Published in final edited form as:
*J Immunol Methods*. 2012 January 31; 375(1-2): 176–181. doi:10.1016/j.jim.2011.10.007.

# Translocation Capture Sequencing: A Method for High Throughput Mapping of Chromosomal Rearrangements

**Thiago Oliveira**[1,4], **Wolfgang Resch**[2], **Mila Jankovic**[1], **Rafael Casellas**[2,3], **Michel C. Nussenzweig**[1,5], and **Isaac A. Klein**[1,6]

[1]Laboratory of Molecular Immunology, The Rockefeller University, New York, New York 10065, USA.

[2]Genomics & Immunity, NIAMS, National Institutes of Health, Bethesda, MD 20892, USA.

[3]Center for Cancer Research, NCI, National Institutes of Health, Bethesda, MD 20892, USA.

[4]Medical School of Ribeirao Preto/USP, Department of Genetics, 8 National Institute of Science and Technology for Stem Cells and Cell Therapy and Center for Cell-based Therapy, Ribeirao Preto, SP 14051-140, Brazil.

[5]Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10065, USA.

## Abstract

Chromosomal translocations require formation and joining of DNA double strand breaks (DSBs). These events disrupt the integrity of the genome and are involved in producing leukemias, lymphomas and sarcomas. Translocations are frequent, clonal and recurrent in mature B cell lymphomas, which bear a particularly high DNA damage burden by virtue of Activation Induced Cytidine Deaminase (AID) expression. Despite the ubiquity of genomic rearrangements, the forces that underlie their genesis are not well understood. Here, we provide a detailed description of a new method for studying these events, Translocation Capture Sequencing (TC-Seq). TC-Seq provides the means to document chromosomal rearrangements genome-wide in primary cells, and to discover recombination hotspots. Demonstrating its effectiveness, we successfully estimate the frequency of *c-myc/IgH* translocations in primary B cells, and identify hotspots of AID-mediated recombination. Furthermore, TC-Seq can be adapted to generate genome-wide rearrangement maps in any cell type and under any condition.

### Keywords

Chromosomal translocations; genomic instability; Activation-Induced Cytidine Deaminase; B lymphocytes; Lymphoma



[6]Corresponding author. iklein@rockefeller.edu. Address: The Rockefeller University, 1230 York Ave. Box 220. NY, NY, 10065. Telephone: 212-327-8068.



**Accession Numbers**
The TC-Seq datasets are deposited in SRA (http://www.ncbi.nlm.nih.gov/sra) under accession number SRA039959

## 1. Introduction

Gross chromosomal rearrangements, including translocations, are common cytogenetic abnormalities in cancer (Kuppers, 2005; Nussenzweig and Nussenzweig, 2010; Tsai and Lieber, 2010; Tsai et al., 2008; Zhang et al., 2010). These events require formation and joining of DNA double strand breaks (DSBs) and may be oncogenic by several mechanisms (Nussenzweig and Nussenzweig, 2010; Richardson and Jasin, 2000; Zhang et al., 2010). They may juxtapose proto-oncogenes to constitutively active promoters, delete tumor suppressors, or produce chimeric oncogenes (Rabbitts, 2009). In mature B cell cancers, rearrangements are frequent, clonal and recurrent. For example, the *c-myc/IgH* translocation, a hallmark of human Burkitt's lymphoma, deregulates *c-myc* expression by bringing it under the control of Immunoglobulin (*Ig*) gene transcriptional regulatory elements (Casellas et al., 2009; Gostissa et al., 2009; Kuppers, 2005).

Mature B cells are particularly prone to DNA damage by virtue of Activation-Induced Cytidine Deaminase (AID) Expression (Nussenzweig and Nussenzweig, 2010). During B cell activation, AID normally diversifies antibody genes by initiating *Ig* class switch recombination (CSR) and somatic hypermutation (SHM) (Muramatsu et al., 2000; Revy et al., 2000). In the current model, AID deaminates cytosine residues in single-stranded DNA (ssDNA) exposed by stalled RNA polymerase II during transcription (Chaudhuri and Alt, 2004; Pavri et al., 2010; Storb et al., 2007). The resulting U:G mismatches are then processed by one of several repair pathways to yield mutations or DSBs, which are obligate intermediates in CSR (Di Noia and Neuberger, 2007; Honjo, 2002; Peled et al., 2008; Stavnezer et al., 2008).

Although AID has a strong preference for targeting *Ig* genes, it has been shown to target many non-*Ig* loci for mutation (Gordon et al., 2003; Liu et al., 2008; Pasqualucci et al., 2001; Pavri et al., 2010; Robbiani et al., 2009 ; Shen et al., 1998; Yamane et al., 2011). And, while AID over-expression in $p53^{-/-}$ B cells results in widespread genomic instability and translocation-associated lymphoma (Robbiani et al., 2009), *c-myc* is the only gene conclusively shown to suffer AID-mediated rearrangement (Robbiani et al., 2008). AID's role in forming genomic abnormalities remains unclear because until recently, an assay to map DSBs and chromosomal rearrangements did not yet exist (Klein et al., 2011). As such, the forces that underlie the generation of chromosomal rearrangements generally are also not well understood. Sequencing of cancer genomes has enabled the mapping of chromosomal rearrangements in some cases, but genome-wide analysis of single-cell events in non-selected, primary cells presents a major challenge.

While the unbiased amplification of rearrangements has been achieved, the small number of events obtained limited the usefulness of the methods used (Mahowald et al., 2009). Here, we provide a detailed description of Translocation-Capture Sequencing (TC-Seq), a new method for documenting a large number of rearrangements in primary cells (Klein et al., 2011). TC-Seq is able to detect known AID-dependent rearrangements in B cells and identify novel rearrangement hotspots. We propose that TC-Seq may be adapted for use in any cell type to study translocation biology in that tissue.

## 2. Materials and Methods

### 2.1. Mice

Mice bearing I-SceI recognition sites in the first intron of *c-myc* ($Myc^{I\text{-}SceI/I\text{-}SceI}$, hereafter referred to as $Myc^{I}$ (Robbiani et al., 2008)) were used for AID sufficient experiments (Figure 1A). I-SceI is a yeast-derived meganuclease that recognizes an 18bp sequence

absent from the mouse genome. These mice were crossed to $AID^{KO}$ mice to generate $Myc^{I}AID^{KO}$ mice for AID deficient experiments.

### 2.2. Infection and sorting

Resting B lymphocytes were isolated from mouse spleens by immunomagnetic depletion with anti-CD43 MicroBeads (Miltenyi Biotech) and cultured at $0.5 \times 10^6$ cells/ml in RPMI supplemented with L-glutamine, sodium pyruvate, antibiotic/antimycotic, HEPES, 50 μM 2-mercaptoethanol (all from GIBCO-BRL), and 10% fetal calf serum (Hyclone). B cells were stimulated in the presence of 500ng/ml RP105 (BD Pharmingen), 25 μg/ml lipopolysaccharide (LPS) (Sigma) and 5 ng/ml mouse recombinant IL-4 (Sigma). Retroviral supernatants were prepared by cotransfection of BOSC23 cells with pCL-Eco and pMX-IRES-GFP-derived plasmids encoding for I-SceI-mCherry or AID-GFP with Fugene 6, 72 hr before infection. At 20 and 44 hr of lymphocyte culture, retroviral supernatants were added, and B cells were spinoculated at 1150 g for 1.5 hr in the presence of 10 μg/ml polybrene. For dual infection, separately prepared retroviral supernatants were added simultaneously on both days. After 4 hr at 37°C, supernatants were replaced with LPS and IL-4 in supplemented RPMI. At 96 hr from the beginning of their culture, singly infected B cells were collected and frozen in 10 million cell pellets at −80C. Dually infected B cells were sorted for double positive cells with a FACSAria instrument (Becton Dickson) then frozen down.

### 2.3. Primers

LinkerTop - 5’GCAGCGGATAACAATTTCACACAGGACGTACTGTGGCGCGCCT3’.

LinkerBottom - 5’Phospho-GGCGCGCCACAGTACTTGACTGAGCTTTA-ddC3’.

Double-stranded asymmetric linker was generated by annealing LinkerTop and LinkerBottom at 50pmol/ul each in annealing buffer (10mM Tris, 50mM NaCl, 1mM EDTA pH8).

pLinker- 5’GCAGCGGATAACAATTTCACACAGGAC3’.

Site-specific primers for semi-nested PCR on the *c-myc* locus. (Figure 1A)

pMycF1 – 5’Biotin- CAAAATTGGGACAGGGATGTGACC3’.

pMycR1 – 5’Biotin-GGTGTCAAATAATAAGAGACACCTCCC3’.

pMycF2 – 5’CTTGGGGGAAACCAGAGGGAATC3’.

pMycR2 – 5’TACACTCTAAACCGCGACGCC3’.

### 2.4. TC-Seq

See Figure 1B for a schematic of the TC-Seq protocol.

**2.4.1. Genomic DNA extraction and sonication—**Five aliquots of 10 million B cells each were thawed on ice. Cells were resuspended in 100ul of phosphate buffered saline (PBS) each and transferred to a 15ml conical tubes containing 5mL Proteinase K buffer (100mM Tris pH8, 0.2% SDS, 200mM NaCl, 5mM EDTA) and 50ul of 20mg/ml Proteinase K. Cells were incubated overnight at 55C. To extract DNA, addition of 5mL of phenol-chloroform was followed by 30s of gentle inversion and 20 minutes of centrifugation at 2,700g. The aqueous phase was transferred to 50ml conical tubes containing 12.5mL of 100% ethanol, and the mixture was centrifuged for 15 minutes at 2,700g at 4C. The pellets were washed twice with 5mL 70% ethanol, transferred to a microcentrifuge tube, air dried

for 5 minutes, resuspended in 100uL 10mM Tris pH8 with 1uL of 0.5mg/ml DNAse-free RNAse (Roche) and incubated at 25C for 1 hour. Genomic DNA concentration was adjusted to 167ng/uL with 10mM Tris pH8 and further divided into microcentrifuge tubes containing 300uL each. Genomic DNA was fragmented by sonication at low power for 7 cycles (30" on/30" off) in a Bioruptor (Diagenode) to yield a 500–1350bp distribution of DNA fragments. DNA was again divided into 50 aliquots of 30ul (5ug) in 1.5mL microcentrifuge tubes.

**2.4.2. Genomic DNA fragment preparation, linker ligation and elimination of unrearranged loci—**Each experiment consisted of genomic DNA from 50 million B cells in 50 aliquots of 5ug for a total of 250ug of fragmented genomic DNA. The following reactions were performed individually on 5ug aliquots. DNA was blunted by End-It DNA Repair Kit (Epicentre) then purified by Qiagen PCR purification column and eluted in 43uL of Buffer EB. Blunted DNA was adenosine-tailed by addition of 5ul 10x NEB buffer 2, 1uL 10mM dATP and 2uL of 5,000U/ml Klenow fragment 3->5' $exo^-$ (NEB) followed by incubation at 37C for 1hr. Each reaction was purified by Qiagen PCR purification kit and eluted in 40uL of Buffer EB. Each aliquot of blunted, A-tailed DNA fragments was ligated to 200pmol of linkers by addition of 4uL linker, 5uL NEB T4 DNA ligase buffer and 1uL of $2\times10^6$U/ml (high concentration) T4 DNA ligase (NEB) followed by a 1hr incubation at 25C then overnight at 16C. Ligase was inactivated by incubation at 70C for 20 minutes, reactions were purified by Qiagen PCR purification column and eluted in 44uL Buffer EB. Native loci were eliminated by adding 5uL 10x NEB I-Sce1 buffer, 0.5uL 100x BSA and 1uL of 5,000U/ml I-Sce1 (NEB) then incubating the reaction at 37C for 2 hours. Each reaction was purified by Qiagen PCR purification column and eluted in 50uL Buffer EB. Finally, all 50 reactions were pooled.

**2.4.3. Round 1 PCR—**Pooled linker-ligated DNA was divided into 2 equal parts for semi-nested ligation-mediated PCR using either forward or reverse primers. Forward and reverse enrichment streams are kept separate for the entire remainder of the protocol. DNA was divided into 1ug aliquots and each aliquot was mixed with 20ul 5x Phusion HF buffer, 3ul 10mM dNTPs, 1ul 20uM biotinylated pMycF1 or pMycR1, 1uL of 2,000U/mL Phusion Taq (NEB) and $H_2O$ to 100uL. Single-primer PCR reactions were run − 1x(98C-1min) 12x(98C-15sec, 65C-30sec, 72C-45sec) 1x(72C-1min) 1x(4C-forever) - then each tube was spiked with 1ul 20uM pLinker and subjected to additional cycles of PCR - 1x(98C-1min) 35x(98C-15sec, 65C-30sec, 72C-45sec) 1x(72C-5min) 1x(4C-forever). Forward and reverse PCR reactions were pooled separately, mixed, and 400uL of each was purified in 4 Qiagen PCR purification columns, each eluted in 50uL EB. All 200ul of each sample was run on a 2% agarose gel until well separated and appropriately sized fragments (0.3–1 kb) were excised (Figure 2A). DNA was purified in Qiagen gel purification columns, gel-based size selection and purification was repeated once. 100uL of washed T1 magnetic streptavidin beads (Invitrogen) were resuspended in 400ul 2x B&W buffer (10mM Tris pH7.5, 1mM EDTA, 2M NaCl), 200ul was added to each forward and reverse PCR1 products and the mixture was incubated for 1hr with gentle nutation at room temperature. Beads were magnetically isolated, washed 3x in 500uL 1x B&W buffer, 1x in $H_2O$ and resuspended in 87ul $H_2O$. Unrearranged loci were further eliminated by adding 10ul 10x NEB I-Sce1 buffer, 1ul 100x BSA and 2uL of 5,000U/mL I-SceI (NEB), then incubating for 2 hours at 37C. Beads were washed 3 times in 1x B&W, once in $H_2O$ and resuspended in 50uL $H_2O$.

**2.4.4. Round 2 PCR—**20ul of beads from each of the forward and reverse PCR1 was separately mixed with 40ul 5x Phusion HF buffer, 6ul 10mM dNTPs, 2ul 20uM pMycF2 or pMycR2, 2uL 20uM pLinker, 2ul of 2,000U/mL Phusion Taq and 128ul $H_2O$ then separated into 50ul aliquots and subjected to PCR - 1x(98C-1min) 35x(98C-10sec, 65C-30sec,

72C-40sec) 1x(72C-5min) 1x(4C-forever). PCR reactions were magnetically separated from beads and each reaction was run on one lane of a 2% agarose gel until appropriate separation was achieved (Figure 2B). Desired fragment sizes were excised (0.27–1 kb) and purified by Qiagen gel purification column. Size selection was repeated once more in 4 Qiagen gel purification columns for each forward and reverse; each column was eluted with 30uL EB.

**2.4.5. Preparing enriched DNA for paired-end high throughput sequencing—** 120ul of isolated DNA was mixed with 13uL 10x NEB Buffer 4, 2uL of 10,000 U/mL Asc1 (NEB) and $H_2O$ to 150ul then incubated at 37C for 2 hours to cleave the linker. Restriction digests were purified by Qiagen PCR purification column and redigested with Asc1 to ensure complete linker removal. Linker digestion leaves a 6-nucleotide barcode (CGCGCC) on DNA fragments indicating linker-dependent amplification. Desired fragment sizes (0.22–1 kb) were isolated by electrophoresis on a 2% agarose gel and purified by Qiagen gel purification column (Figure 2C). DNA was blunted by End-It DNA Repair Kit (Epicentre), column purified, A-tailed as described above then purified by Qiagen MinElute column and eluted in 24ul of EB. Ligation to Illumina paired-end adapters was performed by addition of 3ul 10x T4 DNA ligase buffer, 1ul PE adapter mix (Illumina) and 3ul of 400,000U/mL T4 DNA ligase (NEB) followed by a 20 minute incubation at 25C then 2 hours at 16C. Ligation reactions were purified by Qiagen PCR purification column then run on a 2% agarose gel and desired fragment sizes were excised (0.3–1kb) then purified by Qiagen gel extraction column. Adapter ligated fragments were enriched by adding 10ul 5x Phusion HF buffer, 1.5ul 10mM dNTPs, 1ul Illumina primer PE1.0, 1ul Illumina primer PE2.0, 21uL $H_2O$ and 0.5uL 2,000U/mL Phusion Taq (NEB), then performing 25 cycles of PCR - 1x(98C-1min) 25x(98C-10sec, 65C-30sec, 72C-40sec) 1x(72C-5min) 1x(4C-forever). Desired fragment sizes were isolated on a 2% agarose gel (0.38–1 kb) (Figure 2D). Samples were confirmed to be enriched for rearranged loci by TOPO cloning (Invitrogen) and screening 10 forward- and 10 reverse-enrichment colonies. Finally, appropriate size distribution was confirmed by Agilent Bioanalyzer. Forward and reverse libraries for the same sample were mixed in an equimolar ratio and sequenced by 36×36 or 54×54 paired-end deep sequencing on an Illumina GAII.

## 2.5. Computational analysis

**2.5.1. Data mining—**All alignments were performed using Bowtie software allowing two mismatches and reporting only unique alignments (option -m1). First, all sequences were aligned against the site-specific primer sequence (bait alignment). Partner sequences were selected based on successful bait alignment and the presence of a linker barcode sequence. Sequences with barcode were trimmed of barcode and designated “target” sequences. Target sequences were aligned against the mouse genome (NCBI 37/mm9) allowing up to 2 mismatches and requiring unique alignments in the best alignment stratum (command line options: -v2 --all --best --strata -m1). Exactly identical alignments (same position, same strand) were combined into a single putative event. “Target” sequence reads begin at the sonication breakpoint; thus similar events from different cells will be associated with different alignments and amplified events from the same cell will be associated with identical alignments. Events supported by a single alignment were not considered in any analyses. We also removed events closer than 1 kb to their respective bait. Translocation positions were given as the position of the 5’ end of the read in the alignment (Figure 3).

**2.5.2. Hotspot discovery and evaluation—**For hotspot discovery, reads aligning within 50 kb of their site-specific bait primer were removed. If translocations were distributed randomly in the genome, then the distances between any two neighboring translocations could be modeled as a negative binomial variable. Therefore, we used a negative binomial test to find any neighboring pairs that were closer than expected by the

negative binomial model (cutoff P(X <= distance| r=1, p) < 0.01; p = number of translocation events / genome size; r = number of translocation events). A set of translocation events in which each consecutive neighbor pair passed the cutoff was considered a hotspot candidate. These candidates were filtered by the following criteria: (a) The number of translocation events had to be at least 3; (b) translocation events originating from either forward or reverse primer had to account for at least 10% of all translocation events in the hotspot, with a minimum of 1; (c) the overall probability of observing a cluster of r translocation events within n nucleotides as determined by the negative binomial model had to be less than $10^{-8}$; (d) for AID-dependent hotspot lists, only those clusters with 10-fold enrichment over the same genomic location in the AID-deficient samples were accepted; (e) Hotspots with a large degree (>80%) of overlap with repeat regions were discarded; (f) Hotspots with small footprints (<100nt) were removed, whereas reads originating from sonicated DNA are unlikely to fall within a small genomic region (Figure 3).

## 3. Results

### 3.1. *c-myc/IgH* translocations detected by TC-Seq

AID initiates DSBs in *c-myc* that partner with breaks in *IgH* to form the *c-myc/IgH* translocation (Robbiani et al., 2008). To determine whether TC-Seq can detect translocation of the I-SceI mediated break in *c-myc* to AID mediated breaks at *IgH*, we examined the *IgH* locus in $Myc^{I}$ B cells expressing retrovirally encoded AID. Indeed, we observe extensive translocation to the switch regions in the presence of AID. In 100 million B cells we detected a total of 5,963 translocations from $Myc^{I}$ to *IgH* with 3,463 to Igμ, 253 to Igγ3 and 2,209 to Igγ1, conversely, we only detect 8 translocations to *IgH* in the absence of AID (Klein et al., 2011). TC-Seq estimates the translocation rate of homozygous I-SceI induced *c-myc* DSBs to AID-initiated DSBs in *IgH* as ~1/17,000. In a previous study that measured the frequency of *c-myc/IgH* translocations in $Myc^{I/+}Igh^{I/+}AID^{-/-}$ B cells infected with I-SceI, the translocation rate between these loci was estimated to be ~1/10,000 (Robbiani et al., 2008). Thus, we conclude that TC-Seq can document rearrangements from I-SceI-induced DSBs in *c-myc* to DSBs in the *IgH* locus at rates that approximate actual translocation frequencies.

### 3.2. AID-dependent rearrangement hotspots detected by TC-Seq

Next, we sought out AID-dependent rearrangement hotspots in non-immunoglobulin loci. We searched the B cell genome for accumulations of reads in AID deficient and sufficient samples (see methods). We found 60 AID-dependent hotspots in 37 genes captured by $Myc^{I}AID^{RV}$ in 100 million B cells. We identified several established AID mutation targets such as mir142 and Gas5 (Klein et al., 2011; Pavri et al., 2010; Robbiani et al., 2009), suggesting that these rearrangements represent bona fide DSB targets of AID. When compared to AID ChIP-Seq data we find that genes bearing AID-dependent hotspots were found to load AID (Klein et al., 2011; Yamane et al., 2011). Thus, TC-Seq and its associated hotspot-finding algorithm can effectively discover non-*Ig* AID targets.

## 4. Discussion

To date, no method had yet been developed to map a large number of chromosomal rearrangements in primary cells. As such, it has been difficult to examine the forces that underlie the formation of these genomic abnormities. For example, while AID has been shown to initiate breaks in *c-myc* and generate widespread genomic instability, the retinue of loci targeted by AID for DSBs remains unknown (Robbiani et al., 2008; Robbiani et al., 2009). To examine chromosomal rearrangements to a fixed DSB in primary cells we have

developed Translocation-Capture Sequencing (TC-Seq) (Klein et al., 2011). Using this technique we have generated a genome-wide rearrangement map to I-SceI sites in *c-myc*, both in the absence and presence of AID. Thereby, we detect AID-mediated rearrangements between *c-myc* and *IgH*. Importantly, TC-Seq detects these events at rates that approximate physiological translocation rates, demonstrating the assay's accuracy.

Additionally, we have developed an associated hotspot-finding algorithm, which can identify genomic sites enriched in rearrangements. Thereby, we identify 60 AID-dependent hotspots that rearrange to DSBs in *c-myc*. Among these are several established mutation targets of AID, and genes shown to suffer AID-mediated rearrangement such as *mir-142* (Robbiani et al., 2009). When all AID-dependent hotspots are examined for AID loading by ChIP-Seq we find that these loci are significantly enriched (Klein et al., 2011; Yamane et al., 2011). Thus, TC-Seq can be successfully applied to discover novel rearrangement hotspots genome-wide. Finally, we propose that TC-Seq is a useful genome-wide method for the isolation, enrichment and analysis of chromosomal rearrangements. When applied to the study of genomic abnormalities in other cell types, this assay will be useful in elucidating the mechanisms by which other cancer-associated translocations arise.

### Highlights

1. A new method, TC-Seq, to map chromosomal rearrangements genome-wide.
2. TC-Seq estimates the actual frequency of c-*myc/IgH* translocations in B cells.
3. TC-Seq discovers AID-dependent rearrangement hotspots.

## Acknowledgments

We thank all the members of the Nussenzweig and Casellas labs for valuable input and advice, Klara Velinzon and Svetlana Mazel for FACSorting and David Bosque and Thomas Eisenreich for animal management. We also thank Scott Dewell of the Rockefeller Genomics Resource Center and Gustavo Gutierrez of the NIAMS genome facility for high-throughput sequencing and guidance. This work was supported by NIH grant #AI037526 to M.C.N., NYSTEM #C023046, The Starr Cancer Consortium and the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health. M.C.N. is an HHMI investigator. I.A.K. was supported by NIH MSTP grant GM07739, and is a Cancer Research Institute Predoctoral Fellow and a William Randolph Hearst Foundation Fellow.

## References

Casellas R, Yamane A, Kovalchuk AL, Potter M. Restricting activation-induced cytidine deaminase tumorigenic activity in B lymphocytes. Immunology. 2009; 126:316–328. [PubMed: 19302140]

Chaudhuri J, Alt FW. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol. 2004; 4:541–552. [PubMed: 15229473]

Di Noia JM, Neuberger MS. Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem. 2007; 76:1–22. [PubMed: 17328676]

Gordon MS, Kanegai CM, Doerr JR, Wall R. Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). Proc Natl Acad Sci U S A. 2003; 100:4126–4131. [PubMed: 12651942]

Gostissa M, Yan CT, Bianco JM, Cogne M, Pinaud E, Alt FW. Long-range oncogenic activation of Igh-c-myc translocations by the Igh 3' regulatory region. Nature. 2009; 462:803–807. [PubMed: 20010689]

Honjo T. Does AID need another aid? Nat Immunol. 2002; 3:800–801. [PubMed: 12205466]

Klein IA, Resch W, Jankovic M, Oliveira T, Yamane A, Nakahashi H, Di Virgilio M, Bothmer A, Nussenzweig A, Robbiani DF, et al. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell. 2011; 147:95–106. [PubMed: 21962510]

Kuppers R. Mechanisms of B-cell lymphoma pathogenesis. Nat Rev Cancer. 2005; 5:251–262. [PubMed: 15803153]

Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG. Two levels of protection for the B cell genome during somatic hypermutation. Nature. 2008; 451:841–845. [PubMed: 18273020]

Mahowald GK, Baron JM, Mahowald MA, Kulkarni S, Bredemeyer AL, Bassing CH, Sleckman BP. Aberrantly resolved RAG-mediated DNA breaks in Atm-deficient lymphocytes target chromosomal breakpoints in cis. Proc Natl Acad Sci U S A. 2009; 106:18339–18344. [PubMed: 19820166]

Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell. 2000; 102:553–563. [PubMed: 11007474]

Nussenzweig A, Nussenzweig MC. Origin of chromosomal translocations in lymphoid cancer. Cell. 2010; 141:27–38. [PubMed: 20371343]

Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Kuppers R, Dalla-Favera R. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature. 2001; 412:341–346. [PubMed: 11460166]

Pavri R, Gazumyan A, Jankovic M, Di Virgilio M, Klein I, Ansarah-Sobrinho C, Resch W, Yamane A, Reina San-Martin B, Barreto V, et al. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell. 2010; 143:122–133. [PubMed: 20887897]

Peled JU, Kuang FL, Iglesias-Ussel MD, Roa S, Kalis SL, Goodman MF, Scharff MD. The biochemistry of somatic hypermutation. Annu Rev Immunol. 2008; 26:481–511. [PubMed: 18304001]

Rabbitts TH. Commonality but diversity in cancer gene fusions. Cell. 2009; 137:391–395. [PubMed: 19410533]

Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Forveille M, Dufourcq-Labelouse R, Gennery A, et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell. 2000; 102:565–575. [PubMed: 11007475]

Richardson C, Jasin M. Frequent chromosomal translocations induced by DNA double-strand breaks. Nature. 2000; 405:697–700. [PubMed: 10864328]

Robbiani DF, Bothmer A, Callen E, Reina-San-Martin B, Dorsett Y, Difilippantonio S, Bolland DJ, Chen HT, Corcoran AE, Nussenzweig A, et al. AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell. 2008; 135:1028–1038. [PubMed: 19070574]

Robbiani DF, Bunting S, Feldhahn N, Bothmer A, Camps J, Deroubaix S, McBride KM, Klein IA, Stone G, Eisenreich TR, et al. AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Mol Cell. 2009; 36:631–641. [PubMed: 19941823]

Shen HM, Peters A, Baron B, Zhu X, Storb U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science. 1998; 280:1750–1752. [PubMed: 9624052]

Stavnezer J, Guikema JE, Schrader CE. Mechanism and Regulation of Class Switch Recombination. Annu Rev Immunol. 2008; 26:261–292. [PubMed: 18370922]

Storb U, Shen HM, Longerich S, Ratnam S, Tanaka A, Bozek G, Pylawka S. Targeting of AID to immunoglobulin genes. Adv Exp Med Biol. 2007; 596:83–91. [PubMed: 17338178]

Tsai AG, Lieber MR. Mechanisms of chromosomal rearrangement in the human genome. BMC Genomics. 2010; 11(Suppl 1):S1. [PubMed: 20158866]

Tsai AG, Lu H, Raghavan SC, Muschen M, Hsieh CL, Lieber MR. Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell. 2008; 135:1130–1142. [PubMed: 19070581]

Yamane A, Resch W, Kuo N, Kuchen S, Li Z, Sun HW, Robbiani DF, McBride K, Nussenzweig MC, Casellas R. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat Immunol. 2011; 12:62–69. [PubMed: 21113164]

Zhang Y, Gostissa M, Hildebrand DG, Becker MS, Boboila C, Chiarle R, Lewis S, Alt FW. The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. Adv Immunol. 2010; 106:93–133. [PubMed: 20728025]

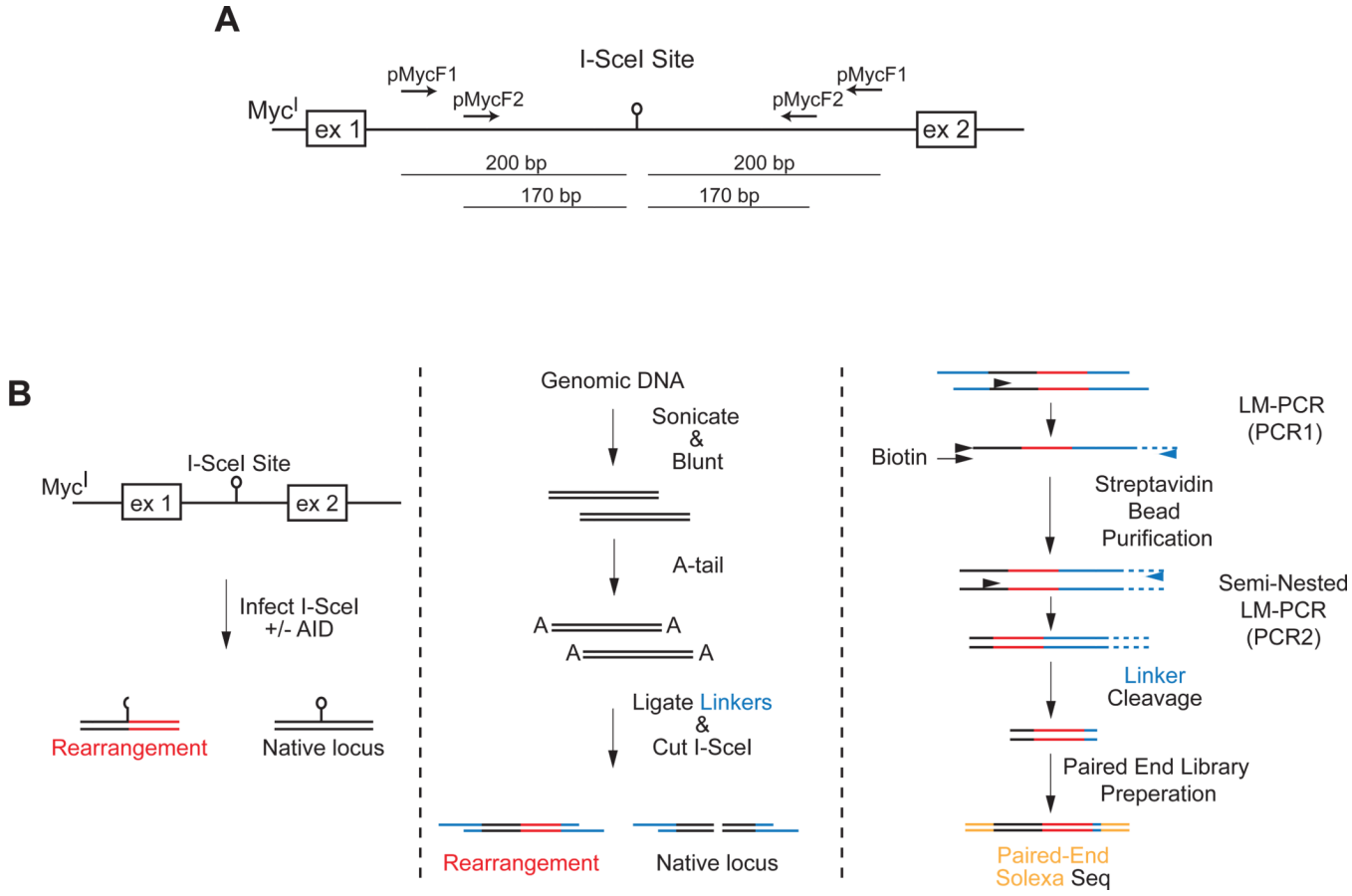


**Figure 1. Translocation Capture Sequencing (TC-seq) Schematic**

(A) The *c-myc* locus in Myc$^{I}$ B cells with associated TC-Seq primers. Primers are shown with their approximate distance from the I-SceI site, which lies between *c-myc* exons 1 and 2. (B) Workflow of TC-Seq. Left panel - rearrangements to the I-SceI site are generated by infecting activated primary B cells with retroviruses expressing I-SceI in the presence or absence of AID. Center panel – genomic DNA from infected and sorted cells are sonicated, blunted, A-tailed, ligated to double-stranded asymmetric linkers and cut with I-SceI to enrich for rearranged loci. Right panel – linker-ligated genomic DNA fragments are first subjected to single-primer PCR to extend the linker in the case of site-specific primer annealing. These products are amplified by linker-mediated PCR with a biotinylated site-specific primer. PCR product is isolated on magnetic streptavidin beads and semi-nested linker-mediated PCR is performed. Finally, linkers are cleaved off enriched fragments then fragments are assembled into a paired-end Illumina library (Klein et al., 2011).

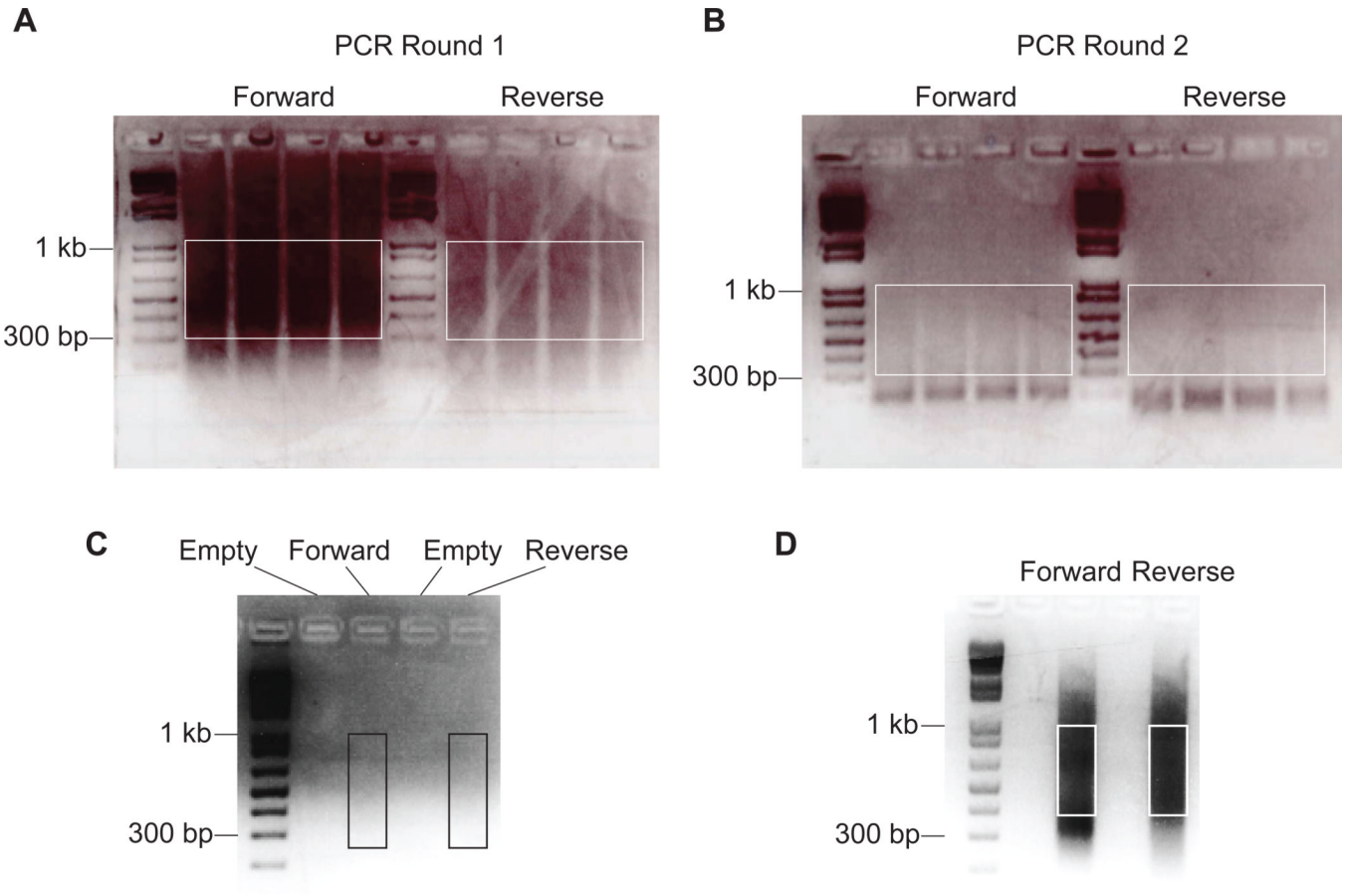


**Figure 2. Purification of High Molecular Weight Products**

At each step high molecular weight products were isolated to ensure that the majority of species contained sufficient sequence to identify a rearrangement partner. All gels are 2% agarose stained with ethidium bromide visualized on an ultraviolet light box. (A) After PCR1 a gel band from 0.3 – 1kb was excised for both forward and reverse enrichments. (B) After PCR2 a gel band from 0.27–1 kb was excised for both forward and reverse enrichments. (C) After linker cleavage a gel band from 0.22–1 kb was excised for both forward and reverse enrichments. At this step the product may not be visible. (D) After PCR with paired-end Illumina primers a band from 0.38–1 kb was excised for both forward and reverse enrichments.

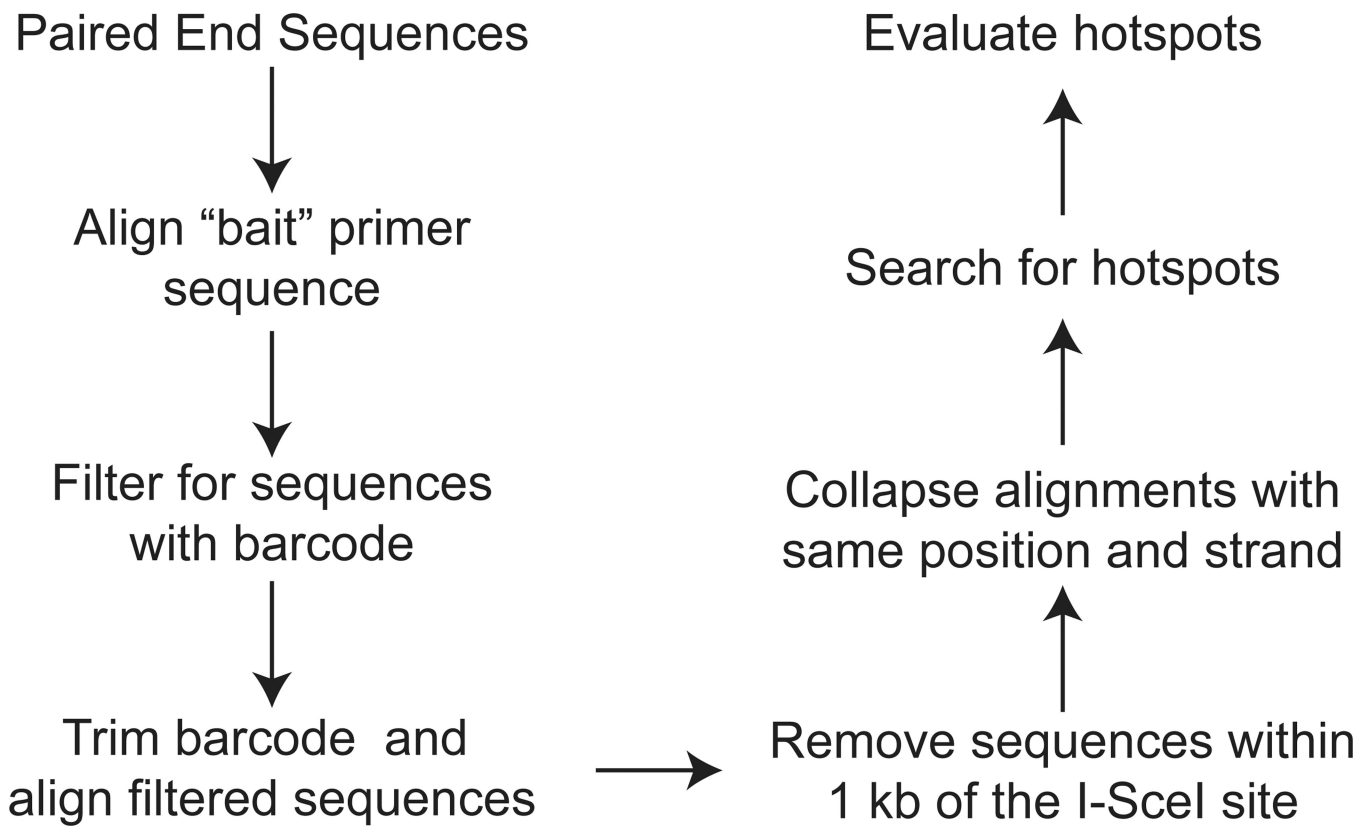


**Figure 3. Workflow Schematic for TC-Seq Computational Analysis**

Paired-end sequences were aligned against the site-specific “bait” primer. Species that contained this primer were checked for a linker barcode on the opposite arm and termed “valid”. Valid reads were trimmed of barcode and aligned to the mouse genome. Reads within 1 kb of the I-SceI site were termed “proximal” and removed from further analyses. Alignments with the same position and strand were assumed to derive from the same biological event and were collapsed into 1 read. Finally, a genome-wide search for hotspots was followed by a hotspot evaluation algorithm that removed supposed artifacts and AID-independent events.