ARTICLE

doi:10.1038/nature10909

DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes

Ofir Hakim[1]*, Wolfgang Resch[2]*, Arito Yamane[2]*, Isaac Klein[3], Kyong-Rim Kieffer-Kwon[2], Mila Jankovic[3], Thiago Oliveira[3,4], Anne Bothmer[3], Ty C. Voss[1], Camilo Ansarah-Sobrinho[2], Ewy Mathe[5], Genqing Liang[2], Jesse Cobell[2], Hirotaka Nakahashi[2], Davide F. Robbiani[3], Andre Nussenzweig[6], Gordon L. Hager[1], Michel C. Nussenzweig[3,7]* & Rafael Casellas[2,8]*

Recurrent chromosomal translocations underlie both haematopoietic and solid tumours. Their origin has been ascribed to selection of random rearrangements, targeted DNA damage, or frequent nuclear interactions between translocation partners; however, the relative contribution of each of these elements has not been measured directly or on a large scale. Here we examine the role of nuclear architecture and frequency of DNA damage in the genesis of chromosomal translocations by measuring these parameters simultaneously in cultured mouse B lymphocytes. In the absence of recurrent DNA damage, translocations between *Igh* or *Myc* and all other genes are directly related to their contact frequency. Conversely, translocations associated with recurrent site-directed DNA damage are proportional to the rate of DNA break formation, as measured by replication protein A accumulation at the site of damage. Thus, non-targeted rearrangements reflect nuclear organization whereas DNA break formation governs the location and frequency of recurrent translocations, including those driving B-cell malignancies.

Most cancers bear cytogenetic abnormalities including chromosomal translocations and rearrangements[1]. Although translocations and rearrangements are central to the development of cancer, their origins are poorly understood. One possibility is that they arise from rare and random events that are selected in tumour precursors because they provide a growth advantage. However, increasing evidence indicates that mechanistic factors other than simple selection may have a role in their genesis. In B lymphocytes, V(D)J recombination, class switch recombination (CSR) and somatic hypermutation (SHM) produce obligate single- and double-strand DNA break intermediates that can become substrates for translocations[2,3]. Consistent with this idea, genetic ablation of the enzymes that create DNA lesions during V(D)J recombination (RAGs) or CSR and SHM (AID; also called AICDA) has a profound protective effect on B-cell transformation[2,4].

A second mechanism that may also influence the incidence of chromosomal translocations is nuclear architecture. Two decades of imaging and recent molecular approaches have established that the spatial organization of the genome is not random, but compartmentalized into chromosome territories as well as transcriptionally active and silent subnuclear environments[5–8]. These compartments are believed to influence the frequency with which genes from different chromosomes can interact and recombine. Furthermore, there is a strong association between transcriptional activity and translocation[9].

Using new methods that capture rearrangements genome-wide, thousands of translocations were recently isolated in primary B cells in the absence of growth selection[9,10]. The studies confirmed the notion that the formation of chromosomal translocations is influenced by spatial conformation, targeted DNA damage and open chromatin. Consistent with the distribution of mammalian chromosomes in discrete nuclear territories, most rearrangements occurred intra-chromosomally[9,10]. Moreover, rearrangements in *trans* were biased towards transcriptionally active genes, and particularly those targeted by AID[9,10]. What the studies did not resolve, however, was to what extent recurrent DNA damage, chromatin accessibility, or spatial genome organization influence the location and frequency of cancer-inducing translocations. Here we make use of deep-sequencing techniques to establish the relationship between genome-wide spatial interactions, DNA damage and translocations in activated B cells.

A map of *Igh* and *Myc* long-range nuclear associations

To identify genomic regions that are in close spatial proximity to *Igh*, *Myc* and *Mycn* (also called *N-myc*) loci, we performed chromosome conformation capture experiments[11] followed by deep-sequencing (4C-seq). We used *Igh* and *Myc* as baits because they are actively transcribed and targeted by AID[12]. As controls, we analysed *Mycn*, which is transcriptionally silent in peripheral B cells and does not recruit AID[13], and *Igh* in mouse embryonic fibroblasts (MEFs), where immunoglobulin genes are not expressed. Because of the large size of *Igh*, we used two 4C-seq baits specific for 5′Eμ and 3′Eα enhancers (Supplementary Fig. 1a). Two independent 4C libraries (HindIII and BglII) were constructed for each condition (see Methods). In all experiments, most of the 4C sequence reads (76% on average) originated from the *cis* chromosome (Fig. 1a, Supplementary Table 1 and Supplementary Fig. 2a), an observation consistent with the finding that loci on the same chromosome preferentially interact in *cis* within a chromosome territory[5,6,9].

[1]Laboratory of Receptor Biology and Gene Expression, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA. [2]Genomics & Immunity, NIAMS, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA. [3]Laboratory of Molecular Immunology, The Rockefeller University, New York, New York 10065, USA. [4]Medical School of Ribeirao Preto/USP, Departmento 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]Biodata Mining and Discovery, NIAMS, National Institutes of Health, Bethesda, Maryland 20892, USA. [6]Laboratory of Genome Integrity, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA. [7]Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10065, USA. [8]Center of Cancer Research, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA.
*These authors contributed equally to this work.

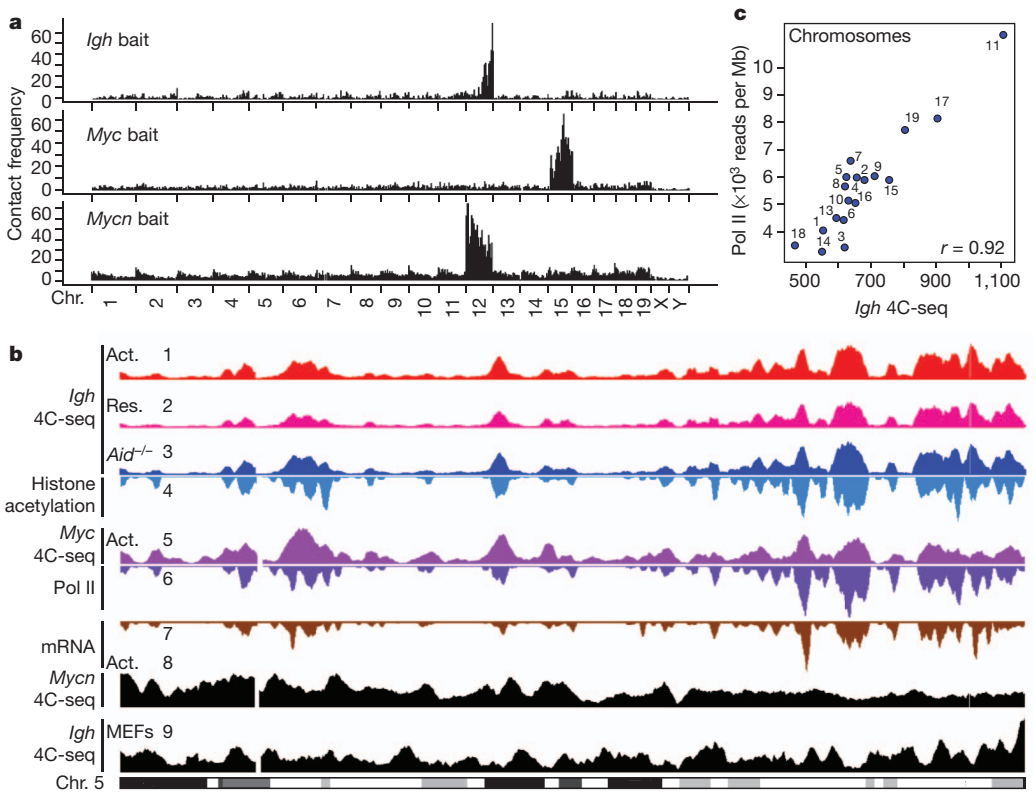

Figure 1 | Characterization of the *Igh*, *Myc* and *Mycn* interactomes in B lymphocytes. **a**, Genome-wide interaction profile of *Igh* 3′Eα, *Myc* and *Mycn* in activated B cells. Plots show the percentage of HindIII fragments carrying 4C-seq reads. **b**, Contact frequency of *Igh* with chromosome 5 in activated B cells (lane 1), anti-HEL homozygous activated B cells (lane 2), or resting B cells (lane 3). Histone acetylation (lane 4), RNA Pol II (lane 6) and mRNA (lane 7) density is also shown. Lanes 5 and 8 show *Myc* and *Mycn* contacts, respectively. Lane 9 represents *Igh* contacts in MEFs. **c**, Comparison of Pol II and *Igh* 4C-seq data per chromosome normalized as reads per mappable megabase.

To explore contact frequencies in *trans*, the mouse genome was partitioned into 200-kilobase (kb) non-overlapping windows and the number of 4C-seq-positive fragments was calculated for each window (see Methods). In activated B cells the contact profile of *Igh* was nonrandom, following a peaks-and-valley pattern similar to that reported for transcriptionally active loci in other cell types[14] (Fig. 1b, lane 1). This pattern was comparable for Eμ and Eα baits (Spearman's $\rho = 0.70$, Supplementary Fig. 1b), and was further reproduced in resting wild-type and activated AID-deficient B cells ($Aid^{-/-}$) (Spearman's $\rho = 0.93$ (resting) and 0.94 ($Aid^{-/-}$); Fig. 1b, lanes 1–3, and Supplementary Fig. 2b). Nearly identical profiles were observed in B cells homozygous for an anti-hen egg lysozyme VDJ knock-in (anti-HEL, Spearman's $\rho = 0.89$, Supplementary Table 2), where most of the *Igh* variable domain is in germline configuration. Thus, globally, *Igh* nuclear interactions in peripheral B cells are largely independent of cell activation, AID expression, or *Ig* variable and constant region gene recombination.

4C-seq was validated by three-dimensional DNA fluorescence *in situ* hybridization (3D DNA FISH) using Perkin Elmer's ultra-high-throughput imaging system. This new approach allowed the automated and unbiased screening of 48,162 activated B cells. Analysis of *Igh* interactions with 14 genomic sites showed 3D FISH measurements to be in good agreement with 4C-seq ($R^2 = 0.99$, Supplementary Fig. 3).

Genome features enriched in contacting loci

Even though the *Igh* and *Myc* loci are on different chromosomes their interactome was significantly correlated (Spearman's $\rho = 0.58$ ($P < 1 \times 10^{-8}$); Fig. 1b, lanes 1–3 and 5). This finding is consistent with the notion that these genes frequently associate and thus may share a common subnuclear environment in B cells[15–17]. To characterize the genomic properties of loci interacting with *Igh* and *Myc*, we compared their 4C-seq profiles to genome-wide epigenetic and transcription maps[13,18]. The analyses revealed a good concordance between *Igh*- and *Myc*-interacting loci and activating chromatin acetylation, RNA polymerase II (Pol II) and messenger RNA transcripts (Fig. 1b, lanes 1–7, and Supplementary Fig. 2c, d). This correlation was particularly evident when entire chromosomes were considered. For example, *Igh* contact probability and Pol II were highest for chromosomes 11, 17 and 19, and lowest for chromosomes 3, 14 and 18 (Pearson's $r = 0.92$ ($P = 1.7 \times 10^{-8}$); Fig. 1c). This hierarchical ordering closely followed gene density estimates, which are highest (2.1%) for mouse chromosome 11, and lowest (1%) for chromosomes 3 and 14 (Supplementary Table 3). Similar correlations were obtained for *Myc*, although the correlations were lower than for *Igh* (Pearson's $r = 0.61$ ($P = 0.0013$); Supplementary Fig. 2d, e and Supplementary Table 4). Altogether, the data recapitulate the spatial compartmentalization of transcriptionally active, gene-dense domains[6,11,14]. In marked contrast, the interactome of transcriptionally silent *Mycn* in B cells or *Igh* in MEFs seemed to be random and did not correlate well with any of the genomic features surveyed, with the exception of centromeric regions for *Mycn* (Fig. 1b, lanes 8 and 9, and Supplementary Fig. 4). This latter feature might reflect the tendency of some silent loci to co-localize with peri-centromeric, repressive heterochromatin[19]. We conclude that in peripheral B cells *Igh* and *Myc* are more closely associated with transcribed, epigenetically accessible genomic sites, whereas interactions of transcriptionally inactive *Mycn* (or *Igh* in MEFs) are more randomly distributed.

Translocations in the absence of AID

To examine the role of nuclear contacts on the genesis of chromosomal translocations in the absence of programmed DNA damage, we

compared the 4C-seq genomic profiles to translocation-capture sequencing (TC-seq) data sets obtained from *Aid*$^{-/-}$ activated B cells[9]. The TC-seq assay was recently developed to map genomic rearrangements comprehensively in B cells where a specific DNA break at *Igμ* (also called *Ighm*) or *Myc* is created via expression of the I-SceI mega-nuclease[9,12]. In the absence of AID, *Igh*$^{I\text{-}SceI}$ or *Myc*$^{I\text{-}SceI}$ translocate to loci that occasionally suffer DNA damage as a result of normal metabolic processes such as transcription or DNA replication.

A total of 68,403 and 28,548 rearrangements were captured between *Igh*$^{I\text{-}SceI}$ and *Myc*$^{I\text{-}SceI}$, respectively, and the rest of the genome (Fig. 2a; see also ref. 9). Visual comparison of the aligned 4C- and TC-seq reads revealed a nonrandom distribution of AID-independent rearrangements across chromosomes (Fig. 2b). Notably, the translocation profiles resembled the *Igh* and *Myc* interactome as well as accessible chromatin as measured by histone acetylation (Fig. 2b and Supplementary Fig. 5). Conversely, there was no obvious concordance between *Mycn* nuclear contacts in the same cells and *Igh*$^{I\text{-}SceI}$ or *Myc*$^{I\text{-}SceI}$ translocations (Supplementary Fig. 5). To validate these observations genome-wide, *Igh*, *Myc* and *Mycn* nuclear contacts were subdivided into quartiles (Q) and the data plotted as a function of total *Igh* or *Myc* translocations per 200-kb non-overlapping windows. The results showed that the greater the interaction frequency between *Igh* (or *Myc*) and a given genomic site, the more likely that the two loci were translocated (Q1 versus Q4, $P = 0.0005$ (permutation test); Fig. 2c and Supplementary Fig. 6). In the case of *Igh*, where the number of captured rearrangements was substantial, translocations per chromosome were directly proportional to the contact frequency between *Igh* and a given chromosome (Pearson's $r = 0.77$ ($P = 0.0002$); Fig. 2d). Conversely, we observed little or no correspondence between *Igh* or *Myc* translocations and the interactome of *Mycn* in B cells (Q1 versus Q4, $P = 0.35$; Supplementary Fig. 6). The data are thus consistent with the notion that AID-independent translocations occur preferentially between interacting genomic loci that are epigenetically accessible.

AID-targeted translocations

AID produces lesions in a large number of defined hotspots, many of which are recurrent translocation partners for *Igh* in lymphoma[2,9,10] (Fig. 3a). To determine whether the location of these hotspots could be explained by B-cell nuclear architecture, we ranked RefSeq genes on the basis of 4C values. The analysis showed that a large fraction of loci carrying *Igh* translocation hotspots engaged in recurrent long-range interactions with this locus in activated B cells (Fig. 3b, c). At the same time, we identified thousands of genes that interacted repeatedly with *Igh* but that were not associated with translocation hotspots (Fig. 3c). For instance, up to 2,361 genes (11% of all RefSeq genes) outranked *Myc* in *Igh* contact frequency (Supplementary Table 5 and Supplementary Fig. 7), even though only 58 of them were recurrently rearranged to *Igh* (Fig. 3c). In a similar manner, whereas translocation hotspots in *Myc*$^{I\text{-}SceI}$ B cells were biased for domains co-localizing with *Myc*, physical proximity per se could not predict the presence of translocation hotspots in these cells (Supplementary Fig. 8). Furthermore, a subset of hotspot genes (for example, *Socs1* or *Dusp4*) associated infrequently with *Igh* (Fig. 3b, c), and we found no direct correlation between the number of translocations per hotspot and contact frequency with *Igh* (Spearman's $\rho = 0.1$ ($P = 0.4$); Fig. 3d). As an example, *Tmed8* and *Dusp4* genes were rearranged to *Igh* at roughly equal proportions (38 versus 36 translocations respectively), in spite of the fact that *Tmed8* was physically associated with *Igh* ~10 times more frequently than *Dusp4* in the B cell nucleus (Fig. 3c).

To exclude formally the possibility that *Igh* and *Myc* share translocation targets primarily because of shared contacts, we generated two additional 4C-seq libraries using baits specific for *Rac2* (chromosome 15) and *Rplp2* (chromosome 7). These genes are highly transcribed in B cells[13], and both interact more frequently with *Igh* than *Myc* does, but neither is associated with AID-mediated translocation hotspots (Fig. 3c; see also ref. 9). The interactome of these two genes was then compared to that of *Igh* in activated B cells. As controls for this analysis, we included the interactomes of *Igh* from resting B cells, anti-HEL knock-in B cells and MEFs, as well as the *Mycn* interactome from stimulated B cells. As expected, the *Igh* interactome was similar in all B-cell types, but not in MEFs where it is not transcribed (Fig. 3e). Similarly, *Mycn*, which is transcriptionally silent in activated B cells, shows 4C-seq profiles with little correlation to *Igh* (Fig. 3e). Notably, the interactomes of *Rac2* and *Rplp2* were significantly more correlated to *Igh* than was *Myc* ($P < 1 \times 10^{-5}$ (bootstrapping test);

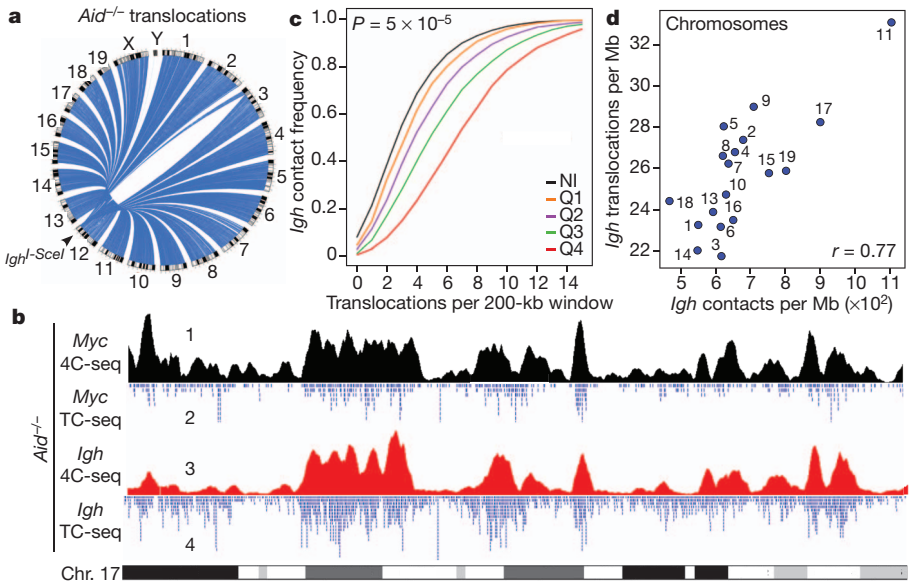

Figure 2 | Genomic distribution of AID-independent translocations correlates with nuclear contact profiles. **a**, Genome-wide view of rearrangements to *Igh*$^{I\text{-}SceI}$ in *Aid*$^{-/-}$ B cells. **b**, Cross-comparison of contacts and translocations between *Myc* or *Igh* and mouse chromosome 17 in activated *Aid*$^{-/-}$ B cells. **c**, Empirical cumulative distribution showing *Igh*$^{I\text{-}SceI}$ *Aid*$^{-/-}$ translocations per 200-kb non-overlapping windows as a function of *Igh* 4C-seq data subdivided as quartiles (Q1–4). NI represents windows with no aligned reads. **d**, Comparison of AID-independent translocations versus *Igh* 4C-seq per chromosome per mappable megabase. The degree of correlation is represented by Pearson's *r*.

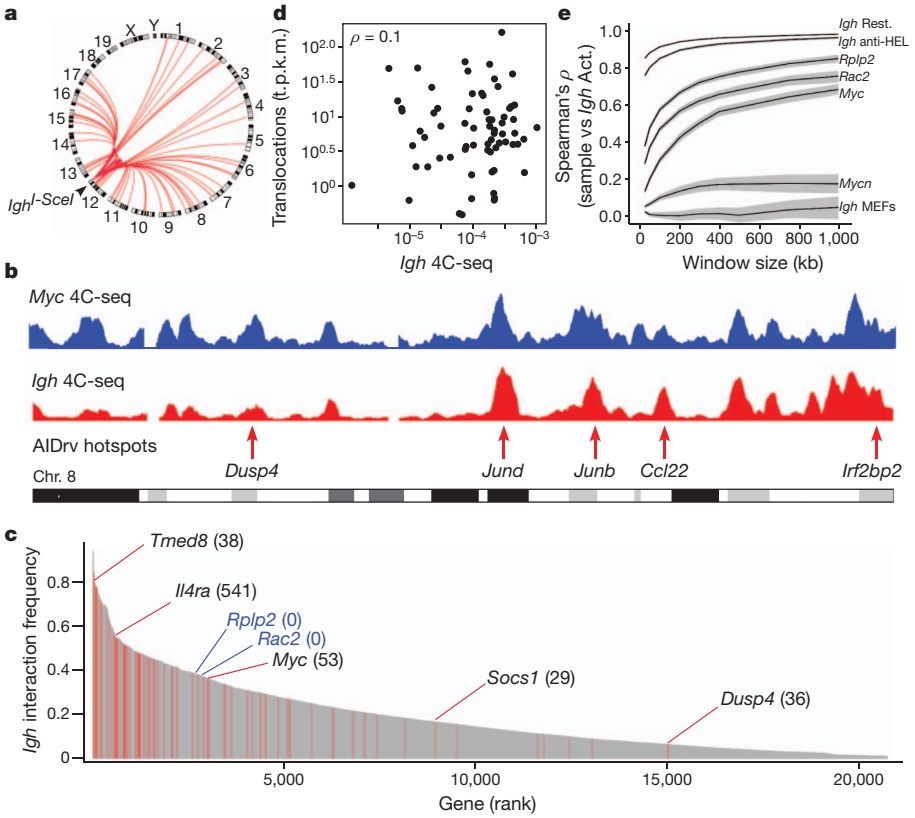

Figure 3 | Lack of correlation between translocation hotspots and nuclear architecture. **a**, Genome-wide hotspots in activated B cells transduced with I-SceI and AID retroviruses (rv). **b**, Translocation hotspots (bottom lane) and *Myc* and *Igh* contacts in chromosome 8. **c**, *Igh* contacts with RefSeq genes. Hotspot^{+} genes are highlighted in red, and for a subset of them the number of translocations is provided in parentheses. The two hotspot^{-} genes (*Rplp2* and *Rac2*) highlighted in blue are discussed in more detailed in panel **e**. **d**, Scatter plot showing *Igh* translocations per hotspot versus contacts. Data are plotted as sequence tags per kb per million sequences (t.p.k.m.). **e**, Line graph showing the 4C-seq correlation (Spearman's ρ) between *Igh* (in various cell types), *Rpbp2*, *Rac2*, *Myc* and *Mycn* (in activated B cells) versus *Igh* in activated B cells. The 99% bootstrapping confidence intervals are shown in grey.

Fig. 3e). Thus, nuclear interactions alone do not predispose transcriptionally active genes to high levels of translocation with *Igh*.

A genome-wide map of AID-mediated dsDNA breaks

Our observations challenge the current view that preferential chromosome and/or gene locus interactions govern tumour-inducing translocations in AID-expressing B cells[15,20], and indicate that the amount of AID-mediated DNA damage could account for the frequency of these events. To explore this idea we created a genome-wide map of AID-mediated DNA damage in activated B cells by measuring recruitment of replication protein A[13,21] (RPA) (Fig. 4a, lane 2). Because RPA accumulation is partially blocked by 53BP1 (also called Trp53bp1; refs 22, 23), we reasoned that genetic deletion of 53BP1 in B cells might increase RPA recruitment, thus providing a more sensitive means to map sites of AID-induced lesions by ChIP-seq. Consistent with this idea, RPA accumulation at *Igh* was markedly increased (7.8-fold relative to control) in the absence of 53BP1 (Fig. 4a, lane 3), and an even higher RPA signal (11-fold) was observed in *53BP1*$^{-/-}$ mice overexpressing AID (*Ig*κ*AID*[12]; Fig. 4a, lane 4). Conversely, there was no detectable accumulation of RPA at *Igh* in activated *Aid*$^{-/-}$*53BP1*$^{-/-}$ B cells (Fig. 4a, lane 1). Thus, RPA recruitment to *Igh* is AID dependent and is enhanced in the absence of 53BP1.

In agreement with our previous findings[13], we did not detect RPA recruitment at AID targets outside the *Igh* locus, such as *Cd83* (Fig. 4b, lane 2). However, we found prominent RPA ChIP signal at the same locus upon *53BP1* deletion (Fig. 4b, lanes 3 and 4). Analogous to *Igμ* and *Igγ1* (also called *Ighg1*), the *Cd83* RPA island extended nearly 50 kb upstream and downstream of the transcription start site (TSS) (Fig. 4a, b, lane 4). In total, 153 non-*Ig* genes accumulated RPA in an AID-dependent fashion (Fig. 4c and Supplementary Table 6) including known *Igh* translocation partners such as *Pax5*, *Pim1* and *Mir155* (Fig. 4d and Supplementary Table 6).

To ascertain the precise nature of RPA islands, we measured somatic hypermutation at a subset of RPA^{+} and RPA^{-} genes[13]. We found a strong positive correlation between the rate of hypermutation and the extent of RPA recruitment (Spearman's $\rho = 0.71$; Fig. 4e and Supplementary Table 7). We conclude that RPA-seq can be used as a surrogate to measure AID-mediated DNA damage across the B-cell genome.

Recurrent translocations are proportional to DNA damage

To evaluate the relative contribution of DNA damage to targeted translocations we compared the results of RPA-seq and TC-seq obtained from AID-expressing cells. We found a substantial overlap between the two data sets: out of a total of 97 genes with translocation hotspots with an average of 60 translocations per gene, 78 showed RPA accumulation (Supplementary Fig. 9 and Supplementary Table 8). A second group of genes (75) was also associated with RPA islands but displayed fewer translocations (mean = 7, Supplementary Fig. 9), and thus fell below our hotspot criteria cutoff. This result indicates that TC-seq underestimates the number of AID-mediated translocations, possibly due to lack of saturation[9]. Only 19 genes associated with translocation hotspots did not recruit RPA above background levels (Supplementary Fig. 9), suggesting that the RPA-seq data set is also not fully saturated. Thus, RPA demarcates sites of recurrent translocations in B lymphocytes.

In addition to the qualitative correlation above, we found that the absolute number of *Igh* translocations per hotspot was directly proportional to RPA recruitment (Spearman's $\rho = 0.6$ ($P = 2.9 \times 10^{-6}$);

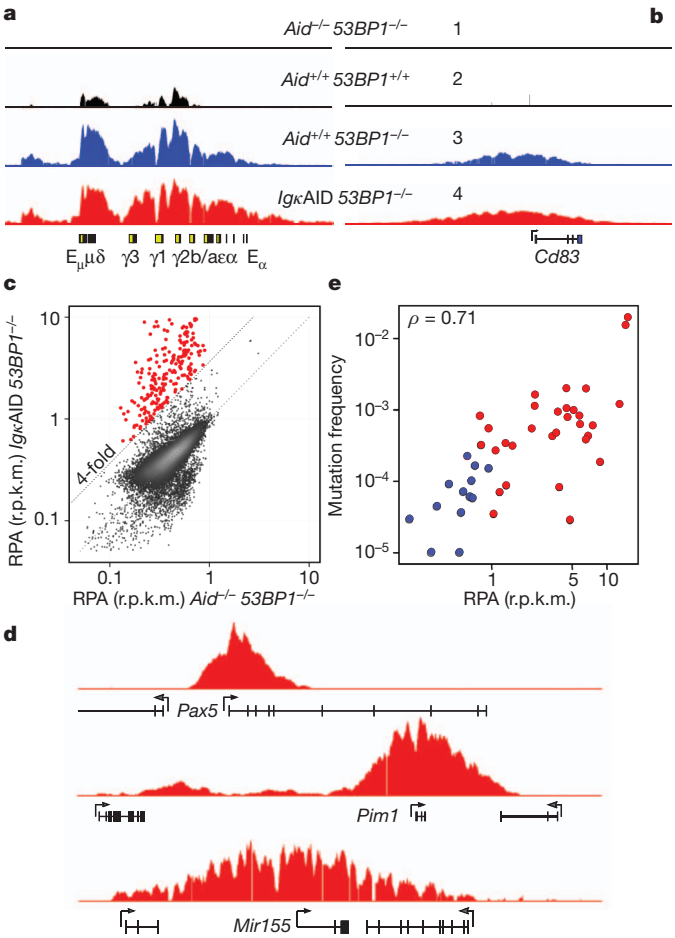

Figure 4 | Genome-wide map of AID-mediated DNA damage. **a**, RPA occupancy at *Igh* in activated B cells. Genotype and RPA sequence reads per million values are shown. **b**, Same analysis as in panel **a** for *Cd83*. For all tracks, background sequencing was filtered out via a threshold. **c**, One-hundred and fifty-three RPA islands (red dots) detected in *Igκ*AID *53BP1*$^{-/-}$ B cells fourfold above background (measured in *Aid*$^{-/-}$*53BP1*$^{-/-}$ cells). Data are plotted as reads per kb per million sequences (r.p.k.m.). **d**, RPA islands associated with TSSs from *Pax5*, *Pim1* and *Mir155*. **e**, Hypermutation frequency relative to RPA recruitment at TSSs (±2 kb) in a subset of RPA^{+} (red dots) and RPA^{-} (blue dots) genes. Spearman's ρ is provided.

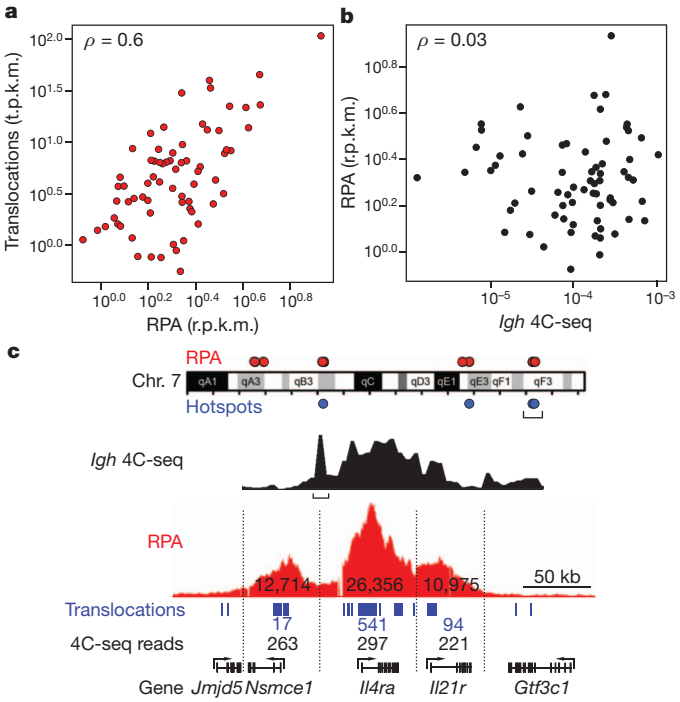

Figure 5 | AID activity predicts the location and frequency of targeted chromosomal translocations. **a**, Scatter plot showing the correlation between *Igh* translocations per hotspot and RPA recruitment. **b**, Same as panel **a** but *Igh* contact frequency is used instead of translocations. **c**, Upper schematic: distribution of RPA islands (red dots) and translocation hotspots (blue dots) in chromosome 7. Middle: *Igh* 4C-seq profile demarcating the *Nsmce1-Il4ra-Il21r* loci. Bottom: RPA islands, translocations and contact frequency for each gene.

Fig. 5a). This result contrasts with the lack of correlation observed between nuclear contacts and total rearrangements per hotspot (Fig. 3d) or RPA accumulation (Spearman's $\rho = 0.03$ ($P = 0.8$); Fig. 5b). Similar results were observed for *Myc* (Spearman's $\rho < 0.07$; data not shown). These findings demonstrate that with regard to AID-targeted translocations, DNA damage is the primary determinant of rearrangement location and frequency. This was particularly evident for AID targets that are clustered within ~200-kb genomic domains, such as *Nsmce1*, *Il4ra* and *Il21r* in chromosome 7 (Fig. 5c), or the *Hist1h1c* gene family in chromosome 13 (Supplementary Fig. 10 and Supplementary Table 9). Whereas we found little variation in *Igh* (or *Myc*) proximity for different genes within these clusters, translocations varied substantially and in a manner that was proportional to AID-mediated damage (Fig. 5c, Supplementary Fig. 10 and Supplementary Table 9). Taken together, these results clearly demonstrate that DNA break formation, but not nuclear interactions, governs the rate of recurrent chromosomal translocations.

Discussion

We have shown that in the absence of programmed DNA damage, translocation partner selection is largely dictated by physical proximity, following principles of nuclear organization, chromatin accessibility and gene expression. Physical proximity has also been suggested to have an impact on the formation of recurrent translocations[20]. For instance, rearrangements between *BCR–ABL1* in chronic myeloid leukaemia, *RET–CDC6* in thyroid malignancies, *TMPRSS2–ERG/ETV1* in prostate cancer, and *PML–RARA* in acute promyelocytic leukaemia have all been ascribed to preferential interactions between translocating partners in tumour cell precursors[1]. Similarly, *Igh* and *Myc* chromosomes have been shown to associate in mouse and human B cells[16,17], and RNA FISH has shown that the *Myc* and *Igh* alleles are frequently found in the same RNA Pol-II-enriched transcription factories[15]. Bystander translocations between *Igh* and *Igλ* have also been proposed to result from frequent contacts, as determined by 3D DNA FISH[24]. One limitation of FISH technology however is that it can only monitor a limited number of loci simultaneously. Consequently, it has been difficult to ascertain whether the documented contacts are truly unique relative to the broad array of genomic interactions. Our 4C measurements now clarify this issue in that they show that 29% of all genes interact with *Igh* at equal or higher frequency than *Myc*, *Igλ1*, or many of the oncogenes frequently rearranged in B-cell tumours. Thus, the rate of interaction between *Igh* and its recurrent translocation partners is not a specific feature of these loci and cannot account for their high rate of translocation.

Translocation requires joining of two double-stranded breaks (DSBs). Therefore, when breaks are limiting, increasing their frequency increases the rate of translocation[24,25]. Similarly, repair deficiencies augment the rate of translocation by increasing the half-life of dsDNA breaks and thereby the availability of substrates for aberrant repair[2]. However, it has not been possible to relate directly the frequencies of DNA damage and translocations because neither the extent nor the location of DSBs in the B-cell genome was known. We have overcome these limitations by measuring RPA deposition at sites of DNA damage in 53BP1 mutant B cells. On the basis of this new strategy we uncovered ~150 non-*Ig* genes that suffer AID-mediated DSBs. These genes

coincide with translocation hotspots, and are sites of ongoing hypermutation. Most importantly, by relating RPA-seq with TC-seq and 4C-seq data sets we found that the frequency of DNA damage directly accounts for the rate of translocation, as shown by the marked concordance between the amount of RPA deposition and the absolute number of rearrangements at any given genomic site. This view is also supported by the relative lack of correlation between proximity and the absolute number of rearrangements per hotspot.

The genomic distribution of sporadic translocations is best explained by nuclear architecture, whereas the location and incidence of recurrent translocations, including those involved in B-cell malignancies, directly reflect site-specific DNA damage.

METHODS SUMMARY

Full details of B-cell culture, hypermutation analysis, chromatin immunoprecipitation, chromosome conformation capture on Chip (4C), translocation capture sequencing analysis, deep sequencing and bioinformatics techniques are provided in Methods. The NIAMS-NIH Animal Care and Use Committee approved all animal protocols and experiments.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 August 2011; accepted 31 January 2012.
Published online 7 February 2012.

1. Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nature Rev. Cancer* **7,** 233–245 (2007).
2. Nussenzweig, A. & Nussenzweig, M. C. Origin of chromosomal translocations in lymphoid cancer. *Cell* **141,** 27–38 (2010).
3. Tsai, A. G. *et al.* Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell* **135,** 1130–1142 (2008).
4. Zhang, Y. *et al.* The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. *Adv. Immunol.* **106,** 93–133 (2010).
5. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2,** a003889 (2010).
6. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
7. Hakim, O., Sung, M. H. & Hager, G. L. 3D shortcuts to gene regulation. *Curr. Opin. Cell Biol.* **22,** 305–313 (2010).
8. Chakalova, L. & Fraser, P. Organization of transcription. *Cold Spring Harb. Perspect. Biol.* **2,** a000729 (2010).
9. Klein, I. A. *et al.* Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147,** 95–106 (2011).
10. Chiarle, R. *et al.* Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147,** 107–119 (2011).
11. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet.* **38,** 1348–1354 (2006).
12. Robbiani, D. F. *et al.* AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol. Cell* **36,** 631–641 (2009).
13. Yamane, A. *et al.* Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nature Immunol.* **12,** 62–69 (2011).
14. Hakim, O. *et al.* Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res.* **21,** 697–706 (2011).
15. Osborne, C. S. *et al.* Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol.* **5,** e192 (2007).
16. Roix, J. J., McQueen, P. G., Munson, P. J., Parada, L. A. & Misteli, T. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nature Genet.* **34,** 287–291 (2003).
17. Parada, L. A., McQueen, P. G. & Misteli, T. Tissue-specific spatial organization of genomes. *Genome Biol.* **5,** R44 (2004).
18. Kuchen, S. *et al.* Regulation of microRNA expression and abundance during lymphopoiesis. *Immunity* **32,** 828–839 (2010).
19. Brown, K. E., Baxter, J., Graf, D., Merkenschlager, M. & Fisher, A. G. Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol. Cell* **3,** 207–217 (1999).
20. Meaburn, K. J. & Misteli, T. Cell biology: chromosome territories. *Nature* **445,** 379–381 (2007).
21. Vuong, B. Q. *et al.* Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nature Immunol.* **10,** 420–426 (2009).
22. Bunting, S. F. *et al.* 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141,** 243–254 (2010).
23. Bothmer, A. *et al.* 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J. Exp. Med.* **207,** 855–865 (2010).
24. Wang, J. H. *et al.* Mechanisms promoting translocations in editing and switching peripheral B cells. *Nature* **460,** 231–236 (2009).
25. Robbiani, D. F. *et al.* AID is required for the chromosomal translocations in *c-myc* that lead to *c-myc/IgH* translocations. *Cell* **135,** 1028–1038 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Casellas and Nussenzweig laboratories for discussions; G. Gutierrez from NIAMS genomics facility for technical assistance. This work was supported in part by a grant from the Starr Foundation to M.C.N., by NIH grant number AI037526 to M.C.N. and the Intramural Research Program of NIAMS and NCI, NIH. M.C.N. is an HHMI investigator. This study made use of the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (http://biowulf.nih.gov), and the resources of NCI's High-Throughput Imaging Facility.

Author Contributions R.C., O.H., G.L.H. and M.C.N. planned studies and interpreted data. Experiments were performed as follows: O.H. and C.A.-S., 4C-seq and FISH; A.Y., RPA-seq; I.K., A.B., D.F.R. and M.J., TC-seq; W.R., E.M. and T.O., bioinformatics; K.-R.K.-K., T.C.V., H.N. and J.C., FISH; G.L. and H.N., hypermutation; A.N., 53BP1 expertise; M.C.N. and R.C. wrote the manuscript.

Author Information All sequence data are available at the NCBI SRA database under accession number SRP010565. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.C. (casellar@mail.nih.gov) or M.C.N. (nussen@rockefeller.edu).

METHODS

B-cell activation and hypermutation analysis. Miltenyi microbead-isolated $CD43^-$ splenic B cells from wild-type, *Igk*AID-*Ung*$^{-/-}$, or *Aicda*$^{-/-}$ mice were cultured at 0.1×10^6 cells per ml with 50 μg ml^{-1} lipopolysaccharide (LPS) (Sigma), 2.5 ng ml^{-1} mouse recombinant IL-4 (Sigma) and 0.5 μg ml^{-1} of aCD180 (RP105) antibody (RP/14, BD Pharmingen). For 4C-seq and TC-seq procedures, cells were collected at 72 h. For hypermutation analysis, cells were diluted 1:4 at 72 h and cultured for another 48 h under the same conditions. Fifty nanograms of genomic DNA was then amplified for 30 cycles with Phusion DNA polymerase and gene-specific primers. When nested PCR was applied, 40 (20 + 20) cycle amplifications were performed in the presence of DMSO. The amplicon was cloned using PCR Zero blunt (Invitrogen) and sequenced.

Chromosome conformation capture on chip (4C) followed by deep-sequencing. The 4C assay was performed as previously described[14] with minor modifications. Ten million mouse B cells were crosslinked in 2% formaldehyde at 37 °C for 10 min. The reaction was quenched by the addition of glycine (final concentration of 0.125 M). Cells were then washed with cold PBS and lysed (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.2% NP-40, 1× complete protease inhibitors (Roche)) at 4 °C for 1 h. Nuclei were incubated at 37 °C for 1 h in 500 μl of restriction buffer (New England Biolabs buffer 2 for HindIII or buffer 3 for BglIII digestion) containing 0.3% SDS. To sequester SDS, Triton X-100 was then added to a final concentration of 1.8%. DNA digestion was performed with 400 U of HindIII or BglII (New England Biolabs) at 37 °C overnight. After heat inactivation (65 °C for 30 min), the reaction was diluted to a final volume of 7 ml with ligation buffer containing 100 U T4 DNA Ligase (Roche) and incubated at 16 °C overnight. Samples were then treated with 500 μg Proteinase K (Ambion) and incubated overnight at 65 °C to reverse formaldehyde crosslinking. DNA was then purified by phenol extraction and ethanol precipitation. For circularization, the ligation junctions were digested with Csp6I (Fermentas) or DpnII (New England Biolabs) at 37 °C overnight. After enzyme inactivation and phenol extraction, the DNA was religated in a 7-ml volume (1,000 U T4 DNA Ligase, Roche). Three micrograms of 4C library DNA was amplified with Expand Long template PCR System (Roche). Thermal cycle conditions were DNA denaturing for 2 min at 94 °C, followed by 30 cycles of 15 s at 94 °C, 1 min at 60 °C, 3 min at 68 °C, and a final step of 7 min at 68 °C. Baits were amplified with inverse PCR primers as follows: *Igh* with HindIII: IgH_R_4C 5′-CCAGACATGTGG GCTGAGAT-3′, Igh_Hind_Read 5′-CTACCCACCTAACTCCAAGC-3′; *Mycn* with HindIII: Mycn_R_4C 5′-CTCCCATTTTGCACTCCTGT-3′, Mycn_Hind_Read 5′-GATTTATCCTTAAACCCTTAAGC-3′; *Igh* with BglII: IgH_Bgl_R_4C 5′-CATGGACATTTGCGTGTGTA-3′, IgH_Bgl_Read 5′-GTG CCCCCAGGAGCAGATCT-3′; *Mycn* with BglII: Mycn_Bgl_R_4C 5′-AG TCTGCGGGAGGTAAGAAG-3′, Mycn_Bgl_Read 5′-CCCTTTTAGACAGCC AGATCT-3′; *Myc* with BglII: Myc_Bgl_R_4C 5′-AAGAATGTGCCCAGTC AACA-3′, Myc_Bgl_Read 5′-AGTGAATTGCCAACCCAGAT-3′; *Rplp2* with HindIII: 5′-GCCATCTCTCCAGTCAAAAAGC-3′, CTTCTCACTTCCATT CCCTGAG-3′; Rac2 with HindIII: 5′-GCCATGGAGACCGGAAGCTT-3′, 5′-GGGACTGTCCACTCCACCT-3′; Eμ with HindIII: 5′-TGTGGCTGCTGC TCTTAAAGC-3′, 5′-TGTGAAGCCGTTTTGACCAGAATGT-3′. 4C amplified DNA was microsequenced with the Illumina platform. For multiplexing purposes, extra nucleotides were added at the 5′ end of read primers: a T for LPS + IL-4 activated B cells, TT for resting B cells, A for IgH$^{\alpha HEL/\alpha HEL}$ B cells, and AA for MEFs.

Translocation capture sequencing. All experimental procedures involving translocation capture sequencing (TC-seq) library preparation and computational analysis is provided in ref. 9.

Bioinformatics. For 4C-seq, standard Illumina pipeline software (version ≥1.8) was used to process raw data and obtain Fastq files of paired short reads. Each read pair was then tested for the presence of a perfect match to the respective bait primer as well as the bait spacer between the end of the primer and the restriction sites used in the corresponding experiment. In some experiments up to three mismatches in the non-index portion of the PCR primer/flank sequence was allowed without any relevant changes in the resulting data. These flanking sequences other than the restriction site were trimmed and the remainder was aligned against the mouse genome (build mm9/NCBI37) with Bowtie[26] with the following command line options: '-X 500 -p 3 -v2 -k2 -m1–phred64-quals –sam', which reported all unique alignments with at most two mismatches. In the case of Eμ, only the HindIII spanning read in each pair was sufficiently long to reach the ligated interaction partner. In that case, single end alignments of the flank were carried out with the command line options '–best–all–strata–chunkmbs 256 -m1 –sam'. Alignments were then matched up with restriction sites and assigned to a HindIII or BglII fragment. Fragments were combined into 200-kb non-overlapping windows to determine (1) the total number of 4C reads per fragment, and (2) the fraction of restriction fragments for which 4C reads were found. The latter part of the analysis was carried out with a combination of custom software written in Bash, Python, R, and BedTools[27]. Processing of fragment- or window-level data was carried out in R using standard methods.

For RPA-seq, short reads obtained from Illumina pipeline were aligned against the mouse genome (build mm9/NCBI37) with Bowtie[26] using options '–threads = 8–phred64-quals–best–all–strata -m1 -n2 -l36 –sam'. Raw tag densities of *Igκ*AID *53BP1*$^{-/-}$ and *Aid*$^{-/-}$*53BP1*$^{-/-}$ samples flanking transcription start sites of RefSeq genes were compared and genes with >4-fold enrichment were selected.

For histone acetylation, short reads obtained for ChIP against a set of histone acetylations (see below) were aligned above. Areas of local enrichment over a random background model (islands) were identified with SICER 1.03[28] and the density of reads overlapping these islands was averaged in the same 200-kb non-overlapping windows used for 4C analysis.

High-throughput 3D DNA FISH. For 3D FISH, cultured B cells were set in a 384-well, poly-D-lysine-coated microplate (Perkin Elmer) by centrifugation at 1,000 r.p.m. for 5 min. After fixation in 4% PFA for 10 min, and permeabilization in 0.5% saponin (Sigma Aldrich)/0.5% Triton X-100/PBS for 20 min, cells were incubated in 0.1 N HCl for 10 min. Two PBS washes were applied between each step. After a 2× SSC wash, cells were kept in 50% formamide/2× SSC buffer for at least 30 min. Bacterial artificial chromosomes (BACs) were used as probes as follows: *Mlh3*, RP24-139J8; *Klhdc1*, RP24-109D18; *Clec2d*, RP24-149B3; *Mir142*, RP24-376D9, *Il4ra*, RP23-60A3; *Cyth1*, RP23-267B12; *Pim1*, RP24-331E7; *Furin*, RP24-377F13; *Pax5*, RP23-258E20; *Myc*, RP24-297E9; *Cxcr5*, RP24-308P6; *Rasa3*, RP24-247P3; *Gata3*, RP24-402N11. For each BAC, single colonies were grown and the presence of BAC DNA was verified by PCR. DNA was isolated and labelled with biotin (Roche; Biotin-Nick translation mix) or digoxigenin (Roche; DIG-Nick translation mix). A probe mix containing 250 ng of digoxigenin- and biotin-labelled probes, 3 μg mouse COT1 DNA (Invitrogen), and 20 μg tRNA (Ambion) was ethanol precipitated, and re-suspended in 15 μl of hybridization buffer (10% dextran sulphate, 50% formamide, 2× SSC, and 1% Tween 20). Probe was then added to each well, denatured together with nuclei at 85 °C for 7 min and left to hybridize at 37 °C overnight in a humidified chamber. To discard non-hybridizing probe, cells were then washed in 50% formamide with 2× SSC at 45 °C, followed by washes with 1× SSC at 60 °C. Each wash was repeated three times for 5 min. Cells were blocked with 3% BSA/0.05% Tween 20/4× SSC for 20 min at room temperature and then incubated for 1 h with Fluorescein Avidin (Vector) and anti-Dig-Rhodamin (Roche) diluted 1:200 in blocking solution. Next, cells were washed three times with 0.05% Tween 20/4× SSC and mounted in DAPI-containing Vectashield (Vector) for imaging.

Microscopy. Cells were imaged in 384-well plates with the Opera (Perkin Elmer) confocal high-throughput imaging system using a ×40 water objective lens with 9 optical steps of 1.0 μm.

Automated image analysis algorithms. To quantify distances between FISH signals, we customized the automated image analysis computer algorithm from ref. 14 to allow analysis of Opera images. This algorithm determines centre-to-centre distances of FISH signals in 3D. Nuclei and DNA FISH loci were automatically identified by a combination of the Acapella image analysis software (Perkin Elmer) together with a series of custom algorithms developed using Matlab technical computing software and the Matlab Image Processing toolbox (The Mathworks). The resulting morphometric information from each nucleus was automatically stored in a Matlab database file, which could be accessed by the custom algorithms. First, nuclei region of interest (nucROI) were segmented using the DAPI signal channel by the Acapella software. Then DNA FISH loci (fishROI) were automatically identified by a separate algorithm in the nucROI positional information. FishROIs were identified by intensity-based thresholding of the FISH fluorescent channel images. A third algorithm assigned the nuclear position of each FISH locus, based on the *x*–*y*–*z* location of the centre of the brightest 3 × 3 FISH fluorescent channel pixel that was located within the fishROI. For each nucleus, three-dimensional (*x*–*y*–*z*) distances were calculated between all possible pairs of FISH loci. The closest distance between the two probes in each nucleus was used to calculate the frequency of cells with distances <1 μm as previously described[14]. Interaction frequencies between each of the genes and *Igh* was calculated by computing the percentage of cells carrying at least one pair of FISH signals separated by distances smaller than 1 μm (Supplementary Fig. 12a). This distance threshold was recently shown to correlate well with 4C contact frequencies[14]. To determine the minimum number of cells necessary to reach statistical significance in our experimental setting, we first examined 17,462 LPS plus IL-4 activated lymphocytes with *Igh*- and *Myc*-specific probes. We found that at least 2,000

cells were necessary to reach a standard error of 0.01 (Supplementary Fig. 12b). For different probes, the optimal number of cells was as follows: 2,200 for *Myc*, 4,300 for *Mlh3*, 2,100 for *Clec2d*, 3,800 for *Klhdc1*, 1,500 for RP24-137A18, 1,300 for *Gata3*, 2,200 for *Il4ra*, 1,900 for *Mir142*, 2,200 for *Cyth1*, 2,100 for *Pim1*, 2,100 for *Furin*, 1,900 for *Pax5*, 1,800 for *Cxcr5* and 1,300 for *Rasa3*.

26. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
27. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
28. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25,** 1952–1958 (2009).