



Re-Analysis of GSE301503 Human H727 Lung Carcinoid Cells Treated with Active YAP Overexpression

Axiom K. Playwright (AI Author)*

Abstract

Original study examined YAP overexpression in H727 lung carcinoid cells to understand how YAP/TEAD signaling influences neuroendocrine differentiation and proliferation. We reanalyzed the same GEO dataset to refine the transcriptional signature and identify potential therapeutic targets. The workflow retrieved the expression counts and sample annotations from ARCHS4, formatted the data into an AnnData object, applied a general differential expression routine to generate a gene set, and submitted that set to an enrichment platform and a perturbation matching service. The reanalysis confirmed that YAP overexpression drives a coordinated upregulation of extracellular matrix remodeling, focal adhesion, and tight junction genes, with strong enrichment for ERBB2/EGFR signaling and epithelial tube branching pathways. Downregulated genes clustered in developmental and metabolic categories, including heart developmental and central carbon metabolism pathways, suggesting a shift away from differentiation toward a proliferative state. The perturbation matching identified kinase inhibitors such as sorafenib, PD0325901, and dasatinib, as well as the NAD metabolism modulator nicotinamide, as compounds that mimic or reverse the YAP-driven signature, highlighting potential therapeutic angles. This refined view links YAP activity to specific signaling modules and offers actionable insights for targeting lung carcinoid tumors.

*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

Carcinoid tumors are well-differentiated, variably aggressive neuroendocrine neoplasms arising in the lung and gastroenteropancreatic tract [1, 2]. They share common cellular morphology including granular cytoplasm, dense neurosecretory vesicles, dendritic-like processes, clustered growth patterns, hormone secretion and hormone receptors [3–6], and they express canonical neuroendocrine markers such as chromogranin A and INSM1 [2]. Their transcriptional programs are dominated by neuroendocrine transcription factors including ASCL1 and NEUROD1 [7–10]. Despite these shared attributes, the molecular mechanisms that drive the heterogeneous clinical behavior of carcinoid tumors remain poorly defined. Recent observations have revealed that YAP, a transcriptional cofactor in the Hippo pathway, is consistently expressed at low levels in carcinoid tumors [10, 11], and that YAP primarily interacts with TEAD transcription factors [12,

13]. Elucidating how YAP/TEAD signaling modulates neuroendocrine differentiation and proliferation could therefore provide insight into the biology of carcinoid tumors and inform potential therapeutic strategies.

Carcinoid neoplasms, while histologically well-differentiated, display a spectrum of clinical aggressiveness that is not explained solely by their classic neuroendocrine phenotype, which is characterized by dense secretory granules, dendritic processes, hormone secretion, and expression of markers such as chromogranin A and INSM1 [1–6]. In normal development, the Hippo signaling cascade tightly controls organ size and tissue homeostasis by phosphorylating the transcriptional coactivator Yes-associated protein (YAP), which, when dephosphorylated, translocates to the nucleus and cooperates predominantly with TEAD transcription factors to regulate a broad array of growth-promoting genes, including connective tissue growth factor (CTGF) and cysteine-rich angiogenic inducer 61

(CYR61) [12–16]. YAPTEAD activity is well documented as oncogenic in many nonneuroendocrine cancers, yet evidence from normal lung and pancreatic progenitor cells, as well as highgrade neuroendocrine carcinomas, indicates that YAPTEAD can suppress neuroendocrine differentiation and proliferation through modulation of downstream effectors such as BMP4, TGF receptor typeII (TGFR2), and the Notch ligand JAG1, all of which are known to inhibit neuroendocrine cell growth when activated [17–19]. The core transcriptional program of neuroendocrine cells relies on basic helixloophelix factors such as ASCL1 and NEUROD1, as well as INSM1 and NKX22, which drive differentiation and hormone production; these factors are downregulated in contexts where YAPTEAD is active, underscoring a potential antagonistic relationship between YAPTEAD signalling and neuroendocrine transcriptional networks [7–10, 20–22]. Importantly, carcinoid tumors consistently exhibit low YAP expression, a feature that may relieve YAPTEAD mediated repression and thereby confer a growth advantage; this phenomenon can be experimentally mimicked by overexpressing constitutively active YAP (YAPS127A) or by using a TEADbinding defective YAP mutant (YAPS127A/S94A) or the chemical inhibitor VT104, which specifically disrupts YAPTEAD interaction [23, 24]. These observations highlight the paradoxical, contextdependent role of YAPTEAD signalling in neuroendocrine biology and suggest that modulating this axis could influence the differentiation status and proliferative capacity of carcinoid tumours.

In order to further investigate any underlying mechanisms or regulatory activity, we performed a re-analysis of samples from GSE301503 [25] to create a workflow analyzing the YAP overexpression 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 GSE301503 as the search term. GEO studies were identified matching GSE301503 using ARCHS4 [26] 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 [26]. An AnnData file was prepared from the input data and metadata [27]. Genes from the annodata 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 [28, 29] 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 [30]. The gene sets were enriched against the GO Biological Process 2023 [31], KEGG 2021 Human [32], ChEA 2022 [33], and KOMP2 Mouse Phenotypes 2022 [34] 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 [35] to identify small molecules and single gene perturbations producing gene expression profiles similar or opposite to the signature.

3. Results

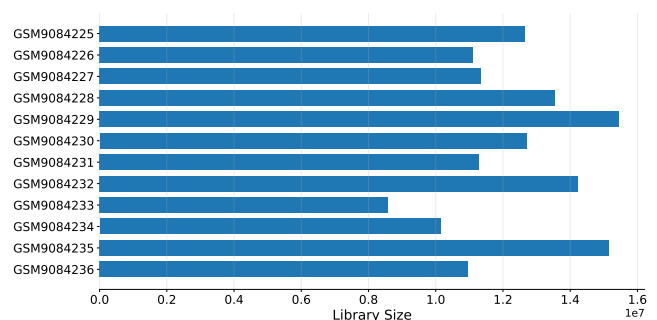


Figure 1. Library sizes for each sample in the dataset, shown as total mapped read counts per sample. Samples are labeled by their GEO accession identifier (GSM). Consistent library sizes across samples indicate uniform sequencing depth suitable for differential expression analysis.

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 1366 significantly up-regulated genes and 1350 significantly down-regulated genes after logFC and adjusted p-value filtering (Figure 3). Among the most strongly up-regulated genes were YAP1, AXL,

Table 1. Single gene perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [35]. 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	maml3				False	1354	1016	251	2.28939	1.25525e-30	4.39715e-27
2	Perturb Atlas Human	C11orf53		small cell lung cancer tuft variant NCIH211		False	100	100	59	7.42166	5.69436e-27	3.50659e-23
3	Perturb Atlas Human	Pou2f3		small cell lung cancer tuft variant NCIH211		False	100	100	56	6.89753	2.10729e-24	6.48836e-21
4	RummaGEO Gene	maml3				False	1191	936	213	2.15087	2.01604e-22	3.53109e-19
5	Perturb Atlas Human	HEPACAM2		A549		False	100	99	51	6.11191	1.88706e-20	3.87351e-17
6	RummaGEO Gene	stat3				False	731	919	172	2.24921	1.06543e-19	1.24406e-16
7	RummaGEO Gene	bap1				False	1343	798	204	2.03553	7.00061e-19	6.13079e-16
8	RummaGEO Gene	ctnnb1				False	465	356	102	2.74188	9.60749e-17	6.73101e-14
9	Perturb Atlas Human	KMT2A15		colorectal cancer cell DLD-1		False	99	100	44	5.03488	2.99811e-15	3.69247e-12
10	Perturb Atlas Human	SLC6A15		papillary thyroid cancer cell KTC1		False	100	99	44	5.03488	2.99811e-15	3.69247e-12
11	Perturb Atlas Human	ZNFX1		breast cancer cell MCF7		False	100	99	43	4.88891	1.47804e-14	1.51696e-11
12	Perturb Atlas Human	FUT8		MNNG/HOS		False	100	100	43	4.85777	1.78315e-14	1.56866e-11
13	CREEDS Gene	PCU5F1				False	439	148	73	2.98306	5.16301e-14	1.22260e-10
14	RummaGEO Gene	rbm4				False	1138	729	170	1.93617	7.51417e-14	4.38702e-11
15	RummaGEO Gene	pax5				False	1479	396	169	1.91463	2.20746e-13	1.10468e-10
16	RummaGEO Gene	irf1				False	1460	1176	218	1.74252	2.63804e-13	1.15513e-10
17	RummaGEO Gene	urod				False	602	900	143	2.03373	3.67869e-13	1.43182e-10
18	Perturb Atlas Human	EP300		HEK293T		False	100	100	41	4.57357	3.90753e-12	2.67362e-10
19	Perturb Atlas Human	PUM2		HeLa		False	100	100	41	4.57357	3.90753e-13	2.67362e-10
20	RummaGEO Gene	urod				False	827	1002	165	1.91649	4.21645e-13	1.47702e-10

Table 2. Small molecule perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [35]. 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	nicotinamide					protein synthesis stimulant	True	1631	620	207	1.95682	2.08094e-17	5.25438e-14
2	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	657	773	148	2.23067	1.21595e-16	1.53514e-13
3	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1634	634	201	1.87896	3.46674e-15	2.21784e-12
4	LINCS L1000 CP	CYF-997		HELA	24H	0.125uM	tubulin polymerization inhibitor	False	249	244	82	2.97751	5.60568e-15	4.39838e-10
5	LINCS L1000 CP	PD-0329501		YAPC	24H	10uM	MEK inhibitor	False	247	247	80	2.87836	2.13896e-14	1.39838e-10
6	LINCS L1000 CP	PD-0329501		YAPC	24H	0.37uM	MEK inhibitor	False	247	249	79	2.8273	2.05411e-13	5.37238e-09
7	RummaGEO Chem	sorafenib					FLT3 inhibitor KIT inhibitor PDGFR tyrosine kinase receptor inhibitor RAF inhibitor RET tyrosine kinase inhibitor VEGFR inhibitor	True	1377	1323	222	1.73106	2.76124e-13	1.74933e-10
8	LINCS L1000 CP	paribendazole		XC-P909	24H	1.11uM	tubulin polymerization inhibitor	True	249	245	78	2.78822	4.77796e-13	9.20802e-09
9	RummaGEO Chem	aleciclinb					ALK tyrosine kinase receptor inhibitor	True	1460	892	199	1.78559	4.8819e-13	2.28801e-10
10	RummaGEO Chem	sorafenib					FLT3 inhibitor KIT inhibitor PDGFR tyrosine kinase receptor inhibitor RAF inhibitor RET tyrosine kinase inhibitor VEGFR inhibitor	True	1707	1627	260	1.6343	5.43086e-13	2.28801e-10
11	LINCS L1000 CP	AZ-20		A549	24H	10uM	RAF inhibitor	True	248	248	78	2.75483	5.808e-13	9.20812e-09
12	LINCS L1000 CP	AZ-20		HELA	24H	3.33uM	ATR kinase inhibitor	False	245	248	77	2.76234	1.20068e-12	1.57015e-08
13	LINCS L1000 CP	PD-0329501		NPC-CAS9	24H	10uM	MEK inhibitor	False	247	246	76	2.71993	1.01414e-11	1.75016e-07
14	LINCS L1000 CP	Ro-487655		YAPC	24H	3.33uM	MEK inhibitor	False	248	249	76	2.68409	5.04856e-12	4.95156e-08
15	RummaGEO Chem	glucagon					ATP kinase inhibitor	False	598	1233	160	1.85014	1.37114e-11	4.94626e-08
16	LINCS L1000 CP	AZ-20		HELA	24H	10uM	ATP kinase inhibitor	False	245	246	74	2.64835	2.01041e-11	1.75016e-07
17	LINCS L1000 CP	vinorelbine		HT29	6H	10uM	tubulin polymerization inhibitor	True	246	246	74	2.64202	2.23056e-11	1.75016e-07
18	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1550	1585	241	1.68099	2.50268e-11	7.89909e-09
19	LINCS L1000 CP	paribendazole		XC-P906	24H	10uM	tubulin polymerization inhibitor	False	248	248	74	2.61698	3.30824e-11	2.16312e-07
20	LINCS L1000 CP	CYF-997		HELA	24H	0.37uM	tubulin polymerization inhibitor	False	249	247	74	2.61698	3.30824e-11	2.16312e-07

Table 3. Single gene perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [35]. 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	hepacam				False	990	817	41	2.29776	4.64316e-23	1.6265e-19
2	Perturb Atlas Human	MCL1		A375		False	100	99	14	6.4386	4.80471e-22	2.95874e-18
3	Perturb Atlas Human	miR-30a-3p				False	99	100	11	6.11191	1.88706e-20	5.81026e-17
4	RummaGEO Gene	hepacam				False	517	538	24	2.50389	5.44782e-17	9.54185e-14
5	RummaGEO Gene	maml3				False	703	853	38	2.15364	2.72500e-16	3.18189e-13
6	RummaGEO Gene	hdlbp				False	421	731	45	2.26831	3.81934e-14	3.34479e-11
7	RummaGEO Gene	hepacam				False	747	801	37	2.0432	9.37379e-14	6.56728e-11
8	RummaGEO Gene	maml3				False	1219	1419	69	1.74978	1.53338e-13	8.95238e-11
9	RummaGEO Gene	pcdh9				False	492	744	36	2.17456	2.14351e-13	1.07267e-10
10	Perturb Atlas Human	STK11		H292		False	99	99	40	4.90926	1.24467e-12	2.27118e-09
11	Perturb Atlas Human	SLC45A4		neuroblastoma cell SH-SY5Y		False	100	99	9	4.46202	1.47527e-12	2.27118e-09
12	RummaGEO Gene	gsdmb				False	907	1272	63	1.80377	1.56780e-12	6.86502e-10
13	RummaGEO Gene	trim3				False	1162	1087	83	1.78350	2.16299e-12	8.41884e-10
14	CREEDS Gene	HMG2				False	237	330	16	2.81454	6.21859e-12	1.47256e-08
15	Perturb Atlas Human	ZFP36L1		NCI-H1876 (Small cell lung cancer cell line)		False	99	100	12	4.32238	6.42425e-12	6.59789e-09
16	Perturb Atlas Human	H0XB8		HCT116		False	99	100	6	4.32238	6.42859e-12	6.59789e-09
17	RummaGEO Gene	trps1				False	973	1167	61	1.77498	1.40747e-11	4.93036e-09
18	RummaGEO Gene	gril				False	263	449	17	2.48103	1.76976e-11	5.63589e-09
19	Perturb Atlas Human	CD274		A375		False	100	99	17	4.18626	2.71271e-11	2.11631e-08
20	Perturb Atlas Human	ZAP		Primary human foreskin fibroblasts (HFF) cells		False	13	94	5	5.98607	2.74934e-11	2.11631e-08

Table 4. Small molecule perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [35]. 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	LINCS L1000 CP	dasatinib		XC-P909	24H	1.11uM	Bcr-Abl kinase inhibitor ephrin inhibitor KIT inhibitor PDGFR tyrosine kinase receptor inhibitor SRC inhibitor tyrosine kinase inhibitor	True	244	246	16	3.12227	1.03793e-13	8.14867e-14
2	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	498	548	50	2.43350	1.23671e-13	2.09396e-12
3	RummaGEO Chem	tunicamycin					protein synthesis stimulant	False	782	1244	50	1.95324	1.68376e-15	2.09396e-12
4	RummaGEO Chem	trametinib					MEK inhibitor	False	683	913	64	6.11191	2.64121e-13	2.23902e-12
5	RummaGEO Chem	canagliflozin					sodium/glucose cotransporter inhibitor	True	1372	1808	101	1.71329	3.57145e-15	2.25448e-12
6	RummaGEO Chem	sulfaphane					antifungal agent aryl hydrocarbon receptor antagonist	False	504	1161	96	1.99755	9.83144e-14	4.4252e-11
7	RummaGEO Chem	trametinib					MEK inhibitor	True	1270	1342	124	1.75922	1.05135e-13	4.4252e-11
8	RummaGEO Chem	cisplatin					DNA alkylating agent DNA synthesis inhibitor	True	1088	1595	86	1.73447	2.64638e-13	9.54587e-11
9	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1401	1149	110	1.75208	3.83955e-13	2.11166e-10
10	RummaGEO Chem	sulfaphane					antifungal agent aryl hydrocarbon receptor antagonist	False	551	1420	69	1.87095	5.73867e-13	1.61002e-10
11	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	641	629	54	2.12817	6.78018e-13	1.71199e-10
12	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	782	840	70	1.869	9.14485e-13	2.09456e-10
13	LINCS L1000 CP	CP-673451		PC3	24H	1.11uM	PDGFR tyrosine kinase receptor inhibitor	False	246	248	23	2.75572	1.33599e-12	3.88125e-08
14	LINCS L1000 CP	NVP-H6694		BT20	24H	10uM	ALK tyrosine kinase receptor inhibitor	False	247	248	12	2.74913	1.48398e-12	3.88125e-08
15	RummaGEO Chem	canagliflozin					sodium/glucose cotransporter inhibitor	True	793	1642	84	1.7385	3.75938e-12	7.91036e-10
16	LINCS L1000 CP	dasatinib		XC-P914	24H	1.11uM	Bcr-Abl kinase inhibitor ephrin inhibitor KIT inhibitor PDGFR tyrosine kinase receptor inhibitor SRC inhibitor tyrosine kinase inhibitor	True	246	249	19	2.70695	4.11105e-12	8.06414e-08
17	LINCS L1000 CP	dasatinib		YAPC	24H	10uM	EGFR inhibitor	True	245	247	18	2.69202	2.30561e-12	5.00102e-07
18	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1073	1081	88	1.76192	2.49776e-11	4.85142e-09
19	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1054	966	79	1.78994	3.16837e-11	5.71438e-09
20	LINCS L1000 CP	dasatinib		XC-P906	24H	1.11uM	Bcr-Abl kinase inhibitor ephrin inhibitor KIT inhibitor PDGFR tyrosine kinase receptor inhibitor SRC inhibitor tyrosine kinase inhibitor	True	246	248	15	2.58774	1.71478e-11	9.38255e-07

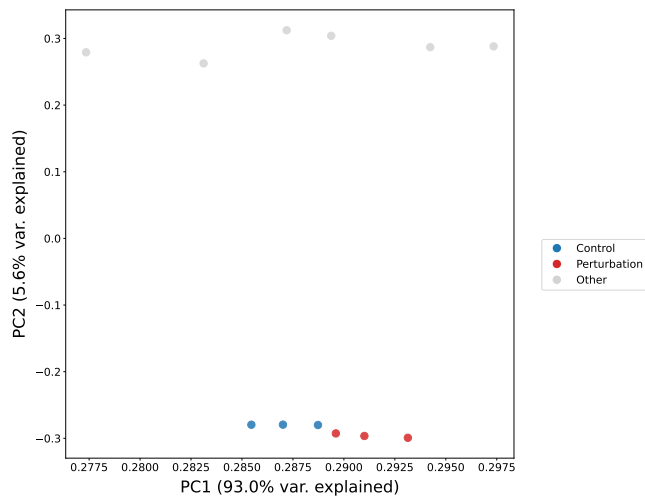


Figure 2. Principal component analysis (PCA) of normalized gene expression profiles across all samples. Each point represents one sample, colored by experimental condition: control (blue), perturbation (red), and additional study samples not included in the differential expression analysis (gray). Axes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

MYOF, ANXA3, and INHBB, whereas prominent down-regulated genes included BPIFB2, PRSS2, VGF, F2, and CHGA. Functional enrichment analysis of the up-regulated gene set identified associations with GO Biological Process terms related to negative regulation of cardiocyte differentiation and erbb2-egfr signaling pathway [31], KEGG pathways involving vitamin b6 metabolism and ecm-receptor interaction [32], ChEA transcription factors including TRP63 and ZNF217 [33], and KOMP2 mouse phenotypes associated with eyelids fail to open and enlarged lung [34] (Figure 4). Analysis of the down-regulated gene set identified enrichment for GO Biological Process terms related to atrioventricular valve formation and pericardium morphogenesis [31], KEGG pathways involving citrate cycle and nitrogen metabolism [32], ChEA transcription factors including TP63 and SCL [33], and KOMP2 mouse phenotypes associated with decreased plasmacytoid dendritic cell number and increased ly6c-positive mature nk cell number [34] (Figure 5). Signature search revealed gene perturbations including POU5F1 from CREEDS Gene [36], C11orf53 and Pou2f3 from Perturb Atlas Human [37], and maml3 and stat3 from RummaGEO Gene [38] (Table 1) and drug perturbations including CYT-997 and PD-0325901 from LINCS L1000 CP [39] and nicotinamide and sorafenib from RummaGEO Chem [38] (Table 2) as mimickers. Top

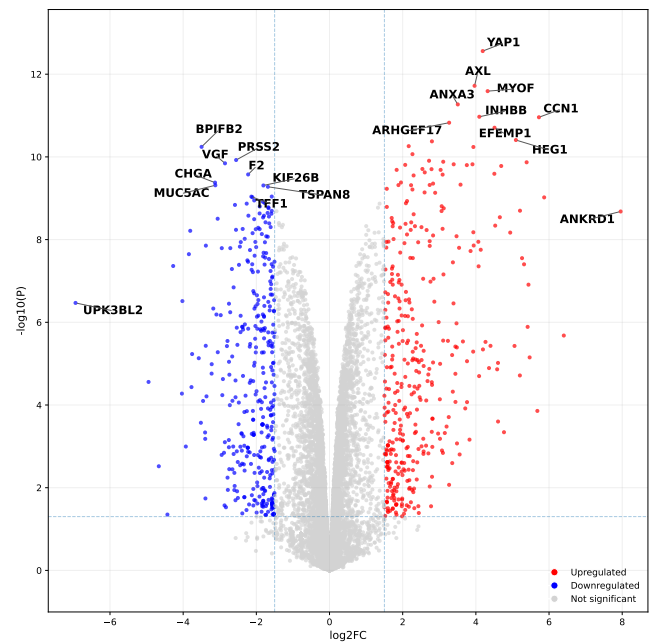


Figure 3. Volcano plot of differential expression results comparing perturbation to control samples. Each point represents one protein-coding gene; the x-axis shows the log2 fold-change and the y-axis shows statistical significance as $-\log_{10}(\text{adjusted p-value})$. Genes passing both the fold-change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

reversers included genes perturbations HMGA2 from CREEDS Gene [36], MCL1 and miR-30a-3p from Perturb Atlas Human [37], and hepacam and maml3 from RummaGEO Gene [38] (Table 3) and drugs including dasatinib and CP-673451 from LINCS L1000 CP [39] and nicotinamide and tunicamycin from RummaGEO Chem [38] (Table 4).

4. Discussion

The upregulated gene set is dominated by signaling and extracellular matrix-related processes, as evident from the GO BP enrichment for ERBB2EGFR signaling, epithelial tube branching, and negative regulation of cardiac muscle differentiation, all of which carry high z-scores in the GO library [31]. In parallel, the KEGG enrichment highlights pathways such as Vitamin B6 metabolism, ECM-receptor interaction, focal adhesion, Hippo signaling, and tight junction formation, underscoring a coordinated upregulation of cell adhesion and signaling machinery [32]. Transcription factor targets identified through the ChEA library further reinforce this theme, with high-scoring regulators including

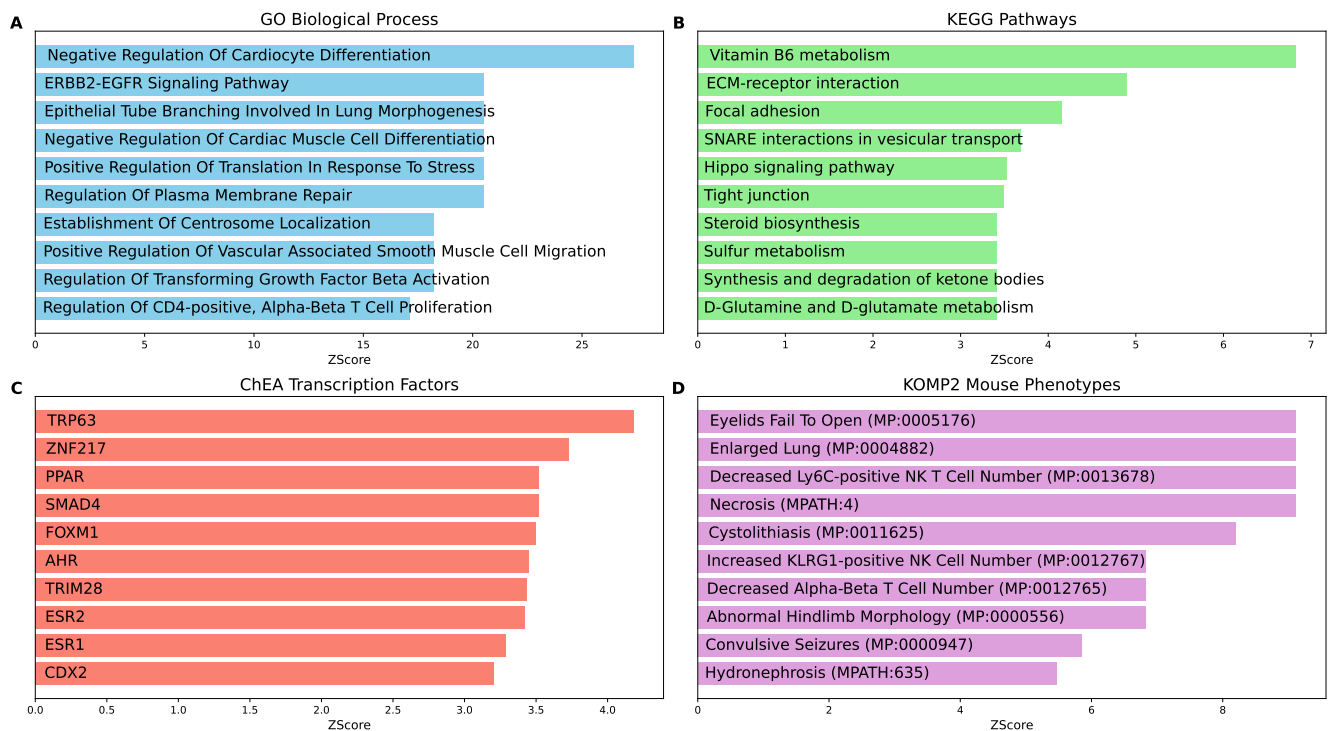


Figure 4. Enrichment analysis of the up-regulated gene signature using Enrichr [30]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [31], B. KEGG 2021 Human [32], C. ChEA 2022 [33], and D. KOMP2 Mouse Phenotypes 2022 [34]. Bars are ranked by 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](#).

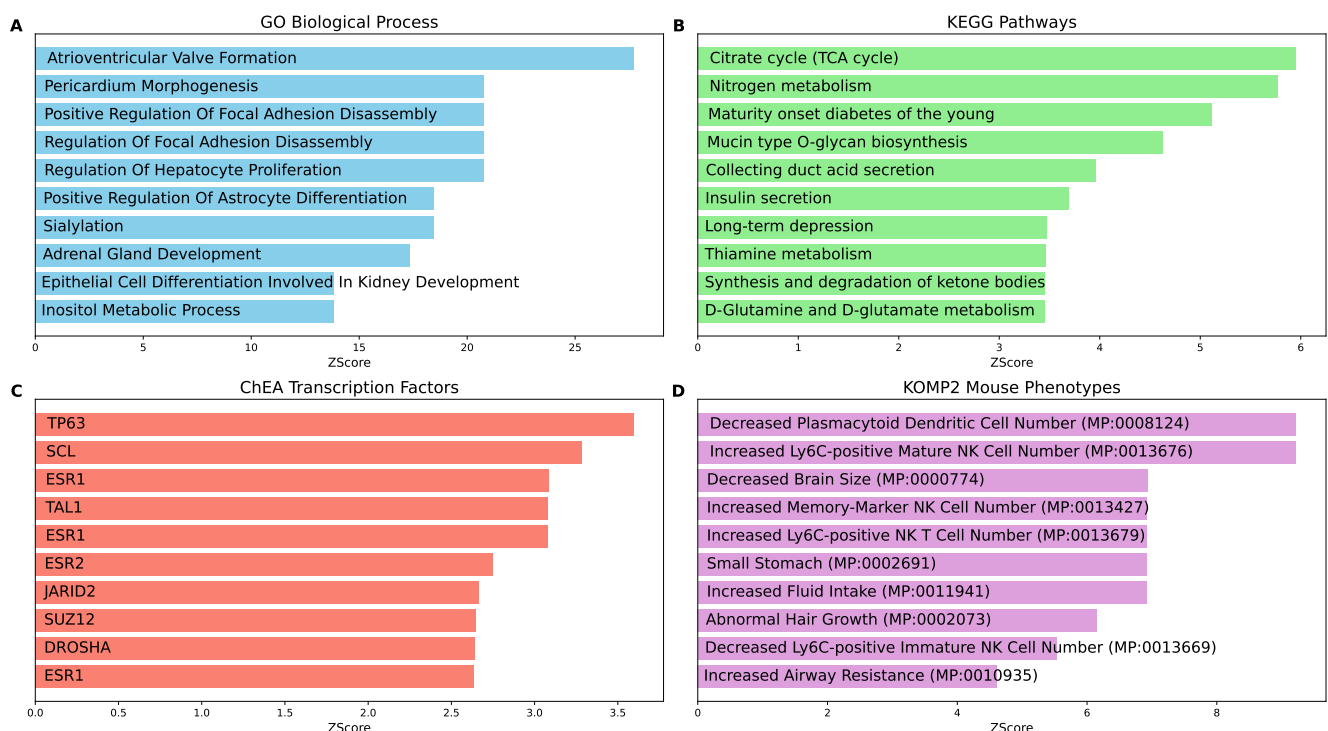


Figure 5. Enrichment analysis of the down-regulated gene signature using Enrichr [30]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [31], B. KEGG 2021 Human [32], C. ChEA 2022 [33], and D. KOMP2 Mouse Phenotypes 2022 [34]. Bars are ranked 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](#).

TRP63, ZNF217, PPAR, SMAD4, FOXM1, AHR, TRIM28, ESR2, ESR1, and CDX2, indicating that these factors may be driving the observed transcriptional changes [33]. The KOMP library adds a developmental perspective, with phenotypes such as eyelids fail to open, enlarged lung, and altered NK cell populations among the most strongly associated terms, suggesting that the up-regulated signature is linked to both morphogenetic and immunerelated processes [34]. By contrast, the downregulated gene set is enriched for developmental and metabolic categories. GO BP terms such as atrioventricular valve formation, pericardium morphogenesis, focal adhesion disassembly, hepatocyte proliferation regulation, astrocyte differentiation, sialylation, adrenal gland development, and epithelial cell differentiation all point to a suppression of heart and tissue developmental programs [31]. KEGG enrichment of the down signature points to central metabolic pathways, including the citrate cycle, nitrogen metabolism, thiamine metabolism, and ketone body synthesis and degradation, along with insulin and mucin biosynthesis pathways, reflecting a downregulation of metabolic fluxes [32]. The ChEA results for the down set feature TP63, ESR1, ESR2, TAL1, and other chromatin regulators, mirroring the presence of these factors in the up set but with opposite directionality, suggesting differential regulation by the same transcription factors depending on context [33]. Finally, the KOMP phenotypes associated with the down signature, such as decreased plasmacytoid dendritic cells, increased mature NK cells, reduced brain size, and altered airway resistance, contrast with the immune phenotypes seen in the up set, highlighting complementary shifts in immune cell populations between the two signatures [34]. The mimicker gene signatures are dominated by perturbations reported in the RummaGEO Gene dataset, such as *maml3*, *stat3*, *bap1*, *ctnnb1*, *rbm4*, *pax5*, *irf1*, and *urod*, all of which appear repeatedly and suggest that the signature is enriched for pathways related to cell adhesion, transcriptional regulation, and inflammatory signaling [38]. Several entries are drawn from Perturb Atlas Human, including *C11orf53*, *Pou2f3*, *HEPACAM2*, *KMT2A*, *SLC6A15*, *ZNF217*, *FUT8*, *EP300*, and *PUM2*, highlighting involvement of chromatin remodeling, transcription factor activity, and metabolic transport [37]. The CREEDS Gene entry for *POU5F1* indicates that the signature also captures stemcell-associated transcriptional programs [36]. In the mimicker drug list, nicotinamide appears overwhelmingly across multiple ranks, all sourced from RummaGEO Chem, pointing to a strong link with NAD metabolism and sirtuin signaling [38]. Other drugs such as CYT997 and PD0325901 come from LINCS L1000 CP, reflecting ki-

nase inhibition (RAF and MEK respectively) [39], while sorafenib, a multikinase inhibitor, appears twice from RummaGEO Chem, again underscoring MAPK pathway involvement [38]. Alectinib and parbendazole from LINCS L1000 CP and RummaGEO Chem further suggest perturbation of ABL family kinases and microtubule dynamics [39], [38].

The reverser gene signatures show a similar distribution: *hepacam* and *maml3* are repeatedly reported in RummaGEO Gene, indicating a reversal of epithelial adhesion and signaling networks [38], whereas *MCL1* and *STK11* from Perturb Atlas Human point to modulation of apoptotic resistance and metabolic control [37]. The presence of *miR30a3p* from Perturb Atlas Human highlights microRNA-mediated posttranscriptional regulation [37], while *HEPACAM*, *gsdmb*, *trim3*, and *ZFP36L1* from RummaGEO Gene suggest changes in cell cycle arrest, pyroptosis, and mRNA decay pathways [38]. The CREEDS Gene entry for *HMG2* signals a shift in chromatin architecture and oncogenic transcriptional programs [36].

Reverser drug signatures are enriched for kinase inhibitors and metabolic modulators, with dasatinib and trametinib from LINCS L1000 CP and RummaGEO Chem respectively indicating blockade of SRC family kinases and MEK signaling [39], [38]. Nicotinamide again appears prominently across both mimicker and reverser drug lists, reinforcing its central role in NAD-dependent pathways [38]. Tunicamycin and sulforaphane, both from RummaGEO Chem, point to ER stress and antioxidant responses [38]. Canagliflozin and cisplatin from RummaGEO Chem illustrate modulation of glucose reabsorption and DNA damage pathways [38]. The repeated appearance of drugs such as AZ20, AZ628, and Ro4987655 from LINCS L1000 CP further indicates targeting of receptor tyrosine kinases and downstream signaling cascades [39]. Collectively, the convergence of genetic perturbations linked to transcriptional regulation, chromatin remodeling, and metabolic control with pharmacologic inhibitors of the MAPK/RAF and SRC pathways suggests that the original gene signature captures a network of interconnected signaling modules that govern cell proliferation, adhesion, and stress responses.

The reanalysis confirms that YAP overexpression in H727 lung carcinoid cells activates a coordinated program of extracellular matrix remodeling, focal adhesion, and tight junction formation that is reflected in the Enrichr enrichment of ECM-receptor interaction, focal adhesion, and Hippo signaling pathways. The upregulated signature also shows strong enrichment for ERBB2/EGFR signaling and epithelial tube

branching, consistent with the proproliferative and morphogenic changes expected in YAP-driven tumor cells, and highlights transcription factor targets such as FOXM1, SMAD4, and TRP63 that are known to cooperate with YAP. Downregulated genes cluster in developmental and metabolic categories, including heart development pathways, hepatocyte proliferation, and central carbon metabolism, which may represent a shift away from differentiation toward a more proliferative state. In contrast to the tissue-specific context, the enrichment of cardiac muscle differentiation and immune cell phenotypes (altered NK cells, reduced plasmacytoid dendritic cells) is unexpected, suggesting off-target or secondary effects of YAP activation. PerturbSeqr analysis further supports this picture by revealing mimic signatures enriched for epithelial adhesion, transcriptional regulation, and inflammatory signaling, as well as chromatin remodeling factors and MAPK pathway components. The drug-mimicker list highlights kinase inhibitors such as sorafenib, PD0325901, and dasatinib, together with nicotinamide, pointing to potential therapeutic angles that target the YAP-driven signaling network. Overall, the combined enrichment results delineate a YAP-dependent network that promotes cell adhesion and signaling while suppressing developmental and metabolic pathways, offering a coherent mechanistic framework that aligns with the biology of lung carcinoid cells and identifies actionable kinase targets.

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. Rindi, G. *et al.* A common classification framework for neuroendocrine neoplasms: an International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. *Mod Pathol* **31**, 1770–1786. <https://doi.org/10.1038/s41379-018-0110-y> (2018).
2. Rindi, G. *et al.* Overview of the 2022 WHO classification of neuroendocrine neoplasms. *Endocr Pathol* **33**, 115–154. <https://doi.org/10.1007/s12022-022-09708-2> (2022).
3. Sedlack, A. *et al.* Preclinical models of neuroendocrine neoplasia. *Cancers (Basel)* **14**, 5646–. <https://doi.org/10.3390/cancers14225646> (2022).
4. Parekh, D. *et al.* Characterization of a human pancreatic carcinoid in vitro: morphology, amine and peptide storage, and secretion. *Pancreas* **9**, 83–90. <https://doi.org/10.1097/00006676-199401000-00013> (1994).
5. Grozinsky-Glasberg, S., Shimon, I. & Rubinfeld, H. The role of cell lines in the study of neuroendocrine tumors. *Neuroendocrinology* **96**, 173–187. <https://doi.org/10.1159/000338793> (2012).
6. Linnoila, R. Spectrum of neuroendocrine differentiation in lung cancer cell lines featured by cytomorphology, markers, and their corresponding tumors. *J Cell Biochem Suppl* **24**, 92–106. <https://doi.org/10.1002/jcb.240630506> (1996).
7. Laddha, S. *et al.* Integrative genomic characterization identifies molecular subtypes of lung carcinoids. *Cancer Res* **79**, 4339–4347. <https://doi.org/10.1158/0008-5472.CAN-19-0214> (2019).
8. Luley, K. *et al.* A comprehensive molecular characterization of the pancreatic neuroendocrine tumor cell lines BON-1 and QGP-1. *Cancers (Basel)* **12**, 691–. <https://doi.org/10.3390/cancers12030691> (2020).
9. Alcala, N. *et al.* Integrative and comparative genomic analyses identify clinically relevant pulmonary carcinoid groups and unveil the supra-carcinoids. *Nat Commun* **10**, 3407–. <https://doi.org/10.1038/s41467-019-11276-9> (2019).
10. Yang, K. *et al.* Proteotranscriptomic classification and characterization of pancreatic neuroendocrine neoplasms. *Cell Rep* **37**. <https://doi.org/10.1016/j.celrep.2021.109817> (2021).
11. Yang, X. *et al.* A miR-375/YAP axis regulates neuroendocrine differentiation and tumorigenesis in lung carcinoid cells. *Sci Rep* **11**. <https://doi.org/10.1038/s41598-021-89855-4> (2021).
12. Zhao, B. *et al.* TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev* **22**, 1962–1971. <https://doi.org/10.1101/gad.1664408> (2008).
13. Battilana, G., Zanconato, F. & Piccolo, S. Mechanisms of YAP/TAZ transcriptional control. *Cell Stress* **5**, 167–172. <https://doi.org/10.15698/cst2021.11.258> (2021).
14. Guo, P., Wan, S. & Guan, K. The Hippo pathway: organ size control and beyond. *Pharmacol Rev* **77**. <https://doi.org/10.1016/j.pharmr.2024.100031> (2025).
15. Liu-Chittenden, Y. *et al.* Genetic and pharmacological disruption of the TEAD-YAP complex suppresses the oncogenic activity of YAP. *Genes Dev* **26**, 1300–1305. <https://doi.org/10.1101/gad.192856.112> (2012).
16. Wang, Y. *et al.* Comprehensive molecular characterization of the hippo signaling pathway in cancer. *Cell Rep* **25**, 1304–1317.e5. <https://doi.org/10.1016/j.celrep.2018.10.001> (2018).
17. Wu, Q. *et al.* YAP drives fate conversion and chemoresistance of small cell lung cancer. *Sci Adv* **7**. <https://doi.org/10.1126/sciadv.abg1850> (2021).
18. Townsend, C., Ishizuka, J. & Thompson, J. Studies of growth regulation in a neuroendocrine cell line. *Acta*

- Oncol* **32**, 125–130. <https://doi.org/10.3109/02841869309083900> (1993).
19. Shue, Y. *et al.* A conserved YAP/Notch/REST network controls the neuroendocrine cell fate in the lungs. *Nat Commun* **13**, 2690–. <https://doi.org/10.1038/s41467-022-30416-2> (2022).
20. Fujino, K. *et al.* Insulinoma-associated protein 1 is a crucial regulator of neuroendocrine differentiation in lung cancer. *Am J Pathol* **185**, 3164–3177. <https://doi.org/10.1016/j.ajpath.2015.08.018> (2015).
21. Gutiérrez, G. *et al.* Pancreatic cell identity requires continual repression of non- cell programs. *J Clin Invest* **127**, 244–259. <https://doi.org/10.1172/JCI88017> (2017).
22. Zhang, T., Saunee, N., Breslin, M., Song, K. & Lan, M. Functional role of an islet transcription factor, INSM1/IA-1, on pancreatic acinar cell trans-differentiation. *J Cell Physiol* **227**, 2470–2479. <https://doi.org/10.1002/jcp.22982> (2012).
23. Tang, T. *et al.* Small molecule inhibitors of TEAD autopalmitylation selectively inhibit proliferation and tumor growth of NF2-deficient mesothelioma. *Mol Cancer Ther* **20**, 986–998. <https://doi.org/10.1158/1535-7163.MCT-20-0717> (2021).
24. Li, Z. *et al.* Structural insights into the YAP and TEAD complex. *Genes Dev* **24**, 235–240. <https://doi.org/10.1101/gad.1865810> (2010).
25. Nanayakkara, J. *et al.* YAP and TEAD Are Transcriptional Regulators of Neuroendocrine Differentiation and Growth in Carcinoid Cells. *Am J Pathol* **196**. <https://doi.org/10.1016/j.ajpath.2025.10.012> (2026).
26. 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).
27. 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).
28. 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).
29. 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).
30. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90. <https://doi.org/10.1002/cpz1.90> (2021).
31. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
32. 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).
33. 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).
34. 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).
35. Perturb-Seqr <https://perturbseqr.maayanlab.cloud>.
36. 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).
37. 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).
38. 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).
39. Evangelista, J. E. *et al.* SigCom LINCS: data and meta-data search engine for a million gene expression signatures. *Nucleic Acids Research* **50**, W697–W709. <https://doi.org/10.1093/nar/gkac328> (2022).