps molecular abundance measurements in each 'omics data set to the dimension of pathways and builds an ensemble of models to

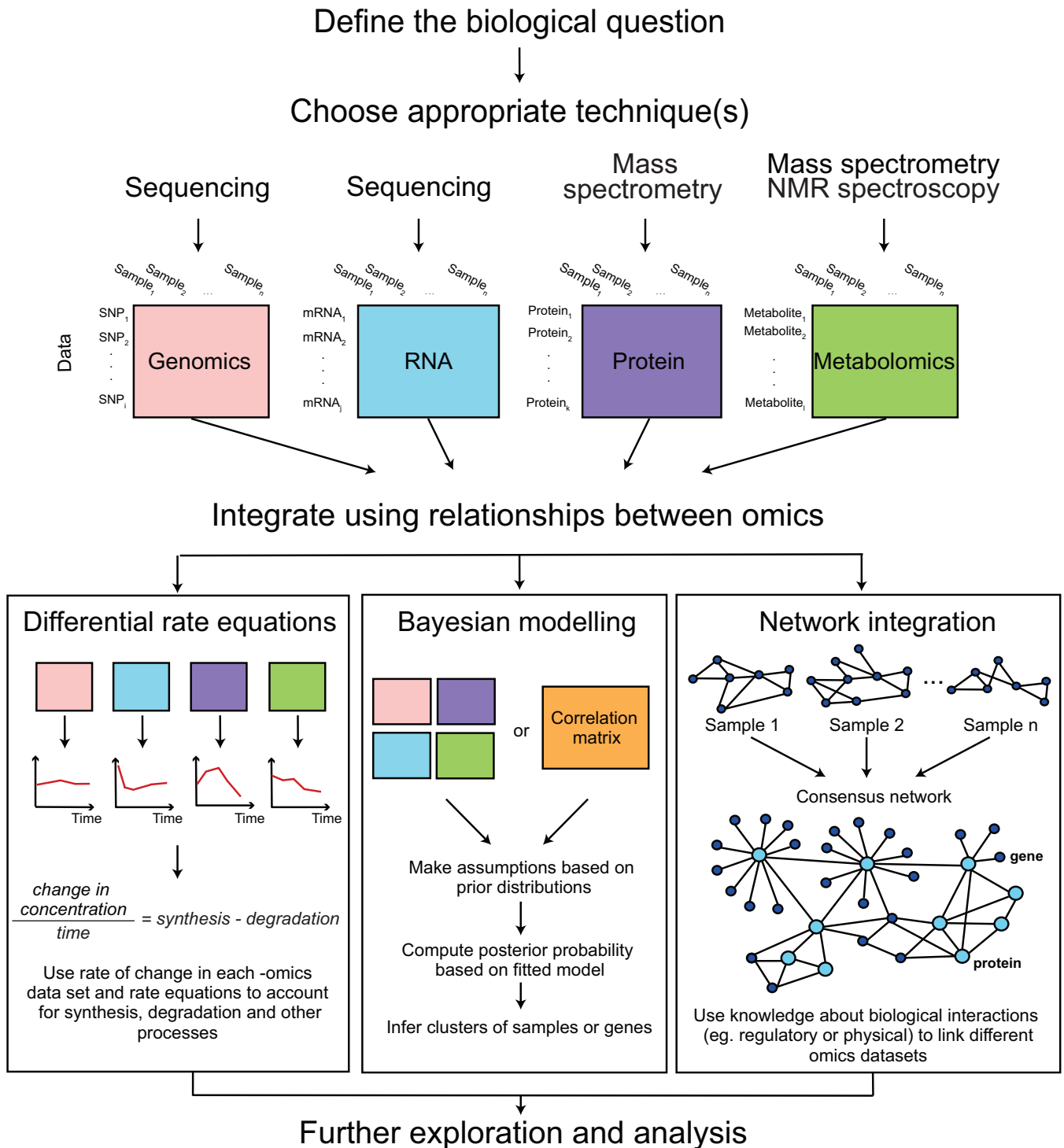


FIG. 2. Steps toward productive integromics.

predict a phenotypic outcome via a neural network algorithm (73). Another tool called the OmicsIntegrator finds putative pathways using a network optimization algorithm and derives a subnetwork of the multiomics signature that best explains the expression data sets (74). Using a similar approach, the inteGREAT tool discovers co-expression networks within the

transcriptome and the proteome and performs network analysis of differentially abundant genes, e.g. clinical biomarkers, incorporating network topology information for each gene from both 'omics levels (75).

Accounting for Specific Properties of Proteomic Data— Many integration tools treat proteomics data as if it is just

another dimension in the 'omics repertoire, ignoring the fact that protein abundance and other quantitative features, e.g. post-translational modifications, are the ultimate output in gene expression regulation. Nevertheless, those tools do not necessarily inform on the underlying processes and determinants of the output, such as translation or degradation. Examining changes in protein numbers in a static condition does not necessarily stem from a shift in translation, as such a change could also be caused by altered protein degradation, localization or even formation of isoforms. Although some tools such as PARADIGM reflect the hierarchical nature and directionality among data from different 'omics platforms (76), incorporation of inter-omics relationships in statistical modeling has been mostly overlooked. In addition, parallel measurements of concentration changes and the underlying processes are still very rare, restricted to the few examples described above.

Further, as illustrated above for mammalian cells responding to an outside stimulus (10–12), gene expression control often results in large fold-changes in the transcriptome, but only small changes in the proteins. Such properties can skew joint analysis of the data in favor of the part of the data showing changes of a greater magnitude, e.g. mRNA expression. As such, integration of data across different experimental platforms requires appropriate normalization of data, e.g. transformation of data into scores on a standardized scale before performing joint analysis. Also, one can consider weighting the data sources with quality or informativity scores so that each 'omics data type can contribute to the analysis in equal proportions.

For some specific biological questions, such tools and techniques already exist and enable the user to normalize the data, estimate statistical parameters of interest, and interpret the results. For example, we and others have developed models to estimate protein synthesis and degradation from dynamic data, e.g. for the response to lipopolysaccharide or ER stress (10–12). In these cases, the relationship between the two paired data types, protein and mRNA concentrations, is a direct result of the central dogma of biology, and rate equations serve to quantify the parameters. The analyses provide precise estimates of significant regulatory events and time points.

We recently expanded this approach to provide a comprehensive tool to infer different regulatory parameters from any dual-omics time series experiment (77). The tool is called Protein Expression Control Analysis (PECA) and consists of autoregressive sub-models, each mimicking a first-order ordinary differential equation of abundance dynamics for a protein. The key integrative aspect of the approach comes from the incorporation of time series measurements and a flexible statistical method to estimate time-varying rate parameters with assessment of the noise in the data (false discovery rates). Indeed, we successfully used PECA to parse time-series transcriptomics measurements or even evaluate ribo-

some binding and dissociation with mRNA (12). In this case, significant changes in association of ribosomes with mRNAs as reported by PECA served as a readout for changes in translation of the respective gene. PECA is easily accessible as a plugin for the PERSEUS toolbox which allows for visualization, normalization, and interpretation of proteomics and other large-scale data (78).

Future Directions: How to Exploit the Synergy of Productive Integrative Work

Steps Toward Productive Integromics—“Integromics” is an emerging form of proteomic analysis that exploits large-scale measurements of concentrations, modifications, and interactions of proteins with themselves and other molecules. The new challenges lie in experts from individual disciplines to become “multilingual” to understand the properties, information types, and limitations of data from other domains. Such multidirectional understanding is necessary to model the relationships between the data and exploit the information gained (Fig. 1).

What might be crucial steps toward such integrative work? As for all research, the first step should lie in defining the biological question, and all subsequent steps should always go back to this question (Fig. 2). As an integrative systems biologist, it will be important to step outside one's comfort zone and specific research field. True discovery might lie in exploration of new areas and techniques, such as the combination of proteomics with genetic information, the metabolome, the lipidome, translation measures or cellular melting, as discussed with the examples above.

The second step involves consideration as to what type of information is needed and what are the appropriate technologies (Fig. 2). If a question can be answered by studying just one molecular type, then there is no use in acquiring other data. For example, it is tempting to combine genome-wide analysis of RNA concentrations with the measurements of corresponding proteins. However, proteomics is typically more labor-intensive, assesses only a fraction of the expressed genome, and delivers noisier measurements. Therefore, mapping the proteome might only be worthwhile if there is evidence for regulation beyond what is captured by the transcriptome—warranting integrative analyses that involve multiple data types. For example, the response to protein misfolding stress is well-known to affect not only transcription, but also translation and RNA and protein degradation, therefore such integrative analyses are warranted.

Further, the choice of technique should involve consideration of the type of information to be gained. For example, proteomics techniques that provide absolute concentration estimates, such as parts-per-million or molecules per cell, are different from those estimating fold-changes in a specific condition *versus* a control. The former can be obtained by label-free quantitation or use of tags such as iTRAQ or TMT,

the latter can be obtained from SILAC experiments. For example, when discussing the correlation between mRNA and protein concentrations, one needs to acquire absolute concentration estimates (1). When aiming to analyze rates of synthesis and degradation, time-resolved measurements of either absolute concentrations or fold-changes work, but the two different data types require different modeling (7, 10, 77).

In addition, the choice of adequate technique depends on the method's limitations. Proteomics is always biased toward peptides that are amenable to proteolytic digestion, solubility, and mass spectrometry. Therefore, a hunt for new splice variants may require use of alternative proteases to increase the proteome space. Further, ribosome footprinting provides genome-scale estimates of ribosome locations along mRNAs, but only indirect measures of translation efficiency. If the biological question asks for translation rates or evidence exist for extensive stalling of ribosomes without producing proteins, like for example during inhibition of the proteasome (79), ribosome footprinting may not be the method of choice.

A third step in integrative analyses lies in the careful planning of the experiment, keeping the anticipated statistical model in mind (Fig. 2). It should include considerations as to the number of genes that are involved and the number of biological or technical replicates to ensure enough statistical power. For multiomics experiments, it is important to ensure that the measurements are matched between techniques with respect to time points, conditions, and genes. If not, then how can the differences in experimental conditions be accounted for in the statistical model?

After data acquisition and normalization, the fourth step warrants the appropriate modeling of the relationship between the proteomics data and the other data types (Fig. 2). These relationships are the key to extracting information beyond what is gained from parallel analysis of individual data sets. As discussed above for tools, such modeling should include the hierarchical relationship between proteomic and other information, reflecting the proteins' position in the cellular system. Because "integromics" is a new field, there is no one type of recommendation yet to model such relationships; many different tools exist.

For example, in some cases the relationships can be expressed by rate equations, such as has been done for dynamic analyses of protein and RNA concentration changes in response to treatment with lipopolysaccharide or ER stressors (10, 12) (Fig. 2): changes in protein concentrations are expressed as a function of translation based on the mRNA concentration and protein degradation. In other cases, the relationships can be modeled in form of a network. An example of such studies examines oligodendrocytes in the context of Alzheimer's disease (80). Using transcriptomic and proteomic data, genome-wide association data, a protein-protein interaction network, and Bayesian analysis, the study identified Cnp as a main regulator that control the aspects of myelin and mitochondrial gene expression dysregulation. The rela-

tionship between proteins was modeled through known physical interactions that are part of signaling pathways.

A fifth and perhaps most enjoyable step of integrative analysis lies in letting go of all the restrictions mentioned above and play (Fig. 2). The benefit of large-scale studies lies in unbiased description of connections that can discover entirely unexpected and new patterns. Key to such exploration is visualization of the data in various ways, and several helpful techniques and tools are discussed above. Of course, an analysis should be guided with the original biological question in mind - but because the data commonly covers hundreds to thousands of genes, and because the relationships are usually based on general models and not manually curated information, it is quite possible to move beyond what is known and unravel new links. For example, in our own work, a large number of translationally upregulated genes during the ER stress response had been observed before, but after mapping these changes to different pathways of the mitochondrial energy metabolism, we discovered an entirely new trend: translation upregulation seemed to support a shift in energy metabolism from the tricarboxylic acid cycle to one-carbon metabolism (12).

Future Directions—Proteomic analysis and systems biology in general have exciting times ahead. We predict that future work will expand existing approaches in multiple ways (Fig. 1). Such studies might include actual experimental measurements of rates of transcription, translation, RNA and protein degradation all in parallel—to complement what has been modeled based on concentration measurements. Methods to measure these processes exist already, but their integration is still rare. Apart from computational models (81) or steady-state analysis (7), there is no study in which all four major synthesis and degradation rates have been measured simultaneously in one dynamic system. Results from such work would inform us on the relationship between the different processes, feedback and coupling between them, or different regulatory goals for different genes.

Future work might also involve more analyses of multiple modifications per protein or peptide—moving away from studying one post-translational modification at a time. When including temporal information, it will be possible to extract hypotheses on causal relationships, e.g. resolve which phosphorylation might trigger subsequent SUMOylation or ubiquitination on a specific protein. Such sequential modifications have been described for individual pathways, e.g. FEN1 (82) but are unknown with respect to the global cellular role.

Future integrative proteomics may also involve increased integration with other molecule types, such as those of the metabolome. Like the examples described above, such analyses will expand our knowledge of both binding sites in protein structures but also the impact of metabolites on protein stability and function (40, 54). To help such new analyses, we foresee more and more studies in which proteomics is com-

bined with other technologies, such as the thermal profiling or polysome profiling mentioned above (52, 54).

And finally, future integrative proteomics might transcend into entirely new areas, e.g. the analysis of very small samples or single cells (83–87). Small sample analyses can reveal local relationships between cells in a tissue, as has been shown for HeLa cells (88). Such information could open many new applications ranging from analysis of rare cell populations to those of limited clinical specimens. Single cell analysis will also expand our views on the relationship between protein and mRNA concentrations (83), as has been discussed for bacteria many years ago (89).

In our view, the essential components to such integrative work will remain not only in combination of proteomics with other techniques and data types, but also the careful and quantitative modeling of the relationships between data as new findings lie in these relationships. Nothing in biology works in an isolated manner: protein changes are both impacted by various processes, but also affect multiple processes in the cell. The time is ripe to move toward the next level of systems biology and consider processes in combination, gaining new insights at their interface.

*The work was supported by the NIH/NIGMS grant 1R35GM127089-01 (to C.V.) and Singapore Ministry of Education grant MOE2016-T2-1-001 (to H.C.). B.V. acknowledges funding by American Heart Association grant 18PRE33990254.

** These authors contributed equally to this work.

|| To whom correspondence should be addressed. E-mail: cvogel@nyu.edu.

Author contributions: B.V., H.W.K., F.M.K., S.M., J.R., H.C., and C.V. analyzed data; B.V., H.W.K., F.M.K., S.M., J.R., H.C., and C.V. wrote the paper.

REFERENCES

- Liu, Y., Beyer, A., and Aebersold, R. (2016) On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550
- Vogel, C., and Marcotte, E. M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232
- Ruggles, K. V., Krug, K., Wang, X., Clauser, K. R., Wang, J., Payne, S. H., Fenyő, D., Zhang, B., and Mani, D. R. (2017) Methods, tools and current perspectives in proteogenomics. *Mol. Cell. Proteomics* **16**, 959–981
- Rodriguez, H., and Pennington, S. R. (2018) Revolutionizing precision oncology through collaborative proteogenomics and data sharing. *Cell* **173**, 535–539
- de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M., and Vogel, C. (2009) Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* **5**, 1512–1526
- Vogel, C., de Sousa Abreu, R., Ko, D., Le, S.-Y., Shapiro, B. A., Burns, S. C., Sandhu, D., Boutz, D. R., Marcotte, E. M., and Penalva, L. O. (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.* **6**, 400
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature* **473**, 337–342
- Li, J. J., Bickel, P. J., and Biggin, M. D. (2014) System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2**, e270
- McManus, J., Cheng, Z., and Vogel, C. (2015) Next-generation analysis of gene expression regulation – comparing the roles of synthesis and degradation. *Mol. Biosyst.* **11**, 2680–2689
- Jovanovic, M., Rooney, M. S., Mertins, P., Przybylski, D., Chevrier, N., Satija, R., Rodriguez, E. H., Fields, A. P., Schwartz, S., Raychowdhury, R., Mumbach, M. R., Eisenhaure, T., Rabani, M., Gennert, D., Lu, D., Delorey, T., Weissman, J. S., Carr, S. A., Hacohen, N., and Regev, A. (2015) Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038
- Cheng, Z., Teo, G., Krueger, S., Rock, T. M., Koh, H. W. L., Choi, H., and Vogel, C. (2016) Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. *Mol. Syst. Biol.* **12**, 855
- Rendleman, J., Cheng, Z., Maity, S., Kastelic, N., Munschauer, M., Allgoewer, K., Teo, G., Zhang, Y. B. M., Lei, A., Parker, B., Landthaler, M., Freeberg, L., Kuersten, S., Choi, H., and Vogel, C. (2018) New insights into the cellular temporal response to proteostatic stress. *Elife* **7**, e39054
- Cheng, Z., Otto, G. M., Powers, E. N., Keskin, A., Mertins, P., Carr, S. A., Jovanovic, M., and Brar, G. A. (2018) Pervasive, coordinated protein-level changes driven by transcript isoform switching during meiosis. *Cell* **172**, 910–923.e16
- Edfors, F., Danielsson, F., Hallström, B. M., Käll, L., Lundberg, E., Pontén, F., Forsström, B., and Uhlén, M. (2016) Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol. Syst. Biol.* **12**, 883
- Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmaier, A., Faerber, F., and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587
- Khan, Z., Ford, M. J., Cusanovich, D. A., Mitran, A., Pritchard, J. K., and Gillad, Y. (2013) Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science* **342**, 1100–1104
- Laurent, J. M., Vogel, C., Kwon, T., Craig, S. A., Boutz, D. R., Huse, H. K., Nozue, K., Walia, H., Whiteley, M., Ronald, P. C., and Marcotte, E. M. (2010) Protein abundances are more conserved than mRNA abundances across diverse taxa. *Proteomics* **10**, 4209–4212
- Schrimpf, S. P., Weiss, M., Reiter, L., Ahrens, C. H., Jovanovic, M., Malmström, J., Brunner, E., Mohanty, S., Lercher, M. J., Hunziker, P. E., Aebersold, R., von Mering, C., and Hengartner, M. O. (2009) Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biol.* **7**, e48
- Friendly, M., Monette, G., and Fox, J. (2013) Elliptical insights: understanding statistical methods through elliptical geometry. *Stat. Sci.* **28**, 1–39
- Fortelny, N., Overall, C. M., Pavlidis, P., and Cohen Freue, G. V. (2017) Can we predict protein from mRNA levels? *Nature* **547**, E19–E20
- Franks, A., Airolidi, E., and Slavov, N. (2017) Post-transcriptional regulation across human tissues. *PLoS Comput. Biol.* **13**, e1005535
- Liu, Y., Borel, C., Li, L., Müller, T., Williams, E. G., Germain, P.-L., Buljan, M., Sajic, T., Boersema, P. J., Shao, W., Faini, M., Testa, G., Beyer, A., Antonarakis, S. E., and Aebersold, R. (2017) Systematic proteome and proteostasis profiling in human Trisomy 21 fibroblast cells. *Nat. Commun.* **8**, 1212
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szijarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., and Pontén, F. (2015) Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419
- Omenn, G. S., Lane, L., Lundberg, E. K., Overall, C. M., and Deutsch, E. W. (2017) Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project. *J. Proteome Res.* **16**, 4281–4287
- Cifani, P., Dhabaria, A., Chen, Z., Yoshimi, A., Kawaler, E., Abdel-Wahab, O., Poirier, J. T., and Kentsis, A. (2018) ProteomeGenerator: A framework for comprehensive proteomics based on de novo transcriptome assembly and high-accuracy peptide mass spectral matching. *J. Proteome Res.* **17**, 3681–3692
- Liu, Y., González-Porta, M., Santos, S., Brazma, A., Marioni, J. C., Aebersold, R., Venkitesan, A. R., and Wickramasinghe, V. O. (2017) Impact of alternative splicing on the human proteome. *Cell Rep.* **20**, 1229–1241
- Tress, M. L., Abascal, F., and Valencia, A. (2017) Alternative splicing may not be the key to proteome complexity. *Trends Biochem. Sci.* **42**, 98–110

28. Blencowe, B. J. (2017) The Relationship between alternative splicing and proteomic complexity. *Trends Biochem. Sci.* **42**, 407–408
29. Tay, A. P., Pang, C. N. I., Twine, N. A., Hart-Smith, G., Harkness, L., Kassem, M., and Wilkins, M. R. (2015) Proteomic validation of transcript isoforms, including those assembled from RNA-Seq data. *J. Proteome Res.* **14**, 3541–3554
30. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., and Weissman, J. S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223
31. Van Dalen, K. M., Hodapp, S., Keskin, A., Otto, G. M., Berdan, C. A., Higdon, A., Cheunkarndee, T., Nomura, D. K., Jovanovic, M., and Brar, G. A. (2018) Global proteome remodeling during ERF stress involves Hac1-driven expression of long undecoded transcript isoforms. *Dev. Cell* **46**, 219–235.e8
32. Ho, Y.-H., Shishkova, E., Hose, J., Coon, J. J., and Gasch, A. P. (2018) Decoupling yeast cell division and stress defense implicates mRNA repression in translational reallocation during stress. *Curr. Biol.* **28**, 2673–2680.e4
33. Schwanhäusser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009) Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9**, 205–209
34. Liu, T.-Y., Huang, H. H., Wheeler, D., Xu, Y., Wells, J. A., Song, Y. S., and Wiita, A. P. (2017) Time-resolved proteomics extends ribosome profiling-based measurements of protein synthesis dynamics. *Cell Syst* **4**, 636–644.e9
35. Albert, F. W., Treusch, S., Shockley, A. H., Bloom, J. S., and Kruglyak, L. (2014) Genetics of single-cell protein abundance variation in large yeast populations. *Nature* **506**, 494–497
36. Albert, F. W., Bloom, J. S., Siegel, J., Day, L., and Kruglyak, L. (2018) Genetics of -regulatory variation in gene expression. *Elife* **7**, e35471
37. Foss, E. J., Radulovic, D., Shaffer, S. A., Ruderfer, D. M., Bedalov, A., Goodlett, D. R., and Kruglyak, L. (2007) Genetic basis of proteome variation in yeast. *Nat. Genet.* **39**, 1369–1375
38. Foss, E. J., Radulovic, D., Shaffer, S. A., Goodlett, D. R., Kruglyak, L., and Bedalov, A. (2011) Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. *PLoS Biol.* **9**, e1001144
39. Picotti, P., Clément-Ziza, M., Lam, H., Campbell, D. S., Schmidt, A., Deutsch, E. W., Röst, H., Sun, Z., Rinner, O., Reiter, L., Shen, Q., Michaelson, J. J., Frei, A., Alberti, S., Kusebauch, U., Wollscheid, B., Moritz, R. L., Beyer, A., and Aebersold, R. (2013) A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494**, 266–270
40. Suhre, K., Arnold, M., Bhagwat, A. M., Cotton, R. J., Engelke, R., Raffler, J., Sarwath, H., Thareja, G., Wahl, A., DeLisle, R. K., Gold, L., Pezer, M., Lauc, G., El-Din Selim, M. A., Mook-Kanamori, D. O., Al-Dous, E. K., Mohamoud, Y. A., Malek, J., Strauch, K., Grallert, H., Peters, A., Kastenmüller, G., Gieger, C., and Graumann, J. (2017) Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357
41. Sun, W., Kechris, K., Jacobson, S., Drummond, M. B., Hawkins, G. A., Yang, J., Chen, T.-H., Quibrera, P. M., Anderson, W., Barr, R. G., Basta, P. V., Bleecker, E. R., Beaty, T., Casaburi, R., Castaldi, P., Cho, M. H., Comellas, A., Crapo, J. D., Criner, G., Demeo, D., Christenson, S. A., Couper, D. J., Curtis, J. L., Doerschuk, C. M., Freeman, C. M., Gouskova, N. A., Han, M. K., Hanania, N. A., Hansel, N. N., Hersh, C. P., Hoffman, E. A., Kaner, R. J., Kanner, R. E., Kleerup, E. C., Lutz, S., Martinez, F. J., Meyers, D. A., Peters, S. P., Regan, E. A., Rennard, S. I., Scholand, M. B., Silverman, E. K., Woodruff, P. G., O’Neal, W. K., Bowler, R. P., SPIROMICS Research Group, and COPDGene Investigators. (2016) Common genetic polymorphisms influence blood biomarker measurements in COPD. *PLoS Genet.* **12**, e1006011
42. Jiang, L.-G., Li, B., Liu, S.-X., Wang, H.-W., Li, C.-P., Song, S.-H., Beatty, M., Zastrow-Hayes, G., Yang, X.-H., Qin, F., and He, Y. (2018) Characterization of proteome variation during modern maize breeding. *Mol. Cell. Proteomics* **18**, 263–276
43. Di Narzo, A. F., Telesco, S. E., Brodmerkel, C., Argmann, C., Peters, L. A., Li, K., Kidd, B., Dudley, J., Cho, J., Schadt, E. E., Kasarskis, A., Dobrin, R., and Hao, K. (2017) High-throughput characterization of blood serum proteomics of IBD patients with respect to aging and genetic factors. *PLoS Genet.* **13**, e1006565
44. Krahmer, N., Najafi, B., Schueder, F., Quagliarini, F., Steger, M., Seitz, S., Kasper, R., Salinas, F., Cox, J., Uhlenhaut, N. H., Walther, T. C., Jungmann, R., Zeigerer, A., Borner, G. H. H., and Mann, M. (2018) Organellar proteomics and phospho-proteomics reveal subcellular reorganization in diet-induced hepatic steatosis. *Dev. Cell* **47**, 205–221.e7
45. Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., Wang, X., Qiao, J. W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J. T., Gatzka, M. L., Wilkerson, M., Perou, C. M., Yellapantula, V., Huang, K.-L., Lin, C., McLellan, M. D., Yan, P., Davies, S. R., Townsend, R. R., Skates, S. J., Wang, J., Zhang, B., Kinsinger, C. R., Mesri, M., Rodriguez, H., Ding, L., Paulovich, A. G., Fenyö, D., Ellis, M. J., Carr, S. A., and NCI CPTAC. (2016) Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62
46. Park, J.-M., Park, J.-H., Mun, D.-G., Bae, J., Jung, J. H., Back, S., Lee, H., Kim, H., Jung, H.-J., Kim, H. K., Lee, H., Kim, K. P., Hwang, D., and Lee, S.-W. (2015) Integrated analysis of global proteome, phosphoproteome, and glycoproteome enables complementary interpretation of disease-related protein networks. *Sci. Rep.* **5**, 18189
47. Mertins, P., Qiao, J. W., Patel, J., Udeshi, N. D., Clauser, K. R., Mani, D. R., Burgess, M. W., Gillette, M. A., Jaffe, J. D., and Carr, S. A. (2013) Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat. Methods* **10**, 634–637
48. Grimes, M., Hall, B., Foltz, L., Levy, T., Rikova, K., Gaiser, J., Cook, W., Smirnova, E., Wheeler, T., Clark, N. R., Lachmann, A., Zhang, B., Hornbeck, P., Ma’ayan, A., and Comb, M. (2018) Integration of protein phosphorylation, acetylation, and methylation data sets to outline lung cancer signaling networks. *Sci. Signal.* **11**, eaaq1087
49. Britten, R. J., and Roberts, R. B. (1960) High-resolution density gradient sedimentation analysis. *Science* **131**, 32–33
50. Aviner, R., Hofmann, S., Elman, T., Shenoy, A., Geiger, T., Elkon, R., Ehrlich, M., and Elroy-Stein, O. (2017) Proteomic analysis of polyribosomes identifies splicing factors as potential regulators of translation during mitosis. *Nucleic Acids Res.* **45**, 5945–5957
51. Imami, K., Milek, M., Bogdanow, B., Yasuda, T., Kastelic, N., Zauber, H., Ishihama, Y., Landthaler, M., and Selbach, M. (2018) Phosphorylation of the ribosomal protein RPL12/uL11 affects translation during mitosis. *Mol. Cell* **72**, 84–98.e9
52. Slavov, N., Semrau, S., Airoldi, E., Budnik, B., and van Oudenaarden, A. (2015) Differential stoichiometry among core ribosomal proteins. *Cell Rep.* **13**, 865–873
53. Becher, I., Andrés-Pons, A., Romanov, N., Stein, F., Schramm, M., Baudin, F., Helm, D., Kurzawa, N., Mateus, A., Mackmull, M. T., Typas, A., Müller, C. W., Bork, P., Beck, M., and Savitski, M. M. (2018) Pervasive protein thermal stability variation during the cell cycle. *Cell* **173**, 1495–1507.e18
54. Martínez Molina, D., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E. A., Dan, C., Sreekumar, L., Cao, Y., and Nordlund, P. (2013) Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84–87
55. Dai, L., Zhao, T., Bisteau, X., Sun, W., Prabhu, N., Lim, Y. T., Sobota, R. M., Kaldis, P., and Nordlund, P. (2018) Modulation of protein-interaction states through the cell cycle. *Cell* **173**, 1481–1494.e13
56. Schopper, S., Kahraman, A., Leuenberger, P., Feng, Y., Piazza, I., Müller, O., Boersema, P. J., and Picotti, P. (2017) Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nat. Protoc.* **12**, 2391–2410
57. Feng, Y., De Franceschi, G., Kahraman, A., Soste, M., Melnik, A., Boersema, P. J., de Laureto, P. P., Nikolaev, Y., Oliveira, A. P., and Picotti, P. (2014) Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **32**, 1036–1044
58. Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P. J., von Mering, C., Claassen, M., and Picotti, P. (2017) Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, eaai7825
59. Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U., and Picotti, P. (2018) A map of protein-metabolite interactions reveals principles of chemical communication. *Cell* **172**, 358–372.e23
60. Reinhard, F. B. M., Eberhard, D., Werner, T., Franken, H., Childs, D., Doce, C., Savitski, M. F., Huber, W., Bantscheff, M., Savitski, M. M., and Drewes, G. (2015) Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. *Nat. Methods* **12**, 1129–1131

61. Rey, G., Milev, N. B., Valekunja, U. K., Ch, R., Ray, S., Silva Dos Santos, M., Nagy, A. D., Antrobus, R., MacRae, J. I., and Reddy, A. B. (2018) Metabolic oscillations on the circadian time scale in cells lacking clock genes. *Mol. Syst. Biol.* **14**, e8376
62. Stefely, J. A., Kwiecien, N. W., Freiberger, E. C., Richards, A. L., Jochem, A., Rush, M. J. P., Ulbrich, A., Robinson, K. P., Hutchins, P. D., Veling, M. T., Guo, X., Kemmerer, Z. A., Connors, K. J., Trujillo, E. A., Sokol, J., Marx, H., Westphall, M. S., Hebert, A. S., Pagliarini, D. J., and Coon, J. J. (2016) Mitochondrial protein functions elucidated by multiomic mass spectrometry profiling. *Nat. Biotechnol.* **34**, 1191–1197
63. Lapointe, C. P., Stefely, J. A., Jochem, A., Hutchins, P. D., Wilson, G. M., Kwiecien, N. W., Coon, J. J., Wickens, M., and Pagliarini, D. J. (2018) Multi-omics reveal specific targets of the RNA-binding protein puf3p and its orchestration of mitochondrial biogenesis. *Cell Syst.* **6**, 125–135.e6
64. Jha, P., McDevitt, M. T., Gupta, R., Quiros, P. M., Williams, E. G., Gariani, K., Sleiman, M. B., Diserens, L., Jochem, A., Ulbrich, A., Coon, J. J., Auwerx, J., and Pagliarini, D. J. (2018) Systems analyses reveal physiological roles and genetic regulators of liver lipid species. *Cell Syst.* **6**, 722–733.e6
65. Williams, E. G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C. A., Houten, S. M., Amariuta, T., Wolski, W., Zamboni, N., Aebersold, R., and Auwerx, J. (2016) Systems proteomics of liver mitochondria function. *Science* **352**, aad0189
66. Huang, S., Chaudhary, K., and Garmire, L. X. (2017) More is better: recent progress in multi-omics data integration methods. *Front. Genet.* **8**, 84
67. Haider, S., and Pal, R. (2013) Integrated analysis of transcriptomic and proteomic data. *Curr. Genomics* **14**, 91–110
68. Pedersen, H. K., Forslund, S. K., Gudmundsdottir, V., Petersen, A. Ø., Hildebrand, F., Hyötyläinen, T., Nielsen, T., Hansen, T., Bork, P., Ehrlich, S. D., Brunak, S., Oresic, M., Pedersen, O., and Nielsen, H. B. (2018) A computational framework to integrate high-throughput “-omics” data sets for the identification of potential mechanistic links. *Nat. Protoc.* **13**, 2781–2800
69. Kuo, T.-C., Tian, T.-F., and Tseng, Y. J. (2013) 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. *BMC Syst. Biol.* **7**, 64
70. Zhou, G., and Xia, J. (2018) OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. *Nucleic Acids Res.* **46**, W514–W522
71. Zeng, I. S. L., and Lumley, T. (2018) Review of statistical learning methods in integrated omics studies (An Integrated Information Science). *Bioinform. Biol. Insights* **12**, 1177932218759292
72. Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., Buettner, F., Huber, W., and Stegle, O. (2018) Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**, e8124
73. Kim, D., Li, R., Lucas, A., Verma, S. S., Dudek, S. M., and Ritchie, M. D. (2017) Using knowledge-driven genomic interactions for multi-omics data analysis: metadimensional models for predicting clinical outcomes in ovarian carcinoma. *J. Am. Med. Assoc.* **318**, 577–587
74. Tuncbag, N., Gosline, S. J. C., Kedaigle, A., Soltis, A. R., Gitter, A., and Fraenkel, E. (2016) Network-based interpretation of diverse high-throughput datasets through the omics integrator software package. *PLoS Comput. Biol.* **12**, e1004879
75. Schwartz, G. W., Petrovic, J., Zhou, Y., and Faryabi, R. B. (2018) Differential integration of transcriptome and proteome identifies pan-cancer prognostic biomarkers. *Front. Genet.* **9**, 205
76. Sedgewick, A. J., Benz, S. C., Rabizadeh, S., Soon-Shiong, P., and Vaske, C. J. (2013) Learning subgroup-specific regulatory interactions and regulator independence with PARADIGM. *Bioinformatics* **29**, i62–70
77. Teo, G., Bin Zhang, Y., Vogel, C., and Choi, H. (2018) PECAPlus: statistical analysis of time-dependent regulatory changes in dynamic single-omics and dual-omics experiments. *NPJ Syst. Biol. Appl.* **4**, 3
78. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740
79. Liu, T.-Y., Huang, H. H., Wheeler, D., Xu, Y., Wells, J. A., Song, Y. S., and Wiita, A. P. (2017) Time-resolved proteomics extends ribosome profiling-based measurements of protein synthesis dynamics. *Cell Syst* **4**, 636–644.e9
80. McKenzie, A. T., Moyon, S., Wang, M., Katsyv, I., Song, W.-M., Zhou, X., Dammer, E. B., Duong, D. M., Aaker, J., Zhao, Y., Beckmann, N., Wang, P., Zhu, J., Lah, J. J., Seyfried, N. T., Levey, A. I., Katsel, P., Haroutunian, V., Schadt, E. E., Popko, B., Casaccia, P., and Zhang, B. (2017) Multi-scale network modeling of oligodendrocytes reveals molecular components of myelin dysregulation in Alzheimer’s disease. *Mol. Neurodegener.* **12**, 82
81. Tchourine, K., Poultney, C. S., Wang, L., Silva, G. M., Manohar, S., Mueller, C. L., Bonneau, R., and Vogel, C. (2014) One third of dynamic protein expression profiles can be predicted by a simple rate equation. *Mol. Biosyst.* **10**, 2850–2862
82. Guo, Z., Kanjanapangka, J., Liu, N., Liu, S., Liu, C., Wu, Z., Wang, Y., Loh, T., Kowolik, C., Jansen, J., Zhou, M., Truong, K., Chen, Y., Zheng, L., and Shen, B. (2012) Sequential posttranslational modifications program FEN1 degradation during cell-cycle progression. *Mol. Cell* **47**, 444–456
83. Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018) SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161
84. Lombard-Banek, C., Reddy, S., Moody, S. A., and Nemes, P. (2016) Label-free quantification of proteins in single embryonic cells with neural fate in the cleavage-stage frog (*Xenopus laevis*) embryo using capillary electrophoresis electrospray ionization high-resolution mass spectrometry (CE-ESI-HRMS). *Mol. Cell. Proteomics* **15**, 2756–2768
85. Genshaft, A. S., Li, S., Gallant, C. J., Darmanis, S., Prakadan, S. M., Ziegler, C. G. K., Lundberg, M., Fredriksson, S., Hong, J., Regev, A., Livak, K. J., Landegren, U., and Shalek, A. K. (2016) Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* **17**, 188
86. Bendall, S. C., Simonds, E. F., Qiu, P., Amir, E.-A. D., Krutzik, P. O., Finck, R., Bruggner, R. V., Melamed, R., Trejo, A., Ornatsky, O. I., Balderas, R. S., Plevritis, S. K., Sachs, K., Pe’er, D., Tanner, S. D., and Nolan, G. P. (2011) Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696
87. Onjiko, R. M., Moody, S. A., and Nemes, P. (2015) Single-cell mass spectrometry reveals small molecules that affect cell fates in the 16-cell embryo. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6545–6550
88. Zhu, Y., Piehowski, P. D., Zhao, R., Chen, J., Shen, Y., Moore, R. J., Shukla, A. K., Petyuk, V. A., Campbell-Thompson, M., Mathews, C. E., Smith, R. D., Qian, W.-J., and Kelly, R. T. (2018) Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882
89. Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538