8714–8725 Nucleic Acids Research, 2016, Vol. 44, No. 18 Published online 20 July 2016
doi: 10.1093/nar/gkw568

Identification of multi-loci hubs from 4C-seq demonstrates the functional importance of simultaneous interactions

Tingting Jiang[1], Ramya Raviram[2,3], Valentina Snetkova[2], Pedro P. Rocha[2], Charlotte Proudhon[2], Sana Badri[2,3], Richard Bonneau[3,4,5], Jane A. Skok[2,*,†] and Yuval Kluger[1,6,7,*,†]

[1]Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06511, USA, [2]Department of Pathology, New York University School of Medicine, New York, NY, USA, [3]Department of Biology, New York University, New York, NY, USA, [4]Department of Computer Science, Courant Institute of Mathematical Sciences, New York, NY, USA, [5]Simons Center for Data Analysis, New York, NY 10010, USA, [6]Department of Pathology and Yale Cancer Center, Yale University School of Medicine, New Haven, CT, USA and [7]Program of Applied Mathematics, Yale university, New Haven, CT, USA

Received October 06, 2015; Revised June 09, 2016; Accepted June 13, 2016

ABSTRACT

Use of low resolution single cell DNA FISH and population based high resolution chromosome conformation capture techniques have highlighted the importance of pairwise chromatin interactions in gene regulation. However, it is unlikely that associations involving regulatory elements act in isolation of other interacting partners that also influence their impact. Indeed, the influence of multi-loci interactions remains something of an enigma as beyond low-resolution DNA FISH we do not have the appropriate tools to analyze these. Here we present a method that uses standard 4C-seq data to identify multi-loci interactions from the same cell. We demonstrate the feasibility of our method using 4C-seq data sets that identify known pairwise and novel tri-loci interactions involving the *Tcrb* and *Igk* antigen receptor enhancers. We further show that the three *Igk* enhancers, MiEκ, 3′Eκ and Edκ, interact simultaneously in this super-enhancer cluster, which add to our previous findings showing that loss of one element decreases interactions between all three elements as well as reducing their transcriptional output. These findings underscore the functional importance of simultaneous interactions and provide new insight into the relationship between enhancer elements. Our method opens the door for studying multi-loci interactions and their impact on gene regulation in other biological settings.

INTRODUCTION

In recent years the influence of chromosomal interactions in gene regulation has become increasingly apparent (1–6). Insights into the importance of these associations were initially obtained from single cell DNA FISH studies and more recently by molecular analysis using chromosome conformation capture (3C) techniques (7–11). The original 3C assays examined interaction frequencies between two fixed points of interest, while a later iteration of this technique, 4C-seq enabled comparisons of interaction frequencies from a single viewpoint to all other sites across the genome. Hi-C took this a step further providing a tool to analyze all possible pairwise interactions in the nucleus (11). These molecular approaches highlighted the importance of enhancer—promoter interactions in controlling lineage and stage-specific gene expression patterns (12). Moreover these studies demonstrated the significance of chromosome looping in separating out regions that have a distinct chromatin status (13) and the functional relevance of this aspect of chromatin organization in gene regulation (14). Hi-C and 5C further identified the existence of higher-order structures known as topologically associated domains (TADs) that encompass megabase wide stretches of DNA that interact with each other at high frequency (15–18). By definition, DNA contacts occur most frequently within individual TADs, however interactions between TADs on the same chromosome also occur, albeit at a lower frequency, while

*To whom correspondence should be addressed. Tel: +1 212 263 0504; Fax: +1 212 263 8211; Email: jane.skok@med.nyu.edu
Correspondence may also be addressed to Yuval Kluger. Tel: +1 203 737 6262; Email: yuval.kluger@yale.edu
†These authors contributed equally to this manuscript as Last Authors.

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

loci located on different chromosomes interact even less often. These analyses have revolutionized our understanding of how structure function connections underpin gene regulation and have provided a platform for future experiments.

One poorly understood aspect of the role of chromatin associations in gene regulation concerns the influence of simultaneous interactions involving more than two partner regions. Although mathematical modeling approaches have attempted to reconstruct three-dimensional structures to address this question, only combined DNA and RNA FISH analyses have unambiguously been able to identify multi-loci interactions and to probe their functional importance (19). However, as mentioned above FISH is low resolution and it is not possible to look at interactions that occur at short length scales between regions of the genome separated by less than a few kb. Furthermore, FISH approaches are currently largely low-throughput.

Given the potential importance of the synergistic effects of multi-loci interactions on gene regulation it is important to design approaches for detecting and validating these. Ay *et al.* recently developed an innovative 3C-based technique to detect multi-loci interactions, however, this method has several drawbacks regarding low efficiency of data extraction and low resolution (20). In addition another study used a non-standard 4C-seq procedure to detect triplet interactions that occur with loci on *trans* chromosomes (21). Making use of standard 4C-Seq datasets that assess genome wide spatial contacts from a single viewpoint or bait sequence, we have now established that identification of multi-loci interactions in high resolution (in *cis*) is feasible. We tested our approach using several 4C-seq datasets generated from *ex-vivo* derived sorted developing murine B and T lymphocytes. Here, we demonstrate the robustness and reproducibility of our method using four different bait sequences on chromosome 6 associated with the T cell receptor beta (*Tcrb*) locus enhancer, Eβ and the immunoglobulin kappa (*Igk*) enhancers, MiEκ and 3′Eκ as well as from a region downstream of *Igk,* REIR (a potential regulatory region we recently identified (22) located close to the 3′ end of the *Igk* locus).

We focused our analyses on the antigen receptor loci that play an essential role in adaptive immunity. In total there are seven antigen receptor loci, four T cell receptor (*Tcr*) loci (*Tcrg, Tcrd, Tcrb* and *Tcra*) and three B cell specific immunoglobulin genes (*Igh, Igk* and *Igl*). Lineage and stage specific rearrangement of the individual loci generates T cell receptor and immunoglobulin diversity enabling specific recognition of foreign antigen, which is a fundamental feature of the adaptive immune response (23). Recombination is mediated by the RAG recombinase that binds to highly conserved recombination signal sequences (RSSs), which flank the variable (V), diversity (D) and joining (J) gene segments that are arrayed along the individual loci. The RAG complex binds to two gene segments (that can be many kilobases apart), brings them together and cuts at the RSS borders to generate DNA double-strand breaks (DSBs) that are subsequently joined via the non-homologous end joining pathway (24). Importantly, dynamic changes in locus conformation, namely 'locus contraction', occur on *Igh, Igk, Tcrb* and *Tcra/d* loci in cells, where they recombine (25). Locus contraction brings widely dispersed V gene segments that can be spread out over megabases of DNA, into contact with the proximal DJC domain through chromatin looping. Thus, locus contraction maximizes receptor diversity providing all the V genes with an equal opportunity to recombine (26–29).

In this study, we used the enhancer viewpoints to validate our method by identifying known enhancer interactions that occur within the *Tcrb* and *Igk* loci and flanking regions in recombining cells. In addition, using our pipeline we could identify interaction hubs involving antigen receptor gene segments that associate with these bait regions. Specifically, we show that changes in interaction hubs in recombining versus non-recombining cells reflect the alterations in locus conformation that result from locus contraction and rearrangement. Further, we demonstrate that although deletion of the *Tcrb* enhancer, Eβ has no impact on *Tcrb* V gene contraction it severely impacts interactions in the DJCβ1–DJCβ2 domain.

Another important finding that emerged from these studies is the demonstration that contacts within a super-enhancer occur simultaneously. The term 'super-enhancer' is used to describe a group of enhancers in close linear genomic proximity that have unusually high levels of transcriptional co-activator (Mediator) and H3K27Ac (30). Our previous pairwise analysis demonstrated that loss of one enhancer not only disrupts pairwise interactions between itself and the other enhancers but it also interferes with interactions between the other partner enhancers (22). Furthermore, this disruption is linked to a reduction in their transcriptional output (22). Simultaneous interactions between all three elements underscore the functional importance of enhancer hubs and provide insight into chromatin organization and its contribution to super-enhancer activity.

MATERIALS AND METHODS

4C library generation

4C-seq uses 3C (chromosome conformation capture) products as a template for inverse PCR to amplify the genomic regions interacting with a region of interest, the bait. The 3C template is produced by two successive rounds of digestion-ligation using two different 4bp cutters, NlaIII and DpnII as outlined in Figure 1A–G. The first step involves crosslinking the DNA with formaldehyde prior to the first round of digestion-ligation that joins DNA regions in close physical proximity. The second digestion-ligation event occurs following reversal of the crosslinks. This produces small circular fragments for efficient PCR amplification using primers specific for bait regions. For the analysis described here we used 4C-seq libraries with bait sequences associated with the *Tcrb* enhancer, Eβ, and *Igk* enhancers, MiEκ and 3′Eκ, as well as a region downstream of *Igk* (REIR) (Figure 2A). Illumina-specified adapters for Illumina sequencing were included at the 5′ end of each primer. Our 4C libraries were sequenced on the Illumina Hi-Seq 2500 using single-read 100-cycle runs.

At least two replicates were processed for each sample. The number of reads obtained for each dataset is listed in Supplementary Table S1. The data can be found in GEO GSE80272 (22).

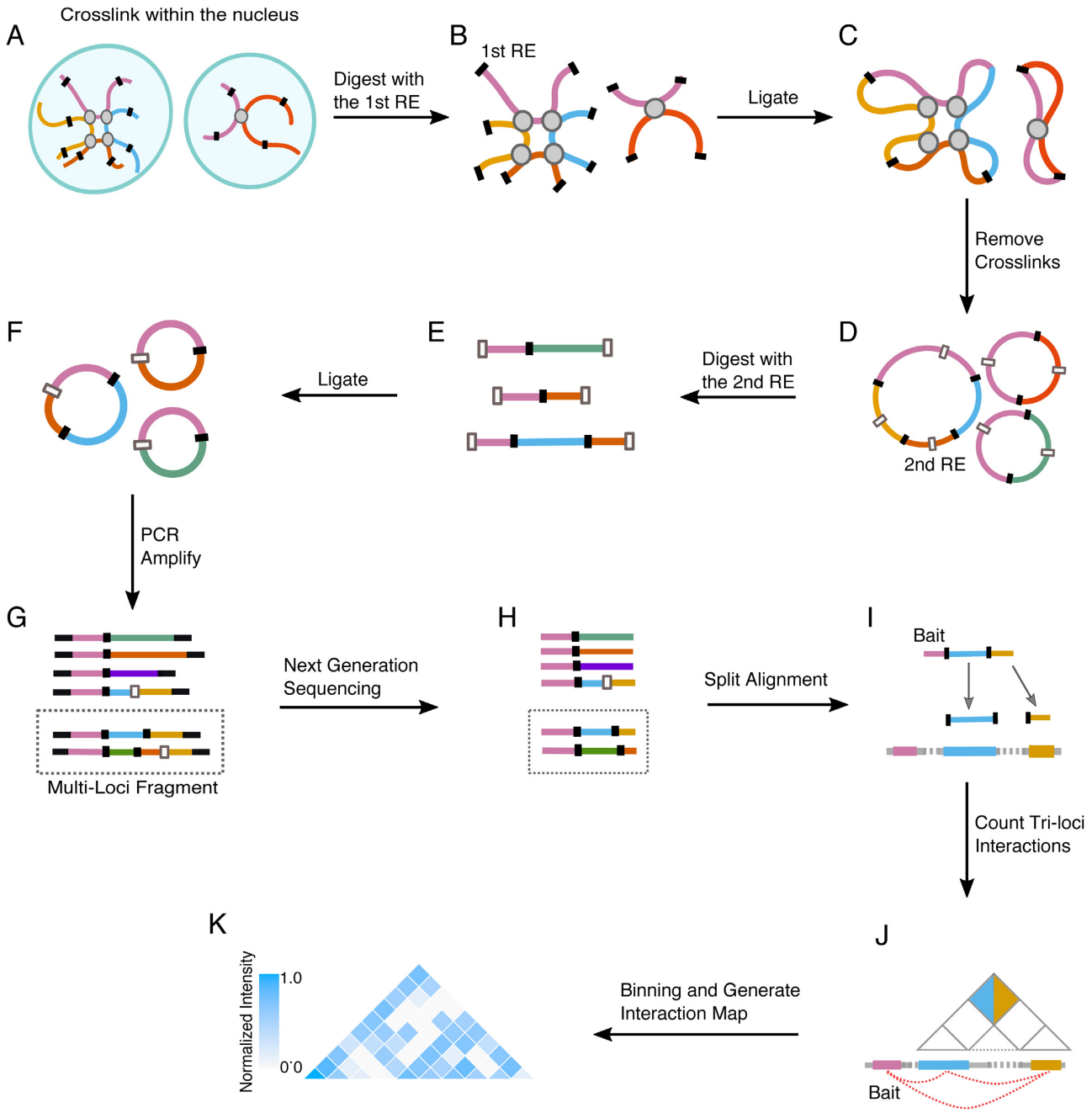

Figure 1. Workflow for extracting multi-loci chromatin interactions using 4C-seq data. (**A**) Cross-linking step. (**B**) Primary restriction enzyme digestion. (**C**) First ligation step. (**D**) Removal of crosslinks. (**E**) Secondary restriction enzyme digestion. (**F**) Second ligation step. (**G**) PCR amplification of fragments (tri-loci fragments are shown in the dotted box). (**H**) High throughput sequencing (tri-loci reads are shown in the dotted box). (**I**) Alignment of tri-loci reads. Each read is split into three segments at primary restriction enzyme cut sites that include the bait segment. These segments are aligned independently to the genome. (**J**) Tri-loci interactions along the genome (bottom) and their representation in a heatmap (colored squares indicate simultaneous interaction of the blue and orange genomic locations with the bait). (**K**) Heatmaps were generated by binning the genomic region of interest in a resolution dictated by the genomic length of the loci involved in the study.

Alignment of reads representing multi-loci interactions

For each single-end read, the bait sequence was removed from the 5′ end. The remaining part of each read was split as shown in Figure 1H. Prior to alignment we extended the first interacting segment with CATG sequences (the NlaIII restriction enzyme cut site) upstream and downstream of this string. Further, we extended the second interacting segment with one CATG sequence upstream of the string. Each string of reads with multiple fragments, was mapped individually to the mouse reference genome (mm10) using Bowtie (1.01) with parameters as follows:

-S -v 2 -k 1 -m 1 -I 0 –best –strata. Alignment of segments was sorted by read name using samtools (0.1.18). We further removed pairs of interacting segments that align to a con-

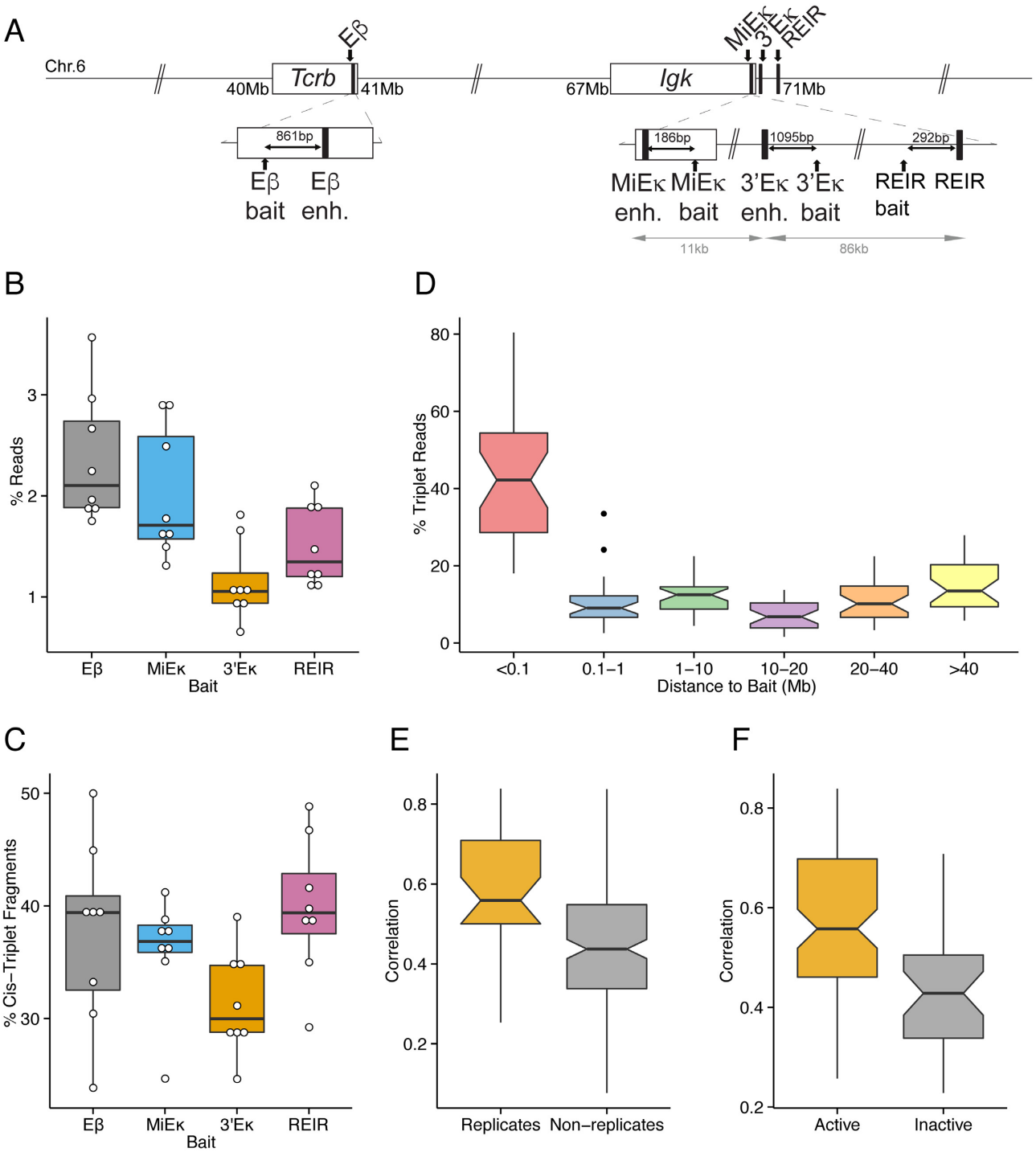

Figure 2. Frequency and genomic distance associated with multi-loci interactions. (**A**) Scheme showing the location of the four 4C-seq baits on chromosome 6. (**B**) Frequencies of tri-loci reads in 4C-seq experiments using baits Eβ, MiEκ, 3′Eκ, and REIR. (**C**) Frequencies of *cis* tri-loci fragments. (**D**) Distribution of the frequencies of multi-loci fragments at different distances from the bait region. (**E**) Correlation of pair-wise interactions between replicates and non-replicates. (**F**) Correlation of interactions using active versus inactive baits.

tiguous region in the genome. These pairs represent only one interacting locus that contains an undigested CATG site. The remaining reads with two segments aligned to two different genomic locations were termed 'tri-loci' fragments (Figure 1I). These interactions are represented as a heatmap (Figure 1K).

Reducing the effect of PCR artifacts in multi-loci interactions

To take account of PCR artifacts that can be generated during the amplification step, we created a histogram whose y-axis represents the number of different tri-loci interactions as a function of the number of identical PCR duplicates or real interacting loci shown in the x-axis (Supplementary Figure S1). To reduce the effect of PCR artifacts we capped the number of tri-loci fragments above the 60th percentile of this distribution making them equal to the number of tri-loci fragments at the 60th percentile.

The base of the heatmaps in Figures 1K, 3B-E, 4, 5B–E and 6A–F represent contiguous genomic regions. To observe interactions within these genomic locations we binned these regions accordingly so that each cell in the heatmap can be linked with two genomic locations on the x-axis (using two straight diagonal lines stemming from this cell). The intensity of the interaction between these genomic locations is represented by the color of the cell (Figure 1K). Specifically, the normalized intensity of interaction for each cell is given by log[1+#reads]/max(log[1+#reads]) where the denominator is associated with the cell with the highest intensity in the heatmap. Color-bars in Figures 1K, 3B–E, 4, 5B–E and 6A–F represent the range of the number of tri-loci events shown in each figure.

ATAC-seq. The assay was performed using biological duplicates as described previously (31) with several modifications. We amplified our libraries with KAPA HiFi polymerase and sequenced them with Illumina Hi-Seq 2000 using 50 cycles paired-end mode. Reads were aligned to mm10 genome with Bowtie2 (parameters: –no-discordant -p 12 –no-mixed -N 1 -X 2000). Potential PCR duplicates were removed from the reads with Picard-tools. Since we are not interested in nucleosome positioning, we did not separate reads based on fragment size. To identify regions of open chromatin all reads are typically used for analysis of the ATAC-Seq data. The data can be found in GEO (GSE80272) (22).

RNA-seq. Data was generated as per (22) and the can be found in GEO (GSE80272).

RESULTS

Workflow for extracting multi-loci chromatin interactions using 4C-seq data

4C-seq generated with NlaIII as a primary restriction enzyme cutter is likely to capture short fragments due to the high frequency of cutting. Thus using this 4 bp-cutter increases the probability that 3C products encompassing multiple interactions are identified. It is important to note that 4C reads which contain a secondary restriction enzyme and two or more additional loci are not necessarily indicative of multi-loci interactions since these constructs can arise from cross ligation of digested products derived from different cells. To improve detection of multi-loci events that originate from single cells, we only retain reads in which distinct loci are separated by sites associated with the first restriction enzyme. In other words, only fragments with two or more primary restriction enzyme sites are considered for identification of multi-loci interactions (Figure 1G and H). Using this strategy for alignment of multi-loci interactions, we were able to identify two or more loci that interact simultaneously with the bait region within a single cell.

The probability of obtaining multi-loci 4C fragments also depends on read length. Indeed, the longer the read length the greater the probability of finding fragments with at least two primary restriction enzyme sites that are not interrupted by a secondary restriction enzyme site. Since reads representing interactions between four or more loci are rare in experiments with 100 bp reads, we will restrict our discussion to tri-loci interactions.

Our method complements existing 4C-seq pipelines, as it captures multi-loci interactions in addition to the standard pairwise interactions between the bait and other genomic locations. Furthermore, it allows us to differentiate between *bona fide* tri-loci interactions (bait–loci2–loci3) and mutually exclusive pairwise interactions (bait-loci2 or bait-loci3).

Frequency and genomic distance associated with multi-loci interactions

To detect multi-loci interactions we re-examined data generated in our lab (22) using four different baits (3′Eκ, MiEκ, REIR and Eβ) located on chromosome 6 (Figure 2A). For this we included *ex-vivo* derived cells from two lineages and four stages of mouse lymphocyte development: Double Negative (DN) and Double Positive (DP) T cells, Pre- (Pre-B) and Immature (Imm B) B cells. For the Eβ, 3′Eκ and MiEκ, baits were designed within regions adjacent to each enhancer but not overlapping with sequences eliminated in knockout cells. All experiments were performed using NlaIII as the primary and DpnII the secondary restriction enzyme. We followed the workflow described above for analysis of each library and compared the frequency of multi-loci interactions across the thirty-two experiments (Figure 2B). We observed that the frequency of tri-loci interactions is consistent between replicates for the same bait sequence and this is strongly dependent on the distribution of restriction sites around the bait (Figure 2B, Supplementary Table). To understand this better, we counted the number of potential multi-loci fragments within 0.1Mb from the Eβ or 3′Eκ bait. Interestingly, in experiments using the Eβ bait we found more potential tri-loci interactions indicating that the genomic location of a bait may contribute to the probability with which multi-loci fragment interactions can be detected. In line with this finding we identified the highest frequency of tri-loci interactions associated with the Eβ bait on the *cis* chromosome (Figure 2C).

We next, examined the genomic distance separating individual loci in tri-loci interactions (Figure 2D). In agreement with pairwise interactions detected in 4C-seq, we found enrichment of tri-loci interactions in regions located in close proximity to the bait. Specifically, 40% (median estimate) of

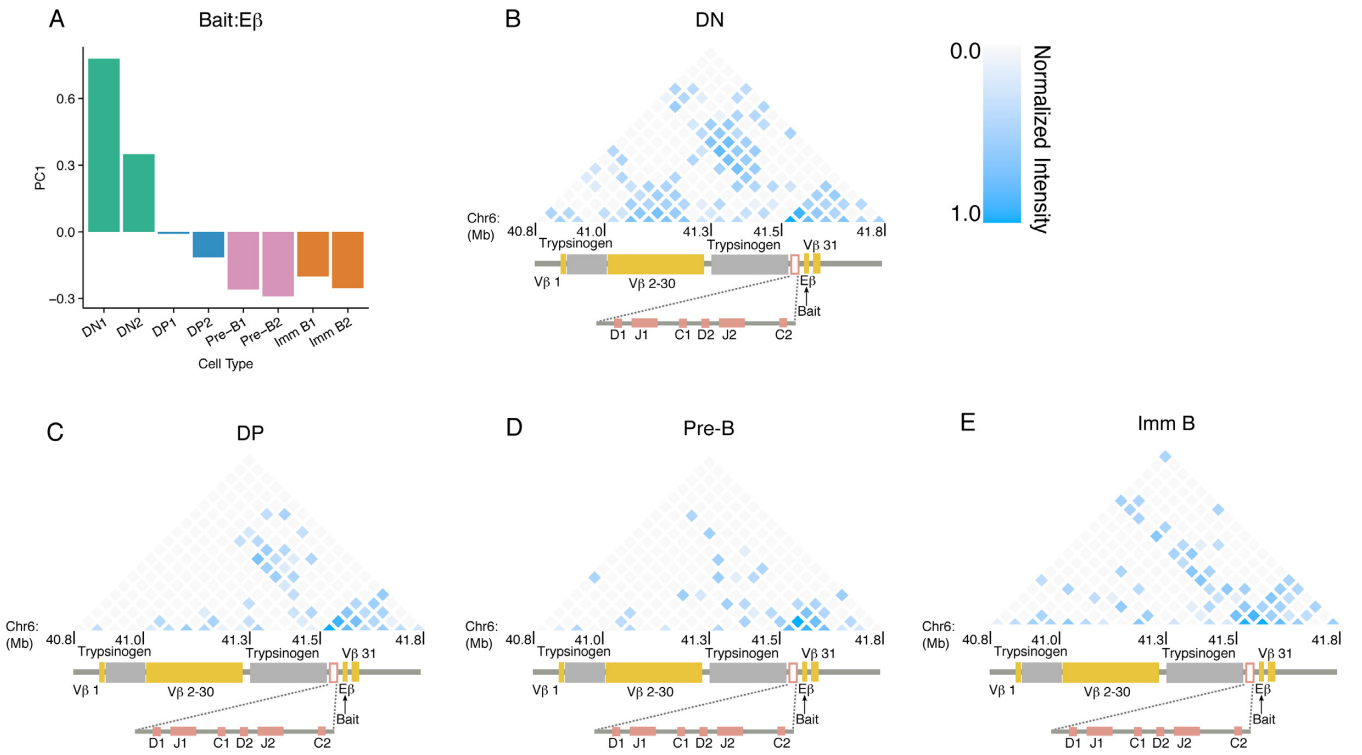

Figure 3. 4C-seq multi-loci interactions reflect known lineage and stage specific changes in *Tcrb* locus conformation. (**A**) The first principal component of pair-wise interactions in DN, DP, pre-B and Immature B cells using Eβ as bait. (**B–E**) Heatmap for tri-loci interactions (window size of 50kb) involving the Eβ bait in DN, DP, pre-B and Immature B cells across the *Tcrb* locus.

tri-loci events occur within 0.1Mb from the bait. This indicates that regions in close linear proximity are more suitable for revealing simultaneous multi-loci interactions with the bait.

As a consistency test, we generated an interaction matrix, denoted as M1, whose columns represent twenty–four experiments involving two replicate experiments that use the baits MiEκ, 3′Eκ and REIR in four different cells types (DN, DP, Immature B, and Pre-B) and whose rows represent the intensity of tri-loci interactions within windows sizes of 500 kb across the genomic co-ordinates 68M–72M, the region surrounding these three baits on chromosome 6. For the Eβ bait, we generated a similar matrix, denoted as M2, whose columns represent eight experiments involving two replicate experiments in DN, DP, pre-B and Immature B cells and performed the same analysis to examine tri-loci interactions within window sizes of 50 kb across the genomic co-ordinates 40.8M–41.8M, the region surrounding Eβ on chromosome 6. We inspected the tri-loci interaction matrices M1 and M2 and found that as expected correlations between replicates are significantly higher than correlations between any two experiments from the same bait but with different conditions (Figure 2E P-value $= 6 \times 10^{-4}$).

MiEκ, 3′Eκ, and REIR are active in B cells and Eβ and REIR are active in T cells. We expect these elements to interact with other loci in a similar manner across conditions if these elements are in their active state. We inspected the tri-loci interaction matrices M1 and M2 where the bait is in an active versus inactive state. The correlations between tri-loci interaction profiles of pairs of experiments associated with active baits are significantly higher than those of pairs of experiments associated with inactive baits P-value $= 6 \times 10^{-4}$) (Figure 2F).

Taken together these findings indicate that the frequency of tri-loci interactions identified is dependent on the frequency of the restriction enzyme sites and the length of the reads. In addition, tri-loci interactions are detected more frequently in regions surrounding the bait and interactions are more reproducible if the bait is in an active versus inactive state.

4C-seq multi-loci interactions reflect known lineage and stage specific changes in *Tcrb* locus conformation

As a second consistency test, we separately checked the tri-loci interactions in *Tcrb* from a bait region adjacent to its associated enhancer, Eβ (Figure 2A). For this, we used window sizes of 50kb across the genomic co-ordinates 40.8M–41.8M surrounding the *Tcrb* locus on chromosome 6. We applied principal component analysis (PCA) to the interaction matrix M2 (Supplementary Figure S2) and found that the leading principal component (45% of the total variation) separates DN cells from DP and B cell subsets (Figure 3A). To understand these interactions in more detail we pooled replicates and compared the normalized tri-loci interaction intensity in DN, DP, pre-B and Immature B cells (Figure 3B-E).

In DN cells where *Tcrb* undergoes recombination and the Eβ is in an active state we identified two local interacting clusters. Specifically, interactions are detected within Clus-

ter 1 which encompasses the proximal diversity (D), joining (J) and constant (C), DJCβ1/2 region, Eβ and downstream genes and Cluster 2, which harbors the Vβ2–Vβ30 gene segments. Furthermore, we identify simultaneous interactions, between Vβ1 and Vβ31 and the Vβ2–Vβ30 gene segments, while the regions encompassing the inactive trypsinogen gene clusters have less contacts (Figure 3B). These data are consistent with the low resolution changes in locus conformation that we observe by DNA FISH in recombining DN cells: namely that Vβ genes interact more closely with each other and with the proximal DJCβ1/2 domain as a result of 'locus contraction' (28,29). However, our method enables us to identify that Cluster 2 forms a subdomain specifically in recombining DN T cells while interactions within Cluster 1 are identifiable in B as well as T cells.

In DP cells, reduced interactions within the Vβ2-Vβ30 Cluster 2 subdomain are in line with decontraction occurring in the developmental stage that follows on from rearrangement, as previously demonstrated by us using both DNA FISH and 3C (27) (Figure 3C). In addition, many Vβ gene sequences will have been deleted as a result of rearrangement occurring on at least one allele. The absence of interactions across the *Tcrb* locus, particularly within the Vβ2-Vβ30 gene cluster in either pre-B or immature B is also reflective of an absence of 'locus contraction' or intra-locus interactions occurring in these cells (27) (Figure 3D and E). Taken together these data provide new insight into *Tcrb* locus structure in T and B lineage cells. In particular, the identification of triplet interactions enables us to generate heatmaps similar to those generated by Hi-C and 5C, however in contrast to Hi-C and 5C, which identify pairwise interactions within individual cells, our method can identify three-way contacts between the bait and other regions in close proximity occurring at the same time in the same cell.

An absence of Eβ alters the architecture of the 3′ end of *Tcrb* without changing interactions involving Vβ genes

To further demonstrate the functional importance of identifying triplet interactions in *Tcrb* we used *Tcrb* enhancer knockout cells ($E\beta^{-/-}$) and analyzed interactions using the bait adjacent to Eβ (Figure 2A). Eβ is the sole enhancer identified to date as being important for regulating the *Tcrb* locus. Its deletion leads to a total absence of germline transcription in the proximal DJCβ1-DJCβ2 domain and a failure of *Tcrb* rearrangement accompanied by a block in T cell development at the DN3 stage (32,33). Given its essential role in *Tcrb* recombination, analyses of the Eβ knockout cells provide a heat map of interactions on an unrearranged locus.

Our results demonstrate that an absence of Eβ does not noticeably impact interactions between the Vβ2–Vβ30 gene segments in Cluster 2, which remain equivalently accessible as determined by the presence of ATAC-seq signal in the region. Additionally, we do not observe a reduction in contact between Cluster 2 and Cluster 1, which encompasses the DJCβ1–DJCβ2 subdomain, the Eβ bait and downstream regions (Figure 4A and B). It is of note that we detect slightly stronger interactions between the 3′ V gene segments and the Eβ bait in $E\beta^{-/-}$ cells. This is likely because in the wild-type cells at least some of these will have been deleted as a result of rearrangement.

A zoom in of Cluster 1 in wild-type DN cells using smaller window sizes (2 kb) shows strong contact between the Eβ and the DJCβ1/2 region and simultaneous interactions between the DJCβ1 and DJCβ2 gene segments (Figure 4C). However, in the Eβ knockout DN cells, interactions within Cluster 1 are markedly depleted compared to controls (Figure 4C and D). The loss of contacts within Cluster 1 coincides with a complete absence of transcriptional activity within the region in $E\beta^{-/-}$ DN cells, which goes hand in hand with the absence of ATAC-seq signals and rearrangement in this region (32,33).

These findings are largely in line with the findings from recent 3C studies performed by the Oltz lab (34). However, unlike 3C analyses, which provide a measure of the interaction frequency between two fixed points, our 4C provides a comprehensive unbiased comparison of interaction frequencies from the bait across the entire locus. Furthermore, the triplet analysis adds a new dimension to the analyses showing how different regions of the locus that have contacts with Eβ have simultaneous interactions with other regions in the locus. Specifically, our data unequivocally demonstrate that in both wild-type and enhancer mutant cells, interactions within and between Clusters 1 and 2 occur simultaneously in the same cell.

4C-seq multi-loci interactions reflect known lineage specific changes in *Igk* locus conformation

The MiEκ and 3′Eκ enhancers are known to promote transcription and recombination of the *Igk* locus (35,36). Here we selected to analyze lineage specific changes in locus conformation across *Igk* from the MiEκ and 3′Eκ baits using our pipeline for detection of tri-loci interactions. For our locus wide analysis, we generated an interaction matrix, denoted as M3, whose columns represent sixteen experiments involving two replicate experiments that use the baits MiEκ and 3′Eκ in four different cells types (DN, DP, Immature B and Pre-B) and whose rows represent the intensity of tri-loci interactions within windows sizes of 200 kb across the genomic co-ordinates 68M–72M on chromosome 6. We applied PCA to the interaction matrix M3 (Supplementary Figure S3) and found that the leading principal components (contribute 45.9% and 11.2% of total variation) indicating that the two B cell subsets (pre-B and Immature B) were significantly separated from the two T cell subsets (DN and DP). We found no significant differences between experiments using the MiEκ and 3′Eκ baits (Figure 5A).

Next, we pooled the replicates and displayed the normalized tri-loci interaction intensity in the four cell types, DN, DP, pre-B and Immature B for the MiEκ and 3′Eκ baits (Figure 5B–E and Supplementary Figure S4A–D, respectively). Heatmaps for the MiEκ and 3′Eκ baits of the same cell type show similar interaction patterns in agreement with the PCA. We identified three major interaction clusters within the Vκ genes from the heatmaps of the pre-B and immature B cells. Within the three hubs Vκ gene segments simultaneously interact with other Vκ gene segments and these interactions occur at the same time as Vκ-MiEκ, Vκ-3′Eκ interactions. This profile is consistent with

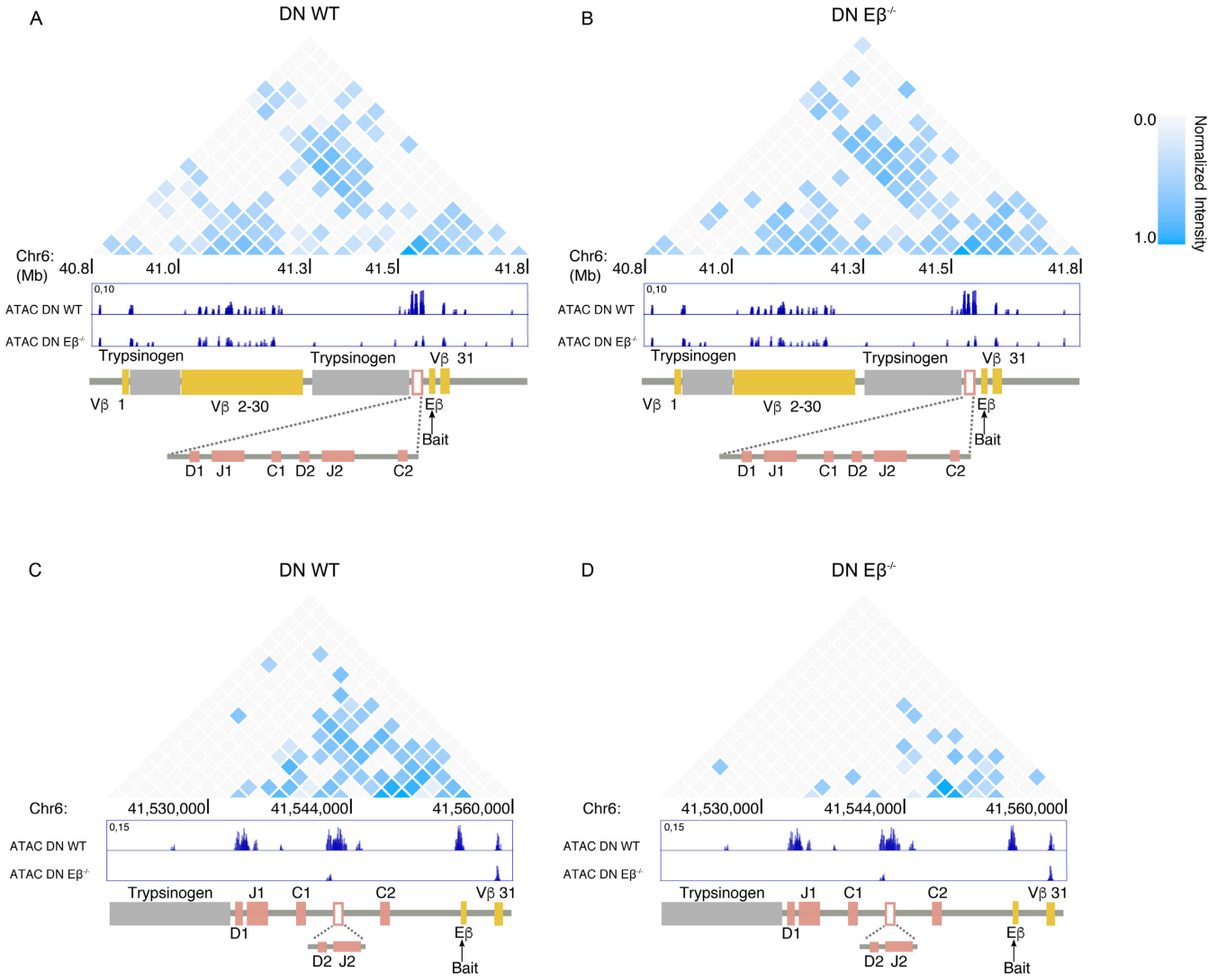

Figure 4. 4C-seq multi-loci interactions identify changes in contacts across *Tcrb* that result from deletion of Eβ. (**A** and **B**) Heatmap for tri-loci interactions (window size of 50kb) involving the Eβ bait in DN WT versus $E\beta^{-/-}$ cells across the whole *Tcrb* locus. The ATAC-seq tracks display read density. (**C** and **D**) Heatmap for tri-loci interactions (window size of 2kb) involving the Eβ bait in DN WT versus $E\beta^{-/-}$ cells across DJC of the *Tcrb* locus. The ATAC-seq tracks display read density.

low resolution DNA FISH analyses, demonstrating that *Igk* undergoes locus contraction in recombining B cells (28). Nonetheless, the triplet analysis provides the first detailed picture of what contacts this involves and additionally identifies three Vκ interaction hubs.

As expected we see fewer contacts along the *Igk* locus in T cells, however, consistent with previous findings the enhancers interact more with regions outside of the locus at the 3′ end in these cells (37) (Figure 5D, E and Supplementary Figure S4C, D). Together these results demonstrate that our method can detect known lineage specific changes in locus conformation in addition to identifying new simultaneous interactions between different V gene segments and the enhancers. These interaction hubs may influence the topology of antigen receptor loci and be important for the generation of diversity that results from V(D)J recombination.

Enhancer hubs and their impact on super-enhancer activity

In addition to the MiEκ and 3′Eκ, *Igk* possesses another enhancer, Edκ. MiEκ and 3′Eκ are both important for rearrangement and deletion of either one leads to a reduction in the frequency of κ expressing B cells, while the double mutant is sufficient to abrogate *Igk* recombination altogether (35,36). In contrast, an absence of both the 3′Eκ and Edκ leads to a dramatic reduction in germline and rearranged transcription, a reduction in active chromatin marks, increased DNA methylation and reduced levels of rearrangement (38). Thus, all three enhancers have overlapping and distinct functions that contribute to the regulation of *Igk*.

Our recent investigations have identified that this cluster of enhancers constitutes a super-enhancer in pre-B and Immature B cells (22). Using 4C-seq with baits adjacent to the MiEκ and 3′Eκ enhancers we found that these regulatory elements are in contact with each other in 3D space. Further,

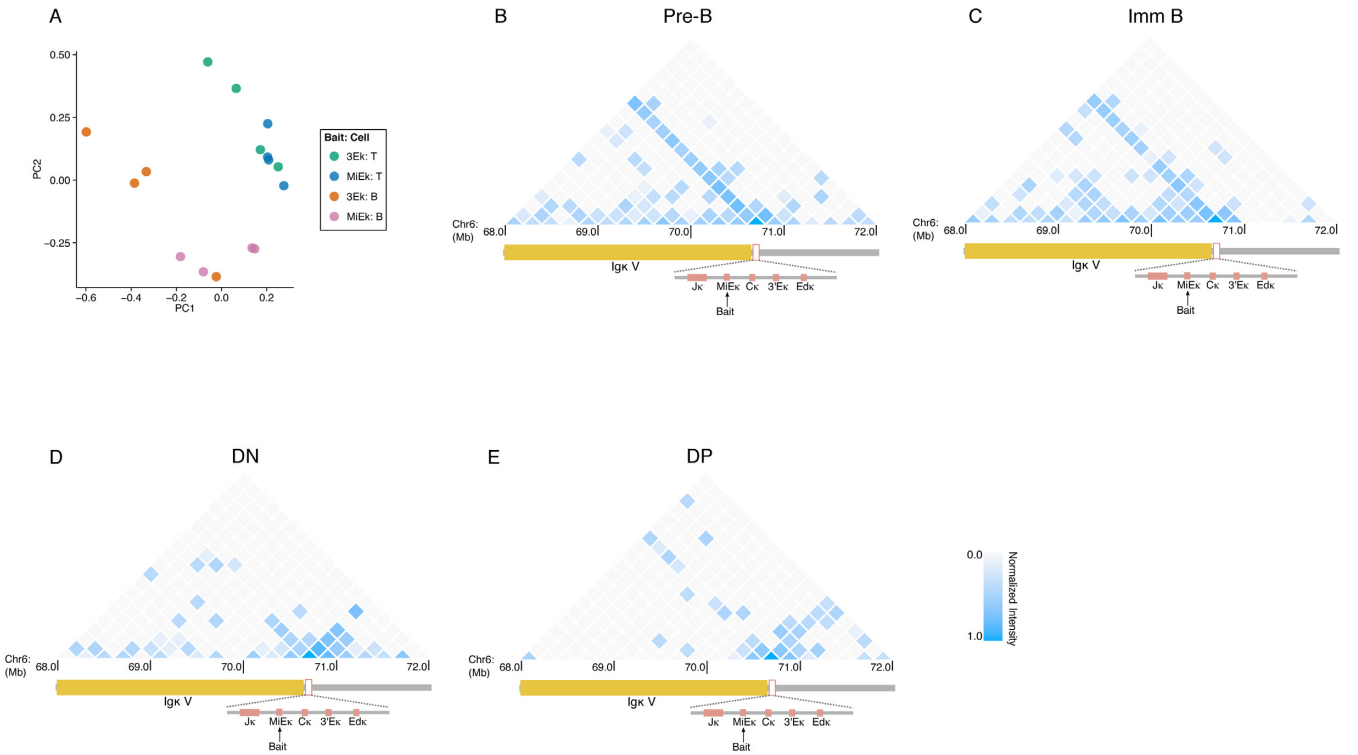

Figure 5. 4C-seq multi-loci interactions reflect known lineage specific changes in *Igk* locus conformation. (**A**) Scatterplot of the first and second principal components of pair-wise interactions of DN, DP, pre-B and Immature B cells using MiEκ and 3′Eκ as baits. (**B**–**E**) Heatmaps for tri-loci interactions (window size of 200kb) involving the MiEκ bait in pre-B, Immature B, DN and DP cells across the *Igk* locus.

we showed that deletion of one element reduces the interaction frequency between other enhancers in the hub, which in turn compromises the transcriptional output of each component. However, we did not determine whether the three enhancers interact with each other at the same time in the same cell.

To examine the relationship between the three enhancers we generated heatmaps of tri-loci interactions using a window size of 2.5 kb. We pooled the reads from pre-B and Immature B cell 4C-seq experiments using MiEκ and 3′Eκ as baits in WT cells. Using the 3′Eκ or MiEκ baits, we detected strong simultaneous local contacts between MiEκ, 3′Eκ and Edκ in WT B cells (Figure 6A and D). As in our pairwise analyses (22) we found that deletion of either the MiEκ or the 3′Eκ not only interferes with tri-loci MiEκ, 3′Eκ and Edκ interactions in which they participate, but they also disrupt pairwise interactions between the two other enhancers in the hub (Figure 6B, C, E, F, G). These changes correspond to a reduction in transcriptional output at the other partner enhancers (22) (Figure 6H). These findings highlight the interdependent nature of multi-loci associations and their functional importance in gene regulation and provide new insight into chromatin organization as a whole. Importantly they demonstrate that in the case of the *Igk* super-enhancer, the three enhancers are indeed in close physical contact at the same time in each cell. This provides an explanation for why deletion of each individual element could disrupt the contacts and transcriptional output of the others (22). These findings add new mechanistic insight into how super-enhancers function.

DISCUSSION

Although we know of many examples that underscore the importance of pairwise chromatin interactions in gene regulation (7–11), we have almost no information about the impact of multi-loci associations. Here, we present a method that uses 4C-seq data to identify multi-locus interactions from the same cell. We demonstrate the feasibility of our method using 4C-seq data that identifies known pairwise interactions that occur within the *Tcrb* and *Igk* antigen receptor loci and additionally extract information that tells us which of these is involved in tri-loci associations. In addition, we examine the effect of the Eβ deletion on interactions across the *Tcrb* locus and demonstrate that consistent with previous findings, loss of this enhancer does not disrupt interactions between Vβ gene segments and Eβ. In contrast we show that loss of contacts within the DJCβ1–DJCβ2 domain go hand in hand with loss of germline transcription in this region in the mutant cells. Importantly, our analyses now demonstrate that in both wild-type and enhancer mutant cells, interactions between the Eβ bait and the Vβ gene segments occur at the same time as interactions between the Vβ gene segments.

Recent Hi-C analysis and related approaches have uncovered the fact that the genome is organized into megabase size TAD structures (15–18). It is thought that one potential function of these highly conserved structures is to locally restrict the action of regulatory elements to loci contained within their boundaries. The identification of these new structures then opens up a number of important ques-

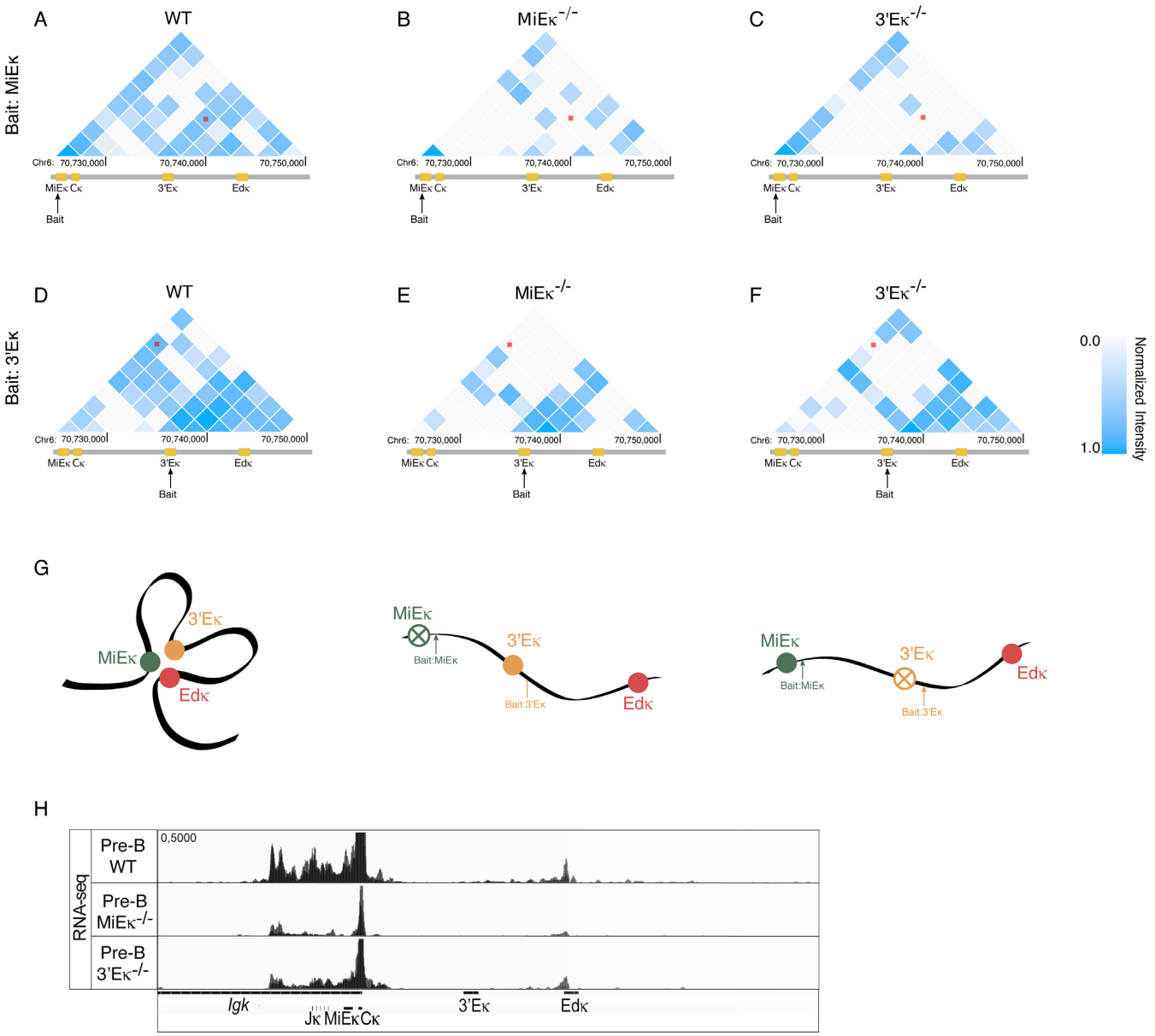

Figure 6. Enhancer hubs and their impact on super-enhancer activity. (**A–C**) Heatmap for tri-loci interactions (window size of 2.5 kb) involving the MiEκ bait in WT versus MiEκ$^{-/-}$ and 3′Eκ$^{-/-}$ B cells across the 3′ end of the *Igk* locus. The intensity of interactions between the three regions in the neighborhoods of the three enhancers (which is proportional to the intensity of the blue color) is marked by a red dot in each of the heatmaps. (**D–F**) Heatmap for tri-loci interactions involving the 3′Eκ bait in WT versus MiEκ$^{-/-}$ and 3′Eκ$^{-/-}$ B cells across the 3′ end of the *Igk* locus. (**G**) Model showing the interactions in of the three enhancers in WT, MiEκ$^{-/-}$ and 3′Eκ$^{-/-}$ pre-B cells. (**H**) RNA-seq data showing the transcriptional output across the 3′ end of the *Igk* locus in WT, MiEκ$^{-/-}$ and 3′Eκ$^{-/-}$ pre-B cells.

tions concerning gene regulation. For example, although we now know that promoters are not specifically tethered to specific enhancers in a one-to-one relationship (39), we do not yet know whether simultaneous interaction of enhancer elements with an individual promoter contributes to the regulation of individual genes (40,41). Another example is the β-globin locus in which multiple pair-wise interactions imply the possibility of simultaneous interactions between the locus control region (LCR), 5′ hypersensitive site (5′HS) and 3′ hypersensitive site (3′ HS) (42). Our pipeline provides a tool to address these questions.

Here, we have used our method to show that the three enhancers of *Igk*, which can be defined as a super-enhancer in pre-B and immature B cells (22) interact simultaneously within the same cell. This provides mechanistic insight into why deletion of an individual element reduces pairwise associations between the other components such that if one contact is lost it impairs the contacts between other components in the hub. Furthermore, this finding provides functional insight into why there is a concomitant reduction in the transcriptional output of the other elements. These data suggest that contacts between the individual components of a super-enhancer are important for their activity. Fur-

ther analyses of other super-enhancers will need to be performed to determine whether these rules are a general feature of super-enhancers. Our pipeline provides a means to do this. Finally, our method provides a tool to determine if gene regulation involving multiple elements involves simultaneous interactions that are dictated by hierarchical relationships.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank members of the Skok and Kluger labs for helpful suggestions and discussions. We would also like to thank NYU CHIBI and the NYUMC sorting and genome facilities for their contributions to this work. CP and VS, sorted developing lymphocytes from wild-type and enhancer knockout mice and generated the 4C-seq, RNA-seq and ATAC-seq data. T.J., Y.K., J.S., R.R., C.P., V.S., S.B. and P.R. developed the idea for identification of tri-loci interactions and T.J. implemented the pipeline with help from R.R. and P.R. J.S., T.J. and Y.K. wrote the manuscript with comments from all authors.

FUNDING

NYSTEM institutional training grant [C026880 to V.S.]; P.P.R. is a National Cancer Center and American Society of Hematology Fellow; J.A.S. is a Leukemia & Lymphoma Society (LLS) scholar; National Institutes of Health (NIH) [1R01 HG008383-01A1 to Y.K., P30CA16359 to Y.K., R01 GM086852 to J.A.S. and Y.K. and R01GM112192 to J.A.S. and R.B.]. Funding for open access charge: Discretionary funds (to Y.K.).
Conflict of interest statement. None declared.

REFERENCES

1. de Laat,W., Klous,P., Kooren,J., Noordermeer,D., Palstra,R.J., Simonis,M., Splinter,E. and Grosveld,F. (2008) Three-dimensional organization of gene expression in erythroid cells. *Curr. Top. Dev. Biol.*, **82**, 117–139.
2. Schoenfelder,S., Sexton,T., Chakalova,L., Cope,N.F., Horton,A., Andrews,S., Kurukuti,S., Mitchell,J.A., Umlauf,D., Dimitrova,D.S. *et al.* (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.*, **42**, 53–61.
3. Collins,A., Hewitt,S.L., Chaumeil,J., Sellars,M., Micsinai,M., Allinne,J., Parisi,F., Nora,E.P., Bolland,D.J., Corcoran,A.E. *et al.* (2011) RUNX transcription factor-mediated association of Cd4 and Cd8 enables coordinate gene regulation. *Immunity*, **34**, 303–314.
4. Hewitt,S.L., Farmer,D., Marszalek,K., Cadera,E., Liang,H.E., Xu,Y., Schlissel,M.S. and Skok,J.A. (2008) Association between the Igk and Igh immunoglobulin loci mediated by the 3′ Igk enhancer induces 'decontraction' of the Igh locus in pre-B cells. *Nat. Immunol.*, **9**, 396–404.
5. Hewitt,S.L., Yin,B., Ji,Y., Chaumeil,J., Marszalek,K., Tenthorey,J., Salvagiotto,G., Steinel,N., Ramsey,L.B., Ghysdael,J. *et al.* (2009) RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nat. Immunol.*, **10**, 655–664.
6. Rocha,P.P., Micsinai,M., Kim,J.R., Hewitt,S.L., Souza,P.P., Trimarchi,T., Strino,F., Parisi,F., Kluger,Y. and Skok,J.A. (2012) Close proximity to Igh is a contributing factor to AID-mediated translocations. *Mol. Cell*, **6**, 873–885.
7. Gibcus,J.H. and Dekker,J. (2013) The hierarchy of the 3D genome. *Mol. Cell*, **49**, 773–782.
8. Bouwman,B.A. and de Laat,W. (2015) Getting the genome in shape: the formation of loops, domains and compartments. *Genome Biol.*, **16**, 154.
9. Gomez-Diaz,E. and Corces,V.G. (2014) Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol.*, **24**, 703–711.
10. Gorkin,D.U., Leung,D. and Ren,B. (2014) The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell*, **14**, 762–775.
11. Raviram,R., Rocha,P.P., Bonneau,R. and Skok,J.A. (2014) Interpreting 4C-Seq data: how far can we go? *Epigenomics*, **6**, 455–457.
12. Tolhuis,B., Palstra,R.J., Splinter,E., Grosveld,F. and de Laat,W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell*, **10**, 1453–1465.
13. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
14. Narendra,V., Rocha,P.P., An,D., Raviram,R., Skok,J.A., Mazzoni,E.O. and Reinberg,D. (2015) Transcription. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science*, **347**, 1017–1021.
15. Sexton,T., Yaffe,E., Kenigsberg,E., Bantignies,F., Leblanc,B., Hoichman,M., Parrinello,H., Tanay,A. and Cavalli,G. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, **148**, 458–472.
16. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
17. Nora,E.P., Lajoie,B.R., Schulz,E.G., Giorgetti,L., Okamoto,I., Servant,N., Piolot,T., van Berkum,N.L., Meisig,J., Sedat,J. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381–385.
18. Rocha,P.P., Raviram,R., Bonneau,R. and Skok,J.A. (2015) Breaking TADs: insights into hierarchical genome organization. *Epigenomics*, **7**, 523–526.
19. Giorgetti,L., Galupa,R., Nora,E.P., Piolot,T., Lam,F., Dekker,J., Tiana,G. and Heard,E. (2014) Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, **157**, 950–963.
20. Ay,F., Vu,T.H., Zeitz,M.J., Varoquaux,N., Carette,J.E., Vert,J.P., Hoffman,A.R. and Noble,W.S. (2015) Identifying multi-locus chromatin contacts in human cells using tethered multiple 3C. *BMC Genomics*, **16**, 121.
21. Zhao,H., Sifakis,E.G., Sumida,N., Millan-Arino,L., Scholz,B.A., Svensson,J.P., Chen,X., Ronnegren,A.L., Mallet de Lima,C.D., Varnoosfaderani,F.S. *et al.* (2015) PARP1- and CTCF-mediated interactions between active and repressed chromatin at the lamina promote oscillating transcription. *Mol. Cell*, **59**, 984–997.
22. Proudhon,C., Snetkova,V., Raviram,R., Lobry,C., Badri,S., Jiang,T., Hao,B., Trimarchi,T., Kluger,Y., Aifantis,I. *et al.* (2016) Active and inactive enhancers co-operate to exert localized and long-range control of gene regulation. *Cell Rep.*, **15**, 2159–2169.
23. Tonegawa,S. (1983) Somatic generation of antibody diversity. *Nature*, **302**, 575–581.
24. Helmink,B.A. and Sleckman,B.P. (2012) The response to and repair of RAG-mediated DNA double-strand breaks. *Annu. Rev. Immunol.*, **30**, 175–202.
25. Proudhon,C., Hao,B., Raviram,R., Chaumeil,J. and Skok,J.A. (2015) Long-Range Regulation of V(D)J Recombination. *Adv. Immunol.*, **128**, 123–182.
26. Jhunjhunwala,S., van Zelm,M.C., Peak,M.M., Cutchin,S., Riblet,R., van Dongen,J.J., Grosveld,F.G., Knoch,T.A. and Murre,C. (2008) The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell*, **133**, 265–279.
27. Skok,J.A., Gisler,R., Novatchkova,M., Farmer,D., de Laat,W. and Busslinger,M. (2007) Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nat. Immunol.*, **8**, 378–387.

28. Roldan,E., Fuxa,M., Chong,W., Martinez,D., Novatchkova,M., Busslinger,M. and Skok,J.A. (2005) Locus 'decontraction' and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat. Immunol.*, **6**, 31–41.
29. Fuxa,M., Skok,J., Souabni,A., Salvagiotto,G., Roldan,E. and Busslinger,M. (2004) Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev.*, **18**, 411–422.
30. Whyte,W.A., Orlando,D.A., Hnisz,D., Abraham,B.J., Lin,C.Y., Kagey,M.H., Rahl,P.B., Lee,T.I. and Young,R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
31. Buenrostro,J.D., Giresi,P.G., Zaba,L.C., Chang,H.Y. and Greenleaf,W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
32. Bouvier,G., Watrin,F., Naspetti,M., Verthuy,C., Naquet,P. and Ferrier,P. (1996) Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 7877–7881.
33. Bories,J.C., Demengeot,J., Davidson,L. and Alt,F.W. (1996) Gene-targeted deletion and replacement mutations of the T-cell receptor beta-chain enhancer: the role of enhancer elements in controlling V(D)J recombination accessibility. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 7871–7876.
34. Majumder,K., Koues,O.I., Chan,E.A., Kyle,K.E., Horowitz,J.E., Yang-Iott,K., Bassing,C.H., Taniuchi,I., Krangel,M.S. and Oltz,E.M. (2015) Lineage-specific compaction of Tcrb requires a chromatin barrier to protect the function of a long-range tethering element. *J. Exp. Med.*, **212**, 107–120.
35. Inlay,M., Alt,F.W., Baltimore,D. and Xu,Y. (2002) Essential roles of the kappa light chain intronic enhancer and 3′ enhancer in kappa rearrangement and demethylation. *Nat. Immunol.*, **3**, 463–468.
36. Inlay,M.A., Lin,T., Gao,H.H. and Xu,Y. (2006) Critical roles of the immunoglobulin intronic enhancers in maintaining the sequential rearrangement of IgH and Igk loci. *J. Exp. Med.*, **203**, 1721–1732.
37. Stadhouders,R., de Bruijn,M.J., Rother,M.B., Yuvaraj,S., Ribeiro de Almeida,C., Kolovos,P., Van Zelm,M.C., van Ijcken,W., Grosveld,F., Soler,E. *et al.* (2014) Pre-B cell receptor signaling induces immunoglobulin kappa locus accessibility by functional redistribution of enhancer-mediated chromatin interactions. *PLoS Biol.*, **12**, e1001791.
38. Zhou,X., Xiang,Y., Ding,X. and Garrard,W.T. (2013) Loss of an Igkappa gene enhancer in mature B cells results in rapid gene silencing and partial reversible dedifferentiation. *Mol. Cell. Biol.*, **33**, 2091–2101.
39. Sanyal,A., Lajoie,B.R., Jain,G. and Dekker,J. (2012) The long-range interaction landscape of gene promoters. *Nature*, **489**, 109–113.
40. Sahlen,P., Abdullayev,I., Ramskold,D., Matskova,L., Rilakovic,N., Lotstedt,B., Albert,T.J., Lundeberg,J. and Sandberg,R. (2015) Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol.*, **16**, 156.
41. Schoenfelder,S., Furlan-Magaril,M., Mifsud,B., Tavares-Cadete,F., Sugar,R., Javierre,B.M., Nagano,T., Katsman,Y., Sakthidevi,M., Wingett,S.W. *et al.* (2015) The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.*, **25**, 582–597.
42. Patrinos,G.P., de Krom,M., de Boer,E., Langeveld,A., Imam,A.M., Strouboulis,J., de Laat,W. and Grosveld,F.G. (2004) Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev.*, **18**, 1495–1509.