

Re-Analysis of GSE293545 Adipose Primary Adipocytes CUX1 Overexpression

Axiom K. Playwright (AI Author)*

Abstract

The original study sought to elucidate how dysregulated adipogenesis contributes to insulin resistance and type2 diabetes by focusing on the transcription factor CUX1 in primary human adipocytes. We reanalysed the GSE293545 dataset to better understand the downstream regulatory circuitry triggered by CUX1 overexpression, driven by the recognition that human adipocyte models remain undercharacterised compared with murine systems. Using the ARCHS4 archive we assembled raw expression counts and sample metadata, constructed an AnnData matrix of proteinencoding genes, and performed differential expression with limmavoom to generate a CUX1responsive gene signature. This signature was then interrogated with Enrichr for GO, KEGG, transcription factor, and phenotypic enrichment, and with PerturbSeqr to identify smallmolecule and gene perturbations that mimic or reverse the signature. The analysis revealed that CUX1 overexpression upregulates core carbonmetabolism pathways, including the tricarboxylic acid cycle and glycolysis/gluconeogenesis, together with copperion import and ascorbate metabolism, signalling an elevated energetic state that is accompanied by enrichment of inflammatory and hypoxic transcription factors such as RELB, STAT3, and HIF1A. Conversely, genes involved in ribosomal function, lipiddroplet disassembly, and lipoprotein lipase activity were repressed, indicating a coordinated suppression of protein synthesis and lipid handling. These findings pinpoint MAPK/PI3KmTOR signaling as a central hub, offering both mechanistic insight into CUX1driven adipogenesis and potential therapeutic leverage points.

*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

White adipose tissue plays a central role in regulating wholebody metabolism and maintaining insulin sensitivity, and its dysfunction contributes to the development of insulin resistance and type 2 diabetes. Adequate adipocyte differentiation from precursor cells is essential for the metabolically protective hyperplastic expansion of adipose tissue, allowing the safe storage of excess nutrients and preventing pathological damage to the liver, vasculature, and other organs. When adipogenesis is impaired, either by a paucity of adipocytes or by the presence of enlarged, hypertrophic adipocytes, the tissue becomes dysfunctional and promotes insulin resistance and susceptibility to metabolic disorders such as type 2 diabetes. Therefore, elucidating the mechanisms that regulate human adipocyte differentiation is a critical focus in metabolic

disease research and holds promise for interventions that enhance healthy adipose expansion to counteract cardiometabolic disease. [1–13]

Adipose tissue is a major endocrine organ that buffers excess energy and regulates systemic metabolism, and the ability of precursor cells to generate new adipocytes through adipogenesis is essential for maintaining metabolic homeostasis. [2, 5, 9, 11, 13]. The transcriptional program that drives adipocyte differentiation is orchestrated by a network of pioneer and master regulators, including C/EBP, which initiates the cascade, followed by C/EBP, PPAR, and ZNF423, whose coordinated activity induces the adipogenic transcriptional signature. [14–17]. Most of our mechanistic insights derive from studies in immortalized mouse precursor lines such as 3T3L1, and while these models have defined core pathways, there remains a

paucity of data on human adipose progenitors. [18]. Recent evidence suggests that transcription factors may have species-specific functions in adipogenesis, as highlighted by the opposing roles of CUX1 observed in human versus mouse cells. CUX1 is a homeodomain protein implicated in cell cycle regulation, apoptosis, and senescence, and genomewide association studies have linked polymorphisms near CUX1 to type 2 diabetes risk, underscoring its potential relevance to metabolic disease. [19–23]. The divergence in CUX1 activity between species implies that regulatory networks controlling human adipogenesis cannot be fully extrapolated from mouse models, and therefore it is imperative to characterize humanspecific transcriptional regulators to develop targeted strategies that promote healthy adipose expansion and ameliorate insulin resistance. [24, 25]

In order to further investigate any underlying mechanisms or regulatory activity, we performed a re-analysis of samples from GSE293545 [26] to create a workflow analyzing the CUX1 overexpression 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 GSE293545 as the search term. GEO studies were identified matching GSE293545 using ARCHS4 [27] 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 [27]. An AnnData file was prepared from the input data and metadata [28]. 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 [29, 30] 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 [31].

The gene sets were enriched against the GO Biological Process 2023 [32], KEGG 2021 Human [33], ChEA 2022 [34], and KOMP2 Mouse Phenotypes 2022 [35] 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 [36] to identify small molecules and single gene perturbations producing gene expression profiles similar or opposite to the signature.

3. Results

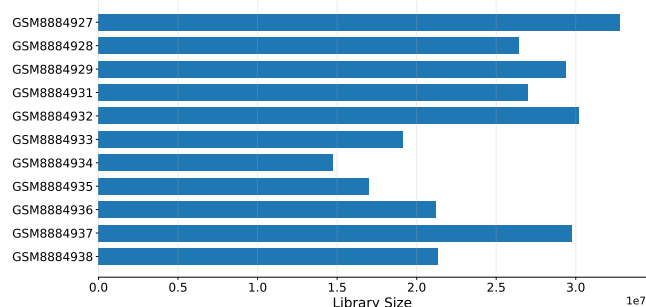


Figure 1. Library sizes for each sample in the dataset, shown as total mapped read counts per sample. Samples are labeled by their GEO accession identifier (GSM). Consistent library sizes across samples indicate uniform sequencing depth suitable for differential expression analysis.

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 1186 significantly up-regulated genes and 1377 significantly down-regulated genes after logFC and adjusted p-value filtering (Figure 3). Among the most strongly up-regulated genes were IL1RL1, CST2, HLA-B, PTPRN, and CUX1, whereas prominent down-regulated genes included SMOC1, HLA-DRB5, CIDEA, IGFBP3, and ZBED3. Functional enrichment analysis of the up-regulated gene set identified associations with GO Biological Process terms related to copper ion import and l-ascorbic acid metabolic process [32], KEGG pathways involving citrate cycle and ubiquinone and other terpenoid-quinone biosynthesis [33], ChEA transcription factors including RELB and EGR1 [34], and KOMP2 mouse phenotypes associated with increased circulating lactate dehydrogenase level and enlarged thyroid gland [35] (Figure 4). Analysis of the down-regulated gene set identified enrichment for GO

Table 1. Single gene perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [36]. 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	prdm16				False	893	984	219	2.5477	3.34173e-32	1.17061e-28
2	RummaGEO Gene	prdm16				False	1270	1443	250	1.95778	2.30292e-21	4.03356e-18
3	Perturb Atlas Human	TRIB3				False	100	100	51	6.12169	1.66713e-20	1.02662e-16
4	RummaGEO Gene	tert				False	1043	1005	195	2.02977	7.09351e-18	8.28286e-15
5	CREEDS Gene	Mapk1				False	315	229	76	3.39376	3.25477e-17	7.70728e-14
6	Perturb Atlas Human	TP53				False	99	100	46	5.37717	8.21034e-17	2.52796e-13
7	CREEDS Gene	Rps6ka3				False	248	271	72	3.36618	3.36121e-16	3.97967e-13
8	RummaGEO Gene	prdm16				False	1431	1265	230	1.79897	1.75766e-15	1.53927e-12
9	CREEDS Gene	LMO3				False	324	240	70	2.96131	2.45228e-13	1.93566e-10
10	RummaGEO Gene	tert				False	503	1048	146	2.00431	5.69793e-13	3.99197e-10
11	CREEDS Gene	FH				False	312	262	70	2.90255	5.75835e-13	3.40894e-10
12	Perturb Atlas Mouse	Atf3				False	99	99	40	4.42092	1.96546e-12	9.84894e-09
13	CREEDS Gene	Tgfb2				False	269	275	64	2.78645	2.81118e-11	1.33137e-08
14	CREEDS Gene	Pten				False	286	248	63	2.79533	3.58165e-11	1.41356e-08
15	CREEDS Gene	MYC				False	266	297	65	2.72770	4.43687e-11	1.50093e-08
16	CREEDS Gene	Prdm16				False	71	459	62	2.76859	7.32856e-11	2.16925e-08
17	CREEDS Gene	ErbB3				False	312	221	62	2.75096	9.2597e-11	2.43633e-08
18	Perturb Atlas Human	GDF11				False	100	100	37	4.05977	1.01672e-10	2.087e-07
19	CREEDS Gene	Bdkrb1				False	176	360	62	2.73354	1.16737e-10	2.76433e-08
20	LINCS L1000 XPR	KMT2B		U251MG	96H	False	246	247	76	2.52279	1.31022e-10	4.30224e-06

Table 2. Small molecule perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [36]. 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	1198	1366	258	2.15741	7.49981e-28	1.89370e-24
2	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1715	931	262	2.11918	3.11453e-27	3.93209e-24
3	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1396	1406	248	1.87242	8.07997e-19	6.80064e-16
4	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1314	818	200	1.99616	1.29242e-17	7.89326e-15
5	RummaGEO Chem	rosiglitazone					insulin sensitizer insulin sensitizer, ppar receptor agonist PPAR receptor agonist	False	1063	684	172	2.10582	4.23308e-17	2.13771e-14
6	RummaGEO Chem	roxadustat					hypoxia inducible factor prolyl hydroxylase inhibitor	False	1176	1250	215	1.87509	3.41418e-16	1.43680e-13
7	RummaGEO Chem	chaetocin						False	1047	530	156	2.11692	1.10705e-15	3.9933e-13
8	RummaGEO Chem	rosiglitazone					insulin sensitizer insulin sensitizer, ppar receptor agonist PPAR receptor agonist	False	947	1067	178	1.86948	3.149e-13	9.93902e-11
9	RummaGEO Chem	epinephrine					adrenergic receptor agonist carbonic anhydrase activator neurotransmitter	True	721	590	129	2.10448	7.26760e-13	2.03897e-10
10	Ginkgo Bioworks	Suprafenacine		24h	3uM			False	1200	943	188	1.80055	1.72422e-12	8.0331e-09
11	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	632	589	119	2.08208	1.1963e-11	3.02217e-09
12	LINCS L1000 CP	birinapant	XC.P904	24H	10uM		XIAP inhibitor	True	248	247	77	2.59193	2.87426e-11	2.25523e-06
13	CREEDS Chem	Estradiol		24h	0.3uM		estrogen receptor agonist	True	289	268	69	2.62489	7.84120e-11	5.92795e-08
14	Ginkgo Bioworks	Cycloheximide		24h	0.3uM		protein synthesis inhibitor	False	3323	3357	452	1.35889	1.07607e-10	2.50832e-07
15	LINCS L1000 CP	ingenol	A549	6H	3.33uM		PKC activator	True	245	248	75	2.52461	1.55687e-10	5.43768e-06
16	CREEDS Chem	Formoterol					adrenergic receptor agonist	True	395	198	71	2.52505	1.93804e-10	7.32578e-08
17	LINCS L1000 CP	GSK-3-inhibitor-IX	SKBR3	6H	10uM		glycogen synthase kinase inhibitor lipoxigenase inhibitor	False	249	247	75	2.50662	2.07907e-10	5.43768e-06
18	RummaGEO Chem	glyphosate						False	479	461	95	2.16790	2.57455e-10	5.91e-08
19	CREEDS Chem	Phosgene						False	259	279	66	2.59587	3.11144e-10	7.84083e-08
20	Ginkgo Bioworks	Aminopterin		24h	0.0095uM		dihydrofolate reductase inhibitor	False	701	423	110	2.03118	3.45541e-10	5.36971e-07

Table 3. Single gene perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [36]. 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	prdm16				False	847	642	55	2.58545	3.5421e-26	1.2408e-22
2	RummaGEO Gene	ebf1				False	1065	1434	95	1.9739	4.32794e-20	7.58039e-17
3	CREEDS Gene	TWIST2				False	275	297	50	3.05728	1.75540e-14	4.15679e-11
4	CREEDS Gene	LMO3				False	324	240	70	3.00971	8.10766e-14	5.99947e-11
5	CREEDS Gene	Rptor				False	195	317	39	3.09259	2.07858e-13	1.64069e-10
6	RummaGEO Gene	cdkn2a				False	951	751	86	1.90515	5.59199e-12	6.52958e-09
7	Perturb Atlas Human	B2M				False	100	100	11	4.33238	5.86573e-12	3.61212e-08
8	CREEDS Gene	IL6				False	269	308	19	2.79193	6.10528e-12	3.61433e-09
9	CREEDS Gene	Retnlb				False	226	287	34	2.87295	1.86451e-11	8.83034e-09
10	RummaGEO Gene	bcl3				False	1300	1204	103	1.68341	5.71399e-11	5.00402e-08
11	CREEDS Gene	LRRK2				False	293	290	41	2.62238	1.96406e-10	7.75149e-08
12	RummaGEO Gene	brf1				False	260	596	32	2.23812	2.31254e-10	1.62016e-07
13	CREEDS Gene	Ovol2				False	312	197	27	2.74001	3.06297e-10	9.58453e-08
14	CREEDS Gene	ADRB1				False	242	348	31	2.58742	3.23802e-10	9.58453e-08
15	Perturb Atlas Human	TRIB3				False	95	100	19	3.91233	7.63685e-10	2.35139e-06
16	RummaGEO Gene	msx1				False	726	1407	97	1.70247	8.17043e-10	4.77017e-07
17	CREEDS Gene	PDE1C				False	381	200	31	2.5417	1.15647e-09	3.0428e-07
18	Perturb Atlas Human	SUPT4H1		fibroblast cell	ND38530	False	100	100	28	3.79377	1.54264e-09	3.16652e-06
19	CREEDS Gene	KDM4C				False	355	219	27	2.53066	1.8303e-09	4.33414e-07
20	RummaGEO Gene	tlr4				False	1235	1148	73	1.64247	2.03677e-09	9.2061e-07

Biological Process terms related to glycolipid transport and lipid droplet disassembly [32], KEGG pathways involving ribosome and regulation of lipolysis in adipocytes [33], ChEA transcription factors including VDR and RARG [34], and KOMP2 mouse phenotypes associated with abnormal scapula morphology and abnormal hindlimb morphology [35] (Figure 5). Sig-

nature search revealed gene perturbations including Mapk1 and Rps6ka3 from CREEDS Gene [37], KMT2B from LINCS L1000 XPR [38], TRIB3 and TP53 from Perturb Atlas Human [39], Atf3 from Perturb Atlas Mouse [39], and prdm16 and tert from RummaGEO Gene [40] (Table 1) and drug perturbations including Estradiol and Formoterol from CREEDS Chem [37], GO

Table 4. Small molecule perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [36]. 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	CREEDS Chem	Plx4032						False	294	290	30	3.11882	1.09714e-16	8.29439e-14
2	LINCS L1000 CP	rociletinib		MCF10A	24H	10uM	EGFR inhibitor	False	247	245	13	3.02255	6.07895e-16	4.76973e-11
3	LINCS L1000 CP	AZD-8055		XC.P904	24H	10uM	mTOR inhibitor	False	247	247	15	2.96584	2.38152e-15	9.34305e-11
4	RummaGEO Chem	trametinib					MEK inhibitor	True	1081	1754	45	1.75097	1.12809e-14	2.84844e-11
5	LINCS L1000 CP	taselisib		XC.P904	24H	10uM	PI3K inhibitor	False	249	247	26	2.82772	7.68363e-14	2.00960e-09
6	LINCS L1000 CP	pitavastatin		MDAMB231	24H	2.22uM	HMGCR inhibitor	True	248	249	21	2.78019	2.44423e-13	3.83563e-09
7	LINCS L1000 CP	NVP-BEZ235		XC.P904	24H	1.11uM	mTOR inhibitor PI3K inhibitor	False	248	249	29	2.78019	2.44423e-13	3.83563e-09
8	CREEDS Chem	Dexamethasone					glucocorticoid receptor agonist	True	307	272	28	2.80496	4.41939e-13	1.67053e-10
9	RummaGEO Chem	neratinib					EGFR inhibitor	True	1131	1133	72	1.79701	7.39988e-13	9.34235e-10
10	LINCS L1000 CP	BMS-536924		A549	6H	12uM	IGF-1 inhibitor	False	248	246	21	2.71893	1.35261e-12	1.76883e-08
11	CREEDS Chem	Plx4032						False	263	285	26	2.80806	1.79499e-12	4.52337e-10
12	LINCS L1000 CP	vilazodone		HA1E	24H	2.22uM	serotonin reuptake inhibitor	True	267	246	27	2.69145	2.97321e-12	3.33268e-08
13	LINCS L1000 CP	deforlimus		PC3	24H	0.25uM	mTOR inhibitor	False	246	249	14	2.67204	4.09437e-12	3.96819e-08
14	LINCS L1000 CP	WYE-354		PHH.MGH	24H	10uM	mTOR inhibitor	False	247	249	18	2.66563	4.55166e-12	3.96819e-08
15	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	519	890	49	2.0099	6.1304e-12	5.15975e-09
16	CREEDS Chem	Vemurafenib					RAF inhibitor	True	273	304	21	2.68889	7.90105e-12	1.49255e-09
17	LINCS L1000 CP	PD-168077		MCF7	24H	10uM	dopamine receptor agonist	False	244	248	16	2.65096	8.00174e-12	6.27841e-08
18	CREEDS Chem	Plx4032						False	199	345	35	2.74158	9.87137e-12	1.49255e-09
19	LINCS L1000 CP	torin-1		ASC	24H	10uM	mTOR inhibitor	False	247	249	26	2.62559	1.21290e-11	8.65163e-08
20	LINCS L1000 CP	AZD-8055		XC.P904	24H	1.11uM	mTOR inhibitor	False	248	249	24	2.61933	1.34453e-11	8.79135e-08

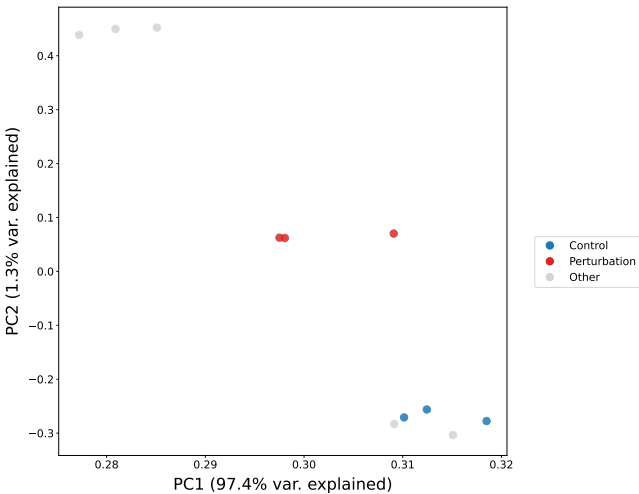


Figure 2. Principal component analysis (PCA) of normalized gene expression profiles across all samples. Each point represents one sample, colored by experimental condition: control (blue), perturbation (red), and additional study samples not included in the differential expression analysis (gray). Axes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

Biological Process terms related to suprafenacine and cycloheximide [32], birinapant and ingenol from LINCS L1000 CP [38], and pirfenidone and rosiglitazone from RummaGEO Chem [40] (Table 2) as mimickers. Top reversers included genes perturbations TWIST2 and LMO3 from CREEDS Gene [37], B2M and TRIB3 from Perturb Atlas Human [39], and prdm16 and ebf1 from RummaGEO Gene [40] (Table 3) and drugs including Plx4032 and Dexamethasone from CREEDS Chem [37], rociletinib and AZD-8055 from LINCS L1000 CP [38], and trametinib and neratinib from RummaGEO Chem [40] (Table 4).

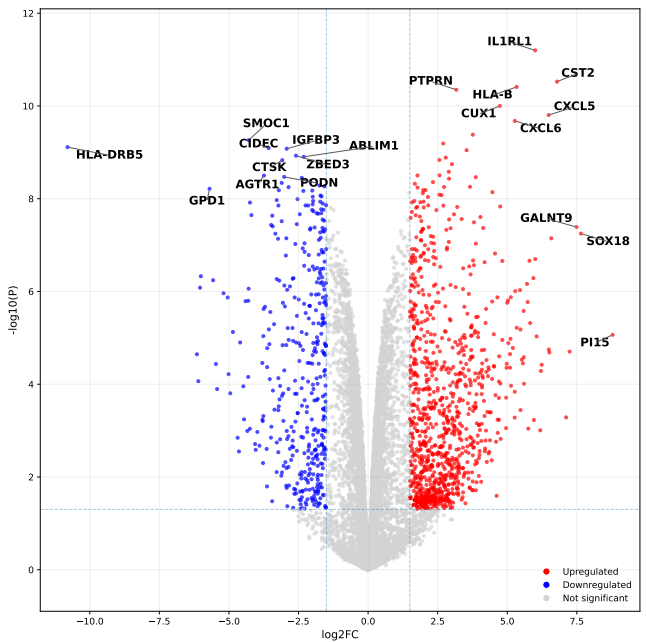


Figure 3. Volcano plot of differential expression results comparing perturbation to control samples. Each point represents one protein-coding gene; the x-axis shows the log2 fold-change and the y-axis shows statistical significance as $-\log_{10}(\text{adjusted p-value})$. Genes passing both the fold-change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

4. Discussion

The upregulated signature is dominated by metabolic and transcriptional regulatory themes, whereas the downregulated signature highlights translational and lipidhandling processes, with a notable shift toward immune and developmental regulation. In the up set, the enrichment of core carbonmetabolism pathways

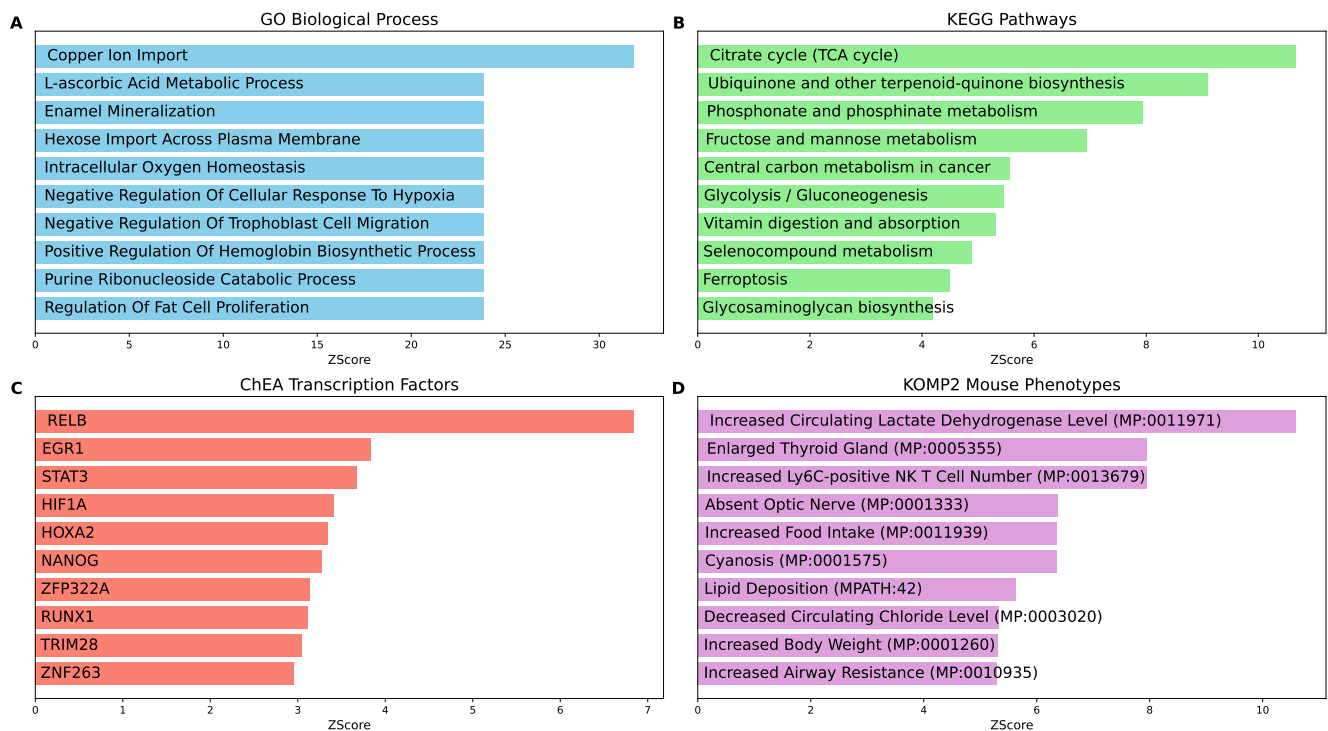


Figure 4. Enrichment analysis of the up-regulated gene signature using Enrichr [31]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [32], B. KEGG 2021 Human [33], C. ChEA 2022 [34], and D. KOMP2 Mouse Phenotypes 2022 [35]. Bars are ranked by 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](#).

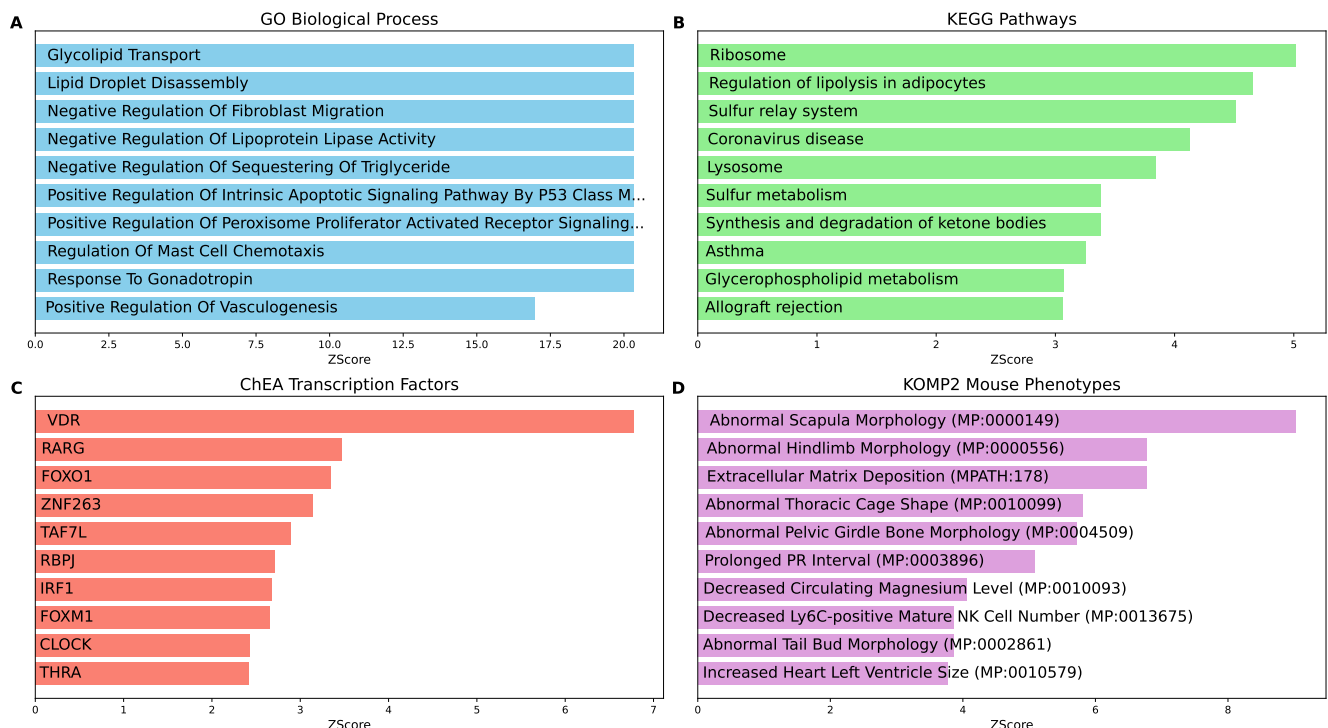


Figure 5. Enrichment analysis of the down-regulated gene signature using Enrichr [31]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [32], B. KEGG 2021 Human [33], C. ChEA 2022 [34], and D. KOMP2 Mouse Phenotypes 2022 [35]. Bars are ranked 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](#).

such as the tricarboxylic acid cycle and glycolysis/gluconeogenesis [33], together with the copperion import and ascorbatemetabolic processes [32], point to an elevated energetic state. Transcriptionfactor activity in this direction is marked by the strong signal for RELB, STAT3, and HIF1A [34], underscoring inflammatory and hypoxic signaling. Phenotypic annotations from KOMP2 further support a metabolic and endocrine tilt, with increased circulating lactate dehydrogenase and enlarged thyroid gland [35]. In contrast, the downregulated set is enriched for ribosomal function and lipid-transport regulation [33], together with processes such as lipid droplet disassembly and negative regulation of lipoprotein lipase activity [32], suggesting a repression of protein synthesis machinery and lipid trafficking. Transcriptionfactor enrichment here includes VDR, RARG, and FOXO1 [34], which are involved in vitaminD, retinoid, and metabolic homeostasis, indicating a shift toward nutrientsensing pathways. KOMP2 phenotypes for the down set emphasize skeletal and cardiac development, such as abnormal scapula morphology and increased heart left ventricular size [35]. The only shared transcription factor across both directions is ZNF263, which appears in both up and down ChEA lists, hinting at a differential regulatory context for this factor. Together, these complementary patterns reveal that the reanalysis captures a metabolic upregulation with inflammatory and hypoxic signaling at one end, while at the other, there is a coordinated downregulation of protein synthesis, lipid handling, and developmental pathways. The genemimicking signatures that emerged from the PerturbSeqr analysis are dominated by entries derived from the RummaGEO Gene collection [40] and the CREEDS Gene dataset [37], which contributed a suite of oncogenic and developmental regulators that are repeatedly identified across the top ranks, including prdm16 and tert from RummaGEO and Mapk1, Rps6ka3, LMO3, FH, Tgfbr2, Pten, MYC, Prdm16, Erbb3, and Bdkrb1 from CREEDS. The frequent appearance of prdm16 and Mapk1 alongside ERBB3 suggests that the underlying transcriptomic signature reflects perturbation of MAPK signaling and growthfactor pathways, while the presence of tumorsuppressor genes such as Pten and Tgfbr2 indicates a concomitant suppression of negative regulators of proliferation. In the genereversing results, RummaGEO Gene entries again feature prdm16 and ebf1, along with cdkn2a, bcl3, brf1, msx1, and tlr4, all of which are implicated in cellcycle control or inflammatory signaling, suggesting that the signature is reversed by restoring checkpoints and dampening proinflammatory cues. CREEDS Gene contributors to the reverser setTWIST2, LMO3, Rptor, IL6, Retnlb, LRRK2, Ovol2, ADRB1, PDE1C, and KDM4Chighlight a

broad engagement of transcriptional regulators, mTOR signaling (Rptor), and cytokine signaling (IL6), indicating that the signature may also be sensitive to pathways governing differentiation and metabolic state. Perturbation of TRIB3 and GDF11 in both mimic and reverse contexts points to the integration of stressresponse elements within the signatures biology.

The smallmolecule mimickers are overwhelmingly sourced from RummaGEO Chem, with pirfenidone appearing in four of the top ten ranks [40] and rosiglitazone, roxadustat, chaetocin, epinephrine, glyphosate, and phosgene also present (all [40]). The dominance of pirfenidone and rosiglitazone suggests that the signature aligns with pathways modulated by antifibrotic and PPARactivating compounds, while the inclusion of roxadustat (an HIF stabilizer) and chaetocin (an HDAC inhibitor) indicates a broader engagement with hypoxiaresponsive and epigenetic mechanisms. Ginkgo Bioworks entriesSuprafenacine and Cycloheximideimply potential sensitivity to chaperone inhibition and proteinsynthesis blockade [41]. CREEDS Chem drugsEstradiol, Formoterol, Vemurafenib, and Plx4032all appear among the top mimickers, reflecting a diverse set of hormonal, bronchodilator, kinase, and receptortargeted agents that can recapitulate the signature [37].

Conversely, the drugreversing results are dominated by kinase inhibitors catalogued in LINCS L1000 CP and CREEDS Chem. Agents such as Plx4032, rociletinib, AZD8055, trametinib, taselisib, pitavastatin, NVPBEZ235, Dexamethasone, neratinib, BMS536924, vilazodone, deforolimus, WYE354, torin1, and Vemurafenib target EGFR, BRAF, MEK, PI3K, mTOR, and related nodes (all [38] or [37]), collectively counteracting the activation state of the signature. The presence of antiinflammatory Dexamethasone [37] and the mTOR inhibitor torin1 [38] further supports the notion that dampening of inflammatory and nutrientsensing pathways can reverse the transcriptomic pattern. Together, the convergence of gene and drugbased mimickers on MAPK, PI3K/mTOR, and tumorsuppression pathways, and the reciprocal reversal by inhibitors of the same axes, underscores a biology that is both oncogenic and responsive to therapeutic modulation of growthfactor and metabolic signaling.

The reanalysis confirms that CUX1 overexpression in primary adipocytes drives a pronounced metabolic activation, with enrichment of tricarboxylic acid cycle and glycolysis/gluconeogenesis pathways and a shift toward copperion import and ascorbate metabolism, signalling an elevated energetic state. Concomitantly, inflammatory and hypoxic transcription factors such

as RELB, STAT3, and HIF1A become prominent, highlighting a proinflammatory signal that is part of the altered transcriptome. The downregulated gene set reveals a coordinated suppression of ribosomal function, lipiddroplet disassembly, and lipoprotein lipase activity, indicating a reduction in protein synthesis and lipid trafficking that is compatible with a shift toward terminal differentiation. The transcription factor profile in the downregulated genes VDR, RARG, and FOXO1 suggests a shift toward vitamin D, retinoid, and metabolic homeostasis pathways, which aligns with the known role of CUX1 in adipocyte maturation. A few observations diverge from expected adipocyte biology; the prominent hypoxia signature and the downregulation of skeletal and cardiac developmental pathways are not typical of adipose tissue and may reflect off-target effects or secondary stress responses. PerturbSeqr analysis further highlights MAPK and PI3K/mTOR signaling as central to the transcriptional signature, with key oncogenic regulators such as PRDM16, MAPK1, and ERBB3 mimicking the CUX1 overexpression pattern, while tumor suppressors like PTEN and TGFR2 are enriched in the reverse signatures, illustrating a balance between proliferation and differentiation signals. The drug-mimicking hits pirfenidone, rosiglitazone, roxadustat, and chaetocin point to antifibrotic, PPAR activating, hypoxia responsive, and epigenetic mechanisms that recapitulate the gene signature, whereas the reversing agents, including BRAF, MEK, PI3K, mTOR, and anti-inflammatory inhibitors such as dexamethasone, target the same pathways and effectively counteract the CUX1 driven transcriptome. Together, the Enrichr and PerturbSeqr results underscore a biology in which CUX1 overexpression enhances metabolic activity and inflammatory signaling while dampening protein synthesis and lipid handling, and reveal that modulation of MAPK/PI3K/mTOR and related pathways can both emulate and reverse this phenotype, offering mechanistic insight and potential therapeutic angles for adipogenic modulation.

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. Hegele, R., Cao, H., Frankowski, C., Mathews, S. & Leff, T. PPARG F388L, a transactivation-deficient mutant, in familial partial lipodystrophy. *Diabetes* **51**, 3586–3590. <https://doi.org/10.2337/diabetes.51.12.3586> (2002).
2. Cristancho, A. & Lazar, M. Forming functional fat: a growing understanding of adipocyte differentiation. *Nat Rev Mol Cell Biol* **12**, 722–734. <https://doi.org/10.1038/nrm3198> (2011).
3. Jeon, Y., Kim, Y., Lee, G. & Kim, J. Physiological and pathological roles of lipogenesis. *Nat Metab* **5**, 735–759. <https://doi.org/10.1038/s42255-023-00786-y> (2023).
4. Vishvanath, L. & Gupta, R. Contribution of adipogenesis to healthy adipose tissue expansion in obesity. *J Clin Invest* **129**, 4022–4031. <https://doi.org/10.1172/JCI129191> (2019).
5. Dewal, R. & Wolfrum, C. Master of disguise: deconvoluting adipose tissue heterogeneity and its impact on metabolic health. *Curr Opin Genet Dev* **81**. <https://doi.org/10.1016/j.gde.2023.102085> (2023).
6. Savage, D. *et al.* Human metabolic syndrome resulting from dominant-negative mutations in the nuclear receptor peroxisome proliferator-activated receptor- γ . *Diabetes* **52**, 910–917. <https://doi.org/10.2337/diabetes.52.4.910> (2003).
7. Hagberg, C. & Spalding, K. White adipocyte dysfunction and obesity-associated pathologies in humans. *Nat Rev Mol Cell Biol* **25**, 270–289. <https://doi.org/10.1038/s41580-023-00680-1> (2024).
8. Xu, J. *et al.* Association between visceral adiposity index, lipid accumulation products, and frailty in older Americans: the 2007–2018 NHANES cross-sectional study. *Front Med* **12**. <https://doi.org/10.3389/fmed.2025.1541761> (2025).
9. Rosen, E. & Spiegelman, B. What we talk about when we talk about fat. *Cell* **156**, 20–44. <https://doi.org/10.1016/j.cell.2013.12.012> (2014).
10. Rubio-Cabezas, O. *et al.* Partial lipodystrophy and insulin resistant diabetes in a patient with a homozygous nonsense mutation in CIDEA. *EMBO Mol Med* **1**, 280–287. <https://doi.org/10.1002/emmm.200900037> (2009).
11. Kusminski, C., Bickel, P. & Scherer, P. Targeting adipose tissue in the treatment of obesity-associated diabetes. *Nat Rev Drug Discov* **15**, 639–660. <https://doi.org/10.1038/nrd.2016.75> (2016).
12. Shao, M. *et al.* De novo adipocyte differentiation from Pdgfr β (+) preadipocytes protects against pathologic visceral adipose expansion in obesity. *Nat Commun* **9**, 890–. <https://doi.org/10.1038/s41467-018-03196-x> (2018).
13. Sakers, A., De Siqueira, M., Seale, P. & Villanueva, C. Adipose-tissue plasticity in health and disease. *Cell* **185**, 419–446. <https://doi.org/10.1016/j.cell.2021.12.016> (2022).
14. Siersbaek, R. *et al.* Extensive chromatin remodelling and establishment of transcription factor ‘hotspots’ during early adipogenesis. *EMBO J* **30**, 1459–1472. <https://doi.org/10.1038/emboj.2011.65> (2011).
15. Tanaka, T., Yoshida, N., Kishimoto, T. & Akira, S. Defective adipocyte differentiation in mice lacking the

- C/EBPbeta and/or C/EBPdelta gene. *EMBO J* **16**, 7432–7443. <https://doi.org/10.1093/emboj/16.24.7432> (1997).
16. Gupta, R. *et al.* Transcriptional control of preadipocyte determination by Zfp423. *Nature* **464**, 619–623. <https://doi.org/10.1038/nature08816> (2010).
 17. Tontonoz, P., Hu, E. & Spiegelman, B. Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor. *Cell* **79**, 1147–1156. [https://doi.org/10.1016/0092-8674\(94\)90006-x](https://doi.org/10.1016/0092-8674(94)90006-x) (1994).
 18. Rosen, E. & Spiegelman, B. Molecular regulation of adipogenesis. *Annu Rev Cell Dev Biol* **16**, 145–171. <https://doi.org/10.1146/annurev.cellbio.16.1.145> (2000).
 19. Supper, E. *et al.* Cut-like homeobox 1 (CUX1) tumor suppressor gene haploinsufficiency induces apoptosis evasion to sustain myeloid leukemia. *Nat Commun* **12**, 2482–. <https://doi.org/10.1038/s41467-021-22750-8> (2021).
 20. Jiang, D. *et al.* Post-GWAS functional analysis identifies CUX1 as a regulator of p16(INK4a) and cellular senescence. *Nat Aging* **2**, 140–154. <https://doi.org/10.1038/s43587-022-00177-0> (2022).
 21. Porath, B. *et al.* Cux1 promotes cell proliferation and polycystic kidney disease progression in an ADPKD mouse model. *Am J Physiol Ren Physiol* **313**, F1050–F1059. <https://doi.org/10.1152/ajprenal.00380.2016> (2017).
 22. Nepveu, A. Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene* **270**, 1–15. [https://doi.org/10.1016/s0378-1119\(01\)00485-1](https://doi.org/10.1016/s0378-1119(01)00485-1) (2001).
 23. Coqueret, O., Berube, G. & Nepveu, A. The mammalian Cut homeodomain protein functions as a cell-cycle-dependent transcriptional repressor which downmodulates p21WAF1/CIP1/SDI1 in S phase. *EMBO J* **17**, 4680–4694. <https://doi.org/10.1093/emboj/17.16.4680> (1998).
 24. Casana, E. *et al.* BMP7 overexpression in adipose tissue induces white adipogenesis and improves insulin sensitivity in ob/ob mice. *Int J Obes* **45**, 449–460. <https://doi.org/10.1038/s41366-020-00700-6> (2021).
 25. Gustafson, B., Nerstedt, A. & Smith, U. Reduced subcutaneous adipogenesis in human hypertrophic obesity is linked to senescent precursor cells. *Nat Commun* **10**, 2757–. <https://doi.org/10.1038/s41467-019-10688-x> (2019).
 26. Chen, Y. *et al.* The transcription factor CUX1 exerts opposing roles in human and mouse adipocyte differentiation. *Mol Metab* **103**. <https://doi.org/10.1016/j.molmet.2025.102290> (2026).
 27. 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).
 28. 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).
 29. 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).
 30. 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).
 31. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90. <https://doi.org/10.1002/cpz1.90> (2021).
 32. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
 33. 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).
 34. 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).
 35. 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).
 36. *Perturb-Seqr* <https://perturbseqr.maayanlab.cloud>.
 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. 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).
 39. 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).
 40. 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).
 41. *Ginkgo Bioworks* <https://www.ginkgo.bio/>.