Tracking rare single donor and recipient immune and leukemia cells after allogeneic hematopoietic cell transplantation using mitochondrial DNA mutations

Penter L., Cieri N., Maurer K., Kwok M., Lyu H., Lu W.S., Oliveira G., Gohil S.H., Leshchiner I., Lareau C.A., Ludwig L.S., Neuberg D.S., Kim H.T., Li S., Bullinger L., Ritz J., Getz G., Garcia J.S., Soiffer R.J., Livak K.J., Wu C.J.

Blood Cancer Discovery
2024 NOV 01 ; 5(6): 442-459
2024 SEP 05 (first published online)
doi: 10.1158/2643-3230.BCD-23-0138

Publisher: American Association for Cancer Research

Tracking rare single donor and recipient immune and leukemia cells after allogeneic hematopoietic cell transplantation using mitochondrial DNA mutations

Livius Penter[1,2,3,4,5], Nicoletta Cieri*[1,2,3], Katie Maurer*[1,2,3], Marwan Kwok*[1,2,3,6], Haoxiang Lyu[1,7], Wesley S. Lu[1,7], Giacomo Oliveira[1,2,3], Satyen H. Gohil[8], Ignaty Leshchiner[2], Caleb A. Lareau[9], Leif S. Ludwig[4,10,11], Donna S. Neuberg[12], Haesook T. Kim[12], Shuqiang Li[1,7], Lars Bullinger[4], Jerome Ritz[1,3], Gad Getz[2,3], Jacqueline S. Garcia[1,3], Robert J. Soiffer[1,3], Kenneth J. Livak[1,7,#], Catherine J. Wu[1,2,3,#,$]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA.
[2]Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA.
[3]Harvard Medical School, Boston, Massachusetts, USA.
[4]Department of Hematology, Oncology, and Tumorimmunology, Campus Virchow Klinikum, Berlin, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany.
[5]Berlin Institute of Health at Charité – Universitätsmedizin Berlin, BIH Biomedical Innovation Academy, BIH Charité Digital Clinician Scientist Program, Charitéplatz 1, 10117 Berlin, Germany
[6]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK
[7]Translational Immunogenomics Lab, Dana-Farber Cancer Institute, Boston, Massachusetts, USA.
[8]Department of Haematology, University College London Hospitals, London, UK.
[9]Memorial Sloan Kettering Cancer Center, New York City, NY.
[10]Berlin Institute of Health at Charité Universitätsmedizin Berlin, Berlin, Germany.
[11]Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, Berlin Institute for Medical Systems Biology, Berlin, Germany.
[12]Department of Data Science, Dana-Farber Cancer Institute, Boston, Massachusetts, USA.

*contributed equally
#senior authors
$corresponding author

Running title: mtDNA-based single-cell leukemia and chimerism tracking

Word count (text): 7,456

Word count (abstract): 148 words

Figures: 7 main figures, 20 supplementary figures, 10 supplementary tables

Corresponding author:

Catherine J. Wu, MD

Professor, Dana-Farber Cancer Institute and Harvard Medical School

Chief, Division of Stem Cell Transplantation and Cellular Therapies

Institute Member, Broad Institute of Harvard and MIT

Telephone: 617-632-5943

Email: cwu@partners.org

Mailing address:

450 Brookline Avenue
Boston, MA 02215

Presented in part as poster at 63rd Annual Meeting of the American Society of Hematology (ASH) in Atlanta, GA from December 11-14, 2021
Presented in part as poster at 65th Annual Meeting of the American Society of Hematology (ASH) in San Diego, CA from December 9-12, 2023

Competing interests

C.A.L. and L.S.L. are named inventors on a patent related to mitochondrial lineage tracing (PCT/US2019/036583).

C.A.L. and L.S.L. are consultants to Cartography Biosciences.

C.J.W. is an equity holder of BioNtech, Inc, receives research funding from Pharmacyclics, and is a SAB member of Repertoire, Aethon Therapeutics, and Adventris.

G.O is consultant for Bicycle Therapeutics.

J.R. receives research funding from Kite/Gilead, Oncternal, and Novartis, serves on a Data Safety Monitoring Committee for AvroBio and on the Scientific Advisory Boards for Akron Biotech, Clade Therapeutics, Garuda Therapeutics, LifeVault Bio, Novartis, Smart Immune, Talaris Therapeutics, and TScan Therapeutics.

J.S.G. reports serving on steering committee and receiving personal fees from AbbVie, Astellas, and Takeda and institutional research funds from AbbVie, Genentech, Prelude, and AstraZeneca.

K.J.L. reports equity in Standard BioTools Inc. and serves on the scientific advisory board for MBQ Pharma Inc.

R.J.S. serves on the Board of Directors for Be the Match/National Marrow Donor Program and DSMB for Juno Therapeutics, Celgene USA, and BMS; reports personal fees from Vor Biopharma, Smart Immune, Daiichi Sankyo Inc, Neovii, Bluesphere Bio, Cugene, and Jasper.

All other authors do not have any relevant conflict of interest.

Summary

Combined tracking of clonal evolution and chimeric cell phenotypes could enable detection of the key cellular populations associated with response following therapy, including after allogeneic hematopoietic stem cell transplantation (HSCT). We demonstrate that mitochondrial DNA (mtDNA) mutations co-evolve with somatic nuclear DNA mutations at relapse post-HSCT and provide a sensitive means to monitor these cellular populations. Further, detection of mtDNA mutations via single-cell ATAC with select antigen profiling by sequencing (ASAP-seq) simultaneously determines not only donor and recipient cells, but also their phenotype, at frequencies of 0.1-1%. Finally, integration of mtDNA mutations, surface markers, and chromatin accessibility profiles enables the phenotypic resolution of leukemic populations from normal immune cells, thereby providing fresh insights into residual donor-derived engraftment and short-term clonal evolution following therapy for post-transplant leukemia relapse. As throughput evolves, we envision future development of single-cell sequencing-based post-transplant monitoring as a powerful approach for guiding clinical decision making.

Keywords: mitochondrial DNA mutations, measurable residual disease, allogeneic hematopoietic stem cell transplantation, microchimerism, clonal evolution, single cell proteogenomics

Statement of significance

Mitochondrial DNA mutations enable single-cell tracking of leukemic clonal evolution and donor-recipient origin following allogeneic hematopoietic stem cell transplantation. This provides unprecedented insight into chimeric cellular phenotypes of early immune reconstitution, incipient relapse, and quality of donor engraftment with immediate translational potential for future clinical post-transplant monitoring and decision making.

Introduction

For the clinical challenges presented by treatment of aggressive blood malignancies, allogeneic hematopoietic stem cell transplantation (HSCT) is an established immunotherapy that can improve long-term outcomes through graft-versus-leukemia (GvL) effects.(1–4) However, post-HSCT relapse is frequent, and immunotherapeutic salvage strategies aimed to reinstate disease control such as rapid immunosuppression tapering (IST), hypomethylating agents, donor lymphocyte infusions (DLI) or, more recently, immune checkpoint blockade (ICB) have mostly limited efficacy with exceptions in specific settings (5–10). This can necessitate a second HSCT, for example from a haploidentical donor, which is increasingly performed (11). A qualitative understanding of disease recurrence through assessment of relapsed leukemia phenotypes, clonal evolution, and potential for recovery of donor engraftment at the time of, or even prior to, molecular relapse would aid in improved selection of appropriate therapeutic salvage strategies. However, while sensitive relapse detection is feasible using flow cytometric or bulk sequencing approaches, the identification of relapse-associated leukemia phenotypes has so far been technically not possible before overt relapse when most salvage therapies become less effective.

Recent advances in single-cell sequencing technologies and analysis enable detection of natural genetic barcodes such as single nucleotide polymorphisms (SNPs) or mutations in mitochondrial DNA (mtDNA) and somatic nuclear DNA. These barcodes, in combination with determination of cell states, could provide a basis for integrated genetic and phenotypic immune cell and leukemia monitoring. In particular, the analysis of SNPs and mtDNA mutations provides an approach for genetic deconvolution of donor and recipient-derived cell populations in post-HSCT samples.(12) The high mtDNA content per cell and the small size of the mitochondrial genome along with its exquisitely polymorphic nature, especially in the control region(13), suggests this approach would be very robust for monitoring reconstituting cell populations following HSCT from non-sibling donors. Further, in contrast to germline SNPs, analysis of mtDNA can resolve clonal evolution of leukemic cell populations, thereby providing information about longitudinal changes in disease state (14–17). However, while evolution of mtDNA mutations after therapeutic bottlenecks has been described(15,17), the relationship of changes in mtDNA and somatic

nuclear DNA mutations following therapeutic bottlenecks remains unknown. Moreover, there are also questions about the ability of such an approach to provide information on leukemia phenotypes during early relapse and the sensitivity for single-cell mtDNA-based detection of rare donor or recipient-derived populations.

Here, we demonstrate that mtDNA mutations co-segregate with somatic nuclear DNA mutations during clonal evolution enabling sensitive identification of rare cell populations and aiding in the resolution of their phenotypes in the context of relapse following HSCT. Together, our analyses suggest that single-cell-based immune and leukemia monitoring have the potential to provide both quantitative and qualitative information for therapeutic decision making in the post-HSCT setting.

Results

Co-evolution of mitochondrial and somatic nuclear DNA mutations

The extent to which mtDNA mutations track with somatic nuclear DNA mutations following therapeutic bottlenecks is largely unknown. We hypothesized that existing bulk datasets of serial chronic lymphocytic leukemia (CLL) samples would provide opportunities to address this question through combined analysis of both classes of mutations. Although whole-exome sequencing (WES) is a standard tool for tracking somatic nuclear DNA mutations, it does not sufficiently cover the mitochondrial genome to detect mtDNA mutations. In contrast, bulk RNA-seq data has high coverage of mitochondrial transcripts beyond 10,000x (**Suppl. Fig. 1A**). We were therefore motivated to extract mtDNA mutations from bulk transcriptomes generated from 110 samples collected longitudinally from 26 CLLs for whom concomitant WES data was available (**Methods; Suppl. Table 1**).(18) The median interval between the first and last sample of each CLL was 2,184 days (range 280-4,402). By WES, we categorized CLLs as genetically stable (n=8) or evolving, either in absence of therapy (n=7) or following fludarabine-based immunochemotherapy (n=11) (**Fig. 1A-B**). To exclude artifacts, we focused on mtDNA mutations with a bulk heteroplasmy >0.5% in at least one sample (**Suppl. Fig. 1B**). Although we identified fewer mitochondrial (median 11, range 2–26) than somatic nuclear (median 63, range 18–168) DNA mutations per CLL (**Suppl. Fig. 1C**), we found that their longitudinal dynamics were consistent with the somatic mutation-defined growth kinetic categories. In particular, similar to somatic nuclear DNA mutations, their frequency increased with follow-up time from the first to the last sample, even as the number of dynamic mtDNA mutations was lower in genetically stable compared to post-therapy evolving CLL cases (median 2 vs. 5 per CLL; $p < 0.001$; *Student t-test*) (**Fig. 1C-D, Suppl. Fig. 1D**). Likewise, heteroplasmy changes were smaller in stable compared to evolving CLL (coefficient of variance 31 vs 57; $p < 0.001$; *Student t-test*) (**Fig. 1E**). Finally, when matching longitudinal dynamics of mtDNA mutations and cancer cell fractions (CCF, which is a measure of involvement by somatic nuclear mutations), we observed high correlation ($r = 0.84$, $p < 0.001$; ***Pearson's** correlation*) (**Suppl. Fig. 1E, Suppl. Fig. 2-5**).

Given the deep coverage of mitochondrial transcripts in bulk RNA-seq data, we wondered whether mtDNA mutations would enable sensitive tracking of dynamic CLL clones. We defined the limit of detection (LOD) for each mtDNA mutation as the average heteroplasmy across the study cohort and observed a median LOD of 0.3% (range 0.008%-73%), which was mostly determined by genomic localization (**Suppl. Fig. 6A-C**). With this approach, we tracked expansion of individual mtDNA mutations such as 2115T>C and 7300T>C (CLL1) or 7929G>A and 8040T>C (CLL19) from low initial heteroplasmies (between 0.02% and 0.13%) to values between 0.7 % and 6.1%. At the same time, some mtDNA mutations such as 2002G>A (CLL1) or 13358T>C (CLL19) collapsed from a heteroplasmy of 8.8% and 1.3% to 0.5% and 0.05% (**Fig. 1F, Suppl. Fig. 6D**). Altogether, these data reveal mtDNA mutations to have similar longitudinal dynamics as somatic nuclear DNA mutations and the potential to sensitively track emerging CLL clones.

Tracking resistance to graft-versus-leukemia responses

Having observed the pronounced dynamics of mtDNA mutations in CLL cases over long follow-up, we asked how these natural barcodes could be also applied to the setting of allogeneic stem cell transplantation. We extracted mtDNA mutations from bulk RNA-seq data generated from 8 CLL cases for which paired peripheral blood mononuclear cell (PBMC) specimens were collected before and after immunochemotherapy (FCR [fludarabine, cyclophosphamide, rituximab]) followed by reduced-intensity conditioning (RIC)-HSCT (follow-up time ranged from 83-1,979 days [median 931 days]) (**Fig. 2A**). Again, we observed an association of follow-up time between samples and the magnitude of change in mtDNA mutations, such that early relapse remained genetically stable while late recurrence displayed clonal evolution (310 vs. 1,082 days, p = 0.034; *Wilcoxon signed-rank test*) (**Fig. 2B-C**), consistent with previous genetic characterization by WES(19). Further, we could sensitively track mtDNA mutations. For example, in CLL5328, we observed clonal replacement evidenced by disappearance of 16247A>G from an initial heteroplasmy of 27% below LOD at relapse, as opposed to expansion of 6426G>A from 0.2% to 54.4% (**Fig. 2D**).

Given the high agreement of parallel analyses of somatic nuclear and mitochondrial DNA mutations, we decided to analyze mtDNA and somatic nuclear DNA mutations within the same cell. We thus performed targeted single-cell DNA sequencing with surface marker capture (scDNA-seq; Tapestri platform) using assays for detection of patient-specific somatic nuclear and mitochondrial mutational profiles. Assay designs were based on information provided by prior genetic characterization from WES (generated pre-FCR and post-HSCT) and single-cell ATAC sequencing (mtscATAC-seq) (15,19). We focused on 3 CLL cases (5328, 5327 and 5335) and generated a primer panel for the detection of 98 mutations across 69 genes, encompassing 67 nuclear and 21 mtDNA amplicons (**Fig. 2E, Suppl. Table 2**). As expected per the assay designs, the coverage of mtDNA was limited to targeted regions and enabled high sequencing depth to optimize detection of even variants with low heteroplasmy (**Suppl. Fig. 7A**). To discriminate among immune cell subtypes, we concurrently captured surface marker staining (Total-seq D) (**Fig. 2F, Suppl. Fig. 7B**).

In total, we obtained 33,796 high-quality scDNA-seq profiles from the 3 PBMC sample pairs. Eight of 8 previously reported mtDNA mutations that defined major CLL subclones and a median of 14 (8-19) of 20 (16-26) somatic mutations from these cases were detected. CLL-specificity of mtDNA and somatic mutations was confirmed by their absence in donor and recipient immune cell populations. For all 3 CLL cases, somatic founder mutations could be identified that were present before and after post-HSCT relapse (**Fig. 2G, Suppl. Fig. 8A-B**). This was also the case for the mtDNA mutation 3538G>A in CLL5328 which persisted throughout therapy. A subset of mtDNA mutations which co-segregated with somatic mutations subdivided the CLL populations further, such as 16247A>G (CLL5328), 5979G>A and 2332C>T with ASL^{Y321C} and MME^{K525N} (CLL5335) or 3830T>C and 3526G>A with $PKDREJ^{V245F}$ and $TP53^{V272M}$ (CLL5327). Clonal shifts in CLL, as defined by somatic mutations that were only discernable before or after FCR/HSCT, tracked with analogous changes in mtDNA mutations. These shifts in clonal architecture ranged from clonal replacement (evident through emergence of 6426G>A and 16290C>T in CLL5328); to selection of a resistant subclone (marked by loss of 2332C>T and 5979G>A in CLL5335); to skewing of the ratio within CLL clones at relapse post-HSCT (in CLL 5327) (**Fig. 2G, Suppl. Fig. 8C-E**). In CLL5328, we identified donor-derived mtDNA mutations based on their absence in recipient cells that

were associated with $CD8^+$ T cells (1918G>A, 5650G>A) or lymphocytes in general (786G>A) highlighting the fact that mtDNA mutations also mark clonal ancestries in non-malignant cell populations (**Fig. 2G, Suppl. Fig. 8F**). Finally, consistent with the detection of 6426G>A at very low heteroplasmy in bulk RNA-seq data pre-HSCT, we could identify 3 of 2,416 (0.12%) leukemic cells from CLL5328 which harbored this mutation alongside *DPCD*^{D41E} and *ZNF215*^{H42P} prior to transplantation, providing direct evidence of the existence of a resistant population identifiable by mitochondrial and somatic nuclear DNA mutations (**Fig. 2H**). In sum, mtDNA mutations are shaped by GvL-mediated selection and directly track with evolving somatic nuclear DNA mutations.

Mitochondrial DNA mutations as markers of donor-recipient pairs

Mixed chimeric states are frequently detected following allogeneic HSCT. Examples include replacement of recipient hematopoiesis with donor cells during the initial post-transplant phase, regrowth of recipient-derived cells at incipient disease recurrence (i.e. measurable residual disease – MRD) or displacement of donor-derived cells at time of overt relapse (i.e. waning residual physiologic hematopoiesis).

To define the potential utility of genetic mtDNA variation for sensitively distinguishing among admixed cells originating from unrelated individuals, a common scenario in >50% of transplants (12,20,21), we compared the performance of donor-recipient deconvolution via maternal mtDNA variants with that from nuclear single nucleotide polymorphism (SNP)-based deconvolution. We thus performed scDNA-seq on two bone marrow samples collected from a relapsed AML after matched unrelated donor HSCT. We reutilized the previously designed CLL mtDNA panel which provided coverage of the control region that spans the first and last 700 bp of the circular mitochondrial genome (chrM:16,024-576, where most mtDNA mutations are located(22)) with a commercial 279-amplicon panel targeting 37 recurrently mutated genes in AML/MDS(23). In this manner, we identified germline SNPs and the known *NRAS*^{G13R} and *SF3B1*^{R775Q} mutations of this case. When we compared donor-recipient deconvolution using 4 maternal mtDNA variants and 10 nuclear germline SNPs, we observed highly consistent annotations in 7,541 of 7,618 (99%) cells (**Suppl. Fig. 9A-B**). AML cells clustered into recipient-derived hematopoietic stem cell (HSC)-like, granulocyte macrophage progenitor (GMP)-like, monocyte-like and erythrocyte-like

cells (**Suppl. Fig. 9C-D**). This highly accurate donor-recipient deconvolution was further supported by the absence of *NRAS*^{G13R} and *SF3B1*^{R775Q} in donor-derived cells in the HSC, GMP, monocyte and erythroid compartments, while they were detectable in 80-98% (*SF3B1*^{R775Q}) and 50-70% (*NRAS*^{G13R}) of recipient-derived cells (coverage >10x) (**Suppl. Fig. 9E-F**). Furthermore, the mtDNA mutation 11736T>C co-segregated with *SF3B1*^{R775Q} and *NRAS*^{G13R}, providing an example of a leukemia-specific mtDNA mutation (**Suppl. Fig. 9G**). Finally, we analyzed T/NK cell donor chimerism, which was highest in NK and lowest in CD4^{+} T cells (**Suppl. Fig. 9H**). Altogether, combined detection of mtDNA and somatic nuclear DNA mutations demonstrated highly consistent results for donor-recipient deconvolution whether with mtDNA or SNPs.

Our detection of co-segregating somatic nuclear and mtDNA mutations along with the high accuracy of mtDNA-based donor-recipient deconvolution prompted us next to systematically evaluate the sensitivity of this approach for resolving mixed chimeric states. To define what percentage of unrelated transplant pairs could be distinguished using maternal mtDNA variants (excluding tumor-specific mtDNA mutations), we simulated the deconvolution of donor-recipient pairings using two external bulk mtDNA-seq datasets comprising 81 individuals without cancer (**Supplementary Data**) (24). We identified a median of 30 homoplasmic mtDNA mutations (detected heteroplasmy >98%) per individual (range 7–78) and a total of 624 mtDNA mutations, of which 141 (22.6%) were found within the most variable region of chrM (15.9 kB to 700bp) (**Suppl. Fig. 10A-C**). Among all 6,480 simulated pairings, the median number of diverging homoplasmic mtDNA mutations was 35 (range 2–94), with at least 5 diverging mtDNA mutations in 99.8% of pairs (**Fig. 3A**). Simulations yielding fewer than 8 diverging mtDNA mutations originated exclusively from pairs within the same mtDNA haplotype (**Suppl. Fig. 10D**). These analyses indicate that donor and recipient cells in unrelated transplants can be effectively distinguished using homoplasmic mtDNA mutations.

Because not all mtDNA mutations localize to expressed regions of the mitochondrial genome, we focused on expressed mtDNA mutations, identified from 108 non-malignant bulk RNA-seq profiles, to assess the feasibility of this approach for single-cell transcriptomic assays(25). We identified 393 expressed

homoplasmic mtDNA mutations, a median of 21 per individual (range 0–61), which provided at least 5 diverging mtDNA mutations in 95.8% of pairings (**Suppl. Fig. 10E-G**), thus supporting the ability to distinguish most donor-recipient pairs using transcriptome data. The number of distinguishing mtDNA variants was highly similar across datasets, although sequencing coverage was more uniform when using DNA- rather than RNA-sequencing (**Suppl. Fig. 10H-J**), consistent with observations from single-cell RNA sequencing (scRNA-seq) analyses (14,26,27). From all 189 individuals, a total of 924 homoplasmic mtDNA mutations were identified. By downsampling the number of analyzed individuals, this number represents only a fraction of homoplasmic mtDNA mutations in the entire human population, consistent with the >19,000 mtDNA mutations documented in MITOMAP (**Fig. 3B**).(28)

Benchmarking of donor-recipient deconvolution in-silico and in-vitro

To assess the potential sensitivity of single-cell donor-recipient deconvolution using mtDNA mutations, we performed an *in-silico* spike-in experiment using previously reported circulating blood mtscATAC-seq profiles of two CLL cases (CLL4, CLL5).(15) We mixed between 1 to 1,000 profiles of CLL4 (randomly sampled from 3,369 profiles) into all available 7,579 mtscATAC-seq profiles of CLL5. CLL4 and CLL5 harbored a total of 45 maternally inherited mtDNA variants (variance-to-mean ratio (VMR) < 0.01; strand concordance > 0.65). Cells from CLL4 and CLL5 could be distinguished by 29 and 8 maternal mtDNA variants, respectively (**Suppl. Fig. 11A**). We calculated a mean heteroplasmy for the distinct maternal mtDNA variants of CLL4 and CLL5 per cell and observed clear separation between the two cell populations (**Suppl. Fig. 11B**). This permitted recovery of between 98.3% and 100% of CLL4 from the *in-silico* mixture, resulting in a theoretical sensitivity of 1:7,580 (0.13%) (**Fig. 3C-top**). In total, we found only 123 cells (1%) across monocytes, T and CLL cells whose origin could not be annotated due to insufficient mtDNA coverage in the combined dataset of CLL4 and CLL5 (11,797 cells) (**Suppl. Fig. 11C-E**). We repeated the analysis with two AML datasets (AML1011, AML1012), and likewise found a theoretical sensitivity of 1:6,586 (0.15%), with only 102 of 12,319 (0.8%) unannotated cells (**Suppl. Fig. 11F**).

To compare the sensitivity of mtDNA- versus SNP-based deconvolution, we mixed scRNA-seq profiles of the same samples, at ratios of 1 to 1,000 CLL4 cells spiked into 10,000 CLL5 cells. We tested SNP-

based deconvolution of scRNA-seq data using the tools *souporcell* (29) and *vireo* (30). When performing germline reference-free deconvolution, we found that deconvolution was reliable only when spike-in was above 3% and 1% for *vireo* and *souporcell*, respectively (**Fig. 3C-bottom left, Suppl. Fig. 11G**). At lower percentages, both methods left 22–80% of cells unannotated and *souporcell* identified hundreds of false-positive doublets (**Suppl. Fig. 11H**). Performance improved markedly for *vireo* when a germline reference for both individuals was provided, which allowed detection down to 1 in 10,000 cells, yet left up to 3% of cells unannotated (**Fig. 3C-bottom right, Suppl. Fig. 11I**). In contrast, mtDNA deconvolution does not require germline samples, as maternal mtDNA variants that distinguish donor and recipient can be extracted directly from chimeric data due to their mutual exclusivity even at low frequencies (for example, using k-means clustering [shown for 10 CLL4 cells (0.13%) spiked into 7,579 CLL5 cells in **Suppl. Fig. 12A**]).

The high rate of unannotated cells was also apparent when applying *vireo* with WES donor/recipient germlines and *souporcell* without germline references on real-world scRNA-seq generated from 18 AML post-HSCT samples.(31) We observed concordance of donor-recipient annotations between both tools, including with X- and Y-chromosomal gene expression in AML1010 (transplanted from a sex-mismatched donor) (**Suppl. Fig. 12B-E**). However, the rate of unassigned cells using *vireo* was much higher than with *souporcell* (5.5–57.1% vs. 1.2–19%) (**Suppl. Fig. 12F**), demonstrating that differences in genome coverage and drop-out are inherent issues with SNP-based genetic deconvolution of single-cell data, a shortcoming much less encountered when using mtDNA mutations with single-cell data that provide high uniform coverage of mtDNA, such as ATAC with select antigen profiling by sequencing (ASAP-seq).

We likewise assessed the ability to distinguish phenotypically distinct immune cell populations derived from different individuals using mtDNA mutations in non-synthetic data. We modeled the detection of antigen-specific T cells through a mixing experiment in which expanded healthy donor T cells (from ‘donor 1’) transduced with a murinized MART1-specific TCR were spiked into unmanipulated PBMCs of a second healthy donor (‘donor 2’) at ratios from 1:3 to 1:300. We applied ASAP-seq to the mixtures to gain information on chromatin accessibility, donor-recipient and cell type annotations (**Fig. 3D-E-top**).(32) In

addition to detecting lineage-defining surface markers (via staining cells with a custom panel of 22 Total-seq B (TSB) oligo-tagged antibodies including an antibody specific for the murinized TCR (mTcrβ)), we utilized MART1-specific tetramers conjugated with 2 distinct Total-seq A (TSA) streptavidin (SAV) oligotags to identify MART1-specific T cells (**Suppl. Table 3-4**). Utilizing this experimental design, we deconvolved 15,116 cells from both donors, which resulted in ratios of donor1 to donor2 that were highly concordant with the mixing ratios of the experiment. In total, we detected 1,449 of 5,345 (27%) donor1 cells at the mixing ratio of 1:3, and 150 in 4,262 (3.5%) and 14 in 4,381 (0.3%) at the ratios of 1:30 and 1:300 (**Fig. 3F-H**).

In addition to the sensitive detection of spiked-in expanded T cells, we could track their phenotype. First, analysis of chromatin accessibility profiles demonstrated profound chromatin remodeling after $CD3^+/CD28^+$-bead activation and IL7/IL15-induced expansion with increased accessibility of *RUNX1/2* and reduced accessibility of the *SPIB* binding motif (**Fig. 3I, Suppl. Fig. 12G**). Secondly, detection of the transduced MART1 TCR sequence (donor1) from chromatin accessibility data delineated 83 cells with integration of the expression vector, none of which belonged to donor2 (**Fig. 3E-bottom**). Finally, we directly identified 185 MART1-specific T cells based on co-expression of SAV oligotags and mTcrβ (**Suppl. Fig. 12H-I**), ranging from 168 (3%) to 5 cells (0.1%), of which 183 (98.9%) were derived from donor1 and 2 (1.1%) from donor2 at frequencies that reflected the donor ratios in the experiment (**Fig. 3J**). Our *in-silico* and *in-vitro* mixing experiments thus demonstrated that mtDNA mutations sensitively resolve donor- and recipient-derived cell populations alongside their specific phenotypes at frequencies well below 1%.

Tracking incipient AML relapse in responders to immunosuppression tapering

To demonstrate how donor-recipient identification can be used to track AML phenotypes even at low frequencies, we applied this strategy to samples from 4 AML patients (IST1-4) after RIC-HSCT (**Suppl. Table 5**), in which an incipient relapse was detected on the basis of worsening bulk chimerism (87–185 days after graft infusion) and was solely treated with rapid taper of immunosuppression (IST)(5). Clinically, responses were associated with abrupt improvements in blood cell counts and circulating

monocytes, resulting in complete clearance of AML within 105–184 days after IST without any further disease relapse (**Fig. 4A, Suppl. Fig. 13A-D**).

From these four responders, we obtained 13,360 bone marrow and 5,783 peripheral blood ASAP-seq profiles before and after IST and identified major immune cell, erythropoietic, and AML/progenitor populations (**Fig. 4B, Suppl. Fig. 14A**). Consistent with the clinical white blood cell counts, our analysis showed a strong increase in circulating monocytes in 3 of 4 patients relative to T cells following IST, likely representing recovering physiologic hematopoiesis after resolution of the incipient relapse (**Fig. 4C, Suppl. Fig. 14B**). We extracted mtDNA mutations from the ASAP-seq data for chimerism analyses and leukemic tracking. In line with clinical chimerism measurements (XY FISH or short-tandem repeat analysis), all 4 patients had incomplete donor chimerism at time of relapse and prior to IST across progenitor (0-0.5%), erythropoietic (64%) or monocytic cells (0-100%). All cell types converted to full donor chimerism following IST (**Fig. 4D, Suppl. Fig. 14C**).

We assumed that non-lymphocytic recipient-derived cells post-HSCT most likely represent AML, and we used the mixed chimeric cell populations to track leukemic clones across their differentiation states (33,34) to delineate differences between physiologic hematopoiesis and its malignant counterparts. In IST1 and IST2, AML cells were detectable within progenitor-like, erythroid-like and monocyte-like populations, and were marked by several non-homoplasmic mtDNA mutations (3919T>C, 5458T>C, 7457G>A, 10776T>C [IST1] and 6701A>G, 10290G>C [IST2]) (**Fig. 4E, Suppl. Fig. 14D-E, 15A-B**). Notably, while circulating monocytes in IST1 were 100% donor, the bone marrow monocyte chimerism was lower (64%), indicating non-circulating AML relapse. Compared to donor-derived monocytes, recipient-derived monocyte-like cells in IST1 had lower surface expression of CD11c and CD14 and lower chromatin accessibility of the *IL1B* gene, which plays a role in AML pathogenesis (35) (**Fig. 4F-G**). We did not detect differences in surface marker expression between recipient- and donor-derived erythropoietic cells in IST1, but observed differences in chromatin accessibility, for example reduced open chromatin of *CD36* in recipient-derived cells (**Suppl. Fig. 15B-C**). In IST3, recipient-derived monocyte-like cells also displayed differences in surface marker expression compared to donor-derived monocytes, including

higher CD33 and lower CD16, as well as differential chromatin accessibility of genes such as chemokine ligands or homeobox protein genes (**Suppl. Fig. 15D-F**).

Among lymphocytes, donor chimerism tended to be higher with incomplete donor chimerism detected only in NK cells of IST4 (0%) and T cells of IST3 and IST4 (41% and 24%), which remained incomplete even after IST (90% and 70%) (**Fig. 4D**). When comparing surface marker expression and chromatin accessibility of T cells in IST3 and IST4, we observed that 94% of recipient-derived T cells were $CD4^+$ T cells, while 61.5% of $CD8^+$ T cells were donor-derived ($p<0.001$; Fisher's exact test), consistent with their generally different immune reconstitution dynamics following HSCT (**Fig. 4H**)(36). Other differences between donor- and recipient-derived T cells such as expression of PD-1 were mainly driven by differences in CD4:CD8 T cell ratio (**Fig. 4I-J, Suppl. Fig. 15G**). Our results clearly demonstrate that, donor-recipient deconvolution of ASAP-seq profiles can identify AML cells at low frequencies and resolve differences in the phenotypes of chimeric leukemia and immune cell populations.

Residual donor engraftment and leukemia clonal evolution in overt AML relapse

While incipient relapse is characterized by a mainly donor-derived hematopoietic system that retains the potential to intensify ongoing GvL effects through IST, overt AML relapse constitutes high circulating disease burden and minimal (or no) residual donor hematopoiesis. The question to what extent residual donor engraftment persists is highly relevant for the success of strategies aimed at reinstating GvL but is not routinely assessed due to lack of sufficiently granular testing modalities.

To this end, we performed ASAP-seq on longitudinal bone marrow samples from post-transplant participants of the ETCTN/CTEP 10026 study (**Suppl. Fig. 16A-F**). In this study, participants received a priming cycle of decitabine followed by combination of decitabine and ipilimumab with the aim to reinvigorate dormant GvL responses.(10,37,38) We assessed residual donor engraftment in three patients (AML1010, AML1011 and AML1026) with unrelated donors. We further evaluated one patient (AML1012) with a related donor who harbored a *del*(5q) to test whether copy number variations (CNV) could also be used to detect AML cells using ASAP-seq (**Fig. 5A**). Chromatin accessibility profiles from all

4 AML cases demonstrated left-shifted, expanded myelo-/erythropoiesis which were annotated as hematopoietic stem cell (HSC)-like, granulocyte macrophage progenitor (GMP)-like, monocytic (Mono), dendritic (DC) and erythroid (Ery) clusters based on surface marker expression (**Suppl. Fig. 16G**). Unexpectedly, donor-recipient deconvolution revealed that donor-derived hematopoiesis was exceedingly rare, with 10–19 (0.1% - 0.6%) cells identified as donor-derived, most of which were of erythroid lineage. In AML1012, we identified AML cells by detection of *del*(5q) or *amp*(22q), which similarly indicated that only 83 of 5,763 (1.4%) myelo-/erythropoietic cells did not harbor CNV changes and thus likely represented remaining donor-derived hematopoiesis (**Fig. 5B-C, Suppl. Fig. 16H-J**). We also calculated mtDNA-based T cell chimerism, which was stable throughout CTLA-4 blockade and correlated well with clinical chimerism measurements (r = 0.97, **Fig. 5D, Suppl. Fig. 17A**). Altogether, AML clones occupied the entire phenotypic range from HSC-like cells to monocytes and erythroid cells within the same patient, consistent with previous reports(16,27,33). Further, mtDNA-based donor-recipient deconvolution can identify rare donor-derived hematopoietic events during overt AML relapse and enables estimation of the potential for a recovery of physiologic hematopoiesis following salvage therapy.

Given the coevolution of mitochondrial with somatic nuclear DNA mutations in CLL, we wondered whether mtDNA mutations could also resolve clonal evolution induced by GvL in AML. Since patients continued on-study for up to 10 months with circulating blast counts (**Suppl. Table 6)**, we hypothesized that decitabine/ipilimumab treatment led to shifts in AML subclones despite absence of objective responses and therefore analyzed serial samples of each patient before study entry, after decitabine priming, and during combined treatment. As a comparator, we analyzed samples from a responding patient (AML1007). Further, we contrasted the impact of ipilimumab on AML subclones defined by mtDNA mutations with changes induced by HSCT through analysis of specimens prior to and following HSCT in 3 patients (AML1007, AML1011, AML1012).

By mtDNA mutation analysis (**Supplementary Data**), we identified a median of 7 AML subclones per case (range 5–12) and found longitudinal dynamics unrelated to the changes in cell type proportions. In AML1010, we observed a decrease in complexity of AML subclones throughout treatment, while

AML1026 and AML1012 revealed an increase in clonal complexity (**Fig. 5E-F-left, Suppl. Fig. 17B-E, 18**), demonstrating that despite an apparent clinical absence of response, treatment with decitabine/ipilimumab exerted a selective pressure that led to changes among AML subclones. The only exception was AML1011, who lacked evidence of such an effect (**Fig. 5F-right**). However, in the responding AML1007, we again found clear clonal selection during ipilimumab-based treatment as evidenced by two collapsing subclones (**Fig. 5G**).

When tracking longitudinal changes following RIC-HSCT (AML1011, AML1012) and haploidentical HSCT (AML1007), the subclonal structure of AML1011 and AML1012 remained practically unchanged throughout transplantation, while AML1007 showed profound clonal evolution with outgrowth of cells marked by 5668G>C, consistent with acquisition of two novel somatic mutations on clinical amplicon-based sequencing (39) (*RUNX1*^{R201Q} and *KRAS*^{L19F}) (**Fig. 5G**). This could reflect the more pronounced immune pressure exerted by GvL in the setting of haploidentical transplantation compared to RIC-HSCT from a matched unrelated donor, an interpretation which would naturally need further validation. These studies demonstrate that mtDNA mutations can identify subclones that shift in response to immune-based therapy, thereby illustrating the utility of this lineage-tracing approach for dissecting therapeutic effects on leukemia populations.

Co-evolution of mitochondrial and somatic nuclear DNA mutations in AML

Having observed longitudinal shifts in AML subclones defined by mtDNA mutations, we sought to directly evaluate their co-segregation and co-evolution with somatic nuclear DNA mutations using single-cell DNA sequencing. For AML1010, 1012 and 1026, we generated a customized 127-primer panel targeting 137 somatic nuclear DNA mutations identified using WES (**Suppl. Table 7-9**). Further, we obtained a second pan-mtDNA panel with 67 amplicons covering the entire mitochondrial chromosome (**Fig. 6A, Suppl. Fig. 19A-B, 20A**). For each AML, we obtained scDNA-seq data at screening and following therapy with decitabine/ipilimumab. In all three cases, we observed a high degree of co-segregation between mitochondrial and somatic nuclear DNA mutations. Consistent with the prior characterization by ASAP-

seq, the amount of subclonal heterogeneity ranged from reduced structure in AML1012 to a complex pattern of main and subclones in AML1010 and AML1026.

AML1010 subdivided into two subclones defined by mutually exclusive homozygous mutations in *TCF12* alongside 14 additional somatic nuclear DNA mutations. While most cells in subclone 1 expressed CD34, more than 50% of cells in subclone 2 displayed a monocytic phenotype, consistent with differential lineage commitment (**Suppl. Fig. 20B**). Across both subclones, we found a total of seven smaller subclones, which were defined by 37 mtDNA mutations (**Fig. 6B**). Throughout treatment with decitabine/ipilimumab, the proportion of these subclones shifted, with two subclones mitochondrial DNA mutations expanding (**Fig. 6C**). While most leukemic cells in AML1010 were *NRAS*-mutated, one of seven subclones harbored exclusively $NRAS^{wt}$ cells. Multiple mtDNA mutations such as 9820G>C co-segregated with this subcluster, providing additional support for this notable subpopulation (**Suppl. Fig. 20C**).

In AML1026, all cells harbored five somatic nuclear DNA mutations including $SRSF2^{P95H}$, $SF3B1^{R775Q}$, $NRAS^{G13R}$ and two mtDNA mutations (3308T>C and 10685G>A). We identified two large subclones that could be further subdivided into five smaller subclones, based on 11 additional somatic nuclear and 37 mitochondrial DNA mutations (**Fig. 6D**). Treatment shifted these subclones markedly, including expansion of cells marked by $COL9A1^{A218S}$ and 9 mtDNA mutations (**Fig. 6E**).

Finally, in AML1012, we found a uniform detection of 12 somatic nuclear DNA mutations across all cells. Nevertheless, mtDNA mutations provided further subdivision into 7 subclones (**Fig. 7A**). One subclone was marked by the 14739G>C mutation and expanded following therapy with ipilimumab/decitabine (**Fig. 7B-C**). We also observed an increase in cells harboring the $FBXW7^{R465C}$ hotspot mutation(40) across six of the subclones. Together with the overall uniform appearance of AML1012 at the level of somatic nuclear DNA mutations, this likely indicates a close genetic and phenotypic relationship amongst the identified subclones and illustrates the ability of mtDNA mutations to identify subtle differences within tumor populations characterized as monoclonal by somatic nuclear DNA mutations. Given the more

dynamic nature of mitochondrial compared to somatic nuclear DNA mutations, future studies are required to definitively assess the relevance of these subpopulations with regard to clonal fitness.

Discussion

Single-cell donor-recipient deconvolution is of great interest for many questions in basic and translational transplant research but has been technically challenging. Here, we report that single-cell mtDNA sequencing can resolve chimeric populations with high sensitivity. By systematically assessing mtDNA-based donor-recipient deconvolution using in-silico and in-vitro mixing experiments, we show that, compared to donor-recipient deconvolution with nuclear germline SNPs, this approach works even for very skewed ratios of donor- and recipient-derived cells. SNP-based identification often fails because the coverage of relevant single nucleotide variants becomes insufficient. For mtDNA variant analysis, the high copy number of mtDNA drastically reduces failures due to allelic drop-out or insufficient coverage. Also, mtDNA variants do not require sequencing of pure donor or recipient populations (i.e., germline controls) because the relevant maternal DNA variants can be deduced from mixed populations.

By analyzing mtDNA mutations in different settings of post-transplant relapse, we gained fresh biological insights. First, we demonstrate co-segregation of mtDNA and somatic nuclear mutations in AML and CLL patients who relapsed years after chemoimmunotherapy followed by stem cell transplantation and confirm the utility of mtDNA variants as long-term markers of clonal ancestries. Further, we show that mtDNA mutations can resolve clonal dynamics in AML during anti-leukemic therapy even over short periods of several months that could likely not be detected with somatic nuclear mutations. These observations underscore the potential of mtDNA mutations as natural genetic barcodes for longitudinal studies of leukemia evolution and point towards complementary utility with other genetic barcodes. While the recurrent, stable nature and low copy number range of many somatic nuclear mutations and chromosomal aberrations permit predictable and automated identification, they are often unable to resolve short-term selective pressures. Tracking of mtDNA mutations, on the other hand, is more dynamic owing to their larger heteroplasmic range and stochastic propagation during mitosis (Research Square 2023, 3083262). With this higher dynamic nature, they can be potentially utilized to identify changes in subclonal structure where somatic nuclear mutations do not have sufficient resolution. Thus, the

combined study of somatic nuclear and mtDNA mutations may enable the discernment of a more precise picture of shorter- and longer-term clonal evolution. Our reanalysis of CLL bulk RNA-seq profiles is an example how analysis of mtDNA mutations in existing datasets is an opportunity to further data mining.

Second, we compare chromatin accessibility profiles of physiologic and malignant hematopoiesis and define leukemia phenotypes in the context of incipient or overt AML post-transplant relapse. We observed that AML recapitulates normal erythro-myeloid hematopoietic differentiation from hematopoietic stem cell-like to maturing monocytic-like and erythroid-like cells, consistent with observations made with single-cell RNA expression profiles.(16,27,33) Using mtDNA mutations, we identified subtle differences between physiologic and malignant hematopoiesis at the level of chromatin accessibility and lineage marker expression, even for small cell populations. In the future, this approach may be used to overcome a diagnostic blind spot by providing an assay capable of supplying qualitative or phenotypic information of disease relapse at time of MRD. In turn, these results may inform better selection of appropriate salvage therapies based on the leukemia phenotypes driving relapse and degree of remaining donor-derived hematopoiesis potential.

Even with these advances, we acknowledge limitations of our approach. Although clinical experience shows that recurrence of recipient-derived cells usually equates with leukemia relapse, donor-recipient deconvolution using mtDNA mutations is unable to differentiate AML relapse from reversal to a precursor state of clonal hematopoiesis. For such analyses, it would be useful to combine donor-recipient deconvolution with analysis of copy number changes if available or to consider additional approaches like GoT-ChA-seq(41) that enable read-out of somatic nuclear mutations alongside mtDNA mutations. Further, while our study provides proof-of-principle, truly sensitive identification of rare donor- or recipient-derived cell populations will require sequencing of 100,000 cells or more per sample, which is economically infeasible with commercial single-cell assays. As throughput of single-cell assays increases, we expect this limitation to become increasingly less relevant. At that point, it will also be crucial to benchmark single-cell-based detection of MRD with current clinical gold standard assays to understand whether clinical decision-making could indeed be guided by single-cell genomics approaches(42).

Methods

Extraction and analysis of mitochondrial DNA mutations from CLL bulk RNA-seq data

Variant calling

Raw reads were aligned against chrM (GRCh38) using bowtie2(43), sorted with samtools(44) and deduplicated using GATK4 MarkDuplicates(45). Mitochondrial DNA mutations were called using mgatk(14) (>10 reads supporting each mtDNA mutation, strand coordination >0.5, >100 total reads for the position) followed by filtering for high confidence variants. Filtering was based on exclusion of frequent false-positive variants (2617A>G, 2617A>T, 13710A>G, 13710A>T, 5746G>A) and a heteroplasmy <0.5% (18) or <1% (19). For mtDNA mutations that passed filtering, variant calls were performed across all samples from the individual in which a particular mtDNA mutation was identified. Further down-stream analyses were performed using custom scripts.

Definition of CLL cohorts

Genetically stable CLL was defined by cancer cell fraction (CCF) changes of <20%, while evolving cases had CCF changes >20%. Naturally progressing CLL was defined as those cases sampled prior to initiation of an initial first line of disease-specific therapy.

Matching of mtDNA mutations to CLL subclones defined by whole-exome sequencing

To match mtDNA mutations to CLL subclones defined by whole-exome sequencing, the relative longitudinal changes from first to last sample were calculated for each mtDNA mutation based on heteroplasmy and each CLL subclone based on cancer cell fractions (CCF). Matching was performed based on the smallest difference between the relative change of the heteroplasmy of a mtDNA and the relative changes of CCF values.

Limit of detection for individual mtDNA mutations

A limit of detection (LOD) was defined for each mtDNA mutation and for each CLL case by calculating the mean heteroplasmy for the mtDNA mutation across the entire analyzed cohort excluding samples from the particular CLL case.

In-silico mixing experiment

Cell barcodes were randomly sampled from two published single cell libraries (CLL4_1 and CLL5_1) (15). Reads associated with the cell barcodes were extracted from the genome-aligned bam file using sinto (https://github.com/timoast/sinto) and merged with samtools. Deconvolution of mixed scRNA-seq data was performed using souporcell(29) or cellsnp-lite(46) and vireo(30). Deconvolution of mixed scATAC-seq data was performed by calling mtDNA mutations using mgatk, identification of maternal mtDNA variants for each individual and deconvolution based on average heteroplasmy for each cell (cut-off >80% for variants from individual1 and <20% from individual2).

Extraction of mitochondrial DNA mutations from published bulk datasets

Datasets were downloaded using SRA Toolkit (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software) developed by the SRA Toolkit Development Team. Raw reads were aligned against chrM (GRCh38) using bowtie2(43), sorted with samtools(44) and deduplicated using GATK4 MarkDuplicates(45). Mitochondrial DNA mutations were discovered using mgatk(14) and custom scripts adopted from Caleb Lareau's github repository (https://github.com/caleblareau).

Sample accrual of peripheral blood and bone marrow mononuclear cells

EDTA-anticoagulated peripheral blood and bone marrow samples (PBMC, BMMC) were obtained from patients treated at Dana-Farber Cancer Institute in Boston who participated in observational immune correlative studies or from patients who participated in the ETCTN/CTEP 10026 study (**Suppl. Table 10**) (10,31). Anonymous healthy donor PBMCs were obtained from flushed apheresis leukoreduction collars following platelet donation at Brigham and Women's Hospital in Boston. Written informed consent was obtained from study participants before study enrollment, and all procedures involving human participants

were carried out in accordance with the Declaration of Helsinki. PBMC and BMMC were subjected to Ficoll density gradient centrifugation (GE Healthcare) followed by cryopreservation with 10% DMSO in FBS (Sigma-Aldrich) and storage in vapor-phase liquid nitrogen until the time of analysis.

Thawing of cryopreserved cells

Prior to sequencing, cryovials were slowly melted in the steam of a cell culture water bath. Following drop-wise addition of PBS (Corning, Cat. no. MT21040CV) with 10% bovine DNase I, grade II (Sigma-Aldrich, Cat. no. 10104159001) and 10% FBS (ThermoFisher Scientific, Cat. no. 10437028), cells were pelleted in a centrifuge (300 g for 5 min) and resuspended in RPMI supplemented with 10% FBS and 10% DNase I, grade I (ThermoFisher Scientific, Cat. no. NC9007308) as previously described.(31)

ATAC with select antigen profiling by sequencing (ASAP-seq)

ASAP-seq was performed as previously described.(47) Briefly, after thawing 1-2 million cells were resuspended in 50 µl PBS with 0.04% ultrapure BSA (ThermoFisher Scientific, Cat. no. AM2618). After addition of 5 µl Human TruStain FcX (BioLegend, Cat. no. 422302) for 10 minutes at 4 ºC, Total-seq A/B antibodies (**Suppl. Tables 3 and 4**) were added (0.5 µl per antibody) and cells were incubated for 30 minutes at 4 ºC. Fixation and cell permeabilization was performed using the mtscATAC-seq protocol(48): following 3 washing cycles with PBS with 0.04% ultrapure BSA, cells were resuspended in 450 µl PBS. Fixation was performed by adding 30 µl 16% formalin (ThermoFisher Scientific, Cat. no. 28906) and incubation for 10 minutes at room temperature. After addition of 26.8 µl of 2.36 M glycine (ThermoFisher Scientific, Cat. no. 15527013) for a final concentration of 0.125 M glycine in a total volume of 506.8 µl, 700 µl ice-cold PBS were added and cells washed with PBS for one more time. Lysis was performed by resuspension in 100 µl lysis buffer for 3 minutes at 4 ºC. After washing with washing buffer, fixated and permeabilized cells were resuspended at approximately 4,500/µl.

Loading of cell suspensions onto a Chromium Chip H (10x Genomics, Cat. no. PN-1000161) (targeted recovery of 7,000 cells) was followed by library preparation using the Chromium Single-cell ATAC Library & Gel Bead Kit (Cat. no. PN-1000175) according to manufacturer instructions. For capture of Total-seq

A/B oligotags 1 µM bridge oligos were added during the barcoding reaction followed by library preparation as described in the original description of ASAP-seq.(32) Following quality control of libraries with a Bioanalyzer High Sensitivity DNA Kit (Agilent). Pooled libraries were sequenced on a NovaSeq S2 platform (Illumina) with 50 bp paired-end reads, 8 bp for index 1 and 16 bp for index 2.

Lysis buffer

1 ml stock

- 10 µl 1 M Tris-HCl (pH 7.4) (Sigma, Cat. no. T2194)
- 2 µl 5 M NaCl (Santa Cruz, Cat. no. SC-295833)
- 3 µl 1 M $MgCl_2$ (ThermoFisher Scientific, Cat. no. AM9530G)
- 10 µl 10 % Nonidet P40 Substitute (Sigma, Cat. no. 74385)
- 100 µl BSA 10 % (Sigma, Cat. no. A3059)
- 875 µl Nuclease-free H_2O (Promega, Cat. no. P1193)

Washing buffer

10 ml stock

- 100 µl 1 M Tris-HCl (pH 7.4) (Sigma, Cat. no. T2194)
- 20 µl 5 M NaCl (Santa Cruz, Cat. no. SC-295833)
- 30 µl 1 M $MgCl_2$ (ThermoFisher Scientific, Cat. no. AM9530G)
- 1 ml BSA 10 % (Sigma, Cat. no. A3059)
- 8.85 ml nuclease-free H_2O (Promega, Cat. no. P1193)

Preparation of TotalSeq-A MART-1-specific tetramers

HLA-A*02:01 easYmers (Immunaware) were first folded with the ELAGIGLTV peptide (MART-1_{26-35}, GenScript) and then tetramerized with 2 distinct TotalSeq-A PE-conjugated streptavidins (Biolegend), following manufacturer instructions. Correct assembly and binding of the tetramers was confirmed by flow

cytometric evaluation of T cells transduced with a MART-1-specific TCR, as previously reported(49). As the TCR construct used was murinized, binding of a monoclonal antibody against the murine TCRβ (PE, clone H57-597, eBioscience) served as internal control for specific tetramer binding. Data were acquired on a Fortessa cytometer (BD Biosciences) and analyzed using Flowjo v10.8 software (BD Biosciences).

Use of hashing with ASAP-seq

Prior to sequencing of samples from AML1010, AML1012 and AML1026 (ETCTN/CTEP 10026), they were stained with 2 Total-seq A hashtags per sample and pooled. Deconvolution was performed based on average expression of hashtags and a sample-specific cut-off (**Suppl. Fig. 16A-F**).

Analysis of ASAP-seq profiles

Raw reads containing fragments of single cell chromatin profiles were aligned and quantified using cellranger-atac (https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac) with a custom GRCh38 reference hard masked for nuclear DNA of mitochondrial origin (nuMTs) to improve mapping of mitochondrial reads against chrM. Downstream analyses were performed using custom scripts in ArchR following best practices described in the ArchR manual.(50) Reads from Total-seq A and B libraries were processed as previously described.(47) In brief, raw reads were converted from a format with 3 to 2 fastq files compatible with kallisto | bustools (51) using “ASAP to kite” (https://github.com/caleblareau/asap_to_kite). After generation of feature count matrices with bustools, they were imported into R and further processed using the Seurat package (52) and custom scripts.

Annotation of donor and recipient origin

Donor and recipient annotation was based on mean heteroplasmy of maternal variants (cut-offs >80% for variants specific for individual and <20% for variants not found in individual, otherwise cells were considered unassigned or doublets) specific for each individual, which were identified by clustering homoplasmic mtDNA variants and manual review of their distribution across myeloid and T/NK cell populations, taking into consideration known T cell chimerisms from routine clinical diagnostics.

Copy-number changes

Copy number changes where identified using a sliding window approach as previously described (https://github.com/caleblareau/mtscATACpaper_reproducibility) using T cell chromatin profiles as normal control.(14,15,17)

Clustering of mtDNA mutations

Analysis of AML subclusters was performed by clustering mtDNA mutations of leukemic cell populations across all samples from one individual using the Seurat package.(52)

Detection of lentiviral expression vector sequence from single cell chromatin profiles

Vector copies encoding the MART1-specific T cell receptor were identified by mapping DNA fragments to the sequences of the CMV promoter, the Woodchuck Hepatitis Virus (WHV) Posttranscriptional Regulatory Element (WPRE) or the murinized alpha and beta T cell receptor sequences using chromap.(53)

Single-cell DNA sequencing (Tapestri)

Total-seq D staining

Cells were stained with the TotalSeq™-D Human Heme Oncology Cocktail, V1.0 (BioLegend, Cat. no. 399906) according to manufacturer instructions. These included resuspension in 40 µl cell staining buffer (BioLegend, Cat. no. 420201) supplemented with 5 µl Blocking Buffer (Mission Bio) and 5 µl Human TruStain FcX (BioLegend, Cat. no. 422302) and incubation for 15 min at 4 ºC followed by addition of 50 µl reconstituted antibody cocktail and incubation for 15 min at 4 ºC. After 3 washing steps with cell staining buffer, cells were resuspended at a concentration of 4,500/µl.

Library preparation and sequencing

After loading of cells onto a Tapestri cartridge, library preparation was performed according to manufacturer instructions. The genomic and mtDNA panels were mixed at a ratio of 10:1. Following

quality control with a Bioanalyzer High Sensitivity DNA Kit (Agilent), amplicon and Total-seq D libraries were pooled and sequenced on a NextSeq 500 or NovaSeq SP (Illumina) with 150 bp for paired-end reads, 8 bp for index 1 and 8 bp for index 2.

Analysis

Raw reads were processed using a local installation of the Tapestri Pipeline version 2.0.2 using GRCh19/38 for somatic nuclear DNA and GRCh38 for mtDNA reads. Variant calls and surface marker expression were exported for each cell barcode into a csv file from the Tapestri Insight software version 3.0.2 and imported into R where the data was further processed and visualized using custom scripts and standard packages including Seurat and ComplexHeatmap (52,54).

Data availability statement

CLL bulk RNA-seq and WES data were previously deposited on dbGap (phs001431.v1.p). CLL scDNA-seq data have been deposited on NCBI Geo (accession number GSE234558). Published datasets for analysis of maternal mitochondrial DNA variants were accessed from NCBI Geo (BioProjects PRJNA486215 (24) and PRJNA563929 for bulk mtDNA data; PRJNA741686 (25) for bulk RNA-seq). CLL in-silico mixing data were accessed from NCBI Geo (GSE165087 for scRNA-seq and GSE163579 for scATAC-seq). ASAP-seq data for the in-vivo mixing experiment and the immunosuppression tapering cohort have been deposited on NCBI Geo (GSE234558). Processed ASAP-seq, scDNA-seq and scRNA-seq data for samples from participants on ETCTN/CTEP 10026 are deposited on NCBI Geo (GSE234558, GSE223844), while the raw data has been deposited on dbGap (phs003015.v2).

Code availability statement

Code of analyses performed for this work are available under https://github.com/liviuspenter/mtDNA_donor_recipient_coevolution

Acknowledgments

We thank Doreen Hearsey, Jared Dorff and all staff from the Ted and Eileen Pasquarello Tissue Bank in Hematologic Malignancies for their excellent technical support with banking and retrieval of clinical samples. We thank patients for their generous contribution of research samples for this study. Computationally intensive data processing was performed on the O2 cluster of the Research Computing Group at Harvard Medical School.

This work was supported by National Institutes of Health, National Cancer Institute grants P01CA229092 and P01CA206978 (CJW), UM1CA186709 (Principal Investigator: Geoffrey Shapiro), National Cancer Institute Cancer Therapy Evaluation Program, Bristol-Myers Squibb, LLS Therapy Accelerator Program, and LLS SCOR-22937-22 (CJW, RJS).
L.P. was a Scholar of the American Society of Hematology, is a participant in the BIH Charité Digital Clinician Scientist Program funded by the DFG, the Charité – Universitätsmedizin Berlin, and the Berlin Institute of Health at Charité (BIH) and is supported by the Max-Eder program of the German Cancer Aid (Deutsche Krebshifle), by the Else Kröner-Fresenius-Stiftung (2023_EKEA.102) and the DKMS John Hansen Research Grant.
N.C. is a Scholar of the American Society of Hematology, an Amy Strelzer Manasevit Scholar, a Leukemia and Lymphoma Society Career Development Program – Special Fellow, and supported by the DKMS John Hansen Research Grant.
K.M. is supported by the ASCO YIA award and the Lubin Family Foundation Scholar Award.
M.K. is supported by a Cancer Research UK clinician scientist fellowship (RCCFELCSF-May21\100002)
G.O. was supported by Claudia Adams Barr Program for Innovative Cancer Research and by DF/HCC Kidney Cancer SPORE P50 CA101942.
S.L. is supported by the National Institutes of Health, National Cancer Institute Research Specialist Award (R50CA251956).
J.S.G. is supported by the Conquer Cancer Foundation Career Development Award, Leukemia and Lymphoma Society Translational Research Program Award, and NIH K08CA245209.
NCI CTEP provided study drug (Ipilimumab) support.
This work was further supported by the CIMAC-CIDC Network. Scientific and financial support for the CIMAC-CIDC Network is provided through National Institutes of Health, National Cancer Institute Cooperative Agreements U24CA224319 (to the Icahn School of Medicine at Mount Sinai CIMAC), U24CA224331 (to the Dana-Farber Cancer Institute CIMAC), U24CA224285 (to the MD Anderson Cancer Center CIMAC), U24CA224309 (to the Stanford University CIMAC), and U24CA224316 (to the CIDC at Dana-Farber Cancer Institute). The CIMAC-CIDC website is found at https://cimac-network.org/.

Author contributions

L.P. performed experiments, processed and analyzed data, and generated figures.
N.C. designed and performed experiments.
K.M. designed experiments and curated clinical cohorts.
M.K. curated clinical cohorts.
H.L. and W.S.L. performed single-cell library preparations.
G.O. provided antigen-specific T cells.
S.H.G., C.L., L.S.L., L.B., G.G., I.L., H.T.K. and D.S.N. interpreted data.
S.L. supervised H.L. and W.S.L. and implemented single-cell assays.
J.R., J.S.G. and R.J.S. provided clinical samples.
J.S.G. is the principal investigator of ETCTN/CTEP 10026.
K.J.L. and C.J.W. supervised the study.
L.P. and C.J.W. wrote the manuscript.

References

1. Cieri N, Maurer K, Wu CJ. 60 Years Young: The Evolving Role of Allogeneic Hematopoietic Stem Cell Transplantation in Cancer Immunotherapy. Cancer Res. American Association for Cancer Research; 2021;81:4373–84.

2. Penter L, Wu CJ. Personal tumor antigens in blood malignancies: genomics-directed identification and targeting. J Clin Invest. American Society for Clinical Investigation; 2020;130:1595–607.

3. Nagler A, Ngoya M, Galimard J-E, Labopin M, Bornhäuser M, Stelljes M, et al. Longitudinal Outcome over Two Decades of Unrelated Allogeneic Stem Cell Transplantation for Relapsed/Refractory Acute Myeloid Leukemia: An ALWP/EBMT Analysis. Clinical Cancer Research. 2022;28:4258–66.

4. Roeker LE, Dreger P, Brown JR, Lahoud OB, Eyre TA, Brander DM, et al. Allogeneic stem cell transplantation for chronic lymphocytic leukemia in the era of novel agents. Blood Advances. 2020;4:3977–89.

5. Kekre N, Kim HT, Thanarajasingam G, Armand P, Antin JH, Cutler C, et al. Efficacy of immune suppression tapering in treating relapse after reduced intensity allogeneic stem cell transplantation. Haematologica. 2015;100:1222–7.

6. Stahl M, DeVeaux M, Montesinos P, Itzykson R, Ritchie EK, Sekeres MA, et al. Hypomethylating agents in relapsed and refractory AML: outcomes and their predictors in a large international patient cohort. Blood Advances. 2018;2:923–32.

7. Takami A, Yano S, Yokoyama H, Kuwatsuka Y, Yamaguchi T, Kanda Y, et al. Donor Lymphocyte Infusion for the Treatment of Relapsed Acute Myeloid Leukemia after Allogeneic Hematopoietic Stem Cell Transplantation: A Retrospective Analysis by the Adult Acute Myeloid Leukemia Working Group of the Japan Society for Hematopoietic Cell Transplantation. Biology of Blood and Marrow Transplantation. 2014;20:1785–90.

8. Daver N, Garcia-Manero G, Basu S, Boddu PC, Alfayez M, Cortes JE, et al. Efficacy, Safety, and Biomarkers of Response to Azacitidine and Nivolumab in Relapsed/Refractory Acute Myeloid Leukemia: A Non-randomized, Open-label, Phase 2 Study. Cancer Discov. 2019;9:370–83.

9. Penter L, Gohil SH, Huang T, Thrash EM, Schmidt D, Li S, et al. Coevolving JAK2V617F+ relapsed AML and donor T cells with PD-1 blockade after stem cell transplantation: an index case. Blood Adv. 2021;5:4701–9.

10. Garcia JS, Flamand Y, Penter L, Keng M, Tomlinson BK, Mendez LM, et al. Ipilimumab plus decitabine for patients with MDS or AML in posttransplant or transplant-naïve settings. Blood. 2023;141:1884–8.

11. Filippini Velázquez G, Labopin M, Tischer J, Raiola AM, Angelucci E, Kulagin AD, et al. Second haploidentical stem cell transplantation (HAPLO-SCT2) after relapse

from a first HAPLO-SCT in acute leukaemia—a study on behalf of the Acute Leukaemia Working Party (ALWP) of the European Society for Blood and Marrow Transplantation (EBMT). Bone Marrow Transplant. Nature Publishing Group; 2023;58:907–15.

12. Penter L, Gohil SH, Wu CJ. Natural Barcodes for Longitudinal Single Cell Tracking of Leukemic and Immune Cell Dynamics. Front Immunol. 2021;12:788891.

13. Bronstein O, Kroh A, Haring E. Mind the gap! The mitochondrial control region and its power as a phylogenetic marker in echinoids. BMC Evolutionary Biology. 2018;18:80.

14. Lareau CA, Ludwig LS, Muus C, Gohil SH, Zhao T, Chiang Z, et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. Nat Biotechnol. 2021;39:451–61.

15. Penter L, Gohil SH, Lareau C, Ludwig LS, Parry EM, Huang T, et al. Longitudinal single-cell dynamics of chromatin accessibility and mitochondrial mutations in chronic lymphocytic leukemia mirror disease history. Cancer Discov. American Association for Cancer Research; 2021;11:3048–63.

16. Beneyto-Calabuig S, Merbach AK, Kniffka J-A, Antes M, Szu-Tu C, Rohde C, et al. Clonally resolved single-cell multi-omics identifies routes of cellular differentiation in acute myeloid leukemia. Cell Stem Cell. Elsevier; 2023;30:706-721.e8.

17. Penter L, Ten Hacken E, Southard J, Lareau CA, Ludwig LS, Li S, et al. Mitochondrial DNA Mutations as Natural Barcodes for Lineage Tracing of Murine Tumor Models. Cancer Res. 2023;83:667–72.

18. Gruber M, Bozic I, Leshchiner I, Livitz D, Stevenson K, Rassenti L, et al. Growth dynamics in naturally progressing chronic lymphocytic leukaemia. Nature. 2019;570:474–9.

19. Bachireddy P, Ennis C, Nguyen VN, Gohil SH, Clement K, Shukla SA, et al. Distinct evolutionary paths in chronic lymphocytic leukemia during resistance to the graft-versus-leukemia effect. Science Translational Medicine. 2020;12:eabb7661.

20. Saber W, Opie S, Rizzo JD, Zhang M-J, Horowitz MM, Schriber J. Outcomes after matched unrelated donor versus identical sibling hematopoietic cell transplantation in adults with acute myelogenous leukemia. Blood. 2012;119:3908–16.

21. Nagler A, Labopin M, Mielke S, Passweg J, Blaise D, Gedde-Dahl T, et al. Matched related versus unrelated versus haploidentical donors for allogeneic transplantation in AML patients achieving first complete remission after two induction courses: a study from the ALWP/EBMT. Bone Marrow Transplant. Nature Publishing Group; 2023;58:791–800.

22. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. Nature Publishing Group; 1999;23:147–147.

23. Morita K, Wang F, Jahn K, Hu T, Tanaka T, Sasaki Y, et al. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. Nat Commun. 2020;11:5327.

24. Hjelm BE, Rollins B, Morgan L, Sequeira A, Mamdani F, Pereira F, et al. Splice-Break: exploiting an RNA-seq splice junction algorithm to discover mitochondrial DNA deletion breakpoints and analyses of psychiatric disorders. Nucleic Acids Res. 2019;47:e59.

25. Hu Z, van der Ploeg K, Chakraborty S, Arunachalam PS, Mori DAM, Jacobson KB, et al. Early immune markers of clinical, virological, and immunological outcomes in patients with COVID-19: a multi-omics study. Elife. 2022;11:e77943.

26. Miller TE, Lareau CA, Verga JA, DePasquale EAK, Liu V, Ssozi D, et al. Mitochondrial variant enrichment from high-throughput single-cell RNA-seq resolves clonal populations. Nat Biotechnol. 2022;40:1030–4.

27. Penter L, Borji M, Nagler A, Lyu H, Lu WS, Cieri N, et al. Integrative genotyping of cancer and immune phenotypes by long-read sequencing. Nat Commun. 2024;15:32.

28. Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, et al. mtDNA Variation and Analysis Using Mitomap and Mitomaster. Curr Protoc Bioinformatics. 2013;44:1.23.1-26.

29. Heaton H, Talman AM, Knights A, Imaz M, Gaffney DJ, Durbin R, et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. Nat Methods. 2020;17:615–20.

30. Huang Y, McCarthy DJ, Stegle O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. Genome Biology. 2019;20:273.

31. Penter L, Liu Y, Wolff JO, Yang L, Taing L, Jhaveri A, et al. Mechanisms of response and resistance to combined decitabine and ipilimumab for advanced myeloid disease. Blood. 2023;141:1817–30.

32. Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Hao Y, Takeshima Y, et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. Nat Biotechnol. 2021;1–13.

33. 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;176:1265-1281.e24.

34. Zeng AGX, Bansal S, Jin L, Mitchell A, Chen WC, Abbas HA, et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. Nat Med. 2022;28:1212–23.

35. Arranz L, Arriero M del M, Villatoro A. Interleukin-1β as emerging therapeutic target in hematological malignancies and potentially in their complications. Blood Reviews. 2017;31:306–17.

36. Ogonek J, Kralj Juric M, Ghimire S, Varanasi PR, Holler E, Greinix H, et al. Immune Reconstitution after Allogeneic Hematopoietic Stem Cell Transplantation. Front Immunol. 2016;7:507.

37. Davids MS, Kim HT, Bachireddy P, Costello C, Liguori R, Savell A, et al. Ipilimumab for Patients with Relapse after Allogeneic Transplantation. N Engl J Med. 2016;375:143–53.

38. Penter L, Zhang Y, Savell A, Huang T, Cieri N, Thrash EM, et al. Molecular and cellular features of CTLA-4 blockade for relapsed myeloid malignancies after transplantation. Blood. 2021;137:3212–7.

39. Kluk MJ, Lindsley RC, Aster JC, Lindeman NI, Szeto D, Hall D, et al. Validation and Implementation of a Custom Next-Generation Sequencing Clinical Assay for Hematologic Malignancies. J Mol Diagn. 2016;18:507–15.

40. Fan J, Bellon M, Ju M, Zhao L, Wei M, Fu L, et al. Clinical significance of FBXW7 loss of function in human cancers. Mol Cancer. 2022;21:87.

41. Izzo F, Myers RM, Ganesan S, Mekerishvili L, Kottapalli S, Prieto T, et al. Mapping genotypes to chromatin accessibility profiles in single cells. Nature. 2024;629:1149–57.

42. Rathgeber AC, Ludwig LS, Penter L. Single-cell genomics-based immune and disease monitoring in blood malignancies. Clin Hematol Int. 2024;6:62–84.

43. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

44. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10:giab008.

45. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

46. Huang X, Huang Y. Cellsnp-lite: an efficient tool for genotyping single cells. Bioinformatics. 2021;37:4569–71.

47. Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Takeshima Y, Luo W, et al. Scalable, multimodal profiling of chromatin accessibility and protein levels in single cells. Genomics; 2020 Sep page 2020.09.08.286914.

48. Lareau CA, Liu V, Muus C, Praktiknjo SD, Nitsch L, Kautz P, et al. Mitochondrial single-cell ATAC-seq for high-throughput multi-omic detection of mitochondrial genotypes and chromatin accessibility. Nat Protoc. Nature Publishing Group; 2023;18:1416–40.

49. Oliveira G, Stromhaug K, Klaeger S, Kula T, Frederick DT, Le PM, et al. Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. Nature. 2021;596:119–25.

50. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nature Genetics. 2021;53:403–11.

51. Melsted P, Booeshaghi AS, Liu L, Gao F, Lu L, Min KH (Joseph), et al. Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat Biotechnol. Nature Publishing Group; 2021;39:813–8.

52. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. Elsevier; 2021;184:3573-3587.e29.

53. Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, et al. Fast alignment and preprocessing of chromatin profiles with Chromap. Nat Commun. Nature Publishing Group; 2021;12:6566.

54. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32:2847–9.

Figure legends

Figure 1. Co-evolution of mitochondrial and somatic nuclear DNA mutations in CLL inferred from WES and RNA-seq data.

A Overview of study design and cohort. Bulk RNA sequencing (RNA-seq) and whole-exome sequencing (WES) data were obtained from 110 serially collected samples of circulating CLL cells from 26 cases (18). Based on WES analysis and calculation of cancer cell fraction (CCF) values, the CLL cases were characterized as genetically stable (n=8; green), evolving during natural disease progression (n=7; light purple) or evolving during fludarabine-based therapy (n=11; dark purple). RNA-seq data was used for identification of mitochondrial DNA (mtDNA) mutations. By matching of mtDNA mutations to CLL subclones identified using WES, co-evolution was assessed.
B Examples of genetically stable (CLL18, left) and evolving CLL (CLL37, right). Changes in CCF and mtDNA heteroplasmy are shown together. Lines indicate longitudinal tracking of CLL subclones (CCF values) or individual mtDNA mutations.
C Number of dynamic somatic nuclear DNA mutations (left) and dynamic mtDNA mutations (right) across genetically stable and evolving CLL cases as defined by changes in CCF values.
D Number of dynamic mtDNA mutations (left) and dynamic somatic nuclear DNA mutations (right) as a function of days between first and last CLL sample.
E Distribution of changes in CCF values (left) and mtDNA heteroplasmy (right) across genetically stable and evolving CLL cases.
F Longitudinal tracking of individual mtDNA mutations above their respective detection limit as indicated by the horizontal lines.

Statistical testing using Student t-test (C, E) or Pearson's correlation coefficient (D).

Figure 2. Co-evolution of mitochondrial and somatic nuclear DNA mutations in CLL at single cell resolution.

A Overview of experimental strategy. Swimmer plot of CLL cases used for analysis. Peripheral blood samples for bulk RNA-seq were obtained before and after therapy with fludarabine, cyclophosphamide and rituximab (FCR) followed by reduced intensity conditioning (RIC) hematopoietic stem cell transplantation (HSCT). Cases indicated by asterisks (*) were also analyzed using single-cell (sc) DNA sequencing (Tapestri platform MissionBio) (E-H).
B Association of the mean increase in heteroplasmy of mitochondrial DNA (mtDNA) mutations and the time from pre-FCR to post-HSCT sample.
C Days from pre-FCR to post-HSCT sample in genetically stable and evolving CLL as characterized by changes in mtDNA mutations.
D Tracking of individual mtDNA mutations in CLL5328.
E Overview of experimental strategy (E-H). To track co-evolution of somatic nuclear and mtDNA mutations, scDNA-seq was performed on 6 peripheral blood samples from 3 patients. Based on previously identified mtDNA mutations from mtscATAC-seq (15) and whole-exome sequencing (19) in these patients, we designed 98 targeted panels for scDNA-seq to track their co-evolution.
F UMAP projections of surface marker expression data (n=5,596) obtained from CLL5328 with 45 Total-seq D antibodies annotated by cell types (left) or identification of CLL-specific mtDNA mutations (3538G>A – common founder mutation, 16247A>G – pre-FCR and 16290C>T – post-HSCT) (top right) and donor immune cell-specific mtDNA mutations (1918G>A and 5650G>A – $CD8^+$ T cells or 786G>A – pan immune cells) (bottom right).
G Single-cell variant allele frequencies (VAFs) and heteroplasmy of mtDNA mutations in CLL5328 pre-FCR and post-HSCT, in recipient-derived (rec.) and donor-derived (donor) immune cells. Fish plot summarizing results (bottom), showing clonal replacement of CLL5328 before and after therapy, marked by 16247A>G (pre-FCR) or 16290T>C and 6426G>A (post-HSCT) alongside 7 (pre-FCR) or 9 (post-HSCT) somatic nuclear DNA mutations.
H Identification of 3 resistant CLL single cells (circled) pre-FCR based on 16290T>C, 6426G>A (both mtDNA), *DPCD*^{D41E} and *ZNF215*^{H42P} (both somatic nuclear) mutations.

Statistical testing with Pearson's correlation coefficient (B) and Wilcoxon signed-rank test (C).

Figure 3. Applicability and sensitivity of donor-recipient deconvolution using mitochondrial DNA mutations.

A Simulation of donor-recipient pairing using 81 published (24) bulk mitochondrial DNA (mtDNA)-seq profiles. A median of 35 (range 2 – 94) homoplasmic mtDNA mutations distinguished simulated donor-recipient pairings.

B Identification of a total of 924 homoplasmic mtDNA mutations across 189 individuals from 3 published datasets (24,25). Shown is the number of total mtDNA mutations as a function of the number of individuals analyzed.

C In-silico mixing experiment using published (15) mitochondrial single-cell chromatin accessibility (mtscATAC-seq) and single-cell RNA sequencing (scRNA-seq) profiles of peripheral blood obtained from pre-transplant CLL patients (CLL4 and CLL5) to benchmark mitochondrial DNA (mtDNA)-based deconvolution with single nucleotide polymorphisms (SNPs).
Top: Correctly annotated CLL4 cells after deconvolution of mixed data for conditions from 1 to 1,000 CLL4 cells spiked into 7,579 CLL5 cells.
Bottom: Germline-free deconvolution of mixed scRNA-seq data with *souporcell*. For fewer than 100 (*souporcell*) CLL4 cells spiked into 10,000 CLL5 cells annotations are randomly assigned (left). Deconvolution using *vireo* and scRNA-seq germlines (right).

D In-vitro mixing experiment. CD28/CD3-bead stimulated and IL7/IL15-expanded T cells transduced with a lentiviral expression vector encoding a MART1-specific murinized T cell receptor (TCR) from donor1 were mixed with unmanipulated peripheral blood mononuclear cells (PBMCs) from donor2 at ratios of 1:3, 1:30 and 1:300. Total-seq B (TSB) oligotags were used for detection of surface marker expression with ASAP-seq. Flow cytometry plot shows the percentage of MART1-specific T cells in donor1 after T cell expansion and transduction with lentiviral expression vector.

E UMAP projection of 13,988 single-cell chromatin accessibility profiles annotated by cell types (top left), donor1 (orange) or donor2 (purple) (top right) and detected fragments of the lentiviral expression vector (red) (bottom left).

F-G Identification of donor1- or donor2-derived cells, doublets and unannotated cells with maternal mtDNA variants.

H Correlation of identified cells from donor1 with experimental mixing conditions.

I Effect of T cell expansion protocol on chromatin profiles of $CD8^+$ T cells demonstrated by differential chromatin accessibility of transcription factor binding motifs.

J Absolute number of MART1-specific T cells detected across donor1 and donor2 across the 3 dilution steps.

Figure 4. Tracking of AML in post-HSCT AML relapse responsive to immunosuppression tapering (IST).
A Overview of experimental strategy. Bone marrow and peripheral blood samples of 4 patients with AML (IST1-4) who had an incipient relapse post-HSCT treated with rapid IST (left) were analyzed by ASAP-seq to capture chromatin accessibility, surface protein expression and mitochondrial DNA (mtDNA) profiles. Surface protein expression was detected using Total-seq B (TSB) antibodies (right).
B UMAP representation of 19,143 single-cell chromatin accessibility profiles annotated by cell types derived from manual annotation (left) and donor-recipient annotation (right).
C Longitudinal dynamics of cell types before (pre, light blue) and after (post, light green) IST in bone marrow (BM, black) or peripheral blood (PB, grey).
D Donor chimerism across various cell types pre and post-IST.
E Surface protein expression (TSB), gene scores, transcription factor motif activity (both inferred from chromatin accessbility) and mtDNA mutations in recipient- and donor-derived monocytes in IST1. The specificity of mtDNA mutations for AML (top), donor-derived immune cells (middle) and donor-recipient deconvolution (bottom) is indicated.
F Surface protein levels of CD11c and CD14 in recipient- and donor-derived monocytes in IST1.
G Chromatin accessibility of *IL1B* in recipient- and donor-derived monocytes in IST1.
H CD4 and CD8 surface protein levels of donor-or recipient-derived T cells in IST3 and IST4.
I Level of surface PD1 in donor- or recipient-derived T cells in IST3.
J Surface protein expression (TSB), gene scores and activity of EOMES transcription factor motif (both inferred from chromatin accessibility) in recipient- and donor-derived T cells in IST3.

Figure 5. Identification of residual donor-derived hematopoiesis and tracking of AML subclones.
A Overview of experimental strategy. ASAP-seq was performed on serial bone marrow samples from 5 participants of the ETCTN/CTEP 10026 study (decitabine and ipilimumab for post-transplant AML relapse).
Study time points (black) included screening, after one cycle of decitabine, following ipilimumab/decitabine combination (C1, C2, C4, C10) and end of treatment. Additional samples for select patients are indicated and include complete remission (CR) and relapse prior to first or following second stem cell transplantation (HSCT1, HSCT2).
Total-seq A (TSA) was used for hashing of samples while Total-seq B (TSB) was used for detection of surface protein expression.
B UMAP projection of single-cell chromatin accessibility profiles annotated into cell types using ASAP-seq (top) and annotated by identified donor-derived non-T cells (black) that represent residual donor-derived hematopoiesis within disease relapse (grey) (bottom). For AML1012, residual hematopoiesis was detected based on absence of copy number variations (CNV).
C Percentage of residual donor-derived hematopoiesis across cell subsets in AML1010, AML1011 and AML1026.
D Correlation of clinical T cell bulk chimerism from short-tandem repeat analysis with mtDNA-based single-cell T cell chimerism.
E Identification of subclones in AML1010 using 71 mtDNA mutations (left) and longitudinal tracking throughout the ETCTN/CTEP 10026 study across cellular compartments (right).
F Longitudinal tracking of mtDNA-derived subclones in AML1012 (left) and AML1011 (right) at indicated timepoints.
G Longitudinal tracking of mtDNA-derived subclones in AML1007 throughout ETCTN/CTEP 10026 and at relapse after a second haploidentical allogeneic hematopoietic stem cell transplantation (left). UMAP representations showing cell types in AML1007 (top left UMAP) and heteroplasmy of 5668G>C at the indicated timepoints (right).

Statistical testing with Pearson's correlation coefficient (D).

Figure 6. Co-evolution of somatic nuclear and mitochondrial DNA (mtDNA) mutations in AML.
A Overview of experimental strategy. Three AML cases were previously genetically characterized using whole-exome sequencing (31).
Through design of an amplicon panel with 127 targets in combination with a pan-mtDNA panel, single-cell DNA sequencing (scDNA-seq) on the Tapestri platform (MissionBio) was used to enable read-out of somatic nuclear and mitochondrial DNA mutations in the same cells. This permitted combined analysis of co-segregation and co-evolution of these two classes of mutations.
B Heatmap of AML1010. Protein expression with Total-seq D (top), somatic nuclear DNA mutations (middle) and mtDNA mutations (bottom) are shown.
C Fish plot of subclonal structure in AML1010 taking into account somatic nuclear and mitochondrial DNA mutations (left). Two mutually exclusive mutations in *TCF* define the main clones of AML1010, while additional mtDNA mutations further refine the identification of subclones. One subclone (brown) is further notable for the absence of $NRAS^{G12D}$.
Changes in subclones of AML1010 across phenotypic compartments during decitabine/ipilimumab treatment on ETCTN/CTEP 10026 (right).
D Heatmap of AML1026. Protein expression with Total-seq D (top), somatic nuclear DNA mutations (middle) and mtDNA mutations (bottom) are shown.
E Fish plot of subclonal structure in AML1026 taking into account somatic nuclear and mitochondrial DNA mutations (left).
While all AML cells harbor mutations in $SRSF2^{P95H}$, $SF3B1^{R775Q}$ and $NRAS^{G13R}$, two main subclones are identified that are driven by $POLR3B^{N101K}$ or $CCDC88B^{A456P}$ and $PLBC2^{F681=}$, respectively.
Changes in subclones of AML1026 during decitabine/ipilimumab treatment on ETCTN/CTEP 10026 (right).

Figure 7. Co-evolution of somatic nuclear and mitochondrial DNA (mtDNA) mutations in AML1012.
A Heatmap of single cell proteogenomic data of AML1012 prior to (Screening) and following four cycles (C4) of decitabine and ipilimumab treatment. Protein expression with Total-seq D (top), identification of somatic nuclear DNA mutations (middle) and mtDNA mutations (bottom) are shown.
B Fish plot of subclonal structure in AML1012 taking into account somatic nuclear and mitochondrial DNA mutations (left). Changes in subclones of AML1012 across phenotypic compartments during decitabine/ipilimumab treatment on ETCTN/CTEP 10026 (right).
C Expansion of $FBXW7^{R465C}$, 13708G>C and 14739G>C in AML1012 after 4 cycles of therapy with decitabine/ipilimumab compared to the screening timepoint. 14739G>C is uniquely detectable in the expanding cluster 4 (orange) (right).

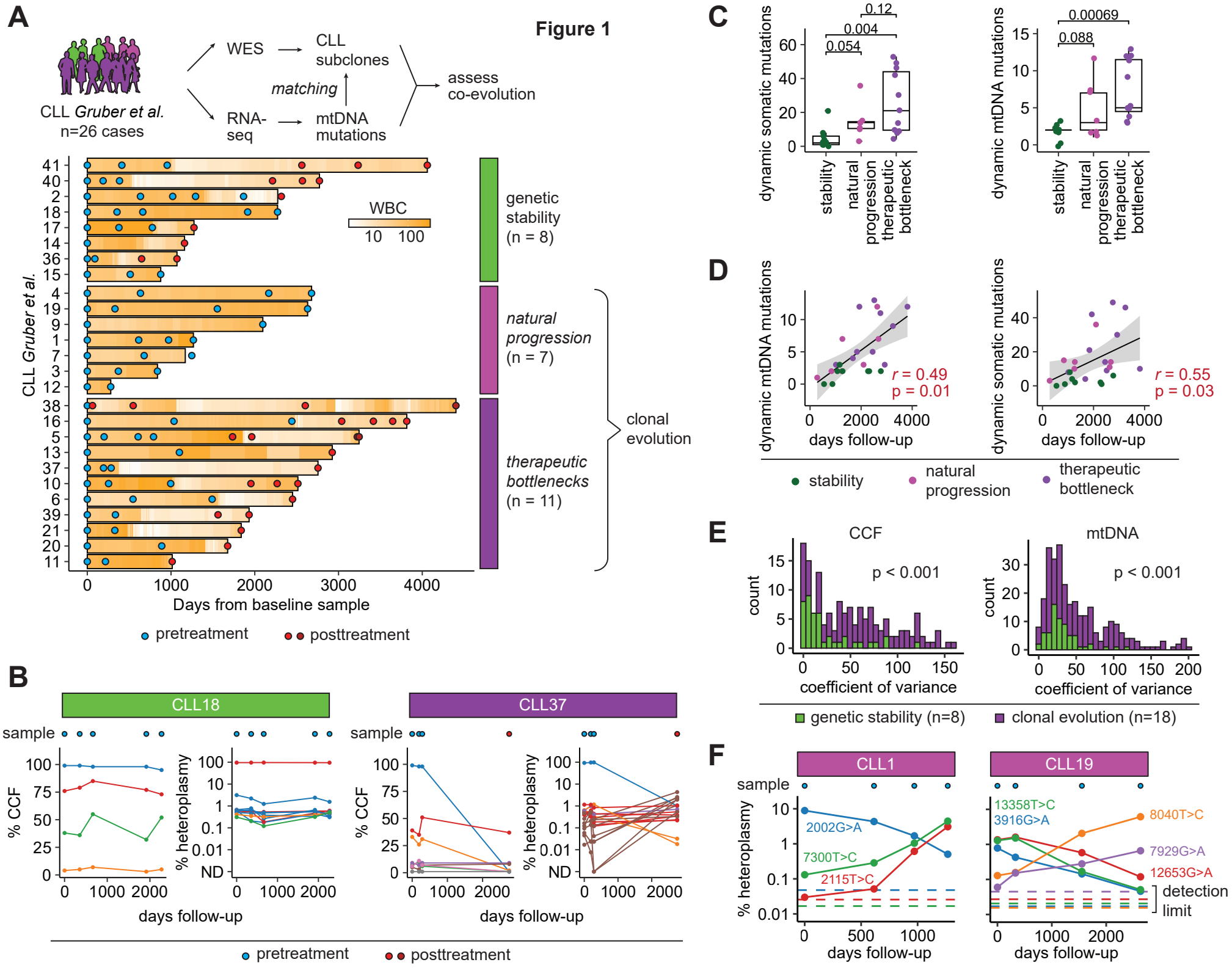


Figure 1
A
CLL Gruber et al.
n=26 cases
WES
CLL subclones
matching
RNA-seq
mtDNA mutations
assess co-evolution
CLL Gruber et al.
WBC
10 100
genetic stability (n = 8)
natural progression (n = 7)
therapeutic bottlenecks (n = 11)
clonal evolution
Days from baseline sample
pretreatment
posttreatment
B
CLL18
CLL37
sample
% CCF
% heteroplasmy
ND
days follow-up
pretreatment
posttreatment
C
dynamic somatic mutations
dynamic mtDNA mutations
0.12
0.004
0.054
0.00069
0.088
stability
natural progression
therapeutic bottleneck
D
dynamic mtDNA mutations
dynamic somatic mutations
r = 0.49
p = 0.01
r = 0.55
p = 0.03
days follow-up
stability
natural progression
therapeutic bottleneck
E
CCF
mtDNA
p < 0.001
count
coefficient of variance
genetic stability (n=8)
clonal evolution (n=18)
F
CLL1
CLL19
sample
% heteroplasmy
2002G>A
7300T>C
2115T>C
13358T>C
3916G>A
8040T>C
7929G>A
12653G>A
detection limit
days follow-up

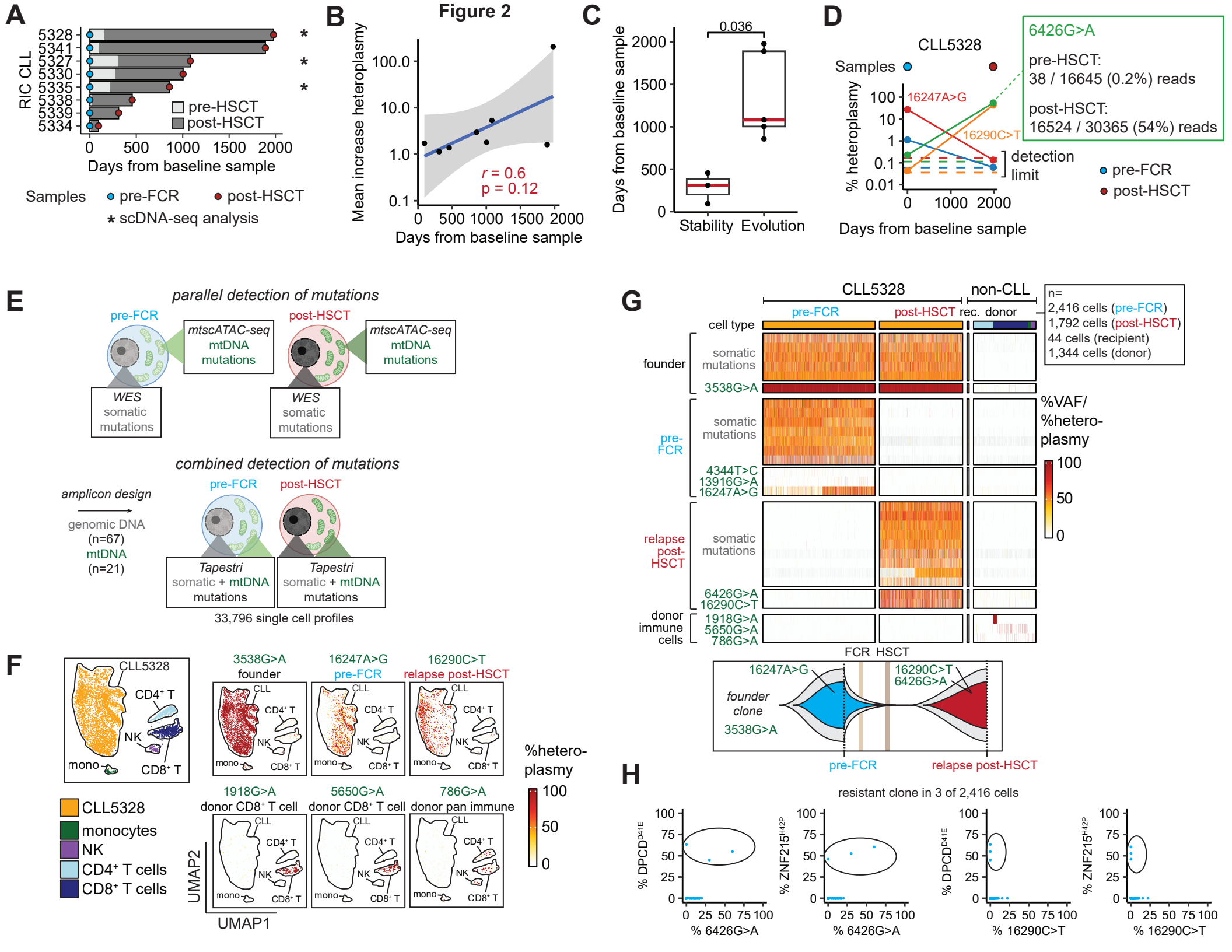


Figure 2
A
RIC CLL
5328
5341
5327
5330
5335
5338
5339
5334
pre-HSCT
post-HSCT
0 500 1000 1500 2000
Days from baseline sample
Samples pre-FCR post-HSCT
scDNA-seq analysis
B
Mean increase heteroplasmy
100.0
10.0
1.0
0.1
r = 0.6
p = 0.12
0 500 1000 1500 2000
Days from baseline sample
C
0.036
Days from baseline sample
2000
1500
1000
500
0
Stability Evolution
D
CLL5328
Samples
% heteroplasmy
100
10
1
0.1
0.01
16247A>G
16290C>T
detection limit
0 1000 2000
Days from baseline sample
pre-FCR
post-HSCT
6426G>A
pre-HSCT:
38 / 16645 (0.2%) reads
post-HSCT:
16524 / 30365 (54%) reads
E
parallel detection of mutations
pre-FCR
post-HSCT
mtscATAC-seq
mtDNA
mutations
WES
somatic
mutations
combined detection of mutations
amplicon design
genomic DNA
(n=67)
mtDNA
(n=21)
Tapestri
somatic + mtDNA
mutations
33,796 single cell profiles
F
CLL5328
CD4+ T
NK
mono
CD8+ T
3538G>A
founder
16247A>G
pre-FCR
16290C>T
relapse post-HSCT
1918G>A
donor CD8+ T cell
5650G>A
donor CD8+ T cell
786G>A
donor pan immune
CLL
mono
%hetero-
plasmy
100
50
0
CLL5328
monocytes
NK
CD4+ T cells
CD8+ T cells
UMAP2
UMAP1
G
CLL5328
non-CLL
pre-FCR
post-HSCT
rec. donor
n=
2,416 cells (pre-FCR)
1,792 cells (post-HSCT)
44 cells (recipient)
1,344 cells (donor)
cell type
founder
somatic mutations
3538G>A
pre-FCR
somatic mutations
4344T>C
13916G>A
16247A>G
relapse post-HSCT
somatic mutations
6426G>A
16290C>T
donor immune cells
1918G>A
5650G>A
786G>A
%VAF/
%hetero-
plasmy
100
50
0
FCR HSCT
16247A>G
16290C>T
6426G>A
founder clone
3538G>A
pre-FCR
relapse post-HSCT
H
resistant clone in 3 of 2,416 cells
% DPCD41E
% ZNF215H42P
% 6426G>A
% 16290C>T
0 25 50 75 100

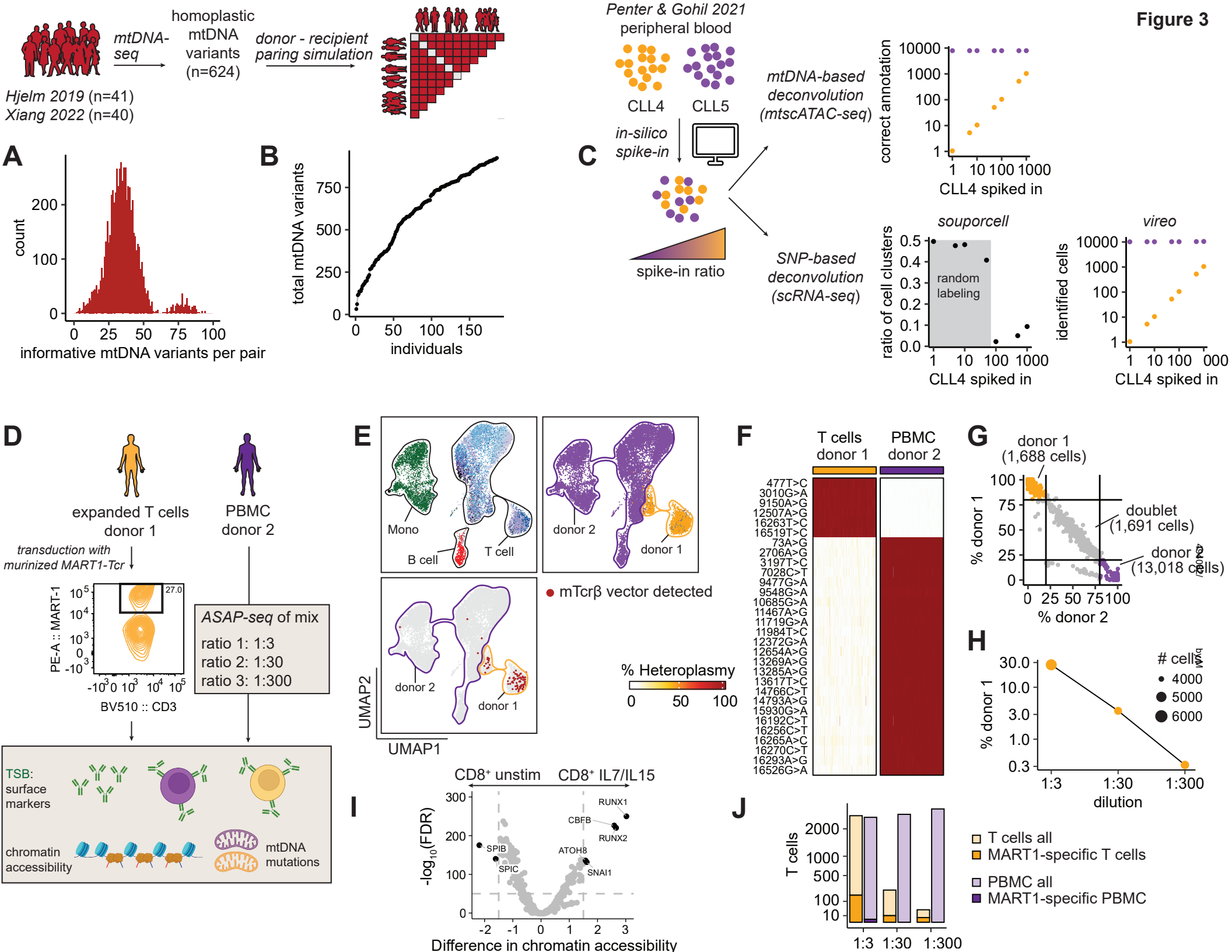


Figure 3
mtDNA-seq
homoplastic mtDNA variants (n=624)
donor - recipient paring simulation
Hjelm 2019 (n=41)
Xiang 2022 (n=40)
A
count
informative mtDNA variants per pair
B
total mtDNA variants
individuals
C
Penter & Gohil 2021
peripheral blood
CLL4
CLL5
in-silico spike-in
spike-in ratio
mtDNA-based deconvolution (mtscATAC-seq)
SNP-based deconvolution (scRNA-seq)
correct annotation
CLL4 spiked in
souporcell
vireo
ratio of cell clusters
identified cells
random labeling
D
expanded T cells donor 1
PBMC donor 2
transduction with murinized MART1-Tcr
PE-A :: MART-1
BV510 :: CD3
ASAP-seq of mix
ratio 1: 1:3
ratio 2: 1:30
ratio 3: 1:300
TSB: surface markers
chromatin accessibility
mtDNA mutations
E
Mono
B cell
T cell
donor 2
donor 1
mTcrβ vector detected
% Heteroplasmy
UMAP1
UMAP2
F
T cells donor 1
PBMC donor 2
G
donor 1 (1,688 cells)
doublet (1,691 cells)
donor 2 (13,018 cells)
% donor 1
% donor 2
H
cells
dilution
I
CD8⁺ unstim
CD8⁺ IL7/IL15
-log10(FDR)
Difference in chromatin accessibility
RUNX1
CBFB
RUNX2
SPIB
SPIC
ATOH8
SNAI1
J
T cells
T cells all
MART1-specific T cells
PBMC all
MART1-specific PBMC

Figure 4

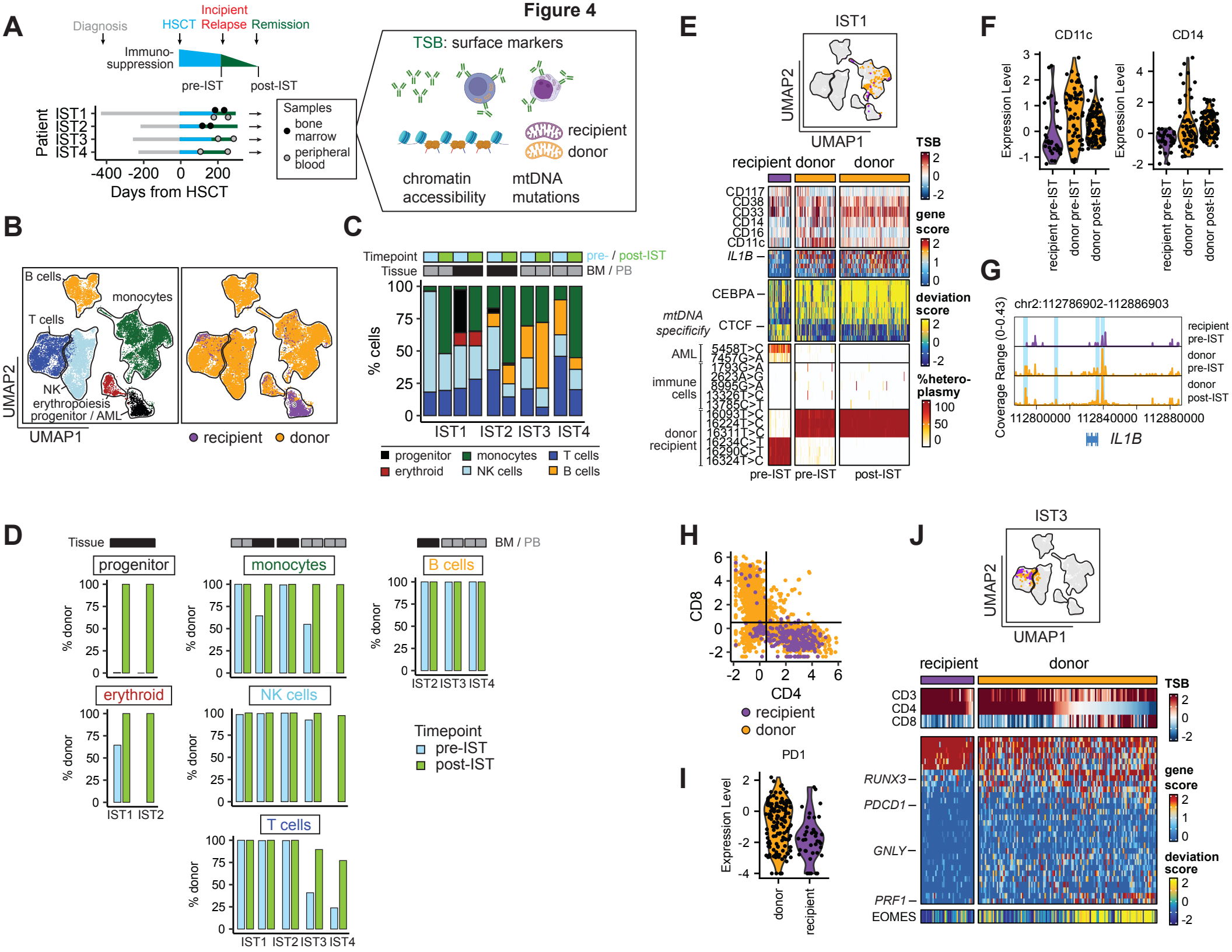

A
Diagnosis
Incipient
HSCT
Relapse
Remission
Immuno-
suppression
pre-IST
post-IST
Patient
IST1
IST2
IST3
IST4
-400 -200 0 200
Days from HSCT
Samples
bone marrow
peripheral blood
TSB: surface markers
recipient
donor
chromatin accessibility
mtDNA mutations
B
B cells
monocytes
T cells
NK
erythropoiesis
progenitor / AML
UMAP2
UMAP1
recipient
donor
C
Timepoint
pre- / post-IST
Tissue
BM / PB
% cells
IST1 IST2 IST3 IST4
progenitor
monocytes
T cells
erythroid
NK cells
B cells
D
Tissue
BM / PB
progenitor
monocytes
B cells
erythroid
NK cells
T cells
% donor
IST1 IST2
IST2 IST3 IST4
IST1 IST2 IST3 IST4
Timepoint
pre-IST
post-IST
E
IST1
UMAP2
UMAP1
recipient donor
donor
TSB
CD117
CD38
CD33
CD14
CD16
CD11c
IL1B
gene score
CEBPA
CTCF
mtDNA specificity
deviation score
AML
5458T>C
7457G>A
1793G>A
2623A>G
8995G>A
immune cells
13326T>C
13785C>T
16093T>C
16224T>C
16311T>C
donor recipient
16234C>T
16290C>T
16324T>C
%hetero-plasmy
pre-IST pre-IST
post-IST
F
CD11c
CD14
Expression Level
recipient pre-IST
donor pre-IST
donor post-IST
G
chr2:112786902-112886903
Coverage Range (0-0.43)
recipient pre-IST
donor pre-IST
donor post-IST
112800000 112840000 112880000
IL1B
H
CD8
CD4
recipient
donor
I
PD1
Expression Level
donor
recipient
J
IST3
UMAP2
UMAP1
recipient
donor
TSB
CD3
CD4
CD8
RUNX3
PDCD1
GNLY
PRF1
EOMES
gene score
deviation score

Figure 5

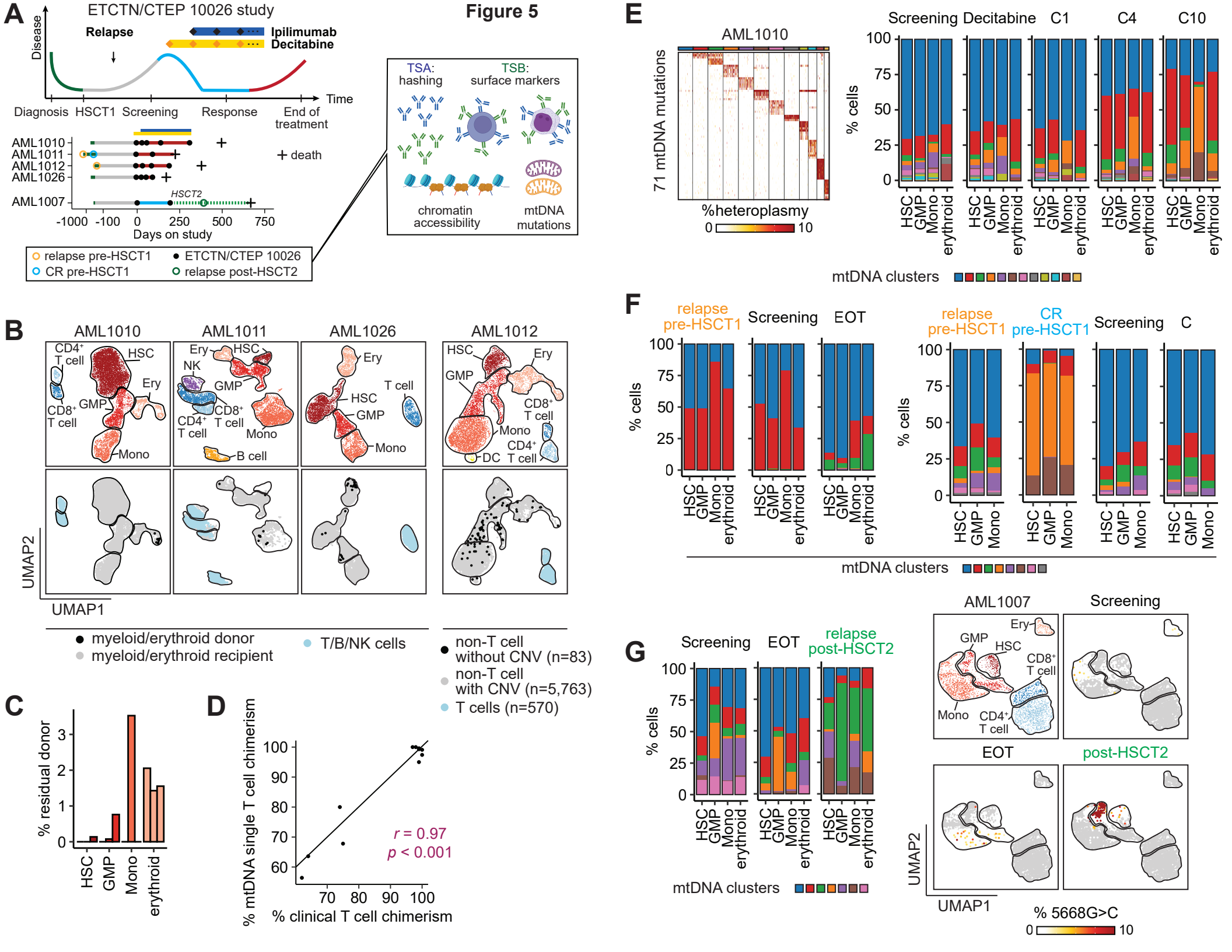

A
ETCTN/CTEP 10026 study
Relapse
Ipilimumab
Decitabine
Disease
Time
Diagnosis HSCT1 Screening Response End of treatment
AML1010
AML1011
AML1012
AML1026
AML1007
HSCT2
death
-1000 -100 0 250 500 750
Days on study
relapse pre-HSCT1
CR pre-HSCT1
ETCTN/CTEP 10026
relapse post-HSCT2
TSA: hashing
TSB: surface markers
chromatin accessibility
mtDNA mutations
B
AML1010
AML1011
AML1026
AML1012
CD4+ T cell
HSC
Ery
CD8+ T cell
GMP
Mono
NK
B cell
DC
UMAP2
UMAP1
myeloid/erythroid donor
myeloid/erythroid recipient
T/B/NK cells
non-T cell without CNV (n=83)
non-T cell with CNV (n=5,763)
T cells (n=570)
C
% residual donor
HSC GMP Mono erythroid
D
% mtDNA single T cell chimerism
% clinical T cell chimerism
r = 0.97
p < 0.001
E
AML1010
71 mtDNA mutations
%heteroplasmy
0 10
Screening Decitabine C1 C4 C10
% cells
HSC GMP Mono erythroid
mtDNA clusters
F
relapse pre-HSCT1
Screening
EOT
CR pre-HSCT1
C
% cells
HSC GMP Mono erythroid
mtDNA clusters
G
Screening
EOT
relapse post-HSCT2
% cells
mtDNA clusters
AML1007
Screening
Ery
GMP
HSC
CD8+ T cell
Mono
CD4+ T cell
EOT
post-HSCT2
UMAP2
UMAP1
% 5668G>C
0 10

Figure 6

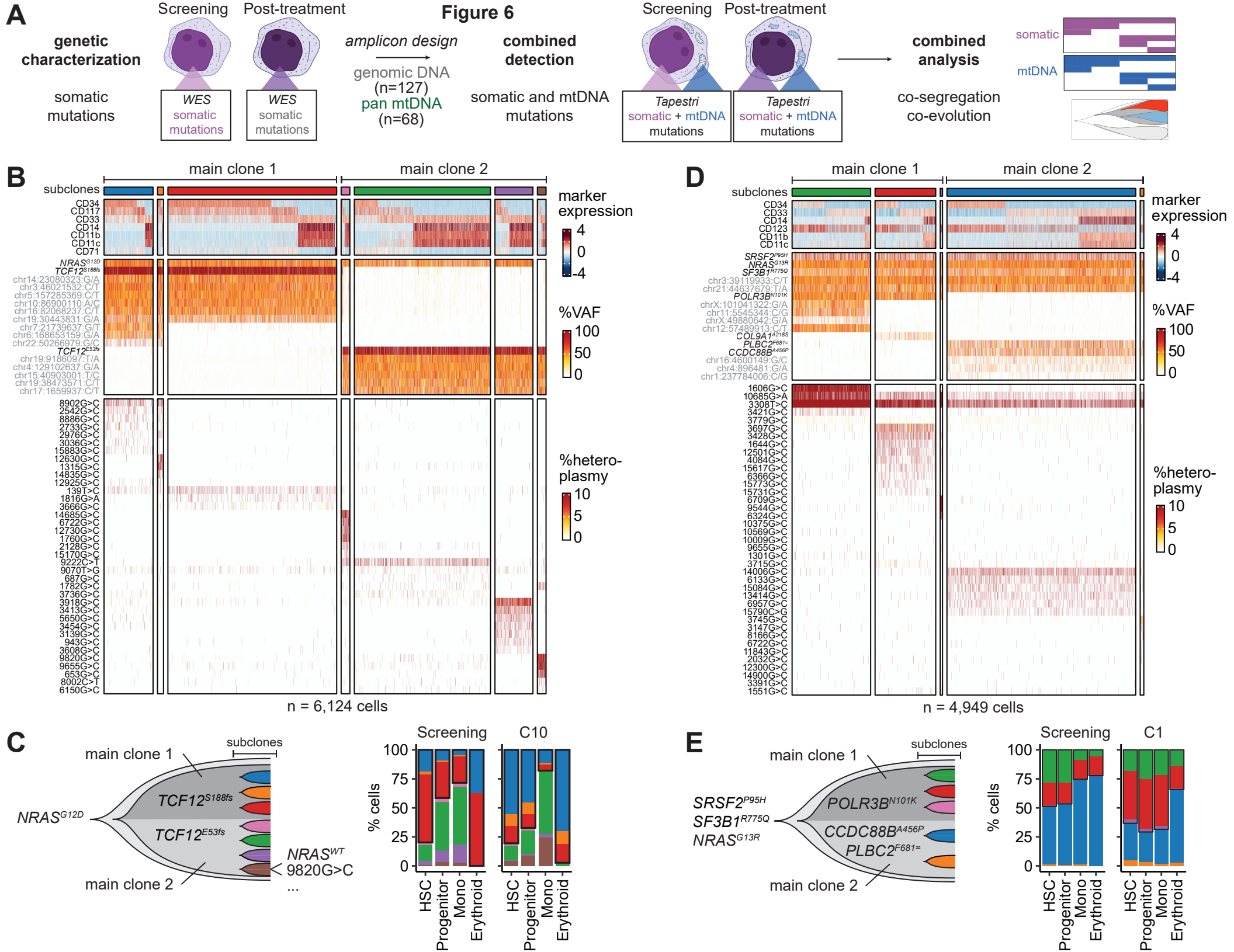
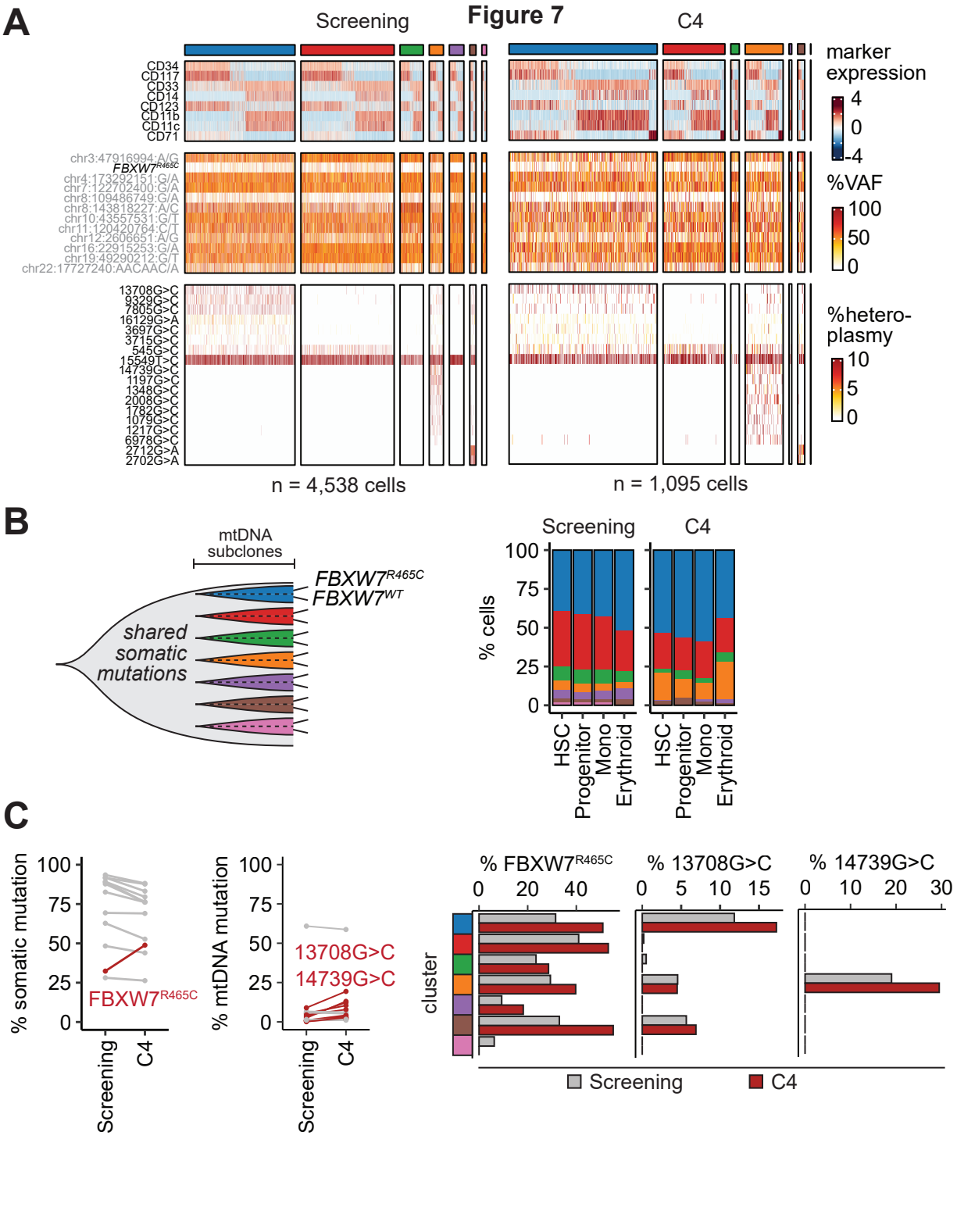

A
genetic characterization
somatic mutations
Screening
Post-treatment
WES somatic mutations
WES somatic mutations
amplicon design
genomic DNA (n=127)
pan mtDNA (n=68)
combined detection
somatic and mtDNA mutations
Screening
Post-treatment
Tapestri somatic + mtDNA mutations
Tapestri somatic + mtDNA mutations
combined analysis
co-segregation
co-evolution
somatic
mtDNA
B
main clone 1
main clone 2
subclones
CD34
CD117
CD33
CD14
CD11b
CD11c
CD71
NRAS G12D
TCF12 S188fs
chr14:23080323:G/A
chr3:46021532:C/T
chr5:157285369:C/T
chr10:86900110:A/C
chr16:82068237:C/T
chr19:30443831:G/A
chr7:21739637:G/T
chr6:168653159:G/A
chr22:50266979:G/C
TCF12 E53fs
chr19:91860097:T/A
chr4:129102637:G/A
chr15:40903001:T/C
chr19:38473571:C/T
chr17:1659937:C/T
8902G>C
2542G>C
8886G>C
2733G>C
2976G>C
3036G>C
15883G>C
12630G>C
1315G>C
14835G>C
12925G>C
139T>C
1816G>A
3666G>C
14685G>C
6722G>C
12730G>C
1760G>C
2128G>C
15170G>C
9222C>T
9070T>G
687G>C
1782G>C
3736G>C
3918G>C
3413G>C
5650G>C
3454G>C
3139G>C
943G>C
3608G>C
9820G>C
9655G>C
653G>C
8002C>T
6150G>C
marker expression
4
0
-4
%VAF
100
50
0
%hetero-plasmy
10
5
0
n = 6,124 cells
D
main clone 1
main clone 2
subclones
CD34
CD33
CD14
CD123
CD11b
CD11c
SRSF2 P95H
NRAS G13R
SF3B1 R775Q
chr3:39119933:C/T
chr21:44637679:T/A
POLR3B N101K
chrX:101041322:G/A
chr11:5545344:C/G
chrX:49880642:G/A
chr12:57489913:G/T
COL9A1 A218S
PLBC2 F681=
CCDC88B A456P
chr16:4600149:G/C
chr4:896481:G/A
chr1:237784006:C/G
1606G>C
10685G>A
3308T>C
3421G>C
3779G>C
3697G>C
3428G>C
1644G>C
12501G>C
4084G>C
15617G>C
6366G>C
15773G>C
15731G>C
6709G>C
9544G>C
6324G>C
10375G>C
10569G>C
10009G>C
9655G>C
1301G>C
3715G>C
14006G>C
6133G>C
15084G>C
13414G>C
6957G>C
15790C>G
3745G>C
3147G>C
8166G>C
6722G>C
11843G>C
2032G>C
12300G>C
14900G>C
3391G>C
1551G>C
marker expression
4
0
-4
%VAF
100
50
0
%hetero-plasmy
10
5
0
n = 4,949 cells
C
main clone 1
subclones
NRAS G12D
TCF12 S188fs
TCF12 E53fs
NRAS WT
9820G>C
...
main clone 2
Screening
C10
% cells
100
75
50
25
0
HSC
Progenitor
Mono
Erythroid
E
main clone 1
subclones
SRSF2 P95H
SF3B1 R775Q
NRAS G13R
POLR3B N101K
CCDC88B A456P
PLBC2 F681=
main clone 2
Screening
C1
% cells
100
75
50
25
0
HSC
Progenitor
Mono
Erythroid

Figure 7


A
Screening
C4
CD34
CD117
CD33
CD14
CD123
CD11b
CD11c
CD71
marker expression
4
0
-4
chr3:47916994:A/G
FBXW7R465C
chr4:173292151:G/A
chr7:122702400:G/A
chr8:109486749:G/A
chr8:143818227:A/C
chr10:43557531:G/T
chr11:120420764:C/T
chr12:2606651:A/G
chr16:22915253:G/A
chr19:49290212:G/T
chr22:17727240:AACAAC/A
%VAF
100
50
0
13708G>C
9329G>C
7805G>C
16129G>A
3697G>C
3715G>C
545G>C
15549T>C
14739G>C
1197G>C
1348G>C
2008G>C
1782G>C
1079G>C
1217G>C
6978G>C
2712G>A
2702G>A
%hetero-plasmy
10
5
0
n = 4,538 cells
n = 1,095 cells
B
mtDNA subclones
FBXW7R465C
FBXW7WT
shared somatic mutations
Screening
C4
% cells
100
75
50
25
0
HSC
Progenitor
Mono
Erythroid
C
% somatic mutation
FBXW7R465C
Screening
C4
% mtDNA mutation
13708G>C
14739G>C
% FBXW7R465C
% 13708G>C
% 14739G>C
cluster
Screening
C4