# Transient Differentiation-State Plasticity Occurs during Acute Lymphoblastic Leukemia Initiation

Vera M. Poort[1,2], Rico Hagelaar[1,2], Markus J. van Roosmalen[1,2], Laurianne Trabut[1,2], Jessica G. C. A. M. Buijs-Gladdines[1], Bram van Wijk[3], Jules Meijerink[1], and Ruben van Boxtel[1,2]

ABSTRACT

Leukemia is characterized by oncogenic lesions that result in a block of differentiation, whereas phenotypic plasticity is retained. A better understanding of how these two phenomena arise during leukemogenesis in humans could help inform diagnosis and treatment strategies. Here, we leveraged the well-defined differentiation states during T-cell development to pinpoint the initiation of T-cell acute lymphoblastic leukemia (T-ALL), an aggressive form of childhood leukemia, and study the emergence of phenotypic plasticity. Single-cell whole genome sequencing of leukemic blasts was combined with multiparameter flow cytometry to couple cell identity and clonal lineages. Irrespective of genetic events, leukemia-initiating cells altered their phenotypes by differentiation and dedifferentiation. The construction of the phylogenies of individual leukemias using somatic mutations revealed that phenotypic diversity is reflected by the clonal structure of cancer. The analysis also indicated that the acquired phenotypes are heritable and stable. Together, these results demonstrate a transient period of plasticity during leukemia initiation, where phenotypic switches seem unidirectional.

Significance: A method merging multicolor flow cytometry with single-cell whole genome sequencing to couple cell identity with clonal lineages uncovers differentiation-state plasticity in leukemia, reconciling blocked differentiation with phenotypic plasticity in cancer.
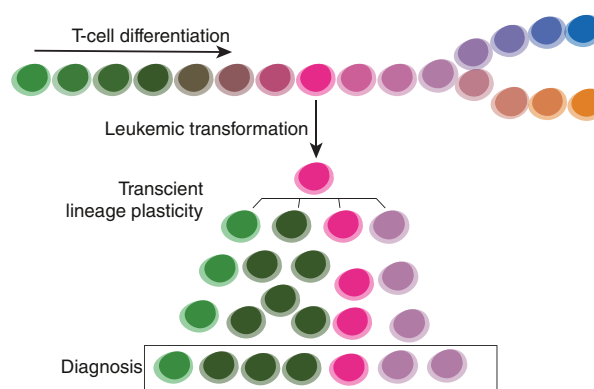
Introduction

Cancer cells exploit phenotypic plasticity to switch between different states (1), such as epithelial and mesenchymal (2), stem-like and differentiated (3), or drug-sensitive and drug-resistance (4), thereby posing significant challenges in cancer treatment (5). Characterizing phenotypic switching between stem-like and differentiated cell states is technically complex because the characteristics of the cell of origin are not necessarily equivalent to the phenotypic characteristics of the cancer stem cell (CSC; ref. 6).

Somatic mutations shared between different cells of the same individual can be used to trace clonal lineages retrospectively (7). This approach allows for timing past events during cancer development in primary human cancer samples. The dynamics of phenotype switching in human cancers can be studied by integrating clonal lineage information with cell identity. For this, adaptations of single-cell transcriptome sequencing protocols that allow for the combined assessment of the cellular state and the clonal composition of cancers have been reported (8, 9). In addition, lineage trees instructed by single-cell ATAC sequencing have revealed the stability of transcriptomic cell states in cancer (bioRxiv 2022.12.28.522128; ref. 10). However, the number of informative genetic variants is low and mainly relies on CNAs or a limited number of cancer driver mutations, which are subjected to selection (9, 11). Ideally, neutral passenger mutations are used as a genetic barcode to trace the clonal lineages of individual cells. To identify these passenger mutations, single-cell whole genome sequencing (scWGS) is required. Clonal expansions have been used to obtain sufficient DNA from a single cell (12–14). However, not all cells, such as leukemic blasts, can clonally expand *in vitro*. Alternatively, whole genome amplification using strand displacement polymerases can be used to obtain sufficient DNA of a single cell. Still, this technology is notorious for introducing amplification biases, artifacts, and allelic dropouts (15). Recently, primary template-directed

[1]Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands. [2]Oncode Institute, Utrecht, the Netherlands. [3]Department of Pediatric Cardiothoracic Surgery, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, the Netherlands.

J. Meijerink and R. van Boxtel contributed equally to this article.

Current address for J. Meijerink: Acerta-Pharma (AstraZeneca), Oss, the Netherlands.

Corresponding Author: Ruben van Boxtel, Princess Máxima Center for Pediatric Oncology, Heidelberglaan 25, Utrecht 3584 CS, the Netherlands. E-mail: r.vanboxtel@prinsesmaximacentrum.nl

amplification (PTA; ref. 16) and complementary bioinformatic approaches (17) improved the accuracy of scWGS and enabled the study of somatic mutagenesis and retrospective lineage tracing of cells with limited proliferation potential *in vitro* (17).

T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive childhood leukemia characterized by genomic rearrangements resulting in a blocked differentiation at specific stages of the T-cell development (18, 19). T-ALL blasts often do not express a T-cell receptor (TCR) on the membrane. However, the clonal V(D)J recombination of the different TCR chain loci is detected in 86% of T-ALL patients (20). These irreversible genetic marks point to the extent of cellular differentiation as each TCR chain is recombined in a different T-cell developmental stage (21). Interestingly, leukemic blast populations spanning different developmental stages are described in ALL (22, 23) and AML (8). Although in AML, phenotypically immature cells are shown to be CSCs (24), CSCs are not identified in ALL. Instead, cells of distinct phenotypes display a leukemia-initiating potential in xenotransplantation experiments (25, 26). Despite the phenotype, these cells reconstitute a phenotypically heterogeneous leukemia in mice (25, 26). However, it remains unclear to what extent lineage plasticity plays a role in human ALL progression.

Therefore, we used T-ALL as a model disease to study phenotypic plasticity and leukemic differentiation stage arrest. We show that diverse mutational processes or driver mutations do not drive phenotypic diversity. Instead, the phylogenetic analysis indicates a brief period of plasticity at leukemia initiation, where cells acquire a variety of phenotypes, which then become fixed. This short period of plasticity could also be detected at relapse initiation. Our work provides insight into the *in-situ* initiation and plastic progression of ALL.

Materials and Methods

Samples

Pediatric thymic tissues from nonleukemic patients undergoing cardiac surgery were obtained. The patient's parents or legal guardians provided written informed consent to use leftover material according to the study protocol TCbio-18-181 approved by the ethical committee and the biobank of the Utrecht University Medical Center (the Netherlands) and in accordance with the Declaration of Helsinki. Briefly, thymus biopsies were mechanically disrupted in an RPMI-1640 medium (#21875034, Gibco) supplemented with fetal bovine serum (#A4766801, Gibco) to obtain a single-cell suspension of thymocytes and immediately used for the flow cytometry analysis. Contaminating red blood cells were removed using an RBC lysis buffer (#420301, BioLegend) per the manufacturer's protocol. If indicated, the cell suspension was partially depleted of the CD4/CD8 double-positive (DP) thymocyte subsets by magnetic bead separation. For this, 7 × 107 cells were stained for 30 min on ice with 20 μL CD4 MicroBeads human (#130-045-101, Miltenyi Biotec), 20 μL CD8 MicroBeads human (#130-045-201, Miltenyi Biotec), and 20 μL CD3 MicroBeads human (#130-050-101, Miltenyi Biotec). Cells were added to LS columns (#130-042-401, Miltenyi Biotec), and flow through was used for the subsequent flow cytometry experiments.

Primary blasts from 33 T-ALL patients were included in this study (Supplementary Table S1). The patient's parents or legal guardians provided written informed consent to use leftover diagnostic material for research, with approval from the institutional review board of the Erasmus MC Rotterdam and in accordance with the Declaration of Helsinki (MEC-2004-203, MEC-2012-287). Leukemia cells were harvested from blood or bone marrow biopsies and enriched to a purity of at least 90%, as described previously (19). Genetic subtypes were reported previously (19).

High-parameter flow cytometry

Experiments were performed using the Bio-Rad ZE5 (Bio-Rad). Hereto, 500,000 cells were washed with PBS and stained for viability for 30 min (Zombie NIR Fixable viability kit, #423105, BioLegend). Then, the cells were stained in Brilliant Stain Buffer (#563794, BD Biosciences) with a cocktail of antibodies for 30 min on ice. Compensation and initial data analysis were performed using FlowJo v10.7.1 (RRID:SCR_008520).

Antibodies used for 17-parameter flow cytometry were as follows: CD4-PE/Cyanine7 (1:500, clone OKT-4, BioLegend Cat# 317414, RRID:AB_571959), CD3-BUV496 (1:500, clone: UCHT1, BD Biosciences #745740, RRID: RRID:AB_2743211), CD11c-BUV661 (1:200, clone: B-ly-6, BD Biosciences Cat# 565067, RRID: AB_2744275), CD5-BUV737 (1:500, clone: UCHT2, BD Biosciences Cat# 564451, RRID:AB_2714177), CD44-BV510 (1:1,000, clone: IM7, BioLegend Cat# 103044, RRID:AB_2650923), CD123-BV605 (1:200, clone: 6H6, BioLegend Cat# 306025, RRID:AB_2562115), CD14-BV650 (1:500, clone: M5E2, BD Biosciences Cat# 563419, RRID:AB_2744286), CD8a-BV711 (1:500, clone: RPA-T8, BioLegend Cat# 301043, RRID:AB_11218793), CD7-BV786 (1:500, clone: MT701, BD Biosciences Cat# 740964, RRID:AB_2740589), CD25-BB515 (1:200, clone: 2A3, BD Biosciences Cat# 564467, RRID:AB_2744340), CD56-BB700 (1:500, clone: MY31, BD Biosciences Cat# 566574, RRID:AB_2743411), CD1a-PE (1:1000, clone HI149, BioLegend Cat# 300105, RRID:AB_314019), CD19-PE/Dazzle594 (1:500, clone: HIB19, BioLegend Cat# 302251, RRID:AB_2563559), TCRα/β-PE/Cyanine5 (1:1000, clone: IP26, BioLegend Cat# 306710, RRID:AB_314648), TCRγ/δ-APC (1:500, clone: 11F2, Miltenyi Biotec Cat# 130-113-500, RRID: AB_2733463), CD45-AF700 (1:1000, clone: 2D1, BioLegend Cat# 368513, RRID:AB_2566373), and Zombie NIR (1:6000, BioLegend Cat# 423105).

Index sorting

Blast populations were purified on an SH800S Cell Sorter (Sony, RRID: SCR_018066). Populations were identified using the following surface markers: DNearly: $CD7^+CD4^-CD8^-CD1a^-$, DN3: $CD7^+CD8^-CD4^-CD1a^+$, iSPCD4: $CD7^+CD8^-CD4^+CD1a^+$, DP: $CD7^+CD8^+CD4^+$, SPCD4: $CD7^+CD8^-CD4^+CD1a^-$, and SPCD8: $CD7^+CD8^+CD4^-$. Single-cell blasts were sorted into DNA low-bind plates, each well-containing 3 μl of PTA cell buffer according to manufacturer's protocol (ResolveDNA V1 Consumables, #100180, Bioskryb Genomics).

Antibodies used for the index sort panel were as follows: CD8a-BV421 (1:400, clone: RPA-T8, BioLegend Cat# 301036, RRID: AB_10960142), CD4-FITC (1:200, clone: VIT4, Miltenyi Biotec Cat# 130-098-160, RRID:AB_2660912), CD1a-PE (1:400, CD7-PE/Cyanine7 (1:200, clone: M-T701, BD Biosciences Cat# 564019, RRID: AB_2738545), CD16-APC (1:200, clone: 3G8, BioLegend Cat# 302012, RRID:AB_314212), TCRγ/δ (1:200), and AnnexinV-APC (1:200, BioLegend Cat# 640920, RRID: AB_2561515).

Flow cytometry data analysis

Flow cytometric .fcs files were loaded in R using the ReadFCS function from the FlowCore package (RRID:SCR_002205). At first,
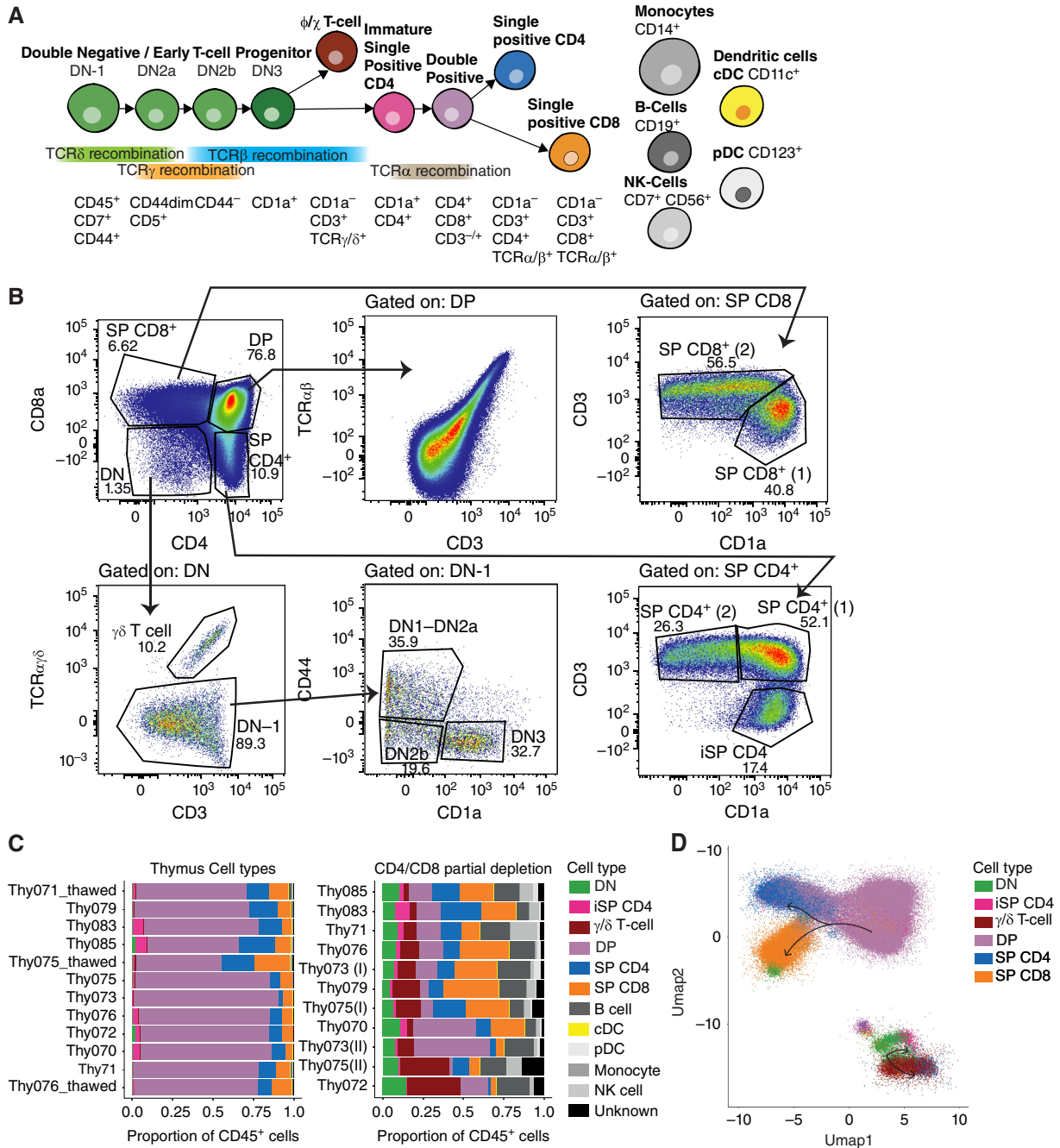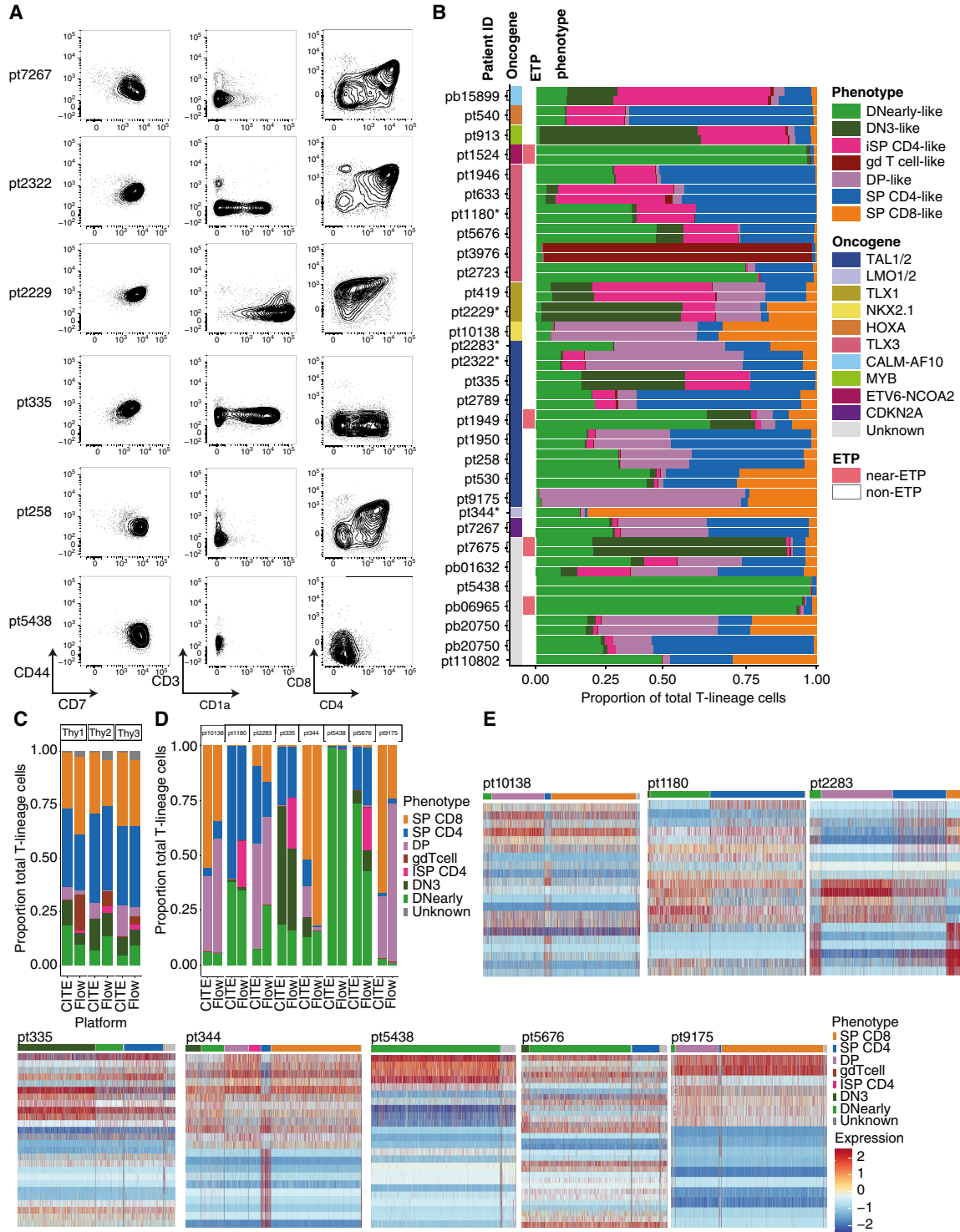
Figure 1.
Thymocyte developmental stages identified by multiparameter flow cytometry. **A**, Schematic overview of the thymocyte development and stages of active V(D)J recombination. Epitope markers used to identify the respective cell types are listed. **B**, Conventional two-dimensional gating strategy to identify the main thymocyte maturation stages. Percentages are based on the parent population, as indicated by the arrows. Cells of the representative sample (Thy070) are shown. **C**, Left, cell types found per sample by semiautomatic gating as the proportion of the total CD45[+] cells. Right, cell types found per sample after partial depletion of CD4/CD8 DP cells. cDC, conventional dendritic cell; DN, CD4/CD8 double-negative T-cell; DP, CD4/CD8 double-positive T-cell; iSP CD4, immature single positive CD4 T-cell; pDC, plasmacytoid dendritic cell; SP CD4, single positive CD4 T-cell; and SP CD8, single positive CD8 T-cell. **D**, Uniform manifold approximation and projection (UMAP) for dimension reduction of t-lineage cell types identified in samples of nine nonleukemic donors. Bar plot of cell types assigned using our semiautomatic gating strategy. Per sample, 5,000 cells were plotted.

quality control was performed on raw flow cytometry data using the package flowAI. After that, manual gates were used to remove debris (FSC-A vs. SSC-A), doublets (FSC-A vs. FSC-H and SSC-A vs. SSC-W), dead cells (ZombieNIR+), and CD45[−] cells. The CD45[+] population was used in further bioinformatic analysis (Supplementary Figs. S1A and S1B).

Figure 2.
T-ALL blasts are immunophenotypically heterogeneous. **A**, Two-dimensional gating plots of six representable T-ALL primary diagnostic samples. Non-T lineage cells were excluded prior to plotting. **B**, Bar graph of 28 primary diagnostic T-ALL samples. Colors represent the proportion per assigned cell type after excluding non-T lineage cells. Samples were sorted on their oncogenic driver identified at diagnosis. **C**, Proportions of cell types found in CITE-seq data set vs. flow cytometry from the same nonleukemic thymus sample. **D**, Proportion of cell types found in data sets acquired with CITE-seq and flow cytometry from the same T-ALL patient. **E**, Differential gene expression of phenotypic subpopulations per sequenced patient. Columns represent cells, whereas rows represent the top five differentially expressed genes per subpopulation.

First, outliers were removed by excluding each channel's top and bottom 0.1% signal. After that, data were biexponentially transformed according to standard flow jo settings using the flowjo_biexp function of the flowWorkspace package (RRID: SCR_001155; channelRange = 4096, maxValue = 300,000, pos = 4.33, neg = 1, widthBasis = −10). Data were randomly down-sampled to 5,000 cells per replicate. Cell types were determined by applying semiautomatic gating using manually assessed cut-off values (Supplementary Tables S1B and S2A). Additionally, di-mensional reduction was done using uniform manifold approxi-mation and projection for dimension reduction. Code and descriptions are available here: https://github.com/ToolsVanBox/. Results were confirmed in two independent data sets for patients for which data were available. First, the EuroFlow diagnostics data ($n = 6$) and expression of epitopes overlapping in both the flow cytometry panel and EuroFlow panel (9/17) were assessed. Addi-tionally, data were compared with the expression data from our Cellular Indexing of Transcriptomes and Epitopes sequencing (CITE-seq) data set.

CITE-seq library generation

CITE-seq was performed using 10× Genomics Immune Pro-filing, including V(D)J amplification for human T-cells kit V3 (1000699, 10× Genomics) according to the manufacturer's pro-tocol. We used TotalSeq-C lyophilized antibody cocktail for epi-tope recognition, including 204 antibodies (Biolegend; Supplementary Table S3). Thymus samples were prepared fresh, and T-ALL samples were thawed. When viability was <90%, dead cells were removed using the Dead Cell Removal Kit (Milteny). Single-cell libraries were sequenced using NovaSeq 6000 (Illu-mina). The Cell Ranger software pipeline (10× Genomics, v5) was used to demultiplex cellular barcodes, map reads to the human genome (GRCh38), and transcriptome using the STAR aligner and produce a matrix of gene counts versus cells V(D)J sequences and CITE-seq counts.

CITE-seq data analysis

Low-quality cell filters were set to exclude cells with <300 genes, <500 UMIs, and <40% mitochondrial reads. Antibody aggregates with >50,000 antibody reads/cell genes in less than three cells were removed. Subsequently, the RNA expression was log-normalized, and the epitope expression was CLR-normalized (setting margin = 2) using Seurat v4 (RRID:SCR_016341). Phenotypic subpopulations were identified using the same funnel principle as used for flow cytometry (Supplementary Fig. S1C). Differential expression analysis was performed using the Seurat v4 option FindAllMarkers. Before dimensional reduction, cell cycle genes and mitochondrial genes were removed from the dataset. Endocytosis score was based on the Gene Set Enrichment Analysis (RRID:SCR_003199) endocytosis gene list: https://www.gsea-msigdb.org/gsea/msigdb/cards/KEGG_ENDOCYTOSIS.

Shannon diversity index

Shannon diversity index was calculated for the number of an-notated T-cell populations (CD7$^+$) per sample using the R package vegan (RRID:SCR_011950). The cut-off value for high or low Shannon index was determined using the R package "cut-off."

Germline controls

Mesenchymal stromal cells (MSC) were cultured from the bone marrow samples in DMEM-F12 medium (#31053044, Gibco) supplemented with 10% FBS (Gibco) and 1% GlutaMax (#35050061, Gibco). MSCs could be harvested when confluent (after approxi-mately 3 weeks).

If MSCs did not grow out, we used sorted innate cells (CD14$^+$ or CD16$^+$) as an alternative for germline control. Because we are only comparing leukemic cells among each other, mutations acquired early in hematopoiesis are not important for our analysis.

DNA isolation and library generation

Bulk DNA was isolated, as previously reported (27). PTA single-cell whole genome amplification was performed according to the manufacturer's protocol (#100180, BioSkryb Genomics) and as previously reported (17). The libraries were sequenced at a depth of 15× for single-cell samples and 30× for bulk samples.

Mutation calling and filtering

Somatic mutation calling samples were performed as previously reported (27). Thereafter, our in-house developed pipeline PTATO (RRID:SCR_025353) was applied to call high-quality somatic mu-tation calls, as well as indel and structural variants with low PTA-induced artifacts (17). Full code and description are found here: https://github.com/ToolsVanBox/.

Driver events

Potential driver events were identified as previously reported (27) and filtered against previously reported events in ALL ge-nomic cohorts (18, 28). All driver SNVs and indels were validated by hand using the Integrative Genomics Viewer (IGV; RRID: SCR_011793).

Mutational load and mutational pattern analysis

Mutation load, single base substitution, and indel counts were normalized to GATK (RRID:SCR_001876) CallableLoci's CALL-ABLE length. The baseline data from previous publications were used (12, 27–29). Mutations were normalized for age by (number_of_mutations–baseline_intercept)/baseline_slope. Muta-tional pattern analysis was performed as previously reported (27) using MutationalPatterns v3.0.1 (RRID:SCR_024247). We identi-fied only signatures previously reported in COSMIC (RRID: SCR_002260).

Constructing phylogenetic trees

Tree shapes were generated using the CellPhy package (30) based on the most likely path, and then manually checked. Thereafter, high-confident somatic variants were assigned to each branch.

For this, filtering was slightly adjusted compared with previous analyses. The root comprised all mutations identified in the bulk sample and missed in <2 single cells. Mutations annotated to sub-branches had to be found in all cells per branch −1. Clonal muta-tions unique to one sample at the exterior branch were additionally filtered to be absent in all other samples or $n − 1$, where 1 is the sub-clonal presence in one other sample. All shared mutations were manually inspected in IGV, and false-positive results were filtered out.

The relative contributions of phenotypic subpopulations are calculated by multiplying the mean variant allele frequency (VAF) of mutations by two and multiplying by the proportion of the subpopulation, as detected by flow cytometry at diagnosis. This is
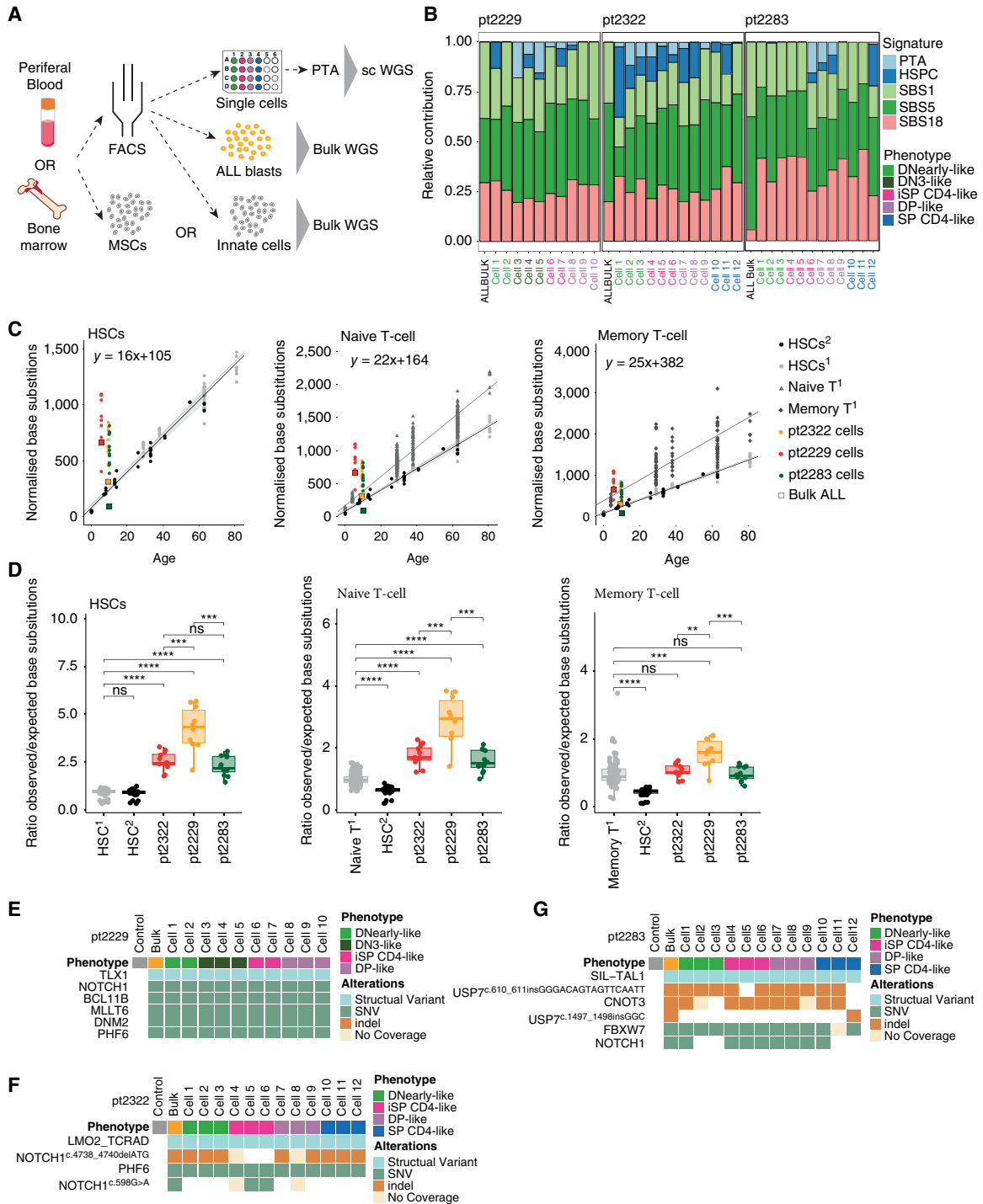
Figure 3.
Assessing somatic mutations in the genomes of single T-ALL blasts. **A**, Schematic overview of sorting and sequencing strategy. **B**, Relative contribution of the COSMIC mutational signatures found per cell of pt2229, pt2322, and pt2283. *X*-axis labels are colored by cell phenotype. **C**, Base substitutions found per T-ALL cell plotted on the age line of HSC, naïve T cells, and memory T cells. Each dot represents a single cell, whereas the squares represent the bulk WGS samples. HSCs2 are hematopoietic stem cells published by Osorio colleagues (12), HSCs1, T-Naïve, and T-memory were published by Machado and colleagues (29). **D**, Quantification of observed mutations in T-ALL patient cells vs. expected mutations in nonleukemic cells corrected for their respective age. Populations were compared using a one-way repeated-measures ANOVA and Bonferroni corrected *t* test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant. **E-G**, Oncoplot showing potential driver structural variants, indels, and SNVs previously described in ALL found per sequenced T-ALL cell.

then normalized to the total sample by dividing by the sum of the contributions.

Recall rate

The recall rate was calculated by assessing the presence of true positive variants, VAF > 0.15 in bulk ALL sample, in single cell PTA samples. We corrected for subclonal diversity by assessing whether a variant was missed by specific branches, where variants that were missed by a complete branch, but were present in another, were labeled as true negatives.

This calculation was not possible for single-cell pt2283-TALL12 as this blast harbored mostly unique mutations; therefore, the lack of mutations could not be confidently called a true negative.

V(D)J recombination analysis

For V(D)J recombination calling, MiXCR v4.3.2 (RRID: SCR_018725) was used with -RNA full-length analysis option, —species hsa —dna to identify reads containing CDR3 regions, their relative abundances, and specific V(D)J gene usage. False positives were filtered out by removing the clone fraction < 0.02 for delta, gamma, and beta sequences. For alpha sequences, the variants were manually checked using BAF and copy number plots. CDR3 sequences with large insertions or deletions (represented as "_") or premature stop codons (represented as "*") are denoted as nonfunctional. Code and descriptions are available here: https://github.com/ToolsVanBox/.

Statistical testing

Unless stated otherwise, all statistical testing procedures were performed in R using the following packages: rstatix (RRID: SCR_021240), ggpubr (RRID:SCR_021139), and ggstatsplot.

Data visualization

If not specified otherwise, all data visualizations were performed using ggplot2 (RRID:SCR_014601) and ggpubr.

Data availability

Raw flow cytometry data are available through FlowRepository.org at Repository ID: FR-FCM-Z5TV (thymocyte subtyping data) and Repository ID: FR-FCM-Z5UY (T-cell leukemia immunophenotyping data).

The WGS data and CITE-seq data are available through The European Genome-phenome Archive (EGA; https://ega-archive.org/) at EGAS00001007446 and EGAS50000000358. All code and pipelines used are available through https://github.com/ToolsVanBox. Additional code is available from the lead contact and will be provided upon request.

Results

Assessing T-lymphoid differentiation stages in pediatric nonleukemic donors

To quantify the differentiation states of T-ALL cells, we first established an immunophenotypic reference dataset of nonleukemic thymocytes. For this, a 17-parameter flow cytometry panel, including myeloid and lymphoid cell markers, was developed and tested on thymocytes of nine independent infant donors (**Fig. 1A**; Supplementary Table S1A). Using this panel, we identified all conventional T-cell populations by manual gating, although some

populations are sparse (<1% of the total population; **Fig. 1B**; Supplementary Fig. S1B).

To better represent the sparse cell populations, the CD4[+]/CD8[+] DP population was partially depleted using magnetic beads (**Fig. 1B and C**). As the epitope densities of CD4 and CD8 combined are higher on DP cells than on SP cells, DP cells were more efficiently depleted. This led to a reduction of 81% of the DP population with an average proportion of 0.17 (range, 0.02–0.47) DP cells left. Correspondingly, the proportions of CD4–CD8– double-negative (DN), γδ T-cell, and immature SP CD4 (iSP CD4) populations, as well as non-T lineage cells, were enhanced (**Fig. 1C**).

To automate differentiation stage annotation for leukemic cells, we developed a semiautomatic gating strategy assigning cell annotations in healthy thymocytes using a funnel principle (Supplementary Fig. S1C; Materials and Methods). We compared the semiautomatic gating results with manually gated populations to test the reliability of this approach. The overall accuracy of our semiautomatic gating strategy was 90.45% (CI, 90.43%–90.47%). If a cell was misclassified, the misclassified cell was most likely to decrease into the population preceding or directly following the actual population in the developmental hierarchy (Supplementary Table S1C). Dimensional reduction analysis shows a continuum of differentiation with cell types transitioning toward the next (**Fig. 1D**; Supplementary Fig. S1D), which suggests that the biological effect of misclassification is likely to be small.

T-ALL harbors multiple phenotypic subpopulations at diagnosis

Next, we used the same flow cytometry panel to analyze 33 blast-enriched primary peripheral blood (PB) or bone marrow samples of pediatric T-ALL patients (>85% blasts). Patient information is available in Supplementary Table S2B. Examination of the expression of the T-lymphocyte development epitopes showed inter-patient differences and partial positivity for CD1a, CD4, and CD8 (**Fig. 2A**). Our semiautomatic gating strategy was used to classify the differentiation states of all CD7[+] cells in the samples (Supplementary Fig. S2A). Of note, patient samples showed a continuum of gradual transitions of blasts over T-cell development stages (**Fig. 2A**), which was overlooked in our annotation strategy. Following the median blast count of 98% (IQR = 94–99), we define immunophenotypic subpopulations as populations with a size of >10%. Intratumor phenotypic heterogeneity was observed in 26 out of the 31 T-ALL patients (83.8%), with an average of 2.4 subpopulations per sample (SD 0.99; **Fig. 2B**). Patients with only one phenotypic blast population had a T-ALL with an immature phenotype ($n = 4$) or a γ/δ T-cell phenotype ($n = 1$). This lack of heterogeneity could be because of the limited epitope markers used to differentiate between immature and γ/δ-developmental stages. The other 26 patients harbor two to four T-lymphoid developmental stages (**Fig. 2B**). Interestingly, the most immature "DN-like" immunophenotype was found in all samples, suggesting that developmental hierarchy was maintained. Unexpectedly, we also noticed intrapatient heterogeneity among patients with the same genetic subtype (**Fig. 2B**).

Additionally, the major subpopulation in each sample was associated with their respective maturation stages as previously defined by Liu and colleagues (Supplementary Table S2; Supplementary Fig. S2B; ref. 18). Diversity in phenotypic populations was higher in patients with cortical and postcortical maturation stages (Supplementary Fig. S2C) and did not predict survival outcomes (Supplementary Fig. S2D). Interestingly, intratumor heterogeneity was

inversely correlated with age in the preand postcortical maturation stages, but increased with age in cortical patients (Supplementary Fig. S2E). Our results were validated using EuroFlow diagnostic panels (31) and the CITE-seq dataset, which showed similar results (Supplementary Fig. S3).

To further characterize the heterogeneity in phenotypic cell states, CITE-seq data were generated for eight T-ALL patients and three nonleukemic thymocyte donors. Phenotypic subpopulations were annotated by applying the same funnel principle to the epitope data set (Supplementary Fig. S1C), and comparable phenotypes were found in CITE-seq data for the thymocyte and T-ALL samples (**Fig. 2C**). Notably, γδ T-cell populations were not found in the thymocytes CITE-seq data (**Fig. 2C**) likely because of the encapsulation method used to generate the libraries. Moreover, "iSP CD4-like" populations were more often found in the patient samples when analyzed with flow cytometry (**Fig. 2D**). Reassuringly, the expression of epitope markers in T-ALL patients accompanied the expression of the respective gene (Supplementary Fig. S4A). Unlike nonleukemic mature T-cells, when plotted in a high-dimensional space, most T-ALL cells do not overlap with developing T cells. This suggests that the intensity and the coexpression of certain epitopes differed from T-ALL to healthy developing T cells (Supplementary Fig. S4B). Similar outcomes were observed between chromatin accessibility of T-ALL and T-cell precursors (32). Endocytosis gene scores were at basal levels and did not increase in more immature cell types (Supplementary Fig. S4C), excluding aberrant endocytosis as an underlying mechanism for the promiscuous expression of epitope markers.

Differential expression analysis of subpopulations across patients revealed mostly patient-specific differences (Supplementary Fig. S4D). Moreover, patient-specific differential expression analysis showed minimal transcriptional heterogeneity in immunophenotypic subpopulations (**Fig. 2E**), albeit immature subset harbor expression of genes found in immature thymocytes *KLRB1* and *FXYD2* in "DNearly-like" cells and *CD1A-D* and *ALDH1A2* in "DN3-like" cells. Likewise, *PTCRA*, *ELOVL4*, and *IL7R* were highly expressed in the more mature "DP-like" and "SPCD4-like" cells and *NKG7* and *CCL5* in the "SPCD8-like" cells. We then extracted a differentiation stage gene set from the thymocyte reference for each subpopulation. Gene set scoring showed a clear correlation between phenotypic subpopulations and their respective differentiation stage in pt2283, pt335, pt1975, and pt10138 (Supplementary Fig. S4E). However, this correlation could not be observed for the other four patients. This could be because of the T-ALL subtype differences as NKX2.1 and TAL1 subtypes correlate, and LMO2 and TLX3 subtypes do not.

Assessing somatic mutations in the genomes of single T-ALL blasts

We next questioned whether genetic determinants drive the immunophenotypic diversity in the T-ALL patients. To address this, we performed WGS of single blasts of the various immunophenotypic populations (**Fig. 3A**). For this, we used PTA to obtain enough DNA of a single blast for the WGS analysis with nucleotide resolution without the need for *in vitro* clonal expansion (16). We selected three patients that were driven by the commonly identified TLX1, LMO2, and TAL1 structural variants (i.e., pt2229, pt2322, and pt2283, respectively; ref. 28). In all three patients, multiple phenotypic subpopulations were identified (**Fig. 2B**). We single-cell sorted four distinct phenotypic subpopulations per patient and analyzed two to three cells per immunophenotypic subpopulation per patient (Supplementary Fig. S5A). In total, we performed a WGS

analysis of 34 single cells. Our in-house developed machine learning-based PTATO tool was used to effectively filter PTA artifacts and obtain accurate catalogs of base substitutions, indels, and structural variants (SV; ref. 17). We observed an average of 823.2 (SD ± 221.6) SNVs in each cell of pt2229 and 651.5 (SD ± 121.7) somatic variants in pt2322 (Supplementary Fig. S5B). The higher mutation load in TLX1 T-ALL was in concordance with previous studies (28). Moreover, the contribution of the distinct PTA artifact signature (33) was low (average of 4.3%, SD ± 5.08), indicating a low fraction of false-positive artifacts introduced by PTA (**Fig. 3B**). We performed a WGS analysis on the bulk T-ALL sample to calculate the recall rate of somatic mutations in each assessed single cell. The median recall rate was 95.8% (IQR = 85.1–97.6), 95.5% (IQR = 91.7–96.7), and 96.0% (IQR = 95.1–96.9) for pt2229, pt2322, and pt2283, respectively (Supplementary Fig. S5C). Additionally, the mutation load in these PTA samples was comparable to the mutation load previously reported in T-ALL samples (Supplementary Fig. S5D). We found comparable mutational load and mutational signature to clonally expanded memory and naïve T-cells (**Fig. 3B–D**; ref. 29). Taken together, these comparisons validated the reliability of our PTA-based WGS data.

Immunophenotypic subpopulations do not harbor additional genetic drivers or distinctive mutational loads

First, we assessed if genetic heterogeneity explains the phenotypic heterogeneity we observed. Besides SVs, 112 additional missense and truncating mutations were identified, of which, mutations in *NOTCH1*, *PHF6*, *BCL11B*, *MLLT6*, *MHS2*, and *DNM2* were previously reported as driver genes in T-ALL (Supplementary Fig. S5E–G; Supplementary Table S4). Importantly, these mutations in known driver genes were mostly clonally present in the T-ALL and not unique to one blast or a subgroup of blasts (**Fig. 3E–G**). A notable exception was *NOTCH1*, in which two independent driver mutations were found in the two major T-ALL subclones of pt2322. This phenomenon indicated convergent evolution and underlined the strong selective pressure on acquiring *NOTCH1* mutations during T-ALL leukemogenesis (34). However, because all cells of this T-ALL harbor an activating *NOTCH1* mutation, we deemed it unlikely that *NOTCH1* (alone) drives the phenotypic diversity we observed here.

Moreover, a subclonal *MSH2* mutation was found in pt2322 (Supplementary Fig. S5F). However, this mutation did not result in a complete loss of the encoding protein, and we found no evidence of a mismatch repair-deficient mutational signature in the mutated cells (**Fig. 3B**). Therefore, it is unlikely that this specific mutation plays a role in the progression of this T-ALL. In this regard, no differences in mutational signatures (**Fig. 3B**) and mutational load (Supplementary Fig. S5B) were observed between the subpopulations. In summary, the driver mutations identified in pt2229, pt2322, and pt2283 were mostly clonal, suggesting a nongenetic mechanism for phenotypic switching.

Combining phylogenic and phenotypic information reveals differentiation state heritability

To elucidate when phenotypic heterogeneity was introduced and how plastic leukemic cells were, we generated phylogenetic trees from the scWGS data to integrate cell state heterogeneity with the cancer evolution model (**Fig. 4**; refs. 12, 13, 35).
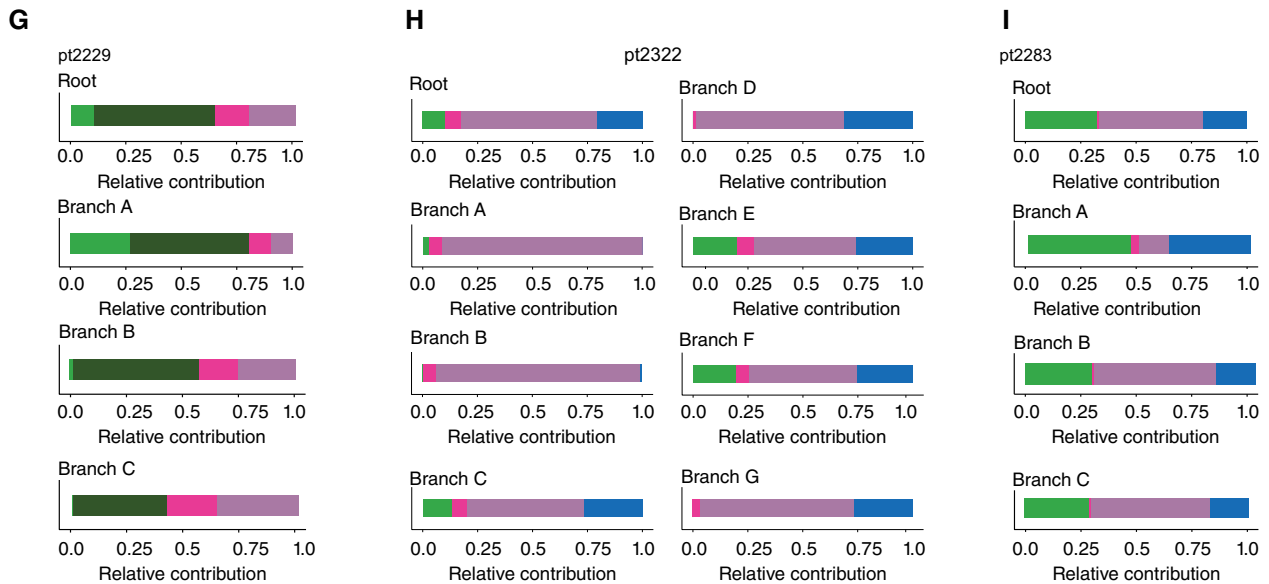
For pt2229, 893 clonal mutations were detected in all single cells and the bulk ALL. These mutations made up the root of the tree (**Fig. 4A**; Supplementary Fig. S6A). The subsequent subclonal branches (branch A & branch B) harbored fewer shared mutations
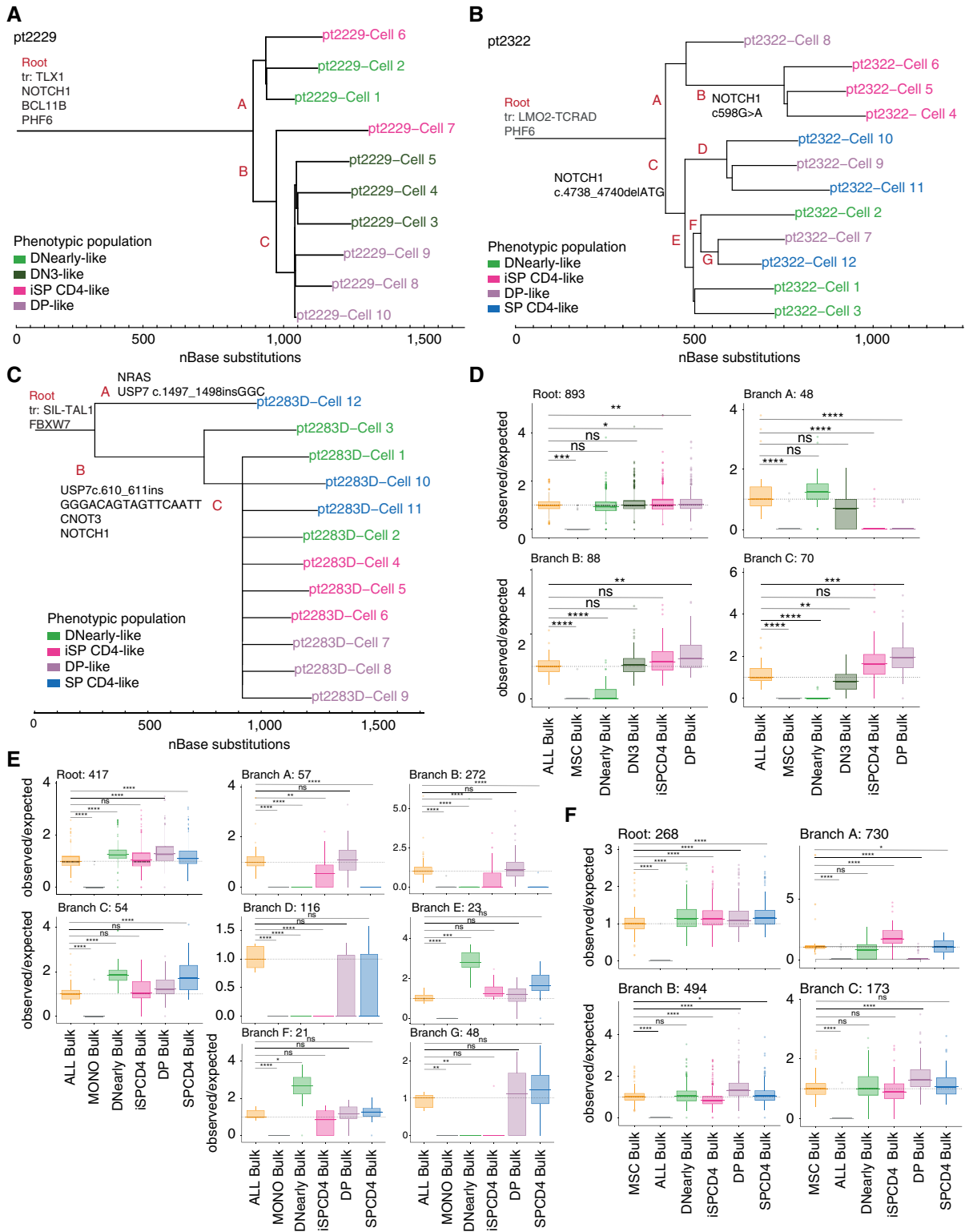
Figure 4.
Lineage tracing reveals heritable differentiation states of leukemic blasts. **A–C**, Phylogenetic tree of T-ALL single blasts of respective patient. Length of the branches indicates the number of somatic base substitutions. Colors indicate the phenotypic subpopulation. Letters address the different branches. **D–F**, VAF contribution of branch-specific mutations in sorted bulk populations normalized to the ALLBULK. Differences in VAF per variant in each population were compared using a one-way repeated-measures ANOVA and Bonferroni corrected *t* test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant. *(Continued on the following page.)*

G

pt2229

Root

Branch A

Branch B

Branch C

H

pt2322

Root

Branch A

Branch B

Branch C

Branch D

Branch E

Branch F

Branch G

I

pt2283

Root

Branch A

Branch B

Branch C

Figure 4.
(*Continued.*) **G–I**, Relative contribution of each phenotypic subpopulation per branch corrected for population size at diagnosis as assessed in **Fig. 2B**.

(48 and 88, respectively). This particular evolution shape corresponded to a punctuated evolution model and was also observed in pt2283 (36). This model suggests that the most recent common ancestor (MRCA) started expanding after a relatively long latency.

For normal T-cell differentiation, we expected a random distribution of differentiation stages among branches because of continuous gradual lineage-specific differentiation. In contrast, T-ALL blasts of the same phenotypes resided closely in the phylogenetic tree. This suggests that phenotypes are heritable, and that plasticity is a transient and early event in leukemia progression (bioRxiv 2022.12.28.522128).

The same observation was made in the other two patients we assessed by scWGS (**Fig. 4B** and **C**; Supplementary Fig. S6B and S6C). Additionally, in pt2322, branch B with "iSPCD4-like" cells harbored a 1.3-fold higher mutation burden than the other blasts. Differences in mutational signatures could not explain this increased mutation load (Supplementary Fig. S6D). In pt2283, we observed a small genetic subclone reflecting 16% of the ALL as an early split of branch A. The phenotypic heterogeneity at the end branches of branch B suggests that phenotypic plasticity occurred after branch A branched off.

To confirm the structure of the phylogenetic trees and validate the nonrandom distribution of immunophenotypes across the branches, we performed bulk WGS on sorted phenotypic populations. We determined what portion of these populations could be attributed to the various branches (Materials and Methods). In line with our single-cell WGS data, the phenotypically distinct bulk populations distributed across the different branches were highly biased (**Fig. 4D–F**). For example, the "DNearly-like" population in pt2229 could almost exclusively be attributed to branch A. In contrast, branch C predominantly contributed to the more differentiated "iSP CD4-like" and "DP-like" populations (**Fig. 4D**). Additionally, the relative contribution of each phenotypic subpopulation toward each branch was calculated by considering the proportions of the phenotypic subpopulations annotated by our semiautomated gating strategy (**Fig. 4G–I**). The bulk sequencing

data confirmed that cells become restricted to their phenotype and seem blocked in different differentiation states.

V(D)J recombination directs toward the MRCA

As discussed above, we observed blocked differentiation of T-ALL blasts in diverse phenotypic cell states. However, knowing that the phenotypic cell state of the MRCA was essential to elucidate whether the observed phenotypic diversity was attributed to arrested differentiation or dedifferentiation.

We, therefore, analyzed the clonality of V(D)J recombination to pinpoint the most likely differentiation stage of the MRCA. Developing T-cells of the α/β lineage recombined four TCR gene loci during specific stages of T-cell development, namely, the *TCR delta* (*TRD*) in the DNearly stage, the *TCR gamma* (*TRG*) in the DN3 stage, and the *TCR beta* (*TRB*) initiated in the DN3 stage and completed in the iSPCD4 stage. Lastly, *TCR alpha* (*TRA*) recombination was completed in the DP stage (**Fig. 1A**; ref. 21). The *TRA* and *TRD* genes were located in the same region of chromosome 14, and as a result of *TRA* recombination, the *TRD* was spliced out (21). Because the chance of getting a productive TCR was ~25% to 30%, many cells had a biallelic recombination (37). In lymphoblastic leukemia, VDJ recombination is often used for minimal residual disease testing and seems clonal (38).

The WGS data were interrogated for productive and nonproductive VDJ recombination, as recombination serves as a proxy for the stage of the developmental arrest of the MRCA. In all patients, *TRD*, *TRG*, and *TRB* had monoclonal rearrangements among the single blasts and bulk-sorted populations (**Fig. 5**; Supplementary Table S5; Supplementary Fig. S6E–G). Interestingly, in pt2229, we observed unique rearrangements of the *TRA* for blasts 8, 9, and 10 with a "DP-like" phenotype and polyclonal *TRA* recombination in the bulk "DP-like" population (**Fig. 5A**; Supplementary Table S5). This indicates that the MRCA rearranged the *TRD*, *TRG*, and *TRB* loci before initiating leukemia, and that the phenotypic cell state of the MRCA was just before the DP developmental stage. This shows that leukemic blasts can further differentiate to
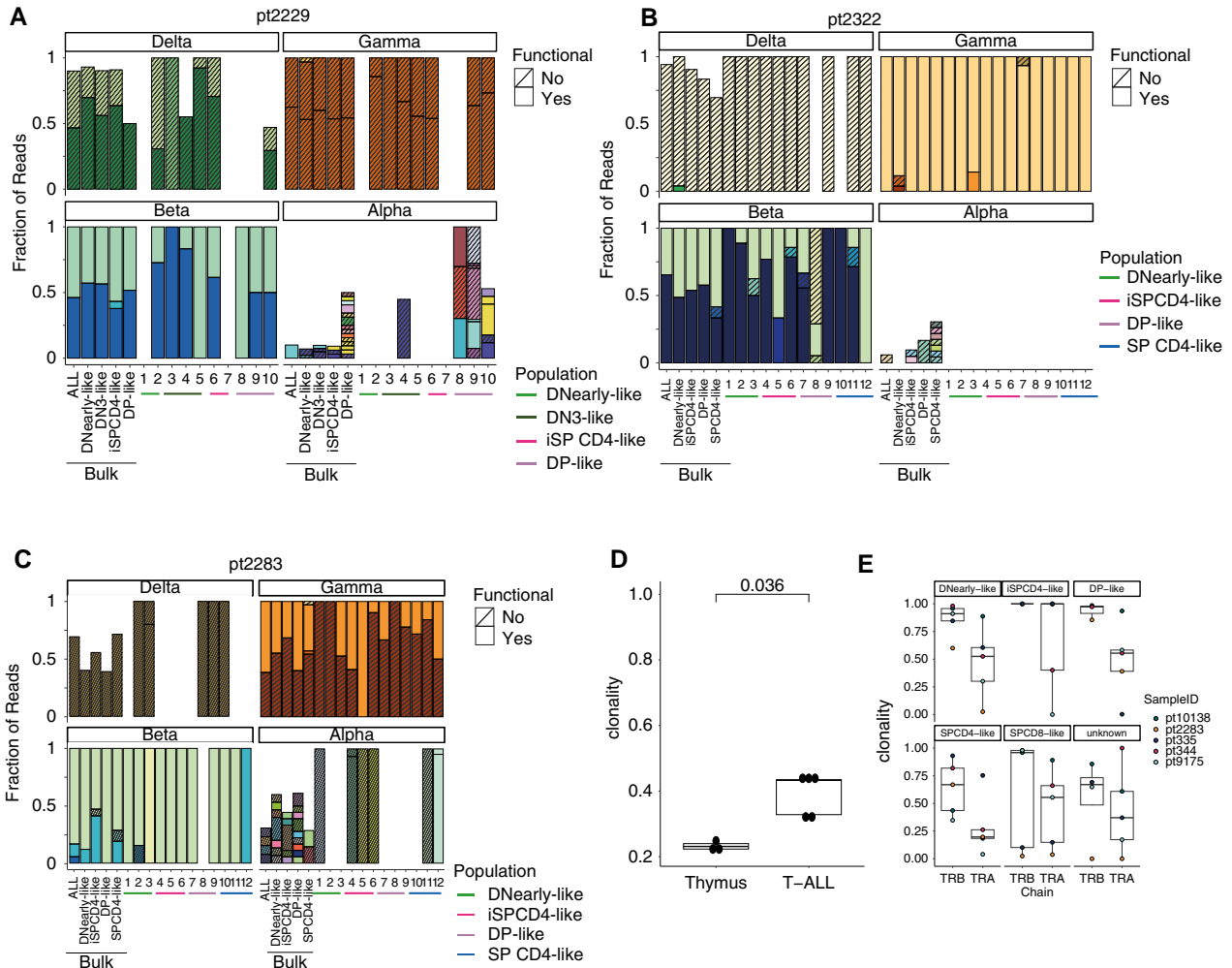
Figure 5.
V(D)J recombination points toward cell state of leukemia initiation. **A–C**, The relative contribution of V(D)J clones per gene segment to sequenced bulk populations and single cells per T-ALL patient. **D**, Clonality (Shannon diversity index) of healthy thymocytes ($n = 3$) vs. T-ALL ($n = 5$) Wilcox test. **E**, Clonality of TRB and TRA chain per phenotypic subpopulation.

generate unique TCR alpha receptors, similar to the continued IgH recombination described in B-ALL (39).

In pt2322, monoclonal *TRD*, *TRG*, and *TRB* were found, and *TRA* recombination was absent, indicating that the MRCA was of an "iSP CD4-like" phenotype (**Fig. 5B**). This sample harbored an oncogenic *TRAD-LMO2* structural variant, which may have prevented continued rearrangements. Erroneous V(D)J recombination was present in 20% to 25% of T-ALL samples (40). The oncogenic rearrangement of *LMO2* frequently occurred in DNearly developmental stages. However, secondary events were needed to fully transform cells. As a result, the MRCA was from the "iSPCD4-like" phenotype. Of note, the small polyclonal fraction in the "SPCD4-like" bulk population may result from contaminating nonleukemic SP CD4 T-cells, which could not be excluded entirely using our sorting strategy.

In pt2283, polyclonal *TRA* recombination was found in a subset of leukemic cells. Interestingly, the rearranged *TRA* of the immature "DNearly-like" cell 1 and bulk populations suggests that cells can

dedifferentiate from an "iSP CD4-like" phenotype to a more immature phenotype (**Fig. 5C**).

Single-cell sequencing of alpha and beta transcripts showed TCR transcripts in 5/8 T-ALL samples and confirmed a higher clonality in T-ALL samples compared with healthy thymocytes (**Fig. 5D**). The TCR clonality was lower in the alpha versus beta chains. Additionally, the TRB recombination was more clonal in the "DNearly-like," "iSPCD4-like," and "DP-like" phenotypes compared with more mature phenotypes, "SPCD4-like" and "SPCD8-like" (**Fig. 5E**).

The identification of the phenotypic cell state of the MRCA confirmed that immunophenotypic heterogeneity in T-ALL originated from plasticity, allowing differentiation and dedifferentiation of leukemic cells.

Preleukemic cell inducing relapse possesses phenotypic plasticity traits

Next, we studied the clonal evolution and plasticity of paired diagnostic and relapse samples. The time between diagnosis and relapse was
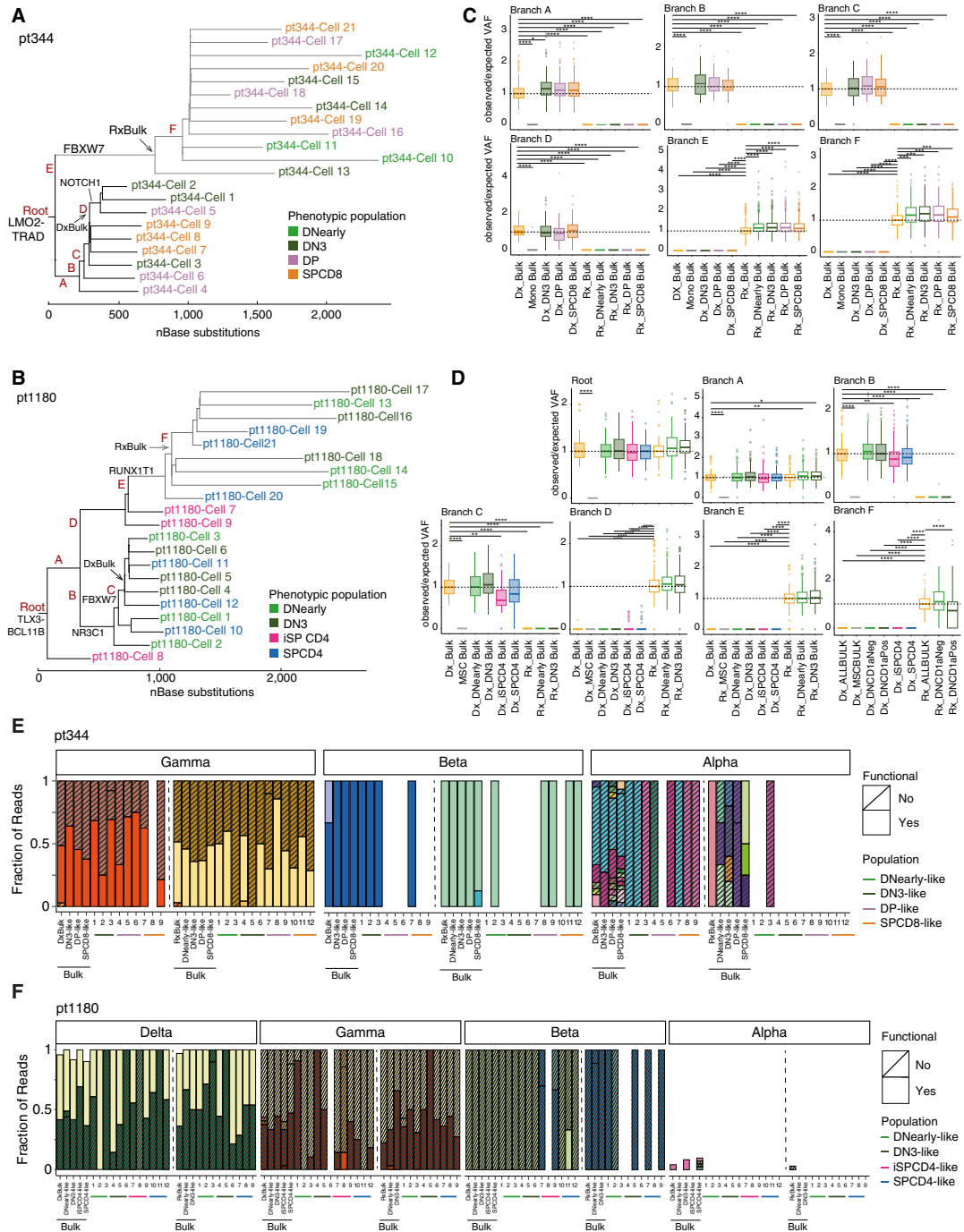
Figure 6.
Phenotypic subpopulations induced at relapse by preleukemic clone. **A** and **B**, Phylogenetic tree of T-ALL single cell blasts of respective patients. Length of branches indicates the number of somatic base substitutions. Colors indicate the phenotypic subpopulation. Letters address the different branches. **C** and **D**, VAF contribution plots of the branch-specific mutation in bulk-sorted phenotypic populations normalized to the diagnostic ALL bulk (Dx_Bulk) or relapse bulk (Rx_Bulk), depending on presence. Differences in VAF per variant in each population were compared using a one-way repeated-measures ANOVA and Bonferroni corrected t test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. **E** and **F**, The relative contribution of V(D)J clones per gene segment to sequenced bulk populations and single cells per T-ALL patient.

4 years for pt344 and 11 months for pt1180. Phylogenetic trees for both patients showed a stringent bottleneck because of therapy, whereafter cells rapidly expanded (**Fig. 6A** and **B**). This indicates that both relapses originated from a single cell. Phenotypic subpopulations identified at diagnosis and relapse were comparable, although the relapse of pt344 contained more "DNearly-like" cells, and pt1180 had markedly less "iSPCD4-like" and "DN3-like" cells (Supplementary Fig. S7A and S7B).

Interestingly, relapse cells showed minimal overlap in SNVs (**Fig. 6C** and **D**; Supplementary Fig. S7C and S7D) and VDJ recombination (**Fig. 6E** and **F**) with the diagnostic, revealing a preleukemic cell present at very low frequencies at diagnosis to increase relapse. In both patients, the colocalization of phenotypically similar cells at diagnosis and relapse indicated that heterogeneity in relapse was induced in the same manner as in the first leukemia. Collectively, these results show that preleukemic cells have a phenotypic plasticity trait.

Discussion

In this study, we investigated the phylogeny of phenotypic subpopulations in pediatric T-ALL and gained valuable insights into the native fate of leukemic cells before diagnosis. By coupling immunophenotyping with WGS data generated from single-cell leukemic blasts, we could reconstruct the phylogenetic lineage tree of leukemia at diagnosis and relapse and reveal the heritability of immunophenotypes. More specifically, we identified a small window of phenotypic plasticity at leukemia initiation. Whereafter, the cells were fixed in their newly acquired phenotype. Additionally, we performed an analysis of the V(D)J status at the single-cell level, allowing for the identification of the phenotype of the MRCA and demonstrating the ability of MRCAs to undergo both dedifferentiation and differentiation. Furthermore, preleukemic clones that induced relapse possessed the same plasticity traits.

The dogma in leukemia has long been that genetic events produce a differentiation block at a developmental stage during hematopoiesis. However, here we show that similar to other leukemias (8, 22), T-ALL recapitulates a differentiation hierarchy manifested as phenotypic intratumor heterogeneity upon diagnosis. Interestingly, immature subpopulations were observed in all samples and previously associated with poor outcomes and relapse (41, 42). However, there was no definitive evidence of CSCs in ALL. Instead, cells of diverse phenotypes grew out in xenograft models (25, 26, 43, 44). In line with these findings, we found immature cells restricted in their phenotypic potential. Furthermore, we did not observe a hierarchical arrangement or asymmetric cell divisions, suggesting CSCs are not present.

The heritability of phenotypic cell states and dedifferentiation was previously reported in glioblastoma (10), supporting our findings. Correspondingly, our data showed that leukemic expansion was initiated by a differentiated cell, and the observed phenotypic diversity was not genetically driven. Moreover, the similarity of individual blasts to the bulk ALL population suggested the fast expansion of the MRCA. The polyclonal V(D)J rearrangement of the alpha genes in cells with the CD4/CD8 DP phenotype indicated continued differentiation of leukemic blasts. This finding aligned with previous studies conducted in B-ALL and AML (39, 45).

Furthermore, we found that the observed phenotypic diversity was predominantly nongenetic, suggesting that epigenetic differences among cancer cells contribute to the phenotypic plasticity. However, we were unable to detect these differences in the transcriptomic analysis.

T-ALL transcriptome appeared relatively homogeneous, as previously reported (46). As the plasticity was initiated at leukemia initiation, the sampling timepoint for CITE-seq could be too late to detect transcriptomic changes, as oncogenic transcription programs may have taken the upper hand. This notion was consistent with previous research demonstrating disrupted epigenetic regulation driving lineage-switching in MLL/AF4 leukemia (47) and consistent with earlier work revealing an aberrant methylation profile in preleukemic cells (48).

This study has potential limitations. The number of samples analyzed was limited. However, we sequenced T-ALL samples from different genetic subtypes to grasp the overall heterogeneity among T-ALL. In addition, the number of cells sequenced in the single-cell WGS approach was between 10 and 40. Using these numbers of cells, the total clonality could be underestimated. Nevertheless, this approach was sufficient to confidently reconstruct the root and subbranches of the tree, which was confirmed by a uniform clonal VAF distribution in the bulk sequenced samples. Moreover, bulk WGS of sorted phenotypic populations was used to validate the contribution of each phenotypic population to the phylogenetic tree. Furthermore, our observations need *in vivo* validation to confirm our findings.

In conclusion, in this study, we developed a framework for understanding leukemogenesis, which provided biological insights into leukemia progression before diagnosis. The improved whole genome amplification and variant calling methods used in this study allowed for the direct measurement of the genomic variation within any human cell at an unprecedented level (16, 17). Utilizing this integrative model of somatic evolution and phenotyping will be particularly insightful in studying leukemias as well as the normal hematopoietic system.

Authors' Contributions

V.M. Poort: Conceptualization, formal analysis, investigation, methodology, writing–original draft, and writing–review and editing. **R. Hagelaar:** Data curation, formal analysis, and investigation. **M.J. van Roosmalen:** Data curation and formal analysis. **L. Trabut:** Methodology. **J.G.C.A.M. Buijs-Gladdines:** Methodology. **B. van Wijk:** Resources. **J. Meijerink:** Conceptualization, supervision, and writing–original draft. **R. van Boxtel:** Conceptualization, supervision, funding acquisition, writing–original draft, and writing–review and editing.

Note

References

1. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic plasticity: driver of cancer initiation, progression, and therapy resistance. Cell Stem Cell 2019;24:65–78.

2. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial–mesenchymal transitions in development and disease. Cell 2009;139:871–90.

3. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell 2011;146:633–44.

4. Boumahdi S, de Sauvage FJ. The great escape: tumour cell plasticity in resistance to targeted therapy. Nat Rev Drug Discov 2020;19:39–56.

5. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov 2022;12:31–46.

6. Visvader JE. Cells of origin in cancer. Nature 2011;469:314–22.

7. Kester L, van Oudenaarden A. Single-cell transcriptomics meets lineage tracing. Cell Stem Cell 2018;23:166–79.

8. van Galen P, Hovestadt V, Wadsworth Ii MH, Hughes TK, Griffin GK, Battaglia S, et al. Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. Cell 2019;176:1265–81.e24.

9. Nam AS, Kim KT, Chaligne R, Izzo F, Ang C, Taylor J, et al. Somatic mutations and cell identity linked by genotyping of transcriptomes. Nature 2019;571:355–60.

10. Chaligne R, Gaiti F, Silverbush D, Schiffman JS, Weisman HR, Kluegel L, et al. Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. Nat Genet 2021;53:1469–79.

11. Van Egeren D, Escabi J, Nguyen M, Liu S, Reilly CR, Patel S, et al. Reconstructing the lineage histories and differentiation trajectories of individual cancer cells in myeloproliferative neoplasms. Cell Stem Cell 2021;28:514–23.e9.

12. Osorio FG, Rosendahl Huber A, Oka R, Verheul M, Patel SH, Hasaart K, et al. Somatic mutations reveal lineage relationships and age-related mutagenesis in human hematopoiesis. Cell Rep 2018;25:2308–16.e4.

13. Lee-Six H, Øbro NF, Shepherd MS, Grossmann S, Dawson K, Belmonte M, et al. Population dynamics of normal human blood inferred from somatic mutations. Nature 2018;561:473–8.

14. Mitchell E, Spencer Chapman M, Williams N, Dawson KJ, Mende N, Calderbank EF, et al. Clonal dynamics of haematopoiesis across the human lifespan. Nature 2022;606:343–50.

15. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell 2012;148:873–85.

16. Gonzalez-Pena V, Natarajan S, Xia Y, Klein D, Carter R, Pang Y, et al. Accurate genomic variant detection in single cells with primary template-directed amplification. Proc Natl Acad Sci U S A 2021;118:e2024176118.

17. Middelkamp S, Manders F, Peci F, van Roosmalen MJ, González DM, Bertrums EJM, et al. Comprehensive single-cell genome analysis at nucleotide resolution using the PTA Analysis Toolbox. Cell Genom 2023;3:100389.

18. Liu Y, Easton J, Shao Y, Maciaszek J, Wang Z, Wilkinson MR, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. Nat Genet 2017;49:1211–8.

19. Homminga I, Pieters R, Langerak AW, de Rooi JJ, Stubbs A, Verstegen M, et al. Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. Cancer Cell 2011;19:484–97.

20. Szczepański T, van der Velden VHJ, Raff T, Jacobs DCH, van Wering ER, Brüggemann M, et al. Comparative analysis of T-cell receptor gene rearrangements at diagnosis and relapse of T-cell acute lymphoblastic leukemia (T-ALL) shows high stability of clonal markers for monitoring of minimal residual disease and reveals the occurrence of second T-ALL. Leukemia 2003;17:2149–56.

21. Dik WA, Pike-Overzet K, Weerkamp F, de Ridder D, de Haas EFE, Baert MRM, et al. New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. J Exp Med 2005;201:1715–23.

22. Good Z, Sarno J, Jager A, Samusik N, Aghaeepour N, Simonds EF, et al. Single-cell developmental classification of B cell precursor acute lymphoblastic leukemia at diagnosis reveals predictors of relapse. Nat Med 2018;24:474–83.

23. Greaves MF, Chan LC, Furley AJW, Watt SM, Molgaard HV. Lineage promiscuity in hemopoietic differentiation and leukemia. Blood 1986;67:1–11.

24. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 1997;3:730–7.

25. Pal D, Blair HJ, Elder A, Dormon K, Rennie KJ, Coleman DJL, et al. Long-term in vitro maintenance of clonal abundance and leukaemia-initiating potential in acute lymphoblastic leukaemia. Leukemia 2016;30:1691–700.

26. Rehe K, Wilson K, Bomken S, Williamson D, Irving J, den Boer ML, et al. Acute B lymphoblastic leukaemia-propagating cells are present at high frequency in diverse lymphoblast populations. EMBO Mol Med 2013;5:38–51.

27. Bertrums EJM, Rosendahl Huber AKM, de Kanter JK, Brandsma AM, van Leeuwen AJCN, Verheul M, et al. Elevated mutational age in blood of children treated for cancer contributes to therapy-related myeloid neoplasms. Cancer Discov 2022;12:1860–72.

28. Brady SW, Roberts KG, Gu Z, Shi L, Pounds S, Pei D, et al. The genomic landscape of pediatric acute lymphoblastic leukemia. Nat Genet 2022;54:1376–89.

29. Machado HE, Mitchell E, Øbro NF, Kübler K, Davies M, Leongamornlert D, et al. Diverse mutational landscapes in human lymphocytes. Nature 2022;608:724–32.

30. Kozlov A, Alves JM, Stamatakis A, Posada D. CellPhy: accurate and fast probabilistic inference of single-cell phylogenies from scDNA-seq data. Genome Biol 2022;23:37.

31. van Dongen JJM, Lhermitte L, Böttcher S, Almeida J, van der Velden VHJ, Flores-Montero J, et al. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. Leukemia 2012;26:1908–75.

32. Erarslan-Uysal B, Kunz JB, Rausch T, Richter-Pechańska P, van Belzen IA, Frismantas V, et al. Chromatin accessibility landscape of pediatric T-lymphoblastic leukemia and human T-cell precursors. EMBO Mol Med 2020;12:e12104.

33. Luquette LJ, Miller MB, Zhou Z, Bohrson CL, Zhao Y, Jin H, et al. Single-cell genome sequencing of human neurons identifies somatic point mutation and indel enrichment in regulatory elements. Nat Genet 2022;54:1564–71.

34. Weng AP, Ferrando AA, Lee W, Morris JP IV, Silverman LB, Sanchez-Irizarry C, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science 2004;306:269–71.

35. Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature 2014;513:422–5.

36. Davis A, Gao R, Navin N. Tumor evolution: linear, branching, neutral or punctuated? Biochim Biophys Acta Rev Cancer 2017;1867:151–61.

37. Huang C, Kanagawa O. Ordered and coordinated rearrangement of the TCR alpha locus: role of secondary rearrangement in thymic selection. J Immunol 2001;166:2597–601.

38. van der Velden VHJ, van Dongen JJM. MRD detection in acute lymphoblastic leukemia patients using Ig/TCR gene rearrangements as targets for real-time quantitative PCR. Methods Mol Biol 2009;538:115–50.

39. Gawad C, Koh W, Quake SR. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. Proc Natl Acad Sci U S A 2014;111:17947–52.

40. Larmonie NSD, Dik WA, Meijerink JPP, Homminga I, van Dongen JJM, Langerak AW. Breakpoint sites disclose the role of the V(D)J recombination machinery in the formation of T-cell receptor (TCR) and non-TCR associated aberrations in T-cell acute lymphoblastic leukemia. Haematologica 2013;98:1173–84.

41. Richter-Pechańska P, Kunz JB, Bornhauser B, von Knebel Doeberitz C, Rausch T, Erarslan-Uysal B, et al. PDX models recapitulate the genetic and epigenetic landscape of pediatric T-cell leukemia. EMBO Mol Med 2018;10:e9443.

42. Turati VA, Guerra-Assunção JA, Potter NE, Gupta R, Ecker S, Daneviciute A, et al. Chemotherapy induces canalization of cell state in childhood B-cell precursor acute lymphoblastic leukemia. Nat Cancer 2021;2:835–52.

43. Blackburn JS, Liu S, Wilder JL, Dobrinski KP, Lobbardi R, Moore FE, et al. Clonal evolution enhances leukemia-propagating cell frequency in T cell acute lymphoblastic leukemia through Akt/mTORC1 pathway activation. Cancer Cell 2014;25:366–78.

44. Blackburn JS, Liu S, Raiser DM, Martinez SA, Feng H, Meeker ND, et al. Notch signaling expands a pre-malignant pool of T-cell acute lymphoblastic leukemia clones without affecting leukemia-propagating cell frequency. Leukemia 2012;26:2069–78.

45. Agarwal A, Bolosky WJ, Wilson DB, Eide CA, Olson SB, Fan G, et al. Differentiation of leukemic blasts is not completely blocked in acute myeloid leukemia. Proc Natl Acad Sci U S A 2019;116:24593–9.

46. De Bie J, Demeyer S, Alberti-Servera L, Geerdens E, Segers H, Broux M, et al. Single-cell sequencing reveals the origin and the order of mutation acquisition in T-cell acute lymphoblastic leukemia. Leukemia 2018;32:1358–69.

47. Tirtakusuma R, Szoltysek K, Milne P, Grinev VV, Ptasinska A, Chin PS, et al. Epigenetic regulator genes direct lineage switching in MLL/AF4 leukemia. Blood 2022;140:1875–90.

48. Roels J, Thénoz M, Szarzyńska B, Landfors M, De Coninck S, Demoen L, et al. Aging of preleukemic thymocytes drives CpG island hypermethylation in T-cell acute lymphoblastic leukemia. Blood Cancer Discov 2020;1:274–89.