Chapter 8
The Interplay Between the Transcriptomics and Proteomics Profiles

John Oluwafemi Teibo, Virgínia Campos Silvestrini, Alessandra P. Vargas, Guilherme Pauperio Lanfredi, and Vítor Marcel Faça

8.1 Introduction

Living organisms have complex physiology, with extremely regulated systems to modulate responses to internal and external stimuli, allowing adaptability in the environment in which they live. These processes involve constant synthesis and degradation of biomolecules as a response to cellular events. The same occurs in pathological situations, where the abnormal stages of development of a disease are carried out by important changes in the set of biomolecules responsible for cell or organism function. Genetic products, mainly mRNAs and proteins, are constantly being modulated in response to normal physiological and pathological cellular events. Therefore, effective monitoring of the cellular or organism complement of mRNAs and proteins, namely the transcriptome and proteome, respectively, are fundamental to understand normal as well as pathological molecular mechanisms in the cells.

Technological advances in genomic sciences, including modern tools for transcriptomics and proteomics, have been recognized as important drivers of biosciences, allowing significant scientific discoveries and biological advances in the last decade. The genomics, transcriptomics, and proteomics now can routinely provide information on cell mutations, disease biomarker, gene therapies, with particular direct impact in personalized medicine applications. More importantly, these approaches can be used coordinately in the same study, providing deep molecular profiles in healthy and diseased situations (Manzoni et al. 2018).

Great scientific advances started with genomics. However, despite being revolutionary in the beginning of the century, the study of the genome does not respond to

J. O. Teibo · V. C. Silvestrini · A. P. Vargas · G. P. Lanfredi · V. M. Faça (✉)
Department of Biochemistry and Immunology and Center for Cell Based Therapy – Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil
e-mail: vitor.faca@fmrp.usp.br

G. A. Passos (ed.), *Transcriptomics in Health and Disease*,
https://doi.org/10.1007/978-3-030-87821-4_8

all the challenges and questions posed by biology especially in human health and disease areas. Many regulatory mechanisms are orchestrated between the genome transcription and the translation of proteins, which are responsible for most of the phenotypic characteristics of an organism (Buccitelli and Selbach 2020). Based on this principle, other technologies and approaches emerged with the aim of identifying a complete set of transcripts and also proteins in a given biological system. The integration of these "omics" approaches can be the key to understanding complex data, helping to generate more complete hypotheses in several areas of biology.

From the earlier studies comparing mRNA and protein abundancy in some biological models, it was obvious the lack of full concordance observed in high-throughput experiments (Anderson and Seilhamer 1997; Gygi et al. 1999). On the biological front, differences could be initially attributed to RNA splicing, differential RNA and protein turnover, post-translational modifications, allosteric protein interactions, and proteolytic processing events. On the experimental front, challenges in experimental design and data interpretation, as well as technological limitations, contributed to some of the differences observed (Hegde et al. 2003).

In this chapter, we will explore the biological and technical factors that affect the concordances and differences already established for the interplay of transcriptomes and proteomes in biological system. After a couple of decades with scientific and technological advances in the "omics" field, researchers have been elucidating some new players and factors responsible for the imbalance between mRNA and proteins, some of the technical limitations has been overcome, data generation and processing and more importantly, how and when data from both transcriptomics and proteomics can be integrated have been improved and properly compared in order to draw hypothesis and conclusions that now can form the basis of pathways involved in health and diseases processes in various biological systems.

8.2 Transcriptomics

The analysis of the entire transcriptome (mRNA, tRNA, rRNA, and miRNA) has become an essential tool in the quantification of gene expression in different tissues, organs, and cells, previously identified only by DNA sequencing. Transcriptomics can routinely provide an overview of the characteristics of gene expression in different samples, determining the presence/absence and quantification of transcripts. These transcripts profiles provide the basis for understanding regulatory pathways that control cell function, growth, and development in different biological systems. This information is also essential for understanding the metabolic and tissue dynamics, especially in comparisons among physiological and pathological states (Jiang et al. 2015).

The dominant contemporary techniques used for transcripts profiling are RNA microarray and RNAseq, both with their distinct advantages and disadvantages. Microarrays measure the abundances of an established set of transcripts via their hybridization with an array of complementary probes, allowing the analysis of

thousands of transcripts at a low cost (Lowe et al. 2017). The abundance of transcription is determined by hybridization of fluorescently labeled transcripts to these know probes. Importantly, this approach is based on a defined set of known sequences, to generate the profiles for the array (Barbulovic-Nad et al. 2006; Lowe et al. 2017). Current commercially available mRNA microarrays can profile virtually the entire transcriptomes for several different organisms.

More recently, RNAseq, which refers to the complete sequencing of the transcriptome, determines the abundance of mRNA from the number of counts from each transcript (Morozova et al. 2009). The first paper published using this technique was in 2006 with 105 transcripts sequenced and that provided sufficient sequence coverage to quantify their relative abundance (Bainbridge et al. 2006). RNAseq became more established and robust from 2008 with the emergence of next generation sequencing and massive sequencing by synthesis (SBS) technology, which now is sufficient for accurate quantitation of the entire human transcriptome (Lappalainen et al. 2013). These approaches have led to rapid expansion of this technology to answer many biological questions revolving around the transcriptomes for health and diseased biological problems.

For both the microarray and RNAseq strategies, it is initially necessary to purify the RNA and convert it into complementary DNA (cDNA). Subsequently, cDNAs are chemically marked with fluorophores and hybridized to probes on the chip to detect present target genes, if the technique used is RNA-microarray. On the other hand, if the strategy used is that of RNAseq, the RNA can also be fragmented to build a library for sequencing analysis. Both strategies must be executed through the platform of choice according to the specific objectives of each experiment or study (Nagalakshmi et al. 2010; Kukurba and Montgomery 2015; Manzoni et al. 2018).

Microarrays are now a robust technique and they are commercially available for complete genomic coverage using optimized sets of probes. However, transcripts not included in the probes will not be observed. More complex organisms have a greater number of exons and also non-coding sequences (introns). In this case, direct sequencing of mRNA molecules can provide more information about transcription products and potential translation products with greater coverage. Based on this principle, RNAseq does not require prior knowledge of the transcripts in the sample, making it possible to compare different sets of genes. This unique feature has enhanced the discovery of novel transcript products from the gene expression process. The microarray technique, on the other hand, has a lower cost and allows a higher number of replicates, necessary for confident new discoveries (Nagalakshmi et al. 2010).

The integration of transcriptomics with other “omics” technologies can help to grasp the complexity of cell life. Transcriptomics tools permit the parallel quantification of thousands of biomolecules and therefore allow for explorative, non-hypothesis-driven studies. However, there are several sources of variability originating from biological and technical causes that can affect the quality of the resulting data, such as biological heterogeneity in the sample, sample collection variations, RNA quantity and quality obtained from preparation steps, technical variation during sample processing, and batch effects, among others. Some of these

issues can be avoided with an appropriate and carefully experimental design that controls for the different sources of variation, but others will be detected only after a quality assessment of the raw data through computational support tools. Therefore, regardless of the technology used to measure gene expression of a cell, ensuring quality control is a critical starting point for any subsequent analysis of the data (Cobb et al. 2005; Larkin et al. 2005; Irizarry et al. 2005; Heber and Sick 2006). As we will discuss ahead, other technologies such as proteomics also present advantages and technical limitations. These potential limitations are the initial key factors to be eliminated or at a least minimized to allow multi-omics platforms data integration and comparison to provide more information toward answering complex biological questions.

8.3 Proteomics

Proteomics consists of suite of techniques that allow proteome analysis, making possible to identify proteins and quantify their abundance and post-translational modifications (PTMs) in a complete and complex set of different samples, including cells, tissues, and fluids, among others (Faça 2017). Unlike the transcriptomics strategies mentioned above, proteomics provides direct measurements on active and post-translationally modified proteins, in addition to their cell expression and localization. This kind of information is essential during the development of several pathologies, since biochemical processes such as splicing, phosphorylation, ubiquitination and other PTMs are usually severely impacted. Thus, proteomics studies also provide information on altered pathways, contributing to the discovery of important biological targets in the emergence of diseases (Silvestrini et al. 2019). Therefore, proteomic strategies are important to complement genomic and transcriptomic information (Aslam et al. 2017; Silvestrini et al. 2019).

When studying the proteome, an increase in the degree of molecular complexity in relation to the genomic study must be considered. The four-nucleotide codes of DNA and mRNA are translated into a complex code of 20 amino acids with different combinations, forming primary sequences that can adopt specific chemical conformations and modifications to produce a functional protein (Manzoni et al. 2018). The proteome is a multidimensional and highly dynamic system, in which each protein has several interconnected properties that together represent the phenotype of a cell or organism.

Although some of the underlying technology for quantifying protein abundance was introduced more than 40 years ago (O'Farrell 1975; Klose 1975), there has been recently a significant advance in the field and the development of new tools. With the advances in mass spectrometry which focus on studying proteomes, cell location, synthesis/degradation, Post-translational modifications, (PTMs) etc. has began to be analyzed in an integrated manner, allowing a better understanding of physiological and cellular processes (Larance and Lamond 2015).

The need to understand cellular changes at the protein level has led to the emergence of more accurate, high-quality proteomic strategies that guarantee sensitivity for the simultaneous identification and quantification of thousands of proteins in a sample. Most common proteomic studies in disease development are based on liquid-chromatography coupled with high-throughput mass spectrometry (LC-MS/MS) technology. In particular, LC-MS/MS has enabled the structural characterization of proteins and protein complexes that have been intractable through other methods, providing experimental evidence with high resolution (Chandramouli and Qian 2009). Currently, two principles of analysis are used: a global and targeted proteomics approach (Chandramouli and Qian 2009). In global proteomics or also called shotgun sequencing or whole proteome analysis, there is no hypothesis of specific proteins to be found in the sample. These approaches have gained interest in clinical applications since a high number of altered proteins are observed in different conditions, and they can be evaluated in a quantitative manner, using a wide range of approaches that are mainly divided in isotopic-labeling or label-free methods (Chandramouli and Qian 2009; Silvestrini et al. 2019, 2020). This strategy currently provides a detailed map of thousands of proteins and their respective abundance, allowing comparison of few different variables in each experiment.

Conversely, in targeted proteomics, a known and specific set of proteins are quantitatively analyzed by mass spectrometry. Panels are created with unique peptide sequences that represent the target protein and that will be accurately monitored during the experiment. This approach is based on the high selectivity of peptide ions filtering to improve the sensitivity and accurate quantification of ions (Faça 2017; Silvestrini et al. 2019). Also, this approach allows faster methods and larger number of samples and variables analyzed, which is still a limitation for high-throughput/shotgun proteomics. In summary, the combination of shotgun and targeted strategies provides additional capabilities to identify and validate protein molecular signatures, for example in patient sample cohorts, since the global analysis identifies the altered proteins and subsequent individual sample are accurately quantified for the set of selected proteins by targeted proteomics (Lanfredi et al. 2021). This is a potent strategy for the discovery of new pathways and disease molecular signatures in various perturbed conditions for a wide variety of biological systems.

8.4 Mechanisms That Regulate mRNA and Protein Levels

For many years, the central dogma of molecular biology stated that RNAs molecules were intermediates between DNA and protein and that the function of RNA was primarily linked to the translation of the genetic material into polypeptide chains (proteins) (Brenner et al. 1961; Jacob and Monod 1961). Therefore, the basic level of understanding of the central dogma of biology supports that protein concentrations in a biological system should then directly correlate with their respective mRNAs levels, since translation is required to produce proteins. In fact, it has been shown that when mRNA levels are low, usually proteins are not detected and the

ability to detect proteins increases significantly at higher levels of mRNA (Vogel and Marcotte 2012).

Considering the last decades' technological advances, the measurement of transcribed mRNA has proven to be very powerful in the discovery of molecular markers and the elucidation of functional biological mechanisms. However, it was also evident that mRNA abundance is not a good predictor of protein abundance in the cell. Many possible points of control and potential interruption for the flow of information coded in DNA sequences until it becomes a functional protein have been elucidated. The synthesis and turnover of cellular proteins require several processes that are interconnected, starting with the transcription, processing, and translation of mRNAs, followed by protein folding, cellular transport and localization, and post-translational modification. Parallel processes such as mRNA degradation, inhibition of mRNA translation and protein degradation also modulate the amount of functional protein available in the cell, tissue or organism, which directly impact their physiological conditions (Vogel and Marcotte 2012). These basic mechanisms are illustrated in Fig. 8.1. Given this high degree of interconnection of processes that affect both mRNA and protein levels, understanding normal physiological

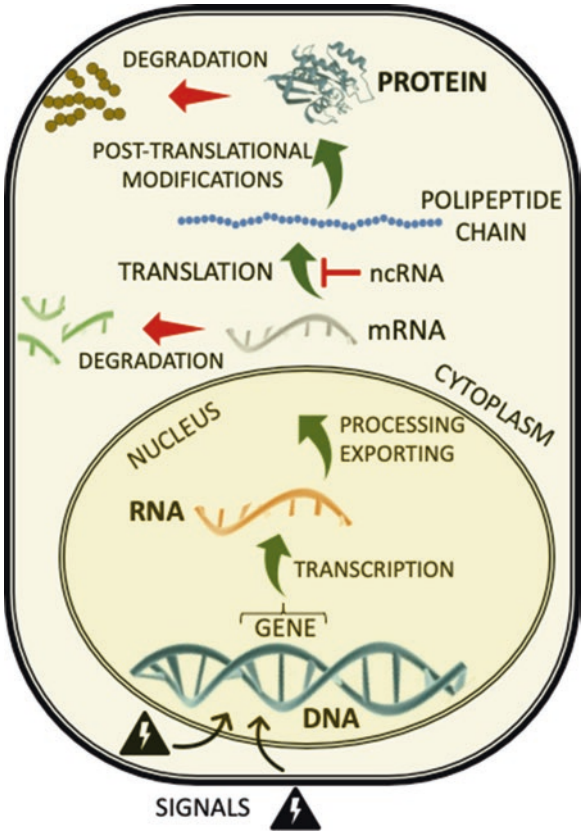

Fig. 8.1 General overview of gene expression. The diagram depicts main processes that are responsible for producing (green arrows) or degrading (red arrows) mRNAs and proteins at a cellular level. Upon intracellular or extracellular signals or stimulus, gene expression is triggered and the fine balance between all these processes is responsible for the maintenance of cellular physiology and defines its normal momentary phenotype

processes that modulate these biomolecules are important to introduce the correlation of mRNA and protein levels in health and disease. Below we discuss some of the most established points in that regard.

8.4.1 mRNA Transcription and Decay

The average abundance of mRNA in a given cell, tissue, or organism is determined by the rates of the transcription versus degradation. Transcriptional regulation occurs at two interconnected levels: the first involves transcription factors and the transcription apparatus, and the second involves chromatin and its regulators. All starts with DNA binding transcription factors that occupy specific sequences and recruit and regulate the transcription by the RNA polymerase II machinery. In eukaryotic systems, there has been extensive study of specific transcription factors and their cofactors, the general transcription apparatus, and various chromatin regulators, leading to current models for specific gene transcription control. The particular set of transcription factors that are expressed in any cell or tissue type at a given moment controls the selective transcription of a subset of genes, correspondent to that cell or tissue expression program. Therefore, the set of genes that are transcribed largely defines the cell phenotype. The gene expression program of a specific cell type includes RNA species from genes that are active in most cells (housekeeping genes) and genes that are active predominantly in one or a limited number of cell types (cell-type-specific genes). Studies of the transcription factors that are key to establishing and maintaining specific cell states suggest that only a small number of the transcription factors that are expressed in cells are necessary to establish cell-type-specific gene expression programs (reviewed by Lee and Young 2013).

On the other end, decay of mRNA can be broadly divided into two classes: mechanisms of quality control that eliminate the production of potentially toxic proteins and mechanisms that lengthen or shorten mRNA half-life for the purpose of changing its abundance, and therefore the availability of functional proteins. Because mRNAs primarily function as templates for protein synthesis, it is logical that cells have evolved translation dependent quality-control mechanisms to dispose of defective mRNAs that synthesize abnormal proteins. Nonsense-mediated mRNA decay (NMD), which, unlike most mRNA decay pathways, appears to be restricted to newly synthesized transcripts, which occurs in all eukaryotes that have been studied, eliminates mRNAs that prematurely terminate translation. This mechanism dampens the potentially toxic effects of defective transcripts that are routinely generated during gene expression of newly synthesized mRNAs. In addition, NMD is inhibited by negative regulators induced by some stresses – such as amino acid starvation and viral infection, among others. Mature mRNAs are degraded by exonucleases acting at both ends of the molecule or endonucleases. Decay rates can be specified by control elements that are usually located within the 3′-untranslated regions (UTRs) of mRNAs and are recognized by various RNA-binding proteins

(RBPs) (Wilusz et al. 2001; Parker and Song 2004). Additionally, degradation of transcripts occurs at distinct cytoplasmic sites in both yeast and human cells indicating that the regulation of mRNA stability is a widespread, tightly regulated, and conserved mechanism for the control of gene expression (Wilusz and Wilusz 2004). Interestingly, little is known about how ribonucleases are regulated, particularly because this class of enzymes is regulated through the proteins they interact with (reviewed by Schoenberg and Maquat 2012).

The time-course measurements of mRNA abundance are, therefore, the key factor to evaluate turnover and stability. Technological advances made the global evaluation of mRNA turnover more common and efficient than it is for proteins. Genome-wide mRNA turnover has been determined in bacteria (Bernstein et al. 2002; Selinger et al. 2003), yeast (Wang et al. 2002; Grigull et al. 2004), plants (Gutiérrez et al. 2002), and humans (Raghavan et al. 2002; Yang et al. 2003) by measuring mRNA levels at different times after RNA polymerase II inactivation. In fact some of these studies brought the concept of timing to describe mRNA stability. Each RNA polymerase II can transcribe about ~100 primary mRNAs per hour from the DNA template. In contrast, ribosomes produce up to 10,000 protein molecules per mRNA per hour (Darzacq et al. 2007; Hausser et al. 2019). We will discuss more about the differences in timing and turnover rates for mRNA and proteins in the next sections.

8.4.2 Regulation of Protein Translation

Cellular functions depend on simultaneous participation of thousands of proteins, which are in a dynamic equilibrium of abundance to maintain homeostasis. As we have been discussing, the cellular processes of protein translation, folding, and degradation together determine the total repertoire of cellular proteins. Protein levels in cells, tissues, and organisms are extremely well regulated in order to reflect the healthy phenotype. Therefore, there should be a very efficient balance between the mechanisms of production and degradation of proteins. In fact, protein translation is the most energy consuming process in the cell, requiring fine modulation before the different stimulus provided by the cellular microenvironment according to the variety of needs of the organism. Starting from the availability of the particular transcriptome of a cell in a given moment, post-transcriptional control takes place during translation, and encompasses both global and transcript-specific mechanisms to regulate protein synthesis (Dever 2002; Gebauer and Hentze 2004). Global regulation, which affects the translation of most transcripts, usually occurs by changes in the phosphorylation state of translation initiation factors and by adjusting the number of available ribosomes (Preiss and Hentze 2003). Transcript-specific regulation, by contrast, modulates the translation of a distinct group of mRNAs and is mediated by a large diversity of mechanisms, such as codon bias or the interaction of the transcript with regulatory elements (Beilharz and Preiss 2004). It involves RNA binding proteins that associate with particular structural features or control

elements present in the UTRs of target transcripts, and are similar to the control of RNA decay which is highlighted earlier.

Among the various processes that coordinate the mRNA translation is the mTOR signaling pathway listed as one of the most studied and understood. The mTOR signaling pathway can rely on various external stimuli to continue the translation regulation. Hormones, growth factors, metabolites, and nutrients can start cell translation machinery (Buttgereit and Brand 1995). mTOR is a Ser/Thr kinase that stimulates anabolic processes through the phosphatidylinositol 3-kinase/protein kinase B (PI3K/AKT) signaling pathway activation by hormone or growth factor via specific receptor tyrosine kinase complexes and its specific substrates. Notably, the entire functional control of these pathways is regulated by post-translational modification, mainly phosphorylation. In addition to regulation by external factors, the mTOR translation modulation pathway is also affected by the cell's internal signaling by conditions such as hypoxia and energy depletion. Another way of modulating the translation extensively studied is the MAPK (mitogen-activated protein kinases); this pathway regulates among others translation parallel to mTOR, interacting with it at several points enabling or inhibiting translational activity. Considering the beginning of MAPK pathway, the upstream event begins with Ras GTPases that can be activated by several external stimuli, also interacting with MAPKs that regulate TSC complex to finally affect mTORC1, or downstream with modulation of translation machinery stimulating its components, such as elf4E (Shaw and Cantley 2006). Despite the detailed understanding of signaling pathways for the components responsible for mRNA translation, such as the regulatory role of the PI3K/mTOR and Ras/MAPK pathways, they are not unique, recent efforts with different analytical techniques show the role of additional signaling pathways in the activation of the translational machinery and even of sensitive or specific transcripts of given pathway (Roux and Topisirovic 2018).

With all these roles assigned, the signaling pathways involved in the translation also prove to be a relevant target for the therapy of diseases, since the imbalance in this adjustment has great potential in the appearance of organism disorders. Comprehensive analysis using a proteogenomics approach of the PI3K/AKT/mTOR pathway showed high activity of these pathways in a significant portion of cancers and despite the great correlation of activity rates, there is in some cases decoupling, showing the regulatory character in the multiple levels of these pathways (Zhang et al. 2017).

8.4.3 Non-coding RNAs Inhibit mRNA Translation

We have been discussing many aspects that affect mRNA translation. However, the development and application of deep sequencing have shown that most of the genome results in transcription to RNAs, but from these only 1–2% of the human genome codes for proteins. Hence, it is possible to divide the transcriptome into two large groups, being coding potential RNAs, that have potential to be translated into

proteins and RNAs without coding potential, not being translated into proteins, non-coding RNAs (ncRNAs). Even though the RNAs were already studied extensively, currently represented mainly by mRNAs, ncRNAs account for the major part of RNAs, holding a great potential for the knowledge of new mechanisms of the processes of expression (Dunham et al. 2012). In fact, the discovery of microRNAs (miRNAs) in 1993 (Lee et al. 1993) followed by developments and discoveries in small RNA and other ncRNA species have redefined the gene regulation landscape. These RNA molecules play a significant role in modulation of an array of physiological and pathological processes that impact directly the balance between mRNA and protein levels (Bhaskaran and Mohan 2014).

Since then, one of the most studied regulatory mechanisms that directly mediate mRNA and protein translation are the miRNAs. These non-coding short RNA molecules inhibit the translation and alter the stability of mRNA by binding to complementary sites on the target mRNAs, usually in the 3′ UTR. With such capabilities, miRNAs are responsible for coordinately controlling genes expression involved in several cellular mechanisms, such as inflammation, cell cycle, apoptosis, migration, and stress, among others pathways involved in disease development (Mollaei et al. 2019). Most importantly, miRNA alterations are evident in several cancer types and correlated with differentiation stages. These molecules can be miRNAs tumor suppressor or oncogenic (oncomiRS). For example, in prostate, pancreatic, bladder cancers, and multiple myeloma the tumor suppressor miR-145 controls targets such as ROCK1, p-AKT, p-PI3K, STAT3, and FOXO1 (Kato et al. 2017; Mollaei et al. 2019). On the other hand, in breast cancer the oncomIR controls PTEN/Akt pathway and contributes to tumorigenesis (Kato et al. 2017; Li et al. 2017).

The main action of ncRNAs widely known is the negative regulation of gene expression by binding a target mRNA through complex formation and induction of its degradation or inhibition of its translation by different mechanisms (Ha and Kim 2014). Regulatory ncRNAs can be divided into microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs) (Ozata et al. 2019), small interfering RNAs (siRNAs), and long non-coding RNAs (lncRNAs) (Yao et al. 2019). The largest quantitative contribution to the group of the non-protein-coding transcripts belongs to the group of lncRNAs, which are arbitrarily considered as about 200 nucleotides in length. Since many of these lncRNAs can also act as primary transcripts for the production of short RNAs, they are involved in the silencing of gene expression (Ponting et al. 2009). In summary, these inhibitory molecules provide possible explanations on how variations can arise between transcriptomics and proteomics profiles in biological systems.

8.4.4 Protein Degradation

On the other side of abundance control, protein half-life can vary significantly depending on a number of different conditions (Glickman and Ciechanover 2002). 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. 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 (Qian et al. 2003).

To keep cellular homeostasis, cells evolved a dynamic and self-regulating quality control processes to maintain protein and to prevent accumulation of damaged molecules. Considering that approximately 240 g protein are synthesized and degraded daily in a 60 kg adult human (Mitch and Goldberg 1996), no wonder a failure on this tight turnover system ultimately leads to disease. In cells, protein degradation is achieved by different degradation systems, of which the ubiquitin–proteasome system (UPS) and autophagy are involved in the degradation of the majority of cellular proteins. Yet another function of proteolytic pathways is selective destruction of proteins whose concentrations must vary with time and alterations in the state of a cell.

The UPS mostly degrades single, unfolded polypeptides able to enter into the narrow channel of the proteasome, and the majority of intracellular proteins are degraded by this process (Zhao et al. 2015). UPS comprises the ubiquitylation system, which involves the activity of specific enzymes that ubiquitylate or deubiquitylate target proteins, and the proteasome system, which degrades ubiquitylated proteins (Collins and Goldberg 2017). Ubiquitylation is a sequential, ATP-consuming process involving a hierarchically acting enzymatic cascade E1, E2, and E3 enzymes, which mediate the covalent attachment of ubiquitin monomers (monoubiquitylation) or chains (polyubiquitylation) to protein substrates. Ubiquitin (Ub) is typically attached via its carboxy-terminus to a lysine residue on a target protein, and it contains seven lysine residues, Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63, which can form up to seven different polyubiquitin chain linkages. The mode of conjugation determines the fate of ubiquitylated proteins, including targeting proteins for degradation, affecting their activity or altering their localization. The proteasome preferentially degrades branched (Lys48-linked) polyubiquitylated proteins, although chains containing nearly all linkages can be recognized and degraded by the proteasome (Meyer and Rape 2014).

The proteasome is the most complex protease in the UPS, has a molecular mass >2.5 MDa, and exists in multiple structural forms but contains two assemblies, a proteolytic chamber formed by the core particle (20S) and a regulatory particle (19S or PA700), which are functionally linked by a gated protein translocation channel, which collectively are known as the 26S complex. Although the roles of many of 26S subunits and associated proteins are still unclear, the 26S proteasome catalyzes the great majority (at least 80%) of the protein degradation in growing mammalian cells. Of note, the proteasome does not degrade proteins to individual amino acids but instead polypeptides are digested to short peptides, which range between 2 and 10 residues in length. The remaining peptides are digested in seconds to amino acids by cytosolic peptidases, but in mammals some serve as precursors for antigenic peptides displayed on MHC-class I molecules (Kisselev et al. 1999; Murata et al. 2018).

As proteolysis is irreversible, intricate multi-level mechanisms have evolved to ensure efficient and selective protein degradation. In this scenario removal of ubiquitin from substrates is tightly controlled by deubiquitylating enzymes (DUBs; also known as deubiquitylases or deubiquitinases). There are four main families of DUBs, and they cleave ubiquitin from proteins and disassemble polyubiquitin chains that are released from substrates before proteasomal degradation, recycling Ub for subsequent ubiquitylation reactions, preventing proteasome congestion and controlling protein turnover by modifying or removing ubiquitin or polyubiquitin chains from the targeted protein (Wilkinson 1997).

Another important, although less specific protein degradation machinery is called autophagy, which is an intracellular pathway for bulk protein degradation and the removal of damaged organelles by lysosomes. It is involved in recycling cellular components like the cytoplasmic proteins; soluble misfolded protein and insoluble misfolded aggregates content for reuse and ensuring that it obeys the rule that "energy can neither be created nor destroyed instead it can be change from one form to another" as the energy required for degradation is high which is also in tandem with notable energy also biosynthesis (Wang et al. 2015). Thus, it helps to change the state of cellular contents to re-useable form to build new cells. There are four pathways identified for the autophagic process which include: the post-translational modification dependent and independent CMA pathways and the ubiquitin dependent and independent macroautophagy pathways (Wang et al. 2015). Autophagy occupies a central position in the maintenance of cellular homeostasis by directing protein degradation, and the process adapts cells to adverse micro-environmental conditions mainly stress such as nutrient/energy starvation, hypoxia, ER stress, hypoxia, and organelle damage (Chen et al. 2019). A precarious balance is essential in protein synthesis as well as turnover so as to prevent the onset of diseases such as neurodegeneration and cancer, which has made autophagy pathway a target in the management of these diseases (Dikic 2017).

Taken together, these major processes and machineries discussed above bring their individual roles into the complex network of events that keep the cellular homeostasis. In the next sections, we will explore how these individual processes contribute to our understanding about when, during cellular events, it is possible to expect a balanced correlation between RNA and protein levels.

8.5 Temporal Correlation of RNA and Protein Levels

In healthy cells and at steady state, RNA and protein synthesis and degradation are well balanced (Harper and Bennett 2016). A given protein abundance can be obtained from infinitely many combinations of these synthesis and degradation rates. In addition, variation in mRNA abundance is frequently buffered on the protein level, meaning that a substantial change in mRNA abundance is not propagated immediately into a corresponding change in protein abundance (Liu et al. 2016). The cell can control the rates of degradation or synthesis for a given protein, and

there is significant heterogeneity even within proteins that have similar functions (Pratt et al. 2002). It is clear that cells control protein production at multiple levels, and the resulting amounts of protein reflect cellular integration of the various regulatory layers, ranging from mRNA production to protein degradation. Although regulation at a single level might prevail in some cases, it is common for cells to coordinately modulate gene expression at several levels.

One important reason for a general lack of correlation between mRNA and protein abundance may be that proteins have very different half-lives as the result of varied protein synthesis and degradation ratios. Given that proteins are on average more stable than mRNAs, proteins can still be present when the mRNA that encoded them is long gone. Therefore, it seems recognizable that the possible reasons why imperfect mRNA–protein correlations arise from the majority of studies for that matter is the factor of "time." Any change in the transcriptional state of a cell will lead to a delay in the response at the protein level simply due to the time it takes to reach a new steady state. Correlations at specific time points during a transition may be uninformative, as changes in mRNA levels in reality correspond to latent changes in protein levels that have yet to occur. Indeed, examples of this are seen in many studies that will be discussed ahead.

As an example, at steady state, the RNA polymerase II machinery can transcribe 2–6 kb/min for a mammalian cell (Maiuri et al. 2011). Considering an average length for an mRNA around 2kb, it takes around a couple of minutes to transcribe one gene. On the protein side, the ribosomal machinery operates at a rate of few to several amino acids per second, generating proteins with average length also in a couple of minutes. Moreover, since many ribosomes can translate the same molecule of mRNA simultaneously, these rates can increase significantly (Riba et al. 2019). Overall, it takes more than an hour to generate 10^6 protein molecules after initiation of transcription from a single locus, for example during cell duplication. A faster means to upregulate proteins is to increase the number of mRNA molecules, amplifying their translation exponentially (Schwanhäusser et al. 2013). However, many factors, such as rates of translation initiation, sequence, folding, and structure of the protein, also significantly affect these rates, again disrupting the correlation of mRNA and protein levels (Riba et al. 2019).

8.6 The Imbalance Between the Transcriptome and Proteomes: Lessons Learned from High-Throughput Studies

Considering what has been discussed so far, it is certain that protein and RNA-based measurements are complementary to provide accurate status of cellular homeostasis. It is important to recognize that many factors can cause imbalances between levels of messenger (transcript) and its final effector (mature protein). As we have seen, several post-transcriptional and post-translational control mechanisms such as

the translation rate or half-lives of mRNAs and proteins are affected by a wide range of factors.

On top of all the biological dynamics, the methods of RNA sequencing and also of protein expression evaluation by mass spectrometry suffer from technical limitations that affect the precision and final accuracy of the quantitative measurements. For example, biases in RNA data can arise during formation of the sequencing library. Also, in mass spectrometry, the shotgun approach relies on proteins digested with enzymes such as trypsin to generate peptides that are the entities identified. The tremendous variety of chemical species generated in shotgun approaches, which are influenced by several physicochemical factors, and the stochastic property of this technique, turn the proteome samples extremely challenging and almost impossible to be completely characterized. Certain approaches such as the use of isotopic labeling for relative quantitation or the run of multiple technical and biological replicates in mass spectrometry turn such proteome complexity into a feasible strategy that can be effectively profiled (Buccitelli and Selbach 2020).

Even with the significant developments in the technologies used to quantify protein abundance over the past couple of years, protein identification and quantification still lag behind the high-throughput experimental techniques used to determine mRNA expression levels. The proteome of a cell or tissue at a specific time point is extremely complex and diverse. The major limits of proteome analysis are associated with the heterogeneity of proteins and the huge differences in abundance (dynamic range). Abundant proteins mask the presence of low abundant proteins. Because no PCR equivalent exists for proteins, low-abundant proteins have a low probability to be detected (Churchill 2002; Larkin et al. 2005). Since it is fundamental to consider at a least some of these drawbacks reported above for a satisfactory comparison between the transcriptome and proteome, several studies have been specifically designed with this particular focus.

Overall, genome-wide studies have shown that the correlation between expression levels of mRNA and protein are marginal, hovering around 40–50% across many studies. One of the seminal studies specifically developed with the purpose to compare the expression profile of active genes in the adult human liver and the protein abundance in human plasma (Kawamoto et al. 1996). The study found a positive correlation between the abundance of the transcript and the protein concentration in the serum. It was also possible to categorize the responsible genes into three groups: those with less than five transcripts (per 1000 mRNA molecules) produce proteins at a level of <0.1 g/100ml, those with 5–20 transcripts produce proteins at 0.1–0.4 g/100 ml, and those with more than 30 transcripts produce proteins at 0.5–4 g/100 ml. This was a pioneering study on a large scale showing that particularly for secreted proteins, the transcript – protein correlation was positive.

Another important study explored for the first time a quantitative comparison of mRNA transcript and protein expression levels for a relatively large number of genes expressed in the same metabolic state in yeast (Gygi et al. 1999). The study concluded that predictions of protein levels from mRNA transcript levels were not feasible. This study particularly relied on 2D-electrophoresis to evaluate the proteome, which itself is a very limited technique in terms of dynamic range for protein

abundance. However, the study found that for a subset of 106 highly abundant proteins, the correlation with mRNA levels was positive. Subsequently, using the same yeast as model system, it was again demonstrated a partial correlation of protein expression after specific perturbations in know pathways, namely the galactose utilization (Ideker et al. 2001). While several genes–proteins ratios correlated well in increased or decreased expression upon perturbation, others know players in the pathway still had poor correlation. More specifically this study attributed the discrepancy in gene–protein expression correlations to post-transcriptional regulatory events. Yet, this study also uncovers that for genes linked by physical interactions in the network tend to have more strongly correlated expression profiles than genes chosen at random. Using modern high-throughput proteomics and accurate relative quantitation based on stable isotopes, another study explored in depth the yeast proteome and transcriptome correlation (de Godoy et al. 2008). This study once again demonstrated the poor overall correlation of transcriptome and proteome, but particularly found that good correlation was found for the subset of genes involved in yeast pheromone pathway components.

Studies based on more complex organisms also provided contrasting mRNA and protein levels. In a detailed comparison of mesenchymal stromal cells obtained from bone marrow or umbilical cord vein, with the overall objective to prove the interchangeability of these sources for cellular therapy, proteomic and gene expression analysis reached a 63% correlation level for those specific set of genes specific for one or the other mesenchymal cell type (Miranda et al. 2012). This dataset is particularly illustrated in the correlation plot (Fig. 8.2a), which indicated that mRNA abundance data (y-axis) presented more spreading in terms of ratios in comparison to proteomic data (x-axis). Using more sophisticated proteomic strategies, the dynamics of protein and mRNA expression levels across the cell cycle in human myeloid leukemia cells using was explored (Ly et al. 2014). Myeloid-specific gene expression and variations in protein abundance, isoform expression, and phosphorylation at different cell cycle stages were dissected for over ~6000 genes individually across the cell cycle, revealing complex, gene-specific patterns. Protein and mRNA correlations were modest across different cell cycle stages, suggesting again greater contribution of post-transcriptional mechanisms in cell cycle control.

Considering that most of the aforementioned studies focused on static of minimally dynamic biological events, the temporal contribution to the lack of correlation between proteome and transcriptome of a cell was still obscure. A breakthrough study shed light in the time variable studying the dynamics of embryonic development (Peshkin et al. 2015). Based on a time-resolved deep quantitative profiling of proteins and mRNA, the study produced an unprecedented dataset that illustrated the turnover of these molecules during the embryo development. As example, Fig. 8.2b demonstrates the normalized curves of mRNA and protein levels over time. It is clear that the initial wave of mRNA expression and accumulation, was followed by a quick decay, while protein levels progressively accumulated for both CAPN8 and LIN28A genes. Obviously, depending on the moment one makes the mRNA and protein measurements for such gene, more or less correlation will be found. On the other hand, the other illustrated genes, DND1 and SPARC, have a

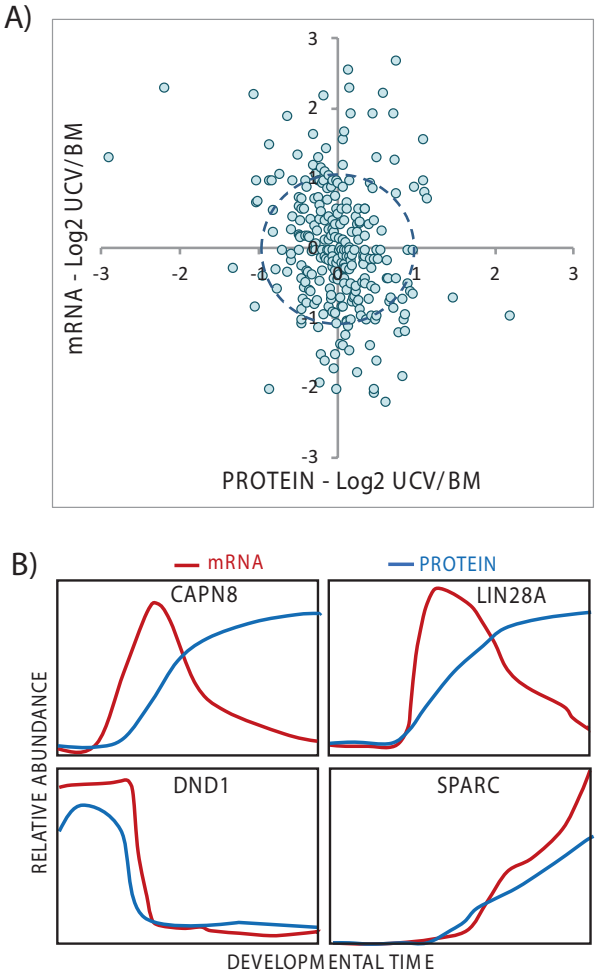

Fig. 8.2 Correlation of gene expression in complex systems. (**a**) The combined analysis of the transcriptome and proteome of mesenchymal stromal cells from bone-marrow (BM and umbilical cord vein (UVC) demonstrated a correlation of 63% of the profiled genes (central circle) and similarity between these two sources of therapeutic cells (reproduced from Miranda et al. 2012). (**b**) Time-course experiments during embryonic development demonstrated the syntheses and decay of mRNA and proteins. While some genes (CAPN8 and LIN28A) have an evident difference in timing for synthesis and degradation, others (DND1 and SPARC) present a completely synchronized and correlated gene expression. (Reproduced from Peshkin et al. 2015)

very tight correlation. Of note, DND1 itself is a RNA-binding factor that positively regulates gene expression by prohibiting miRNA-mediated gene suppression, creating a scape for post-transcriptional regulation. SPARC is a secreted protein, class that has been observed with greater gene expression correlations.

Several other studies support these major findings described above. But particularly for diseases, the context of protein versus mRNA expression becomes particularly important for diagnostics and molecular profiling. As examples, a study correlated the expression of microRNA, mRNA, and proteins in the identification of microRNA-related cancers, particularly in glioblastoma (Seo et al. 2017). For a subset of 146 upregulated genes, mRNA and proteins were positively correlated. These findings are consistent with the hypothesis that the malignant phenotype required additional cancer promoter genes that were coordinately overexpressed. In a similar study, Yang and colleagues used a combination of proteomics and transcriptomics strategies and found potential targets in early colorectal cancer (CRC) (Yang et al. 2019). The study identified 2968 proteins in stage II CRC proteomics data, where most (2846) of these proteins were identified in TGCA transcriptome data. Numerous bioinformatics methods, including differential expression analysis, weighted correlation network analysis, gene ontology, and protein–protein interaction analyses, were used to select a set of 111 key proteins, differentially expressed in terms of proteins and mRNAs, levels. These highly correlated genes can represent a molecular signature for the CRC, and used, for example, to subclassify the tumor types. In summary, the integration of proteomics and transcriptomics data, particular for disease studies, can generate a high-resolution global expression map that can collaborate to discover new biomarkers for several diseases.

8.7 Final Remarks and Perspectives

As we discussed in this chapter, profiling gene expression enables a global physiological picture for a given system in a specific context or moment. When the dynamics of cellular processes is taken into account, several regulatory processes emerge and explain apparent disconnection of the transcriptome and proteome. Unlike the genome, which is virtually static in terms of its composition and size, we gave several examples here that support the dynamics of the genetic cellular programming, which continually changes depending on the phase of the cell cycle, the organ, exposure to drugs or physical agents, aging, diseases such as cancer and autoimmune diseases, and a multitude of other variables.

Several of the factors that modulate abundance of mRNA and proteins have been presented. New features of these molecular mechanisms have been continuously uncovered, mainly promoted by advances in high-throughput deep biomolecular profiling. In addition to the development of modern multidimensional transcriptomics and proteomics strategies, bioinformatics and data integration have become a common basis for translational areas, where complex integrated mRNA and protein molecular signatures are aiding the development of new therapeutic strategies or methods for diseases diagnostics. Ultimately, full and effective integration across the relatively static genomic information with the dynamic transcriptomic and proteomic data will produce complete maps of normal and pathological process to drive personalized medicine.

With the continuous advancing of technology and biology, the interplay of transcriptomics and proteomics profiles in living organisms will become more evident and fundamental to provide answers to many relevant biological questions in health and disease.

Support The authors of this work were supported by grants from the following Brazilian funding agencies: FAPESP; J.O.T -(grant no. 2020/13463-2), CNPq, and CAPES.

References

Anderson L, Seilhamer J (1997) A comparison of selected mRNA and protein abundances in human liver. Electrophoresis 18:533–537

Aslam B, Basit M, Nisar MA, Khurshid M, Rasool MH (2017) Proteomics: technologies and their applications. J Chromatogr Sci 55:182–196

Bainbridge MN, Warren RL, Hirst M, Romanuik T, Zeng T, Go A, Delaney A, Griffith M, Hickenbotham M, Magrini V, Mardis ER, Sadar MD, Siddiqui AS, Marra MA, Jones SJ (2006) Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC Genomics 7:246

Barbulovic-Nad I, Lucente M, Sun Y, Zhang M, Wheeler AR, Bussmann M (2006) Bio-microarray fabrication techniques—a review. Crit Rev Biotechnol 26:237–259

Beilharz TH, Preiss T (2004) Translational profiling: the genome-wide measure of the nascent proteome. Brief Funct Genomic Proteomic 3:103–111

Bernstein JA, Khodursky AB, Lin P-H, Lin-Chao S, Cohen SN (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. PNAS 99:9697–9702

Bhaskaran M, Mohan M (2014) MicroRNAs: history, biogenesis, and their evolving role in animal development and disease. Vet Pathol 51:759–774

Brenner S, Jacob F, Meselson M (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature 190:576–581

Buccitelli C, Selbach M (2020) mRNAs, proteins and the emerging principles of gene expression control. Nat Rev Genet 21:630–644

Buttgereit F, Brand MD (1995) A hierarchy of ATP-consuming processes in mammalian cells. Biochem J 312:163–167

Chandramouli K, Qian P-Y (2009) Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Hum Genomics Proteomics 2009: 239204

Chen R-H, Chen Y-H, Huang T-Y (2019) Ubiquitin-mediated regulation of autophagy. J Biomed Sci 26:80

Churchill GA (2002) Fundamentals of experimental design for cDNA microarrays. Nat Genet 32:490–495

Cobb JP, Mindrinos MN, Miller-Graziano C, Calvano SE, Baker HV, Xiao W, Laudanski K, Brownstein BH, Elson CM, Hayden DL, Herndon DN, Lowry SF, Maier RV, Schoenfeld DA, Moldawer LL, Davis RW, Tompkins RG, Program§§ I and HR to IL-SCR (2005) Application of genome-wide expression analysis to human health and disease. PNAS 102:4801–4806

Collins GA, Goldberg AL (2017) The logic of the 26S proteasome. Cell 169:792–806

Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH (2007) In vivo dynamics of RNA polymerase II transcription. Nat Struct Mol Biol 14:796–806

de Godoy LMF, Olsen JV, Cox J, Nielsen ML, Hubner NC, Fröhlich F, Walther TC, Mann M (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 455:1251–1254

Dever TE (2002) Gene-specific regulation by general translation factors. Cell 108:545–556
Dikic I (2017) Proteasomal and autophagic degradation systems. Annu Rev Biochem 86:193–224
Dunham I, Kundaje A, Aldred SF, et al, The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
Faça VM (2017) Selective reaction monitoring for quantitation of cellular proteins. Methods Mol Biol 1546:213–221
Gebauer F, Hentze MW (2004) Molecular mechanisms of translational control. Nat Rev Mol Cell Biol 5:827–835
Glickman MH, Ciechanover A (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiol Rev 82:373–428
Grigull J, Mnaimneh S, Pootoolal J, Robinson MD, Hughes TR (2004) Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. Mol Cell Biol 24:5534–5547
Gutiérrez R, Ewing R, Cherry J, Green P (2002) Identification of unstable transcripts in Arabidopsis by cDNA microarray analysis: rapid decay is associated with a group of touch- and specific clock-controlled genes. Proc Natl Acad Sci USA 99: 11513-11518
Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19:1720–1730
Ha M, Kim VN (2014) Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol 15:509–524
Harper JW, Bennett EJ (2016) Proteome complexity and the forces that drive proteome imbalance. Nature 537:328–338
Hausser J, Mayo A, Keren L, Alon U (2019) Central dogma rates and the trade-off between precision and economy in gene expression. Nat Commun 10:68
Heber S, Sick B (2006) Quality assessment of affymetrix GeneChip data. OMICS J Integr Biol 10:358–368
Hegde PS, White IR, Debouck C (2003) Interplay of transcriptomics and proteomics. Curr Opin Biotechnol 14:647–651
Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292:929–934
Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JGN, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martínez-Murillo F, Morsberger L, Lee H, Petersen D, Quackenbush J, Scott A, Wilson M, Yang Y, Ye SQ, Yu W (2005) Multiple-laboratory comparison of microarray platforms. Nat Methods 2:345–350
Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3:318–356
Jiang Z, Zhou X, Li R, Michal JJ, Zhang S, Dodson MV, Zhang Z, Harland RM (2015) Whole transcriptome analysis with sequencing: methods, challenges and potential solutions. Cell Mol Life Sci 72:3425–3439
Kato M, Kurozumi A, Goto Y, Nohata N, Arai T, Okato A, Koshizuka K, Kojima S, Ichikawa T, Seki N (2017) Abstract 1459: dual-strand tumor-suppressor microRNA-145 (miR-145-5p and miR-145-3p) are involved in castration-resistant prostate cancer pathogenesis. Cancer Res 77:1459–1459
Kawamoto S, Matsumoto Y, Mizuno K, Okubo K, Matsubara K (1996) Expression profiles of active genes in human and mouse livers. Gene 174:151–158
Kisselev AF, Akopian TN, Woo KM, Goldberg AL (1999) The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes: implications for understanding the degradative mechanism and antigen presentation*. J Biol Chem 274:3363–3371
Klose J (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. Hum Genet 26:231–243
Kukurba KR, Montgomery SB (2015) RNA sequencing and analysis. Cold Spring Harb Protoc 2015:pdb.top084970

Lanfredi GP, Thomé CH, Ferreira GA, Silvestrini VC, Masson AP, Vargas AP, Grassi ML, Poersch A, Candido dos Reis FJ, Faça VM (2021) Analysis of ovarian cancer cell secretome during epithelial to mesenchymal transition reveals a protein signature associated with advanced stages of ovarian tumors. Biochimica et Biophysica Acta (BBA) – Proteins Proteomics 1869:140623

Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, Gonzàlez-Porta M, Kurbatova N, Griebel T, Ferreira PG, Barann M, Wieland T, Greger L, van Iterson M, Almlöf J, Ribeca P, Pulyakhina I, Esser D, Giger T, Tikhonov A, Sultan M, Bertier G, DG MA, Lek M, Lizano E, HPJ B, Padioleau I, Schwarzmayr T, Karlberg O, Ongen H, Kilpinen H, Beltran S, Gut M, Kahlem K, Amstislavskiy V, Stegle O, Pirinen M, Montgomery SB, Donnelly P, MI MC, Flicek P, Strom TM, Lehrach H, Schreiber S, Sudbrak R, Carracedo Á, Antonarakis SE, Häsler R, Syvänen A-C, van Ommen G-J, Brazma A, Meitinger T, Rosenstiel P, Guigó R, Gut IG, Estivill X, Dermitzakis ET (2013) Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501:506–511

Larance M, Lamond AI (2015) Multidimensional proteomics for cell biology. Nat Rev Mol Cell Biol 16:269–280

Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J (2005) Independence and reproducibility across microarray platforms. Nat Methods 2:337–344

Lee TI, Young RA (2013) Transcriptional regulation and its misregulation in disease. Cell 152:1237–1251

Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854

Li N, Miao Y, Shan Y, Liu B, Li Y, Zhao L, Jia L (2017) MiR-106b and miR-93 regulate cell progression by suppression of PTEN via PI3K/Akt pathway in breast cancer. Cell Death Dis 8:e2796–e2796

Liu Y, Beyer A, Aebersold R (2016) On the dependency of cellular protein levels on mRNA abundance. Cell 165:535–550

Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T (2017) Transcriptomics technologies. PLoS Comput Biol 13:e1005457

Ly T, Ahmad Y, Shlien A, Soroka D, Mills A, Emanuele MJ, Stratton MR, Lamond AI (2014) A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. elife 3:e0163

Maiuri P, Knezevich A, De Marco A, Mazza D, Kula A, McNally JG, Marcello A (2011) Fast transcription rates of RNA polymerase II in human cells. EMBO Rep 12:1280–1285

Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, Ferrari R (2018) Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform 19:286–302

Meyer H-J, Rape M (2014) Enhanced protein degradation by branched ubiquitin chains. Cell 157:910–921

Miranda HC, Herai RH, Thomé CH, Gomes GG, Panepucci RA, Orellana MD, Covas DT, Muotri AR, Greene LJ, Faça VM (2012) A quantitative proteomic and transcriptomic comparison of human mesenchymal stem cells from bone marrow and umbilical cord vein. Proteomics 12:2607–2617

Mitch WE, Goldberg AL (1996) Mechanisms of muscle wasting. The role of the ubiquitin-proteasome pathway. N Engl J Med 335:1897–1905

Mollaei H, Safaralizadeh R, Rostami Z (2019) MicroRNA replacement therapy in cancer. J Cell Physiol 234:12369–12384

Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet 10:135–151

Murata S, Takahama Y, Kasahara M, Tanaka K (2018) The immunoproteasome and thymoproteasome: functions, evolution and human disease. Nat Immunol 19:923–931

Nagalakshmi U, Waern K, Snyder M (2010) RNA-Seq: a method for comprehensive transcriptome analysis. Curr Protoc Mol Biol Chapter 4:Unit 4.11.1–13.

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007–4021
Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD (2019) PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet 20:89–108
Parker R, Song H (2004) The enzymes and control of eukaryotic mRNA turnover. Nat Struct Mol Biol 11:121–127
Peshkin L, Wühr M, Pearl E, Haas W, Freeman RM, Gerhart JC, Klein AM, Horb M, Gygi SP, Kirschner MW (2015) On the relationship of protein and mRNA dynamics in vertebrate embryonic development. Dev Cell 35:383–394
Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. Cell 136:629–641
Pratt JM, Petty J, Riba-Garcia I, Robertson DHL, Gaskell SJ, Oliver SG, Beynon RJ (2002) Dynamics of protein turnover, a missing dimension in proteomics. Mol Cell Proteomics 1:579–591
Preiss T, Hentze MW (2003) Starting the protein synthesis machine: eukaryotic translation initiation. BioEssays 25:1201–1211
Qian J, Kluger Y, Yu H, Gerstein M (2003) Identification and correction of spurious spatial correlations in microarray data. BioTechniques 35:42–48
Raghavan A, Ogilvie RL, Reilly C, Abelson ML, Raghavan S, Vasdewani J, Krathwohl M, Bohjanen PR (2002) Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. Nucleic Acids Res 30:5529–5538
Riba A, Nanni ND, Mittal N, Arhné E, Schmidt A, Zavolan M (2019) Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. PNAS 116:15023–15032
Roux PP, Topisirovic I (2018) Signaling pathways involved in the regulation of mRNA translation. Mol Cell Biol 38: e00070-18
Schoenberg DR, Maquat LE (2012) Regulation of cytoplasmic mRNA decay. Nat Rev Genet 13:246–259
Schwanhäusser B, Wolf J, Selbach M, Busse D (2013) Synthesis and degradation jointly determine the responsiveness of the cellular proteome. BioEssays 35:597–601
Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C (2003) Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. Genome Res 13:216–223
Seo J, Jin D, Choi C-H, Lee H (2017) Integration of MicroRNA, mRNA, and protein expression data for the identification of cancer-related MicroRNAs. PLoS One 12:e0168412
Shaw RJ, Cantley LC (2006) Ras, PI(3)K and mTOR signalling controls tumour cell growth. Nature 441:424–430
Silvestrini VC, Lanfredi GP, Masson AP, Poersch A, Ferreira GA, Thomé CH, Faça VM (2019) A proteomics outlook towards the elucidation of epithelial–mesenchymal transition molecular events. Mol Omics 15:316–330
Silvestrini VC, Thomé CH, Albuquerque D, de Souza PC, Ferreira GA, Lanfredi GP, Masson AP, Delsin LEA, Ferreira FU, de Souza FC, de Godoy LMF, Aquino A, Carrilho E, Panepucci RA, Covas DT, Faça VM (2020) Proteomics analysis reveals the role of ubiquitin specific protease (USP47) in Epithelial to Mesenchymal Transition (EMT) induced by TGFβ2 in breast cells. J Proteome 219:103734
Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13:227–232
Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO (2002) Precision and functional specificity in mRNA decay. PNAS 99:5860–5865
Wang D, Peng Z, Ren G, Wang G (2015) The different roles of selective autophagic protein degradation in mammalian cells. Oncotarget 6:37098–37116
Wilkinson KD (1997) Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. FASEB J 11:1245–1256

Wilusz CJ, Wilusz J (2004) Bringing the role of mRNA decay in the control of gene expression into focus. Trends Genet 20:491–497

Wilusz CJ, Wormington M, Peltz SW (2001) The cap-to-tail guide to mRNA turnover. Nat Rev Mol Cell Biol 2:237–246

Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M, Darnell JE (2003) Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res 13:1863–1872

Yang W, Shi J, Zhou Y, Liu T, Zhan F, Zhang K, Liu N (2019) Integrating proteomics and transcriptomics for the identification of potential targets in early colorectal cancer. Int J Oncol 55:439–450

Yao R-W, Wang Y, Chen L-L (2019) Cellular functions of long noncoding RNAs. Nat Cell Biol 21:542–551

Zhang Y, Kwok-Shing Ng P, Kucherlapati M, Chen F, Liu Y, Tsang YH, de Velasco G, Jeong KJ, Akbani R, Hadjipanayis A, Pantazi A, Bristow CA, Lee E, Mahadeshwar HS, Tang J, Zhang J, Yang L, Seth S, Lee S, Ren X, Song X, Sun H, Seidman J, Luquette LJ, Xi R, Chin L, Protopopov A, Westbrook TF, Shelley CS, Choueiri TK, Ittmann M, Van Waes C, Weinstein JN, Liang H, Henske EP, Godwin AK, Park PJ, Kucherlapati R, Scott KL, Mills GB, Kwiatkowski DJ, Creighton CJ (2017) A pan-cancer proteogenomic atlas of PI3K/AKT/mTOR pathway alterations. Cancer Cell 31:820–832.e3

Zhao J, Zhai B, Gygi SP, Goldberg AL (2015) mTOR inhibition activates overall protein degradation by the ubiquitin proteasome system as well as by autophagy. PNAS 112:15790–15797