# Somatic Rearrangement in B Cells: It's (Mostly) Nuclear Physics

Erez Lieberman Aiden[1] and Rafael Casellas[2,*]
[1]The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA
[2]Genomics and Immunity, NIAMS and NCI, National Institutes of Health, 10 Center Drive, Bethesda, MD 20892, USA
*Correspondence: rafael.casellas@nih.gov
http://dx.doi.org/10.1016/j.cell.2015.07.034

We discuss how principles of nuclear architecture drive typical gene rearrangements in B lymphocytes, whereas translocation hot spots and recurrent lesions reflect the extent of AID-mediated DNA damage and selection.

Chromosome rearrangements are essential for human development. At immunoglobulin (*Ig*) gene loci in bone marrow B cells, the recombination-activating gene (RAG) proteins mediate joining of V, D, and J segments to create a diverse array of antibodies. In the periphery, the activation-induced cytidine deaminase AID initiates class switch recombination (CSR) of *Igh* constant domains, leading to different antibody isotypes. These tightly regulated processes lie at the heart of the adaptive immune response.

Non-targeted rearrangements are important for cellular homeostasis and genomic integrity. Spontaneous DNA double-strand breaks (DSBs) are usually repaired in *cis* by non-homologous end joining (NHEJ) or in *trans* by homologous recombination (HR). Although largely beneficial, these processes can generate cancer-causing translocations that juxtapose oncogenes (e.g., *myc*) to potent *Ig* enhancers.

We review three principles of nuclear architecture that influence patterns of recombination in B cells: polymer folding, looping between convergent CTCF motifs, and A-B compartmentalization. We discuss that while *Ig* gene recombination evolved to exploit nuclear architecture, DNA damage and selection determine the location and frequency of recurrent cancer-causing rearrangements.

## The Folding of Chromatin Polymers and Its Role in CSR

A simple estimate for the rearrangement frequency between two loci, A and B, is the probability that they are in close spatial proximity within the cell nucleus. In the absence of local folding features (such as loops), this probability ought to decline monotonically as A and B are positioned further apart in the genome. This decline is evident using Hi-C experiments (Lieberman-Aiden et al., 2009). The decline passes through multiple scaling regimes. For inter-locus distances between 500 kb and 7 Mb, Hi-C experiments showed that the frequency of contact is related to the distance by a power law with an exponent of approximately −1.0.

Polymer theory provides a rationale for these observations. In such studies, theoretical and physical simulations of condensed chromatin are used to deduce the relationship between 1D proximity-in-sequence and 3D proximity-in-space. Presently, the most commonly employed model is the fractal globule, which predicts the frequency of contact (or recombination) between two loci scales roughly as the reciprocal of the distance between them (Lieberman-Aiden et al., 2009). The resulting predictions closely match the empirical values obtained by Hi-C.

Many studies have confirmed polymer model predictions. The simplest, that most DSBs ought to be resolved in *cis*, was confirmed by studies in which breaks were induced by RAGs or by ionizing radiation. More quantitative predictions have been made by using next-generation sequencing of chromosomal rearrangements. These rely on the insertion of restriction sites for I-SceI, a yeast endonuclease, at loci of interest such as *Igh* and *myc*. By transducing activated B cells with I-SceI, DSBs form at the targeted loci, and the resulting rearrangements can be profiled in a high-throughput fashion (Chiarle et al., 2011; Klein et al., 2011). Monotonic declines in I-SceI recombination frequency were characterized by power law scalings (−1.3) resembling those estimated from Hi-C and polymer modeling. Thus, mammalian genomes rearrange in *cis* with a profile that mimics the polymer behavior of chromatin.

CSR benefits from this propensity. During CSR, activated B cells replace the IgM constant domain ($C\mu$) with that of a downstream isotype ($C\gamma$, $C\alpha$, or $C\varepsilon$). In the mouse, $C_H$s are confined to a relatively small region (160 Kb) within the vast *Igh* locus (2.8 Mb, Figure 1A). Recombination is facilitated by transcription of switch (S) regions upstream of each $C_H$ domain, which imparts accessibility to AID and leads to DSBs. In experiments in which switch regions were replaced by I-SceI sites (Gostissa et al., 2014; Zarrin et al., 2007), the induction of I-SceI breaks promoted CSR. Proximal I-SceI breaks recombined at higher frequencies than distant ones, consistent with the idea that, at least in part, the monotonic decline of *Igh* interactions influences these rearrangements. Thus, CSR appears to exploit the polymer behavior of chromosomes, bringing about recombination events between switch regions through the repair of DSBs that come into spatial proximity.

## Long-Range CTCF Looping Facilitates V(D)J Recombination and Ensures Antibody Diversity

V(D)J recombination occurs over a broader range of distances than CSR. The first recombination event joins D and $J_H$ segments separated in the mouse genome by a maximum distance of 100 Kb (Figure 1A). In contrast, $V_H$-$DJ_H$ recombination deletes at least 45–150 Kb for the most proximal V segment
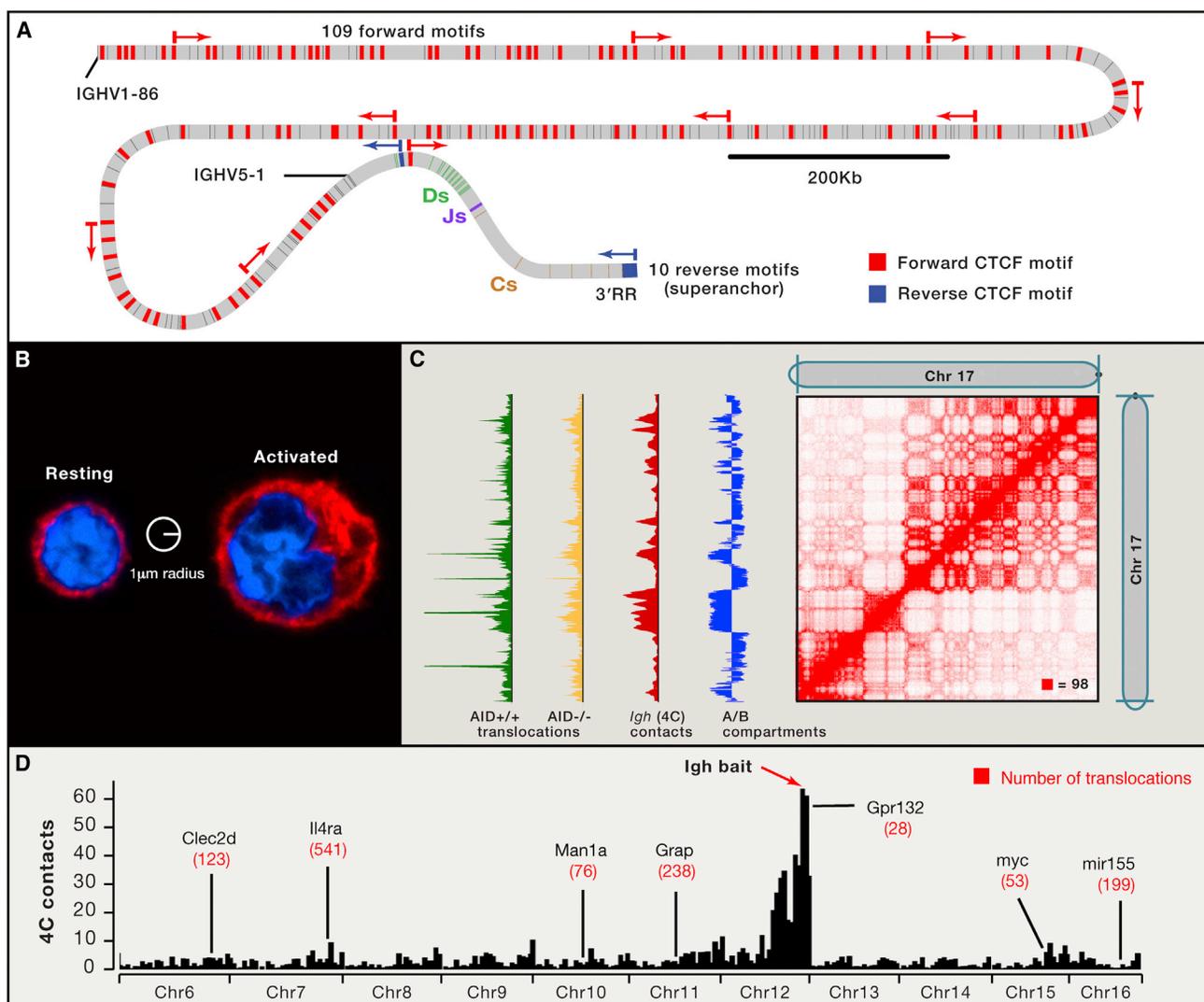
**Figure 1. Patterns of Rearrangement Often Reflect Principles of Nuclear Architecture**

(A) Distribution of CTCF binding motifs (red and blue stripes) at the *Igh* locus. Arrows and colors denote the orientation of the CTCF motifs (forward, red; reverse, blue).

(B) Confocal micrograph comparing a resting to a 72 hr activated B cell. Samples were stained with anti-$\alpha$-tubulin (red) and DAPI (blue) (Kouzine et al., 2013).

(C) Correlation between A-B compartmentalization (defined by Hi-C eigenvector, blue), interactions with *Igh* as measured by 4C (red), chromosomal translocations involving *Igh* and chromosome 17 in the absence of AID (yellow) and in its presence (green). In *AID*$^{+/+}$ cells, recurrent hot spots of translocation are seen. A Hi-C map is also shown at 250 Kb resolution. Hakim et al. (2012) and Rao et al. (2014).

(D) *Igh* 4C-seq profile. 7 (of 236) AID target genes, are highlighted. The total number of *Igh* translocations at each gene, as defined by TC-seq (Klein et al., 2011), is shown.

(IghV5-1) and up to ~2.6 Mb for the most distal one (IghV1-86, Figure 1A). Thus, if $V_H$-$DJ_H$ recombination relied exclusively on the monotonic behavior of chromatin polymers, proximal $V_H$ segments would dominate the mature Ig repertoire, drastically curtailing antibody diversity. However, microscopy studies have shown that the *Igh* locus undergoes conformational changes during V(D)J recombination (Jhunjhunwala et al., 2008). The entire $V_H$ domain associates with D-$J_H$ segments in cells undergoing $V_H$-$DJ_H$ recombination (Subrahmanyam and Sen, 2012). Clearly, mechanisms beyond local polymer folding are at work.

A recent Hi-C map of human B lymphoblastoid cells shed light on these mechanisms revealing ~10,000 loops between pairs of CTCF sites whose motifs are convergent (i.e., facing one another [Rao et al., 2014]). The *Igh* locus illustrates this. Its $V_H$ region contains >100 CTCF sites, all pointing downstream. They face a single motif, pointing upstream, that is situated at the 5′ end of the D region (Figure 1A). In this configuration CTCF looping likely facilitates recombination between distant $V_H$s and rearranged $DJ_H$s. In support of this, deletion of CTCF-binding sites at the D region decreased recombination with distal $V_H$s (Guo et al., 2011; Lin et al., 2015). Thus, V(D)J recombination

appears to rely on CTCF-mediated looping to bring about long-range rearrangements. Besides CTCF, other factors have also been implicated in *Igh* locus contraction, including E2A, Pax5, cohesin, Brg1, and YY-1 (reviewed by Alt et al., 2013). How these factors complement CTCF-mediated looping remains unclear.

## A-B Compartmentalization Shapes Incidental Rearrangements

The DSBs that form during *Ig* gene recombination are not always properly repaired in *cis* and can instead lead to incidental rearrangements, both in *cis* and especially in *trans*. In B cell tumors chromosomal translocations frequently join oncogenes to potent *Ig* enhancers that deregulate their expression. The origin of these events has been debated for several decades; proximity, patterns of DSB formation, and selection have all been proposed as the driving mechanism.

At first, the relationship between nuclear architecture and translocations was explored using microscopy. These studies tested the hypothesis that oncogenes and *Ig* loci translocate frequently because they preferentially associate in B cell nuclei. Fluorescence in situ hybridization (FISH) showed that *Igh* and *myc* are within 1 μm of each other in a fraction of activated B cells, but the overlap frequency varied greatly across studies. When B cells were analyzed shortly after activation (<15 min), *Igh* and *myc* appeared to overlap in >20% of lymphocytes (Osborne et al., 2007), while at later stages of activation (72 hr) or in germinal center B cells, values were in the 3%–5% range (Hakim et al., 2012; Gramlich et al., 2012; Wang et al., 2009). These experiments were interpreted to suggest that, at least under some conditions, *Igh* and *myc* preferentially associate in B cells.

However, the studies did not consider the frequency at which gene loci overlap at random. For 3D-FISH in a diploid nucleus using a 1 μm overlap threshold, the random overlap frequency between two genes is given by the volume of two spheres of radius 1 μm (∼8.4 μm$^3$) divided by the total volume of the nucleus. For B cells at rest or shortly after activation (15 min), a volume of 55 μm$^3$ is typical (Figure 1B), implying that two genes overlap at random in 15% of the cells. In contrast, the average volume of activated B cells >24 hr post-activation is ∼250 μm$^3$ (Figure 1B), resulting in a random overlap frequency of roughly 3.5%. (For 2D FISH, the probabilities are 14% before activation and 5% afterward.) These considerations suggest that the reported differences in *Igh-myc* overlap may be explained by differences in nuclear volume rather than by preferential association.

Recent contact mapping experiments have provided more definitive results. In particular, kilobase resolution Hi-C mapping did not reveal peaks of contact frequency between pairs of loci lying in *trans* (Rao et al., 2014). Furthermore, 4C studies found that *Igh* and *myc* do not form preferential locus-specific associations beyond those between any pair of transcriptionally active genes (Hakim et al., 2012). Crucially, the observation that there are no biases in *trans* between individual gene loci does not imply that there are no biases in *trans* whatsoever. The existence of chromosome territories and their preferential associations is well known, as is the fact that A-B compartmentalization leads to spatial segregation of open and closed chromatin (Lieberman-Aiden et al., 2009; de Laat and Grosveld, 2007). Consequently,

*Igh* and *myc*, on chromosomes 12 and 15 respectively, colocalize with transcriptionally active loci genome wide at similar frequencies (Figure 1C; Hakim et al., 2012).

These broad patterns of spatial proximity influence rearrangement patterns. In *AID*$^{-/-}$ B cells or in irradiated pro-B cells—where the distribution of DNA breaks is relatively uniform across the genome—translocation patterns reflect both chromosome territories and A-B compartmentalization (Hakim et al., 2012). Thus, in the absence of recurrent DNA damage, *incidental* rearrangements—those that do not arise due to targeted mechanisms such as VDJ recombination—mostly follow the broad contours of nuclear compartmentalization.

## Patterns of DNA Damage and Selective Pressure Shape Patterns of Recurrent Rearrangements

The findings discussed so far highlight the significant influence of nuclear architecture on patterns of rearrangement. Yet, another key feature that impacts recombination is the frequency of DNA DSBs. In the presence of DNA-damaging enzymes such as RAGs and AID, the rate of formation of such lesions is not constant across the genome but is instead targeted to specific sites. In cancer, yet another factor must be considered: tumorigenesis may select for or against particular rearrangements.

### Hot Spots of AID-Mediated DNA Damage Become Hot Spots of Rearrangement

A primary role of AID-mediated deamination is to promote the formation of DSBs during CSR. However, AID is also a significant driver of rearrangement hot spots. This tendency has been clearly observed in activated and germinal center B cells. In addition to the rearrangements seen in the *AID*$^{-/-}$ background, activated B cells with intact AID exhibit translocation hot spots at *myc* and other oncogenes implicated in B cell transformation. By monitoring the accumulation of the repair factors RPA and Rad51 at resected DNA breaks, these hot spots were confirmed to be sites of recurrent AID-induced DNA damage. Indeed, there is a direct proportionality between the extent of AID-mediated DSB formation at hot spots and the absolute number of *Igh* translocations. These translocation frequencies dramatically exceed predictions based on nuclear architecture alone. For instance, *Igh* is 20-fold more likely to spatially co-locate with *Gpr132*, which lies 400 kb away on chromosome 12, than with *Il4ra*, which lies, in *trans*, on chromosome 7 (Figure 1D). But *Igh* is 20-fold more likely to rearrange with *Il4ra* than with *Gpr132*. Similarly, *myc* is not particularly likely to spatially co-locate with *Igh*: >2,000 genes are more likely to do so. Yet *myc* is one of a relatively small fraction of genes that recurrently rearrange to *Igh* (Hakim et al., 2012).

Thus, patterns of AID-mediated DSBs are a principal factor in determining sites of recurrent translocation in B cells, irrespective of topology. This also applies to translocations induced by other forms of recurrent DNA damage, such as Rag1/2 activity or the CRISPR-Cas9 system (Frock et al., 2015). These arguments highlight the need to understand the factors that make loci vulnerable to each damage modality. In the case of AID, several proposals have been made, including the presence of super-enhancers, convergent gene transcription, and high interconnectivity between regulatory elements (Alinikula and Schatz, 2014). Different factors are likely to be relevant for each damage

modality. For instance, the mere presence of H3K4me3 appears to be sufficient for RAG recruitment (Teng et al., 2015)

### Architecture Constrains Selection of Cancer-Causing Rearrangements

In tumors, the rate at which a particular rearrangement is found does not reflect its rate of formation in the tumor's cell of origin. Instead, there is strong positive selection for incidental translocations that deregulate oncogene expression. For instance, *Il4ra* translocations form 10-fold more frequently than *myc* translocations in activated B cells (Figure 1D), but—unlike *myc*—*Il4ra* translocations are yet to be reported in B cell tumors.

In *Ig* translocations, oncogene deregulation tends to be the result of the potent *Igκ*, *Igλ*, or *Igh* enhancers. Because these enhancers can work at a distance, they are believed to rely on spatial proximity to drive their biological function. Yet this same architecture imposes constraints on their activity. For instance, when mouse B cells lacking AID-mediated damage were cultured under non-selective conditions, incidental translocations placed the *Igh* 3′ enhancer, known as 3′RR, at a wide range of distances from *myc* (Kovalchuk et al., 2012). Following selection (during mouse plasmacytomagenesis), the same element was found no more than 500 kb from *myc*. Beyond this point, no local changes were observed in the level of Pol II recruitment, gene expression, or epigenetic modifications. Given the above-noted findings about CTCF orientation, it is likely that the transformational potency of the 3′RR enhancer is related to the fact that it lies adjacent to what we dub a CTCF "superanchor": a series of ten CTCF sites, all of which point in the same direction. We predict that the selection of an *Igh* translocation depends on the capacity of this superanchor to form long-range functional loops with translocating CTCF motifs. At least in the setting of *myc* translocations, the upper limit of these loops appears to be half a megabase.

Thus, nuclear architecture influences the location and frequency of rearrangements as well as which of these events are selectively favored during transformation. This observation may have implications for tumors that are not derived from B and T cells and that typically lack targeted, recurrent DNA damage.

In the past 5 years, technological advances have helped to unravel many longstanding issues relating to both physiological and pathological rearrangement in B lymphocytes. As the field of 3D genomics advances, we anticipate that it will continue to illuminate these mechanisms and their implications for tumorigenesis, both within the immune system and beyond.

### REFERENCES

Alinikula, J., and Schatz, D.G. (2014). Cell *159*, 1490–1492.

Alt, F.W., Zhang, Y., Meng, F.L., Guo, C., and Schwer, B. (2013). Cell *152*, 417–429.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Cell *147*, 107–119.

de Laat, W., and Grosveld, F. (2007). Curr. Opin. Genet. Dev. *17*, 456–464.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Nat. Biotechnol. *33*, 179–186.

Gostissa, M., Schwer, B., Chang, A., Dong, J., Meyers, R.M., Marecki, G.T., Choi, V.W., Chiarle, R., Zarrin, A.A., and Alt, F.W. (2014). Proc. Natl. Acad. Sci. USA *111*, 2644–2649.

Gramlich, H.S., Reisbig, T., and Schatz, D.G. (2012). PLoS ONE *7*, e39601.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., et al. (2011). Nature *477*, 424–430.

Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.-R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C., et al. (2012). Nature *484*, 69–74.

Jhunjhunwala, S., van Zelm, M.C., Peak, M.M., Cutchin, S., Riblet, R., van Dongen, J.J., Grosveld, F.G., Knoch, T.A., and Murre, C. (2008). Cell *133*, 265–279.

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Cell *147*, 95–106.

Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K.-R., Bandle, R., Steevenson, N., Nakahashi, H., Awasthi, P., Feigenbaum, L., et al. (2013). Cell *153*, 988–999.

Kovalchuk, A.L., Ansarah-Sobrinho, C., Hakim, O., Resch, W., Tolarová, H., Dubois, W., Yamane, A., Takizawa, M., Klein, I., Hager, G.L., et al. (2012). Proc. Natl. Acad. Sci. USA *109*, 10972–10977.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Science *326*, 289–293.

Lin, S.G., Guo, C., Su, A., Zhang, Y., and Alt, F.W. (2015). Proc. Natl. Acad. Sci. USA *112*, 1815–1820.

Osborne, C.S., Chakalova, L., Mitchell, J.A., Horton, A., Wood, A.L., Bolland, D.J., Corcoran, A.E., and Fraser, P. (2007). PLoS Biol. *5*, e192.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). Cell *159*, 1665–1680.

Subrahmanyam, R., and Sen, R. (2012). Curr. Top. Microbiol. Immunol. *356*, 39–63.

Teng, G., Maman, K., Resch, W., Kim, M., Yamane, A., Qian, J., Kieffer-Kwon, K.-R., Mandal, M., Meffre, E., Clark, M., et al. (2015). Cell *162*, this issue, 751–765.

Wang, J.H., Gostissa, M., Yan, C.T., Goff, P., Hickernell, T., Hansen, E., Difilippantonio, S., Wesemann, D.R., Zarrin, A.A., Rajewsky, K., et al. (2009). Nature *460*, 231–236.

Zarrin, A.A., Del Vecchio, C., Tseng, E., Gleason, M., Zarin, P., Tian, M., and Alt, F.W. (2007). Science *315*, 377–381.