haematologica
Journal of the Ferrata Storti Foundation

Chronic lymphocytic leukemia often arises by a multiclonal selection process

by Davide Bagnara, Andrea N. Mazzarello, Niccolò Cardente, Stefano Vergani, Siddha Kasar, Stacey Fernandes, Gerardo Ferrer, Fabio Ghiotto, Jacqueline C. Barrientos, Jonathan E. Kolitz, Kanti R. Rai, Steve L. Allen, Monica Colombo, Franco Fais, Jennifer R. Brown, Manlio Ferrarini and Nicholas Chiorazzi

Received: August 23, 2024.
Accepted: February 12, 2025.

Citation: Davide Bagnara, Andrea N. Mazzarello, Niccolò Cardente, Stefano Vergani, Siddha Kasar, Stacey Fernandes, Gerardo Ferrer, Fabio Ghiotto, Jacqueline C. Barrientos, Jonathan E. Kolitz, Kanti R. Rai, Steve L. Allen, Monica Colombo, Franco Fais, Jennifer R. Brown, Manlio Ferrarini and Nicholas Chiorazzi. Chronic lymphocytic leukemia often arises by a multiclonal selection process. Haematologica. 2025 Feb 20. doi: 10.3324/haematol.2024.286380 [Epub ahead of print]

Chronic lymphocytic leukemia often arises by a multiclonal selection process.

[*]Davide Bagnara[1,2], [*]Andrea N Mazzarello[1,2], Niccolò Cardente[1], Stefano Vergani[2,3], Siddha Kasar[4,5], Stacey Fernandes[4], Gerardo Ferrer[2,6,7], Fabio Ghiotto[1,8], Jacqueline C Barrientos[2,9], Jonathan E Kolitz[2,9], Kanti R Rai[2,9], Steven L Alen[2,9], Monica Colombo[8], Franco Fais[1,8], Jennifer R Brown[4], Manlio Ferrarini[1], and Nicholas Chiorazzi[2,9]

1 Department of Experimental Medicine, University of Genoa, Genoa, Italy
2 Karches Center for Oncology Research, The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA
3 Miltenyi Biotec B.V. & Co. KG, Bergisch Gladbach, Germany
4 CLL Center, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA
5 Takeda Pharmaceuticals, Cambridge, MA, USA
6 Josep Carreras Leukaemia Research Institute (IJC), Badalona, Catalonia, Spain
7 Centro de Investigación Biomédica en Red Cáncer (CIBERONC), Madrid, Madrid, Spain.
8 IRCCS Ospedale Policlinico San Martino, Genoa, Italy
9 Departments of Molecular Medicine and of Medicine, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA

Running heads: CLL arises from a multiclonal selection process.

Corresponding author: Davide Bagnara - davide.bagnara@unige.it

Data-sharing statement: Raw data are deposited at SRA (BioProject ID PRJNA673787 and PRJNA807871). Processed and annotated data are available upon request to the corresponding author.

Funding: This work was supported in part by philanthropic contributions to NC and KRR from the Karches Family, The Nash Family Foundation, The Muriel Fusfeld Foundation, and the Jean Walton Fund for Leukemia, Lymphoma, and Myeloma Research, FF was supported from the Italian Ministry of Health (Ricerca Corrente 2022-2024), ANM from the European Union Horizon 2020 research and innovation program under the Marie Sklodowska-Curie

grant agreement No 101023721 and from Fondazione Umberto Veronesi grant agreement No 5365.

Author contribution

*DB and ANM contributed equally as co-first authors.

DB and NC conceived the project and designed the experimental approach. DB, ANM ,NCA, SV, MC and NCA performed experiments and data analyses. JRB, SK, SF, KRR, JEK, JCB, and SLA provided clinical samples. GF, FG and FF offered interpretations and conceptual insights. DB, ANM and MF and NC wrote the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of Interest

JRB has served as a consultant for Abbvie, Acerta/Astra-Zeneca, Alloplex Biotherapeutics, BeiGene, Genentech/Roche, Grifols Worldwide Operations, iOnctura, Kite, Loxo/Lilly, Merck, Numab Therapeutics, Pfizer, Pharmacyclics; received research funding from BeiGene, Gilead, iOnctura, Loxo/Lilly, MEI Pharma, TG Therapeutics. KRR is author of CLL chapter in UpToDate and received Royalties from it. SK is employed at Takeda Pharmaceuticals. JCB is a consultant for Jannsen, Abbvie, BeiGene, and AstraZeneca

Abstract.

Although chronic lymphocytic leukemia (CLL) is diagnosed by identifying a circulating B-cell clone that exceeds $5x10^6$/μL, additional distinct clones (ADC) have been identified in various studies. Notably, the numbers of ADC documented in these studies has increased as the various technologies evolved. To better define the frequency and the characteristics of ADCs in CLL, we used a Next Generation Sequencing (NGS) platform that affords high sequencing depth along with steps that limit overcounting to analyze IGHV-IGHD-IGHJ gene rearrangements in circulating $CD5^+$ B cells from 57 patients. Notably, all patients had at least one ADC, in addition to the clinically relevant clone (CRC). , in 46 patients for whom lymphocyte count data were available, 44 had at least one ADC above the threshold of 1 B cell/μL, and remarkably, the average number of ADCs was 12 per patient. Notably, in two patients, the predominant ADC (pADC) qualified clinically as a separate CLL clone and in the in the remaining cases as low/high-count monoclonal B-cell lymphocytosis (MBL) clones. Moreover, in 11 patients studied longitudinally, pADCs were persistent and often increased in number. ADCs in patients with CLL exhibited fourfold more stereotyped IGHV-IGHD-IGH rearrangements than found in $CD5^+$ B cells from healthy individuals, and IGHV use, somatic mutations, and Ig isotype distribution was similar between pADCs and CRCs. Thus, finding multiple expanded clones within the $CD5^+$ B cells is the rule in patients with CLL, indicating that leukemogenesis is a multiclonal process that likely involves competition among B cells with special BCR features.

Introduction

Chronic lymphocytic leukemia (CLL) is defined as a monoclonal expansion of $CD5^+$ B lymphocytes that exceed a defined threshold[1–4]. However, the monoclonal definition can be violated by identifying another B-cell clonal expansion expressing a different IGHV-IGHD-IGHJ (IGHV-D-J) rearrangement. This has been observed at different frequencies in various studies, i.e., 2%[5], 5%[6], 13%[7], and 24%[8]. The presence of additional rearrangements has also been reported for MBL, generally considered the pre-leukemic precursor of CLL[9].

Knowledge accumulated over the past 20^+ years indicates that the structure of the B-cell receptor (BCR) on CLL clones contributes strongly to the development and clinical course of the disease. For example, the mutation status of the expressed IGHV gene divides leukemic clones into two subsets, IGHV-mutated (M-CLL) and IGHV-unmutated (U-CLL)[10,11], with U-CLLs following a more virulent clinical course than M-CLLs[12,13]. Another prominent BCR structural feature of CLL is the non-random association of IGHV, IGHD, IGHJ genes that lead to the presence of remarkably similar VH CDR3 amino acid sequences by different clones[14–17] referred to as "stereotyped BCRs". Moreover, patients whose clones express stereotyped BCRs can exhibit similar characteristic genomic alterations, and experience comparable clinical courses and outcomes[18]. Furthermore, CLL BCRs can signal autonomously, i.e., without engaging antigens extrinsic to the membrane IG, because of BCR homo-dimerization[19]. Finally, specificity studies have indicated that many BCRs display self- and poly-reactivity[4]. These collective findings and the significant clinical benefit from the use of small molecule inhibitors of BCR signalling[20] have led to the conclusion that BCR structure and function are intimately involved in the development of the disease.

In principle, CLL originates from a mature B-cell clone that progressively accumulates a series of transforming mutations, associated with ongoing BCR stimulation. This process could start at an earlier stage of haematopoiesis or B-cell maturation, even prior to the occurrence of IGHV-(D)-J gene rearrangement[21–23] due to a somatic event or an inherited propensity, the latter implied by the strong familial association of CLL[24]. In this situation, many B cells, each with a distinct IGHV-D-J gene rearrangement, would have the opportunity to enter the leukemogenic process, with a series of events eventually favoring the emergence of a dominant, clinically relevant clone (CRC), with other clonal expansions being fingerprints of yet unachieved, and possibly ongoing, transformation.

In this study, we examined the IGHV-D-J gene repertoire of the entire $CD5^+$ B-cell population in 57 patients with CLL, employing a robust NGS protocol that we developed[25]. Our findings show that the $CD5^+$ B cells from CLL patients contain additional clones (ADCs) distinct from the CRC, and the BCRs of the predominant ADCs share features with the BCR repertoire in CLL. Thus, our data indicate that the development of CLL is initiated as a multiclonal process with the dominant clones being select for structural features of the BCRs in CRCs. Our findings also support the notion that leukemogeneisis can begin at earlier stages of B-cell maturation.

Methods

***Samples*:** The study was approved by the Institutional Review Board of Northwell Health in accordance with the Declaration of Helsinki. Written, informed consent was obtained before blood collection from patients with CLL. PBMCs from CLL patients were separated by density gradient centrifugation (Ficoll, GE Healthcare), cryopreserved (10% DMSO 45% FBS and 45%

RPMI), and stored in liquid nitrogen until used. IGH repertoire data derived from the PBMC of 9 aged-matched healthy donors[26], were used as normal controls. A summary of the control data are provided in Supplementary Table 1.

Cell Sorting: Patient PBMCs were incubated with the following anti-human Abs: FITC anti-CD19 (BD Biosciences) and PerCPcy5.5 anti-CD5 (BioLegend). Non-B cells were excluded with efluor-450 anti-CD3 and anti-CD16, and dead cells by Sytox Blue staining (ThermoFisher). Up to 25,000 B cells from each donor were sorted directly into a single 200 µl PCR tube containing 100µl Dynabeads Oligo(dT) (ThermoFisher) in lysis buffer and stored at -80C.

***Library preparation and sequencing*:** Library preparation and subsequent sequencing were performed as described[25]. Briefly, mRNA was isolated from B-cell lysates in 96-well plates using Dynabeads Oligo(dT) (ThermoFisher) according to the manufacturer's protocol. All the bead-immobilized mRNA was reverse transcribed and second-strand synthesis was performed in solid phase using IGHV leader-specific primers; in this step, 13 to 16 random nucleotides (Unique Molecular Identifier - UMI) and partial Illumina adaptor sequences were incorporated. This allowed exponential PCR amplification using universal primers minimizing IGHV gene specific bias. Additionally, the UMIs, which tagged individual mRNA molecules, enabled building consensus sequences and allowing advanced error correction as well mitigating PCR amplification bias. Collectively, these improved the accuracy of quantitative analyses. The PCR product was indexed with Nextera XT kit (Illumina). The library was sequenced with MiSeq Illumina (v3 2 x 300 kit Illumina MS-102-3003) after loading at 12pM with 10% PhiX. The list of the primers and detailed protocol are described in Vergani et.al.[25]. Raw data are deposited at SRA (BioProject ID PRJNA673787 and PRJNA807871).

Bioinformatic analysis: Processing of raw reads was performed using a custom workflow built with pRESTO (REpertoire Sequencing TOolkit)[27]. The obtained IGHV-D-J sequences were submitted to IMGT/HighV-QUEST[28] and analyzed using ChangeO[29], and custom R scripts. Clonal assignment was performed with ChangeO with a distance threshold of 0.07.

CLL stereotype assignment. CLL stereotyped sequences in the ADCs were identified by submitting IGHV-D-J sequences to ARResT/AssignSubsets[30]. Sequences assigned to CLL stereotyped subsets with a confidence of "average" or higher were used in the analyses as was done for the control data in Bagnara et al.[31].

Results

Experimental design. PBMCs were collected from 57 untreated patients with CLL. For 11/57 patients, cells were available at 2 additional time points. $CD19^{+}CD5^{+}$ cells, purified from the PBMCs of each patient or from 9 aged-matched healthy donors[31], were used to determine the full length IGHV-D-J sequence as described[25]. The library preparation employed the use of UMIs and universal primers for the exponential phase of the PCR amplification, minimizing possible IGHV gene specific bias and the effect of PCR amplification for quantitative analysis (see Methods). Each biological sample was aliquoted, and the library prepared in 2 to 8 replicates except for sample CLL1752 (Supplementary Table 2). Only in-frame and productive IGHV-D-J rearrangements were selected for analysis. We included in the analysis only unique sequences considered of high quality, observed in ≥3 different UMIs and derived from the consensus of ≥5 reads. Each unique IGHV-D-J sequence was assigned to a clone, defined as a collection of sequences carrying the same

IGHV-D-J rearrangement (see Methods) including subclonal variants possibly derived by somatic hypermutation. The size of each clone was estimated by the count of mRNA transcripts, inferred from the UMI count. An average of >2,500,000 reads and >400,000 UMI per CLL patient sample was obtained after raw reads processing and quality filter. A summary of the IGHV-D-J repertoire data are provided in Supplementary Table 2.

Presence and frequency of clones distinct from the CRC. Two groups of IGHV-D-J sequences were identified in each CLL patient. The first group, invariably the most abundant in every instance, corresponded to the sequence of the CRC, identified at diagnosis using Sanger methodology; this included subclonal variants differing in various somatic mutations[26]. The other group consisted of a collection of ADCs bearing IGHV-D-J rearrangements that were different from the CRC and its descendants. Here, we have focused on the latter group.

Notably, ADCs were detected in every CLL patient at least at one time point (**Figure 1A**), with an average frequency of 4.8% (0.002 - 73%) relative to the total number of IGHV-D-J transcripts identified (thus including the CRC). The ADC group includes all clones regardless of their level of expansion. Additionally, in each patient the predominant ADC ($pADC_1$) was identified (**Figure 1B**); among the patients, the pADC1s were found at an average frequency of 5.5% (0.012 - 46%) (see Methods).

For 46 patients, white blood cell counts, lymphocyte counts, and flow cytometry data ($CD5^+CD19^+$ phenotype) were available, allowing us to calculate the absolute numbers of cells in the three most dominant ADCs ($pADC_1$, $pADC_2$, and $pADC_3$) by multiplying the frequency of each individual clone by the count of $CD5^+CD19^+$ cells (cell/µl) in the patient's blood. We used

a cut-off for clonal expansion of >1 B cell/µl, a threshold chosen because it represents the median value of B-cell expansion observed in MBL[32]. The $CD5^+$ B cells from the 9 healthy individuals used as control did not exceed this level (**Figure 1C**), whereas 96% (44 out of 46) of the patients exhibited at least one expanded ADC above 1 B cell/µl (**Figure 1D**), with an average of 12 expanded ADCs per patient, ranging from 1 to 120 ADCs. Overall, we observed a correlation between the levels of ADC expansion and their consistent presence in the replicates, although this was not always true in the samples with lower sequencing depths.

Comparison of ADC size with those of CLL and MBL. The diagnosis of CLL, of high-count MBL, and of low-count MBL requires > 5,000, 500 - 5,000, and ~1 - 499 monoclonal B cells/µl[32–34], respectively. In 2 of the 44 patients with expanded ADCs, the size of $pADC_1$ qualified as a second CLL clone by these criteria (5%), as high-count MBL in 8 patients (18%), and as low-count MBL in 34 patients (75%) (**Figure 1C**). Extending the analysis to $pADC_2$ and $pADC_3$ and considering the total number of different time points used for the analysis (see below) indicated that 80% (37/46) of the patients had at least 3 MBL-like expansions at some point during the intervals studied.

Variations of the pADC pattern at different time points. Since samples taken at 2 different time points were available for 11 patients, we next determined if $pADC_1$ frequency changed over time. In ~82% (9/11) of patients, the $pADC_1$ remained the dominant clone over time (**Figure 2A,** black lines), increasing in absolute count in most cases (7/9, 88%). In 22% (2/9) of these patients, the $pADC_1$ decreased with time with one $pADC_1$ falling below the diagnostic level for CLL (**Figure 2A,** dashed black line). Notably, in 2 out of

the 11 patients investigated (18%, dashed red lines), the $pADC_1$ was replaced by a different clone at the second time point.

Moreover, in 3 patients (27%), the $pADC_1$ was and remained the only ADC (**Figure 2B,** overlapping blue dashed line at or near 100%), so in these cases there was a single ADC. Notably, for these patients, the $pADC_1$ had the highest cell/µl count (**Figure 2,** green circle). Also, in 2 of these patients, the absolute count of the $pADC_1$ decreased over time (**Figure 2A,** dashed black line). In 6/11 (54%) patients, the $pADC_1$ remained the major clone, although it increased in relative size (**Figure 2B,** black solid line). However, in the remaining two patients, 18%, (**Figure 2B, -** red dashed line), the $pADC_1$ was outstretched by a different clone at the second time point.

The degree of $pADC_1$ expansion is independent of CRC size. To determine if there was a numerical relationship between the CRC and the $pADC_1$, we correlated the cell counts of the CRCs with the cell counts of the $pADC_1$ (**Figure 3**). Notably, Pearson's correlations failed to demonstrate relationships between $pADC_1$ and CRC, indicating that the level of $pADC_1$ expansion did not directly relate to the size of the CRC (**Figure 3**).

Comparison between CRC and ADC IGHV-D-J rearrangements. To understand whether the IGHV-D-J rearrangements in ADCs shared features with those found in CLL, we qualitatively compared the two groups for IGHV gene use and mutation status, IG isotype display, and the presence of CLL stereotyped IGHV-D-J rearrangements. To do so, when possible, we divided the ADCs into two groups based on clonal frequency: a group with the three most represented pADC ($pADC_{1-3}$) observed in each patient and the remaining referred to as minor ADCs (mADCs)

IGHV gene use. In CLL, IGHV1-69 and IGHV4-34 are two of the most frequently used IGHV genes, with IGHV1-69 being expressed predominantly in U-CLL, and IGHV4-34 more often in M-CLL[11,35]. First, we analyzed the frequencies and characteristics of these two genes detected in $CD5^+$ cells of normal donors and in ADCs and then compared these to those found in a collection of 1,497 CLL cases from our laboratory's database. In contrast to other IGHVs, M-IGHV4-34 and U-IGHV1-69 were present at much lower frequencies in control $CD5^+$ B cells (**Figure 4A,** open dots and white bars) than in CLL, making them outliers in the gene frequency correlation. Likewise, the frequency of use of M-IGHV4-34 and U-IGHV1-69 among mADCs (**Figure 4B,** light blue square dots and bars) diverged considerably from that of CLL. In contrast, $pADC_{1-3}$ were closer to CLL clones in this regard. (**Figure 4A-B,** dark blue triangle dots and bars).

IGHV mutation status. Our internal database of CLL sequences exhibit a median IGHV mutation frequency of about 1.7%, while this frequency was only 0.7% in $pADC_{1-3}$ and virtually null in the mADC and $CD5^+$ normal B cells (**Figure 4C**). IGHV-unmutated and IGHV-mutated sequences, using the 2% mutation threshold, were equally distributed among CLL (52% and 48%) as well as in $pADC_{1-3}$ (55% and 45%) (**Figure 4D**). In contrast, mADCs and $CD5^+$ B cells from healthy controls were unbalanced towards IGHV-unmutated rearrangements (72% and 28% for mADCs, and 84% and 16% for $CD5^+$ controls) (**Figure 4D**). Thus, there were more IGHV-mutated sequences and sequences with higher numbers of mutations in CLL and the $pADC_{1-3}$ compared to mADC and $CD5^+$ B cells from normal controls. Thus, the pADCs were more like CLL using this parameter, and the mADC more like healthy $CD5^+$ B cells.

Although there was a statistically significant concordance in IGHV-mutation status in the CRC and the $pADC_1$ from the same patient (**Figure 4E,** $P<0.007$), nevertheless, an association between IGHV, IGHD, IGHJ use and VH CDR3 sequence was not observed (**Supplementary Table 3**).

IG constant region use. Notably, most rearrangements from all the ADC and $CD5^+$ normal B cells utilized the IgM isotype, like CLL cells. However, a substantial proportion of IGHV4-34 sequences were isotype class switched to IgG in both CLL and $pADC_{1-3}$ (**Figure 4D**). Most IgA were detected in the mutated mADCs, without an apparent enrichment for a specific IGHV gene.

CLL stereotype-like sequences. Twenty-one of the total 5,846 ADC IGHV-D-J rearrangements (0.36%) were assigned to one of the major CLL stereotyped subsets (**Figure 5**). Notably, this frequency was 4-times higher than that in the repertoires from normal $CD5^+$ B cells (0.09%) (**Figure 5A** and ref[31]). Moreover, the stereotyped BCR frequency in the ADCs from the same CLL cases was considerably less than that observed in the CRC (17%; 10/57) (**Figure 5B**). Notably, pADCs were not further enriched in CLL stereotypes (not shown), possibly because of the relatively small cohort of sequences available for analysis. Finally, 2 CLL patients had B cells with a stereotyped rearrangement in both CRC and the ADCs.

Thus, the relative frequency of stereotyped rearrangements increases considerably when moving from a normal $CD5^+$ B-cell population to the ADC population to the CRC.

When counting both CRC and ADCs, we identified the presence of more than one CLL stereotyped IGHV-D-J rearrangement in 4 patients (**Figure 5B**). For 3 (75%) of these patients, at least 2 IGHV-D-J rearrangements were assigned to the same stereotyped subset exclusively in the ADCs or in the ADCs and CRC.

Collectively, the data suggest a selection process favoring the expansion of B cells bearing BCRs which have features shared by the BCRs of the CRCs in the same and other patients.

Discussion

In this study, we used NGS to analyze the IGHV-D-J gene rearrangements of circulating $CD5^+$ B cells from patients with CLL. This revealed the presence of multiple expanded clones, including the CRC, in all patients, although the levels of expansion of these clones varied, indicating that multi-clonality is a common, possibly universal occurrence in CLL and not a feature of only a subset of patients.

Using a 1 B cell/µl cutoff to define an expanded B-cell clone, ~96% of the patients with CLL exhibited at least one ADC and 89% had more than one ADC (**Figure 1C**). Notably, in two cases, the size of pADCs met the criteria defining CLL and in the remaining cases achieved the definition for high- or low-count MBL (**Figure 1D**). These findings place a new perspective on the frequency of multiple expanded clones, independent of the CRC, in patients with CLL. In previous studies, where a second rearrangement was found[5–8], the additional clone could have been the most predominant among a large series of different clonal expansions that were not detected because of the sensitivity of the technology available at the time. Consistent with this possibility, a retrospective NGS analysis revealed the presence of a single, and sometimes of two distinct, expanded IGHV-D-J rearrangements in a few individuals several years prior to the onset of high count MBL or CLL[36].

Our experimental approach minimized the likelihood of over-counting or counting inappropriate variants such as non-productive rearrangements. The latter is relevant for the possibility of the occurrence of allelic inclusion. Although B cells can potentially produce two rearranged IGHV-D-Js, they normally express one. This phenomenon, called allelic exclusion, occurs because only one of the two allelic loci located on chromosome 14 is productively rearranged, while the other remains in the germ line configuration, is partially rearranged (IGHD-IGHJ) or is rearranged non-productively. The

latter possibility raises the question of whether non-productive rearrangements originating from the CRC, as can occur in CLL[37], could be erroneously interpreted as ADCs in our analyses. However, since we excluded non-productive rearrangements, allelic inclusion is not a possibility. Moreover, the presence of numerous ADCs per case (**Figure 1D**) cannot be explained based on the lack of allelic exclusion in the CRCs, especially since allelic inclusion occurs in only 10% of CRCs[37]. In addition, in our studies there was no correlation between the size of the CRC and the presence of $pADC_1$, and changes in $pADC_1$ size over time did not relate to those of CRC, both suggesting independence between the CRC, the $pADC_1$ and probably other ADCs. However, it is possible that other normal/pre-leukemic B cells could lack allelic exclusion and produce two distinct Ig H chains, and our bulk sequencing approach would not distinguish this. To do so, single cell analysis are needed. Notably, our conclusions are consistent with the findings of other investigators[6,7] who used single cell analyses or subclonal fractionation based on kappa or lambda chain expression.

The presence of multiple ADCs might relate to clinical observations in patients with CLL, who develop Richter's Transformation to diffuse large B cell lymphoma or more rarely Hodgkin disease[38,39]. In ~20% of these instances[38,39], the IGHV-D-J rearrangement of the RT clone differs from that of the CLL CRC, suggesting that such lymphomas originated from a distinct ADC present in the patient's clonally expanded B-cell repertoire that was not recognized, most likely based on the sensitivity of the assays used[40–43]. Although these reports are each based on a limited number of cases, in total they support this possibility.

The analysis of IGHV gene use, mutation status, and IG isotype expression indicated that pADCs, more than mADC, exhibit characteristics typical for the repertoire of CLL clones (**Figure 6A**). This was highlighted when focusing on $pADC_{1-3}$ using the IGHV1-69 and IGHV4-34 genes. This showed a distribution of IGHV-mutated and IGHV-unmutated rearrangements, and an isotype use like that observed in the CLL repertoire, which differed substantially from that observed in mADC and in $CD5^+$ B cells from healthy individuals. Moreover, pADC utilizing IGHV4-34 genes were frequently switched to the IgG isotypes as in CLL. Since IGHV4-34-bearing B cells are usually excluded from germinal centres in healthy subjects[44], the presence of isotype switched cells bearing somatically mutated IGHV4-34 is consistent with a defect in B-cell tolerance in people who develop CLL[44,45]. Finally, ADCs showed biased expression of stereotyped rearrangements that is ~4 times greater than in $CD5^+$ normal B cells. We also found distinct IGHV-D-J rearrangements belonging to the same stereotyped subset in the same patient; indeed, in one case (CLL1046) 5 different clones assigned to subset 1 were identified. additionally, IGHV-mutated rearrangements were more abundant in pADC and CRC than in mADC and normal $CD5^+$ B cells. Cells from M-CLL are less apoptosis prone when stimulated via BCR in a variety of experimental conditions *in vitro*, potentially mimicking what occurs *in vivo*[46–50].

Finally, was a significant concordance in IGHV mutation status between CRC and $pADC_1$ in individual cases, suggesting that the trajectories leading to their selection were similar, albeit differing for IGHV-mutated and unmutated cases. Notably, it has been reported that finding ADCs differing from CRC in IGHV mutation status might influence the patient clinical course and outcome[8]. This was observed in subgroups of patients classified as having ADCs by NGS methodologies of lesser sequencing depths.

Collectively, these findings suggest a selection for specific BCR stereotypes resembling those of full-blown CLL or an inherited propensity to generate stereotypes. Additional studies will be needed to corroborate these possibilities, since our findings are liited to a low number of observations. These shared features also strongly suggest that ADCs are part of the leukemogenic process, which appears to involved defined steps (**Figure 6A**): [1] an increase of stereotyped sequences in the mADCs appears to be one of the earliest selection events; [2] IGHV gene use, IGHV-mutation status, and IG isotype use attain a CLL-like fingerprint in the pADCs, concomitant with clonal expansion. The mechanisms behind this IG repertoire modelling are not clear, although selection for BCR structure and signalling are feasible. Additionally, the possibility of molecular events involving pre-BCR rearrangements in early B-cell progenitors or hematopoietic stem cells, which may promote multiclonal B-cell expansions, need to be considered.

In total, our findings strongly suggest that the genesis of CLL involves multiple clones which compete for survival/expansion, with the BCR playing a role in the process. This process could include selection for distinct abilities to respond to certain antigenic stimuli experienced by the competing B-cell clones. The ultimate complete transformation is likely facilitated by genetic alterations, either inherited or acquired early and/or during development, that promote clonality, as well as by microenvironmental inputs that contribute to the survival/expansion of the transforming B cells.

References

1. Rai KR, Jain P. Chronic lymphocytic leukemia (CLL)-Then and now. Am J Hematol. 2016;91(3):330-340.

2. Hallek M. Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment. Am J Hematol. 2019;94(11):1266-1287.

3. Stevenson FK, Forconi F, Kipps TJ. Exploring the pathways to chronic lymphocytic leukemia. Blood. 2021;138(10):827-835.

4. Bagnara D, Mazzarello AN, Ghiotto F, et al. Old and New Facts and Speculations on the Role of the B Cell Receptor in the Origin of Chronic Lymphocytic Leukemia. Int J Mol Sci. 2022;23(22):14249.

5. Plevova K, Francova HS, Burckova K, et al. Multiple productive immunoglobulin heavy chain gene rearrangements in chronic lymphocytic leukemia are mostly derived from independent clones. Haematologica. 2014;99(2):329-338.

6. Brazdilova K, Plevova K, Francova HS, et al. Multiple productive IGH rearrangements denote oligoclonality even in immunophenotypically monoclonal CLL. Leukemia. 2018;32(1):234-236.

7. Kriangkum J, Motz SN, Mack T, et al. Single-Cell Analysis and Next-Generation Immuno-Sequencing Show That Multiple Clones Persist in Patients with Chronic Lymphocytic Leukemia. PloS One. 2015;10(9):e0137232.

8. Stamatopoulos B, Timbs A, Bruce D, et al. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. Leukemia. 2017;31(4):837-845.

9. Klinger M, Zheng J, Elenitoba-Johnson KSJ, Perkins SL, Faham M, Bahler DW. Next-generation IgVH sequencing CLL-like monoclonal B-cell lymphocytosis reveals frequent oligoclonality and ongoing hypermutation. Leukemia. 2016;30(5):1055-1061.

10. Hashimoto S, Dono M, Wakai M, et al. Somatic diversification and selection of immunoglobulin heavy and light chain variable region genes in IgG+ CD5+ chronic lymphocytic leukemia B cells. J Exp Med. 1995;181(4):1507-1517.

11. Fais F, Ghiotto F, Hashimoto S, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. J Clin Invest. 1998;102(8):1515-1525.

12. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood. 1999;94(6):1840-1847.

13. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 1999;94(6):1848-1854.

14. Ghiotto F, Fais F, Valetto A, et al. Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. J Clin Invest. 2004;113(7):1008-1016.

15. Widhopf GF, Rassenti LZ, Toy TL, Gribben JG, Wierda WG, Kipps TJ. Chronic lymphocytic leukemia B cells of more than 1% of patients express virtually identical immunoglobulins. Blood. 2004;104(8):2499-2504.

16. Messmer BT, Albesiano E, Efremov DG, et al. Multiple Distinct Sets of Stereotyped Antigen Receptors Indicate a Role for Antigen in Promoting Chronic Lymphocytic Leukemia. J Exp Med. 2004;200(4):519-525.

17. Tobin G, Thunberg U, Karlsson K, et al. Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. Blood. 2004;104(9):2879-2885.

18. Agathangelidis A, Chatzidimitriou A, Gemenetzi K, et al. Higher-order connections between stereotyped subsets: implications for improved patient classification in CLL. Blood. 2020;137(10):1365-1376.

19. Minden MD, Wossning T, Surova E, Follo M, Köhler F, Jumaa H. Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. Nature. 2012;489(7415):309-312.

20. Burger JA. Treatment of Chronic Lymphocytic Leukemia. N Engl J Med. 2020;383(5):460-473.

21. Marsilio S, Khiabanian H, Fabbri G, et al. Somatic CLL mutations occur at multiple distinct hematopoietic maturation stages: documentation and cautionary note regarding cell fraction purity. Leukemia. 2018;32(4):1041-1044.

22. Damm F, Mylonas E, Cosson A, et al. Acquired Initiating Mutations in Early Hematopoietic Cells of CLL Patients. Cancer Discov. 2014;4(9):1088-1101.

23. Agathangelidis A, Ljungström V, Scarfò L, et al. Highly similar genomic landscapes in monoclonal B-cell lymphocytosis and ultra-stable chronic lymphocytic leukemia with low frequency of driver mutations. Haematologica. 2018;103(5):865-873.

24. Goldin LR, Slager SL, Caporaso NE. Familial chronic lymphocytic leukemia. Curr Opin Hematol. 2010;17(4):350-355.

25. Vergani S, Korsunsky I, Mazzarello AN, Ferrer G, Chiorazzi N, Bagnara D. Novel Method for High-Throughput Full-Length IGHV-D-J Sequencing of the Immune Repertoire from Bulk B-Cells with Single-Cell Resolution. Front Immunol. 2017;14(8):1157.

26. Bagnara D, Tang C, Brown JR, et al. Post-Transformation IGHV-IGHD-IGHJ Mutations in Chronic Lymphocytic Leukemia B Cells: Implications for Mutational Mechanisms and Impact on Clinical Course. Front Oncol. 2021;25(11):640731.

27. Heiden JAV, Yaari G, Uduman M, et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics. 2014;30(13):1930-1932.

28. Alamyar E, Giudicelli V, Li S, Duroux P, Lefranc M-P. IMGT/HighV-QUEST: the IMGT web portal for immunoglobulin (Ig) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. Immunome Res. 2012;8(1):26.

29. Gupta NT, Heiden JAV, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics. 2015;31(20):3356-3358.

30. Bystry V, Agathangelidis A, Bikos V, et al. ARResT/AssignSubsets: a novel application for robust subclassification of chronic lymphocytic leukemia based on B cell receptor IG stereotypy. Bioinformatics. 2015;31(23):3844-3846.

31. Bagnara D, Colombo M, Reverberi D, et al. Characterizing Features of Human Circulating B Cells Carrying CLL-Like Stereotyped Immunoglobulin Rearrangements. Front Oncol. 2022;23(12):894419.

32. Strati P, Shanafelt TD. Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. Blood. 2015;126(4):454-462.

33. Karube K, Scarfò L, Campo E, Ghia P. Monoclonal B cell lymphocytosis and “in situ” lymphoma. Semin Cancer Biol. 2014;24:3-14.

34. Rawstron AC, Ssemaganda A, Tute R de, et al. Monoclonal B-cell lymphocytosis in a hospital-based UK population and a rural Ugandan population: a cross-sectional study. Lancet Haematol. 2017;4(7):e334-e340.

35. Johnson TA, Rassenti LZ, Kipps TJ. Ig VH1 genes expressed in B cell chronic lymphocytic leukemia exhibit distinctive molecular features. J Immunol. 1997;158(1):235-246.

36. Kolijn PMM, Hosnijeh FS, Späth F, et al. high-risk subtypes of Chronic Lymphocytic Leukemia are detectable as early as 16 years prior to diagnosis. Blood. 2022;139(10):1557-1563.

37. Rassenti LZ, Kipps TJ. Lack of Allelic Exclusion in B Cell Chronic Lymphocytic Leukemia. J Exp Med. 1997;185(8):1435-1446.

38. Rossi D, Spina V, Gaidano G. Biology and treatment of Richter syndrome. Blood. 2018;131(25):2761-2772.

39. Mao Z, Quintanilla-Martinez L, Raffeld M, et al. IgVH Mutational Status and Clonality Analysis of Richter’s Transformation. Am J Surg Pathol. 2007;31(10):1605-1614.

40. Salvetti C, Vitale C, Griggio V, et al. Case Report: Sequential Development of Three Mature Lymphoid Neoplasms in a Single Patient: Clonal Relationship and Molecular Insights. Front Oncol. 2022;12:917115.

41. Don MD, Casiano C, Wang H-Y, Gorbounov M, Song W, Ball ED. A Rare Case of Richter Transformation to Both Clonally Unrelated and Clonally Related Diffuse Large B□Cell Lymphoma in the Same Patient. Case Rep Hematol. 2024;2024:7913296.

42. Cheah CY, Spagnolo D, Frost F, Cull G. Synchronous biphenotypic Richter syndrome with Epstein-Barr virus□positive nodal classical Hodgkin lymphoma and bone marrow diffuse large B□cell lymphoma. Histopathology. 2016;69(4):707-710.

43. Bernardo AD, Mussetti A, Aiello A, Paoli ED, Cabras AD. Alternate Clonal Dominance in Richter Transformation Presenting as Extranodal Diffuse Large B-Cell Lymphoma and Synchronous Classic Hodgkin Lymphoma. Am J Clin Pathol. 2014;142(2):227-232.

44. Cappione A, Anolik JH, Pugh-Bernard A, et al. Germinal center exclusion of autoreactive B cells is defective in human systemic lupus erythematosus. J Clin Investig. 2005;115(11):3205-3216.

45. Vergani S, Bagnara D, Agathangelidis A, et al. CLL stereotyped B-cell receptor immunoglobulin sequences are recurrent in the B-cell repertoire of healthy individuals: Apparent lack of central and early peripheral tolerance censoring. Front Oncol. 2023;13(17):1112879.

46. Zupo S, Isnardi L, Megna M, et al. CD38 expression distinguishes two groups of B-cell chronic lymphocytic leukemias with different responses to anti-IgM antibodies and propensity to apoptosis. Blood. 1996;88(4):1365-1374.

47. Chen L, Widhopf G, Huynh L, et al. Expression of ZAP-70 is associated with increased B-cell receptor signaling in chronic lymphocytic leukemia. Blood. 2002;100(13):4609-4614.

48. Lanham S, Hamblin T, Oscier D, Ibbotson R, Stevenson F, Packham G. Differential signaling via surface IgM is associated with VH gene mutational status and CD38 expression in chronic lymphocytic leukemia. Blood. 2003;101(3):1087-1093.

49. Efremov DG, Gobessi S, Longo PG. Signaling pathways activated by antigen-receptor engagement in chronic lymphocytic leukemia B-cells. Autoimmun Rev. 2007;7(2):102-108.

50. Muzio M, Apollonio B, Scielzo C, et al. Constitutive activation of distinct BCR-signaling pathways in a subset of CLL patients: a molecular signature of anergy. Blood. 2008;112(1):188-195.

Figure Legends

Figure 1. Additional Distinct Clones (ADC) quantitative analysis. A. Cumulative frequency of ADCs relative to the total number of IGHV-D-J rearrangements identified in the circulating $CD5^+$ B cells of 57 chronic lymphocytic leukemia (CLL) patients. The values for each patient are expressed by a single dot.

B. Frequency of the predominant ADC ($pADC_1$) among the total number of IGHV-D-J rearrangements including the clinically relevant clone of 57 CLL patients.

C. Absolute cell counts (cells/µl) of the 3 most predominant additional clones ($pADC_1$, $pADC_2$, $pADC_3$) in the 46 CLL patients (black dots) for whom white blood cell counts were available. The plot is divided into 4 horizontal areas based on the interval defining CLL (> 5000 cells/µl), high count monoclonal B-cell lymphocytosis (500 - 5000 cells/µl), and low count monoclonal B-cell lymphocytosis (1 - 499 cells/µl). Open dots indicate data obtained from $CD5^+$ B cells derived from normal donors.

D. Absolute cell counts (cells/µl) of every ADC having >1 cell/µl (44 patients). Data are presented in descending order of the $pADC_1$ expansion of each patient. The numbers on the top of the plot indicate the numbers of ADCs >1 cell/µl in each sample.

Figure 2. Additional Distinct Clones (ADCs) longitudinal analysis. A. Variation in the cell counts (cells/µl) of predominant ADC ($pADC_1$) in 11 chronic lymphocytic leukemia patients over time. The lines connect two samples from the same patient analyzed at different time points. Solid black lines indicate that the $pADC_1$ has the same IGHV-D-J rearrangement in the two samples. A dotted red line indicates that, at the second time point, the $pADC_1$ represented a distinct clone with a different IGHV-D-J rearrangement.

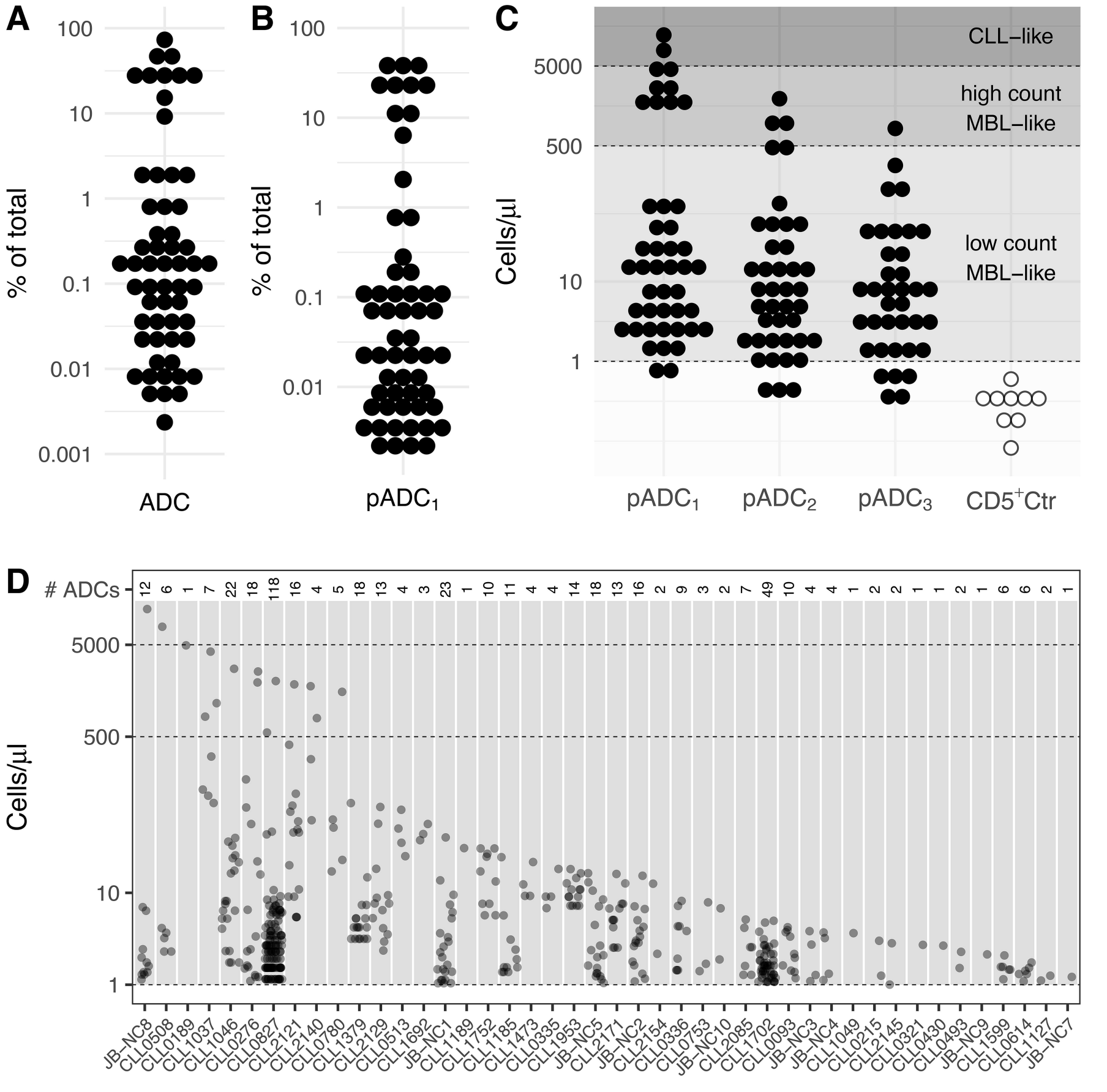
B. Percentage of $pADC_1$ relative to all ADCs at different time points. Lines connect multiple samples from the same patients. Blue dotted line indicates samples where the fraction of $pADC_1$ remained substantially unchanged; solid black line indicates samples where the $pADC_1$ increased over time; red dotted line marks samples in which the $pADC_1$ was a distinct clone, having an IGHV-D-J rearrangement different at the second time point than that seen at the first time point. The green circle highlights the same 3 patients in panel **A** and **B**.

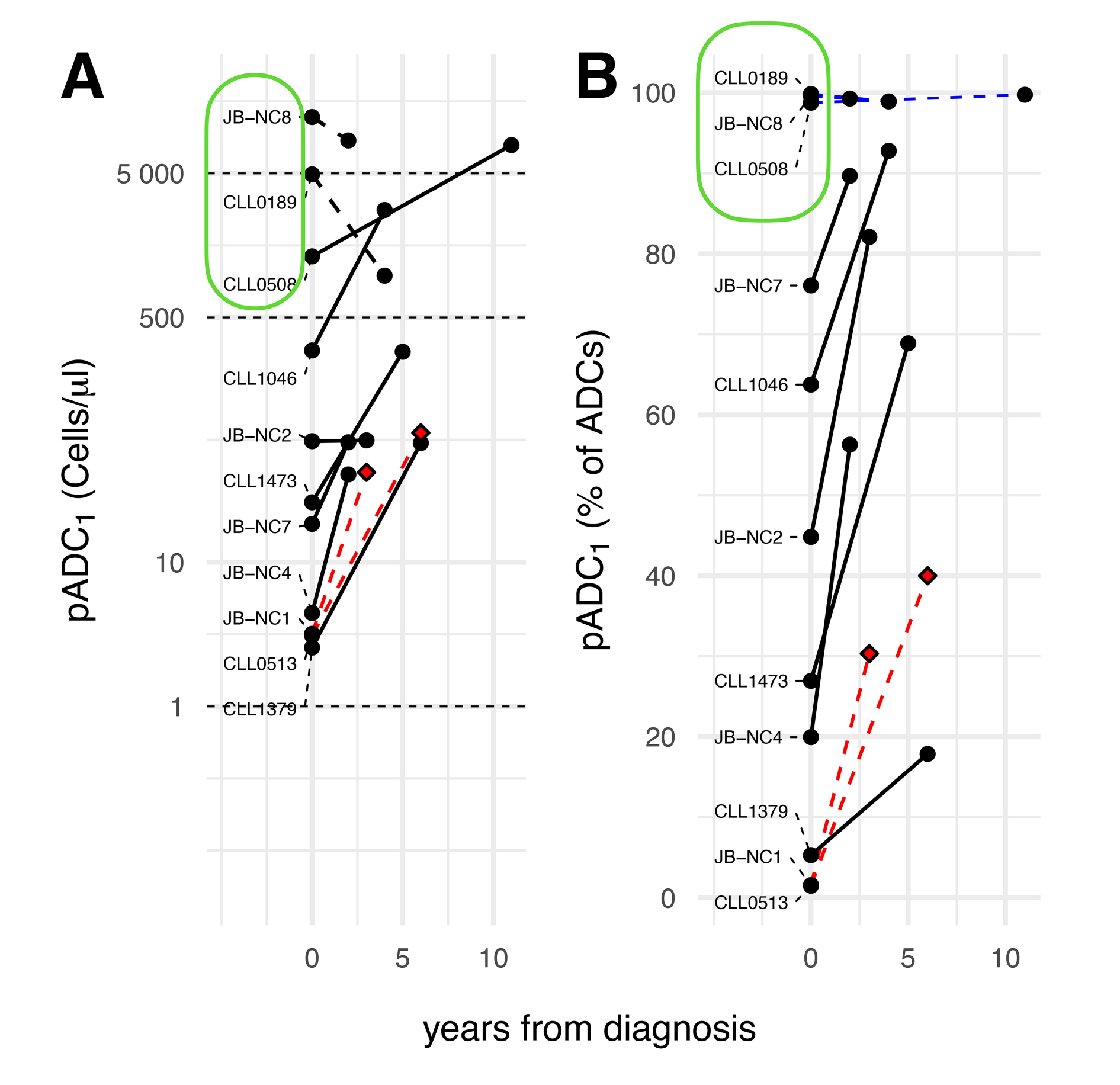
Figure 3. Correlations between Clinically relevant clone (CRC) and $pADC_1$ cell counts (cells/µl) of 46 patients at all available time points. Each dot represents a patient at one time point. The square of Pearson correlation (R^2 <0.01) indicates the absence of correlation between CRC and pADC1 expansion.

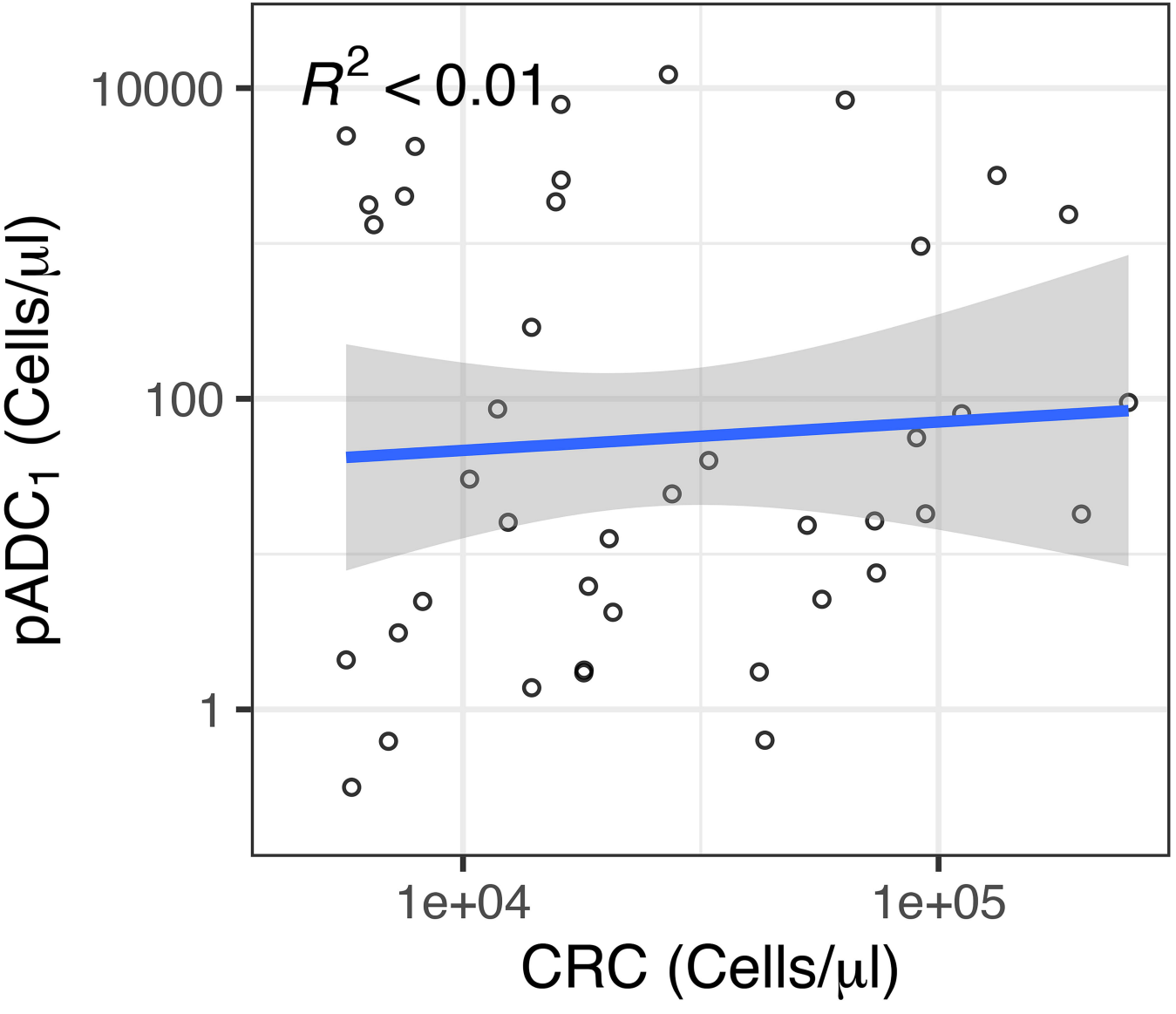
Figure 4. IGHV-D-J repertoire analyses of Additional Distinct Clones (ADCs). A-B. Proportions of clones expressing M-IGHV (**A**) and U-IGHV (**B**) genes. In the lower panels, the x-axes represent the percentages of the indicated populations: dark blue triangles for pADC1-3, light blue squares for mADC, and open circles for $CD5^+$ cells from control subjects. The y-axes show the percentages of chronic lymphocytic leukemia (CLL) clones. The regression line illustrates the correlation between the CLL clones and the other populations. **C.** IGHV mutation distribution in CLL, $pADC_{1-3}$, mADC, and $CD5^+$ B cells from healthy, age-matched controls. In the boxplot, the thick horizontal lines show median values; the box describes the 2 quartiles and the whisker the 95th percentile. **D.** Proportions of mutated (M-) and unmutated (U-) IGHV and IGH isotypes in IGHV4-34 and IGHV1-69 genes in the IGHV-D-J rearrangements from the different clones as indicated. **E.** Association of IGHV-mutation status between the clinically relevant clone and $pADC_1$ (Pearson's chi-squared test, $P < 0.007$).

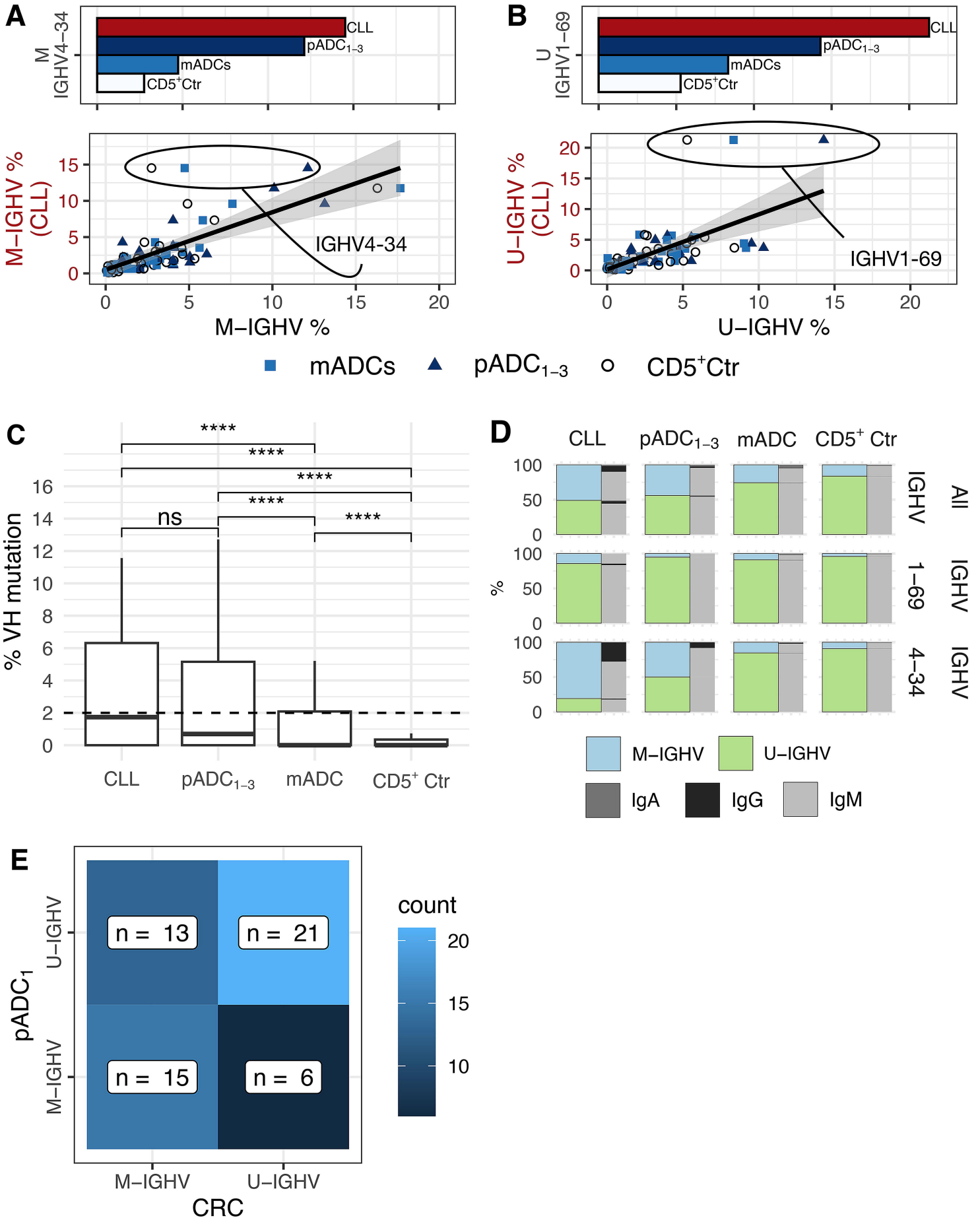
Figure 5. Analyses of the presence of stereotypy in Additional Distinct Clones (ADCs). A. Frequency of chronic lymphocytic leukemia-like stereotyped IGHV-D-J rearrangements detected in ADCs and $CD5^+$ normal B cells from healthy donors; each color indicates a distinct stereotyped subset, as depicted in the graphic legend. **B.** IGHV-D-J clones associated with a chronic lymphocytic leukemia stereotyped subset found in ADCs and clinically relevant clone. Each dot represents a different IGHV-D-J rearrangement, and the colors indicate the stereotyped subset identified.

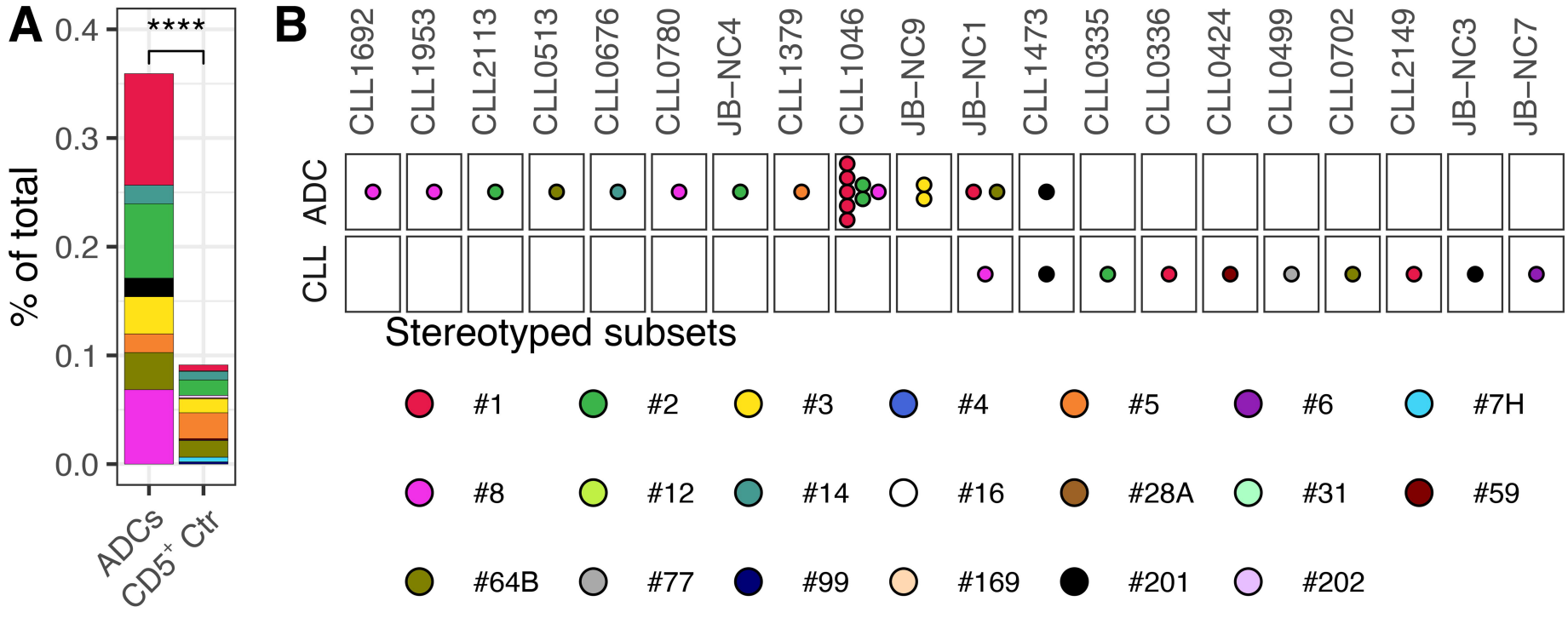
Figure 6. Proposed model. A. Hypothetical steps involved in shaping a patient's $CD5^+$ B-cell repertoire during chronic lymphocytic leukemia (CLL) ontogeny. The major repertoire features and level of clonal expansion are depicted on the left (grey background). The dark blue represents the IGHV-D-J repertoire characteristics of normal $CD5^+$ B cells; dark red represents CLL IGHV-D-J repertoire characteristics; intermediate colors represent the evolution to a CLL-like repertoire. **B.** Hypothetical model of CLL ontogeny in which the initial lesion occurs early in B-cell maturation.

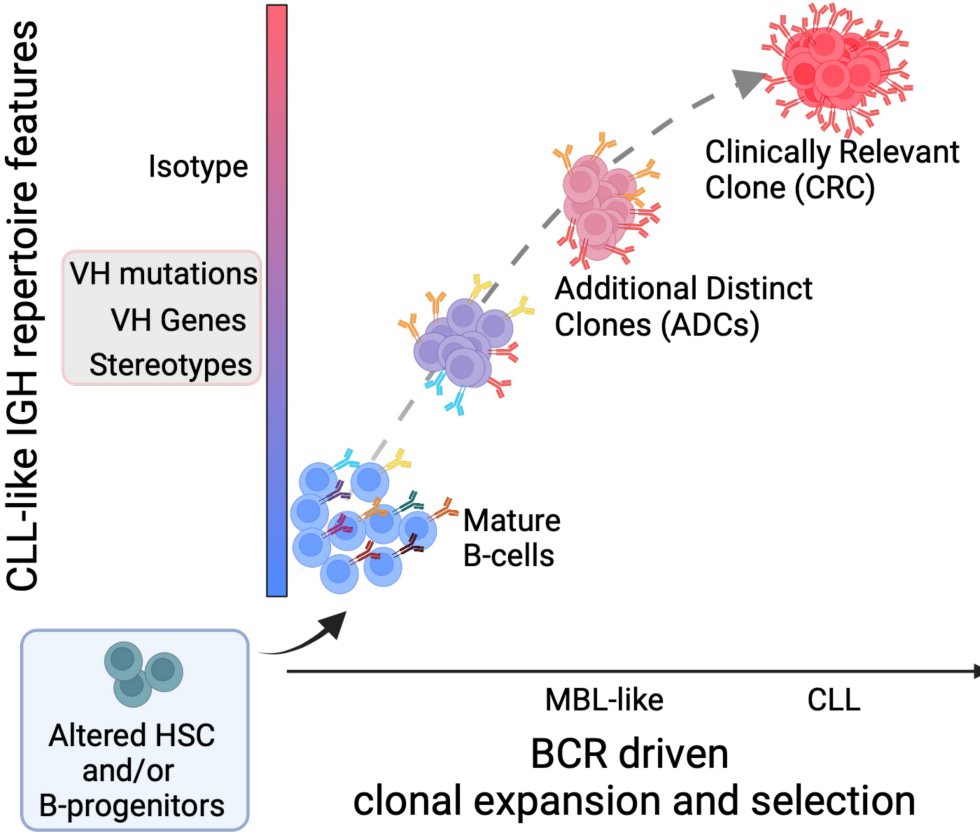

A
B
C
D
% of total
Cells/µl
ADC
pADC1
pADC2
pADC3
CD5+Ctr
CLL–like
high count
MBL–like
low count
MBL–like
ADCs
JB–NC8
CLL0508
CLL0189
CLL1037
CLL1046
CLL0276
CLL0827
CLL2121
CLL2140
CLL0780
CLL1379
CLL2129
CLL0513
CLL1692
JB–NC1
CLL1189
CLL1752
CLL1185
CLL1473
CLL0335
CLL1953
JB–NC5
CLL2171
JB–NC2
CLL2154
CLL0336
CLL0753
JB–NC10
CLL2085
CLL1702
CLL0093
JB–NC3
JB–NC4
CLL1049
CLL0215
CLL2145
CLL0321
CLL0430
CLL0493
JB–NC9
CLL1599
CLL0614
CLL1127
JB–NC7


A
B
JB–NC8
CLL0189
CLL0508
CLL1046
JB–NC2
CLL1473
JB–NC7
JB–NC4
JB–NC1
CLL0513
CLL1379
5 000
500
10
1
pADC1 (Cells/µl)
100
80
60
40
20
0
pADC1 (% of ADCs)
0
5
10
years from diagnosis


$R^2 < 0.01$
10000
100
1
$pADC_1$ (Cells/μl)
1e+04
1e+05
CRC (Cells/μl)


A
M IGHV4–34
CLL
pADC1–3
mADCs
CD5+Ctr
M–IGHV % (CLL)
IGHV4-34
M–IGHV %
B
U IGHV1–69
U–IGHV % (CLL)
IGHV1-69
U–IGHV %
mADCs
pADC1–3
CD5+Ctr
C
% VH mutation
****
ns
CLL
pADC1–3
mADC
CD5+ Ctr
D
CLL
pADC1–3
mADC
CD5+ Ctr
IGHV All
IGHV 1–69
IGHV 4–34
%
M–IGHV
U–IGHV
IgA
IgG
IgM
E
pADC1
U–IGHV
M–IGHV
n = 13
n = 21
n = 15
n = 6
count
CRC


A
B
****
% of total
0.4
0.3
0.2
0.1
0.0
ADCs
CD5+ Ctr
CLL1692
CLL1953
CLL2113
CLL0513
CLL0676
CLL0780
JB–NC4
CLL1379
CLL1046
JB–NC9
JB–NC1
CLL1473
CLL0335
CLL0336
CLL0424
CLL0499
CLL0702
CLL2149
JB–NC3
JB–NC7
ADC
CLL
Stereotyped subsets
#1
#2
#3
#4
#5
#6
#7H
#8
#12
#14
#16
#28A
#31
#59
#64B
#77
#99
#169
#201
#202



Clonal expansion

	CD5⁺ Ctr	mADC	pADC	CRC
Stereotyped BCR				
VH genes usage				
VH mutations				
Isotype-switch				

For Supplementary Tables see Excel files

Supplementary Table 1.
Sequencing data summary of healthy donor IGH repertoire

Supplementary Table 2.
Sequencing data summary of patient with CLL IGH repertoire

Supplementary Table 3.
Summary of CRC and pADC IGHV-D-J sequences