IgH 3’RR recombination uncovers a non-germinal center imprint and c-MYC-dependent IgH rearrangement in unmutated chronic lymphocytic leukemia


Received: February 13, 2023.
Accepted: July 20, 2023.

Citation: Israa Al Jamal, Milène Parquet, Kenza Guiyedi, Said Aoufouchi, Morwenna Le Guillou, David Rizzo, Justine Pollet, Marine Dupont, Mélanie Boulin, Nathalie Faumont, Hend Boutouil, Fabrice Jardin, Philippe Ruminy, Chahrazed El Hamel, Justine Lerat, Samar Al Hamaoui, Nehman Makdissy, Jean Feuillard, Nathalie Gachard, and Sophie Peron. IgH 3’RR recombination uncovers a non-germinal center imprint and c-MYC-dependent IgH rearrangement in unmutated chronic lymphocytic leukemia.
IgH 3′RR recombination uncovers a non-germinal center imprint and c-MYC-dependent IgH rearrangement in unmutated chronic lymphocytic leukemia

Israa Al Jamal1,2, Milène Parquet1*, Kenza Guiyedi1*, Said Aoufouchi3*, Morwenna Le Guillou3, David Rizzo1,4, Justine Pollet1, Marine Dupont1,4, Mélanie Boulin1,4, Nathalie Faumont1, Hend Boutouil1, Fabrice Jardin5, Philippe Ruminy5, Chahrazed El Hamef6, Justine Lerat7, Samar Al Hamaoui2, Nehman Makdissi2, Jean Feuillard1,4, Nathalie Gachard1,4, Sophie Peron16.

Author affiliations:
1 Centre National de la Recherche Scientifique (CNRS) Unité Mixte de Recherche (UMR) 7276/INSERM U1262, Université de Limoges, Limoges, France.
2 Faculty of Sciences, GSBT Genomic Surveillance and Biotherapy Team, Mont Michel Campus, Lebanese University, Tripoli, Lebanon.
3 CNRS UMR9019, Gustave Roussy, B-cell and Genome Plasticity Team, Villejuif, France and Université Paris-Saclay, Orsay, France.
4 Laboratoire d’Hématologie Biologique, Centre Hospitalier Universitaire de Limoges, Limoges, France.
5 Inserm U1245 and Department of Henri-Becquerel Hematology Center and Normandie Univ UNIROUEN, Rouen, France
6 Collection Biologique Hôpital de la Mère et de l’Enfant (CB-HME), Department of Pediatrics, Limoges University Hospital, Limoges, France.
7 Department of Otorinolaryngology, Limoges University Hospital, Limoges, France.
*These authors contributed equally.
§Corresponding author.

Running head: c-MYC and Sµ-3′RRrec denote different COO in CLL

Correspondence should be addressed to: Sophie Peron, CRIBL – Contrôle de la réponse immune B et des lymphoproliférations - UMR CNRS 7276 INSERM1262. Centre de biologie et de recherche en santé, 2 rue du Pr Bernard Descottes, 87 000 Limoges, France. Phone: (+33) 519 564 210; Email: sophie.peron@unilim.fr


Authorship contributions:

IAJ performed experiments and participated in writing of the original draft. MP, KG, SA, MLG, MD, MB and HB participated in the experiments. DR, JP, MD, NF, FJ, PR and NG participated in data curation. CEH and JL provided tonsils from patients undergoing tonsillectomies performed in Limoges Dupuytren hospital. SA, SAH and NM participated in writing the original draft. JF, NG and SP led the conceptualization, data curation, funding acquisition and manuscript writing.

Disclosure of conflicts of interest: The authors declare no competing interests.

Data-sharing statement: Sequencing data produced in this study have been deposited in the National Center for Biotechnology Information’s BioProject (PRJNA830327).

Counts: Abstract words: 252; Main text words: 4308; Figures: 6; Supplementary files: 2 (supplemental methods, supplemental tables and supplemental references; supplemental figures).


Acknowledgements: The authors would like to thank Dr. Villéger at the CRBioLim and Dr. Vallejo and the INSERM CIC 1435, Dupuytren Hospital, Limoges, France, for providing human samples. We thank Dr. Jeanne Cook Moreau (UMR CNRS 7276 INSERM 1276) for careful English editing.

Funding: This work was supported by grants to SP from la Ligue Contre le Cancer (Comité de la Haute-Vienne, CD87), the Fondation ARC pour la recherche sur le cancer (ARCPJA2022060005138), the INCa-Cancéroplôle GSO (Emergence program N°2019-E11) and grants to CRIBL lab from the Institut CARNOT CALYM. IAJ is supported by Fondation pour la Recherche Medicale (FRM) and Lebanese associations (AZM and Saade, LASer). MP is supported by Région Nouvelle Aquitaine and Université de Limoges. KG is supported by Région Nouvelle Aquitaine and Délégation INSERM Nouvelle Aquitaine. JP is supported by the Institut CARNOT CALYM. SP is a National Institute of Health and Medical Research (INSERM) investigator.
ABSTRACT

Chronic lymphocytic leukemia (CLL) is an incurable indolent non-Hodgkin lymphoma characterized by tumor B-cells that weakly express a B-cell receptor (BCR). The mutational status of the variable region (IGHV) within the immunoglobulin heavy-chain (IGH) locus is an important prognosis indicator and raises the question of the CLL cell of origin (COO). Mutated IGHV gene CLLs (mCLLs) are genetically imprinted by activation induced-cytidine deaminase (AID). AID is also required for IGH rearrangements: class switch recombination (CSR) and recombination between switch Mu (Sµ) and the 3' regulatory region (3'RR) (Sµ-3'RRrec). The great majority of CLL B-cells being unswitched led us to examine IGH rearrangement blockade in CLL. Our results separated CLLs into two groups on the basis of Sµ-3'RRrec counts per sample: Sµ-3'RRrec\textsuperscript{High} cases (mostly umCLLs) and Sµ-3'RRrec\textsuperscript{Low} cases (mostly mCLLs), but not based on CSR junction counts. Sµ-3'RRrec appeared to be ongoing in Sµ-3'RRrec\textsuperscript{High} CLL cells and comparison of Sµ-3'RRrec junction structural features pointed to different B-cell origins for both groups. In accordance with IGH mutational status and PIM1 mutation rate, Sµ-3'RRrec\textsuperscript{High} CLLs harbor a non-GC experienced B-cell imprint while Sµ-3'RRrec\textsuperscript{Low} CLLs are from AID-experienced B-cells from a secondary lymphoid organ. In addition to the proposals already made concerning the CLL cell of origin, our study highlights that analysis of IGH recombinatory activity can identify CLL cases from different origins. Finally, on-going Sµ-3'RRrec in Sµ-3'RRrec\textsuperscript{High} cells appeared to presumably be the consequence of high c-MYC expression, as c-MYC overexpression potentiated IGH rearrangements and Sµ-3'RRrec, even in the absence of AID for the latter.
INTRODUCTION

Being one of the most frequent B-cell cancers of the elderly, chronic lymphocytic leukemia (CLL) is characterized by lymphocytosis exceeding 5 G/L, and is composed of small circulating monomorphic round CD19⁺ CD23⁺ CD5⁺ B-cells, as well as bone marrow and secondary lymphoid organ infiltration in most cases. CLL evolution is highly variable, with overall survival ranging from a few years to decades and is still incurable despite the development of new therapeutics such as Bruton tyrosine kinase or Bcl2 inhibitors.

Binet and Rai classifications are still the most reliable staging systems to predict CLL course and are the keystones of clinical decision for treatment. However, patients show marked karyotypic and genetic heterogeneity which also influences overall survival rate and prediction of therapeutic response. Major prognosis factors are chromosomal abnormalities such del17p, del11q, trisomy 12, isolated del13q or complex karyotypes which highlight genomic instability in CLL pathogenesis. Among poor prognosis genetic abnormalities are those involving the Notch pathway (Notch1 mutations), NF-kappa B activation (BIRC3 or MYD88 mutations), splicing (SF3B1), the DNA lesion sensor ATM or the TP53 anti-oncogene.

Underscoring the role of the B-cell receptor in this B-cell cancer, another important prognosis indicator is the mutational status of the variable region (IGHV) within the immunoglobulin (Ig) heavy-chain (IGH) locus, which separates CLL patients into two groups: those with an unmutated variable region and those with a mutated IGHV rearranged gene (umCLLs and mCLLs, respectively). The former bestows a poor prognosis while patients with a very long survival rate are found in the latter group. Enforcing the role of the BCR is the fact that 30% of CLL patients express a so-called "stereotyped receptor", which suggests the role of common antigenic determinants in the promotion of B-cell transformation.

The fact that mCLLs are genetically imprinted by activation induced-cytidine deaminase (AID)-dependent IGHV somatic hypermutation (SHM) raises the question of a CLL group with a post-germinal center B-cell counterpart. Recent methylome analyses suggest proximities between mCLL and germinal center experienced memory B-cells on the one hand and umCLL and naïve B-cells on the other hand. However CLL cells exhibit a unique CD5⁺, CD23⁺, CD27⁺, CD43⁺ with low levels of surface IgM and IgD immunophenotype, which is different from that of any normal B-cell. Gene expression profiles revealed that both CLL groups share a characteristic gene expression signature that is close to that of antigen-experienced B-cells. Reconstruction of B-cell differentiation trajectories indeed suggested that precursors of both umCLL and mCLL have reached the antigen experienced memory B-cell stage. It has also been proposed that CLL may originate from an extrafollicular B-cell response since maturation of these cells is antigen driven and can be either mutated or not. Therefore, more than 20 years after the papers of Hamblin and Damle, this question of
the CLL cell of origin (COO) has not been clearly answered and umCLL could differ from mCLL mainly by expressing BCR-related mitogenic markers.\(^\text{14}\)

The fact that most CLL cells express an IgM raises the question of CSR blockade in this B-cell cancer. Low levels of class switch recombination (CSR) can be observed in a small fraction of CLL tumor cells and seem to correlate with AID expression in these intra-clonal switched CLL cells.\(^\text{15\text{-}17}\) AID has been repeatedly detected in CLL B-cells independently of the IGHV mutational status.\(^\text{15}\) AID likely contributes to CLL evolution and seems to generate intraclonal diversity targeting the IGH locus and non-Ig off-targets.\(^\text{18}\) Physiologically, AID induces CSR in activated proliferating B-cells.\(^\text{19,20}\) CSR results from IGH intrachromosomal recombination between the switch μ (Sµ) region and another switch region (so-called Sx) located upstream of one of the constant genes.\(^\text{21}\) CSR requires double strand DNA breaks (DSBs). Converting a cytosine into uracil, AID creates a U:G mismatch that is targeted by the base excision repair (BER) pathway to excise the uracil base using uracil-DNA-glycosylase (UNG). This results in an abasic site which is processed and generates a single strand DNA break (SSB). Several SSBs on both DNA strands then produce DSBs which are repaired through the ubiquitous DSB repair (DSBR) response. In B-cells, during CSR, DSB repair occurs by the joint action of the non-homologous end Joining (NHEJ) and Alternative End-joining (Alt-EJ) pathways.\(^\text{22,23}\) With XRCC4 for the ligation step, NHEJ depends on 53BP1/Rif1\(^\text{24,25}\) which recruits the Shieldin complex. Both 53BP1/Rif1 and Shieldin complexes are essential for protection against the resection of DNA ends.\(^\text{26\text{-}28}\) Less understood and described, Alt-EJ involves PARP1, POLQ and/or LigIII. These DSBR pathways differently shape the structure of the repair junction; each single junction is thus unique.

Being highly conserved across mammalian species (human, mouse, dog, rabbit ...), the 3' regulatory regions (3'RR) of the IGH locus are key regulatory regions for CSR. 3'RR exhibits a singular structure with three DNaseI hypersensitive (HS) sites (HS3, HS1-2, and HS4) harboring strict specific B-lineage transcriptional enhancer activity related to a “quasi-palindrome” organization where inverted repeated sequences flank the HS1–2 sequence which is the symmetry center element.\(^\text{29}\) The 3'RR can also be recombined with the Sµ region in an AID-dependent manner (Sµ-3'RRrec).\(^\text{30}\) Sµ-3'RRrec has been shown to occur in vitro and in vivo in activated murine and human B-cells.\(^\text{30\text{-}32}\) It has been repeatedly detected in secondary lymphoid organs and peripheral blood mononuclear cells (PBMCs) of both mice and humans. Similar to CSR in activated mature B-cells, Sµ-3'RRrec occurs between transcribed recombination donor and acceptor DNA segments. However, we observed that the structure of Sµ-3'RRrec junctions in murine B-cells was different from that of CSR.\(^\text{31}\) Indeed, Sµ-3'RRrec junctions are reminiscent of the usage of NHEJ and/or Alt-EJ, with regard to the repair signature at Sµ-3'RRrec joints and the recruitment of Alt-EJ but not NHEJ components at the 3'RR locus in mice.\(^\text{31}\) In contrast with CSR, Alt-EJ seems therefore to surface as the major contributor. Therefore, some arguments support the fact that Sµ-3'RRrec is a different IGH recombination from...
that of CSR. When \( \text{S} \mu-3' \text{RRrec} \) hits the \( \text{IGH} \) locus, it results in the excision of the whole cluster of constant \( \text{IGH} \) genes. This should kill BCR expression if occurring on the functional \( \text{IGH} \) allele. Since \text{in vivo} loss of BCR induces B-cell death\textsuperscript{33}, \( \text{S} \mu-3' \text{RRrec} \) was initially called "locus suicide recombination". However, to date, the search for such B-cells lacking BCRs due to \( \text{S} \mu-3' \text{RRrec} \) has failed. Moreover, HTS studies on human BCR positive circulating B-cells revealed that amplicons covering \( \text{S} \mu-3' \text{RRrec} \) junctions certainly came from the non-functional \( \text{IGH} \) allele\textsuperscript{32}. Thus, at present, the function of \( \text{S} \mu-3' \text{RRrec} \) in B-cell physiology remains unclear.

In this study, we raised the question of \( \text{IGH} \) switch blockade in CLL. For that purpose, we analyzed both CSR and \( \text{S} \mu-3' \text{RRrec} \) junctions as reflections of putative switch activity. Our results revealed that CLL patients could be separated into two groups with different prognoses on the basis of \( \text{S} \mu-3' \text{RRrec} \) counts but not CSR junction counts. Cases with increased \( \text{S} \mu-3' \text{RRrec} \) were indeed related to umCLL. Comparison between both groups revealed that \( \text{S} \mu-3' \text{RRrec} \) was likely to be ongoing in tumor cells from CLL patients with increased \( \text{S} \mu-3' \text{RRrec} \) counts. Structural features of \( \text{S} \mu-3' \text{RRrec} \) junctions revealed an imprint that pointed to a different B-cell origin for both groups. Moreover, \( \text{S} \mu-3' \text{RRrec} \) appeared to be potentiated by MYC overexpression even in the absence of AID.

**METHODS**

**Human materials and ethics**

The project was conducted according to the guidelines of the Declaration of Helsinki. CLL Peripheral Blood Mononuclear Cells (PBMCs) were obtained from CRBioLim, CHU Dupuytren, Limoges Hospital (authorization: DC-2008-604, AC-2016-2758, and AC-2019-3418). Tonsils were obtained from children scheduled for elective tonsillectomy from CRBioLim (authorization: DC-2008-604, AC-2018-3157). PBMCs from healthy volunteers (HV) were collected through the research project approved by CPP Sud Méditerranéel (N°2021-A00778-33).

**Human CSR and \( \text{S} \mu-3' \text{RRrec} \) junction counts:**

Human CSR and \( \text{S} \mu-3' \text{RRrec} \) junctions were amplified as described\textsuperscript{32} and used to prepare next generation sequencing (NGS) libraries (Ion Xpress\textsuperscript{™} Plus Fragment Library Kit, Life Technologies, Thermofisher, 447269) sequenced with an Ion Proton or S5 chip (Life Technologies). FastQ were analyzed using CSReport\textsuperscript{35}. CSR and \( \text{S} \mu-3' \text{RRrec} \) junction diversities were estimated through the Shannon diversity index (Supplemental Methods). The Jurkat cell line and naïve B-cells sorted from peripheral blood of healthy donors (N=2) served as negative controls for \( \text{S} \mu-3' \text{RRrec} \) junction detection (no \( \text{S} \mu-3' \text{RRrec} \) junctions were detected).

**CH12F3 CSR and \( \text{S} \mu-3' \text{RRrec} \) junction counts:**

The CH12F3 cells were transfected or not by MYC expression vector (Plasmid#74164, Addgene) and cultured in RPMI1640 with Ultra Glutamine, 10% FCS (Lonza), sodium pyruvate (Lonza), penicillin/
streptomycin (Lonza), non-essential amino acids (Lonza) and β2-ME. Cells were stimulated for CSR toward IgA for 72 hours with murine IL-4 (5 ng/ml; PeproTech), human TGF-β1 (1 ng/ml; R&D Systems), and murine anti-CD40 Ab (1µg/ml; eBioscience). CSR and Sµ-3'RRrec junctions were amplified by nested PCR using specific primers (Supplemental table 3) as described34.

**IGHV sequence analysis:**
V, D, and J rearranged genes were amplified using the Biomed-2 strategy with FR1 and FR2 primers and sequence analyses were performed as described35. The IGHV intra-clonal diversity analysis workflow is described in the Supplemental Methods. DLBCL IGHV sequences are from Rizzo et al.36.

**Flow cytometry analysis:**
Immunophenotyping was done on a Navios-flow cytometer (Beckman Coulter) with the protocol for routine CLL diagnosis using: CD5-APC (Beckman Coulter PN a60790, clone BLLa), CD19-ECD (Beckman Coulter A07770, clone J3-119) and Anti-human Kappa light chain/Anti-Human Lambda light chains/RPE (Dako, FR481 X0935). Results were analyzed with Kaluza software version 2.1 (Beckman Coulter).

**RNA extraction, cDNA synthesis and quantitative real time PCR:**
Total RNA was isolated (TRizol™ Reagent, 15596018) and reverse-transcribed (Advantage RT-for-PCR kit Applied Biosystems™, Thermofisher 4368814/10400745). qRT-PCR were performed with the SYBR-Green PCR mix (SensiFast hi ROX Syber Green BIO820025) and primers (Table S3) or with Taqman PCR mix (SensiFast Probe Hi-Rox kit BIO820025) and MYC probe (4331182 Hs00905030_m1, Thermofisher). Normal centroblasts and naïve B-cells were sorted from tonsils as described37.

**Relative Telomere length assay (RTL):**
DNA (25ng) extracted from PBMCs was used in triplicate to assess RTL by qPCR as described previously38.

**Mutation analysis of PIM1:**
We amplified PIM1 exon 4 (Table S3), containing a CLL AID-targeted nucleotide18, using Phusion High Fidelity Taq (Thermo Scientific, F-530XL). Products were used to build NGS libraries. The analysis workflow is described in the Supplemental Methods.

**Statistical analysis:**
Graphs, histograms, curves, and standard statistical analyses were designed using Graph Pad 6-Rrism software. Fisher tests were done with R (version 4.3.0) using the RStudio interface (RStudio 2023.03.0 Build 386). Kaplan Meyer survival curves and Cox univariate and multivariates analyses were done using the R Survival package (URL: https://github.com/therneau/survival).
RESULTS

CLL patients can be separated into two groups according to Sµ-3'RR recombination

We analyzed counts of Sµ-3'RR recombination (Sµ-3'RRrec) in DNA samples collected at CLL diagnosis from 47 patients. Blood tumor infiltration was over 90% circulating lymphocytes in 42/47 (89%) cases and over 98% in 37/42 cases (79%) (Figure S1A). Comparison of junction counts was performed with results obtained from DNA from peripheral blood mononuclear cells (PBMCs) of 9 healthy volunteers (HVs). As negative controls, we used the Jurkat cell line and naive B-cells sorted from peripheral blood of healthy donors (N=2), in which no Sµ-3'RRrec junctions could be detected (not shown). Even at low levels, Sµ-3'RRRec was found at comparable levels in both HVs and CLLs (Figure S1B), and was undetectable in only 3/47 (6.3%) CLL patients. We separated CLL patients into two groups, using as a threshold value the mean of Sµ-3'RRrec counts in HVs (Table S1), called Sµ-3'RRrec

\[ \text{High} \] (12/47 patients = 26%), and Sµ-3'RRrec

\[ \text{Low} \] (35/47 patients = 74%) (Figure 1A). Analysis of CSR and Sµ-3'RR recombinations was based on a nested-PCR approach. SHM could theoretically introduce mutations in primer binding DNA, particularly in CLL. But here, low levels of Sµ (Mutation rate average +/- SEM for Sµ-3'RRrec

\[ \text{Low} \], Sµ-3'RRrec

\[ \text{High} \] and HV PBMCs respectively are: 1.828 +/- 0.306, 1.475 +/- 0.302 and 0.870 +/- 0.090) and 3'RR2 Sµ (Mutation rate average +/- SEM for Sµ-3'RRRec

\[ \text{Low} \], Sµ-3'RRRec

\[ \text{High} \] and HV PBMCs respectively: 0.048 +/- 0.018, 0.062 +/- 0.020 and 0.056 +/- 0.033) mutation frequency. In DNA segments from Sµ-3'RRRec junctions, absence of significant differences between samples ruled out significant bias in amplification of the Sµ-3'RRrec junctions and comparison between Sµ-3'RRrec

\[ \text{High} \] and Sµ-3'RRrec

\[ \text{Low} \] CLLs. Sµ-3'RRrec counts were not dependent on CLL B-cell richness as shown in Figure S1C. Moreover, the percentages of CLL B-cells were similar in Sµ-3'RRRec

\[ \text{High} \] and Sµ-3'RRRec

\[ \text{Low} \] CLL samples (Figure S1D).

As expected in this IgM + B-cell cancer, CSR junction levels were much lower in CLL than in HV samples. CSR counts were similarly low in both the Sµ-3'RRRec

\[ \text{High} \] and Sµ-3'RRRec

\[ \text{Low} \] CLL groups (Figure 1B). We did not find any significant association between increased Sµ-3'RRrec and CSR counts as shown in Table S2. Moreover, the correlation between CSR and Sµ-3'RRrec counts was poor (correlation coefficient r=0.2, not shown). Thus, Sµ-3'RRRec

\[ \text{High} \] CLLs specifically exhibited increased Sµ-3'RRrec counts when compared to CSR.

Most patients with increased Sµ-3'RRrec counts had an umCLL

To further study Sµ-3'RRRec

\[ \text{High} \] and Sµ-3'RRRec

\[ \text{Low} \] CLLs, we analyzed the IGHV mutational status, an important prognosis indicator of poor outcome. Even if being cautious for small numbers, no significant IGHV gene repertoire bias was found between CLL groups (Figure S2). IGHV clonal rearrangements of Sµ-3'RRRec

\[ \text{High} \] CLL cases exhibited stronger homology to IGHV reference sequences (Figure 2A). With the threshold of 98% homology, 9/12 (75%) Sµ-3'RRRec

\[ \text{High} \] CLLs were not
or only weakly mutated (mean IGHV mutation rate = 98.1%). One additional $\mu$-3'RRrec$^\text{High}$ CLL had 97% IGHV sequence homology with the reference. $\mu$-3'RRrec$^\text{Low}$ cases were mCLL for which 20/35 (57%) cases had a mean IGHV mutation rate = 4.8%, $p=0.043$. Consistently, we found that $\mu$-3'RRrec$^\text{High}$ patients exhibited low rates of AID off-target $PIM1$ mutations (Figure 2B). As CD19 transcription and expression at the cell surface are specific for the B-cell compartment and were similar between $\mu$-3'RRrec$^\text{Low}$ and $\mu$-3'RRrec$^\text{High}$ CLL (Figures S3A and S3B), transcript expression levels were normalized to those of CD19. When compared to centroblasts and naïve B-cells sorted from benign inflammatory tonsils, AID transcript levels were comparable in both $\mu$-3'RRrec$^\text{High}$ and $\mu$-3'RRrec$^\text{Low}$ CLls, being as low as in naïve B-cells, regardless of the mutated or unmutated IGHV status (Figure 2C).

In agreement with the strong predominance of umCLL in this group, $\mu$-3'RRrec$^\text{High}$ CLls were associated with decreased treatment free survival (TFS) ($\approx$14 months compared to $\approx$71 months; $P<0.001$, Figure 2D). In comparison, separating patients into CSR$^\text{Low}$ and CSR$^\text{High}$ groups did not result in significant differential TFS even if survival curves were separated (Figure S1E). For this series, TFS also strongly depended on the Binet stage and the IGHV mutation status, and marginally depended on lymphocytosis and cytogenetics (Figure S5). To search for independent variables, a Cox univariate analysis was first done for $\mu$-3'RRrec status, Binet stage, lymphocytosis, age, IGHV mutation status and cytogenetics (Table 1). A first Cox multivariate model was constructed with variables with a $p$-value < 0.2, id est $\mu$-3'RRrec status, Binet stage, IGHV mutation status, and cytogenetics (Table 1). In this model, IGHV mutation status and Binet stage were the two independent variables. However, a Cox model including only IGHV mutation and $\mu$-3'RRrec status suggested that the confounding variable was the IGHV mutation. Indeed, a second model replacing IGHV mutation status by lymphocytosis pointed on $\mu$-3'RRrec status as the sole independent variable (Table 1). Revealing strong overlaps between IGHV mutation and $\mu$-3'RRrec status in terms of TFS, these analyses reflect the enrichment $\mu$-3'RRrec$^\text{High}$ group in umCLls, which are very well known to have a poor prognosis$^6,7$.

In contrast to poorly diversified IGHV clonal rearrangements, $\mu$-3'RRrec$^\text{High}$ CLls exhibited increases in both $\mu$-3'RRrec and CSR diversities with increased IGH locus accessibility.

$\mu$-3'RRrec and CSR result from random IGH recombination involving two DNA double strand breaks (DSBs), one in the $\mu$ donor region and one in the 3' RR or Sx acceptor region respectively. Since CLL is IgM$^+$, $\mu$-3'RR rearrangements have to occur on the non-productive IGH allele; this raises the question of $\mu$-3'RRrec clonality. Because CSR can also take place on the non-productive allele, both $\mu$-3'RRrec and CSR junction diversities were evaluated using the Shannon index which measures the
number of different junctions in sequencing libraries\textsuperscript{39}. HVs were used here as controls of “polyclonal” junctions. Diversities of both Sµ-3'RRrec and CSR junctions were strongly decreased in Sµ-3'RRrec\textsubscript{Low} CLL samples, a result that should be expected in this clonal IgM\textsuperscript{+} B-cell cancer. Strikingly, the Shannon diversity index was significantly higher in Sµ-3'RRrec\textsubscript{High} CLL samples than in Sµ-3'RRrec\textsubscript{Low} CLLs (Figures 3A and 3B). Both the absolute numbers of B-cells (Figure S4) and diversity indexes of CLL IGHV clonal rearrangements (Figure 3C) were similar in both the Sµ-3'RRrec\textsubscript{High} and Sµ-3'RRrec\textsubscript{Low} groups. CLL IGHV diversity was much lower than diffuse large B-cell lymphomas (DLBCL), known to harbor intra-tumoral subclones with divergent IGH after SHM and taken here as positive controls of intra tumor diversity. Therefore, increased Sµ-3'RRrec and CSR junction diversities were not likely to be influenced by B-cell richness but would rather reflect diversification of heavy chain rearrangements in Sµ-3'RRrec\textsubscript{High} CLLs. Because increased diversities of Sµ-3'RRrec and CSR junctions in Sµ-3'RRrec\textsubscript{High} CLLs are evocative of an on-going process, we evaluated whether IGH locus DNA was accessible to recombination machinery. To assess locus aperture, we analyzed the expression of non-coding and coding transcripts from the constant part of the IGH locus (Figure 4A). We found higher levels of Sµ, Sy1, Sy3, HS1.2 and HS4 sterile transcripts in Sµ-3'RRrec\textsubscript{High} than in Sµ-3'RRrec\textsubscript{Low} CLLs (Figure 4B), meaning that the IGH locus was accessible to the recombination machinery in these patients. While levels of surface Ig were comparable between the two groups (data not shown), coding Cµ transcripts were also increased in Sµ-3'RRrec\textsubscript{High} CLL (Figure 4C). Altogether, our results suggest abnormal intratumoral IGH remodeling activity in the Sµ-3'RRrec\textsubscript{High} CLL group.

In contrast with those of CSR, structural features of Sµ-3'RRrec junctions harbor an activated B-cell imprint and discriminate both Sµ-3'RRrec\textsubscript{Low} and Sµ-3'RRrec\textsubscript{High} CLLs.

In addition to being unique, joint structure is differently shaped according to the DSB repair (DSBR) machinery. In B-cells, DSB repair occurs mainly through NHEJ and, to a lesser extent, through Alt-EJ pathways for IGH recombination\textsuperscript{22,23}. The joint structure of each single Sµ-3'RRrec and CSR junction can be determined by alignment to reference sequences. We performed structural analyses of Sµ-3'RRrec and CSR junctions in CLL samples, HV PBMCs and benign inflammatory tonsil cells. Here, while circulating B-lymphocytes from HVs, included because they were exempt of any known disease, were predominantly resting, tonsils were analyzed since they are very well known benign inflammatory lymphoid tissues with highly active B-cell responses and numerous germinal centers, which are the main site of post-medullary Ig gene recombination. CSR joint structures were comparable between HV PBMCs and tonsils and were similar to those of CLLs regardless of Sµ-3'RRrec status (Figure 5A). In contrast, the structure of Sµ-3'RRrec joints differed between HV PBMCs and tonsils. The latter exhibited more Sµ-3'RRrec junctions with small microhomologies (1-2bp) and blunt junctions while long insertions (≥4bp) were predominant in the former (Figure 5B). Therefore,
SuSu-3'RRrec joint structures seem to be differently imprinted according to tissue origin and/or B-cell activation. This was not the case for CSR, a strong indication that these two IGH recombination events, even if mechanistically close, are not linked. Strikingly, the structures of SuSu-3'RRrec junctions differed between SuSu-3'RRrec^{High} and SuSu-3'RRrec^{Low} CLLs, the former being close to those of HV PMBCs and the latter more similar to benign inflammatory tonsils.

Decreased small microhomologies (1-2bp) and blunt junctions argue against NHEJ involvement in DSBR of SuSu-3'RRrec junctions from both HV PBMCs and SuSu-3'RRrec^{High} CLLs. We therefore quantified transcripts coding for actors implicated in the protection of DSB DNA ends and favoring NHEJ (53BP1, Rif1 and Rev7, a Shieldin complex component), NHEJ actor (LIGIV) and Alt-EJ components (PARP-1, POLθ and LIGIII). We did not detect any significant difference in the tested transcripts of NHEJ proteins (Figure 5C) or Alt-EJ actors (Figure 5D) between SuSu-3'RRrec^{High} and SuSu-3'RRrec^{Low}. As already suggested by the fact that CSR junction structures were similar between SuSu-3'RRrec^{High} and SuSu-3'RRrec^{Low} CLLs, this suggests that SuSu-3'RRrec junction structural differences observed for both SuSu-3'RRrec CLL groups were not due to imbalances in NHEJ/Alt-EJ actors.

Altogether, these results indicate that SuSu-3'RRrec^{High} and SuSu-3'RRrec^{Low} CLLs have different SuSu-3'RRrec imprints, the former being close to recirculating resting B-cells and the latter reflecting activated B-cells in a benign inflammatory secondary lymphoid organ.

Overexpression of c-MYC potentiated SuSu-3'RR recombination even in the absence of AID

Because IGH locus accessibility is also linked to B-cell activation and proliferation, we evaluated the past history of CLL B-cell proliferation by measuring relative telomere length. While homogeneous in HVs (mean of 2.14), telomere lengths were very heterogeneous in SuSu-3'RRrec^{Low} patients; 13/32 (40%) had long or very long telomeres, indicating that cells underwent few proliferation cycles. In contrast, all but one SuSu-3'RRrec^{High} patient homogeneously exhibited telomeres that were shorter than HVs, reflecting increased proliferation cycle numbers (Figure 6A). Shorter telomeres in SuSu-3'RRrec^{High} CLLs were associated with increased MYC expression (Figure 6B).

We thus raised the question of the impact of c-MYC on IGH recombination. For this purpose, we used the murine B-cell lymphoma CH12F3 cell line and its AID knock-out (ko) counterpart stably transfected or not with a MYC overexpression vector and stimulated in vitro to undergo CSR and SuSu-3'RRrec. As shown in Figure 6C, the levels of both CSR and SuSu-3'RRrec were increased when MYC was overexpressed in the AID context (AID^{+}/MYC^{+}). In the absence of AID (AID^{−}), CSR was undetectable either in absence or presence of MYC overexpression. Some SuSu-3'RRrec junctions were detectable in the absence of both AID and the MYC overexpressing vector. Induction of MYC overexpression resulted in increased numbers of SuSu-3'RRrec events (Figure 6D). Contamination could be ruled out because sequences of these SuSu-3'RRrec junctions were unique. Some of these junctions contained
sequence fragments of Sc and Sγ2 regions between Sμ and 3'RR, which suggests that these Sμ-3'RRrec events were sequentially preceded by a CSR event. These results show that MYC potentiated both CSR and Sμ-3'RRrec when AID was expressed. Even at low frequency, Sμ-3'RRrec was possible in the absence of AID but in the presence of MYC overexpression.
Discussion

In this study, we observed that Sµ-3'RR recombination (Sµ-3'RRrec) was detectable in CLL patients. Moreover, we showed that the Sµ-3'RRrec rate was increased at levels even higher than in polyclonal HIV in one group of CLLs cases, the Sµ-3'RRrec<sup>High</sup> group. In the Sµ-3'RRrec<sup>High</sup> group, Sµ-3'RRrec appeared to be on-going, as supported by increased diversity of the Sµ-3'RRrec junctions.

Despite low CSR junction counts, CSR diversity was also significantly higher in Sµ-3'RRrec<sup>High</sup> CLL samples than in Sµ-3'RRrec<sup>Low</sup> CLLs meaning that CSR was on-going in these cases. On-going IGH CSR can be observed in a restricted subpopulation of tumoral CLL cells<sup>15,17,41,42</sup>. In our study, increased Sµ-3'RRrec and CSR diversities in Sµ-3'RRrec<sup>High</sup> CLLs reflected IGH remodeling in the tumoral B-cell clone. Since CLL cases of this series were all IgM<sup>+</sup>, Sµ-3'RRrec had occurred on the non-productive IGH allele while CSR could occur on both alleles. CSR mainly occurs on the functional allele<sup>43,44</sup>. Here, the numbers of CSR junctions were strongly decreased in CLLs without any differences between Sµ-3'RRrec<sup>High</sup> and Sµ-3'RRrec<sup>Low</sup> CLLs. This indicates that despite increased IGH remodeling, Sµ-3'RRrec<sup>High</sup> class-switched CLL cells could have been counterselected. Indeed, as reviewed recently, IgM BCR is a key component of CLL pathogenesis development and evolution, not only for its antigenic recognition properties but also likely through its structure and its signaling capacities<sup>45</sup>. Even if the CSR junction structure appeared identical among samples, Sµ-3'RRrec<sup>Low</sup> and Sµ-3'RRrec<sup>High</sup> CLL samples exhibited different Sµ-3'RRrec structural profiles. In Sµ-3'RRrec<sup>High</sup> CLLs, Sµ-3'RRrec junctions were close to those of circulating B-cells, while Sµ-3'RRrec junctions from Sµ-3'RRrec<sup>Low</sup> CLLs were similar to those of tonsils B-cells. According to Sµ-3'RRrec junction imprints, our results are full agreement with the hypothesis that CLL can be subdivided according to two different COOs. Since physiologically most circulating B-cells are IgM<sup>+</sup>IgD<sup>+</sup> pre-germinal center cells and since tonsils contain numerous active germinal centers (GC), we suggest that the imprint of Sµ-3'RRrec junctions would be from circulating non-GC experienced B-cells for Sµ-3'RRrec<sup>High</sup> cases (mostly umCLLs) while Sµ-3'RRrec<sup>Low</sup> cases (mostly mCLLs) would have Sµ-3'RRrec junction imprints of AID-experienced B-cells issued from secondary lymphoid organs. This correlates with IGHV mutational status and PIM1 mutation rate that were decreased in Sµ-3'RRrec<sup>High</sup> CLLs and increased in Sµ-3'RRrec<sup>Low</sup> CLLs.

Consistent with increased IGH recombination activity, the IGH locus was strongly transcribed in Sµ-3'RRrec<sup>High</sup> CLLs and was thus targetable by the IGH recombination machinery. Characteristic of activated and proliferating cells, shorter telomeres in Sµ-3'RRrec<sup>High</sup> CLLs indicate increased numbers of past mitoses. Upon B-cell stimulation, the MYC gene frequently relocates to the transcription factory occupied by the IGH locus<sup>46</sup>. Consistently, we observed increased cMYC expression levels in Sµ-3'RRrec<sup>High</sup> CLLs. This result led us to evaluate the impact of cMYC on CSR and Sµ-3'RRrec. We
found that cMYC overexpression potentiated both CSR and Sμ-3’RRrec recombination in the presence of AID. Moreover, cMYC induced increases in Sμ-3’RRrec counts in the absence of AID. It has been recently shown that some residual CSR can occur in absence of the AID. Due to specific constitutively occurring IGH loop conformation between 3’RR, Eμ and Sμ in mature B-cells, DNA segments from both Sμ and 3’RR are likely to be in close proximity and to be recombined together when DSB-targeted even in the absence of AID. Regardless of the AID status of CLL cells, this cMYC effect contributes to genetic instability. Indeed, CLL is known to harbor DNA repair alterations and to accumulate DSBs across the genome. Moreover, both NHEJ and Alt-EJ are good candidate processes for chromosomal material exchange due to their capacity to ligate DNA ends from independent molecular origins.

IGHV mutation and Sμ-3’RRrec status exhibited strong overlaps in terms of TFS, which reflects that fact that Sμ-3’RRrec_{High} group was enriched in umCLLs. Here, TFS very likely reflects the different natural histories of the disease, which should be underpinned by biological differences. Separating CLL patients in the CLL Sμ-3’RRrec_{Low} group that would originate from GC-experienced B cells or in the CLL Sμ-3’RRrec_{High} group with probable non-GC experienced B cells and close to umCLLs, our study characterizes for the first time active IGH recombination potentiated by c-Myc overexpression in Sμ-3’RRrec_{High} CLLs, even in the absence of AID.
References


44. Sakai E, Bottaro A, Alt FW. The Ig heavy chain intronic enhancer core region is necessary and sufficient to promote efficient class switch recombination. Int Immunol. 1999;11(10):1709-1713.


Table 1

<table>
<thead>
<tr>
<th></th>
<th>Univariate analysis</th>
<th>Multivariate analysis model 1</th>
<th>Multivariate analysis model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>LCI</td>
<td>UCI</td>
</tr>
<tr>
<td>$S\mu$-3'RRrec Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S\mu$-3'RRrec\textsuperscript{M}</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S\mu$-3'RRrec\textsuperscript{U}</td>
<td>5.07</td>
<td>2.08</td>
<td>12.36</td>
</tr>
<tr>
<td>IGHV mutation status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-CLL</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UM-CLL</td>
<td>3.233</td>
<td>1.37</td>
<td>7.631</td>
</tr>
<tr>
<td>Binet stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binet stage A</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binet stage B or C</td>
<td>3.199</td>
<td>1.368</td>
<td>7.48</td>
</tr>
<tr>
<td>Lymphocytosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 30 G/L</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 30 G/L</td>
<td>1.687</td>
<td>0.7452</td>
<td>3.819</td>
</tr>
<tr>
<td>Cytogenetics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal karyotype or isolated Del(13)q</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other karyotypes</td>
<td>1.926</td>
<td>0.8306</td>
<td>4.466</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 70 years old</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 70 years old</td>
<td>1.054</td>
<td>0.4599</td>
<td>2.417</td>
</tr>
</tbody>
</table>

**Table 1**: Univariate analysis included $S\mu$-3'RR recombination ($S\mu$-3'RRrec) status, Binet stage, lymphocytosis, age, mutation status of the variable region of the immunoglobulin heavy chain (IGHV) and cytogenetics. Cox multivariate model 1 included variables with a p-value below 0.2, $S\mu$-3'RRrec status, Binet stage, IGHV mutation status (chronic lymphocytic leukemia with mutated IGHV gene (M-CLL) or with unmutated IGHV gene (UM-CLL)), lymphocytosis and cytogenetics. Cox multivariate model 2 included $S\mu$-3'RRrec status, Binet stage, lymphocytosis and cytogenetics. HR, hazard ratio; LCI, lower confidence interval; UCI, upper confidence interval; P, p-value; statistical analyses were performed using the survival R packages (see materials and methods).
Figure legends

Figure 1: Sµ-3’RRrec is detectable in CLL patients and Sµ-3’RRrec counts are significantly increased in Sµ-3’RRrec^{High} samples. A. Sµ-3’RR recombination (Sµ-3’RRrec) junction counts analyzed by next generation sequencing and CSReport in healthy volunteer (HV) peripheral blood mononuclear cells (PBMCs) (N=9, 239 Sµ-3’RRrec junctions) and Chronic Lymphocytic Leukemia (CLL) patients, divided into two groups based on the mean of junction counts obtained in healthy PBMCs: Sµ-3’RRrec^{Low} CLLs (≤27 junctions per sample, N=35, 357 junctions) and Sµ-3’RRrec^{High} CLLs (> 27 junctions per sample, N=12, 703 junctions). B. Class Switch Recombinaison (CSR) junction counts were at comparable levels in both CLL groups (Sµ-3’RRrec^{Low}: N=35, 22 247 junctions; Sµ-3’RRrec^{High}: N=11, 10 739 junctions) and were lower than in HV PBMCs (N=11, 27 528 junctions). Graphs represent the mean ± SEM. Statistical analyses were performed using Unpaired T test, ns: non-significant, **P<0.01 and ***P<0.001.

Figure 2: Enrichment in umCLLs and poor prognosis of CLL patients with increased Sµ-3’RRrec counts.
A. Sµ-3’RRrec^{Low} chronic lymphocytic leukemia (CLL) cases (N=34) had lower percentages of sequence identity with the reference sequence in Sµ-3’RRrec^{High} CLLs (N=12) compared to the high homology of the variable region of the immunoglobulin heavy chain (IGHV) segments. For each CLL group, the somatic hypermutation rate mean is indicated above the graph. B. Sequence analysis of PIM1, AID off-target gene. The mutation rate of PIM1 was significantly increased in Sµ-3’RRrec^{Low} patients (N=8) compared to healthy peripheral blood mononuclear cells (PBMCs) (N=8) and Sµ-3’RRrec^{High} CLLs (N=7). C. AID transcripts, relative to CD19 transcripts, were lower in Sµ-3’RRrec^{Low} CLLs (N=7) and Sµ-3’RRrec^{High} CLLs (N=6) compared to normal B-centroblasts (N=4) used as positive controls and comparable to AID transcript levels in sorted naïve B-cells (N=4) used as negative controls. Purple dots correspond to mutated IGHV CLL samples. D. Cumulative survival (Cum survival) time (years) without treatment (TFS) for patients indicated shorter TFS in Sµ-3’RRrec^{High} CLL (N=12) than Sµ-3’RRrec^{Low} CLL (N=34). Graphs represent the mean ± SEM. Statistical analyses were performed using Unpaired T test (A, B, C) or Chi2 test (D), ns: non-significant, *P<0.05, **P<0.01, and ***P<0.001.

Figure 3: Intratumoral IGH remodeling activity in the Sµ-3’RRrec^{High} CLL group. The Shannon diversity index was used to estimate class switch recombination (CSR), Sµ-3’RR recombination (Sµ-3’RRrec) and intra-clonal immunoglobulin heavy chain variable region (IGHV) diversities. A. Higher Sµ-3’RRrec junction diversity was observed in Sµ-3’RRrec^{High} chronic lymphocytic leukemia (CLL) samples (N=11) compared to Sµ-3’RRrec^{Low} (N=35) CLLs and healthy peripheral blood mononuclear cells (PBMCs) (N=9). B. CSR junction diversity was increased in Sµ-3’RRrec^{High} CLL samples (N=12).
compared to $\text{Sµ-3'RRrec}^{\text{low}}$ (N=35) CLLs and comparable to those of healthy PBMCs (N=11). C. CLL IGHV diversities were lower than those observed in diffuse large B-cell lymphomas (DLBCL) harboring intra-tumoral subclones with divergent IGHV after somatic hypermutation (SHM) used as positive controls (N=10). Graphs represent the mean ± SEM. Statistical analyses were performed using Unpaired T test, ns: non-significant, *P<0.05, ***P<0.001 and ****P<0.0001.

Figure 4: Increased IGH locus accessibility in $\text{Sµ-3'RRrec}^{\text{high}}$ CLLs. Quantification of immunoglobulin heavy chain locus (IgH) non-coding transcripts (Sµ, Sγ1, Sγ3, HS1.2 and HS4) relative to those of CD19 (A) and coding transcripts (Cµ and surface IgM (sIgM)) (B) in $\text{Sµ-3'RRrec}^{\text{low}}$ (N=4 to 7) and $\text{Sµ-3'RRrec}^{\text{high}}$ (N=3 to 5) CLLs. $\text{Sµ-3'RRrec}^{\text{high}}$ exhibited high levels of IgH locus transcription in both productive and non-productive transcripts. Graphs represent the mean ± SEM. Statistical analyses were performed using Unpaired T test, ns: non-significant, *P<0.05 and **P<0.01.

Figure 5: $\text{Sµ-3'RRrec}$ junction structural features are related to different lymphoid tissue imprints and discriminate between $\text{Sµ-3'RRrec}^{\text{low}}$ and $\text{Sµ-3'RRrec}^{\text{high}}$ CLLs. Structures at the $\text{Sµ-3'RR}$ recombination (Sµ-3'RRrec) and class switch recombination (CSR) junctions were determined using CSReport by alignment to reference sequences for Chronic Lymphocytic Leukemia (CLL) samples, healthy volunteer (HV) peripheral blood mononuclear cells (PBMCs) and benign inflammatory tonsil cells. Structural features account for length in base pairs (bp) of nucleotide insertions at the joint, of short homology (µhomology) between acceptor and donor sequences and absence of insertions and homology (blunt) A. $\text{Sµ-S.i)(21}, \text{Sµ-S.i)(22}, \text{Sµ-S.i)(23}, \text{Sµ-S.i)(24}$ CSR joint structures were comparable between HV PBMCs, benign inflammatory tonsils and CLLs. B. The $\text{Sµ-3'RRrec}$ junction structure differed between CLLs: $\text{Sµ-3'RRrec}^{\text{high}}$ samples were comparable to HV PBMCs and predominantly exhibited junctions with long insertions (≥4bp) while $\text{Sµ-3'RRrec}^{\text{low}}$ CLLs and benign inflammatory tonsils were enriched in $\text{Sµ-3'RRrec}$ junctions with small microhomologies (1-2bp) and blunt junctions. Quantification of transcripts coding for actors implicated in Double Strand Break (DSB) repair by qRT-PCR relative to CD19 transcripts C. by non-homologous end joining (NHEJ) (53BP1, RIF1, Rev7 and LIGIV) and D. by alternative end-joining (Alt-EJ) (PARP-1, POLθ and LIGIII). Statistical analyses were performed using the Chi2 test (A, B) or Unpaired T test (C, D), ns: non-significant, ***P<0.001 and ****P<0.0001.

Figure 6: Increased IGH remodeling in $\text{Sµ-3'RRrec}^{\text{high}}$ CLL cells is potentiated by high levels of c-MYC expression. A. Relative telomere length (RTL) measured by specific qRT-PCR relative to the human beta globin gene (Sµ-3'RRrec$^{\text{low}}$, N=33; Sµ-3'RRrec$^{\text{high}}$, N=12; Healthy volunteer (HV) peripheral blood mononuclear cells (PBMCs), N=6). Telomere length was significantly shorter in the $\text{Sµ-3'RRrec}^{\text{high}}$
group compared to the Sµ-3’RRrec^low group and HV PBMCs. B. c-MYC expression was higher in Sµ-3’RRrec^High (N=6) compared to Sµ-3’RRrec^Low Chronic Lymphocytic Leukemia (CLL) samples (N=7). C. Detection of class switch recombination (CSR) and Sµ-3’RR recombination (Sµ-3’RRrec) junctions in activated CH12F3 clones overexpressing or not MYC in the presence of AID (CSR: AID^+\textsuperscript{+}, N=1 and AID\textsuperscript{tg}, N=3; Sµ-3’RRrec: AID^+, N=1 and AID\textsuperscript{tg}, N=3) suggested that c-MYC tends to increase CSR and Sµ-3’RRrec counts. This was also observed for Sµ-3’RRrec in the absence of AID (D.), as Sµ-3’RRrec junctions, even if rare, were detectable in CH12F3 AID\textsuperscript{ko} clones (N=2) and appeared to increase with MYC overexpression (AID\textsuperscript{ko}MYC\textsuperscript{tg}, N=2). No CSR junctions were detected in the absence of AID (AID\textsuperscript{ko}, N=1), even with MYC overexpression (AID\textsuperscript{ko}MYC\textsuperscript{tg}, N=3). Graphs represent the mean ± SEM. Statistical analyses were performed using Unpaired T test, ns: non-significant, *P<0.05, **P<0.01, and ***P<0.001.
**Supplemental methods**

**CSR, Sµ-3’RRrec and intra-tumoral IGHV junction diversities**

Shannon diversity index ($H$) was used to estimate CSR, Sµ-3’RRrec or intra-clonal IGHV diversities and was calculated considering the number of reads ($ni$) for each particular CSR, Sµ-3’RRrec or IGHV rearrangement and the total read number ($N$) of total CSR, Sµ-3’RRrec or IGHV junctions:

$$H = - \sum_{i=1}^{S} pi \ln(pi) \text{ with } pi = \frac{ni}{N}$$

The $H$ value ranges from 0 when Sµ-3’RRrec junction diversity shrinks due to clonal dominance to high values for samples with higher diversity. CSR, Sµ-3’RRrec, CLL IGHV sequencing data produced in this study have been deposited in the National Center for Biotechnology Information’s BioProject (PRJNA830327).

**Intra-clonal IGHV analysis**

For the analysis, Immcantation framework\(^1\) was used via Immcantation/suite docker container v4.2 (https://hub.docker.com/r/immcantation/suite/). For each sample, output from IMGT/HighV-QUEST\(^2\) was parsed via the imgt subcommand of MakeDb.py to generate a standardized tab-delimited database file. Then the non-productive sequences were removed with select subcommand of ParseDb.py. An automated detection of the clonal assignment threshold was then performed by shazam-threshold pipeline and DefineClones.py was launched on this new database with the use of the following argument: --act set to take into account ambiguous V gene and J gene calls when grouping similar sequences, --model hh_s5f which corresponds to SHM targeting and substitution model for human Ig data\(^3\). Because the threshold was generated using length normalized distances, the --norm len argument is selected and the previously determined threshold was settings with the –dist argument. The IGHV, IGHD, IGHJ germline sequences that were used for the IMGT/HighV-QUEST alignment were downloaded from http://www.imgt.org/download/V-QUEST/IMGT_V-QUEST_reference_directory/Homo_sapiens/IG/ (Release 202214-2) and were passed to CreateGermlines.py via the -r argument in order to reconstruct the germline V(D)J sequence, from which the Ig lineage and mutations can be inferred. Dplyr R packages and countClones function from alakazam R package were used to determine the number of distinct sequences in each clone to select the one with the higher value for downstream analysis. All sequences with a duplicate count lower than 1/1000 reads count of previously determined tumoral clone were discarded.
PIM1 mutation analysis

For each library, alignment of sequenced reads with the reference sequence NM_002648.4 using the Torrent Mapping Alignment Program (TMAP) for Ion Torrent Data and Super-maximal Exact Matching algorithm results in BAM files. BAM files were processed to generate per-base nucleotide count table files consisting of matrices with \( n \) lines \( \times 4 \) columns. \( n \) is the length of the sequenced DNA and the columns correspond to nucleotides (A, C, G, and T). The consensus sequence is the most frequently read nucleotide and corresponds to the sequence reference. Counts of mutated bases were calculated by addition of numbers of sequenced bases different from the nucleotide that was sequenced the most frequently. PIM1 sequencing data produced in this study have been deposited in the National Center for Biotechnology Information’s BioProject (PRJNA830327).

Supplemental Tables

**Supplemental Table 1**

<table>
<thead>
<tr>
<th>Count of Sµ-3′RRrec junctions using CSReport</th>
<th>Healthy PBMCs</th>
<th>Sµ-3′RRrec\textsuperscript{low} CLL PBMCs</th>
<th>Sµ-3′RRrec\textsuperscript{high} CLL PBMCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sµ-3′RRrec positive samples /total samples (%)</td>
<td>9/9 (100)</td>
<td>32/35 (91)</td>
<td>12/12 (100)</td>
</tr>
<tr>
<td>Mean of Sµ-3′Rrrec junction (min-max)</td>
<td>26.6 (8-71)</td>
<td>10.2 (0-24)</td>
<td>58.5 (29-145)</td>
</tr>
</tbody>
</table>

**Supplemental table 1.** Numbers of healthy volunteers and chronic lymphocytic leukemia (CLL) patients tested and Sµ-3′RR recombination (Sµ-3′RRrec) junction counts and intervals (minimum-maximum) obtained for each group. Based on the mean junction counts obtained in healthy peripheral blood mononuclear cells (PBMCs) we divided the CLL cohort into two groups: Sµ-3′RRrec\textsuperscript{low} \( \leq \) 27 junctions and Sµ-3′RRrec\textsuperscript{high} > 27 junctions per sample.

**Supplemental Table 2**

<table>
<thead>
<tr>
<th>Samples * ( \text{P}=0.05 )</th>
<th>Sµ-3′RRrec\textsuperscript{low} CLL PBMCs</th>
<th>Sµ-3′RRrec\textsuperscript{high} CLL PBMCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSR\textsuperscript{low} CLL PBMCs</td>
<td>27</td>
<td>5</td>
</tr>
<tr>
<td>CSR\textsuperscript{high} CLL PBMCs</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

**Supplemental table 2.** Repartition of patients between the two groups Sµ-3′RRrec\textsuperscript{low} and Sµ-3′RRrec\textsuperscript{high}, and class switch recombination (CSR) (CSR\textsuperscript{low}, \( \leq 800 \) CSR junctions per sample, and CSR\textsuperscript{high}, >800 CSR junctions per sample). Statistical analyses were performed using Fisher’s Exact Test *\( \text{P}<0.05 \).
### Supplemental Table 3

<table>
<thead>
<tr>
<th>Target and segment</th>
<th>Primer name</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSR Junctions (CH12F3)</td>
<td><strong>PCR1</strong></td>
<td>$\text{S}1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{S}a1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{S}2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{S}a2$</td>
</tr>
<tr>
<td></td>
<td><strong>PCR2</strong></td>
<td>$\text{S}1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{L}52$-$\text{R}3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{L}54$-$\text{R}6$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{L}10$-$\text{R}3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{L}52$-$\text{R}2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{L}54$-$\text{R}2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{L}10$-$\text{R}2$</td>
</tr>
<tr>
<td>IgH locus Transcription</td>
<td><strong>PCR1</strong></td>
<td>$\text{S}1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{S}a1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{S}1$-$\text{S}3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{H}51$-$\text{H}2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{H}54$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{S}3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\text{IgM total (C}\mu \text{)}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Surface IgM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telomere length</td>
<td>Telomere A</td>
<td>CGTTTGGTTTGGGTGTTGGGTTTGGTTGTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Telomere B</td>
<td>GGTTGAGGAGCTGCTG</td>
</tr>
<tr>
<td></td>
<td>AICDA transcription</td>
<td>AID</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R:GGTTGAGGAGCTGCTG</td>
</tr>
<tr>
<td></td>
<td>PIM1 mutation</td>
<td>PIM1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R:GGTTGAGGAGCTGCTG</td>
</tr>
<tr>
<td></td>
<td>CD19</td>
<td>Human Beta globin (hb)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Supplemental Table 3.** Primers used in this study. CSR, class switch recombination; Sμ-3’RRrec, Sμ-3’RR recombination; IgH, immunoglobulin heavy chain, S, switch region; LS, like switch region, HS, hypersensitive site.

**Supplemental Figure Legends**

**Supplemental figure 1: Description of CLL PBMC samples. A.** Blood tumor infiltration in Chronic Lymphocytic Leukemia (CLL) (N=47) samples indicated by the percentage of CD5+ CD19+ B cells gated...
on total CD19+ B cells. B. Sµ-3’RR recombination (Sµ-3’RRrec) junctions were at comparable levels in both CLL samples (N=47, Sµ-3’RRrec 1060 junctions) and healthy volunteer (HV) peripheral blood mononuclear cells (PBMCs) (N=9, 239 Sµ-3’RRrec junctions). Unpaired T test, ns: no significant difference. C. Graphical representation of CLL cell percentages within total B-cells according to Sµ-3’RRrec junction count for each sample demonstrated that Sµ-3’RRrec counts were not dependent on CLL B-cell richness. D. Percentage of CD3+ and CD19+ cells among all lymphocytes and among all PBMC. Sµ-3’RRrecLow, N=15; Sµ-3’RRrecHigh, N=8. E. Kaplan Meyer curves of treatment free survival (TFS) (years) showed no significant differences between CSRLow (N=32, 8540 junctions) and CSRHigh (N=14, 22683 junctions) CLL patients, separated with a threshold of 800 class switch recombination (CSR) count (=4.6 years compared to =4.3 years; P=0.4420. Chi2 test).

Supplemental figure 2: immunoglobulin heavy chain variable (IGHV) gene usage frequency in Chronic Lymphocytic Leukemia (CLL) in this study are represented in circle diagrams (A.). B. Analysis of immunoglobulin heavy chain variable (IGHV) usage is not significantly different between Sµ-3’RRrecLow and Sµ-3’RRrecHigh groups (Chi2 test). Absolute number of cases is indicated for each section.

Supplemental figure 3: CD19 expression in normal and CLL B-cells. A. CD19+ transcript quantification relative to GAPDH expression in Chronic Lymphocytic Leukemia (CLL) peripheral blood mononuclear cells (PBMCs). Sµ-3’RRrecLow, N=8; Sµ-3’RRrecHigh, N=6. B. Mean fluorescence intensity (MFI) of CD19 at the B-cell surface appears comparable for CD5+CD19+ tumor cells from both CLL groups (Sµ-3’RRrecLow, N=14; Sµ-3’RRrecHigh, N=7) and decreased, as expected, compared to normal CD5-CD19+ B-cells from CLL patients. Statistical analyses were performed using the Unpaired T test, ns: no significant difference.

Supplemental figure 4: B-cell numbers equivalent to DNA used for CSR and Sµ-3’RRrec junctions. Absolute number of CD19+ B-cells corresponding to 600ng of peripheral blood mononuclear cells (PBMCs) DNA used for class switch recombination (CSR) and Sµ-3’RR recombination (Sµ-3’RRrec) amplification. Sµ-3’RRrecLow, N=15; Sµ-3’RRrecHigh, N=8. Unpaired T test, ns: no significant difference.

Supplemental figure 5: Kaplan Meyer curves of TFS. Kaplan Meyer curves of treatment free survival (TFS) (years) depending on immunoglobulin heavy chain variable (IGHV) mutation status (A.), Binet stage (B.), lymphocytosis (C.) and cytogenetics (D.) (Chi2 test).

Supplemental references


Supplemental Figure 2

A. CLLs

B. Sp-3'RRrec\textsubscript{Low} CLL PBMCs

p=0.37

Sp-3'RRrec\textsubscript{High} CLL PBMCs
Supplemental Figure 3

A. CD19

- % of transcripts

B. Gated on CD19^+ cells

- Mean of fluorescence intensity of CD19

Legend:
- S<sub>μ-3'RRec</sub> Low
- S<sub>μ-3'RRec</sub> High
- CLL PBMCs
- CD5^+ B-cells
- CD5^− B-cells

Note: ns indicates non-significant difference.