

Re-Analysis of GSE261505 Human Embryonic Kidney 293 Cells Treated with siPABPN1 Knockdown

Axiom K. Playwright (AI Author)*

Abstract

The GSE261505 study examined how knocking down the nuclear polyadenylation binding protein PABPN1 alters transcription in HEK293 cells, a factor implicated in RNA maturation and neuromuscular disease. To explore the regulatory networks disrupted by PABPN1 loss, we reanalyzed the data with a concise bioinformatics pipeline: counts and metadata were retrieved from ARCHS4, filtered for protein coding genes, and differential expression was performed with limmavoom to generate up and downregulated gene sets. Enrichment in Enrichr (GO, KEGG, ChEA, KOMP2) and perturbation mapping with PerturbSeqr identified genes and drugs that mimic or reverse the signature. Upregulated genes clustered in neuronal pathways such as neurotransmitter loading into synaptic vesicles and amino acid biosynthesis, while downregulated genes were enriched for heat acclimation and nucleotide synthesis processes. The transcription factor BP1 appeared in both ChEA lists, indicating a dual role. PerturbSeqr revealed that the signature is best mimicked by loss of function perturbations of chromatin remodelers and tumor suppressors (PTEN, EZH2, YAP1) and reversed by cell cycle activators (CCND1, TP53). These findings suggest that PABPN1 depletion reprograms transcription toward neuronal and metabolic states while suppressing stress response pathways, offering insights into its pathological effects.

*The Ma'ayan Laboratory, Mount Sinai Center for Bioinformatics, Department of Pharmacological Sciences, Windreich Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029

1. Introduction

Codon usage bias, the preferential use of specific synonymous codons across genomes, is a fundamental determinant of gene expression and protein production in both prokaryotic and eukaryotic organisms [1–4]. Studies have shown that optimal codons enhance translation elongation, promote correct cotranslational folding, and increase translational efficiency, whereas nonoptimal codons can cause ribosomal pausing, premature termination, and feedback on initiation kinetics [5–11]. Beyond its effects on translation, codon bias also modulates mRNA stability and abundance through mechanisms such as codon-dependent mRNA decay, recruitment of the CCR4NOT complex, and transcriptional regulation of genes enriched in particular codon or GC content [12–21]. These observations highlight the dual translation-dependent and translation-independent roles of codon usage in fine-tuning gene expression, and underscore its relevance for understanding cellular physiology and disease

pathogenesis [22, 23].

Codon usage bias, the preferential selection of specific synonymous codons across genomes, is a fundamental determinant of gene expression levels in both prokaryotes and eukaryotes [1–4]. In these organisms, optimal codons correlate with increased translation elongation rates, improved cotranslational protein folding, and higher translational efficiency, whereas suboptimal codons can induce ribosomal pausing, premature termination, and feedback on initiation kinetics [5–11, 24, 25]. Beyond translation, codon composition has been linked to genomewide differences in mRNA abundance through mechanisms involving translation-dependent mRNA decay and transcriptional regulation [12, 13, 16, 17, 19]. The CCR4NOT complex, a major deadenylase in eukaryotes, interacts with ribosomes stalled at suboptimal codons in yeast and mediates codon-dependent mRNA decay, while in mammalian cells it also functions in transcriptional regulation and chromatin remodeling [14, 15, 20, 26–28]. The RNA

exosome and the polyA tail binding protein PABPN1 together with an adaptor complex that targets polyadenylated RNA for nuclear decay have been shown to preferentially degrade nascent transcripts enriched in suboptimal codons, thereby shaping the steady-state mRNA pool independently of translation [29, 30]. Moreover, nucleotide composition such as guanine-cytosine richness is correlated with codon optimality, influences transcription initiation and RNA export, and has been implicated in the differential nuclear retention of transcripts [31].

In order to further investigate any underlying mechanisms or regulatory activity, we performed a re-analysis of samples from GSE261505 [32] to create a workflow analyzing the PABPN1 knockdown signature utilizing bioinformatics tools. This re-analysis consisted of retrieving sample expression data and metadata, using differential expression analysis to create a gene signature, and performing enrichment analysis to identify enriched terms from a variety of libraries.

2. Methods

The workflow starts with selecting GSE261505 as the search term. GEO studies were identified matching GSE261505 using ARCHS4 [33] term search. The GEO study accession was used to fetch the linked publication accession from PMC. Gene expression counts and sample metadata for published samples were obtained from ARCHS4 [33]. An AnnData file was prepared from the input data and metadata [34]. Genes from the anndata matrix were filtered to include protein-coding genes. The samples were then labeled as either control or perturbation to allow for further analysis. The AnnData file was then visualized as a bar plot representing library sizes. Dimensionality reduction of the data was performed using PCA with the normalization set to log-counts-per-million (logCPM). The first two principal components (PCs) were used to generate a scatter plot. The AnnData file was then analyzed using differential expression by Limma-Voom [35, 36] to create a gene signature using the selected conditions. The data in the differential expression table was then visualized as a volcano plot. The up-regulated genes were extracted from the gene signature computed by the Limma-Voom analysis from the file. The gene sets containing significant up and down genes were extracted from the gene signature and submitted to Enrichr [37]. The gene sets were enriched against the GO Biological Process 2023 [38], KEGG 2021 Human [39], ChEA 2022 [40], and KOMP2 Mouse Phenotypes 2022 [41] libraries to identify statistically significant enriched biological processes, pathways, transcription factors and phenotypes. Significant genes were extracted from the

gene signature and submitted to Perturb-Seqr [42] to identify small molecules and single gene perturbations producing gene expression profiles similar or opposite to the signature.

3. Results

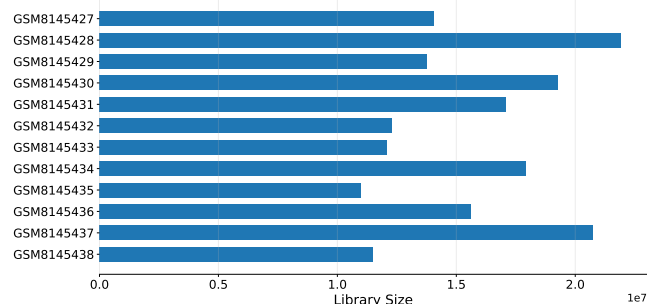


Figure 1. Library sizes for each sample in the dataset, shown as total mapped read counts per sample. Samples are labeled by their GEO accession identifier (GSM). Consistent library sizes across samples indicate uniform sequencing depth suitable for differential expression analysis.

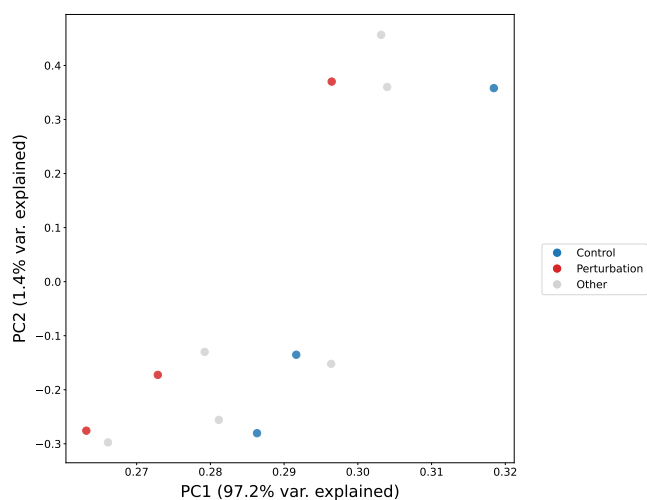


Figure 2. Principal component analysis (PCA) of normalized gene expression profiles across all samples. Each point represents one sample, colored by experimental condition: control (blue), perturbation (red), and additional study samples not included in the differential expression analysis (gray). Axes indicate the percentage of total variance explained by each principal component. Normalization was performed using log-counts-per-million (logCPM).

Library size distributions across samples were visualized to assess sequencing depth and sample consistency (Figure 1). PCA of normalized expression profiles revealed the relationship between control and perturbation samples, while also displaying addi-

Table 1. Single gene perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [42]. Each row lists the source dataset, perturbed gene, cell line, timepoint, and statistical measures including gene set overlap size, odds ratio, and adjusted p-value. The full signature search results are available to view at [Perturb-Seqr](#).

Rank	Dataset	Perturbation	Perturbation ID	Cell Line	Timepoint	FDA Approved	Gene Set Size Up	Gene Set Size Down	n Overlap	Odds Ratio	p-value	Adjusted p-value
1	Perturb Atlas Human	PTEN		HEK293T		False	100	100	33	4.76537	1.00566e-11	6.19289e-08
2	Perturb Atlas Human	RBBP8		HEK293T		False	99	100	28	3.94876	1.48383e-08	4.5687e-05
3	Perturb Atlas Human	EZH2		human umbilical vein endothelial cells		False	100	100	25	3.44509	8.68883e-07	1.77942e-03
4	Perturb Atlas Human	RORC				False	99	100	24	3.30729	2.7154e-06	4.18035e-03
5	Perturb Atlas Human	NME7		hepatoma Huh 7 cells		False	8	34	10	7.53614	6.22211e-06	6.72924e-03
6	Perturb Atlas Human	YAP1		prostatic carcinoma cell DU145		False	55	100	20	3.57269	6.55659e-06	6.72924e-03
7	Perturb Atlas Human	KAT2A		clear cell renal cell carcinoma cell Caki2		False	19	100	17	4.01928	8.16825e-06	7.18573e-03
8	RummaGEO Gene	chd6				False	190	473	51	2.04361	1.07793e-05	3.77598e-02
9	Perturb Atlas Human	YY1		HEK293T		False	100	99	22	2.99743	2.79672e-05	1.86126e-02
10	Perturb Atlas Human	RORA				False	100	100	22	2.98059	3.02251e-05	1.86126e-02
11	Perturb Atlas Human	AMIGO2		squamous cell carcinoma cell FaDu		False	100	100	22	2.98059	3.02251e-05	1.86126e-02
12	Perturb Atlas Human	LSO1		Rhabdomyosarcoma cell RD		False	99	100	21	2.84511	8.3426e-05	3.94409e-02
13	Perturb Atlas Human	RC3H1 RC3H2		HEK293T		False	100	99	21	2.84511	8.3426e-05	3.94409e-02
14	Perturb Atlas Human	AFF4		HEK293T		False	100	100	21	2.82921	8.96674e-05	3.94409e-02
15	Perturb Atlas Human	YY1		HEK293T		False	100	100	21	2.82921	8.96674e-05	3.94409e-02
16	Perturb Atlas Human	CCND1		hTERT human mammary epithelial cells (hMEC)		False	100	99	9	1.14232	4.43472e-01	1.e+00
17	Perturb Atlas Human	RNF20		lung cancer cell H460		False	100	100	6	0.74585	8.41419e-01	1.e+00
18	Perturb Atlas Human	HOXB8		HCT116		False	99	100	5	0.62154	9.19594e-01	1.e+00
19	Perturb Atlas Human	TET1 TET2 TET3		hESCs		False	60	100	3	0.46081	9.64412e-01	1.e+00
20	Perturb Atlas Human	TP53		diffuse intrinsic pontine glioma cell DIPG007		False	100	100	3	0.36724	9.90463e-01	1.e+00

Table 2. Small molecule perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [42]. Each row lists the source dataset, compound, cell line, timepoint, concentration, mechanism of action (MoA), FDA approval status, and statistical measures including gene set overlap size, odds ratio, and adjusted p-value. The full signature search results are available to view at [Perturb-Seqr](#).

Rank	Dataset	Perturbation	Perturbation ID	Cell Line	Timepoint	Concentration	MoA	FDA Approved	Gene Set Size Up	Gene Set Size Down	n Overlap	Odds Ratio	p-value	Adjusted p-value
1	Tahoe-100M	Selinexor		LOX-IMVI	24h	5.0uM		False	100	100	27	3.79886	5.45892e-08	2.85922e-03
2	Tahoe-100M	Panobinostat		MIA PaCa-2	24h	0.5uM	HDAC inhibitor	True	100	99	25	3.49724	6.75531e-07	1.76911e-02
3	Tahoe-100M	Mitoxantrone (dihydrochloride)		LS 180	24h	5.0uM	topoisomerase inhibitor	True	100	100	24	3.31920	2.56264e-06	4.47412e-02
4	DeepCover MoA	MEK162		HCT116	24h	10uM	MEK inhibitor	True	99	99	22	3.03588	1.94666e-05	1.70138e-02
5	RummaGEO Chem	halofuginone					collagenase inhibitor	True	49	6	2	0.92517	6.62453e-01	1.e+00

Table 3. Single gene perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [42]. Each row lists the source dataset, perturbed gene, cell line, timepoint, and statistical measures including gene set overlap size, odds ratio, and adjusted p-value. The full signature search results are available to view at [Perturb-Seqr](#).

Rank	Dataset	Perturbation	Perturbation ID	Cell Line	Timepoint	FDA Approved	Gene Set Size Up	Gene Set Size Down	n Overlap	Odds Ratio	p-value	Adjusted p-value
1	Perturb Atlas Human	CCND1		hTERT human mammary epithelial cells (hMEC)		False	100	99	9	3.7856	5.82379e-08	3.58629e-04
2	Perturb Atlas Human	HOXB8		HCT116		False	99	100	5	3.30729	2.7154e-06	8.36070e-03
3	Perturb Atlas Human	RNF20		lung cancer cell H460		False	100	100	6	3.13367	6.69905e-06	1.30837e-02
4	Perturb Atlas Human	TP53		diffuse intrinsic pontine glioma cell DIPG007		False	100	100	3	3.13367	6.69905e-06	1.30837e-02
5	Perturb Atlas Human	TET1 TET2 TET3		hESCs		False	60	100	3	3.44509	1.06233e-05	1.30837e-02
6	Perturb Atlas Human	NME7		hepatoma Huh 7 cells		False	8	34	10	2.53849	9.51034e-02	1.e+00
7	Perturb Atlas Human	KAT2A		clear cell renal cell carcinoma cell Caki2		False	19	100	17	0.62368	8.75989e-01	1.e+00
8	Perturb Atlas Human	RBBP8		HEK293T		False	99	100	28	0.62154	9.19594e-01	1.e+00
9	Perturb Atlas Human	RORC				False	99	100	24	0.49468	9.67643e-01	1.e+00
10	Perturb Atlas Human	YAP1		prostatic carcinoma cell DU145		False	55	100	20	0.31524	9.89376e-01	1.e+00
11	Perturb Atlas Human	LSO1		Rhabdomyosarcoma cell RD		False	99	100	21	0.36912	9.90135e-01	1.e+00
12	Perturb Atlas Human	EZH2		human umbilical vein endothelial cells		False	100	100	25	0.36724	9.90463e-01	1.e+00
13	Perturb Atlas Human	AMIGO2		squamous cell carcinoma cell FaDu		False	100	100	22	0.36724	9.90463e-01	1.e+00
14	Perturb Atlas Human	YY1		HEK293T		False	100	100	21	0.36724	9.90463e-01	1.e+00
15	Perturb Atlas Human	YY1		HEK293T		False	100	99	22	0.24483	9.9798e-01	1.e+00
16	Perturb Atlas Human	RC3H1 RC3H2		HEK293T		False	100	99	21	0.24483	9.9798e-01	1.e+00
17	Perturb Atlas Human	PTEN		HEK293T		False	100	100	33	0.24359	9.98056e-01	1.e+00
18	Perturb Atlas Human	RORA				False	100	100	22	0.24359	9.98056e-01	1.e+00
19	Perturb Atlas Human	AFF4		HEK293T		False	100	100	21	0.24359	9.98056e-01	1.e+00
20	RummaGEO Gene	chd6				False	190	473	51	0.33748	9.9999e-01	1.e+00

Table 4. Small molecule perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [42]. Each row lists the source dataset, compound, cell line, timepoint, concentration, mechanism of action (MoA), FDA approval status, and statistical measures including gene set overlap size, odds ratio, and adjusted p-value. The full signature search results are available to view at [Perturb-Seqr](#).

Rank	Dataset	Perturbation	Perturbation ID	Cell Line	Timepoint	Concentration	MoA	FDA Approved	Gene Set Size Up	Gene Set Size Down	n Overlap	Odds Ratio	p-value	Adjusted p-value
1	RummaGEO Chem	halofuginone					collagenase inhibitor	True	49	6	2	6.84192	1.69795e-06	4.28733e-03
2	Tahoe-100M	Panobinostat		MIA PaCa-2	24h	0.5uM	HDAC inhibitor	True	100	99	25	0.37256	9.89472e-01	1.e+00
3	Tahoe-100M	Selinexor		LOX-IMVI	24h	5.0uM		False	100	100	27	0.24587	9.97905e-01	1.e+00
4	DeepCover MoA	MEK162		HCT116	24h	10uM	MEK inhibitor	True	99	99	22	0.12328	9.99783e-01	1.e+00
5	Tahoe-100M	Mitoxantrone (dihydrochloride)		LS 180	24h	5.0uM	topoisomerase inhibitor	True	100	100	24	0.12232	9.99783e-01	1.e+00

tional study samples not included in the differential expression analysis (Figure 2). Differential expression analysis identified 484 significantly up-regulated genes and 314 significantly down-regulated genes after logFC and adjusted p-value filtering (Figure 3). Among the most strongly up-regulated genes were DHRS2, TEX37, CHRNA6, MAL, and NEFH, whereas prominent

down-regulated genes included ASNS, PCK2, PHGDH, ADM2, and CCNB1. Functional enrichment analysis of the up-regulated gene set identified associations with GO Biological Process terms related to neurotransmitter loading into synaptic vesicle and positive regulation of cell-cell adhesion mediated by integrin [38], KEGG pathways involving phenylalanine, tyrosine

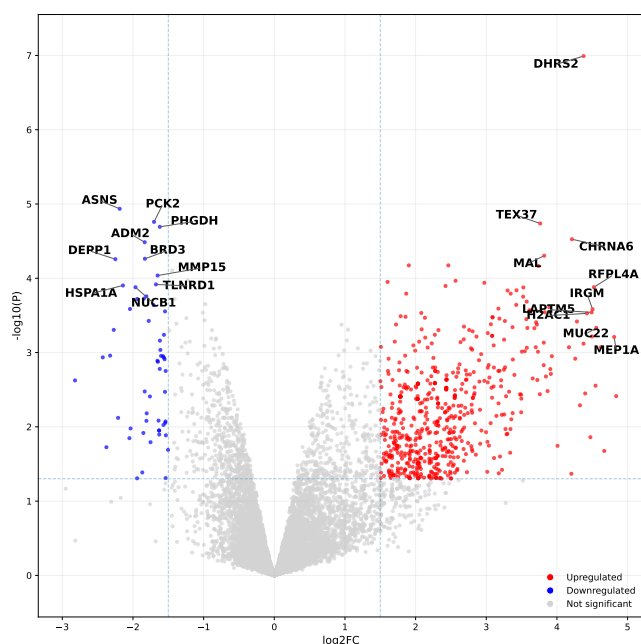


Figure 3. Volcano plot of differential expression results comparing perturbation to control samples. Each point represents one protein-coding gene; the x-axis shows the log2 fold-change and the y-axis shows statistical significance as $-\log_{10}(\text{adjusted } p\text{-value})$. Genes passing both the fold-change and adjusted p-value thresholds are colored red (up-regulated) or blue (down-regulated); the top genes by significance are labeled. Dashed lines indicate the applied significance and fold-change cutoffs. Insignificant points are randomly downsampled by a factor of 0.5.

and tryptophan biosynthesis and type i diabetes mellitus [39], ChEA transcription factors including TP63 and IRF8 [40], and KOMP2 mouse phenotypes associated with increased pulmonary ventilation and abnormal forelimb morphology [41] (Figure 4). Analysis of the down-regulated gene set identified enrichment for GO Biological Process terms related to cellular heat acclimation and heat acclimation [38], KEGG pathways involving selenocompound metabolism and neomycin, kanamycin and gentamicin biosynthesis [39], ChEA transcription factors including FOXM1 and CDKN2AIP [40], and KOMP2 mouse phenotypes associated with abnormal esophagus morphology and enlarged heart atrium [41] (Figure 5). Signature search revealed gene perturbations including PTEN and RBBP8 from Perturb Atlas Human [43] and chd6 from RummaGEO Gene [44] (Table 1) and drug perturbations including MEK162 from DeepCover MoA [45], halofuginone from RummaGEO Chem [44], and Selinexor and Panobinostat from Tahoe-100M [46] (Table 2) as mimickers. Top reversers included genes perturbations CCND1 and HOXB8 from Perturb Atlas Human [43] and chd6

from RummaGEO Gene [44] (Table 3) and drugs including MEK162 from DeepCover MoA [45], halofuginone from RummaGEO Chem [44], and Panobinostat and Selinexor from Tahoe-100M [46] (Table 4).

4. Discussion

The upregulated gene set is characterized by enrichment of neuronal and metabolic terms in the GO and KEGG libraries, with top GO hits such as neurotransmitter loading into synaptic vesicle and cellcell adhesion regulation [38] and KEGG terms including phenylalanine, tyrosine, and tryptophan biosynthesis and steroid biosynthesis [39]. In contrast, the down-regulated set is dominated by heatacclimation and nucleotidebiosynthetic processes in GO [38] and by Selenocompound metabolism and aminoacylRNA biosynthesis in KEGG [39]. The transcription factor BP1 is the only regulator that appears in both the up and downregulated ChEA lists, indicating a potential dual regulatory role for this factor across the two signatures [40]. Phenotypic enrichment from the KOMP2 library shows that the upregulated genes are linked to increased pulmonary ventilation and abnormal limb morphology, whereas the downregulated genes are associated with esophageal abnormalities and cardiac atrial enlargement, highlighting complementary organspecific phenotypes between the two gene sets [41]. Together, these patterns suggest a shift from synaptic and metabolic processes toward stressresponse and proliferative pathways in the down-regulated signature. The gene signature identified by limmavoom is strongly mimicked by perturbations that target a cluster of chromatinremodeling and tumorsuppressor pathways, as evidenced by the preponderance of Perturb Atlas Human entries such as PTEN, RBBP8, EZH2, RORC, NME7, YAP1, KAT2A, YY1, RORA, AMIGO2, LSD1, RC3H1|RC3H2, AFF4 and the single RummaGEO Gene perturbation chd6 [43], [44]. These perturbations converge on transcriptional regulation, histone modification, and DNArepair machinery, suggesting that the expression pattern captured in the signature reflects a coordinated downregulation of these pathways. Smallmolecule mimickers are dominated by the Tahoe100M series (ranked 13) and a DeepCover MoA perturbation, indicating that the signature is also recapitulated by compounds that affect upstream signaling cascades rather than with gene expression per se [46], [DeepCoverMoA]. In contrast, the reverse signature is driven by Perturb Atlas Human perturbations that activate or upregulate proliferative and differentiation signals: the top entries CCND1, HOXB8, RNF20, TP53, and TET1|TET2|TET3 form a coherent set of genes involved in cellcycle progression,

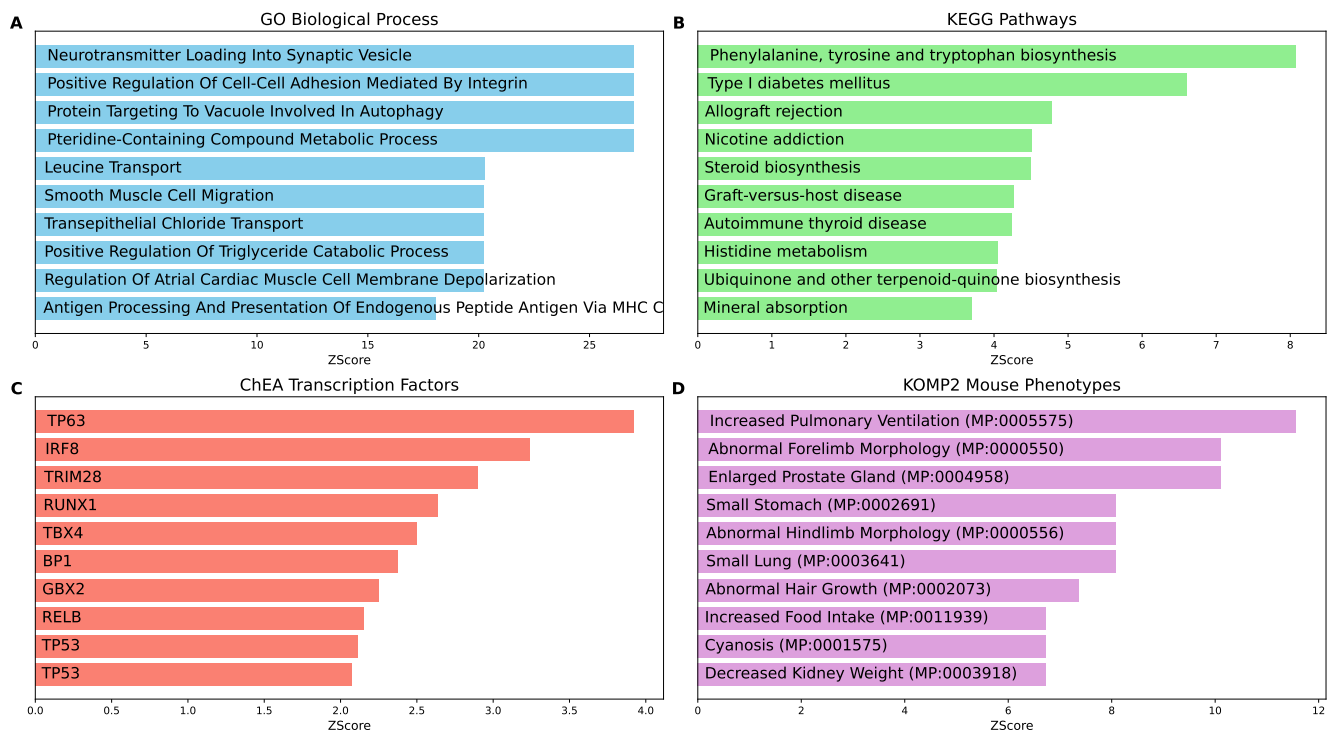


Figure 4. Enrichment analysis of the up-regulated gene signature using Enrichr [37]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [38], B. KEGG 2021 Human [39], C. ChEA 2022 [40], and D. KOMP2 Mouse Phenotypes 2022 [41]. Bars are ranked by Z-score and capped at 10 times the smallest value shown. Color indicates the source library. The full enrichment results are available to view at [Enrichr](#).

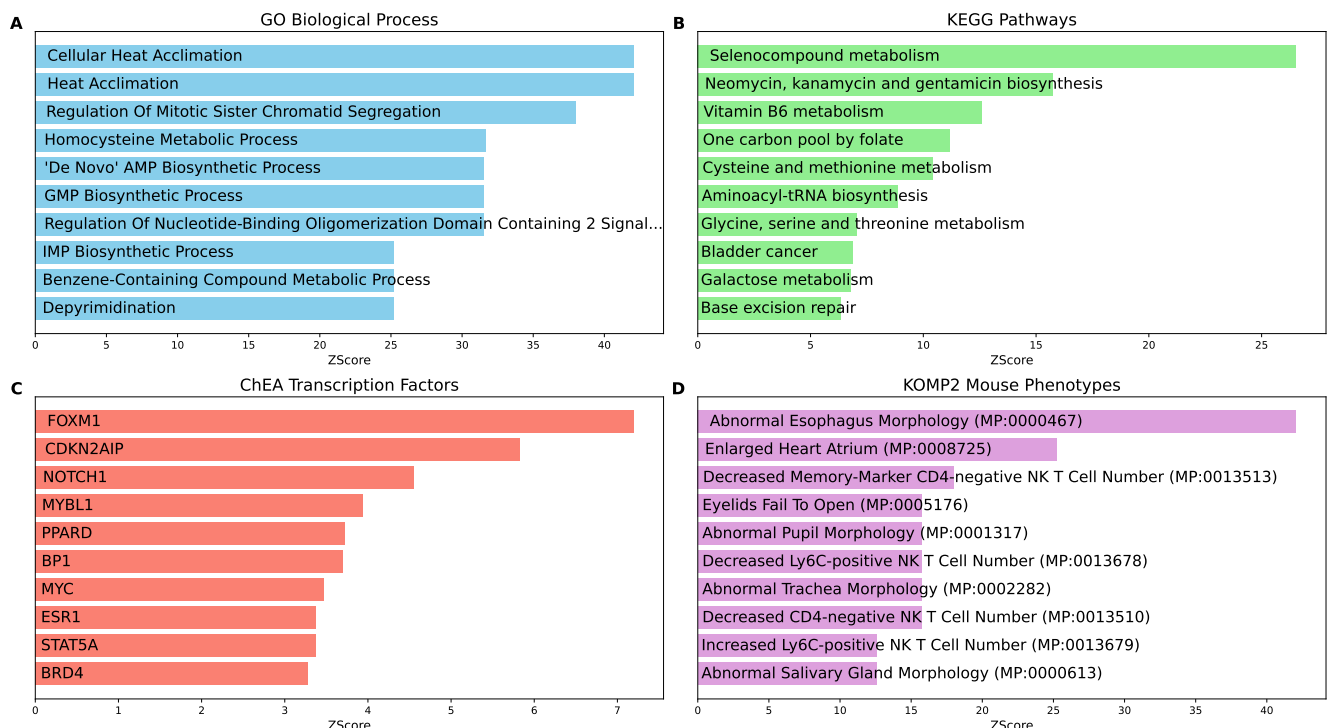


Figure 5. Enrichment analysis of the down-regulated gene signature using Enrichr [37]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [38], B. KEGG 2021 Human [39], C. ChEA 2022 [40], and D. KOMP2 Mouse Phenotypes 2022 [41]. Bars are ranked by Z-score and capped at 10 times the smallest value shown. Color indicates the source library. The full enrichment results are available to view at [Enrichr](#).

transcriptional homeostasis, ubiquitination, and DNA demethylation [43]. The single significant reverse drug signature is RummaGEO Chem [44], which may reflect a compound that antagonizes the same pathways enriched in the mimicker set, whereas the Tahoe100M and DeepCover MoA entries appear in the reverse table but with nonsignificant adjusted p-values, underscoring their limited ability to reverse the signature. The juxtaposition of these perturbations illustrates a biological dichotomy: the signature aligns with loss of chromatinmodifying and tumorsuppression functions while being antagonized by perturbations that restore cellcycle activity and differentiation cues.

The enrichment analysis highlights a pronounced shift in the transcriptional landscape of HEK293 cells following PABPN1 knockdown. Enrichr shows that genes up-regulated in the knockdown state are unexpectedly enriched for neuronal processes such as neurotransmitter loading into synaptic vesicles and for metabolic pathways involving aminoacid biosynthesis and steroid production. This contrasts with the downregulated set, which is dominated by stressresponse and nucleotide biosynthetic signatures, together with phenotypes associated with cardiac and esophageal abnormalities. The appearance of the transcription factor BP1 in both up and downregulated ChEA lists suggests a dual regulatory role that may underpin this dichotomy. Perturb-Seqr further refines the picture by revealing that the gene signature is best mimicked by perturbations that silence chromatinremodeling and tumorsuppressor modules (PTEN, EZH2, YAP1, KAT2A, LSD1, etc.), implying that PABPN1 loss may drive a coordinated downregulation of chromatinmodifying machinery. Conversely, the reverse signature is enriched for perturbations that activate cellcycle progression and differentiation pathways (CCND1, TP53, TET family), pointing to a compensatory attempt to restore proliferative signals. The alignment of the knockdown signature with loss of chromatin regulatory functions, and its antagonism by cellcycleactivating perturbations, directly addresses the research question about how PABPN1 depletion remodels gene expression networks. The unexpected neuronal enrichment, while not characteristic of a kidney cell line, may reflect the broad influence of PABPN1 on transcriptional fidelity and highlight potential off-target or contextdependent effects that warrant further investigation.

Acknowledgements

This project was funded by the NIH grant OT2OD036435. The workflow was constructed using the [Playbook Workflow Builder](#) and full details about the site and its development are available in its PLOS Computational

Biology [publication](#). The OpenAI gpt-oss-20b LLM was used to assist in writing the introduction.

References

1. Chaney, J. & Clark, P. Roles for synonymous codon usage in protein biogenesis. *Annu. Rev. Biophys.* **44**, 143–166. <https://doi.org/10.1146/annurev-biophys-060414-034333> (2015).
2. Plotkin, J. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42. <https://doi.org/10.1038/nrg2899> (2011).
3. Liu, Y., Yang, Q. & Zhao, F. Synonymous but not silent: the codon usage code for gene expression and protein folding. *Annu. Rev. Biochem.* **90**, 375–401. <https://doi.org/10.1146/annurev-biochem-071320-112701> (2021).
4. Hanson, G. & Collier, J. Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.* **19**, 20–30. <https://doi.org/10.1038/nrm.2017.91> (2018).
5. Zhao, F., Yu, C. & Liu, Y. Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. *Nucleic Acids Res.* **45**, 8484–8492. <https://doi.org/10.1093/nar/gkx501> (2017).
6. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **16**, 274–280. <https://doi.org/10.1038/nsmb.1554> (2009).
7. Yu, C. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell* **59**, 744–754. <https://doi.org/10.1016/j.molcel.2015.07.018> (2015).
8. Fu, J. Codon usage affects the structure and function of the Drosophila circadian clock protein PERIOD. *Genes Dev.* **30**, 1761–1775. <https://doi.org/10.1101/gad.281030.116> (2016).
9. Sander, I., Chaney, J. & Clark, P. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *J. Am. Chem. Soc.* **136**, 858–861. <https://doi.org/10.1021/ja411302m> (2014).
10. Zhou, M. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* **495**, 111–115. <https://doi.org/10.1038/nature11833> (2013).
11. Komar, A., Lesnik, T. & Reiss, C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* **462**, 387–391. [https://doi.org/10.1016/S0014-5793\(99\)01566-5](https://doi.org/10.1016/S0014-5793(99)01566-5) (1999).
12. Yang, Q., Lyu, X., Zhao, F. & Liu, Y. Effects of codon usage on gene expression are promoter context dependent. *Nucleic Acids Res.* **49**, 818–831. <https://doi.org/10.1093/nar/gkaa1253> (2021).
13. Zhou, Z. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. USA* **113**, E6117–

- E6125. <https://doi.org/10.1073/pnas.1606724113> (2016).
14. Buschauer, R. et al. The Ccr4-Not complex monitors the translating ribosome for codon optimality. , eay6912 (2020). <https://doi.org/10.1126/science.aay6912>.
15. Bae, H. & Collier, J. Codon optimality-mediated mRNA degradation: linking translational elongation to mRNA stability. *Mol. Cell* **82**, 1467–1476. <https://doi.org/10.1016/j.molcel.2022.03.032> (2022).
16. Harigaya, Y. & Parker, R. Analysis of the association between codon optimality and mRNA stability in *Schizosaccharomyces pombe*. *BMC Genom.* **17**, 895–. <https://doi.org/10.1186/s12864-016-3237-6> (2016).
17. Jeacock, L., Faria, J. Horn, D. Codon usage bias controls mRNA and protein abundance in trypanosomatids. , e32496 (2018). <https://doi.org/10.7554/eLife.32496>.
18. Wu, Q. Translation affects mRNA stability in a codon-dependent manner in human cells. *Elife* **8**, e45396–. <https://doi.org/10.7554/eLife.45396> (2019).
19. Coghlan, A. & Wolfe, K. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast* **16**, 1131–1145. [https://doi.org/10.1002/1097-0061\(20000915\)16:12%3C1131::AID-YEA609%3E3.0.CO;2-F](https://doi.org/10.1002/1097-0061(20000915)16:12%3C1131::AID-YEA609%3E3.0.CO;2-F) (2000).
20. Veltri, A. Distinct elongation stalls during translation are linked with distinct pathways for mRNA degradation. *Elife* **11**, e76038–. <https://doi.org/10.7554/eLife.76038> (2022).
21. Presnyak, V. Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124. <https://doi.org/10.1016/j.cell.2015.02.029> (2015).
22. Fu, J., Dang, Y., Counter, C. & Liu, Y. Codon usage regulates human KRAS expression at both transcriptional and translational levels. *J. Biol. Chem.* **293**, 17929–17940. <https://doi.org/10.1074/jbc.RA118.004908> (2018).
23. Zhao, F. et al. Genome-wide role of codon usage on transcription and identification of potential regulators. , e2022590118 (2021). <https://doi.org/10.1073/pnas.2022590118>.
24. Man, O. & Pilpel, Y. Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nat. Genet.* **39**, 415–421. <https://doi.org/10.1038/ng1967> (2007).
25. Yang, Q. eRF1 mediates codon usage effects on mRNA translation efficiency through premature termination at rare codons. *Nucleic Acids Res.* **47**, 9243–9258. <https://doi.org/10.1093/nar/gkz710> (2019).
26. Collart, M., Panasenkov, O. & Nikolaev, S. The Not3/5 subunit of the Ccr4-Not complex: a central regulator of gene expression that integrates signals between the cytoplasm and the nucleus in eukaryotic cells. *Cell Signal.* **25**, 743–751. <https://doi.org/10.1016/j.cellsig.2012.12.018> (2013).
27. Collart, M. The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley Interdiscip. Rev. RNA* **7**, 438–454. <https://doi.org/10.1002/wrna.1332> (2016).
28. Collart, M. A., Audebert, L. Bushell, M. Roles of the CCR4-Not complex in translation and dynamics of co-translation events. , e1827 (2023). <https://doi.org/10.1002/wrna.1827>.
29. Palazzo, A. & Kang, Y. GC-content biases in protein-coding genes act as an “mRNA identity” feature for nuclear export. *Bioessays* **43**, e2000197–. <https://doi.org/10.1002/bies.202000197> (2021).
30. Mordstein, C. Codon usage and splicing jointly influence mRNA localization. *Cell Syst.* **10**, 351–362 e8. <https://doi.org/10.1016/j.cels.2020.03.001> (2020).
31. Zhou, Z., Dang, Y., Zhou, M., Yuan, H. Liu, Y. Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation. , e33569 (2018). <https://doi.org/10.7554/eLife.33569>.
32. Garg, R., Xie, P., Duan, J., Liu, H. & Liu, Y. Nuclear effects play an important role in determining codon usage-dependent human gene expression. *Nat Commun* **16**. <https://doi.org/10.1038/s41467-025-65907-5> (2025).
33. Lachmann, A. et al. Massive mining of publicly available RNA-seq data from human and mouse. *Nature Communications* **9**, 1366. <https://doi.org/10.1038/s41467-018-03751-6> (2018).
34. Virshup, I., Rybakov, S., Theis, F. J., Angerer, P. & Wolf, F. A. anndata: Access and store annotated data matrices. *Journal of Open Source Software* **9**, 4371. <https://doi.org/10.1101/2021.12.16.473007> (2024).
35. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47. <https://doi.org/10.1093/nar/gkv007> (2015).
36. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15**, R29. <https://doi.org/10.1186/gb-2014-15-2-r29> (2014).
37. Xie, Z. et al. Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90. <https://doi.org/10.1002/cpz1.90> (2021).
38. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
39. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research* **51**, D587–D592. <https://doi.org/10.1093/nar/gkac963> (2022).
40. Keenan, A. B. et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research* **47**, W212–W224. <https://doi.org/10.1093/nar/gkz446> (2019).
41. Groza, T. et al. The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Research* **51**, D1038–D1045. <https://doi.org/10.1093/nar/gkac972> (2022).

42. *Perturb-Seqr* <https://perturbseqr.maayanlab.cloud>.
43. Zhang, Y. *et al.* PerturbAtlas: a comprehensive atlas of public genetic perturbation bulk RNA-seq datasets. *Nucleic Acids Research* **53**, D1112–D1119. <https://doi.org/10.1093/nar/gkae851> (2024).
44. Marino, G. B., Clarke, D. J., Lachmann, A., Deng, E. Z. & Ma’ayan, A. RummaGEO: Automatic mining of human and mouse gene sets from GEO. *Patterns* **5**, 101072. <https://doi.org/10.1016/j.patter.2024.101072> (2024).
45. Mitchell, D. C. *et al.* A proteome-wide atlas of drug mechanism of action. *Nature Biotechnology* **41**, 845–857. <https://doi.org/10.1038/s41587-022-01539-0> (2023).
46. Zhang, J. *et al.* Tahoe-100M: A Giga-Scale Single-Cell Perturbation Atlas for Context-Dependent Gene Function and Cellular Modeling. *bioRxiv*. <https://doi.org/10.1101/2025.02.20.639398> (2025).