# Supplementary Material

**Gene coexpression calculation:** In our construction of cancer-specific graphs, we add links to the PPI graph to form a unified one. Here, we describe how data processing is performed.

We download gene expression data from The Cancer Genome Atlas (TCGA) via the Genomic Data Commons (GDC). For each type of cancer, we have gene expression from a set of patients. We process such data following the procedure used in the recommendations from [1]. We represent each gene as a gene expression vector whose coordinates are gene expression across all patients. We calculate the Pearson correlation coefficient (PCC) between each pair of genes. The correlation scores obtained are used to form the KNN graph.

**Multi-omics features:** We adopt four biological features previously used in EMOGI [2] as node features, capturing key molecular mechanisms commonly altered in cancer.

- Somatic mutation rate: Calculated from SNV data in TCGA following the HotNet2 preprocessing pipeline [3]. Samples with abnormal mutation counts (ultra-mutated) were removed. The SNV frequency of a gene in each type of cancer was defined as the number of nonsilent SNVs divided by the exonic length of the gene.

- Copy number alteration (CNA): CNA data were obtained from TCGA using the GISTIC2 tool [4]. Both gene amplifications and deletions were considered, with ultramutated samples excluded. The CNA rate of a gene was defined as the number of times it was altered (amplified or deleted) in a specific cancer cohort. All CNA types were aggregated into a single feature per gene.

- DNA methylation in promoter regions: Methylation data were collected from Illumina 450k arrays in TCGA (including both tumor and adjacent normal tissues). For each gene, the promoter was defined as the region $\pm 1,000$ base pairs around the transcription start site, according to GENCODE v28 [5]. The average methylation of the promoter was calculated by averaging the values of $\beta$ of the CpG sites in this region, after batch effect correction using ComBat [6]. Differential methylation was defined as the difference between tumor and normal tissues, averaged in all samples.

- Gene expression: Gene expression data was derived from the Wang et al. dataset [7], which includes tumor samples from TCGA and normal samples from GTEx [8]. The data were quantile normalized and batch-corrected using ComBat. Differential expression was calculated as the $\text{logarithmic}_2$ fold change between tumor and matched normal samples and averaged for each gene.

**Pathway-derived features:** Pathway membership was encoded as binary vectors using pathway data from the KEGG database [9]. For each gene, a value of 1

indicates its inclusion in a given pathway, while 0 denotes its absence. This representation enables the model to incorporate functional context from curated signaling and metabolic pathways, enhancing its ability to capture biologically meaningful interactions.

**Table S1**. Training hyperparameters for different models

| Model | Hidden_channels | # Epochs | Patience | Lr | Weight_decay | Dropout |
|-------|-----------------|----------|----------|-------|--------------|---------|
| GCN | 64 | 300 | 30 | 0.01 | 5e–4 | 0.5 |
| GCN-AE | 64 | 200 | - | 0.001 | - | - |
| Pretrain | 64 | 300 | - | 0.001 | 5e–4 | - |
| Finetune | 64 | 300 | 30 | 0.001 | 5e–3 | 0.5 |

**Table S2**. Summary statistics across 12 cancer types. The first three rows report the number of mutation-derived, pathway-derived, and total gene features. The last two rows show the number of driver and passenger genes.

| BRCA | BLCA | LUAD | THCA | LIHC | LUSC | ESCA | PRAD | STAD | COAD | UCEC | CESC |
|------|------|------|------|------|------|------|------|------|------|------|------|
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 8 | 6 | 7 | 5 | 11 | 7 | 5 | 8 | 8 | 9 | 8 | 14 |
| 12 | 10 | 11 | 9 | 15 | 11 | 9 | 12 | 12 | 13 | 12 | 18 |
| 202 | 95 | 179 | 72 | 82 | 26 | 89 | 57 | 94 | 155 | 78 | 18 |
| 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 | 2187 |

**Table S3**. Performance comparison on cancer type–specific driver gene prediction

| Cancer | Ours | GAT | Chebnet | EMOGI | MTGCN | DGHNN |
|--------|------|-----|---------|-------|-------|-------|
| BRCA | 0.7346 | 0.4387 | 0.6507 | 0.6482 | 0.6583 | 0.6058 |
| BLCA | 0.6853 | 0.2892 | 0.5394 | 0.5485 | 0.6568 | 0.6324 |
| LUAD | 0.6416 | 0.3398 | 0.5771 | 0.5591 | 0.6279 | 0.5897 |
| LIHC | 0.5215 | 0.2340 | 0.4215 | 0.3845 | 0.4645 | 0.3575 |
| THCA | 0.3823 | 0.0976 | 0.2860 | 0.2587 | 0.2907 | 0.2413 |
| LUSC | 0.4114 | 0.0673 | 0.2401 | 0.2222 | 0.3099 | 0.2326 |
| ESCA | 0.5536 | 0.2478 | 0.4288 | 0.4174 | 0.4721 | 0.3132 |
| PRAD | 0.6784 | 0.1777 | 0.4910 | 0.5233 | 0.6143 | 0.6310 |
| STAD | 0.6329 | 0.1717 | 0.4730 | 0.4604 | 0.5931 | 0.5478 |
| COAD | 0.5113 | 0.1932 | 0.3490 | 0.3459 | 0.3816 | 0.2957 |
| UCEC | 0.5148 | 0.1658 | 0.4021 | 0.4002 | 0.4954 | 0.4207 |
| CESC | 0.5762 | 0.0839 | 0.5261 | 0.5547 | 0.5972 | 0.4036 |

**Table S4**. Ablation study: performance comparison with and without pathway features, gene expression (GE), and pretraining

| Cancer | No Pathway | With Pathway | No GE | With GE | No Pretrain | With Pretrain |
|--------|-----------|-------------|-------|---------|-------------|---------------|
| BRCA | 0.7056 | 0.7310 | 0.7152 | 0.7310 | 0.7310 | 0.7346 |
| BLCA | 0.6721 | 0.6808 | 0.6756 | 0.6808 | 0.6808 | 0.6853 |
| LUAD | 0.6146 | 0.6458 | 0.6185 | 0.6458 | 0.6458 | 0.6416 |
| LIHC | 0.4415 | 0.4558 | 0.4205 | 0.4558 | 0.4558 | 0.5215 |
| THCA | 0.3362 | 0.3556 | 0.3156 | 0.3556 | 0.3556 | 0.3823 |
| LUSC | 0.3208 | 0.3846 | 0.3185 | 0.3846 | 0.3846 | 0.4114 |
| ESCA | 0.5385 | 0.5523 | 0.4858 | 0.5523 | 0.5523 | 0.5536 |
| PRAD | 0.5363 | 0.6825 | 0.6010 | 0.6825 | 0.6825 | 0.6784 |
| STAD | 0.5824 | 0.5662 | 0.5310 | 0.5662 | 0.5662 | 0.6329 |
| COAD | 0.5005 | 0.4945 | 0.4926 | 0.4945 | 0.4945 | 0.5113 |
| UCEC | 0.4372 | 0.4189 | 0.4452 | 0.4189 | 0.4189 | 0.5148 |
| CESC | 0.4050 | 0.5185 | 0.5320 | 0.5185 | 0.5185 | 0.5762 |

**Table S5**. Top 100 prioritized genes ranked by the model (with scores)

| Rank | Gene | Score | Rank | Gene | Score | Rank | Gene | Score | Rank | Gene | Score |
|------|------|-------|------|------|-------|------|------|-------|------|------|-------|
| 1 | SUN5 | – | 26 | WNT5A | 0.35 | 51 | CDK1 | 0.35 | 76 | PTK2 | 0.45 |
| 2 | TTN | 0.2 | 27 | FZD1 | 0.25 | 52 | FZD7 | 0.35 | 77 | WNT11 | 0.2 |
| 3 | PCBP4 | – | 28 | WNT3A | 0.2 | 53 | EGF | 0.45 | 78 | TLE1 | 0.2 |
| 4 | SUN3 | – | 29 | GRB2 | 0.35 | 54 | DKK1 | 0.35 | 79 | IMPG2 | – |
| 5 | DVL1 | 0.2 | 30 | MUC12 | – | 55 | CD44 | 0.45 | 80 | CDK2 | 0.35 |
| 6 | LRP5 | 0.25 | 31 | MUC6 | 0.25 | 56 | DNAH1 | 0.15 | 81 | WIF1 | 0.3 |
| 7 | PIK3R3 | 0.25 | 32 | MAPK9 | 0.35 | 57 | KDR | 0.45 | 82 | MUC7 | 0.1 |
| 8 | PIK3R2 | 0.1 | 33 | PTPRB | 0.15 | 58 | PRKCB | 0.35 | 83 | STAT3 | 0.45 |
| 9 | DVL2 | 0.2 | 34 | FZD4 | 0.2 | 59 | FZD9 | – | 84 | MAP4K3 | – |
| 10 | SYNE3 | – | 35 | MUC20 | – | 60 | SFRP1 | 0.35 | 85 | CTBP1 | 0.35 |
| 11 | DVL3 | 0.2 | 36 | FZD6 | 0.2 | 61 | WNT4 | 0.25 | 86 | WNT7A | 0.2 |
| 12 | SRC | 0.45 | 37 | FZD5 | 0.1 | 62 | RAC2 | 0.2 | 87 | MUC21 | 0.25 |
| 13 | LRP6 | 0.35 | 38 | FN1 | 0.35 | 63 | JUN | 0.35 | 88 | ANK1 | – |
| 14 | MUC16 | 0.35 | 39 | MUC17 | 0.2 | 64 | WNT1 | 0.35 | 89 | SHC1 | 0.35 |
| 15 | GSK3B | 0.35 | 40 | TP53I3 | – | 65 | FZD3 | 0.2 | 90 | MUC5AC | 0.35 |
| 16 | HDAC1 | 0.4 | 41 | ITGB1 | 0.35 | 66 | MUC5B | 0.25 | 91 | VANGL2 | 0.2 |
| 17 | DIRAS1 | 0.2 | 42 | CALM3 | 0.2 | 67 | ZNRF3 | 0.05 | 92 | TCF7L2 | 0.4 |
| 18 | MAPK1 | 0.45 | 43 | MAML2 | 0.1 | 68 | DLL1 | 0.2 | 93 | H3C12 | – |
| 19 | PRKACA | 0.35 | 44 | RNF43 | 0.15 | 69 | PLCG2 | 0.2 | 94 | CALML3 | 0.25 |
| 20 | PRKACB | – | 45 | FZD2 | 0.15 | 70 | PRKCA | 0.35 | 95 | SOS1 | 0.25 |
| 21 | PRKACG | 0.1 | 46 | FZD8 | 0.2 | 71 | WNT2 | 0.75 | 96 | CDKN1A | 0.5 |
| 22 | MAPK3 | 0.35 | 47 | MUC13 | – | 72 | TLE3 | 0.2 | 97 | MAML1 | 0.25 |
| 23 | RBPJ | 0.2 | 48 | PDGFRB | 0.35 | 73 | PLCG1 | 0.25 | 98 | WNT5B | 0.2 |
| 24 | MAPK8 | 0.45 | 49 | BCL2 | 0.45 | 74 | NOTCH4 | 0.35 | 99 | SLITRK2 | – |
| 25 | TRIM55 | – | 50 | TEK | 0.35 | 75 | RAC3 | 0.3 | 100 | WNT6 | – |

**Table S6**. A summary of the number of nodes and edges for each cancer-specific graph

| Cancer | #Nodes | #Edges |
|--------|--------|--------|
| BRCA | 12 809 | 148 969 |
| BLCA | 12 809 | 148 915 |
| LUAD | 12 809 | 149 129 |
| LIHC | 12 809 | 148 278 |
| THCA | 12 809 | 149 244 |
| LUSC | 12 809 | 148 789 |
| ESCA | 12 809 | 149 319 |
| PRAD | 12 809 | 149 756 |
| STAD | 12 809 | 149 327 |
| COAD | 12 809 | 148 993 |
| UCEC | 12 809 | 150 037 |
| CESC | 12 809 | 148 798 |

# References

[1] Chai, H., Wang, Y., Yang, X., Huang, D., Wang, J., Wu, F., Li, Y., Liu, Q.: Deepdriver: Predicting cancer driver genes based on somatic mutations using deep convolutional neural networks. Frontiers in Genetics **12**, 618553 (2021)

[2] Li, Y., Xu, J., Xiong, L., Huang, K., Liu, M., Zeng, J.: Emogi: An expressive model for predicting cancer driver genes using graph neural networks. Bioinformatics **37**(S1), 232–240 (2021) https://doi.org/10.1093/bioinformatics/btab309

[3] Leiserson, M.D., Vandin, F., Wu, H.-T., Dobson, J.R., Eldridge, J.V., Thomas, J.L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M., *et al.*: Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nature genetics **47**(2), 106–114 (2015)

[4] Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., Getz, G.: Gistic2. 0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome biology **12**(4), 41 (2011)

[5] Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., *et al.*: Gencode reference annotation for the human and mouse genomes. Nucleic acids research **47**(D1), 766–773 (2019)

[6] Johnson, W.E., Li, C., Rabinovic, A.: Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics **8**(1), 118–127 (2007)

[7] Wang, Q., Armenia, J., Zhang, C., Penson, A., Reznik, E., Zhang, L., et al.: Data descriptor: Unifying cancer and normal RNA sequencing data from different sources. Scientific Data. 2018

[8] Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., *et al.*: The genotype-tissue expression (gtex) project. Nature genetics **45**(6), 580–585 (2013)

[9] Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M.: Kegg: integrating viruses and cellular organisms. Nucleic Acids Research **51**(D1), 587–592 (2023) https://doi.org/10.1093/nar/gkac963