Proteomic and interactomic insights into the molecular basis of cell functional diversity

Isabell Bludau[1,2] *and Ruedi Aebersold*[1,3]✉

Abstract | The ability of living systems to adapt to changing conditions originates from their capacity to change their molecular constitution. This is achieved by multiple mechanisms that modulate the quantitative composition and the diversity of the molecular inventory. Molecular diversification is particularly pronounced on the proteome level, at which multiple proteoforms derived from the same gene can in turn combinatorially form different protein complexes, thus expanding the repertoire of functional modules in the cell. The study of molecular and modular diversity and their involvement in responses to changing conditions has only recently become possible through the development of new 'omics'-based screening technologies. This Review explores our current knowledge of the mechanisms regulating functional diversification along the axis of gene expression, with a focus on the proteome and interactome. We explore the interdependence between different molecular levels and how this contributes to functional diversity. Finally, we highlight several recent techniques for studying molecular diversity, with specific focus on mass spectrometry-based analysis of the proteome and its organization into functional modules, and examine future directions for this rapidly growing field.

Interactome
The whole set of physical interactions between molecules in a cell, here specifically referring to protein–protein interactions.

A fundamental goal of molecular life science research is to understand the complex relationship between genotype and phenotype. The pioneering work of Beadle and Tatum on the genetic root of biochemical reactions in *Neurospora* spp.[1] established the paradigm of 'one gene, one protein, one function'. This paradigm makes two fundamental assumptions: that there is a single link between a gene and the function of its corresponding protein product (thus implying that the functional diversity of a biological system is directly encoded in its protein-coding genes), and that the expression of a gene has no impact on the expression or function of any other gene product.

With the technological advances of the genomic age and the ability to decipher the molecular make-up of cells at increasing precision and resolution, it has become clear that the 'one gene, one protein, one function' paradigm does not fully explain the complex functional phenotypes of organisms. Upon sequencing the human genome, the International Human Genome Sequencing Consortium reported ~20,000 protein-coding genes[2], a number markedly lower than pre-genomic estimates. This finding suggested that the degree of functional diversity of an organism is not directly related to the number of protein-coding genes. This suggestion was further evident in the limited progress of large-scale screening methods, such as genome-wide association studies and RNA interference screens, to provide insight into direct links between genotype and phenotype[3,4]. Although single-gene defects with high penetrance have been discovered using these techniques, the genetic basis of many studied phenotypes is more complex than expected, and is often underpinned by a network of genomic changes or mechanisms that involve other molecular layers, such as the transcriptome, proteome and interactome, as well as non-linear crosstalk between them[3–6].

The appreciation of the functional complexity of living systems has led to a paradigm shift towards a systems biology view of genotype–phenotype relationships. The underlying assumptions of systems biology are that the functional diversity of a cell results from multiple layers beyond the genome and that, to gain a comprehensive understanding of complex biological processes, molecules and their relationships must be studied as part of integrated systems[7]. The systems-based approach is dependent on the detection and precise quantification of the molecular diversity of the cell at the levels of the transcriptome, proteome and interactome (FIG. 1). In this Review, we describe known mechanisms that determine molecular diversity beyond the genome, on the transcriptome, proteome

[1]*Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland.*
[2]*Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany.*
[3]*Faculty of Science, University of Zurich, Zurich, Switzerland.*
✉*e-mail: aebersold@imsb.biol.ethz.ch*

https://doi.org/10.1038/s41580-020-0231-2

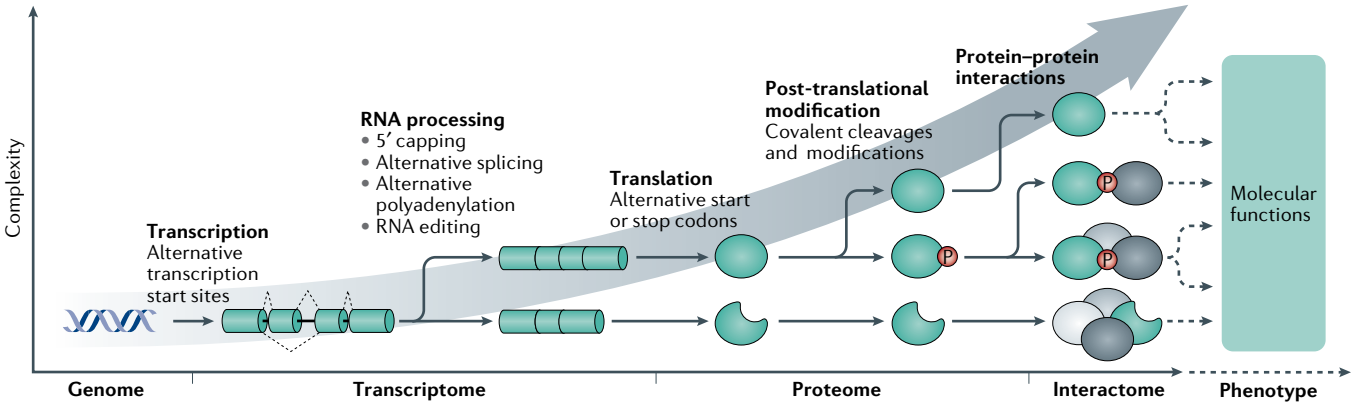


Fig. 1 | **The generation of functional diversity at different molecular levels.** In contrast to the 'one gene, one protein, one function' paradigm, cellular complexity arises from many mechanisms that expand molecular diversity beyond that encoded by the protein-coding genome. These mechanisms include an increase in coding potential using alternative transcription start sites as well as 5′ capping, alternative splicing, alternative polyadenylation and RNA editing at the co-transcriptional or post-transcriptional level. The diversity of proteins is further increased using alternative start and stop codons during translation. A high degree of diversification is introduced by post-translational modifications, which include covalent cleavages and covalent modifications (such as phosphorylation (P)). Finally, proteins can interact with each other to form multiple distinct functional units that can potentially perform various downstream functions. Although recent technological advances provide headway towards fully characterizing the transcriptome, proteome and interactome and their relationships in any given state, the assessment of their functional impact and the phenotype is a challenge that still remains to be fully explored (dashed lines).

and interactome levels, with a focus on the proteome and interactome. Furthermore, we discuss non-linear information flow between the different molecular layers, including crosstalk and feedback loops that enable the fine-tuned regulation of the different 'omes'. Finally, we highlight recent techniques for protein identification and quantification, proteoform characterization and interactome analysis by mass spectrometry and further point towards future directions in the context of studying molecular and functional diversity.

The transcriptome directs the proteome

The functional diversity intrinsically embedded in the genome of a species is largely reflected by the number of protein-coding genes it contains. With the exception of epigenetic modifications and conformational changes to its three-dimensional organization, the genome of an individual cell, and even within an entire organism, is mostly stable. The ability of a cell to have different phenotypes and to dynamically adjust to environmental perturbations therefore primarily originates from variation in the transcriptome and beyond. The transcriptome of a biological system includes protein-coding mRNA, as well as non-coding tRNA, microRNA, ribosomal RNA and long non-coding RNA. Unlike the genome, the transcriptomes of different cells in the same organism can be considerably different and they can dynamically change in response to internal or external stimuli. Recent studies estimate that 56–84% of protein variance can be explained by mRNA variance in cells at steady state[8–10]. It was further shown that mRNA transcripts with particularly high expression levels are more efficiently translated[11]. These results support the notion that mRNA expression can serve as a useful surrogate for protein expression at steady state in most cases.

Modulating mRNA expression levels. Cells carefully adjust mRNA expression levels via multiple regulatory mechanisms. The interaction of *cis*-regulatory elements with a combination of *trans*-acting factors creates a complex network that regulates RNA polymerase-mediated gene transcription in response to changing conditions and, ultimately, the composition and quantity of the transcriptome[12,13]. In addition to transcription itself, several post-transcriptional mechanisms regulate the transcriptome composition of eukaryotic cells. First, aberrantly processed mRNA transcripts can be targeted for degradation by the RNA exosome[14–16]. This degradation is a key component for quality control at the transcriptional level, and as the exosome is present in the nucleus as well as the cytosol, it can target nascent and mature mRNA transcripts. Second, aberrantly processed mRNAs can also be degraded through nonsense-mediated decay (NMD)[17]. Although mainly associated with the degradation of aberrantly processed mRNA, cells are known to exploit both the exosome and NMD mechanisms to fine-tune gene expression and, thereby, to directly influence functional diversity on the proteome level. This is exemplified by the selective targeting of specific splice isoforms for NMD during granulocyte differentiation[18,19]. Third, RNA can be post-transcriptionally modified. There are >100 distinct RNA chemical modifications that do not change the nucleotide sequence but alter the epitranscriptome[20]. The most prominent of these modifications is the methylation of adenosine to produce N^6-methyladenosine (m^6A). This modification can influence the recruitment of effector proteins and the formation of RNA secondary structures, which in turn modulates mRNA metabolism, including maturation, translation and decay[21–23]. According to current estimations, ~0.1–0.4% of all adenosine nucleotides in mammalian mRNA are subjected to m^6A epitranscriptomic

Proteoform
A protein species with a unique combination of amino acid sequence and post-translational modifications. Multiple alternative proteoforms can originate from the same gene locus.

Isoforms
Alternative mRNA transcripts originating from the same gene locus.

Epitranscriptome
The whole set of biochemical modifications of RNAs in a cell, with the methylation of adenosine being the most prominent modification.

modifications[21,24], thereby affecting ~25% of human transcripts[21]. Although little is known about the regulation of mRNA modifications and their overall impact on protein expression, several disease phenotypes are associated with m^{6}A and its regulatory factors[21–24].

Regulating the coding diversity of the transcriptome. In addition to quantitative variations, eukaryotic cells have also evolved mechanisms to expand transcript sequence diversity[25]. A first level of diversification is introduced by the selection of alternative transcription start sites. By selecting alternative *cis*-regulatory promoter regions during transcription initiation, different protein-coding precursor mRNAs (pre-mRNAs) of varying lengths can be transcribed, potentially including different open reading frames and alternative first exons that can increase the number of translated proteins[26]. For example, the pyruvate kinase isoforms PKL and PKR, encoded by *PKLR*, are a product of tissue-specific promoter usage that produces alternative mRNA transcripts from the same gene with different first exons (FIG. 2a). Human genes have, on average, more than four transcription start sites, indicating that this process can introduce considerable diversity to the proteome[27,28].

Pre-mRNA diversity is further increased by multiple processes that alter the nucleic acid sequence of the transcript. Before a pre-mRNA transcript can be transported out of the nucleus and translated, it must undergo three (co-transcriptional) processing events that can potentially expand its coding potential. These processes are acquisition of a 7-methylguanosine cap at the 5′ end, splicing to remove intragenic regions (introns) and ligate expressed regions (exons), and cleavage and addition of a poly(A) tail at the 3′ end. These processing steps ensure protection of the transcript from degradation, assist with translation initiation and can introduce considerable molecular diversification[25,29]. Whereas 5′ capping rarely generates variability, splicing is a major source of transcript diversity in eukaryotes. RNA sequencing studies indicate that 92% of human multi-exon genes undergo alternative splicing[30,31], which amounts to 86% of all human genes being estimated to produce two or more distinct mRNA isoforms[30]. Some genes can generate more than ten alternatively spliced transcripts[32]. These splice isoforms might contain different numbers of exons, mutually exclusive exons, retained introns or alternative 5′ or 3′ splice sites.

With the exception of a few self-splicing introns, pre-mRNA splicing is commonly catalysed by a large ribonucleoprotein complex known as the spliceosome[33]. The choice between alternative 5′ and 3′ splice sites can, among other mechanisms, be regulated by interactions with non-spliceosomal RNA-binding proteins, such as heterogeneous nuclear ribonucleoproteins (hnRNPs) and serine/arginine-rich (SR) proteins[34]. Alternative splicing is essential during development and mRNA isoforms can be specific to certain cell types, tissues and species[35]. Consequently, errors in splicing are associated with disease phenotypes, including cancer, neurodegenerative diseases and muscle dystrophies[36,37]. The pyruvate kinase gene *PKM* is an example of a gene with two alternatively spliced isoforms: PKM1 and PKM2 (FIG. 2b). These two isoforms contain mutually exclusive exons, which regulate the activity of the enzyme. Multiple studies have associated changes in the *PKM* alternative splicing pattern and the subsequent elevated expression of the PKM2 isoform with numerous different types of cancer (reviewed in REF.[38]). Overall, it is expected that

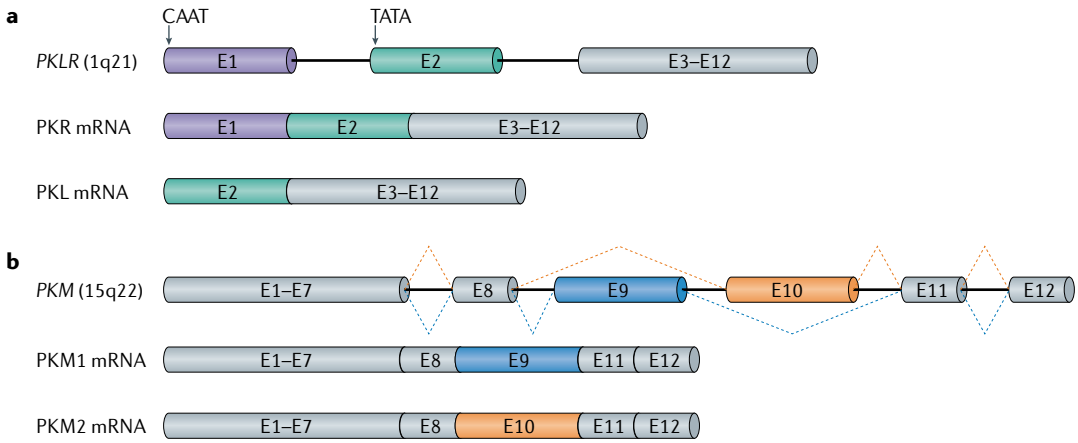


Fig. 2 | **Tissue-specific promoter usage and alternative splicing lead to different pyruvate kinase isoforms.** Pyruvate kinase (PK) is a glycolytic enzyme that catalyses the final step in glycolysis. There are four isoenzymes of PK in mammals (PKR, PKL, PKM1 and PKM2) that are encoded by *PKLR* and *PKM*[38]. **a** | *PKLR* can generate the isoforms PKR and PKL depending on tissue-specific promoter usage. Whereas the PKR isoform, which is transcribed from the CAAT box in the canonical promoter, contains all of the exons (E), that is, E1–E12, the PKL isoform is expressed from a gene-internal TATA promoter and therefore lacks E1 (purple) and starts at E2 (green). PKL is expressed in response to hormones and nutrient availability and is specific to liver and kidney tissue. **b** | The two isoforms of the *PKM* gene are generated by alternative splicing, resulting in the mutually exclusive inclusion of E9 (blue) in PKM1 or E10 (orange) in PKM2. Both E9 and E10 correspond to a regulatory region of the protein: inclusion of E9 in PKM1 enables the enzyme to be constitutively tetrameric and active, and inclusion of E10 in PKM2 allows the regulation of its activity by nutrient availability. PKM1 is predominantly expressed in tissues with high energetic demands, such as the brain and muscle, whereas PKM2 is expressed in embryonic cells and cancer cells. Adapted with permssion from REF.[38], Elsevier.

functional differences between alternative isoforms frequently involve the inclusion or exclusion of binding motifs or specific post-translational modification (PTM) sites[39].

Although alternative splicing is considered the main contributor to transcript diversification, some additional variability is introduced by alternative polyadenylation of the transcript's 3′ end. Alternative polyadenylation can affect either the length of the transcript's coding region or the length of the transcript's 3′ untranslated region, which can in turn impact transcript location, stability and the efficiency with which it is translated. In mammalian genomes, at least 70% of genes have multiple polyadenylation sites that appear to be tissue specific and tightly regulated during development and cell differentiation, thereby implying their regulatory impact on proteome and functional diversity[26,28].

In addition to these three main processing steps, some pre-mRNAs are further edited by the selective deamination of adenosines and cytosines, which usually happens prior to splicing[40,41]. As most RNA-editing sites in human pre-mRNA occur in intronic regions and mutations are frequently synonymous substitutions, RNA editing is expected to have only a limited effect on the overall coding potential of a human cell[40]. Nevertheless, crosstalk between RNA editing and the splicing machinery as well as several non-synonymous RNA-editing events with effects on protein function and associated disease phenotypes have been reported[40,42–44]. For example, adenosine-to-inosine editing in RNA encoding glutamate receptor 2, which causes a Glu-to-Arg substitution, changes the calcium permeability and recovery of the receptor[44,45]. However, a global picture of the functional impact of RNA- editing events still remains to be elucidated.

Overall, starting from ~20,000 annotated human genes, the human transcriptome is currently estimated to comprise >83,000 protein-coding mRNA isoforms that are annotated in the GENCODE database, highlighting the relevance of the aforementioned mechanisms in generating transcriptome diversity[26]. It is still heavily debated to what extent the richness of mRNA isoforms is a mere product of stochastic noise[46,47] and which fraction of alternative transcript isoforms is actually translated into functional protein isoforms. These questions cannot be addressed by exclusively studying the transcriptome but require further insights into the proteome level.

The diversity of the proteome

The proteome of a biological system consists of all of its expressed protein molecules. Like transcriptomes, the proteomes of different cells in the same organism can vary greatly and are capable of dynamically adapting to internal and external stimuli.

Modulating protein expression levels. Although protein synthesis is directly dependent on the expression of a corresponding mRNA transcript, factors beyond transcript concentration contribute to protein expression levels and, thereby, to the functional landscape of the cell[48].

First, the translation rate is influenced by the mRNA sequence itself through mechanisms such as codon bias, epitranscriptomic modifications, the interaction of the transcript with regulatory elements (such as microRNAs, which cause post-transcriptional gene silencing) and the availability of tRNAs and uncharged ribosomes (a lack of which can reduce translation efficiency). Second, the proteome is modulated by protein degradation rates, which are influenced by protein localization, stability, the three-dimensional conformation and their integration into stable protein complexes. Proteins can also be directly targeted for ubiquitin-mediated degradation, or become targets of autophagy and, consequently, lysosomal-mediated degradation[49]. The amino-terminal and carboxy-terminal composition of a protein can determine a protein's half-life through the recognition of degron sequences by proteolytic systems that cause degradation via N-degron pathways or C-degron pathways, respectively[50]. In eukaryotes, these pathways include degradation both by the proteasome and by autophagy.

In healthy cells and at steady state, protein synthesis and degradation are well balanced[51] (FIG. 3a). In addition, variation in mRNA abundance is frequently buffered on the protein level, meaning that a substantial change in mRNA abundance is not propagated into a corresponding change in protein abundance[48]. Whereas genomic copy number variations in cancer commonly affect mRNA expression levels, many of these changes do not affect expression of the corresponding protein, suggesting that these aberrations do not notably contribute to the cancer phenotype[52]. Interestingly, protein-level buffering is especially pronounced for proteins that form stable protein complexes, suggesting that stoichiometric ratios of complex subunits can direct protein-level abundance variation[48,53]. The effect of protein complex-mediated buffering illustrates a prime example of how cells can reduce an undesired functional impact of variation on genome and transcriptome levels, for example, induced by stochastic events or by potentially disease-promoting genotypes.

Despite these effects, mRNA abundance generally provides a good estimate for protein-level abundance at steady state (also see 'The transcriptome directs the proteome' above). However, in cells during transition phases (for example, throughout the cell cycle or differentiation processes) or rapidly adapting to stimuli, the agreement between mRNA and protein levels can temporarily decrease. This decrease can be due to delays in signal transmission from the gene, to transcript to protein level, or due to mechanisms that allow rapid adaptation of the quantitative proteome without affecting mRNA levels; one such mechanism is the elevated translation of existing mRNA transcripts, known as 'translation on demand'[54]. Another mechanism is the rapid degradation of expressed proteins, for example, by the ubiquitin–proteasome pathway, foregoing the necessity to repress transcription of the respective genes. Interestingly, studies indicate that the overall protein concentration of the cell remains fairly constant across conditions, meaning that if a few transcripts are massively upregulated, then the increase in the respective protein concentration is compensated by a decrease in the concentration of other proteins[55,56].

Alternative polyadenylation
Deviations from the standard process of polyadenylation, including the usage of different poly(A) start sites or variable length of the transcript's 3′ untranslated region.

Synonymous substitutions
DNA base substitutions in an exon of a protein-coding gene that do not cause a change in the translated protein sequence.

N-degron pathways
A set of proteolytic systems that can recognize proteins containing N-degrons, thereby causing the degradation of these proteins (formerly 'N-end rule pathways'). N-degrons are degradation signals in the amino-terminal region of a protein, which relate its in vivo half-life to the identity of its amino-terminal composition.

C-degron pathways
A set of proteolytic systems that can recognize proteins containing C-degrons, thereby causing the degradation of these proteins. C-degrons are degradation signals in the carboxy-terminal region of a protein, which relate its in vivo half-life to the identity of its carboxy-terminal composition.

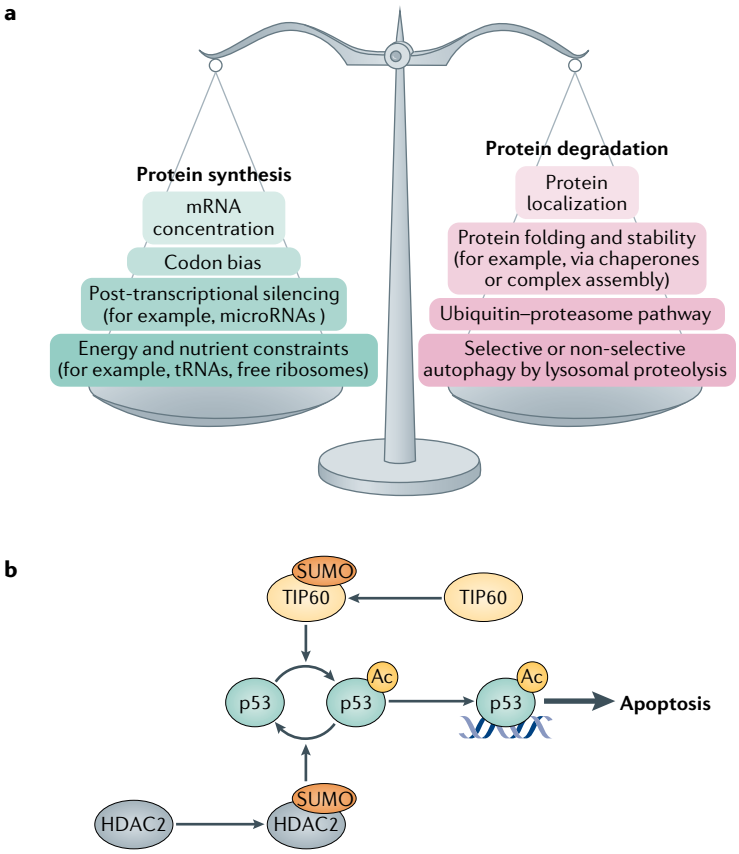


Fig. 3 | **Proteome balance and post-translational modification crosstalk.** The cell tightly regulates both a protein's abundance and its functional activity by means of multiple interdependent mechanisms. **a** | Homeostasis is maintained through a fine balance between protein synthesis and protein degradation. The translation efficiency of mRNA transcripts is influenced by codon bias, by post-transcriptional regulation, such as microRNA-mediated gene silencing, or by energy and nutrient constraints, including the availability of tRNAs and free ribosomes. Protein degradation is influenced by the protein's localization, the efficiency of protein folding (a process that frequently requires chaperones or integration into a protein complex), ubiquitin-mediated degradation and lysosomal proteolysis following selective and non-selective autophagy. Proteome degradation rates are expected to buffer unexpected changes in mRNA transcript abundance, such as those caused by gene amplifications in cancer. **b** | Post-translational modifications are important for regulating protein activity, and post-translational modifications on different proteins are often not independent of each other. The activity of the tumour suppressor p53 in initiating apoptosis, for example, is modulated by its acetylation (Ac) status. The Ac state and activity of p53 is increased by the acetyltransferase TIP60, which is itself activated by SUMOylation. Conversely, SUMOylation of the histone deacetylase HDAC2 promotes the deacetylation and inactivation of p53 (REF.[68]).

Modulating diversity at the protein level. In addition to quantitative protein levels that can directly influence the functional landscape of a cell, other processes further increase proteome diversity. The first level of proteome diversification can be attributed to the use of alternative start or stop codons during translation. These diversification mechanisms are rare, as initiation at 5′ AUG start codons occurs in ~90–95% of cases and termination efficiency at the first in-frame stop codon is ~90–99%[57]. However, the use of an alternative downstream CUG start codon in *MRPL18*, a gene that encodes a mitochondrial large ribosomal subunit protein, provides an example where an alternative start codon has a distinct functional impact, in this case causing mislocalization of the protein and formation of 'hybrid' ribosomes that promote increased heat shock resistance[58]. Errors in translation, estimated to occur at ~0.01–0.1% per amino acid[59], also contribute to protein sequence variation that might expand proteome diversity in stress and ageing. However, most of the stochastically occurring errors are expected to result in non-functional, often misfolded, proteins that are then degraded by the ubiquitin–proteasome system[15].

Protein PTMs, which can be covalent cleavages or covalent modifications[60], introduce the most diversity at the proteome level. Covalent cleavage of the protein's primary amino acid sequence is commonly catalysed by specific proteases or, less frequently, mediated by autocatalytic cleavage[61]. Cleavage can modify protein function, for example, by changing protein localization and activity, as exemplified by the cleavage of pro-caspase proteins during caspase-mediated apoptosis[62]. Many proteases are stored as inactive proenzymes that are only activated upon their own proteolytic cleavage[61]. To date, human cells are known to express 460 different catalytically active proteases[63].

Covalent modifications are catalysed by enzymes that add a specific chemical group to the amino acid side chains or to the carboxy terminus or amino terminus of a protein. Although some PTMs are permanent, others are reversible and can be dynamically altered. Fifteen of the 20 common proteinogenic amino acids can be modified and there are ~400 different PTMs[64,65]. The three most prevalent covalent PTMs are currently estimated to be *N*-linked glycosylation of Asn residues, phosphorylation of Ser, Thr and Tyr residues, and acetylation of Lys residues[64,65].

The largest class of known PTM enzymes regulates protein phosphorylation. Whereas >500 kinases catalyse protein phosphorylation, only another ~140 protein phosphatases catalyse dephosphorylation[66]. Rough estimates indicate that at least 10,000 distinct molecular forms of phosphorylated proteins can be produced and selectively regulated in humans[60]. However, this number is probably an underestimation because phosphorylation events often occur in combination on the same protein. Among other functions, covalent modifications are a crucial component of cell signalling pathways such as receptor tyrosine kinase signalling, and perturbation of the covalent modification machinery, and of kinases and phosphatases specifically, can cause severe disease phenotypes[67]. Different PTM sites on one protein or across multiple proteins can be connected via a complex network of molecular crosstalk. The acetylation status of the key transcriptional regulator p53, for example, is regulated by acetyltransferases and deacetylases that are themselves regulated by SUMOylation[68] (FIG. 3b).

Taken together, the mechanisms that introduce variation on the transcriptome and proteome levels are currently estimated to generate >1 million different proteoforms[65] based on unique combinations of different amino acid sequences and PTMs. The analytical challenge in proteomics is therefore not only to detect and quantify all protein-coding genes that are expressed but also to correctly identify and quantify each proteoform. Although recent technical and methodological

SUMOylation
The post-translational modification of a protein by covalently attached small ubiquitin-like modifier (SUMO) proteins.

developments have enabled the almost complete enumeration and quantification of the human proteome[69,70], the global identification of proteoforms still remains a challenge[65,71]. However, recent developments in the field of top-down proteomics have enabled the parallel identification of >3,000 unique proteoforms in human samples[72,73]. Although the molecular function of some proteoforms, as well as the phenotypic traits that they are associated with, have successfully been annotated[65,74], the systematic assessment of proteoform-specific functions remains challenging.

The functional capacity of the proteome is further enhanced by the fact that the same proteoform might exist in multiple three-dimensional conformations and thus associate with different macromolecular assemblies. The organization of a protein's secondary and tertiary structure can affect its stability (for example, by the exposure of N-degrons or C-degrons), localization and molecular function. Currently, most system-wide proteomic studies do not consider these structural differences and treat proteins as unstructured molecules. By contrast, structurally focused studies are usually performed from a targeted prospective, focusing on a single protein or on a small subset of proteins and their structures. Methods for probing protein conformations in larger proteome fractions have been developed recently based on cross-linking mass spectrometry[75,76], thermal proteome profiling[77] or coupling limited proteolysis to targeted proteomics[78]. Owing to the limited availability of system-wide approaches, to what extent structural differences impact proteome and functional diversity on a global scale remains largely unknown.

The dynamic interactome

Many proteins perform biochemical functions as part of multimolecular assemblies. These assemblies form a large network of molecular interactions, including interactions among molecules of the same type, for example, protein–protein interactions (PPIs), or among molecules of different types, for example, protein–DNA, protein–RNA, protein–lipid or protein–metabolite interactions. The vast array of molecular interactions in the cell is referred to as its interactome. Due to their central role in many biological processes and the focus of this Review on functional complexity beyond the protein-coding genomic sequence, we specifically focus on PPIs and their arrangement in a complex interaction network.

Overview of the cellular interactome. The interactome of a cell is more diverse and can adapt more rapidly to environmental cues than the transcriptome and proteome. The main reason for the versatility of the interactome is that it does not necessitate the synthesis of new molecules. This is evident from the transient nature of some PPIs, for example, the interaction between kinases and their target proteins. By contrast, other interactions are more stable, with participating proteins forming distinct functional units known as macromolecular protein complexes. Prominent examples of these macromolecular 'machines' are the ribosome and proteasome. Many proteins can be functionally distinct in their monomeric or complex bound forms, exemplified by 14-3-3 proteins[79]. Whereas 14-3-3 monomers are associated with a chaperone-like activity[79], 14-3-3 dimers primarily operate as a phosphorylation-dependent protein scaffold that is an important component of many molecular signalling pathways[80].

Over the past decades, several experimental and bioinformatics strategies have been developed to map the interactome of different cellular systems. The most comprehensive single-assay interaction map contains >56,000 associations among >10,000 human proteins and was generated through the systematic analysis of human open reading frames by affinity purification coupled to tandem mass spectrometry (AP-MS)[81,82]. In addition to binary PPI networks, curated databases containing information about well-defined protein complexes and their associated biochemical functions have been generated; for example, the human protein complex map (hu.MAP) reports 4,659 protein complexes[83]. Other protein complex databases include CORUM[84,85] and the Complex Portal[86,87]. Although current databases provide a broad overview of the molecular interactome of a human cell, the observable interaction space is expected to markedly increase with technological advances. It is important to note that PPI and protein complex databases only provide a static, generic description of proteome connectivity and do not resolve the cell-type-specific and state-specific PPIs that indicate the acute biochemical state of the cell.

Factors influencing diversity at the interactome level. The main prerequisite for PPIs or the assembly of a protein complex is that the proteins involved are co-expressed at the required stoichiometry and colocalized within the same cellular compartment. Each protein may need to be in a specific three-dimensional conformation that allows the interaction to form and to be energetically favourable[88,89]. Interactions can depend on the presence of assembly chaperones, which can be proteins or other molecules that assist the formation of an interaction, for example, by acting as scaffolds or by causing a conformational change in either of the interacting subunits[90]. PPIs and complex assemblies often dynamically change in response to stimuli, for example, upon growth factor exposition, or generally over the course of molecular processes. One exemplary process is the dynamically changing composition of the spliceosome at different steps of the splicing reaction[33]. Considering its many rearrangements and different catalytic activities, the spliceosome machinery alone provides an impressive illustration of functional diversity orchestrated on the interactome level.

One central consideration when investigating functional diversification on the interactome level is that interactions might be proteoform-dependent, and vice versa. An interaction might only form when a protein subunit has a specific amino acid sequence or PTMs. In addition, certain PTMs may only occur upon complex formation. Proteoform-dependent rearrangement of the interactome is a common mechanism for the cell to rapidly adapt its functional landscape in response to changing environmental conditions (FIG. 4a). This is illustrated by the phosphorylation-dependent binding

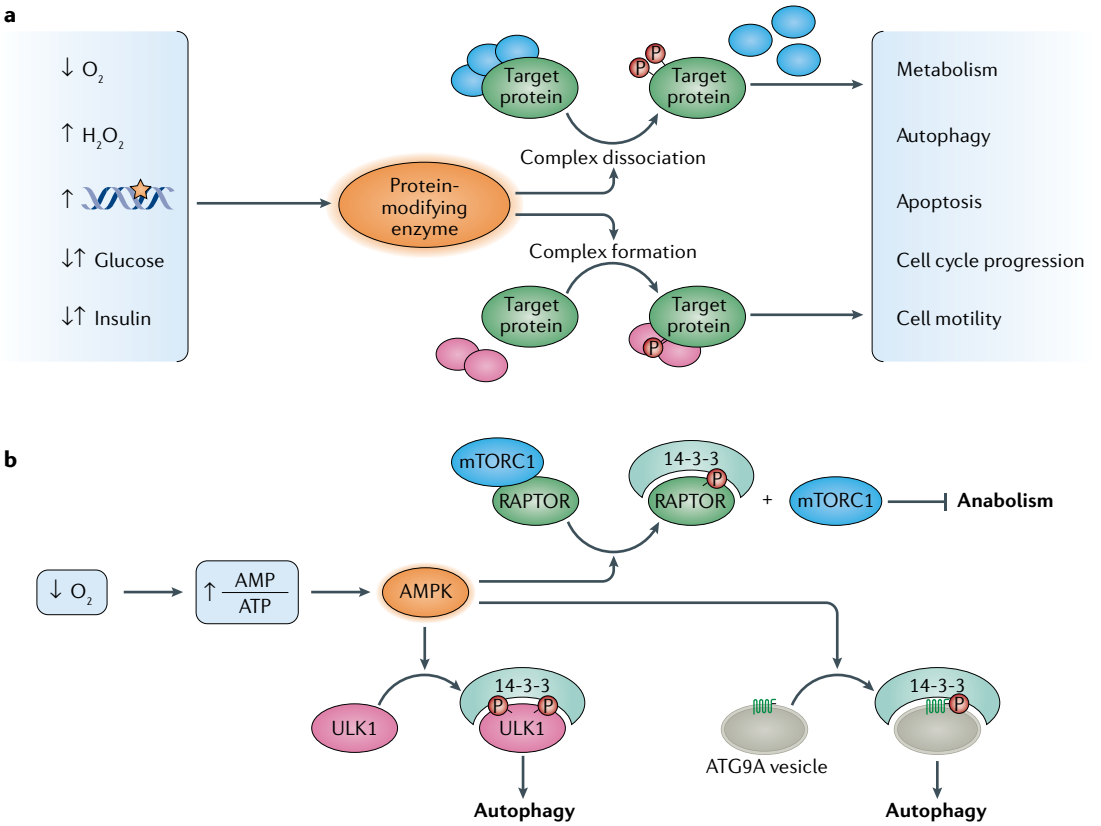


Fig. 4 | **Protein assembly dynamics and proteoform-specific complex formation.** The dynamic assembly and disassembly of protein complexes, which is often specific to certain proteoforms, is a key mechanism that mediates functional diversity. **a** | When a cell is exposed to a change in conditions, such as hypoxia, oxidative stress, DNA damage, changes in nutrient availability (for example, low glucose) or stimulation by other signalling molecules (for example, insulin), this often causes the activation of specific protein-modifying enzymes. The subsequent post-translational modification of the enzyme's target proteins is frequently associated with the formation or disassembly of condition-specific protein complexes, thereby influencing the cell's functional landscape by regulating adaptive shifts between anabolic and catabolic metabolism, autophagy, apoptosis, cell cycle progression and cell motility. **b** | Proteoform-specific rearrangement of the interactome and its functional impact are exemplified by the phosphorylation (P)-dependent binding of 14-3-3 to its target proteins. The increase in the cellular AMP to ATP ratio under hypoxic conditions and subsequent activation of AMP-activated protein kinase (AMPK) results in a phosphorylation-dependent rearrangement of the 14-3-3 interactome. AMPK phosphorylates RAPTOR, a critical binding partner of mTOR complex 1 (mTORC1), which subsequently binds to 14-3-3. This binding causes the release of mTORC1 from RAPTOR, which contributes to mTORC1 inhibition and promotes an adaptive shift from anabolic to catabolic metabolism. The 14-3-3 protein also interacts with other AMPK targets downstream of mTORC1, including both Unc51-like autophagy activating kinase (ULK1) and autophagy-related protein 9A (ATG9A). The interaction of 14-3-3 with ULK1 and ATG9A promotes autophagy to sustain cellular metabolism under starvation conditions. A comprehensive review on phosphorylation-dependent remodelling of the 14-3-3 interactome in different cellular conditions is provided elsewhere[80].

of 14-3-3 protein dimers to their target proteins, which must usually harbour one or two phosphorylated Ser or Thr residues to interact with the 14-3-3 protein scaffold[80,91,92]. One exemplary cause for extensive 14-3-3 interactome rewiring is the hypoxia-induced activation of AMP-activated protein kinase (AMPK), resulting in the phosphorylation of several AMPK targets that can subsequently interact with 14-3-3, thereby causing an adaptive shift from anabolic to catabolic metabolism and promoting autophagy[80] (FIG. 4b).

To date, no system-wide information about the cross-dependence between protein complex formation and specific proteoforms is available. From a functional perspective, it would be interesting to study the extent to which alternative protein function can be attributed to actual changes in the interactome, proteome and transcriptome or combinations thereof. From a gene-centric perspective, it would be equally interesting to investigate to what degree alternative proteoforms derived from the same gene locus perform different functions, and what fraction of splice isoforms or PTMs are a product of stochastic noise without functional impact. Experimental strategies to systematically approach such questions have recently started to emerge; we discuss these methods later in the Review.

Crosstalk between molecular layers

Many examples of crosstalk and feedback loops have been discovered over the past few decades, where events traditionally considered 'downstream' affect 'upstream' targets. These examples include both positive and negative autoregulation, in which a gene product either

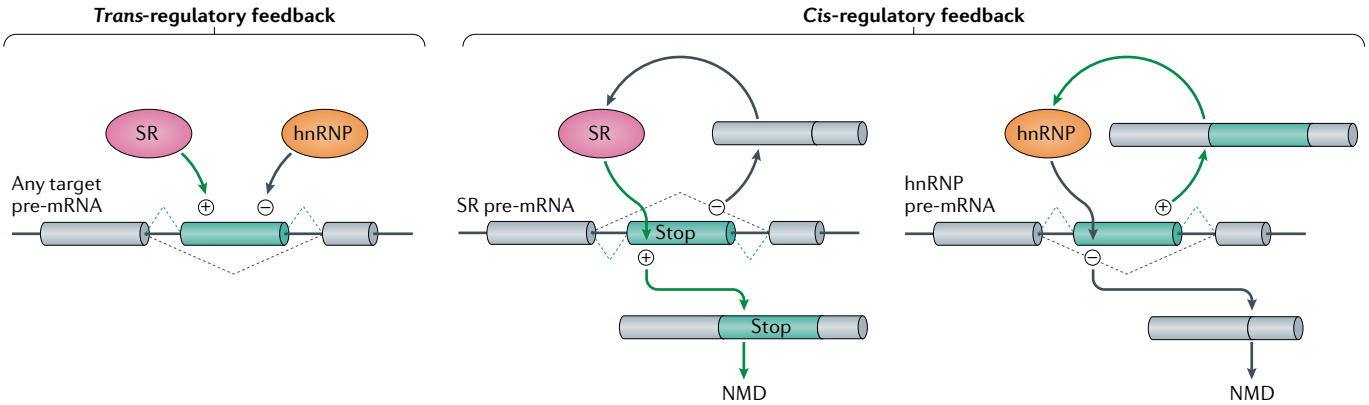


Fig. 5 | ***Cis*-regulatory and *trans*-regulatory feedback loops modulate alternative splicing.** Splicing is regulated by spliceosomal RNA-binding proteins such as serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs)[137]. Whereas SR proteins generally act as splicing activators, hnRNPs function as splicing repressors. Both types of proteins can regulate splicing of different target genes via *trans*-regulatory feedback. In addition, they can also regulate splicing of their own pre-mRNAs via negative *cis*-regulatory feedback. SR proteins can activate the splicing of an exon in their own pre-mRNA that carries a premature stop codon, thereby causing nonsense-mediated mRNA decay (NMD). hnRNPs repress splicing of an internal exon in their own pre-mRNA, which causes a frameshift that triggers NMD.

directly or indirectly regulates its own production. A prime example of a gene regulatory network that includes multiple transcriptional feedback loops is the circadian gene network in mammals[93]. Protein products of the core regulatory genes *PER1*, *PER2*, *CRY1* and *CRY2* interact with each other, forming a protein complex that represses their own transcription. Another mechanism for negative autoregulation is based on alternative splicing-coupled nonsense-mediated mRNA decay, a phenomenon expected to affect ~10–30% of mammalian genes[94]. SR proteins and hnRNPs, which have various roles in RNA processing and splicing, can bind to their own transcripts, for example, causing the production of a splice variant that is subjected to NMD, thereby downregulating protein levels and maintaining homeostatic protein expression (*cis*-regulatory feedback; FIG. 5). Even more prominent than *cis*-regulatory feedback loops are *trans*-regulatory mechanisms, exemplified by the activation or repression of splicing by SR proteins and hnRNPs that target pre-mRNAs of other genes (FIG. 5). Transcription factors that can regulate the expression of different target genes exemplify the most prominent form of *trans*-regulatory feedback. To date, only a small fraction of the regulatory networks inside the cell have been studied on a detailed mechanistic level and the relationships between different molecular layers are just beginning to be explored. Recent studies investigating quantitative trait loci on different molecular levels have revealed interesting insights into how multilayered molecular networks mediate the effects of genomic variants[95–98]. In the future, similar strategies comparable with quantitative trait locus analyses are likely to be used to investigate the expansion of molecular and functional diversity within a cell or organism. To enable the comprehensive mapping of such regulatory networks and to better understand the interdependence and crosstalk between different molecular layers and how these contribute to phenotypic diversity, high-quality measurements of all 'omics' layers are necessary.

Alternative splicing-coupled nonsense-mediated mRNA decay
A regulatory mechanism in which alternative splicing events introduce premature termination codons that lead to the downregulation of the transcript through the process of nonsense-mediated mRNA decay.

Quantitative trait loci
DNA loci that correlate with the variation of a quantitative trait in the phenotype.

Methods for generating 'omics' data

On the basis of the mechanisms explained above, eukaryotic cells can generate a large variety of mRNA isoforms. The comprehensive and quantitative profiling of these transcripts at single-base resolution has been made possible by technological advancements generally referred to as 'next-generation sequencing'. Recent developments in next-generation sequencing-based transcriptome profiling technologies have been extensively reviewed elsewhere[99–102]. Instead, we focus on the protein level, comparing established and recent proteomics-based and interactomics-based techniques.

Mass spectrometry-based proteomics. Most proteomic studies are based on mass spectrometry[103,104], with two main approaches: top-down and bottom-up proteomics. In top-down proteomics, intact proteins are chromatographically separated, ionized and injected into a mass spectrometer. The mass spectra of both the intact proteins and their fragmentation products are subsequently analysed, enabling the direct deduction of individual proteoforms with unique primary protein sequences and PTMs. However, top-down proteomic analysis of complex protein mixtures is challenging both analytically and in terms of required expertise because of limitations in proteoform separation techniques, issues with mass spectrometry analysis of high-molecular-mass ions, highly complex fragment ion patterns and difficulty with spectral interpretation using available analysis software[71,105]. Thus, for proteome-wide analyses, bottom-up proteomic workflows are more frequently applied. In bottom-up proteomics, proteins are first enzymatically digested into smaller peptide sequences, which are subsequently separated by liquid chromatography, ionized and analysed by tandem mass spectrometry, with both the peptide and the fragment ion spectra recorded. As the original connectivity between peptides and proteins is lost during enzymatic digestion, a main limitation in bottom-up proteomics is the required in

silico inference step, which maps measured peptide signals back to individual proteins. The requirement of protein inference further limits the ability of bottom-up proteomic workflows to differentiate proteoforms with the same primary protein structure but different combinations of PTMs. Although it is possible to gain useful information about the presence and localization of individual PTMs or sequence variants, as well as 'averaged' proteoforms (that is, a group of peptides mapping to the same canonical protein sequence, including alternative sequence variants or PTMs, the unique combination of which is not resolved), a robust strategy capable of confidently assigning and distinguishing unique proteoforms has, to our knowledge, not yet been reported for bottom-up proteomic approaches. Furthermore, different protein inference methods can vary in their assumptions[106] and can therefore only provide precise answers within the selected model and do not necessarily reflect reality. Nevertheless, the high sensitivity, reproducibility and scalability of bottom-up proteomic approaches still make them the methods of choice for proteome-wide studies. Different strategies for bottom-up proteomics have been reviewed previously[107–109] and further details on bottom-up proteomics and related opportunities for the derivation of proteoform-resolved information are given in Supplementary Box 1.

Complementation assays and mass spectrometry-based interactomics. Over the past few decades, several different methods have been developed to analyse transient and stable PPIs and protein complexes. Although many of these methods, for example, fluorescence resonance energy transfer[110] and bimolecular fluorescence complementation[111], are based on targeting a small proteome subset with a relatively low throughput, we specifically focus on recent system-wide strategies developed for the high-throughput detection of stable interactions and protein assemblies. To date, five main technologies are used: yeast two-hybrid screens (Y2H screens) and related complementation assays, AP-MS, proximity labelling approaches exemplified by BioID[112] and APEX[113], cross-linking mass spectrometry and protein co-fractionation coupled to mass spectrometry (CoFrac-MS). Recently, thermal proteome profiling and limited proteolysis-coupled mass spectrometry have emerged as strategies for the parallelized detection of multiple protein complexes in one experiment[77,114,115]. Here, we focus on the more established strategies listed above, summarized in FIG. 6a.

Until recently, Y2H screens were the main technique for systematically mapping PPI networks[116]. Despite many successful applications and methodological extensions to the classical Y2H screen assay, the system has several limitations. First, it requires that proteins are expressed in the yeast cell. Expression in the yeast intracellular environment may affect the protein structure, localization and PTM patterns, which could affect PPIs. Second, the fractions of false-positive and false-negative interactions in Y2H screens are hard to determine, and screens require validation assays in which a positive reference set and randomly selected protein pairs are used as controls[116,117]. Third, gaining a comprehensive network of PPIs requires a substantial number of experiments as each possible protein pair requires an independent experiment. Last, Y2H screens can only generate generic protein interaction maps and cannot be used to determine PPIs in a particular cellular state. To achieve condition-specific protein interaction maps, other techniques are necessary.

With advances in mass spectrometry-based proteomics, AP-MS has become the main technique for condition-resolved, large-scale PPI studies[81,118]. In AP-MS, a protein of interest (bait) is purified from a cell lysate using an antibody specific for the bait (which allows the use of non-engineered cells) or specific for an affinity tag fused to the bait. Pull-down is performed under mild, non-denaturing conditions, aiming to preserve stable PPIs between the bait protein and its interaction partners (prey proteins). The bait and its co-purified prey proteins are subsequently analysed by bottom-up mass spectrometry to identify and quantify the bait's stable interaction partners. AP-MS approaches often utilize background proteins commonly detected in AP-MS datasets to filter out false-positive hits and achieve high selectivity and specificity in determining confident interaction partners[119]. One AP-MS experiment can map multiple interactions in parallel, an advantage over Y2H screens. By combining high sensitivity in detecting PPIs at controlled false discovery rates with reasonable throughput, AP-MS is the method of choice for studying global interaction networks. Nevertheless, AP-MS studies are limited by the fact that they commonly require genetic engineering, thus potentially altering protein structure and interaction sites, or are dependent on the availability of specific antibodies. Additionally, AP-MS experiments can only capture stable interactions that remain intact throughout the experimental protocol. Despite its limitations, the AP-MS technique was applied to create the most comprehensive single-assay interaction map to date, containing >56,000 associations among >10,000 human proteins[81,82].

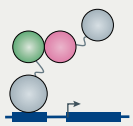
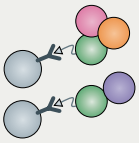
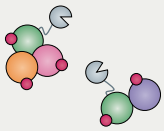
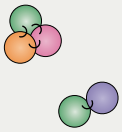
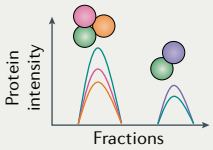
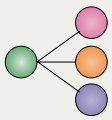
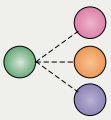
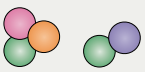
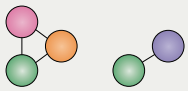
To capture transient interactions or general spatial proximity inside the cell, proximity labelling approaches exemplified by BioID[112] and APEX[113] strategies can be used. Here, the bait protein is fused to an enzyme, for example, a biotin ligase in BioID or a peroxidase in APEX, that labels all proteins in close proximity at both high spatial and high temporal resolution (reviewed in REFS[120,121]). In contrast to the other presented approaches, reported proteins do not have to physically interact with the bait but can simply be in close proximity to it. Similar to AP-MS, proximity labelling techniques can generate a global interaction network at steady state; however, if the whole interactome of a cellular system is of interest, studies focusing on dynamic changes in the proteome-wide PPI network are still limited by the proteome coverage of individual assays and the throughput of the technique. Nevertheless, BioID (combined with a complementary AP-MS analysis) has recently been applied to study the dynamic interaction landscape of the human centrosome–cilium interface during ciliogenesis, mapping >1,700 unique components and >7,000 interactions[122]. Recent method developments further improved the ease of combing AP-MS and BioID by using a single construct, termed MAC-tag, to investigate

Yeast two-hybrid screens
(Y2H screens). A strategy to identify protein–protein interactions in which one target protein is fused with a DNA-binding domain and a second target protein is fused with a respective transcriptional activation domain. When co-expressed in yeast, if the two target proteins interact in the nucleus, they initiate the transcription of a reporter gene.

Co-fractionation coupled to mass spectrometry
(CoFrac-MS). A strategy for the global mapping of protein–protein interactions and protein complexes, involving the native extraction and subsequent separation of protein complexes according to their physicochemical properties, followed by mass spectrometry analysis.

False discovery rates
A metric used for the control of the error rate in experiments affected by the multiple testing problem.

a

	Yeast two-hybrid screens	Affinity purification coupled to mass spectrometry	Proximity labelling (BioID, APEX)	Cross-linking coupled to mass spectrometry	Co-fractionation coupled to mass spectrometry
Screening approach					
Measurement	Yeast phenotype	MS/MS	MS/MS	MS/MS	MS/MS
Requirements	Genetic engineering of fusion genes	Genetic tagging or available antibody	Genetic tagging with enzyme	None	None
Type of interactions	Binary PPIs	Binary PPIs and complexes (only via reciprocal pull-downs)	Proximal proteins	Binary PPIs, complexes and structural information	Binary PPIs and complexes
	Stable interactions	Very stable interactions	Transient interactions (down to ~1 min labelling speed)	Short-term interactions (~1 h cross-linking time)	Very stable interactions
Information from single experiment					
Proteome coverage	Only two proteins per assay; high-throughput multiplexing is possible to achieve high proteome coverage	Proteome coverage depends on selected bait proteins	Proteome coverage depends on selected bait proteins	Good proteome coverage in a single assay	Good proteome coverage in a single assay

b

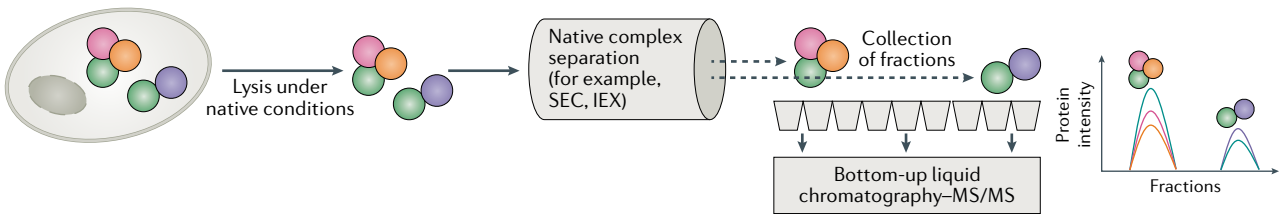


Fig. 6 | **Strategies for generating proteome-wide interactome maps.** Over the past few decades, several different methods have been developed to analyse transient and stable protein–protein interactions (PPIs) and protein complexes. **a** | Traditionally, PPIs (indicated by solid lines between proteins) were studied at medium throughput via yeast two-hybrid strategies. With advances in mass spectrometry-based proteomics, affinity purification coupled to mass spectrometry is now the state-of-the-art method for large-scale interactome mapping. To capture more transient interactions and spatial proximity, proximity labelling strategies such as BioID or APEX can be employed to identify proteins in close spatiotemporal proximity (indicated by dashed lines between proteins). Cross-linking mass spectrometry offers opportunities to detect protein complexes as well as their structural arrangement (indicated by proteins touching each other). Recently, protein co-fractionation coupled to mass spectrometry has been developed as a promising new strategy for the highly parallelized detection of protein complexes from a single fractionation experiment. **b** | For co-fractionation coupled to mass spectrometry experiments, cells are first lysed under native conditions to keep protein complexes intact. The complexes are subsequently separated and fractionated according to their physicochemical properties, for example, by size-exclusion chromatography (SEC) or ion-exchange chromatography (IEX). All sampled fractions are separately analysed by bottom-up mass spectrometry (liquid chromatography with tandem mass spectrometry (MS/MS)). PPIs and protein complexes can be inferred from the local co-elution and high correlation of proteins along the chromatographic dimension. Solid lines indicate the order of processing steps, and dashed lines illustrate the separation process.

interactions, cellular localization and spatial distances within a protein complex in parallel[123].

Cross-linking mass spectrometry has been employed mostly to gain low-resolution structural information from purified proteins or protein complexes via the generation of distance constraints, which provide upper limits to the distance between amino acids in a molecule. Recently, the technique was expanded to more complex applications, including cross-linking the full proteome of *Escherichia coli*, the human HeLa cell line[75] and *Drosophila melanogaster* embryos[76], to identify PPIs and to gain structural information on a system-wide level.

An alternative approach to gaining proteome-wide connectivity information is based on CoFrac-MS[124–130] (FIG. 6b). In typical CoFrac-MS workflows, cells are lysed at near-native conditions to keep most PPIs and protein complexes intact. The resulting cell lysate is subsequently separated and fractionated according to the physicochemical properties of the protein assemblies, while aiming to keep protein complexes intact throughout the fractionation process. Examples of separation techniques include ion-exchange chromatography, for separation by charge, and size-exclusion chromatography, for separation by hydrodynamic radius. For bottom-up proteomic

Graph partitioning
The process of dividing a graph (for example, a protein–protein interaction network) into smaller connected subnetworks.

analysis, each of the sampled fractions is enzymatically digested into peptides and subsequently analysed by liquid chromatography with tandem mass spectrometry. The underlying hypothesis in CoFrac-MS is that interacting proteins at least partially comigrate during chromatographic separation, and that they will have similar quantitative protein profiles based on liquid chromatography with tandem mass spectrometry detection and quantification. Interacting proteins can therefore be identified based on their highly correlated elution signal. The main advantage of CoFrac-MS workflows over AP-MS studies is that, in principle, they allow the analysis of thousands of proteins and their interactions without the need for genetic engineering or specific antibodies. With accurate relative quantification, the approach allows differential analysis between samples and the identification of changes in the detected PPIs across cellular states. CoFrac-MS datasets are most commonly employed for de novo PPI and protein complex detection, inferred from the pairwise correlation of protein elution profiles and the use of graph partitioning algorithms that predict protein complex assembly[131,132]. However, the sensitivity and selectivity of such correlation-based de novo protein complex assignments are limited by the peak capacity of the chosen chromatographic separation technique, as non-interacting proteins also chromatographically co-elute if the number of detected proteins markedly exceeds the peak capacity of the column. To overcome the challenge of limited selectivity and to enable a more quantitative assessment of protein complex assembly states in CoFrac-MS datasets, we have recently presented a novel, targeted, complex-centric analysis concept[133]. Here, the CoFrac-MS data are directly queried for evidence of a priori-defined protein complexes using data from public protein complex databases, thereby markedly increasing selectivity. This analysis strategy provides a promising new approach for probing and quantifying the rewiring of protein assemblies across different conditions[134].

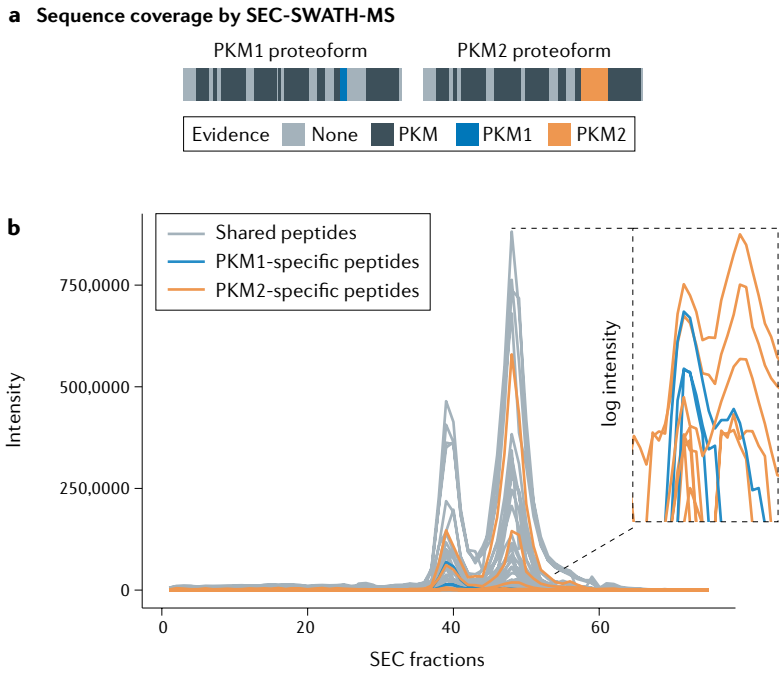


Fig. 7 | **Proteoform-specific assembly characteristics of PKM.** Peptide-level analysis of data from native complex co-fractionation coupled to mass spectrometry confirms proteoform-specific assembly characteristics of PKM. The data shown were generated by the fractionation of a native human cell line lysate by size-exclusion chromatography followed by data-independent acquisition mass spectrometry (SEC-SWATH-MS). **a** | Assessing the two pyruvate kinase PKM isoforms at the protein level reveals high sequence coverage in the SEC-SWATH-MS data. Unique peptides for both PKM1 (blue) and PKM2 (orange) can be detected. Sequence regions that were not detected in the acquired data are light grey, and regions with shared peptide evidence for both PKM1 and PKM2 are dark grey. **b** | Elution profiles of detected PKM peptides along the chromatographic dimension show two distinct elution peaks. The two peaks probably correspond to the PKM dimer (right) and the PKM tetramer (left). Whereas unique peptides for PKM2 (orange) contribute to both assembly states, unique peptides for PKM1 (blue) are only detected in the tetrameric conformation. This figure was generated based on previously unpublished in-house data generated by a similar method for SEC-SWATH-MS data acquisition and analysis as described in REFS[133,134].

Proteoform-specific analysis of protein interactions and assemblies. One of the major questions in protein biology is to what extent specific proteoforms and the proteins' interactome are connected or interdependent. As most large-scale interactome studies have focused on bottom-up proteomic strategies, the inference of completely characterized proteoforms (that is, the characterization of the precise amino acid sequence and the location of all PTMs) and their individual interaction landscape are still elusive. However, individual modifications, such as PTMs at a specific peptide or even a specific amino acid site, have already been associated with the protein's interaction network in large-scale studies. For example, an AP-MS study of human kinases integrated information on the interaction landscape of 57 human kinases with the phosphorylation state of these kinases and their interaction partners[135]. Furthermore, the effect of caspase-mediated proteolysis on the interactome of Jurkat cells was analysed using a CoFrac-MS approach[136]. CoFrac-MS strategies are a promising tool for systematically assessing the relationship between proteoforms and specific protein complex assembly states of a protein. By performing a peptide-level analysis of the co-elution patterns along the dimension of protein complex separation, groups of similarly behaving peptides that potentially originate from alternative proteoforms can be distinguished (FIG. 7). However, a systematic study of expressed proteoforms and their global effect on functional diversification, as can be indicated by the assembly into different protein complexes, has not yet been performed.

Conclusions and perspectives

Research over the past few decades has revealed that the ability of living systems to adjust their functional landscape originates from multiple molecular layers beyond the protein-coding genome. In addition to regulating the quantitative composition of both the transcriptome and the proteome, which have been the intense focus of large-scale research efforts, cells can also adjust the diversity of transcripts and proteoforms and the integration of proteoforms into different protein complexes. Technological developments, especially in the area of mass spectrometry, recently paved the way for enabling a systematic analysis of the molecular diversity

of a cellular system on all different analyte levels, both in the form of generic maps and in the form of differential analyses to detect state-specific differences.

Although each gene is estimated to have ~100 proteoforms[65] on average, the link between diversity at the proteome level and the functional diversity of the cell has not yet been systematically assessed. A promising direction towards evaluating the effect of different proteoforms is to investigate their interdependence with other properties, such as protein localization or protein complex formation. This approach could, for example, be achieved by peptide-level analysis of CoFrac-MS datasets, resulting in data that should contain valuable information about the crosstalk between the protein assembly state and specific proteoforms.

Although protein complex-level information is expected to provide the best estimation for the causes of functional diversity, there is still a lack of methods to systematically probe functional diversity directly. With further improvements in 'omics' technologies and the convergence and integration of methods, such as top-down and bottom-up proteomic approaches, we can speculate that the majority of proteoforms and protein complexes will be mapped within the next decade. This mapping will enable a more systematic evaluation of the interdependence and crosstalk between the different molecular layers and the reconstruction of complex, non-linear regulatory networks. The overarching goal will remain to resolve the question of how molecular diversity translates to functional diversity in order to better understand the relationship between genotype and phenotype under different experimental and clinical conditions.

Published online 31 March 2020

1. Beadle, G. W. & Tatum, E. L. Genetic control of biochemical reactions in neurospora. *Proc. Natl Acad. Sci. USA* **27**, 499–506 (1941).
2. Collins, F. S., Lander, E. S., Rogers, J. & Waterson, R. H. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
3. Carter, H., Hofree, M. & Ideker, T. Genotype to phenotype via network analysis. *Curr. Opin. Genet. Dev.* **23**, 611–621 (2013).
4. Rauscher, B., Valentini, E., Hardeland, U. & Boutros, M. Phenotype databases for genetic screens in human cells. *J. Biotechnol.* **261**, 63–69 (2017).
5. Amberger, J., Bocchini, C. & Hamosh, A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). *Hum. Mutat.* **32**, 564–567 (2011).
6. Arkin, A. P. & Schaffer, D. V. Network news: innovations in 21st century systems biology. *Cell* **144**, 844–849 (2011).
7. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
8. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
9. Li, J. J., Bickel, P. J. & Biggin, M. D. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2**, e270 (2014).
10. Jovanovic, M. et al. Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038 (2015).
11. Marguerat, S. et al. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151**, 671–683 (2012).
12. Thompson, D., Regev, A. & Roy, S. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annu. Rev. Cell Dev. Biol.* **31**, 399–428 (2015).
13. Bernadskaya, Y. & Christiaen, L. Transcriptional control of developmental cell behaviors. *Annu. Rev. Cell Dev. Biol.* **32**, 77–101 (2016).
14. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* **17**, 227–239 (2016).
15. Magnuson, B., Bedi, K. & Ljungman, M. Genome stability versus transcript diversity. *DNA Repair* **44**, 81–86 (2016).
This review discusses sources of error from the genome to the proteome and how errors can contribute to molecular diversity and evolution.
16. Schneider, C., Kudla, G., Wlotzka, W., Tuck, A. & Tollervey, D. Transcriptome-wide analysis of exosome targets. *Mol. Cell* **48**, 422–433 (2012).
17. Popp, M. W. & Maquat, L. E. Nonsense-mediated mRNA decay and cancer. *Curr. Opin. Genet. Dev.* **48**, 44–50 (2018).
18. Wong, J. J. L. et al. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* **154**, 583–595 (2013).
19. Kurosaki, T., Popp, M. W. & Maquat, L. E. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* **20**, 406–420 (2019).
20. Machnicka, M. A. et al. MODOMICS: a database of RNA modification pathways — 2013 update. *Nucleic Acids Res.* **41**, D262–D267 (2012).
21. Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.* **18**, 31–42 (2017).
22. Ranjan, N. & Leidel, S. A. The epitranscriptome in translation regulation: mRNA and tRNA modifications as the two sides of the same coin? *FEBS Lett.* **593**, 1483–1493 (2019).
23. Zaccara, S., Ries, R. J. & Jaffrey, S. R. Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* **20**, 608–624 (2019).
24. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
25. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175 (2014).
This comprehensive review discusses the mechanisms and interdependencies of mRNA transcription and processing.
26. de Klerk, E. & 't Hoen, P. A. C. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet.* **31**, 128–139 (2015).
27. Garieri, M. et al. The effect of genetic variation on promoter usage and enhancer activity. *Nat. Commun.* **8**, 1358 (2017).
28. Reyes, A. & Huber, W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* **46**, 582–592 (2018).
29. Proudfoot, N. J., Furger, A. & Dye, M. J. Integrating mRNA processing with transcription. *Cell* **108**, 501–512 (2002).
30. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
31. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
32. Tress, M. L. et al. The implications of alternative splicing in the ENCODE protein complement. *Proc. Natl Acad. Sci. USA* **104**, 5495–5500 (2007).
33. Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* **18**, 655–670 (2017).
34. Lee, Y. & Rio, D. C. Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* **84**, 291–323 (2015).
35. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
36. Costa, V., Aprile, M., Esposito, R. & Ciccodicola, A. RNA-seq and human complex diseases: recent accomplishments and future perspectives. *Eur. J. Hum. Genet.* **21**, 134–142 (2013).
37. Pistoni, M., Ghigna, C. & Gabellini, D. Alternative splicing and muscular dystrophy. *RNA Biol.* **7**, 441–452 (2010).
38. Chaneton, B. & Gottlieb, E. Rocking cell metabolism: revised functions of the key glycolytic regulator PKM2 in cancer. *Trends Biochem. Sci.* **37**, 309–316 (2012).
39. Buljan, M. et al. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* **46**, 871–883 (2012).
40. Yablonovitch, A. L., Deng, P., Jacobson, D. & Li, J. B. The evolution and adaptation of A-to-I RNA editing. *PLoS Genet.* **13**, e1007064 (2017).
41. Picardi, E., D'Erchia, A. M., Lo Giudice, C. & Pesole, G. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* **45**, D750–D757 (2017).
42. Maas, S., Kawahara, Y., Tamburro, K. M. & Nishikura, K. A-to-I RNA editing and human disease. *RNA Biol.* **3**, 1–9 (2006).
43. Tang, S. J. et al. *Cis*- and *trans*-regulations of pre-mRNA splicing by RNA editing enzymes influence cancer development. *Nat. Commun.* **11**, 799 (2020).
44. Zipeto, M. A., Jiang, Q., Melese, E. & Jamieson, C. H. M. RNA rewriting, recoding, and rewiring in human disease. *Trends Mol. Med.* **21**, 549–559 (2015).
45. Benne, R. The long and the short of it. *Nature* **380**, 391–392 (1996).
46. Tress, M. L., Abascal, F. & Valencia, A. Alternative splicing may not be the key to proteome complexity. *Trends Biochem. Sci.* **42**, 98–110 (2017).
47. Wan, Y. & Larson, D. R. Splicing heterogeneity: separating signal from noise. *Genome Biol.* **19**, 86 (2018).
48. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
This study discusses the relationship between mRNA abundance and protein levels.
49. Dikic, I. Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.* **86**, 193–224 (2017).
50. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl Acad. Sci. USA* **116**, 358–366 (2019).
51. Harper, J. W. & Bennett, E. J. Proteome complexity and the forces that drive proteome imbalance. *Nature* **537**, 328–338 (2016).
This review discusses the regulation and balancing of protein synthesis and degradation.
52. Zhang, B. et al. Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382–387 (2014).
53. Liu, Y. et al. Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat. Biotechnol.* **37**, 314–322 (2019).
54. Beyer, A., Hollunder, J., Nasheuer, H.-P. & Wilhelm, T. Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. *Mol. Cell. Proteomics* **3**, 1083–1092 (2004).
55. Malmström, J. et al. Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* **460**, 762–765 (2009).
56. Milo, R. What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays* **35**, 1050–1055 (2013).
57. Brar, G. A. Beyond the triplet code: context cues transform translation. *Cell* **167**, 1681–1692 (2016).
58. Kearse, M. G. & Wilusz, J. E. Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.* **31**, 1717–1731 (2017).

59. Loftfield, R. B. & Vanderjagt, D. The frequency of errors in protein biosynthesis. *Biochem. J.* **128**, 1353–1356 (1972).
60. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342–7372 (2005).
61. Wolan, D. W., Zorn, J. A., Gray, D. C. & Wells, J. A. Small-molecule activators of a proenzyme. *Science* **326**, 853–858 (2009).
62. Shalini, S., Dorstyn, L., Dawar, S. & Kumar, S. Old, new and emerging functions of caspases. *Cell Death Differ.* **22**, 526–539 (2015).
63. Puente, X. S., Sánchez, L. M., Overall, C. M. & López-Otín, C. Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.* **4**, 544–558 (2003).
64. 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.* **1**, 90 (2011).
65. Aebersold, R. et al. How many human proteoforms are there? *Nat. Chem. Biol.* **14**, 206–214 (2018).
This study presents the current estimation of how many proteoforms are expressed in humans.
66. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
67. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355–365 (2001).
68. Wu, Z., Huang, R. & Yuan, L. Crosstalk of intracellular post-translational modifications in cancer. *Arch. Biochem. Biophys.* **676**, 108138 (2019).
69. Kim, M.-S. et al. A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
70. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
71. Toby, T. K., Fornelli, L. & Kelleher, N. L. Progress in top-down proteomics and the analysis of proteoforms. *Annu. Rev. Anal. Chem.* **9**, 499–519 (2016).
72. Tran, J. C. et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254–258 (2011).
73. Anderson, L. C. et al. Identification and characterization of human proteoforms by top-down LC-21 Tesla FT-ICR mass spectrometry. *J. Proteome Res.* **16**, 1087–1096 (2017).
74. Kelemen, O. et al. Function of alternative splicing. *Gene* **514**, 1–30 (2013).
75. Liu, F., Lössl, P., Scheltema, R., Viner, R. & Heck, A. J. R. Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.* **8**, 15473 (2017).
76. Götze, M., Iacobucci, C., Ihling, C. H. & Sinz, A. A simple cross-linking/mass spectrometry workflow for studying system-wide protein interactions. *Anal. Chem.* **91**, 10236–10244 (2019).
77. Savitski, M. M. et al. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784 (2014).
78. Feng, Y. et al. Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **32**, 1036–1044 (2014).
79. Sluchanko, N. N. & Gusev, N. B. Moonlighting chaperone-like activity of the universal regulatory 14-3-3 proteins. *FEBS J.* **284**, 1279–1295 (2017).
80. Pennington, K., Chan, T., Torres, M. & Andersen, J. The dynamic and stress-adaptive signaling hub of 14-3-3: emerging mechanisms of regulation and context-dependent protein–protein interactions. *Oncogene* **37**, 5587–5604 (2018).
81. Huttlin, E. L. et al. The BioPlex network: a systematic exploration of the human interactome. *Cell* **162**, 425–440 (2015).
82. Huttlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).
This study describes the largest human PPI network generated by AP-MS to date.
83. Drew, K. et al. Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol. Syst. Biol.* **13**, 932 (2017).
84. Ruepp, A. et al. CORUM: the comprehensive resource of mammalian protein complexes — 2009. *Nucleic Acids Res.* **38**, D497–D501 (2009).
85. Giurgiu, M. et al. CORUM: the comprehensive resource of mammalian protein complexes — 2019. *Nucleic Acids Res.* **47**, D559–D563 (2019).
86. Gaulton, A. et al. The Complex Portal — an encyclopaedia of macromolecular complexes. *Nucleic Acids Res.* **43**, D479–D484 (2014).
87. Casanova, E. B. et al. Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. *Nucleic Acids Res.* **47**, D550–D558 (2018).
88. Marsh, J. A. & Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **84**, 551–575 (2015).
89. Levy, E. D. & Pereira-Leal, J. B. Evolution and dynamics of protein interactions and networks. *Curr. Opin. Struct. Biol.* **18**, 349–357 (2008).
90. Ellis, R. J. Molecular chaperones: assisting assembly in addition to folding. *Trends Biochem. Sci.* **31**, 395–401 (2006).
91. Chen, S., Synowsky, S., Tinti, M. & MacKintosh, C. The capture of phosphoproteins by 14-3-3 proteins mediates actions of insulin. *Trends Endocrinol. Metab.* **22**, 429–436 (2011).
92. Collins, B. C. et al. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat. Methods* **10**, 1246–1253 (2013).
93. Takahashi, J. S. Transcriptional architecture of the mammalian circadian clock. *Nat. Rev. Genet.* **18**, 164–179 (2017).
94. Jangi, M. & Sharp, P. A. Building robust transcriptomes with master splicing factors. *Cell* **159**, 487–498 (2014).
95. Großbach, J. et al. Integration of transcriptome, proteome and phosphoproteome data elucidates the genetic control of molecular networks. Preprint at https://doi.org/10.1101/703140 (2019).
This recent study investigates quantitative trait loci on different molecular levels and how they mediate the effects of genomic variants in multilayered molecular networks.
96. Williams, E. G. et al. Systems proteomics of liver mitochondria function. *Science* **352**, aad0189 (2016).
97. Wu, Y. et al. Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. *Cell* **158**, 1415–1430 (2014).
98. Picotti, P. et al. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494**, 266–270 (2013).
99. Hrdlickova, R., Toloue, M. & Tian, B. RNA-seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA* **8**, e1364 (2017).
100. Levy, S. E. & Myers, R. M. Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* **17**, 95–115 (2016).
101. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol. Cell* **58**, 586–597 (2015).
102. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
103. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
104. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
105. Schaffer, L. V. et al. Identification and quantification of proteoforms by mass spectrometry. *Proteomics* **19**, 1800361 (2019).
This recent review discusses the current state of proteoform identification and quantification by top-down proteomics.
106. He, Z., Huang, T., Zhao, C. & Teng, B. in *Modern Proteomics — Sample Preparation, Analysis and Practical Applications* (eds Mirzaei, H. & Carrasco, M.) 237–242 (Springer, 2016).
107. Han, X., Aslanian, A. & Yates, J. R. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **12**, 483–490 (2008).
108. Bantscheff, M., Lemeer, S., Savitski, M. M. & Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **404**, 939–965 (2012).
109. Gillet, L. C., Leitner, A. & Aebersold, R. Mass spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing. *Annu. Rev. Anal. Chem.* **9**, 449–472 (2016).
This in-depth review discusses current mass spectrometry techniques for bottom-up proteomics.
110. Bunt, G. & Wouters, F. S. FRET from single to multiplexed signaling events. *Biophys. Rev.* **9**, 119–129 (2017).
111. Hu, C. D., Chinenov, Y. & Kerppola, T. K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **9**, 789–798 (2002).
112. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
113. Martell, J. D. et al. Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30**, 1143–1148 (2012).
114. Becher, I. et al. Pervasive protein thermal stability variation during the cell cycle. *Cell* **173**, 1495–1507. e18 (2018).
115. Piazza, I. et al. A map of protein–metabolite interactions reveals principles of chemical communication. *Cell* **172**, 358–372.e23 (2018).
116. Rolland, T. et al. A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).
117. Braun, P. et al. An experimentally derived confidence score for binary protein–protein interactions. *Nat. Methods* **6**, 91–97 (2009).
118. Hein, M. Y. et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
119. Mellacheruvu, D. et al. The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).
120. Trinkle-Mulcahy, L. Recent advances in proximity-based labeling methods for interactome mapping. *F1000Res.* **8**, 135 (2019).
121. Gingras, A. C., Abe, K. T. & Raught, B. Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Curr. Opin. Chem. Biol.* **48**, 44–54 (2019).
This excellent review discusses the current state of proximity labelling techniques to analyse protein complexes.
122. Gupta, G. D. et al. A dynamic protein interaction landscape of the human centrosome–cilium interface. *Cell* **163**, 1484–1499 (2015).
123. Liu, X. et al. An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat. Commun.* **9**, 1188 (2018).
124. Liu, X., Yang, W., Gao, Q. & Regnier, F. Toward chromatographic analysis of interacting protein networks. *J. Chromatogr. A* **1178**, 24–32 (2008).
125. Dong, M. et al. A "tagless" strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking. *J. Proteome Res.* **7**, 1836–1849 (2008).
126. Kristensen, A. R., Gsponer, J. & Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nat. Methods* **9**, 907–909 (2012).
127. Kristensen, A. R. & Foster, L. J. in *Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC)* (ed. Warscheid, B.) 263–270 (Humana Press, 2014).
128. Havugimana, P. C. et al. A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012).
129. Wan, C. et al. Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344 (2015).
130. Salas, D., Stacey, R. G., Akinlaja, M. & Foster, L. J. Next-generation interactomics: considerations for the use of co-elution to measure protein interaction networks. *Mol. Cell. Proteomics* **19**, 1–10 (2020).
This recent review discusses the techniques, limitations and possibilities of co-fractionation mass spectrometry approaches for PPI and protein complex mapping.
131. Scott, N. E., Brown, L. M., Kristensen, A. R. & Foster, L. J. Development of a computational framework for the analysis of protein correlation profiling and spatial proteomics experiments. *J. Proteomics* **118**, 112–129 (2015).
132. Hu, L. Z. et al. EPIC: software toolkit for elution profile-based inference of protein complexes. *Nat. Methods* **16**, 737–742 (2019).
133. Heusel, M. et al. Complex-centric proteome profiling by SEC-SWATH-MS. *Mol. Syst. Biol.* **15**, e8438 (2019).
134. Heusel, M. et al. A global screen for assembly state changes of the mitotic proteome by SEC-SWATH-MS. *Cell Syst.* **10**, 133–155.e6 (2020).

135. Varjosalo, M. et al. The protein interaction landscape of the human CMGC kinase group. *Cell Rep.* **3**, 1306–1320 (2013).
136. Scott, N. E. et al. Interactome disassembly during apoptosis occurs independent of caspase cleavage. *Mol. Syst. Biol.* **13**, 906 (2017).
137. Ni, J. Z. et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).

Acknowledgements
The authors thank all members of the Aebersold laboratory who provided input to the content and presentation of the Review. The authors further thank B. Collins, L. Sieverling, J. Čuklina, C. Dörig, U. Kutay and M. Claassen for reading and providing feedback on different parts of the Review. R.A. acknowledges funding support from the SystemsX.ch projects PhosphoNetX PPM and TbX, as well as from the European Research Council (ERC-20140AdG 670821). I.B. acknowledges funding support from the Swiss National Science Foundation (31003A_166435).

Author contributions
I.B. researched data for the review. I.B. and R.A. wrote and edited the manuscript before submission.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Molecular Cell Biology thanks Ivan Marazzi and the other, anonymous, reviewers for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41580-020-0231-2.

RELATED LINKS
Complex Portal: https://www.ebi.ac.uk/complexportal/home
CORUM: https://mips.helmholtz-muenchen.de/corum/
GENCODE database: https://www.gencodegenes.org/human/stats.html