Transcription factor-mediated supervision of global genome architecture maintains B cell identity

Authors: Timothy M. Johanson[1,2*], Aaron T. L. Lun[1,2*], Hannah D. Coughlan[1,2*], Tania Tan[1,2], Gordon K. Smyth[1,3], Stephen L. Nutt[1,2*] & Rhys S. Allan[1,2,4*]

Affiliations: [1]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, 3052, Australia.

[2]Department of Medical Biology, [3]School of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria, 3010, Australia.

*These authors contributed equally to this work

Correspondence

Rhys S Allan (rallan@wehi.edu.au)

Stephen L Nutt (nutt@wehi.edu.au)

[4] Lead contact

Ph: +61 3 9345 9222 fax: +61 3 9347 0852

Recent studies have elucidated cell lineage-specific three-dimensional genome organization; however, it is unclear how such specific architecture is established or maintained. We hypothesized that lineage-defining transcription factors maintain cell identity via global control of genome organization. These factors control cell fate determination by activating lineage-specific and repressing inappropriate genes. However, they also bind many more genomic sites outside of the genes they directly regulate, potentially implicating them in three-dimensional genome organization. Using chromosome conformation capture techniques, here we show that the transcription factor Paired box 5 (Pax5) is critical for the establishment and maintenance of the global lineage-specific architecture of B cells. Pax5 supervises genome architecture throughout B differentiation, until the plasmablast stage, where Pax5 is naturally silenced and B cell specific genome structure is lost. Crucially, we find that Pax5 was not reliant on ongoing transcription to organize the genome. These results implicate sequence-specific DNA-binding proteins in global genome organization to establish and maintain lineage fidelity.

Keywords: genome organization; Pax5; transcription factor; lineage commitment; B cell differentiation

Introduction

The regulators of the distinct transcriptional and chromatin accessibility profiles of different cell lineages are well defined (Thurman et al., 2012). By contrast, the regulators of the three-dimensional, non-random and cell type specific genome organization known to exist in distinct cell populations are less well understood (Dixon et al., 2015; Javierre et al., 2016; Krijger et al., 2016; Lin et al., 2012). For example, CTCF and cohesin complexes play a critical role in genome organization; however, these proteins are ubiquitously expressed, thus alone cannot regulate cell lineage specific genome architecture. It has been speculated that lineage-restricted transcription factors may establish and maintain genome organization in specific lineages (Natoli, 2010). These factors can be lineage-specific, bind many thousands of sites in the genome (the majority outside of gene promoters (Heinz et al., 2010; Revilla et al., 2012)) and in some cases have been shown to regulate genome organization at particular loci (Harju et al., 2002; Montefiori et al., 2016). The immune system represents an excellent model to examine this question, with numerous immune cell lineages arising from a common hematopoietic stem cell under the influence of lineage-specifying transcription factors.

Pax5 is a prime example of a lineage-specifying transcription factor. It is critical for establishing and maintaining B cell lineage commitment (Nutt et al., 1999) via activation of B cell-specific genes and repression of genes promoting other immune cell lineages (Delogu et al., 2006; Holmes et al., 2006). In the absence of Pax5, mouse B cell progenitors (known as pro-B cells) are unable to differentiate into mature B cells, but also gain stem cell-like pluripotential. Similarly, reduction of Pax5 levels in mature B cells leads to lineage de-differentiation, and frequently leukaemia in both mice and humans (Cobaleda et al., 2007; Kuiper et al., 2007; Mullighan et al., 2007).

Pax5 binds to ~20,000 sites in the murine B cell genome; however only 47% of these sites are in genes or promoters (Revilla et al., 2012). We hypothesize that many of these Pax5 binding sites regulate global three-dimensional genome organization. Here we used genome-wide high-resolution chromosome conformation capture to characterize the distinct genome architecture of immune cell lineages and to link Pax5 occupancy to the establishment and maintenance of B cell-specific lineage architecture throughout B cell differentiation. Furthermore, we show that Pax5 establishes this structure independent of transcription.

Results

In situ HiC detects distinct immune cell lineage genome organization.

To elucidate the genome organization of different immune cell lineages we performed *in situ* Hi-C (Rao et al., 2014), an iteration of chromosome conformation capture, on splenic $CD4^+$ T cells, B cells and granulocytes (Fig 1 A, Fig S1 A-C). After confirming library quality (Fig S1 D-G) we used our *diffHic* software package (Lun and Smyth, 2015) to compare the genome organization between the lineages. *diffHic* allocates the ~200 million generated read pairs of each library (Supplementary Table 1) into pairs of 50 kB genomic bins, and uses the bin pair counts to identify differential interactions (DIs) between cell types genome-wide in a statistically rigorous manner, relative to variability between biological replicates. To examine the organizational relationship between the three cell lineages, we first performed a multi-dimensional scaling analysis using the leading log-fold changes in interaction intensity between samples. The analysis separated the cell types into three distinct and highly reproducible clusters (Fig 1 B). Quantifying these differences we found 15152 DIs between $CD4^+$ T and B cells, 22979 between B cells and granulocytes, and 22203 between $CD4^+$ T cells and granulocytes (FDR < 0.05, Supplementary Table 2, Fig 1 B, C). To put these numbers into context we

also performed *in situ* Hi-C on the common progenitors of these lineages, the bone marrow-derived stem cell-enriched lineage marker$^-$ Sca1$^+$ c-Kit$^+$ population (LSK cells)(Fig. S1 H) and embryonic fibroblasts (MEFs), a representative non-haematopoietic cell type. The organizational difference between the LSK cells and each immune lineage was comparable to the difference between lineages, and each population was clearly distinct from MEFs (Fig. S1 I, Supplementary Table 3). Thus, the three immune cell lineages possess distinct genome organization and have more in common with each other, and their shared stem cell progenitor, than with fibroblasts. In addition to examining differential interactions between the different cell populations we also identified the topologically associated domains (TADs) of each population, using the *TADBit* software package (Serra et al., 2016). TADs are an intermediate scale (megabase) unit of genome organization that are largely insulated from neighbouring DNA (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). We found that each population contained between 1500-2000 TADs with a mean size of 1-2 megabase (Fig. 1 D, Fig S1 J-K, Supplementary Table 4). This is consistent with previous findings suggesting that TAD boundaries vary minimally between immune cell populations (Rao et al., 2014).

We next examined whether the DIs occurred in regions containing lineage-specific genes. For example, we identified elaborate DNA structures connecting the T cell lineage commitment gene, *Bcl11b* (Li et al., 2013), to surrounding regions in CD4$^+$ T cells (Fig. 1 E). This organization was entirely absent in B cells and granulocytes. By overlaying our Hi-C data with p300 ChIP-Seq and H3K27ac ChIP-Seq from T cells (Arvey et al., 2015; Vahedi et al., 2015) and B cells (Choukrallah et al., 2015; Kieffer-Kwon et al., 2013), we confirmed this interaction is anchored in the same region that had been identified previously using targeted techniques (Fig S1 L) (Li et al., 2013). Similar to the *Bcl11b* region in T cells, the region containing the B cell lineage specifying transcription factor, *Ebf1*, was highly organized in B cells, but not in T cells or granulocytes. Lineage-specific organization was also seen around the *Il1f9* gene in granulocytes, in which it exhibited specific expression and function (Ericson et

al., 2014; Kato et al., 2009). To determine whether lineage-specific genes were associated with DIs genome-wide, we examined the association between the 1000 most differentially expressed (DE) genes (Heng et al., 2008) (Supplementary Table 5) and the DIs between each of the three groups. We found that, compared to all genes in the mouse genome, DE genes between any two of the three cell lineages are significantly more likely to associate with the DIs that exist between these two lineages (Fig 1 F).

Genome organization during B cell differentiation is dynamic and associated with Pax5 binding

We next explored how genome organization is established and maintained within a given lineage. To this end, we compared six key stages of B cell differentiation; LSK cells, pro-B cells, transitional and marginal zone B cells (Tr/MZ B cells), follicular B cells ('B cell' libraries of previous figures), activated B cells and plasmablasts (Fig 2 A, Fig S2 A-C). These stages encompass the B cell differentiation pathway, from uncommitted progenitor (LSK cells) to the terminally differentiated antibody-producing plasmablast. The number of DIs between the differentiation stages suggest dramatic changes as LSKs become pro-B cells (20903), pro-B cells become Tr/MZ B cells (28372 DIs) and during B cell activation (11251 DIs). This is in contrast to the lower number of changes observed as Tr/MZ B cells become follicular (461 DIs) or activated B cells differentiate into plasmablasts (4913 DIs) (Fig 2 B-C, Supplementary Table 6). These results are consistent with transcriptional profiling that found activation induces more than half of all transcriptional changes that occur during B cell differentiation (Shi et al., 2015). A number of the organizational changes observed alter regions known to be important to the function and/or development of particular stages, such as *Sox4* and *Prdm1* (Fig 2 D), suggesting developmental importance. In addition to DIs, we also examined TAD boundaries and A/B compartment switching (Lieberman-Aiden et al., 2009). We found minimal changes in both size and number of TADs (Fig S2 D-E, Supplementary Table 6) and A/B compartment switching (Fig S2 F) during B cell differentiation. Compartment switching that did occur predominantly occurs during the earliest stages of B cell differentiation (Supplementary Table 7).

To determine the relationship between Pax5 and the genome organizational changes that occur during B cell differentiation, we examined the association between Pax5 binding to the genome of pro-B cells and follicular B cells (Revilla et al., 2012), and the 23556 DIs that exist between these stages (Supplementary Table 8). Comparing the Pax5 binding profiles of pro-B and follicular B cells, we found 12117 pro-B cell specific binding sites, 14383 follicular B cell specific sites, and 10679 binding sites common to the two stages (Fig 3 A, B, Supplementary Table 9). We then examined the association of Pax5 binding with the strengthening or weakening of genome organization. We found that DIs that contain a pro-B cell specific Pax5 binding site in their anchors (regions of DNA that interact to generate loops in the intervening sequence) were significantly more likely to be weakened as B cell differentiation progresses (69%), compared to those that contained a follicular B cell specific binding site (36%, $p<1e^{-16}$) or a site common to the two stages (42%, $p<1e^{-16}$, Fig 3 C). This suggests that if Pax5 disassociates with a region of DNA as a cell differentiates from a pro-B cell to a follicular B cell, DNA structure in this region will likely be lost. In addition, DIs that overlapped follicular B cell specific Pax5 binding sites were significantly more likely to be strengthened as the transition from pro-B cell to follicular B cell occured ($p=1.9e^{-10}$, Fig 3 C). These data suggest that the physical association of Pax5 with DNA structure during B cell differentiation is highly correlated with the maintenance and establishment of organization genome-wide. In line with this, both the natural up regulation of Pax5 as pro-B cells differentiate from LSK cells, and the down regulation of Pax5 as B cells differentiate into plasmablasts (Fig 3 E), impact genome organization. As such, examining the DIs between LSK cells and pro-B cells and their association with pro-B cell Pax5 binding sites (Supplementary Table 9), we found that as differentiation reached the pro-B cell stage those DIs bound by Pax5 were significantly more likely to be strengthened or created than DIs not physically associated with Pax5 ($p<1e^{-16}$, Fig 3 D). In contrast, the 10827 DIs between follicular B cells and plasmablasts (Supplementary Table 10) that were associated with follicular B cell Pax5 binding sites (Supplementary Table 9) were more likely

to be weakened or removed as differentiation reached the plasmablast stage, and Pax5 is silenced (Fig 3 E), than DIs not physically associated with follicular B cell Pax5 binding sites ($p<1e^{-16}$, Fig 3 F). Thus through out B cell differentiation the physical association of Pax5 with DNA is correlated with maintained and strengthened organization of the genome.

Pax5 establishes and maintains global pro-B cell genome organization

To determine if genetic deletion of Pax5 in pro-B cells had similar impacts on DNA structure as the natural down regulation of Pax5 during plasmablast differentiation, we performed *in situ* Hi-C on wild type and *Pax5*$^{-/-}$ pro-B cells (Fig S3 A-D). Examining TAD boundaries and A/B compartment switching in the two groups we found minimal changes in TAD size and number or compartments (Fig S3 E-F, Supplemental Table 7). In contrast, we identified 7810 DIs between the two groups (Supplementary Table 11). 25 of these DIs fell within the *Igh* locus (Fig. 4 A, Supplementary Table 12), confirming previously identified Pax5-dependent genome organization of this locus (Medvedovic et al., 2013). Examination of the 7810 DIs observed between the wild type and *Pax5*$^{-/-}$ pro-B cells revealed that the majority (83%) of DIs were weakened or removed in the absence of Pax5 (Fig. 4 B), providing strong evidence that Pax5 also functions on a genome-wide scale to maintain B cell nuclear organization. Comparison of the genome organization of wild type and *Pax5*$^{-/-}$ pro-B cells to LSK cells showed that the *Pax5*$^{-/-}$ pro-B cells are not simply de-differentiating (Fig S3 G, Supplementary Table 13). Examining whether the loss of genome architecture was a direct effect of Pax5 deletion, we grouped the wild type and *Pax5*$^{-/-}$ pro-B cell DI bin pairs into those with anchors that were Pax5 bound or unbound in wild type pro-B cells (Supplementary Table 11). DIs that contained no Pax5 binding site had a roughly equal likelihood of being strengthened (40%) or weakened (60%) in the absence of Pax5 (Fig. 4 C). In striking contrast, DIs that contained at least one Pax5 binding site were almost exclusively weakened upon Pax5 deletion (96%, $p<1e^{-16}$). To ensure this impact of Pax5 removal on

191 genome organization was not due simply to the vast number of Pax5 binding sites, we repeated this
192 analysis, and in fact all other analyses containing Pax5 peak calling, with more stringently called Pax5
193 binding sites (score threshold of 50) (Supplementary Table 14). In every case results using more
194 stringently called Pax5 binding sites were consistent with findings derived using less stringently called
195 peaks (Fig S3 H-I).
196
197 We next examined whether reintroducing Pax5 into a *Pax5*$^{-/-}$ nucleus, using a hormone-inducible Pax5-
198 estrogen receptor (ER) fusion protein (Nutt et al., 1998) (Fig S3 J-K), would alter genome organization.
199 We found that 6 hour and 24 hour reintroduction of Pax5 induced significant (p=0.03 and p=3.2e$^{-13,}$
200 respectively) strengthening or creation of interactions in regions anchored by Pax5 in wild type pro-B
201 cells, compared to unbound regions, with the impact most apparent after 24 hours (Fig. 4 D, Fig S3 L,
202 Supplementary Table 15-16). Overall, these data demonstrate that the physical interaction of Pax5 with
203 anchor regions of specific DNA structures is critical to establish and maintain genome organization in
204 developing B cells.
205
206 We next examined whether the process of transcription was required for Pax5 to organize the genome.
207 As such, we performed the six-hour Pax5 reintroduction experiment in the presence of α-amanitin, a
208 potent inhibitor of the binding and function of RNA polymerase II (Hug et al., 2017)(Fig 5 A). The six-
209 hour time point was selected as this short exposure to α-amanitin had minimal impact on pro-B cell
210 viability, unlike a 24-hour treatment (Fig S3 M), while inducing an almost complete removal of Pax5
211 target gene (*Dntt*, *Spi1*) and *Actb* transcripts (Fig 5 B) in treated, compared to untreated cells,
212 suggesting complete inhibition of transcription. Comparison of the genome organisation of untreated
213 cells to those in which Pax5 has been reintroduced in the presence, or absence of transcription revealed
214 5144 and 4754 DIs, respectively (Fig 5 C, Supplemental Table 16). DIs were observed around known
215 Pax5 target genes (for example *Bcar3, Lef1, Dntt and Spi1*) (Fig 5 D, Fig S3 N) (McManus et al.,

2011). Strikingly, 49% of the DIs shared exactly the same genomic regions, while the anchors of 73% of DIs fell within 100kB of each other (Fig 5 E). Unsurprisingly, given the near identical changes induced by Pax5 reintroduction with or without transcription, when we compare the genome organization of Pax5 reintroduced cells in the presence or absence of transcription we found no DIs (Fig 5 C). Thus DNA structures that form, or are removed, upon the reintroduction of Pax5 do so largely independent of transcription, suggesting that it is the binding of Pax5 to DNA, not transcription itself, which drives genome organisational change. In addition, the blocking of transcription reinforces the direct nature of the role for Pax5 in establishing and maintaining genome organisation, as no Pax5 regulated intermediary could be transcribed.

Discussion

Since the initial description of cell type-specific genome organization it has remained unclear how such architecture is established and maintained, and how this underlies gene expression programs and cell identity. It has been speculated that cell identity is preserved through global supervision of genomic organization by lineage determining transcription factors (Natoli, 2010); however until now this has not been clearly demonstrated. Given its large number of binding sites outside of gene promoters and its known organizational function at the *Igh* locus (Ebert et al., 2011), we tested the hypothesis that Pax5 plays a broad role in the organization of the B cell genome. Our data shows that the down regulation or removal of Pax5 results in the weakening and loss of global nuclear organization normally associated with B cell commitment and that reintroduction of Pax5 activity rapidly restores much of this lineage-specific genome structure.

Our data intimates that in addition to the canonical role of a transcription factor of recognizing specific DNA motifs and altering the expression of target genes, Pax5, also regulates genome organization.

241 Strikingly, Pax5 could alter genome architecture even when the act of transcription was inhibited.
242 While these findings are surprising, transcription-independent formation of genome structure has
243 recently been observed during embryonic development (Hug et al., 2017; Ke et al., 2017), while
244 conversely, total loss of structure can have minimal impact on transcription (Rao et al., 2017).
245 However, to our knowledge, this is the first time that a transcription factor has been directly shown to
246 organize the genome independent of transcription and suggests that a primary role of lineage-
247 specifying transcription factors is to modify genome structure prior to gene activation.
248
249 As Pax5 has no intrinsic chromatin remodelling capacity it likely co-opts a number of interacting
250 partners to regulate genome organization. These include, as outlined above, the CTCF and cohesin
251 complex, which is a fundamental regulator of chromatin structure throughout the genome (Busslinger
252 et al., 2017), and may cooperate with Pax5 in regions beyond the *Igh* locus (Medvedovic et al., 2013).
253 Pax5 also interacts with a number of Snf2 ATP-dependent nucleosome remodellers (McManus et al.,
254 2011). One such remodeller, Brahma-related gene-1 (Brg1), is known to be rapidly recruited to
255 chromatin upon introduction of Pax5 (McManus et al., 2011), is important for B cell development
256 (Bossen et al., 2015) and requires ATP, which is known to be critical for genome organization (Kieffer-
257 Kwon et al., 2017; Vian et al., 2018).
258
259 Many transcription factors are thought have the ability to activate some genes while repressing others
260 in the same nucleus. One way a factor could achieve this dual role would be through the regulation of
261 chromosome architecture. One could envisage that as a consequence of Pax5 organizing the genome to
262 activate transcription, genes intervening the interacting anchors may be repressed by being 'looped out'
263 and therefore less accessible to transcriptional machinery. In line with this, Pax5 is more frequently
264 found at the promoters of active target genes compared to inactive (Revilla et al., 2012). Such a
265 configuration could be related to the recently described phenomenon of indirect gene repression by

transcription factor 'theft' during early T cell development (Hosokawa and Rothenberg, 2018). In this situation, PU.1, an important regulator of early T cell development (Champhekar et al., 2015), actively recruits other transcription factors, such as Satb1 and Runx1, from their target sites. These 'stolen' transcription factors contribute to gene activation at PU.1 target genes, while their 'theft' causes reduced expression of their previous targets, thus indirect gene repression by PU.1 (Hosokawa and Rothenberg, 2018). Further work is needed to determine whether such structural mechanisms play a role in regulating gene expression in these and other settings, although it is noteworthy that large scale reorganisation of the genome has been shown to occur in developing T cells in a manner analogous to the changes we describe for developing B cells (Hu et al., 2018).

Given the critical role of Pax5 in lineage commitment, it is unsurprising that it has been implicated as a tumour suppressor (Dang et al., 2015; Mullighan et al., 2007). Our work suggests that the more fluid chromatin state of the *Pax5*$^{-/-}$ B cell progenitors may underlies their ability to de-differentiate in response to alternative lineage stimuli and may contribute to subsequent B cell leukaemia (Cobaleda et al., 2007; Nutt et al., 1999). Recent technical advances allow targeted creation or removal of genome architecture, potentially allowing therapeutic manipulation of genome architecture (Deng et al., 2014; Sanborn et al., 2015; Wienert et al., 2015). With this in mind it is increasingly important to understand the mechanisms and co-factors involved in, and nature of, chromatin regulation by Pax5, and also whether this genome organising activity extends to other lineage-specifying transcription factors.

Accession codes: The data reported in this paper are tabulated in the Supplementary Materials and archived on the GEO database under accession number GSE99163.

Acknowledgments:

We thank T. Mason for animal husbandry and N. Iannarella for technical support. This work was supported by grants and fellowships from the National Health and Medical Research Council of Australia (S.N. #1054925, #1058238, G.S. #1058892, A.L. and G.S. #1054618, R.A. and T.J. #1049307, #1100451, T.J. #1124081) and the Australian Research Council (R.A. #130100541). This study was made possible through Victorian State Government Operational Infrastructure Support and Australian Government NHMRC Independent Research Institute Infrastructure Support scheme.

Author Contributions: T.J designed the research, performed experiments, analysed data and wrote the paper. A.L. and H.C. designed critical software packages, performed bioinformatics experiments and analysed data. T.T. performed experiments and analysed data. G.S. designed and performed bioinformatics analysis and contributed to writing the paper. S.N. and R.A. designed and supervised the research and wrote the paper.

Competing financial interests: The authors declare no conflicts of interest.

References:

Arvey, A., van der Veeken, J., Plitas, G., Rich, S.S., Concannon, P., and Rudensky, A.Y. (2015). Genetic and epigenetic variation in the lineage specification of regulatory T cells. Elife *4*, e07571.

Bossen, C., Murre, C.S., Chang, A.N., Mansson, R., Rodewald, H.R., and Murre, C. (2015). The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth. Nat Immunol *16*, 775-784.

Busslinger, G.A., Stocsits, R.R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., and Peters, J.M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature *544*, 503-507.

Champhekar, A., Damle, S.S., Freedman, G., Carotta, S., Nutt, S.L., and Rothenberg, E.V. (2015). Regulation of early T-lineage gene expression and developmental progression by the progenitor cell transcription factor PU.1. Genes Dev *29*, 832-848.

Choukrallah, M.A., Song, S., Rolink, A.G., Burger, L., and Matthias, P. (2015). Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. Nat Commun *6*, 8324.

Cobaleda, C., Jochum, W., and Busslinger, M. (2007). Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. Nature *449*, 473-477.

Dang, J., Wei, L., de Ridder, J., Su, X., Rust, A.G., Roberts, K.G., Payne-Turner, D., Cheng, J., Ma, J., Qu, C.*, et al.* (2015). PAX5 is a tumor suppressor in mouse mutagenesis models of acute lymphoblastic leukemia. Blood *125*, 3609-3617.

Delogu, A., Schebesta, A., Sun, Q., Aschenbrenner, K., Perlot, T., and Busslinger, M. (2006). Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. Immunity *24*, 269-281.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A.*, et al.* (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell *158*, 849-860.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W.*, et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331-336.

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.

Ebert, A., McManus, S., Tagoh, H., Medvedovic, J., Salvagiotto, G., Novatchkova, M., Tamir, I., Sommer, A., Jaritz, M., and Busslinger, M. (2011). The distal V(H) gene cluster of the Igh locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. Immunity *34*, 175-187.

Ericson, J.A., Duffau, P., Yasuda, K., Ortiz-Lopez, A., Rothamel, K., Rifkin, I.R., Monach, P.A., and ImmGen, C. (2014). Gene expression during the generation and activation of mouse neutrophils: implication of novel functional and regulatory pathways. PLoS One *9*, e108553.

Harju, S., McQueen, K.J., and Peterson, K.R. (2002). Chromatin structure and control of beta-like globin gene switching. Exp Biol Med (Maywood) *227*, 683-700.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Heng, T.S., Painter, M.W., and Immunological Genome Project, C. (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol *9*, 1091-1094.

Holmes, M.L., Carotta, S., Corcoran, L.M., and Nutt, S.L. (2006). Repression of Flt3 by Pax5 is crucial for B-cell lineage commitment. Genes Dev *20*, 933-938.

Hosokawa, H., and Rothenberg, E.V. (2018). Cytokines, Transcription Factors, and the Initiation of T-Cell Development. Cold Spring Harb Perspect Biol *10*.

Hu, G., Cui, K., Fang, D., Hirose, S., Wang, X., Wangsa, D., Jin, W., Ried, T., Liu, P., and Zhu, J. (2018). Transformation of accessible chromatin and 3D nucleome underlies lineage commitment of early T cells. Immunity *48*, 227-242. e228.

Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell *169*, 216-228 e219.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Varnai, C., Thiecke, M.J.*, et al.* (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell *167*, 1369-1384 e1319.

Kato, A., Chustz, R.T., and Schleimer, R.P. (2009). Regulation and function of newly-recognized IL-1 family cytokines in human bronchial epithelial cells (98.18). The Journal of Immunology *182*, 98.18-98.18.

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L.*, et al.* (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. Cell *170*, 367-381 e320.

Kieffer-Kwon, K.R., Nimura, K., Rao, S.S.P., Xu, J., Jung, S., Pekowska, A., Dose, M., Stevens, E., Mathe, E., Dong, P.*, et al.* (2017). Myc Regulates Chromatin Decompaction and Nuclear Architecture during B Cell Activation. Mol Cell *67*, 566-578 e510.

Kieffer-Kwon, K.R., Tang, Z., Mathe, E., Qian, J., Sung, M.H., Li, G., Resch, W., Baek, S., Pruett, N., Grontved, L.*, et al.* (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell *155*, 1507-1520.

Krijger, P.H.L., Di Stefano, B., de Wit, E., Limone, F., van Oevelen, C., de Laat, W., and Graf, T. (2016). Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming. Cell stem cell *18*, 597-610.

Kuiper, R.P., Schoenmakers, E.F., van Reijmersdal, S.V., Hehir-Kwa, J.Y., van Kessel, A.G., van Leeuwen, F.N., and Hoogerbrugge, P.M. (2007). High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. Leukemia *21*, 1258-1266.

Li, L., Zhang, J.A., Dose, M., Kueh, H.Y., Mosadeghi, R., Gounari, F., and Rothenberg, E.V. (2013). A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. Blood *122*, 902-911.

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.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat Immunol *13*, 1196-1204.

Lun, A.T., and Smyth, G.K. (2015). diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. BMC Bioinformatics *16*, 258.

McManus, S., Ebert, A., Salvagiotto, G., Medvedovic, J., Sun, Q., Tamir, I., Jaritz, M., Tagoh, H., and Busslinger, M. (2011). The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. EMBO J *30*, 2388-2404.

396 Medvedovic, J., Ebert, A., Tagoh, H., Tamir, I.M., Schwickert, T.A., Novatchkova, M., Sun, Q., Huis In
397 't Veld, P.J., Guo, C., Yoon, H.S., *et al.* (2013). Flexible long-range loops in the VH gene region of the
398 Igh locus facilitate the generation of a diverse antibody repertoire. Immunity *39*, 229-244.
399 Montefiori, L., Wuerffel, R., Roqueiro, D., Lajoie, B., Guo, C., Gerasimova, T., De, S., Wood, W.,
400 Becker, K.G., Dekker, J., *et al.* (2016). Extremely Long-Range Chromatin Loops Link Topological
401 Domains to Facilitate a Diverse Antibody Repertoire. Cell Rep *14*, 896-906.
402 Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K.,
403 Mathew, S., Ma, J., Pounds, S.B., *et al.* (2007). Genome-wide analysis of genetic alterations in acute
404 lymphoblastic leukaemia. Nature *446*, 758-764.
405 Natoli, G. (2010). Maintaining cell identity through global control of genomic organization.
406 Immunity *33*, 12-24.
407 Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum,
408 N.L., Meisig, J., Sedat, J., *et al.* (2012). Spatial partitioning of the regulatory landscape of the X-
409 inactivation centre. Nature *485*, 381-385.
410 Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid
411 lineage depends on the transcription factor Pax5. Nature *401*, 556-562.
412 Nutt, S.L., Morrison, A.M., Dorfler, P., Rolink, A., and Busslinger, M. (1998). Identification of BSAP
413 (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments.
414 EMBO J *17*, 2319-2333.
415 Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L.,
416 Machol, I., Omer, A.D., Lander, E.S., *et al.* (2014). A 3D map of the human genome at kilobase
417 resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

418 Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn,
419 A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D.*, et al.* (2017). Cohesin Loss Eliminates All Loop
420 Domains. Cell *171*, 305-320 e324.
421 Revilla, I.D.R., Bilic, I., Vilagos, B., Tagoh, H., Ebert, A., Tamir, I.M., Smeenk, L., Trupke, J., Sommer,
422 A., Jaritz, M.*, et al.* (2012). The B-cell identity factor Pax5 regulates distinct transcriptional
423 programmes in early and late B lymphopoiesis. EMBO J *31*, 3130-3146.
424 Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D.,
425 Chinnappan, D., Cutkosky, A., Li, J.*, et al.* (2015). Chromatin extrusion explains key features of
426 loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A.
427 Serra, F., Baù, D., Filion, G., and Marti-Renom, M.A. (2016). Structural features of the fly chromatin
428 colors revealed by automatic three-dimensional modeling. bioRxiv, 036764.
429 Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay,
430 A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the
431 Drosophila genome. Cell *148*, 458-472.
432 Shi, W., Liao, Y., Willis, S.N., Taubenheim, N., Inouye, M., Tarlinton, D.M., Smyth, G.K., Hodgkin, P.D.,
433 Nutt, S.L., and Corcoran, L.M. (2015). Transcriptional profiling of mouse B cell terminal
434 differentiation defines a signature for antibody-secreting plasma cells. Nat Immunol *16*, 663-673.
435 Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C.,
436 Stergachis, A.B., Wang, H., Vernot, B.*, et al.* (2012). The accessible chromatin landscape of the
437 human genome. Nature *489*, 75-82.
438 Vahedi, G., Kanno, Y., Furumoto, Y., Jiang, K., Parker, S.C., Erdos, M.R., Davis, S.R., Roychoudhuri, R.,
439 Restifo, N.P., Gadina, M.*, et al.* (2015). Super-enhancers delineate disease-associated regulatory
440 nodes in T cells. Nature *520*, 558-562.

441 Vian, L., Pekowska, A., Rao, S.S.P., Kieffer-Kwon, K.R., Jung, S., Baranello, L., Huang, S.C., El
442 Khattabi, L., Dose, M., Pruett, N.*, et al.* (2018). The Energetics and Physiological Impact of Cohesin
443 Extrusion. Cell *173*, 1165-1178 e1120.
444 Wienert, B., Funnell, A.P., Norton, L.J., Pearson, R.C., Wilkinson-White, L.E., Lester, K., Vadolas, J.,
445 Porteus, M.H., Matthews, J.M., Quinlan, K.G.*, et al.* (2015). Editing the genome to introduce a
446 beneficial naturally occurring mutation associated with increased fetal globin. Nat Commun *6*,
447 7085.
448
449

Figure Legends:

Fig. 1. Immune cell lineages possess distinct genome organization.

a, Schematic of simplified haematopoietic tree showing relationship between $CD4^+$ T cells, B cells and granulocytes. **b,** Multi-dimensional scaling (MDS) plot showing relationship between the interaction profiles of $CD4^+$ T cells, B cells and granulocytes. Distances on the plot represent leading $\log_2$-fold-changes in interaction intensity. The number of DIs between each cell population pair is shown. **c,** Arc plots of DIs between $CD4^+$ T cells, B cells and granulocytes across chromosome 1. Inset shows select region. Arcs represent interaction between DI (FDR < 0.05) anchors. Z-score is $-\log_{10}$(*p*-value). **d,** Arc plots of TADs in $CD4^+$ T cells, B cells and granulocytes across chromosome 1. Inset shows select region. Arcs represent TADs called using the *TADbit* package. **e,** *In situ* Hi-C contact matrices of select regions containing differential interactions overlapping lineage-specific genes in $CD4^+$ T cells, B cells and granulocytes, overlaid with RNA-Seq data from each population (Supplementary Table 17). Coordinates determined by boundaries of statistically determined DIs (FDR<0.05). Color scale indicates number of reads per bin pair. **f,** Percentage of DE genes between two of the three immune cell lineages that overlap DIs between the same two lineages, compared to total mouse genes that overlap the same DIs.

Fig. 2. Genome organization is dynamic during B cell differentiation

a, Schematic of B cell differentiation. **b,** MDS plot showing relationship between the interaction profiles of LSK cells, pro-B cells, Tr/MZ B cells, follicular B cells ('B cell' libraries of Figure 1), activated B cells and plasmablasts. Distances on the plot represent leading $\log_2$-fold-changes in interaction intensity. Arrows indicate the inferred B cell differentiation pathway. The number of DIs between each differentiation stage are shown. **c,** Arc plots of DIs between the successive stages of B cell development across chromosome 1. Arcs represent interaction between DI (FDR < 0.05) anchors.

Z-score is -$\log_{10}$(*p*-value). **d,** *In situ* Hi-C contact matrices of select regions containing differential interactions overlapping stage-specifically expressed genes in LSK cells, pro-B cells, Tr/MZ B cells, follicular B cells, activated B cells and plasmablasts. Coordinates determined by boundaries of statistically determined DIs (FDR<0.05). Color scale indicates number of reads per bin pair.

Fig. 3. Genome organizational changes during B cell differentiation are associated with Pax5 binding

a, Venn diagram of numbers of ChIP-seq determined pro-B cell specific (light blue) and follicular B cell specific Pax5 binding sites (light purple) and those sites common to the two stages (blue). **b,** Pax5 ChIP-seq tracks and MACS called peaks from pro-B cells and follicular B cells showing pro-B cell specific, common and follicular B cell specific Pax5 binding examples. **c,** Percentage of DI bin pairs between follicular B cells versus pro-B cells and with anchors containing pro-B cell specific, common or follicular B cell specific Pax5 binding sites that are strengthened or weakened as pro-B cells differentiate into follicular B cells. Numbers of unclustered DIs shown. **d,** Percentage of DI bin pairs between pro-B cells versus LSK cells with anchors unbound or bound by Pax5 in pro-B cells that are strengthened or weakened as LSK cells differentiate into pro-B cells. Numbers of unclustered DIs shown. **e,** RNA-Seq derived expression data for *Pax5* in pro-B cells, Tr/MZ B cells, follicular B cells, activated B cells and plasmablasts. Mean+/-SD shown. **f,** Percentage of DI bin pairs between plasmablasts versus follicular B cells with anchors unbound or bound by Pax5 in follicular B cells that are strengthened or weakened as follicular B cells differentiate into plasmablasts. Numbers of unclustered DIs shown.

Fig. 4. Pax5 establishes and maintains pro-B cell genome organization

a, *In situ* Hi-C contact matrices of wild type and *Pax5*$^{-/-}$ pro-B cells across the *Igh* locus. Color scale indicates number of reads per bin pair. Differential interaction (DI) arcs represent interaction between

DI (FDR < 0.05) anchors. **b,** Percentage of DIs between *Pax5*$^{-/-}$ versus wild type pro-B cells strengthened or weakened in the absence of Pax5. Numbers of unclustered DIs shown. **c,** Percentage of DI bin pairs between *Pax5*$^{-/-}$ versus wild type pro-B cells with anchors unbound or bound by Pax5 that are strengthened or weakened in the absence of Pax5. MACS2 peak calling with a score threshold of 10 called 22796 Pax5 binding sites. Numbers of unclustered DIs shown. **d,** Percentage of DI bin pairs between 24 hour β-estradiol treated *Pax5*$^{-/-}$ Pax5:ER pro-B cells versus untreated with anchors unbound or bound by Pax5 that are strengthened or weakened upon the nuclear reintroduction of Pax5. β-estradiol treatment induces translocation of cytoplasmic restricted Pax5:ER fusion protein into the nucleus. MACS2 peak calling with a score threshold of 50 called 5805 Pax5 binding sites. Numbers of unclustered DIs shown.

Fig. 5. Pax5 establishes genome organization in the absence of transcription

a, Schematic of treatment timing of Pax5 reintroduction with transcription blocking. **b,** Reverse transcription qPCR of *Dntt*, *Spi1* and *Actb* transcripts in six-hour Pax5 reintroduced pro-B cells in the presence or absence of 2.5μM α-amanitin. Levels of transcripts normalized to non- α-amanitin treated group as 1. Mean +/- SD of duplicates shown. **c,** Numbers of DIs observed between 6-hour Pax5 reintroduced pro-B cells, with or without α-amanitin treatment, and untreated cells. **d,** *In situ* Hi-C contact matrices of regions around the Pax5 target genes *Lef1* and *Bcar3* in untreated, six-hour Pax5 reintroduced and six-hour Pax5 reintroduced with α-amanitin treatment pro-B cells. Color scale indicates number of reads per bin pair. **e,** Arc plots of DIs between untreated and Pax5 reintroduced pro-B cells (upper panels) and untreated and Pax5 reintroduced with α-amanitin treatment pro-B cells (lower panels) across chromosome 3. Insets show select regions. Arcs represent interaction between DI (FDR < 0.05) anchors. Red arcs represent DIs strengthened upon Pax5 reintroduction, grey are weakened. Z-score is $-\log_{10}(p\text{-value})$.

Online Methods

Mice and Cell isolation

All experiments were performed using male animals at age 6-12 w. All mice were maintained at The Walter and Eliza Hall Institute Animal Facility under specific pathogen–free conditions. All males were randomly chosen from the relevant pool. All experiments were approved by The Walter and Eliza Hall Institute Animal Ethics Committee and performed under the Australian code for the care and use of animals for scientific purposes. Results were analysed without blinding of grouping.

Splenocytes and bone marrow cells were obtained from mechanically homogenized organs. Flow cytometric analyses were performed on BD FACSCanto with sorting performed on the BD Influx (BD Bioscience). Purity was checked and always exceeded 96%. Antibodies were purchased from BD Bioscience or eBioscience (Supplementary Table 17). Lineage marker$^-$ Sca-1$^+$ cKit$^+$ cells (LSK cells) were isolated by immunomagnetic sorting (anti-CD117, Miltenyi) from bone marrow followed by flow cytometric sorting.

Pro-B cell expansion

B220+ cells from wild type and *Pax5*$^{-/-}$ bone marrow were expanded on an OP9 cell layer for 7 days in MEM+Glutamax (Gibco) supplemented with 1% IL-7 supernatant, 10mM HEPES, 1mM Sodium Pyruvate, 1x non essential amino acids (Sigma) and 50μM β-mercaptoethanol (Sigma). At day 7 the IgM- fraction was isolated using immunomagnetic depletion, following manufacturer's

instructions. *Pax5*$^{-/-}$ Pax5:ER pro-B cells (Holmes et al., 2006) were identically cultured, with 1μM β-estradiol (Sigma) being added to the culture for 24 hours if Pax5 nuclear reintroduction was required.

B cell activation

B220+ cells of the spleen were isolated by immunomagnetic sorting (Miltenyi) and seeded at a density of 1×10^5 cells per ml in DMEM containing 10% FCS (Sigma) and 25 μg/ml LPS (Sigma) and cultured for 4 days.

Mouse embryonic fibroblast culture

Mouse embryonic fibroblasts (MEFs) derived from C57BL/6 male mice at E14.5 were transformed with SV40 large T antigen then maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS), 250 μM L-asparagine and 55 μM β-mercaptoethanol.

Flow cytometric DNA content analysis

Ethanol fixed cells are treated with RNase (Sigma) then 10μg propidium iodide before flow cytometric analysis to determine DNA content.

***In situ* Hi-C**

In situ Hi-C was performed as previously described (Rao et al., 2014). Primary immune cell libraries were generated in biological duplicate, while wild type and *Pax5*$^{-/-}$ libraries were in biological triplicate. Libraries were sequenced on an Illumina NextSeq 500 to produce 75 bp paired-end reads. Between 53 million and 375 million valid read pairs were generated per sample (Supplementary Table 1).

Reads from each sample were aligned using the presplit_map.py script in the *diffHic* package v1.4.0 (Lun and Smyth, 2015). Briefly, reads were split into 5' and 3' segments if they contained the *Mbo*I ligation signature (GATCGATC), using cutadapt v0.9.5 (Martin, 2011) with default parameters. Segments and unsplit reads were aligned to the mm10 build of the mouse genome using bowtie2 v2.2.5 (Langmead and Salzberg, 2012) in single-end mode. All alignments from a single library were pooled together and the resulting BAM file was sorted by read name. The FixMateInformation command from the Picard suite v1.117 (https://broadinstitute.github.io/picard/) was applied to synchronize mate information for each read pair. Alignments were resorted by position and potential duplicates were marked using the MarkDuplicates command, prior to a final resorting by name. This was repeated for each library generated from each sample in the data set. Each BAM file was further processed to identify the *Mbo*I restriction fragment that each read was aligned to. This was performed using the preparePairs function in *diffHic*, after discarding reads marked as duplicates and those with mapping quality scores below 10. Thresholds were applied to remove artefacts in the libraries, (Supplementary Table 1). Read pairs were ignored if one read was unmapped or discarded, or if both reads were assigned to the same fragment in the same orientation. Pairs of inward-facing reads or outward-facing reads on the same chromosome separated by less than a certain distance (min.inward and min.outward respectively) were also treated as dangling ends and were

removed. For each read pair, the fragment size was calculated based on the distance of each read to the end of its restriction fragment. Read pairs with fragment sizes above ~1200 bp (max.frag) were considered to be products of off-site digestion and removed. In this manner, approximately 60-70% of read pairs were successfully assigned to restriction fragments in each library. An estimate of alignment error was obtained by comparing the mapping location of the 3' segment of each chimeric read with that of the 5' segment of its mate. If the two segments were not inward-facing and separated by less than ~1200 bp (chim.dist), then a mapping error was considered to be present. Of all the chimeric read pairs for which this evaluation could be performed, around 1-4% were estimated to have errors, indicating that alignment was generally successful. Technical replicates of the same library from multiple sequence runs were then merged with the mergePairs function of *diffHic*.

Detecting differential interactions

Differential interactions (DIs) were detected using the *diffHic* package (Lun and Smyth, 2015). Read pairs were counted into 50 kB bin pairs (with bin boundaries rounded up to the nearest *Mbo*I restriction site) using the squareCounts function. This yielded a matrix of read-pair counts for each bin pair in each library. Filtering was performed using the filterDirect function, where bin pairs were only retained if they had average interaction intensities more than 5-fold higher than the background ligation frequency (estimated from the bulk of inter-chromosomal bin pairs). For the retained bin pairs, counts were normalized between libraries using a loess-based approach to account for abundance-dependent biases. This was performed using the normOffsets function to obtain a matrix of offsets. Tests for differential interactions (DIs) were performed using the quasi-likelihood (QL) framework (Lund et al., 2012) of the edgeR package v3.14.0,

which assesses statistical significance relative to biological variation between the replicate libraries. The design matrix was constructed using a one-way layout that specified the cell lineage to which each library belonged. Using the counts and offsets for all bin pairs, a mean-dependent trend was fitted to the negative binomial dispersions with the estimateDisp function. A generalized linear model (GLM) was fitted to the counts for each bin pair (McCarthy et al., 2012), and the QL dispersion was estimated from the GLM deviance with the glmQLFit function. The QL dispersions were then squeezed toward a second mean-dependent trend, using a robust empirical Bayes strategy (Phipson et al., 2016) to share information between bin pairs. Finally, a p-value was computed for each bin pair using the QL F-test, representing the evidence against the null hypothesis, i.e., no difference in counts between groups. Significant bin pairs were aggregated into clusters using the diClusters function, where each cluster represents a single differential interaction. This reduces redundancy in the results by merging adjacent bin pairs in the two-dimensional interaction space. Two bin pairs were placed in the same cluster if they overlapped in the interaction space, to a maximum cluster size of 500 kB to mitigate chaining effects. The significance threshold for each bin pair was defined such that the cluster-level false discovery rate (FDR) was controlled at 5%. Cluster statistics were computed using the combineTests and getBestTest functions from the csaw package v1.6.0 (Lun and Smyth, 2016).

The above DI detection pipeline was repeated for seven different design matrices. One analysis used all the B cell, T cell and granulocyte libraries (Fig 1). The second analysis used all these libraries plus those for LSK cells and MEFs (Fig S1 I). The third analysis used libraries for the five B cell development stages and the LSK cells (Fig 2). The fourth analysis used the wild type and *Pax5*$^{-/-}$ pro-B cells libraries and the fifth used the wild type and *Pax5*$^{-/-}$ pro-B cells libraries along with the LSK cells The sixth used the *Pax5*$^{-/-}$ Pax5:ER untreated and 24 hour β-estradiol treated pro-B

cells libraries and the last used the *Pax5*$^{-/-}$ Pax5:ER untreated, 6 hour β-estradiol treated and 6 hour β-estradiol treated plus α-amanitin pro-B cells libraries

Detecting TAD boundaries

TAD boundaries were detected with the *TADbit* v0.2.0.5 python based software (Serra et al., 2016). Read pairs were counted into 50 kB bin pairs (with bin boundaries rounded up to the nearest *Mbo*I restriction site) using the squareCounts function of *diffHic* with no filter. This yielded a count matrix containing a read pair count for each bin pair in each library. The count matrix was converted into a contact matrix for each somatic chromosome with the inflate function of the InteractionSet package (Lun et al., 2016). Replicate contact matrices were summed. TAD boundaries were detected for each chromosome with the function find_tad. Only TADs boundaries with a score of 7 or higher were included in the results.

Detecting A/B compartments

A/B compartments were identified at a resolution of 50 kB using the method outlined by Lieberman-Aiden et al (Lieberman-Aiden et al., 2009) using the *HOMER* HiC pipeline (Heinz et al., 2010). Briefly, from the filtered and aligned reads HOMER creates a binned and normalized contact matrix (normalized by library size and distance between bins). A correlation matrix between bins is generated from the contact matrix and principle component analysis (PCA) is applied. The first principle component generally represents the chromosome A/B compartments where arbitrarily positive values correspond to the active A compartment and negative values to

669 the repressed B compartment. With a genome provided, HOMER will aim to overlap a majority of
670 transcriptional start sites with the active compartment.
671
672 After processing with the *diffHic* pipeline libraries were converted to the HiC summary format
673 using R. Then input tag directions were created for each library with the makeTagDirectory
674 function specifying the genome (mm10) and restriction enzyme cute site (GATC). Biological
675 replicates tag directories for each cell type were summed. The runHiCpca.pl function was used on
676 each library with -res 50000 and specifying the genome (mm10) to perform PCA to identify
677 compartments. This was performed over a range of resolutions (50 kb to 200 kb) and at each
678 resolution the result was consistent. Each chromosome was manually checked for the sign on the
679 compartments and where needed the sign was flipped by multiplying by -1. This was only
680 required for chromosome 17.
681
682 To identify changes in A/B compartments between libraries, the getHiCcorrDiff.pl function was
683 used to directly calculate the difference in correlation profiles. Then findHiCCompartments.pl
684 was used to determine which regions flipped in compartment.
685
686 **Detecting looping interactions**
687
688 Looping interactions were detected using a method similar to that described by Rao et al. (Rao et
689 al., 2014). Specifically, read pairs were counted in bin pairs for all libraries of a given cell type or
690 condition. For each bin pair, the log-fold change over the average abundance of each of several
691 neighbouring regions was computed. Neighbouring regions in the interaction space included a
692 square quadrant of sides 'x+1' that was closest to the diagonal and contained the target bin pair

693 in its corner; a horizontal stripe of length '2x+1' centred on the target bin pair; a vertical stripe of
694 '2x+1', similarly centred; and a square of sides '2x+1', also containing the target bin pair in the
695 centre. The enrichment value for each bin pair was defined as the minimum of these log-fold
696 changes, i.e., the bin pair had to have intensities higher than all neighbouring regions to obtain a
697 large enrichment value. These enrichment values were calculated using the neighborCounts
698 function in *diffHic*, with 'x' set to 5 bin sizes (i.e., 250 kB). Putative loops were then defined as
699 those with enrichment values above 0.5, with average count across libraries greater than 10, and
700 that were more than 1 bin size away from the diagonal.
701
702 **Associating interacting loci with gene annotation**
703
704 Overlaps between bin pairs and genomic intervals were performed using the findOverlaps
705 function in the InteractionSet package (Lun et al., 2016). This considers an overlap to be present
706 if one of the genomic intervals of interest overlaps either bin in the pair. Similarly, the
707 linkOverlaps function was used to identify bin pairs where both bins are overlapped by the
708 genomic intervals. This approach was used to identify overlaps between the interacting loci and
709 any genomic interval, including gene bodies and transcription factor binding sites. Note that bin
710 pairs were generally used to identify overlaps, rather than the clustered interactions, as the bin
711 pairs are regularly shaped and more comparable between analyses.
712
713 **Visualisation of Hi-C results**
714
715 Multi-dimensional scaling (MDS) plots were constructed using the plotMDS function in the limma
716 package (Ritchie et al., 2015) applied to the filtered and normalized log2-counts-per-million

values of each bin pair for each library. The distance between each pair of samples was the "leading log-fold-change", defined as the root-mean-square average of the 50,000 largest log2-fold-changes between that pair of samples. Plaid plots were constructed using the contact matrices and the plotHic function from the *Sushi R* package (Phanstiel, 2015). The range of color intensities in each plot was scaled according to the library size of the sample, to facilitate comparisons between plots from different samples. Differential interaction (DI) arcs were plotted with the plotBedpe function of the Sushi package. The z-score shown on the vertical access was calculated as $-\log_{10}(p\text{-value})$. ChIP-seq peaks were plotted with the plotBed function of the Sushi package.

ChIP-seq profiling of binding sites and histone marks

ChIP-seq profiles of CTCF, p300 and Pax5 binding sites and of H3K27ac histone marks were downloaded from the Gene Expression Omnibus (GEO) repository (Supplementary Table 18). Where possible, lists of called peaks were obtained from the supplementary data of the relevant published study. Published mm9 were lifted over to the mm10 coordinates using the liftOver function from the rtracklayer package v1.32.2 (Lawrence et al., 2009). If peak calls were not available, the raw read sequence data was downloaded and aligned to the mm10 genome using Subread v1.4.6 (Liao et al., 2013) with unique alignment. Peaks were called using MACS2 v2.1.0 (Zhang et al., 2008) with default parameters, using all available replicate libraries. Unless otherwise stated, a score threshold of 10 was applied to peak calling.

Reverse Transcription qPCR (RT-qPCR) to determine transcriptional activity

741 Isolation of RNA was performed by running one million cells through the RNeasy Mini Kit
742 (Qiagen) according to manufacturers instructions. To assess relative levels of transcript by qPCR,
743 1 ,g of RNA was used for cDNA synthesis using iScript™ Reverse Transcription Supermix (Bio-
744 Rad) according to manufacturers instructions. Quantification of mRNA levels was performed
745 using iTaq™ Universal SYBR® Green Supermix according to manufacturers instructions (Bio-
746 Rad). The relative standard curve method was used for relative quantitation of RNA abundance.
747 Primer sequences: bActin F – GTGACGAGGCCCAGAGCAAGAG, R – AGCGGCCGGACTCATCGTACTC,
748 Dntt F – GAAGATGGGAACAACTCGAAGAG, R – CAGGTGCTGGAACATTCTGGGAG, PU.1 F-
749 GCACACATGCGTGTTTGTGGATGCT, R - GTGCTTCCTTGGGAGTCTGGCGCT
750
751 **RNA-seq profiling of WT and *Pax5*$^{-/-}$ pro-B cells**
752
753 Total RNA was isolated from wild type and *Pax5*$^{-/-}$ pro-B cells using the RNeasy Plus Mini Kit
754 (Qiagen). Two *Pax5*$^{-/-}$ samples were prepared and one wild type. Samples were sequenced on an
755 Illumina Hiseq 2000 at the Beijing Genomics Institute to produce 90bp paired-end reads. Reads
756 were aligned to the mm10 genome using Subread with unique alignment. The number of read
757 pairs mapped to the exonic regions of each gene in the NCBI RefSeq annotation was counted
758 using the featureCounts function (Liao et al., 2014) from the Rsubread package v1.16.1. Low-
759 abundance genes with an average log-count per million below 0 were filtered out. Sex-linked
760 genes Xist and genes on the Y-chromosome and Xist were removed, as were variable
761 immunoglobulin gene segments. Normalisation was performed using the trimmed mean of M-
762 values method (Robinson and Oshlack, 2010). DE genes were detected using the edgeR QL
763 method described above. The Benjamini-Hochberg method was used to control the FDR below
764 5%. DE genes were overlapped with Pax5 binding sites and DIs.

765

766 **RNA-seq profiling of B-cell differentiation stages**

767

768 RNA-seq profiles of pro-B cells, Tr/MZ B cells, follicular B cells, activated B cells, plasmablasts
769 and marginal zone B cells were downloaded from GEO series GSE60927 and GSE72018
770 (Supplementary Table 18). Paired-end reads were aligned to the mm10 genome using Subread
771 with unique alignment. Read counts were obtained for NCBI Entrez Gene IDs using the
772 featureCounts function (Liao et al., 2014) of the Rsubread package v1.16.1. Low-abundance genes
773 with average log-count per million below 1 were filtered. Genes located on the Y-chromosome
774 and Xist were removed to avoid confounding sex effects. Highly variable immunoglobulin chains
775 were removed along with non-coding genes. Normalisation was performed using the trimmed
776 mean of M-values method (Robinson and Oshlack, 2010). Fragments per kilobase per million
777 (FPKM) values were calculated from the normalized and filtered gene counts. The
778 removeBatchEffect function of the limma package was used to correct for batch effects between
779 the two GEO series.

780

781 **Differentially expressed genes between immune cell lineages from the ImmGen database**

782

783 Lists of DE genes between pairs of immune cell lineages were obtained from the Immunological
784 Genome Project website (http://rstats.immgen.org/PopulationComparison) (Heng et al., 2008).
785 The spleen follicular B cell, spleen naïve T cell, and bone marrow granulocyte V1 datasets were
786 compared. Genes were ranked by fold change.

787

788 **Statistical analysis**

Pearson's chi-squared test with Yates' continuity correction was used to determine significance of differences in proportions.

Methods Only References

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Heng, T.S., Painter, M.W., and Immunological Genome Project, C. (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol *9*, 1091-1094.

Holmes, M.L., Carotta, S., Corcoran, L.M., and Nutt, S.L. (2006). Repression of Flt3 by Pax5 is crucial for B-cell lineage commitment. Genes Dev *20*, 933-938.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25*, 1841-1842.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res *41*, e108.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

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.

Lun, A.T., Perry, M., and Ing-Simmons, E. (2016). Infrastructure for genomic interactions: Bioconductor classes for Hi-C, ChIA-PET and related experiments. F1000Res *5*, 950.

Lun, A.T., and Smyth, G.K. (2015). diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. BMC Bioinformatics *16*, 258.

819 Lun, A.T., and Smyth, G.K. (2016). csaw: a Bioconductor package for differential binding analysis
820 of ChIP-seq data using sliding windows. Nucleic Acids Res *44*, e45.
821 Lund, S.P., Nettleton, D., McCarthy, D.J., and Smyth, G.K. (2012). Detecting differential expression
822 in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. Stat Appl
823 Genet Mol Biol *11*.
824 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
825 EMBnet journal *17*, pp. 10-12.
826 McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor
827 RNA-Seq experiments with respect to biological variation. Nucleic Acids Res *40*, 4288-4297.
828 Phanstiel, D.H. (2015). Sushi: Tools for visualizing genomics data.
829 Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S., and Smyth, G.K. (2016). Robust
830 Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect
831 Differential Expression. Ann Appl Stat *10*, 946-963.
832 Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L.,
833 Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase
834 resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.
835 Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers
836 differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res
837 *43*, e47.
838 Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential
839 expression analysis of RNA-seq data. Genome Biol *11*, R25.
840 Serra, F., Baù, D., Filion, G., and Marti-Renom, M.A. (2016). Structural features of the fly chromatin
841 colors revealed by automatic three-dimensional modeling. bioRxiv, 036764.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Supplemental Figure Legends

Supplemental Figure 1 | a, Flow cytometry of homogenised C57BL/6 Pep^{3b} mouse spleen stained with antibodies against TCRβ, CD4, CD8, CD62L, CD44. $CD4^+$ T cells were isolated as $TCR\beta^+$ $CD4^+$ $CD8^-$ $CD62L^+$ $CD44^-$. **b,** Flow cytometry of homogenised C57BL/6 Pep^{3b} mouse spleen stained with antibodies against TCRβ, CD19, B220, IgM, IgD. B cells were isolated as $TCR\beta^-$ $CD19^+$ $B220^+$ IgM^+ IgD^+. **c,** Flow cytometry of homogenised C57BL/6 Pep^{3b} mouse bone marrow stained with antibodies against TCRβ, CD19, B220, Ly6C, Ly6G. Granulocytes were isolated as $TCR\beta^-$ $CD19^-$ $Ly6C^{int}$ $Ly6G^+$. **d,** Representative MA plots comparing bin pair counts between $CD4^+$ T cell, B cell and granulocytes *in situ* Hi-C libraries, before and after normalization using a loess-based approach with the normOffsets function of the *diffHic* package. Y-axis shows $\log_2$ fold change in interaction intensity while the x-axis shows average $\log_2$ intensity.**e,** Representative histogram of distribution of fragment lengths in *in situ* Hi-C libraries. **f,** Representative plot of strand orientation of fragments in *in situ* Hi-C libraries with respect to the log-insert size (i.e. distance between paired reads on the same chromosome). **g,** Percentage of DNA loop anchor bin pairs overlapped by CTCF binding sites in $CD4^+$ T cells, B cells and granulocytes (Supplementary Table 18). **h,** Flow cytometry of homogenised cKit-enriched C57BL/6 Pep^{3b} mouse bone marrow stained with antibodies against cKit and Sca-1. LSK cells were isolated as $cKit^+$ $Sca\text{-}1^+$. **i,** Multi-dimensional scaling plot showing the relationship between the interaction profiles of $CD4^+$ T cells, B cells, granulocytes, LSK cells and MEFs. Distances on the plot represent leading $\log_2$-fold-changes in interaction intensity. **j,** Using the change-point detection algorithm in *TADbit*, we determined the number and **k,** size of TADs in $CD4^+$ T cells, B cells, granulocytes, LSK cells and MEFs. Mean +/- SD shown. **l,** *In situ* Hi-C contact matrices of $CD4^+$ T cells and B cells were overlaid with H3K27 acetylation ChIP-seq and p300 ChIP-seq (Supplementary

Table 18) from each cell type. The boundaries of statistically significant DIs (FDR < 0.05) determined the coordinates shown. Color scale indicates number of reads per bin pair.

Supplemental Figure 2 | a, Flow cytometry of C57BL/6 Pep^{3b} pro-B cell culture stained with antibodies against IgM and CD19. Pro-B cells were isolated as IgM^-. **b,** Flow cytometry of homogenised C57BL/6 Pep^{3b} mouse spleen stained with antibodies against TCRβ, CD19, B220, IgM, and IgD. Tr/MZ B cells were isolated as $TCR\beta^-$ $CD19^+$ $B220^+$ IgM^+ IgD^-. Follicular B cells were $TCR\beta^-$ $CD19^+$ $B220^+$ IgM^+ IgD^+ (as seen in Supplemental Figure 1 B). **c,** Flow cytometry of LPS-stimulated C57BL/6 Pep^{3b} mouse $B220^+$ splenic cell culture stained with antibodies against CD138 and CD22. Activated B cells were isolated as $CD138^-$ $CD22^+$. Plasmablasts were isolated as $CD138^+$ $CD22^-$. **d,** Using the change-point detection algorithm in *TADbit*, we determined the number and **e,** size of TADs in pro-B cells, Tr/MZ B cells, follicular B cells (as seen in Supplemental Figure 1 B), activated B cells and plasmablasts. Mean +/- SD shown. **f,** A/B compartmental interaction plots of total chromosome 1 at six stages of B cell differentiation. 50kB resolution shown.

Supplemental Figure 3 | a, Flow cytometry of wild type and *Pax5*$^{-/-}$ pro-B cells stained with antibodies against IgM and CD19. Pro-B cells are isolated at IgM^{-}. Wild type also shown in Supplemental Figure 2 A. **b,** DNA content examination by flow cytometry of ethanol fixed wild type and *Pax5*$^{-/-}$ pro-B cells stained with propidium iodide. The largest peak represents cells in G0/1 of the cell cycle, the second peak G2/M, and the intervening region is made up of cells in S phase. **c,** Percentage of DNA loops in wild type and *Pax5*$^{-/-}$ pro-B cells with anchors overlapping at least one CTCF binding site. **d,** Multi-dimensional scaling plot constructed from the interaction intensities in wild type and *Pax5*$^{-/-}$ pro-B cells. Distances between samples represent the leading $\log_2$-fold change between samples for the top 50000 bin pairs with the largest $\log_2$-fold changes. **e,** Using the change-point detection algorithm in *TADbit*, we determined the number and **f,** size of TADs in wild type and *Pax5*$^{-/-}$ pro-B cells. Mean +/- SD shown. **g,** Multi-dimensional scaling plot constructed from the interaction intensities in wild type and *Pax5*$^{-/-}$ pro-B cells and LSK cells. Distances between samples represent the leading $\log_2$-fold change between samples for the top 50000 bin pairs with the largest $\log_2$-fold changes. **h,** As shown in Figure 3 C and F, percentage of DI bin pairs strengthened or weakened between follicular B cells versus pro-B cells and plasmablasts versus follicular B cells with anchors containing either pro-B cell specific, follicular B cell specific or common high (MACS2 peak threshold of 50) or low (MACS2 peak threshold of 10) stringency Pax5 binding sites. Numbers of unclustered DIs shown. **i,** As shown in Figure 4 C-D, percentage of DI bin pairs between *Pax5*$^{-/-}$ versus wild type pro-B cells and treated *Pax5*$^{-/-}$ Pax5:ER pro-B cells versus untreated with anchors unbound or Pax5 bound (high and low stringency) that are strengthened or weakened in the absence or 24-hour reintroduction of Pax5. Numbers of unclustered DIs shown. **j,** Flow cytometry of *Pax5*$^{-/-}$ Pax5:ER untreated and 24 hour β-estradiol treated pro-B cells stained with antibodies against Flt3 and CD19. Untreated *Pax5*$^{-/-}$ pro-B cells were isolated as Flt3^{high} CD19^{-}. 24 hour treated pro-B cells (Pax5

reintroduced) were Flt3^{low} CD19^{+}. **k,** DNA content examination by flow cytometry of ethanol fixed *Pax5*$^{-/-}$ Pax5:ER untreated and 24 hour β-estradiol treated pro-B cells stained with propidium iodide. **l,** Percentage of DI bin pairs between treated *Pax5*$^{-/-}$ Pax5:ER pro-B cells versus untreated with anchors unbound or Pax5 bound that are strengthened or weakened after 6-hour reintroduction of Pax5. Numbers of unclustered DIs shown. **m,** Flow cytometry of *Pax5*$^{-/-}$ Pax5:ER pro-B cell cultures with or without β-estradiol (βES) and/or α-amanitin. Numbers indicate the percentage of viable cells in each gate. **n,** *In situ* Hi-C contact matrices of regions around the Pax5 target genes *Spi1* and *Dntt* in untreated, six-hour Pax5 reintroduced and six-hour Pax5 reintroduced with α-amanitin treatment pro-B cells. Color scale indicates number of reads per bin pair.