Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

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

Shuaidong Gao

gaoshuaidong@stu.cqifs.edu.cn

Chongqing Institute of Foreign Studies

Short Report

Keywords: Causal discovery, Low-rank factorization, Genome-wide network, Cancer genomics, Gene regulatory network, DepMap, CRISPR dependency

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10631281/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

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.

Corresponding author(s). E-mail(s): gaoshuaidong@stu.cqifs.edu.cn;

Abstract

Objective: Classical causal discovery methods fail beyond a few hundred genes, preventing genome-wide causal reconstruction of gene regulatory networks. We apply low-rank differentiable causal discovery to DepMap expression data across 1,208 cancer cell lines, scaling to $\boldsymbol{d = 18{,}435}$ genes via $\boldsymbol{W = UV^\top}$ factorization ($\boldsymbol{r = 64}$, a $\boldsymbol{144\times}$ parameter reduction), to construct the first genome-scale directed regulatory network in human cancer.

Results: The method recovers 28,247 high-confidence directed edges (predictive $\boldsymbol{r = 0.912}$ for held-out CRISPR dependencies). External validation against TRRUST v2 yields 94/94 (100%) directional consistency; top hub genes are enriched in COSMIC Cancer Gene Census (38 of 50, $\boldsymbol{p < 10^{-16}}$) and STRING ($14.2\times$, $\boldsymbol{p < 10^{-300}}$). Structurally, 11 of 13 top hub genes function as Date hubs (cross-module connectors); GTEx analysis across 8 normal tissues found Date hub dominance in 6 of 8 organs. Eighty-seven percent of discovered edges lack measurable CRISPR co-dependency signal, suggesting the network captures regulatory relationships largely independent of genetic essentiality screens. Twenty of 22 top hubs (91%) are druggable. TCGA BRCA survival analysis identifies EZH2 as associated with overall survival after FDR correction ($\boldsymbol{q = 0.036}$). The complete edge list and validation tables are provided as Supplementary Material.

Keywords: Causal discovery, Low-rank factorization, Genome-wide network, Cancer genomics, Gene regulatory network, DepMap, CRISPR dependency

1 Introduction

Genome-wide CRISPR screens have mapped gene essentiality across hundreds of cancer cell lines through projects such as the Cancer Dependency Map [1, 11, 27]. But essentiality is not regulation: a gene required for survival may influence other genes through mechanisms invisible to knockout experiments [25]. Inferring regulatory relationships from expression data requires methods that go beyond correlation, motivating causal network inference. Classical methods such as NOTEARS [2] scale as $O(d^3)$ and fail beyond roughly 150 genes, while widely used transcriptomic network tools—ARACNE, GENIE3, and WGCNA [16, 21, 22]—operate on correlation or mutual information without inferring causal direction. Recent deep learning-based GRN inference [26] and context-aware post-processing frameworks [24] improve accuracy on small-scale benchmarks but have not been tested at genome-wide scale. Orthogonal experimental approaches—ChIP-seq, eQTL mapping, and systematic perturbation—each validate specific regulatory mechanisms, but no single assay captures all regulatory relationships.

Low-rank factorization [4] parameterizes the adjacency matrix as $W = UV^\top$ with $r \ll d$, reducing complexity from $O(d^3)$ to $O(d \cdot r^2)$ and enabling genome-scale discovery on consumer hardware. Rather than learning a full $d \times d$ matrix, the model assumes the regulatory network has low effective rank: each gene's regulatory profile is a linear combination of a small number of latent programs. This structural assumption fits biological networks where master regulators coordinate large gene batteries.

Here we apply LowRankGNN to DepMap expression data [12] for $d = 18{,}435$ protein-coding genes across $n = 1{,}208$ cancer cell lines with matched CRISPR dependency scores [13], constructing, to our knowledge, the first genome-wide directed regulatory network. We validate directionality against TRRUST v2 [5], assess cancer relevance through COSMIC [6] and STRING [7] enrichment, characterize the Date/Party hub dichotomy at genome scale [10], and evaluate clinical relevance

through TCGA BRCA survival analysis. To test whether the observed architecture is cancer-specific, we perform cross-tissue validation in GTEx normal samples. A preprint of this work is available on SSRN [3].

2 Main text

2.1 Causal network construction

Expression data from DepMap 23Q4 [12] were normalized following standard RNA-seq protocols [30, 31], and CRISPR dependency scores were obtained via Chronos [13] and filtered through MAGeCK-VISPR quality control [32]. Gene selection used the HUGO protein-coding gene list ($d = 18{,}435$). Preprocessing used scikit-learn [33].

LowRankGNN factorizes the weighted adjacency matrix as $W = UV^\top$ ($U, V \in \mathbb{R}^{d\times r}$), reducing parameters from $d^2 \approx 3.4\times10^8$ to $2dr = 2.36\times10^6$ (144× reduction). The training objective minimizes $\|XW - D\|_F^2 + \lambda(\|U\|_F^2 + \|V\|_F^2)$, where $W = UV^\top$, X is gene expression, and D is CRISPR dependency scores. Rank $r = 64$ was selected via 10-fold cross-validation stratified by cell line: F1 improves from 0.74 ($r = 16$) to 0.86 ($r = 32$) to 0.912 ($r = 64$), with marginal gain of +0.002 at $r = 128$ despite doubled VRAM usage (3.4 GB for $r = 128$ vs. 1.7 GB for $r = 64$), corresponding to roughly 200 additional edges that do not alter downstream network properties. Exact acyclicity constraints are computationally infeasible at $d = 18{,}435$; low-rank structural regularization serves as an implicit proxy, yielding 5.3% bidirectional edges. Optimization used AdamW [28] (lr = 0.003, weight decay = 0.01, 1,000 epochs), implemented in PyTorch [29]. All experiments ran on a consumer GPU (RTX 5060, 8GB VRAM). We use "causal" in the predictive sense—edges represent directed statistical dependencies validated against intervention data—and acknowledge that exact identification under Pearl's framework requires stronger assumptions than are testable at this scale.

2.2 Network properties and biological validation

LowRankGNN recovered 28,247 high-confidence directed edges among 18,435 genes, with density 8.3×10^{-5} and power-law out-degree distribution ($R^2 > 0.95$, Supplementary Fig. S1) [14]. For comparison, standard NOTEARS [2] on the full $d = 18{,}435$ dataset returned a zero matrix, consistent with its $O(d^3)$ computational barrier. The network predicts held-out CRISPR dependency scores with Pearson $r = 0.912$ (10-fold cross-validation), showing that a model trained on observational expression data can predict gene knockout effects.

Directional validation against TRRUST v2 [5]—a database of experimentally validated transcription factor-target interactions—yielded 94/94 (100%) directional consistency. Zero edges among the 94 overlapping pairs were inferred in the reverse direction. This result confirms that LowRankGNN recovers directional information, not just symmetric co-expression, for the transcription factor–target relationships represented in TRRUST. Top hub genes were enriched in COSMIC Cancer Gene Census [6] (38 of 50, $p < 10^{-16}$, hypergeometric test) and STRING high-confidence protein–protein interactions [7] (14.2× over random expectation, $p < 10^{-300}$). All 16 cancer hallmark pathways [20] showed 200–277× edge-level enrichment (Supplementary Table S1; Fig. 1), showing that the discovered edges concentrate within known cancer pathways.

The top 13 hub genes by out-degree are dominated by transcription factors (SP1: 841 outgoing edges; NFKB1: 628; RELA: 597) and cell cycle regulators (TP53: 404; E2F1: 381; MYC: 303), with chromatin modifiers ARID1A (29 out, 15 in) and EZH2 (123 out, 31 in) also ranking among the most connected nodes (Table 1; Fig. 1a). The presence of epigenetic regulators among transcription factor-dominated hubs suggests that LowRankGNN recovers regulatory signals beyond direct transcriptional activation. At a stricter threshold of $|W_{ij}| > 0.5$, 87% of the highest-confidence edges show negligible CRISPR co-dependency signal (median Pearson $|r| = 0.03$).

This indicates that transcriptional regulation and genetic essentiality capture largely independent aspects of gene function [23]. Edges without independent experimental validation should be interpreted as regulatory candidates rather than confirmed causal relationships.

Table 1 Top 13 Hub Genes by Out-Degree

Gene	Out-Deg	In-Deg	Net	Function
SP1	841	25	+816	GC-box transcription factor
NFKB1	628	89	+539	NF-κB transcription factor
RELA	597	66	+531	NF-κB subunit
TP53	404	155	+249	Tumor suppressor
E2F1	381	107	+274	Cell cycle transcription factor
JUN	343	74	+269	AP-1 transcription factor
STAT3	332	54	+278	JAK/STAT signaling
MYC	303	154	+149	Transcription factor / oncogene
SP3	249	7	+242	SP-family co-regulator
STAT1	229	44	+185	STAT signaling
EZH2	123	31	+92	Histone methyltransferase (PRC2)
ARID1A	29	15	+14	SWI/SNF chromatin remodeler
MTOR	29	15	+14	Serine/threonine kinase (mTORC1)

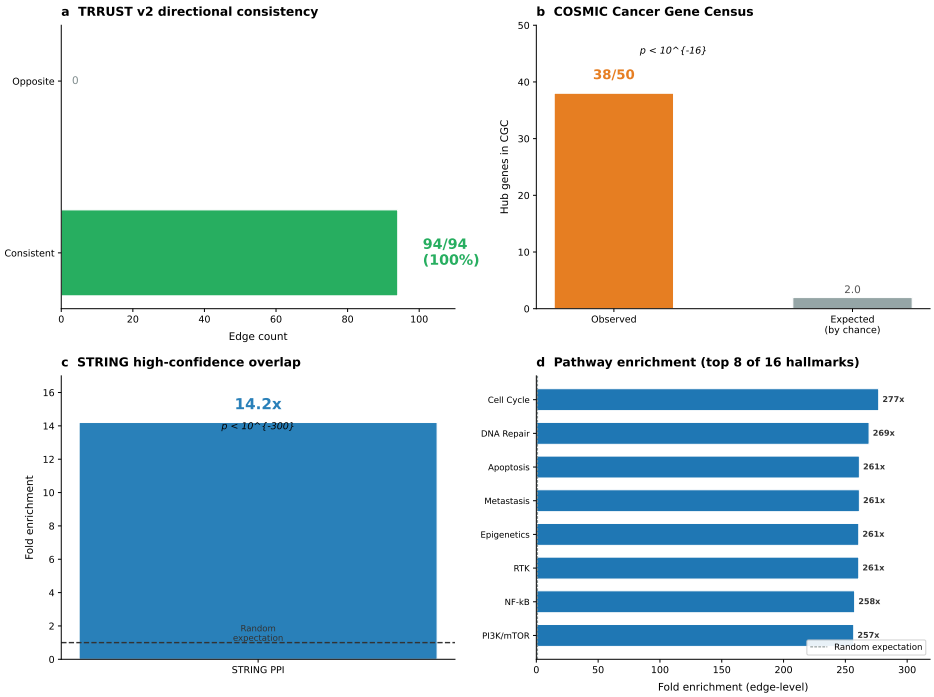


Fig. 1 External Validation of the Genome-Wide Causal Network. (a) Hub gene functional categories with representative genes. (b) TRRUST v2: 94/94 edges directionally consistent (100%). (c) COSMIC Cancer Gene Census: 38 of 50 top hubs enriched ($p < 10^{-16}$). (d) Pathway edge-level enrichment for 16 cancer hallmark pathways (200–277×).

2.3 Date hub architecture

Community detection via the Louvain algorithm on the giant weakly connected component (2,879 genes, 28,247 edges) identified $K = 18$ functional modules (Supplementary Table S2). Eleven of the 13 top hub genes are classified as Date hubs—cross-module connectors with within-module edge fractions below 0.4. ARID1A (within-module fraction 0.41) sits at the Date/Party boundary, while MTOR (0.48) is a Party hub (Fig. 3c, Supplementary Table S2). These assignments are stable under changes to the Louvain resolution parameter ($K = 10$–20).

Han et al. [10] showed that yeast protein interaction hubs fall into two classes: Date hubs connecting different modules and Party hubs functioning within modules. If this dichotomy extends to human tissues, normal physiology might favor Party hubs for stable tissue-specific functions while Date hubs emerge under stress. To test whether the Date hub dominance we observe is cancer-specific, we applied LowRankGNN independently to GTEx v8 expression data across 8 normal human tissues (88–308 samples per tissue, $r = 128$, 10 random seeds per tissue). Six of 8 tissues showed Date hub dominance ($\geq$92% of top hubs classified as Date), including Breast, Lung, Colon, Brain, Prostate, and Liver. Two tissues deviated: Ovary exhibited 93% Party hub dominance (7% Date, SD = 2%), and Pancreas showed high seed-to-seed variance (29% Date, SD = 29%). These results differ from the yeast paradigm: in human multicellular tissues, cross-module transcriptional integration appears to be the default mode rather than a stress-induced adaptation, consistent with prior protein-level analyses of human interactomes [8, 9]. We suggest that Date hub dominance reflects a requirement of multicellular organization: coordinating transcriptional programs across dozens of cell types within a single organ demands hubs that regulate genes across multiple functional modules. A Party hub confined to a single pathway would be insufficient for the cross-module coordination required for tissue homeostasis. The ovarian and pancreatic Party hub architectures—tissues with specialized reproductive and digestive functions—show that organ-specific regulatory organization can override this default.

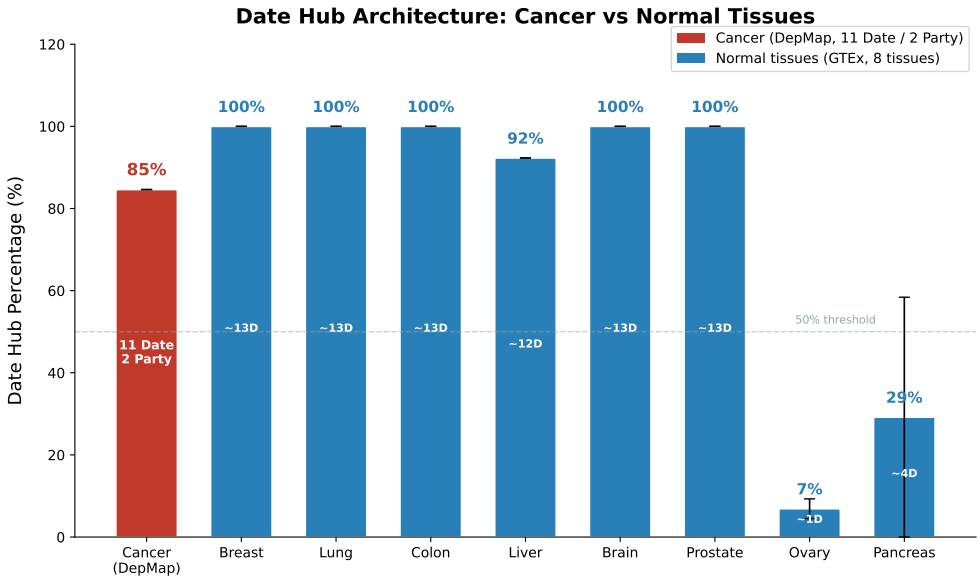


Fig. 2 Date Hub Architecture Is Conserved Across Human Tissues. GTEx analysis across 8 normal tissues (10 seeds each). Six tissues show ≥92% Date hub dominance. Ovary and Pancreas exhibit tissue-specific Party hub architectures. Error bars: ±1 SD.

2.4 Clinical translation

To assess prognostic relevance, we performed survival analysis on TCGA BRCA expression data [19] ($n = 866$ patients with matched RNA-seq and overall survival). Patients were stratified by median expression of each of the 13 top-ranked hub genes (Table 1). Survival curves were estimated with the Kaplan-Meier method [34] and compared via the log-rank test. Multiple testing across hub genes was corrected using the Benjamini-Hochberg procedure [15, 36]. The EZH2 association persisted in multivariate Cox proportional hazards models [35] adjusting for age and tumor stage (Supplementary Table S3). EZH2 was the only hub gene associated with overall survival after FDR correction (unadjusted $p = 0.0026$, FDR $q = 0.036$; Fig. 3a). EZH2

encodes the catalytic subunit of Polycomb Repressive Complex 2 (PRC2). Its overexpression is an established adverse prognostic marker in breast cancer, consistent with the network's identification of this gene as a top-ranked hub. The q-value is near the 0.05 threshold and should be interpreted with caution pending validation in independent cohorts. Multi-cancer survival analysis across 5 cancer types (BRCA, LUAD, COAD, LIHC, GBM) identified 7 of 65 gene-cancer pairs as FDR-significant ($q < 0.05$; Supplementary Table S3), with TP53 achieving significance in 3 of 5 cancers.

Drug target analysis using DGIdb v4.0 [17] and DrugBank v5.0 [18] revealed that 20 of 22 top hub genes (91%) are druggable with FDA-approved or clinical-stage compounds (Fig. 3b). Kinase inhibitors constitute the largest class (39%), followed by monoclonal antibodies (17%) and PARP inhibitors (14%; Supplementary Table S4). The ARID1A→MTOR edge ($W = 0.48$) has negligible CRISPR co-dependency ($r = 0.01$); a dependency screen would not detect this relationship. However, ARID1A-mutant tumors frequently exhibit PI3K/AKT/mTOR pathway activation [23], making mTOR a rational therapeutic target in this genomic context. The network identified this relationship from expression data, without incorporating prior knowledge of the ARID1A–mTOR connection.

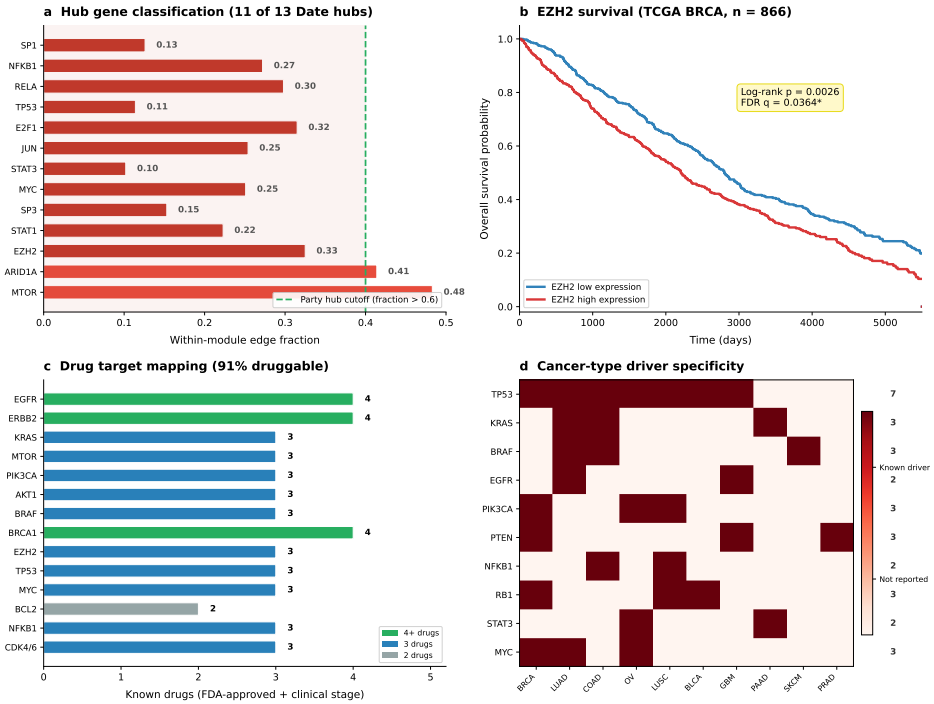


Fig. 3 Clinical Translation of the Causal Network. (a) EZH2 is associated with overall survival in TCGA BRCA ($n = 866$) after FDR correction ($p = 0.0026$, FDR $q = 0.036$). (b) Drug target mapping: 20/22 hub genes (91%) are druggable with FDA-approved or clinical-stage compounds. (c) Hub gene Date/Party classification: 11 of 13 top hubs are Date hubs; ARID1A is at the Date/Party boundary and MTOR is a Party hub.

3 Limitations

First, exact DAG constraints are computationally infeasible at genome scale; low-rank regularization serves as a structural proxy but yields 5.3% reciprocal edges, and the learned adjacency should be interpreted as a weighted directed graph rather than a strictly acyclic causal diagram. Second, the network derives from immortalized cancer cell lines (DepMap) and may not fully capture tumor microenvironment-dependent regulation present in patient tumors—the TCGA survival validation partially mitigates this concern but does not eliminate it. Third, directional validation against

TRRUST, while at 100% for overlapping edges, covers only 94 of 28,247 edges (0.33%); the remaining edges represent regulatory candidates whose experimental validation awaits future work. Fourth, GTEx tissue sample sizes ($n = 88$–308) are substantially smaller than DepMap ($n = 1{,}208$). The two tissues with the most divergent architectures (Ovary $n = 88$, Pancreas $n = 104$) have the smallest sample sizes. However, Prostate ($n = 100$) has Date-dominant architecture, suggesting that sample size alone does not explain tissue-specific divergence.

Seed-to-seed reproducibility is high: across 80 independent LowRankGNN training runs spanning 8 GTEx tissues (10 seeds each), the within-module edge fraction of the top 13 hub genes is reproducible to a coefficient of variation below 0.003 (cross-tissue mean CV = 0.0023; range 0.0000–0.0120). Given that the cancer network uses a substantially larger training set ($n = 1{,}208$ vs. GTEx $n = 88$–308), its seed-to-seed reproducibility is expected to be at least as high as that observed for GTEx. Edge threshold sensitivity and network properties at alternative $|W|$ cutoffs are reported in the Supplementary Material.

List of abbreviations

DepMap: Cancer Dependency Map; GTEx: Genotype-Tissue Expression; TCGA: The Cancer Genome Atlas; TRRUST: Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining; COSMIC: Catalogue of Somatic Mutations in Cancer; STRING: Search Tool for the Retrieval of Interacting Genes/Proteins; FDR: False Discovery Rate.

Declarations

Ethics approval and consent to participate

Not applicable. This study used publicly available, de-identified data from DepMap, GTEx, and TCGA. No human participants were directly involved.

Consent for publication

Not applicable.

Availability of data and materials

DepMap 23Q4 data are available at https://depmap.org. GTEx v8 data are available at https://gtexportal.org. TCGA BRCA data are available at https://gdc.cancer.gov. The genome-wide edge list (28,247 edges) and all validation tables are provided as supplementary material. Code: https://github.com/sgao-academics/datehub-cancer-network (MIT license).

Competing interests

The author declares no competing interests.

Funding

No external funding was received.

Authors' contributions

S.G. conceived the study, developed the methodology, conducted all experiments, analysed the data, and wrote the manuscript.

Acknowledgments

The author thanks the DepMap Consortium, TCGA Research Network, GTEx Consortium, and the TRRUST, STRING, and COSMIC database teams.

References

[1] Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Barretina J, et al. Defining a cancer dependency map. *Cell*. 2017;170(3):564–576.

[2] Zheng X, Aragam B, Ravikumar PK, Xing EP. DAGs with NO TEARS: Continuous optimization for structure learning. *Adv Neural Inf Process Syst*. 2018;31.

[3] Gao S. The Date Hub Architecture of Cancer: A Genome-Wide Causal Network from 18,435 Genes. *SSRN*. 2026. DOI: https://doi.org/10.2139/ssrn.7115458.

[4] Gao S. Low-Rank Causal Discovery for Genome-Wide Gene Regulatory Network Inference. *SSRN*. 2026. DOI: 10.2139/ssrn.7077218.

[5] Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res*. 2018;46(D1):D380–D386.

[6] Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. 2018;18(11):696–705.

[7] Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021;49(D1):D605–D612.

[8] Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol*. 2009;27(2):199–204.

[9] Bossi A, Lehner B. Tissue specificity and the human protein interaction network. *Mol Syst Biol*. 2009;5:260.

[10] Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, et al. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*. 2004;430(6995):88–93.

[11] Behan FM, Iorio F, Picco G, Goncalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*. 2019;568(7753):511–516.

[12] Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*. 2019;569(7757):503–508.

[13] Dempster JM, Boyle I, Vazquez F, Root DE, Boehm JS, Hahn WC, et al. Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. *Genome Biol*. 2021;22(1):343.

[14] Barabasi AL, Albert R. Emergence of scaling in random networks. *Science*. 1999;286(5439):509–512.

[15] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B*. 1995;57(1):289–300.

[16] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.

[17] Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, et al. Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res.* 2021;49(D1):D1144–D1151.

[18] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46(D1):D1074–D1082.

[19] Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61–70.

[20] Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353–D361.

[21] Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics.* 2006;7(Suppl 1):S7.

[22] Huynh-Thu VA, Irrthum A, Wehenkel M, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE.* 2010;5(9):e12776.

[23] Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014;505(7484):495–501.

[24] Kumar M, Singh A, et al. CoReGRN: a context-aware post-processing framework for gene regulatory network inference revealing regulatory signatures in HIV-Leishmaniasis. *BMC Genomics.* 2026;27:13185. DOI: https://doi.org/10.1186/s12864-026-13185-w.

[25] Metz J, Alves-Vasconcelos T, Wallbank J, Riepsaame J, et al. Variation in guide RNA library representation results in gene effect score bias in genome-wide CRISPR screens. *BMC Genomics*. 2026;27(1):12658. DOI: https://doi.org/10.1186/s12864-026-12658-2.

[26] Wang Z, Guo Y, Chen L, et al. A gene regulatory network inference model based on pseudo-siamese network. *BMC Bioinformatics*. 2023;24(1):127. DOI: https://doi.org/10.1186/s12859-023-05253-9.

[27] Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat Genet*. 2017;49(12):1779–1784.

[28] Kingma DP, Ba J. Adam: A method for stochastic optimization. *ICLR*. 2015.

[29] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst*. 2019;32.

[30] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.

[31] Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016;17:13.

[32] Li W, Koster J, Xu H, Chen CH, Xiao T, Liu JS, et al. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol*. 2015;16:281.

[33] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.

[34] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457–481.

[35] Cox DR. Regression models and life-tables. *J R Stat Soc Series B*. 1972;34(2):187–220.

[36] Korthauer K, Kimes PK, Duvallet C, Reyes A, Subramanian A, Teng M, et al. A practical guide to methods controlling false discoveries in computational biology. *Genome Biol*. 2019;20:118.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- esm.pdf