Tumor Resource Allocation using Transcriptomes and AI

March 13, 2026

Abstract

Cancer cells survive by reprogramming their gene expression to support growth, metabolism, and immune evasion, but they must balance these competing demands. Here, we analyze transcriptome data from twelve cancer types to determine what tumors are biologically optimized for. We apply a machine-learning framework combining a variational autoencoder with classifiers to learn compact representations of gene expression and accurately distinguish cancer types, achieving about 98% accuracy. The learned latent features reveal dominant biological pathways, including metabolic, immune, and proliferative programs. By viewing cancer as a system that must make trade-offs to survive, this approach helps identify the biological pathways that each tumor relies on most and could be targeted by treatment.

Introduction

Cancer is one of the leading causes of death worldwide, affecting millions of individuals and placing a great burden on global health care systems. Beyond this impact, cancer poses significant economic costs due to long-term treatment, hospitalization, and lost productivity. Despite major advances in different genomic technologies and therapies, patient outcomes remain inconsistent. This reflects the diversity of tumor behavior across different cancer types. Understanding the underlying causes of this behavior remains a challenge in cancer biology.

A foundational framework for understanding cancer complexity is the hallmarks of cancer, which describe the core biological capabilities that tumors acquire during progression. The hallmarks include sustained proliferative signaling, resistance to cell death, evasion of growth suppressors, replicative immortality, angiogenesis, invasion and metastasis, immune evasion, and metabolic reprogramming. Instead of arising from a single genetic altercation, cancer develops through activation of multiple hallmarks, where each hallmark contributes to tumor survival and expansion (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011; Hanahan, 2022). The framework highlights cancer as a systems-level disease that interacts with biological processes rather than specific mutations [1, 2, 3].

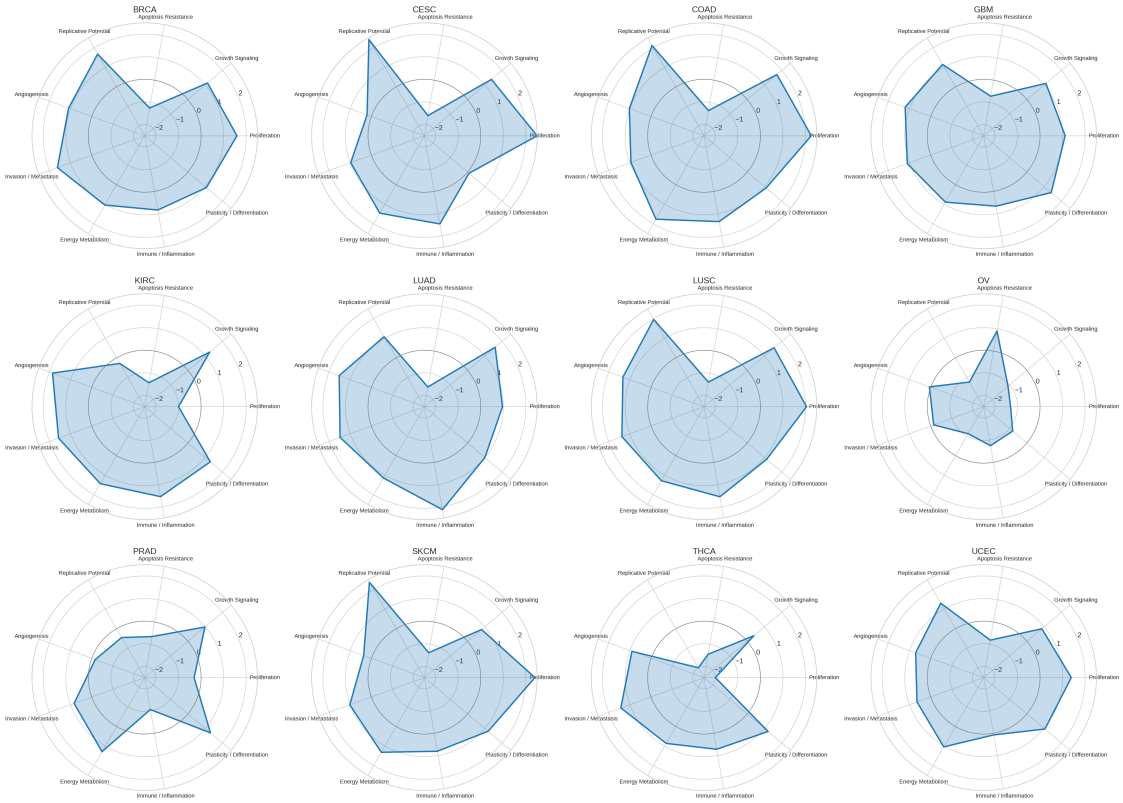

Figure 1: **Hallmark Activation Patterns Across Cancers**: The hallmark scores were computed by aggregating normalized enrichment scores (NES) from MSigDB Hallmark gene sets into nine canonical cancer hallmarks. Each radar plot represents one cancer type. Values greater than 0 indicate enrichment (activation) of the corresponding biological program relative to other cancer types, whereas values less than 0 indicate relative suppression.

It is important to note that tumors must perform several of these hallmarks simultaneously, often under limited biological resources (energy, space, immune tolerance, etc.). As a result, tumors face trade-offs between competing functions. For example, a tumor that is optimized for rapid proliferation may have to sacrifice invasive ability or immune evasion, suggesting that tumors cannot maximize all the hallmarks at once. They are specialized in particular biological tasks, which explains why tumors with similar genetic drivers can behave differently to treatment. This evolutionary trade-off perspective emphasizes tumor heterogeneity and specialization in performing key biological functions [4, 5].

references

Methods

Data Preprocessing

We analyzed gene expression data from twelve cancer types. Only genes present in all tumor types were retained to ensure fair comparison.

Missing values were replaced using the median expression of each gene. For gene-level comparisons, expression values were standardized using z-score normalization. This allows genes to be compared on the same scale.

For each cancer type, we compared its samples against all other cancers combined. We computed a two-sample statistic to measure how strongly each gene was overexpressed or underexpressed. Genes were then ranked from most upregulated to most downregulated. These ranked gene lists were used for pathway analysis.

Gene Set Enrichment Analysis (GSEA)

We performed Gene Set Enrichment Analysis (GSEA) using the MSigDB Hallmark gene sets [6, 7]. GSEA evaluates whether groups of biologically related genes show coordinated activation.

For each cancer type, we calculated normalized enrichment scores (NES) for every hallmark gene set. NES values allow comparison of pathway activity across cancers.

Aggregation into the nine Hallmarks of Cancer

To connect transcriptomic activity to biological meaning, we grouped MSigDB gene sets into nine conceptual cancer hallmarks as described by Hanahan and Weinberg [1, 2, 3]. These include proliferation, immune evasion, angiogenesis, metabolic reprogramming, plasticity, and others. For each hallmark category, we averaged the NES values of its associated gene sets:

$$H_{c,k} = \frac{1}{m_k} \sum_{j=1}^{m_k} \mathrm{NES}_{c,j}.$$

Positive values indicate relative activation in that tumor type. Negative values indicate relative suppression. This step converts complex gene-level data into interpretable biological programs.

Hallmark Space Clustering and Heatmap Visualization

To compare cancers in hallmark space, we constructed a cancer-by-hallmark matrix using aggregated hallmark scores. Values were standardized across cancers (z-scored per hallmark) to highlight relative differences.

Hierarchical clustering was performed using correlation distance and average linkage. Row order was determined from the clustering dendrogram, and cancers were visualized using a heatmap representation. Symmetric color scaling was applied to ensure interpretability of relative activation and suppression patterns.

Comparison of Glandular and Solid Tumors

Tumors were categorized into glandular (e.g., breast, colon, lung adenocarcinoma, ovarian, prostate, thyroid, renal) and non-glandular solid tumors (e.g., cervical squamous, lung squamous, glioblastoma, melanoma).

For each hallmark, group means were computed. Statistical differences between groups were evaluated using two-sample t-test. p-values were reported for each hallmark category.

Identification of Top Upregulated and Downregulated Genes

To better understand tumor-specific transcriptional identity, we calculated the mean normalized expression of each gene within each cancer type. For every tumor, we identified the five most upregulated and five most downregulated genes relative to the global baseline. Upregulated genes highlight active transcriptional programs. Downregulated genes often reflect loss of differentiation or suppression of specific cellular states.

Transcriptome-Based Cancer Classification

To test whether cancer types can be distinguished using gene expression alone, we trained a supervised machine learning classifier. We used the full gene expression matrix without pathway aggregation. The data were split into training (80%) and test (20%) sets using stratified sampling to preserve class balance. Expression values were standardized. We then reduced dimensionality using Principal Component Analysis (PCA), retaining 100 components. Classification was performed using a support vector machine (SVM) with a radial basis function (RBF) kernel. Class weights were balanced to account for unequal sample sizes. Model performance was evaluated on the held-out test set. The PCA+SVM model achieved 98% accuracy, demonstrating that cancer types occupy clearly separable regions in transcriptomic space.

Results

Hallmark-Level Functional Profiling Across Cancer Types

We first asked whether different cancer types exhibit distinct patterns across the canonical Hallmarks of Cancer. For each tumor type, we compared gene expression against all other cancers in the cohort and ranked genes using differential expression statistics. We then performed Gene Set Enrichment Analysis (GSEA) using the MSigDB Hallmark gene sets [6, 7].

Figure 2: Top 5 Upregulated and Downregulated Genes per Cancer (z-score)

GSEA identifies coordinated activation of biological programs. For each cancer type, we computed normalized enrichment scores (NES), which quantify the relative activation or suppression of each gene set compared to the remaining cancers.

To improve biological interpretability, we grouped related gene sets into nine conceptual hallmark categories described by Hanahan and Weinberg [1, 2, 3]. These include proliferation, growth signaling, apoptosis resistance, replicative potential, angiogenesis, invasion and metastasis, deregulated metabolism, immune/inflammatory signaling, and phenotypic plasticity.

Within each conceptual category, NES values were averaged to obtain a single hallmark activity score per cancer. Positive values indicate that a hallmark is more active relative to other cancers, whereas negative values indicate relative suppression.

Cancer Phenotype Clustering in Hallmark Space

Projecting each tumor type into this nine-dimensional hallmark space reveals clear structure across cancers. Hierarchical clustering based on hallmark activity profiles demonstrates that tumors group according to shared biological programs rather than random similarity (Figure 3).

The largest variation across cancers occurs along proliferative and replicative axes. Certain tumors show markedly elevated cell-cycle and mitotic activity, while others display comparatively lower proliferative signatures but stronger angiogenic or immune-related programs. These results indicate that cancers occupy distinct regions in conceptual hallmark space, reflecting different biological strategies for growth and survival.

Metabolic reprogramming is broadly active across tumor types, consistent with the idea that altered energy utilization is a near-universal feature of cancer. Immune and inflammatory signaling varies substantially between cancers, highlighting differences in tumor–microenvironment interactions.

Hallmark Differences Between Glandular and Solid Tumors

To investigate whether broader anatomical classes of tumors exhibit systematic biological differences, cancers were grouped into glandular tumors (e.g., breast, colon, lung adenocarcinoma, prostate, ovarian, thyroid, renal) and solid non-glandular tumors (e.g., cervical squamous, lung squamous, glioblastoma, melanoma). Mean hallmark scores were compared between groups (Figure 4).

Solid tumors displayed significantly higher activity in both *Proliferation* ($p = 0.017$) and *Replicative Potential* ($p = 0.018$) compared to glandular tumors. The magnitude of these differences was substantial, with proliferative activity nearly two units higher on average in solid cancers. This suggests that enhanced cell-cycle machinery is a defining feature of non-glandular solid tumors in this dataset.

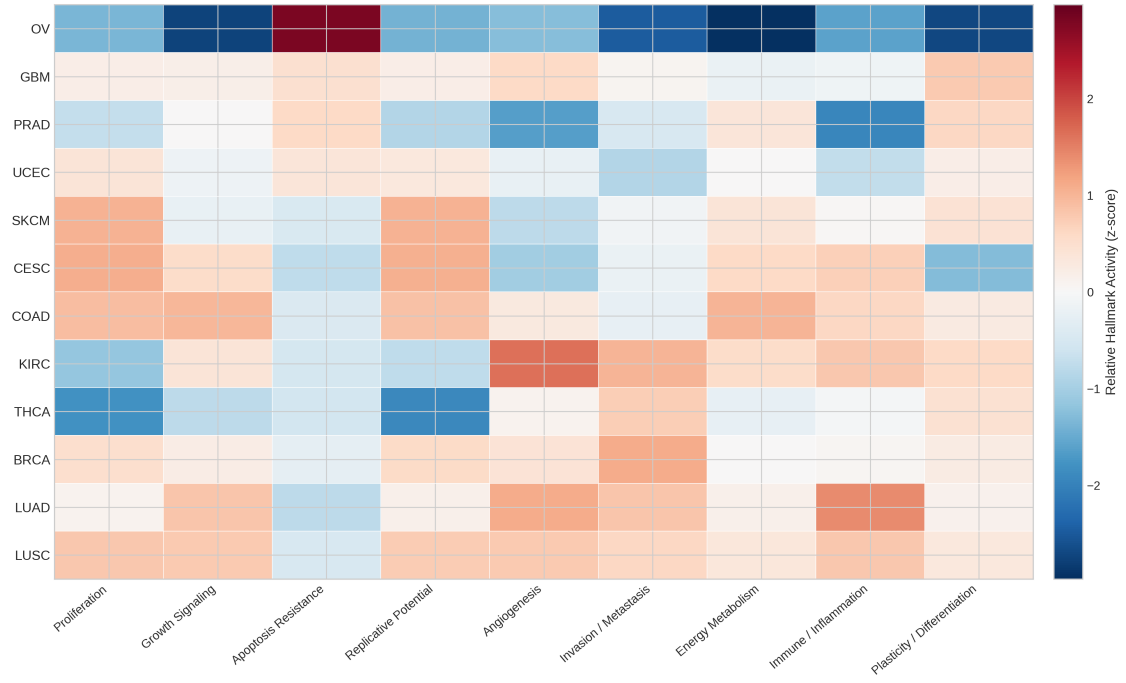

Figure 3: Cancer Phenotype Clustering in Conceptual Hallmark Space. Each row represents a cancer type and each column represents an aggregated hallmark category. Colors denote z-scored relative hallmark activity across cancers.

Other hallmark differences were more modest. Immune and inflammatory signaling tended to be elevated in solid tumors, although variability limited statistical significance. Glandular tumors showed slightly higher apoptosis resistance and angiogenesis scores, but these differences did not reach statistical significance. Metabolic and plasticity programs were broadly active in both groups.

Overall, proliferative and replicative mechanisms represent the strongest axis separating glandular and solid tumors in hallmark space.

Gene-Level Drivers of Hallmark Activity

To understand the gene-level drivers underlying hallmark patterns, we examined standardized expression values and identified the five most upregulated and five most downregulated genes for each cancer relative to the global mean (Figure 2).

Upregulated genes reflect tumor-enriched transcriptional programs. Downregulated genes often indicate loss of differentiation or suppression of lineage-specific functions. Several cancers display strong lineage signals. For example, BRCA overexpresses TFDP3, supporting its proliferative phenotype. Prostate cancer (PRAD) strongly expresses PCA3 and PCGEM1, androgen-responsive transcripts consistent with growth signaling. Kidney cancer (KIRC) shows elevated metabolic transporters, reinforcing its metabolic signature.

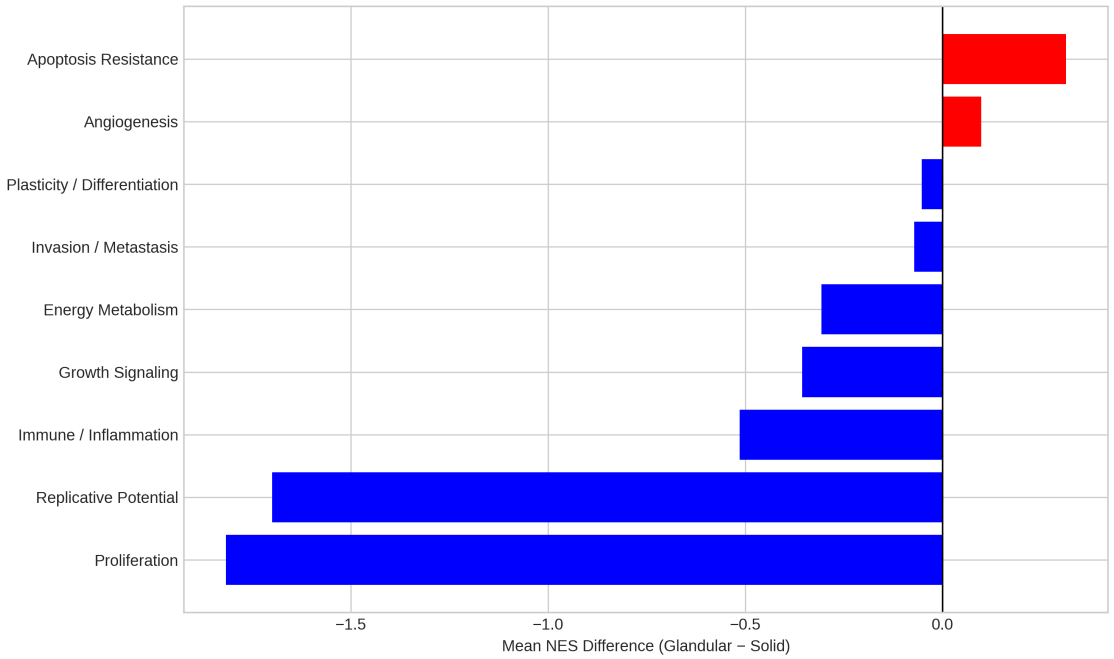

Figure 4: Hallmark Differences Between Glandular and Solid Tumors. Bars show mean NES differences (Glandular Solid) for each conceptual hallmark category.

Glioblastoma (GBM) overexpresses developmental regulators, consistent with increased plasticity.

Downregulated genes provide complementary insight. In cervical cancer (CESC), reduced protocadherin expression suggests diminished cell–cell adhesion, supporting invasion programs. In melanoma (SKCM) and lung squamous carcinoma (LUSC), reduced epithelial marker expression reflects altered differentiation states.

Transcriptome-Based Classification of Cancer Types

Finally, we tested whether global transcriptomic data alone are sufficient to distinguish cancer types. A supervised support vector machine (SVM) classifier trained on principal components of the expression matrix achieved 98% accuracy on held-out test data. The confusion matrix (Figure 5) shows minimal misclassification.

This high separability aligns with the hallmark-based clustering. Cancers that differ in biological programs also separate cleanly in global transcriptomic space, indicating that hallmark-level phenotypes capture meaningful structure within the underlying gene-expression data.

Conclusion

We analyzed gene expression data from twelve different cancer types and translated these data into a more interpretable biological framework. Instead of looking at thousands of individual genes separately, we grouped

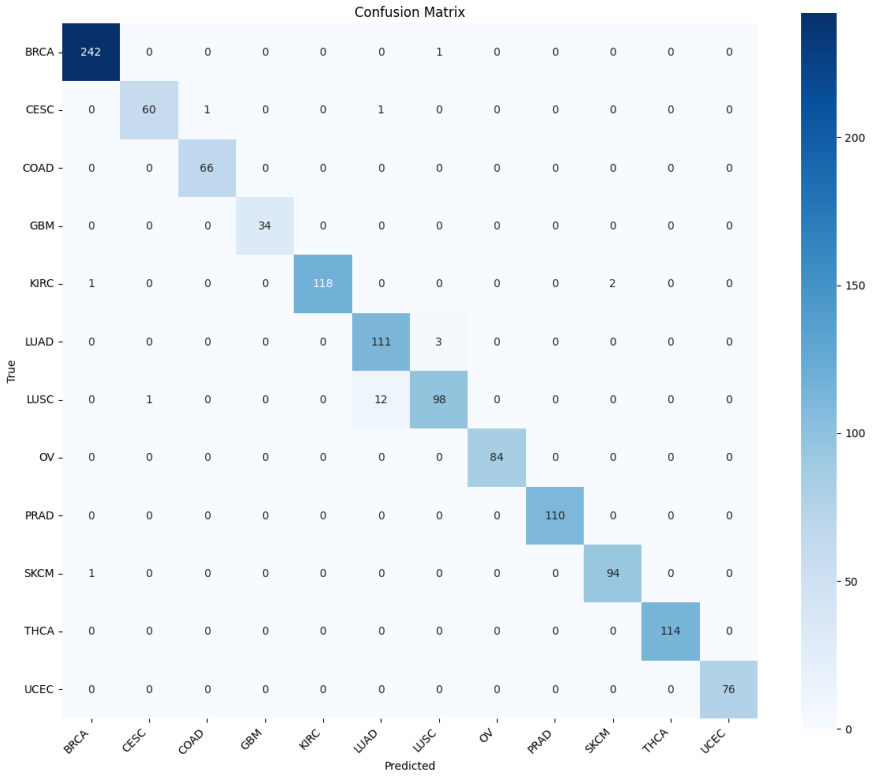

Figure 5: Confusion Matrix for Transcriptome-Based Cancer Classification.

them into nine well-known cancer hallmarks, such as proliferation, immune signaling, metabolism, angiogenesis, and invasion.

This approach allowed us to compare cancers based on their underlying biological programs rather than only their tissue of origin. We found that cancers cluster into distinct regions in hallmark space, meaning that different tumor types rely on different biological strategies for growth and survival. In particular, proliferative and replicative hallmarks were major sources of variation across cancers.

When we compared glandular tumors to non-glandular solid tumors, we observed significant differences in proliferative activity. Solid tumors showed stronger activation of cell-cycle and replication-related pathways, suggesting a greater dependence on rapid cell division. Other hallmark programs, such as immune signaling and metabolism, also varied between groups, though more subtly.

Importantly, this hallmark-based view aligns with the strong separability observed using machine learning classification. This indicates that the biological programs captured by hallmark aggregation reflect meaningful and consistent differences between tumor types.

This way of analyzing cancer has practical implications for treatment. Many modern therapies target biological pathways rather than specific tissues. For example, drugs may inhibit cell-cycle regulators, block angiogenesis, suppress PI3K/mTOR signaling, or activate immune responses. By identifying which hallmark

programs are most active in a given cancer type, we can better understand which therapeutic strategies may be most effective.

Rather than treating cancers only according to where they originate in the body, a hallmark-based framework helps characterize tumors by their functional behavior. This systems-level view may support more precise treatment selection and improved stratification of patients based on dominant biological programs.

Our results show that complex transcriptomic data can be distilled into clear, biologically meaningful patterns. Mapping cancers into hallmark space provides a structured and interpretable way to compare tumors and may contribute to future efforts in precision oncology.

References

[1] Douglas Hanahan and Robert A. Weinberg. "The Hallmarks of Cancer". In: *Cell* 100.1 (2000), pp. 57–70. DOI: 10.1016/S0092-8674(00)81683-9.

[2] Douglas Hanahan and Robert A. Weinberg. "Hallmarks of Cancer: The Next Generation". In: *Cell* 144.5 (2011), pp. 646–674. DOI: 10.1016/j.cell.2011.02.013.

[3] Douglas Hanahan. "Hallmarks of Cancer: New Dimensions". In: *Cancer Discovery* 12.1 (2022), pp. 31–46. DOI: 10.1158/2159-8290.CD-21-1059.

[4] Jean Hausser and Uri Alon. "Tumour heterogeneity and the evolutionary trade-offs of cancer". In: *Nature Reviews Cancer* 20 (2020), pp. 247–257. DOI: 10.1038/s41568-020-0241-6.

[5] Amy M. Boddy, Wei Huang, and Athena Aktipis. "Life History Trade-Offs in Tumors". In: *Current Pathobiology Reports* 6.4 (2018), pp. 201–207. DOI: 10.1007/s40139-018-0188-4.

[6] Arthur Liberzon et al. "The Molecular Signatures Database Hallmark Gene Set Collection". In: *Cell Systems* 1.6 (2015), pp. 417–425.

[7] Aravind Subramanian et al. "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences* 102.43 (2005), pp. 15545–15550.