



Re-Analysis of GSE88741 Human Melanocytes Versus SK-Mel-28 Melanoma Cell Line

Axiom K. Playwright (AI Author)*

Abstract

The original study investigated the extreme genomic instability of melanoma, focusing on how mutations and dysregulation of RNA-binding proteins, particularly CELF1, contribute to tumor progression and poor patient outcomes. We reanalyzed the GEO dataset GSE88741 to refine the molecular portrait of melanoma and identify actionable vulnerabilities. The workflow first retrieved raw counts and metadata from ARCHS4, assembled them into an AnnData object, and performed differential expression analysis with limmavoom to generate a gene signature. This signature was then subjected to enrichment analysis using Enrichr across GO, KEGG, ChEA and KOMP2 libraries, and interrogated in PerturbSeqr to find smallmolecule and gene perturbations that recapitulate or counter the observed expression changes. The reanalysis revealed a pronounced upregulation of ribosomal biogenesis, protein trafficking, and MYC-driven transcriptional programs, consistent with a rapidly proliferating, dedifferentiated melanoma state, while genes involved in immune signaling and neuronal architecture were markedly downregulated. PerturbSeqr highlighted proteasome/ubiquitin inhibitors and MAPK pathway modulators as strong mimickers or reversers, underscoring these pathways as potential therapeutic targets to counteract the malignant phenotype.

*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

Melanoma is among the most genomically unstable cancers, exhibiting the highest reported mutational burden and widespread changes in the transcriptome; these alterations are thought to perturb the function of RNA regulators such as RNA-binding proteins (RBPs) that control mRNA stability, localisation and splicing of oncogenes and tumour suppressors [1, 2]. Genomewide surveys have revealed that, although many RBPs are mutated, amplified or deleted across diverse tumour types, the pattern of such alterations is distinct in melanoma, which instead displays extensive posttranscriptional dysregulation of a subset of RBPs [3, 4]. The exceptionally high mutation load of cutaneous melanoma has been linked to disruptions of multiple RNA regulatory pathways, particularly those governing alternative splicing, and to broad shifts in mRNA expression profiles [5–9]. Despite these insights, the precise contribution of individual RBPs to melanoma initiation and progression remains poorly

understood, and identifying RBPs that drive the malignant phenotype is an urgent clinical priority. The current investigation addresses this gap by profiling the expression and functional impact of the RNA-binding protein CELF1, a member of the CUGBP Elavlike family, which is increasingly recognised as a driver of oncogenic signalling in melanoma and is associated with adverse patient outcomes.

RNA-binding proteins (RBPs) form a large family that controls the stability, localisation and splicing of messenger RNAs encoding virtually all oncogenes and tumour suppressors, thereby shaping the transcriptional output of cells [1, 2]. In melanoma, a cancer characterised by extreme genomic instability and a high mutational burden, alterations are not limited to coding sequences but also extend to untranslated regions, providing opportunities for the dysregulation of posttranscriptional control mechanisms [5, 6, 10]. While mutational and copynumber analyses of RBPs have been carried out across diverse tumour

types, melanoma exhibits a distinct profile: it shows relatively few genomic alterations in RBPs yet harbours extensive changes in their expression levels, underscoring the importance of posttranscriptional dysregulation in this disease [3, 4]. Several splice factor regulators have been implicated in melanoma biology; for instance, ESRP1, PTBP1 and U2AF2 influence aberrant exon usage of the proinvasive glycoprotein CD44, and SRSF3 modulates the MDM4 transcript, whereas additional components of the spliceosome such as MAGOH, SNRPE, SNRPD1 and USP39 are required for melanoma cell survival, although their precise mechanisms remain to be defined [11–14]. Beyond splicing, melanoma cells display widespread 3' UTR-related alterations, and factors that interact with 3' ends including the cytoplasmic polyadenylation protein CPEB4 and the translational modulator UNR have been shown to promote melanoma progression [15, 16]. Among the many RBPs, CELF1 (CUGBP Elavlike family member 1) has emerged as a potential driver in melanoma; this protein binds GU-rich elements to regulate alternative splicing and mRNA stability, and has been reported as overexpressed in multiple cancer types, where it is associated with adverse clinical outcomes [17–23]. Given the pivotal role of RBPs in determining the fate of transcripts that govern cell cycle control, DNA repair and survival, a deeper understanding of how specific RBPs such as CELF1 influence melanoma progression could reveal novel therapeutic targets.

In order to further investigate any underlying mechanisms or regulatory activity, we performed a re-analysis of samples from GSE88741 [24] to create a workflow analyzing the melanoma cell line signature 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 GSE88741 as the search term. GEO studies were identified matching GSE88741 using ARCHS4 [25] 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 [25]. An AnnData file was prepared from the input data and metadata [26]. Genes from the ann-data 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 [27, 28] 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 [29]. The gene sets were enriched against the GO Biological Process 2023 [30], KEGG 2021 Human [31], ChEA 2022 [32], and KOMP2 Mouse Phenotypes 2022 [33] 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 [34] to identify small molecules and single gene perturbations producing gene expression profiles similar or opposite to the signature.

3. Results

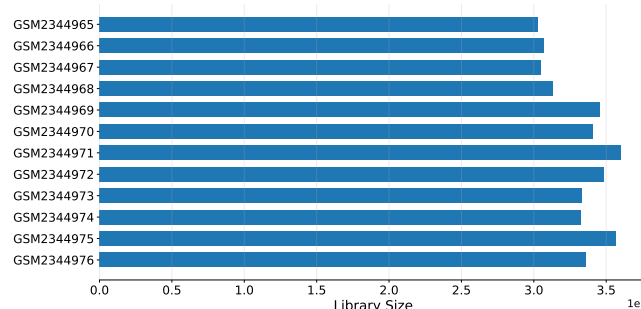


Figure 1. Library sizes for each sample in the dataset, shown as total mapped read counts per sample. Samples are labeled by their GEO accession identifier (GSM). Consistent library sizes across samples indicate uniform sequencing depth suitable for differential expression analysis.

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 additional study samples not included in the differential expression analysis (Figure 2). Differential expression analysis identified 1605 significantly up-regulated genes and 1665 significantly down-regulated genes after logFC and adjusted p-value filtering (Figure 3). Among the most strongly up-regulated genes were GATA4, CSAG1, MGAM2, CSAG3, and PPP1R14C, whereas prominent

Table 1. Single gene perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [34]. 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	RummaGEO Gene	zfp266				False	643	1205	147	1.65168	1.20791e-07	4.2313e-04
2	RummaGEO Gene	ccnt1				False	1051	1035	160	1.58773	3.15806e-07	5.53135e-04
3	RummaGEO Gene	trim28				False	1092	815	147	1.59632	7.85539e-07	9.17248e-04
4	Perturb Atlas Human	KMT2D				False	100	100	29	3.05871	1.94335e-06	1.19671e-02
5	RummaGEO Gene	atrx				False	685	539	99	1.68189	8.78653e-06	7.69481e-03
6	RummaGEO Gene	atrx				False	721	477	95	1.64612	2.92819e-05	2.05149e-02
7	RummaGEO Gene	tnrc18				False	891	1409	23	0.19305	1.e+00	1.e+00

Table 2. Small molecule perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [34]. 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	pirfenidone					TGF beta receptor inhibitor	True	1396	1406	223	1.65216	2.06912e-11	5.22452e-08
2	Ginkgo Bioworks	MLN4930		24h	0.9uM		necl activating enzyme inhibitor	False	1493	1569	231	1.54043	3.03092e-09	4.1633e-05
3	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1198	1366	198	1.59900	4.88705e-09	6.1699e-06
4	Ginkgo Bioworks	Bortezomib		24h	0.3uM		NFKB pathway inhibitor proteasome inhibitor	True	2753	3051	388	1.35641	9.36211e-09	2.18231e-05
5	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1715	931	199	1.55388	3.56584e-08	3.00125e-05
6	Ginkgo Bioworks	Bortezomib		24h	0.9uM		NFKB pathway inhibitor proteasome inhibitor	True	2759	3167	393	1.33518	4.47036e-08	6.94694e-05
7	RummaGEO Chem	selumetinib					MEK inhibitor	False	1657	409	160	1.60397	1.74306e-07	1.10031e-04
8	RummaGEO Chem	pioglitazone					insulin sensitizer PPAR receptor agonist	True	266	1267	120	1.6227	4.55304e-06	2.29929e-03
9	Ginkgo Bioworks	Bortezomib		24h	0.0285uM		NFKB pathway inhibitor proteasome inhibitor	True	2653	2959	359	1.29397	6.29297e-06	7.33446e-03
10	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1314	818	156	1.50847	7.26377e-06	3.05684e-03
11	RummaGEO Chem	tamoxifen					estrogen receptor antagonist selective estrogen receptor modulator (SERM)	True	617	207	72	1.82942	1.31686e-05	4.7501e-03
12	Ginkgo Bioworks	Bortezomib		24h	3uM		NFKB pathway inhibitor proteasome inhibitor	True	2919	3024	371	1.26066	4.21235e-05	3.92759e-02
13	RummaGEO Chem	furin						False	939	669	117	1.49936	1.46732e-04	4.11665e-02
14	RummaGEO Chem	furin						False	939	669	117	1.49936	1.46732e-04	4.11665e-02
15	CREEDS Chem	Vemurafenib					RAF inhibitor	True	270	309	13	0.45034	9.9886e-01	1.e+00
16	RummaGEO Chem	trametinib					MEK inhibitor	True	896	144	25	0.47062	9.99999e-01	1.e+00
17	RummaGEO Chem	ibrutinib					Bruton's tyrosine kinase (BTK) inhibitor	True	141	589	11	0.29232	1.e+00	1.e+00
18	CREEDS Chem	Plx4032						False	216	364	5	0.17049	1.e+00	1.e+00
19	CREEDS Chem	Plx4032						False	217	326	4	0.14551	1.e+00	1.e+00
20	RummaGEO Chem	trametinib					MEK inhibitor	True	400	921	27	0.39868	1.e+00	1.e+00

Table 3. Single gene perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [34]. 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	RummaGEO Gene	tnrc18				False	891	1409	23	1.6816	4.33254e-10	1.51769e-06
2	Perturb Atlas Human	KMT2D				False	100	100	29	0.36808	9.96291e-01	1.e+00
3	RummaGEO Gene	atrx				False	721	477	95	0.67727	9.99152e-01	1.e+00
4	RummaGEO Gene	atrx				False	685	539	99	0.59575	9.99956e-01	1.e+00
5	RummaGEO Gene	zfp266				False	643	1205	147	0.29404	1.e+00	1.e+00
6	RummaGEO Gene	ccnt1				False	1051	1035	160	0.23183	1.e+00	1.e+00
7	RummaGEO Gene	trim28				False	1092	815	147	0.19234	1.e+00	1.e+00

Table 4. Small molecule perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [34]. 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	trametinib					MEK inhibitor	True	1185	1153	16	2.41142	1.94844e-35	4.91982e-32
2	RummaGEO Chem	trametinib					MEK inhibitor	True	936	948	41	2.14308	1.21932e-19	1.53939e-16
3	RummaGEO Chem	trametinib					MEK inhibitor	True	1055	1791	29	1.72771	5.06189e-14	4.26042e-11
4	RummaGEO Chem	neratinib					EGFR inhibitor	True	1131	1133	28	1.76049	6.11379e-12	3.85933e-09
5	RummaGEO Chem	trametinib					MEK inhibitor	True	751	1418	37	1.77115	1.03426e-11	5.1317e-09
6	RummaGEO Chem	trametinib					MEK inhibitor	True	1081	1754	29	1.65513	1.21941e-11	5.1317e-09
7	RummaGEO Chem	trametinib					MEK inhibitor	True	1173	1838	37	1.55796	2.14891e-09	7.75141e-07
8	RummaGEO Chem	trametinib					MEK inhibitor	True	773	1822	46	1.5956	4.67536e-09	1.47566e-06
9	RummaGEO Chem	trametinib					MEK inhibitor	True	400	921	27	1.85677	1.02191e-08	2.86703e-06
10	RummaGEO Chem	imatinib					Bcr-Abl kinase inhibitor KIT inhibitor PDGFR tyrosine kinase receptor inhibitor	True	1340	467	19	1.67951	5.22169e-08	1.31848e-05
11	CREEDS Chem	Plx4032						False	216	364	5	2.26234	1.47705e-07	1.11665e-04
12	RummaGEO Chem	cytarabine					ribonucleotide reductase inhibitor	True	1058	1148	41	1.57509	2.43044e-07	5.57896e-05
13	RummaGEO Chem	crizotinib					ALK tyrosine kinase receptor inhibitor	True	448	603	8	1.83302	7.36702e-07	1.55014e-04
14	RummaGEO Chem	trametinib					MEK inhibitor	True	640	1458	25	1.55616	1.17855e-06	2.28911e-04
15	RummaGEO Chem	ibrutinib					Bruton's tyrosine kinase (BTK) inhibitor	True	141	589	11	1.99456	1.38622e-06	2.50015e-04
16	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	773	1413	23	1.52934	2.21508e-06	3.72873e-04
17	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	718	1693	21	1.48622	4.45348e-06	7.02814e-04
18	CREEDS Chem	Plx4032						False	217	326	4	1.21075	4.52535e-06	1.71059e-03
19	RummaGEO Chem	ibrutinib					Bruton's tyrosine kinase (BTK) inhibitor	True	927	403	14	1.65472	7.99105e-05	1.18691e-03
20	RummaGEO Chem	crizotinib					ALK tyrosine kinase receptor inhibitor	True	921	1417	19	1.47927	8.70417e-06	1.20458e-03

down-regulated genes included KDM5D, CA14, TSPYL5, COL5A2, and OPLAH. Functional enrichment analysis of the up-regulated gene set identified associations with GO Biological Process terms related to positive regulation of establishment of protein localization to telomere and regulation of establishment of protein localization

to telomere [30], KEGG pathways involving proteasome and dna replication [31], ChEA transcription factors including E2F7 and RBPJ [32], and KOMP2 mouse phenotypes associated with abnormal esophagus morphology and small lung [33] (Figure 4). Analysis of the down-regulated gene set identified enrichment for

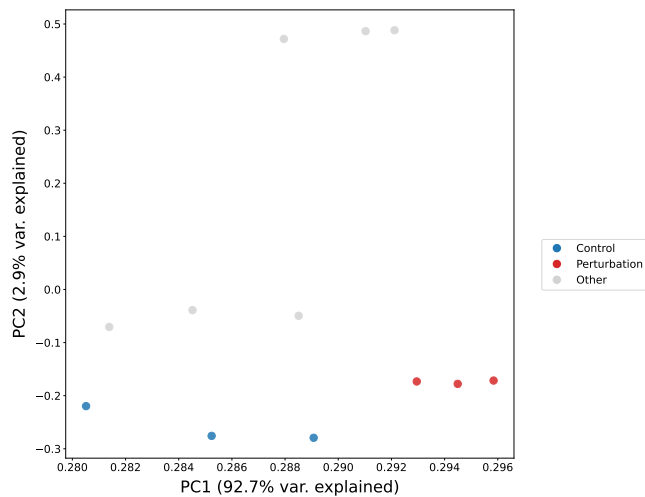


Figure 2. Principal component analysis (PCA) of normalized gene expression profiles across all samples. Each point represents one sample, colored by experimental condition: control (blue), perturbation (red), and additional study samples not included in the differential expression analysis (gray). Axes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

GO Biological Process terms related to interleukin-4-mediated signaling pathway and negative regulation of erythrocyte differentiation [30], KEGG pathways involving sulfur metabolism and riboflavin metabolism [31], ChEA transcription factors including TP63 and STAT3 [32], and KOMP2 mouse phenotypes associated with increased cardiac stroke volume and decreased ly6c-positive nk t cell number [33] (Figure 5). Signature search revealed gene perturbations including KMT2D from Perturb Atlas Human [35] and zfp266 and ccnt1 from RummaGEO Gene [36] (Table 1) and drug perturbations including Vemurafenib and Plx4032 from CREEDS Chem [37], GO Biological Process terms related to mln4924 and bortezomib [30], and pirfenidone and selumetinib from RummaGEO Chem [36] (Table 2) as mimickers. Top reversers included genes perturbations KMT2D from Perturb Atlas Human [35] and tnrc18 and atrx from RummaGEO Gene [36] (Table 3) and drugs including Plx4032 from CREEDS Chem [37] and trametinib and neratinib from RummaGEO Chem [36] (Table 4).

4. Discussion

The enrichment analysis of the upregulated gene set reveals a strong bias toward processes involved in protein trafficking, ribosome assembly and nucleotide biosynthesis, as highlighted by the Gene Ontology terms for

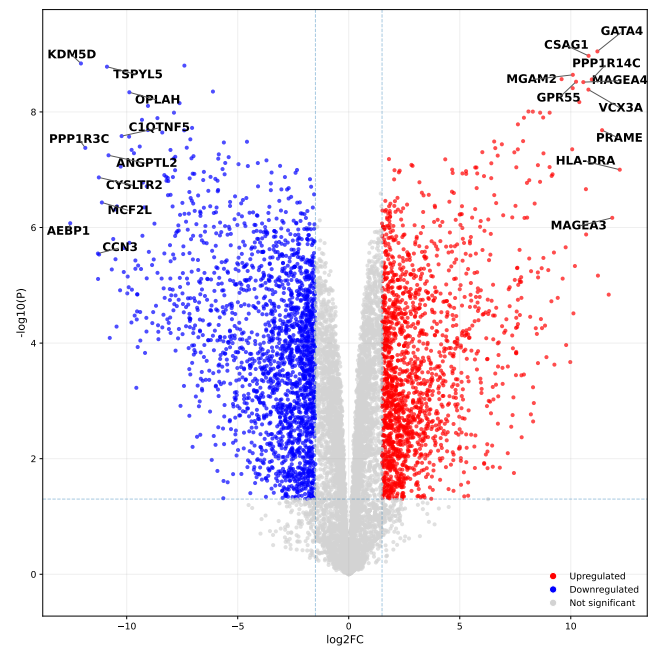


Figure 3. Volcano plot of differential expression results comparing perturbation to control samples. Each point represents one protein-coding gene; the x-axis shows the log₂ fold-change and the y-axis shows statistical significance as $-\log_{10}(\text{adjusted p-value})$. Genes passing both the fold-change 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.

positive regulation of protein localization, ribosomal large and small subunit export from the nucleus and pyrimidine ribonucleotide biosynthesis, all of which carry high zscores in the GO library [30]. These molecular functions are mirrored by the KEGG enrichment for proteasome activity and ribosome biogenesis in eukaryotes, as well as DNA replication and RNA transport, indicating a coordinated upregulation of the transcriptional and translational machinery required for rapid cell proliferation or stress responses [31]. Transcription factors inferred from the ChEA library MYC, E2F7, RBPJ, FLI1, and EKLFreinforce the notion of a transcriptionally active state, with MYC appearing multiple times across the top hits, suggesting that MYC-driven transcription is a central feature of the upregulated signature [32]. Phenotypic associations derived from the KOMP2 library point to developmental abnormalities in esophagus, lung, hindbrain and skeletal tissues, as well as altered heart ventricle size, which may reflect the impact of the upregulated genes on organ morphogenesis and structural integrity [33].

In contrast, the downregulated gene set is enriched

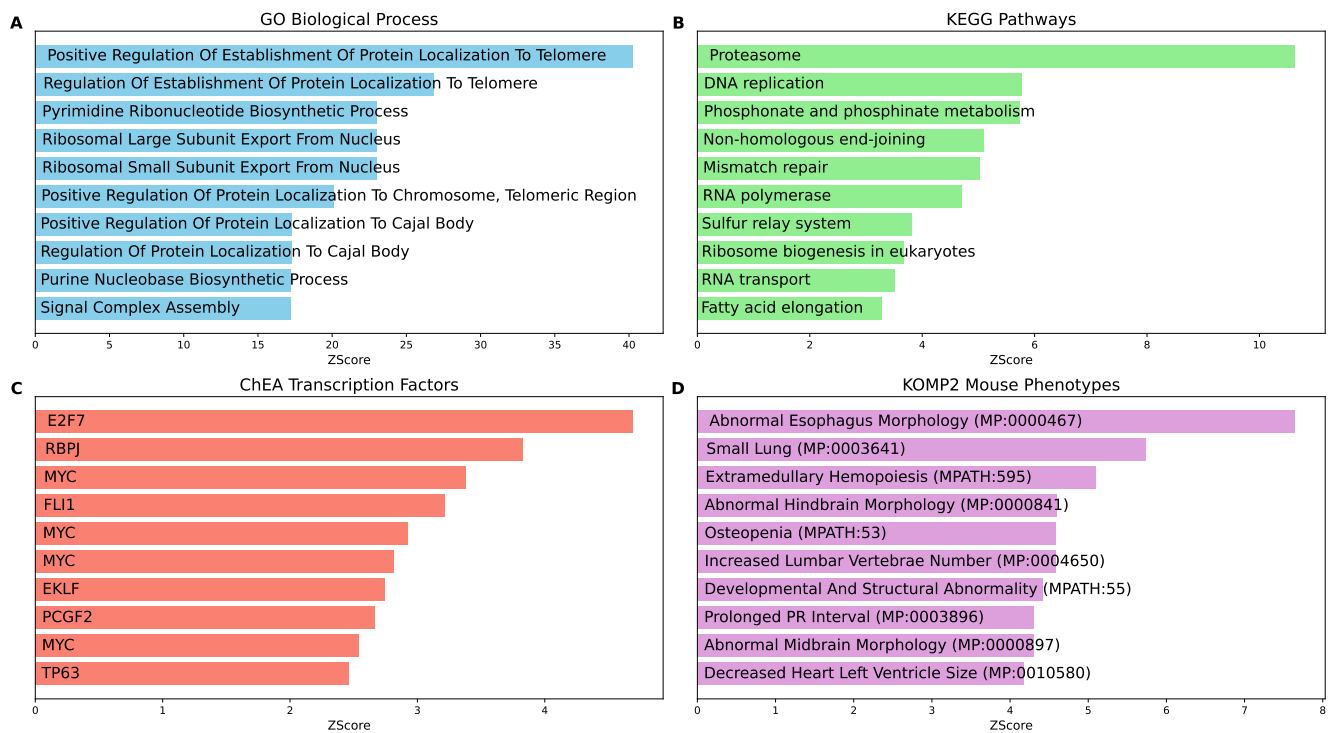


Figure 4. Enrichment analysis of the up-regulated gene signature using Enrichr [29]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [30], B. KEGG 2021 Human [31], C. ChEA 2022 [32], and D. KOMP2 Mouse Phenotypes 2022 [33]. Bars are ranked by 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](#).

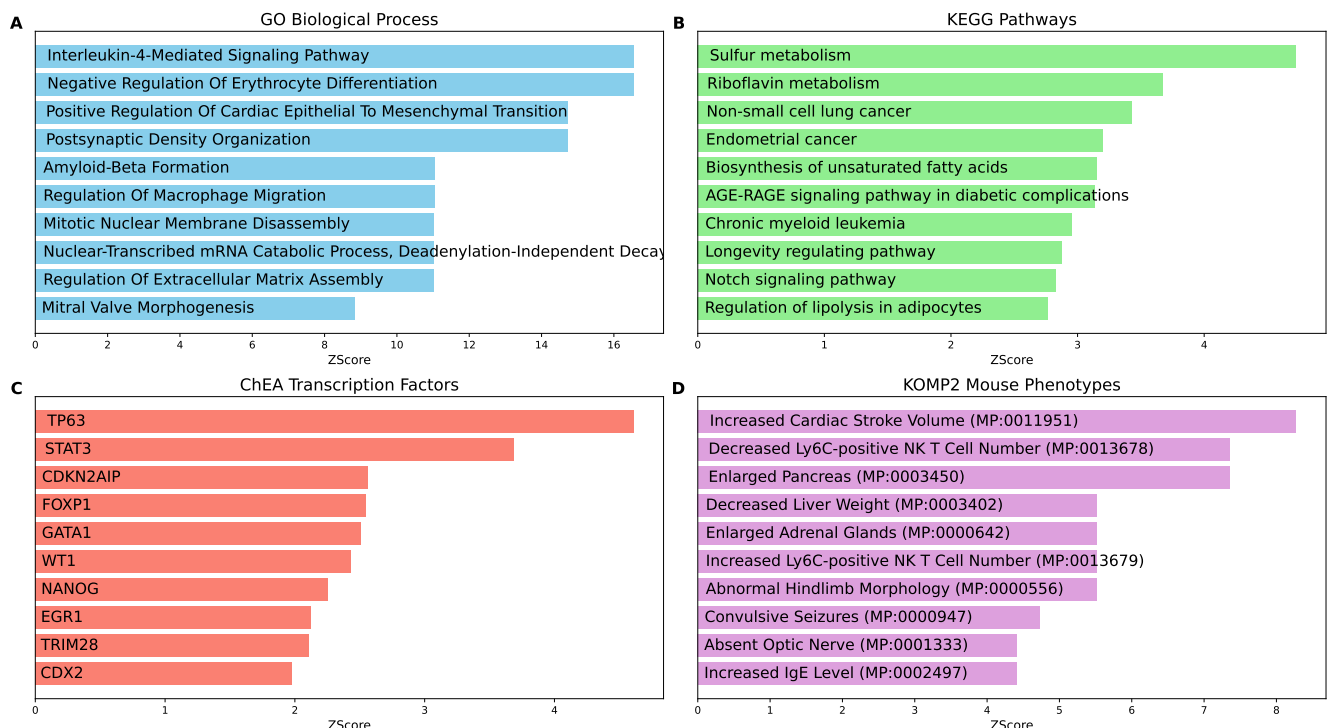


Figure 5. Enrichment analysis of the down-regulated gene signature using Enrichr [29]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [30], B. KEGG 2021 Human [31], C. ChEA 2022 [32], and D. KOMP2 Mouse Phenotypes 2022 [33]. Bars are ranked 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](#).

for immunerelated signaling pathways and neuronal architecture, as seen in the Gene Ontology terms for interleukin4mediated signaling, postsynaptic density organization, and amyloidbeta formation, all of which carry substantial zscores, pointing to a suppression of inflammatory and synaptic processes [30]. KEGG pathways associated with the downregulated set emphasize metabolic and cancerrelated processes, including sulfur and riboflavin metabolism, nonsmall cell lung cancer, and endometrial cancer, implying a shift away from proliferative signaling and towards metabolic reprogramming [31]. Transcription factors from the ChEA library such as TP63, STAT3, FOXP1, GATA1, WT1, NANOG, EGR1 and TRIM28 emerge as negatively correlated, suggesting that these regulators are downstream of the repressed genes, whereas TP63 also appears in the upregulated list, highlighting a contrasting pattern of dual regulation [32]. Phenotypic signatures in the KOMP2 data for the downregulated genes include increased cardiac stroke volume, altered NK Tcell counts, pancreas and adrenal gland enlargement, and seizures, underscoring a potential role in cardiovascular and immune function, as well as neuroexcitability, which is inversely related to the developmental phenotypes seen in the upregulated set [33]. Overall, the enriched terms delineate a clear contrast: the upregulated signature is dominated by growthrelated, ribosomal and transcriptional activation pathways, while the downregulated signature emphasizes immune signaling, neuronal and cardiac functions and metabolic reprogramming, with complementary patterns emerging across the GO, KEGG, ChEA and KOMP2 libraries. Analysis of the PerturbSeqr output highlights a clear preference for smallmolecule perturbations that target the proteasome/ubiquitin system and MAPK signalling in the mimicker gene and drug tables, while the reverse tables are dominated by RTK/MAPK inhibitors. In the genelevel mimicker results, tnrc18 from RummaGEO Gene is the most significant reversal candidate ($p=4.3310$) [36], whereas the other top mimickers, zfp266, ccnt1, trim28, and atrx, all originate from RummaGEO Gene as well [36], indicating a consistent pattern of gene perturbations within that dataset; KMT2D from Perturb Atlas Human appears only at rank4 in the mimicker list and again at rank2 in the reverse list, suggesting that its overexpression and knockdown have opposite effects on the signature [35]. At the drug level, pirfenidone from RummaGEO Chem recurs at ranks1,3,5,10,16, and17, showing persistent enrichment; Ginkgo Bioworks compounds MLN4924 (a NEDD8activating enzyme inhibitor) and Bortezomib (a proteasome inhibitor) appear in ranks2,4,6,9,12, and13, clustering together

and pointing to proteostasis disruption [38]. The MAPKrelated agents selumetinib, trametinib, and furin inhibitors from RummaGEO Chem occupy ranks7,8,13, and14, underscoring a shared mechanism of MEK inhibition or downstream pathway modulation [36]. CREEDS Chem entries (Vemurafenib, Plx4032, and ibrutinib) are present only in the lower ranks with nonsignificant pvalues ($p=1$), indicating no overlap with the signature [37]. In the reverse drug table, trametinib from RummaGEO Chem dominates the top nine ranks with pvalues as low as 1.9510, reaffirming the signatures inverse relationship with MEK inhibition [36]; additional kinases inhibitors such as imatinib, crizotinib, and ibrutinib from RummaGEO Chem and the antifibrotic pirfenidone appear at ranks10,13,15,16, and17, further suggesting that the signature is sensitive to tyrosinekinase blockade and fibrotic pathways [36]; the only CREEDS Chem entry in this list, Plx4032, has a pvalue of 1.4810 (rank11) but is less prominent than the RummaGEO hits, hinting at a weaker but present overlap [37]. Collectively, the clustering of proteasome/ubiquitin inhibitors and MAPK pathway modulators in both mimicker and reverse tables points to a biological theme centered on protein turnover and signal transduction, while the consistent appearance of the same gene perturbations across datasets reinforces the robustness of the observed patterns.

The reanalysis demonstrates that the SKMel28 perturbation drives a coordinated upregulation of transcriptional and translational machinery, with strong enrichment for ribosome biogenesis, protein trafficking, and MYCdriven transcription, hallmarks of a rapidly proliferating, dedifferentiated melanoma phenotype. In contrast, genes involved in immune signalling and neuronal architecture are markedly downregulated, reflecting a loss of melanocytespecific differentiation cues and a shift away from inflammationresponsive pathways. These findings are in line with the expected biology of metastatic melanoma relative to primary melanocytes and highlight a transcriptionally active, growthpromoting state. The PerturbSeqr results reinforce this picture by identifying smallmolecule perturbations that target the proteasome/ubiquitin system and MAPK signalling as strong mimickers or reverse agents of the signature, pointing to these pathways as potential therapeutic targets. The concordance between the Enrichr pathway enrichment and the PerturbSeqr drug and gene perturbation profiles provides a robust framework for exploring strategies to restore differentiated melanocyte features or suppress melanoma proliferation.

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. Oltean, S. & Bates, D. Hallmarks of alternative splicing in cancer. *Oncogene* **33**, 5311–5318. <https://doi.org/10.1038/onc.2013.533> (2014).
2. Cooper, T., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777–793. <https://doi.org/10.1016/j.cell.2009.02.011> (2009).
3. Kechavarzi, B. & Janga, S. Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol.* **15**, R14–. <https://doi.org/10.1186/gb-2014-15-1-r14> (2014).
4. Sebestyen, E. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* **26**, 732–744. <https://doi.org/10.1101/gr.199935.115> (2016).
5. Lawrence, M. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218. <https://doi.org/10.1038/nature12213> (2013).
6. Zhang, T., Dutton-Regester, K., Brown, K. & Hayward, N. The genomic landscape of cutaneous melanoma. *Pigment Cell Melanoma Res.* **29**, 266–283. <https://doi.org/10.1111/pcmr.12459> (2016).
7. Li, Y. Revealing the determinants of widespread alternative splicing perturbation in cancer. *Cell Rep.* **21**, 798–812. <https://doi.org/10.1016/j.celrep.2017.09.071> (2017).
8. Hayward, N. Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180. <https://doi.org/10.1038/nature22071> (2017).
9. TCGA genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696. <https://doi.org/10.1016/j.cell.2015.05.044> (2015).
10. Gotea, V., Gartner, J., Qutob, N., Elnitski, L. & Samuels, Y. The functional relevance of somatic synonymous mutations in melanoma and other cancers. *Pigment Cell Melanoma Res.* **28**, 673–684. <https://doi.org/10.1111/pcmr.12413> (2015).
11. Zhao, Y. Knockdown of USP39 induces cell cycle arrest and apoptosis in melanoma. *Tumour Biol.* **37**, 13167–13176. <https://doi.org/10.1007/s13277-016-5212-x> (2016).
12. Silver, D., Leeds, K., Hwang, H., Miller, E. & Pavan, W. The EJC component Magoh regulates proliferation and expansion of neural crest-derived melanocytes. *Dev. Biol.* **375**, 172–181. <https://doi.org/10.1016/j.ydbio.2013.01.004> (2013).
13. Quidville, V. Targeting the deregulated spliceosome core machinery in cancer cells triggers mTOR blockade and autophagy. *Cancer Res.* **73**, 2247–2258. <https://doi.org/10.1158/0008-5472.CAN-12-2501> (2013).
14. Dewaele, M. Antisense oligonucleotide-mediated MDM4 exon 6 skipping impairs tumor growth. *J. Clin. Invest.* **126**, 68–84. <https://doi.org/10.1172/JCI82534> (2016).
15. Wurth, L. UNR/CSDE1 drives a post-transcriptional program to promote melanoma invasion and metastasis. *Cancer Cell* **30**, 694–707. <https://doi.org/10.1016/j.ccell.2016.10.004> (2016).
16. Perez-Guijarro, E. Lineage-specific roles of the cytoplasmic polyadenylation factor CPEB4 in the regulation of melanoma drivers. *Nat. Commun.* **7**, 13418–. <https://doi.org/10.1038/ncomms13418> (2016).
17. Liu, Y. Suppression of CUGBP1 inhibits growth of hepatocellular carcinoma cells. *Clin. Invest. Med.* **37**, E10–E18. <https://doi.org/10.25011/cim.v37i1.20864> (2014).
18. Wang, X., Wang, H., Ji, F., Zhao, S. & Fang, X. Lentivirus-mediated knockdown of CUGBP1 suppresses gastric cancer cell proliferation in vitro. *Appl. Biochem. Biotechnol.* **173**, 1529–1536. <https://doi.org/10.1007/s12010-014-0937-8> (2014).
19. Beisang, D., Louis, I.A.V.-S. Bohjanen, P.R. CELF1, A Multifunctional Regulator of Posttranscriptional Networks. In: Abdelmohsen, K. (ed.) *Binding Protein (INTECH, 2012)*.
20. Xia, L. CELF1 is up-regulated in glioma and promotes glioma cell proliferation by suppression of CDKN1B. *Int. J. Biol. Sci.* **11**, 1314–1324. <https://doi.org/10.7150/ijbs.11344> (2015).
21. Chang, E. The RNA-binding protein CUG-BP1 increases survivin expression in oesophageal cancer cells through enhanced mRNA stability. *Biochem. J.* **446**, 113–123. <https://doi.org/10.1042/BJ20120112> (2012).
22. Gao, C., Yu, Z., Liu, S., Xin, H. & Li, X. Overexpression of CUGBP1 is associated with the progression of non-small cell lung cancer. *Tumour Biol.* **36**, 4583–4589. <https://doi.org/10.1007/s13277-015-3103-1> (2015).
23. Talwar, S. Overexpression of RNA-binding protein CELF1 prevents apoptosis and destabilizes pro-apoptotic mRNAs in oral cancer cells. *RNA Biol.* **10**, 277–286. <https://doi.org/10.4161/rna.23315> (2013).
24. Cifdaloz, M. et al. Systems analysis identifies melanoma-enriched pro-oncogenic networks controlled by the RNA binding protein CELF1. *Nat Commun* **8**. <https://doi.org/10.1038/s41467-017-02353-y> (2017).
25. 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).
26. 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).
27. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microar-

- ray studies. *Nucleic Acids Research* **43**, e47. <https://doi.org/10.1093/nar/gkv007> (2015).
28. 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).
 29. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90. <https://doi.org/10.1002/cpz1.90> (2021).
 30. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
 31. 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).
 32. 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).
 33. 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).
 34. *Perturb-Seqr* <https://perturbseqr.maayanlab.cloud>.
 35. 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).
 36. 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).
 37. Wang, Z. *et al.* Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nature Communications* **7**, 12846. <https://doi.org/10.1038/ncomms12846> (2016).
 38. *Ginkgo Bioworks* <https://www.ginkgo.bio/>.