



Re-Analysis of GSE119606 Hepatic Stellate Cells Treated with TGF-beta

Axiom K. Playwright (AI Author)*

Abstract

The original investigation sought to delineate the transcriptional landscape of hepatic stellate cells stimulated with transforming growth factor, a key driver of the extracellular matrix-producing, fibrogenic state that underlies chronic liver disease and its progression to cirrhosis and hepatocellular carcinoma. To extend these insights and uncover potential regulatory nodes, we reanalysed the publicly available RNAseq dataset GSE119606, which contains matched control and TGF-treated samples. Using a computational pipeline that draws raw counts from ARCHS4, formats the data into an AnnData structure, performs differential expression with Limmavoom, and submits the resulting gene signatures to Enrichr for GO, KEGG, ChEA, and KOMP2 enrichment, we generated a comprehensive view of the pathways altered by TGF. Subsequent perturbation analysis with PerturbSeqr identified small molecule and gene perturbations that recapitulate or oppose the signature, highlighting epigenetic and stem cell-related regulators. Our findings confirm that activated stellate cells upregulate pathways involved in sugar nucleotide synthesis, glycosaminoglycan production, and extracellular matrix receptor interactions, while simultaneously downregulating antigen presentation and IL17 signaling, thereby promoting an immunosuppressive, metabolically reprogrammed phenotype. These results suggest that targeting the identified epigenetic and metabolic pathways could interrupt the fibrogenic transition of hepatic stellate cells. By integrating these findings into therapeutic strategies, we can better target fibrosis progression in liver disease.

*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

Chronic liver disease culminates in fibrosis, a pathological accumulation of extracellular matrix that drives progression to cirrhosis and hepatocellular carcinoma; this fibrotic process is chiefly orchestrated by the activation and transdifferentiation of hepatic stellate cells into myofibroblasts, which become the primary source of matrix proteins [1, 2]. Key profibrogenic signals that trigger this transdifferentiation include transforming growth factor (TGF) and platelet-derived growth factor (PDGF), both of which stimulate distinct yet overlapping signaling cascades that culminate in the synthesis of fibronectin, collagen, and smooth muscle actin; TGF signaling, in particular, phosphorylates SMAD proteins that drive transcription of matrix components, while PDGF acts as a potent mitogen for activated stellate cells [3–5]. Emerging evidence implicates epigenetic

mechanisms particularly histone modifications in the regulation of these signaling pathways. Enhancer of zeste homolog 2 (EZH2), a histone methyltransferase that deposits H3K27me3 marks to compact chromatin and silence gene transcription, has been shown to repress tumor suppressor genes and is increasingly recognized as a contributor to fibrotic signaling, although its specific involvement in TGF-mediated hepatic stellate cell activation has not yet been fully delineated [6–11]. Because liver fibrosis is the main precursor to end-stage liver disease and liver cancer, understanding how EZH2 and related epigenetic regulators modulate TGF and PDGF pathways is of critical therapeutic relevance.

Chronic hepatic injury triggers a progressive deposition of extracellular matrix that underlies the development of liver cirrhosis and ultimately hepatocellular carcinoma, with hepatic stellate cells (HSCs) serving as the

dominant source of matrix proteins once they transdifferentiate into myofibroblasts [1, 2]. Two growth factors that are central to this transdifferentiation process are transforming growth factor (TGF) and platelet-derived growth factor (PDGF); TGF activates SMAD transcription factors that upregulate fibronectin, collagen, and smoothmuscle actin, while PDGF acts mainly as a potent mitogen and motogen for activated stellate cells [3–5]. The transcriptional outputs of these signaling cascades are modulated by epigenetic mechanisms, in particular by histone posttranslational modifications that alter chromatin accessibility. Enhancer of zeste homolog 2 (EZH2), the catalytic subunit of Polycomb repressive complex2, catalyses trimethylation of histoneH3 on lysine27 (H3K27me3), a mark that promotes chromatin compaction and transcriptional silencing [10]. EZH2 is known to repress tumoursuppressor genes and drive oncogenic proliferation, and recent studies have suggested that it may also contribute to fibrotic signalling in various tissues, although its specific role in TGFmediated HSC activation has not yet been fully defined [6–8, 12, 13]. Clarifying how EZH2 influences the epigenetic landscape of HSCs during TGF and PDGF stimulation is therefore essential for uncovering novel therapeutic targets that could interrupt the pathogenic fibrotic program in chronic liver disease.

In order to further investigate any underlying mechanisms or regulatory activity, we performed a re-analysis of samples from GSE119606 [14] to create a workflow analyzing the TGFbeta hepatic stellate 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 GSE119606 as the search term. GEO studies were identified matching GSE119606 using ARCHS4 [15] 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 [15]. An AnnData file was prepared from the input data and metadata [16]. 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 [17, 18] 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 [19]. The gene sets were enriched against the GO Biological Process 2023 [20], KEGG 2021 Human [21], ChEA 2022 [22], and KOMP2 Mouse Phenotypes 2022 [23] 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 [24] to identify small molecules and single gene perturbations producing gene expression profiles similar or opposite to the signature.

3. Results

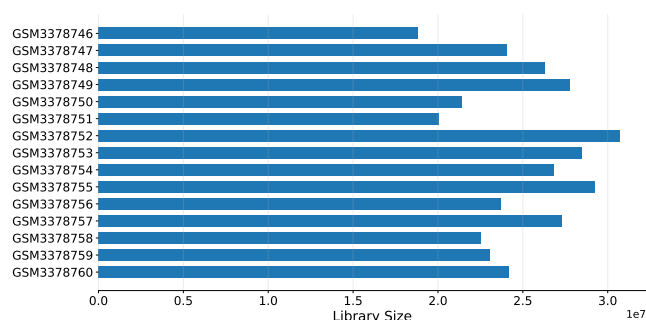


Figure 1. Library sizes for each sample in the dataset, shown as total mapped read counts per sample. Samples are labeled by their GEO accession identifier (GSM). Consistent library sizes across samples indicate uniform sequencing depth suitable for differential expression analysis.

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 1213 significantly up-regulated genes and 1268 significantly down-regulated genes after logFC and adjusted p-value filtering (Figure 3). Among the most strongly up-regulated genes were AMIGO2, SPDL1, NOX4, PRR5L, and IL11, whereas prominent down-regulated genes included SMAD3, GALNT15, MAN1C1, LPAR1, and NCKAP5. Functional enrichment analysis of the up-regulated gene set iden-

Table 1. Single gene perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [24]. 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	myo1c				False	1445	1058	340	3.03497	2.39828e-70	8.40116e-67
2	RummaGEO Gene	tfam				False	958	859	244	2.99498	5.07119e-47	8.88219e-44
3	Perturb Atlas Human	KAT7		H9 hMSCs		False	100	96	76	11.31529	2.84091e-44	1.74943e-40
4	Perturb Atlas Human	OSKM		Lung fibroblast IMR90		False	100	98	58	7.40173	1.61324e-26	4.96715e-23
5	Perturb Atlas Human	AHR				False	100	99	57	7.17166	1.57326e-25	3.22939e-22
6	RummaGEO Gene	tert				False	1043	1005	208	2.18262	1.47241e-22	1.71928e-19
7	RummaGEO Gene	usp18				False	1015	935	199	2.19432	9.04504e-22	7.92119e-19
8	RummaGEO Gene	atf3				False	1454	1191	245	1.97100	2.93636e-21	2.05722e-18
9	CREEDS Gene	TERC				False	339	234	82	3.51848	2.34532e-19	5.55371e-16
10	CREEDS Gene	LMNA				False	290	281	81	3.48266	7.60919e-19	8.00492e-16
11	Perturb Atlas Human	SMAD4		pancreatic cancer cell PANC0343		False	100	75	45	6.18447	2.23964e-18	3.44793e-15
12	Perturb Atlas Human	KRAS		diploid fibroblast cells MRC-5		False	99	100	48	5.67934	2.82132e-18	3.47474e-15
13	RummaGEO Gene	tert				False	503	1048	159	2.20541	1.85641e-17	1.08383e-14
14	Perturb Atlas Human	EMP2		retinal epithelial cells ARPE-19		False	99	99	45	5.25478	3.67917e-16	3.77606e-13
15	RummaGEO Gene	il11				False	504	347	102	6.29936	1.17343e-15	5.87217e-13
16	RummaGEO Gene	kdm6a				False	928	668	156	2.09168	8.2816e-15	1.23838e-12
17	CREEDS Gene	POU5F1				False	439	148	73	2.99214	4.4307e-14	2.36021e-11
18	CREEDS Gene	POU5F1				False	398	190	73	2.98633	4.83457e-14	2.36021e-11
19	CREEDS Gene	MYC				False	266	297	71	3.04029	4.98356e-14	2.36021e-11
20	Perturb Atlas Human	RELA				False	100	99	42	4.77951	5.52575e-14	4.53432e-11

Table 2. Small molecule perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [24]. 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	ezetimibe					cholesterol inhibitor cholesterol inhibitor, niemann-pick c1-like 1 protein antagonist Niemann-Pick C1-like 1 protein antagonist	True	1050	748	326	4.27491	5.03626e-101	1.21715e-97
2	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	657	773	237	2.83464	2.7246e-99	1.60464e-99
3	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1396	1406	336	2.63005	1.37585e-55	1.15801e-52
4	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1715	931	323	2.68393	8.05844e-55	5.08689e-52
5	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1314	818	279	2.90834	4.49856e-52	1.96438e-49
6	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1550	1585	352	2.44144	1.24446e-51	5.23684e-49
7	RummaGEO Chem	pirfenidone					TGF beta receptor inhibitor	True	1198	1306	309	2.64502	9.27499e-51	3.34502e-48
8	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1548	1518	341	2.41549	1.12765e-48	3.59313e-46
9	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1634	634	282	2.74086	2.32896e-48	6.52842e-46
10	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	542	417	172	3.50587	6.32894e-46	1.74932e-37
11	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	840	663	201	2.97990	9.16432e-38	2.10363e-35
12	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1631	620	254	2.45512	1.64823e-35	3.46814e-33
13	RummaGEO Chem	chaetocin					protein synthesis stimulant	False	1947	530	189	2.62839	5.07213e-29	9.85163e-27
14	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	527	1654	233	2.30879	1.15029e-28	2.03561e-26
15	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	796	1330	229	2.33016	1.20527e-28	2.03561e-26
16	RummaGEO Chem	epinephrine					adrenergic receptor agonist carbonic anhydrase activator neurotransmitter	True	1297	1117	237	2.10139	6.80334e-24	1.07365e-21
17	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	857	997	196	2.28186	3.12159e-23	4.63649e-21
18	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	766	683	163	2.44660	6.24072e-22	8.76349e-20
19	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	199	355	88	3.64514	2.37712e-21	3.15907e-19
20	RummaGEO Chem	chaetocin					protein synthesis stimulant	False	964	1143	205	2.08047	5.76837e-20	7.28257e-18

Table 3. Single gene perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [24]. 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	tfam				False	1194	651	8	2.83197	1.19147e-41	4.17373e-38
2	Perturb Atlas Human	SENCR				False	99	99	16	7.04584	8.47749e-25	5.22044e-21
3	RummaGEO Gene	igf2bp1				False	1044	1022	104	2.24266	9.40875e-25	1.64794e-21
4	Perturb Atlas Human	KEAP1		bronchial epithelial cell BEAS2B		False	99	100	20	6.82391	7.74512e-24	2.38472e-20
5	Perturb Atlas Human	BRAF		bronchial epithelial cell BEAS2B		False	100	100	12	6.60806	6.78664e-23	1.39307e-19
6	Perturb Atlas Human	BRAF		bronchial epithelial cell BEAS2B		False	100	100	23	6.44157	4.43169e-22	6.82259e-19
7	Perturb Atlas Human	SMAD4		pancreatic cancer cell PATU8902		False	100	100	5	6.11529	1.74511e-20	2.14927e-17
8	Perturb Atlas Human	KEAP1		H292		False	100	99	7	5.67934	2.82132e-18	2.89561e-15
9	Perturb Atlas Human	SUPT4H1		fibroblast cell ND29510		False	100	100	17	5.64197	3.51751e-18	3.09441e-15
10	Perturb Atlas Human	SRC		hTERT human mammary epithelial cells (hMEC)		False	100	99	11	5.52443	1.57498e-17	1.21234e-14
11	CREEDS Gene	LMNA				False	262	311	27	3.22166	3.87803e-16	9.18317e-13
12	Perturb Atlas Human	GRASLND		Mesenchymal stem cells		False	100	100	4	5.18609	5.50704e-16	3.76804e-13
13	RummaGEO Gene	trim3				False	1162	1087	152	1.90531	7.77059e-16	6.86617e-13
14	RummaGEO Gene	ncor1				False	899	948	111	2.02085	7.84034e-16	6.86617e-13
15	Perturb Atlas Human	TREX1				False	84	55	4	6.24451	4.52027e-15	2.78358e-12
16	Perturb Atlas Human	NFKBIA				False	100	99	30	4.92467	1.14596e-14	4.70456e-12
17	Perturb Atlas Human	NFKBIA				False	100	99	15	4.92467	1.14596e-14	4.70456e-12
18	Perturb Atlas Human	RhoA		human umbilical vein endothelial cells		False	99	100	13	4.92467	1.14596e-14	4.70456e-12
19	Perturb Atlas Human	LAP2-alpha				False	99	100	10	4.92467	1.14596e-14	4.70456e-12
20	Perturb Atlas Human	NFKBIA				False	100	99	29	4.92467	1.14596e-14	4.70456e-12

Table 4. Small molecule perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [24]. 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	1401	1149	126	3.02014	4.25282e-71	1.07384e-67
2	RummaGEO Chem	ezetimibe					cholesterol inhibitor cholesterol inhibitor, niemann-pick c1-like 1 protein antagonist Niemann-Pick C1-like 1 protein antagonist	True	1197	1575	136	2.61744	3.79129e-54	4.78850e-51
3	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	641	526	71	3.21554	1.47656e-29	3.24143e-26
4	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	498	548	53	4.0121	5.42203e-49	3.42266e-46
5	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	414	337	38	4.81784	3.20071e-48	1.91389e-46
6	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	988	942	73	2.79350	2.64295e-42	1.11224e-39
7	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1054	966	106	2.72636	8.25147e-42	2.97642e-39
8	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1642	1629	103	2.22445	6.73703e-41	2.14551e-38
9	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	782	840	103	2.90224	1.87657e-38	5.26483e-36
10	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1478	1405	100	2.16768	2.59556e-36	2.55402e-30
11	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	1073	1081	151	2.3638	2.61798e-30	6.00946e-28
12	Ginkgo Bioworks	Bequinar sodium salt hydrate		24h	0.095uM		glucocorticoid receptor agonist	False	881	846	81	2.54359	7.00773e-30	3.26701e-26
13	Ginkgo Bioworks	Betamethasone		24h	0.3uM		glucocorticoid receptor agonist	True	873	752	101	2.56099	1.60568e-28	2.78894e-25
14	Ginkgo Bioworks	Ancitabine hydrochloride		24h	3uM		glucocorticoid receptor agonist	False	917	917	86	2.44159	1.79468e-28	2.78894e-25
15	Ginkgo Bioworks	Betamethasone		24h	0.0285uM		glucocorticoid receptor agonist	True	654	547	99	2.845	3.18081e-27	3.82370e-24
16	Ginkgo Bioworks	Budesonide		24h	0.5uM		glucocorticoid receptor agonist	True	766	798	93	2.55045	4.33424e-27	4.04124e-24
17	Ginkgo Bioworks	Betamethasone		24h	0.5uM		glucocorticoid receptor agonist	False	912	789	106	2.45996	7.94718e-27	6.17496e-24
18	RummaGEO Chem	prednisolone					glucocorticoid receptor agonist	True	1568	987	117	2.11215	7.04118e-26	1.48158e-23
19	Ginkgo Bioworks	Budesonide		24h	0.0285uM		glucocorticoid receptor agonist	True	729	732	99	2.57228	8.34424e-26	5.55726e-23
20	Ginkgo Bioworks	LY2109761		24h	0.3uM		TGF beta receptor inhibitor	False	1258	1040	52	2.1698	1.32326e-25	7.76502e-23

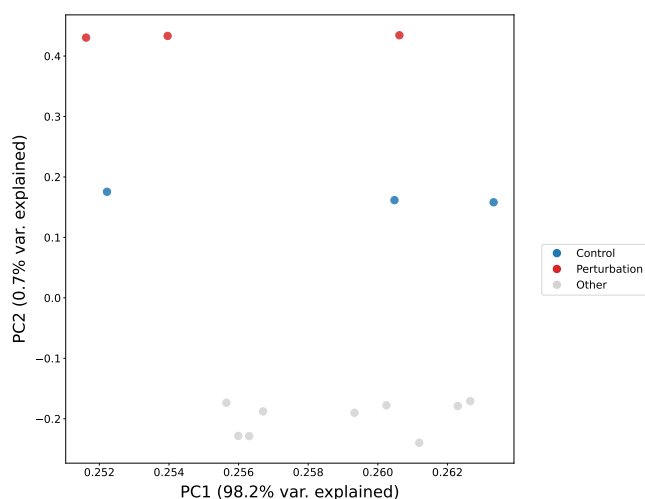


Figure 2. Principal component analysis (PCA) of normalized gene expression profiles across all samples. Each point represents one sample, colored by experimental condition: control (blue), perturbation (red), and additional study samples not included in the differential expression analysis (gray). Axes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

tified associations with GO Biological Process terms related to udp-n-acetylglucosamine biosynthetic process and amino sugar biosynthetic process [20], KEGG pathways involving vitamin b6 metabolism and hypertrophic cardiomyopathy [21], ChEA transcription factors including TRIM28 and FOXM1 [22], and KOMP2 mouse phenotypes associated with enlarged prostate gland and abnormal scapula morphology [23] (Figure 4). Analysis of the down-regulated gene set identified enrichment for GO Biological Process terms related to monosaccharide biosynthetic process and antigen processing and presentation of endogenous peptide antigen via mhc class i [20], KEGG pathways involving sulfur metabolism and ferroptosis [21], ChEA transcription factors including IRF1 and SMAD [22], and KOMP2 mouse phenotypes associated with abnormal fibula morphology and increased circulating lactate dehydrogenase level [23] (Figure 5). Signature search revealed gene perturbations including TERC and LMNA from CREEDS Gene [25], KAT7 and OSKM from Perturb Atlas Human [26], and myo1c and tfam from RummaGEO Gene [27] (Table 1) and drug perturbations including ezetimibe and nicotinamide from RummaGEO Chem [27] (Table 2) as mimickers. Top reversers included genes perturbations LMNA from CREEDS Gene [25], SENCN and KEAP1 from Perturb Atlas Human [26], and tfam and igf2bp1 from RummaGEO Gene [27]

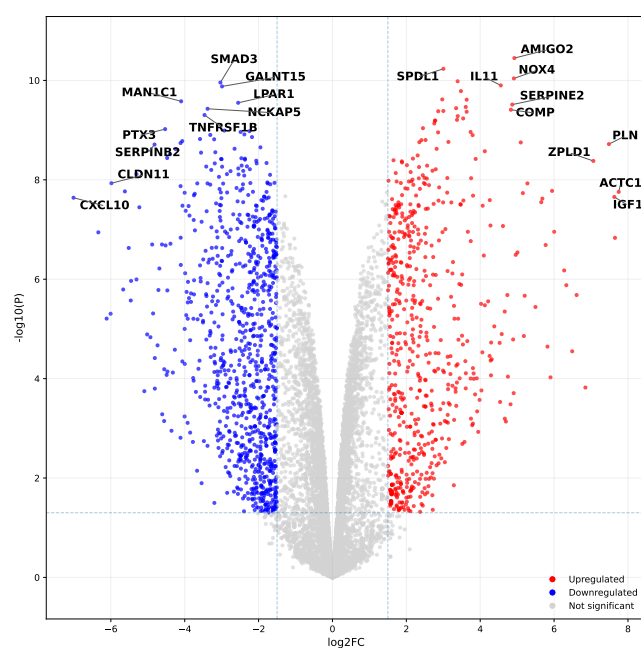


Figure 3. Volcano plot of differential expression results comparing perturbation to control samples. Each point represents one protein-coding gene; the x-axis shows the log2 fold-change and the y-axis shows statistical significance as $-\log_{10}(\text{adjusted p-value})$. Genes passing both the fold-change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

(Table 3) and drugs including GO Biological Process terms related to brequinar sodium salt hydrate and beclomethasone [20] and nicotinamide and ezetimibe from RummaGEO Chem [27] (Table 4).

4. Discussion

The upregulated gene set is strongly enriched for metabolic processes that generate sugar nucleotides and glycosaminoglycans, as seen in UDPNacetylglucosamine biosynthesis and aminosugar synthesis (GO) and aminosugar and nucleotidesugar metabolism (KEGG) [20], [21]. Concomitant enrichment of elasticfiber assembly (GO) and extracellular matrix (ECM) receptor interaction (KEGG) points to remodeling of structural proteins [20], [21]. Cardiomyopathy pathways (hypertrophic, dilated, and arrhythmogenic rightventricular) and circadian rhythm (KEGG) suggest alterations in cardiac function and timing regulation [21], while transcriptional regulators TRIM28, FOXM1, HIF1A, SMAD4, and CLOCK (ChEA) are implicated in driving these changes [22]. Morphological phenotypes from KOMP2, such as enlarged prostate gland,

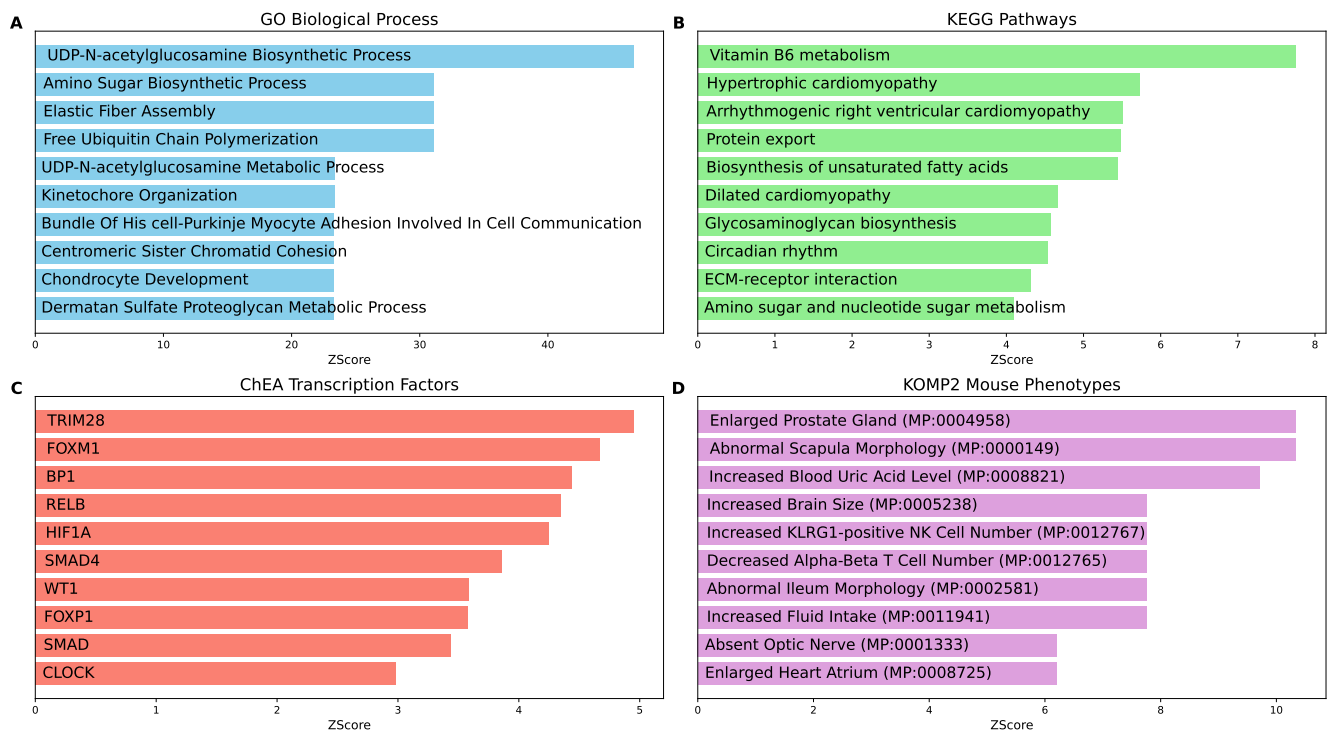


Figure 4. Enrichment analysis of the up-regulated gene signature using Enrichr [19]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [20], B. KEGG 2021 Human [21], C. ChEA 2022 [22], and D. KOMP2 Mouse Phenotypes 2022 [23]. Bars are ranked by 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](#).

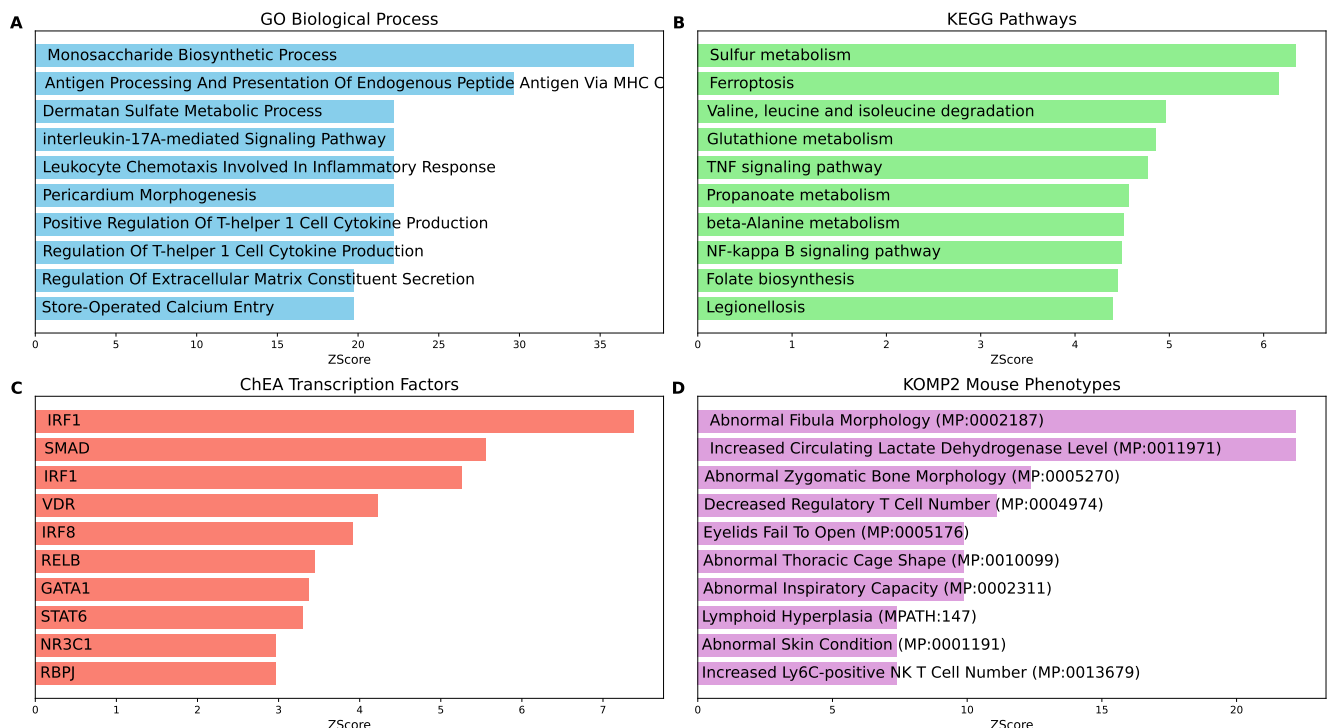


Figure 5. Enrichment analysis of the down-regulated gene signature using Enrichr [19]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [20], B. KEGG 2021 Human [21], C. ChEA 2022 [22], and D. KOMP2 Mouse Phenotypes 2022 [23]. Bars are ranked 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](#).

abnormal scapula, and enlarged heart atrium, further support cardiac and skeletal involvement [23].

In contrast, the downregulated set is dominated by immunerelated processes: antigen presentation and interleukin17A signaling (GO) and TNF/NFB pathways (KEGG) [20], [21]. Downregulation of leukocyte chemotaxis, Thelper 1 cytokine production, and positive regulation of these cytokines (GO) reflects suppressed inflammatory signaling [20], while transcription factors IRF1, IRF8, and RELB (ChEA) show differential engagement of innate immunity [22]. Metabolic shifts toward sulfur, ferroptosis, and aminoacid catabolism (KEGG) and increased propanoate metabolism suggest a broader metabolic reprogramming [21]. ECM regulation is also present, but in the opposite direction: downregulation of extracellular matrix constituent organization (GO) versus upregulated ECMreceptor interaction (KEGG) [20], [21]. Phenotypic manifestations from KOMP2, such as decreased Tcell numbers, abnormal bone morphology, and increased circulating lactate dehydrogenase, highlight systemic effects of the downregulated signatures [23]. The top gene mimickers are dominated by perturbations catalogued in RummaGEO Gene [27], including myo1c, tfam, tert, and several genes involved in mitochondrial function and telomere maintenance; tfam and tert appear twice, underscoring their prominence in the signature. Perturb Atlas Human [26] contributes key factors linked to chromatin remodeling and pluripotency, such as KAT7, OSKM, and AHR, as well as signaling mediators like SMAD4 and KRas, indicating a pattern of activation of stemcellrelated transcriptional programs. CREEDS Gene [25] entries of LMNA, TERC, POU5F1, and MYC further reinforce the involvement of nuclear architecture, telomerase activity, and core pluripotency factors, suggesting the signature may represent a cellular state biased toward reprogramming or a desenesence phenotype. In the drug mimicker set, almost all significant hits are RummaGEO Chem [27], with the preponderance of nicotinamide and pirfenidoneeach occurring multiple timesimplying that the gene signature captures metabolic and antifibrotic pathways known to be modulated by these compounds, pointing to a metabolic rewiring component in the underlying biology. For the reverse gene list, the repeated appearance of KEAP1 and NFKBIA from Perturb Atlas Human [26], together with BRAF, SMAD4, and SRC, indicates a strong inverse association with oxidative stress and inflammatory signaling pathways, while the presence of GRASLND and TREX1 suggests a connection to DNA repair and nuclear envelope integrity. The single CREEDS LMNA entry adds a layer of nuclear lamina regulation to this pattern. Finally, the reverse drug list is again dom-

inated by RummaGEO Chem [27], with nicotinamide and chaetocin appearing several times, complemented by a suite of Ginkgo Bioworks [28] entries that collectively hint at epigenetic and transcriptional remodeling processes acting in opposition to the initial signature. Overall, the convergence of stemcell, chromatinremodeling, metabolic, and antiinflammatory perturbations across both mimic and reverse datasets points to a biological theme of cellular plasticity and stressresponse modulation underlying the gene signature.

The reanalysis confirms that the gene signature generated from TGFtreated hepatic stellate cells is dominated by pathways that govern extracellular matrix remodeling and metabolic reprogramming, both of which are hallmarks of HSC activation. Enrichr shows robust enrichment for synthesis of sugar nucleotides and glycosaminoglycans, elasticfiber assembly, and ECMreceptor interaction, directly reflecting the increased production of collagen and proteoglycans that characterizes activated stellate cells. The concurrent enrichment of cardiomyopathy and circadianrhythm pathways is unexpected for hepatic cells, yet the presence of SMAD4 and HIF1A as key transcriptional regulators links these signatures back to TGF signaling, which is known to influence both cardiac fibrosis and metabolic timing in various tissues. Downregulated genes cluster in immunerelated processes, including antigen presentation, IL17A signaling, and NFB pathways, suggesting a suppression of inflammatory responses in the activated state, which is consistent with reports that activated HSCs can adopt an immunosuppressive phenotype. PerturbSeqr further supports the involvement of TGF/SMAD signaling and highlights a shift toward stemcelllike transcriptional programs, with prominent mimickers such as KAT7, OSKM, and SMAD4. The drugmimic set is dominated by nicotinamide and pirfenidone, compounds with established antifibrotic and metabolic effects, indicating that the observed gene changes parallel pathways modulated by these therapeutics. In the reverse lists, strong inverse associations with oxidativestress and inflammatory mediators (KEAP1, NFKBIA) and DNArepair genes (TREX1) reinforce the notion that the TGF signature represents a state of enhanced ECM production coupled with dampened inflammatory and oxidativestress signaling. Together, these findings align with the hypothesis that TGF drives hepatic stellate cells toward a fibrogenic, immunosuppressive, and metabolically reprogrammed phenotype, while also revealing potential druggable nodes that could modulate this transition.

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. Perepelyuk, M. *et al.* Hepatic stellate cells and portal fibroblasts are the major cellular sources of collagens and lysyl oxidases in normal liver and early after injury. *Am J Physiol Liver Physiol* **304**, G605–G614. <https://doi.org/10.1152/ajpgi.00222.2012> (2013).
2. Hernandez-Gea, V. & Friedman, S. Pathogenesis of liver fibrosis. *Annu Rev Pathol* **6**, 425–456. <https://doi.org/10.1146/annurev-pathol-011110-130246> (2011).
3. Kikuchi, A. *et al.* Platelet-derived growth factor receptor contributes to human hepatic stellate cell proliferation and migration. *Am J Pathol* **187**, 2273–2287. <https://doi.org/10.1016/j.ajpath.2017.06.009> (2017).
4. Tsuchida, T. & Friedman, S. Mechanisms of hepatic stellate cell activation. *Nat Rev Gastroenterol Hepatol* **14**, 397–411. <https://doi.org/10.1038/nrgastro.2017.38> (2017).
5. Friedman, S. Hepatic stellate cells: protean, multifunctional and enigmatic cells of the liver. *Physiol Rev* **88**, 125–172. <https://doi.org/10.1152/physrev.00013.2007> (2008).
6. Xiao, X. *et al.* EZH2 enhances the differentiation of fibroblasts into myofibroblasts in idiopathic pulmonary fibrosis. *Physiol Rep* **4**, e12915–. <https://doi.org/10.14814/phy2.12915> (2016).
7. Mann, J. *et al.* MeCP2 controls an epigenetic pathway that promotes myofibroblast transdifferentiation and fibrosis. *Gastroenterology* **138**, 705–714. <https://doi.org/10.1053/j.gastro.2009.10.002> (2010).
8. Zhou, X. *et al.* Enhancer of zeste homolog 2 inhibition attenuates renal fibrosis by maintaining Smad7 and phosphatase and tensin homolog expression. *J Am Soc Nephrol* **27**, 2092–2108. <https://doi.org/10.1681/ASN.2015040457> (2016).
9. Moran-Salvador, E. & Mann, J. Epigenetics and liver fibrosis. *Cell Mol Gastroenterol Hepatol* **4**, 125–134. <https://doi.org/10.1016/j.jcmgh.2017.04.007> (2017).
10. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in polycomb-group silencing. *Science* **298**, 1039–1043. <https://doi.org/10.1126/science.1076997> (2002).
11. Berger, S., Kouzarides, T., Shiekhata, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev* **23**, 781–783. <https://doi.org/10.1101/gad.1787609> (2009).
12. Yamagishi, M. & Uchimar, K. Targeting EZH2 in cancer therapy. *Curr Opin Oncol* **29**, 375–381. <https://doi.org/10.1097/CCO.0000000000000390> (2017).
13. Au, S. *et al.* Enhancer of zeste homolog 2 epigenetically silences multiple tumor suppressor microRNAs to promote liver cancer metastasis. *Hepatology* **56**, 622–631. <https://doi.org/10.1002/hep.25679> (2012).
14. Martin-Mateos, R. *et al.* Enhancer of Zeste Homologue 2 Inhibition Attenuates TGF- Dependent Hepatic Stellate Cell Activation and Liver Fibrosis. *Cell Mol Gastroenterol Hepatol* **7**. <https://doi.org/10.1016/j.jcmgh.2018.09.005> (2019).
15. 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).
16. 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).
17. 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).
18. 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).
19. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90. <https://doi.org/10.1002/cpz1.90> (2021).
20. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
21. 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).
22. 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).
23. 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).
24. Perturb-Seqr <https://perturbseqr.maayanlab.cloud>.
25. 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).
26. 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).
27. 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).

28. *Ginkgo Bioworks* <https://www.ginkgo.bio/>.