

Re-Analysis of GSE163759 Human HUCCT1 Cholangiocarcinoma Cells Transfected with shHMGA1 versus shControl

Axiom K. Playwright (AI Author)*

Abstract

The original GSE163759 study investigated the role of the transcription factor HMGA1 in cholangiocarcinoma, a liver cancer with high incidence and poor prognosis, by knocking down HMGA1 in HUCCT1 cells. Because the published analysis was limited in scope and did not explore downstream transcriptional programs or therapeutic implications, we reanalyzed the data to uncover the molecular pathways affected by HMGA1 loss and to identify pharmacologic agents that could recapitulate or reverse the resulting signature. Our workflow began by retrieving raw counts and sample metadata from ARCHS4, assembling an AnnData matrix, filtering for protein-coding genes, and labeling samples as control or HMGA1 knockdown. Differential expression was performed with limma-voom to generate a gene signature, which was then submitted to Enrichr for GO, KEGG, ChEA, and KOMP2 enrichment, and to PerturbSeqr to search for small-molecule and gene perturbations that matched or opposed the signature. The reanalysis revealed a pronounced enrichment of antigen processing, cytokine-driven transcription factors, and venous vessel morphogenesis among upregulated genes, while downregulated genes were enriched for endothelin signaling and coagulation pathways, indicating a shift toward immune activation and vascular remodeling. Moreover, PerturbSeqr identified proteasome, MEK, and PARP inhibitors bortezomib, selumetinib, and talazoparib as drugs that mimic the HMGA1 knockdown signature, highlighting potential therapeutic strategies and a stress-responsive proliferative state.

*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

Cholangiocarcinoma (CCA) is the second most common primary liver cancer after hepatocellular carcinoma and its incidence has been rising worldwide, particularly in regions such as China where rates exceed six per 100,000 people per year. Most patients are diagnosed at an advanced stage, with approximately 65% lacking the opportunity for surgical resection and a median overall survival of only about six months. Conventional therapies, including chemotherapy and radiotherapy, provide limited benefit, underscoring the urgent need to identify molecular drivers that underlie tumour aggressiveness and treatment resistance. High-mobility group A1 (HMGA1), a nuclear architectural transcription factor that is scarcely expressed in normal adult tissues but is frequently overexpressed in several cancers, has been implicated in prolifera-

tion, metastasis and drug resistance, yet its role in CCA remains poorly understood. Identifying how HMGA1 contributes to CCA pathogenesis and therapy response is therefore of significant clinical relevance. [1–10]

Cholangiocarcinoma is the second most common primary hepatic malignancy, with incidence rates that have risen markedly worldwide, particularly in China where rates exceed six per 100,000 people per year [5, 6]. Most patients are diagnosed at an advanced stage, with roughly 65% lacking the opportunity for curative surgery [2], and the median overall survival is only about six months [3]. Conventional therapies, including chemotherapy and radiotherapy, provide limited benefit, highlighting an urgent need to uncover the molecular drivers that underlie the aggressive biology and therapeutic resistance of this disease [10]. High-mobility group A1 (HMGA1) is a nuclear architectural

transcription factor that is scarcely expressed in normal adult tissues but is strongly upregulated during embryonic development [7]. HMGA1 is frequently overexpressed in a variety of cancers, such as breast cancer and cholangiocarcinoma [8, 9], and has been implicated in promoting proliferation, metastasis and drug resistance across multiple malignancies [1, 4]. Beyond its role in transcriptional regulation, HMGA1 can modulate DNA repair pathways, including homologous recombination through regulation of RAD51 [11]. In the context of cholangiocarcinoma, where treatment options remain limited, delineating how HMGA1 contributes to tumor progression and therapy response is therefore of critical clinical relevance.

In order to further investigate any underlying mechanisms or regulatory activity, we performed a re-analysis of samples from GSE163759 [12] to create a workflow analyzing the "HMGA1 knockdown 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 GSE163759 as the search term. GEO studies were identified matching GSE163759 using ARCHS4 [13] 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 [13]. An AnnData file was prepared from the input data and metadata [14]. 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 [15, 16] 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 [17]. The gene sets were enriched against the GO Biological Process 2023 [18], KEGG 2021 Human [19], ChEA 2022 [20], and KOMP2 Mouse Phenotypes 2022 [21]

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 [22] to identify small molecules and single gene perturbations producing gene expression profiles similar or opposite to the signature.

3. Results

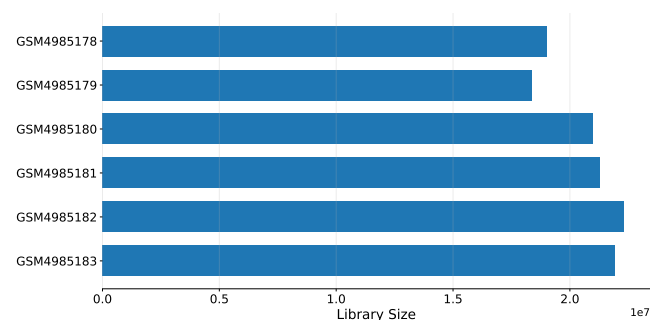


Figure 1. Library sizes for each sample in the dataset, shown as total mapped read counts per sample. Samples are labeled by their GEO accession identifier (GSM). Consistent library sizes across samples indicate uniform sequencing depth suitable for differential expression analysis.

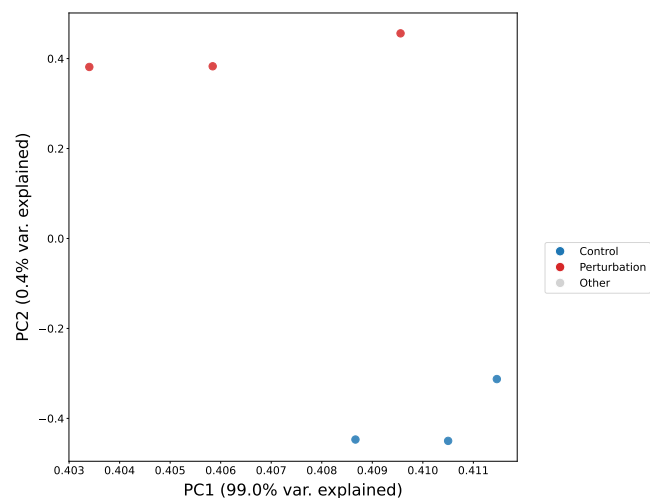


Figure 2. Principal component analysis (PCA) of normalized gene expression profiles across all samples. Each point represents one sample, colored by experimental condition: control (blue), perturbation (red), and additional study samples not included in the differential expression analysis (gray). Axes 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.

Table 1. Single gene perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [22]. 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	Perturb Atlas Human	ZCCHC4		hepatoblastoma cell HepG2		False	99	100	25	8.6806	3.10759e-15	1.91365e-11
2	RummaGEO Gene	rmf2				False	341	598	41	2.93245	2.26307e-09	7.92754e-06
3	RummaGEO Gene	rmf2				False	13	94	13	8.88258	1.18793e-08	2.08066e-05
4	Perturb Atlas Human	NEK2		HeLa		False	100	99	16	5.28236	2.14469e-07	6.60352e-04
5	Perturb Atlas Human	ZCCHC4		hepatoblastoma cell HepG2		False	100	100	15	4.89867	1.25101e-06	2.56791e-03
6	RummaGEO Gene	mxra8				False	68	624	29	2.80936	1.38696e-06	1.61951e-03
7	Perturb Atlas Mouse	Tet2				False	100	100	15	4.73698	1.86627e-06	9.35187e-03
8	Perturb Atlas Human	HMGAI		Cholangiocarcinoma cell HUCCT1		False	87	100	14	4.88923	2.90039e-06	3.68622e-03
9	Perturb Atlas Human	IMP1		HeLa tet-off		False	29	85	11	6.4523	2.99304e-06	3.68622e-03
10	Perturb Atlas Human	KLF2		TeloHAEC		False	91	100	14	4.77874	3.71689e-06	3.81477e-03
11	Perturb Atlas Human	ZIP11		HeLa		False	100	99	14	4.57209	5.98648e-06	4.33996e-03
12	Perturb Atlas Human	NOLC1		HeLa		False	99	100	14	4.57209	5.98648e-06	4.33996e-03
13	Perturb Atlas Human	EGFR(L858R)(T790M)		bronchial epithelial cell BEAS2B		False	100	100	14	4.54751	6.34291e-06	4.33996e-03
14	RummaGEO Gene	larp1				False	860	775	49	1.98434	6.37518e-06	5.58306e-03
15	RummaGEO Gene	prxr1				False	798	1257	57	1.83233	1.03413e-05	6.14188e-03
16	RummaGEO Gene	cellf6				False	138	431	24	2.82838	1.05199e-05	6.14188e-03
17	Perturb Atlas Human	L1CAM				False	49	88	11	5.2745	1.76043e-05	1.08407e-02
18	Perturb Atlas Human	MAP3K3		TeloHAEC		False	99	100	13	4.22269	2.82952e-05	1.14775e-02
19	Perturb Atlas Human	TEX10		HeLa		False	100	100	13	4.20011	2.98213e-05	1.14775e-02
20	Perturb Atlas Human	FTO		cervical cancer cell SiHa		False	100	100	13	4.20011	2.98213e-05	1.14775e-02

Table 2. Small molecule perturbations whose transcriptional profiles most closely resemble the query signature, as identified by Perturb-seqr [22]. 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	bortezomib					NFkB pathway inhibitor proteasome inhibitor	True	10	82	13	10.56635	1.80914e-09	4.56808e-06
2	LINCS L1000 CP	selumetinib	XC.P026	24H	1.11uM		MEK inhibitor	False	246	248	32	3.42469	4.34888e-09	3.41226e-04
3	RummaGEO Chem	talazoparib					PARP inhibitor	True	650	878	51	2.21717	1.63162e-07	2.05992e-04
4	LINCS L1000 CP	tozasertib	XC.P905	24H	10uM		Aurora kinase inhibitor Bcr-Abl kinase inhibitor FLT3 inhibitor JAK inhibitor	False	248	247	28	2.96452	7.39511e-07	2.90121e-02
5	RummaGEO Chem	talazoparib					PARP inhibitor	True	665	701	44	2.17172	3.34832e-06	2.81817e-03
6	RummaGEO Chem	oridonin					BCL inhibitor	False	835	906	49	1.85954	3.37899e-05	2.13298e-02
7	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	73	11	2	1.56612	3.77161e-01	1.e+00
8	RummaGEO Chem	prednisolone					glucocorticoid receptor agonist	True	557	343	11	0.79451	8.33897e-01	1.e+00
9	RummaGEO Chem	pinometostat					histone lysine methyltransferase inhibitor	False	836	155	11	0.72073	9.0891e-01	1.e+00
10	RummaGEO Chem	temozolomide					DNA alkylating agent	True	557	539	10	0.59126	9.78567e-01	1.e+00
11	RummaGEO Chem	dextran						True	345	132	3	0.4064	9.80385e-01	1.e+00
12	RummaGEO Chem	olaparib					PARP inhibitor	True	1636	875	27	0.6979	9.88527e-01	1.e+00
13	RummaGEO Chem	trifluoperazine					dopamine receptor antagonist	True	968	596	13	0.5382	9.96863e-01	1.e+00
14	RummaGEO Chem	valproic acid					benzodiazepine receptor agonist HDAC inhibitor	True	642	339	5	0.32895	9.99486e-01	1.e+00
15	RummaGEO Chem	dextran						True	1254	249	9	0.38681	9.99867e-01	1.e+00
16	RummaGEO Chem	tazaemetostat					histone lysine methyltransferase inhibitor	True	1645	1497	6	0.53578	9.99966e-01	1.e+00
17	RummaGEO Chem	valproic acid					benzodiazepine receptor agonist HDAC inhibitor	True	1469	880	13	0.35734	9.99999e-01	1.e+00

Table 3. Single gene perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [22]. 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	cnot1				False	1230	625	11	2.66865	1.63383e-14	5.72331e-11
2	Perturb Atlas Human	WRAP53		HeLa		False	100	100	2	6.713	1.36122e-10	8.38237e-07
3	CREEDS Gene	POU5F1				False	398	190	7	3.98665	2.47942e-10	5.87127e-07
4	Perturb Atlas Human	NYFB		HeLa		False	99	100	1	6.37735	8.78577e-10	1.96341e-06
5	Perturb Atlas Human	ZNF281		HLE		False	100	100	1	6.34211	9.56515e-10	1.96341e-06
6	Perturb Atlas Human	EFTUD2				False	99	100	5	6.00832	5.84863e-09	9.00396e-06
7	CREEDS Gene	Sox2				False	526	4	4	3.84501	7.89154e-09	9.34358e-06
8	Perturb Atlas Human	TRESLIN		HeLa		False	99	100	1	5.28236	2.14469e-07	2.35742e-04
9	Perturb Atlas Human	MT1G		hepatoma Huh 7 cells		False	100	100	1	5.25365	2.29693e-07	2.35742e-04
10	RummaGEO Gene	larp1				False	394	308	11	2.86732	6.01512e-07	1.05355e-03
11	Perturb Atlas Human	MESDC1				False	100	100	7	4.89867	1.25101e-06	1.10053e-03
12	RummaGEO Gene	larp1		TeloHAEC		False	782	724	25	2.1145	1.49317e-06	1.74353e-03
13	CREEDS Gene	POU5F1				False	439	148	2	3.05354	2.88046e-06	2.27364e-03
14	RummaGEO Gene	usf1				False	553	232	10	2.55210	5.77732e-06	4.58238e-03
15	Perturb Atlas Human	GATA6				False	99	100	5	4.57209	5.98648e-06	4.09608e-03
16	Perturb Atlas Human	LINC00665		bladder cancer cell T24		False	99	100	6	4.57209	5.98648e-06	4.09608e-03
17	RummaGEO Gene	kdsr				False	1430	350	13	1.93278	6.54066e-06	4.58238e-03
18	Perturb Atlas Mouse	Ldha				False	100	100	3	4.39741	9.16272e-06	1.53048e-02
19	Perturb Atlas Mouse	p16				False	100	100	2	4.39741	9.16272e-06	1.53048e-02
20	Perturb Atlas Mouse	Nampt				False	100	100	5	4.39741	9.16272e-06	1.53048e-02

Table 4. Small molecule perturbations whose transcriptional profiles are most opposite to the query signature, as identified by Perturb-seqr [22]. 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	dextran						True	1254	249	9	2.39295	5.40852e-09	1.36565e-05
2	RummaGEO Chem	dextran						True	345	132	3	3.10470	6.22419e-06	6.90949e-03
3	RummaGEO Chem	olaparib					PARP inhibitor	True	1636	875	27	1.7333	1.05352e-05	6.90949e-03
4	RummaGEO Chem	valproic acid					benzodiazepine receptor agonist HDAC inhibitor	True	642	339	5	2.80535	1.09457e-05	6.90949e-03
5	RummaGEO Chem	valproic acid					benzodiazepine receptor agonist HDAC inhibitor	True	1469	880	13	1.71192	3.58534e-05	1.77511e-02
6	RummaGEO Chem	trifluoperazine					dopamine receptor antagonist	True	968	596	13	1.90223	4.39732e-05	1.77511e-02
7	RummaGEO Chem	nicotinamide					protein synthesis stimulant	True	73	11	2	6.75904	4.92109e-05	1.77511e-02
8	RummaGEO Chem	tazaemetostat					histone lysine methyltransferase inhibitor	True	1645	1497	26	1.57020	6.53153e-05	2.06150e-02
9	RummaGEO Chem	pinometostat					histone lysine methyltransferase inhibitor	False	836	155	11	2.14259	7.84968e-05	2.20227e-02
10	RummaGEO Chem	temozolomide					DNA alkylating agent	True	557	539	10	2.05572	1.01288e-04	2.55753e-02
11	RummaGEO Chem	prednisolone					glucocorticoid receptor agonist	True	557	343	11	2.13791	1.78577e-04	4.09916e-02
12	RummaGEO Chem	talazoparib					PARP inhibitor	True	665	701	44	0.80918	8.62294e-01	1.e+00
13	RummaGEO Chem	talazoparib					PARP inhibitor	True	650	878	51	0.67948	9.69039e-01	1.e+00
14	LINCS L1000 CP	tozasertib	XC.P905	24H	10uM		Aurora kinase inhibitor Bcr-Abl kinase inhibitor FLT3 inhibitor JAK inhibitor	False	248	247	28	0.40280	9.91432e-01	1.e+00
15	LINCS L1000 CP	selumetinib	XC.P026	24H	1.11uM		MEK inhibitor	False	246	248	32	0.30210	9.97718e-01	1.e+00
16	RummaGEO Chem	oridonin					BCL inhibitor	False	835	906	49	0.40828	9.99920e-01	1.e+00
17	RummaGEO Chem	bortezomib					NFkB pathway inhibitor proteasome inhibitor	True	10	82	13	0.	1.00000e+00	1.e+00

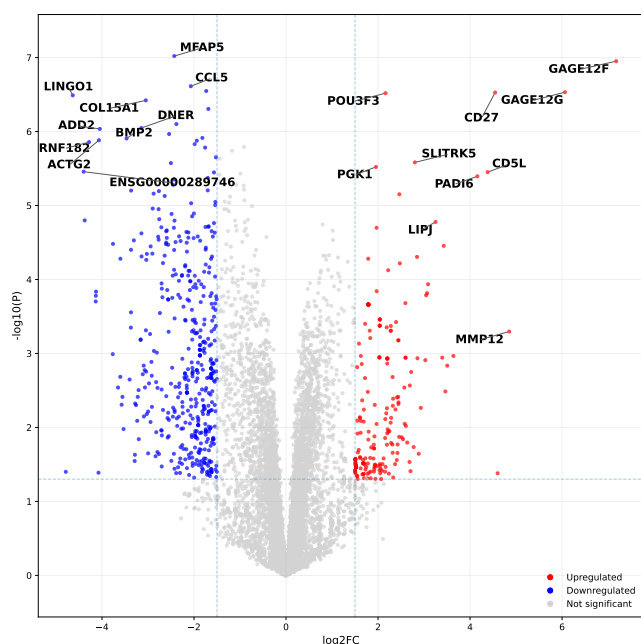


Figure 3. Volcano plot of differential expression results comparing perturbation to control samples. Each point represents one protein-coding gene; the x-axis shows the log2 fold-change and the y-axis shows statistical significance as $-\log_{10}(\text{adjusted } p\text{-value})$. Genes passing both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

tency (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 65 significantly up-regulated genes and 234 significantly down-regulated genes after logFC and adjusted p-value filtering (Figure 3). Among the most strongly up-regulated genes were GAGE12F, GAGE12G, CD27, POU3F3, and SLITRK5, whereas prominent down-regulated genes included MFAP5, CCL5, AMZ1, LINGO1, and COL15A1. Functional enrichment analysis of the up-regulated gene set identified associations with GO Biological Process terms related to antigen processing and presentation of endogenous peptide antigen via mhc class i and peptide transport [18], KEGG pathways involving primary immunodeficiency and histidine metabolism [19], ChEA transcription factors including FOXM1 and TP63 [20], and KOMP2 mouse phenotypes associated with abnormal epidermal immune compartment and increased thermal nociceptive threshold [21] (Figure 4). Analysis of the

down-regulated gene set identified enrichment for GO Biological Process terms related to endothelin receptor signaling pathway and negative regulation of coagulation [18], KEGG pathways involving thiamine metabolism and riboflavin metabolism [19], ChEA transcription factors including VDR and BCL11B [20], and KOMP2 mouse phenotypes associated with increased klrg1-positive nk cell number and abnormal salivary gland morphology [21] (Figure 5). Signature search revealed gene perturbations including ZCCHC4 and NEK2 from Perturb Atlas Human [23], Tet2 from Perturb Atlas Mouse [23], and rnf2 and mxra8 from RummaGEO Gene [24] (Table 1) and drug perturbations including selumetinib and tozasertib from LINCS L1000 CP [25] and bortezomib and talazoparib from RummaGEO Chem [24] (Table 2) as mimickers. Top reversers included genes perturbations POU5F1 and Sox2 from CREEDS Gene [26], WRAP53 and NFYB from Perturb Atlas Human [23], Ldha and p16 from Perturb Atlas Mouse [23], and cnot1 and larp1 from RummaGEO Gene [24] (Table 3) and drugs including tozasertib and selumetinib from LINCS L1000 CP [25] and dextran and olaparib from RummaGEO Chem [24] (Table 4).

4. Discussion

The upregulated gene signature is dominated by immunerelated pathways, as evidenced by the simultaneous enrichment of antigen processing and presentation (GO [18], KEGG [19]) and by transcription factors known to drive immune activation, such as CIITA, IRF1, IRF8 and FOXM1 (ChEA [20]). This immune theme is reinforced in the KOMP2 phenotypic terms, which include abnormal epidermal immune compartment and abnormal sinus arrhythmia (KOMP2 [21]), further pointing to altered skin and cardiovascular immunity. Additionally, several upregulated signatures implicate vascular remodeling (venous blood vessel morphogenesis (GO [18]) and (KOMP2 [21]) and metabolic transport (Lornithine transmembrane transport (GO [18]) and ABC transporters (KEGG [19])). The coordinated response suggests a link between immunity and nutrient handling. In contrast, the downregulated set highlights pathways that govern vascular tone and blood clotting, with endothelin receptor signaling and negative regulation of coagulation (GO [18]) both enriched in GO, and complement and coagulation cascades (KEGG [19]) in KEGG. The same anticoagulatory theme is echoed by the presence of the VDR transcription factor (ChEA [20]) and vitamin D receptor signaling pathway (GO [18]), linking calcium and lipid metabolism to thrombosis control. Moreover, the downregulated profile emphasizes struc-

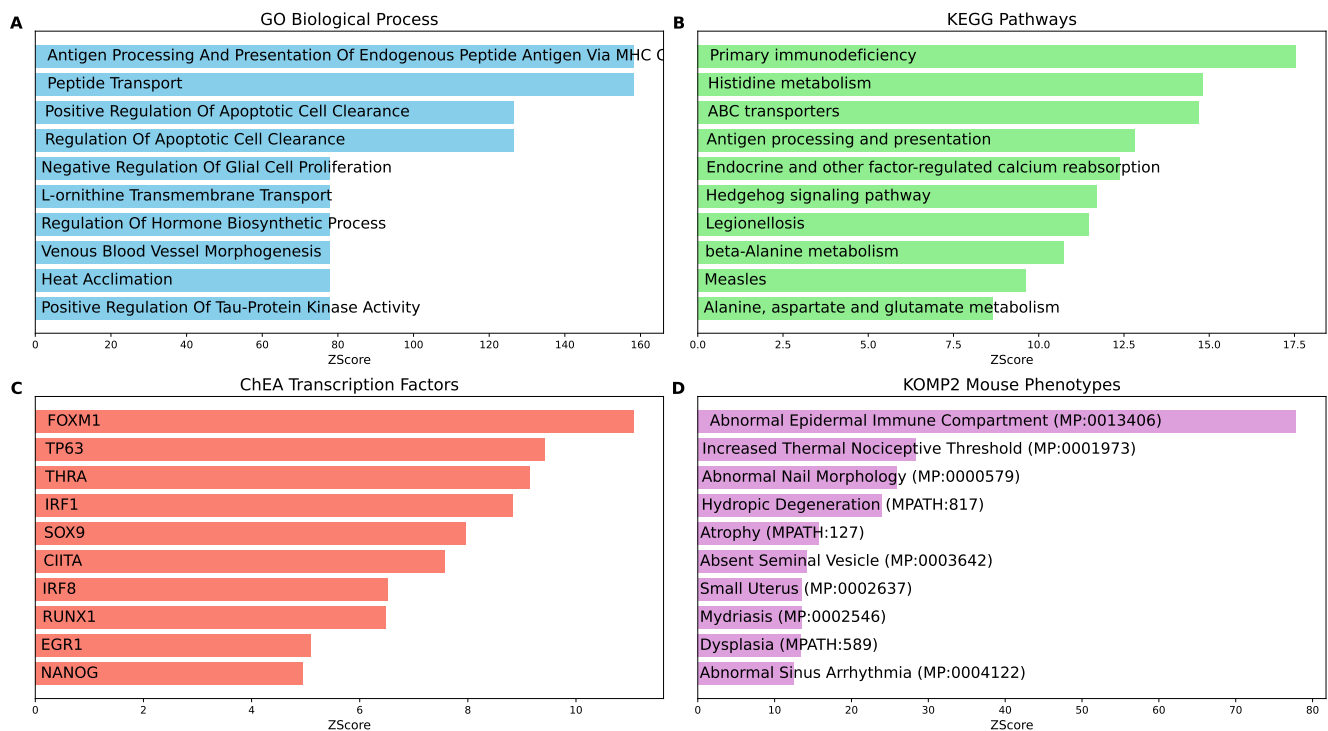


Figure 4. Enrichment analysis of the up-regulated gene signature using Enrichr [17]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [18], B. KEGG 2021 Human [19], C. ChEA 2022 [20], and D. KOMP2 Mouse Phenotypes 2022 [21]. Bars are ranked by 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](#).

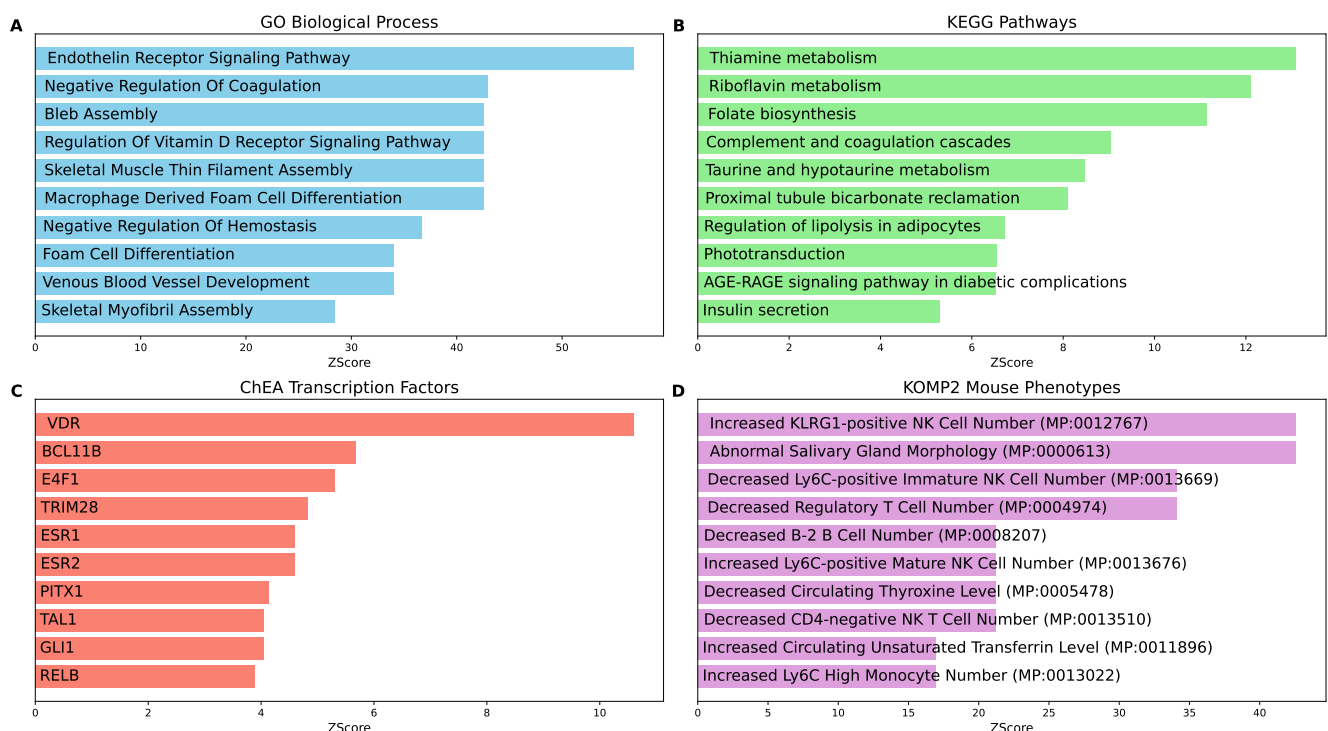


Figure 5. Enrichment analysis of the down-regulated gene signature using Enrichr [17]. Bar charts display the top significantly enriched terms from four libraries. A. GO Biological Process 2023 [18], B. KEGG 2021 Human [19], C. ChEA 2022 [20], and D. KOMP2 Mouse Phenotypes 2022 [21]. Bars are ranked 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](#).

tural and metabolic processes such as skeletal muscle thinfilament assembly (GO [18]), phototransduction (KEGG [19]) and folate biosynthesis (KEGG [19]), indicating a loss of contractile and metabolic function. Together, the contrasting enrichment patterns reveal that upregulated genes are skewed toward immune activation and vascular remodeling, whereas downregulated genes are more focused on hemostasis, vascular tone and tissue structural integrity. The most striking pattern emerges in the drug signatures, where the strongest mimicker hits are a proteasome inhibitor, a MEK inhibitor, and a PARP inhibitor, all of which are known to modulate cellular stress and DNA repair pathways, which aligns with the enrichment of DNA repair related gene perturbations in the reverse gene list. The proteasome inhibitor bortezomib (RummaGEO Chem) [24] appears at the top of the mimicker drug list with a highly significant pvalue, suggesting that suppression of protein degradation recapitulates the signature, while the MEK inhibitor selumetinib (LINCS L1000 CP) [25] and the PARP inhibitor talazoparib (RummaGEO Chem) [24] also achieve highly significant pvalues, indicating that blockade of MAPK signaling and DNA repair also drives the same transcriptional state; tozasertib (LINCS L1000 CP) [25] provides additional support for inhibition of cell cycle checkpoints. In the reverse drug list, the top hit dextran (RummaGEO Chem) [24] is biologically inert but appears with a very low pvalue, and valproic acid (RummaGEO Chem) [24] and trifluoperazine (RummaGEO Chem) [24] appear at slightly lower ranks, pointing to shared effects of histone deacetylase inhibition and dopamine receptor antagonism in counteracting the signature. The mimicker gene list is dominated by Perturb Atlas Human perturbations of genes involved in cell cycle regulation and chromatin remodeling, such as ZCCHC4 (Perturb Atlas Human) [23], NEK2 (Perturb Atlas Human) [23], HMGA1 (Perturb Atlas Human) [23], and KLF2 (Perturb Atlas Human) [23], while RummaGEO Gene perturbations of *rnf2* (RummaGEO Gene) [24] and *mxra8* (RummaGEO Gene) [24] reflect ubiquitination and extracellular matrix processes that are also enriched in the signature. The reverse gene list contains a mix of transcription factors and RNA processing factors, including *cnot1* (RummaGEO Gene) [24], *WRAP53* (Perturb Atlas Human) [23], *POU5F1* (CREEDS Gene) [26], *NFYB* (Perturb Atlas Human) [23], *ZNF281* (Perturb Atlas Human) [23], and *EFTUD2* (Perturb Atlas Human) [23], highlighting a potential requirement for transcriptional reprogramming and splicing to reverse the signature. The convergence of proteasome, MAPK, and PARP inhibitors in the mimicker set and the transcription factor and splicing factor perturbations

in the reverse set together suggest that the signature represents a stress responsive, proliferative state that can be driven by inhibition of protein degradation, signaling, or DNA repair and reversed by modulation of transcription and RNA processing.

The reanalysis confirms that HMGA1 knockdown in HUCCT1 cholangiocarcinoma cells drives a pronounced shift toward immune activation and vascular remodeling, as the upregulated gene signature is enriched for antigen processing and presentation, cytokine driven transcription factors, and venous vessel morphogenesis. This immune theme is accompanied by changes in metabolic transport and nutrient handling pathways, while the downregulated set highlights loss of vascular tone regulation, coagulation control, and structural muscle and metabolic functions. In parallel, PerturbSeqr identifies proteasome, MEK, and PARP inhibitors as drugs that recapitulate the knockdown signature, underscoring a stress responsive, proliferative state that involves impaired protein degradation, MAPK signaling, and DNA repair pathways. Conversely, the reverse drug list points to transcriptional and RNA processing factors that could counteract the signature, suggesting that modulating transcriptional reprogramming and splicing may reverse the effects of HMGA1 loss. The unexpected activation of immune pathways in a cholangiocarcinoma context, together with the loss of coagulation and vascular tone genes, highlights a novel link between HMGA1 suppression and altered tumor microenvironment interactions. These findings directly address the research question of how HMGA1 knockdown rewires cellular signaling and identify potential pharmacologic interventions that mimic or reverse the observed transcriptional state.

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. Piscuoglio, S. *et al.* HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma. *Histopathology* **60**, 397–404. <https://doi.org/10.1111/j.1365-2559.2011.04121.x> (2012).

2. Jarnagin, W. *et al.* Staging, resectability, and outcome in 225 patients with hilar cholangiocarcinoma. *Ann Surg* **234**, 507–17. <https://doi.org/10.1097/0000658-200110000-00010> (2001).
3. Alabraba, E. *et al.* Increased multimodality treatment options has improved survival for Hepatocellular carcinoma but poor survival for biliary tract cancers remains unchanged. *Eur J Surg Oncol* **45**, 1660–7. <https://doi.org/10.1016/j.ejso.2019.04.002> (2019).
4. Hristov, A. *et al.* HMGA1 correlates with advanced tumor grade and decreased survival in pancreatic ductal adenocarcinoma. *Mod Pathol* **23**, 98–104. <https://doi.org/10.1038/modpathol.2009.139> (2010).
5. Banales, J. *et al.* Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nat Rev Gastroenterol Hepatol* **17**, 557–88. <https://doi.org/10.1038/s41575-020-0310-z> (2020).
6. Banales, J. *et al.* Expert consensus document: cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). *Nat Rev Gastroenterol Hepatol* **13**, 261–80. <https://doi.org/10.1038/nrgastro.2016.51> (2016).
7. Mendez, O. *et al.* Extracellular HMGA1 Promotes tumor invasion and metastasis in triple-negative breast cancer. *Clin Cancer Res* **24**, 6367–82. <https://doi.org/10.1158/1078-0432.CCR-18-0517> (2018).
8. Chiappetta, G. *et al.* HMGA1 protein overexpression in human breast carcinomas: correlation with ErbB2 expression. *Clin Cancer Res* **10**, 7637–44. <https://doi.org/10.1158/1078-0432.CCR-04-0291> (2004).
9. Quintavalle, C. *et al.* High mobility group A1 enhances tumorigenicity of human cholangiocarcinoma and confers resistance to therapy. *Mol Carcinog* **56**, 2146–57. <https://doi.org/10.1002/mc.22671> (2017).
10. Wang M, Chen Z, Guo P, Wang Y, Chen G. Therapy for advanced cholangiocarcinoma: current knowledge and future potential. *J Cell Mol Med*. 2021;25:618–28. <https://doi.org/10.1111/jcmm.16151>.
11. Pellarin, I. *et al.* The architectural chromatin factor high mobility group A1 enhances DNA ligase IV activity influencing DNA repair. *PLoS ONE* **11**, e0164258–. <https://doi.org/10.1371/journal.pone.0164258> (2016).
12. Song, J. *et al.* Overexpression of HMGA1 confers radioresistance by transactivating RAD51 in cholangiocarcinoma. *Cell Death Discov* **7**. <https://doi.org/10.1038/s41420-021-00721-8> (2021).
13. 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).
14. 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).
15. 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).
16. 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).
17. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90. <https://doi.org/10.1002/cpz1.90> (2021).
18. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29. <https://doi.org/10.1038/75556> (2000).
19. 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).
20. 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).
21. 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).
22. Perturb-Seqr <https://perturbseqr.maayanlab.cloud>.
23. 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).
24. 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).
25. Evangelista, J. E. *et al.* SigCom LINCS: data and metadata search engine for a million gene expression signatures. *Nucleic Acids Research* **50**, W697–W709. <https://doi.org/10.1093/nar/gkac328> (2022).
26. 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).