Chromatin conformation analysis of primary patient tissue using a low input Hi-C method

Noelia Díaz[1,6], Kai Kruse[1,6], Tabea Erdmann[2], Annette M. Staiger[3,4,5], German Ott[3], Georg Lenz[2], and Juan M. Vaquerizas[1,*]

Affiliations:

1. Max Planck Institute for Molecular Biomedicine, Roentgenstrasse 20, 48149 Muenster, Germany.
2. Department of Medicine A, Hematology, Oncology and Pneumology, University Hospital Muenster, 48149 Muenster, Germany.
3. Department of Clinical Pathology, Robert-Bosch-Hospital, Auerbachstrasse 110, 70376 Stuttgart, Germany.
4. Dr. Margarete Fischer Bosch Institute of Clinical Pharmacology, Auerbachstrasse 112, 70376 Stuttgart, Germany.
5. Eberhard Karls Universität Tübingen, Geschwister-Scholl-Platz, 72074 Tübingen, Germany.
6. These authors have contributed equally to this work.

*** Correspondence to (email):**
Juan M. Vaquerizas (jmv@mpi-muenster.mpg.de)

Keywords:
Chromosome conformation capture; low input Hi-C; Low-C; primary tissue; TAD; chromosome compartments; DLBCL; translocations; cancer.

Abstract

Chromatin conformation constitutes a fundamental level of eukaryotic genome regulation. However, our ability to examine its biological function and role in disease is limited by the large amounts of starting material required to perform current experimental approaches. Here, we present Low-C, a Hi-C method for low amounts of input material. By systematically comparing Hi-C libraries made with decreasing amounts of starting material we show that Low-C is highly reproducible and robust to experimental noise. To demonstrate the suitability of Low-C to analyse rare cell populations, we produce Low-C maps from primary B-cells of a diffuse large B-cell lymphoma patient. We detect a common reciprocal translocation t (3;14) (q27;q32) affecting the BCL6 and IGH loci and abundant local structural variation between the patient and healthy B-cells. The ability to study chromatin conformation in primary tissue will be fundamental to fully understand the molecular pathogenesis of diseases and to eventually guide personalised therapeutic strategies.

Introduction

The three-dimensional organisation of chromatin in the nucleus plays a fundamental role in regulating gene expression, and its misregulation has a major impact in developmental disorders (Lupiáñez et al. 2015; Franke et al. 2016) and diseases such as cancer (Hnisz et al. 2016). The development of chromosome conformation capture (3C) (Dekker et al. 2002) assays and, in particular, their recent high-throughput variants (e.g. Hi-C), have enabled the examination of 3D chromatin organisation at very high spatial resolution (Rao et al. 2014; Lieberman-Aiden et al. 2009). However, the most widely used current experimental approaches rely on the availability of a substantial amount of starting material – on the order of millions of cells – below which experimental noise and low sequencing library complexity become limiting factors (Belaghzal, Dekker, and Gibcus 2017). Thus far, this restricts high-resolution analyses of population Hi-C to biological questions for which large numbers of cells are available and limits the implementation of chromatin conformation analyses for rare cell populations such as those commonly obtained in clinical settings. While single-cell approaches exist (Nagano et al. 2013; Ramani et al. 2017; Stevens et al. 2017; Flyamer et al. 2017), they typically operate on much lower resolutions than population-based approaches and require an extensive set of specialist skills and equipment that might be out of reach for the average genomics laboratory.

Recently, two methods have been developed to measure chromatin conformation using low amounts of starting material (Z. Du et al. 2017; Ke et al. 2017). However, the lack of a systematic comparison of the data obtained with these approaches and conventional *in situ* Hi-C limits our understanding of the technical constraints imposed by the amounts of starting material available. In addition, it remains to be demonstrated whether these methods could be directly applied to samples with clinical interest, such as for example, tumour samples.

Here, we present Low-C, an improved *in situ* Hi-C method that allows the generation of high-quality genome-wide chromatin conformation maps using very low amounts of starting material. We validate this method by comparing chromatin conformation maps for a controlled cell titration, demonstrating that the obtained maps are robust down to 1,000 cells of starting material and are able to detect all conformational features – compartments, topologically associating domains (TADs) and loops– similarly as maps produced with a higher number of cells. Finally, we demonstrate the applicability of Low-C to clinical samples by generating chromatin conformation maps of primary B-cells from a diffuse large B-cell lymphoma (DLBCL) patient. Computational analysis of

80 the data allows us to detect patient-specific translocations and substantial amounts of variation in topological features.

Results

Low-C: A Hi-C method for low amounts of input material

We first sought to develop a Hi-C method for low amounts of input material. To do so, we modified the original *in situ* Hi-C protocol (Rao et al. 2014), which recommends 5-10 million (M) starting cells, to allow for much smaller quantities of input material. The modifications are subtle, involving primarily changes in reagent volume and concentrations, as well as timing of the individual experimental steps (Fig. 1a, Methods, Supplementary Table 1). The combined changes, however, are highly effective, allowing us to produce high-quality Hi-C libraries from starting cell numbers as low as one thousand (1k) cells.

To assess the feasibility and limitations of Low-C, we prepared libraries for progressively lower numbers of mouse embryonic stem cells (mESC) using two different restriction enzymes (Supplementary Table 2). Each library was deep-sequenced to an average depth of 100-150x10^6 reads and processed using a computational Hi-C pipeline with particular emphasis on the detection and filtering of experimental biases (Methods). The ratios of the number of *cis*- and *trans*-contacts (Lajoie, Dekker, and Kaplan 2015) indicate a high library quality for all samples (Methods) (Supplementary Table 3). Visual inspection of normalised Hi-C maps for 1M to 1k cells revealed a high degree of similarity between Low-C samples, with TADs clearly identifiable at a resolution of 50kb (Fig. 1b, Supplementary Fig. 1a). To determine the degree of similarity between samples, we computed correlations of all contact intensities against the 1M sample, which showed very high levels of reproducibility (Fig. 1b, Pearson correlation coefficient R>=0.95 in all cases). To evaluate the overall level of reproducibility with other protocols, we performed a comparison of a pooled Low-C dataset, merging samples up to 50k cells, to a previously published mESC Hi-C dataset (Dixon et al. 2012), to account for differences in sequencing depth. This comparison revealed a strong contact intensity correlation (R=0.97), that was further confirmed by a principal component analysis that displayed strong clustering of Low-C samples and high similarity of Low-C to other mESC datasets (Supplementary Fig. 1b). In addition, we performed aggregate TAD and aggregate loop analysis (Flyamer et al. 2017) on the 1k and 1M samples (Fig. 1c,e), which revealed highly consistent TAD (Fig. 1d) and loop strengths (Fig. 1f) across datasets. Overall these results suggest that Low-C is a robust method to generate chromatin conformation maps using small amounts of input material.

Low-C data have similar properties as conventional Hi-C

We next wanted to ensure that the number of input cells does not limit the range of observations one can obtain from a Hi-C matrix. In a Hi-C experiment, each DNA fragment can only be observed in a single ligation product, limiting the number of possible contacts of the corresponding genomic region to twice the number of input cells (in a diploid cell line). This raises the concern for Low-C that low-probability contacts – such as those in far-*cis* – would be lost for very small numbers of cells. To test this, we calculated the correlation of contact intensities at increasing distances for the 100k, 10k, and 1k against the 1M sample. Reassuringly, while the expected decrease in correlation with distance was apparent, the decrease in contact correlation is independent of the input cell number (Fig. 2a), indicating that the loss of low-probability contacts was not a limiting factor for input cell numbers as low as one thousand. Furthermore, the remaining differences in correlation disappeared when comparing sub-sampled matrices to the same number of valid pairs (Supplementary Fig. 2a), suggesting that sequencing depth, and not the initial number of cells, is the main determinant of the correlation coefficients. We also confirmed that diversity of Hi-C contacts, measured as the absolute number of unique fragment pairs in a Low-C experiment, is not affected by the amount of input cells, but it is primarily a function of sequencing depth (Supplementary Fig. 2b-c).

To explore the limits of Low-C, we performed an extensive characterisation of the properties of these libraries. Previous work had identified systematic biases in Hi-C data that can serve as read-outs for the efficiency of Hi-C library generation (Yaffe and Tanay 2011; Jin et al. 2013; Cournac et al. 2012) (Fig. 2b, Supplementary Fig. 3a-f). Most notably, PCR duplicates indicate low library complexity – a limitation that has been previously described when trying to scale down the Hi-C protocol (Belaghzal, Dekker, and Gibcus 2017) – while an excess of different types of ligation products, such as self-ligated fragments, can point to problems in the digestion and ligation steps (Methods). Unsurprisingly, given the higher need for amplification, we find that PCR duplicates increase with lower amounts of starting material, with roughly 20% of read pairs identified as duplicates in the 1k sample (Fig. 2b). Ligation errors, however, remained more or less constant across samples, irrespective of the number of cells (Supplementary Fig. 4). Other low-input Hi-C datasets (Z. Du et al. 2017a; Ke et al. 2017) display similar biases (Fig. 2a,b), confirming that decreasing library complexity appears to be the strongest limitation on the lowest number of input cells that is feasible for low-input Hi-C approaches.

Compartments, TADs and loops can be detected in Low-C data

Next, we set out to ensure that not only the Hi-C maps themselves, but also measures
155 derived from them are reproducible and unaffected by differences in input cell number. To do so, we calculated several established and widely used Hi-C measures on the Low-C matrices at 50kb resolution, namely: the profile of expected contacts at increasing distances between genomic regions (Lieberman-Aiden et al. 2009) (Fig. 3a, Supplementary Fig. 5); the correlation matrix and its first eigenvector, used to derive
160 AB compartments (Lieberman-Aiden et al. 2009) (Fig. 3b, Supplementary Fig. 6); and the insulation score (Crane et al. 2015), commonly used to infer TADs and TAD boundaries (Kruse et al. 2016) (Fig. 3c). All three examples of Hi-C measures are consistent with results from conventional Hi-C and showed high reproducibility between the 1M and 1k samples with no apparent dependence on the number of input
165 cells, demonstrating that Low-C libraries are highly consistent and reproducible for input cell numbers as low as one thousand cells.

Generation of Low-C maps for DLBCL primary tissue

Given our ability to obtain high quality chromatin conformation maps using low amounts of input material, we sought to determine whether the technique could be
170 applied in a real-world scenario where the amount of starting material is likely to be the limiting factor in obtaining chromatin contacts maps. To test this, we performed Low-C on a diffuse large B-cell lymphoma (DLBCL) sample and in normal B-cells extracted from a healthy donor as a control (see Methods). Generating chromatin contact maps with low amounts of input material is beneficial not only because it allows to test the
175 3D chromatin conformation directly in the diseased cells, but also since it maximises the availability of tissue for other procedures and minimises patients' burden from having to undergo repeated biopsies to obtain extra material.

Patient and donor CD20+ lymphocytes were isolated from lymph nodes and blood, respectively, using a magnetic microbead-labelled CD20+ antibody and magnetic-
180 activated cell sorting (MACS) (Yan et al. 2009) (Fig. 4a, Methods). We confirmed that the cell fixation procedure did not affect the efficiency of MACS sorting and that we were able to correctly distinguish CD20+ from CD20- cells in a mixture of HBL1 and Jurkat cells (Supplementary Fig. 7) and in the peripheral blood mononuclear cells (PBMCs) from the control sample (Supplementary Fig. 8), respectively before and after
185 formaldehyde fixation. Using the MACS approach, we were able to isolate the majority of B-cells from the control sample and the cell line mixture, although a non-critical fraction of B-cells was lost during the process (Supplementary Fig. 9). The same was

true for the patient sample where fixation did not affect the surface molecules needed for MACS sorting (Supplementary Fig. 10) and where the eluted CD20+ cell population was made up of 95.5% B-cells (Supplementary Fig. 10f). We then performed Low-C on approximately 50k cells from each of the patient and control samples and deep-sequenced the resulting libraries to approximately 500 million (patient) and 300 million (control) reads (Supplementary Table 2). The resulting chromatin maps show a high degree of similarity between the patient and control B-cells (Fig. 4b). TADs (ordinary and loop domains) and loops are clearly distinguishable in the maps, and *de novo* loop calling using HICCUPS (Rao et al. 2014) and subsequent aggregate loop analysis (Fig. 4c) confirms that these can be identified automatically with high confidence. Overall these results confirm that Low-C can be successfully used in a clinical setup to obtain high-quality chromatin conformation maps directly from primary patient tissue.

Identification of structural variation in patient Low-C data

Structural variation and, in particular, genome rearrangements are a characteristic feature in many cancers (Weischenfeldt et al. 2013). Since chromatin contact maps have an intrinsic bias for detecting interactions that happen in the proximal linear sequence (Lieberman-Aiden et al. 2009), Hi-C-like data can be used to detect structural variation (Harewood et al. 2017; Lin et al. 2018; Lupiáñez et al. 2015; Franke et al. 2016; Hnisz et al. 2016; van de Werken et al. 2012; Krijger and de Laat 2016; Zepeda-Mendoza et al. 2015; Simonis et al. 2009). In order to detect potential translocations in the DLBCL sample in a fully automated and unbiased manner, we performed virtual 4C (V4C) for the patient and control data. Specifically, we considered each 25kb bin in the genome in turn as a viewpoint to detect cases that display significant amounts of signal anywhere in the genome of the DLBCL cells that do not appear in control B-cells (Methods). Hi-C maps at locations of putative structural variations were then browsed manually to remove false positives.

Most prominently, this analysis identified two regions of interest on chromosome 3q27 separated by ~8Mb, with significant interactions with chromosome 14q32 (Fig. 5a-c). As expected, a normal V4C profile was observed around the viewpoint in chromosome 3 for both regions in both control and DLBCL cells. In contrast, the V4C profile found for the interacting regions on chromosome 14q32 was only apparent in the patient data, suggesting that the interactions are patient-specific. A closer examination of the genes located in the patient-interacting regions revealed that the first viewpoint (Fig. 5, magenta shaded region) lies directly at the BCL6 gene, a transcription factor known to be affected in DLBCL, while the interacting region on chromosome 14q32 lies at the

immunoglobulin heavy-chain (IGH) locus (Fig. 5d, e), suggesting a t (3q27:14q32) reciprocal translocation. Translocations involving BCL6 are among the most commonly observed rearrangements in DLBCL (C Bastard et al. 1992; Kramer et al. 1998; Offit et al. 1994), with one study reporting a ~30% (14/46) penetrance in DLBCL patients (Christian Bastard et al. 2018). The second viewpoint with significant interactions towards the telomeric end of chromosome 3 (Fig. 5, green shaded region) interacts with a more centromeric location on chromosome 14. The pattern of interaction signal decay over linear distance in the trans-chromosome interactions map suggests a breakpoint around 195.2Mb (Fig. 5d,f, black triangles) and allows us to manually reconstruct the most likely rearrangement of these regions in DLBCL from the Hi-C data: the telomeric ends of both chromosomes are involved in a reciprocal translocation, with breakpoints around chr3:187.7Mb and chr14:105.9Mb (Fig. 5h). To validate our data, we performed a fluorescence *in situ* hybridisation (FISH) analysis that confirmed a rearrangement of the BCL6 gene (Fig 5i), providing orthogonal validation of the Hi-C findings. In addition, the lack of Hi-C signal between the breakpoints in chromosome 3 and 14 suggests that the regions chr3:187.7-195.2Mb and chr14:105.6-105.9Mb have been lost on one pair of chromosomes, generating regions of loss of heterozygosity in the remaining chromosome. Interestingly, we find another smaller rearrangement involving ANXA3 on chromosome 4 and EDAR2 on chromosome X (Supplementary Fig. 11). Misregulation of ANXA3 is known to promote tumour growth, metastasis and drug resistance in both breast cancer (R. Du et al. 2018) and hepatocellular carcinoma (Tong et al. 2015). In summary, our results demonstrate that Low-C can be used directly on primary tissue to detect patient-specific chromosomal rearrangements in an unbiased manner.

Extensive rewiring of chromatin organisation in DLBCL cells

Visual comparison of the patient and control chromatin contact maps revealed numerous local structural differences. For example, the region undergoing loss of heterozygosity reported above (chr3:187.7-195.2Mb; Fig. 5d, arrow) displays a clear gain of TAD structure encompassing the genes TP63, a member of the p53 family of transcription factors that has been previously associated with cancer, and the tumour protein p63 regulated gene-1 (TPRG1) which lies in the same *de novo* established TAD, suggesting their potential co-regulation. To evaluate the overall extent of changes in chromatin conformation at the TAD structure level between the two samples, we used the insulation score (Crane et al. 2015) to determine TAD boundaries in both samples and looked for regions with broad changes in the Hi-C signal (Methods). Using a conservative threshold, we detected 648 regions in the

genome with notable changes in local Hi-C contacts (Supplementary Table 4). Out of these, 37 appear to be *de novo* TADs, which in many cases overlap with known disease-related genes such as PTPRG (LaForgia et al. 1991) (Fig. 6c), APBB2 (Deffenbacher et al. 2012) (Fig. 6d), and TEAD1 (Zhou et al. 2017; Schmid et al. 2015) (Fig. 6e). Overall, we observe the majority of changes in TAD structure to be patient-specific gains, whereas the loss of TADs present in normal B-cells in the patient is a relatively rare event (Fig. 6b). Altogether, our results demonstrate that Low-C can be used to study chromatin contact differences between patient samples at the TAD level and that there are significant differences in TAD structure between DLBCL and normal B-cells.

Discussion

The development of high-throughput genome-wide techniques to measure chromatin conformation has been instrumental to further our understanding of the biological importance of the three-dimensional organisation of chromatin in the nucleus. In addition to providing a local environment where enhancer-promoter interactions can orchestrate the correct deployment of gene expression programmes during development, the three-dimensional chromatin conformation is fundamental to establish proper spatial boundaries, that provide enhancer insulation and limit their function to those genes that need to be regulated. Chromatin conformation at the level of TADs seems to be fairly static for fully differentiated cells (Nora et al. 2012; Dixon et al. 2012, 2015), although dynamic changes in TAD structure can be observed during development in organisms ranging from *Drosophila* to mammals (Hug et al. 2017; Bonev et al. 2017; Z. Du et al. 2017; Ke et al. 2017), highlighting their dynamic behaviour.

A current limitation for our understanding of these dynamic changes and the potential differences in 3D chromatin conformation between tissues or in a disease context is the high amount of material that is usually necessary to perform these experiments. While single-cell Hi-C methods exist, these are usually only able to capture a small fraction of the chromatin contacts that occur across the genome. This results in sparse chromatin maps of low resolution that usually rely on TAD calls made using standard Hi-C maps, limiting their applicability in comparing samples or finding *de novo* TADs.

Here, we introduce Low-C, an improved Hi-C method that allows the generation of high-resolution chromatin contact maps using low amounts of input material. Beyond existing low input Hi-C approaches (Z. Du et al. 2017; Ke et al. 2017), we perform a thorough comparison of Low-C maps and their derived measurements in a controlled environment to systematically demonstrate that Low-C is not affected by biases originating from the amount of starting material. We also show that the method is robust and applicable to mammalian samples down to one thousand cells without compromising the quality of the resulting datasets. Therefore, our results establish Low-C as an efficient method to study chromatin conformation for rare cell populations, where the collection of material currently necessary to perform population-based Hi-C protocols is infeasible. These include transient developmental stages (Hug et al. 2017; Z. Du et al. 2017; Ke et al. 2017), as well as systems of medical relevance, such as primary tissue from patient samples, where an examination of changes in chromatin conformation between healthy and disease cells might shed light on the etiology of the disease.

To demonstrate the usability of this approach in a real-world scenario, we generated Low-C maps for a DLBCL patient sample. Since changes in chromatin contact profiles and genomic rearrangements can be detected very easily through Hi-C approaches (Engreitz et al. 2012), we developed an unbiased approach to systematically detect translocations using these data, uncovering a known reciprocal translocation in this patient biopsy. This, together with recent reports of similar approaches in other tumour types (Harewood et al. 2017) highlights the clinical applicability of this technology. An added benefit of our approach when compared with previous work in primary tissue samples is the generation of high-quality genome-wide chromatin interaction maps, which allows us to examine the level of variability between cells in health and disease. In fact, we detect a large amount of variation at the TAD level, in particular in the DLBCL sample, which gains a significant amount of structure. Interestingly, in several cases the emergence of novel chromatin structural features coincides with the genomic location of genes previously associated with cancer, such as TP63 and ANXA3. Whereas the current maps do not allow us to determine cause or consequence for these changes, a broader examination of these changes in larger cohorts of patient samples, together with an integrative analysis of gene expression and chromatin states might provide insight into the causal relationships between these in a disease-specific and patient-specific manner.

Despite the increased applicability of our method, there are still a number of factors to take into consideration when planning such experiments. First, tissue heterogeneity or the presence of healthy cells in biopsies can become an issue with increasingly lower cell numbers. Specifically, the lower the input cell number, the greater the impact of contaminations or variabilities in sample composition will be on the averaged chromatin structures visible in the Hi-C maps. These might obfuscate or increase the uncertainty about specific structural observations. In our DLBCL analysis, we set out to minimise these effects by coupling our Low-C to efficient cell sorting techniques. Second, decreasing library complexity is still the current limiting factor for low input Hi-C studies (Belaghzal, Dekker, and Gibcus 2017), and a significant amount of PCR duplicates are to be expected when reducing the amount of starting material. Third, a further general limitation for bulk Hi-C methods, regardless the initial cell input, is that long-range three-dimensional contacts between gene promoters and enhancers are likely to be missed, since they usually happen within the context of TAD interactions. Therefore, to study these important interactions, which have been shown to affect gene regulation and are associated with the risk for various types of diseases (Javierre et al. 2016; Martin et al. 2015), it might be useful to couple Low-C with capture or

promoter-capture techniques (Hughes et al. 2014; Dryden et al. 2014; Mifsud et al. 2015; Jäger et al. 2015), that will allow the retrieval of these specific interactions.

In summary, our data demonstrates that it is feasible to obtain high-quality genome-
345 wide chromatin contact maps from low amounts of input material. We anticipate that the robustness and relatively simple implementation will make Low-C an attractive option that will facilitate bringing the analysis of chromatin architecture within reach of personalised clinical diagnostics.

350 # Methods

Low-C protocol

We followed the general protocol for in situ Hi-C as described previously (Rao et al. 2014), which we adapted for use on low cell numbers. Mainly, differences were related to adjustments in the volume of the reactions, a shortening of the digestion step, a
355 removal of biotin from the unligated fragments, and an alternative strategy for size-selection during library preparation. For a detailed step by step protocol please see Fig. 1a and Supplementary Table 1.

Cell culture

mESC OG2 cells were cultured as described previously (Shi et al. 2008), FACS-sorted,
360 selected for positive eGFP expression and collected in PBS. Cells were then pelleted (300 g, 4°C for 10 min) and resuspended in 1 ml PBS.

Patient and Control samples processing

Peripheral blood mononuclear cells (PBMCs) were obtained either from a blood extraction from a healthy donor or from a lymph node biopsy from a DLBCL patient.
365 The patient sample came from the Department of Clinical Pathology at the Robert-Bosch-Hospital in Stuttgart (Germany) and its informed consent for retrospective analysis was approved by the ethics committee of the Medical Faculty, Eberhard-Karls-University and University Hospital Tübingen (reference no. 159/2011BO2). PBMCs from the control came from a donor from the Department of Medicine A, Hematology,
370 Oncology and Pneumology, University Hospital Muenster in Muenster (Germany).

Control PBMCs were isolated from the in-between layer by density gradient centrifugation with Biocoll (Biochrom AG, Germany) and were then frozen at -80°C for preservation. The patient sample came from a biopsy of a lymph node. Briefly, the biopsy was immediately cut into pieces, homogenized and re-suspended generating a
375 cell suspension that was then frozen and kept at -80°C, as previously described (Staiger et al. 2017).

Once the samples were thawed, cells were cross-linked in a 1% formaldehyde and quenched with 2.5M Glycine solution (for details check the detailed Low-C protocol at Supplementary Table 1). A test to ensure that formaldehyde fixation won't affect the
380 surface molecules was performed before and after fixation (Supplementary Figure 8). The viability of the surface molecules on a mixture of HBL1 and Jurkat cells was

assessed by staining with a CD19-PE (FL2) and CD20-FITC (FL1) antibodies (Supplementary Figure 7).

B-cells were then isolated by MACS-sorting (Yan et al., 2009) using a positive selection kit (Miltenyi Biotec, 130-091-104). Briefly, CD20+ cells were labelled using magnetic coated CD20 MicroBeads, the cell suspension was loaded onto a MACS LS column (Miltenyi Biotec, 130-041-306) and placed on a magnetic field generated by a MACS Separator. The CD20+ cells were retained into the column while the flow through (unlabelled cells) was eliminated. Then the column was removed from the MACS Separator, the magnetically retained CD20+ cells were then eluted and collected into a 15 ml Falcon tube. The performance of the MACS sorting was assessed by checking the B-cell presence and its proportions in the flow through as well as in the eluted portion for the control PBMCs (Supplementary Figure 9a-b, e-f), for the mixed cell population sample (Supplementary Figure 9c-d, g-h) and for the patient sample (Supplementary Figure 10).

Once the eluted samples were recovered, we proceeded with the lysis and the rest of the Hi-C library preparation as described in detail in Supplementary Table 1.

Bioinformatics processing of Low-C and Hi-C libraries

Prior to mapping, the two mates of each paired-end reads sample were scanned for MboI ligation junctions, indicating sequencing through a Hi-C ligation product. If a junction was found, the read was split. Reads were then mapped independently to the *M. musculus* reference genome (mm10) using BWA-MEM (0.7.17), which may also result in split reads where the ends map to different locations in the genome. Those reads that did not align uniquely to the genome or that had a mapping quality lower than 3 were filtered out. Read pairs where one read was filtered out are discarded.

For the remaining read pairs, there are three possibilities: (i) none of the two reads in a mate pair was split in the pre-processing or mapping step (see above), (ii) one read in the pair was split, resulting in 3 mapped reads with the same ID, and (iii) one read in a pair was split multiple times or both reads were split at least once, resulting in more than 3 reads with the same ID. In case (iii) the mate pair is filtered out, as the exact interacting genomic location cannot be determined; in case (ii) the pair is considered valid if two reads map to the same genomic location (within 100bp), otherwise it is discarded; case (i) is considered valid.

Restriction fragments in the genome were identified computationally using known restriction sequences of MboI and HindIII, and the remaining pairs of reads were assigned to the restriction fragments.

Obtaining valid pairs of reads

Pairs were filtered out if: i) the mapped reads' distance to the nearest restriction site was larger than 5kb, ii) both reads mapped to the same fragment, or iii) the orientation and distance of reads indicated a ligation or restriction bias (Jin et al. 2013; Cournac et al. 2012). Briefly, paired reads mapping in the same direction on the chromosome likely originate from a pair of fragments that had a cut restriction site between them and that had subsequently ligated – these were considered valid. Paired reads mapping in opposite directions may indicate that the reads map to a single large fragment with one or more uncut restriction sites. In this case, pairs facing inward would have originated from an unligated, pairs facing outward from a self-ligated fragment. At large genomic distances, there are approximately equal numbers of same and opposite orientation pairs. At shorter distances, there is an increased likelihood of uncut restriction sites between two reads, and pairs in opposite direction are filtered out. For every dataset, both the inward and outward ligation cut-offs have been fixed at 10kb.

Finally, pairs were marked as PCR duplicates if another pair existed in the library that mapped to the same locations in the genome, with a tolerance of 2bp. In those cases, only one pair from all duplicate ones for a given locus was retained for downstream processing. Finally, the genome was partitioned into equidistant bins and fragment pairs were assigned to bins using a previously described strategy (Rao et al. 2014). The resulting contact matrix was filtered for low-coverage regions (with less than 10% of the median coverage of all regions) and corrected for coverage biases using Knight-Ruiz matrix balancing as described before (Rao et al. 2014; Knight and Ruiz 2013). Bins that had no contacts due to filtering were marked as "unmappable".

Cis/trans *ratio calculation*

The cis/trans ratio is calculated as the number of valid intra-chromosomal contacts (cis) to the valid inter-chromosomal contacts (trans). When comparing different species, this ratio will be affected by genome size and the number of chromosomes. We therefore also provide a "species-normalised" cis/trans ratio by multiplying the trans value by the ratio of possible intra-chromosomal to inter-chromosomal contacts

f (the ratio of the number of intra-chromosomal pixels in the Hi-C map to the number of inter-chromosomal pixels).

Observed/expected (OE) Hi-C matrix generation

For each chromosome, we obtain the expected Hi-C contact values by calculating the average contact intensity for all loci at a certain distance. We then transform the normalized Hi-C matrix into an observed/expected (OE) matrix by dividing each normalized observed by its corresponding expected value.

Aggregate TAD/loop analysis

In general, average feature analysis is performed by extracting subsets of the OE matrix (can be single regions along the diagonal, or region pairs corresponding the matrix segments off the diagonal) and averaging all resulting sub-matrices. If the sub-matrices are of different size, they are interpolated to a fixed size using “imresize” with the “nearest” setting from the Scipy Python package.

TADs and loop anchors in Fig. 1 have been obtained from (. TADs and loop anchors in Fig. 4 have been called de novo from their respective datasets (see below). The region size for TADs has been chosen as 3x TAD size, centred on the TAD, and aggregate analyses have been performed in 25kb matrices. The region size around loop anchors has been chosen as 400kb in 25kb matrices.

TAD strength is calculated as in (. Briefly, we calculate the sum of values in the OE matrix in the TAD-region and the sum of values for the two neighbouring regions of the same size divided by two. The TAD strength is then calculated as the ratio of both numbers.

Loop strength is calculated as in (. Briefly, we first calculate the sum of all values in the 300kb region of the Hi-C matrix centred on the loop anchors. As a comparison, we calculate the same value for two control regions, substituting one of the loop anchors for an equidistant region in the opposite direction. The loop strength is then calculated as the original sum of values divided by the average sum of values in the two control regions.

Expected values vs. distance

Intra-chromosomal Hi-C matrix entries (50kb resolution) were binned by distance to the diagonal and divided by the total number of possible contacts at each distance. The resulting average counts were plotted against distance in a log-log plot.

AB compartments

480 For each chromosome separately, the Hi-C matrix was converted to an OE matrix (see above). The OE matrix was then converted into a correlation matrix, where each entry (i, j) represents the Pearson correlation between row i and j of the OE matrix. Finally, the signs of the first eigenvector entries were used to call compartments.

Insulation score and TAD boundaries

485 The insulation score was calculated as described before (Crane et al. 2015), by averaging contacts in a quadratic sliding window along the diagonal of the Hi-C matrix. Insulation scores were then divided by the chromosomal average and log2-transformed. Boundaries were calculated from the vector of insulation scores as previously described (Crane et al. 2015; Hug et al. 2017a). Aggregate TAD plots in
490 Fig. 4, and the insulation and TAD intensity difference plots in Fig. 6 use the intervals between two consecutive boundaries as input.

De novo loop calling

Loops in the DLBCL and B-cell samples have been called using an in-house implementation of HICCUPS (Rao et al. 2014). Briefly, for each entry in the Hi-C matrix,
495 HICCUPS calculates several enrichment values over different local neighbourhoods (termed “donut, lower-left, horizontal and vertical – for definition of the neighbourhoods see the original publication). Each enrichment value is associated with an FDR value for assessing statistical significance. We call loops at a matrix resolution of 25kb and perform filtering exactly as described, only retaining loops that (i) are at least 2-fold
500 enriched over either the donut or lower-left neighbourhood, (ii) are at least 1.5-fold enriched over the horizontal and vertical neighbourhoods, (iii) are at least 1.75-fold enriched over both the donut and lower-left neighbourhood, and (iv) have an FDR <= 0.1 in every neighbourhood. We thus obtain 10,093 loops in the DLBCL and 13,213 loops in the B-cell samples – comparable to the number of loops identified originally in
505 GM12878 cells (Rao et al. 2014).

Identification of structural rearrangements in DLBCL

To generate a list of candidate regions that may have undergone structural rearrangements in DLBCL, we performed Virtual 4C (V4C) for each Hi-C bin of the DLBCL matrix at 50kb resolution (viewpoint), looking for peaks of signal away from the
510 original viewpoint (target) that were not present in normal B-cells.

Specifically, in a Hi-C matrix M of size NxN, we examined each bin i, with i ϵ [0, N]. If any of the bins in the interval [i-7, i+7] is unmappable (see above), it is not considered for further analysis, as we found that regions with mappability issues are typically false-positive rearrangements. We then obtained the vector v of Hi-C signal as row i of M. The viewpoint peak height is then given by v_i. An entry v_j, with j ≠ i, is considered a peak if it is larger than $0.15*v_i$ and 99.5% of all other values in v (the latter was introduced to filter out highly noisy V4C profiles). Peaks closer than 50 bins to i are discarded as local enrichment of contacts.

V4C peaks are called as above for the DLBCL and the B-cell samples. We consider a peak as a putative rearrangement if it only occurs in the DLBCL, but not he B-cell sample. The final list of <100 putative rearrangements could then be inspected by eye in the local and inter-chromosomal Hi-C, eliminating highly noisy Hi-C regions and likely false-positives. Finally, this left just 14 peaks, of which 4 could be attributed to the ANXA3, and 10 to the t (3,14) rearrangements discussed in the manuscript.

Hi-C difference matrices

Plots highlighting differences between DLBCL and B-cell samples (Figure 6) have been obtained by subtracting B-cell from DLBCL Hi-C matrices at 50kb resolution. Pixels without signal in either datasets are removed for clarity.

TAD intensity difference calculations

To quantify the changes in TAD formation and intensity that occur from B-cell to DLBCL (Fig. 6a), we first merged boundaries in both samples (see above), and then calculated the average Hi-C signal between all possible pairs of contacts in-between two consecutive boundaries. This was done separately for the two datasets, and the TAD intensity difference for each region was calculated as the difference in average Hi-C signal of DLBCL and B-cell.

Correlations

All reported correlations are Pearson correlations. Corresponding plots were made using the “hexbin” plotting function on log-transformed counts from the matplotlib library version 2.0.0 in Python (matplolib.org).

The distance correlations in Fig. 2a have been obtained as follows: All intra-chromosomal contacts in a Hi-C map are first binned by distance. Bins are defined as [0-250kb), [250kb, 500kb), [500kb, 750kb), … in the 50kb resolution maps, [0-500kb), [500kb-1Mb), [1.5Mb-2Mb), … in the 100kb resolution maps, and [0-1Mb), [1Mb-2Mb),

[2Mb-3Mb), … in the 250kb resolution maps. For each library (100k, 10k, 1k, Dixon et al., Du et al.) correlations to the 1M sample between all corresponding contact strengths in each bin are calculated. The x axis has been scaled to omit very large distances at which correlations become erratic due to the sparsity of the Hi-C matrix.

Fluorescent in situ hybridisation analysis

Interphase-FISH for BCL6 (Vysis Break apart FISH probe kit, Abbot Molecular Diagnostics, Germany) was performed on 4 µm thick tissue sections cut from FFPE archival tissue blocks as previously described (Horn et al. 2014).

Data availability

The *in situ* Hi-C data generated in this study have been deposited in ArrayExpress and will be available upon publication.

Previously published Hi-C datasets used in this study are available in Gene Expression Omnibus (GEO; Rao et al. 2014: GSE63525; Dixon et. al 2012: GSE35156; Du et al. 2017: GSE82185) and Genome Sequence Archive (GSA) (Ke et al. 2017: PRJCA000241).

Genome annotations have been downloaded from GENCODE, version 27.

References

Bastard, C, H Tilly, B Lenormand, C Bigorgne, D Boulet, A Kunlin, M Monconduit, and H Piguet. 1992. "Translocations Involving Band 3q27 and Ig Gene Regions in Non-Hodgkin's Lymphoma." *Blood* 79 (10): 2527–31. http://www.ncbi.nlm.nih.gov/pubmed/1586704.

Bastard, Christian, Clotilde Deweindt, Jean Pierre Kerckaert, Bernard Lenormand, Annick Rossi, Francesco Pezzella, Christophe Fruchart, Christian Duval, Mathieu Monconduit, and Herve Tilly. 2018. "LA23 Rearrangements in Non-Hodgkin's Lymphoma: Correlation With Histology, Immunophenotype, Karyotype, and Clinical Outcome in 217 Patients." Accessed May 17. http://www.bloodjournal.org/content/bloodjournal/83/9/2423.full.pdf?sso-checked=true.

Belaghzal, Houda, Job Dekker, and Johan H. Gibcus. 2017. "Hi-C 2.0: An Optimized Hi-C Procedure for High-Resolution Genome-Wide Mapping of Chromosome Conformation." *Methods*. doi:10.1016/j.ymeth.2017.04.004.

Bonev, Boyan, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L Papadopoulos, Yaniv Lubling, Xiaole Xu, et al. 2017. "Multiscale 3D Genome Rewiring during Mouse Neural Development." *Cell* 171 (3). Elsevier: 557–572.e24. doi:10.1016/j.cell.2017.09.043.

Cournac, Axel, Hervé Marie-Nelly, Martial Marbouty, Romain Koszul, and Julien Mozziconacci. 2012. "Normalization of a Chromosomal Contact Map." *BMC Genomics* 13 (1). BioMed Central: 436. doi:10.1186/1471-2164-13-436.

Crane, Emily, Qian Bian, Rachel Patton McCord, Bryan R. Lajoie, Bayly S. Wheeler, Edward J. Ralston, Satoru Uzawa, Job Dekker, and Barbara J. Meyer. 2015. "Condensin-Driven Remodelling of X Chromosome Topology during Dosage Compensation." *Nature* 523 (7559): 240–44. doi:10.1038/nature14450.

Deffenbacher, K. E., J. Iqbal, W. Sanger, Y. Shen, C. Lachel, Z. Liu, Y. Liu, et al. 2012. "Molecular Distinctions between Pediatric and Adult Mature B-Cell Non-Hodgkin Lymphomas Identified through Genomic Profiling." *Blood* 119 (16): 3757–66. doi:10.1182/blood-2011-05-349662.

Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. "Capturing Chromosome Conformation." *Science (New York, N.Y.)* 295 (5558): 1306–11. doi:10.1126/science.1067799.

Dixon, Jesse R., Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E. Antosiewicz-

Bourget, Ah Young Lee, Zhen Ye, et al. 2015. “Chromatin Architecture Reorganization during Stem Cell Differentiation.” *Nature* 518 (7539). Nature Publishing Group: 331–36. doi:10.1038/nature14222.

Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. “Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions.” *Nature* 485 (7398). Nature Publishing Group: 376–80. doi:10.1038/nature11082.

Dryden, Nicola H., Laura R. Broome, Frank Dudbridge, Nichola Johnson, Nick Orr, Stefan Schoenfelder, Takashi Nagano, et al. 2014. “Unbiased Analysis of Potential Targets of Breast Cancer Susceptibility Loci by Capture Hi-C.” *Genome Research* 24 (11): 1854–68. doi:10.1101/gr.175034.114.

Du, Ruikai, Bingjie Liu, Lei Zhou, Dong Wang, Xueyan He, Xiaojun Xu, Lixing Zhang, Chaoshi Niu, and Suling Liu. 2018. “Downregulation of Annexin A3 Inhibits Tumor Metastasis and Decreases Drug Resistance in Breast Cancer.” *Cell Death & Disease* 9 (2). Nature Publishing Group: 126. doi:10.1038/s41419-017-0143-z.

Du, Zhenhai, Hui Zheng, Bo Huang, Rui Ma, Jingyi Wu, Xianglin Zhang, Jing He, et al. 2017. “Allelic Reprogramming of 3D Chromatin Architecture during Early Mammalian Development.” *Nature* 547 (7662). Nature Publishing Group: 232–35. doi:10.1038/nature23263.

Engreitz, Jesse M., Vineeta Agarwala, Leonid A. Mirny, K Cibulskis, and C Sougnez. 2012. “Three-Dimensional Genome Architecture Influences Partner Selection for Chromosomal Translocations in Human Disease.” Edited by Shawn Ahmed. *PLoS ONE* 7 (9). Public Library of Science: e44196. doi:10.1371/journal.pone.0044196.

Flyamer, Ilya M., Johanna Gassler, Maxim Imakaev, Sergey V. Ulyanov, Nezar Abdennur, Sergey V. Razin, Leonid A. Mirny, et al. 2017. “Single-Nucleus Hi-C Reveals Unique Chromatin Reorganization at Oocyte-to-Zygote Transition.” *Nature* 544 (7648). Nature Publishing Group: 110–14. doi:10.1038/nature21711.

Franke, Martin, Daniel M. Ibrahim, Guillaume Andrey, Wibke Schwarzer, Verena Heinrich, Robert Schöpflin, Katerina Kraft, et al. 2016. “Formation of New Chromatin Domains Determines Pathogenicity of Genomic Duplications.” *Nature* 538 (7624). Nature Publishing Group: 265–69. doi:10.1038/nature19800.

Harewood, Louise, Kamal Kishore, Matthew D. Eldridge, Steven Wingett, Danita Pearson, Stefan Schoenfelder, V. Peter Collins, and Peter Fraser. 2017. “Hi-C as

a Tool for Precise Detection and Characterisation of Chromosomal Rearrangements and Copy Number Variation in Human Tumours." *Genome Biology* 18 (1). BioMed Central: 125. doi:10.1186/s13059-017-1253-8.

Hnisz, Denes, Abraham S. Weintraub, Daniel S. Day, Anne-laure Valton, Rasmus O. Bak, Charles H. Li, Johanna Goldmann, et al. 2016. "Activation of Proto-Oncogenes by Disruption of Chromosome Neighborhoods." *Science* 351 (6280): 1454–58. doi:10.1126/science.aad9024.

Horn, Heike, Julia Bausinger, Annette M Staiger, Maximilian Sohn, Christopher Schmelter, Kim Gruber, Claudia Kalla, M Michaela Ott, Andreas Rosenwald, and German Ott. 2014. "Numerical and Structural Genomic Aberrations Are Reliably Detectable in Tissue Microarrays of Formalin-Fixed Paraffin-Embedded Tumor Samples by Fluorescence in-Situ Hybridization." *PloS One* 9 (4). Public Library of Science: e95047. doi:10.1371/journal.pone.0095047.

Hug, Clemens B., Alexis G. Grimaldi, Kai Kruse, and Juan M. Vaquerizas. 2017. "Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription." *Cell* 169 (2). Elsevier: 216–228.e19. doi:10.1016/j.cell.2017.03.024.

Hughes, Jim R, Nigel Roberts, Simon McGowan, Deborah Hay, Eleni Giannoulatou, Magnus Lynch, Marco De Gobbi, Stephen Taylor, Richard Gibbons, and Douglas R Higgs. 2014. "Analysis of Hundreds of Cis-Regulatory Landscapes at High Resolution in a Single, High-Throughput Experiment." *Nature Genetics* 46 (2). Nature Publishing Group: 205–12. doi:10.1038/ng.2871.

Jäger, Roland, Gabriele Migliorini, Marc Henrion, Radhika Kandaswamy, Helen E. Speedy, Andreas Heindl, Nicola Whiffin, et al. 2015. "Capture Hi-C Identifies the Chromatin Interactome of Colorectal Cancer Risk Loci." *Nature Communications* 6 (1): 6178. doi:10.1038/ncomms7178.

Javierre, Biola M, Oliver S Burren, Steven P Wilder, Roman Kreuzhuber, Steven M Hill, Sven Sewitz, Jonathan Cairns, et al. 2016. "Lineage-Specific Genome Architecture Links Enhancers and Non-Coding Disease Variants to Target Gene Promoters." *Cell* 167 (5). Elsevier: 1369–1384.e19. doi:10.1016/j.cell.2016.09.037.

Jin, Fulai, Yan Li, Jesse R. Dixon, Siddarth Selvaraj, Zhen Ye, Ah Young Lee, Chia-An Yen, Anthony D. Schmitt, Celso A. Espinoza, and Bing Ren. 2013. "A High-Resolution Map of the Three-Dimensional Chromatin Interactome in Human

Cells." *Nature* 503 (7475). Nature Publishing Group: 290–94. doi:10.1038/nature12644.

Ke, Yuwen, Yanan Xu, Xuepeng Chen, Songjie Feng, Zhenbo Liu, Yaoyu Sun, Xuelong Yao, et al. 2017. "3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis." *Cell* 170 (2). Elsevier: 367–381.e20. doi:10.1016/j.cell.2017.06.029.

Knight, P. A., and D. Ruiz. 2013. "A Fast Algorithm for Matrix Balancing." *IMA Journal of Numerical Analysis* 33 (3). Oxford University Press: 1029–47. doi:10.1093/imanum/drs019.

Kramer, M H, J Hermans, E Wijburg, K Philippo, E Geelen, J H van Krieken, D de Jong, E Maartense, E Schuuring, and P M Kluin. 1998. "Clinical Relevance of BCL2, BCL6, and MYC Rearrangements in Diffuse Large B-Cell Lymphoma." *Blood* 92 (9): 3152–62. http://www.ncbi.nlm.nih.gov/pubmed/9787151.

Krijger, Peter Hugo Lodewijk, and Wouter de Laat. 2016. "Regulation of Disease-Associated Gene Expression in the 3D Genome." *Nature Reviews Molecular Cell Biology* 17 (12). Nature Publishing Group: 771–82. doi:10.1038/nrm.2016.138.

Kruse, Kai, Clemens B. Hug, Benjamín Hernández-Rodríguez, and Juan M. Vaquerizas. 2016. "TADtool: Visual Parameter Identification for TAD-Calling Algorithms." *Bioinformatics* 32 (20). Oxford University Press: 3190–92. doi:10.1093/bioinformatics/btw368.

LaForgia, S, B Morse, J Levy, G Barnea, L A Cannizzaro, F Li, P C Nowell, L Boghosian-Sell, J Glick, and A Weston. 1991. "Receptor Protein-Tyrosine Phosphatase Gamma Is a Candidate Tumor Suppressor Gene at Human Chromosome Region 3p21." *Proceedings of the National Academy of Sciences of the United States of America* 88 (11): 5036–40. http://www.ncbi.nlm.nih.gov/pubmed/1711217.

Lajoie, Bryan R., Job Dekker, and Noam Kaplan. 2015. "The Hitchhiker's Guide to Hi-C Analysis: Practical Guidelines." *Methods* 72 (January): 65–75. doi:10.1016/j.ymeth.2014.10.031.

Lieberman-Aiden, Erez, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. "Comprehensive Mapping of Long-Range Interactions Revelas Folding Principles of the Human Genome." *Science (New York, N.Y.)* 326 (October): 289–93. doi:10.1126/science.1181369.

Lin, Da, Ping Hong, Siheng Zhang, Weize Xu, Muhammad Jamal, Keji Yan, Yingying

Lei, et al. 2018. “Digestion-Ligation-Only Hi-C Is an Efficient and Cost-Effective Method for Chromosome Conformation Capture.” *Nature Genetics* 50 (5). Nature Publishing Group: 754–63. doi:10.1038/s41588-018-0111-2.

Lupiáñez, Darío G., Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, et al. 2015. “Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions.” *Cell* 161 (5): 1012–1025. doi:10.1016/j.cell.2015.04.004.

Martin, Paul, Amanda McGovern, Gisela Orozco, Kate Duffus, Annie Yarwood, Stefan Schoenfelder, Nicholas J. Cooper, et al. 2015. “Capture Hi-C Reveals Novel Candidate Genes and Complex Long-Range Interactions with Related Autoimmune Risk Loci.” *Nature Communications* 6 (November): 10069. doi:10.1038/ncomms10069.

Mifsud, Borbala, Filipe Tavares-Cadete, Alice N Young, Robert Sugar, Stefan Schoenfelder, Lauren Ferreira, Steven W Wingett, et al. 2015. “Mapping Long-Range Promoter Contacts in Human Cells with High-Resolution Capture Hi-C.” *Nature Genetics* 47 (6): 598–606. doi:10.1038/ng.3286.

Nagano, Takashi, Yaniv Lubling, Tim J Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean, Ernest D Laue, Amos Tanay, and Peter Fraser. 2013. “Single-Cell Hi-C Reveals Cell-to-Cell Variability in Chromosome Structure.” *Nature* 502 (7469). Nature Publishing Group: 59–64. doi:10.1038/nature12593.

Nora, Elphège P, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, et al. 2012. “Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre.” *Nature* 485 (7398): 381–85. doi:10.1038/nature11049.

Offit, Kenneth, Francesco Lo Coco, Diane C. Louie, Nasser Z. Parsa, Denis Leung, Carol Portlock, Bihui H. Ye, et al. 1994. “Rearrangement of the Bcl-6 Gene as a Prognostic Marker in Diffuse Large-Cell Lymphoma.” *New England Journal of Medicine* 331 (2). Massachusetts Medical Society : 74–80. doi:10.1056/NEJM199407143310202.

Ramani, Vijay, Xinxian Deng, Ruolan Qiu, Kevin L Gunderson, Frank J Steemers, Christine M Disteche, William S Noble, et al. 2017. “Massively Multiplex Single-Cell Hi-C.” *Nature Methods* 14 (3). Nature Research: 263–66. doi:10.1038/nmeth.4155.

Rao, Suhas S.P. P, Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D.

Bochkov, James T. Robinson, Adrian L. Sanborn, et al. 2014. “A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.” *Cell* 159 (7). Elsevier Inc.: 1665–80. doi:10.1016/j.cell.2014.11.021.

Schmid, Corina A, Mark D Robinson, Nicole A Scheifinger, Sebastian Müller, Sergio Cogliatti, Alexandar Tzankov, and Anne Müller. 2015. “DUSP4 Deficiency Caused by Promoter Hypermethylation Drives JNK Signaling and Tumor Cell Survival in Diffuse Large B Cell Lymphoma.” *The Journal of Experimental Medicine* 212 (5). The Rockefeller University Press: 775–92. doi:10.1084/jem.20141957.

Shi, Yan, Jeong Tae Do, Caroline Desponts, Heung Sik Hahm, Hans R. Schöler, and Sheng Ding. 2008. “A Combined Chemical and Genetic Approach for the Generation of Induced Pluripotent Stem Cells.” *Cell Stem Cell* 2 (6): 525–28. doi:10.1016/j.stem.2008.05.011.

Simonis, Marieke, Petra Klous, Irene Homminga, Robert-Jan Galjaard, Erik-Jan Rijkers, Frank Grosveld, Jules P P Meijerink, and Wouter de Laat. 2009. “High-Resolution Identification of Balanced and Complex Chromosomal Rearrangements by 4C Technology.” *Nature Methods* 6 (11): 837–42. doi:10.1038/nmeth.1391.

Staiger, Annette M., Marita Ziepert, Heike Horn, David W. Scott, Thomas F E Barth, Heinz-Wolfram Bernd, Alfred C. Feller, et al. 2017. “Clinical Impact of the Cell-of-Origin Classification and the MYC/ BCL2 Dual Expresser Status in Diffuse Large B-Cell Lymphoma Treated Within Prospective Clinical Trials of the German High-Grade Non-Hodgkin’s Lymphoma Study Group.” *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 35 (22): 2515–26. doi:10.1200/JCO.2016.70.3660.

Stevens, Tim J., David Lando, Srinjan Basu, Liam P. Atkinson, Yang Cao, Steven F. Lee, Martin Leeb, et al. 2017. “3D Structures of Individual Mammalian Genomes Studied by Single-Cell Hi-C.” *Nature* 544 (7648): 59–64. doi:10.1038/nature21429.

Tong, Man, Tsun-Ming Fung, Steve T. Luk, Kai-Yu Ng, Terence K. Lee, Chi-Ho Lin, Judy W. Yam, et al. 2015. “ANXA3/JNK Signaling Promotes Self-Renewal and Tumor Growth, and Its Blockade Provides a Therapeutic Target for Hepatocellular Carcinoma.” *Stem Cell Reports* 5 (1). Elsevier: 45–59. doi:10.1016/j.stemcr.2015.05.013.

Weischenfeldt, Joachim, Orsolya Symmons, François Spitz, and Jan O. Korbel. 2013.

"Phenotypic Impact of Genomic Structural Variation: Insights from and for Human Disease." *Nature Reviews Genetics* 14 (2). Nature Publishing Group: 125–38. doi:10.1038/nrg3373.

Werken, Harmen J G van de, Gilad Landan, Sjoerd J B Holwerda, Michael Hoichman, Petra Klous, Ran Chachik, Erik Splinter, et al. 2012. "Robust 4C-Seq Data Analysis to Screen for Regulatory DNA Interactions." *Nature Methods* 9 (10). Nature Publishing Group: 969–72. doi:10.1038/nmeth.2173.

Yaffe, Eitan, and Amos Tanay. 2011. "Probabilistic Modeling of Hi-C Contact Maps Eliminates Systematic Biases to Characterize Global Chromosomal Architecture." *Nature Genetics* 43 (11). Nature Publishing Group: 1059–65. doi:10.1038/ng.947.

Yan, Hang, Chen-guang Ding, Pu-xun Tian, Guan-qun Ge, Zhan-kui Jin, Li-ning Jia, Xiao-ming Ding, Xiao-ming Pan, and Wu-jun Xue. 2009. "Magnetic Cell Sorting and Flow Cytometry Sorting Methods for the Isolation and Function Analysis of Mouse CD4+ CD25+ Treg Cells." *Journal of Zhejiang University. Science. B* 10 (12). Zhejiang University Press: 928–32. doi:10.1631/jzus.B0920205.

Zepeda-Mendoza, Cinthya J., Swagatam Mukhopadhyay, Emily S. Wong, Nathalie Harder, Erik Splinter, Elzo de Wit, Melanie A. Eckersley-Maslin, et al. 2015. "Quantitative Analysis of Chromatin Interaction Changes upon a 4.3 Mb Deletion at Mouse 4E2." *BMC Genomics* 16 (1): 982. doi:10.1186/s12864-015-2137-5.

Zhou, Y, T Huang, J Zhang, C C Wong, B Zhang, Y Dong, F Wu, et al. 2017. "TEAD1/4 Exerts Oncogenic Role and Is Negatively Regulated by MiR-4269 in Gastric Tumorigenesis." *Oncogene* 36 (47): 6518–30. doi:10.1038/onc.2017.257.

Acknowledgments

This research was funded by the Max Planck Society. We thank Caitlin McCarthy and Hans Schöler for kindly providing the OG2 mouse stem cells (B6; CBA-Tg (Pou5f1-EGFP)2Mnn/J; stock number 004654) used in this experiment and the Genomics core facility of the Medical Faculty Muenster for sequencing.

Author contributions

J.M.V. conceived and supervised the study. N.D. performed *in situ* Hi-C experiments. K.K. performed computational analyses. T.E., G.O. and G.L provided clinical samples and assisted with B-cell isolation. A.S. performed FISH on clinical samples. N.D., K.K. and J.M.V. analysed and interpreted the data. N.D., K.K. and J.M.V. wrote the manuscript. All authors participated in the discussion of the results, and commented on and approved the manuscript.

Competing financial interests

The authors declare no competing financial interests.

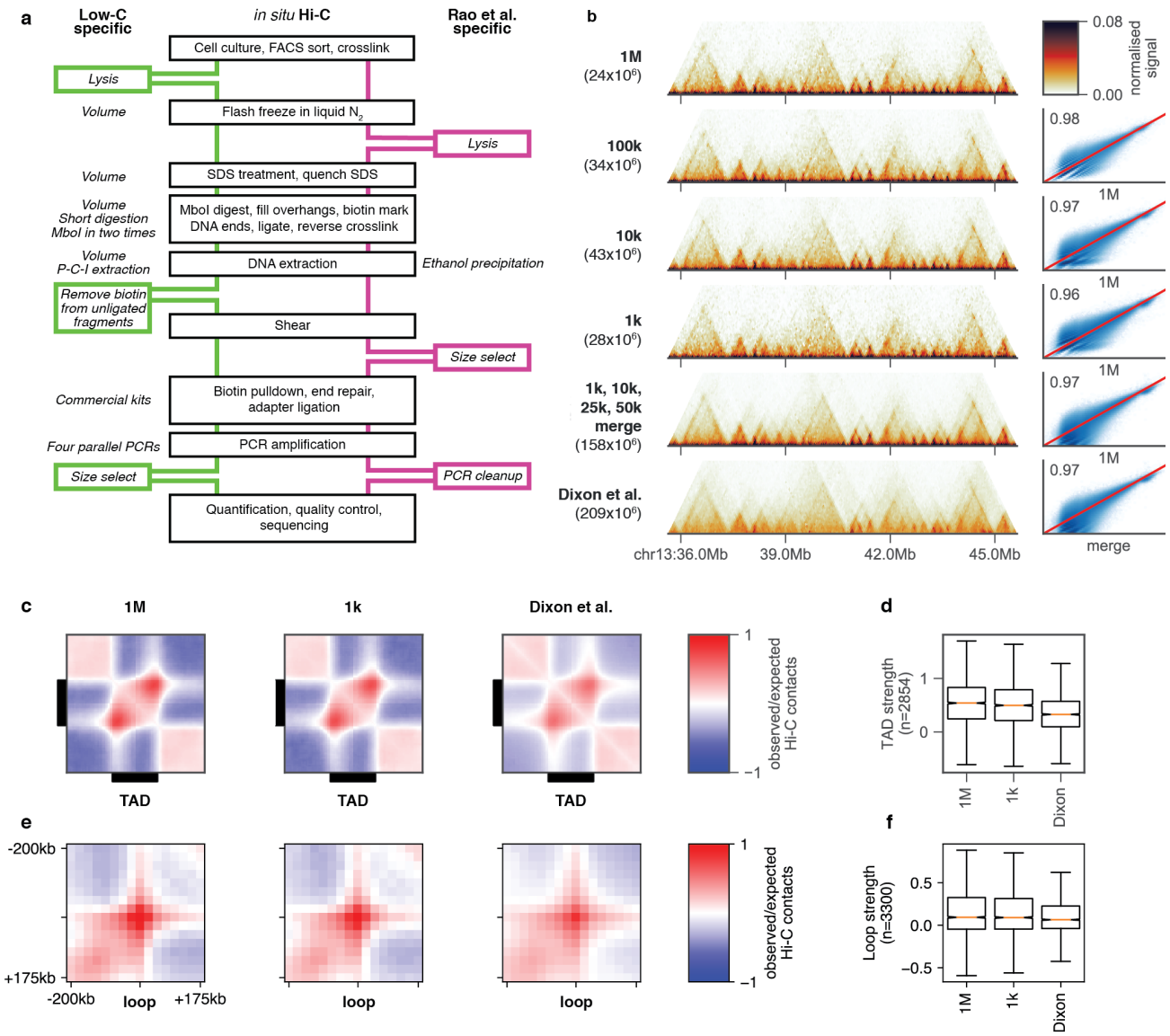


Figure 1. Low-C enables the examination of chromatin architecture for samples with low amounts of input material. (a) Schematic overview of the Low-C protocol and comparison with the previously published *in situ* Hi-C protocol form Rao et al., (2014) (Rao et al. 2014). Black boxes denote common steps in both protocols. Green and magenta boxes denote additional steps in the Low-C and *in situ* Hi-C protocol, respectively. Italicized text marks protocol-specific differences regarding the step next to it. (P-C-I=Phenol-Chloroform-Isoamyl alcohol) (b) Low-C matrices for a 10Mb region on chromosome 13. Input cell numbers for the Hi-C matrices shown span four orders of magnitude (1M to 1k cells). Pixel intensity corresponds to normalized counts. The bottom two Hi-C matrices display data from a merge of Low-C samples (1k, 10k, 25k, and 50k), and a previously published ESC dataset from Dixon et al. (2012) (Dixon et

al. 2012) with similar sequencing depth as the merged sample. In brackets below the sample label we list the number of valid read pairs in each Hi-C library. Next to the Hi-C matrices scatter plots and Pearson correlation coefficient of the contact intensities in the 50kb resolution maps of each sample on the left against the 1M sample are shown. The correlation and scatter plots next to the Dixon et al. dataset correspond to a comparison with the merged sample. Red line indicates identity. (c) Aggregate TAD analysis of the 1M, 1k, and Dixon et al. Hi-C maps. Shown is the average observed/expected ratio of Hi-C signal for regions around all TADs as determined by Rao et al. (2014) (Rao et al. 2014). (d) Comparison of TAD strength (Methods) for the 1M, 1k, and Dixon samples. (e) Aggregate loop analysis showing the average observed/expected Hi-C signal at all loop regions as determined by Rao et al. (2014) (Rao et al. 2014). (f) Comparison of loop strength (Methods) for the 1M, 1k, and Dixon samples.

835

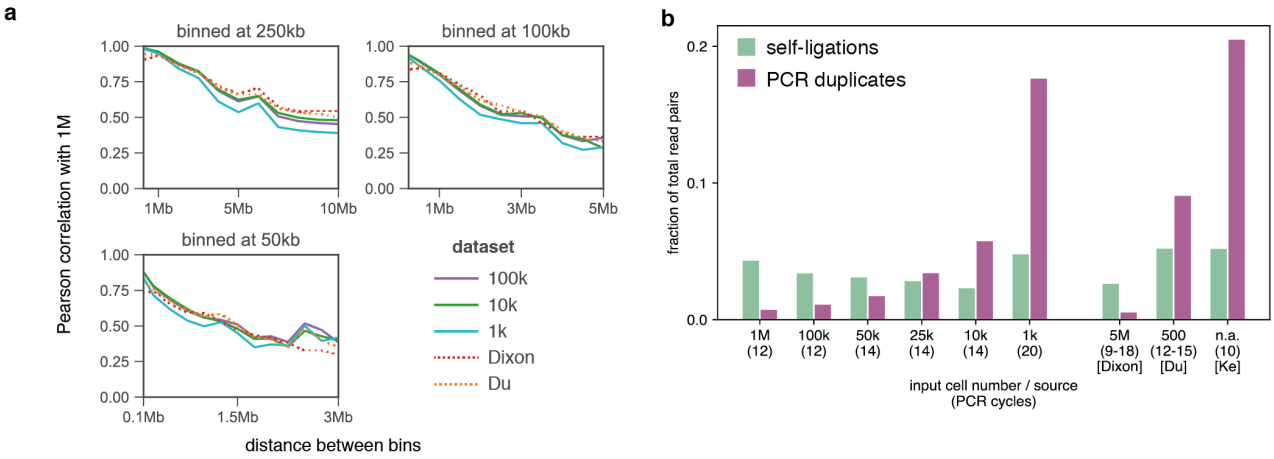


Figure 2. Analysis of experimental biases and quantitative properties of Low-C libraries**.** (a) Pearson correlation coefficient for contact intensities in bin pairs at increasing distances from the diagonal. Each plot represents different bin sizes,
840 indicated above the plot. Colours correspond to input cell number or sample source, respectively. All reported correlations are with the 1M sample. (b) Fraction of fragment pairs affected by and filtered out due to self-ligated fragments or PCR duplicates. Input cell numbers and PCR cycles (brackets) are indicated on the bottom of the plot. Note that all samples are in mESC, except for the Ke et al. (2017) sample on the right, which
845 is in zygote (PN5).

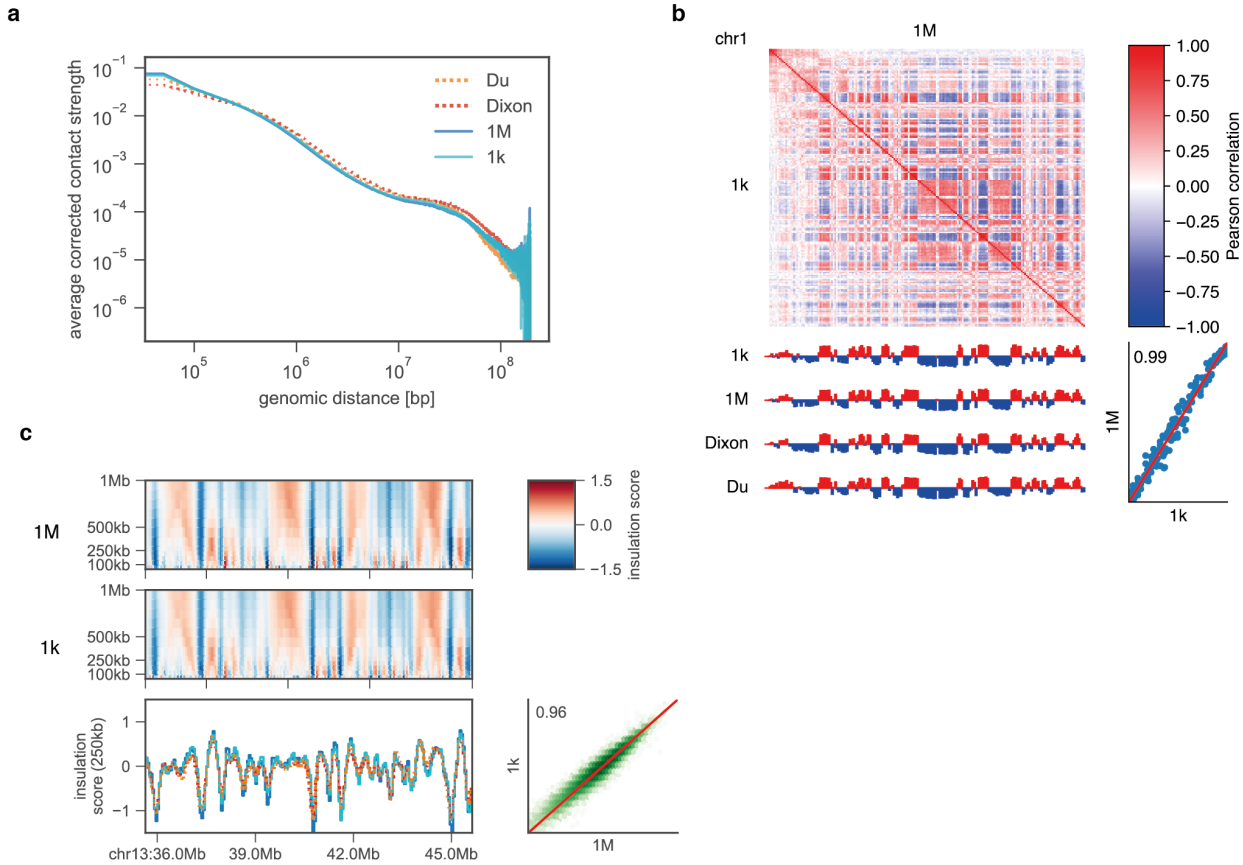
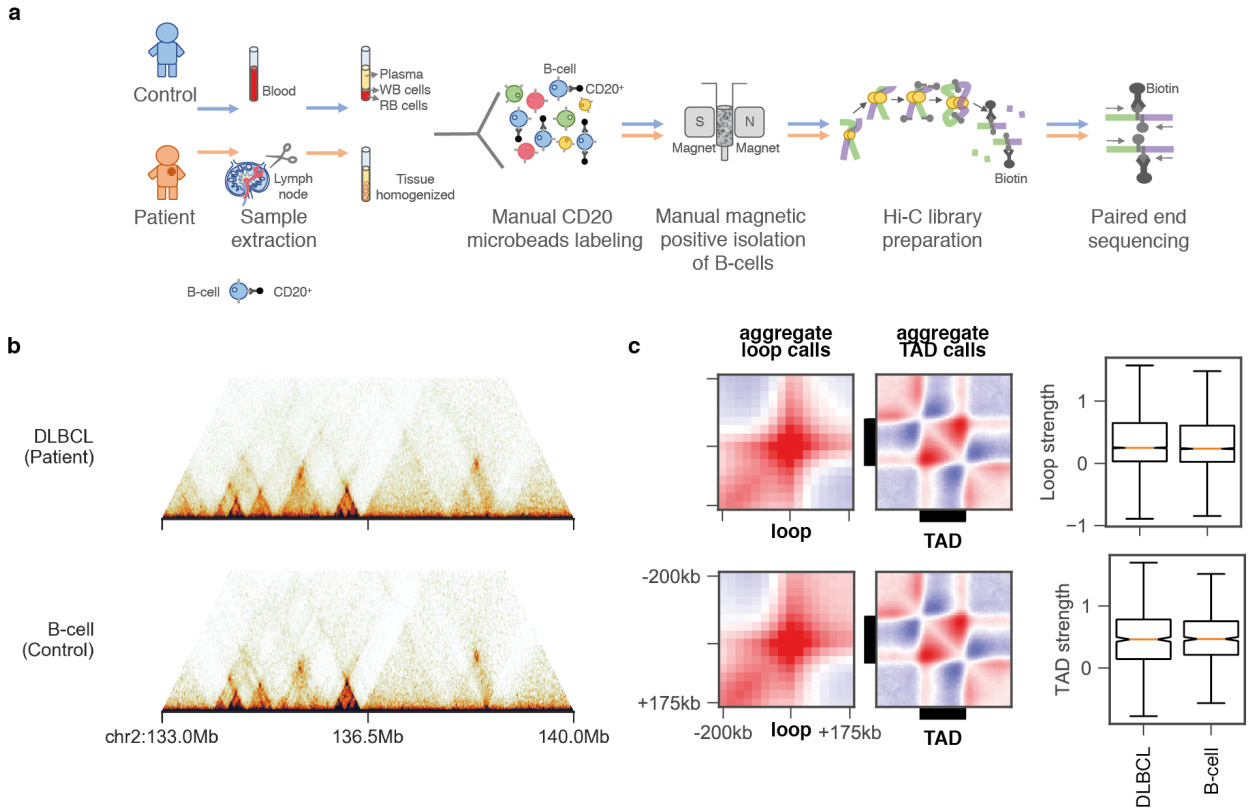


Figure 3. Compartments, TADs and loops can be detected and are highly reproducible in Low-C data. (a) Log-log "distance decay" plot for chromosome 1 showing the decrease in contact probability between 50kb bins with increasing distance for the 1M and 1k, as well as the Dixon et al. (2012) and Du et al. (2017) samples. (b) AB compartment comparison for chromosome 1 binned at 1Mb. Contact correlation matrices for the 1M and 1k samples (top) and the corresponding first eigenvector (coloured according to the sign of the eigenvector entries) for the samples listed in (a) are shown on the left. Bottom right shows a scatter plot of first eigenvector values of the 1M vs. the 1k sample with Pearson correlation coefficient shown in the plot. Red line indicates identity. (c) Insulation score (Kruse et al. 2016) comparison for the region on chromosome 13 shown in Fig. 1b. Heatmaps display insulation score values for a range of window sizes, line plots highlight the insulation index for a window size of 250kb (see panel a for line colours). Next to it is a scatter plot with the Pearson correlation coefficient of the complete insulation index vectors for the 1k and 1M samples. Red line indicates identity.



865 **Figure 4**. Generation of Low-C maps for DLBCL primary tissue and healthy B-cells. (a) Schematic overview of DLBCL (patient) and B-cell (control) sample extraction using MACS sorting of CD20+ microbead-labelled cells extracted from primary tissue and Low-C library generation. (b) Sample region on chromosome 2 highlighting TADs, loops and loops domains in the Low-C maps for DLBCL and B-cells. (c) Aggregate
870 TAD and loops analysis (see Figure 1) using de novo loop and TAD boundary calls (Methods). On the right is a comparison of loop and TAD strengths between patient and control.

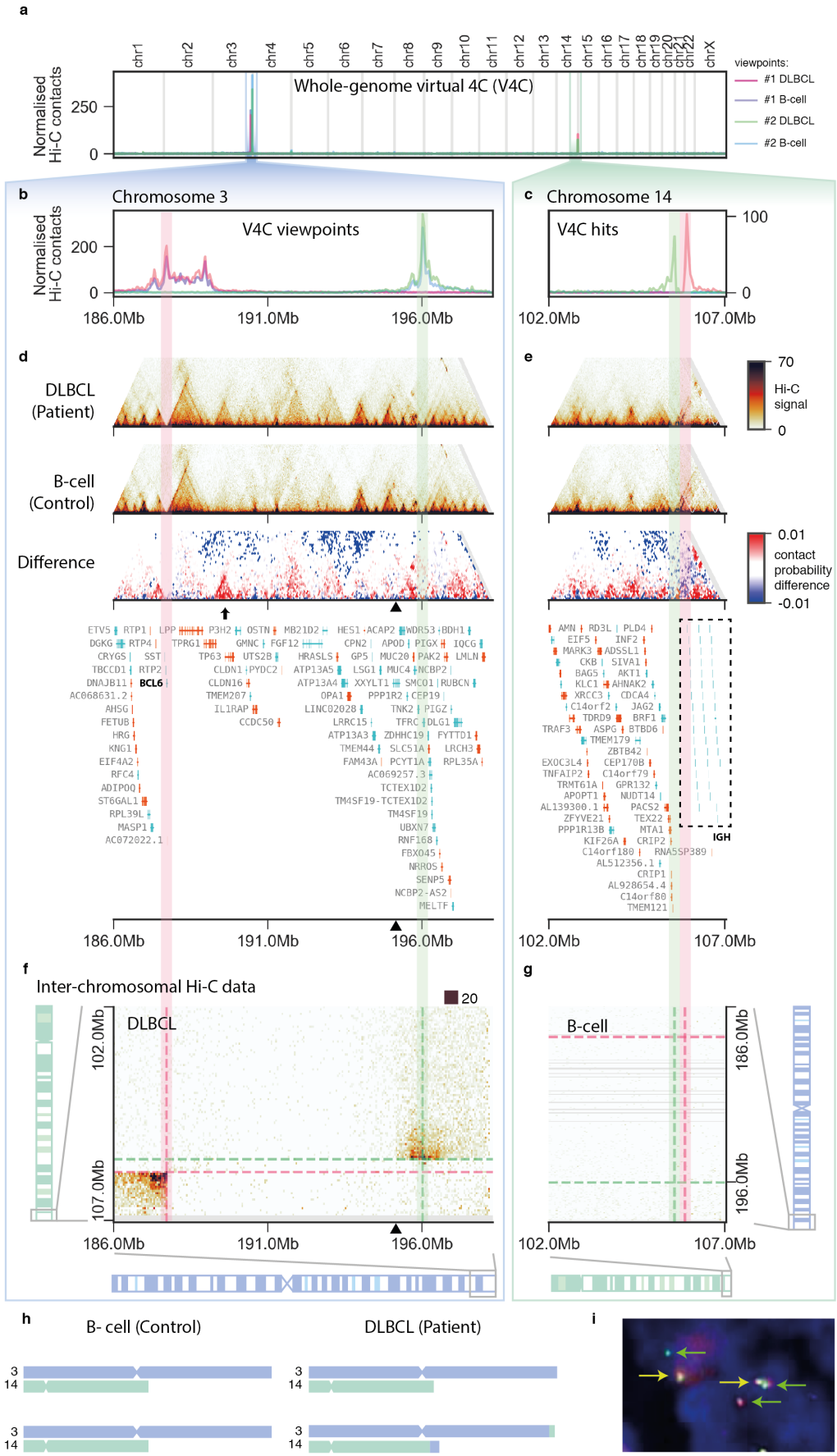


Figure 5. Unbiased detection and characterisation of t (3q27:14q32) reciprocal translocation in the DLBCL sample. (a) Whole-genome virtual 4C plot for two

viewpoints on chromosome 3 (magenta, green). (b, c) Zoom-in of the virtual 4C plots to the viewpoint (b) and target (c) regions. (d, e) Local Hi-C maps of the viewpoint (d) and target (e) regions. (f, g) Inter-chromosomal Hi-C maps of target vs. viewpoint region in the patient (f) and viewpoint vs target region in control (g), highlighting the
880 telomeric ends of the corresponding chromosomes. (d, f) Black triangles at the x axis indicate most likely breakpoint region as determined visually from the inter-chromosomal Hi-C maps in f. (h) Schematic representation of the reciprocal translocation as interpreted from the Hi-C data. (i) Confirmation of BCL6 translocation by FISH using a BCL6 Dual Color, Break Apart Rearrangement Probe hybridized to
885 the patient‘s tumour sample. A translocation of the *BCL6* gene is indicated by separated red and green signals (green arrows). In addition, two non-rearranged red/green fusion signals are seen (yellow arrows).

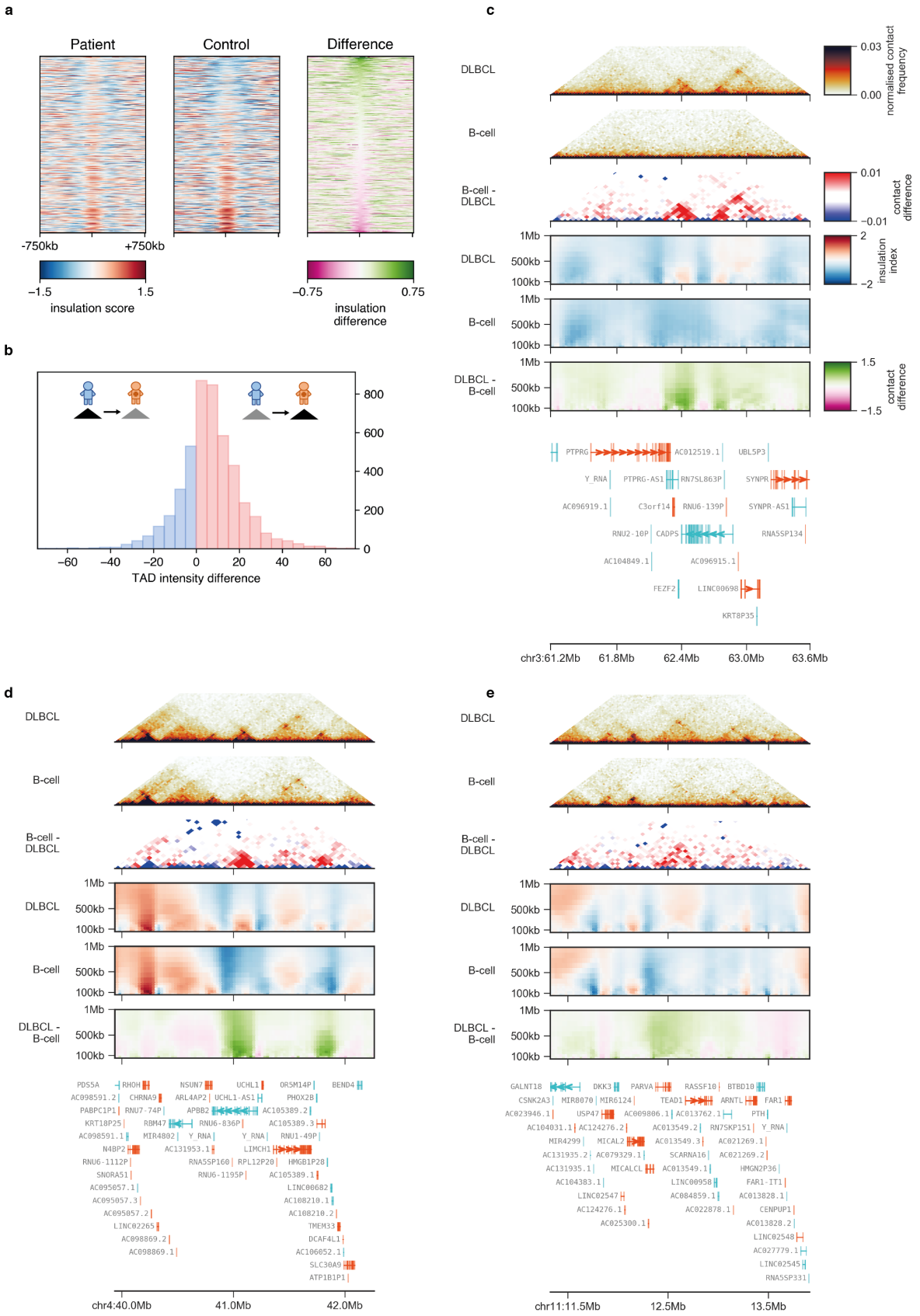
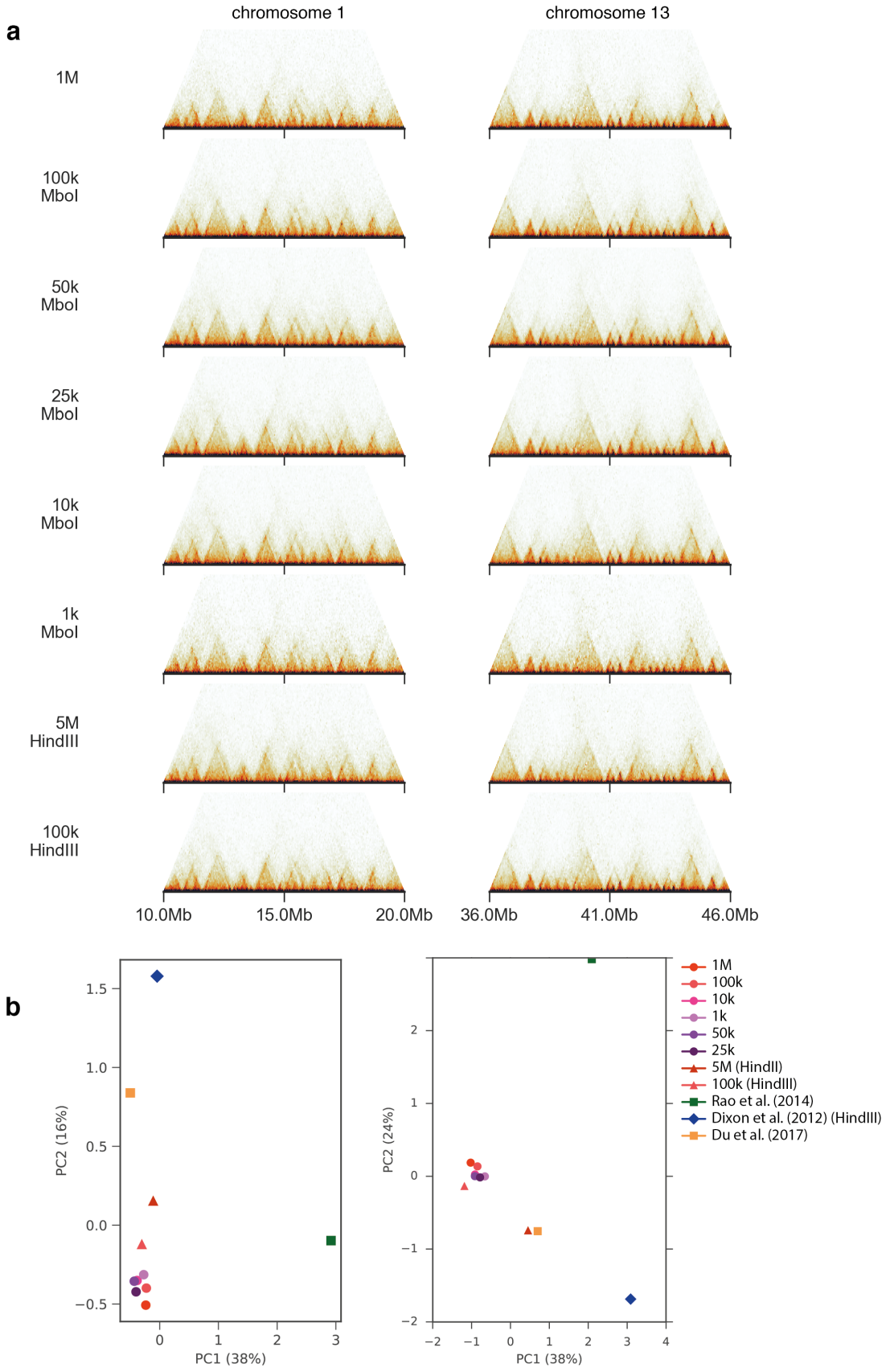


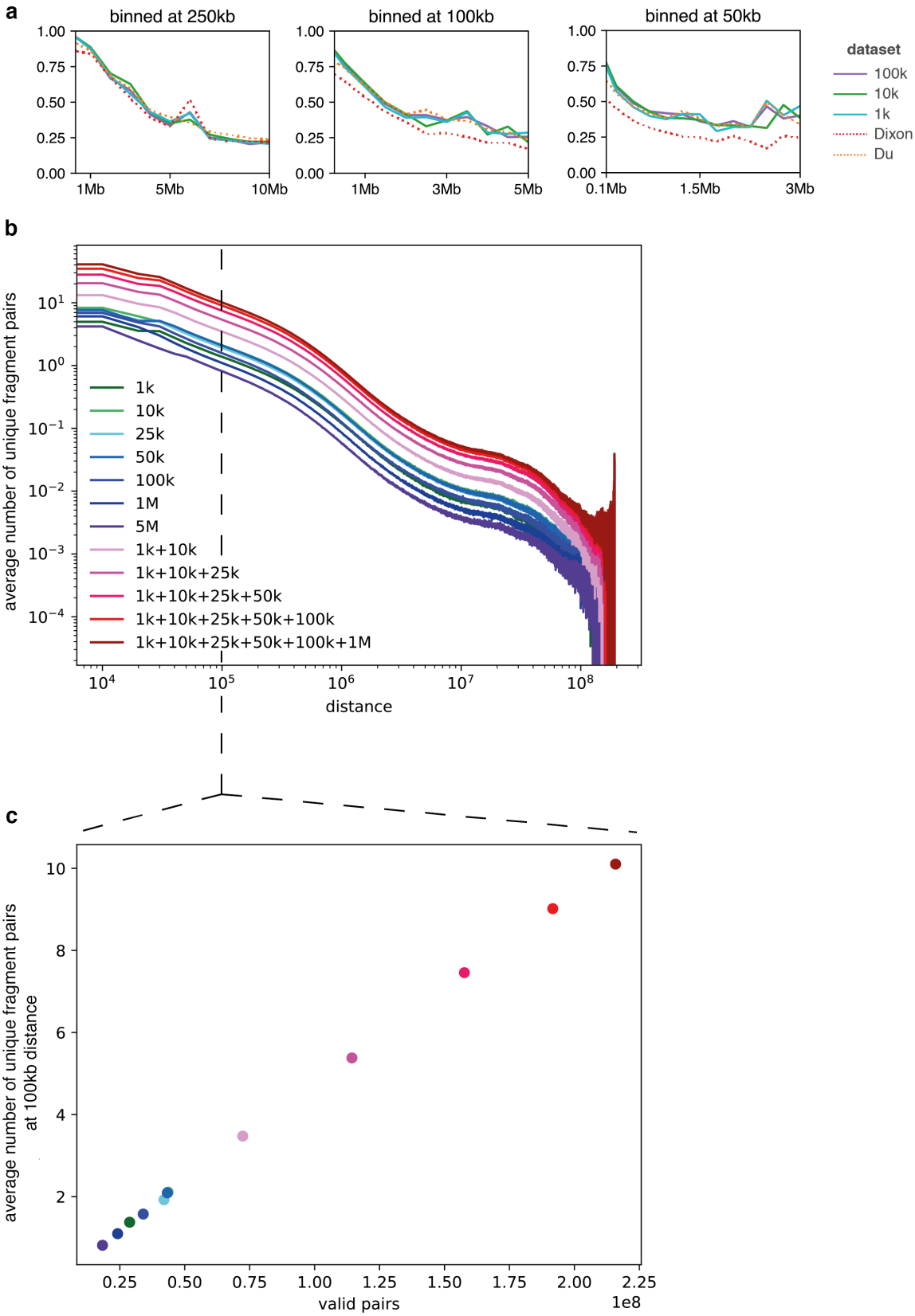
Figure 6. Extensive rewiring of chromatin organisation in DLBCL cells. (a) Insulation score changes between DLBCL (patient) and B-cell (control) for all regions in-between

two consecutive TAD boundaries. (b) Histogram of the differences in average Hi-C signal between two consecutive TAD boundaries from patient to control. (c, d, e) Examples of de novo TADs emerging in DLBCL showing (from top to bottom: Local DLBCL Hi-C, local B-cell Hi-C, difference in local Hi-C (DLBCL – B-cell), insulation
895 scores for different window sizes in DLBCL, insulation score in B-cell, difference in insulation score (DLBCL – B-cell), genes in the region (GENCODE).



Supplementary Figure 1. Low-C results are highly similar for different input cell numbers. (a) Visual comparison of Hi-C matrices at 50kb resolution from all Low-C

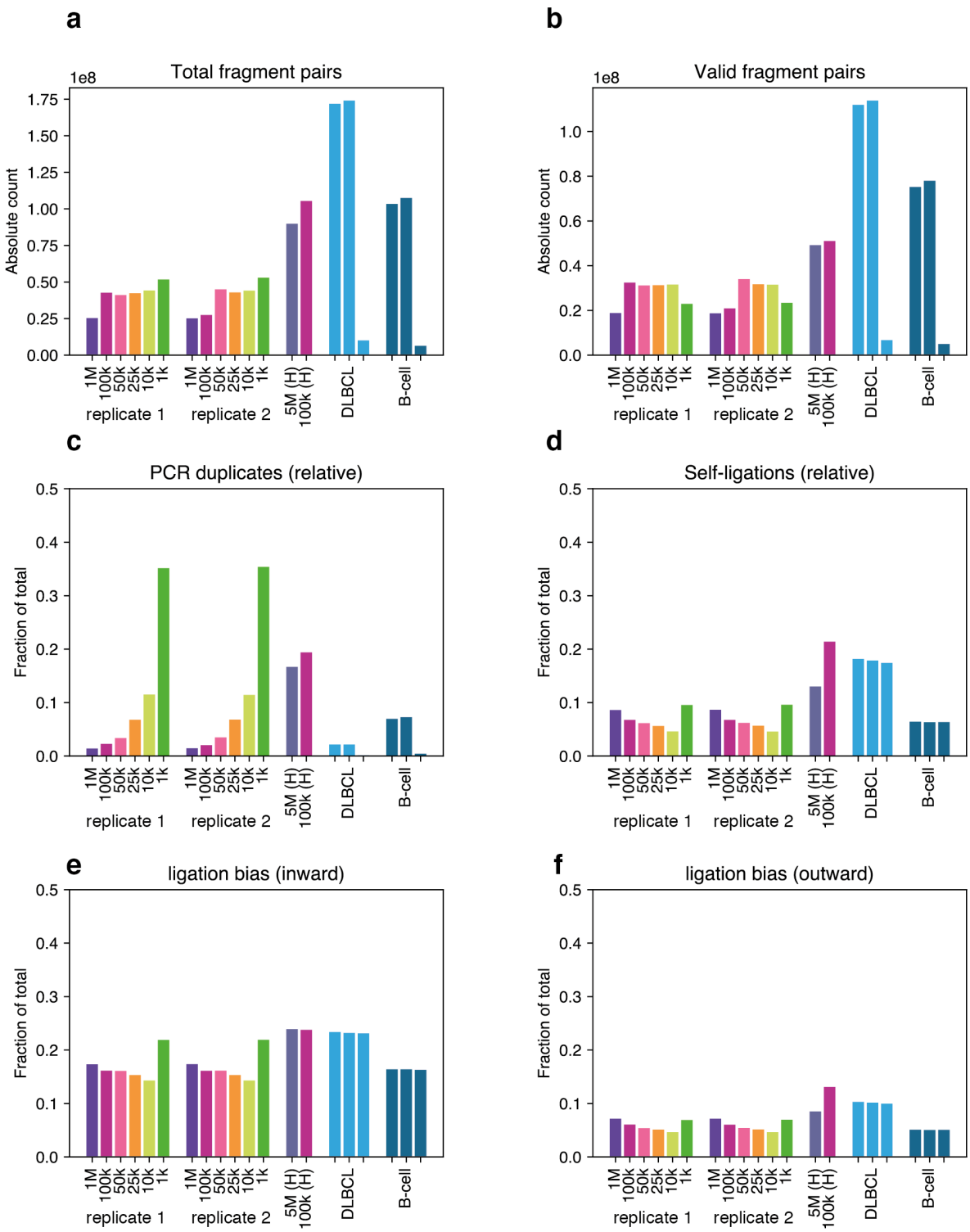
900 libraries in this study. Two different genomic regions are shown as examples. Input cell numbers are reported to the left of each matrix. (b) PCA on the top 50,000 most variable entries in the 100kb resolution Hi-C matrices of contacts between 200kb and 2Mb (left) and 50kb and 1Mb (right) for all Low-C samples. The plot also includes the mESC Hi-C datasets published by Dixon et al. (2012) (Dixon et al. 2012), Du et al
905 (2017) (Z. Du et al. 2017a), and the mouse CH12-LX dataset from Rao et al. (2014) (Rao et al. 2014). Circles denote MboI, triangles denote HindIII in the Low-C samples. Dixon et al. used HindIII, Rao et al. and Du et al. used MboI.



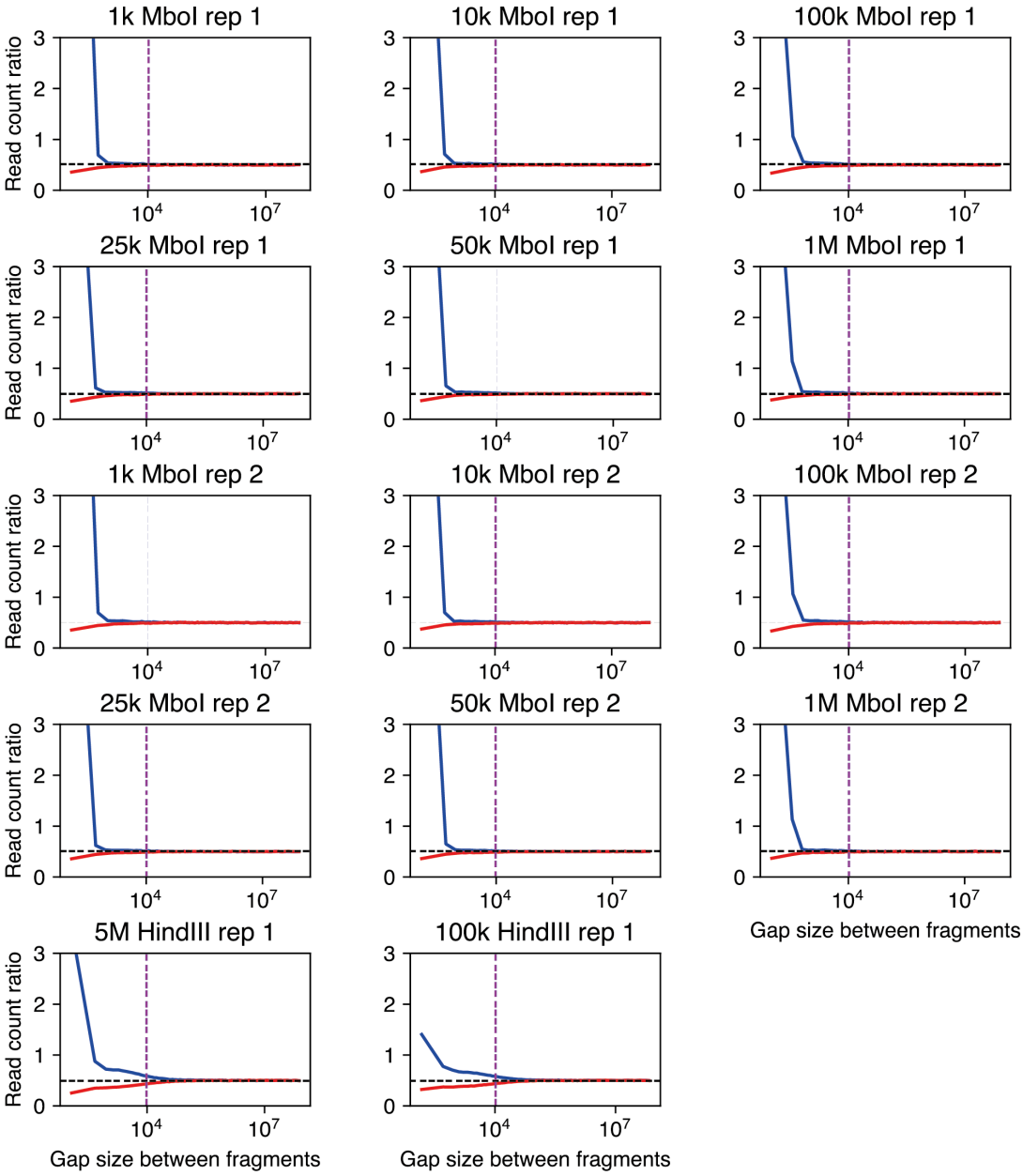
Supplementary Figure 2. Low-C library complexity is not affected by input cell number. (a) Same plot as Fig. 2a, but all datasets have been down sampled to the number of valid pairs in the 1M Low-C library for comparison. (b) Line plot of the average number of unique fragment pairs observed at increasing fragment distances

for all Low-C libraries and, additionally, merged Low-C libraries. (c) Average number of unique fragment pairs at 100kb distance as a function of the number of valid fragment pairs in each library.

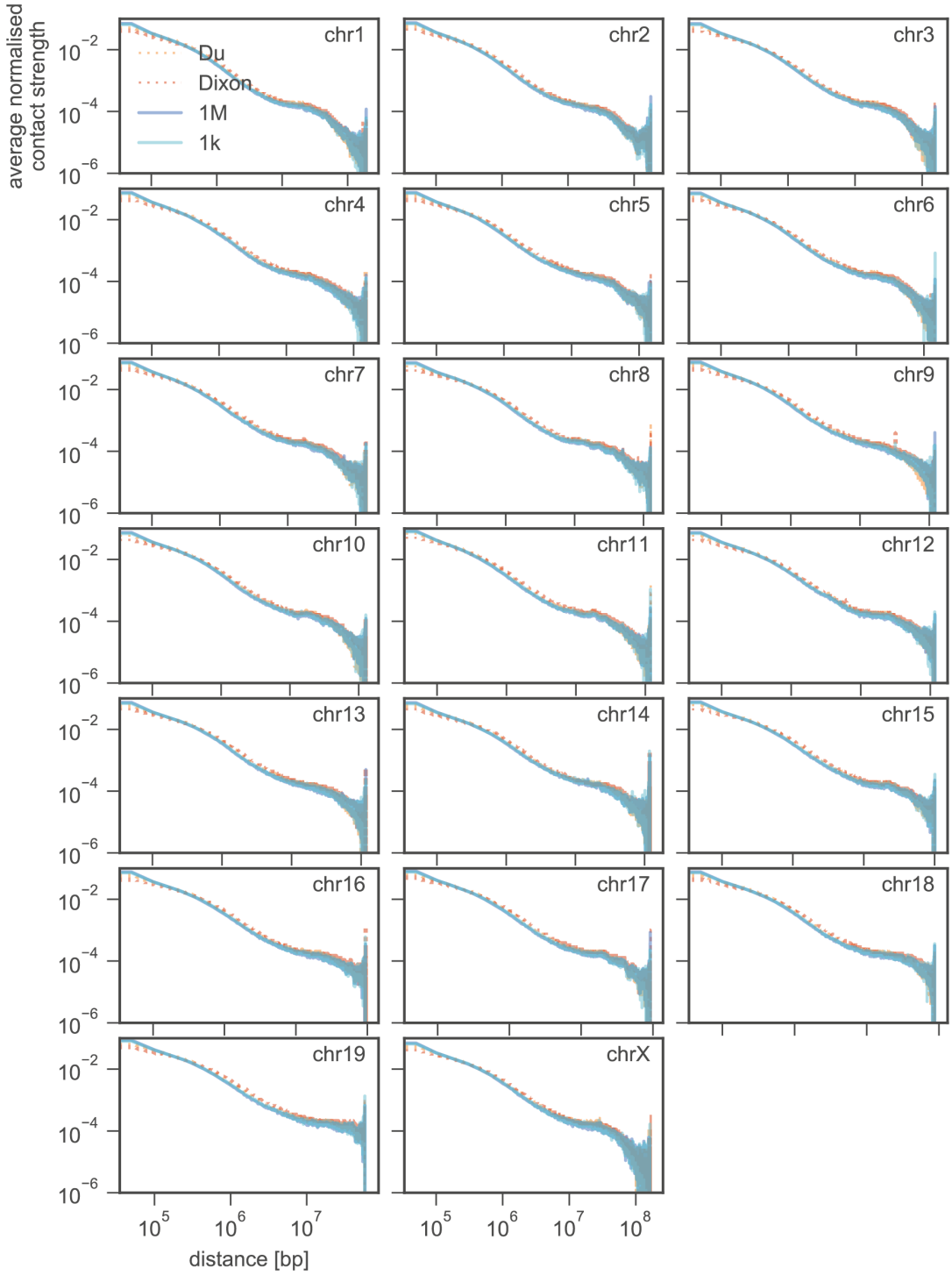
915



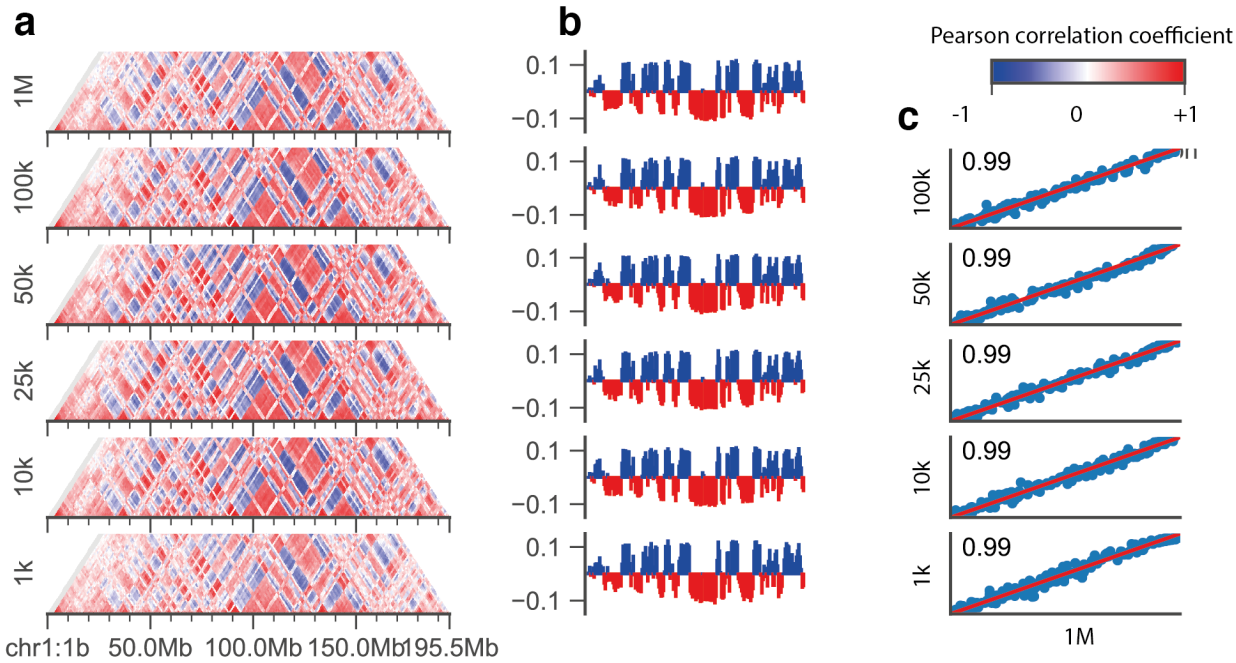
Supplementary Figure 3. Low-C library statistics**.** (a-b) Low-C fragment pair statistics (a) Total number of fragment pairs in each library. (b) Total number of valid fragment pairs after filtering for biases. (c-f) Different types of biases (see Methods) for each Low-C library expressed in fraction of total pairs. (c) PCR duplicates. (d) Self-ligated fragments. (e) Inward ligation error. (f) Outward ligation error. H=HindIII.



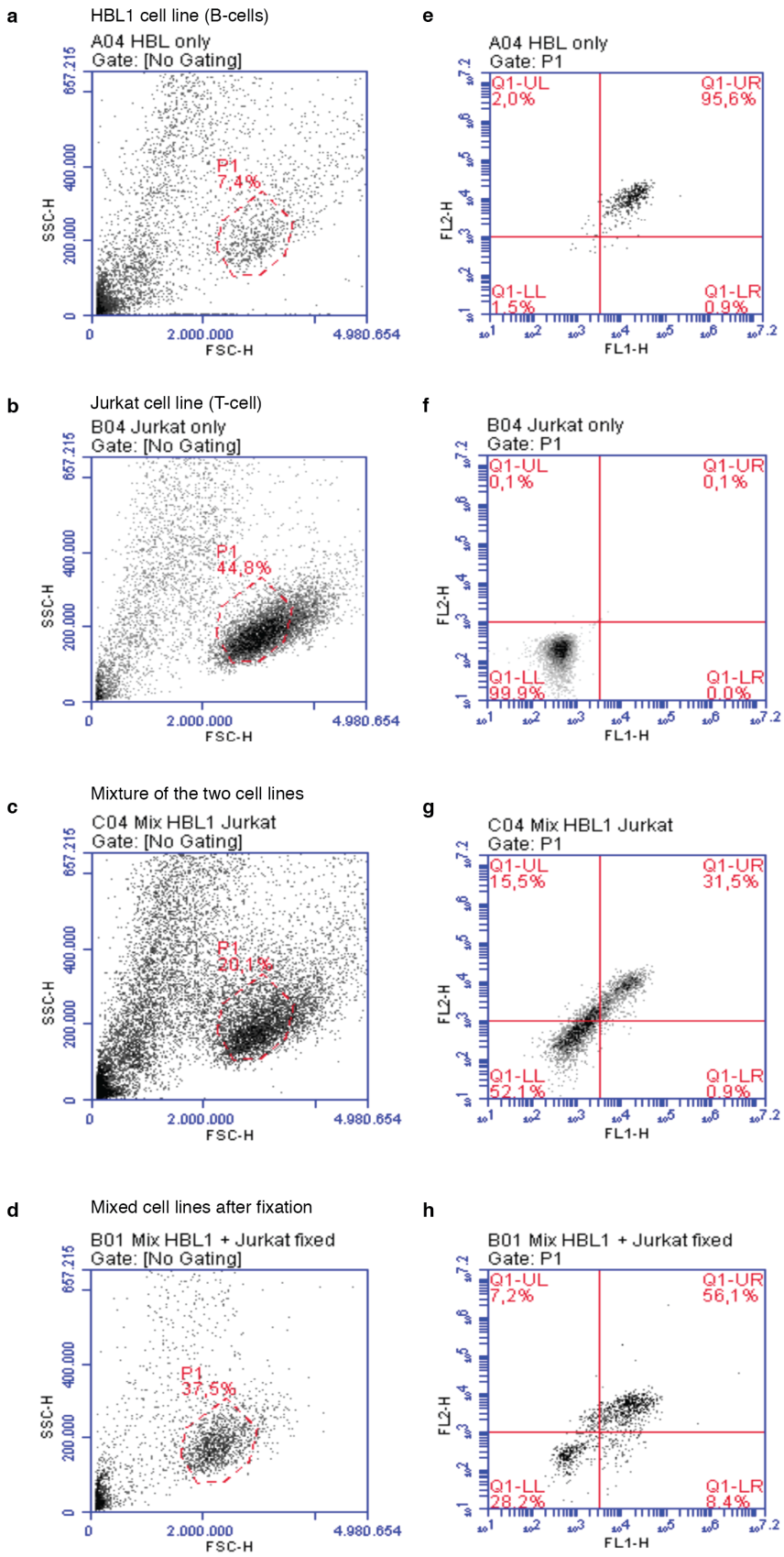
Supplementary Figure 4. Ligation error plots of Low-C samples using differing amounts of input material. Shown are ratios of paired reads facing towards each other (red; “inward”-facing, i.e. the first read is on the + strand and the second on the – strand) or away from each other (blue; “outward”-facing, i.e. the first mapped read is on the - strand, the second mapped read is on the + strand) to reads facing in the same direction (-/- or +/+) (Cournac et al. 2012; Jin et al. 2013) Dotted lines represent cutoffs for filtering read pairs at distances <10kb.



930 **Supplementary Figure 5.** Contact probability decay with distance (log-log plots) for each chromosome in the 1k and 1M samples, as well as Dixon *et al*. (2012) (Dixon et al. 2012) and Du *et al.* (2017) (Z. Du et al. 2017a) mESC Hi-C maps.



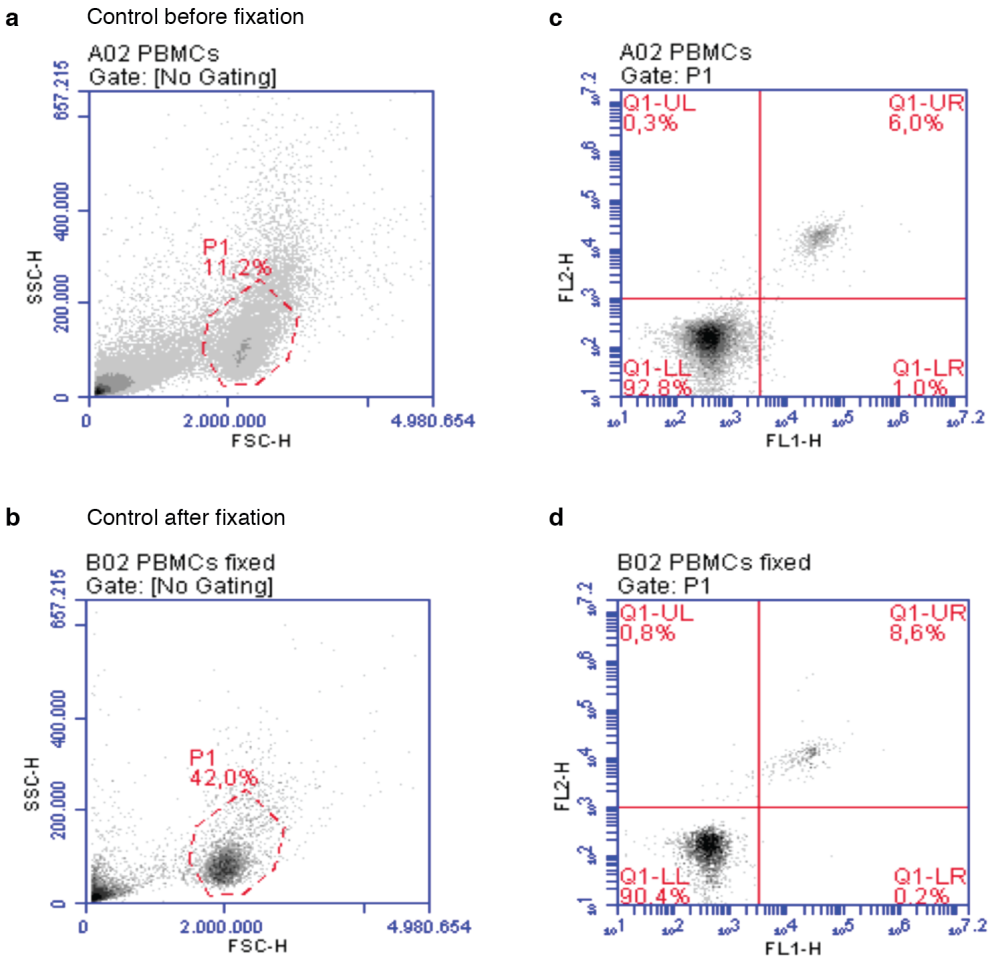
Supplementary Figure 6. AB compartment reproducibility for each Low-C MboI sample. (a) Correlation matrices for chromosome 1. (b) First eigenvector of the matrices in (a), positive values are coloured blue, negative values red. (c) Scatter plots of the eigenvector values in (b) against the eigenvector of the 1M sample with Pearson correlation coefficient R indicated in the top left corner. Red line indicates identity.



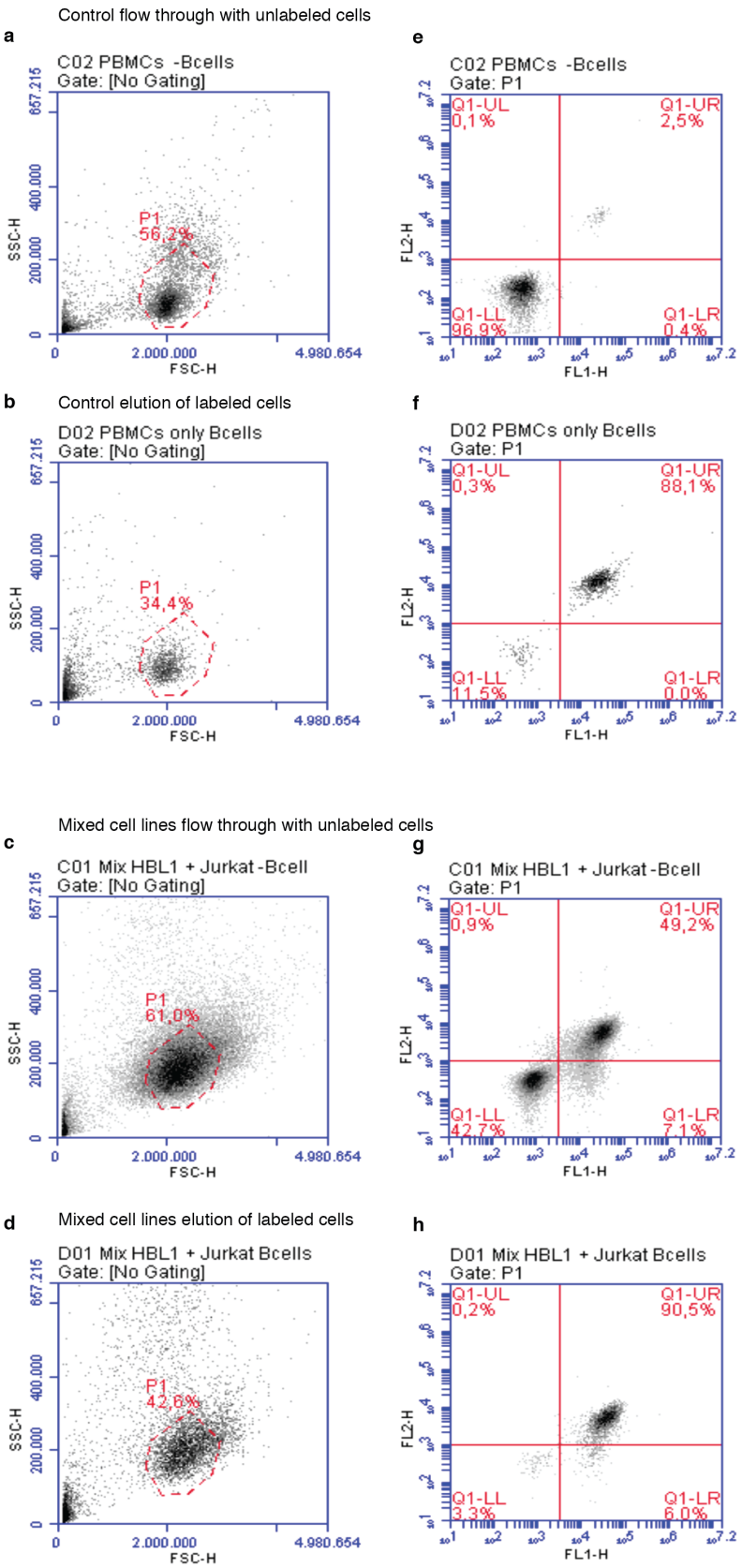
940

Supplementary Figure 7. Flow cytometry analysis for HBL1 and Jurkat cells. **(**a-d) Scatter density plots for a HBL1 (B-cell) cell line (a), Jurkat (T-cell) cell line (b), and a

mixed HBL1-Jurkat cell population before (c) and after (d) formaldehyde fixation. Discontinuous red circles demarcate the cell population of interest. (e-h) Scatter plots for double-staining with CD20 (channel one; FL1-H) and CD19 (channel two; FL2-H) antibodies for HBL1 (e), Jurkat (f) and a mixed HBL1-Jurkat cell populations before (g) and after (h) formaldehyde fixation. Red lines demarcate the quadrants for positive/negative cell labelling.

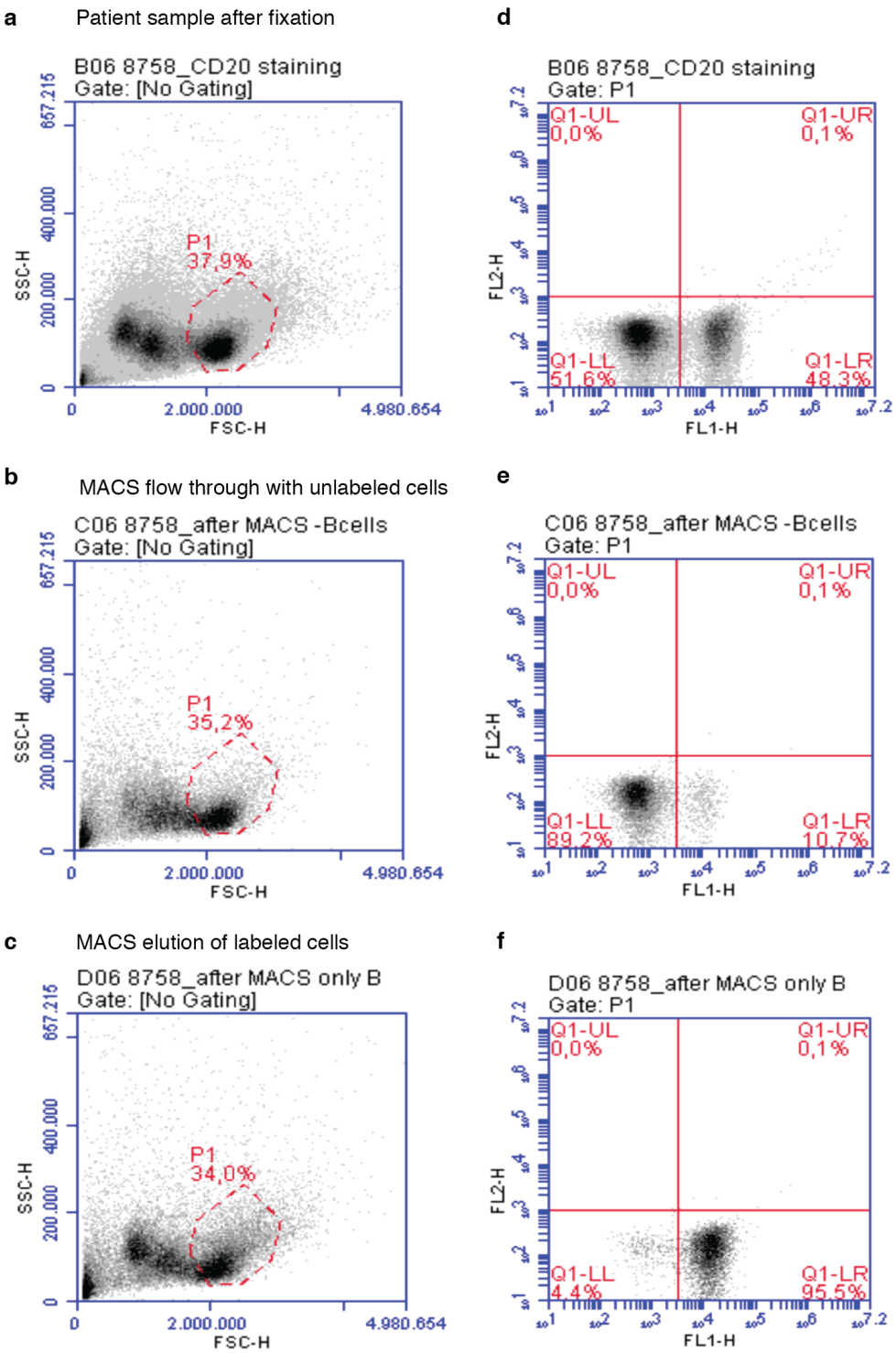


Supplementary Figure 8. Flow cytometry analysis of PBMCs. (a, b) Scatter density plots for the PBMC population before (a) and after (b) cell fixation with formaldehyde. Among PBMCs, lymphocytes are shown enclosed on a red discontinuous circle. (c, d) Scatter plots for double-staining of PBMCs with CD20 (channel one; FL1-H) and CD19 (channel two; FL2-H) antibodies, before (c) and after (d) formaldehyde fixation. Red lines demarcate the quadrants for positive/negative cell labelling.



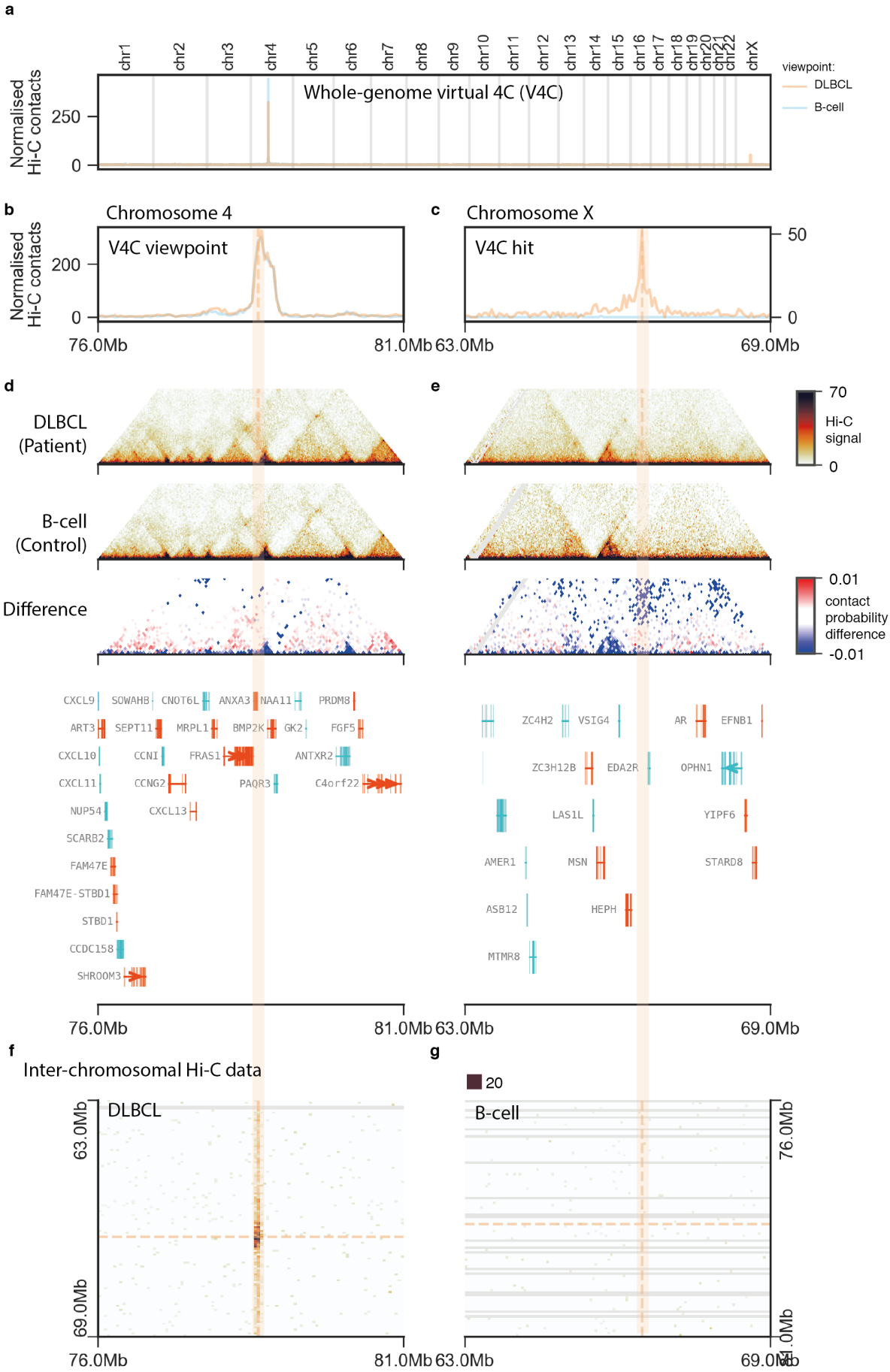
Supplementary Figure 9. Flow cytometry analysis on the MACS sorted populations from a healthy donor's PBMCs and a mixture of B- and T-cell population. (a-d)

Scatter density plots for the healthy donor PBMCs without B-cells (a, MACS unlabelled CDC20- flow through), the healthy donor B-cells (b, MACS elution of labelled CD20+ cells), Jurkat (T-cells) from a mixed HBL1-Jurkat mixed cell population (c, MACS unlabelled CDC20- flow through), and HBL1 (B-cells) from the same HBL1-Jurkat mixed population (d, MACS elution of labelled CD20+ cells). Discontinuous red circles demarcate the cell population of interest. (e-h) Scatter-plots for double-staining of each cell population with CD20 (channel one; FL1-H) and CD19 (channel two; FL2-H) antibodies. Scatter plot for the healthy donor PBMCs without B-cells (e, MACS unlabelled CDC20- flow through), the healthy donor B-cells (f, MACS elution of labelled CD20+ cells), Jurkat (T-cells) from a mixed HBL1-Jurkat mixed cell population (g, MACS unlabelled CDC20- flow through), and HBL1 (B-cells) from the same HBL1-Jurkat mixed population (h, MACS elution of labelled CD20+ cells). Red lines demarcate the quadrants for positive/negative cell labelling.



Supplementary Figure 10. Flow cytometry analysis of the patient sample after fixation with formaldehyde and MACS sorting. (a-c) Scatter density plots for the patient lymphocytes after formaldehyde fixation (a), patient MACS sorted
985 lymphocytes without B-cells (b, MACS unlabelled CDC20- flow through), and patient B-cells (c, MACS elution of labelled CD20+ cells). Lymphocytes are shown enclosed on a red discontinuous circle. (d-f) Scatter-plots for double-staining of each cell

population with CD20 (channel one; FL1-H) and CD19 (channel two; FL2-H) antibodies, for the patient lymphocytes after formaldehyde fixation (d), patient MACS
990 sorted lymphocytes without B-cells (e, MACS unlabelled CDC20- flow through), and patient B-cells (f, MACS elution of labelled CD20+ cells). Red lines demarcate the quadrants for positive/negative cell labelling.



Supplementary Figure 11. Unbiased detection and characterisation of structural rearrangement at ANXA3 in the DLBCL sample. (a) Whole-genome virtual 4C plot for

995 two viewpoints on chromosome 4 (orange, control in blue). (b, c) Zoom-in of the virtual 4C plots to the viewpoint (b) and target (c) regions. (d, e) Local Hi-C maps of the viewpoint (d) and target (e) regions. (f, g) Inter-chromosomal Hi-C maps of target vs. viewpoint region in the patient (f) and viewpoint vs. target region in control (g).

1000

Supplementary Table 1. Step-by-step Low-C experimental protocol.

Table provided as a separate Excel file.

1005 **Supplementary Table 2**. Overview of Low-C samples in this study and their respective numbers of read pairs.

Sample	*Technical replicate*	*Restriction enzyme**	*Read pairs*	*Mapped read pairs*	*Valid read pairs (after filtering)*	*Total valid pairs*
1M	1	M	45373559	25251575	18798179	37499845
1M	2	M	45521764	25140413	18701666	
100k	1	M	82904353	42649971	32408508	53300749
100k	2	M	53221423	27414542	20892241	
50k	1	M	60840152	44930010	33976593	65130281
50k	2	M	66595079	41112150	31153688	
25k	1	M	62322544	42866712	31680705	62953595
25k	2	M	63250992	42279423	31272890	
10k	1	M	65602065	44080826	31497073	63027306
10k	2	M	65489378	44080826	31497073	
1k	1	M	86616647	51722407	22937947	46290053
1k	2	M	89231005	52922902	23352106	
5M	1	H	123453028	89807393	49189133	49189133
100k	1	H	140373791	105378368	51030522	51030522
DLBCL	1	M	250881396	171870902	111898247	
DLBCL	2	M	252231324	174080064	113844746	232446650
DLBCL	3	M	14289185	10030808	6703657	
B-Cell	1	M	150626741	103327894	75175780	
B-Cell	2	M	155140165	107468078	77957359	158093470
B-Cell	3	M	9001104	6340367	4960331	

**M = MboI, H = HindIII*

Supplementary Table 3. Numbers of valid read pairs in *cis*/*trans* as a measure of quality control.

Sample	*Restriction enzyme*	*Cis*	*Trans*	*Fraction cis*	*Cis/trans*	*Fraction cis (corrected for genome size)*
5M	HindIII	42641237	6500291	0.868	6.560	0.991
100k	HindIII	43690020	7297432	0.857	5.987	0.991
1M	MboI	30513663	6949072	0.815	4.391	0.987
100k	MboI	43711297	9549094	0.821	4.578	0.988
50k	MboI	55375829	9698299	0.851	5.710	0.990
25k	MboI	52464537	10439922	0.834	5.025	0.989
10k	MboI	54891458	8080917	0.872	6.793	0.992
1k	MboI	39768358	6476292	0.860	6.141	0.991
DLBCL	MboI	142651132	89401751	0.615	1.596	0.966
B-cell	MboI	98490584	59345608	0.624	1.660	0.967

Supplementary Table 4. List of genomic regions and genes ranked by magnitude of conformational differences between the DLBCL and B-cell Hi-C maps.

Table provided as a separate Excel file.