**Online Supplementary Appendix for:**

**Longitudinal single-cell dynamics of chromatin accessibility and mitochondrial mutations in chronic lymphocytic leukemia mirror disease history**

**Supplementary methods**

*Whole-exome sequencing and somatic copy number identification*

Sequencing of CLL and RT samples was performed at the Broad Institute of MIT and Harvard. Copy number changes, somatic nuclear mutations and clonal dynamics were identified as previously described (1). Whole exome sequencing data of CLL1 had previously been analyzed (1).

*Reanalysis of public bulk ATAC-seq data*

Analysis was performed as previously described (2). In short, raw sequencing data was downloaded from Gene Expression Omnibus (GEO) accession number GSE111015. Alignment to the GRCh38 reference genome was performed using bowtie2 (3). Extraction of reads mapping to the mitochondrial genome was done using Samtools (4). Heteroplasmy was calculated using mgatk (5) and custom scripts. Quality filters were coverage of variant >10x and coverage of position >200x.

*Reanalysis of public mtscATAC-seq data*

Re-analysis of mtscATAC-seq profiles of CLL A and B was based on preprocessed output files from cellranger and mgatk which were kindly provided by Caleb Lareau. Mitochondrial DNA mutations were called from combined mtscATAC-seq profiles of CD19+ and CD19- FACS sorted cells for each patient using the function call\_mutations\_mgatk() published along with the original manuscript (https://github.com/caleblareau/mtscATACpaper\_reproducibility) as well as custom scripts.

*Identification of CD5- B cells*

Physiologic CD5- B cells were identified based on imputed gene activity scores of CD5 and clustering of single cell chromatin accessibility data using ArchR::addClusters() with *reducedDims* = 'IterativeLSI', *method* = 'Seurat', *name* = 'Clusters', *resolution* = 1 (6). Estimated purity of CLL cells was calculated as percentage of cells within CLL cluster amongst all CLL/B cells. For comparison with a healthy donor, single cell chromatin accessibility data from a healthy donor were downloaded from 10x Genomics (atac\_pbmc\_10k\_v1) and processed using Cell Ranger ATAC version 1.2.0 with a custom GRCh38 reference genome.

**Supplementary Tables**

**Supplementary Table 1.** Sample characteristics.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CLL Patient ID** | **Sample** | **Tissue** | **Condition at time of sample** | **# scATAC-seq profiles** | **# cells with**  **mtDNA coverage > 20x** | **# scRNA-seq profiles** | **WBC at time of sampling [109/L]** | **% CD19+ CD5+** |
| 1 | 1 | PB | W&W | 6,676 | 5,721 | 7,195 | 30.8 | 72.30 |
| 1 | 2 | PB | W&W | 6,014 | 5,451 | N/A | 80.3 | N/A |
| 1 | 3 | PB | Pretreatment FCR | 4,125 | 3,592 | 6,016 | 439 | 100.00 |
| 1 | 4 | PB | Remission | 4,111 | 263\* | N/A | 9.6 | 0.00 |
| 1 | 5 | PB | Relapse post-FCR | 1,715 | 569 | 8,094 | 110.5 | 42.40 |
| 2 | 1 | PB | Pretreatment FCR | 5,725 | 3,231 | 14,144 | 142 | 100.00 |
| 2 | 2 | PB | Relapse post-FCR | 1,717 | 1,142 | 13,156 | 101.7 | N/A |
| 3 | 1 | PB | Pretreatment FCR | 3,862 | 303\* | 19,933 | 42.1 | 100.00 |
| 3 | 2 | PB | Relapse post-FCR | 4,567 | 3,566\* | 13,551 | 12.2 | N/A |
| 4 | 1 | PB | Pretreatment FCR | 3,447 | 2,710 | 43,384 | 110.4 | 100.00 |
| 4 | 2 | PB | Relapse post-RIC | 2,810 | 2,272 | 27,085 | 30.62 | N/A |
| 5 | 1 | PB | Pretreatment FCR | 7,790 | 4,554 | 45,560 | 57.5 | 99.00 |
| 5 | 2 | PB | Relapse post-RIC | 4,698 | 3,384 | 41,099 | 20.8 | 88.50 |
| 6 | 1 | PB | Pretreatment FCR | 5,125 | 4,061 | 9,983 | 65 | 100.00 |
| 6 | 2 | PB | Relapse post-RIC | 6,398 | 4,949 | 10,901 | 13.3 | 83.40 |
| 6 | 3 | PB | Ibrutinib 1 month | 8,550 | 5,206 | 11,264 | 33.6 | 97.50 |
| 6 | 4 | PB | Ibrutinib 21 months | 9,633 | 2,862 | 10,626 | 12 | 100.00 |
| 7 | 1 | PB | Pretreatment Ibrutinib CD19- cells | 4,219 | 2,855 | 5,850 | 12.1 | 99.90 |
| 7 | 1 | PB | Pretreatment Ibrutinib CD19+ cells | 3,004 | 2,085 | 12.1 | 99.90 |
| 7 | 2 | PB | Ibrutinib 6 months CD19- cells | 5,558 | 2,308 | 7,752 | 16.4 | N/A |
| 7 | 2 | PB | Ibrutinib 6 months CD19+ cells | 4,152 | 1,208 | 16.4 | N/A |
| 7 | 3 | PB | Ibrutinib 16 months | 2,597 | 2,072 | 31,521 | 6.7 | N/A |
| 8 | 1 | PB | Pretreatment Ibrutinib | 501 | 269 | 1,527 | 435.3 | 93.80 |
| 8 | 2 | PB | Ibrutinib 7 months | 4,194 | 1,721 | 10,022 | 607 | N/A |
| 8 | 3 | PB | Ibrutinib 14 months | 4,477 | 941 | 35,135 | 603.6 | N/A |
| 9 | 1 | PB | Peripheral blood CLL phase | 4,609 | 3,473 | 14,113 | 22.12 | 56.00 |
| 9 | 2 | PB | Peripheral blood Richter's phase | 5,692 | 4,160 | 5,373 | 45.56 | N/A |
| 9 | 3 | BM | Bone marrow Richter's phase | 7,480 | 4,948 | 17,085 | 45.56 | 74.00 |
| 9 | 4 | LN | Lymph node Richter's phase CD19+ cells | 24,945 | 15,215 | N/A | 45.56 | 100.00 |
| 9 | 4 | LN | Lymph node Richter's phase CD19- cells | 4,888 | 2,599 | N/A | 45.56 | 100.00 |

PB: peripheral blood; BM: bone marrow; LN: lymph node; W&W: watchful waiting; FCR: fludarabine, cyclophosphamide, rituximab; RIC: reduced-intensity conditioning; mtDNA: mitochondrial DNA; WBC: white blood cell count; N/A: not available

\*Due to low coverage mtDNA cut-off of 10x was used

**Supplementary Figures**

**Suppl. Fig. 1**. Single cell chromatin profiles of CLL cells.

Diagram

Description automatically generated

**(A-C)** Transcription start site (TSS) enrichment score and number of unique fragments per cell. The example of CLL9 is shown at time of Richter’s syndrome for peripheral blood, bone marrow and lymph node.  **(D)** Browser tracks of *CD5* gene shown for healthy donor (10k peripheral blood mononuclear cells from 10x Genomics) and CLL3 after treatment with fludarabine, cyclophosphamide, rituximab (FCR) (post-FCR).   
**(E)** Estimated CLL purity using clustering of single cell chromatin accessibility data and gene activity scores of *CD5* in CLL/B cell compartment across samples of cohort.   
**(F)** UMAP representation of single cell chromatin accessibility profiles obtained from CLL3.  
**(G)** Gene activity scores for *PAX5* and *CD5* imputed from single cell chromatin accessibility profiles from CLL3.   
**(H)** Differential analysis of gene activity scores between B cells in CLL cells from CLL3.   
**(I)** Comparison of heteroplasmy of mitochondrial DNA mutations in CD5- B cells and CLL cells in CLL3.

**Suppl. Fig. 2**. Identification and analysis of mtDNA mutations in study cohort.

Diagram

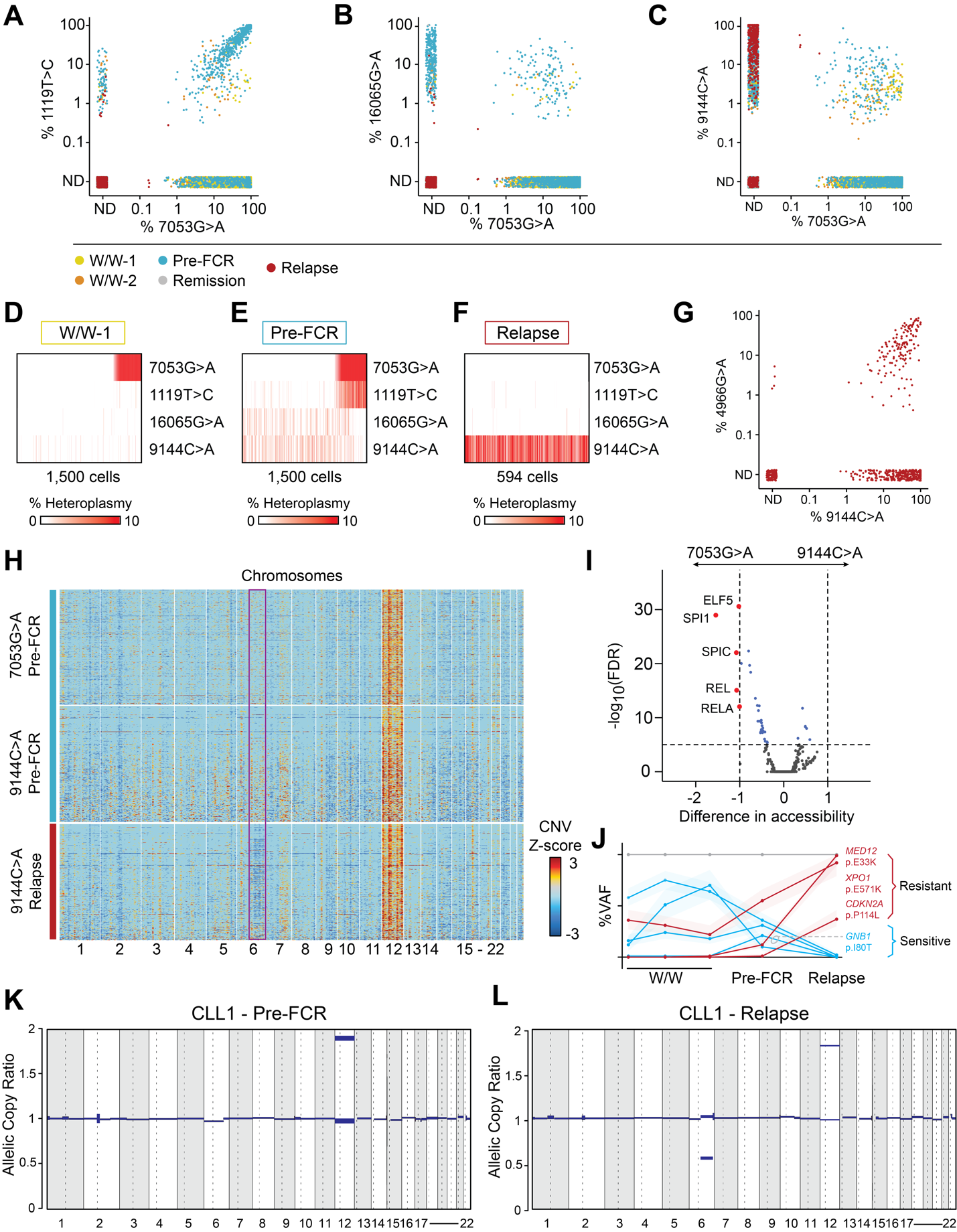
Description automatically generated

**(A)** Substitution rate (observed over expected) of mtDNA mutations (y-axis) in each category of mononucleotide and trinucleotide modifications color-coded by the mtDNA heavy (H) and light (L) strand.  
**(B-D)** Number of detected mtDNA mutations versus age at diagnosis of patients (B), number of mtDNA mutations versus number of sequenced CLL cells (C) and evenness of the distribution of mtDNA mutations (normalized Shannon Index) versus number of detected mtDNA mutations (D). Grey areas indicate 95% confidence interval. Correlation coefficients *R* were calculated using Spearman’s correlation.   
**(E)** Mean heteroplasmy of recurrent mtDNA mutations (black) in the CLL cell compartment from 11 CLL patients. Grey - mtDNA mutations considered non-recurrent (presence in <4 patients or mean heteroplasmy <1%).

**(F)** mtDNA mutations detected in %T cells versus %monocytes. Significantly enriched mutations (p < 0.05; Fisher’s exact test) are colored for T cells (blue) and monocytes (green).

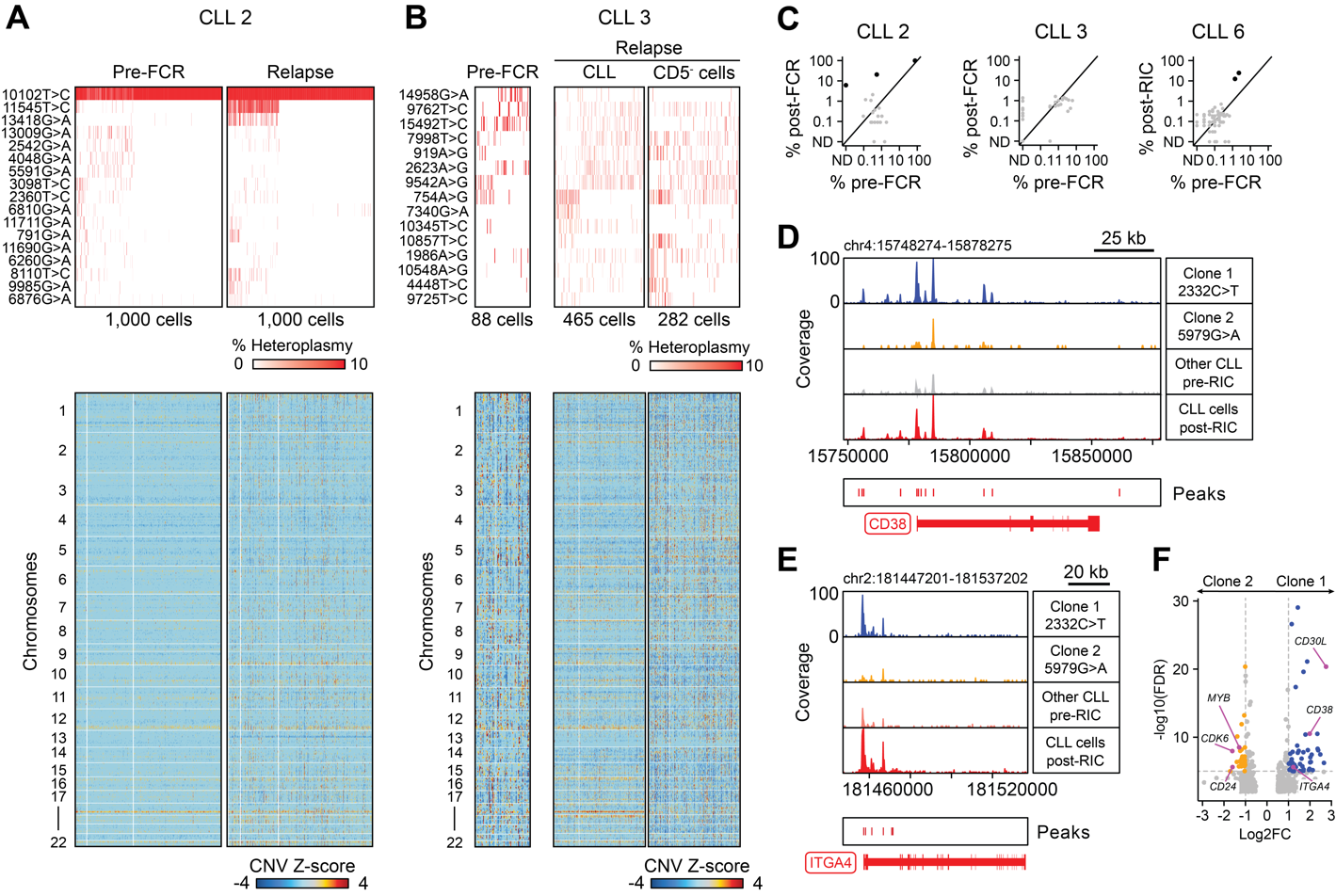
**(G, H)** Distribution of %T cells (blue) or %monocytes (green) marked by mtDNA mutations with significant enrichment in either T cells (G) or monocytes (H).

**Suppl. Fig. 3.** Mitochondrial DNA mutations reveal subclonal structure of CLL1.



**(A-C)** Heteroplasmy of 7053G>A versus 1119T>C, 16065G>A and 9144C>A in CLL cells across all 5 timepoints (color-coded as shown in legend).  
**(D-F)** Heteroplasmy of 7053G>A, 1119T>C, 16065G>A and 9144C>A in CLL cells during watchful waiting (W/W-1), before chemoimmunotherapy with fludarabine, cyclophosphamide and rituximab (FCR) (Pre-FCR) and at relapse. Only cells with at least one detectable mutation are shown.  
**(G)** Heteroplasmy of 9144G>A versus 4966G>A in CLL cells at relapse after FCR.  
**(H)** Inferred copy number changes from scATAC-seq data in cells marked by 7053G>A or 9144C>A before FCR (Pre-FCR) or marked by 9144C>A at relapse. The rectangle demonstrates partial loss of chromosome 6q at relapse.   
**(I)** Differential accessibility of transcription factor motifs before chemotherapy (Pre-FCR) in cells marked by 7053G>A (left) or 9144C>A (right).   
**(J)** Variant allele frequency (VAF) of mutations detected using whole-exome sequencing (WES) in CLL1 at indicated timepoints. Data were previously generated and analyzed (1). Red – enriched at relapse, Blue – eradicated by FCR  
**(K, L)** Copy number changes detected using WES in CLL1 at indicated timepoints.

**Suppl. Fig. 4.** Changes in mtDNA mutations at relapse following chemoimmunotherapy and stem cell transplantation.



**(A, B)** Mitochondrial DNA mutations and inferred copy number changes from scATAC-seq data in CLL cells (CLL2) and CLL versus physiologic CD5- B cells (CLL3) before fludarabine, cyclophosphamide, rituximab (FCR) (Pre-FCR) and at relapse after FCR (Relapse). Only cells with at least one detectable mutation are shown.

**(C)** Frequency of CLL cells marked by mtDNA mutations before (%pre-FCR) and at relapse after FCR (%post-FCR) or reduced intensity conditioning allogeneic stem cell transplantation (%post-RIC). Corresponding heatmaps with mtDNA mutations and copy number changes of CLL6 are shown in **Fig. 4**.

**(D, E)** Browser tracks showing differential chromatin accessibility of *CD38* and *ITGA4* (encoding CD49d) between 2 subclones marked by 2332C>T and 5979G>A (CLL5).   
**(F)** Differential gene scores between single cells from clone 1 marked by 2332C>T and clone 2 marked by 5979G>A (CLL5).

**Suppl. Fig. 5.** Changes in chromatin accessibility and gene expression at relapse after chemoimmunotherapy.

Diagram

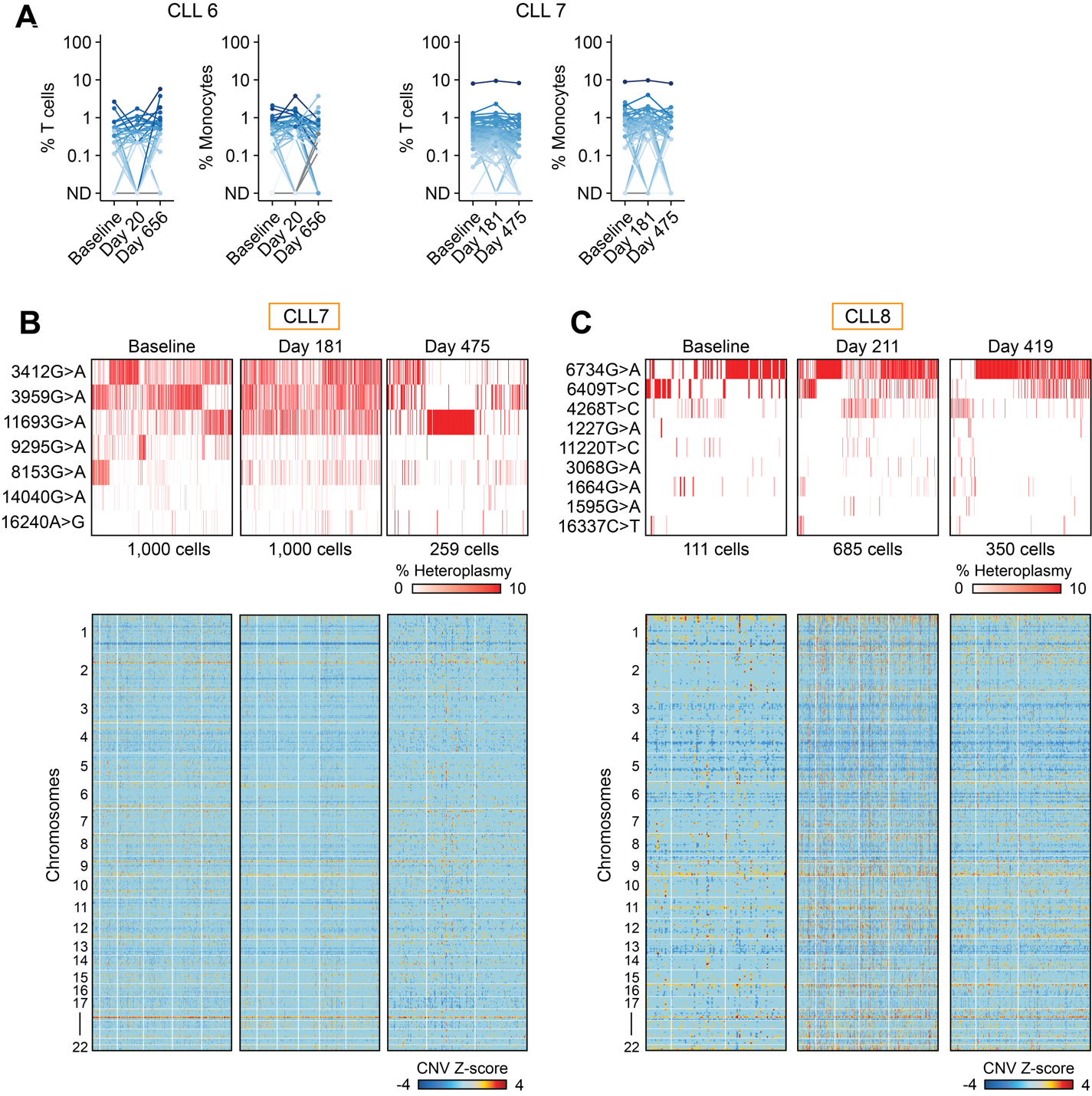
Description automatically generated with low confidence

**(A)** Deviation scores of SPIB, SPI1, BCL11B, BCL11A and IRF1 in CLL1-6 before chemotherapy with fludarabine, cyclophosphamide and rituximab (FCR) (pre) and at relapse after FCR or allogeneic stem cell transplantation (RIC allo-HSCT) (post). Shown are 100 cells per CLL and timepoint.   
**(B)** UMAP plot of scRNA-seq profiles of 72,401 CLL/B cells from CLL1-6 used for differential gene expression analysis. Color shading indicates timepoint (light = pre-FCR; dark = relapse after FCR or RIC allo-HSCT).

**(C)** Differential gene expression analysis based on scRNA-seq profiles shown in (B) pre-FCR (pre) versus relapse after FCR or RIC allo-HSCT (post).

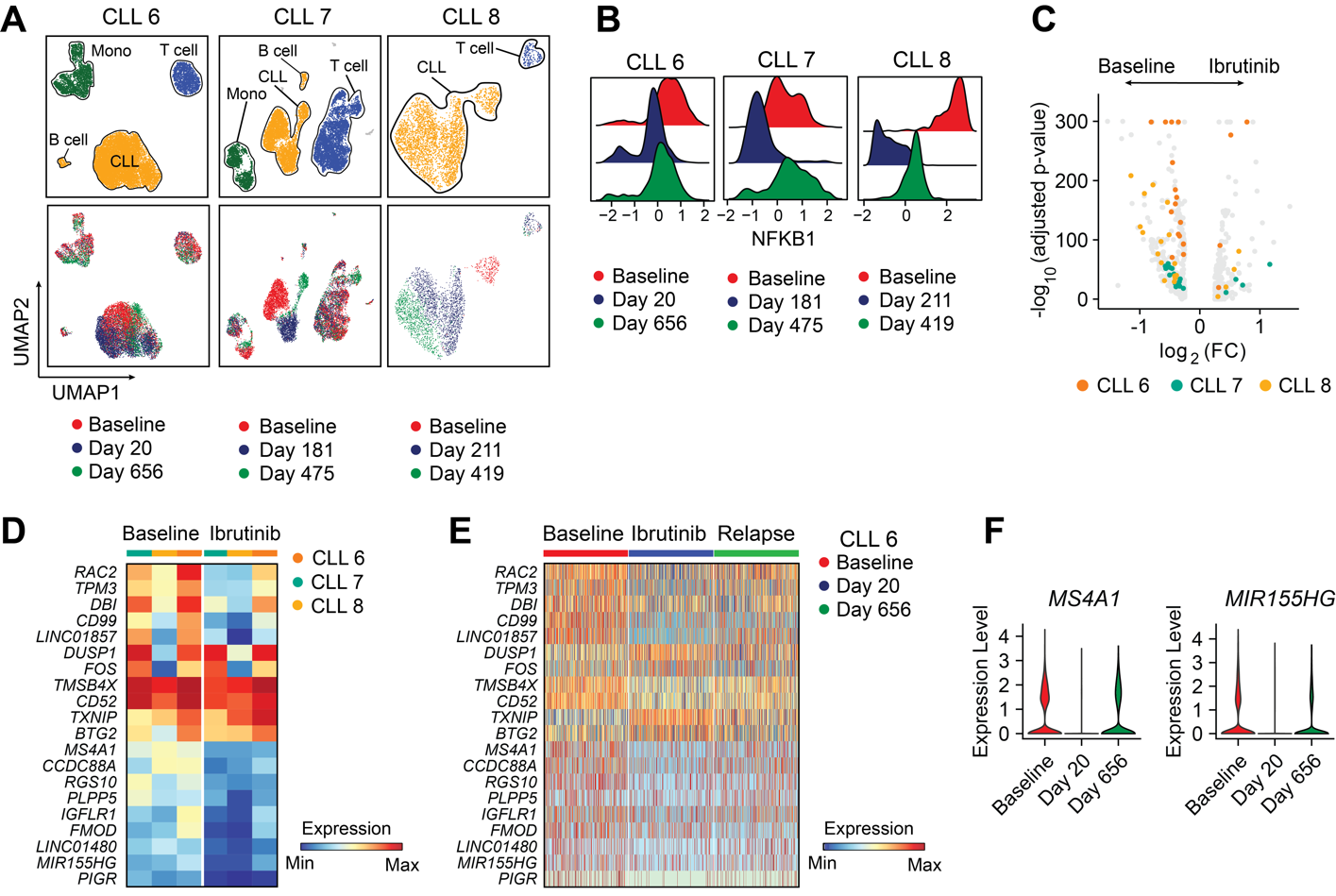
**(D, E)** Expression of genes downregulated at relapse (*CD24, MEF2C*) or upregulated at relapse (*CXCR4, RGS1/2*) after FCR or RIC allo-HSCT across CLL1-6.

**Suppl. Fig. 6.** Changes in mtDNA mutations during ibrutinib treatment.



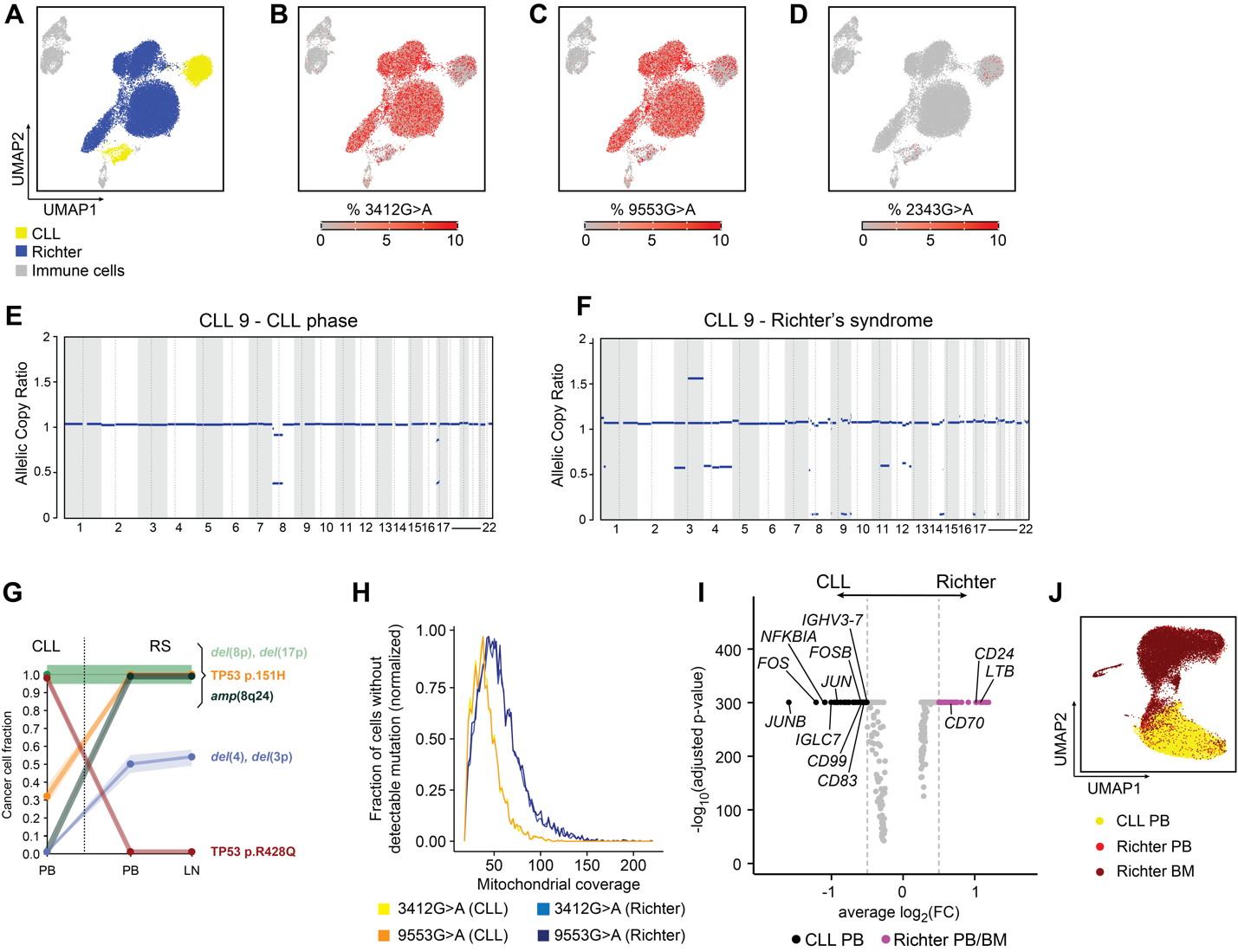
**(A)** Percentage of T cells and monocytes marked by different mtDNA mutations during ibrutinib treatment in CLL6 and CLL7.   
**(B, C)** Mitochondrial DNA mutations and inferred copy number changes from scATAC-seq data in CLL cells (CLL7 and CLL8) before and during ibrutinib treatment. Only cells with at least one detectable mutation are shown.

**Suppl. Fig. 7.** Gene expression and chromatin state changes during ibrutinib treatment.



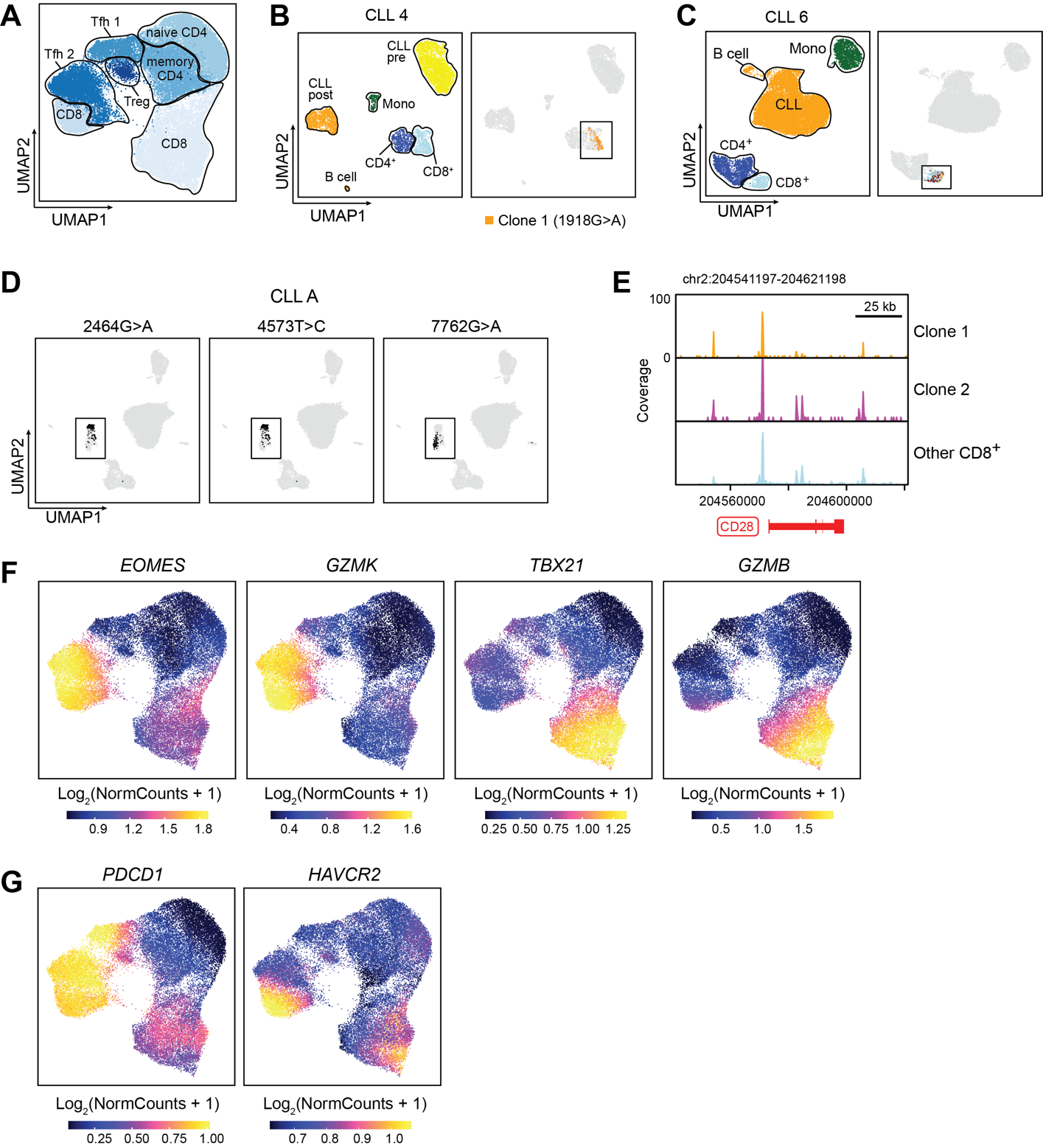
**(A)** UMAP plots based on single cell chromatin accessibility profiles (CLL6-8) showing cell types (top) and time of sampling (bottom).   
**(B)** NFKB1 transcription factor motif binding before ibrutinib treatment (red), after 20-211 days (blue) and after 419-656 days (green) of ibrutinib treatment.   
**(C, D)** Differential gene expression analysis (scRNA-seq) of CLL/B cells during ibrutinib treatment comparing baseline samples with samples obtained after 20-211 days.   
**(E, F)** Gene expression in CLL/B cells before, during ibrutinib treatment (day 20) and at relapse (day 656) (CLL6).

**Suppl. Fig. 8.** Distribution of disease-specific mtDNA mutations within CLL and RS cells in CLL9.

****

**(A-D)** Distribution of 3412G>A, 9553G>A and 2343G>A across CLL (yellow), Richter’s (blue) and immune cells (grey) in CLL9.  
**(E, F)** Copy number alterations calculated from whole-exome sequencing (WES) data of CLL and Richter’s cells.  
**(G)** Clonal dynamics inferred from nuclear somatic mutations from WES. RS – Richter’s syndrome.  
**(H)** Distribution of mtDNA coverage in cells without detectable 3412G>A and 9553G>A mutation in CLL (yellow and orange) and Richter’s cells (light blue and dark blue).  
**(I)** Differential gene expression (scRNA-seq) of CLL/B cells from peripheral blood (PB) during CLL phase versus PB and bone marrow (BM) during transformation to Richter’s syndrome. **(J)** UMAP of scRNA-seq profiles of CLL cells from CLL PB (yellow), Richter’s PB (red) or Richter’s BM (brown).

**Suppl. Fig. 9.** Phenotypes of lymph node-associated T cells and mtDNA mutations as markers of clonal T cell expansion.



**(A)** Cluster annotation of UMAP plot showing 33,573 scATAC-seq profiles of T cells from CLL1-9. Tfh - T follicular helper cell, Treg - regulatory T cell

**(B)** Identification of clonally expanded CD8+ T cells marked by 1918G>A (CLL4).

**(C)** Identification of two CD8+ T cell clones marked by 2647G>A (clone 1; yellow) and 10408T>C (clone 2; brown) in CLL6.  
**(D)** Identification of two CD8+ T cell clones marked by 2464G>A and 4573T>C (clone 1) or 7762G>A (clone 2) in CLLA.

**(E)** Chromatin tracks of *CD28* across T cell clones defined by mtDNA mutations in CLLA.

**(F, G)** Gene activity scores of *EOMES*, *GZMK*, *TBX21*, *GZMB*, *PDCD1* (encoding for PD-1) and *HAVCR2* (TIM-3) in CLL1-9.

**Supplementary References**

1. 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.

2. Lareau CA, Ludwig LS, Sankaran VG. Longitudinal assessment of clonal mosaicism in human hematopoiesis via mitochondrial mutation tracking. Blood Adv. American Society of Hematology; 2019;3:4161–5.

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

4. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

5. Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, et al. Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. Cell. 2019;176:1325-1339.e22.

6. 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.