Chronic lymphocytic leukemia includes a tumor subset resembling memory B cells that develop early and persist

Running title: Tumor cell subsets and subclonal diversity in CLL

Bettina Budeus,[1] Maria Dampmann,[1,2] Artur Kibler,[1] Martina Brauser,[1] Kevin Bronischewski,[1] Ekaterina Homp,[1] Sarah Taudien,[1] Patricia Johansson,[1] Johannes Bloehdorn,[3] Stephan Stilgenbauer,[3] Julia von Tresckow,[2] Hans Christian Reinhardt,[2] Ralf Küppers,[1,4] Jan Dürig[2] and Marc Seifert[1,*]

[1]Institute of Cell Biology (Cancer Research), Medical Faculty, University of Duisburg-Essen, Essen, Germany; [2]Department of Hematology and Stem Cell Transplantation, Medical Faculty, University of Duisburg-Essen, Essen, Germany; [3] Department of Internal Medicine III, University of Ulm, Ulm, Germany; [4]German Cancer Consortium (DKTK), Essen, Germany

*Corresponding author:
PD Dr. rer. nat. Marc Seifert
Immunology and Lymphomagenesis
Institute of Cell Biology (Cancer Research)
University Hospital Essen
Virchowstraße 173, 45147 Essen
Germany
Phone: +49 201 7233389
e-mail: marc.seifert@uni-due.de

The authors have declared that no conflict of interest exists.

Abstract

Single-cell RNAseq/VDJseq of tumor cells and normal residual B (NRB) cells from peripheral blood of chronic lymphocytic leukemia (CLL) patients identified three distinct tumor subsets according to phenotype, transcriptome, and immunoglobulin-V-gene (IgV)-mutations. Two major subsets share a typical CLL phenotype but differ in signaling, metabolism and cell cycle control, indicating that the circulating CLL pool is shaped by two states of activity. The third CLL subset shows the phenotype, proliferation capacity and extensive IgV-mutation diversity of normal $CD5^+$ memory B cells. This previously unrecognized CLL tumor subset, which intermingles with NRB cells, was confirmed in 33 IgV-mutated (M-CLL) and IgV-unmutated (U-CLL) cases. Longitudinal IgV-mutation phylogenetics suggest that these NRB-associated CLL cells are generated pathogenetically early, mostly in germinal center reactions, and archive the individual IgV-diversification program, which is conserved throughout CLL course. Our study suggests that diversity is established early in CLL, that each tumor is composed of multiple subclonal expansions, and subclonal evolution can be depicted by IgV-mutation phylogenetics.

Introduction

Chronic lymphocytic leukemia (CLL) is the most common leukemia in adults and to date remains an incurable disease (Hallek, 2019) with highly variable clinical course (Gohil and Wu, 2019; Zenz et al., 2010). The pathogenesis of CLL is unclarified or debated in many aspects. We aimed to investigate the heterogeneity of circulating CLL cells on single-cell level to understand whether functionally distinct tumor cell subsets could be identified throughout the course of the disease, focusing on the analysis of memory B cell features and immunoglobulin (Ig)-V gene mutation patterns in CLL cells.

Somatic CLL mutations can occur already in hematopoietic precursor cells (Marsilio et al., 2018), and transfer of hematopoietic stem cells from CLL patients gives rise to oligoclonal mature B cells with CLL-like phenotype (Damm et al., 2014; Kikushige et al., 2011). The normal counterpart of CLL has been suspected among different B cell subsets, ranging from putative B-1 cells (Caligaris-Cappio et al., 1982; Stall et al., 1988), naïve (Forconi et al., 2010), memory (Klein et al., 2001) to marginal zone B cells (Chiorazzi and Ferrarini, 2003; Chiorazzi and Ferrarini, 2011). Our own study suggested mature $CD5^+$ B cell subsets as the cellular origin of CLL, and the associated pre- or post-GC gene expression patterns are reflected in Ig-unmutated CLL (U-CLL) and IgV-mutated CLL (M-CLL), respectively (Seifert et al., 2012). This is in line with the widely accepted view that U-CLL cases derive from a T cell-independent origin, and M-CLL cases from GC-experienced B cells (Bosch and Dalla-Favera, 2019; Kulis et al., 2012). The analysis of epigenetic reprogramming events during GC maturation showed that the normal counterpart of CLL is distributed along a continuum of normal maturation stages, which correlate with clinical outcome (Oakes et al., 2016). A hallmark of GC maturation is somatic hypermutation (SHM) of IgV genes, mediated through activation-induced cytidine deaminase (AID). IgV gene diversity in CLL (Bagnara et al., 2006; Campbell et al., 2008; Gurrieri et al., 2002; Oscier et al., 1997; Patten et al., 2012; Volkheimer et al., 2007) is potentially affected by ongoing SHM (Huemer et al., 2014). BCR repertoire sequencing and IgV-mutation phylogenetics allow the reconstruction of GC-maturation and -selection processes (Kibler et al., 2021; Seifert and Küppers, 2009), also in CLL (Campbell et al., 2008; Gurrieri et al., 2002; Lanasa et al., 2010; Oscier et al., 1997; Patten et al., 2012; Volkheimer et al., 2007).

CLL is preceded by the asymptomatic precursor state monoclonal B cell lymphocytosis (MBL) (Marti et al., 2007). Some MBL populations are oligoclonal and few cases show evidence of intraclonal IgV-diversity (Lanasa et al., 2010). This argues for a stepwise development of MBL into CLL, and it was proposed that CLL progression appears to occur either through ongoing clonal diversification (e.g., by antigen selection) or an expansion of a clonally heterogeneous population (Rawstron et al., 2007).

CLL expansion is dependent on the (re-)circulation of tumor cells between tissue and peripheral blood (PB) (Stevenson et al., 2014). Tissue homing (Burger et al., 1999), microenvironmental interactions (Aguilar-Hernandez et al., 2016; Endo et al., 2007; Lu et al., 2004; Lutzny et al., 2013; Yu

et al., 2016) and autoantigen exposure in proliferation centers (Herishanu et al., 2011) drive tumor growth, and transiently acquired anergy then promotes tissue egress and CLL dissemination. Two subsets of CLL cells have been described that are closely associated with these processes: recently divided tissue emigrants ($IgM^{low}CXCR4^{low}CD5^{high}$) as well as quiescent cells, which are prepared for tissue reentry, reside in the PB of CLL patients (Calissano et al., 2011; Kipps et al., 2017).

CLL is characterized by intraclonal genomic heterogeneity, and there are indications that this is partly established early in pathogenesis (Baliakas et al., 2019; Gaiti et al., 2019; Lanasa et al., 2010). Genome-wide high-throughput analyses of genetic lesions (Gohil and Wu, 2019; Landau et al., 2013; Landau et al., 2015) and analysis of mitochondrial mutations (Penter et al., 2021) showed that CLL progress is associated with clonal evolution, and disease outcome is often shaped by genetic lesions that were already present in earlier, slow-growing stages (Gruber et al., 2019).

The present study is based on the hypothesis that mature $CD5^+$ B cells, potentially predisposed by initial transforming events or expression of stereotyped CLL BCRs, undergo GC-diversification and selection. We further assume that this early phase, which occurs before manifestation of a major MBL/CLL population, may be a critical determinant of CLL plasticity and evolution. We present a single-cell level analysis of normal residual B (NRB) cells and tumor cells in U-CLL and M-CLL patients by combined single-cell RNAseq and single-cell VDJseq. We identify three distinct CLL tumor cell subsets, which differ substantially in gene expression and proliferation capacity. IgV-mutation phylogenetics of the BCR genes revealed that the three subsets can be distinguished according to their IgV-mutation patterns. One novel subset, which phenotypically intermingles with the NRB cells, is generated early in pathogenesis, and persists with characteristics of malignant $CD5^+$ memory B cells, forming an archive of the highly individual IgV-diversification pattern in every M-CLL and U-CLL tumor.

The longitudinal BCR repertoire deep sequencing of 12 CLL cases showed that CLL pathogenesis is not a linear process, but from early on multiple IGHV diversified tumor subclones are generated in parallel. The intraclonal IgV-diversity in CLL is reduced by treatment, but fully and conservatively reconstructed upon progression and relapse. Thus, IgV-diversity in CLL is suitable to determine subclonal dynamics in CLL.

Results

Identification of distinct CLL subsets on single cell level. Flow cytometric analysis of the PB from 60 CLL patients (Table 1) confirmed the presence of a small population of NRB cells ($CD20^{high}CD5^{-/+}CD23^{-/+}CD27^{-/+}$, range 0.01 - 4.7%, average 0.1% of total $CD20^{+}CD3^{-}$ B cells, Figure S1A and B) and a major CLL population ($CD20^{low}CD5^{high}CD23^{high}CD27^{high}$) (Criado et al., 2018). We investigated the intraclonal IgV gene diversity and cellular composition of CLL tumor cells and the associated NRB cell compartment by single-cell RNA-sequencing and VDJ-sequencing (scRNAseq/scVDJseq) from four CLL cases (CLL 01-04, Table 2). We isolated up to 100,000 NRB cells and an equal amount of CLL cells from CLL cases 01 and 02 by fluorescence activated cell sorting (Figure 1A, B and S1A, C) and processed each aliquot separately in 10X Chromium assays. The cells from the NRB gate were highly enriched and received sequencing-indices different from the cells in the CLL gate. Between 700 and 3,500 cell equivalents with full VDJ-rearrangement information (scVDJseq) were retrieved after scRNAseq quality control (exclusion of doublets, damaged and dying cells) per aliquot.

Uniform Manifold Approximation and Projection (UMAP) clustering of the scRNAseq/scVDJseq datasets visualized NRB and CLL cell heterogeneity. In both patients, scVDJseq identified cells with CLL IgV-rearrangement among all processed samples. Among the cells from the CLL-gating strategy, two major CLL subpopulations were detected (referred to as “CLL cluster 1 and 2”, Figure 1 A and B, top right). Among the cells from the NRB-gating strategy, UMAP separated polyclonal NRB cells from a variable number of cells with CLL-*IGH*- and/or *IGL*-rearrangement (Figure 1A and B, lower left). The integrated analysis of all cells with CLL rearrangement revealed that only a small fraction of the NRB-derived CLL cells colocalizes with CLL cluster 1 or 2 (“contaminating CLL cells”). Instead, the majority of the NRB-derived CLL cells represent a distinct tumor cell population (Figure 1A and B, lower right); in one patient potentially more than one NRB-associated CLL subset was identified (Figure 1A, lower left). To validate the existence of distinct CLL subsets we integrated additional scRNAseq/scVDJseq profiles from CLL cases 03 and 04. From both donors the NRB cells were enriched by cell sorting (Figure S1C), processed together with a similar amount of corresponding CLL cells, and all cells belonging to the tumor clone identified by expression of the respective CLL rearrangement (Figure 1C, D and Table 3). The integrated analysis of the four samples confirmed that in four different patients the same three CLL subsets could be identified based on common expression signatures (Figure 1C, D, and Table 3). The integrated analysis confirmed an overall higher similarity of CLL clusters 1 and 2, forming the major CLL population, and one highly dissimilar NRB-associated CLL subset (Figure 1C and D).

Our single-cell study revealed that circulating CLL cells are not a homogenous population, but can be distinguished into three transcriptionally distinct subsets, including one previously unrecognized subset that intermingles with the normal B cell population.

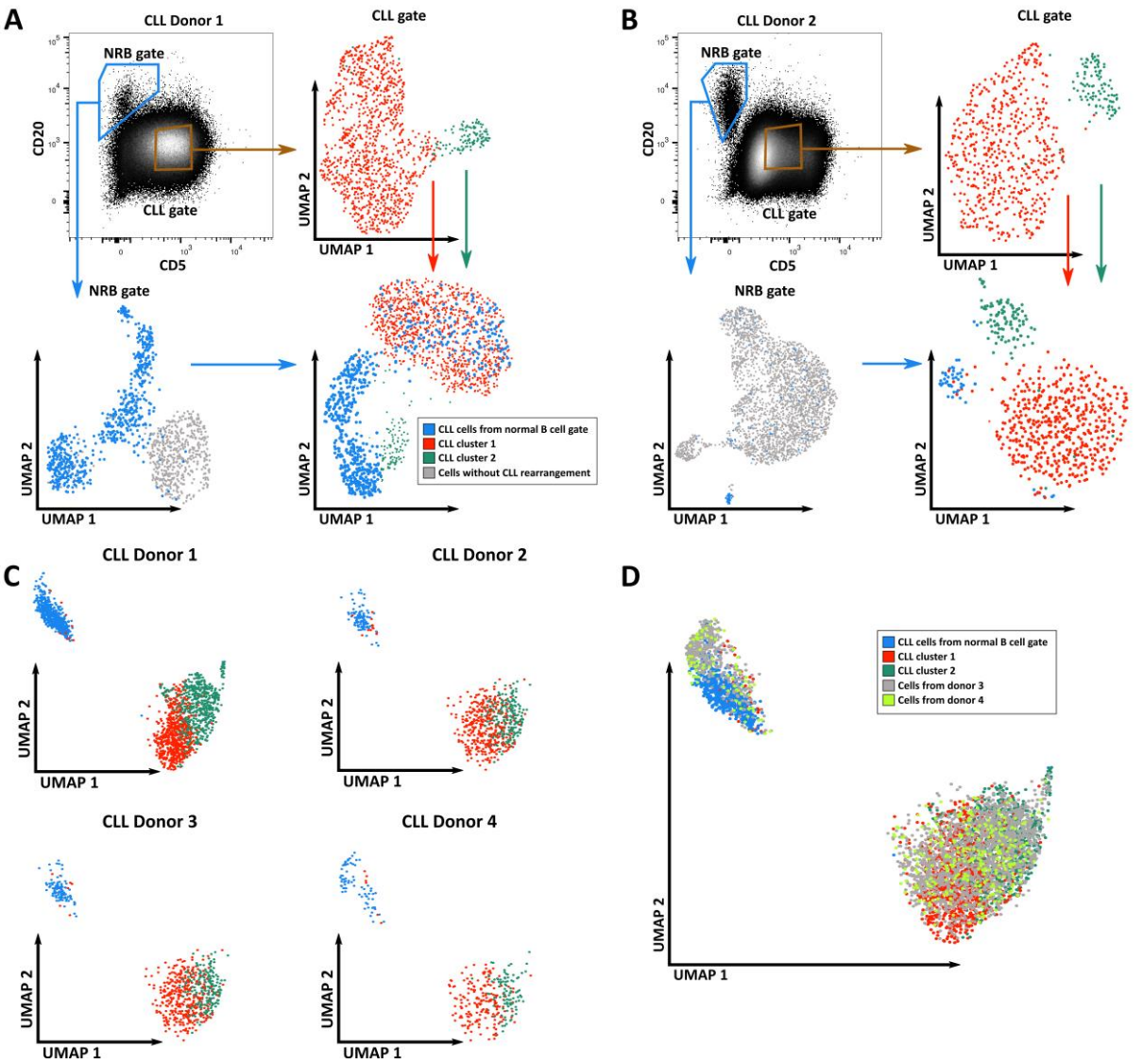

Figure 1: scRNAseq/scVDJseq analysis of PB samples from four CLL cases. From CLL patient 01 (A) and CLL patient 02 (B) the gating strategy for CLL cells (main CLL population, orange) and NRB cells (blue) is shown in the top left plot. Top right: UMAP of single cells from the CLL gate (orange). CLL cluster 1 (red) and CLL cluster 2 (green) were identified among cells expressing the known CLL-IgH/IgL rearrangement and clustering (Seurat). Lower left: UMAP of cells from the NRB gate (blue). Cells expressing the known CLL-IgH/IgL rearrangement ("NRB-associated CLL cells") are shown in blue, polyclonal NRB cells are shown in grey. Lower right: integrated UMAP from all cells expressing the known CLL-IgH/IgL rearrangement. Coloring according to the top right/lower left plot. Blue events colocalizing with CLL clusters 1 or 2 (red/green) likely represent cell-sorting contaminations. C) and D): Integrated analysis of scRNAseq/scVDJseq data from tumor cells of four CLL cases (patients 01-04). CLL cells were identified by expression of the respective CLL-IgH/IgL rearrangement among the NRB gate and the main CLL population (CLL cluster 1 and 2) and UMAP and clustering was performed. Individual cases are shown in C), all four samples are shown in D). Coloring of CLL_01 and CLL_02 is according to the indexed NRB-gate and CLL-gate analysis, CLL subsets of cases 03 and 04 were determined by unsupervised clustering (see Figure S3B and C) and colored according to gene expression signatures of donors 01 and 02 (see Figure 2A).

Characterization of distinct CLL subsets by single cell transcriptomics. The scRNAseq gene expression profiling identified a total number of 9,673 genes, of which 9,464 were repeatedly identified among the CLL cells from all four patients. A number of 1,329 genes (14%) was found differentially expressed between the three major CLL clusters (Figure 2A). According to Gene Ontology pathway analysis, CLL cluster 1 represents a resting, but metabolically active tumor subset (Figure 2B and S2A), whereas CLL cluster 2 is characterized by transcriptional and translational activity and a weak signature of proliferative activity (Figure S2B). CLL clusters 1 and 2 can be distinguished according to the gene sets defining recently divided and older/quiescent cells (Figure S2C), which is also reflected by their differential expression of CXCR4 and CD5 (Figure 2C). In Figure 2C we provide normalized expression plots of selected genes associated with CLL phenotype and pathogenesis. Whereas CLL clusters 1 and 2 show differential expression of *JUNB*, *FCER2*, *MYC* and *TGFB1*, the NRB-associated CLL subset shows highest expression of *BCL2*, *ROR1*, *CDK14* and *MCM6*, amongst others (Figure 2C). According to surface marker transcript levels, NRB-associated CLL cells are $CD20^{high}CD23^{low/-}CD5^{+}CD27^{+/-}$ (Figure 2C), more reminiscent of mature $CD5^{+}$ B cells (naïve or memory) than typical CLL cells. No surface marker combination proved to be exclusively expressed on NRB-associated CLL cells, hampering specific isolation. GO pathway analysis revealed that NRB-associated CLL cells are characterized by telomere maintenance, cell cycle checkpoints, DNA integrity and a large number of nuclear and intracellular vesicle transport associated genes (Figure 2D and S2D-G).

According to SingleR projection analysis, NRB-associated CLL cells show expression signatures of typical CLL tumor cells (Figure 2E), either more similar to M-CLL or U-CLL signatures, reflecting the Ig mutation status (Figure S2H). Compared to normal B cell subsets alone, they predominantly show a $CD5^{+}$ memory B cells signature (Figure 2E).

We conclude that our scRNAseq analysis of four CLL cases (2 M-CLL / 2 U-CLL) consistently identified three distinct CLL subsets. These subsets differ in their relative expression of pathways associated with cell metabolism, signaling, and protein transport/turnover. Central features enriched among the NRB-associated subset are DNA-replication, -maintenance, and cell cycle control.

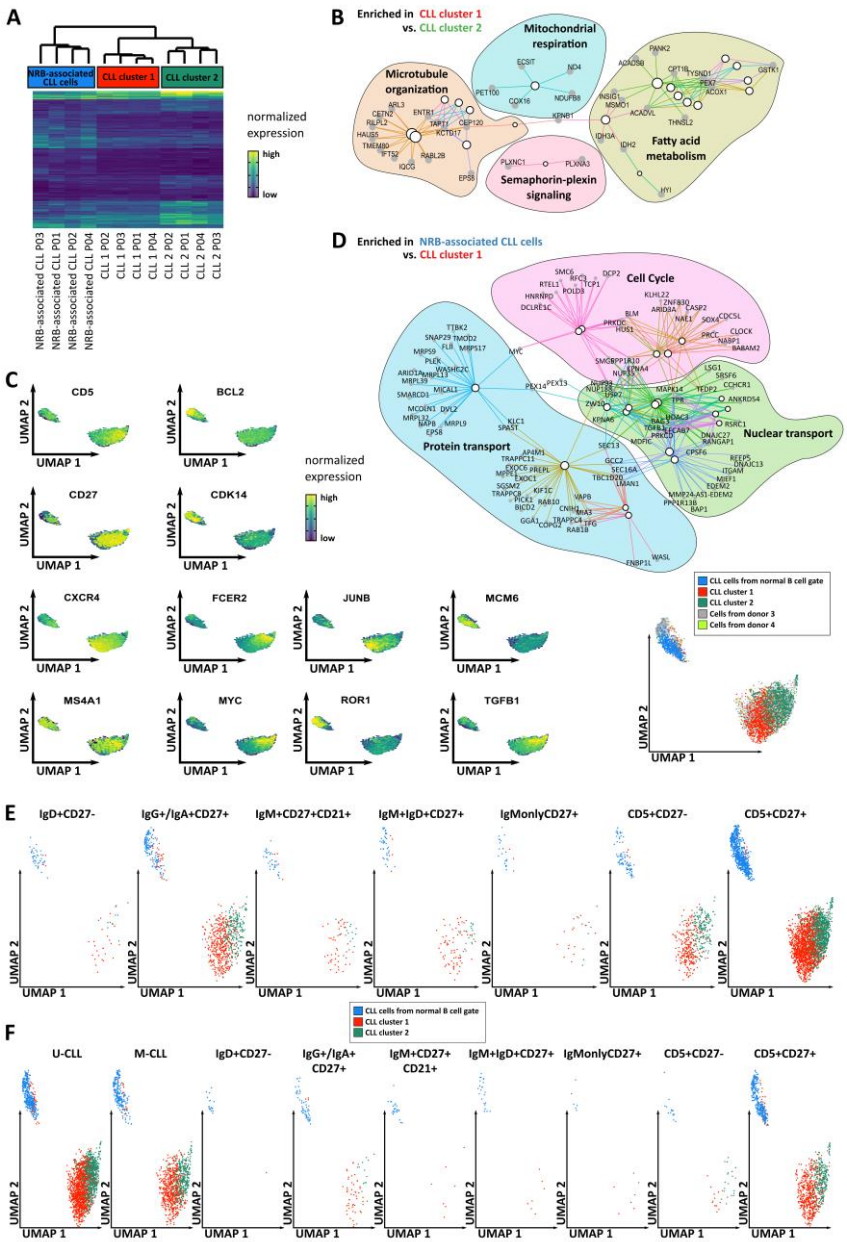

Figure 2: scRNAseq profiling of CLL subsets. A) Heatmap of the statistically significant differentially expressed genes between the three CLL clusters (n=1,329), repeatedly identified in all four patients. B) Gene Ontology pathway analysis of the genes enriched in CLL cluster 1 compared to CLL cluster 2, repeatedly identified in all four patients. C) Normalized expression plots of selected genes associated with CLL surface expression and pathogenesis among the integrated UMAP and clustering analysis of all four patients (UMAP from Figure 1D). D) Gene Ontology pathway analysis of the genes enriched in NRB-associated CLL cells compared to CLL cluster 1. E) All cells with CLL-IgV-rearrangement (black, upper left plot) were integrated from donors 01-04, and their global similarity to normal B cell subset signatures or to U-CLL/M-CLL signatures published in Seifert et al., 2012 was determined by reference based single-cell RNAseq annotation (SingleR). For each cell only one (the predominant) assigned cell type is allowed and the number of cells with that signature shown in separate plots. Top row: only normal B cell signatures were tested, bottom row: M-CLL- and U-CLL-signatures are allowed (See also Figure S2H).

Functional analysis of the proliferation potential of CLL subsets. The scRNAseq gene expression profiling suggested that NRB-associated CLL cells in particular, and to a lower extent also CLL cluster 2 cells, were enriched for gene sets associated with mitosis, cell cycle control, DNA repair and telomere maintenance. We hence aimed to test whether actively proliferating cells could be identified among the subsets, or whether they differed in their proliferation capacities. Cells from the NRB compartment and the main CLL population (Figure S1A) were isolated, marked with cell tracking dye and stimulated *in vitro* (CpG and CD40L). After 5 days the proliferating fraction was sort-purified (Figure 3A) and the fraction of proliferating CLL cells was determined by IGHV gene PCR, cloning and Sanger sequencing, using the CLL-IGHV-rearrangement as a specific marker for tumor cells. Unstimulated NRB cells or CLL cells did not proliferate, cells of the main CLL population died under unstimulated conditions, suggesting that neither population included actively proliferating cells. Stimulated cells from the main CLL population (> 10^6 cells of CLL cluster 1 and 2) barely showed proliferating cells in 9 patients analyzed (Figure 3A). In contrast, stimulation induced proliferation among 30,000 to 50,000 isolated NRB cells in 9/9 patients (Figure 3B). From 6/9 patients we could isolate between 2,000 and 10,000 proliferating cells and obtain sufficient material for IGHV gene PCR and cloning. Among 4 samples, a substantial amount of CLL IGHV rearrangements (55/147 total, range 5 to 19 per patient) was detectable (Figure 3C), in addition to polyclonal, mostly mutated IGHV rearrangements of non-CLL origin. Due to the lower NRB-cell abundance among U-CLL patients, this cohort includes only one informative U-CLL case.

We conclude that NRB-associated CLL cells likely do not actively divide, but can respond to stimulation *in vitro*, indicating that this CLL subset shows a higher proliferation propensity than the main CLL population.

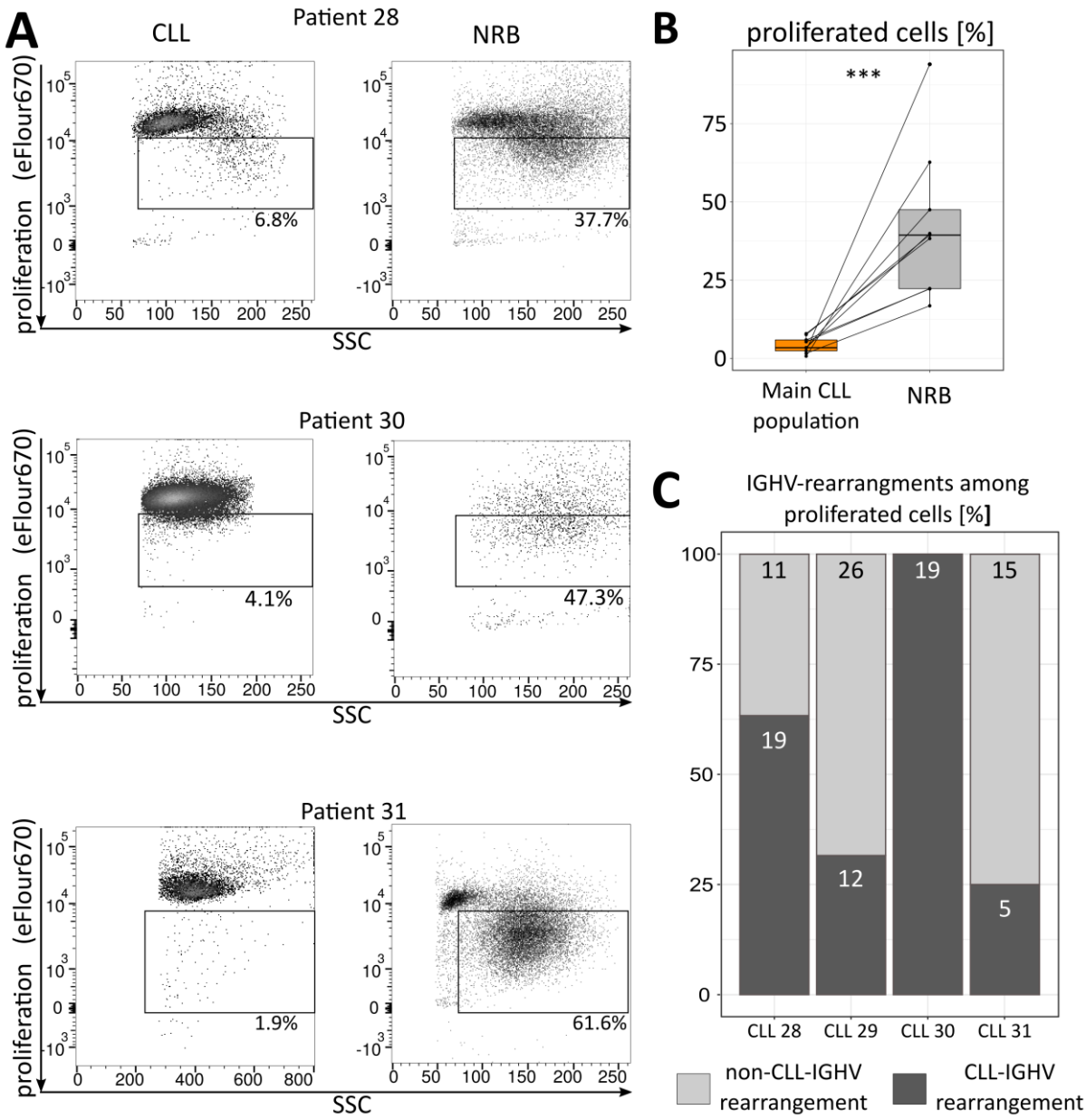

Figure 3: Proliferation potential of CLL subsets. A) Exemplary FACS-plots of proliferating CLL or NRB cells from stimulated samples from three patients (CLL 28, 30, 31). Proliferated cells are identified by dilution of the cell tracking dye eFlour670 and isolated by cell-sorting (black gates). **B)** Percentage of proliferated cells among the main CLL population and NRB cells ($p < 0.01$, Wilcoxon test, n=9). **C)** Percentage of the CLL-IGHV-rearrangement (dark grey) and non-CLL-rearrangements (light grey) among the total number of examined proliferated cells. Numbers give the amount of CLL- or non-CLL-IGHV-rearrangements identified per sample (n=4).

VDJ sequence analysis of CLL subsets. The scRNAseq/scVDJseq profiling identified sequence variants of the CLL-IgV-rearrangements in both M-CLL and both U-CLL cases. For a reliable comparison of intraclonal diversity, equal numbers of sequences were randomly drawn from the two CLL clusters and the NRB cell-associated CLL clone members. Among all four patients, the IgV-sequences from CLL cluster 2 (proliferation signature) barely showed intraclonal diversity and the vast majority formed one large node ("center of mass", Figure 4A-B). The IgV-sequences from CLL cluster 1 (resting) showed an intermediate level of intraclonal diversity with multiple branches, but most (80% - 97%) of the sequences located at the center of mass (Figure 4C). The NRB cell-associated CLL subset showed highest intraclonal diversity, i.e., formed the most branched dendrograms, and fewer sequences locating to the center of mass (Figure 4D and S3A-C).

To confirm the higher intraclonal diversity among NRB-associated CLL cells, and to exclude that this is biased by technical artifacts, we performed IGHV deep sequencing using unique molecular identifiers (UMIs). From four CLL cases (CLL 05-08) aliquots of 50,000 NRB- and CLL-gated cells (together "main CLL population") were separated, genomic DNA isolated, and bulk VDJ sequencing performed. From each NRB- or CLL-aliquot between 20,000 and 40,000 sequences (after UMI collapsing) encoding the known CLL-IGHV-rearrangement were obtained. The phylogenetic analyses of IGHV-mutation patterns confirmed the high intraclonal diversity among the NRB-associated CLL cells, always exceeding that of the corresponding main CLL population (Figure 4F-I). In 3/4 patients the IGHV deep sequencing analysis suggested that the NRB-associated CLL cells formed a diverse "framework" dendrogram, of which the IGHV-variants from the main CLL population represented a subsection. The NRB-associated CLL cells are mostly "unique", with usually very few sequences per node, whereas the main CLL population in all cases included at least one, sometimes several subclonal expansions, i.e., single large nodes including thousands of identical sequences (but distinct UMIs).

In line with their $CD20^{high}CD5^{+}CD27^{+}$ isolation strategy, the NRB-associated CLL cells from the four M-CLL cases analyzed here show the extensive IGHV mutation diversity seen among GC-experienced memory B cell clones (Kibler et al., 2021; Seifert and Küppers, 2009; Seifert et al., 2015). The main CLL-population represents a selection of these IGHV-variants, including few major subclonal expansions.

So far, our analyses were focused on M-CLL cases (6/8) and restricted to only a few ten thousand CLL cells. To obtain sufficient data for a statistical evaluation of dendrogram parameters among M-CLL and U-CLL cases, we analyzed 21 additional CLL patients (12 M-CLL and 9 U-CLL) with diverse clinical staging (Table 2) and isolated on average 300,000 (range 13,000-600,000) cells of the main CLL population and on average 15,000 (range 900-65,000) NRB cells ($CD20^{high}CD5^{low}$) and performed IGHV deep sequencing (InvivoScribe). Intraclonal IGHV-diversity was detected in all 21 cases and was always increased among the NRB-associated CLL subset (Figure S3D and E). Based on

the relative frequencies of CLL- and non-CLL rearrangements, we calculated that the NRB-associated CLL subset on average represents 25% (range 0.13 to 99%) of the NRB cell pool, ranging from 100 to 20,000 sequences in our analysis (for detailed information on clinical status, cell frequencies, purified cell numbers, sequencing depth and retrieved sequences see Table 2).

The analysis of phylogenetic parameters (intraclonal IGHV gene diversity, number of branching nodes, number of consecutive mutations per branch etc., Figure 4J) revealed that the IgV-mutation dendrograms among NRB-associated CLL subset were significantly more complex than those from the main CLL population. In other words: each dendrogram includes white nodes (including NRB-CLL and main CLL population sequences) and blue nodes (only NRB-CLL sequences), but orange nodes (only main CLL population sequences) are rare in most cases (Figure 4K and S3D and E).

This finding is independent of M-CLL or U-CLL subtype: the IgV-mutation dendrograms from 9/9 U-CLL patients consistently showed structures typical for memory B cells (Figures S3D). The IgV-mutation patterns were not random, but consecutive, significantly enriched in AID-hotspot-motifs ($p < 0.01$, Wilcoxon signed rank test) and nonsense mutations were counter-selected ($p < 0.001$, Wilcoxon signed rank test). The average mutation frequency of the 5% IGHV sequences with highest mutation load does not differ significantly between M-CLL and U-CLL (Figure 4L). The major phylogenetic difference between M-CLL vs. U-CLL is the location of the “center of mass”: In U-CLL cases the tumor mass is located in close vicinity to the root (IGHV germline sequence), whereas in M-CLL the tumor mass is distributed along the dendrogram (Figure 4M and S3D and E).

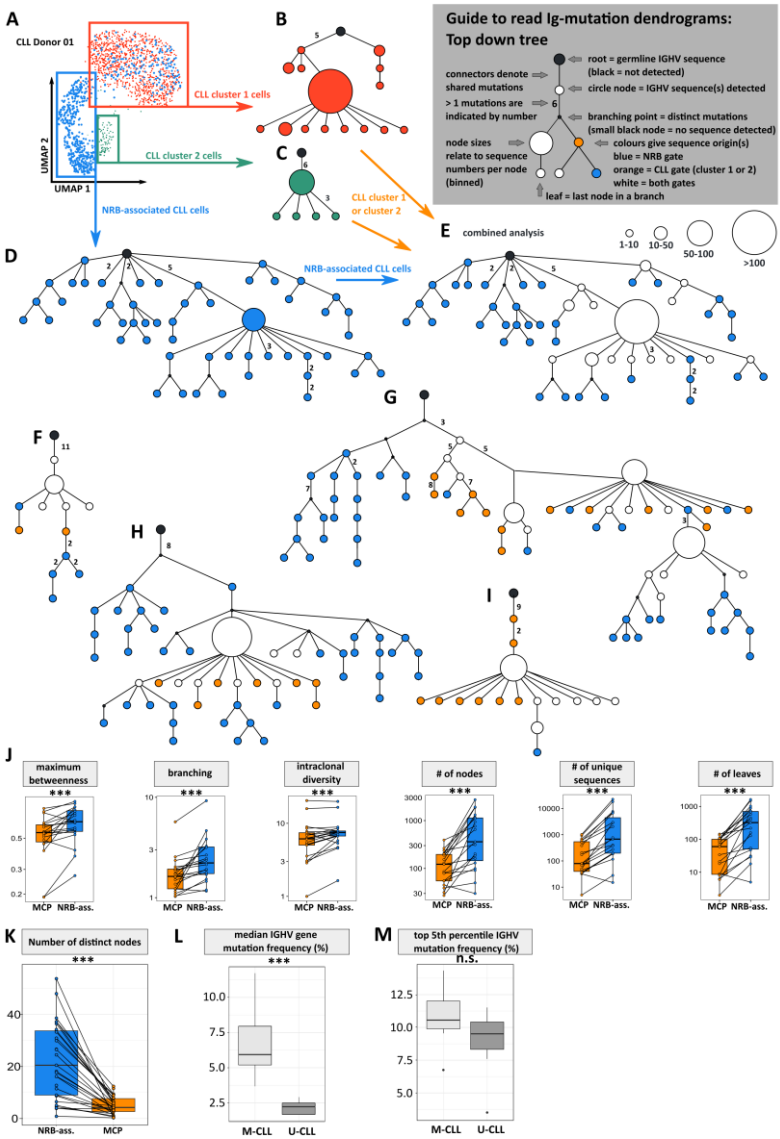

Figure 4: Phylogenetic analysis of IGHV-mutation patterns among CLL subsets. The grey box in the upper right describes key elements of Ig-mutation top-down dendrograms. A) UMAP and clustering of CLL cells from patient 01 (Figure 1A, lower right) is shown to visualize subset selection. **B)** IgV-mutation dendrogram of tumor cells from CLL cluster 1 (red cells in A) and **C)** cells from CLL cluster 2 (green cells in A) and **D)** NRB-associated CLL cells (blue cells in A) from the scRNAseq/scVDJseq of CLL donor 01 are shown. **E)** Combined dendrogram of all tumor cells from CLL patient 01 shown in B)-C). Note that CLL cluster 2 cells, which are >>99% located at one single node are pooled with CLL cluster 1 and colored in orange (or white when NRB-associated CLL sequences locate to the same node). **F) – I)** Top-down dendrograms of four additional CLL patients analyzed by VDJ-deep-sequencing and UMI-clustering. **J)** IGHV-mutation dendrogram parameters from 21 CLL cases analyzed with commercially available bulk VDJ-sequencing strategy in replicates. MCP: main CLL population, NRB-ass.: NRB-associated CLL cells. **K)** Number of nodes containing only sequences from NRB-associated CLL cells (blue), versus those containing only sequences from the main CLL population (orange) per dendrogram (n=21). **L)** Median IGHV gene identity among M-CLL and U-CLL cases (n=21). **M)** IGHV gene identity of the top 5th percentile in M-CLL and U-CLL cases (= the highest IGHV gene identity value of the 5% of the sequences with highest mutation load of each clone start, n=21). (***$p < 0.001$, n.s. not significant, Wilcoxon test).

Longitudinal analysis of IGHV-diversity among CLL subsets. We performed a longitudinal IGHV deep sequencing analysis from 12 CLL cases with consecutive life frozen PB samples spanning up to 16 years. From 7 of these patients, NRB-associated and main CLL population cells were analyzed separately across several timepoints, the remaining 5 samples only at one timepoint due to limited cell numbers. Based on this analysis we could distinguish characteristic signatures in the IgV-mutation dendrograms that relate to the course of the disease.

CLL cases with stable course showed stable dendrogram structures and the relative size of the nodes barely changed (Figures 5A and S4A). Slowly progressing cases showed an increasing number of nodes and additional branches among the NRB-associated and main CLL population cells, suggesting that during disease progression IGHV-diversity increases. However, some of the "newly acquired" IGHV-variants among the main CLL population were already present years before among the NRB-associated CLL cells (Figures 5B and S4B).

The occurrence of IGHV-variants among the main CLL population, which were present earlier in NRB-associated CLL cells is also detectable among rapidly progressing/relapsing cases (Figures 5C and S4C and D). Each cycle of treatment (Table 4) and relapse in the rapidly progressing cases always affects the dendrogram shapes from both, NRB-associated and main CLL population cells. However, the structure of the NRB-associated CLL dendrograms remains more stable, whereas the main CLL population dendrograms change, in particular in their distribution of the tumor mass (compare "the location of large nodes", e.g., Figure 5C timepoints 2014 and 2018a and 2018b). Importantly, at all timepoints, the NRB-associated CLL cell dendrograms are more diverse and always encompass the less diverse, but highly expanded main CLL population (Figure 4). In total, even in cases with long remission (Figures 5D and S4E-H) barely additional "new" IGHV-variants arise, but the arrangement and sequence of branches, nodes, and leaves is highly conserved.

Taken together, our data indicate that subclonal IgV-diversity is generated early in CLL pathogenesis, presumably in a GC-dependent fashion. This IgV-diversity is conserved, as can be deduced from a small subset of NRB-associated CLL cells, which form a kind of "archive" of the IgV-diversification program. This subclonal IgV-diversity can be used to monitor subclonal composition and dynamics in CLL by phylogenetic evaluation.

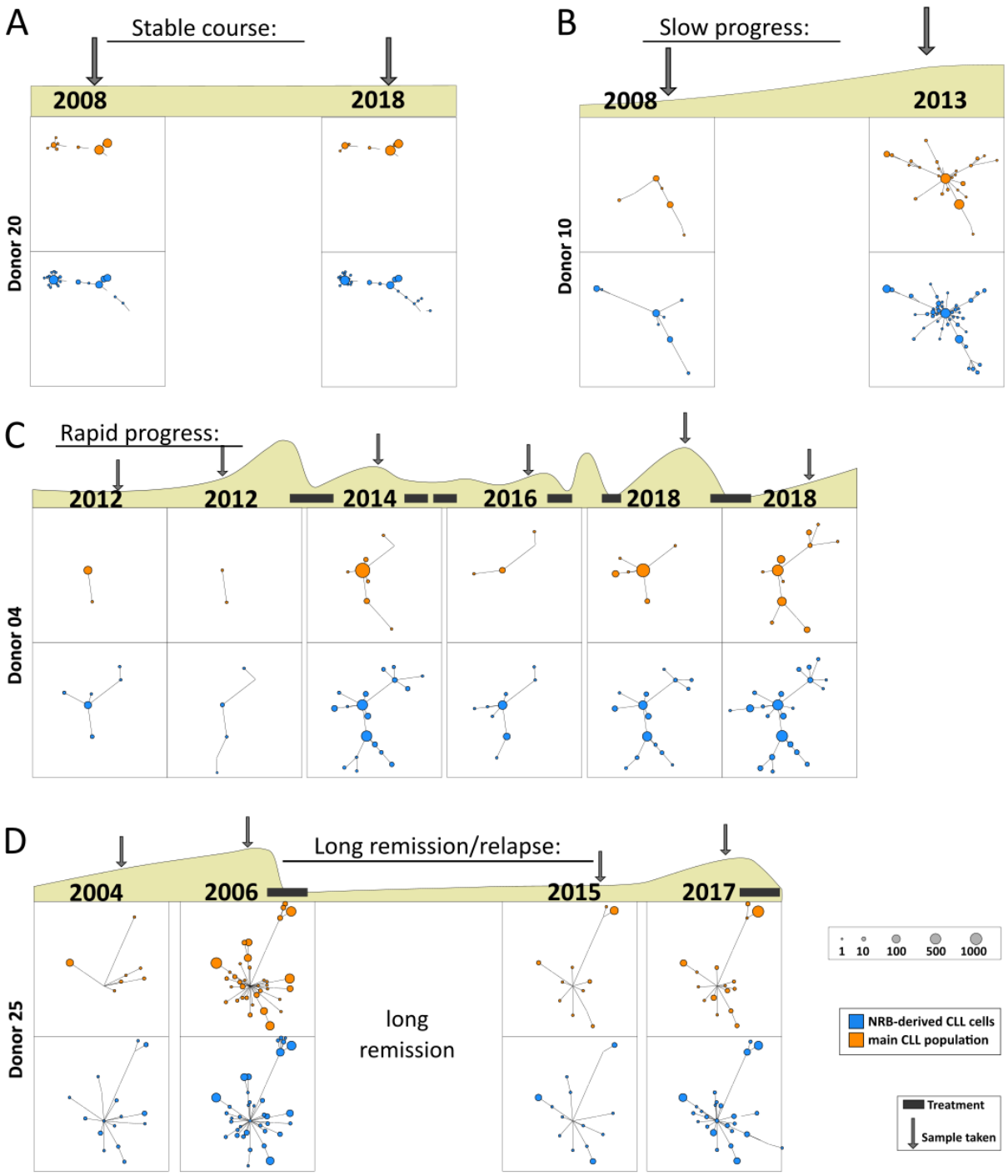

Figure 5: Longitudinal analysis of IGHV-mutation phylogenetics among CLL subsets. Representative dendrogram structures from sequential samples from four patients are shown. Lymphocyte counts are shown as area plots for each patient. Treatment phases are marked as black bars. For each IGHV deep sequencing timepoint (marked by grey arrow) the number of sequences displayed from the main CLL population is identical to the number of sequences displayed from NRB-associated CLL cells. The node size (exponentially displayed, see scale legend) reflects the number of distinct, non-collapsed sequences per node. For treatment information, see Table 4. **A)** Stable course, **B)** Slow progress, **C)** Rapid progress/relapse, and **D)** Relapse after long remission.

Analysis of chromosomal aberrations among CLL subsets. The IGHV-mutation phylogenetics strongly suggested a pathogenetically early diversification and conservation of IgV-subclones in CLL, and that the generation of NRB-associated CLL cells is closely connected to this early phase of IgV-diversification. To support this hypothesis, we used the frequency of expressed genes from neighbored gene loci to calculate *de novo* copy number variation (CNV) in our four scRNAseq profiles with the R-package CONICS for each of the three subsets separately. This analysis was able to predict the known 17p deletion (diagnostic FISH analysis) among all tumor cells from patient 03, according to significantly reduced expression signatures of 17p gene loci (Figure 6 A). In contrast, we observed a significant discrepancy of expressed gene loci from chromosome 12 in patient 02, known to carry trisomy 12 from diagnostic FISH analysis. Notably, main CLL population cells showed a gain in expression of 12q gene loci, but this was clearly absent from the majority of the NRB-associated CLL subset, indicating that most of these CLL cells evolved at a timepoint before the outgrowth of this subclonal driver mutation in this donor.

Taken together, as both chromosomal aberrations were found to be largely monoclonal, this analysis suggests that the NRB-associated CLL cells in patient 03 were generated after, but those from patient 02 before the timepoint where the mutations occurred.

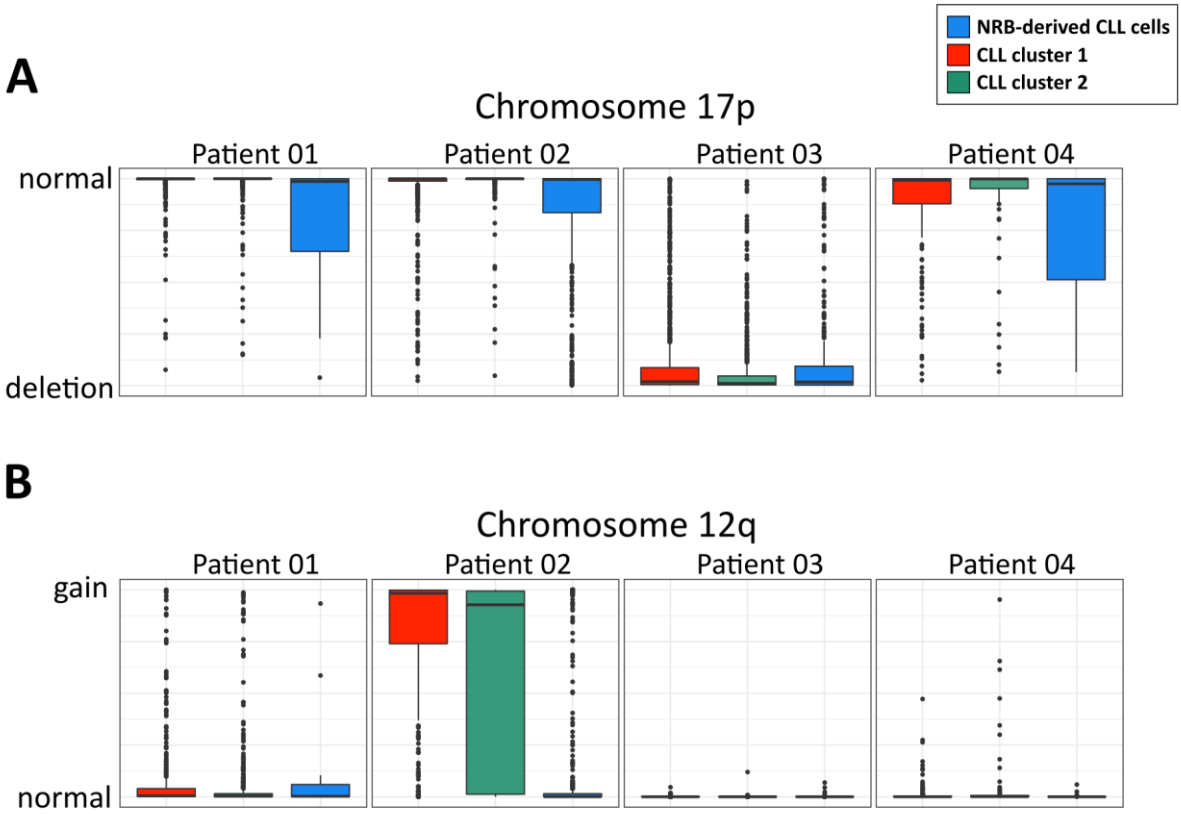

Figure 6: CNV analysis among CLL subsets according to scRNAseq data from four CLL patients. The estimated copy numbers of selected gene loci located on chromosome 17p (A) and chromosome 12q (B) among the total number of CLL cells from each subset in four patients, as determined by the R-package CONICS.

Discussion

Our single cell analysis identified three tumor subsets in the PB of CLL patients, sharing the CLL heavy and light chain IgV-rearrangements, and expressing typical CLL-related transcription patterns. The three CLL subsets differ in the expression of several hundred genes and the relative expression of CD20, CD5, CD23, CXCR4 and CD27. According to these differences, and according to the cell isolation strategy, two of the CLL subsets (termed "CLL cluster 1 and 2") are typical CLL tumor cells, with $CD20^{low}$ and $CD19^{+}CD5^{+}CD23^{+}$ expression (Zenz et al., 2010). CLL cluster 1 and 2 differ in the inverse expression pattern of CXCR4 and CD5, which has been associated with a recently proliferated and a resting, anergic state of CLL cells during tissue homeostasis (Calissano et al., 2011). Our scRNAseq analysis validated the existence of such functional states in CLL, and moreover, shows that they exist as distinct subsets, and that the underlying gene expression patterns can be assigned to unique transcriptional activities, mainly including DNA replication and cell cycle checkpoint in cluster 2 (proliferation) and enhanced metabolism and microtubule organization in cluster 1 (resting) (Figures 2 and S2).

We identified a third CLL subset by enrichment and scRNAseq/scVDJseq analysis of the *in vivo* underrepresented NRB cells. This "NRB-associated CLL cell" subset has not been described previously in single-cell studies (Gaiti et al., 2019; Pastore et al., 2019; Wang et al., 2017; Zhao et al., 2016), and we focused our study on this novel tumor cell subset.

NRB-associated CLL cells differ from the asymptomatic precursor state MBL by their lower frequency in PB, lack of the typical CLL-like immunophenotype and higher IgV-diversity (Lanasa et al., 2010; Shanafelt et al., 2010; Vardi et al., 2013). The *in vitro* proliferative capacity further supports that NRB-associated CLL cells deviate from the mostly unresponsive, and apoptosis-prone behavior of bulk MBL/CLL cells, unless these are cultured with feeder cells (Pascutti et al., 2013; Schleiss et al., 2019).

The most intriguing feature of NRB-associated CLL cells is their high IgV-diversity, which is detectable in both, M-CLL and U-CLL cases. Although the NRB-associated CLL subset represents only a minimal fraction of the entire tumor (the isolation strategy enriched the NRB cells up to 10,000-fold in comparison to the main CLL population), their IgV-diversity is higher than that of the main CLL population. Phylogenetic analyses showed that this IgV-diversity is in numbers and dendrogram structure similar to human memory B cell clones (Kibler et al., 2021; Seifert and Küppers, 2009). Memory B cells can principally also be generated in GC-independent (Kaji et al., 2012) or in T cell-independent responses (Defrance et al., 2011), but then do not show the high IgV-mutation load that is associated with GC-dependent responses. Also in humans there is evidence for IgV-diversification outside of GCs (Weill et al., 2009), however, this is most prominent in young children, and the IgV-mutation load and the frequency of mutated cells are considerably lower than in typical GC-experienced memory B cells (Seifert and Küppers, 2009). The NRB-associated CLL cells are highly IgV-diversified, showing multiple consecutive mutations (long dendrogram branches). Such patterns

require extensive proliferation and selection, but also extensive cell death (counterselection of unfavorable mutations), hallmarks of normal GC-reactions.

Still, our study is inconclusive about a general role of GCs in CLL pathogenesis. In particular some U-CLL cases in our study showed IgV-mutation dendrograms showing an "undirected" mutation pattern compatible with largely random diversification (e.g., cases 13, 14, 23, Figure S3). Moreover, even highly diversified cases sometimes included unmutated clone members, which may have bypassed the GC reaction.

Further features of NRB-associated CLL cells support their similarity to memory B cells: the enrichment of cell-cycle pathways, telomere- and DNA-maintenance point to a high proliferation capacity. We could not validate active proliferation among the cells (at least not in our limited experimental setting), but observed proliferation upon *in vitro* stimulation among both, non-tumor cells and NRB-associated CLL cells. We purified the proliferating cell fraction and found multiple polyclonal IgV-rearrangements deriving from non-tumor cells. As the polyclonal IgV-sequences were all mutated, we assume that naïve B cell proliferation was below detection limit, as can be expected from the short stimulation period (Tangye et al., 2003). Therefore, the proliferating NRB-associated CLL cells that were detectable in 4/6 cases rather followed the quick response dynamics related to normal memory B cells (Good et al., 2009).

According to unbiased cell type recognition (Figure 2E and F), all NRB-associated CLL cells carry traits of either M-CLL- or U-CLL-derived expression signatures (Figure 2F). Within the same settings, also a similarity to $CD5^+$ memory B cells (but not other normal PB B cells) is observed, which is significantly enhanced when the CLL-profiles are excluded from the recognition algorithm (Figure 2E). Flow cytometric information on the surface phenotype is so far mostly available from our sorting strategy, which only confirmed $CD20^{high}$ expression. However, from four M-CLL cases, we specifically isolated $CD20^{high}CD5^+CD27^+$ B cells and repeatedly identified the highly Ig-diversified NRB-associated CLL cells among this normal $CD5^+$ memory B cell population (Figure 4F-I). We cannot exclude that the novel CLL subset is also present among other NRB subsets, but the combined evidence supports that a major fraction of NRB-associated CLL cells shares characteristics of $CD5^+$ memory B cells.

IgV-mutation phylogenetics can also be used to analyze clonal dynamics between sequential/related samples (Kibler et al., 2021; Meng et al., 2017). To this end, we performed IgV-mutation phylogenetics of longitudinal samples from 12 CLL cases, always separating NRB cells from the main tumor population. This analysis showed that throughout the disease course, the IgV-mutation patterns can change, and sometimes new IgV-variants develop. This suggests that in an established CLL tumor IgV-diversity can increase, presumably in CLL proliferation centers in lymph nodes or bone marrow. However, new IgV-variants are rare, and we consistently observed that the majority of the IgV-variants that exist at one timepoint, often covered by millions of sequences among the main CLL

population, are already present years earlier among a small number of NRB-associated CLL cells. This indicates that a substantial number of IgV-variants are generated early and persist among the CLL tumor clone. This persistence is supported by the analysis of relapsing cases, which show that the majority of IgV-diversity is conserved in CLL, as it is conservatively reconstituted even after long and deep remission. The idea that subclonal diversity is created early in CLL pathogenesis is also supported by epimutation studies (Gaiti et al., 2019) and subclonal variants in the CLL precursor MBL (Lanasa et al., 2010). Importantly, in all cases analyzed the NRB-associated CLL cells comprehensively cover the IgV-subclone diversity, ranging from IGHV germline configuration in U-CLL cases to the clone members with highest mutation load in U-CLL/M-CLL. Therefore, their emergence has begun with the first diversifying events, and the entire IgV diversification history is preserved among the NRB-associated subset, which persist along the typical CLL cells of the tumor mass, in principle just like memory B cells. As this is consistently detectable in all CLL cases in our cohort, we propose that IgV-diversification and memory formation (GC-dependent or -independent) may generally occur during early CLL pathogenesis, in U-CLL as well as in M-CLL. Both tumor entities can show highly complex IgV-mutation dendrograms (e.g., in the number of branching events or branch length). The phylogenetic difference between these cases is that U-CLL tumors contain an on average higher fraction of unmutated or lowly mutated sequences, therefore the tumor mass distributes closer to the dendrogram root.

The persistence of these early generated cells is supported by our analysis of chromosomal CNVs. Chromosomal aberrations are early, driving events in CLL pathogenesis (Baliakas et al., 2019; Landau et al., 2014). The prediction of chromosomal CNVs in our four scRNAseq/scVDJseq data sets confirmed the common presence of a known, monoclonal 17p deletion in one patient, but suggested the absence of a known trisomy 12 only in NRB-associated CLL cells of another patient. These findings give some more insight into the nature of NRB-associated CLL cells: the absence of trisomy 12 in the latter patient strongly supports their pathogenetically early generation, i.e., before acquisition of trisomy 12. The 17p-deletion in the other patient, however, shows that NRB-associated CLL cells may be malignant, carrying powerful driving mutations (Döhner et al., 2000). To assess whether NRB-associated CLL cells may have clinical relevance, i.e., may represent a pool of malignant “precursor cells” which can contribute to the outgrowth of new CLL subclones, additional studies are required. In our view such studies require a comprehensive comparison of the genomic aberration landscape between NRB-associated and main CLL population cells, and subsequent validation by targeted single-cell approaches. We currently aim to develop a direct enrichment strategy for the NRB associated CLL cells, as deep sequencing of the bulk $CD20^{high}$ population is insufficient to reliably determine unique or shared somatic mutations among NRB-associated CLL cells. Note that contaminations from the main CLL population can only be excluded by single-cell RNAseq/VDJseq but would massively affect the detection of false positive “shared” mutation events in bulk sequencing approaches.

Taken together, our study identified a previously unrecognized tumor cell subset in M-CLL and U-CLL cases, which developed early in pathogenesis, is characterized by high intraclonal IgV-diversity, carries traits of normal $CD5^+$ memory B cells, and persist along the main CLL tumor mass. During disease onset, these "a-priori" subclones grow out in parallel, but likely compete with each other: CLL pathogenesis is from early on multicentric, very often including multiple large, expanded IgV-variants. These subclonal dynamics can be traced by VDJ deep sequencing among CLL subsets during the course of the disease.

Materials and methods

Sample preparation

B cells were isolated by Ficoll density centrifugation (density 1.077 g/ml, Pan BioTech, Aidenbach, Germany) followed by staining with anti-CD3 (BD Biosciences, Heidelberg, Germany), anti-CD5 (BioLegend, Koblenz, Germany), anti-CD20, anti-CD23, and anti-CD27 (each BD Biosciences) antibodies. Stained cells were analyzed on a CytoFLEX S flow cytometer (Beckman Coulter, Krefeld, Germany) using CytExpert v2.4 (Beckman Coulter) or FlowJo v10.6.2 (BD Biosciences) software. Sort-purification was performed with a FACSAria Fusion cell sorter (BD Biosciences).

Single-cell RNA- and IGHV-sequencing

Combined scRNAseq and scVDJseq was performed with a Chromium Controller and Next GEM technology (10XGenomics, Pleasanton, CA, USA) using the "Chromium Single Cell V(D)J Enrichment Kit, Human B Cell" and "Chromium Next GEM Single Cell 5' Library and Gel Bead Kit v1.1" (10XGenomics) for the recovery of up to 15,000 single cells. NRB and CLL cells from patients 01 and 02 were sort-purified and separately loaded on the chip, NRB and CLL cells from patients 03 and 04 were sort-purified and pooled in equal amounts before loading on the chip. The generated libraries were purified, and paired end sequenced using the MiSeq v3 reagent kit (Illumina, San Diego, CA, USA). Reads were aligned with cellranger v3.1.0 (upon GRCh38 2020 version) and analyzed using Seurat (Stuart et al., 2019) with the following settings: all genes found at least in ten cells, all cells with at least 800 different genes, a count of at least 3000 reads per cell and cells with no more than 7 percent reads from mitochondrial RNA were further considered. Only cells for which an IGHV-rearrangement was obtained were processed as this was used to identify the cells from the CLL tumor clone. Identification of distinct cell phenotypes was carried out by SingleR (Aran et al., 2019) using our own array-based datasets (Seifert et al., 2012).

Heatmap

Differentially expressed genes between the three CLL subsets from four patients in the scRNAseq analysis were identified by comparing the respective mean counts plus/minus the standard deviation of the gene in one population to the maximum value of the same gene in the other population. This was done in pairwise comparisons of each CLL subset from all cells of all patients. Clustering of genes and samples was calculated via the R-tool hclust from the distance matrix.

Pathway analysis

Pathway analysis among differentially expressed genes between CLL subsets was performed via clusterProfiler and enrichR (Wu et al., 2021).

CNV detection

Copy numbers were detected using CONICSmat (Muller et al., 2018) and compared with known copy number aberrations from FISH analyses.

Library generation and BCR repertoire sequencing

UMI strategy (custom): We sort-purified $CD5^+$ memory B cells and CLL cells from four CLL cases with known IGHV rearrangement in aliquots of 5,000 to 10,000 cells each and extracted genomic DNA (Gentra Puregene Core Kit, Qiagen, Hilden, Germany). Library generation was performed as described (Budeus 2021). In short, rearranged IGHV genes were amplified by PCR for the MiSeq Illumina platform (Illumina, San Diego, California, USA). In the first two PCR cycles 12 nucleotide UMIs are introduced and amplified in the second PCR. Sequences with identical UMI were classified as PCR duplicates and reduced to the longest detected consensus sequence. Each ambiguous nucleotide was replaced by the most frequent (majority vote) nucleotide.

Replicate strategy (InvivoScribe): Genomic DNA was extracted from sort-purified cells from 22 patients by QiaAmp DNA micro kit (Qiagen). For reliable identification of expanded clones, each B-cell population was analyzed in two replicates. IGHV gene rearrangements were amplified using the LymphoTrack IGH FR1 Assay kit (InvivoScribe, San Diego, CA, USA). Amplified DNA was processed for HTSeq according to the LymphoTrack protocol, amplified with a 1-step multiplex PCR for 29 cycles using subgroup-specific IGH framework region 1 primers and IGHJ primers, incorporating platform-specific adaptors. PCR products were purified and paired-end sequenced using the MiSeq v3 reagent kit (Illumina, San Diego, CA, USA). Reads with quality score of ≥ 30 and mean quality score ≥ 35 were further processed. Ambiguities between forward and reverse reads were denoted as 'N'. Identical sequences were reduced to the longest detected sequence (collapsing) and each 'N' in a sequence was replaced by the most frequent (majority vote) nucleotide. Only sequences detected more than once were further processed.

Determination of clonal sequences

Sequences were considered as clonally related when using the same IGHV gene and sharing at least 95% complementarity determining region III (CDRIII) nucleotide sequence homology, including up to 5% length variation. Identification of the IGHV gene was performed in R (www.R-project.org/) using rBLAST and the international ImMunoGeneTics information system (IMGT) database (www.imgt.org). Mutation dendrograms were calculated with IgTree (Barak et al., 2008).

Random drawing for phylogenetic dendrogram structure identification

For each patient, a fixed number of CLL IgV gene sequences, equal to the number of total NRB-associated CLL cell IgV gene sequences from the same timepoint, was drawn randomly from the uncollapsed number of sequences. To remove putative sequence contaminations (e.g., by PCR artefacts or barcode hopping), a Gaussian Mixture Model was used and collapsed sequences that did not belong with 100% certainty to the NRB-associated CLL cell distribution were excluded from the calculation. A Mixture Model is a mixture of k component distributions that collectively make a mixture distribution

$$f(x) = \sum_{k=1}^{K} a_k f_k(x)$$

$a_k$represents a mixing weight for the kth component where $\sum_{k=1}^{K} \quad a_k = 1$. The R-package mixtools (Benaglia et al., 2009) was used to calculate the probability with a log-likelihood approach.

Phylogenetic analysis

All calculations (e.g., tree parameter etc.) were based on the whole un-simplified dendrograms including all available (drawn) sequences. Note that only for visualization in the figures, the full dendrograms were "simplified" by community calculations from the igraph R package (Csardi and Nepusz, 2006), as otherwise the number of nodes and branches would be overwhelming and impossible to be displayed in a suitable resolution.

Maximum betweenness centrality

The Maximum betweenness centrality is calculated from the number of shortest paths through a node:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

$v, s, t = nodes, \sigma_{st} =$ total number of shortest paths between s and t, $\sigma_{st}(v)$ = number of those paths which path through v. For each dendrogram the maximum number is given. This is a measurement of importance of nodes.

Branching

Branching calculates the mean number of outgoing edges (branches) per sample.

Intraclonal diversity

The intraclonal diversity is calculated from the approximate string distance between each sequence. The distance is a generalized Levenshtein (edit) distance, giving the minimal possibly weighted number of insertions, deletions and substitutions needed to transform one string into another. For each dendrogram, the maximum number is given. This is a measurement of the variability of the dendrogram.

In vitro proliferation assay

NRB and CLL cells were sort purified and stained with 5 µM eFluor 670 cell proliferation dye (ThermoFischer, Oberhausen, Germany). eFluor670^{high} cells were cultured in RPMI 1640 medium supplemented with 20% fetal bovine serum, 100 U/ml penicillin and 100 µg/ml streptomycin (all from Pan Biotech, Aidenbach, Germany) at 37°C and 5% CO_2. Stimulation was performed using 0.03 µg/µl anti-immunoglobulin (anti-Ig, Jackson ImmunoResearch, Newmarket, UK) and 1 µg/ml CD40-ligand-HA with 5 ng/ml anti-HA antibodies (both R&D Systems, Minneapolis, USA), and Type B CpG OND D-SL01 (InvivoGen, San Diego, USA). After 5 days of *in vitro* culture, proliferated (eFluor670^{low}) and non-proliferated (eFluor670^{high}) fractions were sort purified and RNA extraction was performed (miRNeasy micro-Kit, Qiagen). After cDNA synthesis (High capacity cDNA Reverse Transcription Kit, Applied Biosystems) two rounds of multiplex PCR using a mixture of custom IGHV- and IGHJ-primers (Budeus et al., 2015) with a total of 75 cycles were performed. PCR products were ligated into pGEM-T easy vectors (Promega, Walldorf, Germany) and transformed into JM109 E. coli (Promega, Walldorf, Germany). Bacteria were grown over night on LB/amp/IPTG/X-gal plates and non-colored colonies were selected for plasmid preparation. Plasmids were extracted by mi-Plasmid Miniprep Kit (Metabion, Planegg, Germany) and processed with Sanger sequencing. CLL tumor cells were identified by their respective IGHV-rearrangement.

Statistics

Dendrogram parameters and proliferation capacities were analyzed with a Wilcoxon signed rank test, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Study approval

PB samples were obtained after written informed consent was received from participants prior to inclusion in the study according to the Declaration of Helsinki, and approval by the ethics committee of the Medical Faculty at the University of Duisburg-Essen, Germany (BO-10-4380).

Author contribution

M.S. developed the concept. M.S., B.B., R.K., and J.D. designed and developed experimental strategies, contributed to data evaluation, and prepared the manuscript. B.B., M.B., A.K., K.B., M.D., E.H., and S.T. designed and carried out most of the experiments. B.B. developed and performed bioinformatical and statistical analysis. P.J., J.B., J.v.T., C.R., S.S., and J.D. contributed samples, clinical data evaluation and helpful discussion.

Acknowledgements

We thank Julia Jesdinsky-Elsenbruch for excellent technical assistance and Klaus Lennartz for his valuable engineering support. We thank Ludger Klein-Hitpass, Rene Scholtysik and the staff members of the Genomics and Transcriptomics Facility at the Institute of Cell Biology in Essen. This work was supported by the Deutsche Forschungsgemeinschaft through SE1885/2-1, SE1885/2-2, SE1885/4-1, SFB1074 Subprojects B1 and B2, the DFG Clinician Scientist Program UMEA, FU 356/12-1, the intramural gender equity measure "Willkommen zurück", the Dr. Werner Jackstädt-Stiftung, and the Deutsche Krebshilfe through grant 70112628.

References

Aguilar-Hernandez, M.M., M.D. Blunt, R. Dobson, A. Yeomans, S. Thirdborough, M. Larrayoz, L.D. Smith, A. Linley, J.C. Strefford, A. Davies, P.M. Johnson, N. Savelyeva, M.S. Cragg, F. Forconi, G. Packham, F.K. Stevenson, and A.J. Steele. 2016. IL-4 enhances expression and function of surface IgM in CLL cells. *Blood* 127:3015-3025.

Aran, D., A.P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, S. Chak, R.P. Naikawadi, P.J. Wolters, A.R. Abate, A.J. Butte, and M. Bhattacharya. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 20:163-172.

Bagnara, D., V. Callea, C. Stelitano, F. Morabito, S. Fabris, A. Neri, S. Zanardi, F. Ghiotto, E. Ciccone, C.E. Grossi, and F. Fais. 2006. IgV gene intraclonal diversification and clonal evolution in B-cell chronic lymphocytic leukaemia. *British journal of haematology* 133:50-58.

Baliakas, P., S. Jeromin, M. Iskas, A. Puiggros, K. Plevova, F. Nguyen-Khac, Z. Davis, G.M. Rigolin, A. Visentin, A. Xochelli, J. Delgado, F. Baran-Marszak, E. Stalika, P. Abrisqueta, K. Durechova, G. Papaioannou, V. Eclache, M. Dimou, T. Iliakis, R. Collado, M. Doubek, M.J. Calasanz, N. Ruiz-Xiville, C. Moreno, M. Jarosova, A.C. Leeksma, P. Panayiotidis, H. Podgornik, F. Cymbalista, A. Anagnostopoulos, L. Trentin, N. Stavroyianni, F. Davi, P. Ghia, A.P. Kater, A. Cuneo, S. Pospisilova, B. Espinet, A. Athanasiadou, D. Oscier, C. Haferlach, K. Stamatopoulos, and t.E.R.I.o.C.L.L. Eric. 2019. Cytogenetic complexity in chronic lymphocytic leukemia: definitions, associations, and clinical impact. *Blood* 133:1205-1216.

Barak, M., N.S. Zuckerman, H. Edelman, R. Unger, and R. Mehr. 2008. IgTree: creating Immunoglobulin variable region gene lineage trees. *J Immunol Methods* 338:67-74.

Benaglia, T., D. Chauveau, D.R. Hunter, and D.S. Young. 2009. mixtools: An R Package for Analyzing Mixture Models. *2009* 32:29.

Bosch, F., and R. Dalla-Favera. 2019. Chronic lymphocytic leukaemia: from genetics to treatment. *Nat Rev Clin Oncol* 16:684-701.

Budeus, B., S. Schweigle de Reynoso, M. Przekopowitz, D. Hoffmann, M. Seifert, and R. Küppers. 2015. Complexity of the human memory B-cell compartment is determined by the versatility of clonal diversification in germinal centers. *Proc Natl Acad Sci U S A* 112:E5281-5289.

Burger, J.A., M. Burger, and T.J. Kipps. 1999. Chronic lymphocytic leukemia B cells express functional CXCR4 chemokine receptors that mediate spontaneous migration beneath bone marrow stromal cells. *Blood* 94:3658-3667.

Caligaris-Cappio, F., M. Gobbi, M. Bofill, and G. Janossy. 1982. Infrequent normal B lymphocytes express features of B-chronic lymphocytic leukemia. *J Exp Med* 155:623-628.

Calissano, C., R.N. Damle, S. Marsilio, X.J. Yan, S. Yancopoulos, G. Hayes, C. Emson, E.J. Murphy, M.K. Hellerstein, C. Sison, M.S. Kaufman, J.E. Kolitz, S.L. Allen, K.R. Rai, I. Ivanovic, I.M. Dozmorov, S. Roa, M.D. Scharff, W. Li, and N. Chiorazzi. 2011. Intraclonal complexity in chronic lymphocytic leukemia: fractions enriched in recently born/divided and older/quiescent cells. *Mol Med* 17:1374-1382.

Campbell, P.J., E.D. Pleasance, P.J. Stephens, E. Dicks, R. Rance, I. Goodhead, G.A. Follows, A.R. Green, P.A. Futreal, and M.R. Stratton. 2008. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A* 105:13081-13086.

Chiorazzi, N., and M. Ferrarini. 2003. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. *Annu Rev Immunol* 21:841-894.

Chiorazzi, N., and M. Ferrarini. 2011. Cellular origin(s) of chronic lymphocytic leukemia: cautionary notes and additional considerations and possibilities. *Blood* 117:1781-1791.

Criado, I., E. Blanco, A. Rodriguez-Caballero, M. Alcoceba, T. Contreras, M.L. Gutierrez, A. Romero, P. Fernandez-Navarro, M. Gonzalez, F. Solano, C. Gomez, M. Perez-Andres, J.J.M. van Dongen, J. Almeida, A. Orfao, P.I.D.G. EuroFlow, and M.B.L. The Primary Health Care Group of Salamanca for the Study of. 2018. Residual normal B-cell profiles in monoclonal B-cell lymphocytosis versus chronic lymphocytic leukemia. *Leukemia* 32:2701-2705.

Csardi, G., and T. Nepusz. 2006. The igraph software package for complex network research. *InterJournal* Complex Systems:1695.

Damm, F., E. Mylonas, A. Cosson, K. Yoshida, V. Della Valle, E. Mouly, M. Diop, L. Scourzic, Y. Shiraishi, K. Chiba, H. Tanaka, S. Miyano, Y. Kikushige, F. Davi, J. Lambert, D. Gautheret, H. Merle-Beral, L. Sutton, P. Dessen, E. Solary, K. Akashi, W. Vainchenker, T. Mercher, N. Droin, S. Ogawa, F. Nguyen-Khac, and O.A. Bernard. 2014. Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer discovery* 4:1088-1101.

Defrance, T., M. Taillardet, and L. Genestier. 2011. T cell-independent B cell memory. *Curr Opin Immunol* 23:330-336.

Döhner, H., S. Stilgenbauer, A. Benner, E. Leupolt, A. Krober, L. Bullinger, K. Dohner, M. Bentz, and P. Lichter. 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* 343:1910-1916.

Endo, T., M. Nishio, T. Enzler, H.B. Cottam, T. Fukuda, D.F. James, M. Karin, and T.J. Kipps. 2007. BAFF and APRIL support chronic lymphocytic leukemia B-cell survival through activation of the canonical NF-kappaB pathway. *Blood* 109:703-710.

Forconi, F., K.N. Potter, I. Wheatley, N. Darzentas, E. Sozzi, K. Stamatopoulos, C.I. Mockridge, G. Packham, and F.K. Stevenson. 2010. The normal IGHV1-69-derived

B-cell repertoire contains stereotypic patterns characteristic of unmutated CLL. *Blood* 115:71-77.

Gaiti, F., R. Chaligne, H. Gu, R.M. Brand, S. Kothen-Hill, R.C. Schulman, K. Grigorev, D. Risso, K.T. Kim, A. Pastore, K.Y. Huang, A. Alonso, C. Sheridan, N.D. Omans, E. Biederstedt, K. Clement, L. Wang, J.A. Felsenfeld, E.B. Bhavsar, M.J. Aryee, J.N. Allan, R. Furman, A. Gnirke, C.J. Wu, A. Meissner, and D.A. Landau. 2019. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* 569:576-580.

Gohil, S.H., and C.J. Wu. 2019. Dissecting CLL through high-dimensional single-cell technologies. *Blood* 133:1446-1456.

Good, K.L., D.T. Avery, and S.G. Tangye. 2009. Resting human memory B cells are intrinsically programmed for enhanced survival and responsiveness to diverse stimuli compared to naive B cells. *J Immunol* 182:890-901.

Gruber, M., I. Bozic, I. Leshchiner, D. Livitz, K. Stevenson, L. Rassenti, D. Rosebrock, A. Taylor-Weiner, O. Olive, R. Goyetche, S.M. Fernandes, J. Sun, C. Stewart, A. Wong, C. Cibulskis, W. Zhang, J.G. Reiter, J.M. Gerold, J.G. Gribben, K.R. Rai, M.J. Keating, J.R. Brown, D. Neuberg, T.J. Kipps, M.A. Nowak, G. Getz, and C.J. Wu. 2019. Growth dynamics in naturally progressing chronic lymphocytic leukaemia. *Nature* 570:474-479.

Gurrieri, C., P. McGuire, H. Zan, X.J. Yan, A. Cerutti, E. Albesiano, S.L. Allen, V. Vinciguerra, K.R. Rai, M. Ferrarini, P. Casali, and N. Chiorazzi. 2002. Chronic lymphocytic leukemia B cells can undergo somatic hypermutation and intraclonal immunoglobulin V(H)DJ(H) gene diversification. *J Exp Med* 196:629-639.

Hallek, M. 2019. Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment. *Am J Hematol* 94:1266-1287.

Herishanu, Y., P. Perez-Galan, D. Liu, A. Biancotto, S. Pittaluga, B. Vire, F. Gibellini, N. Njuguna, E. Lee, L. Stennett, N. Raghavachari, P. Liu, J.P. McCoy, M. Raffeld, M. Stetler-Stevenson, C. Yuan, R. Sherry, D.C. Arthur, I. Maric, T. White, G.E. Marti, P. Munson, W.H. Wilson, and A. Wiestner. 2011. The lymph node microenvironment promotes B-cell receptor signaling, NF-kappaB activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood* 117:563-574.

Huemer, M., S. Rebhandl, N. Zaborsky, F.J. Gassner, S. Hainzl, L. Weiss, D. Hebenstreit, R. Greil, and R. Geisberger. 2014. AID induces intraclonal diversity and genomic damage in CD86(+) chronic lymphocytic leukemia cells. *European journal of immunology* 44:3747-3757.

Kaji, T., A. Ishige, M. Hikida, J. Taka, A. Hijikata, M. Kubo, T. Nagashima, Y. Takahashi, T. Kurosaki, M. Okada, O. Ohara, K. Rajewsky, and T. Takemori. 2012. Distinct cellular pathways select germline-encoded and somatically mutated antibodies into immunological memory. *J Exp Med* 209:2079-2097.

Kibler, A., B. Budeus, E. Homp, K. Bronischewski, V. Berg, L. Sellmann, F. Murke, A. Heinold, F.M. Heinemann, M. Lindemann, I. Bekeredjian-Ding, P.A. Horn, C.J. Kirschning, R. Küppers, and M. Seifert. 2021. Systematic memory B cell archiving and random display shape the human splenic marginal zone throughout life. *The Journal of experimental medicine* 218:

Kikushige, Y., F. Ishikawa, T. Miyamoto, T. Shima, S. Urata, G. Yoshimoto, Y. Mori, T. Iino, T. Yamauchi, T. Eto, H. Niiro, H. Iwasaki, K. Takenaka, and K. Akashi. 2011. Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. *Cancer Cell* 20:246-259.

Kipps, T.J., F.K. Stevenson, C.J. Wu, C.M. Croce, G. Packham, W.G. Wierda, S. O'Brien, J. Gribben, and K. Rai. 2017. Chronic lymphocytic leukaemia. *Nat Rev Dis Primers* 3:16096.

Klein, U., Y. Tu, G.A. Stolovitzky, M. Mattioli, G. Cattoretti, H. Husson, A. Freedman, G. Inghirami, L. Cro, L. Baldini, A. Neri, A. Califano, and R. Dalla-Favera. 2001. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J Exp Med* 194:1625-1638.

Kulis, M., S. Heath, M. Bibikova, A.C. Queiros, A. Navarro, G. Clot, A. Martinez-Trillos, G. Castellano, I. Brun-Heath, M. Pinyol, S. Barberan-Soler, P. Papasaikas, P. Jares, S. Bea, D. Rico, S. Ecker, M. Rubio, R. Royo, V. Ho, B. Klotzle, L. Hernandez, L. Conde, M. Lopez-Guerra, D. Colomer, N. Villamor, M. Aymerich, M. Rozman, M. Bayes, M. Gut, J.L. Gelpi, M. Orozco, J.B. Fan, V. Quesada, X.S. Puente, D.G. Pisano, A. Valencia, A. Lopez-Guillermo, I. Gut, C. Lopez-Otin, E. Campo, and J.I. Martin-Subero. 2012. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nature genetics* 44:1236-1242.

Lanasa, M.C., S.D. Allgood, A.D. Volkheimer, J.P. Gockerman, J.F. Whitesides, B.K. Goodman, J.O. Moore, J.B. Weinberg, and M.C. Levesque. 2010. Single-cell analysis reveals oligoclonality among 'low-count' monoclonal B-cell lymphocytosis. *Leukemia* 24:133-140.

Landau, D.A., S.L. Carter, P. Stojanov, A. McKenna, K. Stevenson, M.S. Lawrence, C. Sougnez, C. Stewart, A. Sivachenko, L. Wang, Y. Wan, W. Zhang, S.A. Shukla, A. Vartanov, S.M. Fernandes, G. Saksena, K. Cibulskis, B. Tesar, S. Gabriel, N. Hacohen, M. Meyerson, E.S. Lander, D. Neuberg, J.R. Brown, G. Getz, and C.J. Wu. 2013. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 152:714-726.

Landau, D.A., K. Clement, M.J. Ziller, P. Boyle, J. Fan, H. Gu, K. Stevenson, C. Sougnez, L. Wang, S. Li, D. Kotliar, W. Zhang, M. Ghandi, L. Garraway, S.M. Fernandes, K.J. Livak, S. Gabriel, A. Gnirke, E.S. Lander, J.R. Brown, D. Neuberg, P.V. Kharchenko, N. Hacohen, G. Getz, A. Meissner, and C.J. Wu. 2014. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* 26:813-825.

Landau, D.A., E. Tausch, A.N. Taylor-Weiner, C. Stewart, J.G. Reiter, J. Bahlo, S. Kluth, I. Bozic, M. Lawrence, S. Bottcher, S.L. Carter, K. Cibulskis, D. Mertens, C.L. Sougnez, M. Rosenberg, J.M. Hess, J. Edelmann, S. Kless, M. Kneba, M. Ritgen, A. Fink, K. Fischer, S. Gabriel, E.S. Lander, M.A. Nowak, H. Dohner, M. Hallek, D. Neuberg, G. Getz, S. Stilgenbauer, and C.J. Wu. 2015. Mutations driving CLL and their evolution in progression and relapse. *Nature* 526:525-530.

Lu, D., Y. Zhao, R. Tawatao, H.B. Cottam, M. Sen, L.M. Leoni, T.J. Kipps, M. Corr, and D.A. Carson. 2004. Activation of the Wnt signaling pathway in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 101:3118-3123.

Lutzny, G., T. Kocher, M. Schmidt-Supprian, M. Rudelius, L. Klein-Hitpass, A.J. Finch, J. Dürig, M. Wagner, C. Haferlach, A. Kohlmann, S. Schnittger, M. Seifert, S. Wanninger, N. Zaborsky, R. Oostendorp, J. Ruland, M. Leitges, T. Kuhnt, Y. Schafer, B. Lampl, C. Peschel, A. Egle, and I. Ringshausen. 2013. Protein kinase c-beta-dependent activation of NF-kappaB in stromal cells is indispensable for the survival of chronic lymphocytic leukemia B cells in vivo. *Cancer Cell* 23:77-92.

Marsilio, S., H. Khiabanian, G. Fabbri, S. Vergani, C. Scuoppo, E. Montserrat, E.J. Shpall, M. Hadigol, P. Marin, K.R. Rai, R. Rabadan, S. Devereux, L. Pasqualucci, and N. Chiorazzi. 2018. Somatic CLL mutations occur at multiple distinct hematopoietic maturation stages: documentation and cautionary note regarding cell fraction purity. *Leukemia* 32:1041-1044.

Marti, G., F. Abbasi, E. Raveche, A.C. Rawstron, P. Ghia, T. Aurran, N. Caporaso, Y.K. Shim, and R.F. Vogt. 2007. Overview of monoclonal B-cell lymphocytosis. *British journal of haematology* 139:701-708.

Meng, W., B. Zhang, G.W. Schwartz, A.M. Rosenfeld, D. Ren, J.J.C. Thome, D.J. Carpenter, N. Matsuoka, H. Lerner, A.L. Friedman, T. Granot, D.L. Farber, M.J. Shlomchik, U. Hershberg, and E.T. Luning Prak. 2017. An atlas of B-cell clonal distribution in the human body. *Nat Biotechnol* 35:879-884.

Muller, S., A. Cho, S.J. Liu, D.A. Lim, and A. Diaz. 2018. CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. *Bioinformatics* 34:3217-3219.

Oakes, C.C., M. Seifert, Y. Assenov, L. Gu, M. Przekopowitz, A.S. Ruppert, Q. Wang, C.D. Imbusch, A. Serva, S.D. Koser, D. Brocks, D.B. Lipka, O. Bogatyrova, D. Weichenhan, B. Brors, L. Rassenti, T.J. Kipps, D. Mertens, M. Zapatka, P. Lichter, H. Döhner, R. Küppers, T. Zenz, S. Stilgenbauer, J.C. Byrd, and C. Plass. 2016. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nature genetics* 48:253-264.

Oscier, D.G., A. Thompsett, D. Zhu, and F.K. Stevenson. 1997. Differential rates of somatic hypermutation in V(H) genes among subsets of chronic lymphocytic leukemia defined by chromosomal abnormalities. *Blood* 89:4153-4160.

Pascutti, M.F., M. Jak, J.M. Tromp, I.A. Derks, E.B. Remmerswaal, R. Thijssen, M.H. van Attekum, G.G. van Bochove, D.M. Luijks, S.T. Pals, R.A. van Lier, A.P. Kater, M.H. van Oers, and E. Eldering. 2013. IL-21 and CD40L signals from autologous T cells can induce antigen-independent proliferation of CLL cells. *Blood* 122:3010-3019.

Pastore, A., F. Gaiti, S.X. Lu, R.M. Brand, S. Kulm, R. Chaligne, H. Gu, K.Y. Huang, E.K. Stamenova, W. Beguelin, Y. Jiang, R.C. Schulman, K.T. Kim, A. Alonso, J.N. Allan, R.R. Furman, A. Gnirke, C.J. Wu, A.M. Melnick, A. Meissner, B.E. Bernstein, O. Abdel-Wahab, and D.A. Landau. 2019. Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in CLL. *Nat Commun* 10:1874.

Patten, P.E., C.C. Chu, E. Albesiano, R.N. Damle, X.J. Yan, D. Kim, L. Zhang, A.R. Magli, J. Barrientos, J.E. Kolitz, S.L. Allen, K.R. Rai, S. Roa, P.K. Mongini, T. MacCarthy, M.D. Scharff, and N. Chiorazzi. 2012. IGHV-unmutated and IGHV-mutated chronic lymphocytic leukemia cells produce activation-induced deaminase protein with a full range of biologic functions. *Blood* 120:4802-4811.

Penter, L., S.H. Gohil, C. Lareau, L.S. Ludwig, E.M. Parry, T. Huang, S. Li, W. Zhang, D. Livitz, I. Leshchiner, L. Parida, G. Getz, L.Z. Rassenti, T.J. Kipps, J.R. Brown, M.S. Davids, D.S. Neuberg, K.J. Livak, V.G. Sankaran, and C.J. Wu. 2021. Longitudinal Single-Cell Dynamics of Chromatin Accessibility and Mitochondrial Mutations in Chronic Lymphocytic Leukemia Mirror Disease History. *Cancer Discov*

Rawstron, A.C., F. Bennett, and P. Hillmen. 2007. The biological and clinical relationship between CD5+23+ monoclonal B-cell lymphocytosis and chronic lymphocytic leukaemia. *British journal of haematology* 139:724-729.

Schleiss, C., W. Ilias, O. Tahar, Y. Guler, L. Miguet, C. Mayeur-Rousse, L. Mauvieux, L.M. Fornecker, E. Toussaint, R. Herbrecht, F. Bertrand, M. Maumy-Bertrand, T. Martin, S. Fournel, P. Georgel, S. Bahram, and L. Vallat. 2019. BCR-associated factors driving chronic lymphocytic leukemia cells proliferation ex vivo. *Sci Rep* 9:701.

Seifert, M., and R. Küppers. 2009. Molecular footprints of a germinal center derivation of human IgM+(IgD+)CD27+ B cells and the dynamics of memory B cell generation. *J Exp Med* 206:2659-2669.

Seifert, M., M. Przekopowitz, S. Taudien, A. Lollies, V. Ronge, B. Drees, M. Lindemann, U. Hillen, H. Engler, B.B. Singer, and R. Küppers. 2015. Functional capacities of human IgM memory B cells in early inflammatory responses and secondary germinal center reactions. *Proc Natl Acad Sci U S A* 112:E546-555.

Seifert, M., L. Sellmann, J. Bloehdorn, F. Wein, S. Stilgenbauer, J. Dürig, and R. Küppers. 2012. Cellular origin and pathophysiology of chronic lymphocytic leukemia. *J Exp Med* 209:2183-2198.

Shanafelt, T.D., P. Ghia, M.C. Lanasa, O. Landgren, and A.C. Rawstron. 2010. Monoclonal B-cell lymphocytosis (MBL): biology, natural history and clinical management. *Leukemia* 24:512-520.

Stall, A.M., M.C. Farinas, D.M. Tarlinton, P.A. Lalor, L.A. Herzenberg, S. Strober, and L.A. Herzenberg. 1988. Ly-1 B-cell clones similar to human chronic lymphocytic leukemias routinely develop in older normal mice and young autoimmune (New Zealand Black-related) animals. *Proc Natl Acad Sci U S A* 85:7312-7316.

Stevenson, F.K., F. Forconi, and G. Packham. 2014. The meaning and relevance of B-cell receptor structure and function in chronic lymphocytic leukemia. *Seminars in hematology* 51:158-167.

Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W.M. Mauck, 3rd, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija. 2019. Comprehensive Integration of Single-Cell Data. *Cell* 177:1888-1902 e1821.

Tangye, S.G., D.T. Avery, E.K. Deenick, and P.D. Hodgkin. 2003. Intrinsic differences in the proliferation of naive and memory human B cells as a mechanism for enhanced secondary immune responses. *J Immunol* 170:686-694.

Vardi, A., A. Dagklis, L. Scarfo, D. Jelinek, D. Newton, F. Bennett, J. Almeida, A. Rodriguez-Caballero, S. Allgood, M. Lanasa, A. Cortelezzi, E. Orlandi, S. Veronese, M. Montillo, A. Rawstron, T. Shanafelt, A. Orfao, K. Stamatopoulos, and P. Ghia. 2013. Immunogenetics shows that not all MBL are equal: the larger the clone, the more similar to CLL. *Blood* 121:4521-4528.

Volkheimer, A.D., J.B. Weinberg, B.E. Beasley, J.F. Whitesides, J.P. Gockerman, J.O. Moore, G. Kelsoe, B.K. Goodman, and M.C. Levesque. 2007. Progressive immunoglobulin gene mutations in chronic lymphocytic leukemia: evidence for antigen-driven intraclonal diversification. *Blood* 109:1559-1567.

Wang, L., J. Fan, J.M. Francis, G. Georghiou, S. Hergert, S. Li, R. Gambe, C.W. Zhou, C. Yang, S. Xiao, P.D. Cin, M. Bowden, D. Kotliar, S.A. Shukla, J.R. Brown, D. Neuberg, D.R. Alessi, C.Z. Zhang, P.V. Kharchenko, K.J. Livak, and C.J. Wu. 2017. Integrated single-cell genetic and transcriptional analysis suggests novel drivers of chronic lymphocytic leukemia. *Genome Res* 27:1300-1311.

Weill, J.C., S. Weller, and C.A. Reynaud. 2009. Human marginal zone B cells. *Annu Rev Immunol* 27:267-285.

Wu, T., E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, and G. Yu. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (N Y)* 2:100141.

Yu, J., L. Chen, B. Cui, G.F. Widhopf, 2nd, Z. Shen, R. Wu, L. Zhang, S. Zhang, S.P. Briggs, and T.J. Kipps. 2016. Wnt5a induces ROR1/ROR2 heterooligomerization to enhance leukemia chemotaxis and proliferation. *J Clin Invest* 126:585-598.

Zenz, T., D. Mertens, R. Küppers, H. Döhner, and S. Stilgenbauer. 2010. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nat Rev Cancer* 10:37-50.

Zhao, Z., L. Goldin, S. Liu, L. Wu, W. Zhou, H. Lou, Q. Yu, S.X. Tsang, M. Jiang, F. Li, M. McMaster, Y. Li, X. Lin, Z. Wang, L. Xu, G. Marti, G. Li, K. Wu, M. Yeager, H. Yang, X. Xu, S.J. Chanock, B. Li, Y. Hou, N. Caporaso, and M. Dean. 2016. Evolution of multiple cell clones over a 29-year period of a CLL patient. *Nat Commun* 7:13765.

Tables

Table 1: CLL patient information.

Number of CLL patients		60
Age	*Mean ± SD*	64.3 ± 12.6
	Range	37 - 89
Gender	*Male*	41
	Female	19
IGHV mutation status	*Mutated*	22
	Unmutated	16
	n/a	22
Binet stage (at sampling)	*A*	45
	B	9
	C	6
High risk cytogenetics	*Del17p or del 11q*	12
	No del17p/del11q	35
	n/a	13
Lymphocyte doubling time	*< 12 months*	13
	> 12 months	38
	n/a	9
Time to first treatment	*Patients with treatment*	24
	Median (months)	34.5
	Range (months)	0 - 131
Follow up	*Median (months)*	67
	Range (months)	0 - 278

n/a, not available

Table 2: Characteristics of the BCR repertoire sequencing CLL cohort. IGHV mutation status was determined by IGHV gene sequencing at diagnosis and confirmed by IGHV repertoire deep sequencing. Cases with an average IGHV mutation frequency < 2% were regarded as uCLL. Cytogenetic abnormalities, age in years at diagnosis (age), and lymphocyte doubling time (LDT) determined during the first 12 months after diagnosis are given. Unavailable information is denoted as n.a..

ID	IgV status	Rearranged IGHV gene	IGH CDR III amino acid sequence	Cytogenetics and selected mutations	Age (years)	LDT (months)	Used in:
01	M	1-18	ATEGGYCSGGSCYRIDY	n.a.	49	> 12	scRNAseq
02	U	3-23	AKDLRLTYYDFWSGHTLAFDI	tris12	59	> 06	scRNAseq
03	U	3-30	ARDKALPYDMYYYGSGSYKD	del17p	73	n.a.	scRNAseq/NGS (replicates)
04	M	2-5	ASLYSSSWYNWHDP	del13q14	30	n.a.	scRNAseq/NGS (replicates)
05	M	4-34	ARGYPDTPMMKRYYYYGFDV	del13q14	45	> 12	NGS (UMIs)
06	M	4-39	AWARGTNSWGRFDY	del13q14	67	> 12	NGS (UMIs)
07	M	3-23	AKDVAGPGGSGYYYPFDY	del13q14	60	> 12	NGS (UMIs)
08	M	3-23	AKSQSGSGAYYSGGEFDF	normal	57	> 12	NGS (UMIs)
09	U	1-69	ARASIFGVVNLGPPDYYYGMDV	n.a.	78	n.a.	NGS (replicates)
10	U	5-51	APKSRYFDWLLPASSEWFFDY	del13q14	63	< 12	NGS (replicates)
11	M	3-30	AKDDAIYYDSSGYDAYYYYGMAV	tris12	63	> 12	NGS (replicates)
12	U	4-34	ARGRRIVDYYDSSGYYLGYYYYGMDV	del 11q, TP53mut	41	n.a.	NGS (replicates)
13	U	2-26	ARNADPLLRYFDWLLLDFDY	del11q	47	< 12	NGS (replicates)
14	U	1-69	ARDCYDYVWGSYLRAFDI	normal, SF3B1mut	75	n.a.	NGS (replicates)
15	U	3-23D	ARSNSYYDFWSGYHVDYYYYMDV	tris12, t (11;14)	56	> 12	NGS (replicates)
16	U	3-11	AREEVLYYGSGSYYNWFDP	del11q, del13q	73	< 12	NGS (replicates)
17	M	3-53	ARDPPHIEIQCFGY	del13q14	56	> 12	NGS (replicates)
18	M	3-30	ARPVGGRWLAKRIVLLCLL	del13q	78	> 12	NGS (replicates)
19	M	5-51	ARPSFSGSYLDAFDI	del13q14	37	> 12	NGS (replicates)
20	M	4-34	ARRQ	del13q14	61	> 12	NGS (replicates)
21	M	3-74	ARVYYSDNTDRRNFDH	n.a.	54	> 12	NGS (replicates)
22	M	3-15	AIANCGASCAGLNWFDP	del13q14	50	< 12	NGS (replicates)
23	U	1-69	ARVLDCSSTSCYTPHFWYYYYGMDV	del17p, del11q, TP53mut	53	< 06	NGS (replicates)
24	M	1-2	ARVCIVAVGTTGTTP	n.a.	56	> 12	NGS (replicates)
25	M	3-33	VKEAIDETTSGAFDV	del13q14, del17p, TP53mut	58	< 12	NGS (replicates)
26	M	3-23D	AKSENYDYNFWSGYSDPFDI	del13q14	74	> 12	NGS (replicates)

27	M	3-23D	AKNCQNIVATYPDY	del11q, del13q	78	> 12	NGS (replicates)
28	U	1-3	ARVDTTMTPFDY	normal	49	> 12	Proliferation
29	M	3-48	ARGGHEAVPGTPFDY	normal	65	> 12	Proliferation
30	M	3-21	ADVPYTSSWFAFDI	normal	51	> 12	Proliferation
31	M	1-69	ARDGDLGYCSGGSCYWVYY	tris12	63	> 12	Proliferation
32	M	3-21	ARGLRFLPARGDAFDI	normal	32	> 12	Proliferation
33	M	3-48	ADSPEIPVWG	del13q14	56	> 12	Proliferation
34	M	2-5	AHRLGGYNWNHGYFDY	del13q14	68	> 12	Proliferation
35	U	3-71	AHSTTAYYYGMDV	normal	72	> 12	Proliferation
36	M	3-21	ARVAAANLSAAAGVFDY	del13q14	58	> 12	Proliferation

Table 3: Cell equivalents retrieved from scRNA-seq/VDJ-seq profiles

		Donor 01	Donor 02	Donor 03	Donor 04
Retrieved cells (total)		2239	4068	2130	1981
Informative NRB cells[a]		539	341	20	1616
Informative CLL cells[b]	total	1700	667	2110	365
	# CLL cluster 1	615 (36%)	409 (61%)	1141 (54%)	207 (57%)
	# CLL cluster 2	577 (34%)	198 (30%)	637 (30%)	103 (28%)
	# NRB-associated	508 (30%)	60 (9%)	332 (16%)	55 (15%)

[a] cell equivalents with non-CLL-Ig-rearrangement

[b] cell equivalents with CLL-Ig-rearrangement

Table 4: Treatment of patients included in Figure 5 and S4.

Patient	Time	Therapy
03	2013-2014	Chemo-Immunotherapy
03	2014	Radiotherapy
03	2014-2015	BCL2 Inhibitor
03	2016-2018	BTK-Inhibitor + anti-CD20 AB
03	2018	PI3K-Inhibitor
03	2018	Immunomodulatory therapy
04	2010	Chemo-Immunotherapy
10	2008	Chemo-Immunotherapy
10	2013	Chemo-Immunotherapy
12	2007-2008	Chemo-Immunotherapy
12	2011-2013	Immunomodulatory therapy
12	2014	Chemo-Immunotherapy + BTK-Inhibitor
12	2014-2018	BTK-Inhibitor
13	2010	Chemo-Immunotherapy
13	2012-2013	Chemo-Immunotherapy
13	2016-2017	Chemo-Immunotherapy
25	2006	Chemotherapy
25	2017	Chemo-Immunotherapy
25	2017-2018	BTK-Inhibitor