



Workability of mRNA Sequencing for Predicting Protein Abundance

Elena A. Ponomarenko¹, George S. Krasnov², Olga I. Kiseleva¹, Polina A. Kryukova¹, Viktoriia A. Arzumanian¹, Georgii V. Dolgalev¹, Ekaterina V. Ilgisonis¹, Andrey V. Lisitsa¹ and Ekaterina V. Poverennaya^{1,*}

- ¹ Institute of Biomedical Chemistry, Moscow 119121, Russia
- ² Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991, Russia; gskrasnov@mail.ru
- * Correspondence: k.poverennaya@ibmc.msk.ru

Abstract: Transcriptomics methods (RNA-Seq, PCR) today are more routine and reproducible than proteomics methods, i.e., both mass spectrometry and immunochemical analysis. For this reason, most scientific studies are limited to assessing the level of mRNA content. At the same time, protein content (and its post-translational status) largely determines the cell's state and behavior. Such a forced extrapolation of conclusions from the transcriptome to the proteome often seems unjustified. The ratios of "transcript-protein" pairs can vary by several orders of magnitude for different genes. As a rule, the correlation coefficient between transcriptome–proteome levels for different tissues does not exceed 0.3–0.5. Several characteristics determine the ratio between the content of mRNA and protein: among them, the rate of movement of the ribosome along the mRNA and the number of free ribosomes in the cell, the availability of tRNA, the secondary structure, and the localization of the transcript. The technical features of the experimental methods also significantly influence the levels of the transcript and protein of the corresponding gene on the outcome of the comparison. Given the above biological features and the performance of experimental and bioinformatic approaches, one may develop various models to predict proteomic profiles based on transcriptomic data. This review is devoted to the ability of RNA sequencing methods for protein abundance prediction.

Keywords: NGS; transcriptome; proteome; protein abundance; gene expression; mRNA-to-protein ratio; RNA-Seq; mass spectrometry

1. The Attractiveness of Transcriptomic Methods for Assessing Gene Expression

Similar to the social sciences, the number of models, theories, and knowledge about mechanisms in molecular biology is significantly inferior to the total amount of experimental data obtained. These features distinguish molecular biology from mathematics or astronomy [1]. High-throughput methods of molecular profiling domesticated over the past two decades make it possible to assess gene expression at the transcriptomic and proteomic levels and screen for mutations and chromosomal rearrangements in a genome-wide mode. The profiling of epigenetic changes, such as DNA methylation and histone modifications, has also become routine.

Such data can be used for systematic study and subsequent multiomics modeling [2] of quantitative relationships between different "omics layers": transcriptomic and epigenomic, and transcriptomic and proteomic [3]. For example, by studying the mRNA level, it is possible to reveal the mechanisms and patterns responsible for gene expression regulation at the epigenetic level by profiling histone marks and methylation of the corresponding DNA regions [4].

Data on gene expression at the mRNA and protein levels are used to identify the mechanisms and characteristics of the gene that determine the degree of dependence of the protein content on the abundance of the corresponding transcript [5].



Citation: Ponomarenko, E.A.; Krasnov, G.S.; Kiseleva, O.I.; Kryukova, P.A.; Arzumanian, V.A.; Dolgalev, G.V.; Ilgisonis, E.V.; Lisitsa, A.V.; Poverennaya, E.V. Workability of mRNA Sequencing for Predicting Protein Abundance. *Genes* **2023**, *14*, 2065. https://doi.org/10.3390/ genes14112065

Academic Editors: Jinghua Zhao and Jiadong Ji

Received: 7 October 2023 Revised: 3 November 2023 Accepted: 7 November 2023 Published: 11 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

Developing a high-precision bioinformatics tool, which utilizes transcript content data for protein profile prediction, seems attractive for several reasons. Firstly, the protein, not the transcript, is the final "effector" of the entire process of gene expression, starting from the transcription initiation. Differences at the proteomic rather than the transcriptomic level are primary when evaluating the influence of specific factors (e.g., drug therapy or development of pathology) on cellular processes. Secondly, the ratios between gene expression products at the transcript and protein level can drastically (by several orders of magnitude) differ between distinct genes. Therefore, transferring findings drawn from transcriptomic data to the proteome level can often lead to incorrect conclusions [6]. The proteomic picture is additionally complicated by the presence of post-translational modifications (PTMs), which may completely change the activity and functions of the protein. Thirdly, transcriptomic profiling methods (RNA-Seq, microarrays, PCR) have significantly higher performance, convenience, and sensitivity compared to proteomic studies. The indicated advantages relate to panoramic (LC-MS/MALDI-TOF) and targeted (MRM and SRM) mass spectrometric approaches, as well as immunochemical methods [7]. On average, transcriptomic analysis (particularly RNA-seq) provides information on a significantly larger number of genes than proteomic analysis [8,9]. Thus, it is possible to quantify the expression level of ca. 10–18 thousand human mRNAs [10–12]. For comparison, routine panoramic mass spectrometric tissue analysis provides information about 3–5 thousand proteins [13].

Another advantage of RNA-Seq is the ability to provide comprehensive information about the complete transcript sequence, including point mutations, insertions, and deletions, and simultaneously identify cases of alternative splicing [14], including de novo analysis. The situation is entirely different with proteomic methods. In mass spectrometric analysis, protein detection and quantification are carried out mainly by analyzing small proteotypic peptides. In panoramic mode, non-specific amino acid substitutions or other protein sequence changes are usually not considered. The "Brute force" approach, which involves the enumeration of all possible aberrations, often expands the search space, increasing the fraction of false positives [15–17]. The same can be said about immunochemical methods: introducing a few substitutions in the protein sequence can significantly affect the affinity-binding constants of a protein and the corresponding antibody [18].

A deep and quantitative assessment of gene expression at the level of transcripts is much easier to perform than at the level of proteins [19]. The main reasons for the insufficient effectiveness of proteomic approaches are:

- the absence of an analogue of direct amplification of protein sequences, in contrast to genes and transcripts [20];
- the high diversity of proteinogenic amino acids (23 proteinogenic blocks and modified amino acids [21]) compared to the four-letter nucleotide composition of RNAs (even considering possible PTMs);
- the lack of high-specific enzyme selection to cleave protein sequences (compared to a rich set of polymerases, DNases, RNases, and ligases).

In mass spectrometric proteomic analysis, non-standard technical solutions can increase the number of identifications. Nevertheless, such solutions make the analysis time-consuming and inapplicable for the routine assessment of protein profiles. One example is proteolysis by multiple proteases [22] or multidimensional fractionation (2DE, 2D-LC). It was shown that two-dimensional alkaline fractionation doubled the coverage of protein sequences (from 23% to 54%) in a shotgun MS experiment [23].

The high value of knowledge about proteins coupled with imperfect proteomic technologies makes it essential to develop methods for predicting proteomic profiles based on transcriptomic data for a cell or tissue, and to study the patterns connecting the transcriptome and proteome.

2. Key Points of Transcriptome-to-Proteome Research

The central dogma of molecular biology links information flows between DNA, RNA, and protein [24]. Although the DNA sequence generally determines the sequence of transcribed mRNA, which defines the arrangement of amino acids in a protein, there is no trivial relationship between the abundance of the transcript and the corresponding protein. Only indirect relationships between the abundance of transcripts and proteins were revealed for model objects. However, in principle, there is a possibility to predict the protein content based on genomic and transcriptome data [25]. It has been shown that the amount of protein is not a linear function of the amount of the corresponding mRNA [26]. The level of mRNA by itself was also demonstrated to be insufficient to predict protein representation and explain the relationship between genotype and phenotype [27].

Initially, based on the data on individual genes, it was believed [28] that the levels of mRNA and protein strongly correlate, but the development of large-scale and high-throughput transcriptomic and proteomic methods refuted this hypothesis (see Figure 1). Such global analyses have shown that transcripts with similar abundance levels can have corresponding proteins with widely varying concentrations. Some pioneering works [29], dedicated to the investigation of the relationship between the transcriptome and the proteome of *S. cerevisiae*, a popular object for developing predictive models, revealed that the protein content could vary as much as 20-fold at the same levels of corresponding RNAs.



Figure 1. Timeline of mRNA-to-protein correlation coefficient based on transcriptome and translatome data.

Further investigations have shown that the coefficient of determination $R^2 = 0.58$ between mRNA and protein levels can appear after log transformation of the abundance values, which often yields a Gaussian distribution of abundance data [30].

Another paper shows a considerably higher correlation between the number of mRNA copies and the absolute number of protein copies at $R^2 = 0.73$ [31] (Figure 1). However, gene expression values and protein concentrations were obtained by at least two technologies and then averaged [31]. Thus, averaging across technologies removes technology-specific errors [31].

Other studies using direct quantitative estimates demonstrate significantly more modest correlation coefficients between the levels of transcripts and proteins. This conclusion was made in the study [32] of the transcriptome (RNA-Seq) and proteome (LC-MS) of liver tissues of 100 mice descending from representatives of inbred lines. The abundances of 5000 peptides and 22,000 transcripts were evaluated, from which around 500 proteins and 7000 mRNAs with the most accurate measurement results were selected for correlation analysis [32]. The study indicated a pronounced positive correlation between the levels of proteins and transcripts (with an averaged Pearson correlation coefficient of 0.27) (Figure 1).

In other works on plants (sowing rice and maize [33,34]), very weak correlations between transcriptome and proteome were also observed, both indicating a Pearson coefficient of linear relationship lower than 0.4 (Figure 1).

Various approaches for deriving transcriptomic and proteomic data in these studies represent the existing prosperity of methods. For example, the first study explored the quantitative relationship between the abundances of mRNA, derived from RNA-seq (levels reported in RPKM), and protein, gained via MS label-free relative quantification (intensity-based absolute quantification, or iBAQ) [34]. The pairs of values presented transcriptomic and proteomic profiles en masse to illustrate tissue-specific patterns of gene expression. The second study aimed the investigation of changes in gene expression during the development and specification of leaf vascular systems, and took, as units, differentially expressed proteins (DEPs), derived from MS isobaric group labeling for relative and absolute quantification (iTRAQ), and genes (DEGs) [33]. Transcript abundances were acquired through newly introduced high-throughput tag-sequencing for digital gene expression (DGE) analysis.

Interestingly, the authors of [34] noted that the expression of critical genes responsible for the morphology and function of maize leaf tissues was regulated precisely at the transcription level, depending on time and localization in the plant. In other cases, a low correlation indicates divergence in regulation on transcriptional and translational levels.

Meta-analysis and the creation of combined transcriptomic and proteomic datasets with subsequent correlation analysis were used in the study conducted on yeast [28]; it was shown that the correlation coefficient calculated for the combined data set was higher than the corresponding values calculated for individual data sets. In another work, non-linear optimization was used to estimate undetectable *D. vulgaris* proteins based on mRNA level data [35]. The method is based on the maximization of the objective function that describes the relationship between the transcriptomic and proteomic networks without considering network changes over time.

3. Gene-Centric Approach: The mRNA-Protein Ratio Varies Greatly between Different Genes but Is Conserved in Different Tissues and Cell Types

Several scientific groups have found that the ratio of transcripts to proteins for corresponding genes is relatively stable in various cell lines [36] and tissues [37] (Figure 1). This observation gave rise to the assumption that the protein content in any tissue can be predicted by the amount of mRNA [37,38]. Various factors can be responsible for the ratio of protein and mRNA levels; among them are:

- (a) constant gene-specific and context-independent (i.e., codon selection, secondary structures of transcript and mRNA, protein tertiary structure, molecular weight, tRNA repertoire);
- (b) depending on the specific state of the cell (the number of available ribosomes and translation initiation factors, the availability of tRNA, the rate of protein degradation).

The factors listed above can be included in a single regression model that makes it possible to link the mRNA level with the content of the encoded protein [39]. This approach was used by Frederic Edfors et al. [3]. The mRNA-to-protein ratio (RTP) for a particular gene is relatively constant between different tissues or cell lines. Thus, RTP is determined to a greater extent by the constant parameters of the gene or protein itself (ribosome advancement rate, tRNA preference, mRNA secondary structure, protein stability), rather than by a dynamic biological context that depends on the type of cell or tissue, i.e., the specific state of the cell [40]. The RTP values themselves between different genes can vary with a range of several orders of magnitude. This distribution differs between groups of genes involved in different biological processes and is associated with protein size in inverse dependence. In general, while the Pearson correlation coefficient between mRNA-

protein levels estimated by the authors averaged 0.6–0.7, such a correction for RTP made it possible to increase it to values exceeding 0.9 [3,41].

According to other researchers, the value of 0.9 is overestimated: the high variability between genes in different tissues (3×10^6) , as well as between different genes in the same tissue (1.7×10^5) , creates a high correlation for observed and predicted protein levels; even if for individual genes, this correlation is weak [38]. This effect is similar to Simpson's paradox: if two groups show the same trend, then the wrong combination of these data can change the direction of the relationship. To prove this assumption, Franks et al. [42] decided to reanalyze the raw data from the previous studies on human tissues [3,37,43]. Their results suggest a poor correlation (R = 0.33 for all measured mRNA across 12 tissues) between protein and mRNA contents for the same genes across various tissues, which they attribute to the extensive post-transcriptional regulation (Figure 1).

Interesting results of the group of Schwanhäusser et al. [5] shed light on the ratio of the cellular content of mRNA and protein. In 2011, the analysis of the content and turnover levels of mRNA and proteins corresponding to 5000 genes demonstrated that the half-lives of mRNA and protein do not correlate, in contrast to the levels of mRNA and protein levels. Moreover, the lifetime of proteins and mRNA is associated with the biological process in which the gene is involved.

In addition, it was shown that the cellular protein content is controlled mainly at the translation level [5]. Nevertheless, more than 85% of the variability in protein copy numbers (between samples, cells, or tissues) is determined by variability at the mRNA level when taking into account the gene-specific rate of translation and degradation, i.e., mRNA concentration remains a crucial factor for predicting protein levels [5].

For the first time, gene-centric coefficients of the ratio between mRNA and protein levels were obtained by Futcher et al. back in 1999 for a limited set of genes [30]. For example, on average, a single yeast (*S. cerevisiae*) cell contained only 54 mRNA molecules encoding actin (ACT1) and 160,000–205,000 actin protein molecules. For cytosolic aldehyde dehydrogenase (ALD6), this ratio was even more dramatic: three mRNA molecules and 160,000–180,000 protein molecules. Hence, the authors derived the following conclusion: the average doubling time of yeast colony is ca. 2 h, and one actin mRNA molecule accounts for ca. 4000 protein molecules. Therefore, the translation of each transcript is initiated approximately every 2 s. In turn, this means that if the average mRNA carries ten ribosomes involved in translation, then each ribosome completes translation in 20 s, assuming that the average protein has about 450 amino acid residues. Thus, it can be concluded that yeast is characterized by translating about 20 amino acids per second [30]. Interestingly, the same indicator for mammals is lower and amounts to approximately three to eight amino acids per second [44]. Following these observations, we can expect intertaxon variability in the correlation between mRNA and protein levels for different organisms.

Regarding relationships in the dynamics of protein and mRNA, the work of Cheng et al. [27] is of interest. The authors showed that in response to cellular stress (exposure to dithiothreitol), mRNA content increases abruptly, followed by a gradual increase in the concentration of encoded proteins (within several hours). At the same time, the protein content increases by a much greater amplitude than the mRNA level. This emphasizes that protein content is regulated both at the level of gene transcription and mRNA translation [27]. The authors of the study used a previously developed algorithm for isolating (deconvolution) the features of the regulation of protein and mRNA synthesis based on data on temporal profiling of the levels of transcripts and encoded proteins (PECA) [45]. Unlike experimental procedures such as SILAC [46], the PECA algorithm does not separate the intensities of protein synthesis and degradation.

One of the most striking works in the analysis of the relationship between transcriptomic and proteomic parameters is the study by Matthias Wilhelm, published in 2014 [37]. This is a landmark work on the drafting of the human proteome. The authors evaluated the gene-specific translation rates by median values in tissues [38]. A model for the quantitative relationship between protein level (iBAQ) and mRNA (RNA-Seq) was built based

on expression profiles for 12 human tissues. The resulting Spearman correlation coefficient between the measured levels of protein and mRNA (from R = 0.41 for the thyroid gland to R = 0.55 for the kidney) turned out to be lower than previously shown for cell lines [37] (Figure 1). In addition, the authors showed that cell lines indeed inherit the main features of the expression profiles of both genes and proteins from their progenitor tissues. Still, original tissues have greater variability in expression than the cell lines derived from them [37].

The most complete model based on many factors and linking the transcriptomic and proteomic levels of the organization of living systems is described in [47]. This work is a bioinformatics processing of data obtained in a large-scale study [10]. In the proposed model, the authors obtained mRNA concentration values by RNA sequencing of 29 healthy human tissues (in total, more than 11.5 thousand protein-coding genes were analyzed at the transcriptome and proteome levels).

The resulting kinetic model included the following parameters: the level of mRNA concentration, the level of protein concentration, the number of free ribosomes in the cell, the rate of translation, and the half-life of proteins and mRNA. The result of the work is the predicted level of TPR (the ratio of mRNA to the level of the corresponding protein) for 11.5 thousand genes. It was shown that the variability of this indicator between genes in one tissue significantly exceeds the variability between the same gene in different tissues, which proves the previous statements about the gene-specificity of TPR.

4. Regulation of Gene Expression

Gene expression is regulated at multiple levels, including transcription, translation, and post-translational modification. These processes encompass RNA synthesis (which includes epigenetic and transcriptional regulation), RNA degradation, protein synthesis (or translational control), and protein degradation. Collectively, they determine the protein pool in a cell, as outlined by the central dogma of molecular biology. While protein synthesis and degradation generally play a more significant role than RNA synthesis and degradation, all these processes are incorporated into predictive models.

The protein level is related to the level of the coding transcript by several "constant" factors determined by the mRNA sequence, protein stability, etc. That is why the difference in mRNA-protein levels undergoes more significant variability between genes (within the same tissue or cell line) than between different tissues for the same gene [10]. However, relying on "constant" coefficients alone is insufficient to predict protein levels [41]. The authors [41] showed that the concentration of the CD81 protein (transmembrane protein mediating signal transduction and developing complex with integrins) in different tissues varies by two orders of magnitude at the same mRNA concentrations. At the same time, the concentration of MEF2D (transcription factor) mRNA varies up to tenfold at the same protein concentrations.

Thus, RNA/protein synthesis and degradation depend on "constant" factors and the ones that depend on the state of the cell. So, these factors—DNA methylation [48], mRNA modifications [49], changes in the histone code [50], (de)condensation of chromatin [51], binding of various transcription factors [52], alternative splicing [53], the ratio of exon length to the total length of the transcribed region [54], and polyadenylation processes [55,56]—can be identified using special technology of DNA or RNA sequencing and bioinformatics algorithms.

The importance of integrating various data for all ways of gene expression regulation can be illustrated by the example of the influence of the composition of transcript isoforms encoding the protein on the rate of protein translation and degradation rate. Floor and Doudna [57] performed dissemination on the influence of transcript structure on protein translation efficiency. Their analysis strategy, Transcript Isoforms in Polysomes sequencing (TrIP-seq), combined polysome profiling with global gene expression analysis. TrIP-seq reveals transcript-isoform-specific translation patterns. In line with the previous findings, the authors confirmed that transcripts that contain the same ORF but different UTRs could have strikingly different translation rates. A few years later, Salovska et al. [58] showed that individual protein isoforms of the same genes can have different degradation rates, significantly impacting the protein abundance levels.

The epitranscriptome presents a complex landscape of post-transcriptional modifications that can significantly impact protein expression levels [49]. One prominent example is the m6A modification, which has been shown to both enhance [59], and inhibit translation processes [60]. Traditional methods for analyzing m6A modification sites in transcripts include m6A-seq, MeRIP-seq, miClip, and m6A-CLIP [61]. Alternatively, direct RNA sequencing performed on Oxford Nanopore technology can be processed through bioinformatic tools to identify m6A modification sites. Understanding the effects of such modifications on protein expression levels is critical for advancing our knowledge of cellular processes and disease mechanisms [62].

Another mechanism of translation regulation is RNA-interference, caused by microRNA (miRNA) and small interfering RNA (siRNA). There are different functions of gene regulation for these types of RNAs and one of them is to inhibit protein translation of the target mRNA [63–66]. For example, the expression of several microRNAs leads to translation inhibition, which in turn leads to the development of fibrosis [67]. Small RNA-seq (sRNA-seq), also called microRNA-seq (miRNA-seq), allows identifying miRNA and siRNA as a bulk mode [68], so as single-cell analysis [69].

In the context of translation regulation, it is also worth mentioning QTLs, which can be detected by DNA sequencing. So, for 30% of the 199 unique protein approved by the FDA as biomarkers, pQTLs were found [70]. The pQTLs explain the reports of different basic contents of these biomarkers in people of different races [71].

5. Translatome

Previously, it has been elucidated that the abundance of proteins results from a complex cascade of gene expression, which is meticulously regulated at each stage. The impact of regulatory factors extends beyond generating a specific set of transcripts. Numerous studies suggest that merely 40% of protein variability is determined at the transcriptome level [5,19].

Presumably, the major part of regulation comes from the process of translation [72]. Thus, the true predictive power remains in the translates, which refers to the functionally activated set of mRNAs undergoing protein synthesis [73].

Indeed, genes are transcribed and translated with varying intensity depending on the cell's current needs and context [74]. This is true for both complex cellular differentiation processes and maintaining homeostasis [75]. For example, for several genes involved in cellular development, an energetically unfavorable combination of active transcription and translational repression mechanisms (translational control) mediated by RNA-binding proteins has been observed [76]. The correlation will be negative until the cell requires the synthesis of that particular protein [77].

There are compensatory mechanisms in the evolution of genetic expression that maintain a stable protein composition [78]. Translational control does not always "sustain" the direction of the transcriptional one [79]. For example, post-transcriptional level mutations can restore the level of protein synthesis that was reduced due to transcriptional regulation [80]. From this perspective, the translatome is more significant as it is closer to the proteome in the chain of gene expression processes.

A meaningful study of this level of genetic expression would have been impossible without the development of sophisticated approaches. The earliest method, polysome profiling, was developed back in the 1960s. It is based on the assumption that ribosomes, the largest macromolecular complexes in most cells, will sediment in a sucrose gradient faster than other organelles [72].

The revolutionary method named ribosome profiling (Ribo-seq) implements deep sequencing of mRNA fragments, protected from the RNases' "attack" by the binding of ribosome complexes, and therefore called 'footprints' [81]. As the control sample pro-

ceeds traditional mRNA sequencing, on the output researchers gain an overview of how ribosomes move along transcripts. This is characterized by the translation efficiency (TE) rate [81], which gives the idea of ribosome occupancy per mRNA. It can vary widely for different transcripts of life stages [82]. Meanwhile, as indicated in a study on drosophila as a model object in steady state (via TRAP methodology), most abundant transcripts are not those carrying the main amount of ribosomes [83].

Another, already noted approach is to tag polysomes via protein fusion to gain specific affinity (ribosome affinity purification (RAP) or translating RAP (TRAP)) [68]. Ribosome profiling is characterized by short readouts (depending on the chosen RNase [84,85]), which makes this method sensitive to noise. However, today it is the most common method for translatome research. With the ability of this method to provide information on the "location" of ribosomes down to the nucleotide level, it can be used to study and characterize both canonical and non-canonical ORFs [86]. The discovery of these ORFs significantly expands the number of coding sequences (CDS).

A number of studies have compared the correlation between the translatome and the transcriptome with the proteome in lower eukaryotes [82,87]. It has been generally shown that the translatome correlates more with the proteome than the transcriptome. Thus, the Spearman correlation between the proteome and the translatome (Ribo-seq) in normal *S. cerevisiae* cells was 0.77 [87]. Contrariwise, the correlation between the proteome and transcriptome was 0.46.

In contrast to what has been discussed earlier for transcriptome, translatome is characterized by far less divergence in gene expression between tissues [79]. It indicates the evidence of post-transcriptional buffering [88].

In the study conducted on human tissues, the correlation between the translatome and proteome was measured for 9642 genes present in every sample. For organs such as the brain, liver and testis, Spearman's correlation coefficient yielded values $\rho = 0.65$, 0.69, and 0.60, respectively. In comparison, the corresponding correlation ratios between the transcriptome and proteome were $\rho = 0.57$, 0.61, and 0.42 [10,89].

As previously mentioned, a high correlation is associated with the contribution of post-transcriptional regulation. For instance, in an experiment on a mouse cell line, it is independently implemented for 20% of differentially expressed genes [79].

The primary mechanisms of translational control are specifically directed towards regulating initiation [90], although they can also be applied to post-initiation steps, elongation, and termination of translation.

Interesting findings were obtained in an experiment focused on the effects of hypoxic stress on cardiomyocytes [91]. It appeared that with the increase in the intensity of the stress stimulus, cells may change their preference of a strategy aimed at enhancing the expression of specific genes. For example, short-term hypoxia results in changes at the translatome level by increasing ribosome recruitment on mRNA (upregulation via binding of NCBP3 protein to 5'-UTR). Indeed, enhancing translation is the fastest way to increase the representation of necessary proteins within a cell. Translational load focuses on genes associated with the HIF-1 signaling cascade. This cascade induces rapid changes in cellular physiology [92]. On the contrary, prolonged hypoxia leads to enhanced synthesis of a specific subset of transcripts, involving genes associated with autophagy, apoptosis, and cell proliferation. These are the processes associated with long-term effects or novel cell functions, in which implementation requires the activation of previously silent genes. Post-translational regulation adjusts pre-existing protein composition to a new cellular context [26].

Translational control also takes precedence when regulation at the transcriptional level is not manifested. This is the case in developmental processes [93]. Another curious example is the case of trypanosomatids' differentiation into an infective form, completely regulated post-transcriptionally [82]. There is evidence of changes in the translatome during disease or malignant transformation of an organ [94].

Translatomics and transcriptomics in this paradigm should complement each other. The perfect model for protein abundance prediction probably considers the strength of their respective contributions.

6. Single-Cell Transcriptomics-Proteomics

The previous discussion was dedicated only to insights derived from studies of cell populations at the bulk level. However, a question arises of how well the observed relationship between the transcriptome and the proteome carries over to the level of individual cells. Thanks to the recent advancements in the field of single-cell transcriptomics, the picture of the regulation of gene expression at the level of RNA is starting to become clear [95]. First of all, it is now well established that gene expression in individual living cells occurs in the form of stochastic bursts, also called transcriptional bursts [96]. This phenomenon introduces a significant amount of variation in the cellular levels of individual mRNAs, which can be difficult to separate from other regulated factors affecting gene regulation, such as the activity of transcriptional factors, but approaches to mitigating this problem are being developed [97]. Additional challenges for single-cell-level quantitative analysis include the detection of products of low-expressed genes [98], and the accurate quantification of transcript isoforms [99]. Despite these and other challenges [100], singlecell transcriptomics is progressing rapidly. It has already provided high-quality portraits of gene expression for various complex cell populations such as tissues, revealing important details about vital biological processes [101–104].

The robustness of single-cell transcriptomics techniques is largely attributed to the availability of methods for amplifying individual DNA molecules, which, as mentioned at the beginning of the article, is currently unavailable for proteomics. Despite this, approaches to adapt mass spectrometry-based identification of proteins for single cells are underway [105]. As these methods feature significant downscaling, most of the published methods of this category so far require drastic modifications to the typical mass spectrometry protocol, and, as a consequence, the capability of such methods currently lags behind the respective transcriptomics techniques [106]. Still, as these approaches are constantly refined and updated [107], we are now starting to see the first published global comparisons of transcriptomes and proteomes for single cells. In a recent paper by Brunner et al., a novel single-cell mass spectrometry workflow (T-SCP) was separately complemented by two established single-cell RNA-Seq techniques (SMARTSeq2 and Drop-Seq) to investigate the proteogenomics of the HeLa cell line, and products of 1672 genes were detected by all three methods [108]. Surprisingly, HeLa cells demonstrated a much better correlation between each other at the proteome, rather than the transcriptome level, once again highlighting the aforementioned highly stochastic nature of gene expression in individual cells. To investigate transcriptome-to-proteome correlation, the coefficients of variation for shared genes between all data sets were compared, revealing that variation at the gene level did not correlate well between the transcriptome and the proteome, in line with the insights from the bulk-level analysis. One more study, currently available as a preprint, reports the development of a protocol for the parallel measurement of transcriptomes and proteomes in single cells using nanodroplet splitting. After successfully applying this protocol to C10 and SVEC cells, a subsequent correlation analysis between transcript and protein abundances produced Pearson's correlation coefficients ranging from 0.31 to 0.56, with values once again similar to previous analyses. In conclusion, even though transcriptional regulation is significantly different at the level of single cells than at the bulk level, presently available data suggests that the correlation between the transcriptome and the proteome remains at the same, rather modest levels.

Parallel profiling at several omics levels of proteins and nucleic acids isolated from a single cell eliminates the problem of genotypic or phenotypic heterogeneity of the bulk sample. The transition from an "average" cell to a "single" cell is an unprecedented opportunity to unambiguously determine the correlation between the transcriptome and the proteome and, subsequently, with the phenotype [109].

To understand how a single cell relates to its averaged portrait obtained from cells in bulk is a difficult but manageable task [100]. In recent years, several methods have been developed to robustly and safely isolate individual cells and quantify their content (i.e., laser capture microdissection [110], robotic micromanipulation [111], microfluidics coupled with RNA-Seq or proteomics techniques [112] and fluorescent methods—fluorescence-activated cell sorting (FACS) [113] or fluorescence microscopy [114]). One of the most straightforward and elegant approaches is via a proximity-based assay, such as the proximity ligation assay (PLA) or proximity extension assay (PEA) [115]. These assays allow amplifying the signal from a single protein (or a pair of interacting proteins) with a PCR-like mechanism, greatly increasing the sensitivity. Briefly, the PEA features two protein-specific antibodies, which are tagged with complementary oligonucleotides. When both of these antibodies interact with the same protein molecule, and are thus in proximity (hence the method name), complementary oligonucleotides hybridize, creating a site that can be amplified with specifically designed primers, thus increasing the copy number of this nucleotide segment, which can be detected via RT-PCR or NGS methods.

7. Conclusions

While the genome contains information about liability to disease, and the transcriptome can be reliably quantified, only the accurate measurements of the proteome and the metabolome can most fully reveal the current state of the organism, i.e., create a digital profile of the organism for a particular time point [116]. Thus, there is a need to develop a high-quality model of the relationship between transcriptomic and proteomic data since the current capabilities of RNA-seq are much superior to proteomic methods. Assessing the correlation between different layers of information transfer inside the cell is important for multi-omics analyses, allowing for addressing the missing data or even fully predicting one omics layer from another [117].

The knowledge of transcript abundance is insufficient for predicting protein abundance due to many factors. A cohort of large-scale investigations aimed at comparing the levels of transcripts and proteins at a genome-wide scale have led to the identification of several key factors affecting the quantitative relationship between the transcriptome and the proteome. These include the rate of translation (which in turn depends on the mRNA sequence), the activity of the regulatory elements (miRNAs and other non-coding RNAs), the relative availability of ribosomes, and protein degradation rate. Notably, these factors are highly dynamic and often specific for a particular protein, which makes the development of global predictive models extremely challenging.

However, current capabilities of nucleic acid sequencing methods provide a wide scope for studying the processes of expression regulation. Taking into account a number of factors, the modifications of DNA and RNA nucleotides, the level of expression of transcripts and small RNAs, and the level of translation, etc., obtained through specialized sequencing approaches, allows us to adjust the mRNA-protein model, achieving high prediction rates.

Author Contributions: Conceptualization, E.A.P., E.V.P. and O.I.K.; writing—original draft preparation, G.S.K., O.I.K., P.A.K., V.A.A., G.V.D., E.V.I. and E.V.P.; writing—review and editing, P.A.K., A.V.L. and E.V.P.; visualization, O.I.K. and P.A.K.; supervision, E.A.P. and E.V.P.; project administration, E.A.P., A.V.L. and E.V.P. All authors have read and agreed to the published version of the manuscript.

Funding: The study was performed employing "Avogadro" large-scale research facilities, and was financially supported by the Ministry of Education and Science of the Russian Federation, Agreement No. 075-15-2021-933, unique project ID: RF00121X0004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Aebersold, R.; Blattmann, P. Mass Spectrometric Exploration of the Biochemical Basis of Living Systems. *Chimia* 2019, 73, 540–548. [CrossRef] [PubMed]
- Vitrinel, B.; Koh, H.W.L.; Mujgan Kar, F.; Maity, S.; Rendleman, J.; Choi, H.; Vogel, C. Exploiting Interdata Relationships in Next-Generation Proteomics Analysis. *Mol. Cell Proteom. MCP* 2019, *18*, S5–S14. [CrossRef] [PubMed]
- Edfors, F.; Danielsson, F.; Hallström, B.M.; Käll, L.; Lundberg, E.; Pontén, F.; Forsström, B.; Uhlén, M. Gene-Specific Correlation of RNA and Protein Levels in Human Cells and Tissues. *Mol. Syst. Biol.* 2016, *12*, 883. [CrossRef] [PubMed]
- Spainhour, J.C.; Lim, H.S.; Yi, S.V.; Qiu, P. Correlation Patterns Between DNA Methylation and Gene Expression in The Cancer Genome Atlas. *Cancer Inform.* 2019, 18, 1176935119828776. [CrossRef]
- 5. Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global Quantification of Mammalian Gene Expression Control. *Nature* 2011, 473, 337–342. [CrossRef] [PubMed]
- Smyczynska, U.; Stanczak, M.; Kuljanin, M.; Włodarczyk, A.; Stoczynska-Fidelus, E.; Taha, J.; Pawlik, B.; Borowiec, M.; Mancias, J.D.; Mlynarski, W.; et al. Proteomic and Transcriptomic Landscapes of Alström and Bardet–Biedl Syndromes. *Genes* 2022, 13, 2370. [CrossRef] [PubMed]
- Archakov, A.; Aseev, A.; Bykov, V.; Grigoriev, A.; Govorun, V.; Ivanov, V.; Khlunov, A.; Lisitsa, A.; Mazurenko, S.; Makarov, A.A.; et al. Gene-Centric View on the Human Proteome Project: The Example of the Russian Roadmap for Chromosome 18. *Proteomics* 2011, 11, 1853–1856. [CrossRef]
- Poverennaya, E.V.; Ilgisonis, E.V.; Ponomarenko, E.A.; Kopylov, A.T.; Zgoda, V.G.; Radko, S.P.; Lisitsa, A.V.; Archakov, A.I. Why Are the Correlations between mRNA and Protein Levels so Low among the 275 Predicted Protein-Coding Genes on Human Chromosome 18? J. Proteome Res. 2017, 16, 4311–4318. [CrossRef]
- Zgoda, V.G.; Kopylov, A.T.; Tikhonova, O.V.; Moisa, A.A.; Pyndyk, N.V.; Farafonova, T.E.; Novikova, S.E.; Lisitsa, A.V.; Ponomarenko, E.A.; Poverennaya, E.V.; et al. Chromosome 18 Transcriptome Profiling and Targeted Proteome Mapping in Depleted Plasma, Liver Tissue and HepG2 Cells. *J. Proteome Res.* 2013, *12*, 123–134. [CrossRef]
- 10. Wang, D.; Eraslan, B.; Wieland, T.; Hallström, B.; Hopf, T.; Zolg, D.P.; Zecha, J.; Asplund, A.; Li, L.-H.; Meng, C.; et al. A Deep Proteome and Transcriptome Abundance Atlas of 29 Healthy Human Tissues. *Mol. Syst. Biol.* **2019**, *15*, e8503. [CrossRef]
- Fagerberg, L.; Hallström, B.M.; Oksvold, P.; Kampf, C.; Djureinovic, D.; Odeberg, J.; Habuka, M.; Tahmasebpoor, S.; Danielsson, A.; Edlund, K.; et al. Analysis of the Human Tissue-Specific Expression by Genome-Wide Integration of Transcriptomics and Antibody-Based Proteomics. *Mol. Cell Proteom. MCP* 2014, *13*, 397–406. [CrossRef] [PubMed]
- 12. Ramsköld, D.; Wang, E.T.; Burge, C.B.; Sandberg, R. An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data. *PLoS Comput. Biol.* **2009**, *5*, e1000598. [CrossRef] [PubMed]
- Wang, T.; Cui, Y.; Jin, J.; Guo, J.; Wang, G.; Yin, X.; He, Q.-Y.; Zhang, G. Translating mRNAs Strongly Correlate to Proteins in a Multivariate Manner and Their Translation Ratios Are Phenotype Specific. *Nucleic Acids Res.* 2013, 41, 4743–4754. [CrossRef] [PubMed]
- 14. Kiseleva, O.; Ponomarenko, E.; Poverennaya, E. Empowering Shotgun Mass Spectrometry with 2DE: A HepG2 Study. *Int. J. Mol. Sci.* 2020, *21*, 3813. [CrossRef]
- 15. Lisitsa, A.; Moshkovskii, S.; Chernobrovkin, A.; Ponomarenko, E.; Archakov, A. Profiling Proteoforms: Promising Follow-up of Proteomics for Biomarker Discovery. *Expert Rev. Proteom.* **2014**, *11*, 121–129. [CrossRef]
- 16. Song, C.; Wang, F.; Cheng, K.; Wei, X.; Bian, Y.; Wang, K.; Tan, Y.; Wang, H.; Ye, M.; Zou, H. Large-Scale Quantification of Single Amino-Acid Variations by a Variation-Associated Database Search Strategy. *J. Proteome Res.* **2014**, *13*, 241–248. [CrossRef]
- 17. Cao, R.; Shi, Y.; Chen, S.; Ma, Y.; Chen, J.; Yang, J.; Chen, G.; Shi, T. dbSAP: Single Amino-Acid Polymorphism Database for Protein Variation Detection. *Nucleic Acids Res.* 2017, 45, D827–D832. [CrossRef]
- 18. Yang, Y.; Nan, Y.; Cai, J.; Xu, J.; Huang, Z.; Cai, X. The Thr to Met Substitution of Amino Acid 118 in Hepatitis B Virus Surface Antigen Escapes from Immune-Assay-Based Screening of Blood Donors. *J. Gen. Virol.* **2016**, *97*, 1210–1217. [CrossRef]
- 19. Liu, Y.; Beyer, A.; Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **2016**, *165*, 535–550. [CrossRef]
- Archakov, A.; Zgoda, V.; Kopylov, A.; Naryzhny, S.; Chernobrovkin, A.; Ponomarenko, E.; Lisitsa, A. Chromosome-Centric Approach to Overcoming Bottlenecks in the Human Proteome Project. *Expert Rev. Proteom.* 2012, 9, 667–676. [CrossRef]
- Yuan, Z.; Liu, X.; Liu, C.; Zhang, Y.; Rao, Y. Recent Advances in Rapid Synthesis of Non-Proteinogenic Amino Acids from Proteinogenic Amino Acids Derivatives via Direct Photo-Mediated C-H Functionalization. *Molecules* 2020, 25, 5270. [CrossRef] [PubMed]
- 22. Giansanti, P.; Tsiatsiani, L.; Low, T.Y.; Heck, A.J.R. Six Alternative Proteases for Mass Spectrometry-Based Proteomics beyond Trypsin. *Nat. Protoc.* 2016, *11*, 993–1006. [CrossRef] [PubMed]
- Ilgisonis, E.V.; Kopylov, A.T.; Ponomarenko, E.A.; Poverennaya, E.V.; Tikhonova, O.V.; Farafonova, T.E.; Novikova, S.; Lisitsa, A.V.; Zgoda, V.G.; Archakov, A.I. Increased Sensitivity of Mass Spectrometry by Alkaline Two-Dimensional Liquid Chromatography: Deep Cover of the Human Proteome in Gene-Centric Mode. *J. Proteome Res.* 2018, *17*, 4258–4266. [CrossRef] [PubMed]
 Control F. C., Control F. M., Control F. M., Control F. C., C., Control F. C., C., C.,
- 24. Crick, F. Central Dogma of Molecular Biology. Nature 1970, 227, 561–563. [CrossRef] [PubMed]
- Chick, J.M.; Munger, S.C.; Simecek, P.; Huttlin, E.L.; Choi, K.; Gatti, D.M.; Raghupathy, N.; Svenson, K.L.; Churchill, G.A.; Gygi, S.P. Defining the Consequences of Genetic Variation on a Proteome-Wide Scale. *Nature* 2016, 534, 500–505. [CrossRef]

- 26. Vogel, C.; Marcotte, E.M. Insights into the Regulation of Protein Abundance from Proteomic and Transcriptomic Analyses. *Nat. Rev. Genet.* **2012**, *13*, 227–232. [CrossRef]
- 27. Cheng, Z.; Teo, G.; Krueger, S.; Rock, T.M.; Koh, H.W.L.; Choi, H.; Vogel, C. Differential Dynamics of the Mammalian mRNA and Protein Expression Response to Misfolding Stress. *Mol. Syst. Biol.* **2016**, *12*, 855. [CrossRef]
- Greenbaum, D.; Colangelo, C.; Williams, K.; Gerstein, M. Comparing Protein Abundance and mRNA Expression Levels on a Genomic Scale. *Genome Biol.* 2003, 4, 117. [CrossRef]
- Gygi, S.P.; Rochon, Y.; Franza, B.R.; Aebersold, R. Correlation between Protein and mRNA Abundance in Yeast. *Mol. Cell Biol.* 1999, 19, 1720–1730. [CrossRef]
- Futcher, B.; Latter, G.I.; Monardo, P.; McLaughlin, C.S.; Garrels, J.I. A Sampling of the Yeast Proteome. *Mol. Cell Biol.* 1999, 19, 7357–7368. [CrossRef]
- Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E.M. Absolute Protein Expression Profiling Estimates the Relative Contributions of Transcriptional and Translational Regulation. *Nat. Biotechnol.* 2007, 25, 117–124. [CrossRef] [PubMed]
- Ghazalpour, A.; Bennett, B.; Petyuk, V.A.; Orozco, L.; Hagopian, R.; Mungrue, I.N.; Farber, C.R.; Sinsheimer, J.; Kang, H.M.; Furlotte, N.; et al. Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLoS Genet.* 2011, 7, e1001393. [CrossRef] [PubMed]
- Peng, X.; Qin, Z.; Zhang, G.; Guo, Y.; Huang, J. Integration of the Proteome and Transcriptome Reveals Multiple Levels of Gene Regulation in the Rice Dl2 Mutant. *Front. Plant Sci.* 2015, *6*, 351. [CrossRef] [PubMed]
- 34. Jia, H.; Sun, W.; Li, M.; Zhang, Z. Integrated Analysis of Protein Abundance, Transcript Level, and Tissue Diversity to Reveal Developmental Regulation of Maize. *J. Proteome Res.* **2018**, *17*, 822–833. [CrossRef] [PubMed]
- Torres-García, W.; Zhang, W.; Runger, G.C.; Johnson, R.H.; Meldrum, D.R. Integrative Analysis of Transcriptomic and Proteomic Data of Desulfovibrio Vulgaris: A Non-Linear Model to Predict Abundance of Undetected Proteins. *Bioinformatics* 2009, 25, 1905–1914. [CrossRef]
- Lundberg, E.; Fagerberg, L.; Klevebring, D.; Matic, I.; Geiger, T.; Cox, J.; Algenäs, C.; Lundeberg, J.; Mann, M.; Uhlen, M. Defining the Transcriptome and Proteome in Three Functionally Different Human Cell Lines. *Mol. Syst. Biol.* 2010, *6*, 450. [CrossRef]
- Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A.M.; Lieberenz, M.; Savitski, M.M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; et al. Mass-Spectrometry-Based Draft of the Human Proteome. *Nature* 2014, 509, 582–587. [CrossRef]
- Fortelny, N.; Overall, C.M.; Pavlidis, P.; Freue, G.V.C. Can We Predict Protein from mRNA Levels? *Nature* 2017, 547, E19–E20.
 [CrossRef]
- 39. Nie, L.; Wu, G.; Zhang, W. Correlation between mRNA and Protein Abundance in Desulfovibrio Vulgaris: A Multiple Regression to Identify Sources of Variations. *Biochem. Biophys. Res. Commun.* **2006**, *339*, 603–610. [CrossRef]
- 40. Santos, F.B.; Del-Bem, L.-E. The Evolution of tRNA Copy Number and Repertoire in Cellular Life. Genes 2023, 14, 27. [CrossRef]
- Silva, G.M.; Vogel, C. Quantifying Gene Expression: The Importance of Being Subtle. *Mol. Syst. Biol.* 2016, 12, 885. [CrossRef] [PubMed]
- Franks, A.; Airoldi, E.; Slavov, N. Post-Transcriptional Regulation across Human Tissues. *PLoS Comput. Biol.* 2017, 13, e1005535. [CrossRef] [PubMed]
- 43. Kim, M.-S.; Pinto, S.M.; Getnet, D.; Nirujogi, R.S.; Manda, S.S.; Chaerkady, R.; Madugundu, A.K.; Kelkar, D.S.; Isserlin, R.; Jain, S.; et al. A Draft Map of the Human Proteome. *Nature* 2014, *509*, 575–581. [CrossRef] [PubMed]
- 44. Hershey, J.W.B.; Sonenberg, N.; Mathews, M.B. Principles of Translational Control. *Cold Spring Harb. Perspect. Biol.* 2019, 11, a032607. [CrossRef]
- 45. Teo, G.; Vogel, C.; Ghosh, D.; Kim, S.; Choi, H. PECA: A Novel Statistical Tool for Deconvoluting Time-Dependent Gene Expression Regulation. *J. Proteome Res.* 2014, *13*, 29–37. [CrossRef] [PubMed]
- Doherty, M.K.; Hammond, D.E.; Clague, M.J.; Gaskell, S.J.; Beynon, R.J. Turnover of the Human Proteome: Determination of Protein Intracellular Stability by Dynamic SILAC. J. Proteome Res. 2009, 8, 104–112. [CrossRef]
- Eraslan, B.; Wang, D.; Gusic, M.; Prokisch, H.; Hallström, B.M.; Uhlén, M.; Asplund, A.; Pontén, F.; Wieland, T.; Hopf, T.; et al. Quantification and Discovery of Sequence Determinants of Protein-per-mRNA Amount in 29 Human Tissues. *Mol. Syst. Biol.* 2019, 15, e8513. [CrossRef]
- 48. Besser, D.; Götz, F.; Schulze-Forster, K.; Wagner, H.; Kröger, H.; Simon, D. DNA Methylation Inhibits Transcription by RNA Polymerase III of a tRNA Gene, but Not of a 5S rRNA Gene. *FEBS Lett.* **1990**, *269*, 358–362. [CrossRef]
- Arzumanian, V.A.; Dolgalev, G.V.; Kurbatov, I.Y.; Kiseleva, O.I.; Poverennaya, E.V. Epitranscriptome: Review of Top 25 Most-Studied RNA Modifications. *Int. J. Mol. Sci.* 2022, 23, 13851. [CrossRef]
- Jimeno-González, S.; Payán-Bravo, L.; Muñoz-Cabello, A.M.; Guijo, M.; Gutierrez, G.; Prado, F.; Reyes, J.C. Defective Histone Supply Causes Changes in RNA Polymerase II Elongation Rate and Cotranscriptional Pre-mRNA Splicing. *Proc. Natl. Acad. Sci.* USA 2015, 112, 14840–14845. [CrossRef]
- Chambeyron, S.; Bickmore, W.A. Chromatin Decondensation and Nuclear Reorganization of the HoxB Locus upon Induction of Transcription. *Genes Dev.* 2004, 18, 1119–1130. [CrossRef] [PubMed]
- Transcription Factors—ScienceDirect. Available online: https://www.sciencedirect.com/science/article/abs/pii/B97801280123 83054660 (accessed on 6 October 2023).
- Alternative RNA Splicing and Editing: A Functional Molecular Tool Directed to Successful Protein Synthesis in Plants | SpringerLink. Available online: https://link.springer.com/chapter/10.1007/978-3-030-68828-8_5 (accessed on 6 October 2023).

- 54. Hildyard, J.C.W.; Piercy, R.J. When Size Really Matters: The Eccentricities of Dystrophin Transcription and the Hazards of Quantifying mRNA from Very Long Genes. *Biomedicines* **2023**, *11*, 2082. [CrossRef] [PubMed]
- 55. Park, J.-E.; Yi, H.; Kim, Y.; Chang, H.; Kim, V.N. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Mol. Cell* **2016**, *62*, 462–471. [CrossRef] [PubMed]
- Chang, H.; Lim, J.; Ha, M.; Kim, V.N. TAIL-Seq: Genome-Wide Determination of Poly(A) Tail Length and 3' End Modifications. Mol. Cell 2014, 53, 1044–1052. [CrossRef]
- 57. Floor, S.N.; Doudna, J.A. Tunable Protein Synthesis by Transcript Isoforms in Human Cells. eLife 2016, 5, e10921. [CrossRef]
- Salovska, B.; Zhu, H.; Gandhi, T.; Frank, M.; Li, W.; Rosenberger, G.; Wu, C.; Germain, P.-L.; Zhou, H.; Hodny, Z.; et al. Isoform-Resolved Correlation Analysis between mRNA Abundance Regulation and Protein Level Degradation. *Mol. Syst. Biol.* 2020, 16, e9170. [CrossRef]
- 59. Barbieri, I.; Kouzarides, T. Role of RNA Modifications in Cancer. Nat. Rev. Cancer 2020, 20, 303–322. [CrossRef]
- 60. Fernandez Rodriguez, G.; Cesaro, B.; Fatica, A. Multiple Roles of m6A RNA Modification in Translational Regulation in Cancer. *Int. J. Mol. Sci.* **2022**, 23, 8971. [CrossRef]
- 61. Zhu, W.; Wang, J.-Z.; Xu, Z.; Cao, M.; Hu, Q.; Pan, C.; Guo, M.; Wei, J.-F.; Yang, H. Detection of N6-methyladenosine Modification Residues (Review). *Int. J. Mol. Med.* **2019**, *43*, 2267–2278. [CrossRef]
- 62. Zhong, Z.-D.; Xie, Y.-Y.; Chen, H.-X.; Lan, Y.-L.; Liu, X.-H.; Ji, J.-Y.; Wu, F.; Jin, L.; Chen, J.; Mak, D.W.; et al. Systematic Comparison of Tools Used for m6A Mapping from Nanopore Direct RNA Sequencing. *Nat. Commun.* **2023**, *14*, 1906. [CrossRef]
- 63. Williams, A.E. Functional Aspects of Animal microRNAs. Cell Mol. Life Sci. 2008, 65, 545–562. [CrossRef] [PubMed]
- 64. Hu, X.; Yin, G.; Zhang, Y.; Zhu, L.; Huang, H.; Lv, K. Recent Advances in the Functional Explorations of Nuclear microRNAs. *Front. Immunol.* **2023**, *14*, 1097491. [CrossRef] [PubMed]
- 65. Gu, S.; Rossi, J.J. Uncoupling of RNAi from Active Translation in Mammalian Cells. RNA 2005, 11, 38–44. [CrossRef] [PubMed]
- Neumeier, J.; Meister, G. siRNA Specificity: RNAi Mechanisms and Strategies to Reduce Off-Target Effects. Front. Plant Sci. 2020, 11, 526455. [CrossRef]
- Mullenbrock, S.; Liu, F.; Szak, S.; Hronowski, X.; Gao, B.; Juhasz, P.; Sun, C.; Liu, M.; McLaughlin, H.; Xiao, Q.; et al. Systems Analysis of Transcriptomic and Proteomic Profiles Identifies Novel Regulation of Fibrotic Programs by miRNAs in Pulmonary Fibrosis Fibroblasts. *Genes* 2018, 9, 588. [CrossRef]
- Pantaleão, L.C.; Ozanne, S.E. Small RNA Sequencing: A Technique for miRNA Profiling. *Methods Mol. Biol.* 2018, 1735, 321–330.
 [CrossRef]
- Hücker, S.M.; Fehlmann, T.; Werno, C.; Weidele, K.; Lüke, F.; Schlenska-Lange, A.; Klein, C.A.; Keller, A.; Kirsch, S. Single-Cell microRNA Sequencing Method Comparison and Application to Cell Lines and Circulating Lung Tumor Cells. *Nat. Commun.* 2021, 12, 4316. [CrossRef]
- Suhre, K.; McCarthy, M.I.; Schwenk, J.M. Genetics Meets Proteomics: Perspectives for Large Population-Based Studies. *Nat. Rev. Genet.* 2021, 22, 19–37. [CrossRef]
- Sjaarda, J.; Gerstein, H.C.; Kutalik, Z.; Mohammadi-Shemirani, P.; Pigeyre, M.; Hess, S.; Paré, G. Influence of Genetic Ancestry on Human Serum Proteome. Am. J. Hum. Genet. 2020, 106, 303–314. [CrossRef]
- 72. Zhao, J.; Qin, B.; Nikolay, R.; Spahn, C.M.T.; Zhang, G. Translatomics: The Global View of Translation. *Int. J. Mol. Sci.* 2019, 20, 212. [CrossRef]
- 73. Ingolia, N.T. Ribosome Footprint Profiling of Translation throughout the Genome. Cell 2016, 165, 22–33. [CrossRef] [PubMed]
- 74. Teixeira, F.K.; Lehmann, R. Translational Control during Developmental Transitions. *Cold Spring Harb. Perspect. Biol.* 2019, 11, a032987. [CrossRef] [PubMed]
- Buszczak, M.; Signer, R.A.J.; Morrison, S.J. Cellular Differences in Protein Synthesis Regulate Tissue Homeostasis. *Cell* 2014, 159, 242–251. [CrossRef] [PubMed]
- Snee, M.; Benz, D.; Jen, J.; Macdonald, P.M. Two Distinct Domains of Bruno Bind Specifically to the Oskar mRNA. RNA Biol. 2008, 5, 49–57. [CrossRef]
- Chang, J.S.; Tan, L.; Schedl, P. The Drosophila CPEB Homolog, Orb, Is Required for Oskar Protein Expression in Oocytes. *Dev. Biol.* 1999, 215, 91–106. [CrossRef] [PubMed]
- Stadler, M.; Fire, A. Conserved Translatome Remodeling in Nematode Species Executing a Shared Developmental Transition. PLoS Genet. 2013, 9, e1003739. [CrossRef] [PubMed]
- 79. Wang, H.; Wang, Y.; Yang, J.; Zhao, Q.; Tang, N.; Chen, C.; Li, H.; Cheng, C.; Xie, M.; Yang, Y.; et al. Tissue- and Stage-Specific Landscape of the Mouse Translatome. *Nucleic Acids Res.* **2021**, *49*, 6165–6180. [CrossRef]
- Khan, Z.; Ford, M.J.; Cusanovich, D.A.; Mitrano, A.; Pritchard, J.K.; Gilad, Y. Primate Transcript and Protein Expression Levels Evolve under Compensatory Selection Pressures. *Science* 2013, 342, 1100–1104. [CrossRef]
- Ingolia, N.T.; Ghaemmaghami, S.; Newman, J.R.S.; Weissman, J.S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* 2009, 324, 218–223. [CrossRef]
- Smircich, P.; Eastman, G.; Bispo, S.; Duhagon, M.A.; Guerra-Slompo, E.P.; Garat, B.; Goldenberg, S.; Munroe, D.J.; Dallagiovanna, B.; Holetz, F.; et al. Ribosome Profiling Reveals Translation Control as a Key Mechanism Generating Differential Gene Expression in Trypanosoma Cruzi. *BMC Genom.* 2015, 16, 443. [CrossRef]
- 83. Thomas, A.; Lee, P.-J.; Dalton, J.E.; Nomie, K.J.; Stoica, L.; Costa-Mattioli, M.; Chang, P.; Nuzhdin, S.; Arbeitman, M.N.; Dierick, H.A. A Versatile Method for Cell-Specific Profiling of Translated mRNAs in Drosophila. *PLoS ONE* **2012**, *7*, e40276. [CrossRef]

- Inada, T.; Winstall, E.; Tarun, S.Z.; Yates, J.R.; Schieltz, D.; Sachs, A.B. One-Step Affinity Purification of the Yeast Ribosome and Its Associated Proteins and mRNAs. RNA 2002, 8, 948–958. [CrossRef] [PubMed]
- Jin, H.Y.; Xiao, C. An Integrated Polysome Profiling and Ribosome Profiling Method to Investigate In Vivo Translatome. *Methods Mol. Biol.* 2018, 1712, 1–18. [CrossRef] [PubMed]
- Ruiz Cuevas, M.V.; Hardy, M.P.; Hollý, J.; Bonneil, É.; Durette, C.; Courcelles, M.; Lanoix, J.; Côté, C.; Staudt, L.M.; Lemieux, S.; et al. Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Rep.* 2021, 34, 108815. [CrossRef] [PubMed]
- Blevins, W.R.; Tavella, T.; Moro, S.G.; Blasco-Moreno, B.; Closa-Mosquera, A.; Díez, J.; Carey, L.B.; Albà, M.M. Extensive Post-Transcriptional Buffering of Gene Expression in the Response to Severe Oxidative Stress in Baker's Yeast. *Sci. Rep.* 2019, *9*, 11005. [CrossRef]
- Buccitelli, C.; Selbach, M. mRNAs, Proteins and the Emerging Principles of Gene Expression Control. Nat. Rev. Genet. 2020, 21, 630–644. [CrossRef]
- Wang, Z.-Y.; Leushkin, E.; Liechti, A.; Ovchinnikova, S.; Mößinger, K.; Brüning, T.; Rummel, C.; Grützner, F.; Cardoso-Moreira, M.; Janich, P.; et al. Transcriptome and Translatome Co-Evolution in Mammals. *Nature* 2020, 588, 642–647. [CrossRef]
- 90. Gebauer, F.; Hentze, M.W. Molecular Mechanisms of Translational Control. Nat. Rev. Mol. Cell Biol. 2004, 5, 827–835. [CrossRef]
- Shen, Z.; Zeng, L.; Zhang, Z. Translatome and Transcriptome Profiling of Hypoxic-Induced Rat Cardiomyocytes. *Mol. Ther. Nucleic Acids* 2020, 22, 1016–1024. [CrossRef]
- Wang, G.L.; Jiang, B.H.; Rue, E.A.; Semenza, G.L. Hypoxia-Inducible Factor 1 Is a Basic-Helix-Loop-Helix-PAS Heterodimer Regulated by Cellular O2 Tension. *Proc. Natl. Acad. Sci. USA* 1995, 92, 5510–5514. [CrossRef]
- Hu, W.; Zeng, H.; Shi, Y.; Zhou, C.; Huang, J.; Jia, L.; Xu, S.; Feng, X.; Zeng, Y.; Xiong, T.; et al. Single-Cell Transcriptome and Translatome Dual-Omics Reveals Potential Mechanisms of Human Oocyte Maturation. *Nat. Commun.* 2022, 13, 5114. [CrossRef] [PubMed]
- 94. Lian, X.; Guo, J.; Gu, W.; Cui, Y.; Zhong, J.; Jin, J.; He, Q.-Y.; Wang, T.; Zhang, G. Genome-Wide and Experimental Resolution of Relative Translation Elongation Speed at Individual Gene Level in Human Cells. *PLoS Genet.* **2016**, *12*, e1005901. [CrossRef]
- 95. Aldridge, S.; Teichmann, S.A. Single Cell Transcriptomics Comes of Age. Nat. Commun. 2020, 11, 4307. [CrossRef]
- 96. Rodriguez, J.; Larson, D.R. Transcription in Living Cells: Molecular Mechanisms of Bursting. *Annu. Rev. Biochem.* 2020, *89*, 189–212. [CrossRef] [PubMed]
- 97. Gupta, A.; Martin-Rufino, J.D.; Jones, T.R.; Subramanian, V.; Qiu, X.; Grody, E.I.; Bloemendal, A.; Weng, C.; Niu, S.-Y.; Min, K.H.; et al. Inferring Gene Regulation from Stochastic Transcriptional Variation across Single Cells at Steady State. *Proc. Natl. Acad. Sci. USA* 2022, 119, e2207392119. [CrossRef]
- Zheng, Y.; Zhong, Y.; Hu, J.; Shang, X. SCC: An Accurate Imputation Method for scRNA-Seq Dropouts Based on a Mixture Model. BMC Bioinform. 2021, 22, 5. [CrossRef]
- Arzalluz-Luque, Á.; Conesa, A. Single-Cell RNAseq for the Study of Isoforms-How Is That Possible? *Genome Biol.* 2018, 19, 110. [CrossRef]
- 100. Lähnemann, D.; Köster, J.; Szczurek, E.; McCarthy, D.J.; Hicks, S.C.; Robinson, M.D.; Vallejos, C.A.; Campbell, K.R.; Beerenwinkel, N.; Mahfouz, A.; et al. Eleven Grand Challenges in Single-Cell Data Science. *Genome Biol.* 2020, 21, 31. [CrossRef]
- 101. Tabula Muris Consortium; Overall Coordination; Logistical Coordination; Organ Collection and Processing; Library Preparation and Sequencing; Computational Data Analysis; Cell Type Annotation; Writing Group; Supplemental Text Writing Group; Principal Investigators. Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris. *Nature* 2018, 562, 367–372. [CrossRef]
- 102. Park, J.-E.; Botting, R.A.; Domínguez Conde, C.; Popescu, D.-M.; Lavaert, M.; Kunz, D.J.; Goh, I.; Stephenson, E.; Ragazzini, R.; Tuck, E.; et al. A Cell Atlas of Human Thymic Development Defines T Cell Repertoire Formation. *Science* 2020, 367, eaay3224. [CrossRef]
- 103. Wu, F.; Fan, J.; He, Y.; Xiong, A.; Yu, J.; Li, Y.; Zhang, Y.; Zhao, W.; Zhou, F.; Li, W.; et al. Single-Cell Profiling of Tumor Heterogeneity and the Microenvironment in Advanced Non-Small Cell Lung Cancer. *Nat. Commun.* 2021, 12, 2540. [CrossRef] [PubMed]
- 104. Tian, Y.; Carpp, L.N.; Miller, H.E.R.; Zager, M.; Newell, E.W.; Gottardo, R. Single-Cell Immunology of SARS-CoV-2 Infection. Nat. Biotechnol. 2022, 40, 30–41. [CrossRef]
- Petrosius, V.; Schoof, E.M. Recent Advances in the Field of Single-Cell Proteomics. *Transl. Oncol.* 2023, 27, 101556. [CrossRef]
 [PubMed]
- 106. Singh, A. Towards Resolving Proteomes in Single Cells. Nat. Methods 2021, 18, 856. [CrossRef] [PubMed]
- 107. Specht, H.; Emmott, E.; Petelski, A.A.; Huffman, R.G.; Perlman, D.H.; Serra, M.; Kharchenko, P.; Koller, A.; Slavov, N. Single-Cell Proteomic and Transcriptomic Analysis of Macrophage Heterogeneity Using SCoPE2. *Genome Biol.* **2021**, *22*, 50. [CrossRef]
- Brunner, A.-D.; Thielert, M.; Vasilopoulou, C.; Ammar, C.; Coscia, F.; Mund, A.; Hoerning, O.B.; Bache, N.; Apalategui, A.; Lubeck, M.; et al. Ultra-High Sensitivity Mass Spectrometry Quantifies Single-Cell Proteome Changes upon Perturbation. *Mol. Syst. Biol.* 2022, 18, e10798. [CrossRef]
- 109. Chai, J.; Song, Q. Multiple-Protein Detections of Single-Cells Reveal Cell-Cell Heterogeneity in Human Cells. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 30–38. [CrossRef]

- 110. Emmert-Buck, M.R.; Bonner, R.F.; Smith, P.D.; Chuaqui, R.F.; Zhuang, Z.; Goldstein, S.R.; Weiss, R.A.; Liotta, L.A. Laser Capture Microdissection. *Science* 1996, 274, 998–1001. [CrossRef]
- 111. Hu, P.; Zhang, W.; Xin, H.; Deng, G. Single Cell Isolation and Analysis. Front. Cell Dev. Biol. 2016, 4, 116. [CrossRef]
- 112. Yin, H.; Marshall, D. Microfluidics for Single Cell Analysis. *Curr. Opin. Biotechnol.* **2012**, 23, 110–119. [CrossRef]
- 113. Quantifying, E. Coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells | Science. Available online: https://www.science.org/doi/10.1126/science.1188308 (accessed on 6 October 2023).
- Adan, A.; Alizada, G.; Kiraz, Y.; Baran, Y.; Nalbant, A. Flow Cytometry: Basic Principles and Applications. *Crit. Rev. Biotechnol.* 2017, 37, 163–176. [CrossRef] [PubMed]
- Greenwood, C.; Ruff, D.; Kirvell, S.; Johnson, G.; Dhillon, H.S.; Bustin, S.A. Proximity Assays for Sensitive Quantification of Proteins. *Biomol. Detect. Quantif.* 2015, 4, 10–16. [CrossRef] [PubMed]
- 116. Balashova, E.E.; Lokhov, P.G.; Ponomarenko, E.A.; Markin, S.S.; Lisitsa, A.V.; Archakov, A.I. Metabolomic Diagnostics and Human Digital Image. *Pers. Med.* **2019**, *16*, 133–144. [CrossRef] [PubMed]
- 117. Tsepilov, Y.A.; Sharapov, S.Z.; Zaytseva, O.O.; Krumsiek, J.; Prehn, C.; Adamski, J.; Kastenmüller, G.; Wang-Sattler, R.; Strauch, K.; Gieger, C.; et al. A Network-Based Conditional Genetic Association Analysis of the Human Metabolome. *GigaScience* 2018, 7, giy137. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.