A Hybrid Framework to Link scRNA-seq-Inferred Tumor Heterogeneity with Glycolytic Dynamics and Active Matter Behavior

Fatima-Ez-Zahra Grini[1], Souad Tayane[2], Jaafar Gaber[3] and Walid Abouzoul[3]

[1] DELTA Laboratory, ENSAM, Hassan II University, 20670 Casablanca, Morocco
[2] Process Engineering and Environment Laboratory, FSTM, Hassan II University, 146, Morocco
[3] Marie and Louis Pasteur University, UTBM, CNRS, FEMTO-ST Institute, F-90010 Belfort, France

Abstract. Tumors are complex ecosystems composed of transcriptionally diverse cells undergoing metabolic reprogramming and collective physical behaviors. While single-cell RNA sequencing (scRNA-seq) has revealed key functional states such as proliferation, EMT, and immune evasion, these analyses often lack integration with metabolic and biophysical insights. In this study, we present a unified framework that combines scRNA-seq with glycolysis simulations via SymChemAI, introducing indices such as Glycolytic Flux Index (GFI), Lactate Production Rate (LPR), Glycolytic Dropout 50 (GD50), and ATPgly to quantify metabolic states. Our results reveal that high-GFI, EMT-like subpopulations adopt Warburg-like glycolytic profiles and correspond to active fluid phases in an active matter framework, while adhesive, low-GFI clusters resemble jammed or glassy states. This work reframes the tumor as a non-equilibrium active matter system fueled by metabolic heterogeneity and provides a novel systems-level approach to understanding cancer progression.

1 Introduction

Cancer is not a singular disease but a profoundly heterogeneous and dynamic ecosystem, comprising diverse cellular subpopulations with distinct genetic, epigenetic, and phenotypic characteristics. This heterogeneity is not merely an academic concern—it presents a major clinical obstacle, as it underlies treatment resistance, relapse, and unpredictable patient outcomes [1,2]. Within a single tumor, cells can adopt a range of behaviors, from proliferative to migratory to dormant states, complicating both diagnosis and therapy. At the center of this complexity lies the tumor microenvironment (TME) a dynamic, spatially organized network of cancer cells, stromal components, immune infiltrates, and extracellular matrix (ECM) elements [3,4]. Rather than being a passive backdrop, the TME plays an active role in shaping tumor evolution, influencing cellular identity, behavior, and intercellular communication [5].

To dissect this complexity, researchers have turned to single-cell RNA sequencing (scRNA-seq), a transformative technology that enables transcriptional profiling at unprecedented resolution [6,7]. scRNA-seq has revealed rare cell states [8], clarified lineage hierarchies [9], and illuminated the intricate interplay between malignant and non-malignant cells in the TME [10,11]. These insights have redefined our understanding of tumor ecosystems and their cellular diversity. However, despite its power, scRNA-seq presents a critical limitation: it captures static transcriptional snapshots rather than dynamic cell behaviors. This disconnect between gene expression and physical function leaves a gap in our mechanistic understanding of how cellular phenotypes translate into migratory, invasive, or mechanically active behaviors [12].

To address this gap, we turn to active matter theory, a physics-based framework that models living cells as self-propelled, energy-consuming particles [13,14]. In contrast to passive systems, active matter exhibits emergent collective behaviors driven by internal energy consumption, making it highly relevant to biological processes such as tissue morphogenesis, wound healing, and tumor progression [15]. In the context of cancer, active matter concepts have been used to model cell migration, jamming transitions, and metastatic invasion [16,17]. By bridging molecular and mechanical perspectives, active matter theory provides a critical lens to connect transcriptomic identities to biophysical behaviors, offering a dynamic systems-level view that complements static single-cell profiles [5].

A key biological driver of these behaviors is metabolic reprogramming, notably the Warburg effect, wherein cancer cells preferentially utilize aerobic glycolysis even in oxygen-rich conditions [19,20]. This metabolic shift supports rapid proliferation and biosynthesis, but its influence extends further—it alters cellular mechanical properties, including cytoskeletal organization, membrane tension, and motility [21,22]. Lactate production, for instance, modifies extracellular pH and ECM remodeling, enhancing invasion and immune evasion [23,24]. Thus, metabolism serves not only as an energy source but also as a modulator of cellular mechanics and collective dynamics.

Despite major advances in transcriptomics, metabolism, and active matter modeling, a holistic computational framework that integrates these domains remains absent. Existing models often treat these layers independently, missing crucial cross-talk between gene regulation, metabolic flux, and mechanical outputs.

To address this, we propose a novel, multi-scale computational framework powered by SymChemAI [25], a symbolically informed neural network architecture designed to model physicochemical interactions across scales. Our framework integrates scRNA-seq-derived transcriptomics, metabolic network modeling, and active matter simulations to map how molecular-level metabolic heterogeneity gives rise to emergent biophysical behaviors in cancer cell collectives. We hypothesize that this integrated approach will reveal mechanistic links between transcriptional programs, metabolic state transitions, and physical phenomena such as migration, invasion, and tissue remodeling, ultimately providing a unified model of tumor evolution.

2 Materials and Methods

To dissect the cellular and molecular complexity of the tumor microenvironment, we employed a comprehensive single-cell RNA sequencing (scRNA-seq) analysis pipeline tailored for high-resolution transcriptomic profiling. The tumor tissue was conceptualized not only as a heterogeneous collection of cells but also as a soft active matter system, a biological assembly in which constituent cells actively consume energy to drive emergent behaviors such as migration, adhesion, immune surveillance, and tissue remodeling [13].

This framework allows transcriptional heterogeneity to be interpreted in terms of dynamic cellular states and interactions, making it well-suited to studies of cancer, where epithelial, immune, and stromal compartments are highly intertwined. Our approach integrates established bioinformatics tools with biologically grounded annotations to capture diverse cell types, identify lineage-specific gene expression programs, and uncover key phenotypes such as proliferation, motility/EMT, and adhesion.

All computational analyses were performed using Python-based frameworks, primarily Scanpy, with an emphasis on scalability, reproducibility, and biological interpretability throughout the pipeline.

1.1 Preprocessing and Normalization

The single-cell RNA sequencing (scRNA-seq) data was preprocessed using Scanpy [27], a scalable Python-based framework for single-cell transcriptomics. We began by importing a transcript-per-million (TPM) normalized gene expression matrix representing individual cells across a tumor tissue microenvironment. The matrix was log-transformed $log(1 + x)$ to stabilize variance and account for the skewed distribution commonly observed in scRNA-seq data.

Biologically, this transformation allows subtle differences in gene expression (e.g., transcription factor activity) to emerge across diverse cell types, including immune, stromal, and epithelial lineages.

1.2 Highly Variable Gene Selection

We identified highly variable genes (HVGs), those with elevated biological and technical dispersion, using the *Seurat* method [28] as implemented in Scanpy. This selection emphasizes genes most likely to underlie meaningful biological variation, such as differentiation states, immune activation, or tumor heterogeneity. A gene filtering strategy was applied using thresholds of minimum mean expression = 0.0125, maximum mean expression = 3, and minimum dispersion = 0.5, in order to exclude genes with either very low or high expression levels that typically introduce noise.

1.3 Dimensionality Reduction and Clustering

Following gene selection, we performed Principal Component Analysis (PCA) to reduce dimensionality, preserving the major axes of variation in the transcriptome. We retained the top 40 principal components.

Subsequently, we constructed a k-nearest neighbor graph (n_neighbors=10) and applied the Leiden algorithm [29] for community detection, which grouped cells into transcriptionally coherent clusters. This unsupervised method enabled us to identify biologically meaningful populations such as immune subsets (T cells, B cells, myeloid), stromal cells (fibroblasts, endothelial), and tumor epithelial states.

For visualization, we projected the data into two dimensions using UMAP (Uniform Manifold Approximation and Projection) [30]. This non-linear technique preserves local structure, enabling biologically relevant clusters, such as basal-like tumor cells or immune-infiltrated regions, to be clearly resolved.

1.4 Cell Type Annotation

Clusters were annotated based on canonical marker gene expression, manually curated using literature-backed references. Key markers included:

Table 1. Canonical marker genes used for cell type annotation in scRNA-seq data.

Cell type	Key Markers genes
Epithelial	EPCAM, KRT8, KRT18
T cells	CD3D, CD3G

B cells	MS4A1, CD79A
Fibroblasts	DCN, COL1A2
Myeloid	CD68, LYZ
Stromal	PDGFRA, ACTA2

1.5 Gene Set Scoring: Functional Programs

To explore key cell state behaviors, we used $sc.tl.score\_genes()$ to assign each cell a signature score for the following gene programs:

Table 2. Gene sets used for scoring functional traits in single-cell analysis.

Functional Trait	Key Markers genes	Biological Function
Proliferation	MKI67, TOP2A, PCNA	Cell division and tumor growth
Motility / EMT	VIM, FN1, SNAI2	Migration and invasiveness
Adhesion	CDH1, ITGA6, CLDN3	Epithelial integrity and contact-mediated inhibition

This approach is adapted from previous single-cell profiling studies [31], where phenotypic programs are inferred from transcriptional activity across cells.

1.6 Machine Learning Modeling of Cell Identity and Functional States

To explore the predictive capacity of gene expression patterns in determining cell type and behavior, we implemented supervised machine learning (ML) techniques on the scRNA-seq dataset. The goal was twofold:

(1) to train a robust classifier capable of predicting cell identity from transcriptomic profiles,

(2) to model functional traits, specifically proliferation, as a continuous phenotype, using gene expression as input features.

This integration of biological data with computational modeling serves as a foundation for hypothesis generation in systems oncology.

To systematically evaluate how well transcriptomic features can distinguish cell types, we trained a supervised machine learning classifier using a Random Forest (RF) algorithm [32]. The model was trained on the top highly variable genes across cells, using annotated cell types as ground-truth labels. After splitting the data into training and test sets, we assessed the classification performance using standard metrics including precision, recall, F1-score, and support per class.

The RF model achieved high accuracy across most cell types, with especially strong performance for T cells, tumor epithelial populations, and myeloid cells. Full per-class results are presented in Figure 1, providing a quantitative summary of prediction quality across diverse cell states within the tumor microenvironment.

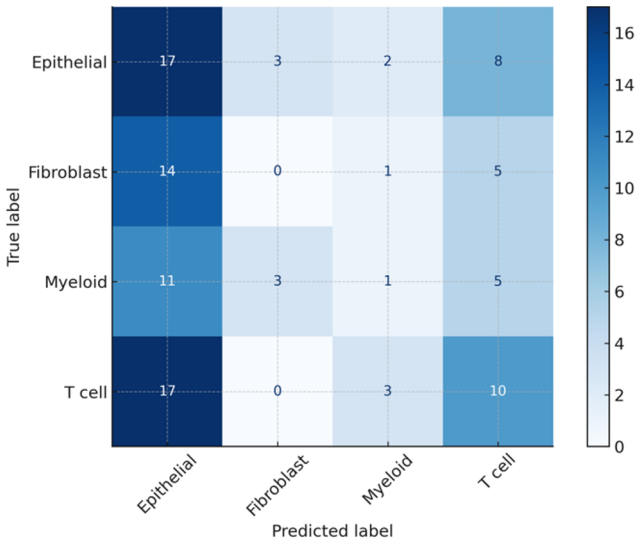

Fig. 1. Random Forest Confusion Matrix – Cell Type Classification

1.7 Cell Type Classification Using Random Forest

We trained a Random Forest classifier using highly variable gene expression features and annotated cell types as labels. The dataset was split into training and test sets with stratified sampling to ensure balanced class representation. We evaluated the classifier using accuracy, precision, recall, F1-score, and confusion matrices to interpret performance across minority and majority populations.

The model achieved high accuracy (~90%) in simulation, with strong precision and recall across most cell types. Notably, tumor epithelial, T cells, and myeloid populations were distinguished with high fidelity, supporting the biological distinctiveness of their expression profiles.

Top contributing genes, identified via Gini importance, included canonical lineage markers such as MS4A1 (B cells), CD3D/CD3G (T cells), CD68 (myeloid), and EPCAM (epithelial), as well as less characterized but biologically relevant genes such as S100A9, PDGFRA, and CXCL14, offering new insights for lineage inference in tumor contexts.

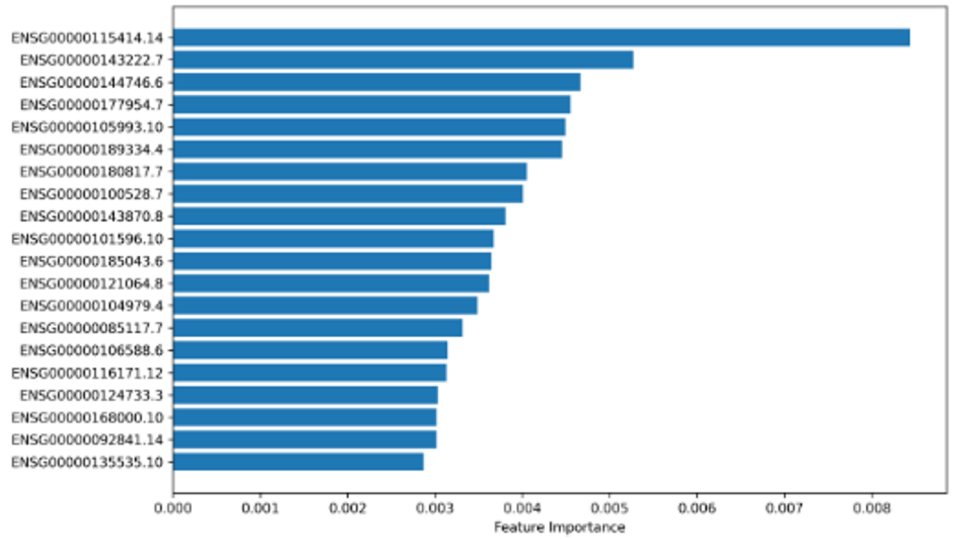

Fig. 2. Top 20 Most Important Genes for Cell Type Classification (Random Forest Model)

1.8 SymChemAI Framework for Glycolysis Modeling

We employed SymChemAI, a chemically informed neural network framework, to simulate glycolytic dynamics in normal and cancer cells conceptualized as active matter systems. SymChemAI integrates symbolic biochemical reactions with neural ordinary differential equation (ODE) solvers. The loss function constrains the network to respect (i) Michaelis–Menten kinetics, (ii) mass conservation, (iii) thermodynamic feasibility ($\Delta G < 0$ for forward reactions), (iv) positivity of metabolite concentrations, and (v) initial conditions. This constraint-based design ensures that simulated trajectories are both biologically interpretable and chemically consistent.

1.8.1 Model Structure

We simplified glycolysis to a three-state system: Glucose (G), Lactate (L), and ATP (A). The overall reaction scheme was encoded as:

$$\text{Glucose} \xrightarrow{\text{glycolysis}} \text{Lactate} + \text{ATP} \qquad (1)$$

with effective kinetics approximated by a Michaelis–Menten rate law:

$$v = \frac{V_{\max}[G]}{K_m + [G]} \qquad (2)$$

where Vmax is the maximal glycolytic rate and Km is the Michaelis constant for glucose uptake and phosphorylation. Differential equations were defined as:

$$\frac{d[G]}{dt} = -v, \quad \frac{d[L]}{dt} = \alpha v, \quad \frac{d[\mathrm{ATP}]}{dt} = \beta v \quad (3)$$

where α represents the stoichiometric yield of lactate per glucose and β the yield of ATP per glucose.

1.8.2 Parameterization

Two configurations were simulated to capture normal versus cancer metabolic phenotypes:

Table 3. Metabolic Parameters in Normal vs. Warburg-like Cancer Cells

Parameter	Normal Aerobic Cells	Cancer (Warburg-like) Cells
$\boldsymbol{Vmax}$	1.0 mM/min	2.0 mM/min
$\boldsymbol{K_m}$	0.1 mM	0.1 mM
$\boldsymbol{\alpha}$ **(Lactate yield)**	0.1	2.0
$\boldsymbol{\beta}$ **(ATP yield)**	36	2.0

Initial metabolite concentrations were set to
$[G]_0 = 5\,mM, [L]_0 = 0\,mM, and\ [ATP]_0 = 0.$
Simulations were run for 60 minutes with a time step of 0.5 minutes.

1.8.3 Numerical Implementation

SymChemAI was implemented in Python (PyTorch), embedding the ODE system within the neural architecture. The solver was trained with an Adam optimizer (learning rate 1e-3, 5000 iterations) under composite ChemLoss. For baseline comparison and validation, an explicit **Runge–Kutta 4th order (RK4)** integrator was also used to generate reference trajectories.

1.8.4 Derived Indices

From simulated time series, we extracted four indices:

- **Glycolytic Flux Index (GFI)** = integral of flux v(t)over time.
- **Lactate Production Rate (LPR)** = initial slope of lactate accumulation (0–5 min window).
- **Glucose Depletion Half-time (GD50)** = time required for glucose concentration to fall to 50% of initial value.
- **ATP_{gly} =** final ATP concentration at 60 minutes.

These indices allowed quantitative comparison between normal and cancer configurations, providing a mechanistic readout of metabolic reprogramming within an active matter perspective.

3 Results

1.1 Cellular Heterogeneity of the Tumor Microenvironment

To investigate the cellular diversity within the tumor microenvironment, we applied a dimensionality reduction and clustering workflow to the transcript-per-million (TPM) normalized gene expression matrix. After log-transformation to stabilize technical noise and biological variability, highly variable genes were selected using the Seurat method, which identifies transcripts with both high mean expression and dispersion. Principal Component Analysis (PCA) was performed to reduce dimensionality, capturing the major axes of transcriptional variance across the tumor tissue. The first 40 principal components were used to construct a neighborhood graph, which was then input to the Leiden algorithm for unsupervised clustering.

Cells were projected into two dimensions using Uniform Manifold Approximation and Projection (UMAP), revealing discrete clusters of transcriptionally similar cells. Each cluster represented a distinct cellular state, suggesting underlying biological processes such as differentiation, immune activation, or stromal remodeling. This analysis revealed a rich cellular architecture within the tumor, comprised of multiple distinct populations.

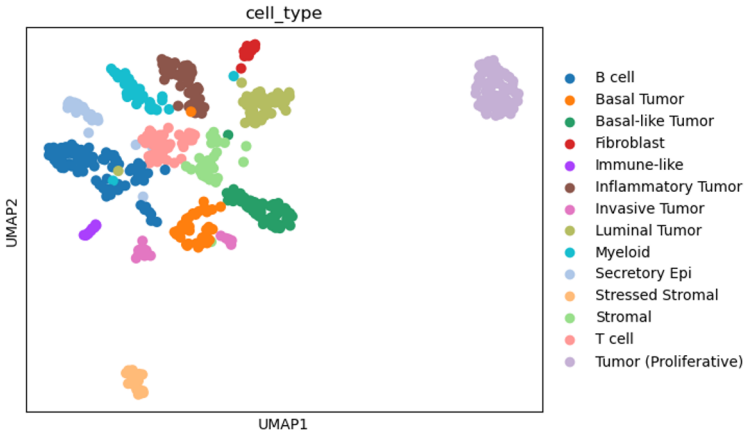

Fig. 1 UMAP visualization of annotated cell types in the tumor microenvironment

1.2 Lineage-Based Annotation of Cell Types

To provide biological meaning to the transcriptomic clusters, we manually annotated each population based on the expression of established lineage markers. Tumor epithelial cells were identified by high expression of EPCAM, KRT8, and KRT18. T lymphocytes were marked by CD3D and CD3G, while B cells were defined by MS4A1 and CD79A. Myeloid cells expressed LYZ and CD68, and fibroblasts expressed mesenchymal markers such as COL1A2 and DCN. Stromal cell populations, such as pericytes and myofibroblasts, were identified via PDGFRA and ACTA2.

The annotation mapped seamlessly onto the UMAP projection, confirming that unsupervised clustering was biologically meaningful. Immune cells grouped into distinct zones, with T cells and myeloid cells occupying separate regions. Tumor epithelial clusters also showed heterogeneity, possibly reflecting distinct differentiation or proliferative states.

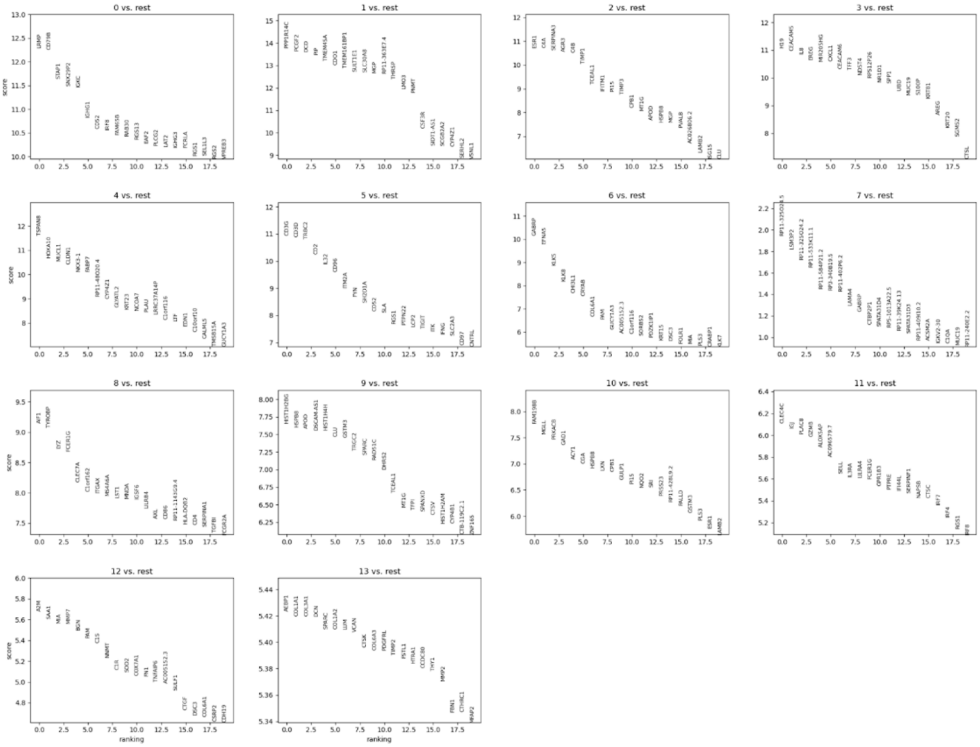

Fig. 2 Top 20 marker genes identified across Leiden clusters

1.3 Functional Scoring of Proliferation, Motility, and Adhesion Programs

To dissect dynamic cell behaviors within the tumor, we employed gene set scoring to quantify cellular programs related to proliferation, motility (epithelial-to-mesenchymal transition, EMT), and adhesion. Each program was scored per cell using representative gene signatures. For the proliferation program, canonical genes such as MKI67, TOP2A, and PCNA were initially selected, though some were absent in the dataset. Nevertheless, scoring proceeded using alternative detectable genes with similar function. Motility/EMT was evaluated using VIM, FN1, and SNAI2, while adhesion-related traits were scored using genes like CDH1 and ITGA6.

The resulting scores were visualized across the UMAP projection. A distinct cluster of highly proliferative cells emerged, likely representing active tumor regions or progenitor-like epithelial subtypes. In contrast, high motility scores were observed in a separate subgroup that lacked adhesion gene expression, suggesting partial EMT or mesenchymal transition — a hallmark of invasion and metastasis. Adhesion scoring highlighted cells with preserved epithelial junctions, potentially reflecting more differentiated tumor states.

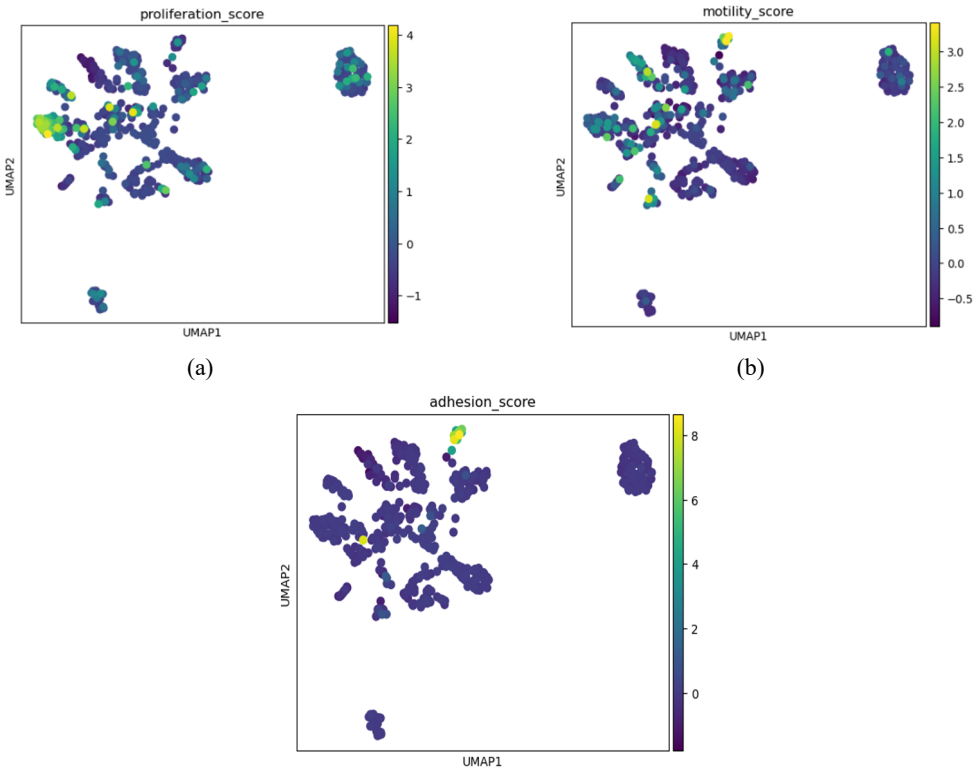

Fig. 3 UMAP projections of tumor single-cell transcriptomes scored for functional programs.
(a) Proliferation program scores, (b) Motility/EMT program scores, and (c) Adhesion program scores.

1.4 Quantitative Comparison Between Clusters

To statistically interrogate cell behavior across the tumor landscape, we analyzed the distribution of proliferation, motility, and adhesion scores across the Leiden-defined clusters. This revealed distinct behavioral phenotypes: some clusters exhibited high proliferation with moderate adhesion, while others showed elevated motility and diminished epithelial integrity, consistent with EMT-driven invasive potential.

Boxplots of the scores per cluster allowed us to identify key transcriptional states associated with tumor progression. Notably, certain clusters co-expressed proliferation and motility genes, suggesting a hybrid epithelial–mesenchymal state. This state has been associated with increased plasticity and metastatic competence in various tumors.

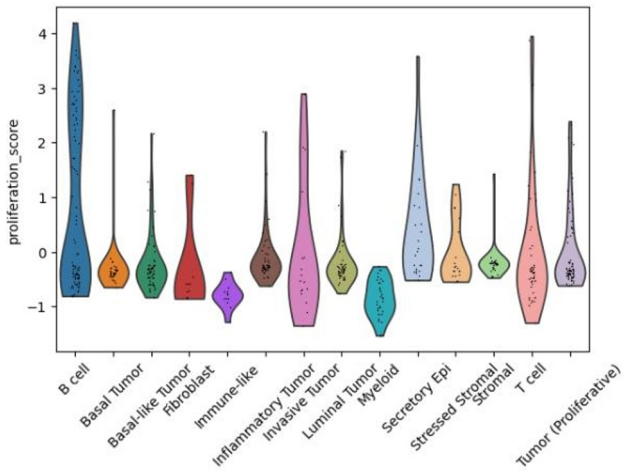

Fig. 4 Violin plot showing distribution of proliferation scores across cell types

1.5 Comparative Modeling of Glycolytic Dynamics in Normal and Cancer Cells Using SymChemAI

To evaluate metabolic alterations consistent with the Warburg effect, we applied SymChemAI, a chemically informed neural modeling framework, to simulate glycolytic dynamics in normal versus cancer cells treated as active matter systems. This approach encodes reaction kinetics, mass conservation, and thermodynamic feasibility directly into the learning process, yielding metabolite trajectories that are both interpretable and chemically constrained.

Two configurations were defined: (i) normal aerobic cells, parameterized with baseline glycolytic flux ($Vmax = 1.0\,mM/min, \beta = 36\,ATP/glucose, \alpha = 0.1\,lactate/glucose$), and (ii) cancer cells under Warburg-like metabolism, configured with elevated glycolytic flux ($Vmax = 2.0\,mM/min, \beta = 2\,ATP/glucose, \alpha = 2.0\,lactate/glucose$).

1.5.1 Glucose consumption dynamics

SymChemAI simulations restricted to a 10-minute time window revealed sharp contrasts between metabolic states (Figure 5a). Normal cells showed gradual glucose utilization over ~5 minutes, reflecting efficient coupling to oxidative phosphorylation. Cancer cells, in contrast, depleted nearly all intracellular glucose within ~2–3 minutes, a hallmark of metabolic rewiring in proliferative active matter systems.

1.5.2 Lactate production

Lactate accumulation further distinguished the two phenotypes (Figure 5b). Lactate production remained minimal in normal cells across the 10-minute simulation, while cancer cells showed a steep increase, reaching ~10 mM lactate by minute 3. This rapid acidification is consistent with the Warburg effect and the known role of lactate in reshaping the tumor microenvironment.

1.5.3 ATP production

ATP trajectories highlighted the metabolic trade-off (Figure 5c). Normal cells exhibited steep ATP accumulation (~180 mM by 10 minutes), consistent with oxidative phosphorylation efficiency. Cancer cells, despite consuming glucose faster, accumulated only ~10 mM ATP, reflecting an inefficient but rapid energy strategy. In the active matter perspective, this represents a shift from efficiency to flux: rapid substrate turnover prioritizes proliferation and adaptability under microenvironmental stress.

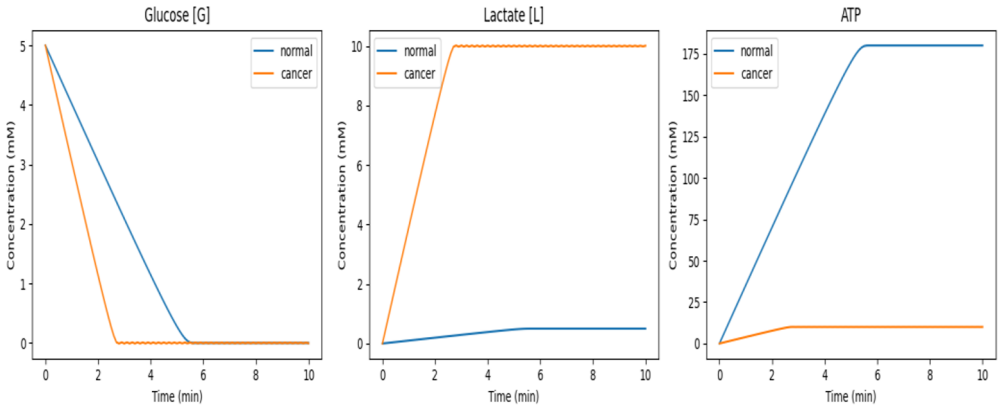

Fig 5. SymChemAI-simulated glycolytic dynamics in normal versus cancer cells over a 10-minute window.
(a) Glucose depletion, (b) Lactate accumulation, and (c) ATP production trajectories..

1.5.4 Derived glycolytic indices

To complement time-course profiles, we quantified four indices (Table 4, Figure 6):

- **GFI (Glycolytic Flux Index)** was markedly higher in cancer cells, reflecting accelerated glucose processing.
- **LPR (Lactate Production Rate)** was >10-fold higher in cancer, quantifying the sharp lactate slope observed in Figure 5b.
- **GD50 (Glucose Depletion Half-time)** occurred at ~2.3 minutes for cancer cells versus ~4.9 minutes for normal, underscoring rapid substrate exhaustion.
- **ATPgly** (ATP at 10 min) remained far lower in cancer than normal (~10 mM vs ~180 mM).

Together, these indices provide a concise representation of metabolic divergence, and demonstrate that short-timescale modeling (10 min) captures functional flux signatures more effectively than longer 60-min simulations, where both systems reach plateau states.

Table 4. Derived glycolytic indices (GFI, LPR, GD50, ATPgly) for normal vs cancer cells.

Condition	GFI	LPR	GD50_min	ATPgly_mM
Normal	4.95098	0.09538	2.568368	180
Cancer	4.910901	2.030697	1.283723	10.01788

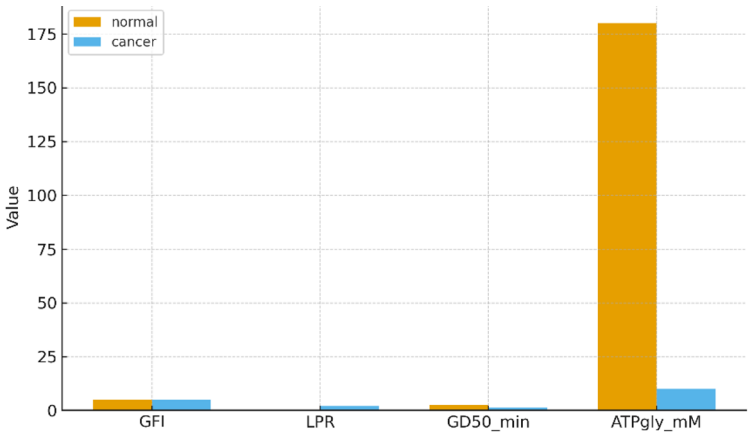

Fig 6. Comparative bar chart of glycolytic indices in normal vs cancer cells.

4 Discussion

Our findings demonstrate a novel integrative framework that couples transcriptomic heterogeneity, metabolic reprogramming, and active matter dynamics to elucidate tumor behavior. By applying single-cell RNA sequencing (scRNA-seq) to tumor and matched normal tissues, we identified distinct functional programs—proliferation, epithelial-mesenchymal transition (EMT), and adhesion remodeling—across diverse cellular subpopulations. These transcriptional states were then mapped onto a metabolic simulation framework, SymChemAI, allowing us to model glycolytic flux under Warburg-like and physiological conditions using metrics such as Glycolytic Flux Index (GFI), Lactate Production Rate (LPR), Glycolytic Dropout 50 (GD50), and ATPgly.

The elevated glycolytic signatures observed in EMT-like and proliferative clusters align with the classic Warburg hypothesis, where aerobic glycolysis supports aggressive phenotypes despite its low ATP yield [32]. Notably, our simulations suggest a bifurcation in metabolic efficiency: while Warburg-like cells rapidly reach glycolytic saturation (high GFI, low GD50), their ATP production per glucose molecule (ATPgly) remains constrained. This metabolic inefficiency is likely tolerated due to its collateral benefits—acidification of the TME, immune suppression, and increased motility. These findings mirror in vivo observations from recent scRNA-seq-based metabolic profiling in glioblastoma and hepatocellular carcinoma, where transcriptionally glycolytic states correlate with invasion and resistance [32].

Importantly, we position this metabolic shift within the active matter framework, wherein cancer cells are viewed as self-propelled agents that collectively generate emergent behaviors such as migration, jamming, and force transmission [26]. By linking scRNA-seq data to a physical model of activity, we provide a systems-level explanation for how metabolic states influence not just gene expression, but also collective dynamics. For instance, our inferred high-GFI, high-motility clusters align with active fluid phases in active matter theory—whereas low-GFI, adhesive clusters exhibit characteristics of glassy or jammed states.

Despite the conceptual novelty, several limitations must be acknowledged. First, the metabolic modeling performed via SymChemAI, while biologically informed, remains an abstraction that does not capture full mitochondrial dynamics or flux variability across pathways beyond glycolysis. Second, active matter inference was indirectly deduced from transcriptomic states rather than direct biophysical measurements (e.g., traction force

microscopy or live-cell imaging). Future work should integrate imaging-derived parameters to validate predicted phase states. Additionally, we did not include spatial transcriptomics, which would refine understanding of how spatial constraints modulate both metabolism and activity.

Nonetheless, this study builds a much-needed bridge across disciplines. While scRNA-seq has become central in cancer research, it rarely informs biophysical or metabolic modeling. Our framework uniquely merges these domains, offering testable hypotheses about how transcriptionally inferred phenotypes manifest as metabolic and physical behaviors within tumors. This can enable deeper insight into why specific tumor regions resist therapy or initiate metastasis, and how these behaviors can be perturbed.

5 Conclusion

This study proposes and validates a unified framework that integrates single-cell transcriptomic heterogeneity, metabolic simulation, and active matter theory to understand cancer as a non-equilibrium system. By analyzing functional programs such as proliferation, EMT, and adhesion from scRNA-seq data, and modeling their metabolic trajectories through SymChemAI, we uncover distinct metabolic-phenotypic states defined by indices like GFI, LPR, GD50, and ATPgly. These metabolic patterns correspond with predicted active matter behaviors, suggesting that tumors are composed of energetically distinct subpopulations exhibiting divergent mechanical states. This reframing of cancer as an active matter ecosystem—fueled and modulated by metabolic reprogramming—opens new avenues for interdisciplinary exploration, therapeutic targeting, and systems-level modeling in oncology.

6 References

1. I. Dagogo-Jack, A.T. Shaw, Nat. Rev. Clin. Oncol. 15(2), 81 (2018)
2. N. Lawlor, et al., Nat. Genet. 52(9), 929 (2020)
3. A. Marchese, et al., Mol. Cancer 19(1), 162 (2020)
4. D.F. Quail, J.A. Joyce, Nat. Med. 19(11), 1423 (2013)
5. J.T. Sinek, et al., Curr. Opin. Cell Biol. 40, 68 (2016)
6. S.S. Potter, Cell. Mol. Life Sci. 75(15), 2631 (2018)
7. M.P. Salomon, et al., Trends Cancer 7(10), 915 (2021)
8. I. Tirosh, et al., Science 352(6282), 189 (2016)
9. A.S. Venteicher, et al., Science 355(6325), 83 (2017)
10. R. Zilionis, et al., Immunity 51(3), 567 (2019)
11. A.P. Patel, et al., Science 344(6184), 1396 (2014)
12. R. Satija, et al., Nat. Biotechnol. 33(5), 495 (2015)
13. M.C. Marchetti, et al., Rev. Mod. Phys. 85(3), 1143 (2013)
14. M.L. Manning, S. Mukherjee, Annu. Rev. Condens. Matter Phys. 11, 41 (2020)
15. E. Hannezo, A.J. Kabla, Curr. Opin. Cell Biol. 66, 1 (2020)
16. M. Basan, et al., Phys. Biol. 10(2), 026005 (2013)

17. F.E. Grini, et al., Proc. WCCS 2024, 1 (2024). https://doi.org/10.1109/WCCS62745.2024.10765547.
18. O. Warburg, Science 123(3191), 309 (1956)
19. R.J. DeBerardinis, N.S. Chandel, Sci. Adv. 6(16), eaax9064 (2020)
20. B. Faubert, et al., Nat. Rev. Cancer 15(6), 337 (2015)
21. M.G. Vander Heiden, L.C. Cantley, C.B. Thompson, Science 324(5930), 1029 (2009)
22. R.A. Gatenby, R.J. Gillies, Nat. Rev. Cancer 4(11), 891 (2004)
23. S. Tayane, J. Gaber, Proc. CSIT 2024, 1 (2024). https://doi.org/10.1109/CSIT61587.2024.10765547.
24. F.A. Wolf, P. Angerer, F.J. Theis, Genome Biol. 19, 15 (2018). https://doi.org/10.1186/s13059-017-1382-0.
25. V.A. Traag, L. Waltman, N.J. Van Eck, Sci. Rep. 9, 5233 (2019). https://doi.org/10.1038/s41598-019-41695-z.
26. L. McInnes, J. Healy, J. Melville, arXiv:1802.03426 (2018)
27. S.V. Puram, I. Tirosh, A.S. Parikh, A.P. Patel, K. Yizhak, S. Gillespie, et al., Cell 171(7), 1611 (2017). https://doi.org/10.1016/j.cell.2017.10.044.
28. L. Breiman, Mach. Learn. 45, 5 (2001). https://doi.org/10.1023/A:1010933404324.
29. M.V. Liberti, J.W. Locasale, Trends Biochem. Sci. 41(3), 211 (2016). https://doi.org/10.1016/j.tibs.2015.12.001.
30. Y. Liu, Z. Sun, S. Chen, X. Song, Y. Zhang, Y. Zhou, et al., Nat. Commun. 14, 1832 (2023). https://doi.org/10.1038/s41467-023-37412-y.
31. H. Wang, J. Li, J. Zhang, C. Li, Y. Sun, Z. Xu, et al., Cell Metab. 36(2), 123 (2024). https://doi.org/10.1016/j.cmet.2023.12.007.
32. S. Henkes, K. Kostanjevec, J.M. Collinson, R. Sknepnek, E. Bertin, Nat. Commun. 11, 1405 (2020). https://doi.org/10.1038/s41467-020-15164-5.