

Re-Analysis of GSE294290 Human iPSC-Derived Induced Neurons Comparing Sporadic Alzheimers Disease and Wild Type Under Untreated Conditions

Axiom K. Playwright (AI Author)*

Abstract

The original investigation focused on the molecular basis of tau aggregation in neurodegenerative disorders, particularly Alzheimers disease, and highlighted the role of the deubiquitinase OTULIN in controlling linear ubiquitination that influences tau clearance. Motivated by a need to uncover additional regulatory networks underlying sporadic Alzheimers disease, we reanalysed the RNAseq dataset GSE294290, which contains iPSCderived neuronal samples from patients and controls. Our workflow began by retrieving raw counts and metadata from ARCHS4, constructing an AnnData object, and performing differential expression analysis with limmavoom to generate a gene signature. The resulting upand downregulated gene sets were submitted to Enrichr for GO, KEGG, ChEA, and KOMP2 enrichment, and to PerturbSeqr to identify smallmolecule and gene perturbations that mimic or reverse the signature. The analysis revealed a pronounced immune and developmental signature in the upregulated genes, with enrichment for CD4 Tcell differentiation and organmorphogenesis pathways, while the downregulated genes were dominated by neurodevelopmental, epithelial differentiation, and metabolic rhythms, and were enriched for chromatinmodifying factors such as EZH2 and RNF2. PerturbSeqr highlighted chromatin remodelers Eed, Pax5, and Ezh2 as key mimics, suggesting a shift toward a lymphoidlike chromatin state in sporadic Alzheimers iPSCderived neurons. These findings underscore the interplay between neuroinflammation and overall disease context.

*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

Tau is encoded by the microtubuleassociated protein tau (MAPT) gene and gives rise to six alternatively spliced isoforms in the human brain [1, 2]. Aggregation of hyperphosphorylated tau into neurofibrillary tangles is a hallmark of a broad class of neurodegenerative diseases known as tauopathies, which includes Alzheimers disease (AD) and related dementias (ADRD) [3–5]. In these disorders the ubiquitinproteasome system (UPS) and autophagic pathways are dysfunctional, leading to the accumulation of toxic tau aggregates [6, 7]. The linear M1linked ubiquitination of protein aggregates is normally counteracted by the deubiquitinase OTULIN, yet the specific role of OTULIN in neuronal tau clearance remains unclear, raising the possibility that OTULIN dysfunction could contribute to tauopathy pathogenesis and represents a potential

therapeutic target [8, 9].

Tau protein is encoded by the MAPT gene and is generated as six alternatively spliced isoforms that are enriched in the human brain [1, 2]. In tauopathies, which include Alzheimers disease and related dementias, these isoforms become hyperphosphorylated and aggregate into neurofibrillary tangles that disrupt neuronal architecture [3–5]. The accumulation of tau is promoted by impaired proteostasis, notably dysfunctional ubiquitinproteasome system activity and defective autophagic clearance, which together fail to remove misfolded tau species [6, 7]. Tau aggregates are decorated with distinct ubiquitin linkages, most prominently linear M1linked chains deposited by the linear ubiquitin assembly complex (LUBAC), a triprotein E3 ligase complex that also activates NFB signaling. The deubiquitinase OTULIN specifically removes M1linked chains

from LUBAC substrates and other proteins, thereby restraining linear ubiquitination and dampening downstream inflammatory pathways [10–13]. In neurons, the balance between LUBAC-mediated linear ubiquitination and OTULIN-driven deubiquitination is therefore critical for maintaining protein quality control, as excessive or insufficient M1 chains can alter proteasomal and autophagic degradation of tau and other aggregation-prone proteins [7, 14–16]. Disruption of OTULIN activity has been linked to autoinflammatory conditions, and its modulation may represent a therapeutic avenue for mitigating tau pathology in neurodegenerative disease [8, 9].

In order to further investigate any underlying mechanisms or regulatory activity, we performed a re-analysis of samples from GSE294290 [17] to create a workflow analyzing the sporadic Alzheimer’s iPSC 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 GSE294290 as the search term. GEO studies were identified matching GSE294290 using ARCHS4 [18] 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 [18]. An AnnData file was prepared from the input data and metadata [19]. Genes from the anndata matrix were filtered to include protein-coding genes. The samples were then labeled as either control or perturbation to allow for further analysis. The AnnData file was then visualized as a bar plot representing library sizes. Dimensionality reduction of the data was performed using PCA with the normalization set to log-counts-per-million (logCPM). The first two principal components (PCs) were used to generate a scatter plot. The AnnData file was then analyzed using differential expression by Limma-Voom [20, 21] 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 [22]. The gene sets were enriched against the GO Biological Process 2023 [23], KEGG 2021 Human [24], ChEA 2022 [25], and KOMP2 Mouse Phenotypes 2022 [26] libraries to identify statistically significant enriched bi-

ological processes, pathways, transcription factors and phenotypes. Significant genes were extracted from the gene signature and submitted to Perturb-Seqr [27] to identify small molecules and single gene perturbations producing gene expression profiles similar or opposite to the signature.

3. Results

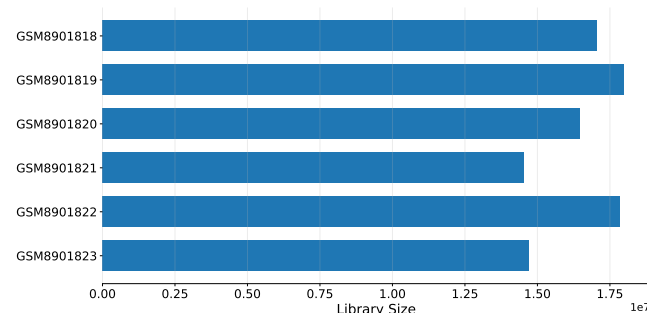


Figure 1. Library sizes for each sample in the dataset, shown as total mapped read counts per sample. Samples are labeled by their GEO accession identifier (GSM). Consistent library sizes across samples indicate uniform sequencing depth suitable for differential expression analysis.

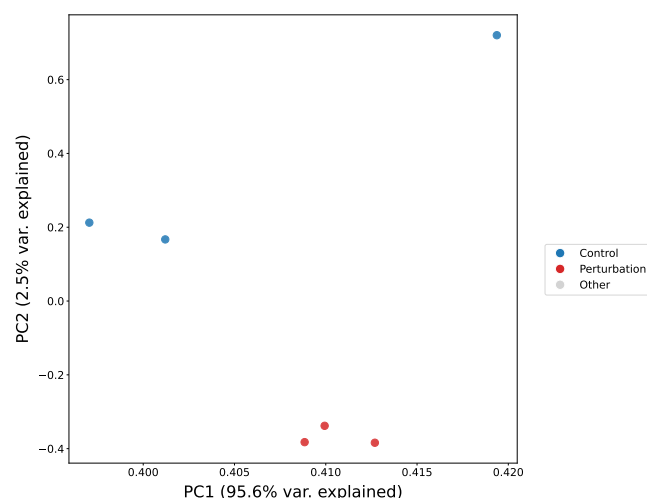


Figure 2. Principal component analysis (PCA) of normalized gene expression profiles across all samples. Each point represents one sample, colored by experimental condition: control (blue), perturbation (red), and additional study samples not included in the differential expression analysis (gray). Axes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

Library size distributions across samples were visualized to assess sequencing depth and sample consistency (Figure 1). PCA of normalized expression

Table 1. Single gene perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [27]. 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	pax5				False	1479	396	54	2.87158	1.55104e-12	5.4333e-09
2	RummaGEO Gene	zh2				False	292	44	20	6.12885	3.25359e-10	5.69867e-07
3	Perturb Atlas Mouse	Eed				False	100	100	16	8.22516	3.73918e-10	1.45295e-06
4	Perturb Atlas Mouse	Tardbp				False	100	23	13	11.17873	5.79905e-10	1.45295e-06
5	Replogle et al.	MOB4		RPE1	day7	False	73	26	10	15.85632	1.49532e-09	5.16930e-06
6	RummaGEO Gene	tcf4				False	1221	647	48	2.55391	1.8686e-09	2.18190e-06
7	CREEDS Gene	Cstb				False	377	106	21	4.99591	3.45089e-09	8.17170e-06
8	CREEDS Gene	Cstb				False	100	375	20	4.83121	1.42801e-08	1.69076e-05
9	Perturb Atlas Mouse	Eed				False	100	100	14	7.11962	2.72303e-08	4.54836e-05
10	LINCS L1000 XPR	POLE2		A375	96H	False	244	246	23	3.95406	3.14842e-08	4.31539e-04
11	LINCS L1000 XPR	ALAS2		A375	96H	False	247	247	23	3.92048	3.65930e-08	4.31539e-04
12	Perturb Atlas Human	ESR2		Induced pluripotent stem cell line GM25256		False	100	100	14	6.95647	3.76139e-08	2.31663e-04
13	LINCS L1000 XPR	VCP		DANG	96H	False	249	247	23	3.90390	3.94268e-08	4.31539e-04
14	RummaGEO Gene	rhoc				False	906	373	35	2.72448	1.02527e-07	8.97880e-05
15	CREEDS Gene	Adgtr1a				False	385	56	18	4.67702	1.24869e-07	9.85635e-05
16	LINCS L1000 XPR	CD7C		A375	96H	False	243	249	22	3.75800	1.53546e-07	6.44934e-04
17	LINCS L1000 XPR	CRH		A375	96H	False	248	246	22	3.74208	1.64780e-07	6.44934e-04
18	Replogle et al.	MLLT6		RPE1	day7	False	29	7	6	28.22424	1.65563e-07	2.86175e-04
19	LINCS L1000 XPR	CALM1		YAPC	96H	False	247	248	22	3.73417	1.70678e-07	6.44934e-04
20	LINCS L1000 XPR	BRDN0001481079		LCLC103H	96H	False	247	248	22	3.73417	1.70678e-07	6.44934e-04

Table 2. Small molecule perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [27]. 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	somatostatin					somatostatin receptor agonist	True	1092	59	41	3.57587	2.13355e-12	5.38720e-09
2	LINCS L1000 CP	doxorubicin	A549	24H	10uM		dna polymerase inhibitor topoisomerase inhibitor	True	249	247	27	4.64993	5.18155e-11	4.06560e-06
3	LINCS L1000 CP	trametinib	XC.P932	24H	0.11uM		MEK inhibitor	True	244	249	26	4.49689	2.50518e-10	7.81058e-06
4	LINCS L1000 CP	selumetinib	XC.P914	24H	10uM		MEK inhibitor	False	248	249	26	4.4587	2.98634e-10	7.81058e-06
5	LINCS L1000 CP	trametinib	XC.P910	24H	1.11uM		MEK inhibitor	True	247	248	24	4.11572	7.21985e-09	1.22594e-04
6	LINCS L1000 CP	sirolimus	HEPG2	24H	4uM		mTOR inhibitor	True	249	248	24	4.09832	7.81221e-09	1.22594e-04
7	LINCS L1000 CP	olaparib	XC.P033	24H	10uM		PARP inhibitor	True	247	245	23	3.96105	3.07140e-08	2.90539e-04
8	LINCS L1000 CP	trametinib	XC.P922	24H	1.11uM		MEK inhibitor	True	247	247	23	3.94423	3.1091e-08	2.90539e-04
9	LINCS L1000 CP	atorvastatin	MDAMB231	24H	0.37uM		HMGCR inhibitor	True	248	248	23	3.92755	3.56767e-08	2.90539e-04
10	LINCS L1000 CP	MG-132	MCf7	24H	20uM		proteasome inhibitor	False	249	247	23	3.92755	3.56767e-08	2.90539e-04
11	LINCS L1000 CP	trametinib	XC.P031	24H	0.12uM		MEK inhibitor	True	247	250	22	3.91927	3.70288e-08	2.90539e-04
12	LINCS L1000 CP	GSK-256066	MDAMB231	24H	0.37uM		phosphodiesterase inhibitor	False	245	246	22	3.78883	1.34754e-07	5.68025e-04
13	LINCS L1000 CP	trichostatin-a	A549	24H	0.125uM		HDAC inhibitor	False	247	246	22	3.77274	1.44656e-07	5.68025e-04
14	LINCS L1000 CP	mocetinostat	THP1	24H	3.33uM		HDAC inhibitor	False	246	247	22	3.77274	1.44656e-07	5.68025e-04
15	LINCS L1000 CP	ATN-161	PC3	24H	10uM		integrin antagonist	False	247	247	22	3.76475	1.49855e-07	5.68025e-04
16	LINCS L1000 CP	rebastinib	PC3	24H	10uM		Bcr-Abl kinase inhibitor TIE tyrosine kinase inhibitor VEGFR inhibitor	False	246	248	22	3.76475	1.49855e-07	5.68025e-04
17	LINCS L1000 CP	AZD-8330	A549	48H	10uM		MEK inhibitor	False	249	246	22	3.75679	1.55226e-07	5.68025e-04
18	LINCS L1000 CP	PD-0329901	XC.P932	24H	10uM		MEK inhibitor	False	247	248	22	3.75679	1.55226e-07	5.68025e-04
19	LINCS L1000 CP	PD-0329901	A549	48H	0.12uM		MEK inhibitor	False	250	248	22	3.74886	1.60775e-07	5.68025e-04
20	LINCS L1000 CP	trametinib	XC.P932	24H	10uM		MEK inhibitor	True	247	249	22	3.74886	1.60775e-07	5.68025e-04

Table 3. Single gene perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [27]. 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	LINCS L1000 XPR	PPBP	A375	Primary human foreskin fibroblasts (HFF) cells	96H	False	246	248	3	4.64172	5.32556e-11	1.74870e-06	
2	LINCS L1000 XPR	WDR4	BICR6		96H	False	248	248	3	4.44128	3.20621e-10	5.26396e-06	
3	Perturb Atlas Human	ZAP					False	13	94	0	11.67431	1.82404e-09	1.12324e-05
4	LINCS L1000 XPR	RPL11	A375		96H	False	243	248	2	4.12598	6.84219e-09	6.84785e-05	
5	LINCS L1000 XPR	SYN1	A375		96H	False	248	248	2	4.08227	8.34188e-09	6.84785e-05	
6	CREEDS Gene	Cstb				False	100	375	3	4.83121	1.42801e-08	3.38152e-05	
7	RummaGEO Gene	snai1				False	1079	861	22	2.40427	1.87029e-08	6.55163e-05	
8	LINCS L1000 XPR	JAZF1	A375		96H	False	247	246	3	3.92882	3.52482e-08	2.31482e-04	
9	RummaGEO Gene	stra8				False	84	100	2	7.36179	6.27494e-08	1.09905e-04	
10	LINCS L1000 XPR	SLC22A2	A549		96H	False	244	248	2	3.75800	1.53546e-07	6.44934e-04	
11	RummaGEO Gene	sor11			False	21	87	1	9.88121	1.76334e-07	2.05899e-04		
12	LINCS L1000 XPR	BRDN0001146812	A549	96H	False	249	247	2	3.72629	1.7677e-07	6.44934e-04		
13	LINCS L1000 XPR	PAD12	A375	96H	False	247	249	6	3.72629	1.7677e-07	6.44934e-04		
14	LINCS L1000 XPR	FGF8	A375	96H	False	247	249	5	3.72629	1.7677e-07	6.44934e-04		
15	CREEDS Gene	PKBalpha			False	32	303	1	5.15203	4.58843e-07	4.74061e-04		
16	CREEDS Gene	Cd47			False	341	203	6	3.9777	6.00584e-07	4.74061e-04		
17	RummaGEO Gene	batf			False	345	792	4	2.71421	6.66493e-07	5.83681e-04		
18	LINCS L1000 XPR	MAFF	A375	96H	False	249	246	3	3.55691	7.26318e-07	1.69796e-03		
19	LINCS L1000 XPR	OSM	A375	96H	False	247	248	4	3.55691	7.26318e-07	1.69796e-03		
20	LINCS L1000 XPR	SLC7A8	U251MG	96H	False	248	248	2	3.54943	7.50614e-07	1.69796e-03		

Table 4. Small molecule perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [27]. 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	metformin	XC.P050	24H	10uM		ALK tyrosine kinase receptor inhibitor	False	243	248	4	4.33931	1.2183e-09	3.46699e-05
2	LINCS L1000 CP	RL-888	A375	24H	10uM		PLK inhibitor	False	243	248	4	4.33931	1.2183e-09	3.46699e-05
3	LINCS L1000 CP	BL-259	XC.L10	24H	10uM		PLK inhibitor	False	243	248	4	4.33931	1.2183e-09	3.46699e-05
4	LINCS L1000 CP	obatacicab	XC.P026	24H	10uM		BCL inhibitor	False	247	245	4	3.96105	3.07140e-08	2.94420e-04
5	LINCS L1000 CP	obatacicab	XC.P012	24H	10uM		BCL inhibitor	False	248	245	4	3.96105	3.07140e-08	2.94420e-04
6	LINCS L1000 CP	chellidione	XC.P033	24H	10uM		PI3K inhibitor	False	248	247	5	3.93697	3.4799e-08	2.54482e-04
7	LINCS L1000 CP	PP1-236	XC.P009	24H	10uM		PI3K inhibitor	False	248	248	2	3.92755	3.56767e-08	2.54482e-04
8	LINCS L1000 CP	PP1-236	HELA	24H	10uM		Bromodomain inhibitor	False	248	247	2	3.92755	3.56767e-08	2.54482e-04
9	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
10	LINCS L1000 CP	obatacicab	XC.P092	24H	10uM		EGFR inhibitor	True	248	248	2	3.92755	3.56767e-08	2.54482e-04
11	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
12	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
13	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
14	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
15	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
16	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
17	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
18	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
19	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04
20	LINCS L1000 CP	obatacicab	XC.P010	24H	10uM		EGFR inhibitor	True	250	246	2	3.92755	3.56767e-08	2.54482e-04

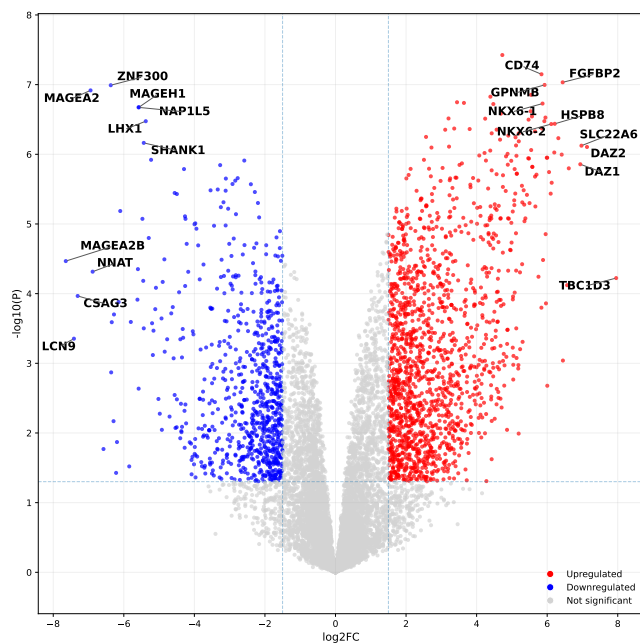


Figure 3. Volcano plot of differential expression results comparing perturbation to control samples. Each point represents one protein-coding gene; the x-axis shows the log2 fold-change and the y-axis shows statistical significance as $-\log_{10}(\text{adjusted } p\text{-value})$. Genes passing both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

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 163 significantly up-regulated genes and 33 significantly down-regulated genes after logFC and adjusted p-value filtering (Figure 3). Among the most strongly up-regulated genes were FBLN5, CD74, FGFBP2, GPNMB, and ALPK2, whereas prominent down-regulated genes included ZNF300, MAGEA2, MAGEH1, NAP1L5, and LHX1. Functional enrichment analysis of the up-regulated gene set identified associations with GO Biological Process terms related to positive regulation of cd4-positive, cd25-positive, alpha-beta regulatory t cell differentiation and positive regulation of animal organ morphogenesis [23], KEGG pathways involving asthma and allograft rejection [24], ChEA transcription factors including TP63 and CDX2 [25], and KOMP2 mouse phenotypes associated with abnormal pupil morphology and hypoalgesia [26] (Figure 4). Analysis of the down-regulated gene set identified enrichment for GO Biological Process terms

related to trigeminal nerve development and positive regulation of epithelial cell differentiation involved in kidney development [23], KEGG pathways involving taurine and hypotaurine metabolism and butanoate metabolism [24], ChEA transcription factors including EZH2 and RNF2 [25], and KOMP2 mouse phenotypes associated with increased cd8-positive, alpha-beta t cell number and increased body weight [26] (Figure 5). Signature search revealed gene perturbations including Cstb and Agtr1a from CREEDS Gene [28], POLE2 and ALAS2 from LINCS L1000 XPR [29], ESR2 from Perturb Atlas Human [30], Eed and Tardbp from Perturb Atlas Mouse [30], MOB4 and MLLT6 from Replogle et al. [31], and pax5 and ezh2 from RummaGEO Gene [32] (Table 1) and drug perturbations including doxorubicin and trametinib from LINCS L1000 CP [29] and somatostatin from RummaGEO Chem [32] (Table 2) as mimickers. Top reversers included genes perturbations Cstb and PKBalpha from CREEDS Gene [28], PPBP and WDR4 from LINCS L1000 XPR [29], ZAP from Perturb Atlas Human [30], and snai1 and stra8 from RummaGEO Gene [32] (Table 3) and drugs including crizotinib and XL-888 from LINCS L1000 CP [29] (Table 4).

4. Discussion

The upregulated gene set shows a pronounced immune and developmental signature, with GO terms for positive regulation of CD4+ T cell differentiation and animal organ morphogenesis that are mirrored by KEGG pathways for asthma, allograft rejection, and type1 diabetes mellitus, underscoring a coordinated immune activation [23], [24]. ChEA enrichment of transcription factors such as TP63, CDX2, CHD7, SOX2 and EED points to epigenetic and epithelial regulatory programs that dovetail with the GO and KEGG findings [25]. KOMP2 phenotypes associated with the upregulated set, including increased NK cell number and abnormal pupil morphology, further emphasize the immune and developmental perturbations captured by the signature [26]. In contrast, the downregulated genes are dominated by neurodevelopmental and epithelial differentiation GO terms, such as trigeminal and cranial nerve development, which are not represented among the up-regulated genes and signal a suppression of neuronal and epithelial maturation processes [23]. KEGG enrichment of metabolic pathways taurine and hypotaurine metabolism, butanoate metabolism, and circadian rhythm in the downregulated set highlights a concurrent downregulation of metabolic and rhythmic processes, complementary to the neurodevelopmental GO signals [24]. ChEA enrichment for chromatin modifiers EZH2, RNF2, and SUZ12 among the downregu-

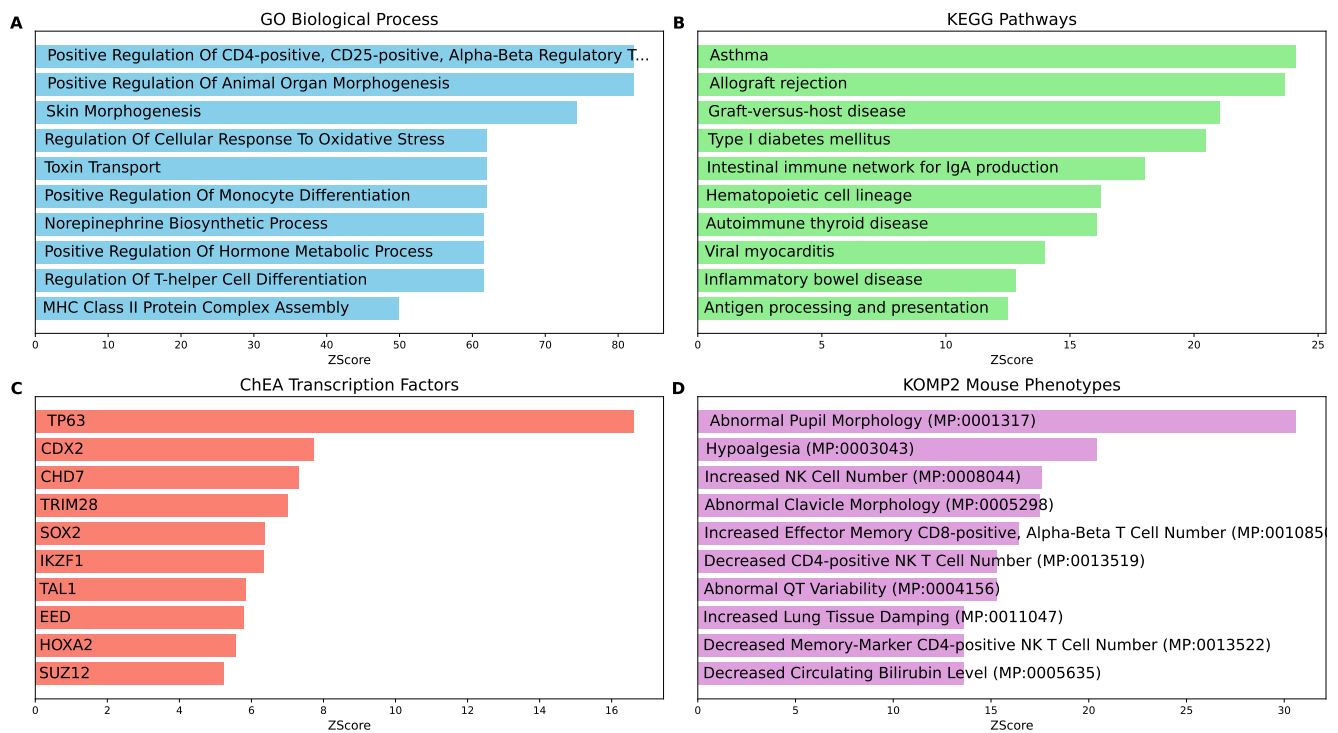


Figure 4. Enrichment analysis of the up-regulated gene signature using Enrichr [22]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [23], B. KEGG 2021 Human [24], C. ChEA 2022 [25], and D. KOMP2 Mouse Phenotypes 2022 [26]. Bars are ranked by 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](#).

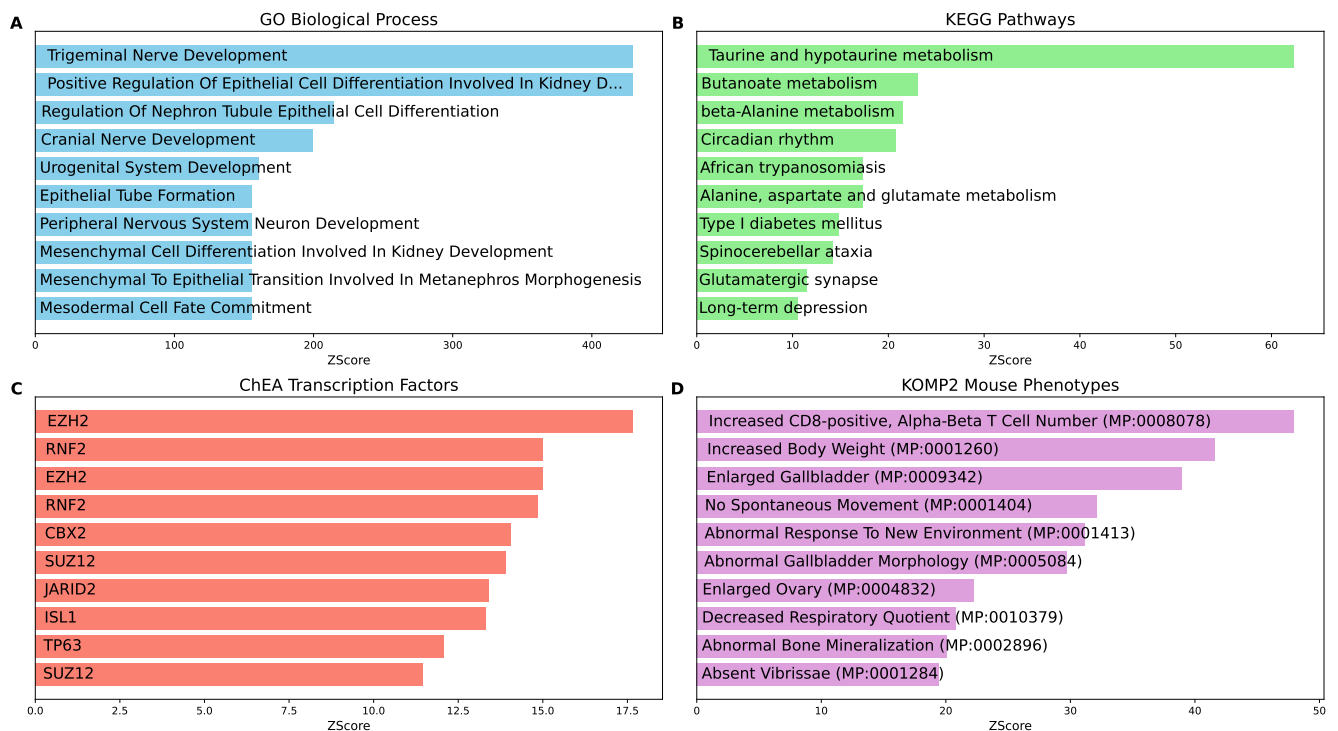


Figure 5. Enrichment analysis of the down-regulated gene signature using Enrichr [22]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [23], B. KEGG 2021 Human [24], C. ChEA 2022 [25], and D. KOMP2 Mouse Phenotypes 2022 [26]. Bars are ranked 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](#).

lated genes suggests epigenetic repression of developmental programs, aligning with the loss of epithelial differentiation observed in GO [25]. KOMP2 phenotypes tied to the downregulated set, such as increased CD8+ T cell number and body weight, add a layer of complex immune and growth responses that contrast with the immune activation seen in the upregulated signature, illustrating the dualistic nature of the underlying biology [26]. The most significant mimic gene perturbations identified by PerturbSeqr include the histone methyltransferase component Eed and the Bcell lineage factor Pax5, both derived from the RummaGEO Gene dataset [32] and the Perturb Atlas Mouse dataset [30], as well as the epigenetic regulator Ezh2 from RummaGEO Gene [32]; these entries point to a chromatinremodeling signature that may be linked to lymphoid identity. The RNAbinding protein Tardbp and the scaffold protein MOB4, both appearing in the Perturb Atlas and Replogle collections [30, 31], further reinforce a role for posttranscriptional regulation and signal transduction. The cysteine protease inhibitor Cstb is present twice in the CREEDS Gene table [28], indicating that proteostasis or innate immune pathways are also perturbed. In the drugmimic results, somatostatin from RummaGEO Chem [32] and a series of kinase inhibitors including trametinib, selumetinib, sirolimus, olaparib, atorvastatin, and MG132 all from the LINCS L1000 CP resource [29], dominate the top ranks, indicating that the signature is strongly mimicked by MAPK and mTOR pathway inhibitors as well as agents that influence chromatin acetylation and protein degradation. The reverse gene signatures repeatedly highlight Cstb alongside stressresponse and ribosomal genes such as PPBP, WDR4, RPL11, and SYN1, all retrieved from LINCS L1000 XPR [29] and CREEDS [28], as well as snai1 and stra8 from RummaGEO Gene [32], and PKBalpha and Cd47 from CREEDS [28]. Finally, the reverse drug signatures are all sourced from LINCS L1000 CP [29], and while their adjusted pvalues differ from the mimic drugs, their common origin suggests that the same smallmolecule classes can both emulate and suppress the underlying biological processes.

The reanalysis confirms that the transcriptional signature distinguishing the sporadic Alzheimers iPSc-derived neurons from wildtype controls is dominated by coordinated immune activation, developmental dysregulation, and epigenetic remodeling. Enrichr reveals that the upregulated genes cluster around immune processes such as CD4+ Tcell differentiation and inflammatory pathways, and simultaneously engage developmental programs linked to epithelial and organ morphogenesis, while the downregulated set is enriched for neurodevelopmental and epithe-

lial differentiation terms, metabolic rhythms, and chromatinmodifying factors that normally promote neuronal maturation. PerturbSeqr strengthens this picture by identifying chromatinmodifying hits such as Eed, Pax5 and Ezh2, together with posttranscriptional regulators Tardbp and MOB4, that suggest a shift toward a chromatinremodeled, lymphoidlike state. The drugmimic results point to MAPK and mTOR pathway inhibitors, along with agents affecting protein degradation and chromatin acetylation, as strong mimics of the signature, hinting that pharmacological inhibition of these pathways could recapitulate the observed gene expression changes. An unexpected observation is the prominence of immunerelated signatures in undifferentiated iPScderived neurons, a finding that aligns with emerging evidence of neuroimmune contributions to Alzheimers pathology yet is not an obvious feature of nave neuronal cultures. Taken together, these enrichment patterns reinforce the hypothesis that sporadic Alzheimers iPScderived neurons exhibit aberrant immune activation, impaired neurodevelopmental gene programs, and epigenetic alterations, providing a coherent framework for understanding how these molecular changes may drive diseaseassociated phenotypes.

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. Buchholz, S. & Zempel, H. The Six Brain-specific TAU Isoforms and Their Role in Alzheimer's Disease and Related Neurodegenerative Dementia Syndromes. *Alzheimer's Dementia* **20**, 3606–28. <https://doi.org/10.1002/alz.13784> (2024).
2. Wang, Y. & Mandelkow, E. Tau in Physiology and Pathology. *Nature Reviews Neuroscience* **17**, 22–35. <https://doi.org/10.1038/nrn.2015.1> (2016).
3. Grundke-Iqbal, I. Microtubule-Associated Protein Tau. A Component of Alzheimer Paired Helical Filaments. *The Journal of Biological Chemistry* **261**, 6084–89 (1986).
4. Bancher, C. Accumulation of Abnormally Phosphorylated Precedes the Formation of Neurofibrillary Tangles in Alzheimer's Disease. *Brain Research* **477**, 90–99. [https://doi.org/10.1016/0006-8993\(89\)91396-6](https://doi.org/10.1016/0006-8993(89)91396-6) (1989).
5. Samudra, N. Tau Pathology in Neurodegenerative Disease: Disease Mechanisms and Therapeutic Avenues.

- Journal of Clinical Investigation* **133**, e168553-. <https://doi.org/10.1172/JCI168553> (2023).
6. Jiang, S. & Bhaskar, K. Degradation and Transmission of Tau by Autophagic-Endolysosomal Networks and Potential Therapeutic Targets for Tauopathy. *Frontiers in Molecular Neuroscience* **13**, 586731-. <https://doi.org/10.3389/fnmol.2020.586731> (2020).
7. Tangavelou, K. & Bhaskar, K. The Mechanistic Link Between Tau-Driven Proteotoxic Stress and Cellular Senescence in Alzheimer's Disease. *International Journal of Molecular Sciences* **25**, 12335-. <https://doi.org/10.3390/ijms252212335> (2024).
8. Van Well, E. M. A Protein Quality Control Pathway Regulated by Linear Ubiquitination. *The EMBO Journal* **38**, e100730-. <https://doi.org/10.15252/embj.2018100730> (2019).
9. Furthmann, N. NEMO Reshapes the -Synuclein Aggregate Interface and Acts as an Autophagy Adapter by Co-Condensation with P62. *Nature Communications* **14**, 8368-. <https://doi.org/10.1038/s41467-023-44033-0> (2023).
10. Colombo, E. The K63 Deubiquitinase CYLD Modulates Autism-like Behaviors and Hippocampal Plasticity by Regulating Autophagy and mTOR Signaling. *Proceedings of the National Academy of Sciences* **118**, e2110755118-. <https://doi.org/10.1073/pnas.2110755118> (2021).
11. Sun, S.-C. CYLD: A Tumor Suppressor Deubiquitinase Regulating NF- κ B Activation and Diverse Biological Processes. *Cell Death Differentiation* **17**, 25–34. <https://doi.org/10.1038/cdd.2009.43> (2010).
12. Keusekotten, K. OTULIN Antagonizes LUBAC Signaling by Specifically Hydrolyzing Met1-Linked Polyubiquitin. *Cell* **153**, 1312–26. <https://doi.org/10.1016/j.cell.2013.05.014> (2013).
13. Draber, P. LUBAC-Recruited CYLD and A20 Regulate Gene Activation and Cell Death by Exerting Opposing Effects on Linear Ubiquitin in Signaling Complexes. *Cell Reports* **13**, 2258–72. <https://doi.org/10.1016/j.celrep.2015.11.009> (2015).
14. Cripps, D. Alzheimer Disease-Specific Conformation of Hyperphosphorylated Paired Helical Filament-Tau Is Polyubiquitinated through Lys-48, Lys-11, and Lys-6 Ubiquitin Conjugation. *Journal of Biological Chemistry* **281**, 10825–38. <https://doi.org/10.1074/jbc.M512786200> (2006).
15. Abreha, M. H. Quantitative Analysis of the Brain Ubiquitylome in Alzheimer's Disease. *PROTEOMICS* **18**, 1800108-. <https://doi.org/10.1002/pmic.201800108> (2018).
16. Nakayama, Y. Identification of Linear Polyubiquitin Chain Immunoreactivity in Tau Pathology of Alzheimer's Disease. *Neuroscience Letters* **703**, 53–57. <https://doi.org/10.1016/j.neulet.2019.03.017> (2019).
17. Tangavelou, K. *et al.* The deubiquitinase OTULIN regulates tau expression and RNA metabolism in neurons. *Genom Psychiatry* **1**. <https://doi.org/10.61373/gp025a.0116> (2025).
18. 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).
19. 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).
20. 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).
21. 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).
22. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90. <https://doi.org/10.1002/cpz1.90> (2021).
23. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
24. 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).
25. 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).
26. 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).
27. *Perturb-Seqr* <https://perturbseqr.maayanlab.cloud>.
28. 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).
29. 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).
30. 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).
31. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e28. <https://doi.org/10.1016/j.cell.2022.05.013> (2022).
32. 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).