Precise single-cell transcriptomic mapping of normal and leukemic cell states reveals unconventional lineage priming in acute myeloid leukemia

Authors: Andy G.X. Zeng[1,2], Ilaria Iacobucci[3]†, Sayyam Shah[1]†, Amanda Mitchell[1], Gordon Wong[1,2], Suraj Bansal[1], Qingsong Gao[3], Hyerin Kim[1,2], James A. Kennedy[4], Mark D. Minden[1,5,6,7], Torsten Haferlach[8], Charles G. Mullighan[3,9], John E. Dick[1,2]

Affiliations:

[1]Princess Margaret Cancer Centre, University Health Network; Toronto, ON, Canada

[2]Department of Molecular Genetics, University of Toronto; Toronto, ON, Canada

[3]Department of Pathology, St Jude Children's Research Hospital, Memphis, TN, USA

[4]Division of Medical Oncology and Hematology, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

[5]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

[6]Department of Medicine, University of Toronto, Toronto, ON, Canada

[7]Division of Medical Oncology and Hematology, University Health Network, Toronto, ON, Canada

[8]MLL Münchner Leukämielabor GmbH, München, Germany

[9]Center of Excellence for Leukemia Studies, St. Jude Children's Research Hospital, Memphis, TN

† contributed equally

Abstract

Initial classification of acute leukemia involves the assignment of blasts to cell states within the hematopoietic hierarchy based on morphological and immunophenotypic features. Yet, these traditional classification approaches lack precision, especially at the level of immature blasts. Single-cell RNA-sequencing (scRNA-seq) enables precise determination of cell state using thousands of markers, thus providing an opportunity to re-examine present-day classification schemes of acute leukemia. Here, we developed a detailed reference map of human bone marrow hematopoiesis from 263,519 single-cell transcriptomes spanning 55 cellular states. Cell state annotations were benchmarked against purified cell populations, and in-depth characterization of gene expression programs underlying hematopoietic differentiation was undertaken. Projection of single-cell transcriptomes from 175 samples spanning acute myeloid leukemia (AML), mixed phenotype acute leukemia (MPAL), and acute erythroid leukemia (AEL) revealed 11 subtypes involving distinct stages of hematopoietic differentiation. These included AML subtypes with notable lymphoid or erythroid lineage priming, challenging traditional diagnostic boundaries between AML, MPAL, and AEL. Quantification of lineage priming in bulk patient cohorts revealed specific genetic alterations associated with this unconventional lineage priming. Integration of transcriptional and genetic information at the single-cell level revealed how genetic subclones can induce lineage restriction, differentiation blocks, or expansion of mature myeloid cells. Furthermore, we demonstrate that distinct cellular hierarchies can co-exist within individual patients, providing insight into AML evolution in response to varying selection pressures. Together, precise mapping of hematopoietic cell states can serve as a foundation for refining disease classification in acute leukemia and understanding response or resistance to emerging therapies.

Introduction

Hematopoietic differentiation is perturbed in acute leukemia, yet leukemia cells still retain features of hematopoietic differentiation. These features play a critical role in initial disease classification, wherein morphological and immunophenotypic markers are used to assign leukemic blasts to broad hematopoietic lineages [1]. Yet, these markers lack precision in discerning between specific cell states within a hematopoietic lineage, particularly at the level of immature blasts. Single-cell RNA sequencing (scRNA-seq) provides thousands of new markers to enable precise determination of leukemia cell state. This provides an opportunity to refine our existing classification of acute leukemia.

This is particularly important within acute myeloid leukemia (AML) due to extensive variation in leukemic blast involvement within the myeloid lineage. It is well established that AML cells within individual patients form cellular hierarchies spanning multiple hematopoietic differentiation stages, sustained by primitive leukemia stem cells (LSCs) at the apex. Pioneering studies mapping AML cells to early scRNA-seq landscapes of human hematopoiesis have revealed variation in leukemia cell hierarchies between patients [2,3] and subsequent work inferring AML hierarchy composition from bulk patient cohorts has linked this variation to survival, relapse, and drug sensitivity [4,5]. With cellular hierarchies emerging as a clinically relevant approach for understanding heterogeneity in AML, there is a need for more precise mapping of leukemia cell states, as this may inform disease-specific targeted therapies.

Present-day purification schemes of human hematopoietic stem and progenitor cell (HSPCs) subsets have been refined through decades of functional studies [6–12]. Deep multi-omic profiles of these HSPC subsets have revealed thousands of genes and chromatin regions underlying each stage of hematopoietic differentiation [13–15]. Recent advances in single-cell RNA-sequencing (scRNA-seq) have further improved our understanding of the downstream differentiation trajectories taken by individual HSPCs [3,16–23]. However, existing single-cell landscapes of human hematopoiesis are limited by a lack of sufficient HSPCs [17–19,24], low overall cell numbers [2,3], or a lack of transcriptome-level information [25]. Furthermore, transcriptional definitions of the most primitive hematopoietic stem cells (HSCs) and their downstream progenitors vary from study to study, posing challenges for the design of functional studies based on findings from single-cell analyses [26].

Here, we present a reference map integrating bulk and CD34-sorted datasets, comprising 263,159 high-quality single-cell transcriptomes. This map offers balanced proportions of early HSPCs and terminally differentiated immune cells. We validated our transcriptionally defined cell states with profiles from immunophenotypically pure populations and constructed a publicly available R package enabling rapid projection and classification of new scRNA-seq data using this reference map. Applied to scRNA-seq profiles from 175 diverse leukemia samples with myeloid lineage involvement, we identify 11 subtypes based on differentiation stage involvement. We also nominate genetic alterations driving unconventional lineage priming towards lymphoid and erythroid lineages. Finally, through integrated scRNA-seq and targeted DNA profiling we find that, on rare occasions, distinct leukemia cell hierarchies can co-exist within individual patients.

Methods

Biological Samples:

All samples were collected with informed consent according to procedures approved by either the University Health Network (UHN) Research Ethics Board or St Jude Children's Research Hospital (SJCRH) Institutional Review Board. Human CB samples were obtained from Trillium Health, Credit Valley, and William Osler Hospitals in Ontario. Primary AML samples obtained from the Munich Leukemia Laboratory (MLL) (n = 12) were processed for scRNA-seq with 10x 5' v1.1 technology at SJCRH, and 2 primary AML samples obtained from the Princess Margaret Hospital (PMH) Leukaemia Bank were processed for scRNA-seq with 10x 3' v3 technology at PMH, per manufacturer instructions.

Computational Analyses:

scRNA-seq preprocessing was performed using seurat [27] and scanpy [28] and customized for each dataset based on quality control (QC) metric distributions. Batch correction of scRNA-seq datasets was performed with harmony [29] and reference map projection was performed using symphony [30]. Differential expression analyses utilized DESeq2 [31] at the level of pseudo-bulk profiles. Gene expression modules were defined by cNMF [32], transcription factor regulons were defined by pySCENIC [33], and signature scoring was performed using AUCell [34]. Pseudo-time inference was performed using Monocle3 [35]. For genotyping analysis from scRNA-seq, expressed mutations and gene fusions were called by cbsniffer (https://github.com/genome/cb_sniffer) while copy number alterations were called by inferCNV (https://github.com/broadinstitute/inferCNV).

Full details regarding additional experimental and computational methods used in this study are provided in the supplemental Methods.

Results

A balanced landscape of human bone marrow hematopoiesis

We sought to develop a high-quality single-cell reference atlas of bone marrow hematopoiesis where there are balanced proportions of rare hematopoietic stem and progenitor cells (HSPCs) alongside terminally differentiated populations. To achieve this, scRNA-seq bone marrow data from two unsorted bulk datasets [17–19] was integrated together with three CD34+ sorted datasets [21–23] and a sixth dataset comprising both unsorted bulk and CD34+ sorted data [3]. After quality control, this resulted in 263,159 high-quality scRNA-seq profiles from bone marrow cells wherein 89,404 cells were derived from CD34+ sorted samples and 173,755 cells were derived from unsorted bulk samples. Following dimensionality reduction and focused clustering (supplemental Methods), we identified 55 hematopoietic cell states, with 40 of those states residing along continuous differentiation trajectories starting from primitive HSCs and ending at terminal cell states spanning erythroblasts (Ery), megakaryocytes (Mk), granulocytes (Gran), monocytes (Mono), conventional dendritic cells (cDC), plasmacytoid dendritic cells (pDC), and B cells. The final reference atlas, alongside annotations of the 55 cell states, is shown in **Figure 1**.

Comparison of transcriptional cell states and functionally defined HSPC subsets

To validate the reference map and cell state assignments, transcriptional cell state annotations were compared against immunophenotypicially pure HSPC populations. First, we projected bulk RNA-seq profiles from flow-sorted populations along the hematopoietic hierarchy [14,36,37], confirming their ordering along our UMAP embedding of hematopoiesis **(Figure 2A)**. Multipotent progenitors (MPPs), megakaryocyte-erythrocyte progenitors (MEPs), and common myeloid progenitors (CMPs) were previously shown to be admixed populations, separable into multilineage F1 subfractions or Mk/Ery committed F2/F3 fractions [38]. Projection of transcriptomes from these subfractions confirmed that multilineage F1 fractions and Mk/Ery restricted F2 and F3 fractions within MPPs, CMPs, and MEPs were positioned in alignment with their known functional characteristics (**Figure 2B**).

Next, we evaluated the projection of scRNA-seq profiles collected following flow sorting of various hematopoietic stem and progenitor fractions from two studies [20,39] (**supplemental Figure 1**). Immunophenotypic fractions including MPPs, lympho-myeloid primed progenitors (LMPPs), CMPs, MEPs, and granulocyte-monocyte progenitors (GMPs) were found to be transcriptionally admixed and span multiple cell states (**supplemental Figure 1**), consistent with prior reports of functional heterogeneity within these fractions [38,40,41]. Yet, variation was observed based on gating strategies. For example stringently gated populations in Karamitros *et al* [40] yielded more homogeneous LMPP, GMP, and multi-lymphoid progenitor (MLP) populations by scRNA-seq projection (**Figure 2C**). Re-analysis of index-sorted scRNA-seq profiles [16] GMPs and MEPs with $CD38^{mid}$ immunophenotypes were enriched for true progenitors while those with $CD38^{high}$ immunophenotypes were enriched for more differentiated precursors (**supplemental Figure 2A-C**). Single-cell functional assays confirmed that individual $CD38^{mid}$ GMPs and MEPs formed colonies more effectively than $CD38^{high}$ GMPs and MEPs (**supplemental Figure 2D-J**).

In contrast to hematopoietic progenitors, immunophenotypically defined HSC fractions were relatively pure and exhibited greater concordance with transcriptional cell state annotations (**supplemental Figure 1A,E and Figure 2D**). Notably, higher concordance between transcriptionally and immunophenotypically defined HSCs with increasing stringency of markers used in HSC purification, reaching 90% concordance among highly pure long-term (LT) HSC fractions defined with additional markers including CD49f+ [42–44] (**Figure 2D-E**). Collectively, these analyses show agreement between transcriptionally defined cell states with functionally defined HSPC fractions and, most importantly, demonstrate a high concordance between transcriptional HSCs within our reference map and functionally defined LT-HSCs.

Uncovering gene expression programs across hematopoietic differentiation

To leverage this transcriptome-level expression information across 55 cell states, gene expression programs underlying hematopoietic differentiation were characterized in an unsupervised manner. Consensus NMF (cNMF) [32] was performed across the reference atlas resulting in the identification of 48 distinct gene expression programs, each representing groups of genes with correlated expression across human hematopoiesis. 36 programs were specific to hematopoietic cell states, while 12 programs captured the activity of biological processes spanning the cell cycle, oxidative phosphorylation, and stress response, among others (**Figure**

2F). To characterize these gene expression programs, we identified key transcription factors (TF) regulons (**supplemental Figure 3A**) and biological pathways (**supplemental Figure 3B**) underlying these molecular programs. Notably, while the G2 and M phases of the cell cycle are typically grouped into a single molecular program [45], NMF decoupled these stages into distinct molecular programs allowing us to observe stepwise activation of S-phase, G2-phase, and M-phase molecular programs within our dataset (**supplemental Figure 3C-K**).

To understand how these gene expression programs vary across the continuum of human hematopoietic differentiation, we performed pseudo-time analysis [35] along each lineage captured within our reference map. Of note, we also developed an approach to project hematopoietic pseudo-time estimates onto query data and validated the accuracy of this approach in an independent test dataset [46] (**supplemental Figure 4A-C**). These analyses revealed the key genes and transcription factors underlying successive waves of molecular programs induced alongside hematopoietic differentiation towards human monocyte and erythroid lineages (**Figure 2G-H and supplemental Figure 4D-I**) as well as B cell, pDC, cDC (**supplemental Figure 5**), megakaryocyte, eosinophil/basophil/mast cell (EoBasoMast), and neutrophil lineages (**supplemental Figure 6**).

Assignment of individual leukemia cells to precise hematopoietic cell states

Initial classification of acute leukemia involves the determination of hematopoietic lineage through a small number of cell surface markers. Now that we had a precise map of cell states spanning normal human hematopoiesis, we asked whether the thousands of gene expression markers available across these hematopoietic states could improve lineage classification in acute leukemia. To validate our transcriptional approach for leukemia cell classification, we projected scRNA-seq profiles from leukemia samples with clearly defined lineage features. As expected, Pro-B cells were expanded in B-acute lymphoblastic leukemia (B-ALL) [47] (**Figure 3A**), multipotent MLPs were expanded in mixed-phenotype acute leukemia (MPAL) [36] (**Figure 3B**), pDCs were expanded in blastic plasmacytoid dendritic cell neoplasm (BPDCN) [48] (**Figure 3C**), Mk precursors (MkP) were expanded in acute megakaryoblastic leukemia (AMKL) [49] (**Figure 3D**), and erythroblasts were expanded in acute erythroid leukemia (AEL) [50] (**Figure 3E**).

To map AML samples to precise cellular states along human hematopoiesis we first performed scRNA-seq profiling on 12 AML patient samples harbouring either myelodysplasia (MDS) related

genetic features or classical alterations involving NPM1 or KMT2A. Next, scRNA-seq profiles from eleven studies were re-analyzed [2,3,36,49–56] comprising an additional 154 adult and pediatric AML samples and 9 MPAL samples. After quality control and exclusion of mature lymphocytes, composition analysis utilizing 600,570 cells mapped to 38 cell states revealed patterns of covariation among leukemia cell states (**Figure 3F, supplemental Figure 7A**). Next, similar cell states which co-occurred across patient samples were collapsed together, thus condensing the 38 precise cell states into 11 broader differentiation stages including HSC/MPP-like, LMPP-like, Early Lymphoid, MEP/MkP-like, EoBasoMast, Early Erythroid, Late Erythroid, GMP-like, ProMono-like, Mono-like, and cDC-like (**supplemental Figure 7B**). We confirmed that projected AML cells retained gene expression programs reflecting their assigned stages along hematopoietic differentiation (**supplemental Figure 7C-D**). We further confirmed that primitive HSC/MPP-like and LMPP-like blasts are enriched for Quiescent LSPC signatures [4] and genes upregulated within sorted LSC+ fractions [57] (**supplemental Figure 7E-G**).

Through single-cell composition analysis of 175 AML and MPAL samples, eleven patient subtypes were defined through differences in leukemia cell state composition (**Figure 3G-J, supplemental Figure 8**). Notably, these AML subtypes span multiple stages of differentiation across multiple lineages, including primitive subtypes enriched for HSC/MPP (s1), Early Myeloid (s2), and Early Lymphoid (s3) populations (**Figure 3H**), novel subtypes enriched for MEP/MkP (s4), and Erythroid (s5,s6) lineage abundance (**Figure 3I**), those involving multiple stages of monocytic or dendritic differentiation (s7,s8), and those specifically enriched for GMP-like (s9,s10) or mature myeloid (s11) populations (**Figure 3J**). Interestingly, some patient samples within subtype 11 appear to lack stem and progenitor involvement entirely, with the earliest AML population corresponding to a ProMono-like or Monocyte state (**supplemental Figure 8**). This lack of shared progenitor states suggests heterogeneous disease origins among patient samples despite a shared clinical diagnosis of AML.

Genetic determinants of lineage priming in acute myeloid leukemia

Given the diversity in lineage priming in AML observed through single-cell composition analysis, we sought to understand how genetic driver alterations influenced lineage priming at the level of patient cohorts. Marker genes were identified for each AML differentiation stage and used to train sparse regression models to estimate their relative abundance from bulk RNA-seq profiles (supplemental Methods**; supplemental Figure 9**). Leveraging bulk RNA-seq of 864 AML patients

spanning three cohorts [4,58–60], associations were identified between inferred differentiation stage abundance and the presence of specific driver mutations (**Figure 4A**) as well as key cytogenetic alterations (**supplemental Figure 10A**). Furthermore, associations between inferred differentiation stage abundance with AML morphology and hierarchy subtype served as validation for our approach (**supplemental Figure 10B-C**).

We next evaluated determinants of lymphoid lineage priming in AML. Within the single-cell data, we observed co-clustering of AML and MPAL samples within subtype 3, which was enriched for Early Lymphoid (MLP and CLP) abundance (**Figure 4B**). This highlights the biological continuum between these disease classes. Through bulk analysis, we confirmed that MPAL is highly enriched for Early Lymphoid cells compared to AML (p=8.8e-14) (**Figure 4C**). Within AML patients, we found high Early Lymphoid abundance to be associated with NUP98-NSD1 fusions (p=0.0036) and RUNX1 mutations (p=1.8e-7) (**Figure 4D-E**).

We next examined erythroid lineage priming. AML and AEL samples co-clustered within subtypes 5 and 6 from single-cell composition analysis, which were enriched for Early Erythroid (CFU-E, Pro-Erythroblast, and Basophilic Erythroblast) abundance (**Figure 4F**). Through bulk RNA-seq, we confirmed that patient samples with French-American-British (FAB) classes assigned as M6 (erythroid) were highly enriched for Early Erythroid cells compared to M0-M5 AMLs (p=0.00067) (**Figure 4G, supplemental Figure 10D**). Considering genetic associations, higher Early Erythroid abundance was observed with TP53 mutations (p=7.6e-15), complex karyotype (p=3.9e-12), and MDS-related cytogenetic alterations including monosomy 7/del (7q) (p=0.0031), monosomy 17/del(17p) (p=0.0077), and monosomy 5/del(5q) (p=0.00064) (**Figure 4H, supplemental Figure 10E-F**). Notably, differential priming towards myeloid and erythroid lineages also captured inter-patient heterogeneity within 136 AEL samples, wherein higher Early Erythroid abundance was associated with TP53 mutations (p=3.6e-9) and Poor cytogenetic risk (p=1.1e-9) while high GMP-like abundance was associated with KMT2A alterations (p=0.00046) and Good/Intermediate risk (p=0.00032) (**Figure 4I-J**).

Collectively, these data highlight that priming towards lymphoid and erythroid lineages in AML is influenced by genomic drivers of the disease. Furthermore, despite clear criteria delineating AML from MPAL and AEL based on a small collection of morphological or immunophenotypic markers, high-resolution maps leveraging thousands of gene expression markers suggest that these diseases exist along a biological continuum with poorly demarcated boundaries.

Impact of genetic subclones on cellular hierarchies in AML

Recognizing that genomic drivers are associated with lineage priming in AML at the patient level, we next asked whether these relationships could be identified at the level of genetic subclones within individual patients. Various approaches were utilized to evaluate this, including the identification of expressed mutations or copy number alterations within a subset of patients, the use of previously annotated genetic information from re-analyzed studies [2,54,56], as well as *de novo* profiling of in-house patient samples with single-cell targeted DNA (scDNA-seq) with immunophenotyping through the Tapestri platform. Using these genetic profiles, the differences in AML lineage priming between genetic subclones within individual patients could be evaluated (**Figure 5A**).

These analyses uncovered examples of mutations influencing AML hierarchies, including lineage skewing from myeloid to erythroid in a subclone that acquired MDS-related chromosomal alterations (**Figure 5B-D**), or differentiation block at the level of HSC/MPPs induced by a subclonal *GATA2* mutation (**Figure 5E-G**). In a sample that was profiled through scRNA-seq as well as Tapestri scDNA-seq + immunophenotype (**Figure 5H)**, a balanced composition comprised of a CD34-expressing primitive population and a CD11b-expressing mature monocytic population could be identified (**Figure 5I-J**). Through scDNA-seq with immunophenotyping, we found that cells which belonged to the ancestral Monosomy 7 and *RUNX1* clone were predominantly primitive (CD34+CD11b-), while cells that belonged to a *KRAS* mutated subclone were predominantly mature (CD34-CD11b+) (**Figure 5K-L**). These data demonstrate that subclonal genetic mutations can induce lineage biases, differentiation blocks, and even mature myeloid cell expansion within AML patients.

These differences in cell composition between genetic subclones suggest that distinct leukemia cell hierarchies can co-exist within a subset of AML patients. Through paired scRNA-seq profiling of a primary AML sample and its corresponding patient-derived xenograft (PDX), we confirmed that the hierarchy of a PDX sample can differ from that of the original primary AML sample (**Figure 5M-N**). Notably, while the primary sample comprised both primitive and mature cells, these primitive cells were absent from the PDX suggesting that the PDX was sustained by a mature myeloid population that had acquired stemness properties. This raises the question of whether distinct leukemia cell hierarchies within an AML patient could be sustained by co-existing primitive and mature populations with functional LSC properties.

Distinct leukemia cell hierarchies can co-exist within individual AML patients

To formally investigate whether distinct LSC-driven hierarchies may co-exist within certain AML patients, we re-analyzed engraftment data from 74 AML patients wherein leukemia cells were sorted into four fractions based on surface CD34 and CD38 and engrafted into NSG mice at varying doses [57]. Functional LSC activity, defined as the ability to initiate leukemic grafts, was observed within the CD34+CD38- fraction for 56 out of 74 patients (76%). Of the remaining 18 patients (18%) without LSC activity in the CD34+CD38- fraction, 7 patients (9%) had LSC activity detected within the CD34+CD38+ fraction while 11 patients (15%) had LSC activity restricted to the CD34- fraction. Considering the majority of patients with LSC activity within the CD34+CD38- fraction, 8 patients (11%) had engraftment restricted to CD34+CD38- cells while 15 patients (20%) had engraftment restricted to CD34+ fractions. These engraftment data are outlined in **supplemental Figure 11**.

Interestingly, 29 patients (39%) had observed functional LSC activity in three or more fractions, spanning CD34+ and CD34- immunophenotypes. The presence of LSC activity across multiple CD34 and CD38 fractions could be related to previously documented immunophenotypic plasticity of CD34 and CD38 marker expression among primitive LSCs [57,61–66], particularly in patients with primitive hierarchies with extensive stem and progenitor involvement [4]. Alternatively, in other patients, this could also be explained by co-existing hierarchies sustained by primitive and mature populations with functional LSC potential. To find examples of the latter, we evaluated immunophenotypes of PDXs generated from each CD34 and CD38 fraction and identified two patient samples with distinct PDX immunophenotypes following engraftment from CD34+ and CD34- primary fractions as likely containing distinct LSC populations.

For these two patients, we performed scRNA-seq on the unsorted primary patient sample as well as sorted CD34+CD38- (34+38-) and CD34-CD38+ (34-38+) fractions. We engrafted these CD34+CD38- and CD34-CD38+ fractions individually and performed scRNA-seq on the resulting PDX cells (**Figure 6A**). Flow cytometry and scRNA-seq analysis on primary AML cells from one of these patients (pt.90240) revealed multiple leukemia cell populations spanning LMPP-like, CLP-like, GMP-like, Pre-DC-like, and Mono-like (**Figure 6B-C, supplemental Figure 12**). Within the CD34+CD38- fraction (0.06% of AML cells) nearly all cells were LMPP-like (**Figure 6D**). PDX samples derived from this fraction recapitulated the complete patient hierarchy by both flow cytometry and scRNA-seq composition (**Figure 6E**). The CD34-CD38+ fraction (11% of AML

cells) comprised primarily of downstream progenitors (**Figure 6F**). PDX samples derived from this fraction were almost fully CD34-, composed entirely of dendritic cell precursors and mature monocytes (**Figure 6G**). Similar results were observed with the second patient (pt.90394) wherein the CD34+CD38- graft recapitulated the full AML hierarchy while the CD34-CD38+ graft was comprised entirely of mature myeloid cells (**supplemental Figures 13 and 14**). Collectively, these data provide proof of principle for a model wherein distinct leukemia cell hierarchies can co-exist within AML individual patients (**Figure 6H**), which may in turn contribute to AML evolution and relapse following selective pressures applied by different therapies (**Figure 6I**).

Discussion

Our map of bone marrow hematopoiesis provides an in-depth characterization of human hematopoietic differentiation, ranging from early HSPC to terminally differentiated populations across multiple lineages. Most importantly, through careful annotation and validation, this curated map constitutes a valuable resource for those planning to utilize scRNA-seq data to study either normal or leukemic hematopoiesis. To maximize the public utility of our resource, we have developed an R package (https://github.com/andygxzeng/BoneMarrowMap) to enable reference mapping and annotation of any normal or leukemic scRNA-seq sample within minutes. Further, an improved understanding of the relationship between transcriptionally defined cell states with immunophenotypically defined cell populations will enable smoother translation of insights from single-cell analyses into functional experiments.

The application of this reference mapping approach to classify leukemia cells, particularly in AML, has allowed us to confront disease heterogeneity in a biologically meaningful way. Strikingly, some AML samples exhibited virtually no overlap in cell state involvement with one another, possibly reflecting disparate cellular origins of their diseases. In contrast, other AML samples shared extensive overlap in cell state involvement with MPAL or AEL samples, challenging traditional diagnostic boundaries between classes of leukemia [1]. Moving towards homogeneous disease classes is critical for properly evaluating therapy response, and the thousands of new markers uncovered through scRNA-seq could be leveraged to refine diagnostic criteria in the classification of acute leukemia.

Last, our findings highlight the complex interplay between genetic alterations and lineage determination in AML, identifying mutations and cytogenetic alterations that are highly associated with the abundance of 11 AML differentiation stages across patient cohorts. By analyzing AML genetics at the single-cell level, we expanded upon previous studies [56,67,68] to show that distinct genetic subclones can differ in their leukemia cell hierarchy composition, and formally demonstrated that distinct AML hierarchies sustained by distinct LSC populations can co-exist within a subset of patients. The co-existence of primitive and mature AML populations with functional LSC properties was previously proposed in a subset of AML patients [69,70], and monocytic LSCs have emerged as an important source of relapse from Venetoclax-based therapy [71,72]. Though this appears to be a rare occurrence, AML samples with co-existing hierarchies may be better poised to evolve in response to targeted therapies and other selective pressures. Finally, charting the specific lineage perturbations caused by each genetic alteration provides a framework for reconciling genetic clonal evolution with changes in leukemia cell hierarchy composition in explaining evolution and relapse in AML.

Resource Availability

An R package for projecting and classifying hematopoietic cells from query data is available at (https://github.com/andygxzeng/BoneMarrowMap). The BoneMarrowMap atlas can be explored interactively through the cellxgene web portal: https://cellxgene.cziscience.com/e/cd2f23c1-aef1-48ae-8eb4-0bcf124e567d.cxg/. Annotated raw count data for BoneMarrowMap and AML samples profiled within this study will be made available at time of journal publication. For earlier access through collaboration, please contact Andy Zeng (andy.zeng@mail.utoronto.ca) and John Dick (john.dick@uhnresearch.ca) for inquiries.

Authorship Contributions

A.G.X.Z.: designed and led the study, performed functional experiments, analyzed data, and wrote the manuscript. I.I.: performed single cell profiling experiments and analyzed data. S.S.: analyzed data and contributed to figure preparation. A.M.: performed functional experiments and contributed to engraftment analyses. S.B.: contributed to engraftment analyses and figure preparation. G.W.: performed functional experiments. Q.G.: analyzed data. H.K.: contributed to functional experiments. J.A.K.: contributed to engraftment analyses. M.D.M.: provided clinical samples. T.H.: provided clinical samples. C.G.M.: provided clinical samples. J.E.D.: contributed to study design, supervised the study, and revised the manuscript for final approval. All authors contributed to manuscript writing and gave final approval.

Acknowledgements

We thank the patients for donating samples of their blood. We thank Mason Boulanger for help with organization of clinical data, and all members of the Dick lab for helpful discussions. A.G.X.Z is supported by a University of Toronto MD/PhD studentship award. J.E.D is supported by Princess Margaret Cancer Foundation; Ontario Institute for Cancer Research through funding provided by the Government of Ontario; Canadian Institutes for Health Research RN380110-409786; International Development Research Centre Ottawa Canada; Canadian Cancer Society 703212; Terry Fox New Frontiers Program project grant 1106; University of Toronto's Medicine by Design initiative with funding from the Canada First Research Excellence Fund; The Ontario Ministry of Health; Canada Research Chair.. C.G.M is supported by the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital; the

Alex's Lemonade Stand Foundation for Childhood Cancer; the National Institutes of Health, National Cancer Institute grants P30 CA021765 and R35 CA197695 ; a St. Baldrick's Foundation Robert J. Arceci Innovation Award; the Henry Schueler 41&9 Foundation.

Disclosure of Conflicts of Interest

J.E.D.: received research funding from BMS/Celgene and IP licenses from Pfizer/Trillium Therapeutics. I.I.: reported consultation honoraria from Arima, travel expenses reimbursed by Mission Bio for invited talk and honoraria from MD Education. C.G.M.: received research funding from AbbVie and Pfizer, honoraria from Amgen and Illumina, royalty payments from Cyrus, and is on an advisory board for Illumina. T.H.: equity ownership of MLL Munich Leukemia Laboratory

References

1. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia*. 2022;36(7):1703–1719.
2. van Galen P, Hovestadt V, Wadsworth MH Ii, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell*. 2019;176(6):1265–1281.e24.
3. Granja JM, Klemm S, McGinnis LM, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* 2019;37(12):1458–1465.
4. Zeng AGX, Bansal S, Jin L, et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nat. Med.* 2022;28(6):1212–1223.
5. Bottomly D, Long N, Schultz AR, et al. Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. *Cancer Cell*. 2022;40(8):850–864.e9.
6. Civin CI, Strauss LC, Brovall C, et al. Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *J. Immunol.* 1984;133(1):157–165.
7. Lansdorp PM, Sutherland HJ, Eaves CJ. Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow. *J. Exp. Med.* 1990;172(1):363–366.
8. Baum CM, Weissman IL, Tsukamoto AS, Buckle AM, Peault B. Isolation of a candidate human hematopoietic stem-cell population. *Proc. Natl. Acad. Sci. U. S. A.* 1992;89(7):2804–2808.
9. Bhatia M, Wang JC, Kapp U, Bonnet D, Dick JE. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* 1997;94(10):5320–5325.
10. Majeti R, Park CY, Weissman IL. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell*. 2007;1(6):635–645.
11. Doulatov S, Notta F, Eppert K, et al. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat. Immunol.* 2010;11(7):585–593.
12. Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. *Cell Stem*

Cell. 2012;10(2):120–136.
13. Laurenti E, Doulatov S, Zandi S, et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat. Immunol.* 2013;14(7):756–763.
14. Xie SZ, Kaufmann KB, Wang W, et al. Sphingosine-1-phosphate receptor 3 potentiates inflammatory programs in normal and leukemia stem cells to promote differentiation. *Blood Cancer Discov*. 2021;2(1):32–53.
15. Takayama N, Murison A, Takayanagi S-I, et al. The Transition from Quiescent to Activated States in Human Hematopoietic Stem Cells Is Governed by Dynamic 3D Genome Reorganization. *Cell Stem Cell*. 2021;28(3):488–501.e10.
16. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* 2017;19(4):271–281.
17. Li B, Kowalczyk M, Slyper M, et al. Census of Immune Cells.
18. Hay SB, Ferchen K, Chetal K, Grimes HL, Salomonis N. The Human Cell Atlas bone marrow single-cell interactive web portal. *Exp. Hematol.* 2018;68:51–61.
19. Oetjen KA, Lindblad KE, Goswami M, et al. Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight*. 2018;3(23.):
20. Pellin D, Loperfido M, Baricordi C, et al. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat. Commun.* 2019;10(1):2395.
21. Setty M, Kiseliovas V, Levine J, et al. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* 2019;37(4):451–460.
22. Mende N, Bastos HP, Santoro A, et al. Unique molecular and functional features of extramedullary hematopoietic stem and progenitor cell reservoirs in humans. *Blood*. 2022;
23. Ainciburu M, Ezponda T, Berastegui N, et al. Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single-cell resolution. *Elife*. 2023;12.:
24. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573–3587.e29.
25. Triana S, Vonficht D, Jopp-Saile L, et al. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat. Immunol.* 2021;22(12):1577–1589.
26. Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. *Nature*. 2018;553(7689):418–426.
27. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic

data across different conditions, technologies, and species. *Nat. Biotechnol.* 2018;36(5):411–420.

28. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19(1):15.
29. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*. 2019;16(12):1289–1296.
30. Kang JB, Nathan A, Weinand K, et al. Efficient and precise single-cell reference atlas mapping with Symphony. *Nat. Commun.* 2021;12(1):5890.
31. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
32. Kotliar D, Veres A, Nagy MA, et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife*. 2019;8.:
33. Van de Sande B, Flerin C, Davie K, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* 2020;15(7):2247–2276.
34. Aibar S, González-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods*. 2017;14(11):1083–1086.
35. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 2014;32(4):381–386.
36. Montefiori LE, Bendig S, Gu Z, et al. Enhancer hijacking drives oncogenic BCL11B expression in lineage ambiguous stem cell leukemia. *Cancer Discov.* 2021;
37. An X, Schulz VP, Li J, et al. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood*. 2014;123(22):3466–3477.
38. Notta F, Zandi S, Takayama N, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*. 2016;351(6269):aab2116.
39. Zhang Y, Xie X, Huang Y, et al. Temporal molecular program of human hematopoietic stem and progenitor cells after birth. *Dev. Cell*. 2022;57(24):2745–2760.e6.
40. Karamitros D, Stoilova B, Aboukhalil Z, et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat. Immunol.* 2018;19(1):85–97.
41. Drissen R, Thongjuea S, Theilgaard-Mönch K, Nerlov C. Identification of two distinct pathways of human myelopoiesis. *Sci Immunol*. 2019;4(35.):
42. Belluschi S, Calderbank EF, Ciaurro V, et al. Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nat. Commun.* 2018;9(1):4100.

43. Kaufmann KB, Zeng AGX, Coyaud E, et al. A latent subset of human hematopoietic stem cells resists regenerative stress to preserve stemness. *Nat. Immunol.* 2021;22(6):723–734.
44. Anjos-Afonso F, Buettner F, Mian SA, et al. Single cell analyses identify a highly regenerative and homogenous human CD34+ hematopoietic stem cell population. *Nat. Commun.* 2022;13(1):1–13.
45. Tirosh I, Izar B, Prakadan SM, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016;352(6282):189–196.
46. Roy A, Wang G, Iskander D, et al. Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. *Cell Rep.* 2021;36(11):109698.
47. Caron M, St-Onge P, Sontag T, et al. Single-cell analysis of childhood leukemia reveals a link between developmental states and ribosomal protein expression as a source of intra-individual heterogeneity. *Sci. Rep.* 2020;10(1):8079.
48. DePasquale EAK, Ssozi D, Ainciburu M, et al. Single-Cell Multiomics Reveals Clonal T-Cell Expansions and Exhaustion in Blastic Plasmacytoid Dendritic Cell Neoplasm. *Front. Immunol.* 2022;13:809414.
49. Lasry A, Nadorp B, Fornerod M, et al. An inflammatory state remodels the immune microenvironment and improves risk stratification in acute myeloid leukemia. *Nature Cancer*. 2022;4(1):27–42.
50. Kuusanmäki H, Dufva O, Vähä-Koskela M, et al. Erythroid/megakaryocytic differentiation confers BCL-XL dependency and venetoclax resistance in acute myeloid leukemia. *Blood*. 2023;141(13):1610–1625.
51. Li K, Du Y, Cai Y, et al. Single-cell analysis reveals the chemotherapy-induced cellular reprogramming and novel therapeutic targets in relapsed/refractory acute myeloid leukemia. *Leukemia*. 2022;37(2):308–325.
52. Naldini MM, Casirati G, Barcella M, et al. Longitudinal single-cell profiling of chemotherapy response in acute myeloid leukemia. *Nat. Commun.* 2023;14(1):1–20.
53. Ennis S, Conforte A, O'Reilly E, et al. Cell-cell interactome of the hematopoietic niche and its changes in acute myeloid leukemia. *iScience*. 2023;26(6):106943.
54. Petti AA, Williams SR, Miller CA, et al. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. *Nat. Commun.* 2019;10(1):1–16.
55. Abbas HA, Hao D, Tomczak K, et al. Single cell T cell landscape and T cell receptor repertoire profiling of AML in context of PD-1 blockade therapy. *Nat. Commun.* 2021;12(1):1–13.

56. Beneyto-Calabuig S, Merbach AK, Kniffka J-A, et al. Clonally resolved single-cell multi-omics identifies routes of cellular differentiation in acute myeloid leukemia. *Cell Stem Cell*. 2023;30(5):706–721.e8.
57. Ng SWK, Mitchell A, Kennedy JA, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature*. 2016;540(7633):433–437.
58. Marquis M, Beaubois C, Lavallée V-P, et al. High expression of HMGA2 independently predicts poor clinical outcomes in acute myeloid leukemia. *Blood Cancer J.* 2018;8(8):68.
59. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. 2018;562(7728):526–531.
60. Cancer Genome Atlas Research Network, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013;368(22):2059–2074.
61. Thomas D, Majeti R. Biology and relevance of human acute myeloid leukemia stem cells. *Blood*. 2017;129(12):1577–1585.
62. Stelmach P, Trumpp A. Leukemic stem cells and therapy resistance in acute myeloid leukemia. *Haematologica*. 2023;108(2):353–366.
63. Eppert K, Takenaka K, Lechman ER, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* 2011;17(9):1086–1093.
64. Sarry J-E, Murphy K, Perry R, et al. Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J. Clin. Invest.* 2011;121(1):384–395.
65. Lagadinou ED, Sach A, Callahan K, et al. BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells. *Cell Stem Cell*. 2013;12(3):329–341.
66. Pabst C, Bergeron A, Lavallée V-P, et al. GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo. *Blood*. 2016;127(16):2018–2027.
67. Miles LA, Bowman RL, Merlinsky TR, et al. Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature*. 2020;587(7834):477–482.
68. Turkalj S, Jakobsen NA, Groom A, et al. GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells. *Cell Stem Cell*. 2023;30(5):722–740.e11.
69. Goardon N, Marchi E, Atzberger A, et al. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell*. 2011;19(1):138–152.
70. Quek L, Otto GW, Garnett C, et al. Genetically distinct leukemic stem cells in human CD34-

acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. *J. Exp. Med.* 2016;213(8):1513–1535.

71. Pei S, Pollyea DA, Gustafson A, et al. Monocytic Subclones Confer Resistance to Venetoclax-Based Therapy in Patients with Acute Myeloid Leukemia. *Cancer Discov.* 2020;
72. Pei S, Shelton IT, Gillen AE, et al. A Novel Type of Monocytic Leukemia Stem Cell Revealed by the Clinical Use of Venetoclax-Based Therapy. *Cancer Discov.* 2023;13(9):2032–2049.

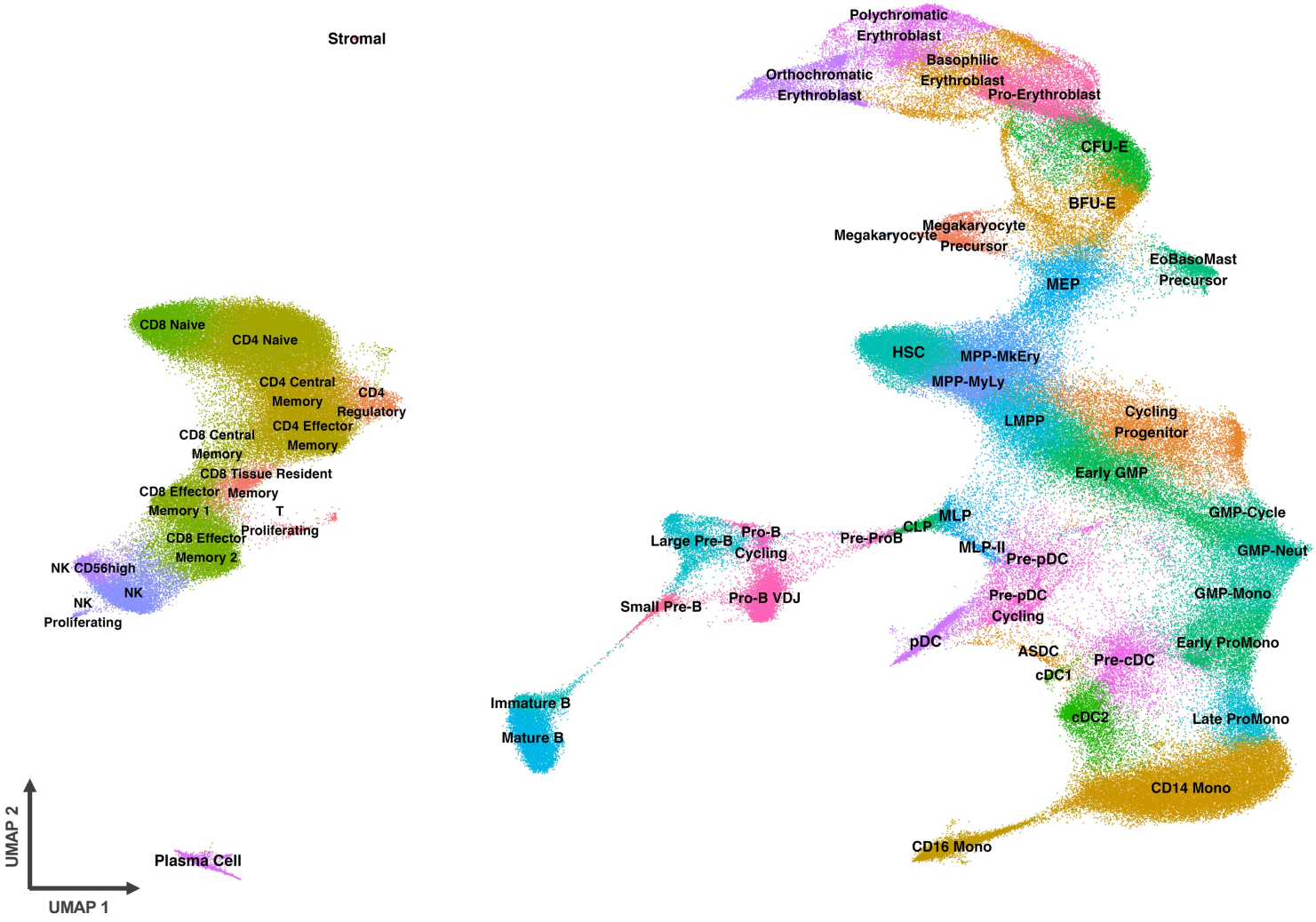

Figure 1. A balanced landscape of human bone marrow hematopoiesis
Single-cell transcriptional atlas of human hematopoiesis with balanced representation of CD34+ stem and progenitor cells together with terminally differentiated populations, comprising 263,159 bone marrow cells spanning 55 cell states across 45 donors from six studies.

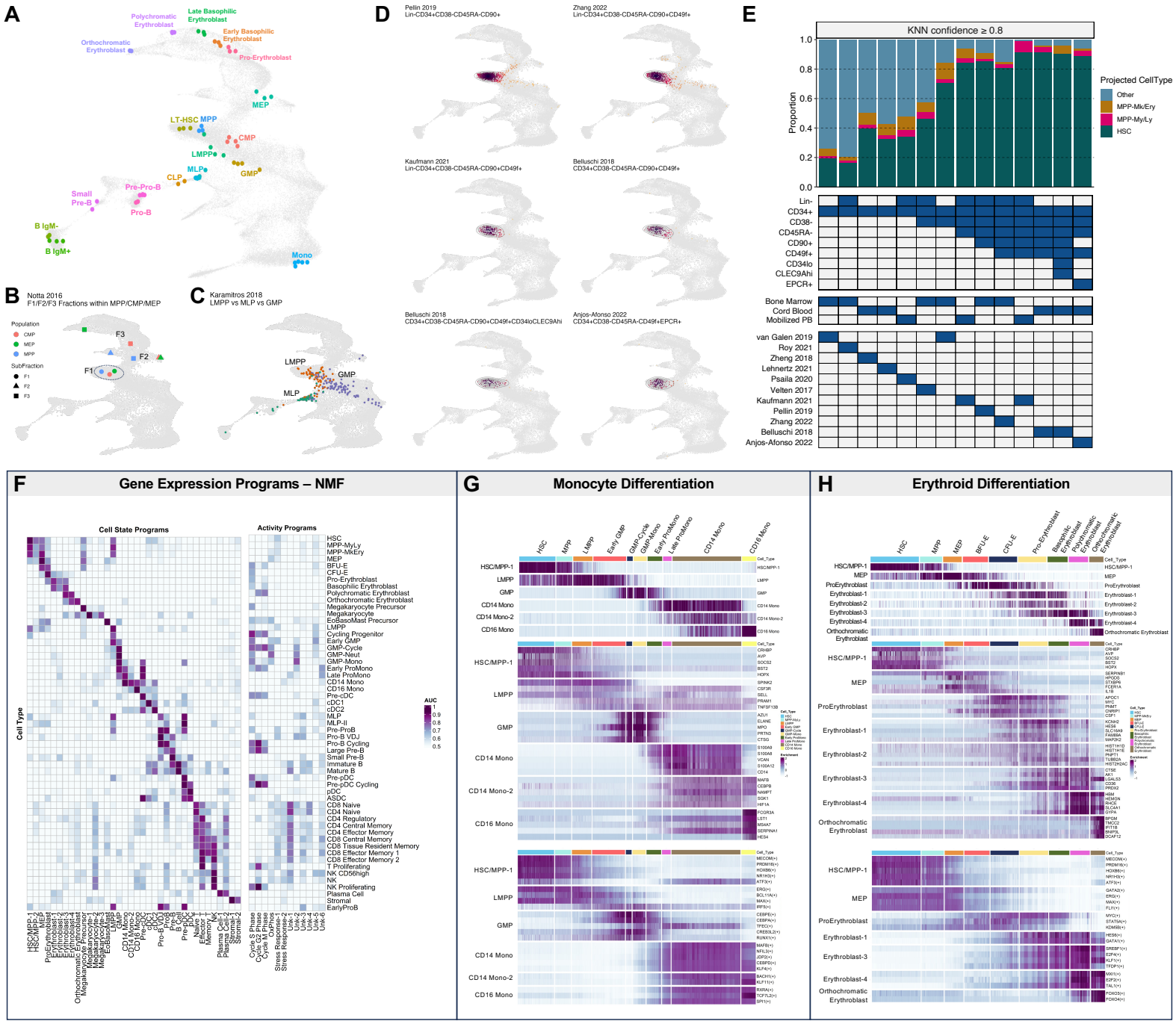

Figure 2. Validation and characterization of BoneMarrowMap

A) Projection of bulk RNA-seq profiles from purified hematopoietic cell populations onto BoneMarrowMap. **B)** Projection of sorted F1, F2, and F3 subfractions of MPP, CMP, and MEP. **C)** Projection of scRNA-seq profiles purified with functionally informed gating of LMPP, GMP, and MLP populations. **D)** Projection of scRNA-seq profiles from LT-HSCs purified using different markers of human HSCs. **E)** Proportion of transcriptional HSCs projected from scRNA-seq data sorted with increasing levels of immunophenotypic purity. The combination of markers used, tissue source, and dataset name are also depicted. **F)** Unsupervised discovery of 48 gene expression programs across human hematopoiesis by cNMF wherein 36 corresponded to hematopoietic cell states and 12 corresponded to other cellular processes. The relative activity of each program within each hematopoietic cell state is depicted through a heatmap of AUC values. **G-H)** Expression of gene expression programs across hematopoietic pseudo-time along **G)** monocyte and **H)** erythroid differentiation trajectories. For each trajectory, the top heatmap shows the enrichment of gene expression programs underlying cell states along each trajectory. The middle heatmap shows the expression of the top 5 genes driving each gene expression program. The bottom heatmap shows the regulon activity of top transcription factors associated with each program. Cells are ordered along hematopoietic pseudo-time.

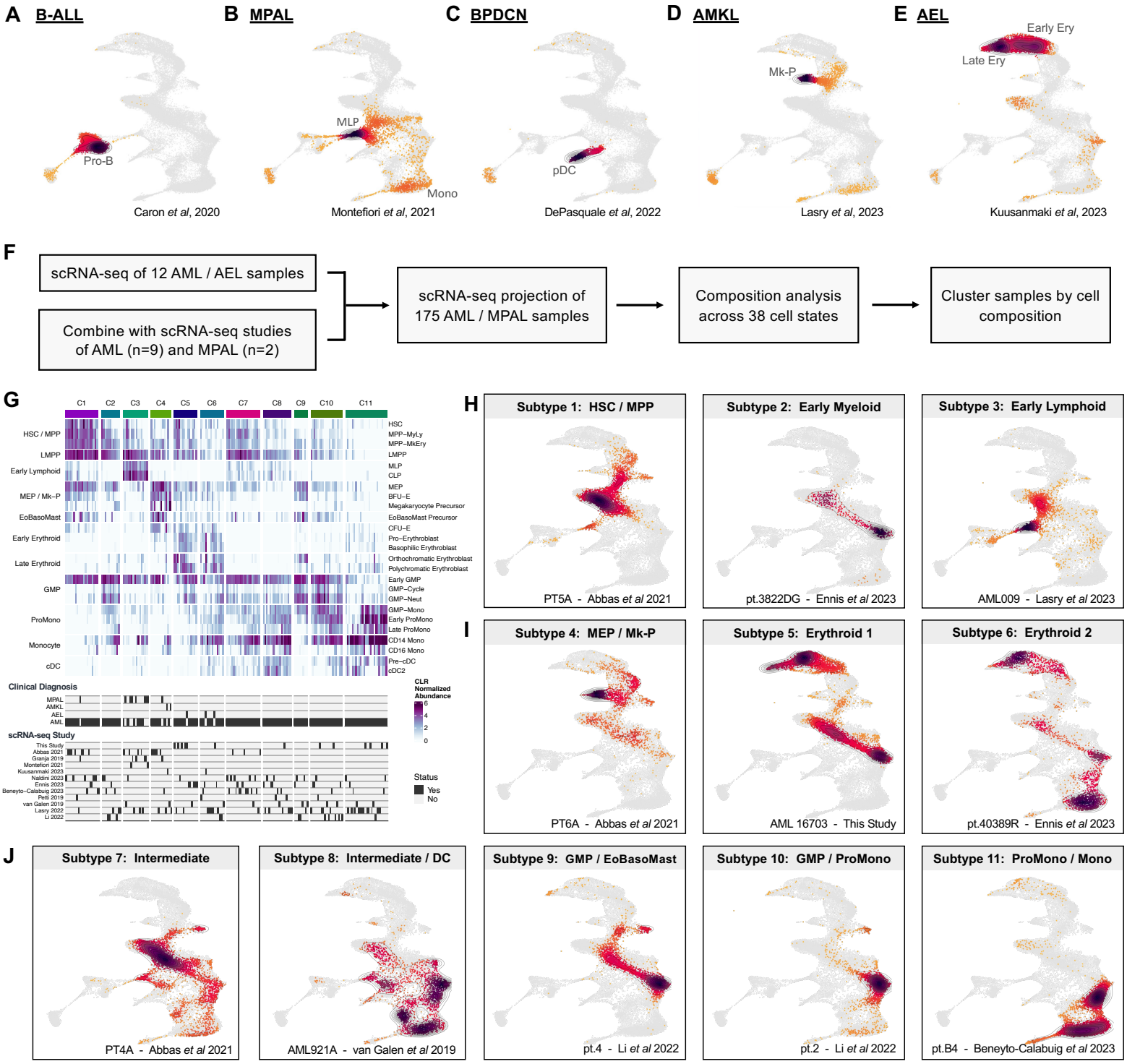

Figure 3. Projection and classification of single cell transcriptomes in acute leukemia
A-E) Projection of scRNA-seq profiles from leukemia samples spanning multiple diagnoses, including **A)** B cell acute lymphoblastic leukemia (B-ALL), **B)** mixed phenotype acute leukemia (MPAL), **C)** blastic plasmacytoid dendritic cell neoplasm (BPDCN), **D)** acute megakaryoblastic leukemia (AMKL), and **E)** acute erythroid leukemia (AEL). **F)** Cell state projection and composition analysis of 166 acute myeloid leukemia (AML) and 9 MPAL samples to define disease subtypes on the basis of leukemia cell composition. **G)** Eleven AML subtypes identified through cell state projection and composition analysis. Relative abundance of specific cell states and broader differentiation stages are depicted through the heatmap; clinical diagnosis and originating study are also annotated for each AML sample. **H-J)** Projection results for representative samples from each AML subtype. AML subtypes were diverse and could exhibit **H)** extensive involvement of early HSPCs, **I)** extensive involvement of Mk/Ery lineages, or **J)** extensive involvement of GMPs and mature myeloid populations.

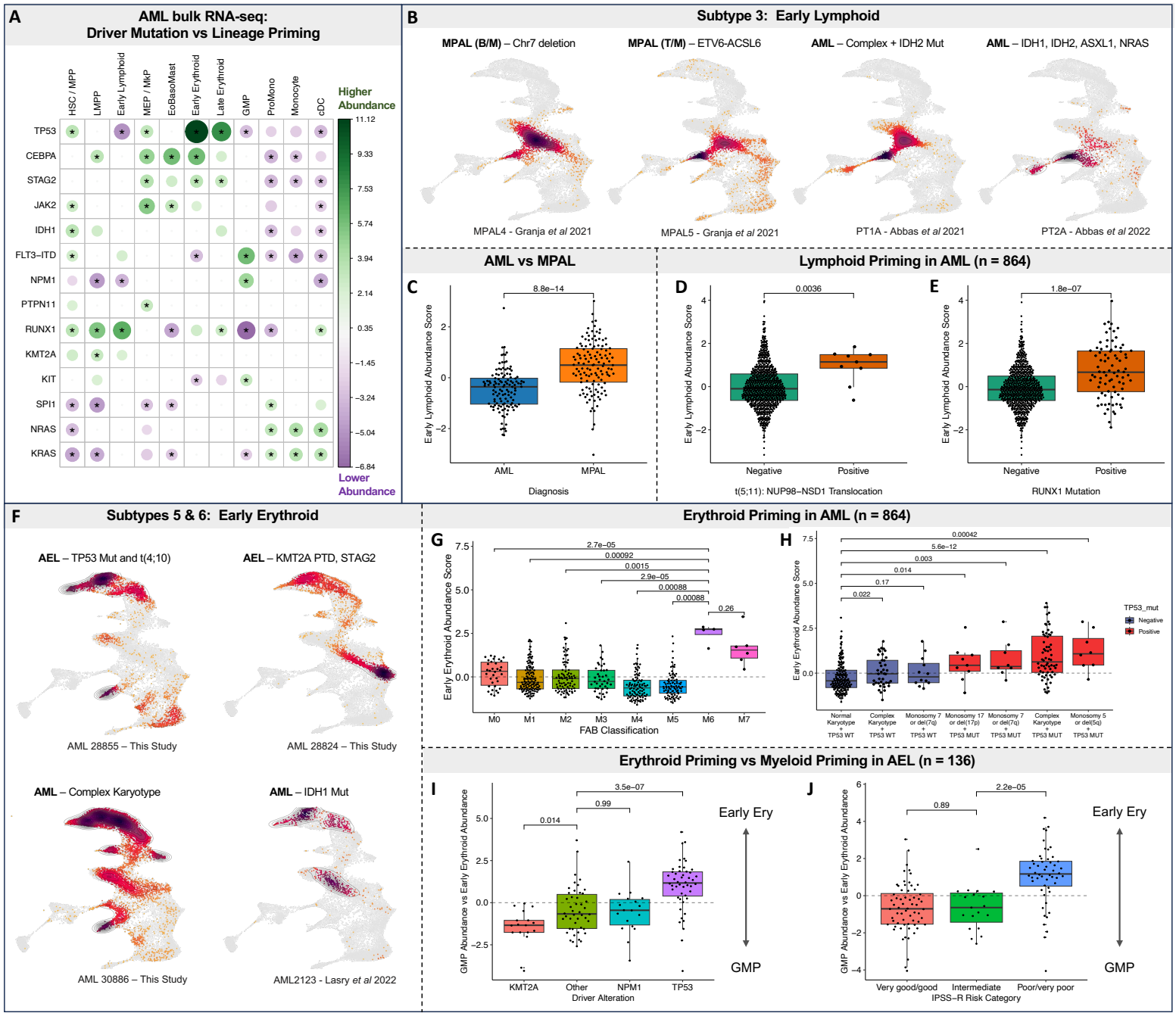

Figure 4. Genetic determinants of lymphoid and erythroid priming in acute leukemia

A) Association between mutation states and inferred abundance of AML differentiation stages from 864 AML samples profiled by RNA-seq. The magnitude of each association, quantified as the –log10(pvalue), is depicted through the size and color intensity of each dot, wherein higher abundance is green and lower abundance is purple. Only associations at uncorrected $p < 0.05$ are shown, those at FDR < 0.05 are starred. **B)** Representative AML and MPAL patient samples from Subtype 3, which is enriched for Early Lymphoid cells. **C)** Comparison of inferred Early Lymphoid abundance between 126 MPALs and 122 AMLs from Montefiori *et al* 2021. **D)** Association between *NUP98-NSD1* fusion status and *RUNX1* mutation status with inferred Early Lymphoid abundance within 864 AML samples. **E)** Representative AML and AEL patient samples from Subtypes 5 and 6, which are enriched for Early Erythroid cells. **G)** Association of M6 FAB morphology with inferred Early Erythroid abundance within 864 AML samples. **H)** Association of TP53 mutation status, complex karyotype, and alterations in chromosomes 5, 7, and 17 with inferred Early Erythroid abundance within 864 AML samples. **I-J)** Erythroid vs myeloid priming within 136 AEL samples, depicted as the difference between the inferred Early Erythroid abundance and GMP abundance in each sample. Lineage priming patterns towards erythroid vs myeloid lineages was associated with **I)** driver mutation status and **J)** clinical risk category.

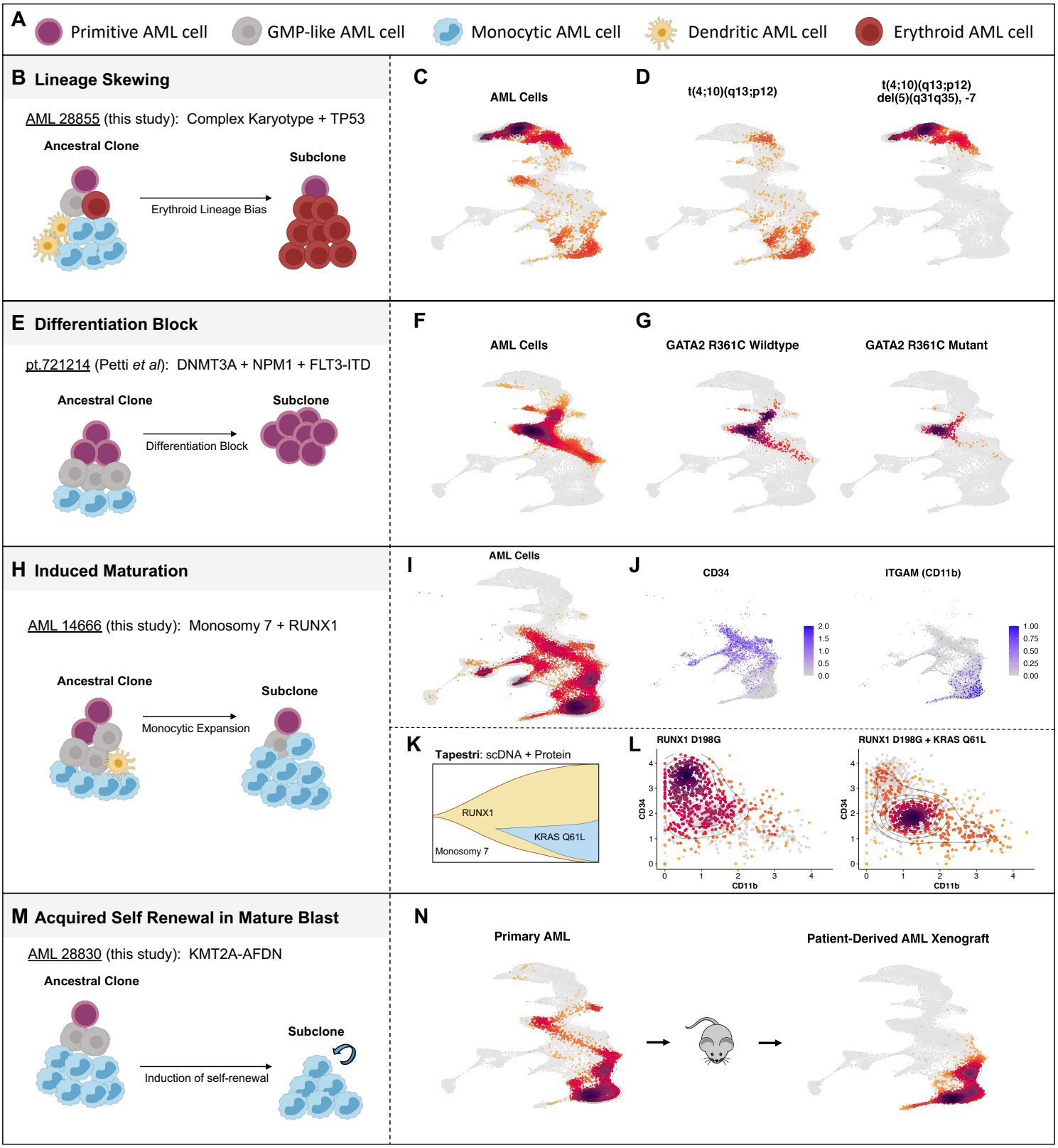

Figure 5. Impact of genetic subclones on cellular hierarchies in AML

A) Graphical representation of broad differentiation stages within AML samples. **B-D)** Example of lineage skewing from myeloid to erythroid at the sub-clonal level, depicting projection results from **C)** the primary AML sample and from **D)** a clone involving t(4;10) and an erythroid-restricted sub-clone involving t(4;10), del(5q), and monosomy 7. **E-G)** Example of a primitive differentiation block at the sub-clonal level, depicting projection results from **F)** the primary AML sample and from **G)** a *GATA2* wildtype clone and an HSC/MPP-restricted *GATA2* mutant sub-clone. **H-L)** Example of induced maturation of AML cells at the sub-clonal level, depicting projection results from **I)** the primary AML sample, wherein **J)** primitive stem and progenitor cells express *CD34* and mature myeloid cells express *ITGAM* (encoding CD11b). **K)** scDNA + protein profiling of this sample by Tapestri reveals an ancestral clone with monosomy 7 and a *RUNX1* mutation with a sub-clonal *KRAS* mutant clone. **L)** Immunophenotype analysis reveals the *RUNX1* mutant ancestral clone to be mostly primitive (CD34+CD11b-) and the *KRAS* mutant sub-clone to be mostly mature (CD34-CD11b+). **M-N)** Example of acquired self-renewal within mature blasts wherein **N)** mature populations from a primary AML sample sustain the patient-derived xenograft.

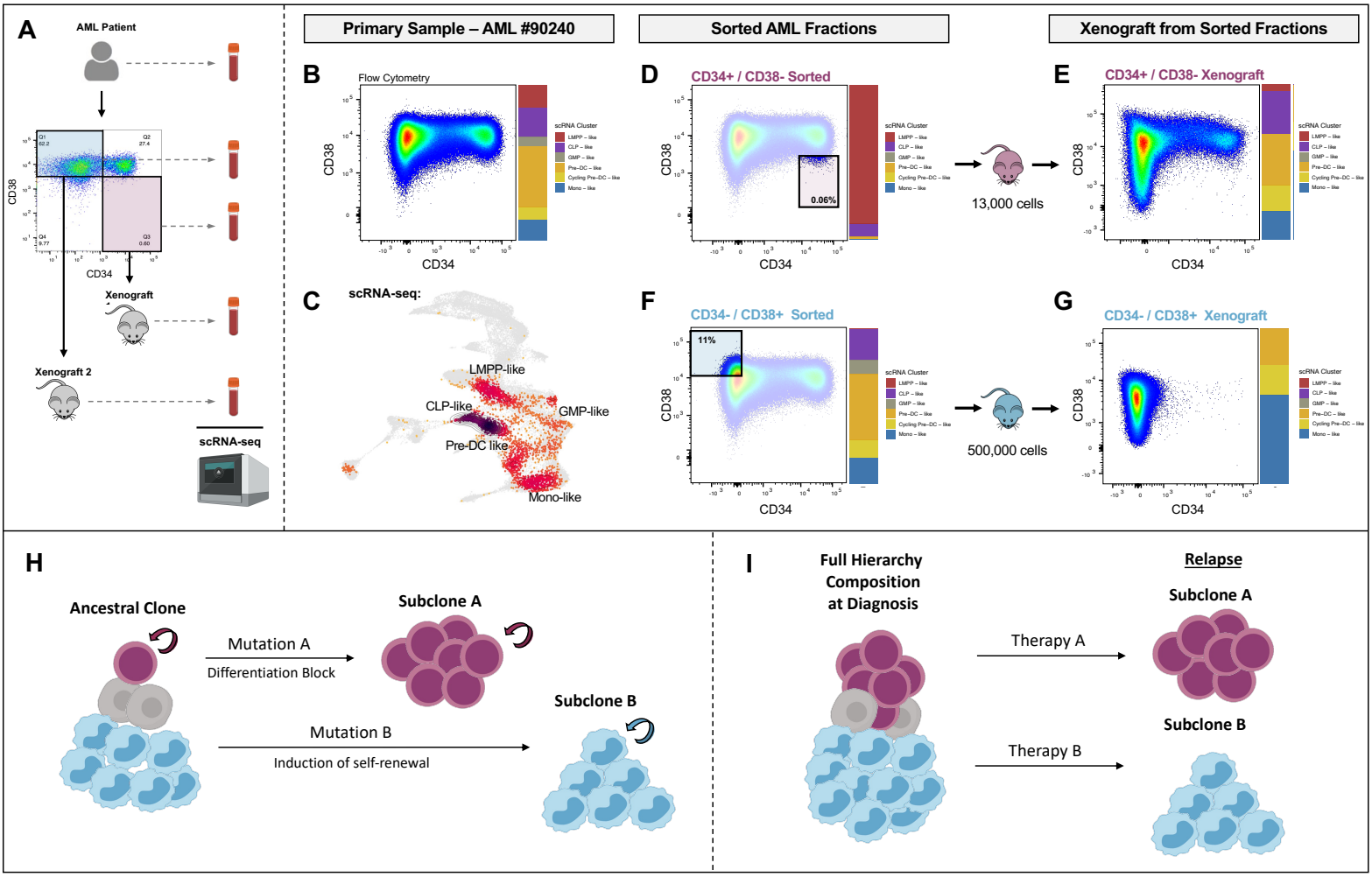

Figure 6. Distinct leukemia cell hierarchies can co-exist within individual AML patients

A) Outline of experimental workflow for identifying co-existing LSC-driven hierarchies in AML. Briefly, the primary AML patient sample, CD34+CD38- fraction, and CD34-CD38+ fraction are profiled by flow cytometry and scRNA-seq. Patient-derived xenografts (PDXs) from the CD34+CD38- fraction and from the CD34-CD38+ fraction are also profiled by flow cytometry and scRNA-seq and the composition of their leukemia cell hierarchies are compared. **B-G)** Results for AML pt #90240. **B)** Immunophenotype and scRNA-seq cell state composition of primary AML. **C)** scRNA-seq projection results of primary AML sample. **D)** Immunophenotype and scRNA-seq composition for CD34+CD38- fraction, representing 0.06% of primary cells. **E)** Immunophenotype and scRNA-seq composition for PDX derived from CD34+CD38- fraction, recapitulating the primary AML hierarchy. **F)** Immunophenotype and scRNA-seq composition for CD34-CD38+ fraction, representing 11% of primary cells. **G)** Immunophenotype and scRNA-seq composition for PDX derived from CD34-CD38+ fraction, restricted to mature AML populations. **H)** Model depicting the co-existence of leukemia cell hierarchies within a subset of AML patients, which could originate from intra-tumoral genetic heterogeneity. **I)** While relapse to conventional therapies is associated with expansion of primitive clones, acquisition of self-renewal at later stages of AML differentiation could enable expansion of self-sufficient monocytic clones at relapse to novel therapies targeting primitive AML cells.