Personal:

My father's death from Hepatitis B Virus (HBV)-driven hepatocellular carcinoma, which remained undiagnosed until it had metastasized, inspired my interest in studying the earliest molecular changes that commit a cell towards cancer. As I learned more about oncology, I became increasingly engrossed by how malignancies like cancer often evolve silently for years before becoming clinically detectable. This pattern also repeated in my extended family: my grandmother's METex14-mutant non-small cell lung cancer was only discovered until she suffered a stroke, and my grandfather's MYD88 mutation that led to his Waldenström lymphoma, was only found after years of searching for the cause of his severe fatigue. As I read about these three cancers, I became intrigued by their specific molecular drivers—HBV as a carcinogen, MET and MYD88 mutations as driver mutations—which committed normal cells to a cancerous path. I was deeply drawn to identifying the biological point just before a patient's cells irreversibly transform to cancer, where therapeutic intervention could prevent disease rather than treat existing disease.

As I began my research journey and explored stem cell biology and cancer genomics, I was drawn to study acute myeloid leukemia (AML), which develops from a specific pre-malignant state called clonal hematopoiesis (CH). Over 200 million adults older than 65 years old worldwide carry CH hematopoietic stem cells which have acquired somatic mutations that "predispose" but not necessarily "commit" them to a malignant fate. While every AML case emerges through CH, most individuals with CH never develop leukemia. I quickly realized that this paradox made AML an ideal system: understanding the earliest cellular changes that distinguish a pre-malignant from a malignant cell could prevent cancer before it was too late. Moreover, my desire to study CH led me to develop scientific questions of how to identify the earliest molecular changes that mark a cell's commitment toward leukemia for both early diagnosis and treatment.

The most notable societal significance of my project lies in closing this gap, with impacts including earlier diagnosis and preventative treatment of acute myeloid leukemia. By identifying cell-surface markers that distinguish pre-leukemic stem cells from healthy stem cells, my work could shift cancer care from late-stage intervention to actionable, early-stage prevention. The markers I have identified would allow clinicians to detect dangerous cellular changes years earlier, monitor disease evolution non-invasively, and develop therapies that eliminate only the cells that are committed to malignancy. Ultimately, therapeutic agents that prevent the clonal hematopoietic population from becoming malignant (for example, by using an antibody to a surface marker specific to CH cells and thus deplete or kill those cells) would help those patients avoid AML.

This work has broader ramifications beyond early detection and preventative treatment of patients at high risk for AML, specifically shedding light into other tumor types which may also be driven by stem cells, including brain, lung, breast, colon, pancreatic, prostate, and ovarian (Marzagalli et al., 2021). AML is the paradigm for other cancer stem cells, where the stem cell is

at the apex of a cellular hierarchy and both initiates and maintains disease (Velten et al., 2021). Thus, the insights gained from this research in detecting and characterizing early cellular shifts could inform many other prevention strategies across multiple tumor types.

Finally, this work has important implications for monitoring and preventing relapse in AML patients. A subset of patients with AML who have gone into complete remission after treatment still experience relapse, even two decades later, because of the capabilities of pre-LSCs in reinitiating disease (Patel et al., 2021). If these rare stem cell populations containing mutations can be identified and monitored using the surface markers identified in this study, relapse could be detected at its earliest molecular stage before overt clinical progression.

Over the past two years, I have dedicated several hundred hours to this research, including conducting (1) large-scale computational analysis that has involved data processing, differential expression analysis, and identifying candidate targets and (2) wet-lab experimentation that has involved experimental design, CRISPR/Cas9 genome editing, multicolor flow cytometry, and rigorous analysis of quantitative results. This work was conducted at the Stanford Institute for Stem Cell Biology and Regenerative Medicine under the supervision of Asiri Ediriwickrema, Thomas Koehnke, and Ravi Majeti.

Research: Please see attached draft research paper

Identification of Novel Targetable Surface Markers on Pre-Leukemic Stem Cells for Early Detection and Therapeutic Intervention of Acute Myeloid Leukemia

Camille Chu

Abstract

Acute myeloid leukemia (AML) is an aggressive hematologic malignancy caused by rare but highly malignant, self-renewing leukemic stem cells (LSCs) that drive cancer progression and resist chemotherapy. In normal hematopoiesis, multipotent hematopoietic stem cells (HSCs) differentiate into mature blood cells. However, in AML, a small number of HSCs acquire driver mutations that cause the abnormal expansion of HSC-derived clones that become pre-leukemic stem cells (pre-LSCs) in the aberrant process of clonal hematopoiesis. These pre-LSCs are predisposed but not yet committed to a cancerous fate. A critical, ongoing problem in the AML field is to better understand the transformation from non-cancerous pre-LSC to malignant LSCs which divide to form leukemic blasts. Currently, in AML research, pre-LSCs cannot be reliably distinguished from healthy HSCs using surface markers, thus making it challenging to detect early disease, track malignant evolution, or selectively eliminate pre-leukemic clones before leukemia develops. In addition, most therapeutics target bulk leukemia cells that are transitioning into blasts, instead of the rare stem cell-like populations responsible for disease initiation and relapse. In this study, novel cell-surface markers uniquely enriched on pre-LSCs are identified and validated. These include CD267 and CD1A, both surface proteins which have been reported to be involved in aspects of hematopoeisis. By enabling disease diagnosis and targeted therapies that specifically eliminate pre-LSCs before they become LSCs, these markers can help prevent high-risk patients from developing AML.

1. Introduction

AML is an aggressive blood cancer of the bone marrow defined by the uncontrolled proliferation of immature myeloid cells that impacts nearly 150,000 people worldwide and has a five-year survival rate of only 30% (Cleveland Clinic, 2023). Although advances have been made in chemotherapy and stem cell transplantation to address leukemia, approximately 50% of AML patients who achieve an initial remission still relapse within two years (Mühleck et al., 2022). Both disease initiation and relapse are driven by leukemic stem cells (LSCs), which are a rare population of malignant cells capable of both self-renewal and disease progression (Majeti et al., 2007).

AML development begins long before clinical diagnosis through the process of clonal hematopoiesis, in which hematopoietic stem cells (HSCs) acquire somatic mutations—most commonly in epigenetic regulators such as *DNMT3A* and *TET2*—that confer a competitive advantage and lead to clonal expansion to pre-leukemic stem cells or pre-LSCs (Shlush et al., 2014). In fact, 10-20% of healthy adults over the age of 70 acquire these mutated but non-malignant cells. Moreover, pre-LSCs can persist for years before additional mutations are acquired that permanently commit them to their leukemic stem cell fate, as illustrated in **Figure 1** (Barreto et al., 2022).

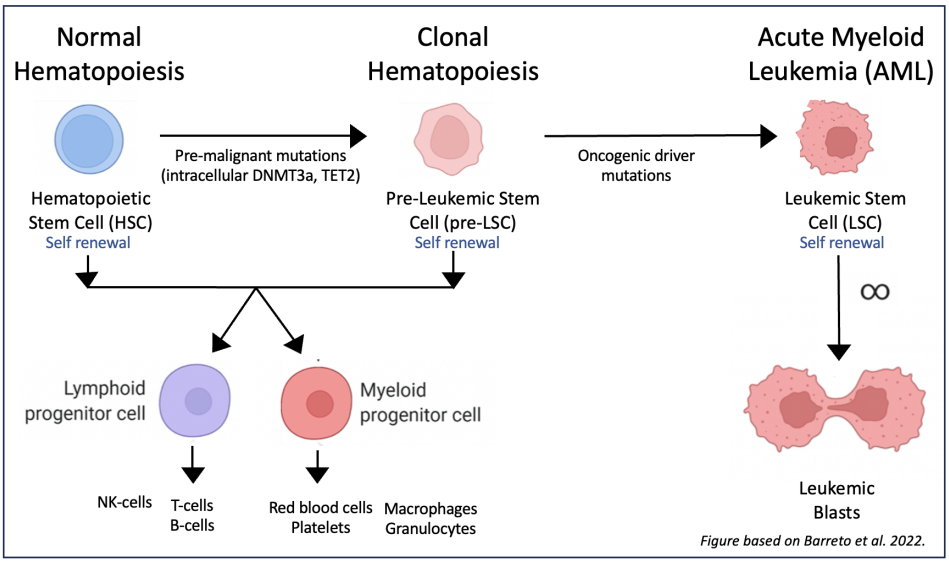

Figure 1: Normal Hematopoiesis: pluripotent hematopoietic stem cells, HSCs, can mature into various blood cell types. **Clonal Hematopoiesis**: intracellular mutations cause HSCs to become pre-leukemic stem cells, pre-LSCs, which can still have a normal cell fate but are predisposed towards becoming malignant LSCs. **Acute Myeloid Leukemia**: malignant LSCs infinitely divide to form leukemic blasts, causing AML.

A critical barrier in the advancement of AML research is the lack of reliable surface markers that can distinguish pre-LSCs from normal HSCs. Because pre-LSCs have many normal stem cell properties, they are currently difficult to isolate and therapeutically target. Most therapeutics that treat AML only target the bulk leukemic cells, allowing for

rare stem-like populations of pre-LSCs and LSCs to survive treatment and drive relapse (Stelmach & Trumpp, 2023).

This research addresses this challenge by investigating whether pre-LSCs express distinct cell-surface markers that can be identified through transcriptomic analysis and validated experimentally through flow cytometry. The identification of such markers would have significant clinical and societal impact by facilitating the development of preventative therapeutic strategies for clonal hematopoiesis patients. This would both increase the survival of AML patients and contribute to a better understanding of the overall field of cancer stem cells for other tumor types.

This study specifically explores the cell-surface proteins that define pre-LSCs. It has been shown that pre-leukemic somatic mutations in the intracellular epigenetic regulators *DNMT3A* and *TET2* result in chromatin remodeling associated with clonal hematopoiesis (Sato et al., 2016). The working hypothesis is that these mutations lead to altered transcriptional programs, including cell-surface protein expression on pre-LSCs, as illustrated in **Figure 2A**. Therefore, if pre-LSCs consistently express specific cell-surface proteins compared to normal HSCs, these markers could serve as effective targets for isolating pre-LSCs and enabling early diagnosis of AML. These findings would also allow for more effective therapeutic prevention of additional oncogenic driver mutations and clonal hematopoiesis progression to malignant LSCs, as shown in **Figure 2B**.

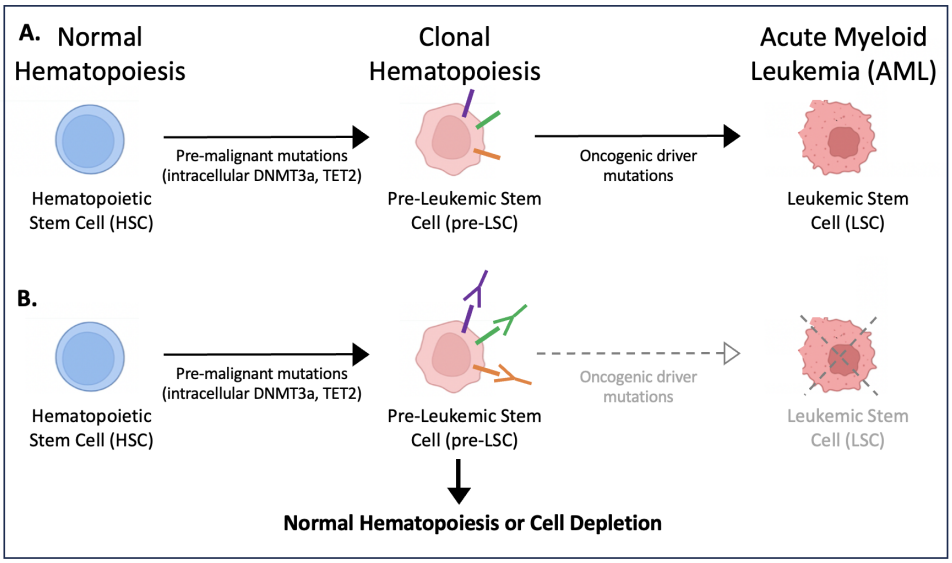

Figure 2: A. Hypothesis that the intracellular mutations that cause Clonal Hematopoiesis result in **cell surface markers** (purple, green, orange lines) to be displayed on pre-LSCs. **B.** Hypothesis that the identification of pre-LSC-specific cell surface markers will allow early detection of pre-LSCs as well as possible therapeutic intervention, pushing pre-LSCs towards normal hematopoiesis or depleting pre-LSCs so malignant LSCs never form.

2. Research Questions and Hypotheses

Two major research questions guided this study:

1. Do pre-leukemic hematopoietic stem cells express a distinct set of cell-surface markers compared with normal HSCs?
2. Can these markers be validated experimentally in patient cells with clonal hematopoiesis or genetically engineered CD34+ cells with pre-leukemic mutations?

The central hypothesis of this research is that pre-LSCs undergoing clonal hematopoiesis, as characterized by mutations in intracellular proteins *DNMT3A* and *TET2*, also express distinct surface markers that could significantly aid in the diagnosis and treatment of AML when compared with normal HSCs. Expected outcomes include identifying candidate surface markers enriched in pre-LSCs using single-cell and bulk transcriptomic analysis, validating these markers in patient-derived cells and CRISPR-engineered human CD34+ stem cells using multicolor flow cytometry, and generating a list of clinically useful surface markers that could enable early detection of AML.

3. Materials and Methods

This study combined extensive single-cell and bulk transcriptomic analysis with experimental validation using primary human hematopoietic stem cells and genetically engineered pre-leukemic models. The workflow was designed to identify mutation-associated surface markers in pre-leukemic stem cells (pre-LSCs) and rigorously validate their protein expression using high-parameter flow cytometry and other technical assays.

3.1 Computational Identification of Novel Surface Markers

The first part of this project involved extensive computational analysis of single-cell and bulk RNA sequencing datasets to identify surface proteins selectively enriched in pre-leukemic stem cells (pre-LSCs). Single-cell and bulk RNA sequencing datasets of human clonal hematopoiesis and aging hematopoietic stem and progenitor cells were obtained from published studies (Jakobsen et al., 2024; Nam et al., 2022). These datasets included hematopoietic stem cells with mutations in the intracellularly-expressed *DNMT3A* or *TET2*, which are the most commonly acquired mutations in clonal hematopoiesis and AML, as well as genetically wild-type HSC controls spanning multiple donor age groups.

All computational analyses were performed using R and the Seurat package within RStudio. Filtering through quality control was first conducted on raw count matrices of the single-cell and bulk RNA sequencing data to remove low-quality cells, potential doublets, and cells with high mitochondrial gene expression, as modeled in **Figure 3**. Cells were filtered based on number of detected genes (nFeature_RNA), total UMI counts (nCount_RNA), and percentage of mitochondrial transcripts. After normalization and scaling, highly variable genes were identified and principal component analysis was

performed using the gene set to reduce dimensionality and identify major sources of variability. Plots and heatmaps of the top gene contributors across principal components were created to evaluate the number of principal components for downstream analysis.

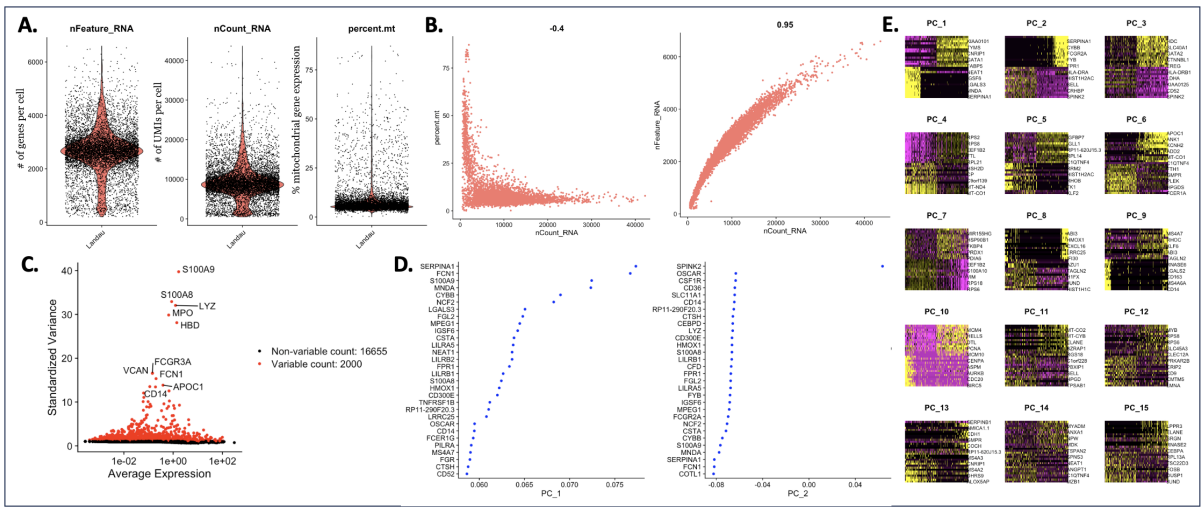

Figure 3: Computational preprocessing and dimensionality reduction workflow for single-cell RNA sequencing analysis. A. Quality control filtering of raw sequencing data based on number of detected genes (nFeature_RNA), total UMI counts (nCount_RNA), and percentage of mitochondrial transcripts (percent.mt) to remove low-quality cells and potential doublets. **B.** Plot of QC feature relationships before filtering. **C.** Identification of highly variable genes using standardized variance across average expression levels. **D.** Principal component analysis performed on scaled expression data. **E.** Heatmaps of top genes across principal components to determine biological signal and PCs for downstream analysis.

Using these principal components, unsupervised graph clustering was performed and distinct cell populations were visualized using Uniform Manifold Approximation and Projections (UMAPs), as modeled in **Figure** 4. Cluster identities were assigned based on established and well-characterized lineage-specific marker genes and visualized with violin plots, feature plots, and annotated UMAP projections. Differential gene expression analysis was subsequently performed between mutant HSC groups (*DNMT3A* and *TET2*) and wild-type HSC populations across multiple independent comparisons stratified by mutation type and donor age. For each gene, the average log2 fold changes (magnitude of differential expression), adjusted p-values (statistical significance of expression), percentage of cells expressing each gene, absolute difference in expression percentage between groups, and additional statistics were calculated. In addition, these genes were visualized through heatmaps of top differentially expressed genes and volcano plots summarizing the fold changes and p-values. To ensure the robustness and biological relevance of the genes, thresholds for filtering were set, including an adjusted p-value less than 0.05, average log2 fold change greater than 1, and a percent expression difference greater than 10%. Thus, genes that met these predefined statistical and biological thresholds were prioritized for downstream filtering, such as surface protein annotation and cross-dataset validation. This enabled the identification of high-confidence candidate markers for experimental assays.

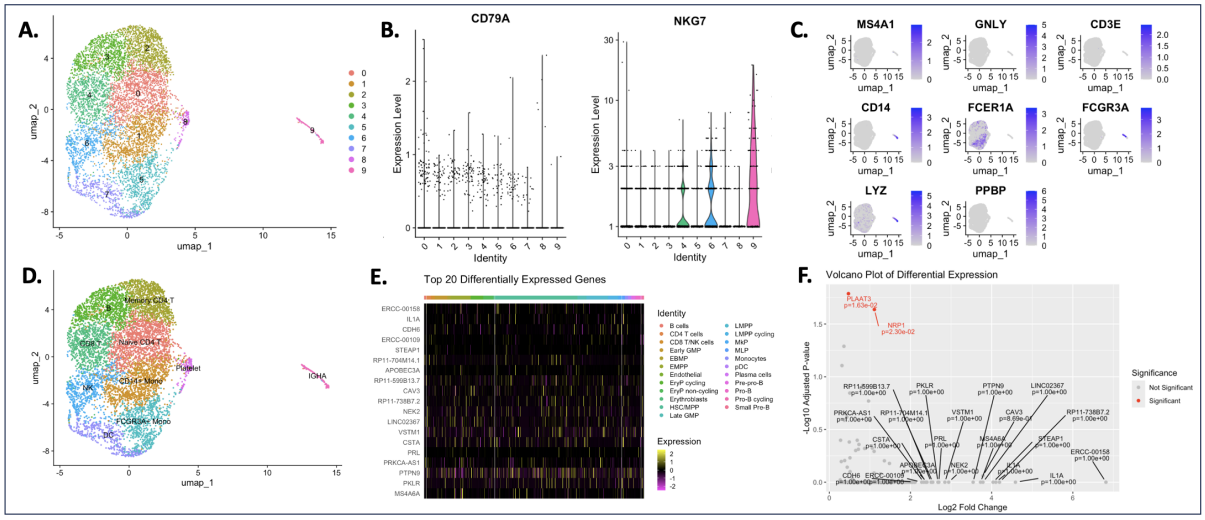

Figure 4: Clustering, annotation, and differential expression analysis workflow. **A.** UMAP visualization of transcriptionally distinct clusters identified after PCA and clustering. **B, C.** Violin and feature plots of markers for cluster annotation. **D.** UMAP with annotated cell identities based on established lineage-specific markers. **E.** Heatmap of top differentially expressed genes across clusters. **F.** Volcano plot of differential expression, displaying log2 fold change versus adjusted p-value, with significant genes (in red) meeting statistical and biological thresholds for downstream filtering.

To prioritize therapeutically actionable targets, differentially expressed genes were further filtered with the Surface Protein Annotation Tool (SPAT), which scores proteins based on their predicted membrane localization, extracellular accessibility, and suitability for antibody targeting. Only genes with SPAT scores greater or equal to 5 were included in further analysis. Candidate markers were then cross-referenced across all mutation types, age groups, and datasets to identify genes that were consistently enriched in the pre-LSCs across varying biological conditions. The resulting genes were prioritized and ranked by absolute log2 fold change (which ranged from 1.5 to 8-fold), statistical significance (p-value < 0.05), and additional criteria such as the availability of antibodies for experimental analysis. This multifaceted computational pipeline allowed for the identification of a refined set of candidate surface markers that were both biologically meaningful and experimentally actionable.

3.2 Experimental Validation

The second step of this project focused on rigorous experimental validation of candidate surface markers using primary human hematopoietic stem cells and genetically engineering pre-leukemic cells. Primary human CD34+ hematopoietic stem and progenitor cells (HSPCs) were obtained through institutional approval, including samples derived from individuals with early-stage clonal hematopoiesis harboring *DNMT3A* mutations, and were expanded under stem cell-supportive culture conditions. These samples represent clinically relevant pre-leukemic states that precede overt AML and are particularly challenging to study due to the rarity and phenotypic similarity of pre-LSCs to normal HSCs.

CRISPR/Cas9-edited human CD34$^+$ HSPCs with knockout mutations in *DNMT3A* or *TET2* were independently generated to establish an isogenic model of pre-leukemic mutations. Paired sgRNA sequences targeting *DNMT3A* (5′-GAGCCAGAGTACGAGGTGAG-3′ and 5′-GCCCGTGGGGTCCGATGCTG-3′) and *TET2* (5′-GCTCATGTGCAGTCACTGTG-3′ and 5′-CAAGTGCTGTTTCAACACTG-3′) were used to create the frameshift knockout mutations. Additionally, a sgRNA pair targeting *AAVS1* (5′-ATATGTCCCAGATAGCACTG-3′ and 5′-GGACGCACCATTCTCACAAA-3′) was used as the neutral control in the experiment. Editing efficiency and sample quality were assessed by calculating the indel frequencies, and genotyping was confirmed using digital droplet PCR. Only samples that met predefined quality and editing-efficiency thresholds were advanced for downstream phenotypic analysis, which ensured that observed differences in surface marker expression could be solely attributed to mutation status rather than technical variability or noise.

Surface protein expression was assessed using high-parameter multicolor flow cytometry. Customized antibody panels with over twenty markers distributed across more than ten fluorescent channels were designed, optimized, and then executed. The panel design required careful consideration of fluorophore compatibility, antigen density, and compensation to allow for accurate detection of low-abundance surface proteins on these rare stem cell populations. The fluorophore-conjugated antibodies targeting candidate surface markers identified computationally were incorporated into these panels for immunophenotyping.

Flow cytometry analysis was conducted following a stringent, multistep gating strategy to ensure the precise isolation of primitive stem cell populations. Cells were first gated to exclude debris based on forward and side scatter, and then subsequently gated to remove doublets and multiplets (FSC-H vs FSC-W and SSC-H vs SSC-W). Non-viable cells were excluded using DAPI staining to restrict the analysis to live cells, and mature lineage-positive populations (CD3+, CD19+, CD20+, CD33+) were removed to enrich for primitive stem/progenitor cells. Within this lineage-depleted population, CD34 and EPCR expression were used to define the CD34+EPCR+ compartment. CD34 is a well-established marker of hematopoietic stem and progenitor cells, while EPCR enriches for a more primitive, functional subset of the CD34+ population (Anjos-Afonso & Bonnet, 2023). Focusing on this CD34+EPCR+ population thus allowed analysis of mutation-associated phenotypic changes within an early progenitor context that is relevant to clonal hematopoiesis and leukemic progression.

To ensure accurate interpretation of marker expression, fluorescence minus one (FMO) controls and unstained controls were included for each experiment. Candidate marker expression was quantified using normalized median fluorescence intensity (MFI), calculated relative to the corresponding FMO control to account for background fluorescence and technical noise. For visualization, histograms of candidate marker expression in mutant and control samples were overlaid directly with FMO controls, which allowed for the precise assessment of shifts in expression. Fluorescence-activated cell sorting (FACS) was performed to isolate defined stem cell populations, and post-sort purity checks were

conducted to confirm sorting accuracy. Droplet digital PCR was used to validate mutation status in sorted populations and link surface phenotype directly to genotype.

Following the experimentation, flow cytometry data was analyzed using FlowJo software. Marker expression patterns were evaluated across multiple independent experiments, biological replicates, mutation types, and model systems. Markers with consistent enrichment in pre-leukemic stem cell populations across the CRISPR-engineered models were considered high-confidence candidates.

4. Results and Analysis

To determine whether early leukemic mutations in epigenetic regulators confer a reproducible and therapeutically targetable phenotype on normal hematopoietic stem cells, a two-part research strategy was implemented: (1) computational identification of candidate surface markers enriched in mutation-associated HSCs, and (2) experimental validation of markers in CRISPR-engineered isogenic HSPCs and primary samples.

Applying the computational pipeline described above, a high-confidence set of surface markers enriched in *DNMT3A* and *TET2* mutant HSCs were identified. Among these surface proteins, CD1A, FLT4 (VEGFR3), TACI (CD267), and CD5 demonstrated consistent enrichment across mutation types and independent donor datasets, as shown in **Figure 5**. These genes showed statistically significant increases in both magnitude of expression (average log2 fold change) and proportion of expressing cells (percent difference), with associated p-values and adjusted p-values indicating their statistical significance. Thus, these four markers exhibit strong enrichment in mutant HSC populations in both *DNMT3A* and *TET2* samples, demonstrating their potential as markers for downstream experimental analysis.

Gene	Protein name	Bulk Average Log2 Fold Change	DNMT3A Average Log2 Fold Change	DNMT3A Percent Difference	DNMT3A P-value	DNMT3A P-value Adjusted	TET2 Average Log2 Fold Change	TET2 Percent Difference	TET2 P-value	TET2 P-value Adjusted
CD1A	CD1a	4.577	4.447	0.168	1.582e-07	6.480E-3	6.283	0.202	7.760e-12	3.178e-07
FLT4	VEGFR3	7.760	3.004	0.100	8.673e-07	3.551E-2	4.119	0.234	1.025e-17	4.196e-13
TNFRSF13B	TACI (CD267)	4.267	1.439	0.086	2.035e-07	8.334E-3	3.159	0.125	1.973e-07	8.077E-3
CD5	CD5	4.473	1.174	0.003	2.408e-11	9.861e-07	1.142	0.015	1.009e-08	4.133E-4

Figure 5: Statistical metrics of candidate surface markers based on computational analysis of single-cell and bulk RNA sequencing datasets. Bulk average log2 fold change reflects overall differences in marker expression across all hematopoietic stem and progenitor cells. Mutation-specific statistics for *DNMT3A* and *TET2* summarize average log2 fold changes (magnitude of expression difference), percent differences in expression (proportion of cells expressing the marker), p-values (statistical significance), and adjusted p-values (corrected significance) for candidate markers CD1A, FLT4, TACI, and CD5, which were the genes prioritized for downstream experimental validation.

To determine whether transcription enrichment translated into detectable protein-level changes (rather than donor-level changes in the primary samples), CD34+ HSPCs with CRISPR/Cas9-induced knockout mutations in *DNMT3A* and *TET2* were generated. *AAVS1*-edited cells served as the neutral controls to directly attribute phenotypic changes to mutation status rather than variability between samples. The CRISPR editing efficiencies, characterized by the indel frequency, ranged up to 87%, with low-quality samples being omitted from further analysis.

High-parameter multicolor flow cytometry panels were designed to examine primitive hematopoietic stem and progenitor compartments with specificity. As shown in **Figure 6**, sequential flow cytometry gating was performed using a gating strategy described above to accurately isolate the CD34+EPCR+ population.

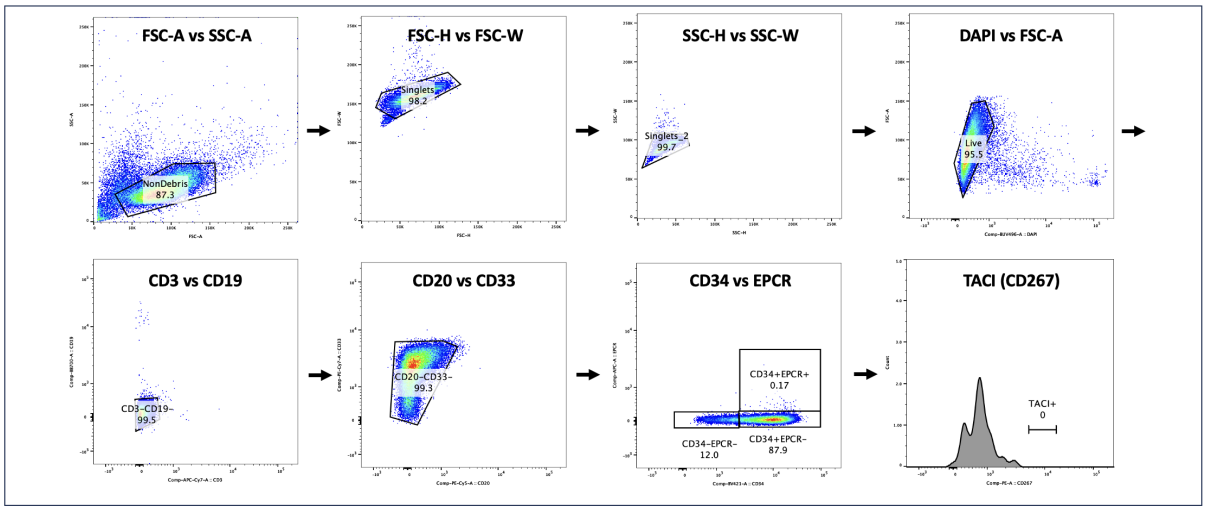

Figure 6: Representative **flow cytometry gating strategy** for isolation and analysis of **CD34+EPCR+ cells**. Cells were first gated to exclude debris based on forward and side scatter, followed by doublet and multiplet discrimination. Dead cells were excluded using DAPI staining to define the live population. Lineage-positive cells were removed by gating CD3-CD19- and CD20-CD33- populations. CD34 and EPCR expression were then used to identify the CD34+EPCR+ stem/progenitor population. Expression of the TACI marker was subsequently assessed within the CD34+EPCR+ gate. Percentages shown indicate the proportion of events within each parent gate. This gating strategy was applied across all samples and markers.

The flow cytometry histograms illustrate the expression patterns of candidate surface markers in CD34+EPCR+ cells across *AAVS1* control, *DNMT3A* knockout, and *TET2* knockout samples, overlaid with the fluorescence-minus-one (FMO) controls to ensure accurate discrimination of shifts in marker expression. Median fluorescence intensity (MFI) was calculated in the CD34+EPCR+ gate for each sample, representing the midpoint of signal intensity for the population of cells, and the value was normalized to the FMO median to quantify marker expression shifts attributed to mutation status.

Several markers demonstrate consistent increases in both (1) the percentage of marker-positive cells and (2) MFI in knockout cells compared to controls, as illustrated in **Figure 7**. In particular, CD1A+ frequency increased from 2.8% to 3.9% from the control and its

normalized MFI increased from 5.56 in *AAVS1* control to 6.68 in *TET2* KO. Similarly, CD267(TACI)+ frequency increased from 0% to 2.40% from the control and normalized MFI increased from 6.76 to 7.37. These results suggest that both intensity and frequency changes for cell-surface proteins CD1A and CD267 are associated with *TET2* loss. These consistent right-shifted fluorescence distributions and increased frequencies of marker positive cells indicate mutation-associated increases in surface protein expression within the CD34+EPCR+ compartment. Notably, these changes reflect both population-level shifts in expression intensity and expansion of marker-positive populations, suggesting that epigenetic disruption alters the phenotypic composition of the progenitor population.

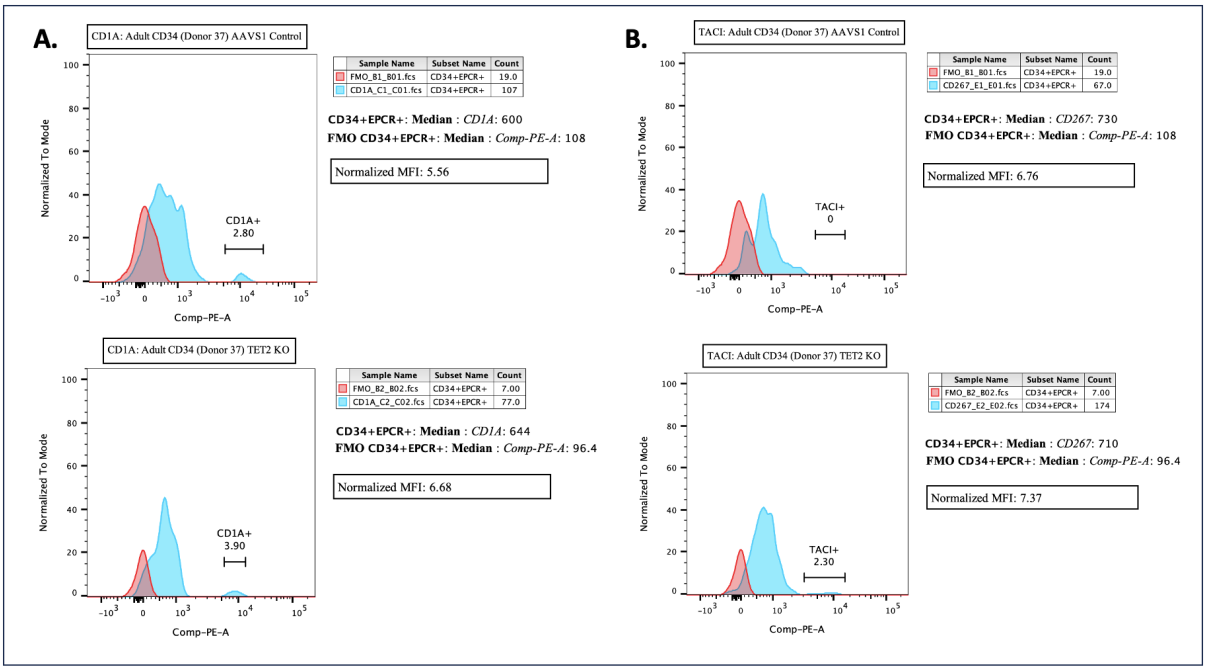

Figure 7: A. Flow cytometry histograms showing **CD1A** expression and **B. TACI** expression in CD34+EPCR+ cells from *AAVS1* control and *TET2* knockout samples. Histograms are overlaid with the corresponding fluorescence-minus-one control (red) to define background signal. Median fluorescence intensity was calculated in the CD34+EPCR+ gate and normalized to the FMO median to quantify subtle shifts in marker expression, with the determined MFI values shown for each sample.

The candidate markers identified computationally and validated experimentally are notable not only for their strong enrichment in mutant HSPCs but also for their established or putative biological roles in hematopoiesis and immune signaling. For instance, CD1A—a transmembrane MHC class I-like glycoprotein that helps present lipid antigens to T cells—has been implicated in early hematopoietic differentiation and may reflect changes in lineage priming associated with epigenetic disruption (Sun et al., 2023). CD267 (TACI/TNFRSF13B) is an alternatively spliced receptor. Initial evidence suggests that its short and long forms may be expressed at different times during normal hematopoiesis and have different roles (Garcia-Carmona et al., 2018). The short form, TACI-s, is involved in B-cell maturation and survival signaling, suggesting that alterations in TACI-s signaling pathways are detectable at the pre-leukemic stage. The long form, TACI-l, is speculated to

promote B cell apoptosis, suggesting increased TACI-l expression could shift pre-LSCs away from normal differentiation and towards LSC. Collectively, these markers demonstrate that *DNMT3A* and *TET2* mutant pre-LSCs exhibit functionally relevant phenotypic differences from wild-type HSCs, linking early epigenetic mutations with the acquisition of targetable surface phenotypes. These results establish a foundation for both prospective detection of pre-leukemic clones and exploration of their functional biology.

5. Conclusion

This study demonstrates that pre-leukemic hematopoietic stem cells (pre-LSCs) have distinct surface phenotypes that can be used to identify cells at the earliest stages of leukemic progression in clonal hematopoiesis. Combining integrated transcriptomic analysis with experimental validation in engineered and patient systems led to the identification of several novel surface markers, including CD1A and CD267, that are selectively enriched in *DNMT3A* and *TET2* mutant HSCs. These findings show that pre-leukemic states are characterized by not only intracellular mutations but also distinct extracellular phenotypes, providing measurable and therapeutically targetable features for early disease. By demonstrating that early epigenetic, pre-cancerous mutations can generate detectable cellular phenotypic changes prior to overt leukemic transformation to malignant LSCs, this work contributes to the biological understanding of leukemic evolution. Most importantly, the identification of these markers establishes a framework for earlier detection and treatment of high-risk hematopoiesis, offering a basis for early intervention and preventive therapeutic strategies before the development of acute myeloid leukemia. Beyond the immediate clinical applications, this work sets a precedent for a combined computational and experimental approach to discover and evaluate actionable phenotypes in rare stem cell populations in other cancers.

Next steps to expand this research include further functional validation with in vitro and in vivo experimentation. For instance, colony-forming unit (CFU) assays could be used to assess whether marker-positive pre-LSC populations show altered differentiation potential, which would further validate this research. Xenotransplantation experiments could determine if these pre-LSC populations exhibit increased self-renewal, clonal fitness, or progression toward malignancy in vivo. Finally, applying this integrated computational and experimental framework to other malignancies driven by cancer stem cells, such as pre-malignant mammary stem cells in breast cancer, could be impactful in better understanding the early stages of cancer development across various tissue types. Ultimately, translating these findings to antibody-based targeting and selective depletion of high-risk pre-leukemic subsets could transform clinical practices in early intervention and treatment of disease.

6. References

Anjos-Afonso, F., & Bonnet, D. (2023). Human CD34+ hematopoietic stem cell hierarchy: how far are we with its delineation at the most primitive level? *Blood*, *142*(6), 509–518. https://doi.org/10.1182/blood.2022018071

American Cancer Society. (2025). What is acute myeloid leukemia (AML)? https://www.cancer.org/cancer/types/acute-myeloid-leukemia/about/what-is-aml.html

Barreto, I. V., Pessoa, F. M. C. d. P., Machado, C. B., Pantoja, L. d. C., Ribeiro, R. M., Lopes, G. S., Amaral de Moraes, M. E., de Moraes Filho, M. O., de Souza, L. E. B., Burbano, R. M. R., Khayat, A. S., & Moreira-Nunes, C. A. (2022). Leukemic stem cell: A mini-review on clinical perspectives. *Frontiers in Oncology*, *12*, 931050. https://doi.org/10.3389/fonc.2022.931050

Cleveland Clinic. (2023). Acute myeloid leukemia (AML): Symptoms, treatment & prognosis. https://my.clevelandclinic.org/health/diseases/6212-acute-myeloid-leukemia-aml

Garcia-Carmona, Y., Ting, A. T., Radigan, L., Athuluri Divakar, S. K., Chavez, J., Meffre, E., Cerutti, A., & Cunningham-Rundles, C. (2018). TACI Isoforms Regulate Ligand Binding and Receptor Function. *Frontiers in Immunology*, *9*, 2125. https://doi.org/10.3389/fimmu.2018.02125

Jaiswal, S., & Ebert, B. L. (2019). Clonal hematopoiesis in human aging and disease. *Science (New York, N.Y.)*, *366*(6465), eaan4673. https://doi.org/10.1126/science.aan4673

Jakobsen, N. A., Turkalj, S., Zeng, A. G. X., Stoilova, B., Metzner, M., Rahmig, S., Nagree, M. S., Shah, S., Moore, R., Usukhbayar, B., Angulo Salazar, M., Gafencu, G. A., Kennedy, A., Newman, S., Kendrick, B. J. L., Taylor, A. H., Afinowi-Luitz, R., Gundle, R., Watkins, B., Wheway, K., … Vyas, P. (2024). Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging. *Cell Stem Cell*, *31*(8), 1127–1144. https://doi.org/10.1016/j.stem.2024.05.010

Majeti, R., Park, C. Y., & Weissman, I. L. (2007). Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell*, *1*(6), 635–645. https://doi.org/10.1016/j.stem.2007.10.001

Mühleck, R., Scholl, S., Hilgendorf, I., Schrenk, K., Hammersen, J., Frietsch, J. J., Fleischmann, M., Sayer, H. G., Glaser, A., Hochhaus, A., & Schnetzke, U. (2022).

Outcome of patients with relapsed or refractory acute myeloid leukemia treated with Mito-FLAG salvage chemotherapy. *Journal of Cancer Research and Clinical Oncology*, *148*(9), 2539–2548. https://doi.org/10.1007/s00432-021-03821-1

Nam, A. S., Dusaj, N., Izzo, F., Murali, R., Myers, R. M., Mouhieddine, T. H., Sotelo, J., Benbarche, S., Waarts, M., Gaiti, F., Tahri, S., Levine, R., Abdel-Wahab, O., Godley, L. A., Chaligne, R., Ghobrial, I., & Landau, D. A. (2022). Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation. *Nature Genetics*, *54*(10), 1514–1526. https://doi.org/10.1038/s41588-022-01179-9

Sato, H., Wheat, J. C., Steidl, U., & Ito, K. (2016). DNMT3A and TET2 in the Pre-Leukemic Phase of Hematopoietic Disorders. *Frontiers in Oncology*, *6*, 187. https://doi.org/10.3389/fonc.2016.00187

Shlush, L. I., Zandi, S., Mitchell, A., Chen, W. C., Brandwein, J. M., Gupta, V., Kennedy, J. A., Schimmer, A. D., Schuh, A. C., Yee, K. W., McLeod, J. L., Doedens, M., Medeiros, J. J. F., Marke, R., Kim, H. J., Lee, K., McPherson, J. D., Hudson, T. J., HALT Pan-Leukemia Gene Panel Consortium, ... Dick, J. E. (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature*, *506*(7488), 328–333. https://doi.org/10.1038/nature13038

Stelmach, P., & Trumpp, A. (2023). Leukemic stem cells and therapy resistance in acute myeloid leukemia. *Haematologica*, *108*(2), 353–366. https://doi.org/10.3324/haematol.2022.280800

Sun, L., Su, Y., Jiao, A., Wang, X., & Zhang, B. (2023). T cells in health and disease. *Signal Transduction and Targeted Therapy*, *8*(1), 235. https://doi.org/10.1038/s41392-023-01471-y

Surface Protein Annotation Tool. (n.d.). https://spat.leucegene.ca