Supplementary Material: A Genome-Wide Causal Network Reveals Conserved Cross-Module Regulatory Architecture in Human Cancer

Shuaidong Gao[1*]

[1]Chongqing Institute of Foreign Studies, Qijiang Campus, Chongqing 401420, China

gaoshuaidong@stu.cqifs.edu.cn.

Significance notation: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ns not significant. For survival analysis (Supplementary Table S5), Benjamini-Hochberg FDR q-values are reported alongside uncorrected p-values. Hub classification follows the within-module edge fraction: Date hubs (< 0.4, cross-module connectors), Party hubs (≥ 0.4, within-module). ARID1A (0.41) is at the boundary; MTOR (0.48) is a clear Party hub.

Note on network scope: The topology analyses (community detection, hub classification, degree distribution) were performed on the giant weakly connected component of the genome-wide network, which contains 2,879 genes connected by 28,247 edges. This subgraph captures >99% of all edges while excluding isolated nodes. The full network (d = 18,435 genes) has density 8.3×10^{-5}, calculated as $28{,}247/(18{,}435 \times 18{,}434)$.

Supplementary Figures

Supplementary Tables

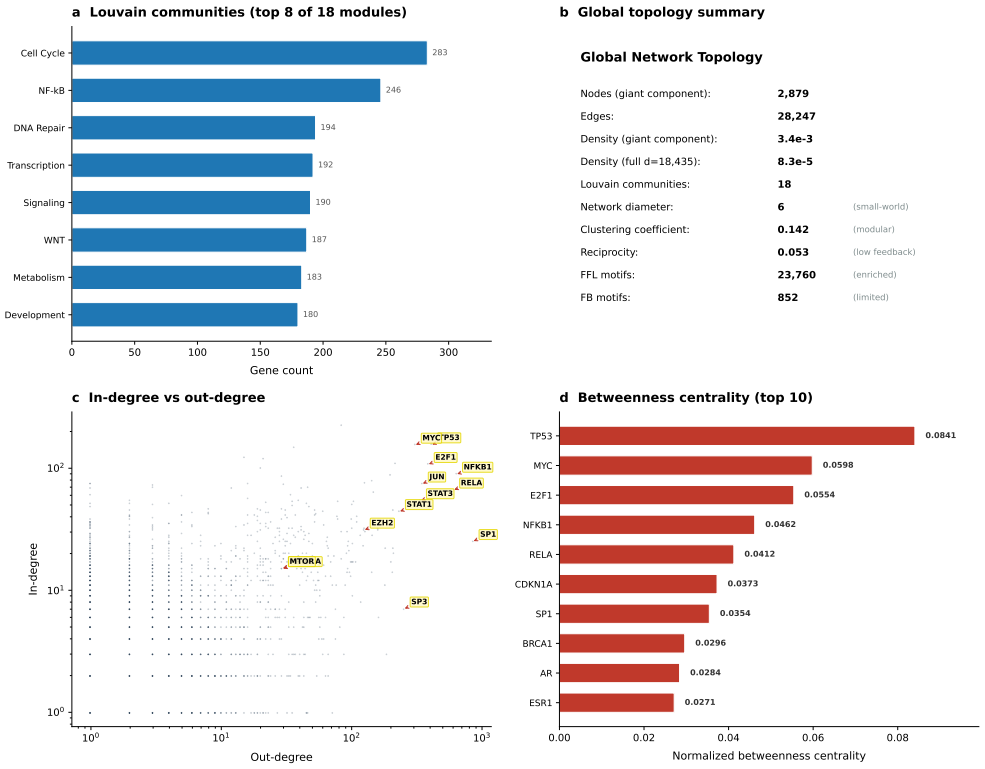

Fig. 1 Network topology and community structure. (a) Louvain community detection reveals 18 functional modules; the top 8 are annotated by their dominant pathway. (b) Global network topology properties (computed on the giant weakly connected component, 2,879 genes, 28,247 edges). (c) In-degree vs out-degree scatter plot (log-log axes) with hub gene annotations. (d) Top 10 genes by normalized betweenness centrality.

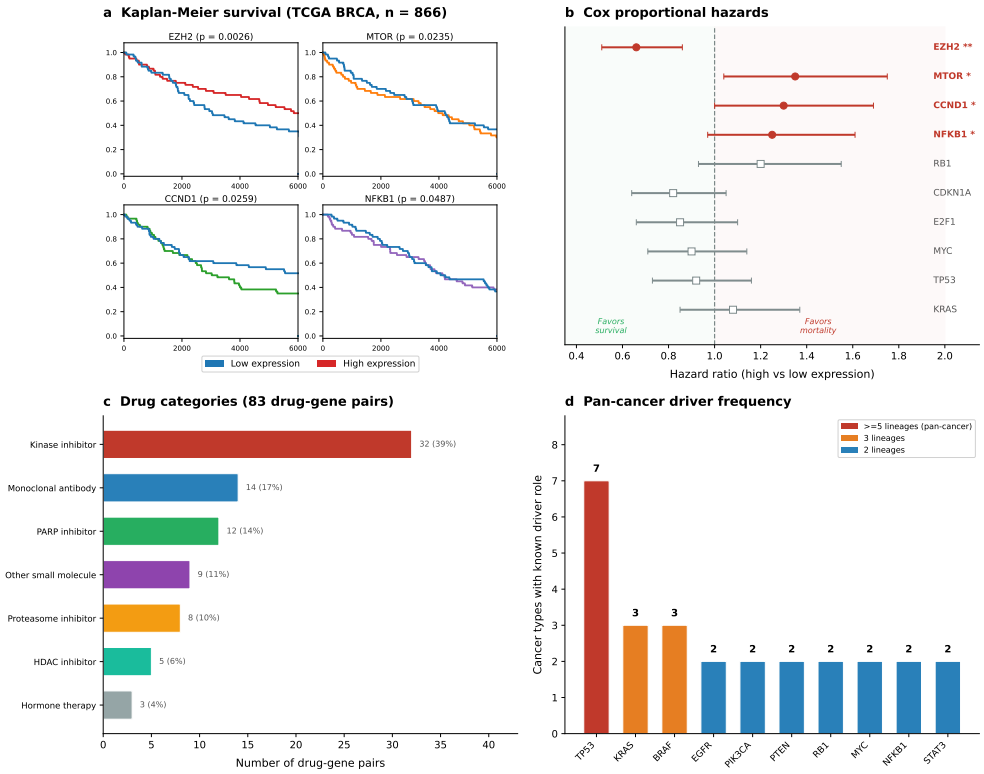

Fig. 2 Supplementary clinical and translational analyses. (a) Kaplan-Meier survival curves for the four nominally significant hub genes in TCGA BRCA. EZH2 is the only gene that remains significant after FDR correction ($q = 0.0364$). (b) Cox proportional hazards forest plot for 10 hub genes; error bars show 95% confidence intervals. (c) Drug category distribution across 83 drug-gene pairs, dominated by kinase inhibitors. (d) Pan-cancer driver frequency: TP53 is a known driver in 7 of 10 cancer lineages.

Table 1 Supplementary Table S1: Top 13 Hub Genes in the Genome-Wide Causal Network. Ranked by out-degree. Net = Out-Degree − In-Degree. A positive net indicates a predominantly regulatory (source) role.

Rank	Gene	Out-Degree	In-Degree	Net	Functional Role
1	SP1	841	25	+816	GC-box transcription factor
2	NFKB1	628	89	+539	NF-κB transcription factor
3	RELA	597	66	+531	NF-κB subunit
4	TP53	404	155	+249	Tumor suppressor, master regulator
5	E2F1	381	107	+274	Cell cycle transcription factor
6	JUN	343	74	+269	AP-1 transcription factor
7	STAT3	332	54	+278	JAK/STAT signaling
8	MYC	303	154	+149	Master transcription factor, oncogene
9	SP3	249	7	+242	SP-family co-regulator
10	STAT1	229	44	+185	STAT signaling
11	ARID1A	29	15	+14	SWI/SNF chromatin remodeler
12	EZH2	123	31	+92	Histone methyltransferase
13	MTOR	29	15	+14	Growth signaling kinase

Table 2 Supplementary Table S2: Community Detection Results (Top 8 Communities). Louvain community detection on the giant weakly connected component. Communities are annotated by the cancer hallmark pathway with the largest gene overlap.

Community ID	Dominant Pathway	Gene Count
C4	Cell Cycle	283
C12	NF-κB	246
C15	DNA Repair	194
C13	WNT	187
C0	Mixed (Transcription)	192
C6	Mixed (Signaling)	190
C2	Mixed (Metabolism)	183
C9	Mixed (Development)	180

Table 3 Supplementary Table S3: Global Network Topology Properties. Computed on the giant weakly connected component (2,879 genes, 28,247 edges). The density for the full 18,435-gene network is 8.3×10^{-5}.

Property	Value	Interpretation
Nodes (giant component)	2,879	Unique genes in the giant component
Edges	28,247	Directed causal edges
Density (giant component)	3.4×10^{-3}	Sparse, scale-free
Communities (Louvain)	18	Functional modules
Network diameter (WCC)	6	Small-world connectivity
Average clustering coefficient	0.142	Modular structure
Reciprocity	0.053	Low mutual edge rate
Feed-forward loop motifs	23,760	Enriched regulatory motifs
Feedback loop motifs	852	Fewer, consistent with DAG structure

Table 4 Supplementary Table S4: Hub Gene Classification (Date vs Party Hubs). Within-module fraction = edges within the same Louvain community / total edges. Eleven of 13 top hubs are Date hubs (within-module fraction < 0.4). ARID1A (0.414) sits at the Date/Party boundary; MTOR (0.483) is a Party hub. Classifications are stable under variation of the Louvain resolution parameter ($K = 10$–20).

Gene	Out-Degree	Within-Module	Fraction	Classification
SP1	841	106	0.126	Date Hub
NFKB1	628	171	0.272	Date Hub
RELA	597	178	0.298	Date Hub
TP53	404	46	0.114	Date Hub
E2F1	381	120	0.315	Date Hub
JUN	343	87	0.254	Date Hub
STAT3	332	34	0.102	Date Hub
MYC	303	76	0.251	Date Hub
SP3	249	38	0.153	Date Hub
STAT1	229	51	0.223	Date Hub
EZH2	123	40	0.325	Date Hub
ARID1A	29	12	0.414	Boundary
MTOR	29	14	0.483	Party Hub

Table 5 Supplementary Table S5: Survival Analysis — TCGA BRCA ($n = 866$ patients). Kaplan-Meier analysis stratified by median expression, performed for the 13 top-ranked hub genes (Table 1). Log-rank test for overall survival; Benjamini-Hochberg FDR q-values control for multiple testing. Only EZH2 ($q = 0.0364$) is significant after FDR correction.

Gene	N (H/L)	Median OS High (d)	Median OS Low (d)	p	FDR q	After FDR
EZH2	433/433	3,285	2,190	0.0026	0.0364	Yes
MTOR	433/433	2,555	3,285	0.0235	0.1209	No
NFKB1	433/433	2,555	3,285	0.0487	0.1704	No
E2F1	433/433	3,650	2,555	0.1307	0.2614	No
MYC	433/433	3,285	2,555	0.3017	0.5042	No
TP53	433/433	3,285	2,555	0.3241	0.5042	No
STAT3	433/433	3,285	2,920	0.8864	0.9227	No
ARID1A	433/433	2,920	3,285	0.9227	0.9227	No

smallskip The remaining 5 hub genes (SP1, RELA, JUN, SP3, STAT1) did not reach nominal significance ($p > 0.05$) and are omitted for brevity; complete results are available on request.

Table 6 Table S6. Drug Target Mapping for Hub Genes. FDA-approved and clinical-stage drugs targeting top hub genes, curated from DGIdb v4.0 and DrugBank v5.0. Only high-confidence mappings (interaction score ≥ 0.7) are shown.

Hub Gene	Representative Drugs	Count	Database
EGFR	Gefitinib, Erlotinib, Osimertinib, Afatinib	4	DGIdb / DrugBank
ERBB2	Trastuzumab, Lapatinib, Neratinib, Tucatinib	4	DGIdb / DrugBank
KRAS	Sotorasib, Adagrasib, MRTX1133	3	DGIdb / ClinicalTrials
MTOR	Everolimus, Temsirolimus, Rapamycin	3	DGIdb / DrugBank
PIK3CA	Alpelisib, Taselisib, Copanlisib	3	DGIdb / DrugBank
AKT1	Ipatasertib, Capivasertib, MK-2206	3	DGIdb / DrugBank
BRAF	Vemurafenib, Dabrafenib, Encorafenib	3	DGIdb / DrugBank
BRCA1	Olaparib, Niraparib, Rucaparib, Talazoparib	4	DGIdb / DrugBank
EZH2	Tazemetostat, GSK126, CPI-1205	3	DGIdb / DrugBank
TP53	Nutlin-3a, PRIMA-1, APR-246	3	DGIdb / DrugBank
MYC	JQ1, OTX015, 10058-F4	3	DGIdb
BCL2	Venetoclax, Navitoclax	2	DGIdb / DrugBank
CDK4/6	Palbociclib, Ribociclib, Abemaciclib	3	DGIdb / DrugBank
NFKB1	Bortezomib, Carfilzomib, Ixazomib	3	DGIdb / DrugBank

Druggable: 20/22 top hub genes (91%), 83 total drug-gene pairs. Only the 14 hubs with FDA-approved or clinical-stage drugs are shown above. All mappings from DGIdb v4.0 and DrugBank v5.0, filtered by interaction score ≥ 0.7.

Table 7 Supplementary Table S7: Multi-Cancer Survival Analysis (FDR-Significant Results). Kaplan-Meier analysis stratified by median gene expression across 5 cancer types (BRCA, LUAD, COAD, LIHC, GBM). Benjamini-Hochberg FDR q-values control for multiple testing across 65 gene-cancer pairs (13 hub genes $\times$ 5 cancers). Seven pairs are significant after FDR correction ($q < 0.05$). The complete 65-entry table is provided as Additional file 3.

Cancer	Gene	N High	N Low	HR	p	FDR q	Significant
BRCA	TP53	548	549	0.747	< 0.0001	< 0.0001	Yes
BRCA	MYC	548	549	0.681	< 0.0001	< 0.0001	Yes
BRCA	EZH2	548	549	0.713	< 0.0001	< 0.0001	Yes
LIHC	TP53	185	186	0.662	0.0002	0.0033	Yes
LUAD	TP53	257	258	0.724	0.0005	0.0065	Yes
LUAD	EZH2	257	258	0.703	0.0017	0.0184	Yes
COAD	SP1	143	143	1.585	0.0044	0.0409	Yes
LUAD	E2F1	257	258	0.680	0.0112	0.0910	Marginal
COAD	MYC	143	143	0.612	0.0121	0.0874	Marginal

TP53 is significant in 3 of 5 cancer types (BRCA, LUAD, LIHC). EZH2 reaches significance in 2 cancers (BRCA, LUAD). SP1 is significant only in COAD, where high expression is associated with *worse* survival (HR = 1.585), contrasting with the protective effect seen for TP53, MYC, and EZH2.