RAG1 and RAG2 non-core regions are implicated in the leukemogenesis and off-target V(D)J recombination in *BCR-ABL1*-driven B cell lineage lymphoblastic leukemia.

Xiaozhuo Yu[1*], Wen Zhou[1*], Xiaodong Chen[1], Shunyu He[1], Mengting Qin[1], Meng Yuan[1], Yang Wang[1], Woodvine otieno Odhiambo[1], Yinsha Miao[2 **] and Yanhong Ji[1,2 **]

[1]Department of Pathogenic Biology and Immunology. School of Basic Medical Sciences, Xi'an Jiaotong University Health Science Center.

[2] Department of Clinical laboratory, Xi'an No. 3 Hospital, the Affiliated Hospital of Northwest University.

*Xiaozhuo Yu and Wen Zhou contributed equally to this work as cofirst anthor

**Correspondences:

Yanhong Ji (jiyanhong@xjtu.edu.cn), Yinsha Miao (miaoyinsha@med.nwu.edu.cn),

Running title: RAG's non-core regions suppress leukemogenesis and off-target V(D)J recombination

Key words: RAG, non-core regions, off-target V(D)J recombination, *BCR-ABL1*$^+$ B-ALL, genomic stability

Abstract

The evolutionary conservation of non-core RAG regions suggests significant roles that might involve quantitative or qualitative alterations in RAG activity. Off-target V(D)J recombination contributes to lymphomagenesis and is exacerbated by RAG2' C-terminus absence in $Tp53^{-/-}$ mice thymic lymphomas. However, the genomic stability effects of non-core regions from both cRAG1 and cRAG2 in *BCR-ABL1*$^{+}$ B-lymphoblastic leukemia (*BCR-ABL1*$^{+}$ B-ALL), the characteristics, and mechanisms of non-core regions in suppressing off-target V(D)J recombination remains unclear. Here, we established three mice models of *BCR-ABL1*$^{+}$ B-ALL in *full-length RAG* (*fRAG*), *core RAG1* (*cRAG1*), and *core RAG2* (*cRAG2*) mice. The *cRAG* (*cRAG1 and cRAG2*) leukemia cells exhibited greater malignant tumor characteristics compared to *fRAG* cells. Additionally, *cRAG* cells showed higher frequency of off-target V(D)J recombination and oncogenic mutations than *fRAG*. We also revealed decreased RAG binding accuracy in *cRAG* cells and a smaller recombinant size in *cRAG1* cells, which could potentially exacerbate off-target V(D)J recombination in *cRAG* cells. In conclusion, these findings indicate that the non-core RAG regions, particularly the non-core region of RAG1, play a significant role in

preserving V(D)J recombination precision and genomic stability in *BCR-ABL1*$^{+}$ B-ALL.

Introduction

V(D)J recombination serves as the central process for early lymphocyte development and generates diversity in antigen receptors. This process involves the double-strand DNA cleavage of gene segments by the V(D)J recombinase, including RAG1 and RAG2. RAG recognizes conserved recombination signal sequences (RSSs) positioned adjacent to V, D, and J gene segments. A bona fide RSS contains a conserved palindromic heptamer (consensus5′-CACAGTG) and A-rich nonamer (consensus 5′-ACAAAAACC) separated by a degenerate spacer of either 12 or 23 base pairs (***Schatz and Ji*,*2011***; ***Hirokawa, et al.*,*2020***). The process of efficient recombination is contingent upon the presence of recombination signal sequences (RSSs) with differing spacer lengths, as dictated by the "12/23 rule" (***Eastman, et al.*,*1996***; ***Banerjee and Schatz*,*2014***). Following cleavage, the DNA ends are joined via non-homologous end joining (NHEJ), resulting in the precise alignment of the two coding segments and the signaling segment (***Rooney, et al.*,*2004***). V(D)J recombination promotes B cell development, but aberrant V(D)J recombination can lead to

precursor B-cell malignancies through RAG mediated off-target effects (***Thomson, et al.,2020***; ***Mendes, et al.,2014***; ***Onozawa and Aplan,2012***).

The regulation of RAG expression and activity is multifactorial, serving to ensure V(D)J recombination and B cell development (***Gan, et al.,2021***; ***Kumari, et al.,2021***). The RAGs consist of core and non-core region. Although recombinant dispensability appears evident, the non-core RAG regions are evolutionarily conserved, indicating their potential significance in vivo that may involve quantitative or qualitative modifications in RAG activity and expression (***Liu, et al.,2022***; ***Curry and Schlissel,2008***; ***Sekiguchi, et al.,2001***). Specifically, the non-core RAG2 region (amino acids 384–527 of 527 residues) contains a plant homeodomain (PHD) that can recognize histone H3K4 trimethylation, as well as a T490 locus that mediates a cell cycle-regulated protein degradation signal in proliferated pre-B cells stage (***Matthews, et al.,2007***; ***Liu, et al.,2007***). The off-target V(D)J recombination frequency is significantly higher when RAG2 is C-terminally truncated, thereby establishing a mechanistic connection between the PHD domain, H3K4me3-modified chromatin, and the suppression of off-target V(D)J recombination (***Lu, et al.,2015***; ***Mijušković, et al.,2015***). RAG2 destruction is linked to the cell cycle through the cyclin-

dependent kinase cyclinA/Cdk2, which phosphorylates T490. Failure to degrade RAG during the S stage poses a threat to the genome (***Zhang, et al.***,***2011***). A T490A mutation at the phosphorylation site contribute to lymphomagenesis in a p53-deficient background. The RAG1' non-core region (amino acids1–383 of 1040 residues) has been identified as a RAG1 regulator. While the core RAG1 maintains its catalytic activity, it's in vivo recombination efficiency and fidelity are reduced in comparison to the full-length RAG1 (fRAG1). In addition, the RAG binding to the genome is more indiscriminate (***Silver, et al.***,***1993***; ***Beilinson, et al.***,***2021***; ***Sadofsky, et al.***,***1993***). The N-terminal domain (NTD), which is evolutionarily conserved, is predicted to contain multiple zinc-binding motifs, including a Really Interesting New Gene (RING) domain (aa 287 to 351) that can ubiquitylate various targets, including RAG1 itself (***Deng, et al.***,***2015***). While the ubiquitylation activity has been characterized in vitro, its in vivo relevance to V(D)J recombination and off-target V(D)J recombination remains uncertain. Furthermore, the N-terminal domain (NTD) contains a specific region (amino acids 1 to 215) that facilitates interaction with DCAF1, leading to the degradation of RAG1 in a CRL4-dependent manner (***Schabla, et al.***,***2018***). Additionally, the NTD plays a role in chromatin binding and the genomic targeting of the

RAG complex (***Schatz and Swanson*,*2011***). Despite increased evidence emphasizing the significance of non-core RAG regions, particularly RAG1's non-core region, the function of non-core RAG regions in off-target V(D)J recombination and the underlying mechanistic basis have not been fully clarified.

Typically, genomic DNA is safeguarded against inappropriate RAG cleavage by the inaccessibility of cryptic RSSs (cRSSs), which are estimated to occur once per 600 base pairs (***Lewis, et al.*,*1997***; ***Teng, et al.*,*2015***). However, recent research has demonstrated that epigenetic reprogramming in cancer can result in heritable alterations in gene expression, including the accessibility of cRSSs (***Khoshchehreh, et al.*,*2019***; ***Becker, et al.*,*2020***; ***Fatma, et al.*,*2022***; ***Goel, et al.*,*2022***). We selected the *BCR-ABL1*$^{+}$ B-ALL model, which is characterized by ongoing V(D)J recombinase activity and BCR-ABL1 gene rearrangement in pre-B leukemic cells (***Schjerven, et al.*,*2017***; ***Wong and Witte*,*2004***). The genome structural variations (SVs) analysis was conducted on leukemic cells from *fRAG, cRAG1*, and *cRAG2, BCR-ABL1*$^{+}$ B-ALL mice to examine the involvement of non-core RAG regions in off-target V(D)J recombination events. The non-core domain deletion in both *RAG1* and *RAG2* led to accelerated leukemia onset and progression, as well as an increased off-target V(D)J recombina-

tion. Our analysis showed a reduction in RAG binding accuracy in *cRAG* cells and a decrease in recombinant size in *cRAG1* cells, which may be responsible for the increased off-target V(D)J recombination in *cRAG* leukemia cells. In conclusion, our results highlight the potential importance of the non-core RAG region, particularly RAG1's non-core region, in maintaining accuracy of V(D)J recombination and genomic stability in *BCR-ABL1*$^{+}$ B-ALL.

Method

Mice

The C57BL/6 mice were procured from the Experimental Animal Center of Xi'an Jiaotong University, while cRAG1 (amino acids 384-1040) and cRAG2 (amino acids 1-383) were obtained from David G. Schatz (Yale University, New Haven, Connecticut, USA). The mice were bred and maintained in a specific pathogen-free (SPF) environment at the Experimental Animal Center of Xi'an Jiaotong University. All animal-related procedures were in accordance with the guidelines approved by the Xi'an Jiaotong University Ethics Committee for Animal Experiments.

Generation of Retrovirus Stocks

The pMSCV-BCR-BAL1-IRES-GFP vector is capable of co-

expressing the human BCR-ABL1 fusion protein and green fluorescence protein (GFP), while the pMSCV-GFP vector serves as a negative control by solely expressing GFP. To produce viral particles, 293T cells were transfected with either the MSCV-BCR-BAL1-IRES-GFP or MSCV-GFP vector, along with the packaging vector PKAT2, utilizing the X-tremeGENE HP DNA Transfection Reagent from Roche (Basel, Switzerland). After 48 hours, the viral supernatants were collected, filtered, and stored at -80°C.

Bone Marrow Transduction and Transplantation

Experiments were conducted using mice aged between 6 to 10 weeks. *BCR-ABL1*$^{+}$ B-ALL was induced by utilizing marrow from donors who had not undergone 5-FU treatment. The donor mice were euthanized through CO_2 asphyxiation, and the bone marrow was harvested by flushing the femur and tibia with a syringe and 26-gauge needle. Erythrocytes were not removed, and 1 ×10^6 cells per well were plated in six-well plates. A single round of co-sedimentation with retroviral stock was performed in medium containing 5% WEHI-3B-conditioned medium and 10 ng/mL IL-7 (Peprotech, USA). After transduction, cells were either transplanted into syngeneic female recipient mice (1 × 10^6 cells each) that had been lethally irradiated (2 × 450 cGy), or cultured in RPMI-

1640 (Hyclone, Logan, UT) medium supplemented with 10% fetal calf serum (Hyclone), 200 mmol/L L-glutamine, 50 mmol/L 2-mercaptoethanol (Sigma, St Louis, MO), and 1.0 mg/ml penicillin/streptomycin (Hyclone). Subsequently, recipient mice were monitored daily for indications of morbidity, weight loss, failure to thrive, and splenomegaly. Weekly assessment of peripheral blood GFP percentage was done using FACS analysis of tail vein blood. Hematopoietic tissues and cells were utilized for histopathology, in vitro culture, FACS analysis, secondary transplantation, genomic DNA preparation, protein lysate preparation, or lineage analysis, contingent upon the unique characteristics of mice under study.

Secondary Transplants

Thawed BM cells were sorted using a BD FACS Aria II (Becton Dickinson, San Jose California, USA). GFP positive leukemic cells (1×10^6, 1×10^5, 1×10^4, and 1×10^3) were then resuspended in 0.4 mL Hank's Balanced Salt Solution (HBSS) and intravenously administered to unirradiated syngeneic mice.

Flow cytometry analysis and sorting

Bone marrow, spleen cells, and peripheral blood were harvested from leukemic mice. Red blood cells were eliminated using NH4Cl

RBC lysis buffer, and the remaining nucleated cells were washed with cold PBS. In order to conduct in vitro cell surface receptor staining, 1× 10^6 cells were subjected to antibody staining for 20 minutes at 4°C in 1×phosphate buffer saline (1×PBS) containing 3% BSA. Cells were then washed with 1×PBS and analyzed using a CytoFLEX Flow Cytometer (Beckman Coulter, Miami, FL) or sorted on a BD FACS Aria II. Apoptosis was analysed by resuspending the cells in Binding Buffer (BD Biosciences, Baltimore, MD, USA), and subsequent labeling with anti-annexin V-AF647 antibody (BD Biosciences) and propidium iodide (BD Biosciences) for 15 minutes at room temperature. The lineage analysis was performed using the following antibodies, which were procured from BD Biosciences: anti-BP-1-PerCP-Cy7, anti-CD19-PerCP-CyTM$^{5.5}$, anti-CD43-PE, anti-B220-APC, and anti-μHC-APC.

BrdU incorporation and analysis

Cells obtained from primary leukemic mice were cultured in six-well plates containing RPMI-1640 medium supplemented with 10% FBS and 50 mg/ml BrdU. After a 30-minute incubation at 37°C, cells were harvested and intranuclearly stained using anti-BrdU and 7-AAD antibodies, as per the manufacturer's instructions.

Western blotting analysis

Over 1×10^6 leukemic cells were centrifuged and washed with ice-cold PBS. The cells were then treated with ice-cold RIPA buffer, consisting of 50 mM Tris-HCl (pH 7.4), 0.15 M NaCl, 1% Triton X-100, 0.5% NaDoc, 0.1% sodium dodecyl sulphide (SDS), 1 mM ethylene diamine tetraacetic acid (EDTA), 1 mM phenylmethane sulphony fluoride (PMSF) (Amresco), and fresh protease inhibitor cocktail Pepstain A (Sigma). After sonication using a Bioruptor TMUCD-200 (Diagenode, Seraing, Belgium), the suspension was spined at 14,000 g for 3 minutes at 4°C. The total cell lysate was either utilized immediately or stored at -80°C. Protein concentrations were determined using DC Protein Assay (Bio–Rad Laboratories, Hercules, California, USA). Subsequently, the protein samples (20 µg) were incubated with α-RAG1 (mAb 23) and α-RAG2 (mAb 39) antibodies, with GAPDH serving as the loading control. The signal was further detected using secondary antibody of goat anti-rabbit IgG conjugated with horseradish peroxidase (Thermo Scientific, Waltham, MA). The band signal was developed with Immobilon™ Western Chemiluminescent HRP substrate (Millipore, Billerica, MA). The band development was analyzed using GEL-PRO ANALYZER software (Media Cybernetics, Bethesda, MD).

Genomic PCR

Genomic PCR was performed using the following primers (***Schlissel, et al.***,***1991***):

DhL-5′-GGAATTCGMTTTTTGTSAAGGGATCTACTACTGTG-3′;

J3-5′-GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG-3′;

VQ52-5′-CGGTACCAGACTGARCATCASCAAGGACAAYTCC-3′;

Vh558-5′-CGAGCTCTCCARCACAGCCTWCATGCARCTCARC-3′;

Vh7183-5′-CGGTACCAAGAASAMCCTGTWCCTGCAAATGASC-3′.

RNA-seq library preparation and sequencing

GFP^+CD19^+ cells were sorted from the spleen of cRAG1 (n=3, 1×10^6 cells /sample), cRAG2 (n=3, 1×10^6 cells /sample), and fRAG (n=3, 1×10^6 cells /sample) B-ALL mice. Total RNA was extracted using Trizol reagent (Invitrogen, CA, USA) following the manufacturer's guidelines. RNA quantity and purity analysis was done using Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, CA, USA) with RIN number >7.0. RNA-seq libraries were prepared by using 200 ng total RNA with TruSeq RNA sample prep kit (Illumina). Oligo(dT)-enriched mRNAs were fragmented randomly with fragmentation buffer, followed by first- and second-strand cDNA synthesis. After a series of terminal repair, the double-

stranded cDNA library was obtained through PCR enrichment and size selection. cDNA libraries were sequenced with the Illumina Hiseq 2000 sequencer (Illumina HiSeq 2000 v4 Single-Read 50 bp) after pooling according to its expected data volume and effective concentration. Two biological replicates were performed in the RNA-seq analysis. Raw reads were then aligned to the mouse genome (GRCm38) using Tophat2 RNA-seq alignment software, and unique reads were retained to quantify gene expression counts from Tophat2 alignment fles. The differentially expressed mRNAs and genes were selected with log2 (fold change) >1 or log2 (fold change) <-1 and with statistical significance (p value < 0.05) by R package. Bioinformatic analysis was performed using the OmicStudio tools athttps://www.omicstudio.cn/tool.

Preparation of tumor DNA samples

GFP^+CD19^+ splenic cells, tail and kidney tissue were obtained from *cRAG1, cRAG2* and *fRAG BCR-ABL1*$^+$ B-ALL mice, and genomic DNA was extracted using a TIANamp Genomic DNA Kit (TIANGEN-DP304). Subsequently, paired-end libraries were constructed from 1 μg of the initial genomic material using the TruSeq DNA v2 Sample Prep Kit (Illumina, #FC-121-2001) as per the manufacturer's instructions. The size distribution of the libraries

was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, #5067-4626), and the DNA concentration was quantified using a Qubit dsDNA HS Assay Kit (Life Technologies, #Q32851). The Illumina HiSeq 4000 was utilized to sequence the samples, with two to four lanes allocated for sequencing the tumor and one lane for the control DNA library of the kidney or liver, each with 150 bp paired end reads.

Read alignment and structural variant calling

Fastq files were generated using Casava 1.8 (Illumina), and BWA 37 was employed to align the reads to mm9. PCR duplicates were eliminated using Picard's Mark Duplicates tool *(sourceforge.net/apps/mediawiki/picard)*. Our custom scripts (http://sourceforge.net/projects/svdetection) were utilized to eliminate BWA-designated concordant and read pairs with low BWA mapping quality scores. Intrachromosomal and inter-chromosomal rearrangements were identified using SV Detect from discordant, quality prefiltered read pairs. The mean insertion size and standard deviation for this analysis were obtained through Picard's InsertSizeMetrics tool (sourceforge.net/apps/mediawiki/picard). Tumor-specific structural variants (SVs) were identified using the manta software

(https://github.com/Illumina/manta/blob/mater/docs/userGuide/README.md#introduction).

Validation of high confidence off-target candidates

The elimination of non-specific structural mutations from the kidney or tail was necessary for tumor-specific structural variants identification. Subsequently, the method involving 21-bp CAC-to-breakpoint was employed to filter RAG-mediated off-target gene. The validation of high confidence off-target candidates was carried out through PCR. Oligonucleotide primers were designed to hybridize within the "linking" regions of SV Detect, in the appropriate orientation. The PCR product was subjected to Sanger sequencing and aligned to the mouse mm9 reference genome using BLAST *(https://blast.ncbi.nlm.nih.gov/Blast.cgi)*.

Statistics.

Statistical analysis was conducted using SPSS 20.0 (IBM Corp.) and GraphPad Prism 6.0 (GraphPad Software). Descriptive statistics were reported as means ± standard deviation for continuous variables. The equality of variances was assessed using Levene's test. Two-group comparisons, multiple group comparisons, and survival comparisons were performed using independent-samples

t-test, one-way analyses of variance (ANOVA) with post hoc Fisher's LSD test, and log-rank Mantel-Cox analysis, respectively. Kaplan-Meier survival curves were utilized to depict the changes in survival rate over time. Statistical significance was set at $P<0.05$.

Results

cRAG give more aggressive leukemia in a mouse model of *BCR-ABL1*$^+$ B-ALL

In order to assess the impact of RAG activity on the clonal evolution of *BCR-ABL1*$^+$ B-ALL through a genetic experiment, we utilized bone marrow transplantation (BMT) to compare disease progression in *fRAG, cRAG1*, and *cRAG2 BCR-ABL1*$^+$ B-ALL (***Yu, et al.*,*2019***). Bone marrow cells transduced with a BCR-ABL1/GFP retrovirus were administered into syngeneic lethally irradiated mice, and CD19$^+$ B cell leukemia developed within 30-80 days (Fig. 1A, Fig. S1). Western blot results confirmed equivalent transduction efficiencies of the retroviral *BCR-ABL1* in all three cohorts (Fig. S2A). In order to investigate potential variances in leukemia outcome across distinct genomic backgrounds, we used Mantel-Cox estimation to assess survival rates in *fRAG, cRAG1*, or *cRAG2* mice that were transplanted with *BCR-ABL1*-transformed bone

marrow cells. Our findings indicate that, compared to *fRAG BCR-ABL1*$^+$ B-ALL mice, *cRAG1* or *cRAG2 BCR-ABL1*$^+$ B-ALL mice exhibited reduced survival rates during the primary transplant phase (median 74.5 days versus 39 or 57 days, $P < 0.0425$, Fig. 1A). This survival rates discrepancy was also observed during the secondary transplant phase, wherein leukemic cells were extracted from the spleens of primary recipients and subsequently purified via GFP$^+$ cell sorting. A total of 10^5,10^4 and 10^3 GFP$^+$ leukemic cells that originated from *fRAG, cRAG1*, or *cRAG2* leukemic mice were transplanted into corresponding non-irradiated immunocompetent syngenetic recipient mice (median survival days 11-26, 10-16, 11-21 days, $P < 0.0023$-0.0299, Fig. S2B). Additionally, the *cRAG* mice exhibited a significantly higher leukemia burden in the bone marrow, spleen, and peripheral blood compared to the *fRAG* mice (Fig. 1B-D). In order to investigate the cellular process underlying the increased growth rate in *cRAG BCR-ABL1*$^+$ B-ALL, a flow cytometry analysis was conducted to examine the cell cycle and cell apoptosis status. The results demonstrated that *cRAG BCR-ABL1*$^+$ B-ALL had elevated proportion of cells in S/G2-M phase compared to fRAG (Fig. 1E). Furthermore, we observed increased growth due to decreased apoptosis in *cRAG* leukemic cells (Fig. S2C). RNA-seq analysis revealed the changes of

cell differentiation and proliferation/apoptotic pathways (Fig. S3) These results suggest that the absence non-core RAG regions of expedites the progression of malignant transformation and leukemic growth, resulting in aggressive disease phenotype in the cRAG *BCR-ABL1*$^{+}$ B-ALL model.

The loss of non-core RAG regions corresponds to a less mature cell surface phenotype but does not impede IgH VDJ recombination

To identify the B cells developmental stages from which the accumulated B leukemic cells originated. We stained single cells with B cell-specific surface markers and analyzed the samples on flow cytometry. We observed that 91%-98% of GFP^{+} cells from *cRAG* mice were CD19^{+}BP-1^{+}B220^{+}CD43^{+}, suggesting that the majority of leukemic cells belonged to the large pre-B cell stage. However, in *fRAG* leukemic mice, there were 65% large pre-B (GFP^{+}CD19^{+}BP-1^{+}B220^{+}CD43^{+}) and 35% small pre-B cells (GFP^{+}CD19^{+}BP-1^{+}B220^{+}CD43^{-}) (Fig. 2A). The lineage results in accordance with the immunoglobulin μ heavy chain (μHC) expression. Specifically, 5% of *fRAG* leukemic cells exhibited μHC expression, while *cRAG* leukemic cells lacked μHC expression, indicating a deficiency in the pre-BCR checkpoint (Fig. 2B). These re-

sults suggest that leukemic cells derived from *cRAG* mice arrest at early B cell developmental stage. Typically, *IgH* rearrangement initiates with *D-J* joining in pro-B cells, followed by *V-DJ* joining in large pre-B cells, and ultimately, *V-J* rearrangements occur at the *IgL* loci in small pre-B cells. Genomic PCR of $GFP^{+}CD19^{+}$ cell's DNA was used to investigate *VDJ* rearrangement. Notably, *cRAG* leukemic cells exhibited a high degree of oligoclonality, as all tumors analyzed consistently displayed the rearrangement of a limited number of VH family members. In contrast, the *fRAG* leukemias exhibited a significant degree of polyclonality, as confirmed by the recurrent rearrangement of multiple *VH* family members investigated, each rearranged to all possible *JH1-3* segments (Fig. S4AB). This finding is consistent with the more aggressive disease phenotype observed in *cRAG BCR-ABL1*$^{+}$ B-ALL. The progression of BCR-ABL-induced leukemia in *cRAG* mice necessitates the selection of secondary oncogenic events, ultimately leading to the emergence of one or a few dominant leukemic clones. Loss of the non-core RAG region results in the emergence of fewer leukemic clones to generate oligoclonal tumors.

The loss of non-core RAG regions highlights genomic DNA damage

The aforementioned findings indicate that leukemic cells derived

from three kind of mice were hindered in the large pre-B phase to varying extents, instead of consistent development. During this phase, typical B cells show reduced RAG1 expression to accommodate DNA replication and rapid cell proliferation. The non-core RAG regions contain the RING finger domain and T490 residue, which are responsible for regulating RAG degradation in B cell development. Therefore, it is imperative to investigate the potential consequences of non-core region deletion on RAG expression and function in these leukemic cells. Western blotting was done to address this question. The results revealed that RAG1 (cRAG1) and RAG2 (cRAG2) were expressed in $GFP^{+}CD19^{+}$ splenic leukemic cells derived from *BCR-ABL1*+ B-ALL mice with varying genetic backgrounds (Fig. 3A). Notably, upregulation of the RAG1 (cRAG1) protein was observed in *cRAG1* leukemic cells compared to *fRAG* (Fig. 3A, Fig. S5A). The in vitro V(D)J recombination assay confirmed that rearrangements mediated by RSS occurred in leukemic cells. This finding suggests that different forms of RAG exhibited cleavage activity (Fig. 3B and Fig. S5B).

To investigate the potential correlation between aberrant RAG activities and increased DNA double-strand breaks (DSBs), we evaluated the levels of phosphorylated H2AX (□-H2AX), a DSB response factor, in *fRAG, cRAG1*, and *cRAG2* leukemic cells (gat-

ed on GFP$^+$). This served as a measure of both DNA DSBs and global genomic instability. Our flow cytometry findings revealed that *cRAG* leukemic cells had increased □-H2AX compared to the *fRAG* compartment (Fig. 3C). This suggests that cRAG play a more significant role in mediating somatic structural variants in *BCR-ABL1*$^+$ B lymphocytes. These results indicate that the stalled B-precursors exhibit high expression of RAG endonucleases and increased DNA damage.

Off-target recombination mediated by RAG in *BCR-ABL1*$^+$ B lymphocytes

Genome-wide sequencing and analysis were performed to compare somatic structural variants (SVs) in *BCR-ABL1*$^+$ B lymphocytes derived from *fRAG, cRAG1,* and *cRAG2* mice. The leukemic cells were sequenced with an average coverage of 25× (Table. S1). The SVs generated by RAG were screened based on two criteria: the presence of a CAC to the right (or GTG to the left) of both breakpoints, and its occurrence within 21 bp from the breakpoint (***Mijušković, et al.*,*2015***). Further elaboration on these criteria can be found in Supplementary Figure 6. Consequently, aberrant *V*-to-*V* junctions and *V* to intergenic regions were encompassed in five validated abnormal rearrangements at *Ig* loci in *cRAG* leukemic mice (Table. S2). Additionally, seven samples had 24 somatic

structural variations, with an average of 3.4 coding region mutations per sample (range of 0-9), which is consistent with the limited number of acquired somatic mutations observed in hematological cancers and childhood malignancies (Fig. 4 and Table S3). The results of the study demonstrate that *fRAG* cells had low SVs (0-1 per sample), *cRAG1* cells exhibited higher SVs (6-9 per sample) while *cRAG2* cells had moderate SVs incidence (1-4 per sample) (Fig. 4, Table S3). These findings suggest that cRAG may lead to an elevated off-target recombination, eventually posing a threat to the *BCR-ABL1*$^{+}$ B lymphocytes genome.

Off-target V(D)J recombination characteristics in *BCR-ABL1*$^{+}$ B lymphocytes

We further analyzed the characteristics of the identified SVs. Specifically, an assessment of the exon-intron distribution profiles of 42 breakpoints engendered by 24 SVs was executed via genome analysis. The results indicated that while 57% of the breakpoints were situated on the gene body, 43% were enriched within the flanking sequence, majority of which were identified as transcriptional regulatory sequence (Figure 5A). P and N nucleotides are recognized as distinctive characteristics of V(D)J recombination. Consequently, the length of P and N nucleotides remains consistent during RSS-to-RSS and cRSS-to-cRSS recombination.

However, the frequency of P and N sequences was 50%/50% (P/N) in RSS-to-RSS recombination, while it was 4%/8% (P/N) in cRSS-to-cRSS recombination (Fig. 5B). The notably reduced frequency of P and N sequences indicating that off-target sites DNA repair in *BCR-ABL1*$^{+}$ B lymphocytes differs from classical V(D)J recombination repair.

The hybrid joints were specifically pronounced in *cRAG1* and *cRAG2* leukemic cells (100% and 93% respectively), indicating that the non-core regions might be involved in suppressing potentially harmful transposition events (Figure 5C). In order to ascertain the potential impact of non-core RAG region deletion on the occurrence of oncogenic mutations, a comparative analysis of cancer genes was performed across three distinct leukemic cell backgrounds. The results revealed that the number of cancer genes produced in the cRAG1 leukemic cell was significantly higher than the other two. This observation aligns with the manifestation of a most aggressive leukemic phenotype and concurrent alterations in mRNA transcription in *cRAG1 BCR-ABL1*$^{+}$ B-ALL mice.

The non-core regions have effects on RAG binding accuracy and off-target recombination size in *BCR-ABL1*$^{+}$ B lymphocytes

Sequence logos were used to visually compare the RSS and cRSSs in *Ig* loci and *non-Ig* loci respectively. The RSS elements in *Ig* loci exhibited the closest match to the canonical RSS (CACAGTG [12/23 spacer] ACAAAAACC), particularly at functionally important positions. It is noteworthy that the first 5 bases (underlined) of the perfect heptamer sequence CACAGTG served as the binding motif of fRAG, while the first 4 bases of the heptamer sequence, the CACA tetranucleotide, were identified as the cRAG binding motif in leukemic cells (Fig. 6A). Although both motifs (CACAG and CACA) corresponding to the highly conserved portion of the RSS heptamer sequence, variations in the cRSSs sequence among off-target genes in *fRAG* and *cRAG* mice indicate that the removal of RAG's non-core region reduces binding precision and increases the off-target recombination in *BCR-ABL1*$^{+}$ B lymphocytes.

Antigen receptor genes are assembled by large-scale deletions and inversions. Our investigation revealed that both *fRAG* and *cRAG2* leukemic cells generated 100% and 92% off-target recombination, respectively, exceeding 10000 bp. However, *cRAG1* leukemic cells exhibited only 6% off-target recombination exceeding 10000

bp, with 48% being <1000bp, 46% being 1000-10000bp (Fig. 6BC). The results indicate that cRAG1 generate minor size of off-target recombination in *BCR-ABL1*$^+$ B lymphocytes, and non-core RAG1 region influences the off-target recombination magnitude. Additionally, non-core RAG1 region deletion results in reduced off-target recombination size, which may account for the higher incidence of off-target V(D)J recombination in *cRAG1* leukemic cells (Fig. 6D).

Discussion

In this study, we have demonstrated that non-core region deletion of both RAG1 and RAG2 leads to accelerated development of leukemia and increased off-target V(D)J recombination in *BCR-ABL1*$^+$ B lymphocytes. Furthermore, we report reduced cRAG binding accuracy and off-target recombination size in cRAG1 leukemia cells, which might contribute to exacerbated off-target V(D)J recombination of *cRAG BCR-ABL1*$^+$ B lymphocytes. These findings suggest that the non-core regions, particularly the non-core region of RAG1, play a crucial role in maintaining accuracy of V(D)J recombination and genomic stability in *BCR-ABL1*$^+$ B lymphocytes.

Our observations indicate that cRAGs leukemic cells exhibit a heightened production of hybrid joints, and that the non-core RAG regions might suppress hybrid joint generation in vivo. Post-

cleavage synaptic complexes (PSCs) consist of the RAGs, coding ends, and RSS ends (***Fugmann, et al.*,*2000***; ***Libri, et al.*,*2021***). It is likely that RAG evolution has resulted in the formation of PSCs with optimal conformation and/or stability for standard coding and RSS end-joining. Conversely, cRAGs PSCs may facilitate RAG-mediated hybrid joints by enabling the close proximity of coding and RS ends or increasing the PSC stability. Furthermore, it is possible for fRAGs to enlist disassembly/remodeling factors to PSCs, a process that may facilitate the involvement of NHEJ factors in the completion of the normal reaction (***Fugmann, et al.*,*2000***). Within this context, cRAGs might exhibit reduced recruitment capacity due to alterations in overall conformation or the absence of specific motifs, resulting in the formation of more stable PSCs and an increased potential for the accumulation of incomplete hybrid joints. However, our results showed that over 90% junction were hybrid joints in *cRAGs* leukemic cells, surpassing the frequency previously reported in literature (***Raghavan, et al.*,*2006***; ***Talukder, et al.*,*2004***). Studies have indicated that deficiency in non-homologous end joining (NHEJ) may lead to chromosomal instability and lymphomagenesis (***Wiegmans, et al.*,*2021***; ***Gaymes, et al.*,*2002***; ***Rassool*,*2003***; ***Scully, et al.*,*2019***). Notably, our findings have revealed significant variations in the NHEJ repair path-

way among leukemic cells with different genetic backgrounds, suggesting aberrant expression of DNA repair pathways in *cRAGs* leukemic cells (Fig. S3B). This observation suggests the potential for cRAGs to generate elevated levels of hybrid joints, particularly in the absence of a competing normal pathway for efficient formation of coding and RSS joins in a NHEJ-aberrant background.

The off-target V(D)J recombination process mediated by RAG has the potential to generate oncogenic rearrangements (***Mijušković, et al.***,***2015***; ***Greaves***,***2018***; ***Thomson, et al.***,***2020***). Our study reveals that non-core RAG regions deletion increased the off-target oncogenic genes, particularly in *cRAG1* leukemic cells. These genes have the capacity to influence cell proliferation, differentiation, or survival. In line with this, the leukemic cells exhibit augmented cell proliferation and reduced cell apoptosis with alterations in the cell cycle pathway, particularly in *cRAG1* leukemic cells (Fig. S3DF). *CDKN2B* is susceptible to recurrent breakage in *cRAG1* leukemic cells, and its function in impeding proliferation and enhancing leukemic cells apoptosis is attributed to its capacity to block *CDK6* (***Lopes-Ventura, et al.***,***2019***; ***Suzuki, et al.***,***1995***). *CDKN2B* reduction and *CDK6* elevation in *cRAG1* leukemic cells have been verified (Fig. S7AB), thereby suggesting that off-target V(D)J recombination generates known or suspected oncogenic

573 mutations. Nevertheless, the degree to which RAG-mediated on-
574 cogenic recombination contributes to leukemia necessitates further
575 examination.

576 In human *ETV6-RUNX1* ALL, the *ETV6-RUNX1* fusion gene is
577 believed to initiate prenatally, yet the disease remains clinically la-
578 tent until critical secondary events occur, leading to leukemic trans-
579 formation-"pre-leukemia to leukemia" (***Mori, et al.*,*2002***; ***Bateman,***
580 ***et al.*,*2010***; ***Bhojwani, et al.*,*2012***). Genomic rearrangement, me-
581 diated by aberrant RAG recombinase activity, is a frequent driver
582 of these secondary events in ETV6-RUNX1 ALL (***Papaemmanuil,***
583 ***et al.*,*2014***). In contrast, RAG mediated off-target V(D)J recombi-
584 nation is also observed in *BCR-ABL1*$^{+}$ B-ALL. These oncogenic
585 structural variations can also be considered as secondary events
586 that promote the transition -"leukemia to aggressive leukemia".
587 The enhancement of *BCR-ABL1*$^{+}$ B-ALL deterioration and pro-
588 gression by cRAG in mouse model was consistent with our previ-
589 ous study that RAG enhances BCR-ABL1 positive leukemic cell
590 growth through its endonuclease activity. Additionally, we showed
591 that non-core RAG1 region deletion leads to increased cRAG1 ex-
592 pression and high RAG expression related to low survival in pedi-
593 atric acute lymphoid leukemia (Fig. 3A and Fig. S8). Therefore,
594 more attention should be be paid to the non-core RAG region mu-

595 tation in *BCR-ABL1*$^{+}$ B-ALL for the role of non-core region in leu-
596 kemia suppression and off-target V(D)J recombination.

597 Disclosure of Potential Conflicts of Interest

598 The authors declare no potential conflicts of interest.

599 Authors' Contributions

600 Yanghong Ji: Conceptualization, resources, data curation, funding
601 acquisition, validation, writing-review, and editing. Xiaozhuo Yu
602 and Wen Zhou: Conceptualization, validation, visualization, meth-
603 odology, writing-original draft, writing-review, and editing. Xiaodong
604 Chen: validation, writing-review, and editing. Shunyu He: method-
605 ology, writing-review, and editing. Mengting Qin: writing-review,
606 and editing. Meng Yuan: validation, writing-review, and editing.
607 Yang Wang: validation, writing-review and editing. Woodvine
608 otieno Odhiambo: writing-review and editing. YinSha Miao: funding,
609 validation, writing-review, and editing.

610 Acknowledgments

611 This study was supported by grants (no. 31170821, no. 31370874
612 and no. 81670157) from the National Natural Scientific Foundation

of China and by a grant (no. 2016JZ030) from the Natural Scientific Foundation of Shaanxi. The authors would like to thank Professor Shaoguang Li from the Division of Hematology/Oncology, University of Massachusetts Medical School, for providing the MSCV-BCR-BAL1-IRES-GFP construct. The authors would also like to thank Mr. Xiaofei Wang (Xi'an Jiaotong University Health Science Centre) for providing expert technical assistance with cell sorting.

References:

Banerjee JK, Schatz DG. 2014. Synapsis alters RAG-mediated nicking at Tcrb recombination signal sequences: implications for the "beyond 12/23" rule. *MOL CELL BIOL* **34**:2566-2580. doi:10.1128/MCB.00411-14

Bateman CM, Colman SM, Chaplin T, Young BD, Eden TO, Bhakta M, Gratias EJ, van Wering ER, Cazzaniga G, Harrison CJ, Hain R, Ancliff P, Ford AM, Kearney L, Greaves M. 2010. Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *BLOOD* **115**:3553-3558. doi:10.1182/blood-2009-10-251413

Becker LM, O'Connell JT, Vo AP, Cain MP, Tampe D, Bizarro L, Sugimoto H, McGow AK, Asara JM, Lovisa S, McAndrews KM, Zielinski R, Lorenzi PL, Zeisberg M, Raza S, LeBleu VS, Kalluri R. 2020. Epigenetic Reprogramming of Cancer-Associated Fibroblasts Deregulates Glucose Metabolism and Facilitates Progression of Breast Cancer. *CELL REP* **31**:107701. doi:10.1016/j.celrep.2020.107701

Beilinson HA, Glynn RA, Yadavalli AD, Xiao J, Corbett E, Saribasak H, Arya R, Miot C, Bhattacharyya A, Jones JM, Pongubala J, Bassing CH, Schatz DG. 2021. The RAG1 N-terminal region regulates the efficiency and pathways of synapsis for V(D)J recombination. *J EXP MED* **218**doi:10.1084/jem.20210250

Bhojwani D, Pei D, Sandlund JT, Jeha S, Ribeiro RC, Rubnitz JE, Raimondi SC, Shurtleff S, Onciu M, Cheng C, Coustan-Smith E, Bowman WP, Howard SC, Metzger ML, Inaba H, Leung W, Evans WE, Campana D, Relling MV, Pui CH. 2012. ETV6-RUNX1-positive childhood acute lymphoblastic leukemia: improved outcome with contemporary therapy. *LEUKEMIA* **26**:265-270. doi:10.1038/leu.2011.227

Curry JD, Schlissel MS. 2008. RAG2's non-core domain contributes to the ordered regulation of V(D)J recombination. *NUCLEIC ACIDS RES* **36**:5750-5762. doi:10.1093/nar/gkn553

Deng Z, Liu H, Liu X. 2015. RAG1-mediated ubiquitylation of histone H3 is required for chromosomal V(D)J recombination. *CELL RES* **25**:181-192. doi:10.1038/cr.2015.1

Eastman QM, Leu TM, Schatz DG. 1996. Initiation of V(D)J recombination in vitro obeying the 12/23 rule. *NATURE* **380**:85-88. doi:10.1038/380085a0

Fatma H, Maurya SK, Siddique HR. 2022. Epigenetic modifications of c-MYC: Role in cancer cell reprogramming, progression and chemoresistance. *SEMIN CANCER BIOL* **83**:166-176. doi:10.1016/j.semcancer.2020.11.008

Fugmann SD, Lee AI, Shockett PE, Villey IJ, Schatz DG. 2000. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *ANNU REV IMMUNOL* **18**:495-527. doi:10.1146/annurev.immunol.18.1.495

Gan T, Wang Y, Liu Y, Schatz DG, Hu J. 2021. RAG2 abolishes RAG1 aggregation to facilitate V(D)J recombination. *CELL REP* **37**:109824. doi:10.1016/j.celrep.2021.109824

Gaymes TJ, North PS, Brady N, Hickson ID, Mufti GJ, Rassool FV. 2002. Increased error-prone non homologous DNA end-joining--a proposed mechanism of chromosomal instability in Bloom's syndrome. *ONCOGENE* **21**:2525-2533. doi:10.1038/sj.onc.1205331

Goel S, Bhatia V, Biswas T, Ateeq B. 2022. Epigenetic reprogramming during prostate cancer progression: A perspective from development. *SEMIN CANCER BIOL* **83**:136-151. doi:10.1016/j.semcancer.2021.01.009

Greaves M. 2018. A causal mechanism for childhood acute lymphoblastic leukaemia. *NAT REV CANCER* **18**:471-484. doi:10.1038/s41568-018-0015-6

Hirokawa S, Chure G, Belliveau NM, Lovely GA, Anaya M, Schatz DG, Baltimore D, Phillips R. 2020. Sequence-dependent dynamics of synthetic and endogenous RSSs in V(D)J recombination. *NUCLEIC ACIDS RES* **48**:6726-6739. doi:10.1093/nar/gkaa418

Khoshchehreh R, Totonchi M, Carlos RJ, Torres R, Baharvand H, Aicher A, Ebrahimi M, Heeschen C. 2019. Epigenetic reprogramming of primary pancreatic cancer cells counteracts their in vivo tumourigenicity. *ONCOGENE* **38**:6226-6239. doi:10.1038/s41388-019-0871-x

Kumari R, Roy U, Desai S, Nilavar NM, Van Nieuwenhuijze A, Paranjape A, Radha G, Bawa P, Srivastava M, Nambiar M, Balaji KN, Liston A, Choudhary B, Raghavan SC. 2021. MicroRNA miR-29c regulates RAG1 expression and modulates V(D)J recombination during B cell development. *CELL REP* **36**:109390. doi:10.1016/j.celrep.2021.109390

Lewis SM, Agard E, Suh S, Czyzyk L. 1997. Cryptic signals and the fidelity of V(D)J joining. *MOL CELL BIOL* **17**:3125-3136. doi:10.1128/MCB.17.6.3125

Libri A, Marton T, Deriano L. 2021. The (Lack of) DNA Double-Strand Break Repair Pathway Choice During V(D)J Recombination. *FRONT GENET* **12**:823943. doi:10.3389/fgene.2021.823943

Liu C, Zhang Y, Liu CC, Schatz DG. 2022. Structural insights into the evolution of the RAG recombinase. *NAT REV IMMUNOL* **22**:353-370. doi:10.1038/s41577-021-00628-6

Liu Y, Subrahmanyam R, Chakraborty T, Sen R, Desiderio S. 2007. A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *IMMUNITY* **27**:561-571. doi:10.1016/j.immuni.2007.09.005

Lopes-Ventura S, Pojo M, Matias AT, Moura MM, Marques IJ, Leite V, Cavaco BM. 2019. The efficacy of HRAS and CDK4/6 inhibitors in anaplastic thyroid cancer cell lines. *J ENDOCRINOL INVEST* **42**:527-540. doi:10.1007/s40618-018-0947-4

Lu C, Ward A, Bettridge J, Liu Y, Desiderio S. 2015. An autoregulatory mechanism imposes allosteric control on the V(D)J recombinase by histone H3 methylation. *CELL REP* **10**:29-38.

690 doi:10.1016/j.celrep.2014.12.001
691 Matthews AG, Kuo AJ, Ramón-Maiques S, Han S, Champagne KS, Ivanov D, Gallardo M, Carney D,
692 Cheung P, Ciccone DN, Walter KL, Utz PJ, Shi Y, Kutateladze TG, Yang W, Gozani O, Oettinger MA.
693 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *NA-*
694 *TURE* **450**:1106-1110. doi:10.1038/nature06431
695 Mendes RD, Sarmento LM, Canté-Barrett K, Zuurbier L, Buijs-Gladdines JG, Póvoa V, Smits WK,
696 Abecasis M, Yunes JA, Sonneveld E, Horstmann MA, Pieters R, Barata JT, Meijerink JP. 2014. PTEN
697 microdeletions in T-cell acute lymphoblastic leukemia are caused by illegitimate RAG-mediated re-
698 combination events. *BLOOD* **124**:567-578. doi:10.1182/blood-2014-03-562751
699 Mijušković M, Chou YF, Gigi V, Lindsay CR, Shestova O, Lewis SM, Roth DB. 2015. Off-Target
700 V(D)J Recombination Drives Lymphomagenesis and Is Escalated by Loss of the Rag2 C Terminus.
701 *CELL REP* **12**:1842-1852. doi:10.1016/j.celrep.2015.08.034
702 Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C, Hows JM, Navarrete C, Greaves M.
703 2002. Chromosome translocations and covert leukemic clones are generated during normal fetal de-
704 velopment. *Proc Natl Acad Sci U S A* **99**:8242-8247. doi:10.1073/pnas.112218799
705 Onozawa M, Aplan PD. 2012. Illegitimate V(D)J recombination involving nonantigen receptor loci in
706 lymphoid malignancy. *Genes Chromosomes Cancer* **51**:525-535. doi:10.1002/gcc.21942
707 Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, Alexandrov LB, Van Loo P, Cooke
708 SL, Marshall J, Martincorena I, Hinton J, Gundem G, van Delft FW, Nik-Zainal S, Jones DR, Rama-
709 krishna M, Titley I, Stebbings L, Leroy C, Menzies A, Gamble J, Robinson B, Mudie L, Raine K,
710 O'Meara S, Teague JW, Butler AP, Cazzaniga G, Biondi A, Zuna J, Kempski H, Muschen M, Ford AM,
711 Stratton MR, Greaves M, Campbell PJ. 2014. RAG-mediated recombination is the predominant driver
712 of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *NAT GENET* **46**:116-
713 125. doi:10.1038/ng.2874
714 Raghavan SC, Tong J, Lieber MR. 2006. Hybrid joint formation in human V(D)J recombination re-
715 quires nonhomologous DNA end joining. *DNA Repair (Amst)* **5**:278-285.
716 doi:10.1016/j.dnarep.2005.09.008
717 Rassool FV. 2003. DNA double strand breaks (DSB) and non-homologous end joining (NHEJ) path-
718 ways in human leukemia. *CANCER LETT* **193**:1-9. doi:10.1016/s0304-3835(02)00692-4
719 Rooney S, Chaudhuri J, Alt FW. 2004. The role of the non-homologous end-joining pathway in lym-
720 phocyte development. *IMMUNOL REV* **200**:115-131. doi:10.1111/j.0105-2896.2004.00165.x
721 Sadofsky MJ, Hesse JE, McBlane JF, Gellert M. 1993. Expression and V(D)J recombination activity of
722 mutated RAG-1 proteins. *NUCLEIC ACIDS RES* **21**:5644-5650. doi:10.1093/nar/21.24.5644
723 Schabla NM, Perry GA, Palmer VL, Swanson PC. 2018. VprBP (DCAF1) Regulates RAG1 Expression
724 Independently of Dicer by Mediating RAG1 Degradation. *J IMMUNOL* **201**:930-939.
725 doi:10.4049/jimmunol.1800054
726 Schatz DG, Ji Y. 2011. Recombination centres and the orchestration of V(D)J recombination. *NAT*
727 *REV IMMUNOL* **11**:251-263. doi:10.1038/nri2941
728 Schatz DG, Swanson PC. 2011. V(D)J recombination: mechanisms of initiation. *ANNU REV GENET*
729 **45**:167-202. doi:10.1146/annurev-genet-110410-132552
730 Schjerven H, Ayongaba EF, Aghajanirefah A, McLaughlin J, Cheng D, Geng H, Boyd JR, Eggesbø
731 LM, Lindeman I, Heath JL, Park E, Witte ON, Smale ST, Frietze S, Müschen M. 2017. Genetic analy-
732 sis of Ikaros target genes and tumor suppressor function in BCR-ABL1(+) pre-B ALL. *J EXP MED*
733 **214**:793-814. doi:10.1084/jem.20160049

Schlissel MS, Corcoran LM, Baltimore D. 1991. Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. *J EXP MED* **173**:711-720. doi:10.1084/jem.173.3.711

Scully R, Panday A, Elango R, Willis NA. 2019. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* **20**:698-714. doi:10.1038/s41580-019-0152-0

Sekiguchi JA, Whitlow S, Alt FW. 2001. Increased accumulation of hybrid V(D)J joins in cells expressing truncated versus full-length RAGs. *MOL CELL* **8**:1383-1390. doi:10.1016/s1097-2765(01)00423-3

Silver DP, Spanopoulou E, Mulligan RC, Baltimore D. 1993. Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. *Proc Natl Acad Sci U S A* **90**:6100-6104. doi:10.1073/pnas.90.13.6100

Suzuki H, Zhou X, Yin J, Lei J, Jiang HY, Suzuki Y, Chan T, Hannon GJ, Mergner WJ, Abraham JM, Et A. 1995. Intragenic mutations of CDKN2B and CDKN2A in primary human esophageal cancers. *HUM MOL GENET* **4**:1883-1887. doi:10.1093/hmg/4.10.1883

Talukder SR, Dudley DD, Alt FW, Takahama Y, Akamatsu Y. 2004. Increased frequency of aberrant V(D)J recombination products in core RAG-expressing mice. *NUCLEIC ACIDS RES* **32**:4539-4549. doi:10.1093/nar/gkh778

Teng G, Maman Y, Resch W, Kim M, Yamane A, Qian J, Kieffer-Kwon KR, Mandal M, Ji Y, Meffre E, Clark MR, Cowell LG, Casellas R, Schatz DG. 2015. RAG Represents a Widespread Threat to the Lymphocyte Genome. *CELL* **162**:751-765. doi:10.1016/j.cell.2015.07.009

Thomson DW, Shahrin NH, Wang P, Wadham C, Shanmuganathan N, Scott HS, Dinger ME, Hughes TP, Schreiber AW, Branford S. 2020. Aberrant RAG-mediated recombination contributes to multiple structural rearrangements in lymphoid blast crisis of chronic myeloid leukemia. *LEUKEMIA* **34**:2051-2063. doi:10.1038/s41375-020-0751-y

Wiegmans AP, Ward A, Ivanova E, Duijf P, Adams MN, Najib IM, Van Oosterhout R, Sadowski MC, Kelly G, Morrical SW, O'Byrne K, Lee JS, Richard DJ. 2021. Genome instability and pressure on non-homologous end joining drives chemotherapy resistance via a DNA repair crisis switch in triple negative breast cancer. *NAR Cancer* **3**:b22. doi:10.1093/narcan/zcab022

Wong S, Witte ON. 2004. The BCR-ABL story: bench to bedside and back. *ANNU REV IMMUNOL* **22**:247-306. doi:10.1146/annurev.immunol.22.012703.104753

Yu X, Zhang H, Yuan M, Zhang P, Wang Y, Zheng M, Lv Z, Odhiambo WO, Li C, Liu C, Ma Y, Ji Y. 2019. Identification and characterization of a murine model of BCR‐ABL1+ acute B‐lymphoblastic leukemia with central nervous system metastasis. *ONCOL REP* **42**:521-532. doi:10.3892/or.2019.7184

Zhang L, Reynolds TL, Shan X, Desiderio S. 2011. Coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis. *IMMUNITY* **34**:163-174. doi:10.1016/j.immuni.2011.02.003

Figure legends

Figure 1. cRAGs give more aggressive leukemia in mice model of *BCR-ABL1*$^{+}$ B-ALL

(A) Kaplan-Meier survival curve for *fRAG* (n=8), *cRAG1* (n=6), and *cRAG2* (n=10) recipient mice. The survival was calculated by Mantel–Cox test (P<0.0425). (B) The spleen weights of *fRAG, cRAG1* and *cRAG2* leukemic mice (*fRAG*, n=8, *cRAG1*, n=7, *cRAG2*, n=9; *fRAG vs cRAG1*, *P*<0.0001,

797 f*RAG vs cRAG2*, *P*=0.1352). (C) The spleen cell numbers of *fRAG, cRAG1*
798 and *cRAG2* leukemic mice (*fRAG*, n=7, *cRAG1*, n=8, *cRAG2*, n=13; *fRAG* vs
799 *cRAG1*, *P*=0.0047, *fRAG* vs *cRAG2*, *P*=0.0180). (D) The percentage of GFP^+
800 cells in peripheral blood (PB, *fRAG*, n=6, *cRAG1*, n=6, *cRAG*2, n =6; *fRAG* vs
801 *cRAG1*, *P*=0.0003, *fRAG* vs *cRAG2*, *P*=0.0035), bone marrow (BM, *fRAG*,
802 n=5, *cRAG1*, n=5, *cRAG2*, n=6; *fRAG* vs *cRAG1*, *P*=0.0341, *fRAG* vs *cRAG2*,
803 *P*=0.0008), and spleen (SP, *fRAG*, n=9, *cRAG1*, n=4, *cRAG2*, n=9; *fRAG* vs
804 *cRAG1*, *P*=0.0016, *fRAG* vs *cRAG2*, *P*<0.0001) of *fRAG, cRAG1* and *cRAG2*
805 leukemic mice. (E) Representative flow cytometry plots of cell cycle arrest of
806 leukemic cells in *fRAG, cRAG1* and *cRAG2* mice. In the graph, the percent-
807 ages of each phase of the cell cycle are summarized below (*fRAG*, n=3,
808 *cRAG1*, n=5, *cRAG2*, n=5; G0/G1, *fRAG* vs *cRAG1*, *P*=0.0082, *fRAG* vs
809 *cRAG2*, *P*=0.0279; S, *fRAG* vs *cRAG1*, *P*=0.0146, *fRAG* vs *cRAG2*, *P*=0.0370;
810 G2/M, *fRAG* vs *cRAG1*, *P*=0.0134, *fRAG* vs *cRAG2*, *P*=0.1507). In figures B,
811 C, D and J, error bars represent the mean ± s.d., *P* values were calculated by
812 Student's t test and **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

813

814 **Figure 2. The non-core RAG region loss corresponds to a less**
815 **mature cell surface phenotype**

816 (A) Flow cytometry analysis of the B cell markers CD19, BP-1, B220, and
817 CD43 on *BCR-ABL1*-transformed *fRAG, cRAG1* and *cRAG2* leukemic bone
818 marrow cells. The percentages of each phase of the B cell stage are summa-
819 rized in the bottom graph (*fRAG*, n=9, *cRAG1*, n=4, *cRAG2*, n=9; Large-preB,
820 *fRAG* vs *cRAG1*, *P*=0.0349, *fRAG* vs *cRAG2*, *P*=0.0017; Small-pre-B, *fRAG*
821 vs *cRAG1*, *P*=0.0141, *fRAG* vs *cRAG2*, *P*=0.0005). The expression of the cy-
822 toplasmic μ chain was analyzed by flow cytometry. Representative samples
823 are shown in (B), and the results from multiple samples analyzed in inde-
824 pendent experiments are summarized in the bottom graph as the fraction of
825 cells expressing cytoplasmic factors (*fRAG*, n=11, *cRAG1*, n=8, *cRAG2*, n=8;

826 *fRAG* vs *cRAG1*, *P*=0.3020, *fRAG* vs *cRAG2*, *P*=0.2267). Error bars represent
827 the mean ± s.d., P values were calculated by Student's t test and **P* < 0.05,
828 ***P* < 0.01, ****P* < 0.001.

829

830 **Figure 3. The non-core RAG region loss highlights genomic**
831 **DNA damage**

832 (A) Western blotting analysis showed RAG1 and RAG2 expression in
833 GFP^+CD19^+ leukemic cells originating from *BCR-ABL1*$^+$ B-ALL in different ge-
834 netic backgrounds. (B) Rearrangement substrate retrovirus was transduced
835 into leukemic cells. Flow cytometry was used to analyze the percentage of
836 CD90.1 and hCD4 positive cells, and the percentage populations are shown
837 in the bottom graph (*fRAG*, n=3, *cRAG1*, n=3, *cRAG2*, n=3; *fRAG* vs *cRAG1*,
838 *P*=0.0002, *fRAG* vs *cRAG2*, *P*=0.5865). (C) Flow cytometry analysis of □-
839 H2AX levels in f*RAG, cRAG1* and *cRAG2* leukemic cells and the percentage
840 of □-H2AX-positive cell populations shown in the bottom graph (*fRAG*, n=11,
841 *cRAG1*, n=8, *cRAG2*, n=8; *fRAG* vs *cRAG1*, *P*=0.0505, *fRAG* vs *cRAG2*,
842 *P*=0.0094). Error bars represent the mean ± s.d., *P* values were calculated by
843 Student's t test and **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

844

845 **Figure 4. Structural alterations in *BCR-ABL1*+ B lymphocytes**

846 (A-C) Circos plot representation of all off-target recombination detected in the
847 genome-wide analyses of *fRAG, cRAG1* and *cRAG2* leukemic cells. See also
848 Table S3.

849

850 **Figure 5. Overview and characteristics of off-target recombi-**
851 **nation in *BCR-ABL1*$^+$ B-ALL leukemic cells from *fRAG* and**
852 ***cRAG* mice**

(A)Exon-intron distribution profiles of 42 breakpoints generated by 24 SVs. Gene body includes exon (n = 9; 17.3%) and intron (n = 20; 38.5%). Flanking sequence includes 3'UTR (n = 6; 11.5%), 5'UTR (n = 2; 3.8%), promoter (n = 6; 11.5%), and downstream (n = 9; 17.3%). (B) the off-target recombination was filtered and verified by whole genomic sequence and PCR respectively. P nucleotides and N nucleotides of RSS to RSS and cRSS to cRSS were calculated in *BCR-ABL1*$^+$ B-ALL. (C) Hybrid joint percentage generated by either *fRAG*, *cRAG1* or *cRAG2* in *BCR-ABL1*$^+$ B-ALL. It was 0, 100%, and 93% in *fRAG*, *cRAG1* or *cRAG2* leukemic cells respectively. (D)The 24 off-target recombination genes were retrieved by COSMIC Cancer Gene Census (*http://cancer.sanger.ac.uk/census/*). 0.5 genes and 0.5 cancers gene average sample in *fRAG* leukemic cells; 8 genes and 4.5 cancer genes average sample in *cRAG1* leukemic cells; 3.3 genes and 0.3 cancer genes average sample in *cRAG2* leukemic cells.

Figure 6. The non-core regions have effects on RAG binding accuracy and recombinat size in *BCR-ABL1*$^+$ B lymphocytes

(A) Sequence logos were used to compare the RSS and cRSS in Ig loci and non-Ig loci. Top panel: V(D)J recombination at *Ig* locus; the next three panels: RAG-mediated off-target recombination at non-Ig locus from *fRAG, cRAG1* and *cRAG2* leukemic cells respectively. The scale of recombinant size was categorized into three ranges: <1000bp, 1000-10000bp, and >10000bp. The distribution of different recombinant sizes in *fRAG, cRAG1*, and *cRAG2* leukemic cells was presented in (B), while the number of different recombinant sizes in *fRAG, cRAG1*, and *cRAG2* leukemic cells was displayed in (C). (D) A schematic depiction of the mechanism of cRAG-accelerated off-target V(D)J recombination was provided. Both RAG1 and RAG2's non-core region deletion decreases RAG binding accuracy in *cRAG1* and *cRAG2*, *BCR-ABL1*$^+$ B

ALL. Additionally, RAG1's non-core region deletion significantly reduces the size and scale of off-target V(D)J recombination in *cRAG1*, *BCR-ABL1*$^+$ B ALL.

Supplementary Figure 1. Construction of *fRAG, cRAG1* and *cRAG2, BCR-ABL1*$^+$ B-ALL mice models using bone marrow transplantation (BMT)

In the establishment of *BCR-ABL1*$^+$ B-ALL mice models, *fRAG, cRAG1* and *cRAG2* recipient mice after syngeneic lethal irradiation were transplanted with corresponding donor bone marrow cells transduced by *MSCV-BCR-BAL1-IRES-GFP* or *MSCV-GFP* retroviral supernatants. (A) Gross appearance of the spleen in *fRAG, cRAG1,* and *cRAG2* leukemic mice and corresponding control mice. (B) Peripheral blood (PB) and bone marrow (BM) lymphoblastic cells were stained by Wright-Giemsa. The scale bars represent 10 μm. (C) Bone marrow cells from *fRAG, cRAG1* and *cRAG2* leukemic mice were examined by flow cytometry for the expression of GFP and CD19.

Supplementary Figure 2. Biological behavior of leukemia in fRAG, cRAG1 and cRAG2 BCR-ABL1+ B-ALL mouse model

(A) BCR-ABL1 expressions in GFP+CD19+ leukemic cells were determined by western. GAPDH protein was used as a loading control. The K562 and 293T cell lines served as the positive control and negative controls, respectively. (B) Survival of secondary transplant setting. Leukemia cells from primary recipients were recovered from the spleens and purified by GFP$^+$ cell sorting. A total of 10^5,10^4 and 10^3 GFP$^+$ leukemia cells originating from *fRAG, cRAG1* or *cRAG2 BCR-ABL1*$^+$ B-ALL were transplanted into corresponding nonirradiated immunocompetent syngenetic recipient mice (*fRAG*, n=3, *cRAG1*, n=3, *cRAG2*, n=3; $P < 0.0023$-0.0299 by Mantel–Cox test). (C) Apop-

908 tosis was measured by flow cytometry (Annexin V and 7-AAD). The Annexin
909 V+ and 7-AAD- cells were defined as early apoptotic cells, while Annexin V+
910 and 7-AAD+ cells were late apoptotic cells (*fRAG*, n=11, *cRAG1*, n=6, *cRAG2*,
911 n=9; early apoptotic cells: *fRAG vs cRAG1*, *P*=0.0002, *fRAG vs cRAG2*,
912 *P*=0.0026; late apoptotic cells, *fRAG vs cRAG1*, *P*=0.0026, *fRAG vs cRAG2*,
913 *P*<0.0001). Error bars represent the mean ± s.d. *P* values were calculated by
914 Student's t test and *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

915

916 **Supplementary Figure 3. The genetic pathways in *fRAG,***
917 ***cRAG1,* and *cRAG2 BCR-ABL1*+ lymphocytes.**

918 mRNA sequence was performed in GFP and CD19 double positive cells. (A)
919 Principal Component Analysis (PCA) showing the distribution of differentially
920 expressed samples of *fRAG, cRAG1,* and *cRAG2, BCR-ABL1*$^+$ B-ALL. (B)
921 Heatmap of representative different expressed genes related to non-
922 homologous end repair. The scale ranges from minimum (blue) to medium
923 (yellow) to maximum (red) relative expression. (C) Volcano plot depicting log2
924 (fold change) (x-axis) and −log10 (p value) (y-axis) for differentially expressed
925 genes (FC > 2, p < 0.05) in GFP$^+$ CD19$^+$ leukemic cells sorted from *fRAG* and
926 *cRAG1, BCR-ABL1*$^+$ B-ALL mice; upregulated (red) and downregulated (blue).
927 n = 3 per group. (D) The Kyoto Encyclopedia of Genes and Genomes (KEGG)
928 analysis was conducted to identify the differentially expressed genes in
929 *cRAG1 BCR-ABL1*$^+$ B-ALL. The top 15 pathways that exhibited significant dif-
930 ferences were listed in this paragraph. The cell proliferation, apoptosis, and
931 differentiation related pathway were highlighted in red squares. (E) Volcano
932 plot depicting log2 (fold change) (x-axis) and −log10 (p value) (y-axis) for dif-
933 ferentially expressed genes (FC > 2, p < 0.05) in GFP$^+$ CD19$^+$ leukemic cells
934 sorted from *fRAG* and *cRAG2, BCR-ABL1*$^+$ B-ALL mice; upregulated (red)
935 and downregulated (blue). n = 3 per group. (D) KEGG analysis was conduct-
936 ed to identify the differentially expressed genes in *cRAG2 BCR-ABL1*$^+$ ALL.

The top 15 pathways that exhibited significant differences were listed in this paragraph. The cell proliferation, apoptosis, and differentiation related pathway were highlighted in red squares.

Supplementary Figure 4. VDJ recombination in leukemic cells with different genetic backgrounds

(A) VDJ recombination was analyzed by genomic PCR in $GFP^{+}CD19^{+}$ cell's DNA from *fRAG, cRAG1* or *cRAG2* leukemic cells.Genomic DNA from RAG1-/- bone marrow cells and WT spleen was used as negative and positive control respectively.

Supplementary Figure 5. RAG protein expression levels and schematic diagram of the recombinant substrate vector

(A)RAG1/cRAG1 and RAG2 protein levels were compared by western blot and ImageJ software in *fRAG1* and *cRAG1*, B-ALL cells. Error bars represent the mean ± s.d. The *P* value was calculated by t test, ***$P<0.001$, ns $P>0.05$. (B) The B-ALL cells were subjected to transformation with the recombinant substrate vector. In the event of expression of RAG recombinase in the leukemic cells, the RSS sequences flanking CD90.1 would be cleaved by RAG, thereby facilitating the positioning and expression of both CD90.1 and hCD4. In the absence of RAG expression, only hCD4 would be expressed.

Supplementary Figure 6. The criteria for identifying off-target recombination.

We adopted the criteria used in previous studies. First, a CAC must exist to the right (or GTG to the left) of both breakpoints, which includes the four RAG-mediated DNA fragmentation cases mentioned above, and second, it must

964 occur within a specified distance from the breakpoint and the CAC distance-
965 to-breakpoint value was set at 21 bp.

966

967 **Supplementary Figure 7. *CDKN2B* and *CDK6* mRNA levels in**
968 **leukemic cells.**

969 (AB) Leukemic cells were harvested from *fRAG* and *cRAG1* B-ALL mice with
970 *CDKN2B* deletion. *CDKN2B* and *CDK6* were determined by mRNA sequence.
971 Error bars represent the mean ± s.d., *fRAG* B-ALL mice, n = 3; *cRAG1* B-ALL
972 mice n=3. **P* <0.05, ***P* < 0.01

973

974 **Supplementary Figure 8. The relationship of RAG1 mRNA lev-**
975 **els and survival of pediatric acute lymphoid leukemia.**

976 The relationship of RAG1 mRNA levels and survival of pediatric acute lym-
977 phoid leukemia was research by cBioPortal *(https://www.cbioportal.org/)*.
978 *RAG1* mRNA levels were studied in pediatric patients with ALL at the time of
979 diagnosis. The patients were separated into two groups based on mRNA lev-
980 els of *RAG1* (mRNA expression z score relative to diploid sample, RNA se-
981 quence RPKM, RAG1 upregulated group, n=8; RAG1 unaltered group,
982 n=146). The *P* values were calculated from the log-rank test, *P*=0.0732.

983

984

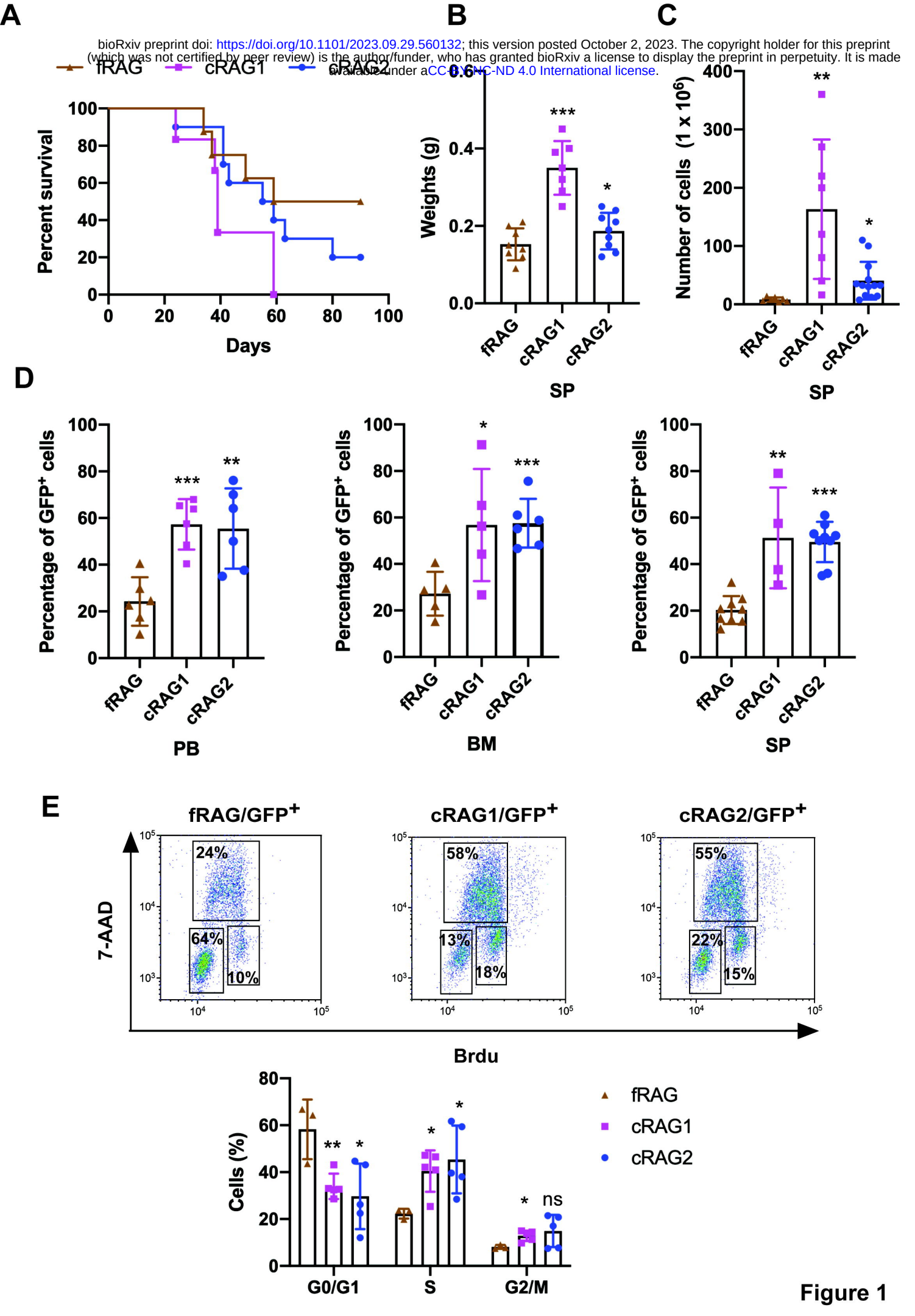


Figure 1

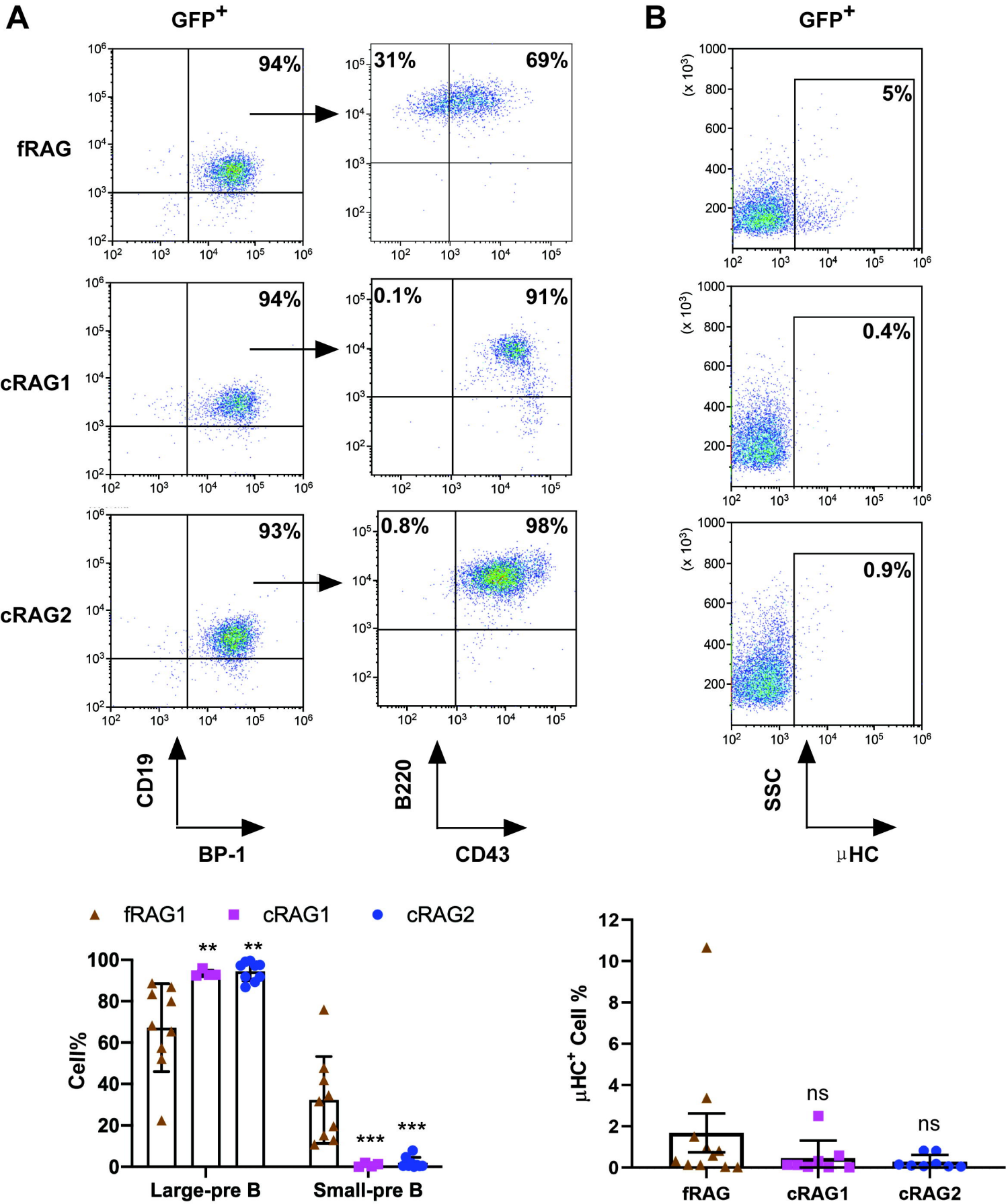

A
GFP+
fRAG
cRAG1
cRAG2
94%
31%
69%
94%
0.1%
91%
93%
0.8%
98%
CD19
BP-1
B220
CD43
B
GFP+
5%
0.4%
0.9%
SSC
µHC
fRAG1
cRAG1
cRAG2
Cell%
Large-pre B
Small-pre B
µHC+ Cell %
ns
ns
fRAG
cRAG1
cRAG2

Figure 2

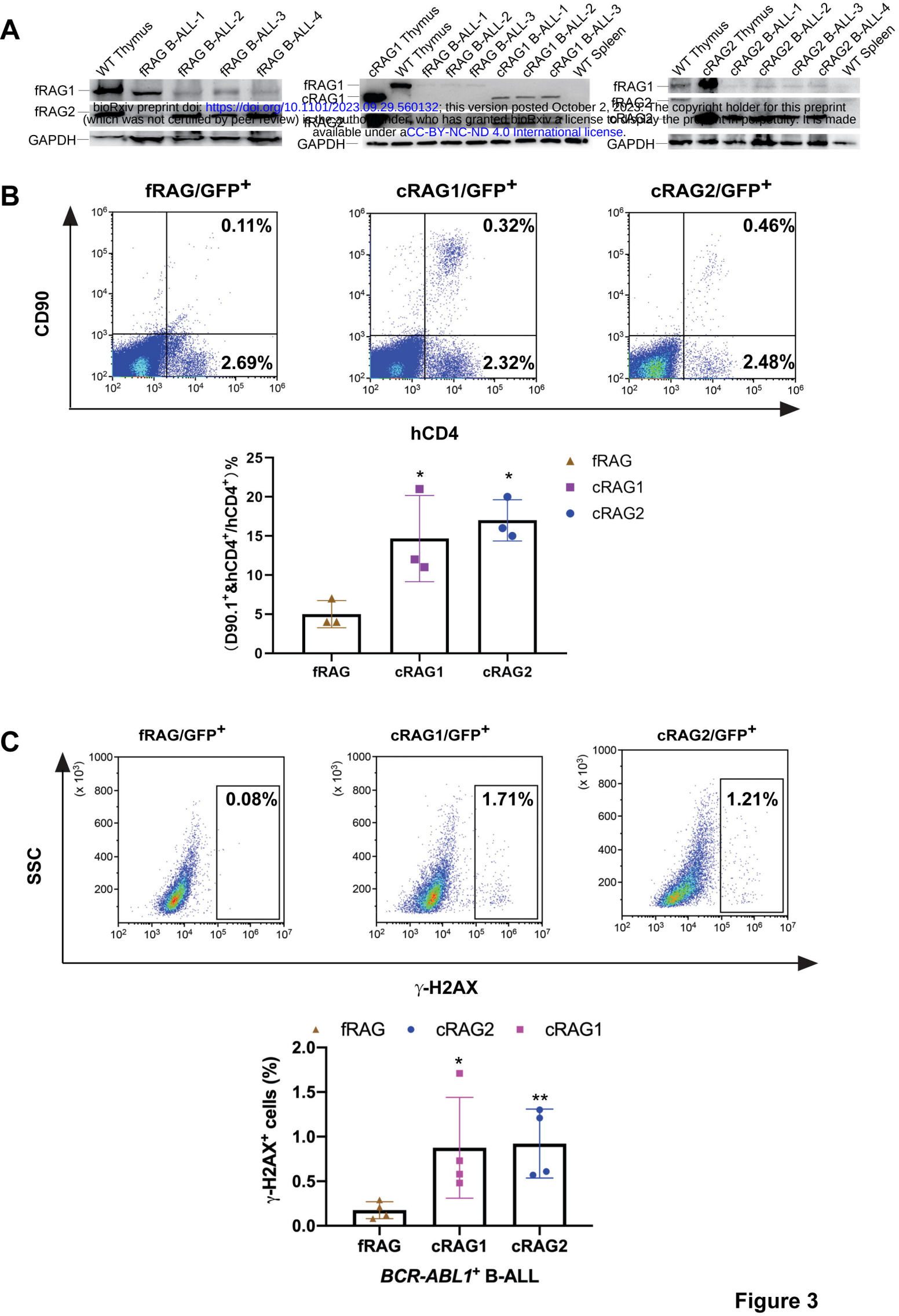
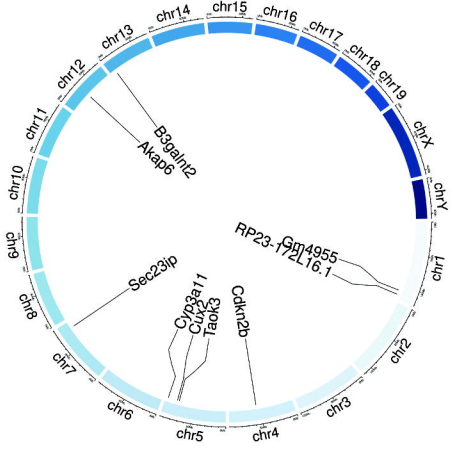
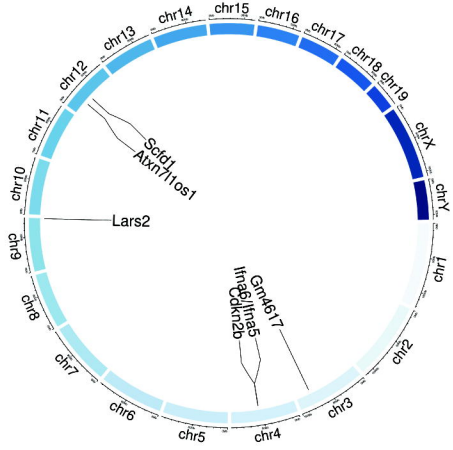


Figure 3

A

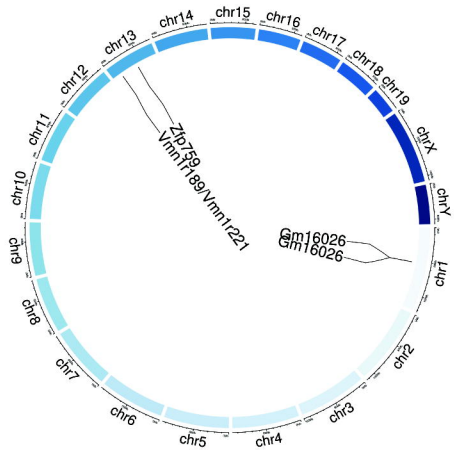


cRAG1 B-ALL-3F

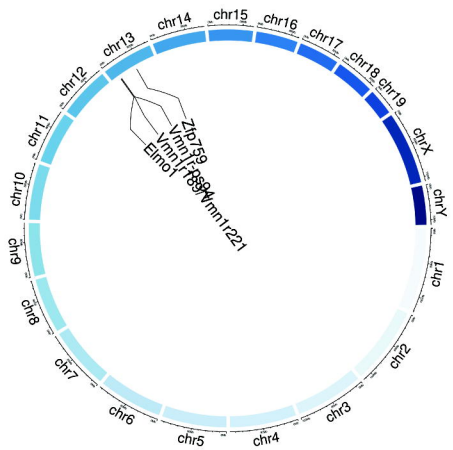


cRAG1 B-ALL-6F

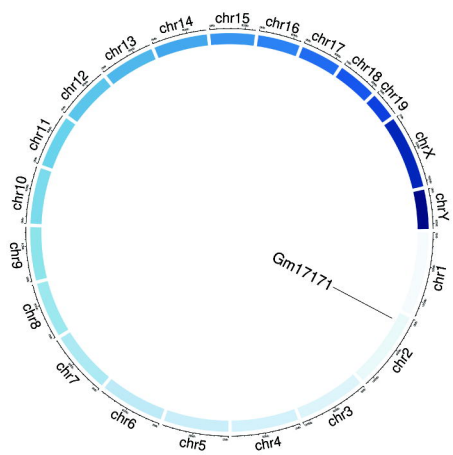
B



cRAG2 B-ALL-3F

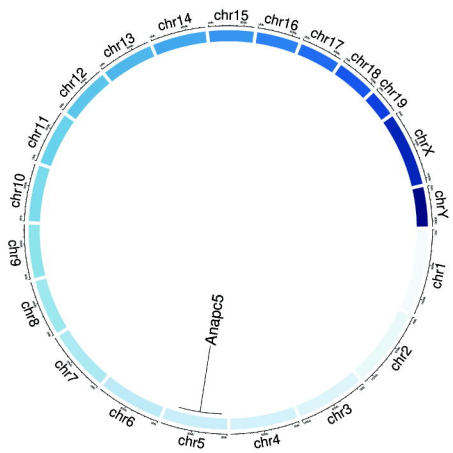


cRAG2 B-ALL-6F

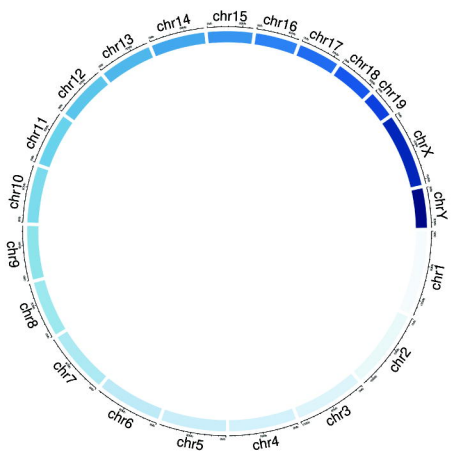


cRAG2 B-ALL-10F

C

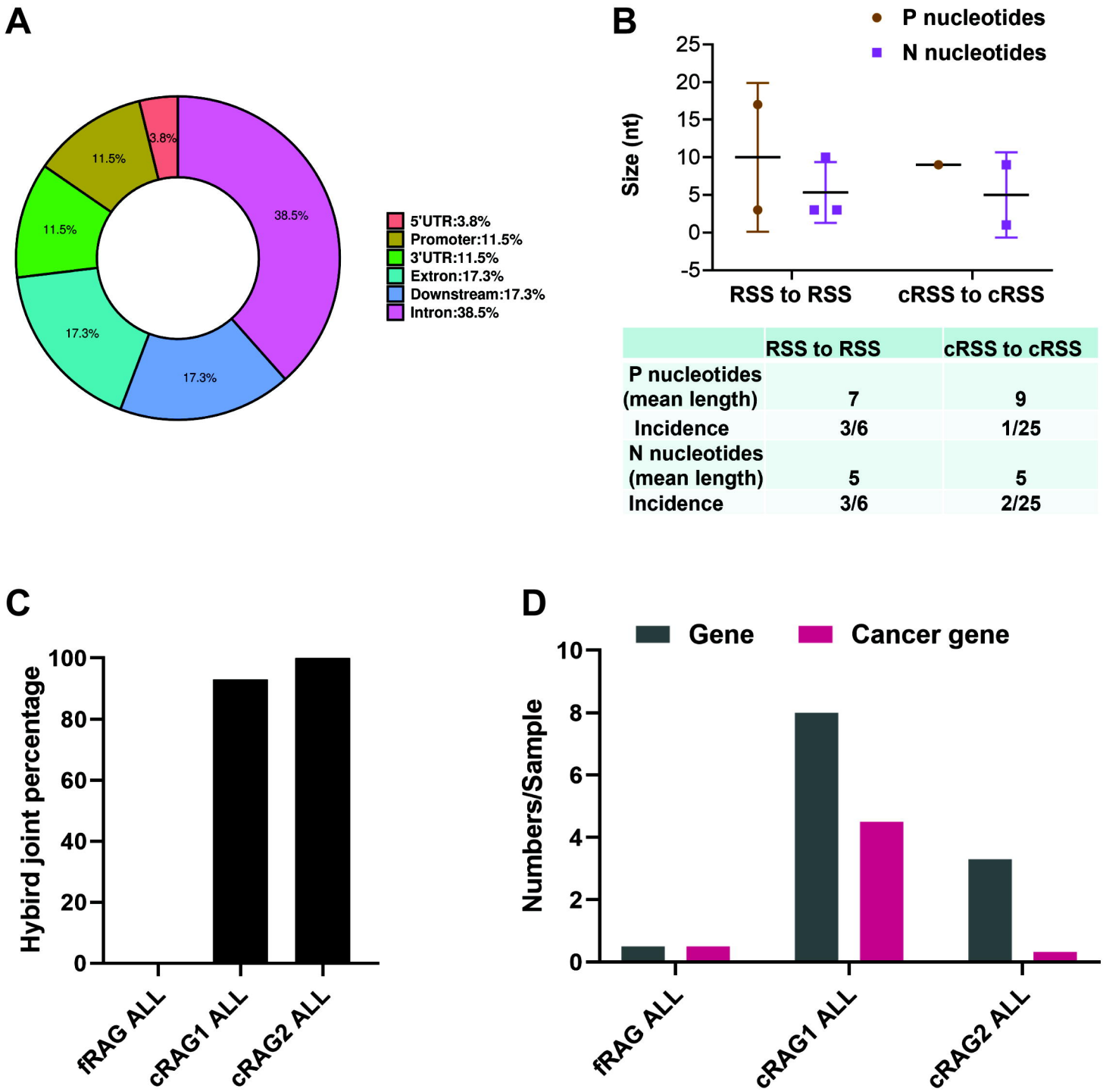


fRAG B-ALL-1F



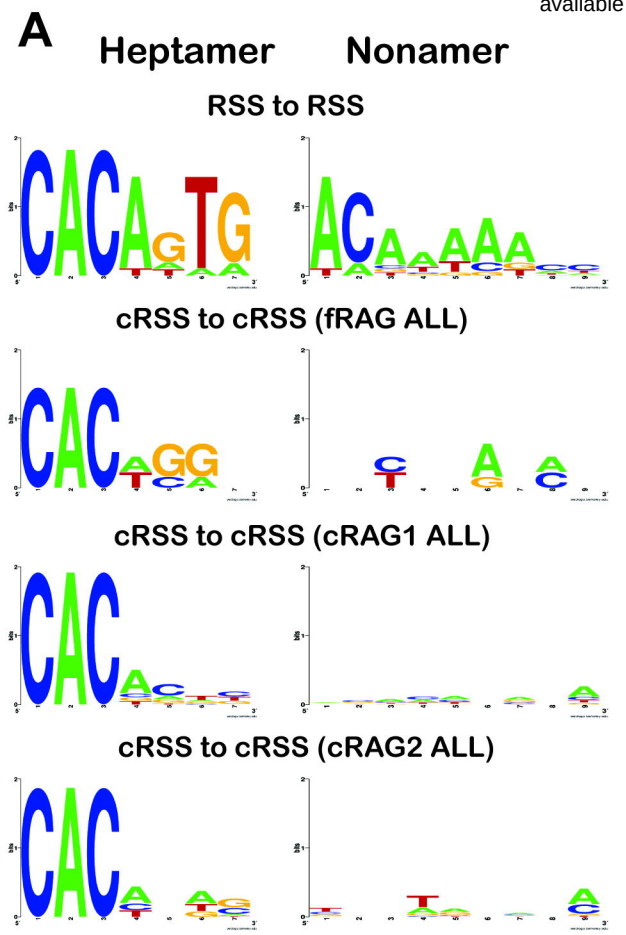
fRAG B-ALL-11F

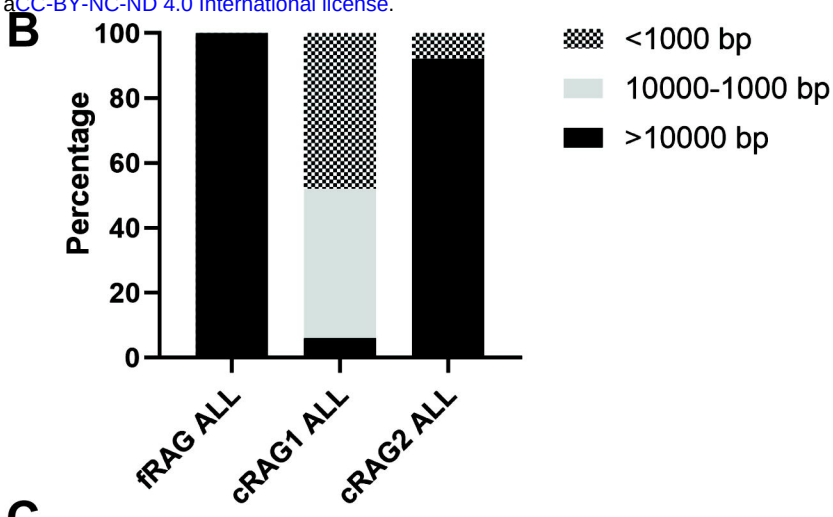
Figure 4

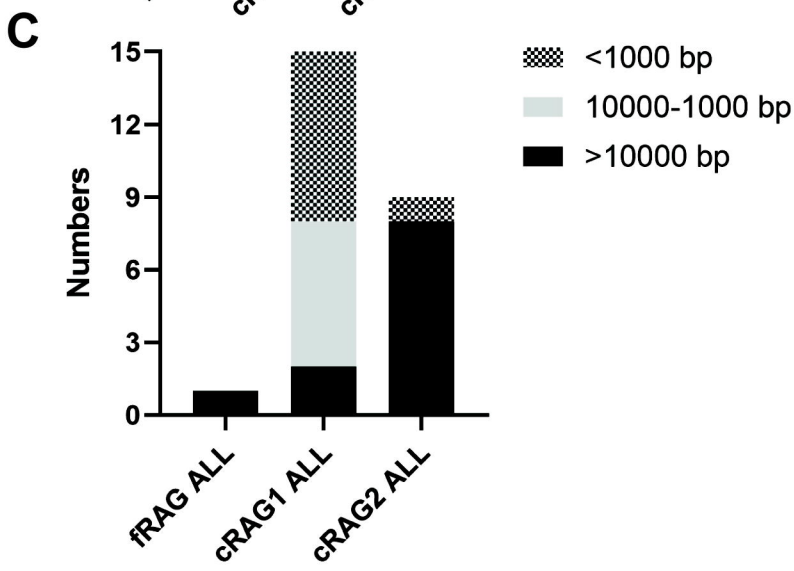


	RSS to RSS	cRSS to cRSS
P nucleotides (mean length)	7	9
Incidence	3/6	1/25
N nucleotides (mean length)	5	5
Incidence	3/6	2/25

Figure 5







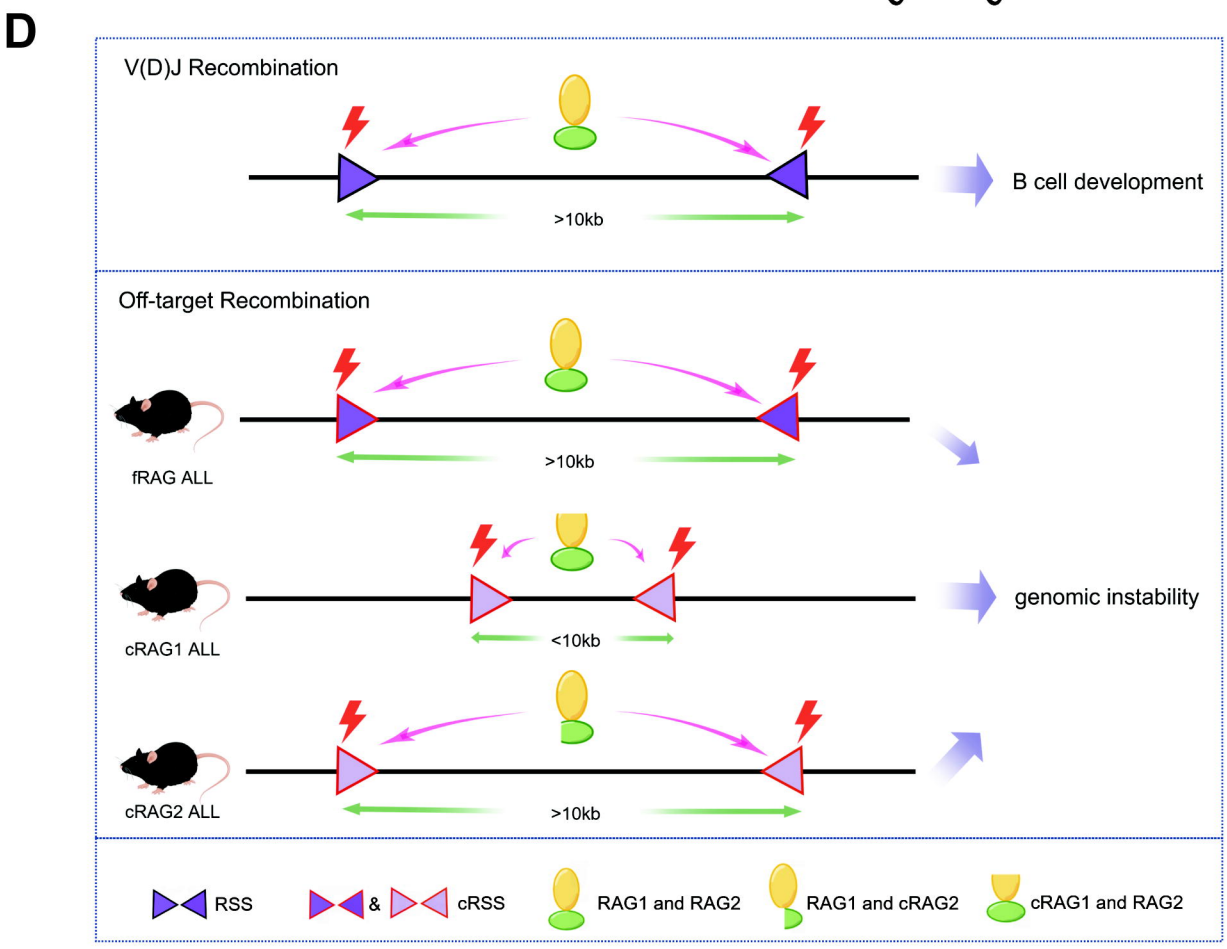


Figure 6