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

Israa Al Jamal,[1,2] Milène Parquet,[1+] Kenza Guiyedi,[1+] Said Aoufouchi,[3,4+] Morwenna Le Guillou,[3] David Rizzo,[1,5] Justine Pollet,[1] Marine Dupont,[1,5] Mélanie Boulin,[1,5] Nathalie Faumont,[1] Hend Boutouil,[1] Fabrice Jardin,[6] Philippe Ruminy,[6] Chahrazed El Hamel,[7] Justine Lerat,[8] Samar Al Hamaoui,[2] Nehman Makdissy,[2] Jean Feuillard,[1,5] Nathalie Gachard[1,5] and Sophie Peron[1]

[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; [4]Université Paris-Saclay, Orsay, France; [5]Laboratoire d'Hématologie Biologique, Centre Hospitalier Universitaire de Limoges, Limoges, France; [6]Department of Henri-Becquerel Hematology Center and Normandie, INSERM U1245, Université de Rouen, Rouen, France; [7]Collection Biologique Hôpital de la Mère et de l'Enfant (CB-HME), Department of Pediatrics, Limoges University Hospital, Limoges, France and [8]Department of Otorinolaryngology, Limoges University Hospital, Limoges, France

[+]MP, KG and SA contributed equally.

Correspondence: S. Peron
sophie.peron@unilim.fr

Received: February 13, 2023.
Accepted: July 20, 2023.
Early view: July 27, 2023.

https://doi.org/10.3324/haematol.2023.282897

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[4] results in BAM files. BAM files were processed to generate per-base nucleotide count table files consisting of matrices with *n* lines × 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 **s**equencing data produced in this study have been deposited in the National Center for Biotechnology Information's BioProject (PRJNA830327).

Supplemental Tables

Supplemental Table1

Count of Sμ-3'RRrec junctions using CSReport	Healthy PBMCs	Sμ-3'RRrec^{Low} CLL PBMCs	Sμ-3'RRrec^{High} CLL PBMCs
Sμ-3'RRrec positive samples /total samples (%)	9/9 (100)	32/35 (91)	12/12 (100)
Mean of Sμ-3'Rrrec junction (min-max)	26.6 (8-71)	10.2 (0-24)	58.5 (29-145)

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^{Low} ≤27 junctions and Sμ-3'RRrec^{High} > 27 junctions per sample.

Supplemental Table 2

Samples *(P=0.05)	Sμ-3'RRrec^{Low} CLL PBMCs	Sμ-3'RRrec^{High} CLL PBMCs
CSR^{Low} CLL PBMCs	**27**	**5**
CSR^{High} CLL PBMCs	**8**	**7**

Supplemental table 2. Repartition of patients between the two groups Sμ-3'RRrec^{Low} and Sμ-3'RRrec^{High}, and class switch recombination (CSR) (CSR^{Low}, ≤800 CSR junctions per sample, and CSR^{High}, >800 CSR junctions per sample). Statistical analyses were performed using Fisher's Exact Test *P<0.05.

Supplemental Table 3

Target and segment		Primer name	Sequence 5'-3'
CSR Junctions (CH12F3)	PCR1	Sμ1	F:TAGTAAGCGAGGCTCTAAAAAGCA
		Sα1	R:CAGCAGTGAGTTTAACAATCC
	PCR2	Sμ2	F:GCTTGAGCCAAAATGAAGTAGACT
		Sα2	R:CCTCAGTGCAACTCTATCTAGGTCT
Sμ-3'RRrec junctions (CH12F3)	PCR1	Sμ1	F:TAGTAAGCGAGGCTCTAAAAAGCA
		LS2-R3	R:AACAAGAGGTGGGGAGTGTG
		LS4-R3	R:CTATAGCCATGTGGGGCTGT
		LS10-R3	R:GGGAGTGCCAGTGTCAACTT
	PCR2	Sμ2	F:GCTTGAGCCAAAATGAAGTAGACT
		LS2-R2	R:TGTCCAGGCTGAGCTACCTT
		LS4-R2	R:TTTACCAATCTCCCCCACTG
		LS10-R2	R:GTGAGTGTGTGGGGTTTGTG
IgH locus Transcription	Sμ	Sμg	F:GGTGTGGGTTTTCACAGCTT
			R:CCTCACCAAGTCCACCAAGT
	Sγ1	Sγ1-3	F:CTGGGATGGAGAAGGGAAGG
			R:CTGGTCTCAAGCACACGTTC
	HS	HS1,2	F:GAGTTTTCGGCATCTCTGGG
			R:ACAGATCAGAGCCCTCACAC
		HS4	F:GTGTGTCTGAGGGTGAGTGA
			R:ACACTGTCACACACTCCACA
	Sγ3	Sγ3	F:AGCTGTGCAACTGGAGTCCT
			R:TGAGCCACCTAATCCAAACC
	IgM total (Cμ)	hIgM CH1	F :CAGAATGCGTCCTCCATGTG
			R :GGTGGACTTGGTGAGGAAGA
	Surface IgM	hsIgM-CH4-For hsIgM-MB-Rev	F: AAGAGGAATGGAACACGGGG R :ACAAGGTGACGGTGGTACTG
Telomere length	Telomere	Telomere A	CGGTTTGTTTGGGTTTGGGTTTGGGTTT GGGTTTGGGTT
		Telomere B	GGCTTGCCTTACCCTTACCCTTACCCTTAC CCTTACCCT
AICDA transcription	AID		F :GAGGCAAGAAGACACTCTGG
			R :GTGACATTCCTGGAAGTTGC
PIM1 mutation	PIM1		F :ATGAGTGGGTGGGGTGAGG
			R :ATCGAGCCAGGCGGCCC
Internal control	CD19		F:AGACTCCTTCTCCAACGGTA
			R:GGTCAGCTCTTCATCCTCGT
	Human Beta globin (hbg)	Hbg1	GCTTCTGACACAACTGTGTTCACTAGC
		Hbg2	CACCAACTTCATCCACGTTCACC

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'$RRrec^{Low}$, N=15; Sμ-3'$RRrec^{High}$, N=8. **E.** Kaplan Meyer curves of treatment free survival (TFS) (years) showed no significant differences between CSR^{Low} (N=32, 8540 junctions) and CSR^{High} (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'$RRrec^{Low}$ and Sμ-3'$RRrec^{High}$ 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'$RRrec^{Low}$, N=8; Sμ-3'$RRrec^{High}$, 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'$RRrec^{Low}$, N=14; Sμ-3'$RRrec^{High}$, 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'$RRrec^{Low}$, N=15; Sμ-3'$RRrec^{High}$, 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

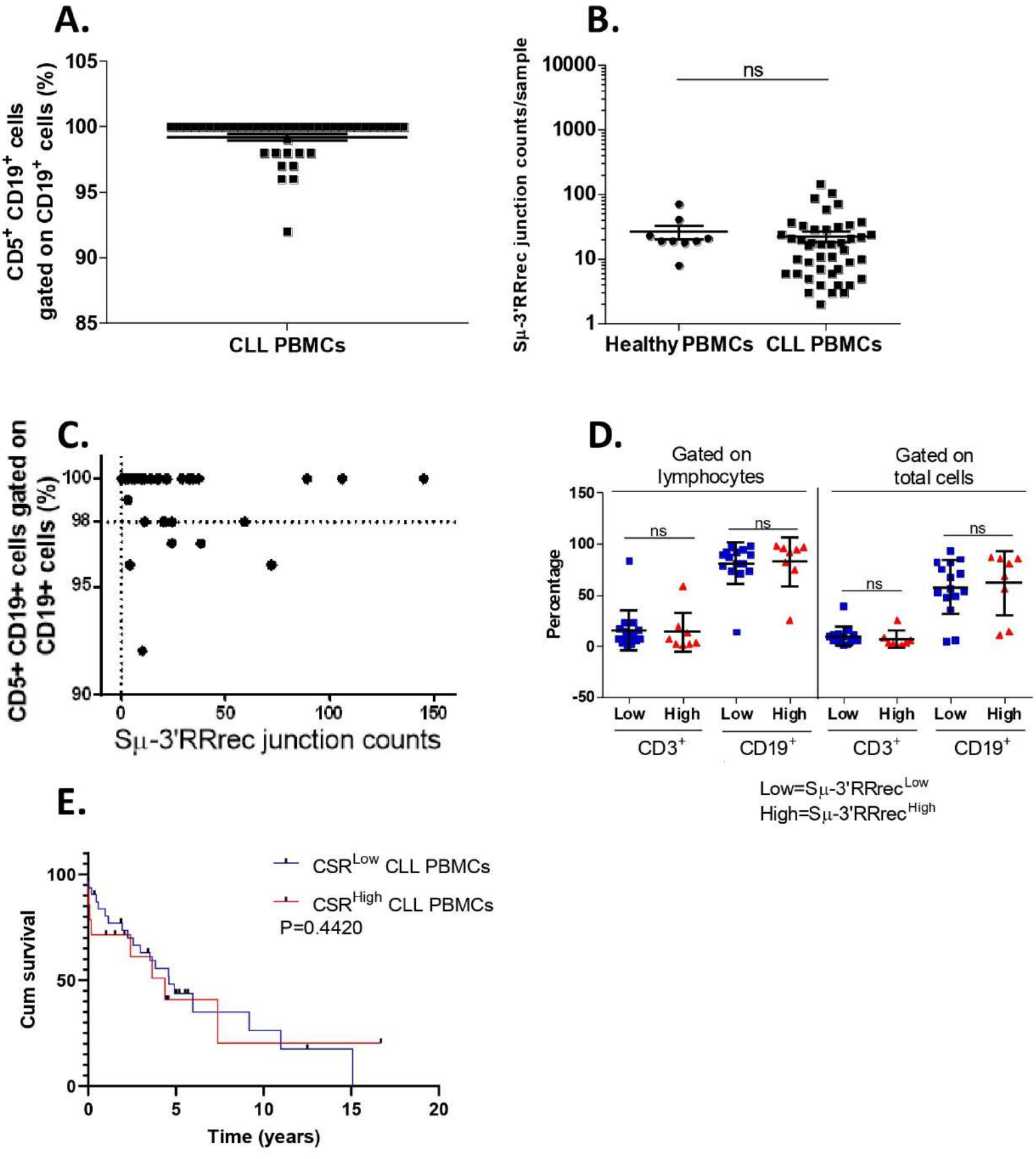
1. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics 2015;31(20):3356–3358.

2. Brochet X, Lefranc M-P, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res 2008;36(Web Server issue):W503-508.

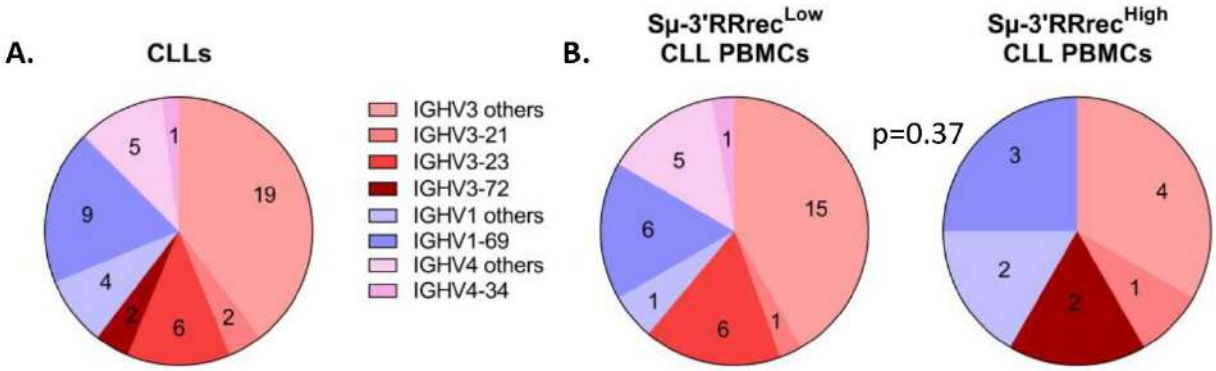
3. Yaari G, Vander Heiden JA, Uduman M, et al. Models of Somatic Hypermutation Targeting and Substitution Based on Synonymous Mutations from High-Throughput Immunoglobulin Sequencing Data. Front Immunol 2013;4358.

4. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25(14):1754–1760.

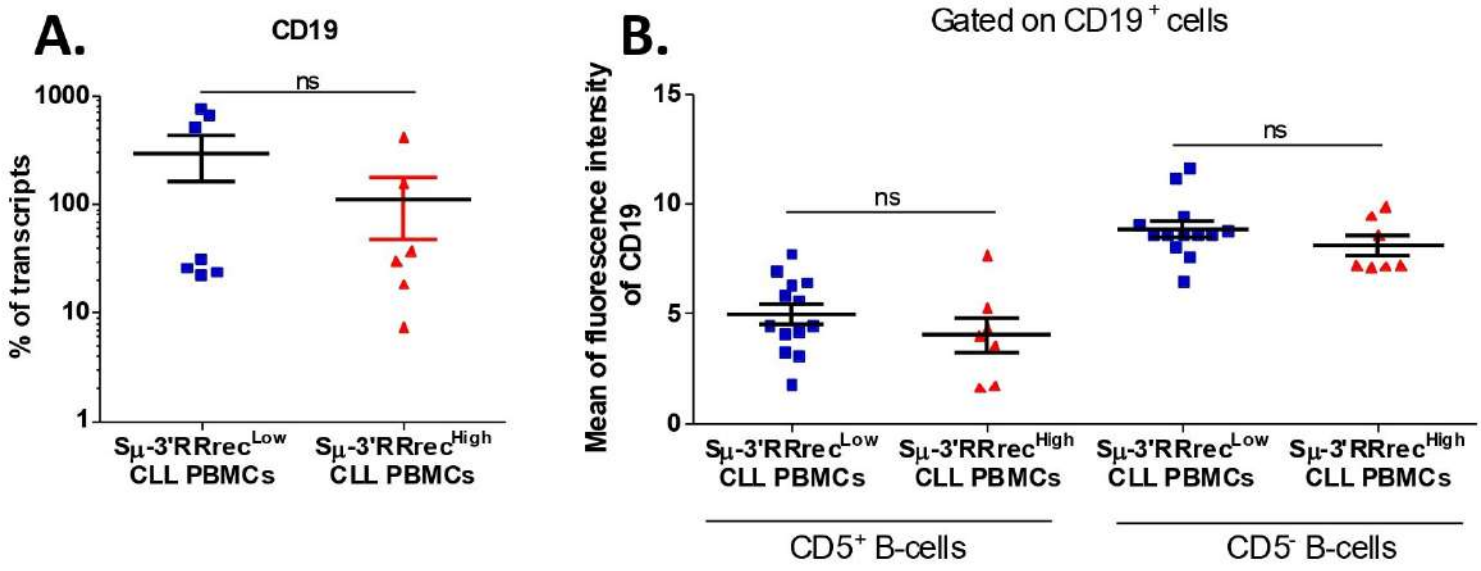
Supplemental Figure 1



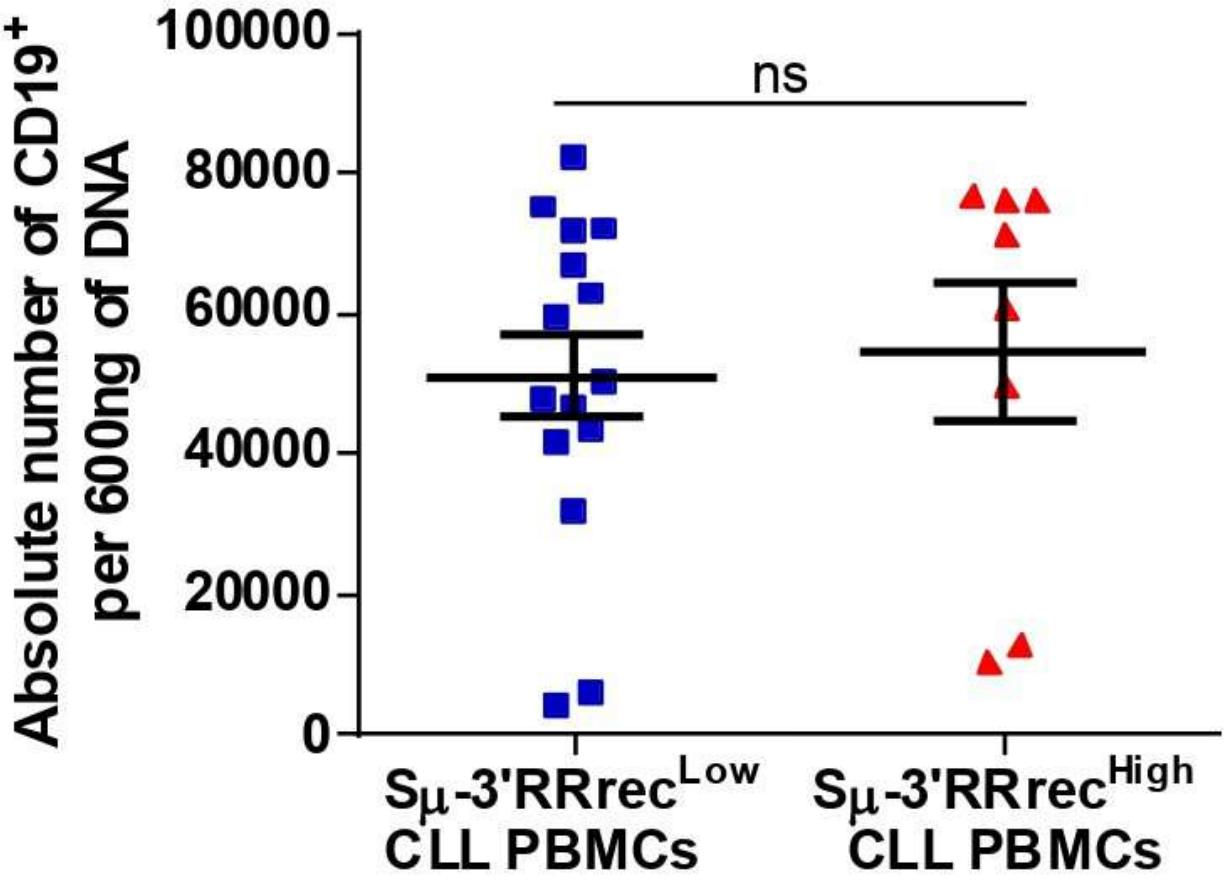
Supplemental Figure 2



Supplemental Figure 3



Supplemental Figure 4



Supplemental Figure 5