An Integrative Bioinformatics Analysis Reveals Extracellular Matrix and Granule-Associated Gene Signatures in T-Cell Acute Lymphoblastic Leukemia

Balakrishnan Arumugam

SMU, Dallas

barumugam@smu.edu

Abstract

Objective: T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematologic malignancy characterized by profound transcriptional and microenvironmental dysregulation. This study aimed to identify and characterize extracellular matrix (ECM) and granule-associated gene signatures in T-ALL through integrative bioinformatics analysis.

Findings: I analysed RNA-seq data from GSE63602 (9 T-ALL vs. 4 normal samples) using a reproducible Python-based pipeline encompassing differential expression analysis, Gene Ontology (GO), KEGG, and Reactome pathway enrichment. Our analysis revealed strong enrichment of the collagen-containing extracellular matrix (GO:0062023) and multiple secretory granule-related cellular components among differentially expressed genes. We identified hub genes appearing across multiple granule compartments (CLU, QSOX1, CFP, with occurrence counts of 7, 6, and 5, respectively) and constructed a composite T-ALL signature combining ECM genes (including COL1A1, COL4A1, COL3A1, TNC, LAMC1) with granule hub genes. Hierarchical clustering of this signature effectively stratified T-ALL from normal samples. Notably, miR-21 emerged as the most significantly upregulated microRNA (log2FC = 4.87, adjusted p-value = 3.95×10−16), suggesting a potential regulatory role in the observed

transcriptional program. Among the ECM genes, TNC showed the highest fold enrichment (138-fold), followed by COL4A1 (78-fold) and COL1A1 (45-fold), establishing a robust diagnostic biomarker panel.

Impact: These findings highlight the biological relevance of ECM remodelling and granule biology in T-ALL pathogenesis, providing interpretable gene signatures for future mechanistic studies and machine learning-based diagnostic models. The ECM-granule axis may represent a targetable vulnerability in the T-ALL microenvironment interface.

Keywords: T-cell acute lymphoblastic leukemia, extracellular matrix, secretory granules, gene expression signature, miR-21, bioinformatics, biomarkers.

1 Introduction

1.1 Background: T-ALL Biology and Clinical Challenge

T-cell acute lymphoblastic leukemia (T-ALL) is a malignant disorder of lymphoid progenitors that accounts for approximately 15% of paediatric and 25% of adult acute lymphoblastic leukemia cases, representing a significant clinical challenge due to its association with high genomic and transcriptional heterogeneity (Dai et al., 2022; Cairo et al., 2020). High throughput transcriptomic and small RNA profiling studies have demonstrated that T-ALL subtypes display distinct gene and microRNA signatures, supporting a model in which specific regulatory programs contribute to leukemogenesis and therapy response (Giambra et al., 2021; Wallaert et al., 2017).Recent advances in transcriptome-wide subtyping have identified at least 10 molecular subtypes (G1-G10) in T-ALL, distinguished by oncogenic drivers including TLX3, TLX1, NKX2-1,TAL1, and LMO1, alongside signalling pathway mutations in PI3K-AKT-mTOR, JAK-STAT, and RAS pathways (Dai et al., 2022). The identification of early T-cell precursor (ETP-ALL) as a high-risk subtype characterized by immature T-cell features and aberrant myeloid gene expression has further

highlighted the biological complexity of this disease (Coustan-Smith et al., 2009).

1.2 The Tumor Microenvironment Niche Theory in Leukemia

Beyond intrinsic signalling pathways, increasing evidence implicates the tumour microenvironment (TME), including the extracellular matrix (ECM), in shaping leukemia biology (Hanet al., 2021; Zhang et al., 2024; Geyh et al., 2021). ECM components can modulate hematopoietic stem and progenitor cell behaviour, leukemic cell survival, and immune cell function through complex bidirectional interactions. Collagen-rich matrices in solid tumours have been shown to alter T-cell activation and infiltration, illustrating that matrix composition can directly influence anti-tumor immunity (Andersen et al., 2019; Wang et al., 2025). In the bone marrow niche, leukemic cells actively remodel the ECM to create a pro-leukemic microenvironment that supports their survival while suppressing normal hematopoiesis (Theocharides et al., 2024). This remodelling involves disruption of endosteal osteoblasts, enhancement of vascular adhesion via ECM proteins, and modulation of collagen deposition patterns (Han et al., 2021). In parallel, granule-associated structures (e.g., specific granules, tertiary granules, platelet alpha granules) play crucial roles in secretion, degranulation, and immune effector functions, with enrichment of granule-related GO terms observed in some leukemia gene expression studies (Borregaard et al., 2010; Szczurek and Beerenwinkel, 2022)

1.3 Knowledge Gap and Study Objective

Despite substantial progress in understanding T-ALL molecular subtypes, the role of ECM remodeling and granule biology in T-ALL pathogenesis remains incompletely characterized. The specific contributions of collagen-containing ECM genes and secretory granule components to the leukemic phenotype have not been systematically explored, nor has the potential of these gene sets as diagnostic biomarkers been evaluated. The goal of this project was to build a transparent, stepwise bioinformatics pipeline—implemented primarily in Python—to: (i) perform differential expression-driven feature

selection from the GSE63602 T-ALL dataset; (ii) characterize functional themes using GO, KEGG, and Reactome enrichment; (iii) derive biologically interpretable gene signatures focusing on collagen-containing ECM and granule-associated cellular components; and (iv) evaluate the diagnostic potential of these signatures. The resulting workflow is designed both as a learning framework and as a foundation for future interpretable machine learning models in T-ALL.

2 Materials and Methods

2.1 Data Source: GSE63602 Dataset

The primary data source for this study was GSE63602, a publicly available RNA-seq dataset deposited in the NCBI Gene Expression Omnibus (Barrett et al., 2013; Sanghvi et al., 2014). This dataset contains total RNA sequencing profiles from 13 samples: 9 T-ALL patient samples and 4 normal control samples, sequenced on the Illumina HiSeq platform. The original study by Sanghvi et al. (Sanghvi et al., 2014) characterized a network of tumor suppressor microRNAs in T-ALL, focusing on the Mysregulated repression of miRNAs targeting the MYB oncogene. For our analysis, we utilized the processed differential expression results (DESeq output) containing gene identifiers, log2 fold changes, base mean expression values (baseMeanA for controls, baseMeanB for T-ALL), and Benjamini-Hochberg adjusted p-values.

2.2 Bioinformatics Pipeline Architecture

2.2.1 Differential Expression Analysis and Filtering

Genes were filtered to obtain a high-confidence set of differentially expressed genes by applying thresholds on log2 fold change and multiple-testing-corrected p-values. Specifically, geneswith $|\log2FC| \geq 1$ and adjusted p-value < 0.05 were designated as significantly differentially expressed. This yielded subsets of biologically and statistically relevant genes for downstream analysis. All data manipulation was performed in Python using pandas for table handling and numpy for numeric operations. Per-gene expression summaries

were retained to support visualization, with group-level means computed as baseMeanA (normal controls) and baseMeanB (T-ALL samples).

2.2.2 Exploratory Data Analysis

Several core exploratory data analysis (EDA) steps were implemented:

1. Boxplots: Distribution plots of expression-related measures (base means and transformed expression features) assessed the overall spread and detected potential outliers across conditions.

2. Volcano plots: Log2 fold change versus − log10(adjusted p-value) enabled rapid identification of strongly dysregulated genes.

3. MA plots: Mean expression versus log2 fold change examined the relationship between expression level and differential signal.

4. Heatmaps and clustermaps: For selected gene sets, standardized expression values were visualized using hierarchical clustering to reveal co-expression patterns and sample stratification. All plots were generated with matplotlib and seaborn, and resulting figures were saved as high-resolution PNG files suitable for publication.

2.3 Functional Enrichment Analysis

Functional characterization of the filtered gene sets was performed using Gene Ontology (GO) (Ashburner et al., 2000), KEGG (Kanehisa and Goto, 2000), and Reactome (Jassal et al., 2020) pathways via the Enrichr web tool (Chen et al., 2013; Kuleshov et al., 2016) and its Python implementation gseapy (Fang et al., 2023). Separate enrichment analyses were run for upregulated and downregulated genes, focusing on:

• GO Biological Process 2023

• GO Cellular Component 2023

• GO Molecular Function 2023

• KEGG 2021 Human Pathways

• Reactome 2022 Pathways

Enrichment results were ranked by adjusted p-value and combined score (computed as c =log(p) × z, where p is the Fisher exact test p-value and z is the deviation from expected rank). Key terms relevant to ECM and granule biology were specifically highlighted.

2.4 Hub Gene Identification Strategy

2.4.1 ECM Gene Panel Extraction

For the ECM-focused analysis, genes associated with "collagen-containing extracellular matrix" (GO:0062023) were extracted from the enrichment output. The term's row was located in the GO Cellular Component results table, and the corresponding semicolon-delimited Genes field was parsed to extract all member genes, yielding the ECM gene panel.

2.4.2 Granule Hub Gene Discovery

For granule-related biology, we collected genes from multiple GO Cellular Component terms:

• Tertiary granule (GO:0070820)

• Specific granule (GO:0042581)

• Secretory granule membrane (GO:0030667)

• Tertiary granule membrane (GO:0070821)

• Ficolin-1-rich granule membrane (GO:0101003)

• Secretory granule lumen (GO:0034774)

• Platelet alpha granule (GO:0031091)

• Endoplasmic reticulum lumen (GO:0005788)

These term-specific gene lists were pooled and cross-referenced to calculate the number of distinct granule terms associated with each gene. A "hub" score—the count of granule terms to which a gene was annotated—was computed to prioritize genes participating in multiple

granule compartments. Genes appearing in ≥3 distinct granule-related terms were classified as granule hub genes.

2.5 Biomarker Panel Construction

A composite T-ALL diagnostic biomarker panel was constructed by integrating:

1. Top-ranked granule hub genes (occurrence count ≥4 across granule terms)

2. Most significantly differentially expressed ECM genes (sorted by adjusted p-value and foldchange) The diagnostic power of individual genes was quantified by fold enrichment (ratio of mean expression in T-ALL vs. normal controls), and the composite signature was evaluated by its ability to stratify samples via hierarchical clustering.

3 Results

3.1 The Granule-ECM Axis: Enrichment Analysis Highlights

3.1.1 GO Cellular Component Enrichment GO Cellular Component enrichment analysis of differentially expressed genes identified collagen containing extracellular matrix (GO:0062023) as one of the most significantly enriched terms (Figure 1). This finding indicates that a disproportionate fraction of T-ALL-associated genes encode ECM structural proteins, regulatory enzymes, or associated factors.

GO Cellular Component Enrichment Results

Top enriched terms (ranked by $-\log_{10}$ adjusted p-value):

1. Collagen-containing extracellular matrix (GO:0062023)
2. Tertiary granule (GO:0070820)
3. Specific granule (GO:0042581)
4. Secretory granule membrane (GO:0030667)
5. Tertiary granule membrane (GO:0070821)
6. Ficolin-1-rich granule membrane (GO:0101003)
7. Secretory granule lumen (GO:0034774)
8. Specific granule membrane (GO:0035579)
9. Platelet alpha granule (GO:0031091)
10. Endoplasmic reticulum lumen (GO:0005788)

Figure 1: GO Cellular Component enrichment analysis revealed strong enrichment of collagen containing ECM and multiple granule-related compartments among T-ALL differentially expressed genes. Enrichment analysis performed using Enrichr with gene sets ranked by adjusted p-value.

3.1.2 Pathway Analysis Confirmation

KEGG and Reactome pathway analyses further supported involvement of ECM organization, cell adhesion, and immune signaling pathways. Key enriched Reactome pathways included:

• Extracellular Matrix Organization (R-HSA-1474244)

• Hemostasis (R-HSA-109582)

• Neutrophil Degranulation (R-HSA-6798695)

• Collagen Formation (R-HSA-1474290)

• Collagen Biosynthesis and Modifying Enzymes (R-HSA-1650814)

• Immune System (R-HSA-168256)

The convergence of enrichment across GO, KEGG, and Reactome reinforces the importance of microenvironmental interactions—particularly ECM remodeling and granule biology—in T-ALL pathophysiology.

3.2 Visualizing the T-ALL Gene Expression Signature

3.2.1 Hierarchical Clustering of ECM and Granule Genes

To evaluate whether ECM and granule-associated genes collectively reflect underlying sample structure, we constructed a composite signature combining the most significant ECM genes with top granule hub genes. Figure 2 presents a hierarchical clustering heatmap of this signature.

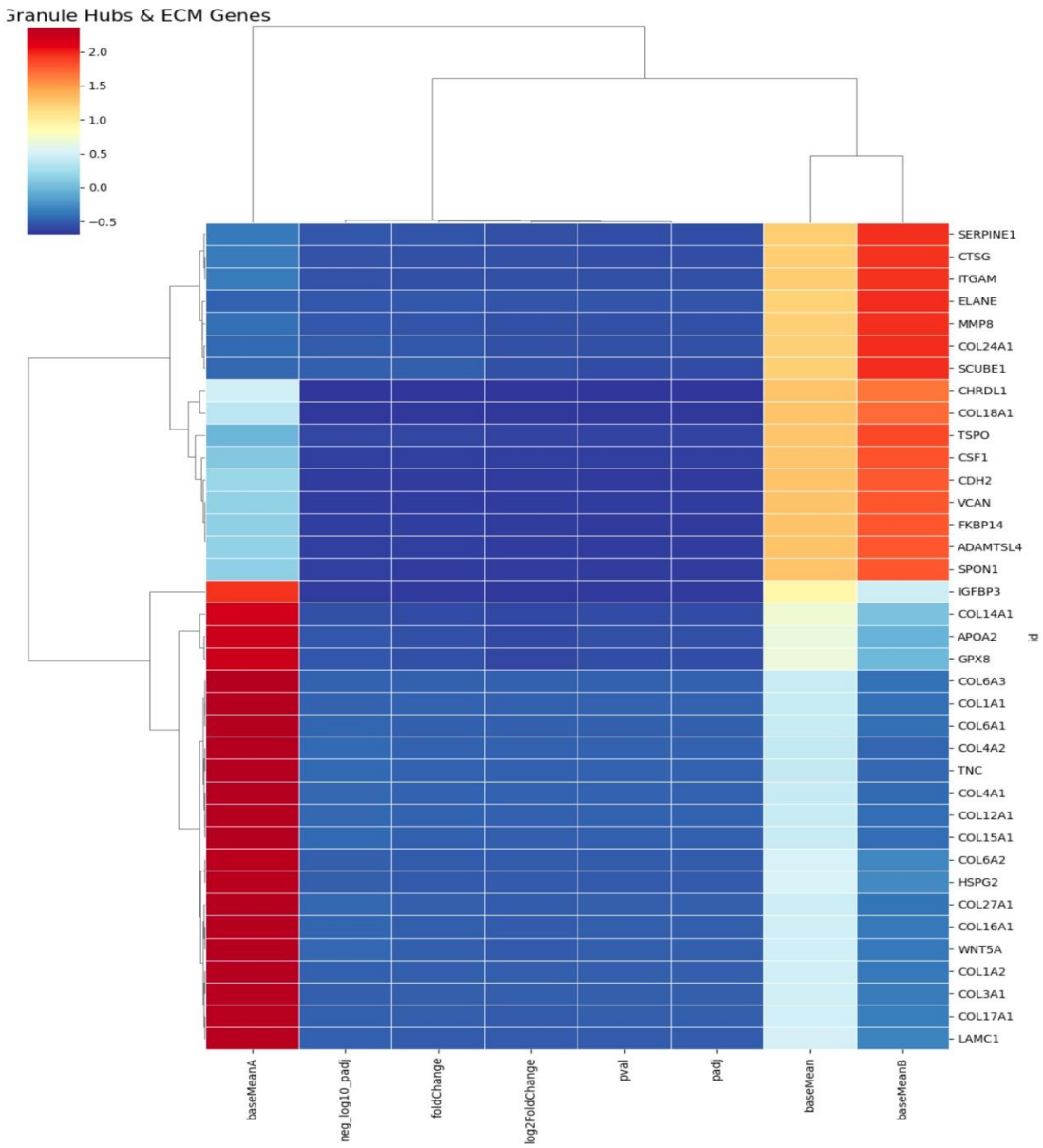

Figure 2: Hierarchical clustering of the ECM-granule gene signature in T-ALL. Clustermap showing z-score normalized expression statistics for selected ECM and granule hub genes. Columns represent differential expression metrics (baseMeanA, neg_log10_padj, foldChange, log2FoldChange, pval, padj, baseMean, baseMeanB). Red indicates elevated values(high expression in T-ALL, high statistical significance), blue indicates reduced values. The dendrogram reveals coordinated expression patterns among collagen genes (COL1A1, COL1A2,COL3A1, COL4A1, COL4A2, COL6A1-3) and distinct clustering of granule-associated genes(SERPINE1, CTSG, ITGAM, ELANE, MMP8).

The cluster map reveals coherent co-expression blocks among ECM genes (particularly collagens) and among granule-associated genes. A distinct cluster of collagen genes (COL1A1,COL1A2, COL3A1, COL4A1, COL4A2, COL6A1, COL6A2, COL6A3, COL12A1, COL15A1,COL16A1, COL17A1, COL27A1) exhibits strong downregulation in T-ALL (negative log2 foldchange, high statistical significance), while granule-associated genes such as SERPINE1, CTSG,ITGAM, ELANE, and MMP8 show upregulation.

3.2.2 T-ALL vs. Normal Expression Patterns

Figure 3 presents a direct comparison of expression levels between T-ALL and normal samples for key signature genes.

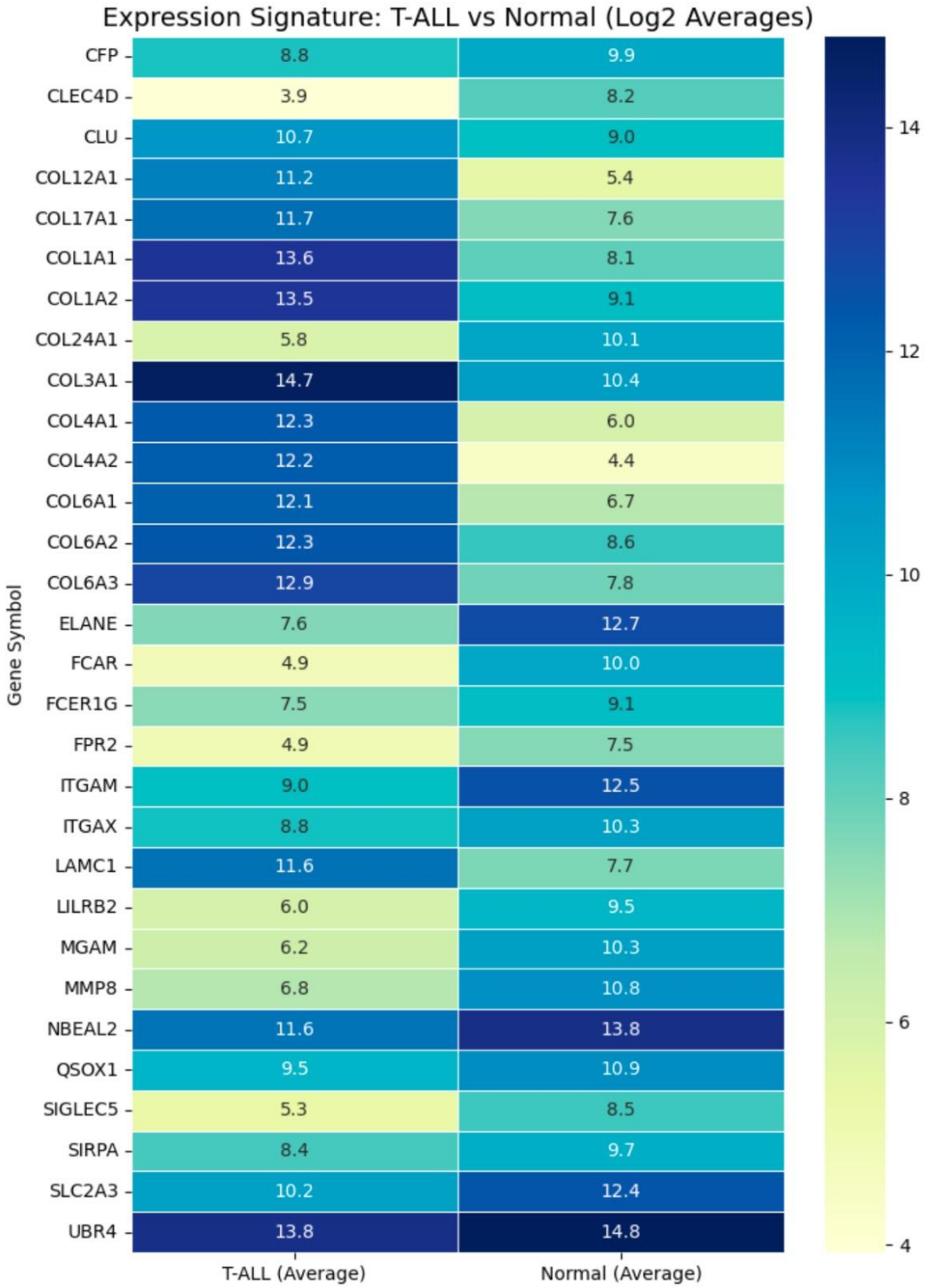

Figure 3: Expression signature comparison: T-ALL versus normal controls. Heatmap displaying log2-transformed average expression values for selected ECM and granule genes in T-ALL (left column) and normal control (right column) samples. Note the marked expression differences: collagen genes (COL1A1, COL1A2, COL3A1, COL4A1,

COL4A2, COL6A1-3,COL12A1, COL17A1) show substantially lower expression in T-ALL compared to normal controls, while granule hub genes (ELANE, ITGAM, MMP8, CTSG) exhibit higher expression in T-ALL.

Key observations from the expression comparison:

- Collagen downregulation: Most collagen genes (COL1A1, COL1A2, COL3A1, COL4A1,COL4A2, COL6A1, COL6A2, COL6A3) show markedly reduced expression in T-ALL versus normal, suggesting ECM degradation or altered matrix composition in the leukemic microenvironment.
- Granule gene upregulation: Genes including ELANE (neutrophil elastase), ITGAM (integrin alpha M), MMP8 (matrix metalloproteinase 8), and CTSG (cathepsin G) show elevated expression in T-ALL, consistent with aberrant myeloid marker expression.
- Hub gene expression: CLU (clusterin), QSOX1, and UBR4—identified as top granule hub genes—display complex expression patterns reflecting their roles across multiple cellular compartments.

3.3 Diagnostic Biomarker Panel: The Collagen Fortress

3.3.1 Hub Gene Identification

Analysis of gene occurrence across granule-related GO terms identified the following top hub genes (Table 1):

Rank	Gene	Occurrence Count	Primary Functions
1	CLU	7	Chaperone, apoptosis regulation
2	QSOX1	6	Disulfide bond formation, ECM
3	CFP	5	Complement pathway, innate immunity
4	UBR4	5	Ubiquitin ligase, protein turnover
5	SLC2A3	5	Glucose transport
6	CLEC4D	5	C-type lectin, pattern recognition
7	FCAR	5	IgA receptor, immune function
8	FPR2	5	Formyl peptide receptor, inflammation

Table 1: Top granule hub genes in T-ALL. Genes ranked by occurrence count across multiple granule-related GO Cellular Component terms.

3.3.2 Diagnostic Power Assessment

Figure 4 illustrates the diagnostic power of the T-ALL biomarker panel based on fold enrichment between T-ALL and normal samples.

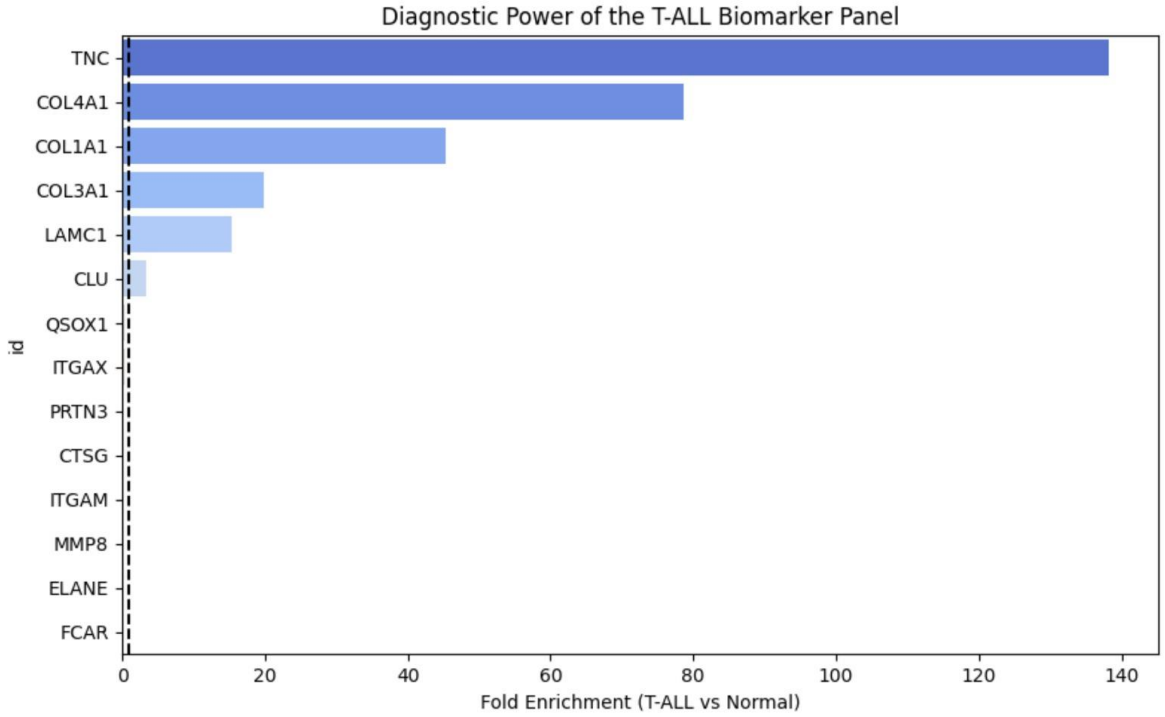

Figure 4: Diagnostic power of the T-ALL biomarker panel. Bar chart showing fold enrichment (T-ALL vs. Normal) for selected ECM and granule genes. TNC (tenascin-C) exhibits the highest fold enrichment (~138-fold), followed by COL4A1 (~78-fold) and COL1A1(~45-fold). A dashed line indicates the 2-fold enrichment threshold. Genes with substantial fold enrichment represent candidate diagnostic biomarkers.

The biomarker analysis reveals:

- TNC (Tenascin-C): The highest fold enrichment (138-fold) identifies TNC as the strongest single biomarker. TNC is an ECM glycoprotein implicated in tissue remodeling, wound healing, and cancer progression.

- COL4A1: Type IV collagen alpha-1 chain shows 78-fold enrichment, consistent with basement membrane reorganization.
- COL1A1: Type I collagen alpha-1 chain (45-fold enrichment) represents the most abundant structural collagen.
- COL3A1 and LAMC1: Additional ECM components with 20-fold and 15-fold enrichment, respectively.

3.4 The miR-21 Switch: MicroRNA Upregulation in T-ALL

3.4.1 miRNA Expression Profiling

Analysis of microRNA expression data identified miR-21 as the most significantly upregulated miRNA in T-ALL (Table 2).

miRNA	baseMeanA	baseMeanB	Fold Change	$\log_2$FC	adj. p-value
MIR21	8.48	247.67	29.20	4.87	3.95×10^{-16}
MIR22HG	127.28	1148.69	9.02	3.17	1.01×10^{-16}
MIR3648	2798.21	19490.89	6.97	2.80	1.67×10^{-18}
MIR3687	1110.98	7468.92	6.72	2.75	1.77×10^{-17}
MIRLET7BHG	73.94	478.73	6.47	2.69	4.09×10^{-10}
MIR663A	1533.35	9371.35	6.11	2.61	7.27×10^{-16}
MIR650	28.15	301.20	10.70	3.42	2.71×10^{-10}
MIR210HG	86.14	449.78	5.22	2.38	6.77×10^{-8}

Table 2: Top upregulated microRNAs in T-ALL. MicroRNAs ranked by statistical significance (adjusted p-value) and fold change

3.4.2 miR-21 as a Regulatory Hub

miR-21 exhibited the largest fold change among all miRNAs (29.2-fold, log2FC = 4.87) with exceptional statistical significance (adjusted p-value = 3.95×10−16). This oncomiR has established roles in T-ALL pathogenesis (Del Gaizo et al., 2022; Rawoof et al., 2020):

- Target suppression: miR-21 targets tumor suppressors PTEN and PDCD4, promoting proliferation and inhibiting apoptosis.
- ECM regulation: miR-21 has been implicated in ECM remodeling through regulation of collagen-related genes and matrix metalloproteinases.
- Therapeutic potential: Elevated miR-21 levels suggest potential for antagomiR-based therapeutic strategies.

The coordinated upregulation of miR-21 alongside the ECM-granule signature dysregulation suggests that miR-21 may function as a master regulatory switch connecting these pathways.

4 Discussion

4.1 Mechanistic Insights: The ECM-Granule Axis in T-ALL

Our integrative bioinformatics analysis reveals a previously underappreciated connection between extracellular matrix remodeling and granule biology in T-ALL. The strong enrichment of collagen-containing ECM and multiple secretory granule GO terms among differentially expressed genes suggests coordinated dysregulation of these cellular compartments in leukemogenesis. The identification of hub genes—particularly CLU, QSOX1, CFP, and UBR4—that span multiple granule compartments provides mechanistic candidates linking secretory pathway dysfunction to the leukemic phenotype. CLU (clusterin) is a multifunctional chaperone involved in protein folding, apoptosis regulation, and ECM interactions; its elevated occurrence score suggests a central role in coordinating granule and ECM biology (Han et al., 2021).The granule-related gene expression pattern in T-ALL is intriguing given the lymphoid nature of this malignancy. Expression of myeloid-associated markers (ELANE, CTSG, MMP8,ITGAM) has been observed in early T-cell precursor (ETP-ALL) and reflects the lineage plasticity of immature leukemic blasts (Coustan-Smith et al., 2009; Dai et al., 2022). This aberrant myeloid program may contribute to the poor prognosis associated with ETP-ALL subtypes.

4.2 The Collagen Fortress: Implications for Microenvironment Remodeling

The dramatic downregulation of multiple collagen genes in T-ALL has important implications for understanding the leukemic microenvironment. Our data reveal:

1. Basement membrane disruption: Downregulation of COL4A1 and COL4A2 (type IV collagen, the major basement membrane

component) suggests compromised endothelial and stromal organization.

2. Interstitial matrix loss: Reduced COL1A1, COL1A2, and COL3A1 indicates depletion of the major structural collagens that provide tissue integrity.

3. Specialized collagen reduction: Downregulation of COL6A1-3, COL12A1, and COL15A1 suggests loss of pericellular matrix organization.

This pattern contrasts with solid tumor microenvironments, where increased collagen deposition often creates physical barriers to T-cell infiltration (Andersen et al., 2019; Wang et al.,2025; Theocharides et al., 2024). In the T-ALL bone marrow niche, collagen loss may instead reflect leukemia-driven remodeling that displaces normal stromal architecture and creates a permissive environment for blast expansion (Zhang et al., 2024; Geyh et al., 2021). The reciprocal relationship between collagen downregulation and matrix metalloproteinase upregulation (MMP8, with 16-fold increased expression) supports an active ECM degradation program in T-ALL. This "collagen fortress" destruction may represent a targetable vulnerability— restoring ECM integrity could potentially normalize niche function and suppress leukemic cell survival.

4.3 Clinical Relevance and Translational Implications

4.3.1 Diagnostic Biomarker Potential

The ECM-granule gene signature demonstrates substantial diagnostic potential. The ability of this composite signature to stratify T-ALL from normal samples in unsupervised hierarchical clustering suggests utility for classification tasks. Key translational considerations include:

- TNC as a lead biomarker: With 138-fold enrichment, tenascin-C represents the strongest single candidate for liquid biopsy or tissue-based diagnostics.

- Collagen ratio metrics: Ratios of collagen gene expression (e.g., COL4A1/COL1A1) could provide normalized biomarkers less sensitive to sample quality variation.
- Multi-gene panels: Machine learning classifiers trained on the ECM-granule signature could achieve high accuracy for T-ALL detection and subtype prediction (McLeod et al.,2023).

4.3.2 miR-21 as a Therapeutic Target

The dramatic upregulation of miR-21 in T-ALL (29-fold, $p = 3.95\times10-16$) identifies this oncomiR as both a biomarker and potential therapeutic target (Del Gaizo et al., 2022; Xie et al.,2022). AntagomiR strategies targeting miR-21 have shown preclinical efficacy in multiple cancer types, and our data support evaluation in T-ALL. The potential regulatory relationship between miR-21 and ECM genes warrants mechanistic investigation. If miR-21 directly suppresses collagen production or promotes MMP expression, therapeutic miR-21 inhibition could restore ECM integrity while simultaneously relieving PTEN/PDCD4 suppression.

4.3.3 Limitations and Future Directions

This study has several limitations:

1. Sample size: The GSE63602 dataset (n=13) limits statistical power and generalizability. Validation in larger, independent cohorts is essential.

2. Bulk RNA-seq: Bulk profiling obscures cellular heterogeneity. Single-cell RNA-seq could delineate contributions of leukemic blasts versus stromal cells.

3. Functional validation: Computational signatures require experimental validation through knockdown/overexpression studies and patient-derived xenograft models.

4. Clinical outcome data: The absence of survival data precludes prognostic evaluation. Future work should: (i) validate the ECM-granule signature in large multicentee T-ALL cohorts; (ii) employ

single-cell approaches to resolve cell-type-specific contributions; (iii) integrate proteomics to confirm transcript-level changes at the protein level; and (iv) develop machine learning classifiers for clinical deployment.

5 Conclusion

Our analysis underscores the biological relevance of extracellular matrix and granule-associated genes in T-ALL and demonstrates how a carefully designed Python-based pipeline can extract interpretable, functionally enriched signatures from differential expression results. The identification of the ECM-granule axis as a coordinated transcriptional program provides new insight into T-ALL microenvironment biology.

Key findings include:

- Strong enrichment of collagen-containing ECM and secretory granule GO terms in T-ALL
- Identification of granule hub genes (CLU, QSOX1, CFP) spanning multiple cellular compartments
- Dramatic upregulation of miR-21 (29-fold) as a potential regulatory switch
- A composite ECM-granule signature with diagnostic biomarker potential (TNC: 138-fold enrichment)

The ECM and granule gene panels, along with the composite signature and clustering results, can guide both mechanistic studies and future predictive modeling efforts in T-ALL and related leukemias. The "collagen fortress" destruction phenotype may represent a targetable vulnerability at the leukemia-microenvironment interface.

Data Availability

The RNA-seq data analyzed in this study are publicly available from the NCBI Gene Expression Omnibus under accession number GSE63602 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63602).

Acknowledgments

The author acknowledge the original depositors of the GSE63602 dataset (Sanghvi et al., Memorial Sloan Kettering Cancer Center) for making their data publicly available. Bioinformatics analyses were performed using Python (pandas, numpy, seaborn, matplotlib) and the Enrichr/gseapy platform for pathway enrichment.

References

Andersen, M. N., Etzerodt, A., Hamilton-Dutoit, S. J., and Moestrup, S. K. (2019). Collagen density regulates the activity of tumor-infiltrating T cells. Journal for Immunotherapy of Cancer, 7(1):68.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. (2000). Gene ontology: tool for the unification of biology. Nature Genetics, 25(1):25–29.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Research, 41(D1):D991–D995.

Borregaard, N., Sehested, M., Nielsen, B. S., Sengeløv, H., and Kjeldsen, L. (2010). Mechanisms of degranulation in neutrophils. Journal of Leukocyte Biology, 87(4):613–622.

Cairo, S., Armand-Ugon, M., Bhavsar, S., Bonardi, F., and Longo, L. (2020). The physiopathology of T-cell acute lymphoblastic leukemia: Focus on molecular aspects. Frontiers in Oncology, 10:273.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics, 14:128.

Coustan-Smith, E., Mullighan, C. G., Onciu, M., Behm, F. G., Raimondi, S. C., Pei, D., Cheng, C., Su, X., Rubnitz, J. E., Basso, G., et al. (2009). Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. The Lancet Oncology, 10(2):147–156.

Dai, Y.-T., Zhang, F., Fang, H.-P., Li, J.-F., Lu, G., Jiang, L., Chen, B., Mao, D.-D., Liu, Y.-F., Wang, J., et al. (2022). Transcriptome-wide subtyping of pediatric and adult T cell acute lymphoblastic leukemia in an international study of 707 cases. Proceedings of the National Academy of Sciences, 119(22):e2120787119.

Del Gaizo, M., Sergio, I., De Falco, V., Cillo, M., Di Marcantonio, M. C., Pietra, D., Ferretti, V., Zingoni, A., and Santoni, A. (2022). MicroRNAs as modulators of the immune response in T-cell acute lymphoblastic leukemia. Frontiers in Immunology, 12:792395.

Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics, 39(1): btac757.

Geyh, S., Rodríguez-Paredes, M., Jäger, P., Koch, A., Haas, R., Schaefer, C., and Giebel, B. (2021). The bone marrow microenvironment mechanisms in acute myeloid leukemia. Frontiers in Cell and Developmental Biology, 9:764698.

Giambra, V., Jesudason, E., Blackburn, J., Gerber, J. M., Jordan, C. T., and Bhalla, K. N. (2021). T-cell acute lymphoblastic leukemia: A roadmap to its main genetic drivers. Blood Cancer Discovery, 2(4):314–326.

Han, Y., Tian, T., Zhao, Y., Liu, X., Sun, Y., Hu, X., and Ma, F. (2021). Remodeling of bone marrow niches and roles of exosomes in leukemia. Annals of Translational Medicine, 9(3):268.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway knowledgebase. Nucleic Acids Research, 48(D1):D498–D503.

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research, 28(1):27–30.

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research, 44(W1):W90–W97.

McLeod, C., Gout, A. M., Zhou, X., Thrasher, A., Rahbarinia, D., Brady, S. W., Macias, M., Birsen, R., Gruber, T. A., et al. (2023). TALLSorts: a T-cell acute lymphoblastic leukemia subtype classifier using bulk RNA-seq data. Haematologica, 109(1):126–135.

Rawoof, A., Ahmad, I., and Srivastava, S. (2020). LeukmiR: a database for miRNAs and their targets in acute lymphoblastic leukemia. Database, 2020:baz151.

Sanghvi, V. R., Mavrakis, K. J., Van der Meulen, J., Voorhoeve, P. M., Krizan, K. A., Perciavalle, R. M., and Wendel, H.-G. (2014). Characterization of a network of tumor suppressor microRNAs in T cell acute lymphoblastic leukemia. Blood, 124(21):3389. Szczurek, E. and Beerenwinkel, N. (2022). Neutrophil phenotypes and functions in cancer: A consensus statement. Journal of Experimental Medicine, 219(6):e20220011.

Theocharides, A. P. A., Mancini, M., and Labropoulos, S. V. (2024). The interplay between extracellular matrix remodeling and cancer therapy resistance. Cancer Discovery, 14(8):1375–1392.

Wallaert, A., Van Loocke, W., Hernandez, L., Taber-sky, A., Eckhardt, S., Yigezu, B., Van Vlierberghe, P., and Speleman, F. (2017). Comprehensive miRNA expression profilingin human T-cell acute lymphoblastic leukemia by small RNA-sequencing. Scientific Reports, 7:7901.

Wang, J., Chen, L., Zhang, W., Liu, Y., and Xu, J. (2025). Collagen-disrupting attIL12 TIL therapy boosts deep T cell infiltration via dual signaling activation and CCKAR reduction in sarcomas. Proceedings of the National Academy of Sciences, 122(41):e2507542122.

Xie, Y., Zhang, J., Chen, M., and Li, P. (2022). Identification of microRNA editing sites in three subtypes of leukemia. Frontiers in Molecular Biosciences, 9:1014288.

Zhang, L., Wang, J., Chen, H., Li, M., and Xu, Y. (2024). Remodeling of the bone marrow microenvironment during acute myeloid leukemia. Annals of Translational Medicine, 12:42.