communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-09156-8

Integrated analysis of post-transcriptional regulations reveals insights into acute myeloid leukemia

Elissa Khadra[1,2,15], Zoaila Iqbal[1,2,15], Morgane Le Gall[1,2], Guilhem Clary[1,2], Francoise Levavasseur[1,3,4], Zubaidan Tuerdi[1,3,4], Charles Dussiau[5,6,7], David Rombaut[1,3,4], Marie Temple[1,3,4,8], Chloe Friedrich[1,3,4,8], Romain Vazquez[1,3,4,8], Justine Decroocq[9], Clément Larrue[10,11,12], Didier Bouscary[1,3,4,9], Olivier Kosmider[1,3,4,8], Nicolas Chapuis[1,3,4,8], Yannick Simoni[1,3,4], Jerome Tamburini[10,13], Evelyne Lauret[1,3,4,14], Michaela Fontenay[1,3,4,8,14], Rudy Birsen[1,3,4,9], Emilie-Fleur Gautier[1,2,14] & Ismael Boussaid[1,2,3,4,8,14] ✉

Post-transcriptional regulations play crucial roles in gene expression, but their global impact in acute myeloid leukemia (AML) remains poorly understood. Here, we perform an integrative analysis of transcriptomic and proteomic data from 44 AML samples to uncover the determinants of protein-to-mRNA-ratios (PTRs), a metric summarizing post-transcriptional regulatory processes. Using over a thousand proxies related to mRNA sequestration, translation efficiency, and RNA/protein stability, multivariate regression reveals key features driving PTRs. Our findings show that PTRs are conserved between AML samples and human tissues, underscoring the universality of underlying mechanisms. Furthermore, we show that the shadow proteome, consisting of genes frequently undetectable by mass spectrometry, can be partially attributed to low predicted PTRs.Due to the lack of existing tools, these post-transcriptional mechanisms remained in the blind spot of proteogenomic analysis. To bridge this gap, we developed **POSTCODE**, a tool for annotating omics datasets with PTR-related proxies, enabling the detection of post-transcriptional regulation changes. Applying this tool to AML datasets enables a better understanding of the mechanisms underlying differential expression, including alterations in proteostasis, translation, RNA stability or transcript localization. Our study provides a comprehensive resource for monitoring post-transcriptional regulation, paving the way for the identification of new precision medicine strategies.

In recent decades, extensive transcriptomic studies have been conducted to unravel deregulated transcriptional programs in acute myeloid leukemia (AML)[1,2]. More recently, post-transcriptional mechanisms of gene expression regulation involving translation[3], RNA decay[4], or mRNA sequestration within stress granules (SGs)[5] and processing bodies (PBs)[6,7] have been deciphered. Additionally, epigenetic writers and erasers of m^1A, m^6A, m^5C, Nm, ac^4C and pseudo-uridylation (ψ) on mRNAs and rRNAs have been implicated in the modulation of translation, sequestration, and decay[8–10]. These multi-layer regulatory processes are mediated by diverse RNA-binding proteins (RBPs), which expression is often deregulated during leukemogenesis[11,12].

The extent to which post-transcriptional regulatory mechanisms affect gene expression in AML remains difficult to contemplate. A comprehensive and integrative approach could provide an overview of this multi-layer process and improve our understanding of oncogenic pathways. A major milestone towards this goal is to explain the discrepancies observed between transcriptional and protein expression levels in AML as in various tissues[13–16]. One approach to globally quantify these regulations is to calculate protein-to-mRNA ratios (PTRs)[17–21]. This metric is essential for understanding the elements that predict post-transcriptional amplification or repression[21]. While calculated in various healthy human tissues[19], PTRs have not yet been established in AML.

To get a global assessment of post-transcriptional regulations, we integrated transcriptomic and proteomic data from 44 AML samples and computed PTRs. We found that PTRs summarizing discrepancies between RNA and protein levels were largely consistent with those in other tissues. To perform a comprehensive analysis, we compiled over a thousand

A full list of affiliations appears at the end of the paper. ✉e-mail: ismael.boussaid@aphp.fr

parameters, termed "proxies," including intrinsic proxies from transcript sequences at untranslated and coding sequence regions (5′UTR, CDS, or 3′ UTR) and from protein sequences. In addition, extrinsic proxies from literature resources measuring post-transcriptional fates such as translation, decay, and sequestration for each gene were incorporated. We also conducted a multivariate regression analysis to generate a proxy-based model that predicts PTRs. Finally, we developed a tool named POSTCODE for annotating and re-analyzing proteomic, ribosome profiling, and transcriptomic datasets based on these proxies. This tool highlights post-transcriptional regulatory changes embedded in differential expression datasets. Indeed, due to the lack of existing tools, these post-transcriptional mechanisms have remained in the blind spot of omics data analysis for a very long time. To fill this gap, we applied this methodology to various proteomics, ribosome profiling, and transcriptomics datasets from AML patients. We found that POSTCODE helped to reveal post-transcriptional regulations and avoid bias in omics data analyses interpretation.

Results

PTRs are highly conserved across 44 AML samples and 30 different human tissues

To assess the global level of post-transcriptional regulation in acute myeloid leukemia (AML), we analyzed transcriptomic and proteomic data from blasts of 44 AML patients obtained from the TCGA database and CD34$^+$ from 3 healthy donors (HD)[22] (Fig. 1a). Analysis of RNA or protein expression taken separately showed high correlations between samples (Supplementary Fig. S1a, b). However, the transcriptomes and proteomes were poorly correlated, indicating that substantial post-transcriptional regulations are at work in AMLs (Fig. 1b and Supplementary Fig. 1c). To quantify these regulations, we calculated the PTRs by dividing the absolute quantity of proteins expressed in copy number per cell using Label-Free Quantification (LFQ) and "proteomic ruler", by the RNA expression expressed in transcripts per million (TPM) for each couple of transcript and protein in 44 AML samples[22,23] and in CD34+ cells from 3 HD (Supplementary Data 2). PTRs were computed for 6669 genes, with values ranging from 0.2159 to 2.4667×10^{14} across the 44 samples, and averaging from 1.8612 to 1.8971×10^{13} (Fig. 1c). The highest PTRs were observed in housekeeping genes, whereas the lowest were found in mitochondrial genes. PTRs of recurrently mutated genes in AML were widely distributed (Fig. 1c). We hypothesized that PTRs reflect multiple levels of regulation, including transcription sequestration, translation, degradation, and protein stability (Fig. 1d). While Spearman's Rho values revealed weak correlations between RNA and protein across AML and HD samples (mean Rho ± SD: 0.3088 ± 0.0410; min-max Rho: 0.2225–0.4081) (Fig. 1e and Supplementary Fig. 1c), PTRs were highly correlated (mean Rho ± SD: 0.7794 ± 0.0722; min-max Rho: 0.4355-0.9235) (Fig. 1e and Supplementary Fig. 1d). PTRs were also highly correlated across French-American-British (FAB) subtypes (mean Rho ± SD: 0.8835 ± 0.0323; min-max Rho: 0.8278–0.9305) and European Leukemia Net (ELN) classification subgroups (mean Rho ± SD: 0.9289 ± 0.0047; min-max Rho: 0.9227–0.9321) (Fig. 1f), indicating a strong conservation of post-transcriptional mechanisms across AML subtypes and HD CD34+ cells. Additionally, we compared PTRs from AML and HD CD34+ to 29 different human tissues issued from the Human Proteome Atlas (HPA)[19]. Despite variations between techniques of RNA sequencing and mass spectrometry techniques and between automated systems, PTRs were very well-conserved between samples (mean Rho ± SD: 0.7520 ± 0.0726; min-max Rho: 0.4719–0.8603) (Fig. 1g).

Taken together, these data confirm the relatively weak correlation between RNA and protein expression levels as previously shown in other studies[13–16]. Our data suggest that the discrepancies between RNA and protein levels, as quantified by PTRs, are highly conserved across different tissues, implying that they are driven by broadly conserved post-transcriptional mechanisms.

5′UTR proxies influencing translation initiation affect PTRs

To investigate the post-transcriptional mechanisms responsible for the conservation of PTRs, we exhaustively detailed the quantifiable characteristics of mRNAs and proteins, further referred to as "proxies" including 5′UTR, CDS, and 3′UTR regions of the mRNA, and the peptide sequences themselves.

Based on an extensive literature review, we first identified several elements within the 5′UTR region that could contribute to variations in PTRs. We investigated how factors such as length, structure, cap dependency, the presence of internal ribosome entry sites (IRES), G quadruplex (rG4), upstream AUGs (uAUG), upstream open reading frames (uORF), other various motifs, and the Kozak consensus sequence were associated with PTR values (Fig. 2a). Our analysis revealed that genes with lower PTR values were associated with longer 5′UTRs (Fig. 2b). Using RNA folding energy (ΔG) of the 5′UTR as a proxy for RNA secondary structure, we observed that genes with lower PTRs are associated with higher global ΔG of their 5′UTR (Fig. 2c). Since the RNA ΔG is directly proportional to the length of the tested sequence (Fig. 2d), and to avoid this bias in quantifying structural impact, we normalized the global ΔG by the length of the 5′UTR sequence and found that lower PTRs remained associated with more structured 5′ UTRs (Fig. 2e). To assess whether this structural effect varies across different regions of the 5′UTR, we analyzed 100-nucleotide (nt) windows (Supplementary Fig. 2a). Our results indicate that structural differences between low and high PTRs were more pronounced at the 5′ ends and middle regions (Supplementary Fig. 2b–d).

RNA folding is dependent on the percentage of GC in a given sequence[24,25]. We therefore found that GC content in the 5′UTR was also associated with PTRs in a similar manner (Supplementary Fig. 2e). Secondary structures in the 5′UTR and GC content have been shown to reduce initiation efficiency due to the need for unwinding during 40S scanning[26]. To test whether this phenomenon could be observed by PTRs measurement, we focused on two key helicases, namely EIF4A1 and DDX3X, which facilitate 5′UTR unwinding[27–32]. Therefore, we used diverse ribosome profiling datasets from cells in which these two helicases were functionally impaired to define their selective targets. Thus, we have shown that irrespective of the use of therapeutic inhibitors[27,32,33] (Hippuristanol and Silvestrol), siRNA[29,31], shRNA[28], or auxin degron constructs[30], EIF4A1 and DDX3X's targets have structured 5′UTRs and corresponding low PTRs (Fig. 2f, g).

Beyond the overall structure, we also explored specific elements within the 5′UTR (Supplementary Data 3). A fast Gene Set Enrichment Analysis (fGSEA) performed on ranked PTRs revealed that transcripts containing an rG4 motif within the 5′UTR[34] were generally associated with low PTRs (NES = −1.65; FDR = 3.11×10^{14}) (Fig. 2h). In contrast, transcripts with IRES sequences did not demonstrate a significant enrichment towards high or low PTRs, which might be attributed to the heterogeneity of IRES-mediated translation initiation[35]. Among the various 5'UTR motifs tested, many motifs reported to associate with low PTRs in many tissues[19] were confirmed here (e.g., GGGGGGG, UUUGUUU, AUACAUU), while motifs like ACACACA and CUGUCCU were associated with high PTRs (Fig. 2h). Our analysis also identified that uORFs and uAUGs were strongly associated with low PTRs (NES = −1.71; FDR = 2.91×10^{-17} and NES = −1.43; FDR = 3.51×10^{-11}, respectively). We noted that the presence of an exon/exon junction complex in the 5′UTR sequence (EJC/EEJ$^{5'UTR}$) also appeared deleterious (NES = −1.35; FDR = 9.12×10^{-11}). Nevertheless, it should be noted that a large number of these genes also have a uORF[36]. Thus, when only genes with EJC/EEJ$^{5'UTR}$ and without associated uORF were considered, statistical differences were no longer observed (NES = −0.79; FDR = 0.99) (Fig. 2h). Conversely, elements such as the oligopyrimidine tract (5′TOP) and transcripts described as cap-independent[37–39] were significantly associated with high PTRs (Fig. 2h).

Since the scanning by the small 40S subunit ends with recognition of the Kozak sequence and dwell time on the start codon[40], we hypothesized that divergence with the consensus Kozak sequence could be reflected in PTRs measurements. While we found that PTRs remain invariant regardless of the number of mismatches in the Kozak sequence compared to the consensus sequence, we also found that specific nucleotide enrichment, like C or U in position −3 is present in low PTR genes (Fig. 2i, j and Supplementary Fig. 2f).

In addition to intrinsic sequence characteristics (length, structure, motifs), we incorporated high-throughput functional data to assess the impact of translation initiation rate on PTRs. We used data from a massively parallel translation assay that determined a proxy for initiation rate called "Mean Ribosome Load" (MRL). This was assessed by inserting a 50-nt preSTART 5′UTR sequence into plasmid constructs for 20,530 human

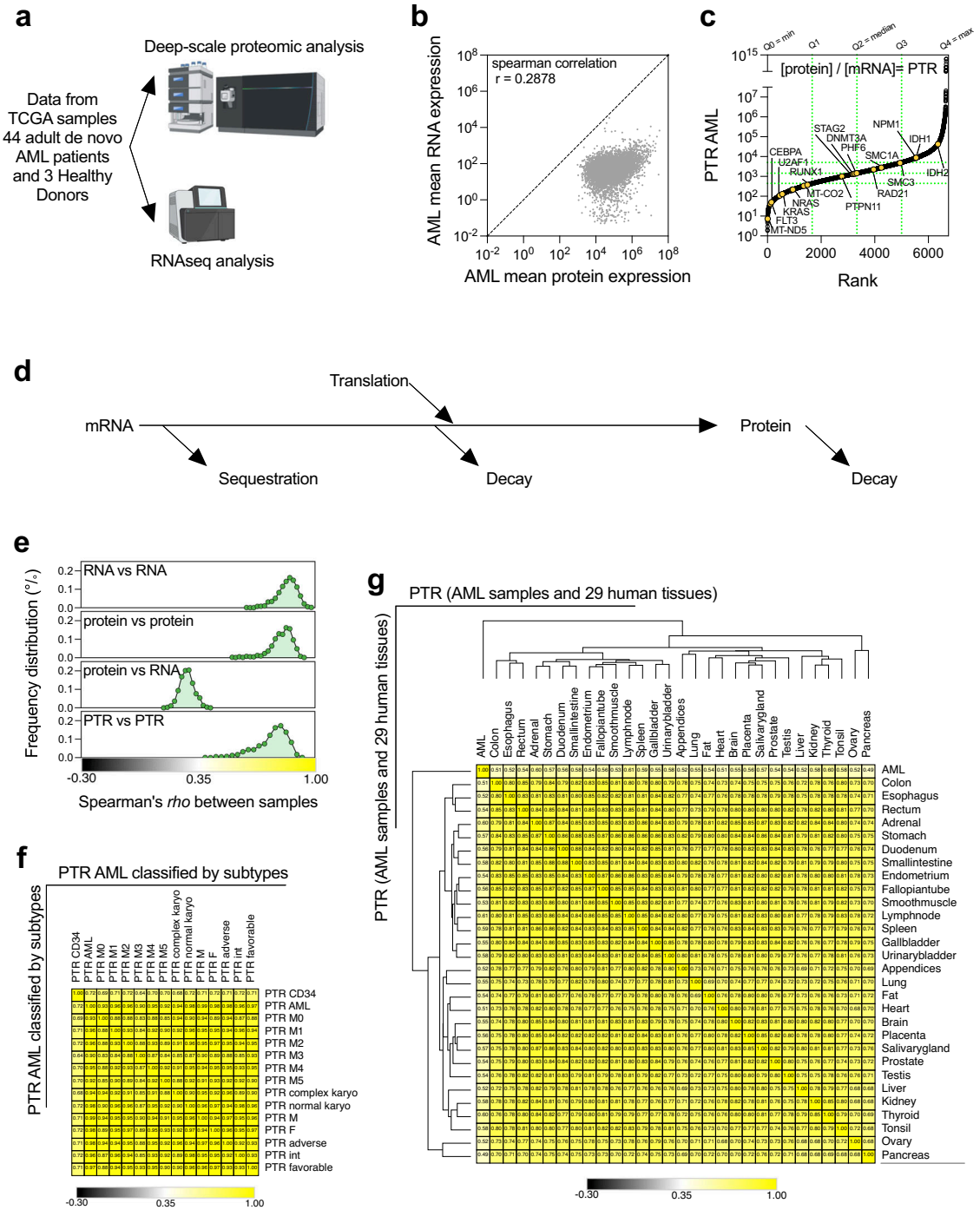


Fig. 1 | Protein to RNA ratios (PTRs) are highly conserved in 44 AML samples and 30 other human tissue samples. a–**c** Schematic workflow used to calculate PTRs from AML transcriptomic and proteomic data. **a** 44 samples from LAML TCGA dataset with transcriptomic and proteomic data were used. Images were created with BioRender.com. **b** Bi-parametric plot of transcriptomic and proteomic mean expression of 44 AML patients at times of diagnosis. RNA was quantified using transcript per million from RNA sequencing. Protein expression was measured using Protein copy numbers per cell via LFQ tandem mass spectrometry and Protein ruler plugin of Perseus by standardization to the total histone MS signal. Spearman correlation coefficient rho is annotated. **c** Ranked PTRs (AML) plot of genes with both proteomic and transcriptomic available quantification in at least one sample. PTR^{min} = Q0; PTR interquartile1 = Q1; Median PTR = Q2; PTR interquartile3 = Q3, and PTR^{max} = Q4 are indicated by horizontal green dot lines. **d** Schematic representation of post-transcriptional regulations influencing PTRs. **e** Frequency distribution curves of Spearman correlations calculated between sample for RNA expression; Protein expression; RNA vs Protein expression; PTRs. **f** Heat map of PTR correlations between samples from different FAB AML subtypes (French–American–British) or ELN 2017 cytogenetic AML subgroups. **g** Heatmap of unsupervised hierarchical clustering of PTR correlations (rows and columns) in AML patient samples and 29 other human tissues. Heatmap is clustered using complete-linkage hierarchical clustering based on Euclidean distances. Colour scale is shared by both heatmaps.

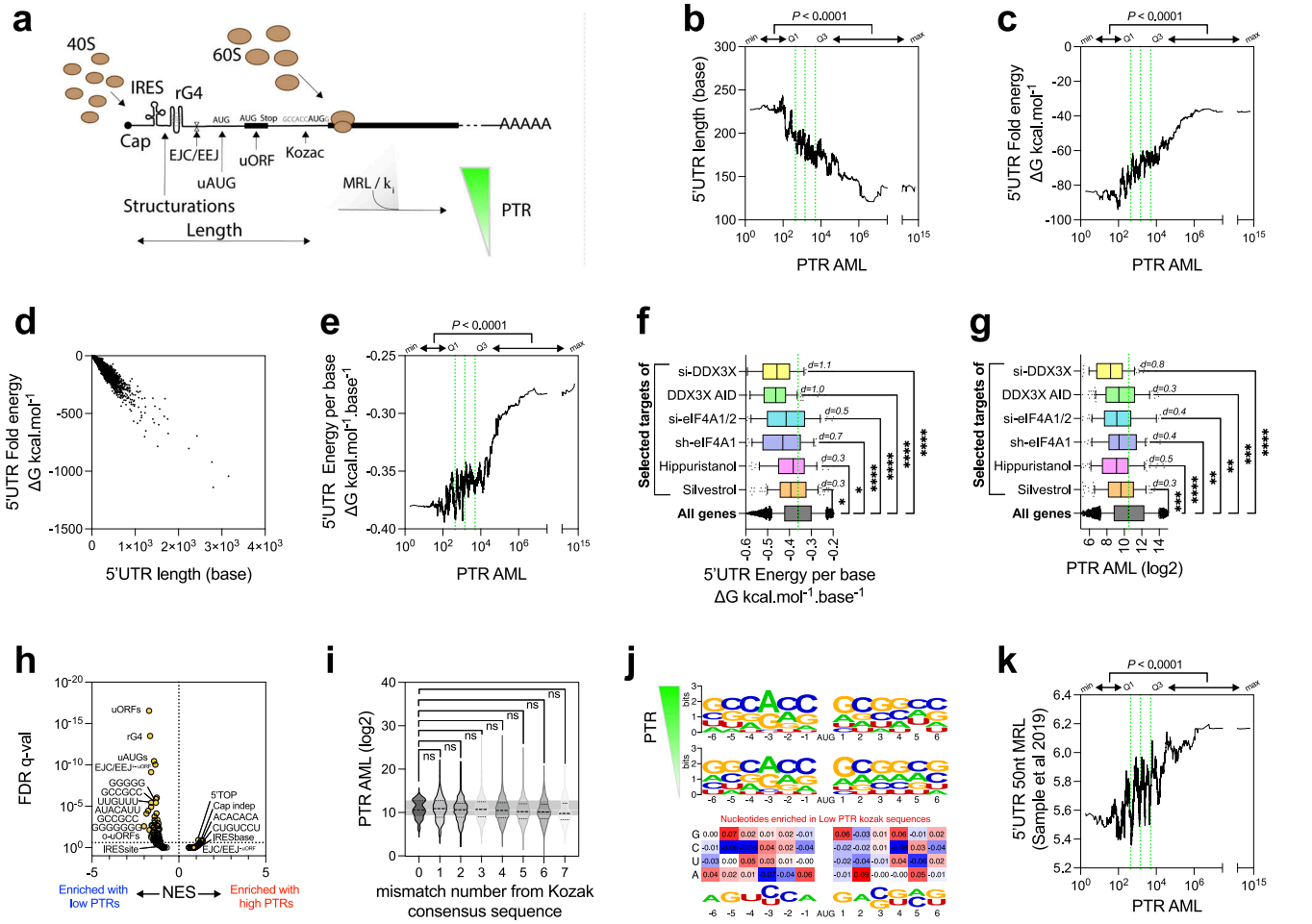


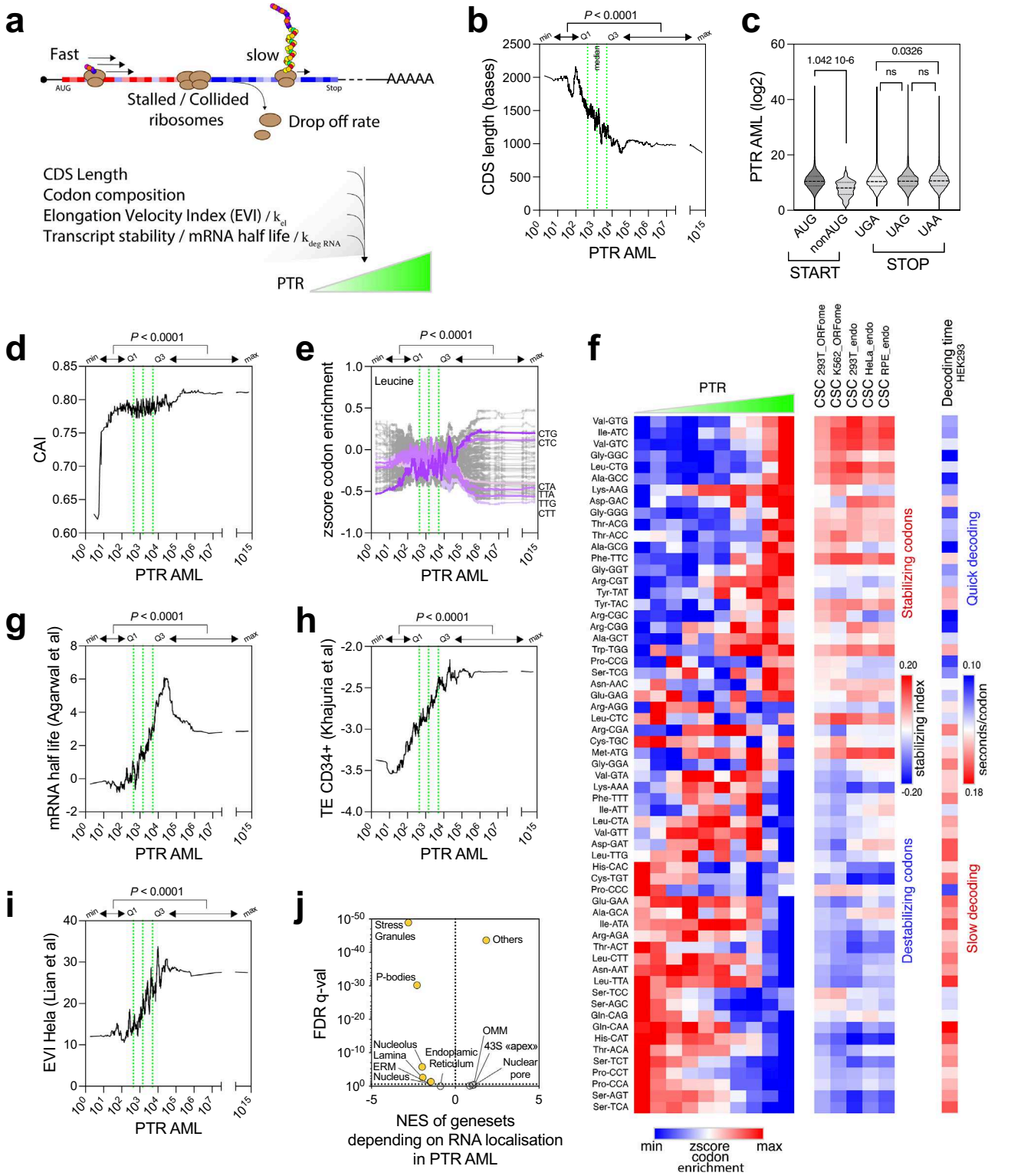
Fig. 2 | The 5′UTR proxies regulating translation initiation influence PTR. **a** Schematic overview of the main proxies identified in the 5′UTR. **b**, **c**, **e**, **k** Curve representing the median value of the proxy (y) computed for each PTR rank (x_n). A sliding window extending from rank $n - 100$ to rank $n + 100$ is used to compute the median. Mann–Whitney tests between low PTRs (min to interquartile Q1) vs high PTRs (interquartile Q3 to max) were performed with row values of each proxy (**b**). Curve representing the median value of the 5′UTR length calculated for each PTR rank (x_n). **c** Curve representing the median value of the 5′UTR folding energy calculated for each PTR rank (x_n). **d** Bi-parametric plot showing the correlation between length and global folding energy of a sequence. **e** Curve representing the median value of the 5′UTR folding energy normalized by the 5′UTR sequence length in ΔG kcal mol^{-1} $base^{-1}$ calculated for each PTR rank (x_n). **f**–**g** Box plots of 5′UTR energy per base values and PTR values of the selective targets of eIF4A and DDX3X identified from multiple studies using ribosome profiling of cells treated with inhibitors of eIF4A and DDX3X or KD strategy with shRNA, siRNA, or auxin-inducible degron strategy. Ordinary one-way ANOVA multiple comparisons between selective targets (si-DDX3X $n = 93$; AID DDX3X $n = 158$; sh-eIF4A1 $n = 82$; si eIF4A1/2 $n = 98$; silvestrol $n = 149$; hippuristanol $n = 103$) compared to all genes $n = 6669$. Violin center line indicate median, the lines above and below indicate interquartiles Q3 and Q1. Absolute Cohen's d values are indicated. **h** Volcano plot representing fGSEA analysis performed on the ranked PTR and showing the enrichment of genes carrying discrete proxy motifs or structural elements such as rG4, IRES, 5′TOP, uAUG, uORF, EJC …. among genes with a low (left) or high (right) PTRs. Gene sets with a significant enrichment (FDR > 0.05) are represented in yellow dots. **i** Violin plots of PTR Log2 values of genes classified in function of the number of mismatches their flanking the start codon sequence presents compared to the consensus kozak sequence. Ordinary one-way ANOVA multiple comparisons of the mean of each group with the mean of no mismatch group. Violin center line indicate median, the lines above and below indicate interquartiles Q3 and Q1. The grey shaded area in the background represents the interquartile range (Q1-Q3) of the 0-mismatch group. **j** WebLogo of nucleotide sequences −6 to +6 flanking the start codon: (on the top) WebLogo realized with sequences of the 200 higher PTR genes and underneath WebLogo realized with sequences of the top 200 lower PTR genes. Heatmap showing position-specific nucleotide enrichment within Kozak sequences when comparing low-PTR to high-PTR genes. For each position, nucleotides preferentially enriched in low-PTR genes are indicated at the bottom. **k** Curve representing the median value of the Mean Ribosome Load (MRL) (Sample et al.[41]) calculated for each PTR rank (x_n).

transcripts corresponding to 12,503 genes[41]. With these data we showed that PTRs reflect variations in translation initiation rates (Fig. 2k).

Overall, these results indicate that PTRs effectively capture the impact of 5′UTR on translation initiation and are expected to vary under conditions that alter initiation rates.

PTRs proxies in Coding Sequence (CDS) reveal the interplay between translation efficiency and transcript stability

To follow the ribosome's journey, we investigated various determinants along the CDS that could contribute to PTRs variations (Fig. 3a). As observed in the 5′UTR region, we found that CDS length was a key factor with longer CDSs associated with lower PTRs (Fig. 3b). This inverse relationship between CDS length and production efficiency has been well-documented in relation to translational efficiency (TE) measured by ribosome profiling experiments[42–48]. The ribosome "drop-off" phenomenon has been suggested as a contributing factor, but the rate of ribosome detachment, estimated at less than 1 detachment per 10,000 codons, is too low to fully account for the differences in TE and therefore in PTRs[49–51] (Supplementary Fig. 3a). Likewise, ribosome frameshifting, which occurs approximately once every 42,000 codons[52], is unlikely to fully account for the observed PTR/CDS length anti-correlation (Supplementary Fig. 3a).



We also found that the presence of non-AUG start codons, which initiate translation less efficiently than AUG codons[53–55], was associated with particularly low PTRs (Fig. 3c). We quantified RNA structuration immediately downstream of the start codon to identify a possible mechanism that could promote recognition of the start codon and initiation by increasing the dwell time of the small subunit[40,56]. Interestingly, contrary to the relationship observed in the 5′UTR (Fig. 2c–e), highly structured regions downstream of the start codon were associated with higher PTR values, though the differences were modest (Supplementary Fig. 3b). This is in sharp contrast with highly structured 5′UTR being associated with a low PTR. We confirmed this pattern by calculating the differential ΔG between 5′UTR and the "Start-CDS-region" (Supplementary Fig. 3c).

We then assessed whether the Stop codon context could impact PTRs and found a significant but weak difference between the UGA and UAA codons (Fig. 3c). Although the differential ΔG between "CDS-Stop-region" and 3′UTR was non-significantly associated with PTRs variations

Fig. 3 | PTRs proxies in coding sequence CDS reveal the interplay between translation efficiency and transcript stability. a Schematic overview of the main proxies identified in the CDS. **b**, **d**–**i** Curve representing the median value of the proxy (y) computed for each PTR rank (x_n). A sliding window extending from rank $n-100$ to rank $n+100$ is used to compute the median. Mann–Whitney tests between low PTRs (min to interquartile Q1) vs high PTRs (interquartile Q3 to max) were performed with row values of each proxy. **b** Curve representing the median value of the CDS length calculated for each PTR rank (x_n). (**c** (*Left*)) Violin plots of PTR Log2 values of genes using a consensus start codon ($n = 6646$) or a non-AUG start codon($n = 24$) (left) Mann–Whitney test. (**c** (*Right*)) Violin plots of PTR Log2 values of genes using respectively UGA ($n = 3109$), UAG ($n = 1385$), or UAA ($n = 1960$) stop codon. One-way ANOVA multiple comparisons Tukey's comparison test of the mean of each type of STOP codon with the mean of other types. Violin center line indicate median, the lines above and below indicate interquartiles Q3 and Q1. **d** Curve representing the median value of CAI calculated for each PTR rank (x_n). **e** Curve representing the median value of the codon enrichment (z-score) calculated for each PTR rank (x_n). In violet are represented the 6 different codons coding leucine. In grey (in the background) are represented all the other codons. **f** (*Left*) Heatmap representing median values of the codon enrichment (z-score) as a function of ranked PTRs subdivided in 10 groups. Each row represents results for a different codon. Each column represents $n = 666$ genes classed depending on ranked PTRs. **f** (*Right*) heatmaps representing two different published metrics: codon stabilization coefficient (CSC) determined in from different models and "codon decoding time". **g** Curve representing the median value of mRNA half-life calculated for each PTR rank (x_n). **h** Curve representing the median value of Translation Efficiency (TE) calculated for each PTR rank (x_n). **i** Curve representing the median value of Elongation Velocity Index (EVI) calculated for each PTR rank (x_n). **j** Volcano plot representing fGSEA analysis performed on ranked PTRs and showing the enrichment of genes according to the cellular localization of their RNA among genes with a low (left) or high (right) PTRs. Gene sets with a significant enrichment (FDR > 0.05) are represented in yellow dots.

(Supplementary Fig. 3d, e). More generally, the impact of the stop codon and its environment was weak compared to that of the Start codon. We then showed that high-PTR genes had a more structured CDS than low-PTR ones (Supplementary Fig. 3f). This level of RNA structuration depends on the GC% of the sequence[57] (Supplementary Fig. 3g), and CDS GC% is further known to be linked to codon bias[58]. We therefore studied the effect of codon composition on PTR. For each sequence, we first determined the Codon Adaptation Index (CAI), a metric used to estimate the average codon bias of the transcripts by giving a higher score to transcripts composed of many optimal codons and, conversely, a low score to the transcript mainly composed of non-optimal codons[59]. We show that the extreme PTRs had different CAIs, with very low PTRs having a low CAI and very high PTRs having high CAI (Fig. 3d). CAI proxy is an average metric summarizing the use of the 64 codons within the transcripts, as a single variable. To improve accuracy, we decided to represent for each amino acid how the codons encoding it were distributed among the different transcripts. We therefore calculated the frequency of use of each codon for all the transcripts and then calculated a z-score expressing the enrichment in a codon for a transcript compared with all the human transcripts. For example, among the codons encoding leucine, we showed that the CTG and CTC codons were particularly enriched in genes associated with a high PTR and conversely depleted in genes with a low PTR (Fig. 3e). Strikingly, CTG is known as the optimal in the sense of codon bias (RSCU) in H. Sapiens. Conversely, non-optimal codons CTT, TTG, TTA, and CTA were depleted from the genes associated with a high PTR (Fig. 3e). This analysis was extended to all amino acids, and a heatmap visualizing codon distribution across 10 increasing PTR bins revealed distinct sets of codons associated with high or low PTRs (Fig. 3f). Previous studies suggest that the CAI/RSCU codon bias influences translation rate, elongation, and transcript stability through mechanisms linking translation and decay[58,60–62]. To understand the differential repartition of codons with the level of PTR, we first added RSCU values for each codon to the right of the heatmap and observed a poor correspondence (Supplementary Fig. 3h). Then we added a metric developed to quantify the stabilizing potential of codons called "Codon Stabilization Coefficient" (CSC) and a metric quantifying the decoding time of each codon[61,63–65], which indicated a clear correspondence between codons enriched in high PTR genes and both stabilizing potential and short decoding time (Fig. 3f and Supplementary Fig. 3i). However, other metrics such as tRNA Adaptation Index (tAI), tRNA genomic copy number as well as tRNA expression[66,67], did not show strong correspondence with PTRs, suggesting that CSC and decoding time are more relevant to PTRs variation (Supplementary Fig. 3h). This codon-based binary distribution led us to evaluate the correlation between codon z-scores. Hierarchical clustering revealed a binary pattern, grouping codons associated with transcript stability and fast decoding versus those linked to instability and slow decoding (Supplementary Fig. 3j). Consequently, we tested whether PTRs and the half-life of the transcripts[68] were two associated variables and we showed a very clear covariation up to a value of PTR of 10^4 and then a weaker correlation up to 10^{15} (Fig. 3g). As mentioned above, since translation and decay are two interrelated processes, we then matched PTRs with data measuring TE in primary CD34+ hematopoietic cells[47] and with data extrapolating the ribosome elongation speed, named Elongation Velocity Index (EVI), the translational rate (TR) and the Density (D = EVI/TR) from ribosome profiling experiments[69]. For all these variables, we showed a very clear covariation, but this time over the entire PTRs range (from 1 to 10^{15}) (Fig. 3h, i and Supplementary Fig. 3k, l).

Finally, we explored the impact of RNA localization on PTRs, using data primarily derived from APEX-seq measurements[70–73] (Supplementary Data 3). Transcripts known to be sequestered in PBs or SGs structures were associated with a very low PTR (Fig. 3j and Supplementary Fig. 3m), as were transcripts localized to the nucleus, nucleolus, and lamina. In contrast, transcripts localized to the outer mitochondrial membrane (OMM) or not localized to the above-mentioned areas exhibited higher PTRs (Fig. 3j).

Altogether, these results indicate that the PTR summarizes multiple post-transcriptional regulatory processes associated with CDS characteristics, including translation start, elongation speed, transcript sequestration, and transcript stability.

PTRs proxies in the 3′UTR reveal a hub for post-transcriptional repression

The 3′UTR is a critical region for post-transcriptional regulation, and we identified several proxies in this region that are likely to influence PTR variation (Fig. 4a). Indeed, in line with our findings in the 5′UTR and CDS, we observed that longer 3′UTRs are associated with lower PTR values, while shorter 3′UTRs correspond to higher PTRs (Fig. 4b). Similarly, a higher global ΔG in the 3′UTR is associated with lower PTR, although this effect was less or not significant when ΔG was normalized by sequence length (Fig. 4c–e and Supplementary Fig. 4a).

One well-established mechanism of post-transcriptional regulation at 3′UTR involves RNA-binding proteins (RBPs). Leveraging data from thousands CLIP-seq experiments across 284 RBPs[74,75], we generated specific gene sets for each RBP based on their binding site number (Supplementary Data 3). Using fGSEA analysis on ranked PTRs, we classified RBPs as either promoting (NES > 0) or repressing (NES < 0) PTR (Fig. 4f). Remarkably, RBPs with NES > 0 are all involved in translation (e.g. *RPS3*, *EIF4A1*, *EIF3B/D/G*, *EIF4G1*, *PCBP1*, *YBX3*, *NIP7*…) or ribosome biogenesis (e.g. *UTP18*, *NOLC1*, *SBDS*, *SDAD1*, *WDR43* …) (Fig. 4f and Supplementary Fig. 4b). In contrast, RBPs with NES < 0 are associated with transcript stability, degradation, sequestration, and repression of gene expression (Fig. 4f and Supplementary Fig. 4c). We then categorized RBPs to the localization of their binding site (intron, 5′UTR, CDS, 3′UTR)[75]. We found that RBPs preferentially binding the CDS or 3'UTR were predominantly linked to low PTRs, underscoring their repressive role (Fig. 4f). To validate these findings, we used RBP motifs from the ATtRACT database[76] for 132 RBPs and additional k-mer motifs identified by others[19] to generate gene sets of transcripts containing these motifs within their 3'UTRs (Supplementary

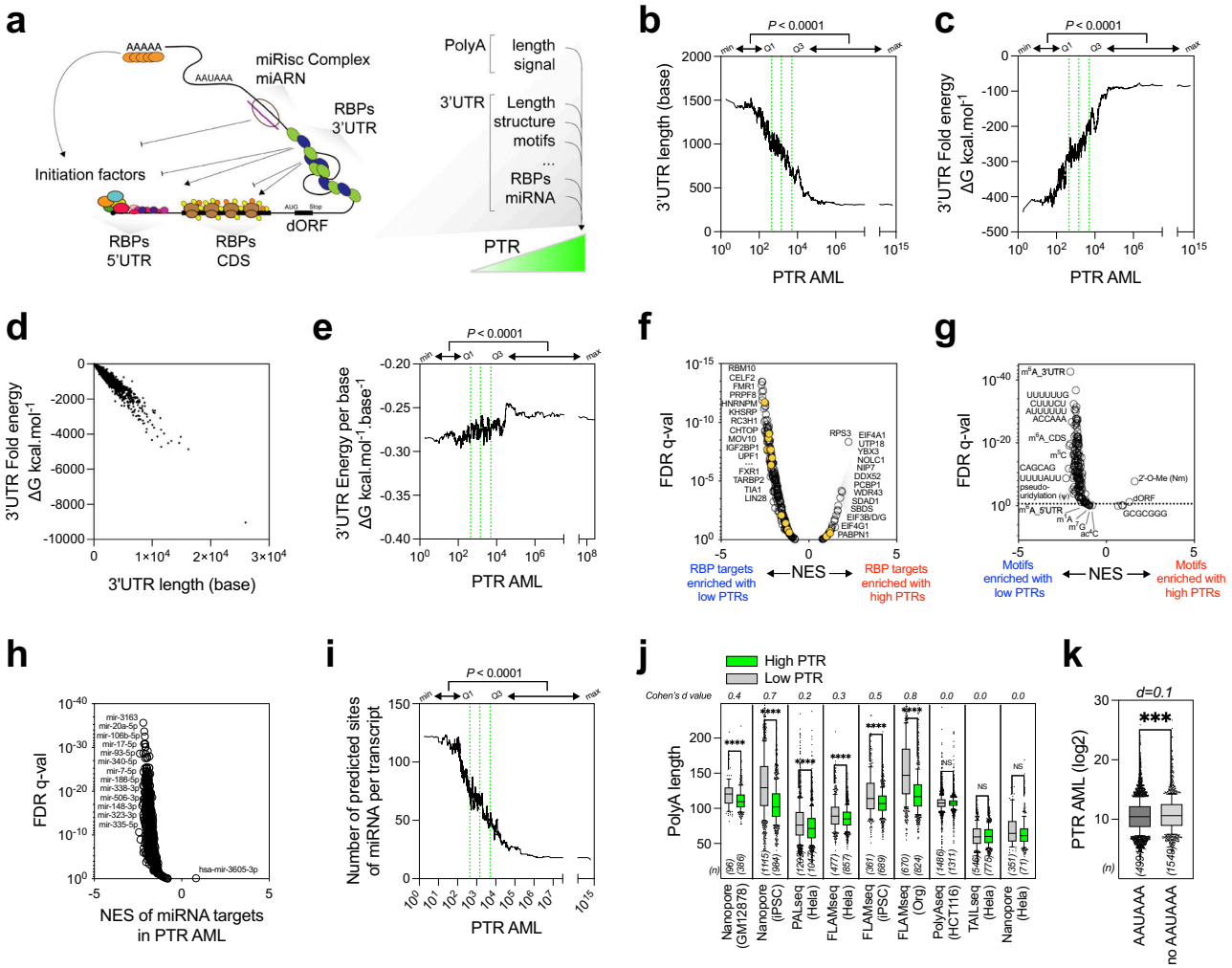


Fig. 4 | PTRs proxies in 3′UTR reveals the presence of a hub for post-transcriptional repression. **a** Schematic overview of the main proxies identified in the 3′UTR. **b**, **c**, **e** Curve representing the median value of the proxy (*y*) computed for each PTR rank (x_n). A sliding window extending from rank $n - 100$ to rank n + 100 is used to compute the median. Mann–Whitney tests between low PTRs (min to interquartile Q1) vs high PTRs (interquartile Q3 to max) were performed with row values of each proxy. **b** Curve representing the median value of the 3′UTR length calculated for each PTR rank (x_n). **c** Curve representing the median value of the 3′ UTR folding energy calculated for each PTR rank (x_n). **d** Bi-parametric plot showing the correlation between length and global folding energy of a sequence. **e** Curve representing the median value of the 3′UTR folding energy normalized by the 3′UTR sequence length in ΔG kcal mol^{-1} $base^{-1}$ calculated for each PTR rank (x_n). **f** Volcano plot representing fGSEA analysis performed on ranked PTRs and showing the enrichment of gene sets, each specific to an RBP, determined from CLIPseq experiments. Each point represents a gene set, so each point represents a different RBP. This graph shows the enrichment of RBP targets among genes with a low (left) or high (right) PTRs. RBPs preferentially binding the CDS or 3′UTR are highlighted in yellow. **g** Volcano plot representing fGSEA analysis performed on ranked PTRs and showing the enrichment of genes carrying in the 3′′UTR, discrete proxy motifs such as dORF or epigenetic marks among genes with a low (left) or high (right) PTRs. **h** Volcano plot representing the fGSEA analysis performed on ranked PTRs and showing the enrichment of gene sets, each specific to a miRNA, determined by multiple target-predicting programs (PITA, RNA22, miRmap, DIANA-microT, miRanda, PicTar, and TargetScan). Each point represents a gene set, so each point represents a different miRNA. This graph shows the enrichment of mRNA targets among genes with a low (left) or high (right) PTRs. **i** Curve representing the median value of the predicted miRNA sites per transcript calculated with the miRNA DB database for each PTR rank (x_n). **j** Box plots of polyA length values of PTR^{high} and PTR^{low} genes in multiple published datasets. **k** Box plots of PTR (log_2) values of genes with ($n = 4991$) or without ($n = 1549$) polyadenylation signal AAUAAA conserved in the 3′UTR. Mann–Whitney test.

Data 3). Consistent with our previous results, transcripts with RBP motifs in their 3′UTRs were typically associated with low PTRs, reinforcing the idea of 3′UTR-mediated repression (Fig. 4g).

Among the 284 RBPs analyzed, we observed that those functioning as epitranscriptomic "writers" or "erasers" targeted mRNA preferentially associated with low PTRs, suggesting a predominant repressive effect (NES < 0) (Supplementary Fig. 4d). To validate this observation, we examined the relationship between PTRs and a range of mRNA epigenetic marks including m^6A, m^1A, m^5C, m^7G, ψ, *Nm* and ac^4C by compiling gene sets from dedicated epitranscriptomic databases[77,78] and landmark studies specific to each modification[8,9,79–84]. Gene set enrichment analysis revealed that transcripts bearing m^6A marks within the 3′UTR or coding sequence (CDS), as well as those modified with m^5C or ψ, were significantly associated with lower PTRs. In contrast, mRNAs with m^6A marks located in the 5′UTR, or modified by m^1A, m^7G, or ac^4C, showed no consistent association with PTRs. Notably, transcripts carrying Nm modifications were enriched among high PTR genes, highlighting a potential promoting role for this modification.

We then investigated the impact of additional open reading frames in the 3′UTRs (dORFs)[85–87] and showed that they were linked to higher PTRs, possibly due to enhanced translation of their canonical ORFs[86] (Fig. 4g).

To further investigate the relationship between miRNA-mediated gene silencing and PTRs, we crossed predicted miRNA target sites with AGO protein binding data obtained from CLIP-seq studies[74,75]. This analysis

yielded 611 miRNA gene sets, each comprising predicted targets of individual miRNAs across multiple target prediction algorithms, including PITA, RNA22, miRmap, DIANA-microT, miRanda, PicTar, and TargetScan (Supplementary Data 3). Gene set enrichment analysis (fGSEA) revealed that all but one miRNA gene set exhibited a negative normalized enrichment score (NES < 0), suggesting that miRNA targets generally have low PTR values, consistent with the established repressive function of miRNAs (Fig. 4h). To extend this analysis, we utilized the miRNA DB database to compute the number of predicted miRNA binding sites per transcript. Our findings revealed a significant negative association between the number of miRNA binding sites and PTR (Fig. 4i). After adjustment for transcript length, we observed that transcripts with a higher density of miRNA binding sites per nt also displayed lower PTR values (Supplementary Fig. 4e). This result was further corroborated using conserved miRNA target site predictions from the TargetScan database (Supplementary Fig. 4f, g). Additionally, we highlighted that two metrics named "Context++" and "predicted KD", developed to quantify miRNA-mediated repression efficiency, were strong predictors of variations in PTR values (Supplementary Fig. 4h, i).

Finally, we explored how polyA tail length was associated with PTRs, utilizing data from nine studies that characterized this parameter across different models and with different methods. Our analysis revealed both significant and non-significant relationships between polyA tail length and PTRs depending on the studies (Fig. 4j and Supplementary Fig. 4j–r). Interestingly, datasets from different models showed weak correlations (Supplementary Information), suggesting that polyA tail length is not a well-conserved feature. However, we did observe that transcripts containing the AAUAAA polyadenylation signal in their 3′UTRs had slightly lower PTRs compared to those lacking this signal, consistent with previous findings[19] (Fig. 4k). However, the magnitude of this difference was small, suggesting that polyadenylation tail length exerted only a modest influence on PTR compared to other proxies.

In summary, this analysis integrates 3′UTR proxies with PTRs on an unprecedented scale, revealing that 3′UTR features frequently mediate post-transcriptional repression. Furthermore, PTRs provide a means to hierarchize the effects of different proxies, enhancing our understanding of the regulatory mechanisms governing gene expression.

The protein, its composition, and structure influence the level of PTR

Up to this point, we have explored the elements affecting pre- and per-translational events. Next, we focused on post-translational factors, particularly those that influence protein stability, which subsequently modulates protein expression levels and, thus, PTRs (Fig. 5a). We investigated how each amino acid was distributed between the different proteins. We therefore calculated the frequency of use of each amino acid for all the genes and then calculated a z-score expressing the enrichment in an amino acid for a given protein compared with the whole human proteome. We found that certain amino acids, such as valine (V), glycine (G), isoleucine (I), alanine (A), lysine (K), phenylalanine (F), methionine (M), and tyrosine (Y), were enriched in high PTR genes. In contrast, amino acids like serine (S), proline (P), histidine (H), glutamine (Q), cysteine (C), and threonine (T) were more prevalent in low PTR genes (Fig. 5b, c). Similar to codon enrichment analysis (Fig. 3f), we added a metric called the Amino Acid Stabilizing Coefficient $AASC^{prot}$, which we calculated using protein half-lives[88] (Fig. 5c). This metric estimates the stabilizing effect of different amino acids on the protein, much like CSC^{RNA} reflects the stabilizing effect of codons on RNA (Fig. 3f). Even more obviously than CSC^{RNA}, $AASC^{prot}$ may explain why certain amino acids were enriched in high PTRs and why others were enriched in low PTRs (Fig. 5c). We further explored the relationship between PTR and protein half-life, showing a strong correlation between the two variables up to a PTR value of ~10^4. The correlation was weaker for higher PTRs (Fig. 5d). This pattern mirrored the PTR-RNA half-life relationship we observed earlier (Fig. 3g). This similarity led us to compare RNA and protein half-lives revealing a strong correlation observed between these two measures[68,88] despite different orders of magnitude (Fig. 5e). Interestingly, protein half-life determinants $AASC^{prot}$ appeared to be very close to $AASC^{RNA}$, a metric estimating the stabilizing effect of different amino acids on mRNA[61,66] (Fig. 5f). This correlation is unlikely to be a result of transcriptional bias since the proteomic techniques used to measure protein half-life are independent of transcription level. An explanation could perhaps be found in evolutionary constraints[13].

We also looked at the decoding rate, but here at amino acid level. Thus, by averaging the codon decoding time for each amino acid, we found that decoding time behaved inversely to the $AASC^{prot}$ stabilizing effect (Fig. 5c).

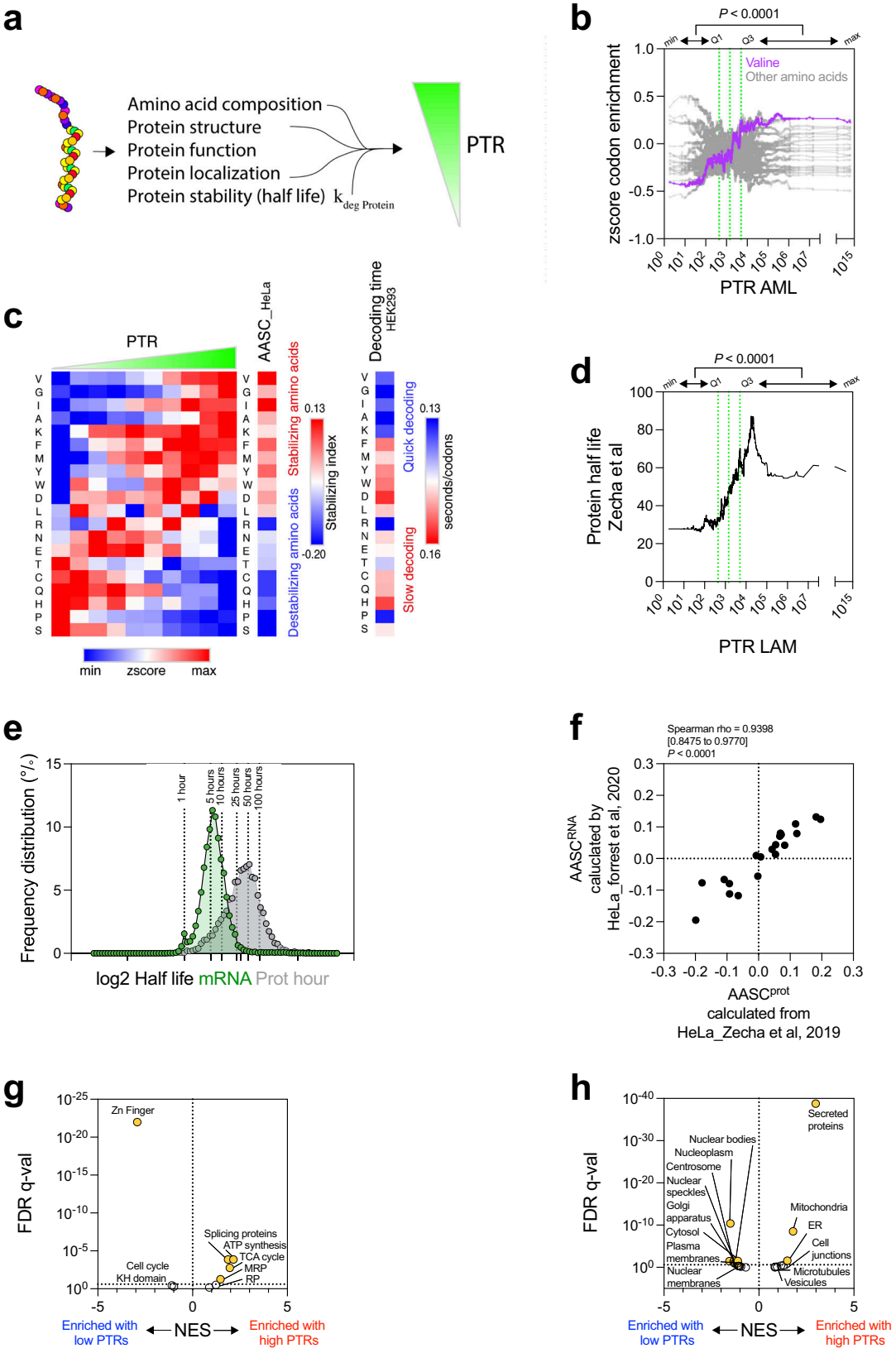
Finally, using fGSEA, we showed that certain structural elements within proteins could predict higher or lower PTRs (Supplementary Data 3). For instance, the presence of a zinc finger motif was associated with lower PTRs (Fig. 5g). Moreover, leveraging data from selected resources and HPA image collection enabling the localization of different proteins within the cell[89,90], we showed that nuclear proteins are generally associated with lower PTRs, whereas secreted proteins or proteins localized to the endoplasmic reticulum or mitochondria tend to have higher PTRs (Fig. 5h).

The integration of proxies enables PTR prediction and explains the shadow proteome

In this study, we identified proxies that reflected intrinsic features of transcript or protein sequences, and extrinsic proxies, based on functional measurements. Comparing all proxies is challenging, as some are discrete (e.g., RBP targets, miRNA targets, motifs, uORFs, dORFs) while others are continuous (e.g., RNA structure, RNA length, TE, half-lives). Among continuous proxies, the strongest associations with low PTRs were observed with 5′UTR, CDS, and 3′UTR length, transcript sequestration in SGs or PBs, and the relative abundance of amino acids like serine or proline (Fig. 6a). In contrast, proxies most strongly associated with high PTRs included RNA and protein half-lives, TE, TR, EVI, and the relative quantity of certain amino acids such as glycine or alanine.

Many post-transcriptional regulations, including translation and decay are interlinked and act synergistically[58,60–62]. To visualize these relationships and determine whether certain proxies may bias conclusions about others, we performed unsupervised hierarchical clustering on a Spearman correlation matrix of continuous proxies. This analysis revealed three distinct clusters. The first cluster, positively correlated with PTRs, included stability-related proxies (e.g., RNA and protein half-lives) as well as translation-related proxies (e.g., MRL, TE, TR, EVI) and EJC density (Fig. 6b). The second cluster, negatively correlated with PTRs, encompassed proxies associated with post-transcriptional repression, such as transcript lengths (5′UTR, CDS, 3′UTR), sequestration (PBs, SGs), miRNA site density, and ribosome density indicating elongation abnormalities (Fig. 6b). The third cluster, which neither correlates nor anti-correlates with PTRs, was strongly influenced by sequence GC content (Fig. 6b). While some proxies appeared redundant, *Uniform Manifold Approximation and Projection* for Dimension Reduction (UMAP) revealed that each proxy contributed uniquely. For instance, genes with long 5′UTRs were distinct from those with long CDS or 3′UTRs (Supplementary Fig. 5a–c). Surprisingly, some proxies showed rather unexpected relationships with each other (Supplementary Fig. 5d–o), such as the structure of the 3′UTR and the optimal codon composition (CAI), with no clear explanation linking these two properties (Supplementary Fig. 5e, j).

We then performed multiple linear regression using 48 selected proxies across 3306 genes (see "Methods"). This model explained 36.5% of the variance in PTR (Fig. 6c). For comparison, an intrinsic model only based on sequence-dependent proxies explained 21.8% of the variance, while an extrinsic model relying only on measured proxies explained 27.3% of the variance (Fig. 6c). The predicted PTR from this model correlated well with observed PTR values (Spearman Rho = 0.62) (Fig. 6d, e). Importantly, this model provides insights into the "shadow proteome" defined as proteins from coding genes that remain undetected in proteomics experiments. Indeed, in the dataset of 44 AML, 19,125 coding genes were quantified at the RNA level, but only 6669 were quantified by label-free proteomics, leaving



12,456 genes without PTR measurements in this context. For these shadow proteins, RNA expression levels were markedly lower, which partly explains the absence of protein detection as previously suggested[91] (Fig. 6f). Our model allowed us to predict PTR values for these genes using the 48 proxies. We demonstrated that genes in the shadow proteome tended to have lower predicted PTRs than those detected in proteomics experiments (Fig. 6g). Even when restricting the analysis to genes with RNA expression levels (Log_{10}TPM) between 0 and 1.2 for which protein expression detection would typically be expected, shadow proteome elements still exhibited lower predicted PTRs (Supplementary Fig. 6a). In fact, because some proxies

Fig. 5 | The protein, its composition, and structure influence the level of PTR. **a** Schematic overview of the main proxies identified at protein level. **b** Curve representing the median value of the amino acid enrichment (z-score) calculated for each PTR rank (x_n). In violet is represented the valine enrichment. In grey (in the background) are represented all the other amino acids. **c** (*Left*) Heatmap representing median values of the amino acid enrichment (z-score) as a function of ranked PTRs subdivided in 10 groups. Each row represents results for a different amino acid. Each column represents $n = 666$ genes bined depending on ranked PTRs. **c** (*Right*) heatmaps representing two different metrics: Amino Acid Stabilization Coefficient ($AASC^{prot}$) and amino acid decoding time. **d** Curve representing the median value of protein half-life calculated for each PTR rank (x_n). A sliding window extending from rank $n - 100$ to rank $n + 100$ is used to compute the median. Mann–Whitney tests between low PTRs (min to interquartile Q1) vs high PTRs (interquartile Q3 to max) were performed with row values of each proxy. **e** Frequency distribution curves of mRNA half-life (green) and protein half-life (grey) in hours. **f** Scatter plot representing values of $AASC^{RNA}$ (calculated with RNA half-life data) vs $AASC^{prot}$ (calculated with protein half-life data), Spearman rho is annotated. **g** Volcano plot representing fGSEA analysis performed on ranked PTRs and showing the enrichment of genes according to the presence of certain structural elements in their protein among genes with a low (*Left*) or high (*Right*) PTRs. Gene sets with a significant enrichment (FDR > 0.05) are represented in yellow dots. **h** Volcano plot representing fGSEA analysis performed on ranked PTRs and showing the enrichment of genes according to the cellular localization of their protein among genes with a low (*Left*) or high (*Right*) PTRs. Gene sets with a significant enrichment (FDR > 0.05) are represented in yellow dots.

require adequate transcript coverage (e.g., TE), the shadow-proteome genes for which a PTR can be predicted display RNA expression levels comparable to those of genes with measured PTR, thereby ruling out any bias due to RNA abundance (Supplementary Fig. 6b). This phenomenon extends beyond AML, as demonstrated by our data annotation of the Human Proteome Map (HPM) and the Human Protein Atlas (HPA)[92,93]. Indeed, across 30 and 29 tissues, respectively, we observed that the less the protein was expressed in the various tissues, the lower was the predicted PTR (Supplementary Fig. 6c, d).

This integrative analysis not only reveals the interdependencies between proxies and post-transcriptional processes but also provides a predictive framework for explaining the shadow proteome.

POSTCODE: a pipeline for deciphering post-transcriptional regulatory entanglements in omics datasets

We sought to determine whether these proxies could highlight post-transcriptional modulations between two conditions. Since PTRs are determined from the [protein]/[mRNA] ratio, we speculated that the previously described proxies have the potential to influence either the [protein] and/or [mRNA] levels. When analyzing omics data, such as transcriptomics, ribosome profiling, or proteomics between two conditions (x and y) representing different cell states, changes in [protein] and/or [mRNA] quantities can be measured. If the changes between conditions are post-transcriptional, the proxies we described can help to highlight these regulatory modifications. To this end, we developed an annotation and graphical analysis pipeline called POSTCODE, which uses these proxies to detect post-transcriptional deregulations in differential gene expression (DGE) analyses (Fig. 7a). We tested the POSTCODE pipeline on several datasets to demonstrate its relevance. While POSTCODE can be applied to various omics levels, it is theoretically more interesting to apply it to proteomic data, as protein expression reflects the cumulative effect of post-transcriptional regulations on transcriptional activity.

First, we used our pipeline to annotate PTRs in 51 proteomics datasets from AML studies (Supplementary Tables 3). For each of these datasets we showed that the median PTR of the quantified genes varies with the number of different proteins quantified (Fig. 7a and Supplementary Fig. 7a). Therefore, this annotation highlights that some proteomic studies are unable to study genes for which strong post-transcriptional repression exists. Datasets with a median PTR greater than 4000 are often the result of analyses for which protein quantities are often limited like primary cells proteomics or Single cell proteomics analysis experiments. These data enable an assessment of whether proteomic analysis reaches sufficient depth to capture certain post-transcriptional repressions, particularly those targeting genes with low PTRs. If not, the analysis may be limited to the tip of the iceberg, capturing only well-expressed genes while overlooking those undergoing post-transcriptional repression.

We then hypothesized that when analysis depth is sufficient, annotating data with our proxies can uncover post-transcriptional regulations driving observed expression changes, and we proposed here multiple examples to illustrate the accuracy and usefulness of POSTCODE. As a first example, we used proteomic analysis comparing AML blasts from low-responders versus high-responder patients and resistant versus sensitive HL60 cell line to the venetoclax-azacytidine combined therapy[94]. In this published analysis, the resistance was attributed to upregulation of FLT3 and MAPK signaling pathways. Applying POSTCODE to these datasets revealed a major proteostasis effect between low- and high-responder cells (Fig. 7b and Supplementary Fig. 7b). Proteins enriched in low-responder AMLs and resistant HL60 cells are known to be highly stable, suggesting a possible proteasome involvement in the resistant phenotype. We further showed that the chosen proteins of MAPK and FLT3 pathways used in the GSE-Analysis of this published study had very long half-life compared to other MAPK and FLT3 gene sets in the literature (Supplementary Fig. 7c), raising the possibility that the observed enrichment in these pathways could be biased. Our approach, therefore, allows a more precise or more nuanced interpretation of the DGE analysis data.

For the second example, we used one of our previous study and applied POSTCODE on proteomic data to reveal the mechanism explaining d the post-transcriptional downregulation of CEBPA in MOLM14 and MV4-11 models treated with the Flt3 inhibitor quizartinib[95]. POSTCODE analysis showed that the fold change (FC) between quizartinib and DMSO treatment in these models was unbalanced towards increased expression of high PTRs genes and decreased expression of low PTRs genes (Supplementary Fig. 7d, e). Proxies related to protein half-life were especially informative, as upregulated proteins were notably stable, while downregulated proteins, including CEBPA, were unstable (Supplementary Fig. 7d, e). This imbalance suggests a proteostasis shift that could be explained by a cessation of proliferation under treatment accompanied by a reduction in transcription and translation during stasis, with continuous degradation of unstable proteins during the 24 h of treatment.

For the third example, we analyzed proteomic data from an AML model where DDX6, the master regulator of PBs formation, was knocked down using an shRNA strategy. As shown in previous studies, DDX6 knockdown leads to the disappearance of PBs structures in cells, thereby halting the sequestration of specific transcripts. This model has been investigated multiple times at the transcriptomic and translatomic levels by various research groups[70,96]. We applied POSTCODE to this proteomic dataset and found that proteins corresponding to the sequestered transcripts are massively downregulated (Fig. 7c), supporting the hypothesis of a protective role for these transcripts, as suggested by previous studies[70] and providing additional insights into DDX6 knockdown-induced post-transcriptional regulations, raising new questions about its impact on leukemic cell fate.

For the third example, the POSTCODE pipeline also detects other forms of post-transcriptional regulation, such as changes in translation and RNA decay. For example, analysis of RNA-seq data examining the effects of JCS-1, a PROTAC targeting the decapping factor DCPS[97], revealed major post-transcriptional effects, particularly those involving miRNA regulation and sequestration in PBs (Fig. 7d). This is noteworthy, as the only other described non-decapping role of DCPS is its involvement in miRNA degradation[98].

Finally, to further assess the utility of POSTCODE in complex clinical settings, we extended our analyses to proteomic data from 44 primary AML

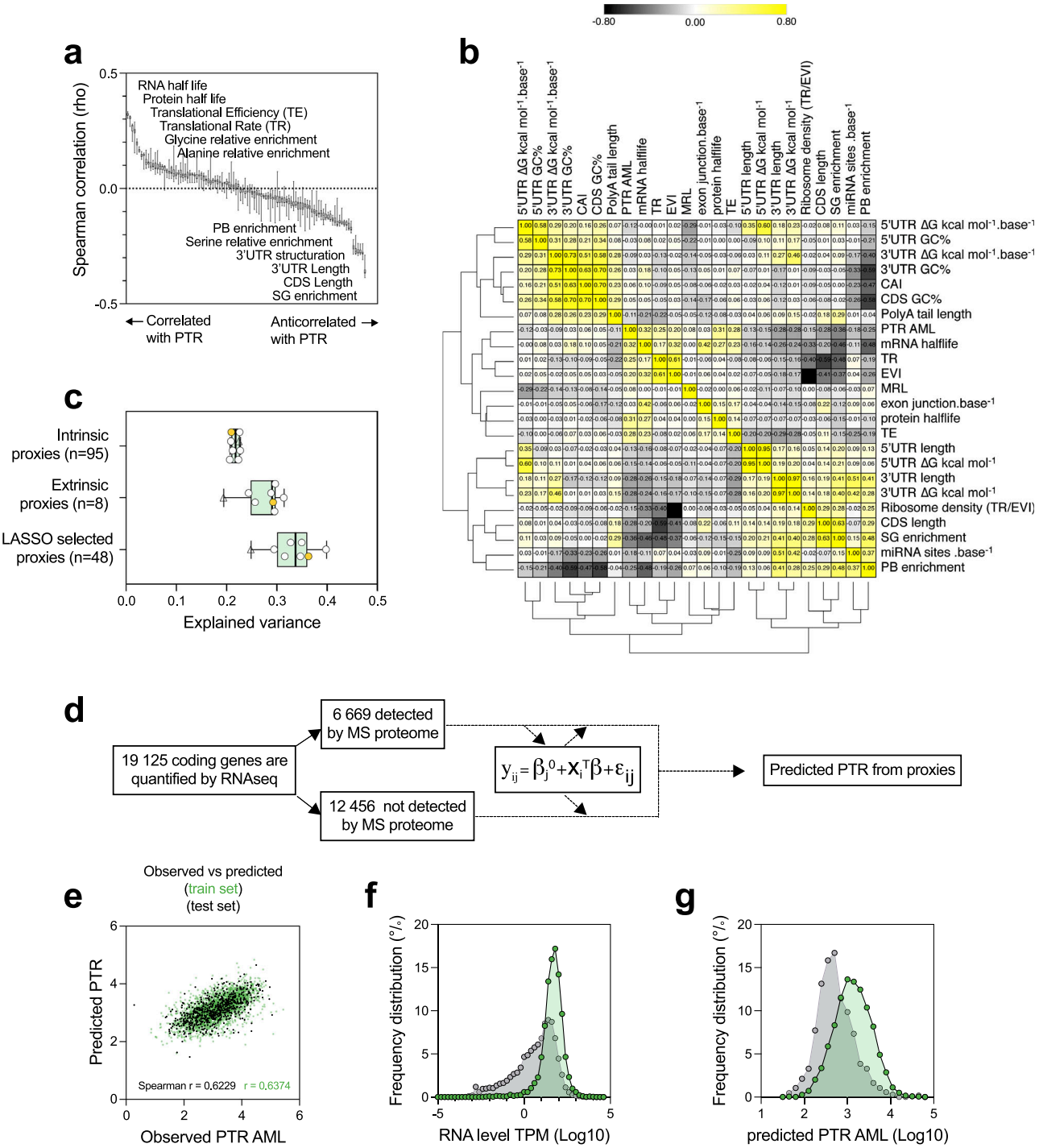


Fig. 6 | The integration of proxies predicts the PTR and explains the shadow proteome. a Spearman correlation (rho) was calculated between PTRs in AML and each proxy and ranked in decreasing order. **b** Heatmap of unsupervised clustering of spearman correlations between proxies and PTRs (rows and columns). Heatmap is clustered using complete-linkage hierarchical clustering based on Euclidean distances. **c** Proportion of the variance explained by selected proxies. Three different models of multiple linear regression have been used: one using all selected proxies, one using intrinsic proxies, and one using extrinsic proxies. Multiple linear regression has been performed for the mean PTR in all AML samples (PTR AML), in FAB subtype, and in the mean PTR of CD34+ samples (PTR CD34+). Box and whisker plots depict the results obtained for this different groups. **d** Schematic overview of the multiple linear regression model used to predict PTRs for genes for which we have all the proxies. where y_{ij} is the subtype/tissue-specific PTR ratio (log10) of gene *i* and tissue *j*, and x_i^T is the ith row of the matrix *X*, which contains 48 proxies. **e** Scatter plot showing observed mean PTR of all AML samples versus predicted PTRs calculated with the multiple linear regression model for $n = 3306$ genes (2644 genes in the train set and 662 genes in the test set) (with 48 proxies available). **f** RNA expression level frequency distribution curves (TPM) of genes whose expression is quantified at protein level by the proteome (green) and of genes whose expression is not quantified at protein level by the proteome (grey). **g** Predicted PTR level frequency distribution curves of genes whose expression is quantified at protein level by the proteome (green) and of genes whose expression is not quantified at protein level by the proteome (grey).

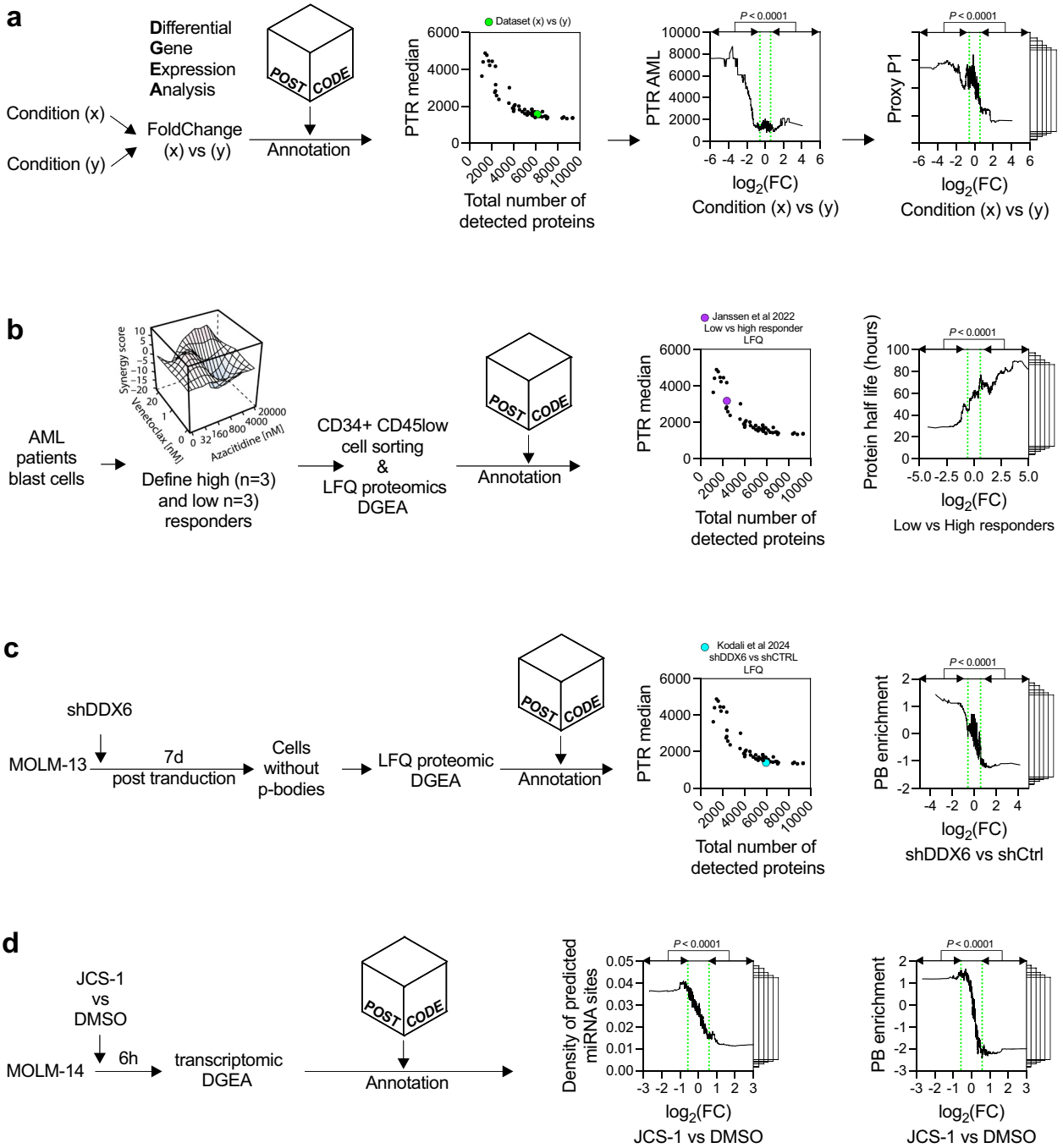


Fig. 7 | POSTCODE: a pipeline for deciphering posttranscriptional regulatory entanglements in omics datasets. a Schematic representation a DGE analysis annotation by POSTCODE pipeline. **b**–**d** POSTCODE workflow: First, calculation of the median PTR of all proteins quantified in the experiment and comparison with 51 published proteomic datasets. Second, curve representing the median value of annotated PTR for each Fold Change rank (x_n). Third, curves representing the median value of annotated proxies for each Fold Change rank (x_n). For these two curves, a sliding window extending from rank $n - 100$ to rank $n + 100$ is used to compute the median. Mann–Whitney tests between FC(up) vs FC(down) were performed with row values of each proxy. **b** POSTCODE analysis applied to proteomic data of AML cells from low and high responders. **c** POSTCODE analysis applied to proteomic data from shDDX6 vs shCtrl MOLM-13 cells. **d** POSTCODE analysis applied to transcriptomic data from JCS-1 vs DMSO in MOLM-14 cells.

patient samples stratified by FAB classification, ELN 2017 prognostic categories, and recurrent molecular alterations. For each classification scheme, we performed differential protein expression analyses (DGEA) comparing each subgroup of interest against all remaining AML samples (e.g., CEPBA mutation vs others, ELN adverse vs others, or FAB M5 vs others). The resulting DGEAs were processed with the POSTCODE pipeline to assess the enrichment of post-transcriptional proxies among up- and downregulated proteins. These analyses revealed extensive subtype-specific enrichment of multiple proxies (Fig. 8a), in contrast to previous comparisons based on controlled experimental perturbations (e.g., drug treatments), where only a limited number of proxies were differentially enriched (Fig. 7a–d). To verify the specificity of these findings, we applied POSTCODE to randomized groups of AML samples and observed no significant proxy enrichment (Fig. 8a, left), confirming that the observed subtype

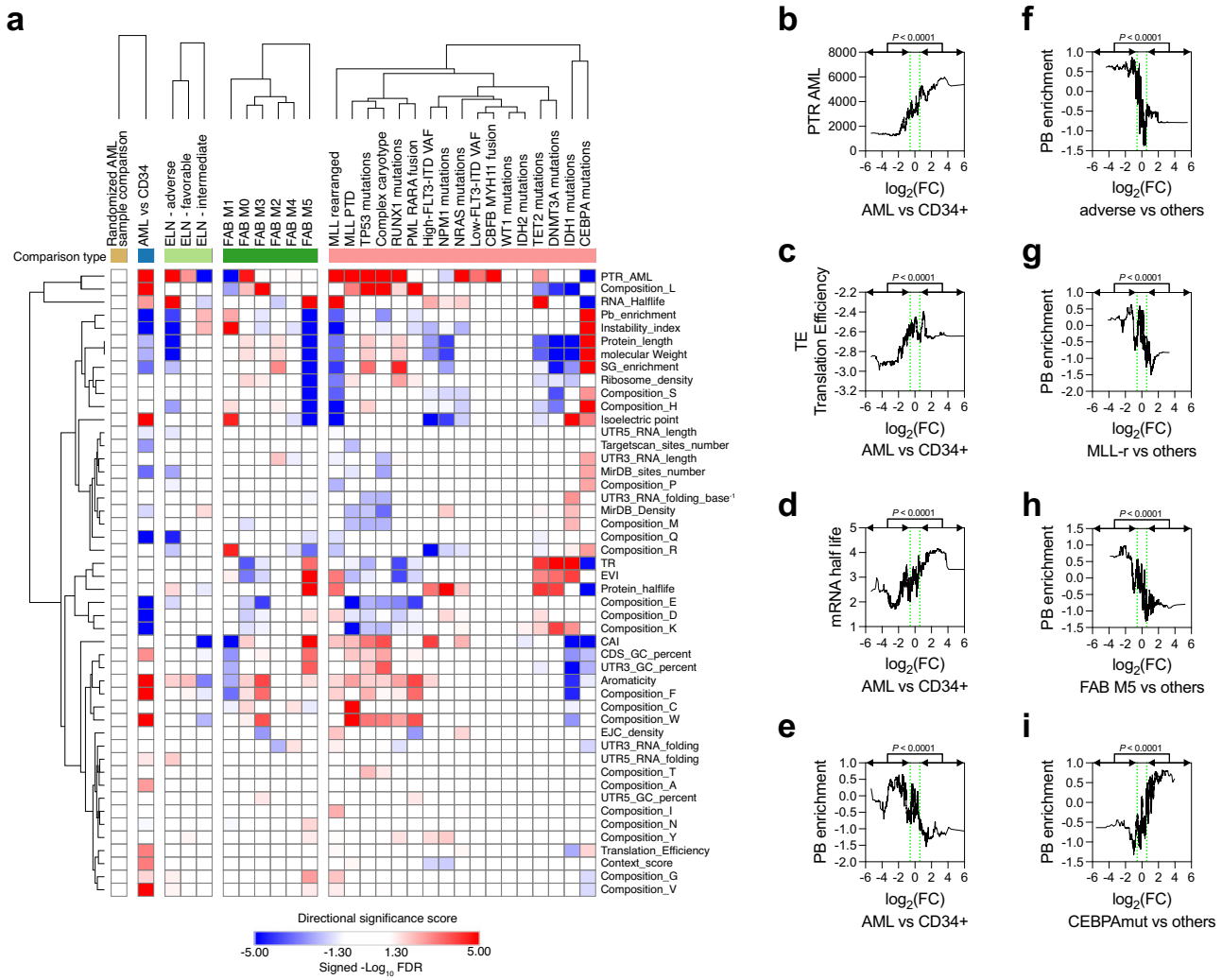


Fig. 8 | AML subtype-specific post-transcriptional programs revealed by POSTCODE. a Heatmap showing unsupervised clustering of directional significance scores for proxies and PTRs across various subtype comparisons in AML. From left to right: light brown, randomized AML sample comparisons; light blue, AML vs. healthy CD34⁺ samples; light green, AML samples stratified by ELN classification; dark green, AML samples stratified by FAB classification; pink, AML samples stratified by mutational background. Each section of the heatmap was clustered independently using complete-linkage hierarchical clustering with Euclidean distance. **b**–**i** Curves representing the median values of annotated proxies across Fold Change ranks (x_n). For each rank, a sliding window from $n - 100$ to $n + 100$ was used to compute the median. Mann–Whitney U tests were performed between the FC(up) and FC(down) groups for each proxy based on raw values. Median values of PTR (**b**), translation efficiency (**c**), mRNA half-life (**d**), and PB enrichment (**e**) across Fold Change ranks for AML vs. healthy CD34⁺ samples. Median PB enrichment values across Fold Change ranks for adverse ELN (**f**), MLL-rearranged (**g**), FAB M5 (**h**), and CEBPA-mutated (**i**) AML samples vs. other subtypes.

patterns reflect true biological variation. For example, when comparing AML proteomes to those of healthy donor-derived CD34+ cells, we observed a consistent upregulation of PTR-high genes in AML. This upregulation was associated with enrichment in proxies linked to high mRNA stability, high translation efficiency, and low sequestration in P-bodies (Fig. 8a, b–e), suggesting coordinated post-transcriptional mechanisms that may support oncogenic protein expression. Further, some proxies revealed clear subtype-specific patterns. For instance, proteins encoded by transcripts known to be sequestered in P-bodies were significantly downregulated in MLL-r, ELN adverse, and FAB M5 AMLs; three biologically related subtypes (Fig. 8f–g). In contrast, CEBPA-mutated AML showed an opposite pattern for this proxy. These findings are consistent with recent reports of decreased P-body abundance in MLL-r leukemia[99], and with our observation of low DDX6 expression across AML subtypes (Supplementary Fig. 8a–d). These results also reinforce observations from our DDX6 knockdown model (Fig. 7c), in which loss of P-bodies led to downregulation of proteins associated with sequestered transcripts, confirming the functional relevance of this regulatory axis. Altogether, these analyses illustrate that POSTCODE can reveal distinct, subtype-specific post-transcriptional regulatory programs in AML. Importantly, the complexity of these signatures also highlights the need for future validation in larger, more uniformly stratified AML cohorts.

In conclusion, we demonstrate the value of using proxies to detect whether post-transcriptional deregulations may be involved in DGE analyses. The POSTCODE pipeline not only helps avoid interpretation biases but can also uncovers molecular pathways and potential therapeutic mechanisms in AML.

Discussion

Our integrated analysis of post-transcriptional regulation in AML focused on identifying the key factors influencing PTRs. The main advantage of this metric, which has been used by other groups[17–21], is that it summarizes the entire spectrum of post-transcriptional processes affecting gene expression within a tissue into a single value. One of the most striking findings was the general consistency of PTRs across diverse tissues, suggesting that many regulatory principles shaping PTR are broadly conserved. Rather than

focusing solely on AML-specific deviations, we first aimed to understand the underlying factors driving this cross-tissue consistency. We then showed that this global conservation does not preclude important context-specific differences. The quantitative effect and functional relevance of individual proxies can vary across biological settings, particularly among AML subtypes. While many previous studies investigated single proxies in isolation, our work builds on these efforts by systematically integrating a broader and biologically informed set of proxies into a unified model and analysis of post-transcriptional regulation.

PTR as a benchmark for known and unknown post-transcriptional proxies

Remarkably, all the proxies we tested produced results consistent with the literature, reinforcing PTRs utility as a metric for extrapolating post-transcriptional amplification or repression of gene expression. For instance, our analysis aligns with previously described translational repressor effect of 5′UTR structuration[40,100], rG4 motifs in the 5′UTR, which are known to not only hamper ribosome scanning, but also actively target mRNA for degradation[101], and also consistent with the described translational repressor effect of uORF and uAUG[36,85,87,102]. Regarding dORFs, recent studies proposed that they may enhance main ORF translation[61,85], while a third study offered a more nuanced view[103]. Our PTR analysis supports the earlier findings, contributing to this ongoing debate. Similarly, the repressive effect of the EJC/EEJ$^{5'UTR}$, recently described[36], was observed but became insignificant when uORF-associated EJC/EEJ$^{5'UTR}$ were excluded, suggesting that uORFs are the primary mediators of this effect[36]. The PTR metric also provides an integrative readout of diverse post-transcriptional regulatory layers, including those mediated by mRNA chemical modifications such as m^6A, m^1A, m^5C, m^7G, ψ, Nm, and ac^4C. These epitranscriptomic marks have been implicated in multiple processes such as translation efficiency, mRNA decay, splicing, SGs sequestration, or EJC deposition, and can exert either repressive or activating effects on gene expression[104–108]. In line with previous studies, our analysis supports a repressive role for m^6A and m^5C, consistent with their reported functions in promoting RNA degradation or inhibiting translation[109,110]. Conversely, we observed that Nm-modified transcripts are associated with higher PTRs, supporting a post-transcriptional activating role, potentially through increased transcript stability[111]. Although the overall Kozak context did not differ dramatically between high- and low-PTR genes, we identified a significant enrichment of C or U at position −3 relative to the AUG in low-PTR transcripts, in line with prior reports linking these nucleotides to reduced translation efficiency[112]. We also demonstrate that the RNA structure downstream of the start codon appears to promote gene expression, possibly by increasing the small subunit's dwell time, allowing better recognition of the start codon, thus promoting initiation, as previously suggested[40,56]. Moreover, we confirmed that using a consensus AUG start codon is associated with optimal protein production, as compared to non-AUG start codons[113]. PTRs also serve as a reliable metric for validating in vitro experimental data[53–55,114,115]. For example, MRL, defined by massively parallel in vitro translation assays, aligns well with in vivo gene post-transcriptional amplification measured by PTRs[41]. Finally, the absence of correlation between poly(A) tail length and PTRs is consistent with previous observations that, in non-embryonic systems, poly(A) length does not robustly influence mRNA translation or decay[116,117].

Interconnection of translation efficiency, decoding speed, and stability

Our analysis of the coding region reveals the extent to which translation efficiency, decoding speed, and elongation rate are intertwined with degradation sensitivity (CSC^{RNA}) or transcript half-life[47,61,63,64,68,69]. Additionally, we provide new insights into the sometimes controversial role of tRNAs by showing that tRNA abundance and optimality are only weakly correlated with PTR levels[45,53,65,118]. In contrast, the Codon Adaptation Index (CAI), a general measure of codon bias, shows a stronger association with PTRs[59]. While CAI reflects the overall codon usage bias, codon-by-codon analysis reveals that codon bias only partially explains PTRs, whereas factors like decoding time and codon-induced stability correlate more strongly with PTRs. The same dynamics hold true when we examine amino acid composition. To this end, we introduce two new metrics: (1) Amino acid decoding time, which corroborates reports that synthesis rates are dependent on amino acid composition[119]; and (2) AASC^{prot}, which may explain the unexpected correlation between mRNA and protein half-lives, despite these two measures differing in magnitude. This correlation likely stems from evolutionary pressures and energetic constraints[13]. Due to energy conservation, the RNA and protein half-lives of housekeeping genes may have been maximized, while those of regulatory genes (such as transcription factors, signaling genes, and chromatin-modifying enzymes) are minimized to enable precise regulation. These selective pressures may explain why AASC^{RNA} correlates so strongly with AASC^{prot}, and why amino acid composition impacts transcript half-life[61,66].

How to interpret length dependencies?

Regardless of the region—5′UTR, CDS, or 3′UTR—transcript length is consistently associated with a low PTR. Several explanations account for this observation. For the 5′UTR, the explanation seems relatively straightforward: the longer the sequence, the longer the scanning time for the ribosome[120]. Additionally, longer 5′UTRs are more likely to contain repressive elements such as uORFs, uAUGs, IRESs, and rG4 motifs, which further impede translation.

For the CDS, the factors contributing to PTRs variation are probably more complex. As with the 5′UTR, longer CDS sequences require longer decoding and translation times to produce a protein. Additionally, phenomena such as ribosome drop-off or frameshifting during translation could hinder protein synthesis, though the occurrence rates described in the literature make these factors relatively minor in explaining the large PTR differences between short and long genes[49–51]. More importantly, the difference in TE between short and long transcripts must be considered. Several ribosome profiling studies have shown that longer transcripts generally have lower TE and translation initiation rates than shorter ones[42–48]. One possible explanation is that translation initiation is influenced by transcript length[121–123]. Indeed, some studies suggest that ribosomes completing a round of translation may re-initiate on the same transcript via a mechanism known as CLAR (Closed Loop Assisted Re-initiation). This is based on the closed-loop model, where the RNA circularizes due to interactions between the cap and polyA tail, bringing the 5'UTR and 3′UTR into proximity[124–129]. Mathematical models have recently shown that this CLAR phenomenon could explain the observed differences in TE between short and long transcripts[51,130]. In essence, shorter CDS transcripts have greater access to this re-initiation process. Thus, a transcript with a 400-nucleotide CDS may experience re-initiation 10 times more frequently than one with a 4000-nucleotide CDS. Data also show more frequent mRNA-eIF4E associations for short transcripts compared to long ones[129], and another study suggests that initiation rates increase progressively with each "round" of translation[131]. This study also indicates that there are optimal lengths for 5′ UTRs and 3′UTRs (80 and 300 nucleotides, respectively) for promoting re-initiation. Finally, it suggests that re-initiation is a different phenomenon from initiation, in particular because it is less dependent on the eIF4A helicase[131]. However, no direct visualization of repeated termination/reinitiation cycles for the same ribosome on the same transcript has yet been described.

Another possible explanation for the length bias observed in ribosome profiling and PTR data is related to the steric hindrance of longer transcripts. Indeed, the topographically condensed mRNA molecule initially has, prior to translation, restricted access to the 5′ cap for eIF4E binding, and this is even more pronounced when the sequence is long due to the secondary and tertiary structure that accumulates[129]. This model suggests that disruption of mRNA structure leads to improved cap recognition and predicts that cap recognition may become more efficient as polysomes form and begin to increase in size. Interestingly, this could, in principle, allow for a cooperative effect on the loading rates of the cap-dependent pre-initiation complex and eIF4F during the early stages of translation.

In addition to "steric hindrance" and "re-initiation," "codon composition" could play a regulatory role in the length dependency of PTRs. A recent report shows that non-opt codon sequences limit protein expression without being linked to steric congestion at the beginning of the sequence, without being linked to the formation of a traffic jam or collision on the sequence, and without decay being able to explain the differences observed[132]. Another group with similar results seems to point to a potential role for eIF3 in this phenomenon[133]. While the exact regulatory mechanism is unclear, non-optimal codon usage reduces initiation efficiency[132–134]. The closed loop model could explain why two sequences of equal length but different optimality lead to different expression. These three hypotheses, "codon optimality," "steric hindrance," and "re initiation", can be combined. Thus, as higher-order polysomes are formed, the translation of mRNA becomes increasingly facilitated both by access to the cap, which is more accessible to initiation factors, and by access to an efficient reinitiation of terminal ribosomes, which locally increases the concentration of free ribosome subunits as described in CLAR models. Although these phenomena are present in all transcripts, they could favour the shortest mRNAs in particular, which could explain the length bias observed via TE and PTRs[129,131].

The direct role of the length of the 3'UTR region on translation is poorly documented. An optimal 3'UTR length of 300 nucleotides has been suggested for promoting re-initiation, and similar lengths have been shown to improve translation efficiency in vitro[131,135]. Beyond this, it is difficult to conceive of a direct mechanism through which 3′UTR length alone regulates translation. However, our data show an anti-correlation between 3′UTR length and proxies such as TE, EVI, and TR, suggesting at least an indirect role. The longer the 3′UTR, the more likely it is to recruit trans-acting mechanisms such as RBP or miRNA binding. These bound factors could explain the reduced translation efficiency of mRNAs with long 3'UTRs. We note indeed that miRNA site density is anti-correlated with TE. Finally, in addition to regulating translation, the 3′UTR length is involved in numerous phenomena that regulate the PTR. In addition to influencing translation efficiency, this proxy is also involved in regulating stability via the recruitment of RBPs and miRNAs, leading to mRNA degradation. A long 3′UTR makes the transcript more susceptible to cleavage by a disintegration mechanism than a transcript with a short 3′UTR. Finally, these same mechanisms in trans can also lead to the sequestration of the transcript in SGs or PBs and down-regulate the PTR[70,71,136].

Taken together, our data and integrated analysis of these proxies highlight the intricate interconnections between post-transcriptional regulations, offering a deeper understanding of the mechanisms underlying PTRs. This approach not only clarifies the main contributors to PTRs but also allows us to quantify their relative effects. Additionally, our analysis of RBPs and miRNAs enables us to classify these regulators according to their repressive or amplifying effects, which, to our knowledge, has not been reported before. The results align well with existing literature on each RBP studied[11].

From descriptive data to predictive analysis

While our analysis aimed to provide a comprehensive description of the key elements influencing PTRs, we also sought to leverage these data for predictive purposes. Using a multiple linear regression model, we predicted PTRs based on the selected proxies, following strategies previously applied by other groups[19,21]. Importantly, the proxies we employed were derived from multiple tissue types. Some of them rely on experimental data from cell lines, for which tissue-specific bias could be a concern. However, we demonstrated that proxies obtained from datasets across distinct tissue subtypes remain highly consistent[6,68,70,71,88,103,137,138]. Therefore, we considered tissue bias to be negligible, and the accurate prediction of PTRs in AML using data not directly derived from AML models further supports the robustness of this approach. Nonetheless, measuring all the post-transcriptional determinants, i.e., determining the mRNA half-life, protein half-life, SGs interactome, PBs interactome, TE, EVI, and TR in AML models, could increase the predictions made here and allow a gain in analytical fine-tuning. To put predicted PTRs to practical use, we introduced the concept of the shadow proteome and showed that it is explained by both low transcriptional and post-transcriptional expression. This provides an additional layer of understanding to data from resources like the Human Proteome Atlas and Human Proteome Map[92,93]. However, the interpretation of shadow proteome genes should not be made solely through the lens of low PTR values. Indeed, non-detection of proteins in bottom-up proteomics may also result from factors such as poor peptide ionization, membrane localization, or other methodological reasons. Therefore, while post-transcriptional repression may contribute to the shadow proteome, it likely acts in combination with these other technical and biological determinants. Furthermore, our model is simple and limited to genes for which all proxies are available. To extend predicted PTRs to more genes, strategies like missing value imputation or more sophisticated modeling approaches will be implemented in future versions of POSTCODE.

To promote accessibility, we developed a user-friendly Shiny interface that facilitates data annotation and visualization, enabling researchers to rapidly evaluate the potential contribution of post-transcriptional regulation in their DGE analyses, analogous to how term enrichment tools are routinely applied. We demonstrated the utility of this framework across a variety of use cases, including drug response (e.g., JCS-1 treatment), gene silencing (e.g., DDX6 knockdown), and clinically annotated AML patient datasets. Across these analyses, we showed that POSTCODE provides critical insights to reduce interpretation bias, and that reporting the median PTR of quantified genes offers a robust estimate of the effective proteomic depth, helping to contextualize which portions of the proteome are being captured in each experiment. The application of POSTCODE to DGEA from 44 primary AML samples stratified by molecular, cytological, and prognostic features revealed its ability and limitations in uncovering post-transcriptional heterogeneity within clinically complex cohorts and underscore its potential for identifying subtype-specific regulatory programs in AML, such as the link between MLL-rearranged leukemias and reduced P-body sequestration, in line with prior reports[99].

Looking ahead, future applications of POSTCODE will benefit from larger, more uniformly stratified patient cohorts to address the intrinsic heterogeneity of AML. Additional directions include the integration of emerging proxy types as new datasets become available, the adaptation of the framework to single-cell transcriptomic and proteomic data for estimating PTR variation at single-cell resolution, and the application of POSTCODE to high-throughput functional screens such as CRISPR or small-molecule libraries to uncover regulators of key post-transcriptional pathways.

Limitations

Despite the strengths of this approach, several limitations should be acknowledged:

First, our analysis was conducted at the gene level rather than the transcript level, for reasons of simplicity and generalizability. Accurately identifying the dominant isoform in each AML sample is challenging due to splicing heterogeneity and the co-expression of multiple isoforms. In addition, many of the proxies used in our model, such as mRNA half-life, translation efficiency (TE), and elongation velocity index (EVI), are only available at the gene level in public datasets. Isoform-level modeling is further limited by the lack of isoform-resolved proteomics data. To address these constraints, we adopted a representative transcript approach, often selecting the isoform with the longest UTR or CDS, as done in previous studies. Consequently, the data we used reflect either the major isoform or an average of the proxies across multiple isoforms for each gene. This simplification may result in underestimation or misattribution of isoform-specific regulatory effects, especially in biologically heterogeneous contexts such as AML. While this choice may not capture all isoform-specific regulatory features, it reduces the risk of missing relevant sequence elements. Future improvements could involve expression-weighted proxies at the isoform level, informed by long-read sequencing.

Second, some of the metrics we employed could be refined through more direct and accurate measurements. For example, RNA secondary structure was estimated using computational folding algorithms, which may not faithfully reflect in vivo structural dynamics. This limitation is particularly relevant under leukemic stress conditions, where RNA folding may differ substantially from predicted equilibrium structures, potentially biasing the impact of structure-related proxies such as 3′UTR accessibility or stability. Incorporating in vivo measurements such as transcriptome-wide structure probing into future versions of our pipeline could improve the reliability of our predictions.

Third, our model assumes additivity on a logarithmic scale, which may oversimplify the actual regulatory dynamics. For instance, log-additive modeling may obscure nonlinear or context-dependant interactions between regulatory elements. These elements often operate in a coordinated manner, influenced by their spatial proximity, distribution along the transcript, or subcellular localization of the mRNA. Such non-additive interactions, like those between RNA-binding proteins and miRNA motifs, may not be captured by our current model, thereby limiting interpretability in genes with intricate regulatory architectures. Additionally, while we identified conserved sequence elements associated with PTRs, these elements might serve other functions that were not directly addressed in our analysis, as high PTRs often correlate with multiple overlapping characteristics. This functional overlap could confound causal interpretations regarding individual regulatory elements. Experimental validation will be necessary to clarify these questions.

Nevertheless, our study provides a valuable resource of conserved sequence elements that warrant further mechanistic exploration, offering a foundation for future functional studies.

Methods

Data collection–source samples for PTR measurement

Proteomic and transcriptomic data for AML samples were sourced from a LAML TCGA study[22]. In this study, bone marrow samples were collected at diagnosis from adult patients with de novo AML under a biobanking protocol approved by the Institutional Review Board (IRB). While 200 of these samples had already been used for transcriptomic analysis in the LAML TCGA study, 44 samples were selected to represent the main cytogenetic and mutational profiles of the cohort. Bone marrow buffy coat cells were cryopreserved immediately post-collection, without further manipulation. Cryovials were thawed in the presence of the cell-permeable, irreversible serine protease inhibitor di-isopropyl fluorophosphate (DFP) to inhibit myeloid serine proteases.

Proteomic and transcriptomic data from healthy control bone marrow cells were derived from three independent donors from the same study[22]. These samples underwent lineage depletion (Lin-) to enrich for progenitor and precursor cells, using the Miltenyi Biotec reagent. Additionally, bone marrow cells from three healthy donors were enriched for $CD34^+$ stem/progenitor cells (CD34+), utilizing the Miltenyi Biotec reagent and autoMACS separator following the manufacturer's protocol.

Proteomic and transcriptomic data from healthy human tissues were collected from two studies[92,93] : the Human Proteome Atlas (HPA) and the Human Proteome Map (HPM). Full details of tissue, donor information, sample preparation, and measurement protocols are provided in these respective studies. To resume, tissue samples for the HPA were obtained from the Department of Pathology, Uppsala University Hospital, Sweden, as part of the Uppsala Biobank. Tissue samples for the HPM were collected post-mortem as part of a rapid autopsy program from three adult donors for each tissue, obtained from the National Disease Research Interchange (NDRI, PA, USA). Haematopoietic cells were isolated from leukopaks of healthy volunteers participating in routine platelet pheresis, with IRB approval from Johns Hopkins University and informed consent obtained from all participants. Haematopoietic cell populations were isolated sequentially from leukopaks using Miltenyi Biotec magnetic beads as per the manufacturer's instructions, targeting CD14+ monocytes, pan CD56+NK cells, CD20+ B cells, CD8+ T cells, and pan CD4+ T cells. Peripheral blood mononuclear cells (PBMCs) were enriched by Ficoll gradient centrifugation, incubated with magnetic beads in modified MACS buffer (0.05% BSA and 2 mM EDTA), and separated using autoMACS (Miltenyi Biotec). Purity was confirmed by flow cytometry. Isolated cells were washed in PBS and stored at −80 °C until use. Platelets were isolated from platelet-rich plasma of healthy donors, pelleted by centrifugation, resuspended, and washed in modified Tyrode's buffer before lysis.

Data collection–respective proteomic and transcriptomics workflow specifications

For the three main datasets used in this study (LAML TCGA - HPM - HPA), raw data are publicly available on GDC Data Portal, ArrayExpress, PRIDE, or MassIVE Database under following accession numbers **E-MTAB-2836, TCGA-LAML, PXD010154, PXD000561, MSV00089012**, and **MSV000089029**. For each of these datasets, the techniques, protocols, and machines used to perform the proteomic and transcriptomic analyses differ. We have used the processed data produced by the authors to avoid the introduction of systematic analytical errors. Details of technical procedures, sample preparation, filtering, and data normalization are therefore given in the source articles[22,92,93]. To calculate PTRs, we used absolute quantification of proteins (LFQ) to obtain number of protein copies per cell via histone normalization using Perseus software version (1.6.14.0) and the "Protein Ruler" approach[23]. Differential gene expression (DGE) analyses comparing samples according to their morphological (FAB), prognostic (ELN), or molecular subgroup were performed using protein copy-per-cell data. For HPM - HPA datasets, Label Free absolute quantification is also used via an "Intensity-Based Absolute Quantification" (iBAQ) normalization. For transcriptomic data, the LAML TCGA dataset used data normalized in transcript per million (TPM), whereas for HPA, normalization was performed in Fragments Per Kilobase Million (FPKM). To study the relationship between RNA and protein levels, we then used Spearman correlations due to the presence of outliers and the non-normal distribution of the transcriptomic and proteomic data[139]. However, the use of Pearson correlations showed the same results and led to the same conclusions.

Gene annotation set

Intrinsic and extrinsic proxies. For extrinsic proxies, data were collected from multiple resources listed in Supplementary Data 4. In the literature, entry varies depending on studies: some use ensemble IDs, others UCSC IDs, others genenames, etc. We used custom R scripts embedding the *merge()* function with the *Tidyverse dplyr* package to compile annotations based on gene names. For each proxy, the methodology chosen to pass from transcript level data to a single value per gene is specified.

For intrinsic proxies, human gene annotations of protein-coding genes were derived from Ensembl v109 (hg38genome build) gene nucleotide sequences[140] and from UniProtKB (Human Swiss-Prot Reviewed) amino acid sequences. Only protein-coding genes were carried forward for analysis. Out of all transcript's isoforms, depending of the region of the transcript studied, we selected the one with the longest CDS, the longest 5′UTR, or the longest 3′UTR as the representative transcript for that gene in order to simplify the analyses, as in some other studies[59,68].

Sequence-based proxies

For all gene names IDs, sequence lengths, GC content, and exon junction density EJC/EEJ were calculated from representative transcripts.

RNA folding energy analyses (secondary structure proxy) were computed via *Vienna-RNAfold* package[141] for all 5′UTR, CDS, and 3′UTR extracted from Ensembl v109 (hg38genome build). Because the relationship between energy vs length is linear, for all those sequences, the folding energy was normalized by the length to obtain a value in $\Delta G/\text{kcal mol}^{-1}\ \text{base}^{-1}$. For genes with multiple isoforms, the minimum, the maximum, and the mean folding energy were computed. To obtain one value per gene, we selected the mean folding energy. Furthermore, using representative transcripts, we add the computation of 100 nt windows in the 5′UTR sequence and around the

start and stop codons that we also calculated and normalized as for the entire sequences.

Gene motif and structural proxies derived from literature

All specific motifs like EJC/EEJ$^{5'UTR}$, 5′TOP, IRES, rG4, uORF, uAUG, dORF are issued from literature. Lists of genes sharing same motifs have been transformed in gene sets to perform fGSEA. All geneset used are listed in (Supplementary Data 3).

EJC/EEJ$^{5'UTR}$ were extracted from a study[36]. that classified the transcripts into four different classes. These classes are defined according to the presence or absence of EJC/EEJ before the Start codon of the canonical ORF and according to the presence or absence of uORFs. We therefore created two gene sets, one for genes with EJC/EEJ$^{5'UTR}$ and without uORF and one for genes with EJC/EEJ$^{5'UTR}$ and with uORF.

Cap-independent transcripts were extracted from an in vitro high-throughput bicistronic assay that quantify cap-independent translation[39]. In this study, the independent cap potential of human 5′UTRs is scored according to the expression of a reporter gene. We selected genes 4 times more translated than the control (score greater than 4).

5′TOP motifs were collected from two studies of reference[37,38]. A geneset was created for each of them.

IRES were collected from different databases and studies, notably the IRESbase database[142] http://reprod.njmu.edu.cn/cgi-bin/iresbase/index.php, the IRESite Database[143] http://www.iresite.org, and from specific studies[144].

rG4s were collected from QUADRaltlas database in which RG4s are distinguished dependending of the experimental (rG4seq + RT-stp profiling) or predicted (G4Hunter + pqsfinder + OGRSmapper) evidences[34] https://rg4db.cibio.unitn.it. We designed three gene sets, one mixing experimental and predicted rG4s, one with only experimentally proved rG4s, and one with predicted (3 tools) rG4s.

uAUGs, uORFs o-uORF, dORF were collected from multiple studies[85–87,102] in which ORFs were predicted from Ribo-seq datasets of different cell lineage models and primary cell types and tissues using RiboTaper[145] v1.3, RiboTISH[146] and PRICE[147]. A geneset was created for each of them.

mRNA epigenetic marks (N^4-Acetylcytidine (ac^4C) N^6-méthyladénosine (m^6A) N^1-méthyladénosine (m^1A) N^5-méthylcytosine (m^5C) N^7-méthylguanosine (m^7G) pseudo-uridylation (ψ) 2′-O-Méthylation (Nm)) were collected from different databases[77,78] and from leading studies for each specific marks[8,9,79–84]. A geneset was created for each of them.

RBP related proxies

Nucleotide motifs were gathered from various studies and databases: the ATtRACT database[76] based in particular on the RBPMAP database[148], and the CISBP RNA database[149] for 132 RBPs. Additional 5′UTR and 3′UTR motifs known to regulate post-transcriptional processes were included from published studies[19]. All k-mer motifs are listed in (Supplementary Data 3). The function *vcountPattern()* from the *Biostrings R* package was used to calculate the number of occurrences of each motif in UTR sequences. K-mer frequencies were computed separately for each of the 5′UTR and 3′UTR sequences (using parameters width = {6 to 9}, depending on k-mer size, and step = 1). K-mers with width parameters of <6 were excluded from analysis, as they generated excessively large gene sets.

RBP gene sets were designed using the ENCORI database[74], leveraging recent advances in high-throughput sequencing of immunoprecipitated RNAs after cross-linking (CLIP-Seq, HITS-CLIP, PAR-CLIP, CLASH, iCLIP) and integrating thousands of these experiments. For 284 RNA-binding proteins (RBPs), we compiled processed data on RBP CLIP peaks. We ranked the targets of each RBP in descending order based on the number of CLIP sites. To generate the gene sets, we selected the top 200 targets of the RBPs with the highest number of CLIP sites, considering these targets as the most specific for each RBP (Supplementary Data 3). For each RBP described in the ENCODE database, we further categorized them according to their preferential binding location (intron/5′UTR/CDS/3′UTR)[75].

miRNA related proxies

miRNA gene sets were designed using ENCORI database[74], which intersects the predicting target sites of 611 miRNAs with binding sites of Ago protein, derived from CLIP-seq data, we created 611 gene sets. Each geneset contains the targets of a specific miRNA, comprising all transcripts predicted as targets by multiple prediction programs (*PITA, RNA22, miRmap, DIANA-microT, miRanda, PicTar*, and *TargetScan*) (Supplementary Data 3).

To calculate the number of predicted miRNA binding sites per transcript, we used two different miRNA target site databases. First, we utilized miRDB[150,151] (https://mirdb.org), from which we downloaded all predicted targets from miRDBv6 (using the prediction tool miRtarget v4 and miRNA sources from miRbase 22). For each NCBI RefSeq transcript, we calculated the occurrence of different miRNA sites and computed the site density (sites per nucleotide) by dividing the number of sites by the transcript length. To obtain a single value per gene, we retained the transcript isoform with the highest number of miRNA sites and the one with the highest density. Secondly, we used the TargetScan database[152,153] (https://www.targetscan.org), from which we downloaded curated and conserved miRNA target sites for each 3′UTR, identified by Ensembl IDs. We utilized the precomputed "Total number of conserved sites in 3′UTR" and calculated site density by dividing the number of sites by the length of the 3′UTR of the specific transcript. To obtain a unique value per gene, we selected the transcript isoform with the highest number of miRNA sites and the one with the highest density.

We also used the TargetScan database to download tables listing miRNA efficiency scores[152,153] ("predicted KD" and "Context++ score") for all miRNA-transcript pairs. To conduct gene-level analyses, we filtered these data to retain only the most negative efficacy value per gene-miRNA pair, thus converting from transcript to gene level. We then applied the same filter across miRNAs targeting the same gene, yielding a single efficacy value per gene. In the original study, "the predicted relative KD" values were calculated using a convolutional neural network (CNN) designed to assess the binding affinity between any mRNA and any 12-nt sequence. This CNN was trained on over a million KD measurements in six AGO2-miRNA complexes, as well as on mRNA fold change data observed after miRNA transfection in HeLa cells. The "context++ score" (CS) for a specific site is derived as the sum of contributions from 14 features, which include: site type, supplementary pairing, local AU content, minimum distance, sRNA1A, sRNA1C, sRNA1G, sRNA8A, sRNA8C, sRNA8G, site8A, site8C, site8G, 3′UTR length, SA, ORF length, ORF 8mer count, 3′UTR offset 6mer count, TA (target site abundance), SPS (seed-pairing stability), and P_{CT} (probability of conserved targeting).

PolyA related proxies

PolyA lengths were collected from processed data of multiple datasets. As with the other proxies, we cross-referenced the datasets to assess their concordance. Here, in contrast to the other proxies, this concordance is poor (Supplementary Information). This may be explained by the diversity of methods used to quantify polyA tail length, and perhaps by the diversity of cell types. Indeed, we listed here the analyses used according to methodology and cell type: data from FLAMseq[154] performed in Hela, iPSC and cerebral organoids (**GSE126465**), data from Nanopore native RNA-Seq performed in GM12878[155] (**NA12878**), Hela[156] (**PRJEB53494**) and iPSC[157] (**GSE127890**), data from TAIL-Seq performed in Hela[158] (**GSE51299**), data from PAL-Seq performed in Hela[116] (**GSE52809**) and data from PolyA-Seq performed in HCT116[159] (**GSE84287**). We also determined the presence of the AAUAAA polyadenylation motif within the 3′UTR sequences using the same methodology as for the other motifs.

Proxies developed from codons and amino acids analysis

Codon frequencies were extracted from CDS sequences from the longest isoform using the *oligonucleotideFrequency* function (parameters width = 3, step = 3, as.prob = TRUE) in the *Biostrings* R package, providing frequencies by normalizing the counts for each codon by the sum of all codon counts for the gene. We then calculated a z-score expressing the enrichment in a codon for a transcript compared with all the human transcripts.

Amino acid frequencies were extracted from CDS sequences in the *Biostrings* R package, using the *alphabetFrequency* function (as.prob = TRUE) to normalize to amino acid sequence length, providing frequencies by normalizing the counts for each amino acid by the sum of all amino acid counts for the protein. We then calculated a z-score expressing the enrichment in an amino acid for a protein compared with all the human proteins.

Synonymous codon selection preferences were assessed through the relative synonymous codon usage (RSCU) method[160]:

$$RSCU_{ij} = \frac{x_{ij}}{\frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}}$$

where x_{ij} is the usage of the jth codon for the ith amino acid, which is encoded by n_i codons.

Codon Adaptation Index was calculated using the longest CDS with the https://ppuigbo.me/ programs *CAI* calculator. The CAI was calculated according to the original study[161]:

$$CAI = \exp\left(\frac{1}{L}\sum_{i=1}^{L}\log\left(\frac{f_i}{\max\left(f_j\right)}\right)\right)$$

where L is the number of codons, f_i is the frequency of the codon i, and $\max(f_j)$ is the maximum codon frequency for that amino acid.

The tRNA Adaptation Index (tAI) of each codon was calculated according to the original study[162]. To calculate this index, the absolute adaptiveness value W_i for each codon i is defined as

$$W_i = \sum_{j=1}^{n_i}\left(1 - s_{ij}\right)\mathrm{tGCN}_{ij}$$

where $_{\mathrm{ni}}$ is the number of tRNA isoacceptors that recognize the ith codon, $\mathrm{tGCN}_{\mathrm{ij}}$ is the gene copy number of the jth tRNA that recognizes the ith codon, and s_{ij} is a selective constraint on the efficiency of the codon–anticodon coupling. For a gene composed of L codons, indexed by $i = 1, 2, \ldots, L$, the tRNA Adaptation Index is defined as the geometric mean of the w_i values

$$\mathrm{tAI}(gene) = \left(\prod_{i=1}^{L} w_i\right)^{1/L}$$

tAI value for HEK293 and HeLa cells was obtained from literature[66,163]. tRNA genomic copy number as well as tRNA expression in iPSC, K562, and HEK293T were obtained from literature[67].

Codon Stability Coefficients on mRNA half-life ($CSCs^{RNA}$) were previously published using several models and several mRNA half-life experiments and show a very high conservation[61,63,66,164]. $CSCs^{RNA}$ were obtained by calculating the Pearson correlation coefficient between transcript half-lives and codon frequency for all 61 non-stop codons using the *cor.test* function in the *R stats* package with method = "pearson".

Amino acid stabilization coefficients on mRNA half-life ($AASCs^{RNA}$) were previously published for the longest sequence per gene when multiple isoforms were described[66]. $AASCs^{RNA}$ were obtained by calculating the Pearson correlation coefficient between transcript half-life and amino acid frequency for all 20 canonical amino acids using the *cor.test* function in the *R stats* package with method = "pearson".

Amino acid stabilization coefficients on protein halflifes ($AASCs^{prot}$) were not previously published. We calculated this metric by calculating the Pearson correlation coefficient between protein half-life[88] and amino acid frequency for all 20 canonical amino acids using the *cor.test* function in the *R stats* package with method = "pearson".

Kozak sequences

The number of mismatches between the sequence flanking the start codon and the Kozak consensus sequence was determined by considering base by base whether the successive positions −6, −5, −4, −3, −2, −1, +1, +2, +3, +4, +5, and +6 correspond (TRUE) or not (FALSE) to the base expected in the consensus sequence.

Sequence logos were generated using (https://weblogo.berkeley.edu/logo.cgi) the WebLogo server for a region encompassing nucleotide sequences -6 to +6 flanking the start codon for the 200 higher PTR genes and for the top 200 lower PTR genes.

Initiation rate related proxy: MRL

Data from a massively parallel translation assay[41] (**GSE114002**) were used to compute a proxy for initiation rate called "Mean Ribosome Load" (MRL). This was assessed by inserting a 50-nucleotide preSTART 5′UTR sequence into plasmid constructs for 20,530 human transcripts corresponding to 12,503 genes. Methodological details are described in the study from which these data were obtained. To obtain a single value per gene, we used processed data and we compute the mean values of different transcript isoforms for a same gene.

mRNA decay related proxies

Data from a meta-analysis[68] of 16 transcriptome-wide mRNA decay rate datasets were used to obtain a proxy for mRNA half-life. Methodological details are described in the study from which these data were obtained. We did not modify the data that had already been transformed by the authors to obtain a single value per gene. This dataset was chosen because the authors demonstrate that PC1 computed is highly conserved whatever the tissue and even between mammalian species. In addition, we used this same dataset to calculate the average value of half-lives in hours across all datasets for our representation of relative frequency in Fig.5e.

Translation-related proxies

Translation Efficiencies (TE), Translation Rate (TR), Elongation Velocity Index (EVI), Ribosome Density (D = TR/EVI) were collected from diverse ribosome profiling studies[47,69]. We used data of TE of hematopoietic model of primary CD34+ cells cultivated shortly into produce erythroid progenitors as a reference[47] (**GSE89183**). We used the pioneer study that has determined TR, EVI, and D with Ribosome profiling in HeLa, HBE, A549, and H1299 cell lines as a reference[69] (**GSE46613**). Both analyses were already provided at gene level. The methodological details were described in the methods sections of the respective studies.

Sequestration-related proxies

PBs and SGs mRNA enrichment were collected at gene level from two pioneer studies[70,71] (**E-MTAB-5477; GSE99304**). For these two types of granules, it exists now a multitude of studies characterizing their specific mRNA via diverse methodology. For PBs mRNA enrichment, several studies of Fluorescence-Activated Particle Sorting FAPS using LSM14aGFP transformed cells have been reported in HEK293 cells[70,96] and HeLa cells[165] (**E-MTAB-5477**; **E-MTAB-12923; HRA002519**). PBs mRNA enrichment was also performed via granules purification with differential centrifugation followed by Immunoprecipitation targeting specific markers like EDC3 in U2OS cells[137] (**GSE138988**) or DCP1A in lymphocytes[138] (**GSE197001**). For SGs mRNA enrichment, several studies of granules purification with differential centrifugation followed by Immunoprecipitation of G3BP1 have been reported in U2OS[71,137] (**GSE99304**; **GSE138988**), 293T[166] (**GSE223295**), Lymphocytes[138] (**GSE197001**). We used the pioneer studies as references because the cross analyses of all those experiments performed in different cell types shows a very good conservation of the mRNA sequestrated (Supplementary Information).

Protein half-life related proxy

Data from a multiplexed pulsed SILAC-TMT dataset[88] were selected as proxy of protein half-lives. Methodological details are described in the study

from which these data were obtained. Analyses were already provided at gene level. This dataset was chosen because the authors showed a good correlation between their dataset and previously published half-lives in three different models: HCA2[167], A549[168], HeLa[169,170]. In addition, we used values of protein half-lives in hours of this same dataset for our representation of relative frequency in Fig.5e because this dataset is representative of all other protein half-life datasets.

Multivariate linear model for predicting protein-to-RNA ratios (PTRs) (interpretable model)

To investigate the contribution of multiple proxies to PTR variation and to construct an interpretable predictive model, a multivariate linear regression framework was implemented with strict control for feature redundancy, generalizability, and robustness.

Model formulation. To predict tissue-independent effects of the different proxies, we employed a multivariate linear model defined as follows:

$$y_{ij} = \beta_j^0 + x_i^T \beta + \varepsilon_{ij}$$

where y_{ij} represents the $\log_{10}$-transformed tissue-specific PTR ratio of gene i in tissue j (e.g., CD34+ and AML FAB subtypes), $\boldsymbol{x}_i^T$ is the vector of proxy values for that gene, β is the vector of regression coefficients, β_j^0 is the intercept varying by tissue, and ε_{ij} is the residual error term.

Proxy selection and dimensionality reduction. The initial feature set included 103 continuous proxies: 94 intrinsic (e.g., codon usage, amino acid composition, UTR/CDS length, folding energy, GC content, CAI, exon-junction density, miRNA site density) and 9 extrinsic (e.g., TE, MRL, TR, EVI, RNA and protein half-lives, P-body and stress granule enrichment). To reduce dimensionality and improve interpretability, LASSO regression with 10-fold cross-validation was applied on 80% of the dataset (training set). This procedure retained 48 non-zero coefficients, corresponding to the most informative and independent features. The list of selected proxies and their coefficients is provided in Supplementary Information.

Multicollinearity diagnostics. Multicollinearity among the 48 selected features was assessed by computing Variance Inflation Factors (VIFs) and generating a correlation matrix with heatmap visualization. All retained features exhibited VIF values < 5, indicating acceptable collinearity. Pairwise correlations were low to moderate (Supplementary Information).

Model evaluation and validation strategy. An 80/20 train-test split was used to evaluate model generalizability. Model training and cross-validation were performed on the training set (80%) ($n = 2644$ genes), and performance was assessed on an independent hold-out test set (20%) ($n = 662$ genes), which was excluded from feature selection and training.

All proxy variables were standardized (z-score transformation) using the mean and standard deviation from the training set. These statistics were subsequently applied to scale the test set to prevent data leakage. Model performance was assessed using the following metrics: Coefficient of determination (R^2) on the test set: 0.365; Spearman correlation between predicted and observed PTRs: $\rho = 0.623$; Root Mean Squared Error (RMSE): 0.547. This RMSE corresponds to a median relative prediction error of approximately 3.5-fold, which is small compared to the biological variability of PTRs (~63.5-fold across the central 80% of genes).

Residual diagnostics. To assess the validity of linear model assumptions, residual diagnostics were performed and showed that residual distributions were centered and symmetric with no indication of heteroscedasticity, and Q-Q plots and normality tests (Shapiro–Wilk for training set, Kolmogorov–Smirnov for test set) indicated only minor deviations (Supplementary Information).

Expanded applicability. Compared to the initial 103-proxy model, the simplified 48-proxy model enabled PTR prediction for a broader set of genes. Predictions were obtained for 4342 genes, including 1036 additional genes from the shadow proteome.

Software. All analyses were conducted using Python 3.11, with the scikit-learn (v1.4) and statsmodels (v0.14) libraries.

GSEA

GSEA was performed using fGSEA[171] (R version 1.27.0) in R with gene lists ranked by PTR values. RBP and miRNA gene sets were designed using ENCORI database see "miRNA related proxies" and "RBP related proxies" methods parts. 5′UTR, 3′UTR gene sets were designed using multiple sources see "Gene Motif and Structural Proxies Derived from Literature" methods part. "RNA Localisation" geneset was designed using multiple APEX seq datasets[72,73] and PBs and SGs mRNA enrichment datasets[70,71]. "Protein localization" geneset was designed using microscopy images classification of the HPA[89] and secreted proteins geneset were designed from a dedicated HPA analysis[90]. Statistical significance was calculated using the adaptive multilevel split Monte Carlo method.

UMAP dimensionality reduction

To project gene associated proxies in two dimensions we performed dimensionality reduction with the Uniform Manifold Approximation and Projection UMAP algorithm with the Embedding Projector standalone tool[172] using default parameters (except *n_neighbors* = 15, We used 19 proxies of the Fig.6b after a z-score transformation and subsequently used each proxy to color the 2D UMAP space.

GO term enrichment analysis

GO term enrichment analysis was performed with PANTHER powered Tool. The annotation data set used was GO biological process complete, the statistical test was the Fisher's Exact Test, followed by calculating the False Discovery Rate (FDR). Only terms with a FDR < 0.05 were included in the final results.

DDX3X and eIF4A selective targets from ribosome profiling experiments

A total of 6 ribosome profiling experiments in models targeting eIF4A or DDX3X expression or activity by different strategies (selective inhibitors[27,32] (**GSE56887**; **GSE151687**), siRNA[29] (**GSE125114 and GSE157063**), shRNA[28], or auxin inducible degron[30] (**GSE203078**) were compiled to annotate the PTRs and 5′UTR structure proxy of the various genes quantified. For each of them, we have used the data processed by the respective authors of these studies.

Proteome compilation, PTRs annotation, and meta-analysis

A total of 38 proteomic experiments performed in AML models were compiled to annotate the PTRs of the various genes quantified[6,22,94,95,173–207]. For each of them, we have used the data processed by the respective authors of these studies. These datasets, their models, and their PRIDE or MassIVE accession numbers are listed Supplementary Data 4.

Statistics and reproducibility

For graphical representation a curve of the median value of the proxy (y) computed for each PTR rank (x_n) was used: A sliding window extending from rank $n - 100$ to rank $n + 100$ was used to compute the median. Mann-Whitney tests between low PTRs (min to interquartile Q1) vs high PTR (interquartile Q3 to max) were performed for each proxy (row values of proxies were used and not the median for statistical tests).

For other analyses, comparison of continuous variables between two groups was performed using Mann–Whitney U-tests (unpaired) for non-parametric data or a t-tests for parametric data. A Shapiro–Wilk test was used to confirm normality. Two-sided tests were used as standard. To visualize both the significance and direction of changes, we used a signed

significance score, defined as the sign of FC (fold change) × – $\log_{10}$(FDR). For multiple comparisons Ordinary one-way ANOVA was performed. Sample sizes are provided in the main text, figure captions, or Supplementary Data 1. To study the relationship between RNA and protein levels, Spearman correlations were used due to the presence of outliers and the non-normal distribution of the transcriptomic and proteomic data[139]. However, the use of Pearson correlations showed the same results and led to the same conclusions. For the multiple linear regression, Data that did not meet test assumptions for homoscedasticity or normality of residuals were log-transformed. All length-related features were log transformed such that: $\hat{y} \leftarrow \log_{10}(y + 0.1)$ to reduce the right skew[68]. For fGSEA analysis, statistical significance was calculated using the adaptive multilevel split Monte Carlo method.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All datasets used are publicly available. Their respective accession number (GSE, HTA, MTAB, MassIVE, or PRIDE number) has been cited in the diverse parts of the methods section and in Supplementary Data 4. The analyzed data are provided in the Supplementary Data 1.

Code availability

Code and documentation are publicly available in GitHub. (https://github.com/POSTCODEadmin/POSTCODE). To facilitate its implementation by future users, a R shiny version of POSTCODE has been created.The exact version of the code used in the paper, which will remain unchanged, is permanently available on Zenodo (https://doi.org/10.5281/zenodo.17356461)[208].

Received: 2 April 2025; Accepted: 30 October 2025;
Published online: 17 December 2025

References

1. van Galen, P. et al. Single-cell RNA-Seq reveals AML hierarchies relevant to disease progression and immunity. *Cell* **176**, 1265–1281.e24 (2019).
2. Severens, J. F. et al. Mapping AML heterogeneity - multi-cohort transcriptomic analysis identifies novel clusters and divergent ex-vivo drug responses. *Leukemia* **38**, 751–761 (2024).
3. Sharon, D. et al. Inhibition of mitochondrial translation overcomes venetoclax resistance in AML through activation of the integrated stress response. *Sci. Transl. Med.* **11**, eaax2863 (2019).
4. Yamauchi, T. et al. Genome-wide CRISPR-Cas9 screen identifies leukemia-specific dependence on a Pre-mRNA metabolic pathway regulated by DCPS. *Cancer Cell* **33**, 386–400.e5 (2018).
5. Biancon, G. et al. Precision analysis of mutant U2AF1 activity reveals deployment of stress granules in myeloid malignancies. *Mol. Cell* **82**, 1107–1122.e7 (2022).
6. Kodali, S. et al. RNA sequestration in P-bodies sustains myeloid leukaemia. *Nat. Cell Biol*. 1–14 https://doi.org/10.1038/s41556-024-01489-6 (2024).
7. Wang, E. et al. Targeting an RNA-binding protein network in acute myeloid leukemia. *Cancer Cell* **35**, 369–384.e7 (2019).
8. Barbieri, I. et al. Promoter-bound METTL3 maintains myeloid leukaemia by m6A-dependent translation control. *Nature* **552**, 126–131 (2017).
9. Vu, L. P. et al. The N6-methyladenosine (m6A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat. Med.* **23**, 1369–1376 (2017).
10. Paris, J. et al. Targeting the RNA m6A reader YTHDF2 selectively compromises cancer stem cells in acute myeloid leukemia. *Cell Stem Cell* **25**, 137–148.e6 (2019).
11. Elcheva, I. A. & Spiegelman, V. S. Targeting RNA-binding proteins in acute and chronic leukemia. *Leukemia* **35**, 360–376 (2021).
12. Naudin, C. et al. PUMILIO/FOXP1 signaling drives expansion of hematopoietic stem/progenitor and leukemia cells. *Blood* **129**, 2493–2506 (2017).
13. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
14. Clark, D. J. et al. Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* **179**, 964–983.e31 (2019).
15. Vasaikar, S. et al. Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* **177**, 1035–1049.e19 (2019).
16. Gautier, E.-F. et al. Comprehensive proteomic analysis of murine terminal erythroid differentiation. *Blood Adv.* **4**, 1464–1477 (2020).
17. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
18. Franks, A., Airoldi, E. & Slavov, N. Post-transcriptional regulation across human tissues. *PLoS Comput. Biol.* **13**, e1005535 (2017).
19. Eraslan, B. et al. Quantification and discovery of sequence determinants of protein-per-mRNA amount in 29 human tissues. *Mol. Syst. Biol.* **15**, e8513 (2019).
20. Giansanti, P. et al. Mass spectrometry-based draft of the mouse proteome. *Nat. Methods* **19**, 803–811 (2022).
21. Vogel, C. et al. Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.* **6**, 400 (2010).
22. Kramer, M. H. et al. Proteomic and phosphoproteomic landscapes of acute myeloid leukemia. *Blood* **140**, 1533–1548 (2022).
23. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A. 'proteomic ruler' for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteom. MCP* **13**, 3497–3506 (2014).
24. Wan, Y. et al. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**, 706–709 (2014).
25. Wayment-Steele, H. K. et al. RNA secondary structure packages evaluated and improved by high-throughput experiments. *Nat. Methods* **19**, 1234–1242 (2022).
26. Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).
27. Wolfe, A. L. et al. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).
28. Modelska, A. et al. The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape. *Cell Death Dis.* **6**, e1603–e1603 (2015).
29. Calviello, L. et al. DDX3 depletion represses translation of mRNAs with complex 5' UTRs. *Nucleic Acids Res* **49**, 5336–5350 (2021).
30. Jowhar, Z. et al. A ubiquitous GC content signature underlies multimodal mRNA regulation by DDX3X. *Mol. Syst. Biol.* **20**, 276–290 (2024).
31. Kito, Y. et al. The ASC-1 complex promotes translation initiation by scanning ribosomes. *EMBO J.* **42**, e112869 (2023).
32. Steinberger, J. et al. Identification and characterization of hippuristanol-resistant mutants reveals eIF4A1 dependencies within mRNA 5′ leader regions. *Nucleic Acids Res.* **48**, 9521–9537 (2020).
33. Lisi-Vega, L. E. et al. Bone marrow mesenchymal stromal cells support translation in refractory acute myeloid leukemia. *Cell Rep*. **44**, 115151 (2025).
34. Bourdon, S. et al. QUADRatlas: the RNA G-quadruplex and RG4-binding proteins database. *Nucleic Acids Res.* **51**, D240–D247 (2023).
35. Martinez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J. & Embarek, A. M. Insights into structural and mechanistic features of viral IRES elements. *Front. Microbiol.* **8**, 2629 (2017).
36. Lim, C. S., Wardell, T., Kleffmann, S. J. & Brown, T. C. M. The exon-intron gene structure upstream of the initiation codon predicts translation efficiency. *Nucleic Acids Res.* **46**, 4575–4591 (2018).

37. Hsieh, A. C. et al. The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61 (2012).
38. Thoreen, C. C. et al. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).
39. Weingarten-Gabbay, S. et al. Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**, aad4939 (2016).
40. Hinnebusch, A. G., Ivanov, I. P. & Sonenberg, N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413–1416 (2016).
41. Sample, P. J. et al. Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.* **37**, 803–809 (2019).
42. 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* **324**, 218–223 (2009).
43. Ciandrini, L., Stansfield, I. & Romano, M. C. Ribosome traffic on mRNAs maps to gene ontology: genome-wide quantification of translation initiation rates and polysome size regulation. *PLoS Comput. Biol.* **9**, e1002866 (2013).
44. Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J. B. Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589–1601 (2013).
45. Weinberg, D. E. et al. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.* **14**, 1787–1799 (2016).
46. Duc, K. D. & Song, Y. S. The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. *PLoS Genet.* **14**, e1007166 (2018).
47. Khajuria, R. K. et al. Ribosome levels selectively regulate translation and lineage commitment in human hematopoiesis. *Cell* **173**, 90–103.e19 (2018).
48. Boussaid, I. et al. Integrated analyses of translatome and proteome identify the rules of translation selectivity in RPS14-deficient cells. *Haematologica* **106**, 746–758 (2021).
49. Sin, C., Chiarugi, D. & Valleriani, A. Quantitative assessment of ribosome drop-off in E. coli. *Nucleic Acids Res.* **44**, 2528–2537 (2016).
50. Bonnin, P., Kern, N., Young, N. T., Stansfield, I. & Romano, M. C. Novel mRNA-specific effects of ribosome drop-off on translation rate and polysome profile. *PLoS Comput. Biol.* **13**, e1005555 (2017).
51. Fernandes, L. D., Moura, A. P. S. de & Ciandrini, L. Gene length as a regulator for ribosome recruitment and protein synthesis: theoretical insights. *Sci. Rep.* **7**, 17409 (2017).
52. Madern, M. F. et al. Long-term imaging of individual ribosomes reveals ribosome cooperativity in mRNA translation. *Cell* **188**, 1896–1911.e24 (2025).
53. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
54. Diaz de Arce, A. J., Noderer, W. L. & Wang, C. L. Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. *Nucleic Acids Res.* **46**, 985–994 (2018).
55. Kameda, T., Asano, K. & Togashi, Y. Free energy landscape of RNA binding dynamics in start codon recognition by eukaryotic ribosomal pre-initiation complex. *PLoS Comput. Biol.* **17**, e1009068 (2021).
56. Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **87**, 8301–8305 (1990).
57. Fischer, J. W., Busa, V. F., Shao, Y. & Leung, A. K. L. Structure-mediated RNA decay by UPF1 and G3BP1. *Mol. Cell* **78**, 70–84.e6 (2020).
58. Hia, F. et al. Codon bias confers stability to human mRNAs. *EMBO Rep.* **20**, e48220 (2019).
59. Puigbò, P., Bravo, I. G. & Garcia-Vallve, S. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol. Direct* **3**, 38 (2008).
60. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421–433 (2015).
61. Wu, Q. et al. Translation affects mRNA stability in a codon-dependent manner in human cells. *eLife* **8**, e45396 (2019).
62. Buschauer, R. et al. The Ccr4-Not complex monitors the translating ribosome for codon optimality. *Science* **368**, eaay6912 (2020).
63. Presnyak, V. et al. Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124 (2015).
64. Bazzini, A. A. et al. Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J.* **35**, 2087–2103 (2016).
65. Dana, A. & Tuller, T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res* **42**, 9171–9181 (2014).
66. Forrest, M. E. et al. Codon and amino acid content are associated with mRNA stability in mammalian cells. *PLoS ONE* **15**, e0228730 (2020).
67. Behrens, A., Rodschinka, G. & Nedialkova, D. D. High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. *Mol. Cell* **81**, 1802–1815.e7 (2021).
68. Agarwal, V. & Kelley, D. R. The genetic and biochemical determinants of mRNA degradation rates in mammals. *Genome Biol.* **23**, 245 (2022).
69. Lian, X. et al. Genome-wide and experimental resolution of relative translation elongation speed at individual gene level in human cells. *PLoS Genet.* **12**, e1005901 (2016).
70. Hubstenberger, A. et al. P-body purification reveals the condensation of repressed mRNA regulons. *Mol. Cell* **68**, 144–157.e5 (2017).
71. Khong, A. et al. The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol. Cell* **68**, 808–820.e5 (2017).
72. Fazal, F. M. et al. Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell* **178**, 473–490.e26 (2019).
73. Padrón, A., Iwasaki, S. & Ingolia, N. T. Proximity RNA labeling by APEX-Seq reveals the organization of translation initiation complexes and repressive RNA granules. *Mol. Cell* **75**, 875–887.e5 (2019).
74. Li, J.-H., Liu, S., Zhou, H., Qu, L.-H. & Yang, J.-H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **42**, D92–D97 (2014).
75. Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
76. Giudice, G., Sánchez-Cabo, F., Torroja, C. & Lara-Pezzi, E. ATtRACT—a database of RNA-binding proteins and associated motifs. *Database* **2016**, baw035 (2016).
77. Xuan, J.-J. et al. RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. *Nucleic Acids Res.* **46**, D327–D334 (2018).
78. Tang, Y. et al. m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. *Nucleic Acids Res.* **49**, D134–D143 (2021).
79. Dai, Q. et al. Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat. Methods* **14**, 695–698 (2017).
80. Lin, S., Choe, J., Du, P., Triboulet, R. & Gregory, R. I. The m(6)A methyltransferase METTL3 promotes translation in human cancer cells. *Mol. Cell* **62**, 335–345 (2016).
81. Arango, D. et al. Acetylation of cytidine in mrna promotes translation efficiency. *Cell* **175**, 1872–1886.e24 (2018).
82. Choe, J. et al. mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis. *Nature* **561**, 556–560 (2018).
83. Yankova, E. et al. Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature* **593**, 597–601 (2021).
84. Kuppers, D. A. et al. N 6 -methyladenosine mRNA marking promotes selective translation of regulons required for human erythropoiesis. *Nat. Commun.* **10**, 4596 (2019).

85. van Heesch, S. et al. The translational landscape of the human heart. *Cell* **178**, 242–260.e29 (2019).
86. Wu, Q. et al. Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J*. https://doi.org/10.15252/embj.2020104763 (2020).
87. Chothani, S. P. et al. A high-resolution map of human RNA translation. *Mol. Cell* **82**, 2885–2899.e8 (2022).
88. Zecha, J. et al. Peptide level turnover measurements enable the study of proteoform dynamics. *Mol. Cell. Proteom. MCP* **17**, 974–992 (2018).
89. Ouyang, W. et al. Analysis of the human protein atlas image classification competition. *Nat. Methods* **16**, 1254–1261 (2019).
90. Uhlén, M. et al. The human secretome. *Sci. Signal.* **12**, eaaz0274 (2019).
91. Jiang, L. et al. A quantitative proteome map of the human body. *Cell* **183**, 269–283.e19 (2020).
92. Kim, M.-S. et al. A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
93. Wang, D. et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).
94. Janssen, M. et al. Venetoclax synergizes with gilteritinib in FLT3 wild-type high-risk acute myeloid leukemia by suppressing MCL-1. *Blood* **140**, 2594–2610 (2022).
95. Sabatier, M. et al. C/EBPα confers dependence to fatty acid anabolic pathways and vulnerability to lipid oxidative stress–induced ferroptosis in FLT3-mutant leukemia. *Cancer Discov.* **13**, 1720–1747 (2023).
96. Safieddine, A. et al. Cell-cycle-dependent mRNA localization in P-bodies. *Mol. Cell* **84**, 4191–4208.e7 (2024).
97. Swartzel, J. C. et al. Targeted degradation of mRNA decapping enzyme DcpS by a VHL-recruiting PROTAC. *ACS Chem. Biol.* **17**, 1789–1798 (2022).
98. Meziane, O. et al. The human decapping scavenger enzyme DcpS modulates microRNA turnover. *Sci. Rep.* **5**, 16688 (2015).
99. Zhu, S. et al. Restoration of microRNA function impairs MYC-dependent maintenance of MLL leukemia. *Leukemia* **34**, 2484–2488 (2020).
100. Leppek, K., Das, R. & Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* **19**, 158–174 (2018).
101. Jia, L. et al. Decoding mRNA translatability and stability from the 5′ UTR. *Nat. Struct. Mol. Biol*. 1–8 https://doi.org/10.1038/s41594-020-0465-x (2020).
102. Johnstone, T. G., Bazzini, A. A. & Giraldez, A. J. Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.* **35**, 706–723 (2016).
103. Zheng, D. et al. Predicting the translation efficiency of messenger RNA in mammalian cells. *Nat. Biotechnol.* 1–14 https://doi.org/10.1038/s41587-025-02712-x (2025).
104. Murakami, S. et al. m6A alters ribosome dynamics to initiate mRNA degradation. *Cell* **0**, 3728–3743 (2025).
105. Linder, B. et al. tRNA modifications tune m6A-dependent mRNA decay. *Cell* **0**, 13 (2025).
106. Uzonyi, A. et al. Exclusion of m6A from splice-site proximal regions by the exon junction complex dictates m6A topologies and mRNA stability. *Mol. Cell* **83**, 237–251.e7 (2023).
107. Yang, X., Triboulet, R., Liu, Q., Sendinc, E. & Gregory, R. I. Exon junction complex shapes the m6A epitranscriptome. *Nat. Commun.* **13**, 7904 (2022).
108. Li, Q. et al. Decoding the interplay between m6A modification and stress granule stability by live-cell imaging. *Sci. Adv.* **10**, eadp5689 (2024).
109. Zhou, Y. et al. m6A sites in the coding region trigger translation-dependent mRNA decay. *Mol. Cell* **84**, 4576–4593.e12 (2024).
110. Schumann, U. et al. Multiple links between 5-methylcytosine content of mRNA and translation. *BMC Biol.* **18**, 40 (2020).
111. Li, Y. et al. 2'-O-methylation at internal sites on mRNA promotes mRNA stability. *Mol. Cell* **84**, 2320–2336.e6 (2024).
112. Cenik, C. et al. Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res.* **25**, 1610–1621 (2015).
113. Kearse, M. G. & Wilusz, J. E. Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.* **31**, 1717–1731 (2017).
114. Boeck, R. & Kolakofsky, D. Positions +5 and +6 can be major determinants of the efficiency of non-AUG initiation codons for protein synthesis. *EMBO J.* **13**, 3608–3617 (1994).
115. Grünert, S. & Jackson, R. J. The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons. *EMBO J.* **13**, 3618–3630 (1994).
116. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508**, 66–71 (2014).
117. Xiang, K. & Bartel, D. P. The molecular basis of coupling between poly(A)-tail length and translational efficiency. *eLife* **10**, e66493 (2021).
118. Qian, W., Yang, J.-R., Pearson, N. M., Maclean, C. & Zhang, J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* **8**, e1002603 (2012).
119. Riba, A. et al. Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc. Natl. Acad. Sci. USA* **116**, 15023–15032 (2019).
120. Wang, J. et al. Rapid 40S scanning and its regulation by mRNA structure during eukaryotic translation initiation. *Cell* **185**, 4474–4487.e17 (2022).
121. Lackner, D. H. et al. A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol. Cell* **26**, 145–155 (2007).
122. Guo, J., Lian, X., Zhong, J., Wang, T. & Zhang, G. Length-dependent translation initiation benefits the functional proteome of human cells. *Mol. BioSyst.* **11**, 370–378 (2015).
123. Jeacock, L., Faria, J. & Horn, D. Codon usage bias controls mRNA and protein abundance in trypanosomatids. *eLife* **7**, e32496 (2018).
124. Amrani, N., Ghosh, S., Mangus, D. A. & Jacobson, A. Translation factors promote the formation of two states of the closed-loop mRNP. *Nature* **453**, 1276–1280 (2008).
125. Costello, J. et al. Global mRNA selection mechanisms for translation initiation. *Genome Biol.* **16**, 10 (2015).
126. Thompson, M. K. & Gilbert, W. V. mRNA length-sensing in eukaryotic translation: reconsidering the 'closed loop' and its implications for translational control. *Curr. Genet*. https://doi.org/10.1007/s00294-016-0674-3 (2016).
127. Vicens, Q., Kieft, J. S. & Rissland, O. S. Revisiting the closed-loop model and the nature of mRNA 5′–3′ communication. *Mol. Cell* **72**, 805–812 (2018).
128. Fakim, H. & Fabian, M. R. Communication is key: 5'-3' interactions that regulate mRNA translation and turnover. *Adv. Exp. Med. Biol.* **1203**, 149–164 (2019).
129. Çetin, B. & O'Leary, S. E. mRNA- and factor-driven dynamic variability controls eIF4F-cap recognition for translation initiation. *Nucleic Acids Res* **50**, 8240–8261 (2022).
130. Rogers, D. W., Böttcher, M. A., Traulsen, A. & Greig, D. Ribosome reinitiation can explain length-dependent translation of messenger RNA. *PLoS Comput. Biol.* **13**, e1005592 (2017).
131. Alekhina, O. M., Terenin, I. M., Dmitriev, S. E. & Vassilenko, K. S. Functional cyclization of eukaryotic mRNAs. *Int. J. Mol. Sci*. **21**, 1677 (2020).
132. Barrington, C. L. et al. Synonymous codon usage regulates translation initiation. *Cell Rep.* **42**, 113413 (2023).
133. Lyons, E. F. et al. Codon optimality modulates protein output by tuning translation initiation. *bioRxiv* 2023.11.27.568910 https://doi.org/10.1101/2023.11.27.568910 (2023).

134. Bicknell, A. A. et al. Attenuating ribosome load improves protein output from mRNA by limiting translation-dependent mRNA decay. *Cell Rep.* **43**, 114098 (2024).
135. Tanguay, R. L. & Gallie, D. R. Translational efficiency is regulated by the length of the 3' untranslated region. *Mol. Cell. Biol.* **16**, 146–156 (1996).
136. Liu, Q. et al. Integrative omics analysis reveals the importance and scope of translational repression in microRNA-mediated regulation*. *Mol. Cell. Proteom.* **12**, 1900–1911 (2013).
137. Matheny, T., Rao, B. S. & Parker, R. Transcriptome-wide comparison of stress granules and P-bodies reveals that translation plays a major role in RNA partitioning. *Mol. Cell. Biol.* **39**, e00313–e00319 (2019).
138. Curdy, N. et al. The proteome and transcriptome of stress granules and P bodies during human T lymphocyte activation. *Cell Rep.* **42**, 112211 (2023).
139. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583**, 3966–3973 (2009).
140. Cunningham, F. et al. Ensembl 2022. *Nucleic Acids Res* **50**, D988–D995 (2022).
141. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
142. Zhao, J. et al. IRESbase: a comprehensive database of experimentally validated internal ribosome entry sites. *Genomics Proteom. Bioinforma.* **18**, 129–139 (2020).
143. Mokrejs, M. et al. IRESite-a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res.* **38**, D131–D136 (2010).
144. Yoon, A. et al. Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. *Science* **312**, 902–906 (2006).
145. Calviello, L. et al. Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods* **13**, 165–170 (2016).
146. Zhang, P. et al. Genome-wide identification and differential analysis of translational initiation. *Nat. Commun.* **8**, 1749 (2017).
147. Erhard, F. et al. Improved Ribo-seq enables identification of cryptic translation events. *Nat. Methods* **15**, 363–366 (2018).
148. Paz, I., Kosti, I., Manuel Ares, J., Cline, M. & Mandel-Gutfreund, Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* **42**, W361 (2014).
149. Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
150. Chen, Y. & Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **48**, D127–D131 (2020).
151. Liu, W. & Wang, X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol.* **20**, 18 (2019).
152. McGeary, S. E. et al. The biochemical basis of microRNA targeting efficacy. *Science* **366**, eaav1741 (2019).
153. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, e05005 (2015).
154. Legnini, I., Alles, J., Karaiskos, N., Ayoub, S. & Rajewsky, N. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nat. Methods* **16**, 879–886 (2019).
155. Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
156. Begik, O. et al. Nano3P-seq: transcriptome-wide analysis of gene expression and tail dynamics using end-capture nanopore cDNA sequencing. *Nat. Methods* **20**, 75–85 (2023).
157. Maier, K. C., Gressel, S., Cramer, P. & Schwalb, B. Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Res.* **30**, 1332 (2020).
158. 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* **53**, 1044–1052 (2014).
159. Yu, F. et al. Poly(A)-seq: a method for direct sequencing and analysis of the transcriptomic poly(A)-tails. *PLoS ONE* **15**, e0234696 (2020).
160. Sharp, P. M., Tuohy, T. M. & Mosurski, K. R. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* **14**, 5125–5143 (1986).
161. Sharp, P. M. & Li, W. H. The codon Adaptation Index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
162. dos Reis, M., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036–5044 (2004).
163. Zheng, G. et al. Efficient and quantitative high-throughput tRNA sequencing. *Nat. Methods* **12**, 835–837 (2015).
164. Narula, A., Ellis, J., Taliaferro, J. M. & Rissland, O. S. Coding regions affect mRNA stability in human cells. *RNA* 073239.119 https://doi.org/10.1261/rna.073239.119 (2019).
165. Shan, T. et al. m6A modification negatively regulates translation by switching mRNA from polysome to P-body via IGF2BP3. *Mol. Cell* **83**, 4494–4508.e6 (2023).
166. Ren, Z., Tang, W., Peng, L. & Zou, P. Profiling stress-triggered RNA condensation with photocatalytic proximity labeling. *Nat. Commun.* **14**, 7390 (2023).
167. Welle, K. A. et al. Time-resolved analysis of proteome dynamics by tandem mass tags and stable isotope labeling in cell culture (TMT-SILAC) hyperplexing. *Mol. Cell. Proteom. MCP* **15**, 3551–3563 (2016).
168. 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.* **8**, 104–112 (2009).
169. Cambridge, S. B. et al. Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J. Proteome Res.* **10**, 5275–5284 (2011).
170. Boisvert, F.-M. et al. A quantitative spatial proteomics analysis of proteome turnover in human cells*. *Mol. Cell. Proteom.* **11**, M111.011429 (2012).
171. Korotkevich, G. et al. Fast gene set enrichment analysis. 060012 Preprint at https://doi.org/10.1101/060012 (2021).
172. Smilkov, D. et al. Embedding projector: interactive visualization and interpretation of embeddings. Preprint at https://doi.org/10.48550/arXiv.1611.05469 (2016).
173. Hansen, J. D. et al. CC-90009: a cereblon E3 ligase modulating drug that promotes selective degradation of GSPT1 for the treatment of acute myeloid leukemia. *J. Med. Chem.* **64**, 1835–1843 (2021).
174. Larrue, C. et al. Targeting ferritinophagy impairs quiescent cancer stem cells in acute myeloid leukemia in vitro and in vivo models. *Sci. Transl. Med.* **16**, eadk1731 (2024).
175. Sellar, R. S. et al. Degradation of GSPT1 causes TP53-independent cell death in leukemia while sparing normal hematopoietic stem cells. *J. Clin. Investig*. **132**, e153514 (2022).
176. Park, S.-M. et al. Dual IKZF2 and CK1α degrader targets acute myeloid leukemia cells. *Cancer Cell* **41**, 726–739.e11 (2023).
177. Grey, W. et al. CKS1 inhibition depletes leukemic stem cells and protects healthy hematopoietic stem cells in acute myeloid leukemia. *Sci. Transl. Med.* **14**, eabn3248 (2022).
178. de Boer, B. et al. Prospective isolation and characterization of genetically and functionally distinct AML subclones. *Cancer Cell* **34**, 674–689.e8 (2018).
179. Hernandez-Valladares, M., Aasebø, E., Berven, F., Selheim, F. & Bruserud, Ø Biological characteristics of aging in human acute myeloid leukemia cells: the possible importance of aldehyde dehydrogenase, the cytoskeleton and altered transcriptional regulation. *Aging* **12**, 24734–24777 (2020).
180. Schmidt, J. R. et al. Pilot study on mass spectrometry-based analysis of the proteome of $CD34^{+}CD123^{+}$ progenitor cells for the identification of potential targets for immunotherapy in acute myeloid leukemia. *Proteomes* **6**, 11 (2018).

181. Zhou, F. et al. A dynamic rRNA ribomethylome drives stemness in acute myeloid leukemia. *Cancer Discov.* **13**, 332–347 (2023).
182. Pauli, C. et al. Site-specific methylation of 18S ribosomal RNA by SNORD42A is required for acute myeloid leukemia cell proliferation. *Blood* **135**, 2059–2070 (2020).
183. Casado, P. et al. Integrative phosphoproteomics defines two biologically distinct groups of KMT2A rearranged acute myeloid leukaemia with different drug response phenotypes. *Signal Transduct. Target. Ther.* **8**, 80 (2023).
184. Emdal, K. B. et al. Phosphoproteomics of primary AML patient samples reveals rationale for AKT combination therapy and p53 context to overcome selinexor resistance. *Cell Rep.* **40**, 111177 (2022).
185. Jayavelu, A. K. et al. The proteogenomic subtypes of acute myeloid leukemia. *Cancer Cell* **40**, 301–317.e12 (2022).
186. Stratmann, S. et al. Proteogenomic analysis of acute myeloid leukemia associates relapsed disease with reprogrammed energy metabolism both in adults and children. *Leukemia* **37**, 550–559 (2023).
187. Aasebø, E. et al. The progression of acute myeloid leukemia from first diagnosis to chemoresistant relapse: a comparison of proteomic and phosphoproteomic profiles. *Cancers* **12**, 1466 (2020).
188. Morales, M. L. et al. Posttranslational splicing modifications as a key mechanism in cytarabine resistance in acute myeloid leukemia. *Leukemia* **37**, 1649–1659 (2023).
189. Cela, I. et al. Proteomic investigation of the role of nucleostemin in nucleophosmin-mutated OCI-AML 3 cell line. *Int. J. Mol. Sci.* **23**, 7655 (2022).
190. Aumer, T. et al. Comprehensive comparison between azacytidine and decitabine treatment in an acute myeloid leukemia cell line. *Clin. Epigenet.* **14**, 113 (2022).
191. Raffel, S. et al. Quantitative proteomics reveals specific metabolic features of acute myeloid leukemia stem cells. *Blood* **136**, 1507–1519 (2020).
192. Pino, J. C. et al. Mapping the proteogenomic landscape enables prediction of drug response in acute myeloid leukemia. *Cell Rep. Med.* **5**, 101359 (2024).
193. Pedicona, F. et al. Targeting the lysine-specific demethylase 1 rewires kinase networks and primes leukemia cells for kinase inhibitor treatment. *Sci. Signal.* **15**, eabl7989 (2022).
194. Matondo, M. et al. Determination of differentially regulated proteins upon proteasome inhibition in AML cell lines by the combination of large-scale and targeted quantitative proteomics. *Proteomics* **17**, 1600089 (2017).
195. Chen, Y. et al. Succinate dehydrogenase deficiency-driven succinate accumulation induces drug resistance in acute myeloid leukemia via ubiquitin-cullin regulation. *Nat. Commun.* **15**, 9820 (2024).
196. Kempf, J. M. et al. Loss-of-function mutations in the histone methyltransferase EZH2 promote chemotherapy resistance in AML. *Sci. Rep.* **11**, 5838 (2021).
197. Lechman, E. R. et al. miR-126 regulates distinct self-renewal outcomes in normal and malignant hematopoietic stem cells. *Cancer Cell* **29**, 602–606 (2016).
198. Xu, B., Zhao, Y., Wang, X., Gong, P. & Ge, W. MZH29 is a novel potent inhibitor that overcomes drug resistance FLT3 mutations in acute myeloid leukemia. *Leukemia* **31**, 913–921 (2017).
199. Pfeiffer, A. et al. Phosphorylation of SHP2 at Tyr62 enables acquired resistance to SHP2 allosteric inhibitors in FLT3-ITD-driven AML. *Cancer Res.* **82**, 2141–2155 (2022).
200. Hernandez-Valladares, M. et al. Proteomic studies of primary acute myeloid leukemia cells derived from patients before and during disease-stabilizing treatment based on all-trans retinoic acid and valproic acid. *Cancers* **13**, 2143 (2021).
201. Schnoeder, T. M. et al. PLCG1 is required for AML1-ETO leukemia stem cell self-renewal. *Blood* **139**, 1080–1097 (2022).
202. Perner, F. et al. YBX1 mediates translation of oncogenic transcripts to control cell competition in AML. *Leukemia* **36**, 426–437 (2022).
203. Wang, L. et al. Discovery of a first-in-class CDK2 selective degrader for AML differentiation therapy. *Nat. Chem. Biol.* **17**, 567–575 (2021).
204. Casado, P. et al. Proteomic and genomic integration identifies kinase and differentiation determinants of kinase inhibitor sensitivity in leukemia cells. *Leukemia* **32**, 1818–1822 (2018).
205. Hospital, M.-A. et al. RSK2 is a new Pim2 target with pro-survival functions in FLT3-ITD-positive acute myeloid leukemia. *Leukemia* **32**, 597–605 (2018).
206. Liu, L. et al. The splicing factor RBM17 drives leukemic stem cell maintenance by evading nonsense-mediated decay of pro-leukemic factors. *Nat. Commun.* **13**, 3833 (2022).
207. Ctortecka, C. et al. Automated single-cell proteomics providing sufficient proteome depth to study complex biology beyond cell type classifications. *Nat. Commun.* **15**, 5707 (2024).
208. Boussaid, Ismael. Rshiny Code used for 'Integrated analysis of post-transcriptional regulations reveals insights into acute myeloid leukemia'. Zenodo https://doi.org/10.5281/zenodo.17356461 (2025).

Acknowledgements

This work has been funded by grants from the Integrated Cancer Research Site for Personalized Medicine SIRIC CARPEM: INCa-DGOS-Inserm-ITMO Cancer_18006 and was supported by the France 2030 program through the Idex Université Paris Cité (ANR-18-IDEX-0001). We would like to thank all the authors contacted to obtain the processed proteomic analysis data or information on these analyses. We would like to acknowledge Dr Franck Letourneur and Benjamin Saint-Pierre from the Platform Genom'IC, Institut Cochin, Alexandre Houy from Institut Curie, Dr Daniel Finley from Harvard Medical School (HMS), and Dr Daniel R. Larson from National Cancer Institute, NIH for discussions that fed this project.

Author contributions

Conceptualization: I.B. Data curation: I.B., E.K., Z.I., M.L., G.C., F.L., and E.F.G. Formal analysis: I.B., E.K., Z.I., M.L., G.C., F.L., and E.F.G. Funding acquisition: I.B. Investigation and validation: I.B., M.L., E.K., G.C., F.L, and E.L, Methodology: I.B. Resources: R.B., J.T., and C.L. Software: I.B., E.K., Z.I., M.L., G.C., F.L., D.R., and C.D. Feedback and tests: Z.T., J.D., C.F., R.V., and M.T. Supervision and Project administration: I.B. Visualization: I.B. Writing—original draft: I.B. Writing— review and editing: I.B., E.F.G., R.B., C.D., M.F., E.L., O.K., N.C., Y.S., and D.B.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-09156-8.

Correspondence and requests for materials should be addressed to Ismael Boussaid.

Peer review information *Communications Biology* thanks Emiliano P. Ricci, Sheng-Yan Lin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Silvia Belluti, Christina Karlsson Rosenthal and Johannes Stortz. [A peer review file is available].

[1]Université Paris Cité, Centre National de la Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM), Institut Cochin, Paris, France. [2]Proteom'IC facility, Université Paris Cité, CNRS, INSERM, Institut Cochin, Paris, France. [3]Équipe labellisée par la Fondation pour la Recherche Médicale, Paris, France. [4]Leukemia Institute Paris Saint-Louis, Paris, France. [5]Division of Experimental Hematology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [6]Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM), Heidelberg, Germany. [7]Molecular Systems Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [8]Assistance Publique-Hôpitaux de Paris.Centre-Université Paris Cité, Hôpital Cochin, Laboratory of Hematology, Paris, France. [9]Assistance Publique-Hôpitaux de Paris.Centre-Université Paris Cité, Hôpital Cochin, Clinical Department of Hematology, Paris, France. [10]Translational Research Centre in Onco-Hematology, Faculty of Medicine, University of Geneva, Swiss Cancer Center Leman, Geneva, Switzerland. [11]Centre de Recherches en Cancérologie de Toulouse, Université de Toulouse INSERM U1037, Toulouse, France. [12]Équipe labellisée Ligue Nationale Contre le Cancer 2023, Toulouse, France. [13]Oncology Department, Geneva University Hospitals, Geneva, Switzerland. [14]Laboratoire d'excellence du Globule Rouge GR-Ex, Université Paris Cité, Paris, France. [15]These authors contributed equally: Elissa Khadra, Zoaila Iqbal. ✉e-mail: ismael.boussaid@aphp.fr