Title
Genome 3D structure regulation of Immune cell development and activation

Permalink
https://escholarship.org/uc/item/72j6d8cp

Author
He, Zhaoren

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Genome 3D structure regulation of immune cell development and activation

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biology

by

Zhaoren He

Committee in charge:

Professor Cornelis Murre, Chair
Professor John Chang
Professor Olga Dudko
Professor Michael Rosenfeld
Professor Yang Xu
Professor Gene Yeo

2020

The Dissertation of Zhaoren He is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank my adviser Cornelis Murre for his extensive mentorship in scientific research and communication. The ideas and writing in this dissertation are the product of countless debates and revisions with Kees.

I would like to thank my dissertation committee for their support and advice. Gene Yeo and John Chang provided deep guidance in my exploration of the single cell sequencing data-based analysis.

I would like to thank my colleagues, Nadia Kurd, Michael Kramer, Brigid Boland, Matthew Tsai, Alexandra Bortnick, Megan Aubrey, Takeshi Isoda and Matthew Denholtz, for their close collaborations and deep insights across several projects. I would like to thank all past and present members of Kees Murre's and John Chang's labs.

I would like to thank my family for their never-ending love and support.

I would like to thank my girlfriend Yang Chen for her overflowing love and encouragement as I treaded through the thick and thin in the journey of pursuing the PhD degree. The adventure we went through together made me believe that I have found the right person in my life.

I would like to thank my close friends from San Diego and UCSD, including Xiangyu Ren, Rongxin Fang, Xinzhu Zhou, Liyang Xiong, Jingqiang Ye, Bili Dong, Shu Zhang, Wenhao Jin, Hanqin Liu, Yuwenbin Li, Jiuyuan Nie, Bingfei Yu, Shuwen Chen, Yupeng He, Ye Gao, Jiaxi Wang; from Thermo Fisher, including Yangyang Chen, Yuting Tseng,; and other friends from my college, high school, middle school and elementary school. Their friendships have inspired me to never give up on my dreams.

Chapter 2, in full, has been submitted for publication of the material as it may appear in *Cell Reports*, 2020. “Elaborate and Intricate Changes in Genome Compartmentalization and Inter-Chromosomal Hubs Orchestrate the Plasma Cell Fate”. Alexandra Bortnick, Zhaoren He, Megan Aubrey, Vivek Chandra, Matthew Denholtz, Kenian Chen, Yin C Lin and Cornelis Murre. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reformatted reprint of the material as it appears as "Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate" in *Cell*, 2017. Isoda, T., Moore, A.J., He, Z., Chandra, V., Aida, M., Denholtz, M., van Hamburg, J.P., Fisch, K.M., Chang, A.N., Fahl, S.P., Wiest, D.L. and Murre, C,. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, has been accepted for publication of the material as it may appear in *Genes & Development*, 2020. “Upon microbial challenge human neutrophils undergo rapid changes in nuclear architecture orchestrate an immediate inflammatory gene program”. Matthew Denholtz, Yina Zhu, Zhaoren He, Hanbin Lu, Takeshi Isoda, Simon Döhrmann, Victor Nizet and Cornelis Murre. The dissertation author was the primary investigator and author of this paper.

Chapter 1 and 5 are, in part, based on the material of the three papers mentioned above.

VITA

EDUCATION

2014 Bachelor of Science, Wuhan University
Biology

2020 Doctor of Philosophy, University of California San Diego
Biology

PUBLICATIONS

Boland B.S.*, **He Z.*,** et al. Single-Cell Analyses Elucidate the Cellular and Molecular Landscape of Ulcerative colitis. (Nature Immunology, submitted)

Kurd N.*, **He Z.*,** et al. Molecular determinants and heterogeneity of circulating and tissue-resident memory CD8+ T lymphocytes revealed by single-cell RNA sequencing. (Science Immunology, in review)

Bortnick A.*, **He Z.*,** et al. Elaborate and intricate changes in genome compartmentalization and inter-chromosomal associations orchestrate the plasma cell fate. (Cell Reports, in review)

Milner J.J.*, Toma C.*, **He Z.,** et al. Functional heterogeneity of tissue-resident memory cells delineated by Blimp1 and Id3. (Immunity, in review)

Matt D.*, Yina Z.*, **He Z.,** et al. Upon microbial challenge human neutrophils undergo rapid changes in nuclear architecture to orchestrate an immediate inflammatory gene program. (Genes & Development, accepted)

Turner S., Morey A., Olshansky M., Collas P., **He Z.** et al. Dynamic changes in 3D spatial chromatin organization underpin virus specific CD8+ T cell differentiation. European Journal of Immunology. (2019).

Quon, S. J., Yu, B., **He, Z**. et al. (2019). CTCF is necessary for CD8+ effector T cell differentiation. The Journal of Immunology. (2019).

Klann, J. E., Kim, S. H., Remedios, K. A., **He, Z.,** et al. Integrin Activation Controls Regulatory T Cell–Mediated Peripheral Tolerance. The Journal of Immunology. (2018)

Isoda T., Moore A.J., **He Z.,** et al. ncRNA transcription-induced changes in nuclear architecture directs with high precision enhancer-promoter interaction. Cell. (2017)

Kakaradov, B.*, Arsenio, J.*, Widjaja, C. E., **He, Z.,** et al. Early transcriptional and epigenetic regulation of CD8+ T cell differentiation revealed by single-cell RNA sequencing. Nature Immunology. (2017)

Zeng, H., Wen, S., Xu, W., **He, Z.,** et al. Highly efficient editing of the actinorhodin polyketide chain length factor gene in Streptomyces coelicolor M145 using CRISPR/Cas9-CodA(sm) combined system. Applied microbiology and biotechnology. (2015)

Jia, H., Liang, T., Wang, Z., **He, Z.,** et al. Multistage regulator based on tandem promoters and CRISPR/Cas. ACS Synthetic Biology. (2014)

*** co-first authors**

ABSTRACT OF THE DISSERTATION

Genome 3D structure regulation of immune cell development and activation

by

Zhaoren He

Doctor of Philosophy in Biology

University of California San Diego, 2020

Professor Cornelis Murre, Chair

Each type of cell in the immune system performs critical function to protect the body and maintain health. In the past few decades, the development and activation of T cell, B cell and Neutrophils have been thoroughly studied. However, it's only until recently

that we developed high throughput technology to study how genome folding structure is regulated in the immune cell development process. Here, I exploited the new technologies to study the development of B cell, T cell, and the activation process of neutrophils.

In B cell activation, the activation from the follicular B to the plasma cell stage is governed by a group of transcriptional regulators. We examined how the expression of these key genes correlate with genome architecture. The study showed that the gain of euchromatic character is closely linked with increase in gene expression, especially for some of the key plasma gene like *Prdm1* and *Atf4*. We found that loss of promoter enhancer interaction in *Ebf1* and *Bcl6* were companied by down-regulation of gene expression, which permit the onset of plasma cell development. Interchromosomally, we also found that coordinated gene expression was correlated with 3D positioning of the genes. Our data indicate that Ebf1 expression inhibits B cell activation, and that the initiation of plasma cell development is orchestrated by changes in intrachromosomal compartmentalization and interchromosomal association.

In T cell differentiation, it is well known that *Bcl11b* specifies T cell fate. Here, we showed during T cell development, the Bcl11b enhancer left the lamina and moved to the nuclear interior. A non-coding RNA named ThymoD (thymocyte differentiation factor) was identified to control this process. The transcription of ThymoD itself is necessary and sufficient for the re-localization. During the process, two CTCF sites in the ncRNA gene body were demethylated and activated. Cohesin was accumulated at the CTCF sites and brought the enhancer to interact with the *Bcl11b* promoter 1Mb downstream. In the meantime, activating epigenetic marks were deposited across the loop domain. These

finding indicate how noncoding transcription orchestrates chromatin folding to direct with high precision enhancer-promoter communication and govern immune cell development.

During neutrophil activation, the cell undergoes large-scale changes in nuclear morphology. The pro-inflammatory genes were rapidly re-localized to the euchromatic compartment upon microbial exposure. The process was accompanied by recruitment of cohesion, which established new enhancer promoter interaction and activated the genes. We also found that the activated enhancers were highly enriched in cistromic elements occupied by PU.1, CEBPB, TFE3, JUN and FOSL2. The data demonstrated how the 3D genome architecture was regulated upon microbial presence by recruiting cohesin to an activated enhancer repertoire to activate the inflammatory genes.

Chapter 1: General introduction to genome organization in immune cell development

1.1 B cell activation

Plasma cells are terminally differentiated B cells. It produces antibody that recognize antigens and protect the body from foreign substance. The protective serum antibody titer is one of the first lines of defense against infection (Slifka et al., 1998). In the meantime, if the antibodies target antigens generated within the same individual, the antibodies become pathogenic auto-antibodies. Plasma cells can also develop into malignancy like plasmacytomas, light chain amyloidosis, monoclonal gammopathies, and multiple myeloma (Hiepe et al., 2011; Chesi et al., 2013). Upon presence of pathogens, naïve B cells rapidly activate and proliferate. The activated B cells then futher differentiate into germinal center B cells or plasmablasts. The lifespan of plasma cells varies a lot depending on the subsets (Chernova et al., 2014; Hammarlund et al, 2017). The long-lived plasma cells are responsible to maintain the enduring antibody titers.

Plasma identity establishment has been actively studied during the past decades. *Pax5*, as shown in a recent paper, is a key regulator for differentiating B cells that organize the genome architecture (Johanson et al, 2018). Studies have shown that Genome structure changes were critical for B cell activation (Kieffer-Kwon et al., 2017; Kieffer-Kwon et al., 2013; Bunting et al., 2016; Park et al., 2014). Repression of *Pax5* is critical for B cell activation as forced expression of Pax5 inhibits antibody secretion (Lin et al., 2002). Other than *Pax5*, Prdm1 and Irf4 have been shown to govern B cell activation in a carefully regulated hierarchical gene network. E2A (Tcf3) and E2-2 (Tcf4) are crucial for plasma cell ((Wöhner et al., 2016; Gloury et al., 2016; Chen et al., 2016). The genes associated with the unfolded protein response (UPR) are also important for plasma cell, including *Ire1*, *Perk*, *Atf6*, *Xbp1* and *Atf4* (Bettigole et al., 2015). IRE1 oligomerizes and activates ribonucleases, which splice XBP1 mRNA. PERK oligomerizes and phosphorylates eIF2. Phosphorylated eIF2 then stops global protein translation while promoting ATF4 translation. ATF4 increases ER capacity in turn. The increased ER capacity is important for the antibody production and secretion process in plasma cell.

1.2 T cell differentiation

In the thymus, T cell differentiation is carefully regulated. Early T cell progenitors differentiate into multipotent DN2a cells upon exposure to Delta-Notch signaling. The DN2a cells then develop into DN2b cells, which are committed to undergo VDJ rearrangement. The TCRb rearrangement initiated in DN3a stage. Once a productive TCRb chain has been produced, the cell expands and differentiates into CD4+CD8+ double-positive (DP) thymocytes. The thymocytes are subjected to stringent selection in the DP compartment. The cell either die by neglect or negative selection or persist through positive selection to differentiate into CD4 or CD8 T cells (Klein et al., 2014; Naito et al., 2011).

A closely regulated group of transcriptional regulators carefully govern the development of T cell. First, the E-proteins are activated to turn on a network of genes involved in Notch signaling (Bain and Murre, 1998; Ikawa et al., 2006; Miyazaki et al., 2017). The Notch signaling then wake up the expression of Bcl11b, GATA-3, and TCF1 (Yui and Rothenberg, 2014).

Specifically, Bcl11b is turned on at the DN2a cell stage to facilitate the transition to the DN2b cell stage. Then, Bcl11b is further activated the lineage determining gene network, with the help from E2A. Bcl11b and E2A will also repress the lineage determining gene for alternative cell fates. (Liu et al., 2010; Ikawa et al., 2010; Li et al., 2010a; Longabaugh et al., 2017). An enhancer in the intergenic locus control region of Bcl11b controls the activation of Bcl11b in DN2 stage. Notch, GATA-3 TCF1 and RUNX1 are all binding to the Major Peak enhancer (Guo et al., 2008; Weber et al., 2011; Garcia-Ojeda et al., 2013; Li et al., 2013). Furthermore, a recent study has shown that the chromatin state transitioning from inactive to active is the rate limiting step of full activation of Bcl11b (Kueh et al., 2016).

1.3 Neutrophil cell activation

Neutrophils are a type of white blood cell that helps heal damaged tissues and resolve infections. They are the first line of immune defense against to such abnormality. Upon injury or infection, neutrophils exit the circulation via extravasation, migrate towards abnormality site, release reactive oxygen and nitrogen species, phagocytose small pathogens, extrude their chromatin as cytotoxic granule-laced extracellular traps (NETs) and rapidly increase cytokine gene expression to coordinate a broader immune response (Ley et al. 2018; Brinkmann et al. 2004). Neutrophils also produce various kind of cytokines and chemokines to facilitate inflammatory response (Ley et al. 2018).

Mature neutrophils are polymorphonuclear. Their nuclei are composed of multiple distinct but internally continuous lobes. Such nucleus structure allows the cells to swiftly migrate between or though endothelial cells by paracellular route or transcellular route

(Rowat et al. 2013; Olins et al. 2009; Muller 2013). Multi-lobular nuclear architecture on neutrophils is heavily controlled by the Lamin B Receptor (LBR) (Shultz et al. 2003; Hoffmann et al. 2002). Deletion of Lbr gene in mouse neutrophils leads to normal nuclear shape and loss of the toroidal nuclei shape (Shultz et al. 2003, Zhu et al. 2017). In human, LBR mutation is linked with Pelger-Huët anomaly. The patients have reduced lobe number in granulocytes (Hoffmann et al. 2002).

Neutrophils employ a suite of different cell surface and endolysosomal receptors to recognize the highly diversified pathogens, for example, Toll-like receptors (TLRs), C-type lectin receptors, and formyl peptide receptors. These receptors activate a variety of downstream signals, which eventually converge on the NF-κB and AP1 transcription factors to activate a wide spectrum of inflammatory genes, including the cytokines and chemokines IL-8/CXCL8, TNFα, IL-1β, IL-17, and IFNγ (Thomas and Schroder 2013; Garcia-Romo et al. 2011; Tecchio et al. 2014).

1.4 The regulation of genome 3D structure in immune cell development

In a human nucleus, there are more than 20,000 genes positioned on a genome of 6469.66 Mb. Genes always start with a genetic regulatory element known as promoter. Promoter provides a platform for transcriptional regulatory element to function. However, the nucleus is extremely crowed place. The total length of all DNA molecules in a single human nucleus is about 3 meters, while average diameter of human nucleus is about 6 micrometers. The cell needs to pack the genome in elegant way to ensure the regulatory molecules, including transcriptional factors, mediators, polymerases, etc. can find their target efficiently in the crowded nucleus.

With the advance in technologies like 3C, 4C, 5C, HiC and HiC-ChIP, we have gained significant knowledge about some basic rules and functions of genome architecture. The chromosomes have their own chromosome territories that rarely intermingle, except for nucleoli (Sexton and Cavalli, 2015; Maharana et al., 2016). Each of the chromosomes are organized by tandem array of topologically associated domains (TADs) which can be roughly classified as euchromatic (A) or heterchromatic (B) compartments (Lieberman-Aiden et al., 2009). The euchromatic compartments usually are positioned in the nuclear interior, while the heterochromatic compartments tend to be associated with the nuclear lamina (Peric-Hupkes et al., 2010; Kind et al., 2015). The domains exhibit strong enrichment of intra-domain interaction and few inter-domain interactions. The purpose of dividing the chromosome into domains is thought to be minimizing unwanted interaction between genes and regulatory elements not in the same domain. During developmental progression, gene elements often reposition from the lamina to the nuclear interior and vice versa to modulate gene expression (Lin et al., 2012; Isoda et al., 2017).

Enhancers are regulatory elements that bound with transcriptional factors. When enhancers interact with promoters, the RNAP loaded on the corresponding promoters starts to actively transcribe the downstream gene. The enhancer-promoter interaction can form large loop on the chromatin and bring distant enhancer into close proximity to the promoter. Enhancer promoter interaction can be regulated by many different factors, including CCCTC-binding factor (CTCF), cohesins and some non-coding RNAs. During cell development, enhancer-promoter interaction can form or break to regulate gene transcription.

For adaptive immunity, the most welly studied folding regulation is the regulation of antigen receptor loci recombination. The diversity relies on the tightly regulated recombination of three pools of gene elements, known as variable (V), diversity (D) and joining (J) genes. The wide variety of the elements is the corner stone for adaptive immunity. B cell utilized three such loci: *Igh*, *Igk* and *Igl* while T cells have four: *Tcrg*, *Tcrd*, *Tcrb* and *Tcra*. Recombination involves multiple rounds of breaking the DNA, repairing and removing the intervening part. Such process is called VDJ recombination. For each locus, a V-gene, a J-gene and a D-gene are selected to construct the final gene. In the recombination process, the DNA will be bended and loop upon itself, which brings the selected elements close together. Therefore, how the DNA loop heavily influence the recombination process. The mouse Igh locus is separated into three sub-domains, one mostly contains the D and J genes, the other two break the V-region in two halves. The shared boundary between the D-J sub-domain and the adjacent V sub-domain is called intergenic control region 1 (IGCR1), which contains multiple transcriptional factor binding and CTCF binding site. Studies have shown that deleting the CTCF binding sites in IGCR1 lead to abnormal merging of the two adjacent sub-domains and abnormal recombination process. In such scenario, the V-D recombination will happen before D-J recombination, while normally D-J recombination always happen first (Guo et al, 2011).

In neutrophils, because of the non-spheroid nuclei, chromosome positioning is quite different from normal cells. Imaging experiments have shown that both X and Y chromosomes are positioned as a nuclear appendage (Lamborot-Manzur et al, 1971; Karni et al, 2001). With the recent advance in sequencing technology, people found that the lobe structure leads to highly enriched remote genomic interactions of heterochromatic regions in neutrophils. Such enrichment of long-distance interaction leads to large-scale chromosome condensation. Terminal differentiation of murine neutrophils is also associated with the relocation of centromeres, pericentromeres, telomeres, LINE elements, and ribosomal DNA from the nuclear interior to the nuclear lamina, a process that requires the Lbr gene (Zhu et al. 2017). The characteristic lobed shape of neutrophils nuclear is a result of the LBR wrapping the malleable nuclear around the heterochromatic component of the neutrophil genome.

In chapter 2, we studied how genome folding changes during B cell activation and found that changes in expression of many genes can be explained by local folding pattern changes. The coordinated expression of genes is also well correlated with inter-chromosomal association between genes during the B cell activation process.

In chapter 3, we studied how the expression of a long non-coding RNA 1 Mb away from the target gene promoter, can regulate the distal enhancer-promoter interaction, determine the localization of the locus, and activate gene expression.

In chapter 4, we studied genome structure changes of human neutrophil during its activation by different stimuli. We found that upon microbe encounter, a subset of modestly euchromatic subdomains, spatially segregated from the highly euchromatic A compartment, displayed strengthening of their euchromatic character, and re-localized from a peri-nuclear envelope position towards the nuclear interior. The rapid loss of insulation at euchromatic subdomain boundaries lead to long-range enhancer promoter interaction that activate inflammatory gene program.

1.5 References

Bain, G., and Murre, C. (1998). The role of E-proteins in B- and T-lymphocyte development. Semin. Immunol. 10, 143–153.

Bettigole, S. E., and Glimcher, L. H. (2015). Endoplasmic reticulum stress in immunity. Annual Rev. Immunol. 33, 107-138.

Bunting, K.L., Soong, T.D., Singh, R., Jiang, Y., Béguelin, W., Poloway, D.W., Swed, B.L., Hatzi, K., Reisacher, W., Teater, M. and Elemento, O. (2016). Multi-tiered reorganization of the genome during B cell affinity maturation anchored by a germinal center-specific locus control region. Immunity 45, 497-512.

Chen, S., Miyazaki, M., Chandra, V., Fisch, K.M., Chang, A.N., and Murre, C. (2016). Id3 orchestrates germinal center B cell development. Mol. Cell. Biol., 36, 2543-2552.

Chernova, I., D. D. Jones, J. R. Wilmore, A. Bortnick, M. Yucel, U. Hershberg, and D. Allman. (2014). Lasting antibody responses are mediated by a combination of newly formed and established bone marrow plasma cells drawn from clonally distinct precursors. J. Immunol. 193, 4971–4979.

Chesi, M., and Bergsagel, P. L. (2013). Molecular pathogenesis of multiple myeloma: basic and clinical updates. Int. J. Hematol. 97, 313-323.

Garcia-Ojeda, M.E., Klein Wolterink, R.G., Lemaıˆtre, F., Richard-Le, Goff, O., Hasan, M., Hendriks, R.W., Cumano, A., and Di Santo, J.P. (2013). GATA-3 promotes T cell specification by repressing B cell potential in pro-T cells in mice. Blood 121, 1749–1759.

Garcia-Romo, G.S., Caielli, S., Vega, B., Connolly, J., Allantaz, F., Xu, Z., Punaro, M., Baisch, J., Guiducci, C., Coffman, R.L. (2011). Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. Sci. Transl. Med. 3, 73.

Gloury, R., Zotos, D., Zuidscherwoude, M., Masson, F., Liao, Y., Hasbold, J., Corcoran, L.M., Hodgkin, P.D., Belz, G.T., Shi, W. and Nutt, S.L. (2016). Dynamic changes in Id3 and E-protein activity orchestrate germinal center and plasma cell development. J. Exp. Med. 213, 1095-1111.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.L., Hansen, E., Despo, O., Bossen, C., Vettermann, C. and Bates, J.G., (2011). CTCF-binding elements mediate control of V (D) J recombination. Nature 477, 424.

Guo, Y., Maillard, I., Chakraborti, S., Rothenberg, E.V., and Speck, N.A. (2008). Core binding factors are necessary for natural killer cell development and cooperate with Notch signaling during T-cell specification. Blood 112, 480–492.

Hammarlund, E., Thomas, A., Amanna, I.J., Holden, L.A., Slayden, O.D., Park, B., Gao, L. and Slifka, M.K. (2017). Plasma cell survival in the absence of B cell memory. Nat. Comm. 8, 1781.

Hiepe, F., Dörner, T., Hauser, A. E., Hoyer, B. F., Mei, H., Radbruch, A. (2011). Long-lived autoreactive plasma cells drive persistent autoimmune inflammation. Nat. Rev. Rheumatol. 7, 170-178.

Hoffmann K, Dreger CK, Olins AL, Olins DE, Shultz LD, Lucke B, Karl H, Kaps R, Müller D, Vayá A. (2002). Mutations in the gene encoding the lamin B receptor produce an altered nuclear morphology in granulocytes (Pelger-Huët anomaly). Nat. Gen. 31, 410–414.

Ikawa, T., Hirose, S., Masuda, K., Kakugawa, K., Satoh, R., Shibano-Satoh, A., Kominami, R., Katsura, Y., and Kawamoto, H. (2010). An essential developmental checkpoint for production of the T cell lineage. Science 329, 93–96.

Ikawa, T., Kawamoto, H., Goldrath, A.W., and Murre, C. (2006). E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. J. Exp. Med. 203, 1329–1342.

Isoda, T., Moore, A.J., He, Z., Chandra, V., Aida, M., Denholtz, M., van Hamburg, J.P., Fisch, K.M., Chang, A.N., Fahl, S.P., et. al. (2017). Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. Cell, 171, 103-119.

Johanson, T. M., Lun, A. T., Coughlan, H. D., Tan, T., Smyth, G. K., Nutt, S. L., and Allan, R. S. (2018). Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. Nat. Immunol., 19, 1257.

Karni, R.J., Wangh, L.J. and Sanchez, A.J., (2001). Nonrandom location and orientation of the inactive X chromosome in human neutrophil nuclei. Chromosoma, 110, 267-274.

Kieffer-Kwon, K.R., Nimura, K., Rao, S.S., Xu, J., Jung, S., Pekowska, A., Dose, M., Stevens, E., Mathe, E., Dong, P. and Huang, S.C. (2017). Myc regulates chromatin decompaction and nuclear architecture during B cell activation. Mol. Cell 67, 566-578.

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

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M. and Fudenberg, G., (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell 163, 134-147.

Klein, L., Kyewski, B., Allen, P.M., and Hogquist, K.A. (2014). Positive and negative selection of the T cell repertoire: what thymocytes see (and don’t see). Nat. Rev. Immunol. 14, 377–391.

Kueh, H.Y., Yui, M.A., Ng, K.K., Pease, S.S., Zhang, J.A., Damle, S.S., Freedman, G., Siu, S., Bernstein, I.D., Elowitz, M.B., and Rothenberg, E.V. (2016). Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. Nat. Immunol. 17, 956–965.

Lamborot-Manzur, M., Tishler, P. and Atkins, L., (1971). Fluorescent drumsticks in male polymorphs. The Lancet 297,973-974.

Ley K, Hoffman HM, Kubes P, Cassatella MA, Zychlinsky A, Hedrick CC, Catz SD. (2018). Neutrophils: New insights and open questions. Sci. Immunol. 3, 4579.

Li, L., Leid, M., and Rothenberg, E.V. (2010a). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science 329, 89–93.

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.

Lin, K. I., Angelin-Duclos, C., Kuo, T. C., and Calame, K. (2002). Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to immunoglobulin M-secreting plasma cells. Mol. Cell. Bio. 22, 4771-4780.

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.

Liu, P., Li, P., and Burke, S. (2010). Critical roles of Bcl11b in T-cell development and maintenance of T-cell identity. Immunol. Rev. 238, 138–149.

Longabaugh, W.J.R., Zeng, W., Zhang, J.A., Hosokawa, H., Jansen, C.S., Li, L., Romero-Wolf, M., Liu, P., Kueh, H.Y., Mortazavi, A., and Rothenberg, E.V. (2017). Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. Proc. Natl. Acad. Sci. USA 114, 5800–5807.

Maharana, S., Iyer, K. V., Jain, N., Nagarajan, M., Wang, Y., and Shivashankar, G. V. (2016). Chromosome intermingling—the physical basis of chromosome organization in differentiated cells. Nucleic Acids Res. 44, 5148-5160.

Miyazaki, M., Miyazaki, K., Chen, K., Jin, Y., Turner, J., Moore, A.J., Saito, R., Yoshida, K., Ogawa, S., Rodewald, H.R., et al. (2017). The E-Id protein axis specifies adaptive lymphoid cell identity and suppressed thymic innate lymphoid cell development. Immunity 46, 818–834.e4.

Muller W. (2013). Getting leukocytes to the site of inflammation. Vet. Pathology 50, 7–22.

Naito, T., Tanaka, H., Naoe, Y., and Taniuchi, I. (2011). Transcriptional control of T-cell development. Int. Immunol. 23, 661–668.

Olins AL, Hoang TV, Zwerger M, Herrmann H, Zentgraf H, Noegel AA, Karakesisoglou I, Hodzic D, Olins DE. (2009). The LINC-less granulocyte nucleus. European Cell Biol. 88, 203–214.

Park, S. K., Xiang, Y., Feng, X., and Garrard, W. T. (2014). Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. Genes Dev. 28, 1159-1164.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M. and Reinders, M. (2010).

Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol. Cell 38, 603-613.

Rowat AC, Jaalouk DE, Zwerger M, Ung W, Eydelnant IA, Olins DE, Olins AL, Herrmann H, Weitz DA, Lammerding J. (2013). Nuclear envelope composition determines the ability of neutrophil-type cells to passage through micron-scale constrictions. J. Biol. Chem. 288, 8610–8618.

Sexton, T., and Cavalli, G. (2015). The role of chromosome domains in shaping the functional genome. Cell, 160, 1049-1059.

Shultz LD, Lyons BL, Burzenski LM, Gott B, Samuels R, Schweitzer PA, Dreger C, Herrmann H, Kalscheuer V, Olins AL. (2003). Mutations at the mouse ichthyosis locus are within the lamin B receptor gene: a single gene model for human Pelger-Huët anomaly. Hum. Mol. Genet. 12, 61–69.

Slifka, M. K., Antia, R., Whitmire, J. K., and Ahmed, R. (1998). Humoral immunity due to long-lived plasma cells. Immunity 8, 363-372

Tecchio, C., Micheletti, A., Cassatella, M.A. (2014). Neutrophil-derived cytokines: facts beyond expression. Front. Immunol. 5, 508.

Thomas, C.J., Schroder, K. (2013). Pattern recognition receptor function in neutrophils. Trends Immunol. 34, 317–28.

Wöhner, M., Tagoh, H., Bilic, I., Jaritz, M., Poliakova, D. K., Fischer, M., and Busslinger, M. (2016). Molecular functions of the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell development. J. Exp. Med. 213, 1201-1221.

Weber, B.N., Chi, A.W., Chavez, A., Yashiro-Ohtani, Y., Yang, Q., Shestova, O., and Bhandoola, A. (2011). A critical role for TCF-1 in T-lineage specification and differentiation. Nature 476, 63–68.

Yui, M.A., and Rothenberg, E.V. (2014). Developmental gene networks: a triathlon on the course to T cell identity. Nat. Rev. Immunol. 14, 529–545.

Zhu, Y., Gong, K., Denholtz, M., Chandra, V., Kamps, M.P., Alber, F., Murre, C. (2017). Comprehensive characterization of neutrophil genome topology. Genes Deve. l31, 141–153.

Chapter 2: Elaborate and Intricate Changes in Genome Compartmentalization and Inter-Chromosomal Associations Orchestrate the Plasma Cell Fate

2.1 Abstract

The transition from the follicular B to the plasma cell stage is orchestrated by an ensemble of transcriptional regulators. Here we examined how changes in expression patterns of key developmental regulators relate to alterations in nuclear positioning. We found that gene activation at the onset of plasma cell development was concomitant with a gain in euchromatic character for an ensemble of genomic regions that dictate plasma cell fate, including the Prdm1 and Atf4 loci. We found that to permit the onset of plasma cell development the Ebf1 locus repositioned to peri-centromeric heterochromatin. Plasma cell differentiation was also associated with increased inter-chromosomal associations concurrent with alterations in gene expression and co-localization of Prdm1, Xbp1 and Atf4 transcripts in nuclear bodies. These data indicate that Ebf1 enforces the follicular B cell fate and that the onset of plasma cell fate is orchestrated by elaborate changes in compartmentalization and inter-chromosomal associations.

2.2 Background

The genome is folded into chromosome territories that, with the exception of nucleoli, rarely intermingle (Sexton and Cavalli, 2015; Maharana et al., 2016). Chromosomes themselves are organized as tandem arrays of loops that interact to establish euchromatic (A) or heterchromatic (B) compartments (Lieberman-Aiden et al., 2009). While the euchromatic (active) compartment is predominantly positioned in the nuclear interior, a large fraction of the heterochromatic (silenced) compartment is associated with the nuclear lamina (Peric-Hupkes et al., 2010; Kind et al., 2015). During developmental progression, regulatory and coding elements often reposition from the lamina to the nuclear interior and vice versa to modulate patterns of gene expression (Lin et al., 2012; Isoda et al., 2017). An additional layer of chromatin architecture involves the assembly of nuclear bodies such as nucleoli and perinucleolar heterochromatin, which, similar to the nuclear lamina, have been associated with inactive gene transcription (Huang et al., 1997; Kind et al., 2013).

Plasma cells are terminally differentiated B cells responsible for maintaining protective serum antibody titers that serve as a first line of defense against infection (Slifka et al., 1998). Plasma cells are also a source of pathogenic auto-antibodies, and in malignancy, the origin of plasmacytomas, light chain amyloidosis, monoclonal gammopathies, and multiple myeloma (Hiepe et al., 2011; Chesi et al., 2013). In response to pathogens, naïve B cells (also referred to as interphase, mature, resting, or quiescent) rapidly proliferate to form a population of activated B cells, which further differentiate into germinal center B cells or extra-follicular antibody-secreting plasma cells (also referred to as plasmablasts). Experiments tracking antigen-specific responses over hundreds of days have revealed plasma cell subsets with varying lifespans (Chernova et al., 2014; Hammarlund et al, 2017). These findings raise questions about the minimum genomic requirements to establish plasma cell identity. One recent study identified Pax5 as an important genome organizer for developing B cells (Johanson et al, 2018). Other studies linking B cell differentiation to chromatin reorganization events have focused on activated B cells (24 hours after stimulation), germinal center B cells, or mixed populations of activated and differentiated cells (Kieffer-Kwon et al., 2017; Kieffer-Kwon et al., 2013; Bunting et al., 2016; Park et al., 2014). Here, we isolate purified populations of Blimp-expressing plasma cells derived both in vitro and in vivo to interrogate the relationship between gene expression and chromatin conformation.

Conventional models hold that enduring antibody titers are maintained by long-lived plasma cells, which take up residence in specialized microenvironments of the bone marrow. Studies in the last five years have shown that, in both mice and people, as many as 40-50% of BM plasma cells are immature and largely short-lived cells. Short-lived plasma cells differ by the expression of the cell surface marker B220 (Chernova et al., 2014; Halliley et al., 2015).

B cell development relies on a highly regulated and hierarchical program of gene expression. In early B cells, the bHLH protein E2A induces the expression of FOXO1 and early B cell factor 1 (EBF1), which in turn activates the expression of FOXO1 in a feed-forward loop, ultimately activating PAX5 expression (Mansson et al., 2012). Numerous transcription factors, including Prdm1 and Irf4 have been shown to orchestrate the onset of plasma cell development. Repression of early B cell genes including Pax5 is also central to initiation of the plasma cell program as forced expression of Pax5 interferes with antibody secretion (Lin et al., 2002). In contrast, sustained combined expression of E2A (Tcf3) and E2-2 (Tcf4) are crucial to the plasma cell program (Wöhner et al., 2016; Gloury et al., 2016; Chen et al., 2016). Plasma cells also constitutively activate the unfolded protein response (UPR), a specializing sensing mechanism for detecting and processing large amounts of protein shuttled through the endoplasmic reticulum (ER). The three sensors known to implement the UPR pathway are inositol-requiring enzyme 1 (IRE 1), PKR-like ER kinase (PERK), and activating transcription factor 6 (ATF6) (Bettigole et al., 2015). IRE1 oligomerizes and activates ribonucleases, which splice x-box-binding protein 1 (XBP1) mRNA. PERK oligomerizes and phosphorylates eukaryotic translation initiation factor 2□ (eIF2□). Phosphorylated eIF2□ inhibits global protein translation while favoring translation of activating transcription factor 4 (ATF4) mRNA. ATF4 increases ER capacity.

Despite their essential role in health and disease our insight into how the genomes of naïve B and plasma cells are organized in 3D-space remains rudimentary. Initial studies in developing B cells demonstrated that during the pre-pro-B to pro-B cell transition and beyond, the genome undergoes large-scale changes in nuclear architecture (Lin et al., 2012; Johanson et al., 2018). Here we have mapped in detail nuclear architectural changes that occur during plasma cell differentiation. We find that Ebf1 enforces the follicular B cell fate and that the onset of plasma cell development is associated with intricate and elaborate refolding of chromatin involving genes encoding for critical developmental regulators and accumulation of transcripts encoding for key developmental regulators into shared transcriptional hubs.

2.3 Results

Single Cell RNA Sequencing of Follicular B and Plasma Cell Transcriptomes Reveals Commonalities among Subsets

To identify transcription signatures associated with differentiating plasma cells we performed single cell RNA-Seq analyses for naïve follicular B cells as well as purified plasma cell populations. Specifically, follicular naïve B cells (CD23+) were isolated from mice that carry a GFP reporter allele within the Prdm1 locus (Kallies et al., 2004), which were then activated to generate in vitro derived plasma cells. Short-lived (GFP+B220+) and long-lived (GFP+B220-) plasma cells were sorted from the bone marrow and spleen derived from Prdm1-GFP mice. We used the 10X Genomics Chromium platform for sequencing, and recuperated 500 to 53931 reads per cell, depending on the cell type, with 88.8~93.2 % of uniquely mapped reads. A total of 9782 genes were detected across the dataset; 7803 genes showed a mean expression of at least 1 transcript per million reads (TPM) per cell . We used an unsupervised t-distributed stochastic neighborhood embedding (tsne) analysis to identify the transcriptomes associated with the isolated populations. Each of the purified populations segregated as distinct clusters (Figure 2.1A). Plasma cells derived from 3-day in vitro LPS-activated follicular B cell cultures displayed substantial overlap with short-lived plasma derived from the spleen or bone marrow (Figure 2.1A). Long-lived plasma cells derived from the bone marrow segregated into two clusters, one of which showed overlap with short-lived plasma cells whereas a second group of long-lived plasma cells clustered independently (Figure 2.1A). We next plotted the expression patterns for a selected group of genes in follicular B, LPS-activated B, short-lived and long-lived plasma cells (Figure 2.1B). As a first approach we compared expression levels derived from scRNA-Seq analysis to those derived from bulk RNA-Seq reads. We found that merging of scRNA-Seq reads yielded comparable expression levels as compared to those obtained from bulk RNA-Seq analysis (Figure 2.S1). Genes encoding for proteins specific to the plasma cell fate, including Prdm1, Tcf4, Xbp1, Irf4 and Atf4 were expressed in the short-lived and long-lived plasma compartments (Figures 1B and S1B-S1C). Tcf3 was expressed in both naïve follicular B and plasma cells consistent with previous observations (Quong et al., 1999; Chen et al., 2016). Ebf1, Bach2, Pax5 and Id3 were primarily expressed in naive follicular B cells but not in short-lived or long-lived plasma cells (Figures 1B and S2A). As expected, Cd86 and c-myc expression were enriched in LPS-activated follicular B cells but downregulated in plasma cells (Figure 2.1B). Gene ontology analysis for the different B cell populations revealed that as expected plasma cell differentiation was closely associated with genes encoding for components of the endoplasmic reticulum as well as the UPR (Figure 2.S2B). Long-lived plasma cells were significantly enriched for genes encoding for components of the ribosome machinery and DNA replication as compared to short-lived plasma cells (Figure 2.S2B). To determine whether and how Prdm1, Xbp1 and Atf4 expression is correlated we plotted the UMI (unique molecular identifiers) per million reads (UPM). We found that Prdm1 expression correlated well with Xbp1 and Atf4 expression (Figure 2.1C). To validate these findings, we performed RNA-FISH. Specifically, Blimp-GFP cells were sorted, formaldehyde-fixed and hybridized with fluorescently labeled intronic Prdm1, Atf4 and Xbp1 probes (Figure 2.1D). Consistent with the scRNA-Seq analysis Xbp1 and Prdm1 as well as Atf4 and Prdm1 showed coordinate expression (Figure1E). To determine the relationship between Prdm1, Xbp1 and Atf4 expression we calculated the

Pearson correlation. We found significant synergy in Xbp1 and Prdm1 expression. Specifically, the Pearson correlation between Prdm1 and Xbp1 transcript foci was 0.601 (Chi-square p-value = 1.21e-6) and 0.442 involving Prdm1 and Atf4 expression (Chi-square p-value = 1.32e-1) (Figure 2.1E). Taken together, these data indicate that the establishment of plasma cell fate is concomitant with the expression of genes associated with the UPR, endoplasmic reticulum biogenesis, ribosome biogenesis and coordinate regulation of genes encoding for key transcriptional regulators.

The Plasma Cell Specific Transcription Signature is Associated with Changes in Compartmentalization

In previous studies we found that the onset of B cell development was closely associated with changes in chromatin topology that correlated well with alterations in gene expression (Lin et al., 2012). To determine whether likewise the onset of a plasma cell specific transcription signature is linked with changes in genome topology, we performed genome-wide tethered chromosomal conformation capture analysis (TCC) (Kalhor et al., 2012). Specifically, follicular naïve B cells (CD23+) were isolated from mice that carry a GFP reporter allele within the Prdm1 locus (Kallies et al., 2004), which were then activated in vitro to generate plasma cells. As expected, the average intra-chromosomal contact probability as a function of genomic distance revealed that the TCC reads for both naïve B and plasma cells decay as a function of genomic distance (Figure 2.2A). Notably, however, compared to follicular B cells, plasma cells were depleted for intra-chromosomal interactions separated more than 10 Mb but enriched for intra-chromosomal interactions that spanned less than 10 Mb (Figure 2.2A). Circos plots derived from contact maps associated with naïve follicular and plasma cells confirmed enrichment for intra-chromosomal genomic interactions that primarily involved euchromatic regions as denoted by PC1+ regions (Figure 2.2B).

Next, we constructed heatmaps for naïve and plasma cell genomes (Figure 2.2C). Heatmaps representing interaction frequencies for chromosome 2 revealed a striking loss of long-range genomic interactions in plasma cells when compared to naïve B cells (Figure 2.2C, upper panel). The differences were particularly pronounced in differential heatmaps constructed by subtracting contact frequencies in plasma cells versus follicular B cells (Figure 2.2C, right upper panel). We further calculated the average contact matrices for all chromosomes and observed this decontracted confirmation. (Figure 2.2C; lower panel).

Previous studies revealed that a wide spectrum of genomic regions switched compartments during the developmental transition from the pre-pro-B to the pro-B cell stage (Lin et al., 2012). To examine whether likewise coding and/or regulatory DNA elements reposition during plasma cell differentiation, we compared the PC1 values associated with genomic regions derived from TCC data for naïve B cells versus plasma cells. During the naïve to plasma cells transition, among the most significantly changed 5% regions, 1949 bins (50kb) lost euchromatic strength while 517 bins gained euchromatic strength, consistent with past findings that plasma cells undergo global gene silencing (Table S1; Shaffer et al., 2002). To more precisely compare compartments in naïve B cells and plasma cells, correlation difference (corrDiff) values were calculated. corrDiff directly measures the interaction profiles of a given locus between two experiments. As corrDiff approaches 1, two regions become more similar. To determine the compartment (A or B) associated with each corrDiff value, we compared the PC1 values of corrDiff low (dissimilar) regions. This analysis revealed that during the naïve to plasma cell transition, genomic regions associated with the Ebf1, Prdm1, gained euchromatic strength whereas regions associated with the Tcrb, Bcl6 and Bc1lla loci lost euchromatic strength (Table S1). Additionally, an ensemble of genomic regions, including genomic regions associated with cell cycle arrest in G1 such as Cdkn2a (p16INK4) and Cdkn2b (p15INK4b), also displayed changes in compartmentalization (Table S1). To determine how changes in compartmentalization relate to function a GO-analysis was performed. We found that a cluster of genes involved in antigen processing and presentation, lymphocyte cell fate and DNA repair gained euchromatic strength upon differentiating from follicular B to plasma cells (Figure 2.2D; left panel). In contrast, the switching of genomic regions from the euchromatic to the heterochromatic compartment was predominantly associated with genes involved in cell migration and chemotaxis (Figure 2.2D; right panel). Taken together, these data indicate that the onset of plasma development is closely associated with global repositioning of genes from the heterochromatic to the euchromatic compartment and vice versa.

Plasma Cell Differentiation is Concurrent with the Repositioning of Regulatory Elements

To examine how changes in compartmentalization relate to changes in gene expression, we plotted PC1 values of TSS in naïve B versus plasma cells and RNA abundance (Figure 2.3A). We found that a significant fraction of genomic regions that switched from compartment A to B and vice versa showed coordinate changes in transcript levels (Figure 2.3A). Next, we segregated genomic regions based on the correlation of HiC-derived interaction patterns between follicular B and plasma cells (Figure 2.3B). The majority of genomic regions that switched from the euchromatic to the heterochromatic compartment showed a modest change in transcriptional silencing while genomic regions that switched from heterochromatin to euchromatin were accompanied by substantial increases in transcription levels (Figure 2.3B and 1.3C). To further examine how during plasma cell differentiation expression relates to nuclear positioning we generated violin plots and contact maps for a selected group of genes (Figure 2.3D and 1.3E). Notably, we found that during the transition from the follicular B to plasma cell fate a selected group of genes, including Atf4, Ell2 and Prdm1 gained euchromatic strength (Figure 2.3E). The alterations in compartmentalization were closely associated with large-scale changes in chromatin folding across the Atf4, Ell2 and Prdm1 loci (Figure 2.3E). Taken together, these observations indicate that the onset of plasma cell development is concomitant with large-scale changes in chromatin folding and compartmentalization across genomic regions encoding for key developmental regulators.

During the Follicular B to Plasma Cell Fate Transition the Ebf1 and Bcl6 Loci Switch Compartments

Our earlier studies demonstrated that the Ebf1 locus is sequestered at the nuclear lamina in pre-pro-B cells but repositions to the nuclear interior in pro-B cells to facilitate Ebf1 gene expression (Lin et al. 2012). Since Ebf1 expression is silenced during the follicular B to plasma cell transition we hypothesized that the Ebf1 locus would reverse compartmentalization upon commitment to the plasma cell fate (Figure 2.S1).

Surprisingly, we found that in differentiating plasma cells a distally located intergenic region adjacent to the Ebf1 gene body repositioned from the A to the B compartment (Figure 2.4A). However, this localized change in compartmentalization was closely associated with large-scale alterations in chromatin folding across the intergenic region (Figure 2.4A). Since plasma cell fate is also concurrent with a decline in Bcl6 expression, we examined whether the Bcl6 genomic region also repositioned in differentiating plasma cells. Consistent with a decline in Bcl6 transcription we found that the onset of plasma cell development was associated with a significant decline in PC1 values for genomic region located immediately upstream (yellow bar) of the Bcl6 locus (Figure 2.4B). Again, this localized change in compartmentalization was accompanied by significantly decreased loss of genomic interactions across the Bcl6 locus (Figure 2.4B). Taken together, these data indicate that the silencing of Ebf1 and Bcl6 expression during plasma cell development is concomitant with changes in compartmentalization.

The Ebf1 Locus Repositions from the Euchromatic Compartment to Peri-centromeric Heterochromatin to Permit Plasma Cell Fate

To validate the repositioning of the Ebf1 intergenic region we used 3D-FISH. To mark distinct nuclear structures, we labeled three transcriptionally repressive compartments, □-satellite DNA, the nuclear lamina and nucleoli (Kieffer-Kwon et al., 2013; Bunting et al., 2016) (Figure 2.5A). Nucleoli were labeled using an antibody directed against nucleophosmin (NPM) (Figure 2.5A). As expected, we found that the development of plasma cells was closely associated with large-scale changes in nucleolar topology (Figure 2.5A). Notably, rather that returning to the nuclear lamina in plasma cells the Ebf1 locus relocated towards heterochromatic regions associated □-satellite DNA and nucleoli (Figure 2.5A and 1.5B). These data indicate that the Ebf1 locus repositions to peri-centromeric heterochromatin plausibly to permit developmental progression from the follicular B to plasma cell fate. To test the possibility that Ebf1 expression acts to enforce the follicular B cell state, we overexpressed Ebf1 in activated B cells. Specifically, we transduced activated follicular B cells that were derived from Prdm1-GFP mice. We measured plasma cell differentiation by Blimp-1-GFP expression 48 hours after infection. We found that Ebf1 overexpression interfered with plasma cell differentiation, decreasing the fraction of Blimp1+ cells by half in three-day cultures without interfering with cell cycle progression (Figure 2.5C). Taken together, these data indicate that Ebf1 expression enforces the follicular B cell checkpoint but upon activation repositions to a heterochromatic environment allowing developmental progression towards the plasma cell fate.

Plasma Cell Fate is Enriched for Inter-Chromosomal Associations

The data described above using RNA-FISH indicate that Xbp1, Atf4 and Prdm1 transcripts co-localize in a significant fraction of plasma cells (Figure 2.1D; Figure 2.S3A). We also found that in plasma cells versus follicular B cells, inter-chromosomal TCC reads were enriched for genomic regions that span the Prdm1, Xbp1 and Atf4 loci (Figure 2.S3B). These observations raise the possibility that inter-chromosomal associations are enriched for genes encoding for proteins associated with the plasma cell fate. To explore this possibility in more detail, we plotted the relative frequencies for inter-chromosomal associations for 1 Mb genomic regions as two-dimensional cloud plots. For both follicular

B and plasma cells, genomic regions enriched for inter-chromosomal association frequencies readily segregated into either compartment A or B (Figure 2.S3C, left and middle panels). Notably, we found that Xbp1 and Atf4, two genes encoding for proteins involved in the UPR pathway, clustered in the inter-chromosomal association cloud within the euchromatic region (PC1>0) in both naïve B and plasma cells (Figure 2.S3C, right panel). In contrast, the Prdm1 locus showed fewer significant associations with the Xbp1 and Atf4 loci, whereas in plasma cells the Prdm1 genomic region was enriched for genomic associations involving genes encoding for key components of the unfolded protein response (UPR) including the Xbp1 and Atf4 loci (Figure 2.S3C, right panel). To determine whether inter-chromosomal associations are concomitant with gene expression patterns we generated correlation matrices for clusters of genes associated with a B-lineage specific gene program (green) and genes involved in ribosome biogenesis (gray) (Figure 2.6A). For comparison, randomly chosen genes (not marked) were included in the analysis (Figure 2.6A). To identify inter-chromosomal associations, genes were grouped as even- or odd- associated chromosomes (Figure 2.6A). Both B-lineage and ribosome-associated gene clusters displayed expression patterns that correlated within the cluster (gray to gray; or green to green) but were anti-correlated between the two clusters (gray to green) (Figure 2.6A). As expected, randomly chosen genes were not correlated (Figure 1.6A). To compare correlated expression patterns to inter-chromosomal associations we generate a second matrix assembled for the same clusters of genes (Figure 1.6A). Notably, we found that enrichment for inter-chromosomal associations was associated with either coordinate or mutually exclusive patterns of gene expression (Figure 2.6A).

To determine whether global gene expression patterns correlated with genome-wide inter-chromosomal associations we generated scRNA-Seq correlation matrices for follicular B and plasma cells. HiC matrices were generated at 1 Mb resolution. To construct scRNA-Seq correlation matrices we analyzed all genes that span a 1 Mb region for all chromosomes and computed the maximum correlation coefficients for each bin (Figure 2.6B). For example, for genomic regions A and B we designated all the genes in region A as gene set GA and all the genes in region B as gene set GB. Next, we defined the correlation coefficients for all possible pairwise correlations between GA and GB and selected the highest absolute values as the scRNA correlation for genomic regions A and B. We repeated this process for all possible pairs and constructed a scRNA-Seq correlation matrix (Figure 2.6B). We then divided paired genomic regions into two categories (high and low) based on HiC reads. For each category, we plotted the distribution of scRNA-Seq correlation coefficients. We predicted that, if scRNA-Seq correlation coefficients were correlated with HiC intensity, a shift in the distribution would readily be observed when plotted as a function of normalized HiC reads (Figure 2.6B; right panel). We then applied this approach to the follicular B and plasma cell TCC data sets. Based on normalized TCC read numbers we categorized pairing into five groups. For both follicular B and plasma cells, we found that inter-chromosomal read numbers were concurrent with scRNA-Seq correlation coefficients (Figure 2.6C). As TCC reads increased between paired genomic regions, scRNA correlation coefficients either positively or negatively became more significant. This bimodal pattern suggests that inter-chromosomal associations not only promote coordinate patterns of gene expression but also facilitate coordinate transcriptional repression. Gene pairs of interest in plasma cells include Id3-Atf4, Prdm1-Xbp1, Pax5-Xbp1, and Tcf3-Xbp1 (Table S2). Taken together, these data indicate that inter-chromosomal proximity correlates with coordinate follicular B and plasma cell gene expression patterns.

2.4 Discussion

It is now well established that upon lineage commitment the genomes of adaptive and innate immune cells undergo large-scale alterations in genome topology including locus contraction, genome contraction and nuclear repositioning. Locus contraction involves the merging of variable regions associated with antigen receptor loci in adaptive immune cells (Jhunjhunwala et al., 2009). Genome contraction involves remote genomic interactions that shape the genomes of neutrophils and spermatozoa (Zhu et al., 2017; Battulin et al., 2015). Here we find that the transition from follicular B cells to plasma cells is accompanied with yet another folding pattern. Specifically, differentiating plasma cell progenitors undergo genome de-contraction as revealed by a decline in remote genomic interactions (>1Mb). Why may plasma cells display a loss of long-range genomic interactions? We suggest that this configuration suits a polarized nuclear morphology characteristic of plasma cells. Plasma cells may adopt an extended chromosome structure to juxtapose genes associated with antibody production within close proximity of the expanded endoplasmic reticulum (Park et al. 2014).

In addition to large-scale genome de-contraction, plasma cell fate is also accompanied with large-scale repositioning of intergenic regions. Notably, we found that genes encoding for factors that dictate germinal center versus plasma cell fate including Prdm1 gain euchromatic strength in differentiating plasma cell progenitors. In contrast, the euchromatic character of the Bcl6 and Ebf1 loci, concomitant with transcriptional silencing, declines upon exiting the follicular B cell compartment. The repositioning of the Ebf1 locus to heterochromatic regions is particularly interesting. To silence Ebf1 expression, the Ebf1 locus does not return to the lamina as observed in early progenitors but rather associates with the peri-centromeric heterochromatic compartment (Lin et al., 2012). The repositioning of genes to the peri-centromeric regions is not unique to the Ebf1 locus. Previous studies demonstrated that the Tcr□ and Ig□ locus reposition to the peri-centromeric regions during the process of VDJ locus rearrangement (Goldmit et al., 2005; Chan et al., 2013). These observations bring into question how genomic regions associate with distinct nuclear bodies, lamina versus peri-centromeric regions, during developmental progression. Previous studies have indicated that changes in compartmentalization is closely associated with alterations in transcription (Tumbar et al., 2001; Chuang et al., 2006). More recent observations revealed that non-coding transcription induced recruitment of cohesin and CTCF promotes elaborate changes in looping that orchestrate the relocation of regulatory elements from the lamina to the nuclear interior (Isoda et al., 2017). Hence the repositioning of genomic regions, from the lamina to the nuclear interior to nucleolar or centromeric regions, may very well involve loop extrusion to instruct nuclear positioning analogous to the Bcl11b locus. The elaborate changes in compartmentalization also raise questions as to why regulatory elements segregate into different compartments? We suggest that the nuclear location of regulatory elements in a heterochromatic environment, such as the lamina or in peri-nucleolar regions, prevents premature and/or stochastic activation of gene expression. This is particularly relevant for genes that dictate lineage specific gene programs such as Ebf1, Bcl11b and Prdm1.

In addition to changes in compartmentalization we found that in follicular B and plasma cells, gene expression and inter-chromosomal associations are closely linked. Inter-chromosomal contacts may arise only in a small fraction of cells within a given moment in time. Here, we suggest that during plasma cell differentiation, genes located on different chromosomes only actively associate during the act of transcription consistent with enriched co-localization of Prdm1, Xbp1 and Atf4 transcripts. Thus, we propose that the Prdm1, Xbp1 and Atf4 loci upon bursting assemble in transcription hubs to be coordinately regulated.

The intermingling of genes interspersed across different chromosomes is not unique to plasma cells. Prominent amongst these is the assembly of the nucleolus, an interaction hub of the rRNA genes located from five different human chromosomes (McStay et al., 2016). Notably, elegant previous studies demonstrated that hundreds of genes showed coregulated ribosomal RNA gene expression contributing to phenotypic variation (Li et al., 2018). Similarly, in neurons, the olfactory receptor genes assemble share an inter-chromosomal hub to coordinate gene expression (Lomvardas et al., 2006; Monahan et al., 2017). Regulatory elements associated with promoters and enhancers mediate pairing of alleles derived from homologous chromosomes (Hogan et al., 2015). The computational analysis of how inter-chromosomal associations relate to transcription revealed additional complexity. Inter-chromosomal associations either preceded co-expressing or mutually exclusive patterns of gene expression. A key factor for this phenomenon may involve the ratio of RNA-polymerase molecules to active promoter regions that share the transcription hub. If the RNA polymerase components are profuse, then all genes located within the hub can be activated simultaneously. In contrast, if only a small fraction of genes located in a hub acquire access to the RNA polymerase machinery, then transcription of genes involved becomes mutually exclusive. Other variables such as the activation state of the enhancers, enhancer abundance, transcription factor abundance, epigenetic modifications and phase separation may also play a role in establishing co-expressing versus mutually exclusive transcriptional hubs. It will be interesting to determine how inter-chromosomal interaction hubs are assembled and whether and how they contribute to establishing immune cell identity and beyond.

2.5 Conclusions

The transition from the follicular B to the plasma cell stage is orchestrated by an ensemble of transcriptional regulators. We found that gene activation at the onset of plasma cell development was concomitant with a gain in euchromatic character for an ensemble of genomic regions that dictate plasma cell fate, including the Prdm1 and Atf4 loci. We found that to permit the onset of plasma cell development the Ebf1 locus repositioned to peri-centromeric heterochromatin. Plasma cell differentiation was also associated with increased inter-chromosomal associations concurrent with alterations in gene expression and co-localization of Prdm1, Xbp1 and Atf4 transcripts in nuclear bodies. These data indicate that Ebf1 enforces the follicular B cell fate and that the onset of plasma cell fate is orchestrated by elaborate changes in compartmentalization and inter-chromosomal associations.

2.6 Availability of data and materials

Genome-wide reads were deposited at GEO. The raw and analyzed single cell RNA sequencing data generated in this study is available at GEO: GSE139565. The bulk RNA-seq, TCC and ATAC-seq data are available at GEO: GSE113014.

2.7 Figures

Figure 2.1. Unique gene expression profiles for in vitro and in vivo derived plasma cells. (A) TSNE plot of single cell RNA-seq dataset. Each dot is a single cell. Color of dot indicates the cell type. (B) Heatmap of manually selected genes. Each row is a cell. Each column is a gene. Red indicates higher transcriptional activity, while blue indicates lower transcriptional activity. Row-side color bar indicates the cell type, with color legend on the left. (C) Dot-plots of gene expression single cell RNA-seq reads, each dot is a single cell. The unit of the plots is UMI per million. Color of dot represents cell type. (D) Locations of probe regions used for RNA-FISH experiments. Example images of RNA-FISH for Prdm1 and Atf4 (top row) or Prdm1 and Xbp1 (bottom row) in day 3 sorted GFP+ plasma cells. (E) Heat maps displaying the number of cells that express different combinations of intronic RNA-FISH spots. Colors displaying the most prevalent (darkest) combination to rarest (lightest) combination. Chi-square test measuring the likelihood that the expression of Prdm1 and Xbp1 are independent of each other: p-value=1.21e-06, Pearson correlation=0.601 (left). Chi-square test measuring the likelihood that the expression of Prdm1 and Atf4 are independent of each other: p-value=1.32e-01, Pearson correlation=0.442 (right).

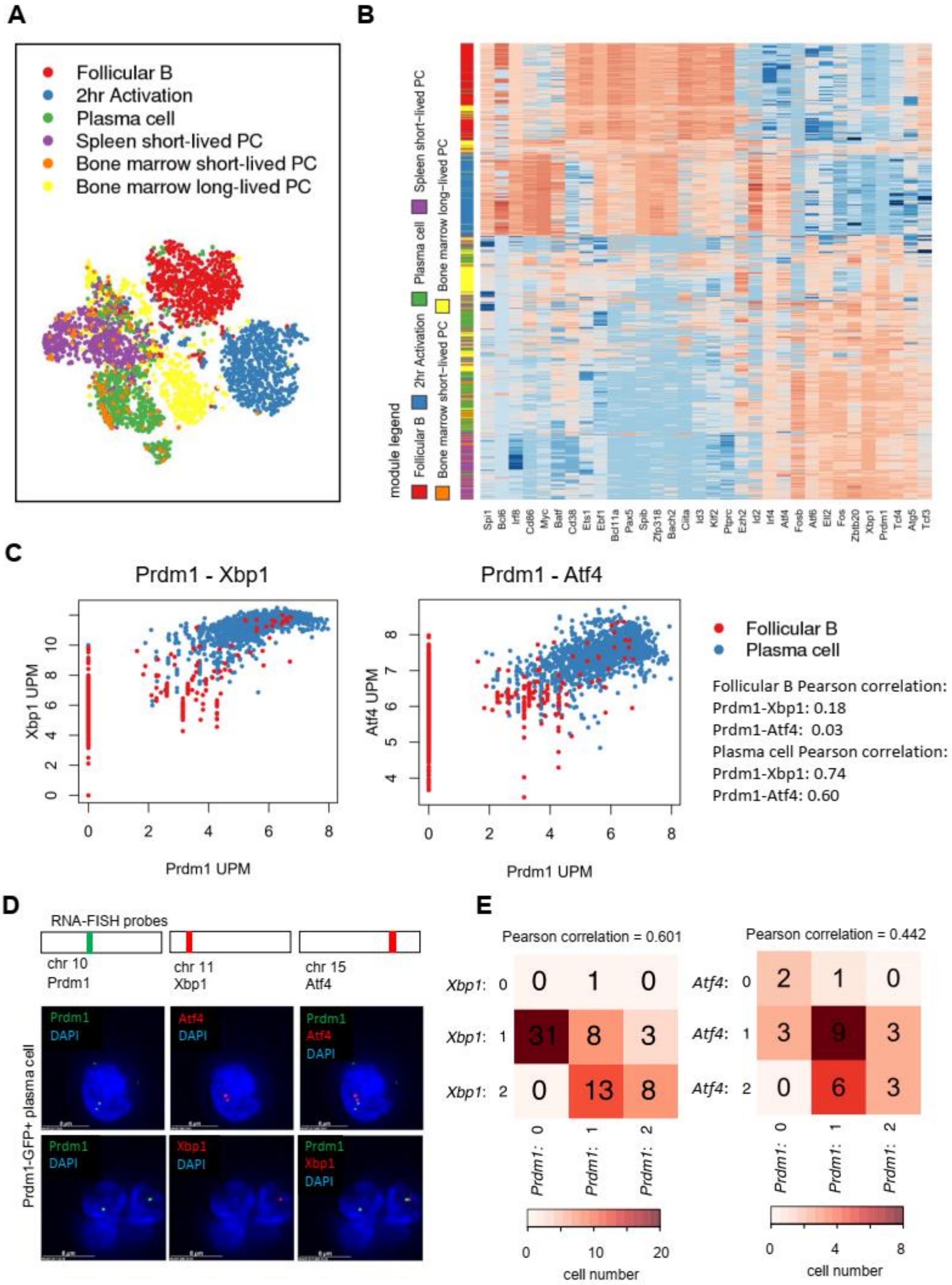

A
Follicular B
2hr Activation
Plasma cell
Spleen short-lived PC
Bone marrow short-lived PC
Bone marrow long-lived PC
B
module legend
Follicular B
2hr Activation
Plasma cell
Spleen short-lived PC
Bone marrow short-lived PC
Bone marrow long-lived PC
Spi1
Bcl6
Irf8
Cd86
Myc
Batf
Cd38
Ets1
Ebf1
Bcl11a
Pax5
Spib
Zfp318
Bach2
Ciita
Id3
Klf2
Ptprc
Ezh2
Id2
Irf4
Atf4
Fosb
Atf6
Elf2
Fos
Zbtb20
Xbp1
Prdm1
Tcf4
Atg5
Tcf3
C
Prdm1 - Xbp1
Xbp1 UPM
Prdm1 UPM
0
2
4
6
8
10
Prdm1 - Atf4
Atf4 UPM
Prdm1 UPM
4
5
6
7
8
Follicular B
Plasma cell
Follicular B Pearson correlation:
Prdm1-Xbp1: 0.18
Prdm1-Atf4: 0.03
Plasma cell Pearson correlation:
Prdm1-Xbp1: 0.74
Prdm1-Atf4: 0.60
D
RNA-FISH probes
chr 10
Prdm1
chr 11
Xbp1
chr 15
Atf4
Prdm1-GFP+ plasma cell
Prdm1
DAPI
Atf4
DAPI
Prdm1
Atf4
DAPI
Prdm1
DAPI
Xbp1
DAPI
Prdm1
Xbp1
DAPI
E
Pearson correlation = 0.601
Xbp1: 0
Xbp1: 1
Xbp1: 2
0
1
0
31
8
3
0
13
8
Prdm1: 0
Prdm1: 1
Prdm1: 2
0
10
20
cell number
Pearson correlation = 0.442
Atf4: 0
Atf4: 1
Atf4: 2
2
1
0
3
9
3
0
6
3
Prdm1: 0
Prdm1: 1
Prdm1: 2
0
4
8
cell number

Figure 2.2. Plasma cell differentiation loses long range intra-chromosomal interactions and gain short and medium range intra-chromosomal interactions. (A) Intra-chromosomal interaction frequencies across the genome at 50 kb resolution for naïve B cells and plasma cells. (B) Circos plot of significant intra-chromosomal interactions within chromosome 2 for naïve B and plasma cells. Interactions with p-values less than 10-9 are shown. PC1, ATAC, and gene positions are shown. Bin size, 50 kb. Numbers at the margins indicate genomic position in megabases. (C) TCC interaction matrix for chromosome 2 (top) and for all the chromosomes (bottom) at 100 kb resolution for naïve B cells and plasma cells. The right panels indicate differential interaction matrices. Red represents enrichment of contact frequencies, whereas purple represents depletion of normalized interaction frequencies in plasma cells. (D) GO-term bar-plots for genes in 50 kb regions genome-wide that have significantly higher PC1 values in plasma cells than in naïve B cells (left plot), and for genes have lower PC1 values in plasma cells than in naïve B cells (right plot).

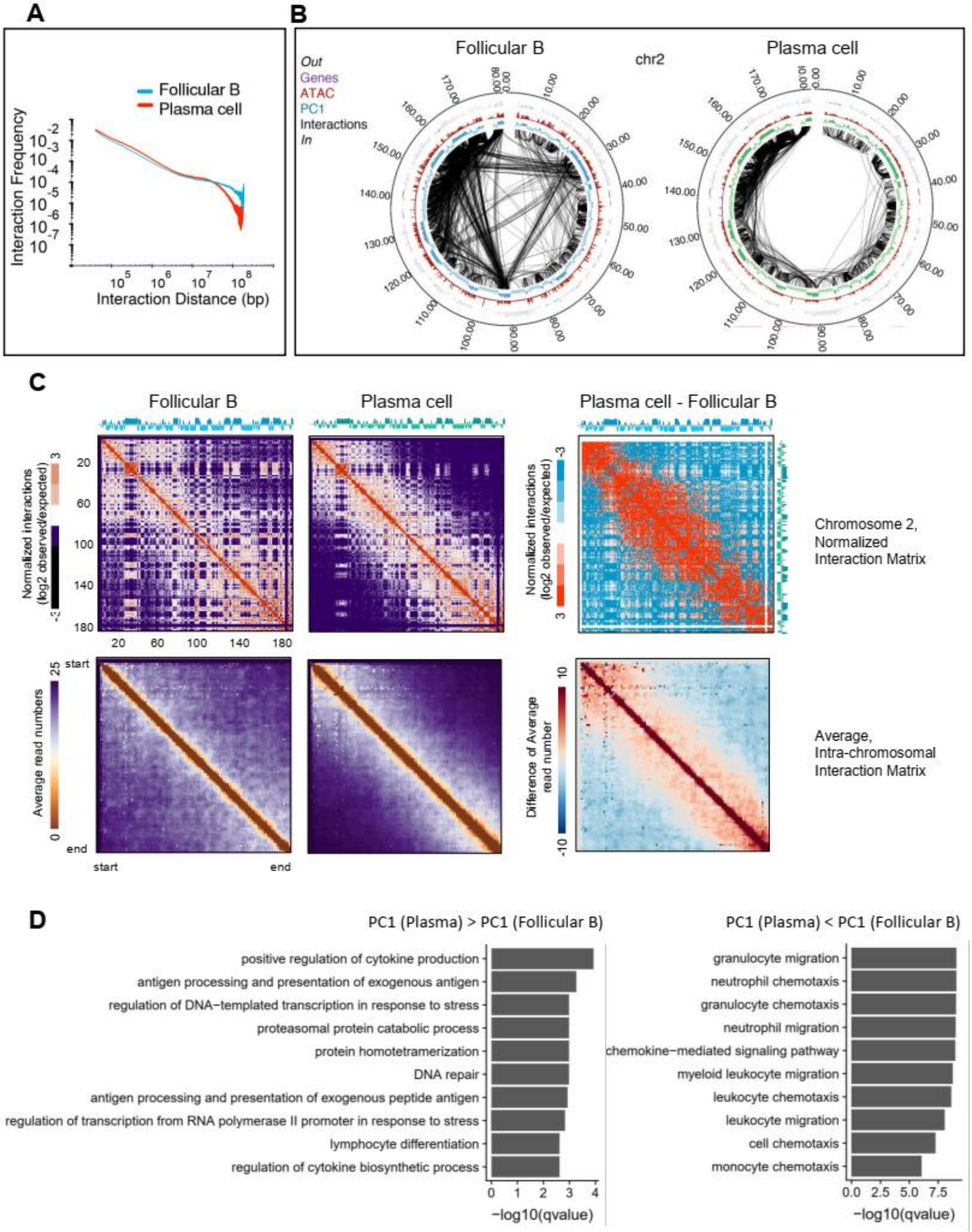

A
Follicular B
Plasma cell
Interaction Frequency
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{5}
10^{6}
10^{7}
10^{8}
Interaction Distance (bp)
B
Follicular B
chr2
Plasma cell
Out
Genes
ATAC
PC1
Interactions
In
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
110.00
120.00
130.00
140.00
150.00
160.00
170.00
C
Follicular B
Plasma cell
Plasma cell - Follicular B
Normalized interactions (log2 observed/expected)
3
-3
20
60
100
140
180
Chromosome 2, Normalized Interaction Matrix
Average read numbers
25
0
start
end
Difference of Average read number
10
-10
Average, Intra-chromosomal Interaction Matrix
D
PC1 (Plasma) > PC1 (Follicular B)
positive regulation of cytokine production
antigen processing and presentation of exogenous antigen
regulation of DNA−templated transcription in response to stress
proteasomal protein catabolic process
protein homotetramerization
DNA repair
antigen processing and presentation of exogenous peptide antigen
regulation of transcription from RNA polymerase II promoter in response to stress
lymphocyte differentiation
regulation of cytokine biosynthetic process
0
1
2
3
4
−log10(qvalue)
PC1 (Plasma) < PC1 (Follicular B)
granulocyte migration
neutrophil chemotaxis
granulocyte chemotaxis
neutrophil migration
chemokine−mediated signaling pathway
myeloid leukocyte migration
leukocyte chemotaxis
leukocyte migration
cell chemotaxis
monocyte chemotaxis
0.0
2.5
5.0
7.5
−log10(qvalue)

Figure 2.3. Changes in folding pattern correlate with changes in gene expression. (A) PC1 comparison between naïve B cells and plasma cells overlaying RNA-seq expression. Each spot represents a gene. PC1 values are calculated as the average PC1 of the 50 kb bins overlapping with the gene. RNA-seq change was calculated as the fold TPM change of the genes in plasma cells and naïve B cells. A 5-fold TPM increase is marked as upregulated (red), a 5-fold decrease is marked as downregulated (blue); otherwise, unchanged (gray). (B) Change in PC1 values as it relates to the correlation difference (corrDiff). Each line represents the distribution of PC1 differences between plasma cell and naïve B of the regions satisfying the criteria indicated. For example, "top 1%" denotes the regions with the top 1% lowest (most dissimilar) corrDiff values. All background means all regions in the genome were considered. (C) Change in gene expression between plasma cells and naïve B cells, shown as the ratio of the log2 of TPM, as it relates to corrDiff. (D) Single cell RNA-seq violin plot for Atf4, Ell2 and Prdm1. X-axis shows cell type. Y-axis indicates UMI per million. (E) TCC contact map for 3Mb regions surrounding Atf4, Ell2 and Prdm1. The position of the three genes and tracks of PC1 are below. Red indicates stronger interaction and blue indicates weaker interaction. Right most column indicates the difference between plasma cells and follicular B cells.

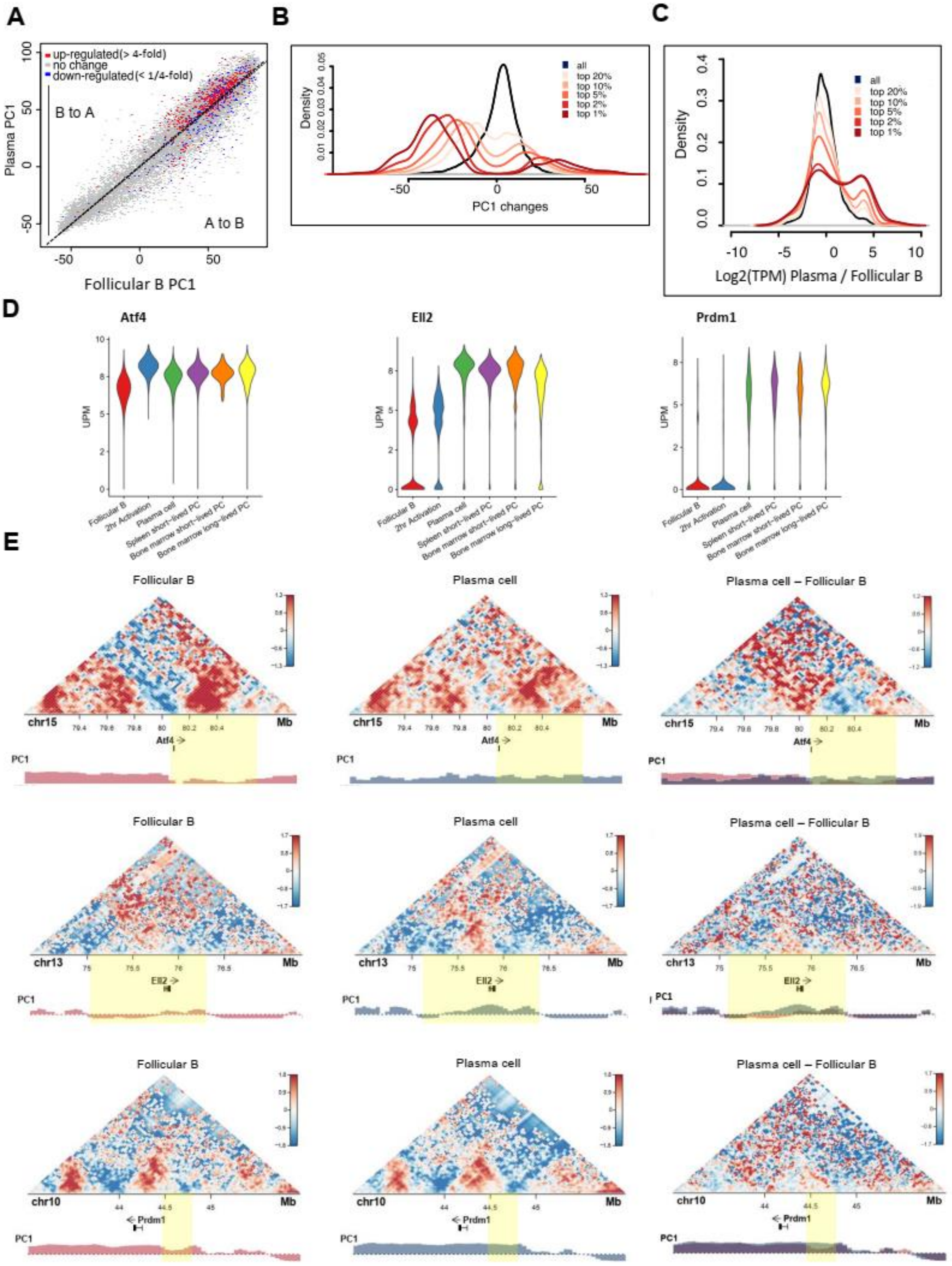
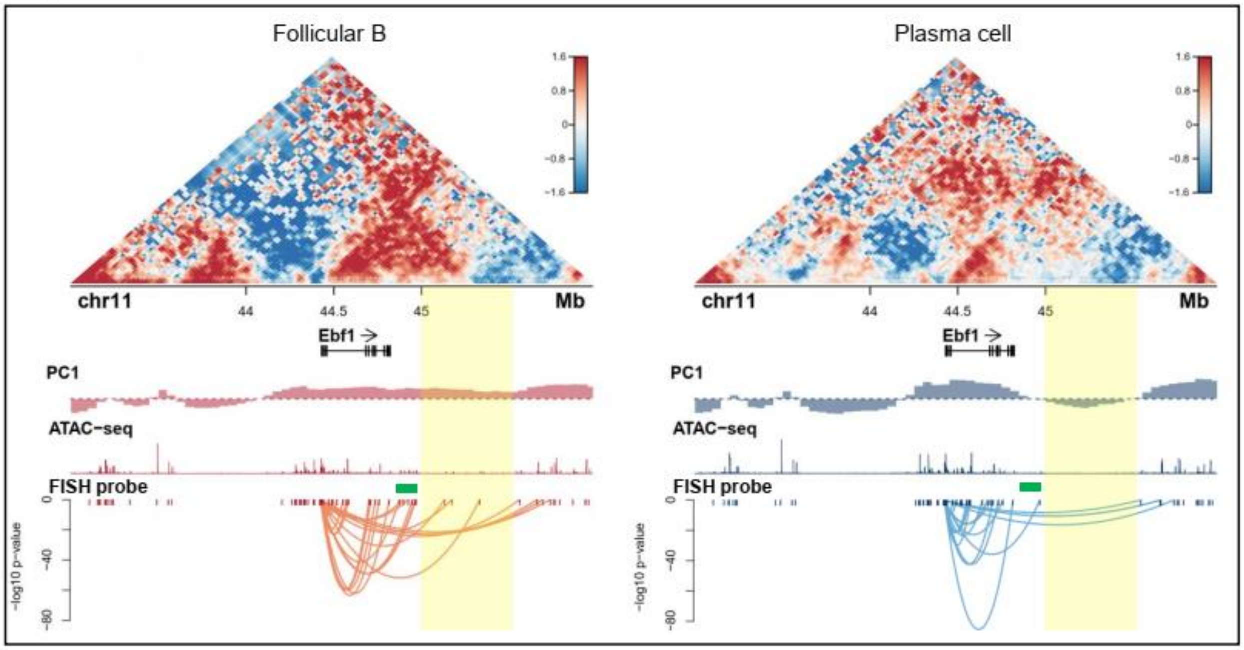

A
up-regulated(> 4-fold)
no change
down-regulated(< 1/4-fold)
B to A
A to B
Plasma PC1
Follicular B PC1
B
Density
PC1 changes
all
top 20%
top 10%
top 5%
top 2%
top 1%
C
Density
Log2(TPM) Plasma / Follicular B
D
Atf4
Ell2
Prdm1
UPM
Follicular B
2hr Activation
Plasma cell
Spleen short-lived PC
Bone marrow short-lived PC
Bone marrow long-lived PC
E
Follicular B
Plasma cell
Plasma cell – Follicular B
chr15
chr13
chr10
Mb
Atf4
Ell2
Prdm1
PC1

Figure 2.4. Plasma cells lose promoter-enhancer interaction at Ebf1 and Bcl6. (A) Contact map for chromosome 11 from 43 to 46 Mb at 50 kb resolution. Gene tracks from Refseq. Sushi plot showing the interaction between Ebf1 promoter and the ATAC peaks in target region. IGV tracks of ATAC-seq and location of Ebf1 DNA-FISH probes (green) for naïve B cells (left) or plasma cells (right). (B) Contact map for chromosome 16 from 22.5 to 25.5 Mb at 50 kb resolution. Gene tracks from Refseq. Sushi plot showing the interaction between Bcl6 intergenic region and the ATAC peaks in target region. IGV tracks of ATAC-seq for follicular B cells (left) or plasma cells (right).

A


Follicular B
Plasma cell
1.6
0.8
0
-0.8
-1.6
chr11
44
44.5
45
Mb
Ebf1
PC1
ATAC-seq
FISH probe
-log10 p-value
0
-40
-80

B

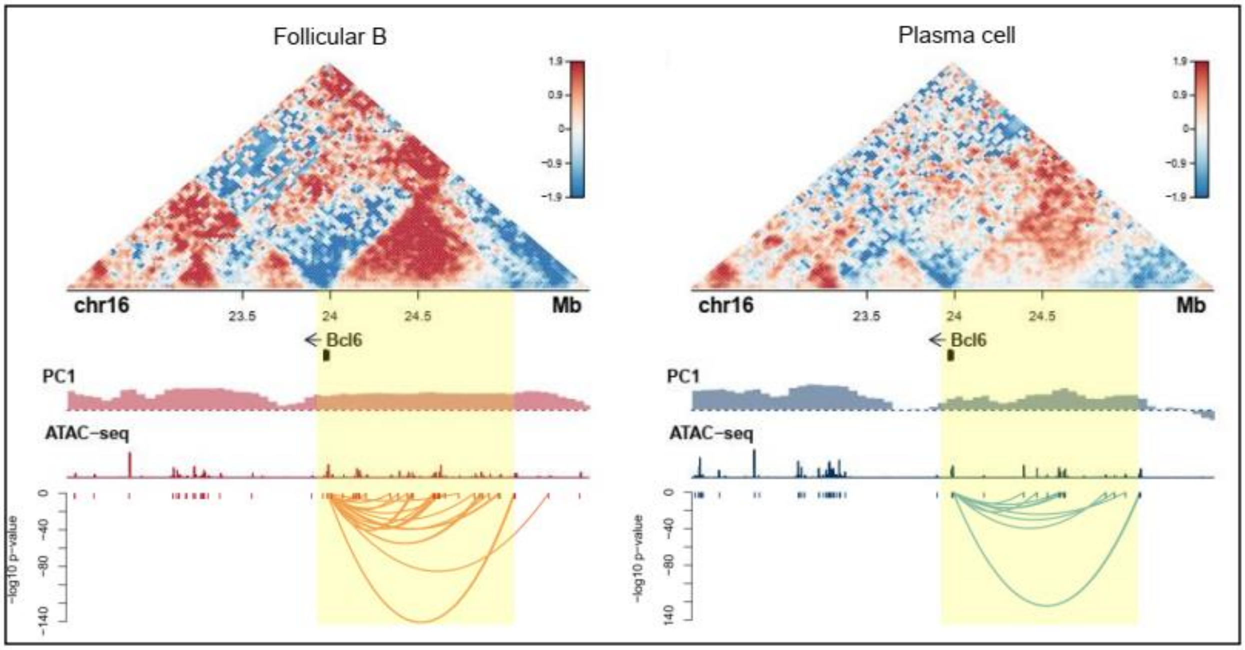

Follicular B
Plasma cell
1.9
0.9
0
-0.9
-1.9
chr16
23.5
24
24.5
Mb
Bcl6
PC1
ATAC-seq
-log10 p-value
0
-40
-80
-140

Figure 2.5. Ebf1 repositions away from the nuclear center towards repressive nuclear structures. (A) Example images from immunofluorescence for Ɣ-satellite (green) combined with DNA-FISH for the Ebf1 locus (red) (top panel). Example images from immunofluorescence for LaminB1 (red) combined with DNA-FISH for the Ebf1 locus (green) (middle panel). Example images from immunofluorescence for NPM (red) combined with DNA-FISH for the Ebf1 locus (green) (bottom panel). Ɣ-satellite were used to label peri-nucleolar heterochromatin (“chromocenters”), LaminB1 were used to label lamin, NPM foci were used to demarcate nucleoli (red) and DAPI was used to demarcate nuclear DNA (blue). (B) Dot plots for distance between Ebf1 locus and Ɣ-satellite, LaminB1 and NPM foci. (C) Representative flow plots of uninfected, empty vector, and EBF1-overexpressing activated B cells 48 hours after infection.

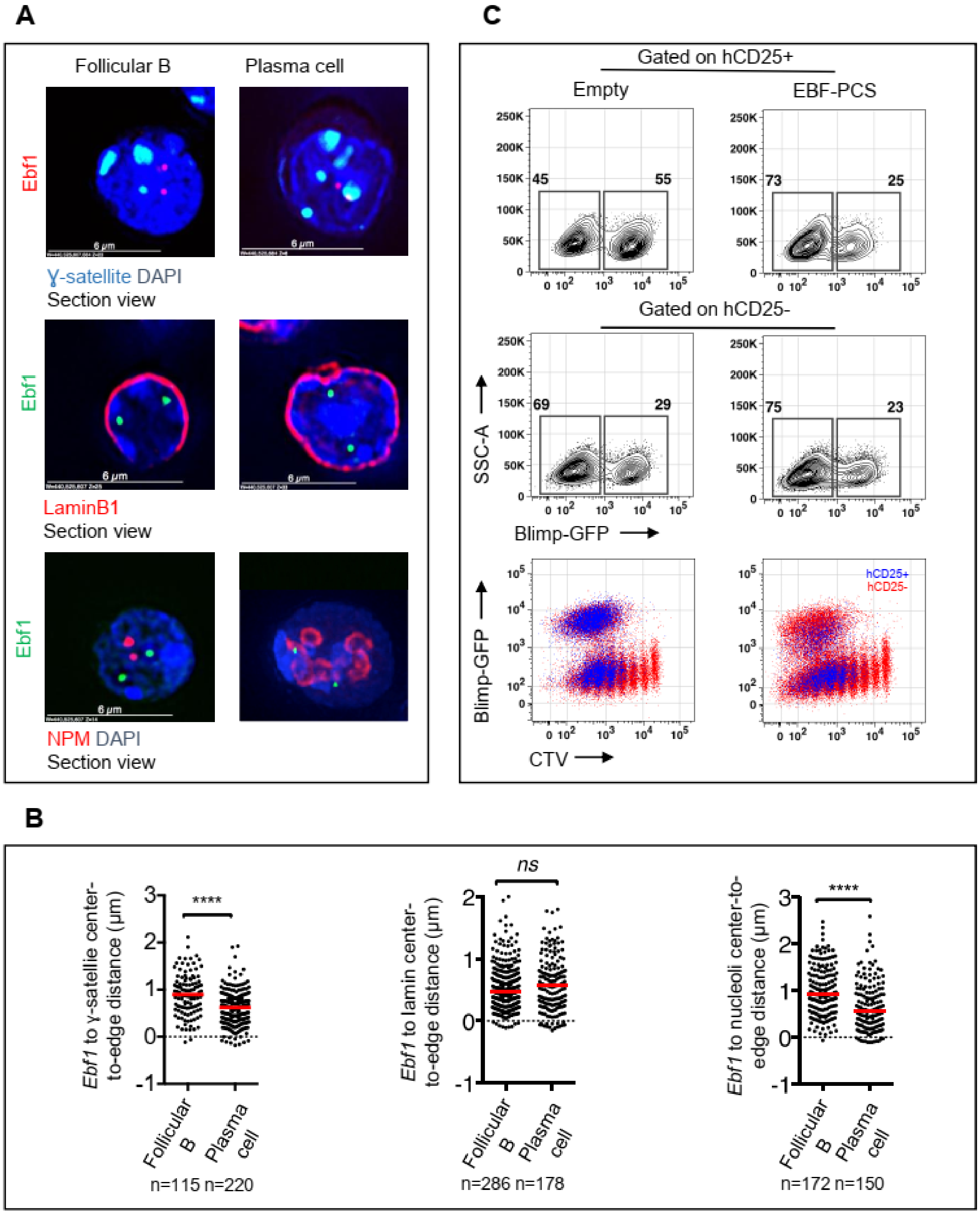

A
Follicular B
Plasma cell
Ebf1
ɣ-satellite DAPI
Section view
Ebf1
LaminB1
Section view
Ebf1
NPM DAPI
Section view
6 μm
C
Gated on hCD25+
Empty
EBF-PCS
45
55
73
25
Gated on hCD25-
69
29
75
23
SSC-A
Blimp-GFP
250K
200K
150K
100K
50K
0
0
10^2
10^3
10^4
10^5
Blimp-GFP
CTV
hCD25+
hCD25-
B
Ebf1 to γ-satellite center-to-edge distance (μm)
****
Follicular B
Plasma cell
n=115 n=220
Ebf1 to lamin center-to-edge distance (μm)
ns
n=286 n=178
Ebf1 to nucleoli center-to-edge distance (μm)
****
n=172 n=150
3
2
1
0
-1

Figure 2.6. Coordinated gene expression correlates with inter-chromosomal associations. (A) Heatmap for the transcriptional correlation between selected genes on different chromosomes in naïve B cells (left). Red indicates positive transcriptional correlation, blue indicates negative transcriptional correlation, while white marks no transcriptional correlation. Heatmap for the interchromosomal TCC read number (right). The intensity of green indicates the number of TCC reads with dark green equals more reads. 1mb bins centered on the TSS of each gene were used to derive the TCC matrix. The ordering of rows and columns are the same for both plots. (B) cartoon outlines how correlation analysis was done. The two matrices on the left are examples of TCC matrix and scRNA correlation matrix. We first selected all the 1mb inter-chromosomal region-pairs that have enriched TCC reads and plotted the distribution of corresponding gene expression correlation as red line on the right. Then we selected all the inter-chromosomal region-pairs that have depleted TCC reads and plotted the distribution of corresponding gene expression correlation as gray line on the right. If TCC reads is correlated with gene expression correlation, then the red and gray lines will significantly separate from each other as shown on the right. (C) Distribution of scRNA correlation between inter-chromosomal region pairs. Each line indicates a subset of region-pairs of specific normalized TCC read number.

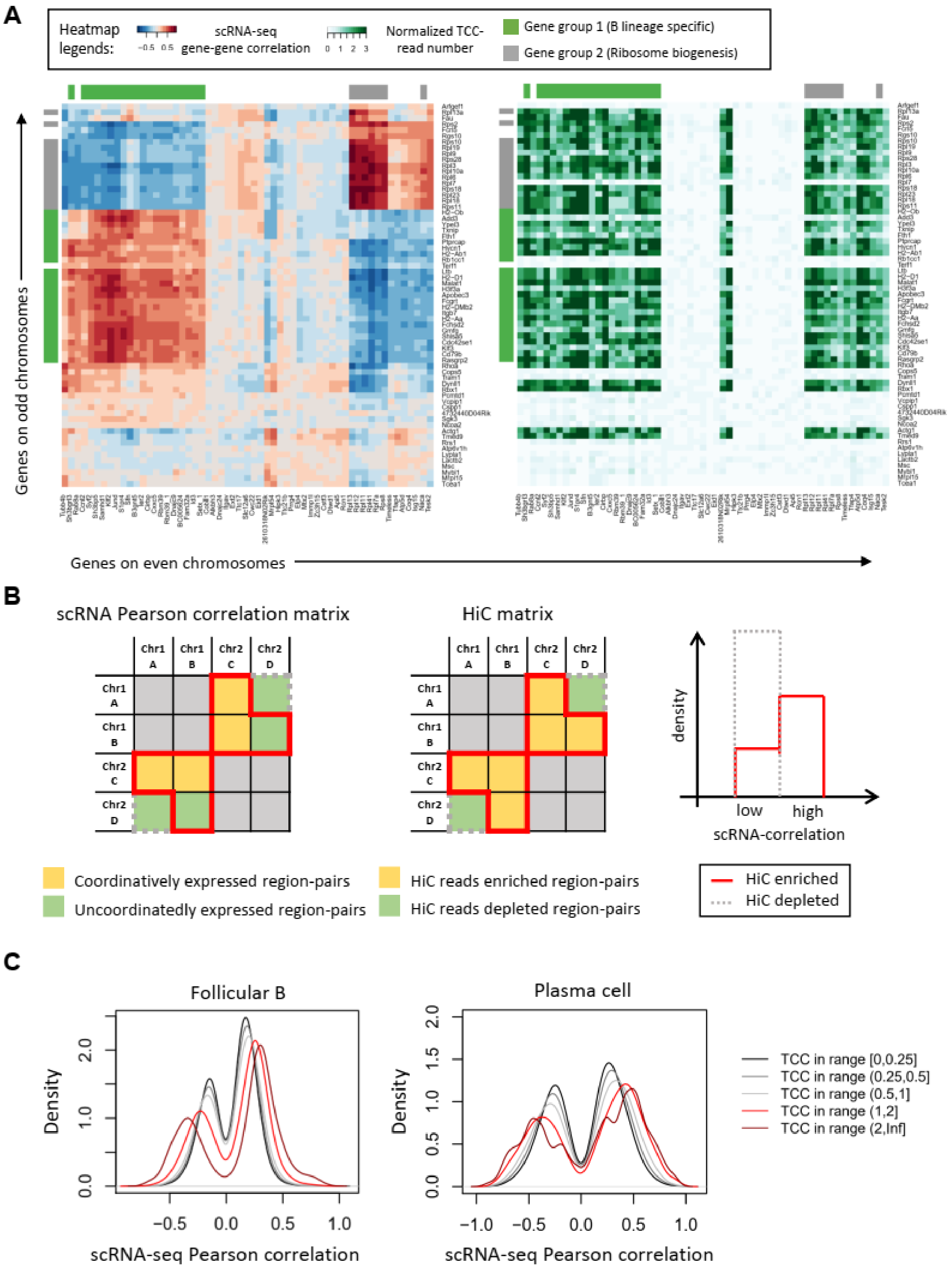

A
Heatmap legends:
-0.5 0.5
scRNA-seq gene-gene correlation
0 1 2 3
Normalized TCC-read number
Gene group 1 (B lineage specific)
Gene group 2 (Ribosome biogenesis)
Genes on odd chromosomes
Genes on even chromosomes
B
scRNA Pearson correlation matrix
HiC matrix
Chr1 A
Chr1 B
Chr2 C
Chr2 D
density
low
high
scRNA-correlation
Coordinatively expressed region-pairs
Uncoordinatedly expressed region-pairs
HiC reads enriched region-pairs
HiC reads depleted region-pairs
HiC enriched
HiC depleted
C
Follicular B
Plasma cell
Density
scRNA-seq Pearson correlation
0.0
1.0
2.0
0.5
1.5
-1.0
-0.5
TCC in range [0,0.25]
TCC in range (0.25,0.5]
TCC in range (0.5,1]
TCC in range (1,2]
TCC in range (2,Inf]

Figure 2.S1. Single cell RNA-seq closely overlaps with bulk RNA-seq. (A) Heatmap of genes expression for each sample. Single cells were merged to get pseudo bulk sample. Each row is a sample. Each column is a gene. Gene expression is measured as Z-normalized UPM. Only differentially expressed genes between naïve B and 3d activated plasma cells are shown. (B) Spearman correlation between bulk RNA-seq and pseudo-bulk RNA-seq generated by merging scRNA-seq. (C) Top panel: Bar-plot of bulk RNA-seq for Atf4, Bcl6, Ebf1, Id2, id, Irf3, Irf4, Pax5, Prdm1 and Xbp1. Bottom panel: Violin plot of single cell RNA-seq Atf4, Bcl6, Ebf1, Id2, id, Irf3, Irf4, Pax5, Prdm1 and Xbp1.

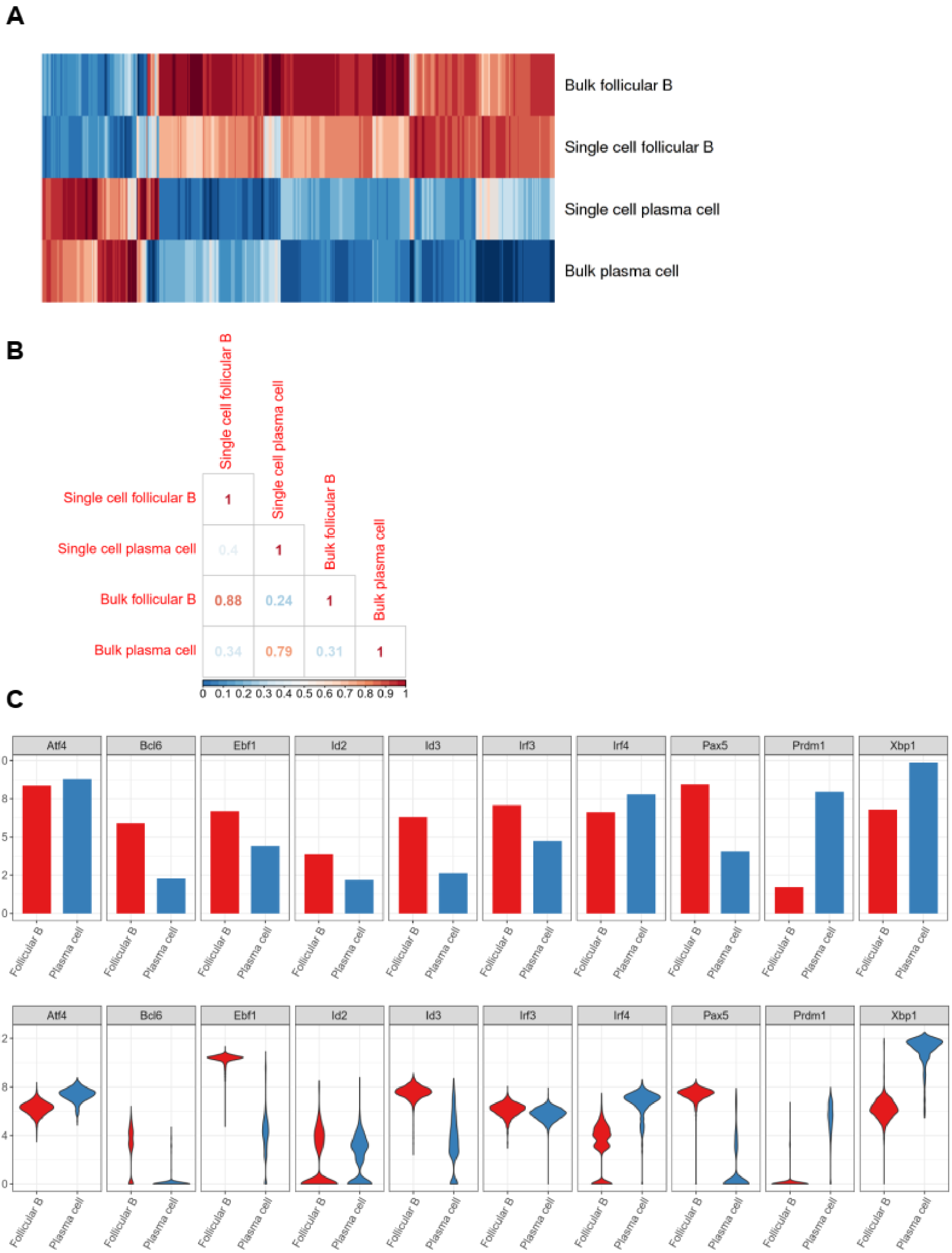

A
Bulk follicular B
Single cell follicular B
Single cell plasma cell
Bulk plasma cell
B
Single cell follicular B
Single cell plasma cell
Bulk follicular B
Bulk plasma cell
1
0.4
1
0.88
0.24
1
0.34
0.79
0.31
1
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
C
Atf4
Bcl6
Ebf1
Id2
Id3
Irf3
Irf4
Pax5
Prdm1
Xbp1
TPM
0
2
5
8
10
Follicular B
Plasma cell
UPM
0
4
8
12

Figure 2.S2. Gene expression and GO analysis on genes of interest. (A) TSNE plot of single cell RNA-seq. Each dot is a single cell. The first TSNE plot is labeled to show the origin of each cell. Color in the rest of the plots indicates the expression level of corresponding gene. Expression level unit is UPM. (B) GO term bar-plot for DE gene between naïve B cell and 3d in-vitro activated plasma cell (left). GO term bar-plot for DE gene between bone marrow B220- plasma cell and bone marrow B220+ plasma cell (right).

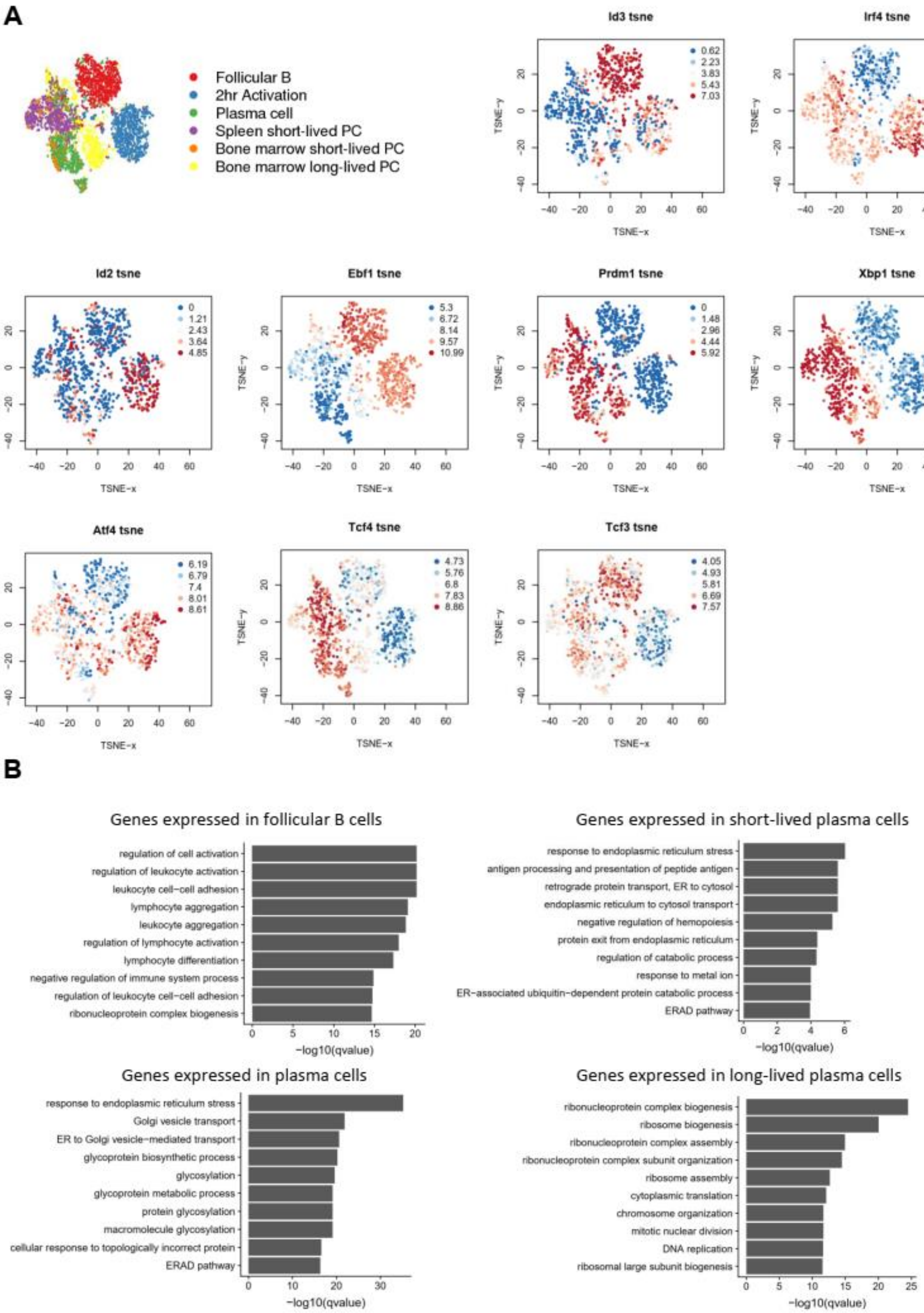
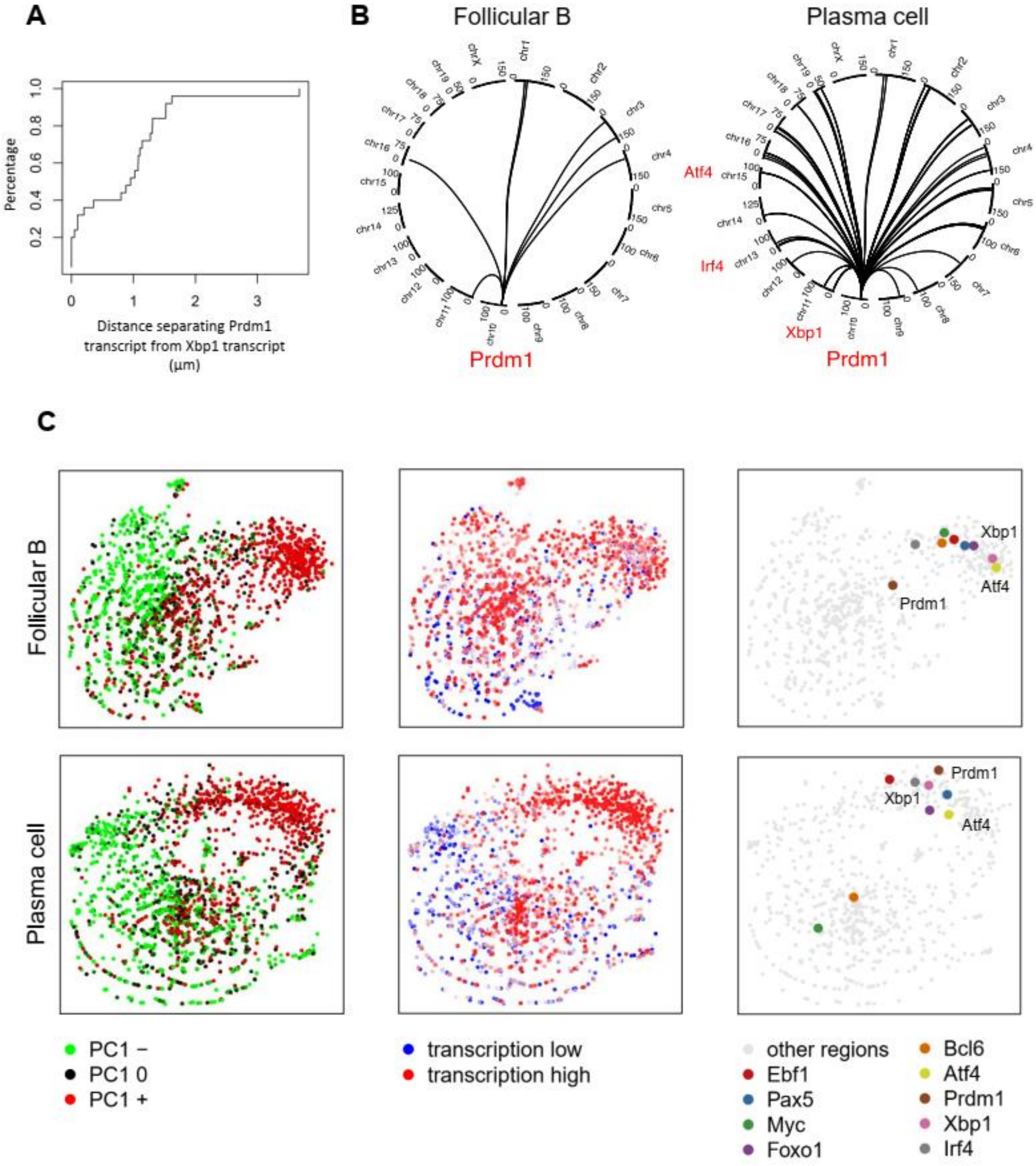
A
Follicular B
2hr Activation
Plasma cell
Spleen short-lived PC
Bone marrow short-lived PC
Bone marrow long-lived PC
Id3 tsne
0.62
2.23
3.83
5.43
7.03
Irf4 tsne
3.23
4.58
5.92
7.27
8.61
Id2 tsne
0
1.21
2.43
3.64
4.85
Ebf1 tsne
5.3
6.72
8.14
9.57
10.99
Prdm1 tsne
0
1.48
2.96
4.44
5.92
Xbp1 tsne
5.5
7.01
8.51
10.02
11.53
Atf4 tsne
6.19
6.79
7.4
8.01
8.61
Tcf4 tsne
4.73
5.76
6.8
7.83
8.86
Tcf3 tsne
4.05
4.93
5.81
6.69
7.57
TSNE-y
TSNE-x
-40
-20
0
20
40
60
B
Genes expressed in follicular B cells
regulation of cell activation
regulation of leukocyte activation
leukocyte cell−cell adhesion
lymphocyte aggregation
leukocyte aggregation
regulation of lymphocyte activation
lymphocyte differentiation
negative regulation of immune system process
regulation of leukocyte cell−cell adhesion
ribonucleoprotein complex biogenesis
0
5
10
15
20
−log10(qvalue)
Genes expressed in short-lived plasma cells
response to endoplasmic reticulum stress
antigen processing and presentation of peptide antigen
retrograde protein transport, ER to cytosol
endoplasmic reticulum to cytosol transport
negative regulation of hemopoiesis
protein exit from endoplasmic reticulum
regulation of catabolic process
response to metal ion
ER−associated ubiquitin−dependent protein catabolic process
ERAD pathway
0
2
4
6
−log10(qvalue)
Genes expressed in plasma cells
response to endoplasmic reticulum stress
Golgi vesicle transport
ER to Golgi vesicle−mediated transport
glycoprotein biosynthetic process
glycosylation
glycoprotein metabolic process
protein glycosylation
macromolecule glycosylation
cellular response to topologically incorrect protein
ERAD pathway
0
10
20
30
−log10(qvalue)
Genes expressed in long-lived plasma cells
ribonucleoprotein complex biogenesis
ribosome biogenesis
ribonucleoprotein complex assembly
ribonucleoprotein complex subunit organization
ribosome assembly
cytoplasmic translation
chromosome organization
mitotic nuclear division
DNA replication
ribosomal large subunit biogenesis
0
5
10
15
20
25
−log10(qvalue)

Figure 2.S3. Interchromosomal association for Prdm1. (A) Cumulative distribution of distance between Prdm1 and Xbp1 intronic RNA-FISH dots in 3d activated plasma cells. (B) Circos plots for inter-chromosomal association centered on Prdm1. Genes contained within the 1 Mb associated regions and that appear in the inter-chromosomal cloud plot are indicated in red. (C) Inter-chromosomal association cloud plots showing the relative inter-chromosomal association preference of all genome regions (computed based on inter-chromosomal TCC reads). Each dot represents a 1 Mb region of the genome. The closer two dots in 2D distance, the more likely they share significant inter-chromosomal association. Dots are colored by PC1 value (left), transcriptional activity (middle) and genes of interest (right).


A
Percentage
0.2
0.4
0.6
0.8
1.0
0
1
2
3
Distance separating Prdm1 transcript from Xbp1 transcript (µm)
B
Follicular B
Plasma cell
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chrX
Prdm1
Atf4
Irf4
Xbp1
Prdm1
C
Follicular B
Plasma cell
Xbp1
Atf4
Prdm1
Prdm1
Xbp1
Atf4
● PC1 −
● PC1 0
● PC1 +
● transcription low
● transcription high
● other regions
● Ebf1
● Pax5
● Myc
● Foxo1
● Bcl6
● Atf4
● Prdm1
● Xbp1
● Irf4

2.8 Methods

Flow cytometry

Single-cell suspensions of splenocytes were prepared, depleted of red blood cells by hypotonic lysis, and stained with optimal dilutions of the indicated antibodies. All of the following reagents were obtained from eBioscience: anti-CD4 (RM4-5), anti-CD8a (53-6.7), anti-Gr-1 (RB6-8C5), anti-F4/80 (BM8), and anti-TER119; anti-IgD (11-26); anti-B220 (RA3-6B2); anti-CD19 (1D3). Doublets were excluded using the combined width and height parameters of the forward and side scatter parameters. Flow cytometric acquisition was performed on a BD LSRII, and analyses were performed using FlowJo 10.1r5 (Tree Star). Cells were sorted with a three-laser FACsAria Fusion.

Cell culture

Splenocytes were incubated with CD23-biotin (eBioscience, clone B3B4, cat. 13-0232-81) and purified using anti-biotin microbeads (Miltenyi Biotec, cat. 130-090-485) on MACs LS columns. Sorted B cells were cultured for 3 days in RPMI 1640 medium supplemented with 10% FCS, antibiotics, 2mM L-glutamine, and □–mercaptoethanol (50 □M) at 37C, 5% CO2, before analysis by flow cytometry. LPS (Escherichia coli) was purchased from Sigma-Aldrich; L2654-1MG and used at 10 □g/mL. For cell division experiments, cells were labeled with CellTrace Violet reagent (Invitrogen, Molecular Probes). Dead cells were excluded using DAPI (Invitrogen, Molecular Probes).

DNA-FISH

Cells were placed on poly-L-lysine coated coverslips for 30 minutes in 37ºC incubator. fixed with paraformaldehyde at the final concentration of 4% for 10 minutes. The 200 kb bacterial artificial chromosome (BAC) probe RP23-118P17 and 40 kb fosmids was obtained from the BACPAC Resource Center (BPRC) at Children's Hospital Oakland Research Institute. Probes were labeled by nick translation using Alexa-488 dUTP, Alexa-568 dUTP, or Alexa-647 dUTP (Invitrogen). Incubated overnight at 37ºC. The following morning, coverslips were washed twice in 50% formamide/2x SSC at 37ºC for 30 minutes on a shaking incubator at 300 rpm. Coverslips were then rinsed in 1x PBS containing DAPI. Coverslips were rinsed once more with 1x PBS.

DNA FISH probe details:

Region of interest	**Chromosome**	**Start (bp)**	**End (bP)**	**BAC-PAC Identifier**
EBF1 promoter	Chromosome 11	44416074	44457525	WIBR1-0109P11
EBF1 distal enhancer	Chromosome 11	45305517	45347678	WIBR1-0350
Ebf1 super-enhancer	Chromosome 11	44941681	45143731	RP23-118P17

Immunofluorescence

Coverslips were permeabilized with 0.1% saponin/0.1% Triton-X in PBS at room temperature for 10 minutes, then incubated for 20 minutes at room temperature with 20% glycerol in 1x PBS. Slides were submerged in liquid nitrogen three times, rinsed once with 1x PBS, and then blocked for 30 minutes with 5% BSA/0.1% Triton-X in 1x PBS for 30 minutes at 37ºC. The nucleolus was stained first with primary antibodies to B23 (Abcam) for 30 minutes in a humidified chamber at 37ºC. Coverslips were washed twice, 10 minutes each, in 0.1% Triton-X in 1x PBS at room temperate on a shaker at 70 rpm.

Secondary staining with performed using donkey antibody to mouse IgG conjugated to Alexa-568 (Invitrogen) for 30 minutes at 37ºC in 5% BSA/0.1% Triton-X and 5% serum in 1x PBS. Coverslips were washed twice for 10 minutes each in 1x PBS/0.1% Triton-X at room temperature with gentle agitation (70 rpm). The nuclear lamina was stained first with primary antibody to Lamin B1 (sc-6217; Santa Cruz Biotechnology), followed by secondary staining.

Single molecule RNA-FISH

Pools of fluorescently labeled oligonucleotide probes per RNA were designed and purchased from LGC Biosearch Technologies using the Stellaris Probe Designer. For intronic Prdm1 and Xbp1, default settings were used. For intronic Atf4, masking level was set to 3, oligo length was set to 18, and nucleotide spacing was set to 1. Cells were fixed and permeabilized as described for DNA-FISH, but were stored overnight at 4C in 70% ethanol rather than PBS. The next day, coverslips were rinsed in wash buffer (10% formamide in 2x SSC) to remove ethanol, then incubated in wash buffer for 5 minutes at room temperature. Coverslips were incubated with probe of interest in hybridization buffer overnight in a 37C humidified hybridization oven. The next morning, coverslips were rinsed in washed buffer and washed twice for 30 minutes each in 37ºC temperature-controller shaker at 100 rpm. During the second wash, DAPI was added to the wash buffer. Slides were rinsed once in 2x SSC, once in 1x PBS, and then mounted with Prolong Gold Anti-Fade reagent. Slides were dried in the dark for at least 6 hours prior to sealing with nail polish and imaging.

RNA FISH probe details

Region of interest	**Region**	**Chromosome**	**Position (bp)**	**# probes**	**Fluorophore**
Intronic Prdm1	Fourth intron	10	44161971-44166597	48	Quasar 670
Intronic Xbp1	First two introns	11	5421204-5421925; 5422023-5423335	36	Cal Fluor 610
Intronic Atf4	All introns (2)	15	80086116-80086636; 80086952-80087063	22	Cal Fluor 610

Imaging

Three-dimensional fluorescent images were acquired using an Applied Precision Inverted Deconvolution Deltavision Microscope with a 100x objective (Nikon 100x/1.40 oil or Olympus 100x/1.4 oil). Optical sections (z-stacks) 0.2µm apart were obtained throughout the cell volume in the DAPI, FITC, Red, and Cy5 channels. Deconvolution was performed using Softworx Software version 5.0 with the following settings: 10 cycles of enhanced ratio (aggressive), camera intensity offset = 50, normalize intensity and apply correction options selected.

Calculating distances between DNA loci

The nuclei of individual cells were identified by DAPI staining, and cells containing two spots per DNA-FISH or intronic RNA-FISH channel were verified manually. Images were cropped to contain single cells. Analysis of cells in three dimensions was performed using the FIJI plugin Tools for Analysis of Nuclear Genome Organisation (TANGO) version 0.94. Distance between two DNA-FISH spots was calculated after segmentation in TANGO by running the “Distance” measurement between center-of-mass of the two loci. Distance between nuclear structures and a DNA-FISH spot was calculated as the minimum distance between the center-of- mass of the DNA-FISH spot to the edge of the corresponding structure. Negative values indicate inclusion of the FISH spot in the nuclear structure.

Overexpression

The GFP cassette of the LMP vector was replaced with hCD25 cassette for EBF1 overexpression. Retroviral supernatant was obtained through 293T transfection by the calcium phosphate method with these constructs in conjunction with the packaging plasmid pCL-Eco. CD23+ cells were MACs selected and stimulated in vitro for 48 hours prior to infection. Cells were harvested 48 hours after infection for analysis by flow cytometry.

Tethered Chromatin Capture (TCC) sample preparation

TCC was performed as previously described (Kalhor et. al., 2012) using 5 million naïve B cells or sorted plasma cells with the following modifications.

Cell lysis and chromatin biotinylation. Cell suspensions were not treated with a Dounce homogenizer. After washing, cells were resuspended in 500 μl of wash buffer and mixed with 190 μl of 2% SDS and incubated at 65C for 10 minutes to solubilize the crosslinked chromatin. Nuclei were spun at RT at 3500 rpm for 2 minutes. Supernatant was discarded and nuclei were resuspended in 245 μl wash buffer. Once the suspension cooled to RT, it was mixed with 105 μl of 25 mM EZlink Iodoacetyl-PEG2-Biotin (IPB) (Thermo Scientific) and rocked at RT for 75 minutes to biotinylate the cysteine residues.

The sample was then mixed with 650 ul of 1xNEBuffer2 and incubated on ice for 5 minutes.

Digestion and dialysis. To start digestion, 85 □l of NEBuffer2, 3 □l 1M DTT, 227 □l H2O, and 35 □l of 25 U/□l Mbo1 (NEB) were incubated O/N at 37C.

Removal of biotin from non-ligated DNA ends and shearing. To shear DNA fragments, the sample size was adjusted to 120 ul by adding water. The sample was transferred to a 6x16mm AFA fiber microtube with snap-cap and sheared to 100-500bp in a Covaris E220 S2 (Biogem) at duty cycle of 5%, intensity of 5 (=175W in E220), 200 cycles/burst, for a total of 180 seconds.

Pull-down of biointinylated DNA and ligation of sequencing adaptors. 15 ul of MyOne Streptavidin C1 beads (Invitrogen) were used. After washing beads with TE, reaction was resuspended in 30 ul of 0.033% Tween20/TE.

PCR amplification. Libraries were amplified using Phusion HotStart II for experimentally determined numbers of PCR cycles with the following program: 98ºC /30 seconds, x-cycles [98ºC/10 seconds, 65ºC for 30 seconds, 72ºC for 20 seconds], 72ºC for 5 minutes

Size selection. The entire PCR product was loaded onto a polyacrylamide gel and run at 200V for 70 minutes. The gel was stained with SybrGold (Invitrogen) and visualized on Gel Doc 1000. Fragments between 300 and 600 bp were excised and purified by spinning the gel through a homemade shredder for 5 minutes at 14,000 rpm at RT. DNA was extracted by incubated sample in 400 ul EB (Qiagen) O/N at 37C shaking at 1400rpm. DNA was ethanol precipitated.

Sequencing. DNA was paired-end sequenced with Illumina HiSeq 2000.

RNA sequencing

A total of 10,000-100,000 B cell subsets were sorted into Buffer RLT (QIAGEN) and RNA was isolated. Total RNA was isolated using RNeasy Mini kit (Qiagen) with on column DNase treatment. RNA was again treated with TURBO DNase (Life Technologies) and mRNA was purified with Dynabeads mRNA purification kit (Life Technologies). First-strand synthesis kit (Life Technologies) in presence of actinomycin D using a combination of random hexamers and oligo(dT). Second-strand synthesis was performed with dUTP instead of dTTP. The ds-cDNA was sonicated to 150-400 bp using the Covaris sonicator. Sonicated cDNA was ligated to adaptors. The resulting DNA was treated with uracil-N-glycosylase prior to PCR amplification with the indexing primers. Following PCR, fragments were size-selected and sequenced on HiSeq 2500.

ATAC sequencing

ATAC-seq was performed as previously described (Buenrostro, J. D. et al, 2013). 50,000 cells were used for library preparation. Cells were treated with transposition mix for 30 minutes at 37ºC. DNA was purified with DNA Clean & Concentrator (Zymo Research). Library fragments were amplified by PCR using 1x NEBnext PCR master mix (NEB) and 1.25 uM of custom Nextera PCR primers 1 and 2 as follows: 72ºC for 5 min., 98ºC for 30 seconds, 98ºC for 10 seconds, 63ºC for 30 seconds and 72ºC for 1 minute. Amplified libraries were selected using SPRI beads. The quality of the library was measured using Agilent TapeStation and the final product was sequenced on an Illumina

HiSeq 4000. Reads were mapped to mouse genome build mm9 using TopHat and converted to track plots by IGV.

Single cell RNA sequencing

Approximately 10,000 cells per sample were loaded into Single Cell A chips (10X Genomics) and partitioned into Gel Bead In-Emulsions (GEMs) in a Chromium Controller (10X Genomics). Single cell RNA libraries were prepared according to the 10x Genomics Chromium Single Cell 3’ Reagent Kits v2 User Guide, and sequenced on a HiSeq 4000.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical test utilization and p-value reporting are presented in the text, figures, and legends.

Analysis of TCC data

TCC reads were mapped by bowtie and processed by HOMER to create tag directory with default setting55. Read numbers of naïve and plasma cells were down-sampled by HOMER, using getRandomReads.pl by reducing the total number of reads to match the smaller sample size (naïve B cells). Inter-chromosomal associations were called by HOMER using 1 Mb bins and a threshold p=10-4 with the other options set as default. Intra-chromosomal interactions were called by HOMER using 25 kb and 50 kb bins, threshold p=10-4 and max distance 3 Mb with the other options set as default. PC1 and corrDiff were calculated by HOMER with 50 kb resolutions. TCC matrices, normalized or raw, were generated by HOMER with default settings using 1 Mb bins for the whole genome and 50 kb bins for individual chromosomes. The raw matrices were normalized by KR-normalization (Rao et al., 2014).

Compartments. To quantify switched domains, the following criteria were used: regions with PC1 change more than 25, regardless of sign, were considered switched. Bin size was 50 kb. Consecutive regions were merged. Two regions separated by only one bin that did not pass threshold were also merged.

corrDiff. As explained by HOMER{Heinz:2010ef}: “PCA analysis is a useful tool for analyzing single TCC experiments. However, comparing PC1 values between two experiments is a little dangerous. PCA is an unbiased analysis of data, and while PC1 values usually correlated with "active" vs. "inactive" compartments, the precise qualitative nature of this association may differ slightly between experiments. Because of this, it is recommended to directly compare the interaction profiles between experiments rather than simply looking at PC1 values. One way to do this is to directly correlate the interaction profile of a locus in one experiment to the interaction profile of that same locus in another experiment. If the locus tends to interact with similar regions in both TCC experiments, the correlation will be high. If the locus interacts with different regions in the two experiments, the correlation will be low. Note that this is NOT PCA analysis, but a direct comparison between two experiments at each locus. By default, this will compare the interaction profiles across each chromosome (inter-chromosomal interactions are generally too sparse for this type of analysis). Supplementary table 1 was generated by the 5% 50 kb bins that have lowest corrDiff value.

Average intra-chromosomal contact maps. The average intra-chromosomal TCC contact maps for naïve B cells and plasma cells (Fig. 2) were generated as follows:

1) calculate the 100 kb resolution raw TCC contact maps for each chromosome (chr1-19 and X, in total 20 matrices); 2) resize each contact matrix to have 100 rows and 100 columns; and, 3) calculate the average contact intensity of all chromosomes.

Analysis of bulk RNA-seq

Single-end sequencing reads were mapped to mouse genome build mm9 using TopHat and analyzed using the Cufflink-cuffdiff pipeline. Transcript quantification was performed using Kallisto with default settings. TPM was used in cross-sample comparison. Fold change was computed by calculating fold changes between TPM+1.

Analysis of scRNA-seq

Reads from single-cell RNA-seq were aligned to mm10 and collapsed into unique molecular identifier (UMI) counts using the 10X Genomics Cell Ranger software (version 2.1.0). Raw cell-reads were then loaded to R using the cellrangerRkit package. The scRNA-seq dataset was then further filtered based on gene numbers and mitochondria gene counts to total counts ratio. Only cells with > 300 genes UMI > 0, and 0.05% ~ 10% of their UMIs mapping to mitochondria genes were kept for downstream analysis. If a sample have more than 2000 cells detected, then only 2000 cells were randomly selected to downstream analysis. Such sampling guaranteed the cell number from each sample were relatively balanced. Only genes that had UMI > 0 in more than 2% of cells of at least one sample were kept for downstream analysis. UMI counts were transformed to UMI per million (UPM) for downstream analysis. To fix the drop-out issue of single cell RNA-seq, we calculated average expression for every 20 neighboring cells for downstream analysis except for TSNE and differential expression analysis. The corrected UPM matched much better with the RNA-FISH (Fig 1C,D,E) than the raw UPM. Differential expression analysis was done based on the original UPM because Wilcox test was used which would not be biased by the drop-out issue. DE gene threshold was set to Bonferroni corrected q-value < 0.05 and (UPM+1) fold change > 2.

Correlation analysis between HiC and scRNA-seq

ScRNA correlation and TCC/HiC matrices at 1Mb resolution were used for the correlation analysis. To construct a scRNA correlation matrix we analyzed all genes within the 1 Mb region across all chromosomes and computed the maximum correlation of gene expression for each bin. For example, for genomic regions A and B we designated all the genes in region A as gene set GA and all the genes in region B as gene set GB. Next, we defined the correlation for all possible pairwise correlations between GA and GB and selected the highest absolute value as the scRNA correlation for A and B genomic regions. We repeated this process for all possible pairs and constructed a scRNA correlation matrix. This method filtered out most of noise introduced by genes that were not expressed and that were expressed independent of any other genes. For TCC/HiC matrix, KR normalized was applied with sparse limit set to 0.95. For all matrices, only bins in even chromosome were kept in rows and only bins in odd chromosome were kept in column, thus only interchromosomal part of the original matrix was considered. We then divided paired genomic regions into multiple categories based on HiC reads intensity. For each category, we plotted the distribution of scRNA correlation coefficients for the region-pairs. If scRNA correlation coefficients were correlated with HiC intensity a shift in the distribution would readily be observed between different categories. In this analysis, genes that have multiple copies in the genome were ignored. Supplementary table 2 was generated by listing all gene pairs that have > 0.5 absolute scRNA correlation and have > 1 normalized TCC reads between the two 1Mb locus.

2.9 Acknowledgments

Chapter 2, in full, has been submitted for publication of the material as it may appear in Cell Reports, 2020. “Elaborate and Intricate Changes in Genome Compartmentalization and Inter-Chromosomal Hubs Orchestrate the Plasma Cell Fate”. Alexandra Bortnick, Zhaoren He, Megan Aubrey, Vivek Chandra, Matthew Denholtz, Kenian Chen, Yin C Lin and Cornelis Murre. The dissertation author was the primary investigator and author of this paper.

2.10 References

Battulin, N., Fishman, V.S., Mazur, A.M., Pomaznoy, M., Khabarova, A.A., Afonnikov, D.A., Prokhortchouk, E.B., and Serov, O.L. (2015). Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. Genome Biol. 16, 77.

Bettigole, S. E., and Glimcher, L. H. (2015). Endoplasmic reticulum stress in immunity. Annual Rev. Immunol. 33, 107-138.

Bunting, K.L., Soong, T.D., Singh, R., Jiang, Y., Béguelin, W., Poloway, D.W., Swed, B.L., Hatzi, K., Reisacher, W., Teater, M. and Elemento, O. (2016). Multi-tiered reorganization of the genome during B cell affinity maturation anchored by a germinal center-specific locus control region. Immunity 45, 497-512.

Caron, G., Hussein, M., Kulis, M., Delaloy, C., Chatonnet, F., Pignarre, A., Avner, S., Lemarié, M., Mahé, E.A., Verdaguer-Dot, N., and Queirós, A.C. (2015). Cell-cycle-dependent reconfiguration of the DNA methylome during terminal differentiation of human B cells into plasma cells. Cell Rep. 13, 1059-1071.

Chan, E.A., Teng, G., Corbett, E., Choudhury, K.R., Bassing, C.H., Schatz, D.G. and Krangel, M.S. (2013). Peripheral subnuclear positioning suppressed Tcrb recombination and segregates Tcrb alleles from RAG2. Proc. Natl. Acad. Sci. USA 110, 4628-4637.

Chen, S., Miyazaki, M., Chandra, V., Fisch, K.M., Chang, A.N., and Murre, C. (2016). Id3 orchestrates germinal center B cell development. Mol. Cell. Biol., 36, 2543-2552.

Chernova, I., D. D. Jones, J. R. Wilmore, A. Bortnick, M. Yucel, U. Hershberg, and D. Allman. (2014). Lasting antibody responses are mediated by a combination of newly formed and established bone marrow plasma cells drawn from clonally distinct precursors. J. Immunol. 193, 4971–4979.

Chesi, M., and Bergsagel, P. L. (2013). Molecular pathogenesis of multiple myeloma: basic and clinical updates. Int. J. Hematol. 97, 313-323.

Chuang, C.H., Carpenter, A.E., Fuchsova, B., Johnson, T., Lanerolle, P., and Belmont, A.S. (2006). Long-range directional movement of an interphase chromosome site. Curr. Biol. 16, 825-831.

Fowler, T., Garruss, A.S., Ghosh, A., De, S., Becker, K.G., Wood, W.H., Weirauch, M.T., Smale, S.T., Aronow, B., Sen, R., and Roy, A.L. (2015). Divergence of transcriptional landscape occurs early in B cell activation. Epigenetics Chromatin 8, 20.

Gass, J. N., Jiang, H. Y., Wek, R. C., and Brewer, J. W. (2008). The unfolded protein response of B-lymphocytes: PERK-independent development of antibody-secreting cells. Cell. Mol. Immunol. 45, 1035-1043.

Gloury, R., Zotos, D., Zuidscherwoude, M., Masson, F., Liao, Y., Hasbold, J., Corcoran, L.M., Hodgkin, P.D., Belz, G.T., Shi, W. and Nutt, S.L. (2016). Dynamic changes in Id3 and E-protein activity orchestrate germinal center and plasma cell development. J. Exp. Med. 213, 1095-1111.

Goldmit, M., Ji, Y., Roldan, E., Jung, S., Cedar, H., and Bergman, Y. (2005). Epigenetic ontogeny of the Ig□ locus during B cell development. Nat Immunol. 6, 198-203.

Halliley, J. L., C. M. Tipton, J. Liesveld, A. F. Rosenberg, J. Darce, I. V. Gregoretti, L. Popova, D. Kaminiski, C. F. Fucile, I. Albizua, et al. (2015). Long-lived plasma cells are contained within the CD19(-)CD38(hi)CD138(+) subset in human bone marrow. Immunity 43, 132–145.

Hammarlund, E., Thomas, A., Amanna, I.J., Holden, L.A., Slayden, O.D., Park, B., Gao, L. and Slifka, M.K. (2017). Plasma cell survival in the absence of B cell memory. Nat. Comm. 8, 1781.

Hiepe, F., Dörner, T., Hauser, A. E., Hoyer, B. F., Mei, H., Radbruch, A. (2011). Long-lived autoreactive plasma cells drive persistent autoimmune inflammation. Nat. Rev. Rheumatol. 7, 170-178.

Hogan, M.S., Parfitt, D.E., Zepeda-Mendoza, C.J., Shen, M.M., and Spector, D.L. (2015). Transient pairing of homologous Oct4 alleles accompanies the onset of embryonic stem cell differentiation. Cell Stem Cell 16, 275–288.

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

Huang, S., Deerinck, T. J., Ellisman, M. H., and Spector, D. L. (1997). The dynamic organization of the perinucleolar compartment in the cell nucleus. J. Cell Biol. 137, 965-974.

Isoda, T., Moore, A.J., He, Z., Chandra, V., Aida, M., Denholtz, M., van Hamburg, J.P., Fisch, K.M., Chang, A.N., Fahl, S.P., et. al. (2017). Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. Cell, 171, 103-119.

Jhunjhunwala, S., van Zelm, M., Peak, M. and Murre, C. (2009). Chromatin architecture and the generation of antigen receptor diversity. Cell 138, 435-448.

Johanson, T. M., Lun, A. T., Coughlan, H. D., Tan, T., Smyth, G. K., Nutt, S. L., and Allan, R. S. (2018). Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. Nat. Immunol., 19, 1257.

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat. Biotechnol., 30, 90-98.

Kallies, A., Hasbold, J., Tarlinton, D. M., Dietrich, W., Corcoran, L. M., Hodgkin, P. D., and Nutt, S. L. (2004). Plasma cell ontogeny defined by quantitative changes in blimp-1 expression. J. Exp. Med., 200, 967-977.

Kieffer-Kwon, K.R., Nimura, K., Rao, S.S., Xu, J., Jung, S., Pekowska, A., Dose, M., Stevens, E., Mathe, E., Dong, P. and Huang, S.C. (2017). Myc regulates chromatin decompaction and nuclear architecture during B cell activation. Mol. Cell 67, 566-578.

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

Li, B., Kremling, K.A., Wu, P., Bukowski, R., Romay, M.C., Xie, E., Buckler, E.S. and Chen, M., (2018). Coregulation of ribosomal RNA with hundreds of genes contributes to phenotypic variation. Genome Res. 28, 1555-1565.

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. and Sandstrom, R. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289-293.

Lin, K. I., Angelin-Duclos, C., Kuo, T. C., and Calame, K. (2002). Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to immunoglobulin M-secreting plasma cells. Mol. Cell. Bio. 22, 4771-4780.

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.

Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R. (2006). Interchromosomal interactions and olfactory receptor choice. Cell 126, 403–413.

Ma, Y., Shimizu, Y., Mann, M. J., Jin, Y., and Hendershot, L. M. (2010). Plasma cell differentiation initiates a limited ER stress response by specifically suppressing the PERK-dependent branch of the unfolded protein response. Cell Stress Chaperones 15, 281-293.

Maharana, S., Iyer, K. V., Jain, N., Nagarajan, M., Wang, Y., and Shivashankar, G. V. (2016). Chromosome intermingling—the physical basis of chromosome organization in differentiated cells. Nucleic Acids Res. 44, 5148-5160.

Masciarelli, S., Fra, A.M., Pengo, N., Bertolotti, M., Cenci, S., Fagioli, C., Ron, D., Hendershot, L.M. and Sitia, R. (2010). CHOP-independent apoptosis and pathway-selective induction of the UPR in developing plasma cells. Mol. Immunol. 47, 1356-1365.

McStay, B. (2016). Nucleolar organizer regions: genomic ‘dark matter’ requiring illumination. Genes Dev. 30, 1598–1610.

Monahan, K., et al. (2017). Cooperative interactions enable singular olfactory receptor expression in mouse olfactory neurons. eLife 6:e28620.

Park, S. K., Xiang, Y., Feng, X., and Garrard, W. T. (2014). Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. Genes Dev. 28, 1159-1164.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M. and Reinders, M. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol. Cell 38, 603-613.

Quong, M. W., Harris, D. P., Swain, S. L., and Murre, C. (1999). E2A activity is induced during B - cell activation to promote immunoglobulin class switch recombination. EMBO. J. 18, 6307-6318.

Sexton, T., and Cavalli, G. (2015). The role of chromosome domains in shaping the functional genome. Cell, 160, 1049-1059.

Shi, W., Liao, Y., Willis, S.N., Taubenheim, N., Inouye, M., Tarlinton, D.M., Smyth, G.K., Hodgkin, P.D., Nutt, S.L. and Corcoran, L.M. (2015). Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. Nat. Immunol. 16, 663.

Shaffer, A.L., Lin, K.I., Kuo, T.C., Yu, X., Hurt, E.M., Rosenwald, A., Giltnane, J.M., Yang, L., Zhao, H., Calame, K. and Staudt, L.M. (2002). Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. Immunity 17, 51-62.

Slifka, M. K., Antia, R., Whitmire, J. K., and Ahmed, R. (1998). Humoral immunity due to long-lived plasma cells. Immunity 8, 363-372.

Tellier, J., Shi, W., Minnich, M., Liao, Y., Crawford, S., Smyth, G.K., Kallies, A., Busslinger, M. and Nutt, S.L. (2016). Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response. Nat. Immunol. 17, 323.

Tumbar, T. and Belmont, A.S. (2001). Interphase movements of a DNA chromosome region modulated by VP16 transcriptional activator. Nat. Cell Biol. 3, 134-139.

Wöhner, M., Tagoh, H., Bilic, I., Jaritz, M., Poliakova, D. K., Fischer, M., and Busslinger, M. (2016). Molecular functions of the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell development. J. Exp. Med. 213, 1201-1221.

Chapter 3: Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T Cell fate

Article

Cell

Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate

Takeshi Isoda,[1] Amanda J. Moore,[1] Zhaoren He,[1] Vivek Chandra,[1] Masatoshi Aida,[1] Matthew Denholtz,[1]
Jan Piet van Hamburg,[1] Kathleen M. Fisch,[2] Aaron N. Chang,[2] Shawn P. Fahl,[3] David L. Wiest,[3] and Cornelis Murre[1,4,*]
[1]Department of Molecular Biology, University of California, San Diego, La Jolla, CA 92093, USA
[2]Center for Computational Biology & Bioinformatics, Institute for Genomic Medicine, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[3]Blood Cell Development and Function, Fox Chase Cancer Center, 333 Cottman Avenue, PA, Philadelphia, PA 19111, USA
[4]Lead Contact
*Correspondence: cmurre@ucsd.edu

http://dx.doi.org/10.1016/j.cell.2017.09.001

SUMMARY

It is now established that Bcl11b specifies T cell fate. Here, we show that in developing T cells, the Bcl11b enhancer repositioned from the lamina to the nuclear interior. Our search for factors that relocalized the Bcl11b enhancer identified a non-coding RNA named ThymoD (thymocyte differentiation factor). ThymoD-deficient mice displayed a block at the onset of T cell development and developed lymphoid malignancies. We found that ThymoD transcription promoted demethylation at CTCF bound sites and activated cohesin-dependent looping to reposition the Bcl11b enhancer from the lamina to the nuclear interior and to juxtapose the Bcl11b enhancer and promoter into a single-loop domain. These large-scale changes in nuclear architecture were associated with the deposition of activating epigenetic marks across the loop domain, plausibly facilitating phase separation. These data indicate how, during developmental progression and tumor suppression, non-coding transcription orchestrates chromatin folding and compartmentalization to direct with high precision enhancer-promoter communication.

INTRODUCTION

The differentiation of T cells is orchestrated in the thymus. Upon exposure to Delta-Notch signaling, early T cell progenitors (ETPs) differentiate into multipotent DN2a cells, which in turn develop into committed DN2b cells. DN2b cells subsequently progress into DN3a cells in which TCRβ VDJ rearrangement is initiated. Once a productive TCRβ chain has been assembled, DN3b cells expand and differentiate into CD4$^+$CD8$^+$ double-positive (DP) thymocytes. In the DP compartment, thymocytes die by either neglect or negative selection or persist through positive selection to differentiate into CD4 single-positive (CD4SP) or CD8SP cells (Klein et al., 2014; Naito et al., 2011).

The developmental progression of T cells is regulated by the combined activities of an ensemble of transcriptional regulators. T-lineage development is initiated by the E-proteins that activate the expression of genes encoding components involved in Notch signaling (Bain and Murre, 1998; Ikawa et al., 2006; Miyazaki et al., 2017). Once instructed to respond to Notch signaling, T cell progenitors activate the expression of Bcl11b, GATA-3, and TCF1 (Yui and Rothenberg, 2014). Specifically, Bcl11b expression is initiated at the DN2a cell stage to promote developmental progression to the DN2b cell stage. At the DN2b cell stage, Bcl11b expression is further elevated and, in concert with E2A, activates a T-lineage-specific program of gene expression and suppresses the expression of genes associated with alternative cell fates (Liu et al., 2010; Ikawa et al., 2010; Li et al., 2010a; Longabaugh et al., 2017). The activation of Bcl11b expression in DN2 cells involves Notch signaling, GATA-3, TCF1, and RUNX1 that bind to an enhancer, named Major Peak, located in the Bcl11b intergenic locus control region (Guo et al., 2008; Weber et al., 2011; García-Ojeda et al., 2013; Li et al., 2013). Recent elegant studies indicated that full activation of Bcl11b expression in developing T cell progenitors requires a rate-limiting transition from an inactive to an active chromatin state (Kueh et al., 2016).

Here, we have examined how Bcl11b expression is activated to establish T cell fate and suppress the development of lymphoid malignancies. We found that, in developing T cell progenitors, the Bcl11b locus control region, containing a well-characterized enhancer, repositioned from the lamina to the nuclear interior. The repositioning of the Bcl11b enhancer was orchestrated by a non-coding RNA, named ThymoD (thymocyte differentiation factor). ThymoD transcription promoted demethylation at sites associated with CTCF occupancy across the transcribed region and activated cohesin-dependent looping, plausibly involving loop extrusion, to bring the Bcl11b promoter and enhancer into a single loop domain. These results are consistent with a model in which non-coding transcription dictates enhancer-promoter communication at multiple levels: (1) demethylation of CpG residues across the ThymoD transcribed region to permit CTCF occupancy, (2) recruitment of the cohesin complex to the transcribed region to activate cohesin-dependent looping, (3) loop extrusion to juxtapose with great precision the enhancer and promoter into a

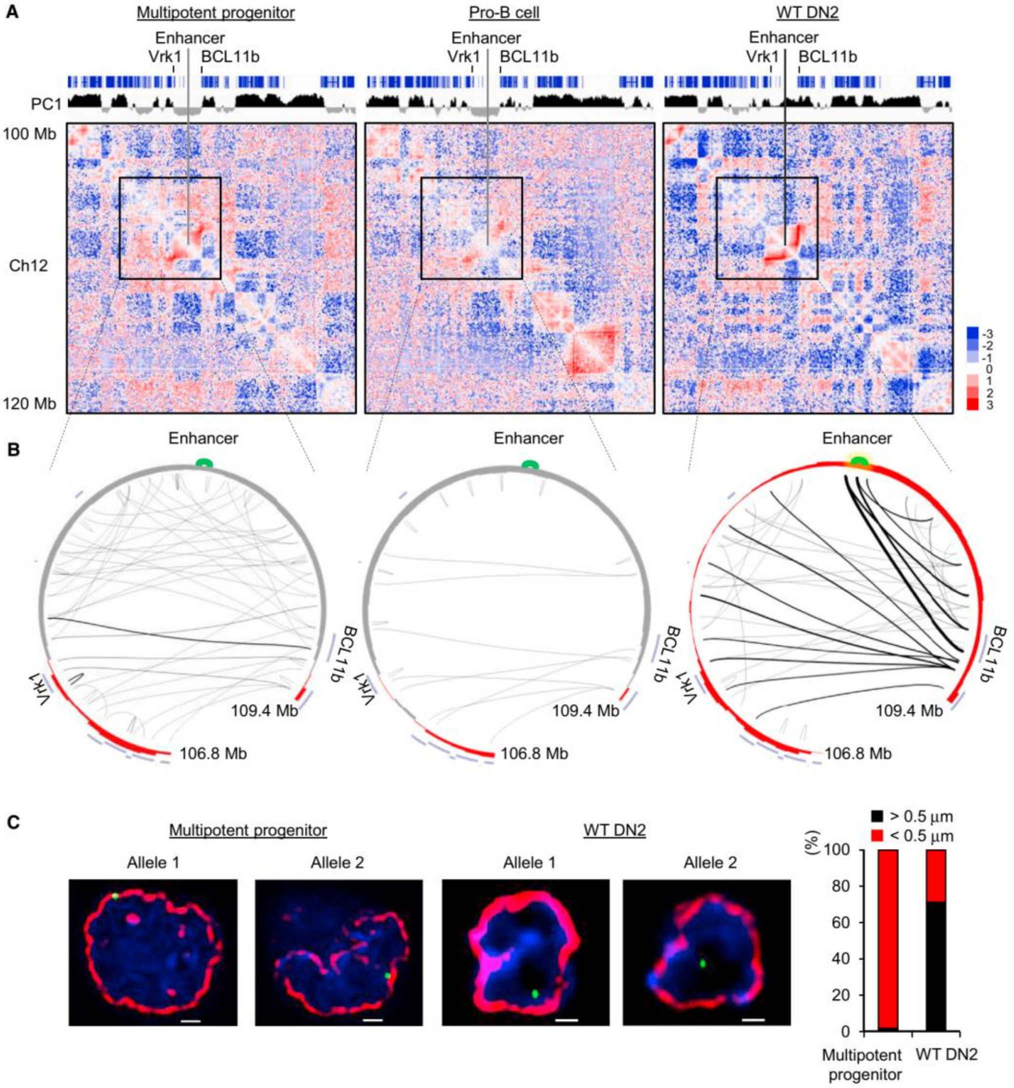


Figure 1. Large-Scale Changes in Nuclear Architecture in Developing T Cell Progenitors

(A) Normalized genome-wide contact matrices for multipotent progenitors, pro-B cells, and DN2 cells of a genomic region located on chromosome 12 are displayed. Top panel shows coding elements (blue) associated with chromosome 12. Positive PC1 values are indicated in black, whereas negative PC1 values are displayed in gray. Heatmap indicates the ratio of observed versus expected frequencies for genomic interactions as revealed by HiC. Numerical and statistical details are described in STAR Methods.

(B) Circos diagrams representing genomic interactions across the Bcl11b intergenic region in multipotent progenitors, pro-B cells, and DN2 cells. Indicated are the Vrk1 and Bcl11b coding regions, as well as the Bcl11b intergenic locus control region. The enhancer is shown in green. Thickness of connecting lines reflects p values associated with the indicated interactions. Positive PC1 values are indicated in red, whereas negative PC1 values are shown in gray. Numerical and statistical details are described in STAR Methods.

(legend continued on next page)

single-loop domain, (4) repositioning the enhancer from a heterochromatic to an euchromatic environment, and (5) permitting the deposition of activating epigenetic marks across the loop domain to facilitate phase separation.

RESULTS

The Bcl11b Locus Control Region Repositions in Developing T Cell Progenitors

In previous studies, we demonstrated that in multipotent progenitors, the Bcl11b intergenic region was associated with the transcriptionally repressive compartment (Lin et al., 2012). To explore the possibility that the Bcl11b locus repositioned in developing thymocytes, adult hematopoietic progenitors were isolated from the bone marrow and cultured on an OP9-DL1 stromal layer in the presence of IL7 and FLT3L (Schmitt and Zúñiga-Pflücker, 2002). Cells were harvested upon reaching the DN2 cell stage and examined for nuclear architecture using HiC (Table S1) (Lieberman-Aiden et al., 2009; Rao et al., 2014). To identify genomic regions that repositioned in developing T cell progenitors, we segregated the DN2 genome into transcriptionally permissive (A) and transcriptionally repressive (B) compartments using principal component analysis. We found that during the transition from the multipotent progenitor to the committed DN2 cell stage, a wide spectrum of genomic regions switched compartments (data not shown). Conspicuous among the genomic regions that repositioned was an intergenic region containing the Bcl11b enhancer (Figure 1A) (Li et al., 2013). Specifically, we found that the Bcl11b intergenic region repositioned from the transcriptionally repressive compartment B to the transcriptionally permissive compartment A (Figure 1A). Notably, the repositioning of the Bcl11b intergenic region was associated with *de novo* genomic interactions involving the Bcl11b enhancer and promoter regions (Figure 1B). To validate these findings, formaldehyde fixed cells were hybridized with fluorescently labeled fosmid probes that span the Bcl11b enhancer. As predicted by the HiC analysis, we found that in a large majority of multipotent progenitor cells, the Bcl11b intergenic region was sequestered at the nuclear envelope, whereas in DN2 cells, it was predominantly localized in the nuclear interior (Figure 1C). These data indicate that in developing thymocytes, the Bcl11b intergenic region harboring the enhancer repositions from the lamina to the nuclear interior to direct the Bcl11b enhancer to the Bcl11b promoter.

Non-coding RNA Transcription Initiated within the Bcl11b Intergenic Region Activates Bcl11b Transcription

Previous studies have suggested that intergenic transcription is associated with the repositioning of genomic regions from heterochromatic to euchromatic genomic regions (Schmitt et al., 2005). As a first approach to identify such transcripts in developing T cell progenitors, we employed assay for transposase-accessible chromatin sequencing (ATAC-seq) and analyzed RNA sequencing (RNA-seq) reads to detect nucleosome depletion and non-coding transcripts across the Bcl11b intergenic region (Table S1) (Buenrostro et al., 2013; Bossen et al., 2015). This analysis revealed a cluster of ATAC-sensitive sites centered on a previously identified enhancer known to regulate Bcl11b expression (Li et al., 2013) (Figure 2A). Four ATAC-sensitive regions associated with this region were identified as CR1, CR2, CR3, and CR4. CR1 corresponds to the Bcl11b enhancer previously named Major Peak (Li et al., 2013). The roles of CR2, CR3, and CR4 were unknown. CR1, CR2, CR3, and CR4 were depleted of nucleosomes in DN2 cells, but not in pro-B cells, indicating lineage specificity (Figure S1A). We next examined this region for developmentally regulated non-coding transcripts whose expression preceded or overlapped with Bcl11b expression (Zhang et al., 2012). Two distinct non-coding RNAs, previously designated as GM16084, initiated from either CR1 (Major Peak) or CR2 (Figure S1B). We will hereafter refer to GM16084 as ThymoD. At the DN1 cell stage, ThymoD transcription was initiated from CR2 immediately prior to the induction of Bcl11b expression (Figure S1B; data not shown). In the DN2 and DN3 cell stages, ThymoD transcription was initiated from both CR1 and CR2 (Figure S1B).

As a first approach to examine potential roles for ThymoD in modulating Bcl11b expression, we inserted a poly(A) site upstream of either CR1 or CR2. As a recipient cell line, we used SCIDadh, representing the DN3 cell stage (Carleton et al., 1999). Upon electroporation of the repair template and gRNAs targeting non-conserved genomic regions either upstream of CR1 or CR2, clones were isolated, expanded, and examined for proper poly(A) site insertion. Clones that harbored the poly(A) insertion on both alleles, flanking either CR1 or CR2, were expanded and examined for ThymoD and Bcl11b expression using real-time PCR. We found that insertion of a poly(A) site upstream of CR1 did not interfere with either ThymoD or Bcl11b transcription (Figure S1C and S1D). In contrast, when inserted upstream of CR2, ThymoD, as well as Bcl11b transcript abundance, significantly declined (Figure S1D). Thus, the insertion of a poly(A) site upstream of CR2, but not CR1, interfered with ThymoD and Bcl11b expression.

Generation of ThymoD p(A)/p(A) Mice

To examine the role of ThymoD in thymocyte development, we generated mice that carried a poly(A) cassette inserted immediately upstream of CR2 (Figure 2A). gRNAs targeting this region in conjunction with the repair template and Cas9 mRNA were injected into C57BL/6 zygotes (Figure S2A) (Yang et al., 2013). Mice that carried at least three poly(A) cassettes in tandem in the ThymoD locus were viable and generated offspring at the expected Mendelian ratios (data not shown). To validate for proper insertion, PCR primers were generated that were located

(C) The Bcl11b intergenic region repositions during the developmental progression from multipotent progenitors to DN2 cells from the nuclear lamina to the nuclear interior. Images represent 3D-FISH analysis using a fosmid probe corresponding to the Bcl11b locus control region. The nuclear envelope was visualized using antibodies directed against the lamina and is indicated in red. The two left images show localization for the Bcl11b enhancer at the nuclear lamina in multipotent progenitors. Right images indicate location of the Bcl11b enhancer in the nuclear interior of DN2 cells. DAPI staining is shown in blue. Bar graph shows fraction of spatial distances (< 500nm) separating the Bcl11b super-enhancer from the nuclear lamina in multipotent progenitor cells (n = 175) and WT DN2 cells (n = 135). Images were digitally magnified. Original magnification was ×100. White bar represents 1 μm.

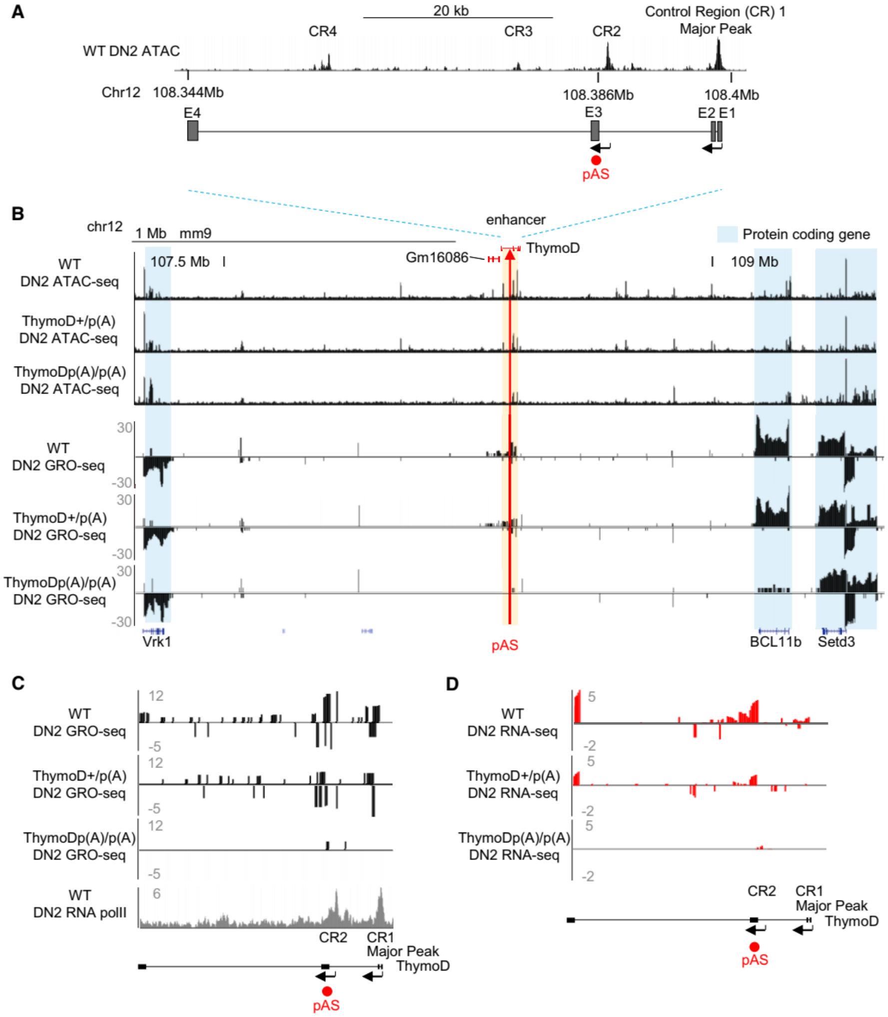


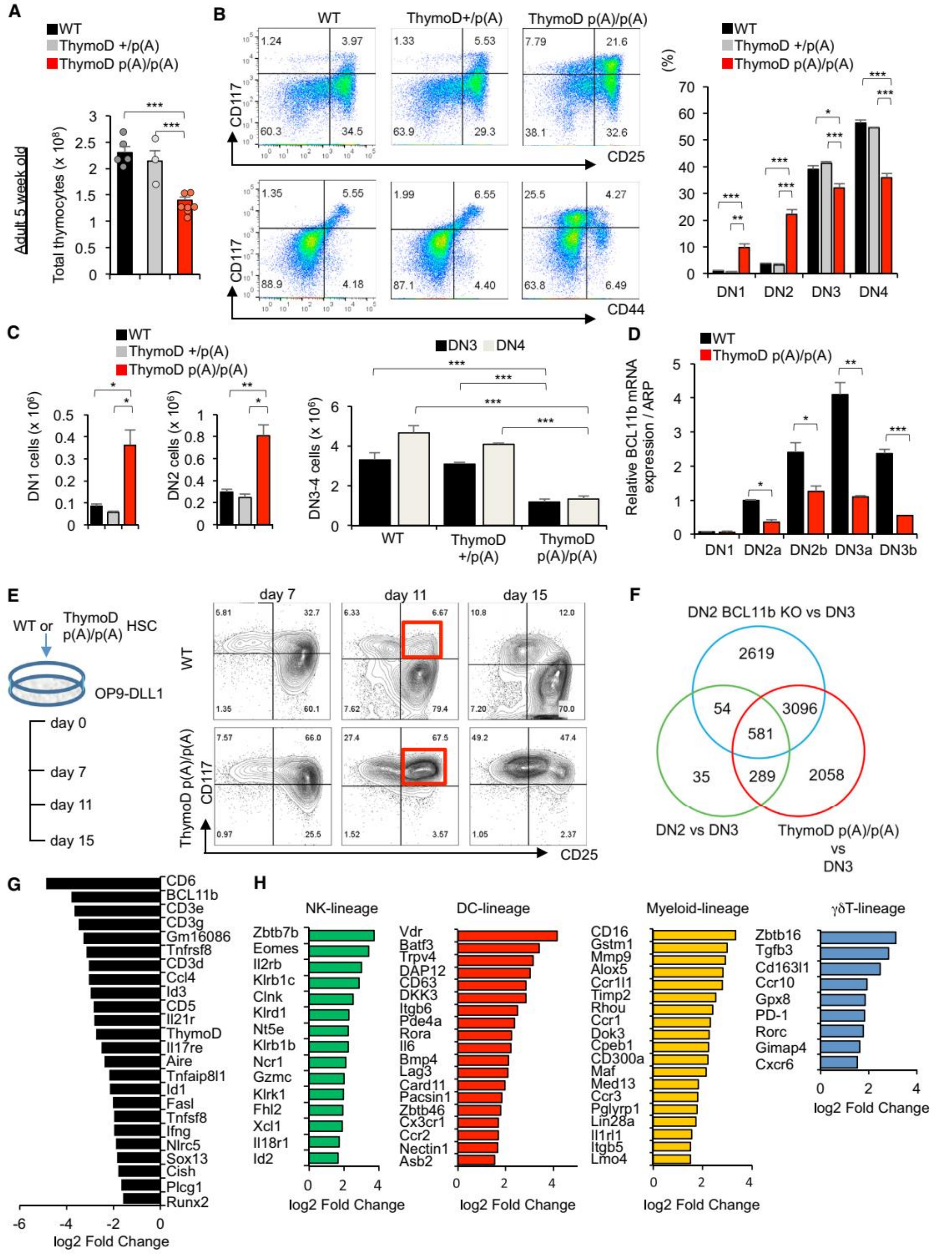
Figure 2. Non-coding RNA ThymoD Activates Bcl11b Expression across Vast Genomic Distances

(A) Schematic diagram depicting the ThymoD region. Upper panel shows ATAC-seq reads for wild-type DN2 cells across the Bcl11b locus control region. CR1 (major peak) indicates the Bcl11b enhancer. CR2, CR3, and CR4 refer to genomic regions depleted for nucleosomes in DN2 cells. Positions of pAS insertions are indicated in red.

(B) ThymoD transcription activates with high precision and selectivity Bcl11b expression. ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells were examined for chromatin accessibility and nascent transcription using ATAC-seq and GRO-seq. ThymoD region is indicated. pAS insertion is shown by red arrow. Enhancer, also named Major Peak or CR1, refers to a previously identified Bcl11b enhancer (Li et al., 2013).

(C and D) pAS insertion abolishes ThymoD transcription. Both GRO-seq and RNA-seq tracks are shown for wild-type, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells.

adjacent and within the targeting construct (Figure S2A). Fragment size of amplified products was as expected (Figure S2B). Sequenced PCR products verified proper insertion of the p(A) sites (Figure S2B; data not shown). DNA sequencing revealed that the putative regulatory elements CR1, CR2, CR3, and CR4 were not mutated by insertion of the poly(A) cassette (data



(legend on next page)

not shown). To validate that the insertion of poly(A) sites interfered with ThymoD transcription, RNA was isolated from ThymoD +/+, ThymoD +/p(A), and ThymoD p(A)/p(A) thymi. We found that ThymoD transcript levels were severely reduced in thymi isolated from ThymoD p(A)/p(A) mice beyond the p(A) insertion site (Figure S2C; data not shown). To examine how ThymoD transcripts were localized in nuclei derived from DN2 cells, we performed RNA-fluorescence *in situ* hybridization (RNA-FISH). ThymoD transcripts were localized at low abundance in wild-type DN2 cells (data not shown).

To determine how p(A) insertion affected the transcription of ThymoD and Bcl11b, we performed global run-on sequencing (GRO-seq). We found that Bcl11b transcripts, as well as transcripts associated with nearby genes, including Vrk1 and Setd3, were readily detectable in wild-type DN2 cells (Figure 2B). To define the transcriptional start sites associated with ThymoD, RNA polymerase II occupancy was analyzed in DN2 cells. Both CR1 and CR2 were enriched for RNA polymerase II occupancy (Figure 2C; lowest panel). Using 3′-RACE (rapid amplification of cDNA ends), we confirmed that CR1 and CR2 act as promoter elements (data not shown). ThymoD sense and anti-sense transcripts were readily detected in wild-type DN2 cells, albeit at relatively low abundance (Figure 2B). However, ThymoD sense and anti-sense transcript levels were severely decreased in ThymoD p(A)/p(A) DN2 cells, indicating that ThymoD transcription is essential to generate a transcriptionally permissive environment (Figures 2C and 2D). Notably, Bcl11b, but not Setd3 and Vrk1, transcript levels were sharply reduced in ThymoD p(A)/p(A) DN2 cells as measured by RNA-seq and GRO-seq (Figure 2B; lowest panel). We conclude that ThymoD transcription activates Bcl11b expression across vast genomic distances with great precision and specificity.

ThymoD Specifies T Cell Fate

To determine whether ThymoD expression specifies T cell fate, thymi were isolated from ThymoD +/+, ThymoD +/p(A), and ThymoD p(A)/p(A) mice and analyzed for developmental abnormalities. We found that thymocyte cellularity in ThymoD p(A)/p(A) mice was reduced 2-fold when compared to thymi derived from wild-type mice (Figure 3A). The proportion and number of the DN1 and DN2 compartments were increased in thymi isolated from ThymoD p(A)/p(A) mice as compared to wild-type mice, whereas the DN3 and D4 compartments were reduced (Figures 3B and 3C). ThymoD p(A)/p(A) thymi also displayed abnormal ratios of DP and SP cells (Figure S3A). The $\gamma\delta$ T cell compartment was reduced in ThymoD p(A)p(A) thymi as compared to wild-type thymi (Figure S3B). We also observed increases in the percentages of NK1.1$^+$, as well as NK1.1$^+$B220$^+$ cells, in thymi isolated from ThymoD p(A)/p(A) mice (Figures S3C and S3D). Aged ThymoD p(A)/p(A) mice showed more pronounced defects that varied and frequently were accompanied by skewed DN ratios (Figure S3E). Ultimately, the majority of aged ThymoD p(A)/p(A) mice succumbed from leukemia and lymphoma (Figures S3F and S3G). The lymphomas and leukemias that developed in ThymoD p(A)/p(A) mice were largely composed of CD4$^+$CD8$^+$ double-positive TCRβ^+ T-lineage cells (data not shown). Since the developmental defects and the development of lymphoid malignancies observed in ThymoD p(A)/p(A) mice closely resembled those observed in radiation-induced lymphoma in Bcl11b heterozygous mice, DN1-3 cells were examined for Bcl11b expression (Liu et al., 2010; Ikawa et al., 2010; Li et al., 2010b; Kamimura et al., 2007; Gutierrez et al., 2011). Indeed, Bcl11b expression in DN2-DN3 cells sorted from ThymoD p(A)/p(A) mice was significantly reduced, as was observed upon insertion of a poly(A) cassette into the ThymoD locus of the SCIDadh cell line (Figure 3D). To further validate the notion that ThymoD and Bcl11b expression were linked, DN2 cells were generated *in vitro* using OP9-DL1 cells. We found that hematopoietic progenitors derived from ThymoD p(A)/p(A) mice failed to develop beyond the DN2 stage upon culture on OP9-DL1 monolayers, consistent with the developmental arrest observed in ThymoD p(A)/p(A) thymi *in vivo* (Figure 3E). Likewise, when cultured on OP9-DL1 cells, ThymoD p(A)p(A) progenitors displayed defects in T (Figure S4A). To examine Bcl11b expression levels in cultured DN2 cells derived from ThymoD p(A)/p(A) mice, RNA was extracted and analyzed by real-time PCR. As was observed in progenitors derived from ThymoD p(A)/p(A) mice, Bcl11b levels were significantly decreased in ThymoD

Figure 3. ThymoD Specifies T Cell Fate

(A) Reduced cellularity in thymi derived from 5-week-old ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) mice. Bar graphs show thymocyte cell numbers in 5-week-old ThymoD+/+, ThymoD+/−, and ThymoD−/− mice.

(B) Thymocyte development is partially blocked at the DN2 cell stage in ThymoD p(A)/p(A) mice. Thymocytes derived from 5-week-old ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) mice were stained for the expression of CD25 versus CD117 and CD44 versus CD117. Cells were gated on the DN compartment. Right panels show bar graphs that display percentages of DN1, DN2, DN3, and DN4 cells in thymi derived from 5-week-old ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) mice.

(C) Left panels show increased numbers of DN1 and DN2 cells in thymi derived from ThymoD p(A)/p(A) mice when compared to thymi isolated from ThymoD+/+ and ThymoD+/p(A) mice. Right panels show decreased numbers of DN3 and DN4 cells in thymi derived from ThymoD p(A)/p(A) mice when compared to thymi isolated from ThymoD+/+ and ThymoD+/p(A) mice.

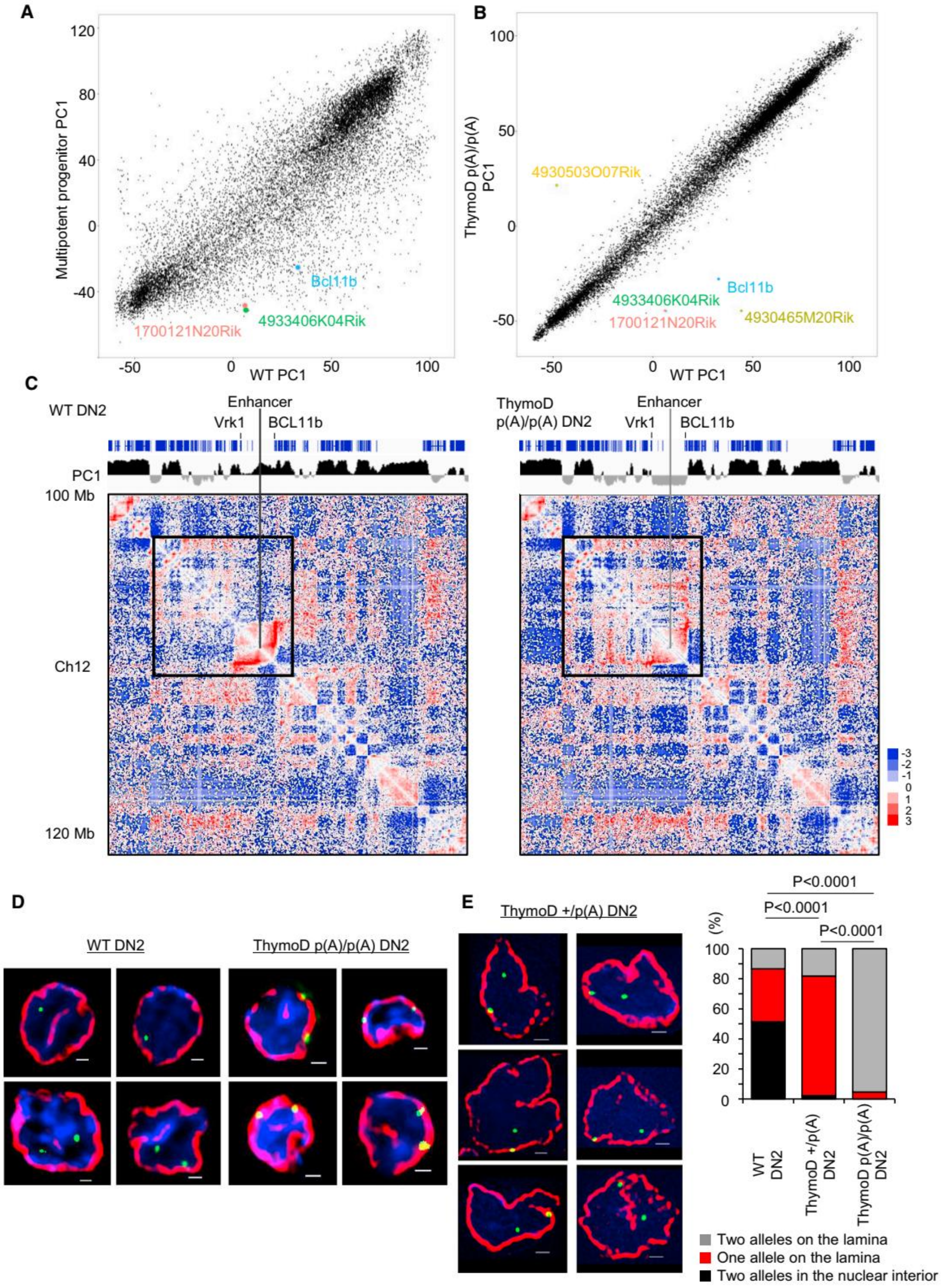
(D) Bar graph indicating relative levels of Bcl11b expression in the DN1-DN3b cells derived from 5-week-old ThymoD+/+ and ThymoD−/− thymi. Results show the mean ± SD (n = 3).

(E) Developmental arrest at the DN2 cell stage in *in vitro* differentiated ThymoD p(A)/p(A) using OP9-DL1 stromal cell cultures. Representative FACS plots for DN compartments are shown.

(F) ThymoD and Bcl11b share a common set of target genes in DN2 cells. Venn diagram is shown displaying differentially expressed genes for indicated genotypes.

(G) ThymoD and Bcl11b are linked in a common pathway to induce a T-lineage specific program of gene expression. Data were obtained from three independent RNA-seq experiments. Table displays a selected group of target genes differentially expressed in sorted ThymoD+/+ and ThymoD p(A)/p(A) DN2 cells. Diagram reveals decreased levels of Bcl11b, Id3, CD3e, and Sox13 expression.

(H) ThymoD acts to suppress the expression of genes associated with the developmental progression of alternative cell lineages. Data were obtained from three independent RNA-seq experiments.



(legend on next page)

p(A)/p(A) DN2 cells cultured *in vitro* (Figure S4B). To determine the extent of overlap between the transcription signatures between ThymoD p(A)/p(A) and Bcl11b-deficient DN2 cells, RNA-seq profiles were compared (Longabaugh et al., 2017). We found that many of the differentially expressed genes in ThymoD p(A)/p(A) were modulated during the DN2 to DN3 transition, and the vast majority overlapped with those differentially expressed in Bcl11b-deficient DN2 cells (Figure 3F). This set of differentially expressed genes was closely linked with specification and commitment of adaptive and innate immune cells (Figures 3G, 3H, and S4C). Prominent among the regulators that characterize adaptive and innate immune cell development were Id2, CD3ε, Runx2, Sox13, Il2rb, Zbtb7b, Fcgr3, Ikzf3, c-kit, Notch3, and Dntt (Figure 3H) (Liu et al., 2010; Ikawa et al., 2010; Li et al., 2010a; Rothenberg et al., 2016). These data directly link ThymoD and Bcl11b into a common pathway to specify T cell fate.

ThymoD Expression Releases the Bcl11b Enhancer from the Nuclear Lamina

Given that the Bcl11b enhancer repositions away from the lamina in developing T cell progenitors and that ThymoD modulates Bcl11b expression, we considered the possibility that ThymoD transcription acts to release the Bcl11b locus from the lamina to the nuclear interior. To address this possibility, we performed HiC analysis on wild-type and ThymoD p(A)/p(A) DN2 cells. Specifically, adult hematopoietic progenitors were isolated from wild-type and ThymoD p(A)/p(A) hematopoietic progenitors, cultured on OP9-DL1 cells, isolated using cell sorting, formaldehyde fixed, and analyzed using HiC. From the HiC reads, we generated contact maps and used principal component analysis to segregate the DN2 genomes into A versus B compartments. Upon comparing PC1 values derived from multipotent progenitors versus wild-type DN2 cells, we found a large ensemble of genomic regions (2260) that switched compartments during the transition from multipotent progenitors to committed DN2 cells (Figure 4A and Table S1). In contrast, a mere 30 genomic regions were associated in different compartments upon comparing ThymoD+/+ and ThymoD p(A)/p(A) DN2 cells (Figure 4B and Table S1). Conspicuous among these was the Bcl11b intergenic region (Figure 4B). The finding that a rather restricted number of genomic regions flipped PC1 values raised the question as how this compares to changes in compartmentalization during the DN2 to DN3 transition. To explore this possibility, we performed HiC on isolated RAG-deficient thymocytes derived from fetal liver cells and cultured on OP9-DL1 cells. Contact maps were generated from DN3 HiC reads and directly compared to those observed for DN2 cells. We found that a total of sixty genomic regions switched compartments during the DN2 to DN3 transition (Tables S2 and S3). We next compared changes in the PC1 distribution across chromosomes for wild-type DN2 versus ThymoD p(A)/p(A) DN2, wild-type DN2 versus wild-type DN3, and wild-type DN2 versus multipotent progenitors. The correlation coefficient of interactions for each locus for the paired cell types was calculated and plotted as the mean correlation coefficient for all loci across each chromosome. We found that most of the differences in PC1 values in wild-type versus ThymoD p(A)/p(A) DN2 cells were restricted to chromosome 12, where the Bcl11 locus is located (Figure S4D). These data are consistent with the notion that ThymoD acts locally to orchestrate compartmentalization. As expected, the changes in compartmentalization in wild-type DN2 versus ThymoD p(A)/p(A) DN2 cells were also closely associated with differences in long-range genomic interactions that span the Bcl11b locus control region (Figure 4C). In wild-type DN2 cells, we observed an elaborate pattern of looping involving the Bcl11b locus control region and the Bcl11b coding region (Figure 4C). In contrast, ThymoD p(A)/p(A) DN2 cells were severely depleted for such interactions. Rather, we observed genomic interactions spanning vast genomic distances and primarily involving neighboring heterochromatic regions (Figure 4C). To validate the changes in compartmentalization, we performed immune-3D-FISH using antibodies directed against the lamina and fluorescently labeled probes corresponding to the Bcl11b intergenic region. As expected, we found that in wild-type DN2 cells, the Bcl11b intergenic region was predominantly localized at the nuclear interior, whereas in ThymoD p(A)/p(A) DN2 cells, the Bcl11b locus was retained for the large majority of cells at or near the nuclear lamina (Figure 4D). Collectively, these observations indicate that ThymoD expression is essential to release the Bcl11b intergenic region from the nuclear lamina.

Nascent ThymoD Transcription Acts in *Cis* to Release the Bcl11b Locus Control Region from the Lamina

While the data described above indicate that ThymoD transcription is essential to release the Bcl11b intergenic region from the nuclear lamina, it remained to be determined whether ThymoD transcription acts in *cis* or in *trans*. To address this

Figure 4. ThymoD Acts in *Cis* to Reposition the Bcl11b Enhancer from the Lamina to the Nuclear Interior

(A) Large ensemble of genomic regions reposition during the transition from multipotent progenitors to differentiated DN2 cells. PC1 values derived from HiC reads are plotted for multipotent progenitors versus differentiated DN2 cells.

(B) ThymoD expression is essential to reposition the Bcl11b locus control region in developing T cell progenitors. Comparison of PC1 values derived from HiC reads generated from ThymoD+/+ versus ThymoD p(A)/p(A) DN2 cells are shown.

(C) ThymoD expression regulates Bcl11b enhancer and promoter compartmentalization. Contact maps derived from HiC reads derived from wild-type and ThymoD p(A)/p(A) DN2 cells are shown. Positive PC1 values are indicated in black, whereas negative PC1 values are displayed in gray. Heatmap indicates the ratio of observed versus expected frequencies for genomic interactions, as revealed by HiC. Numerical and statistical details are described in STAR Methods.

(D) ThymoD expression is essential to release the Bcl11b super-enhancer from the nuclear lamina. Images represent 3D-FISH analysis using a fosmid probe containing the Bcl11b super-enhancer. The nuclear envelope was visualized using antibodies directed against the lamina and is indicated in red. Green signal indicate Bcl11b super-enhancer. Blue reflects DAPI staining.

(E) ThymoD acts in *cis* to release the Bcl11b intergenic region from the lamina. Representative images were taken from ThymoD+/p(A) DN2 cells. Bar graph shows localization of Bcl11b in ThymoD p(A)/p(A) (n = 226), ThymoD+/p(A) (n = 176), and ThymoD p(A)/p(A) DN2 cells (n = 85). Images were digitally magnified. Original magnification was ×100. White bar represents 1 μm. p values are indicated.

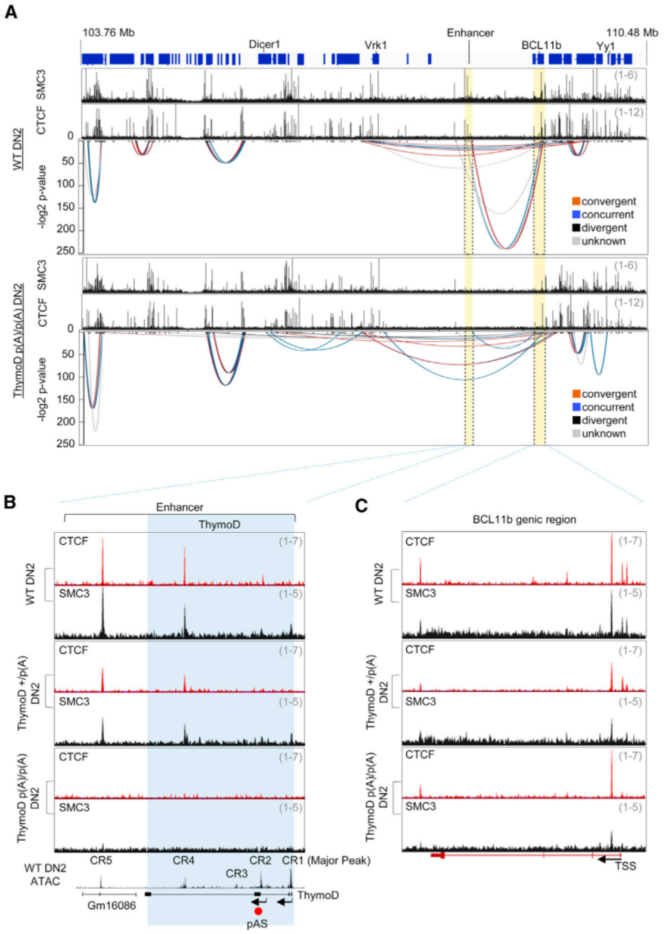


Figure 5. ThymoD Transcription Juxtapose the Bcl11b Super-Enhancer and Promoter Regions into a Single-Loop Domain

(A) ThymoD transcription orchestrates cohesin-dependent looping to bring the Bcl11b promoter and enhancer region into a single-loop domain. Genomic interactions derived from HiC reads across the Bcl11 intergenic and genic regions are shown for wild-type and ThymoD p(A)/p(A) DN2 cells.

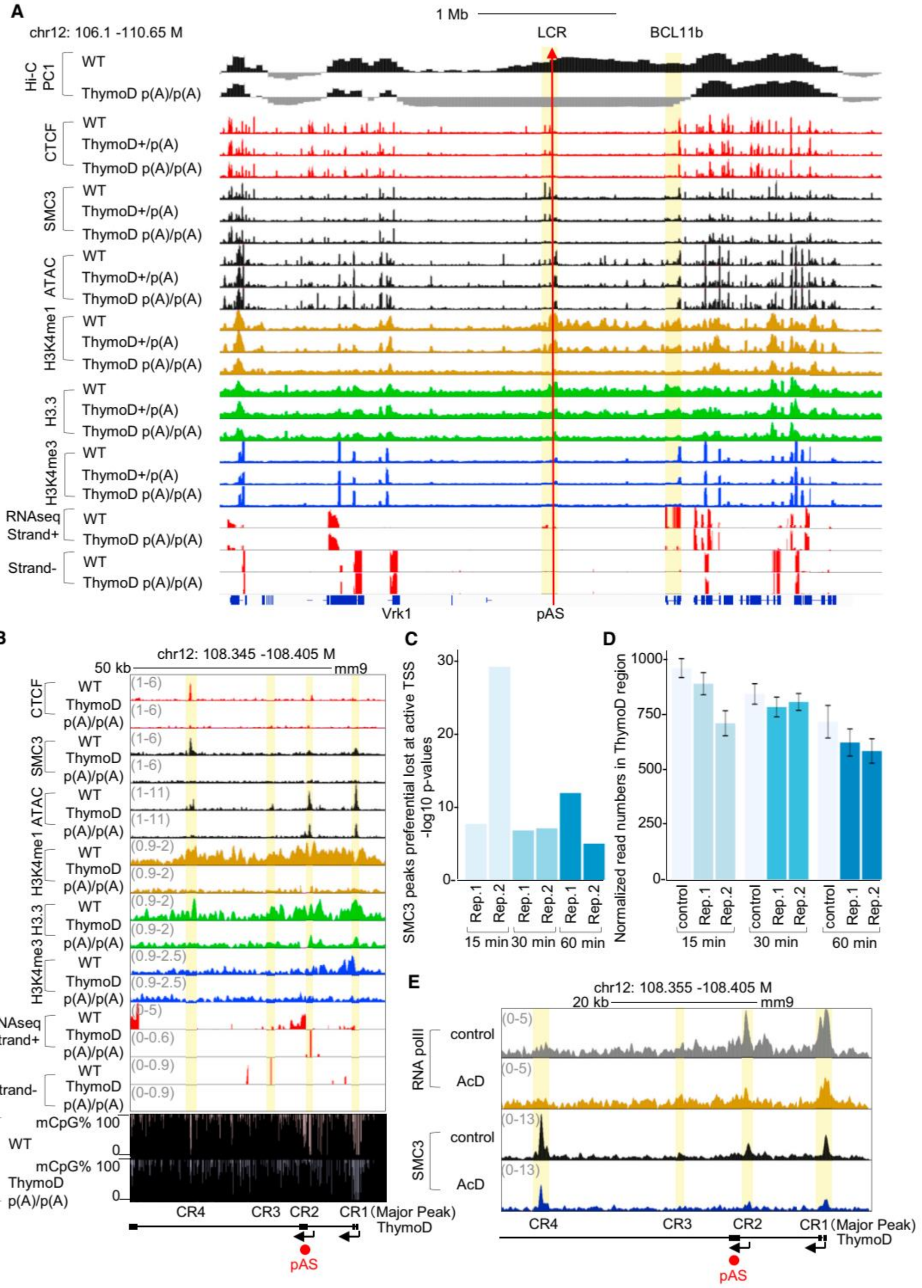
(B) ThymoD p(A)/p(A) DN2 cells are severely depleted for CTCF and SMC3 occupancy across the ThymoD locus. ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells were generated from adult hematopoietic progenitors that were cultured in the presence of OP9DL1 and cytokines. Tracks display CTCF and SMC3 occupancy across a genomic region containing the Bcl11b super-enhancer in ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells. Note that recent mapping data (mm10) indicates that ThymoD and Gm16086 span a single transcription unit. CR1, also named Major Peak, refers to a previously identified Bcl11b enhancer (Li et al., 2013). ChIP-seq reads were corrected for input reads.

(C) ThymoD p(A)/p(A) DN2 cells are severely depleted of CTCF and SMC3 occupancy across the Bcl11b genic region. Tracks display CTCF and SMC3 occupancy across the Bcl11b genic region in ThymoD+/+, ThymoD+/p(A), and ThymoD p(A)/p(A) DN2 cells. CTCF and SMC3 peaks were called using HOMER and marked as red and black tags, respectively. Note that ChIP-seq reads were corrected for input reads.

question, ThymoD+/+ and heterozygous ThymoD+/p(A) DN2 cells were examined for attachment of the Bcl11b intergenic region to the nuclear lamina using immune-3D-FISH. We found that in the vast majority of DN2 +/p(A) cells, a single allele of the Bcl11b super-enhancer was sequestered at the nuclear lamina, whereas the other allele was positioned in the nuclear interior (Figure 4E). Thus, these data indicate that ThymoD expression acts in *cis* to modulate the nuclear location of the Bcl11b locus control region. To test whether forced ThymoD or ThymoD2 expression in ThymoD +/p(A) DN2 cells has the ability to release the ThymoD p(A) allele in *trans*, hematopoietic progenitor cells isolated from ThymoD +/p(A) bone marrow were transduced with virus expressing vector alone, ThymoD (exons 1–3), and ThymoD2 (exons 1, 2, and 4) and were examined for the nuclear positioning of the Bcl11b locus control region (data not shown). We found that the ThymoD p(A) allele remained at the nuclear lamina regardless of ectopic ThymoD or ThymoD2 expression (data not shown). Taken together, these findings demonstrate that (1) nascent ThymoD transcription is required to release the Bcl11b locus control region from the nuclear lamina and (2) ThymoD transcripts generated at wild-type alleles do not have the ability to act in *trans* and release the other Bcl11b allele from the lamina. We conclude that in developing T cell progenitors ThymoD expression acts in *cis* to release the Bcl11b locus control region from the nuclear lamina.

ThymoD-Induced Structural Changes in Chromatin Topology Foster Bcl11b Enhancer-Promoter Communication

To determine how ThymoD transcription modulates local chromatin folding, we examined HiC contact reads derived from wild-type and ThymoD p(A)/p(A) DN2 for genomic interactions across the Bcl11b intergenic and genic regions (Figure 5A). In wild-type DN2 cells, the most significant interactions across a 7 Mb genomic region involved the Bcl11b enhancer and promoter (Figure 5A). Strikingly, genomic interactions associated



(legend on next page)

with the enhancer and promoter were greatly diminished in DN2 cells derived from ThymoD p(A)/p(A) mice (Figure 5A). Rather, we found that ThymoD p(A)/p(A) DN2 cells were highly enriched for loops involving the Bcl11b genic and promoter regions with heterochromatin (Figures 4C and 5A). Thus, ThymoD transcription promotes a pattern of genomic interactions that greatly enriches genomic encounters involving the Bcl11b promoter and enhancer.

To determine how genomic interactions were affected by ThymoD expression, we examined wild-type and ThymoD p(A)/p(A) DN2 cells for CTCF and cohesin (SMC3) occupancy (Figure 5B). We found that in wild-type DN2 cells, CTCF and SMC3 bound to multiple sites that spread across the Bcl11b intergenic and genic regions (Figures 5B and 5C). As expected, the majority of CTCF bound sites overlapped with SMC3 occupancy (Figure 5A). Importantly, both CTCF and SMC3 occupancy across the Bcl11b intergenic region were severely affected in ThymoD p(A)/p(A) DN2 cells (Figure 5B). CTCF and SMC3 occupancy was also affected at the Bcl11b promoter (Figures 5C and S5A).

It could be argued that ThymoD p(A)/p(A) DN2 cells were developmentally distinct from that of wild-type DN2 cells and that the abnormalities observed in CTCF and SMC3 occupancy were caused by a developmental arrest. To address this possibility, we examined CTCF and SMC3 occupancy in ThymoD+/p(A) mice, in which thymocyte development was not perturbed. We found that in ThymoD +/p(A) DN2 cells, CTCF and SMC3 occupancy was reduced 2-fold across the Bcl11b intergenic region, suggesting a direct link between ThymoD transcription and CTCF and SMC3 occupancy (Figure 5B).

As a first approach to determine the mechanism by which ThymoD transcription promotes CTCF and SMC3 occupancy, we examined the ThymoD transcribed region for nucleosome depletion, as well as for the deposition of epigenetic marks (Figures 6A and 6B). We found that ThymoD transcription affected nucleosome depletion and the deposition of epigenetic marks associated with active transcription and the deposition of H3.3 across the Bcl11b promoter, but not genomic regions that flank the Bcl11b regulatory elements (Figures 6A, 6B, and S5A).

The absence of CTCF occupancy in ThymoD p(A)/p(A) DN2 cells raised the question as to why CTCF failed to bind its cognate sites across the ThymoD transcribed region. Since recent studies have demonstrated that non-coding transcription is closely associated with CpG hypomethylation, we considered the possibility that transcription across the ThymoD region acts to recruit members of the Tet protein family (Benner et al., 2015). Hence, we examined DNA isolated from wild-type and ThymoD p(A)/p(A) DN2 cells for cytosine DNA methylation using genome-wide bisulfite sequencing (Figure 6B; bottom tracks). Notably, several CpG residues associated with CTCF binding sites across the ThymoD transcribed region were hypermethylated in ThymoD p(A)/p(A) DN2 cells (Figure 6B; upper and bottom tracks). Taken together, these data indicate that ThymoD transcription is essential to (1) demethylate CpG residues closely associated with CTCF bound sites, (2) recruit members of the cohesin complex to the transcribed region, and (3) reposition the Bcl11b enhancer from a repressive to a transcriptionally permissive neighborhood.

Nascent ThymoD Transcription Is Essential to Recruit Cohesin to the Bcl11b Locus Control Region

The data described above indicate that ThymoD transcription is associated with modulating chromatin topology to bring the Bcl11b enhancer and promoter elements within close spatial proximity. To determine whether nascent ThymoD transcription is required to recruit members of the cohesin complex to the Bcl11b locus control region, wild-type DN2 cells were incubated in the presence of actinomycin D for 15, 30, and 60 min. Treated cells were formaldehyde fixed and immunoprecipitated with antibodies directed against RNA polymerase II and SMC3 followed by genome-wide sequencing. Consistent with previous observations, we found that SMC3 occupancy was significantly depleted at sites associated with RNA polymerase II occupancy (Figure 6C and Table S1) (Izumi et al., 2015; Bhardwaj et al., 2016). We next evaluated whether in DN2 cells nascent ThymoD transcription directly recruits SMC3. We found that DN2 cells treated with actinomycin D showed a significant reduction of SMC3 occupancy across the transcribed region (Figures 6D and S5B). Notably, a brief 15 min incubation with actinomycin D significantly affected SCM3 occupancy (Figure 6E). We conclude that the nascent

Figure 6. ThymoD Transcription Promotes the Deposition of Histone Variant H3.3, Activating Histone Marks and Hypomethyation across the Bcl11b Intergenic Region

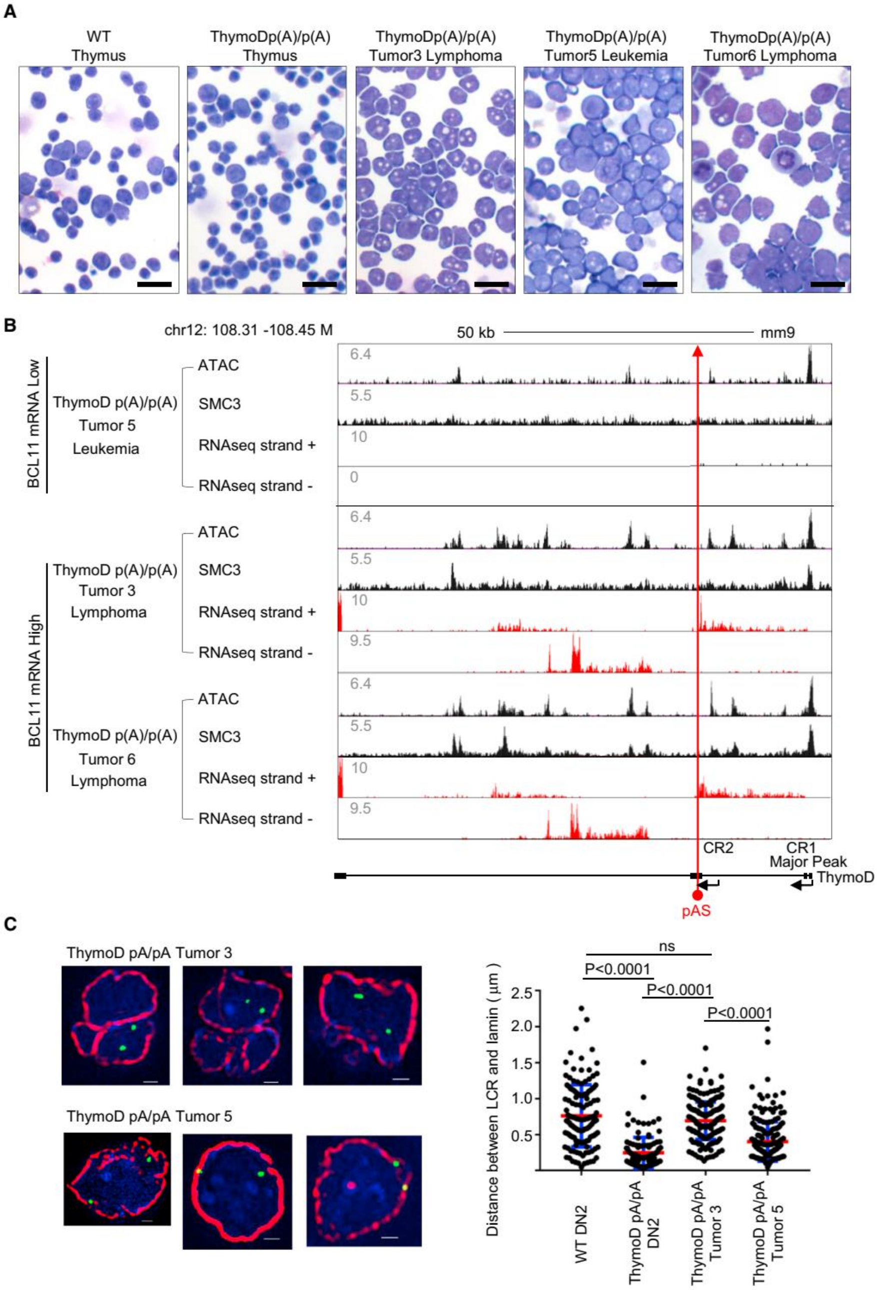
(A) ThymoD expression promotes nucleosome depletion and deposition of epigenetic marks, chromatin accessibility, and H3.3 across the Bcl11b intergenic region. Tracks indicate PC1 values derived from HiC reads, ATAC reads, and CHiP-seq reads for CTCF, SMC3, H3K4me1, H3K4me3, and H3.3 for wild-type and ThymoD p(A)/p(A) DN2 cells. Bottom panels indicated RNA-seq reads across the Bcl11b locus. pAS insertion site is indicated by red arrow. Note that ChIP-seq reads were corrected for input reads.

(B) ThymoD expression modulates chromatin accessibility and the deposition of epigenetic marks across the ThymoD locus. Tracks indicate ATAC reads and ChIP-seq reads for CTCF, SMC3, H3K4me1, H3K4me3, and H3.3 for wild-type and ThymoD p(A)/p(A) DN2 cells. RNA-seq reads across the ThymoD locus for wild-type and ThymoD p(A)/p(A) DN2 cells are indicated. Lower panels show CpG DNA hyper methylation across the CR2 and CR4 regions in ThymoD p(A)/p(A) DN2 cells.

(C) Cohesin occupancy in DN2 cells is closely associated with active transcription start sites. DN2 cells were incubated in the presence of actinomycin D for the indicated time points, formaldehyde fixed, and analyzed for SMC3 occupancy using ChIP-seq. Active transcription start sites were identified using RNA polymerase II ChIP-seq. Plotted is the –log 10 of p values of the Fisher test of the contingency table (Table S1). The analysis indicates preferential loss of cohesin occupancy at active TSS sites.

(D) SMC3 loading across the ThymoD transcribed region requires nascent transcription. SMC reads were determined across the ThymoD locus (chr12: 108.345–108.405 Mb) in DN2 cells treated in the absence or presence of actinomycin D for indicated times and normalized to the total read number. The mean and standard deviation for each experiment were derived from 20 independent samples using boot strapping for all reads across the genome.

(E) Nascent RNA transcription targets SMC3 to the ThymoD locus. DN2 cells were incubated in the absence or presence of actinomycin D for 15 min and analyzed for SMC3 occupancy using ChIP-seq. Representative track shows a reduction of SMC3 occupancy across the ThymoD locus.



(legend on next page)

ThymoD transcription acts in *cis* to directly promote SMC3 loading across the transcribed region.

Compartmentalization and Cohesin-Dependent Looping Is Reversible in ThymoD-Deficient Lymphomas and Leukemias

As described above, ThymoD p(A)/p(A) mice readily develop lymphoma or leukemia. To determine whether the Bcl11b intergenic remained attached to the nuclear lamina in ThymoD p(A)/p(A) lymphomas, long-term cultures were established from the tumors. Tumors isolated from aged ThymoD p(A)/p(A) mice displayed a higher fraction of blasting cells that frequently adopted irregular nuclear morphologies and readily expanded in tissue culture (Figure 7A). Long-term cultures were established either in the absence or presence of OP9-DL1 stromal cells. From these expanding cultures, we established independent cell lines representing either lymphomas or leukemias. To determine whether the expression of ThymoD and Bcl11b remained silent in the lymphomas and leukemias, the tumors were examined for ThymoD and Bcl11b expression using RNA-seq (Figure S6). As expected, ThymoD transcription was blocked at the poly(A) insertion in all of the tumors (Figure S6). Intriguingly, three of the tumors expressed anti-sense transcription at relatively high levels in three of the six tumors (Figure S6). Notably, lymphomas and leukemias that expressed high levels of anti-sense transcription also abundantly expressed Bcl11b (Figure S6). These data indicate that a subset of lymphoid malignancies that develop in ThymoD p(A)/p(A) mice display coordinate ThymoD anti-sense and Bcl11b transcription, whereas in tumors that failed to induce anti-sense ThymoD transcription, Bcl11b expression remained suppressed.

To examine whether the activation of ThymoD and Bcl11b expression in a subset of lymphoma cell lines was associated with chromatin accessibility and SMC3 occupancy across the ThymoD region, three lymphomas (3, 5, and 6) were examined using ATAC-seq and SMC3 ChIP-seq (Figure 7B). As predicted by the analyses described above, lymphomas that were associated with coordinate ThymoD and Bcl11b expression showed widespread ATAC reads across the ThymoD locus and were enriched for SMC3 occupancy (Figure 7B).

To determine whether the induction of ThymoD and Bcl11b expression, as well as the selective increase in SCM3 occupancy, was associated with changes in nuclear location, tumors 3 and 5 were examined by 3D-FISH for compartmentalization. We found that the majority of cells derived from tumor 3, which coordinately expressed high levels of ThymoD and Bcl11b, exhibited localization of the Bcl11b intergenic region that was localized away from the nuclear lamina (Figure 7C). In contrast, the localization of ThymoD in tumor 5, which expresses low levels of ThymoD and Bcl11b, was near or on the nuclear lamina (Figure 7C). These data indicate that the insertion of a poly(A) cassette in the ThymoD locus does not irreversibly attach the ThymoD locus to the nuclear lamina. Instead, ThymoD transcription, either sense or anti-sense, promotes SMC3 occupancy and cohesin-dependent looping, changes in compartmentalization, and the activation of Bcl11b expression to specify T cell fate (Figure S7).

DISCUSSION

Specifying Immune Cell Fate by Non-coding Transcription

While it is now well established that genomic regions reposition during developmental progression, it is not understood how such changes in nuclear location are regulated. Here, we provide insight into the mechanism that underpins the repositioning of regulatory elements. We identify a non-coding RNA, named ThymoD, which acts to reposition the Bcl11 enhancer away from the lamina to the nuclear interior and sequester the Bcl11b promoter and enhancer region into a single-loop domain. How does ThymoD accomplish the repositioning of the Bcl11b enhancer? The phenotype revealed in ThymoD+/p(A) thymocytes strongly suggests that ThymoD acts in *cis* on nearby gene expression, consistent with previous observations that nascent transcription, rather than biochemical activities, associated with a subset of non-coding RNAs executes their function (Melé and Rinn, 2016; Engreitz et al., 2016). Here, we found that the vast majority of DN2+/p(A) cells display one Bcl11b allele located at the nuclear lamina and one in the nuclear interior. Thus, ThymoD acts in *cis* to release the Bcl11b enhancer from the lamina and to active Bcl11b gene expression. How does ThymoD perform this task? We propose a stepwise mechanism. The induction of Bcl11b expression is initiated and maintained in early T cell progenitors by activating a yet-to-be-identified enhancer associated with the ThymoD locus. Next, we suggest that the ThymoD transcribed region directs dioxygenases, Tet proteins, to specific residues that are associated with CTCF occupancy. This explains why, in the absence of ThymoD expression, CTCF occupancy is perturbed across the transcribed region, consistent with previous observations indicating

Figure 7. Compartmentalization and Cohesin-Dependent Looping Is Reversible in ThymoD-Deficient Lymphomas

(A) Development of leukemias and lymphomas in ThymoD p(A)/p(A) mice. Tumors isolated from aged ThymoD p(A)/p(A) mice displayed a higher fraction of blasting cells that frequently adopted irregular nuclear morphologies and readily expanded in tissue culture. DN2 cells derived from wild-type and ThymoD p(A)/p(A) DN2 cells, as well as several lymphomas and leukemias, are shown. Black bar represents 40 μm.

(B) ThymoD expression, either sense or anti-sense transcription, promotes SMC3 occupancy and cohesin occupancy, chromatin accessibility, and the activation of Bcl11b expression. pAS insertion is indicated by red arrow.

(C) ThymoD transcription is closely associated with compartmentalization. Wild-type and ThymoD p(A)/pA) DN2 cells and ThymoD p(A)/pA) lymphomas were examined for nuclear localization of the Bcl11b intergenic region. The two left images show localization for the Bcl11b super-enhancer at the nuclear lamina in multipotent progenitors. Upper images indicate location of the Bcl11b enhancer in the nuclear interior of ThymoD p(A)/p(A) tumor 3. The bottom images show localization of the Bcl11b enhancer at the nuclear lamina in tumor 5. DAPI staining is shown in blue. Bar graph shows fraction of spatial distances (< 500nm) separating the Bcl11b super-enhancer from the nuclear lamina in multipotent progenitor cells (n = 175) and WT DN2 cells (n = 135). ThymoD p(A)/p(A) DN2 cells (n = 120), ThymoD p(A)/p(A) tumor3 (n = 206), ThymoD p(A)/p(A) tumor5 (n = 207). Images were digitally magnified. Original magnification was ×100. White bar represents 1 μm.

that nascent transcription is closely associated with hypomethylated DNA (Benner et al., 2015). Previous observations have also indicated that mice deficient for *Tet2* and *Tet3* expression displayed defects in NKT cell development and lymphoproliferative disease and also exhibited significantly lower abundance of *Bcl11b* expression (Tsagaratou et al., 2017). Although still to be proven, it is conceivable that the absence of *Tet2* and *Tet3* expression in NKT cells leads to hypermethylated CpG residues across the ThymoD transcribed region and in a failure to efficiently activate Bcl11b expression, plausibly resulting in lympho-proliferative disease and defects in NKT cell development. How the Tet proteins target specific CpG residues spanning the transcribed region is an important question that needs to be addressed but may involve distinct secondary DNA structures that are induced either by sense or anti-sense transcription. Regardless of the precise mechanism, these data link non-coding transcription and Tet targeting into a common pathway.

ThymoD transcription was associated with cohesin occupancy consistent with previous studies (Izumi et al., 2015; Bhardwaj et al., 2016; Busslinger et al., 2017). Recruitment of the cohesin complex, in turn, activates an elaborate and coherent pattern of looping, likely involving loop extrusion, to bring the enhancer and promoter region in a singular loop domain (Nasmyth, 2001; Sanborn et al., 2015; Fudenberg et al., 2016). Juxtaposing paired enhancer and promoter regions, separated by vast genomic distances, would yield greatly reduced first-passage times (Lucas et al., 2014). We note that loop extrusion would permit the repositioning of paired enhancers and promoters with great precision even when separated by vast genomic distances. Finally, loop extrusion would sequester the loop domain away from the transcriptionally repressive environment to the euchromatic compartment. We suggest that, once localized in the transcriptionally permissive compartment, transcriptional regulators associated with T-lineage commitment, including GATA-3, TCF1, and RUNX1, would bind enhancer elements across the loop domain to orchestrate the deposition of activating histone marks, as well as the histone variant H3.3, facilitating enhancer-promoter communication that may involve phase separation (Hnisz et al., 2017). Such a sequential order of gene activation is consistent with the notion that ThymoD transcription (DN1 stage) precedes the expression of non-coding RNAs initiated from the enhancer regions (DN2a stage). In sum, we propose a model in which nascent ThymoD transcription dictates enhancer-promoter communication at multiple steps that precede gene activation: (1) demethylation at CpG residues closely associated with CTCF occupancy across the ThymoD genomic region, (2) activation of cohesin-dependent looping to juxtapose the enhancer and promoter into a single-loop domain and to relocalize the enhancer from a heterochromatic to an euchromatic environment, and (3) deposition of activating epigenetic marks across the loop domain once positioned in the euchromatic compartment to facilitate interactions of paired regulatory elements by phase separation.

Mechanisms that Orchestrate Immune Cell Development

Previous studies have established that a subset of genomic regions reposition in the nucleus during the transition from the multipotent progenitor cell stage to committed B lineage cells (Lin et al., 2012; Miyazaki et al., 2014; van Steensel and Belmont, 2017). Switching the nuclear location of genomic regions is often associated with changes in gene expression (Lin et al., 2012; Kind et al., 2015). Prominent among the genomic regions that switch nuclear location in B cell progenitors are genes encoding for transcriptional regulators that specify B cell fate, including the EBF1 and FOXO1 loci (Lin et al., 2012; Mansson et al., 2012). Both loci are associated with the heterochromatic compartment in multipotent progenitors but reposition to the transcriptionally permissive compartment upon developing into committed B lineage cells (Lin et al., 2012). Here, we demonstrate that in developing T cell progenitors, a genomic region associated with the Bcl11b locus that harbors critical regulatory elements repositions from the nuclear lamina to the transcriptionally permissive nuclear interior. These findings point to a common mechanism that specifies B and T cell fate. In developing B cell progenitors, the EBF1 locus switches from the nuclear lamina to the transcriptionally permissive compartment, leading to the induction of EBF1 expression to establish B cell identity. In contrast, in T cell progenitors, the Bcl11b enhancer repositions from the lamina to the nuclear interior to activate Bcl11b expression. Why has such a complex mechanism, involving repression, nuclear repositioning, and activation, evolved to promote adaptive immune cell fate? We suggest that anchoring regulatory elements to the nuclear lamina permits efficient transcriptional repression across a large genomic region preventing stochastic and premature activation of a lymphoid-specific program of gene expression in hematopoietic progenitors. Only upon reaching a distinct developmental stage are progenitor cells instructed to release locus control regions or super enhancers from the nuclear lamina to the transcriptionally active compartment. In B cell progenitors, a repositioned EBF1 locus control region in the transcriptionally permissive compartment would associate with B-lineage-specific transcription factors such as E2A and FOXO1 to induce a B-lineage-specific program of gene expression (Mansson et al., 2012). In T cell progenitors, the Bcl11b locus control region being positioned in the transcriptionally permissive compartment would be able to respond to Notch signaling and transcriptional regulators that include GATA3, TCF1, and RUNX1 to activate a T-lineage-specific transcription signature (Kueh et al., 2016). Thus, specification of adaptive immune cell fate in early lymphoid progenitors is orchestrated by a common mechanism that involves the repositioning of regulatory regions from the lamina to the euchromatic compartment.

Non-coding Transcription and Tumor Suppression

The activation of anti-sense ThymoD transcription associated with a subset of lymphomas that were derived from ThymoD p(A)/p(A) mice was initially puzzling. How is anti-sense ThymoD transcription induced in a subset of lymphomas? Murine T cell lymphomas are frequently associated with mutations that activate the Notch signaling pathway, which activates Bcl11b expression (Weng et al., 2004; Yui and Rothenberg, 2014). Although a detailed characterization of cooperating oncogenic mutations in the ThymoD p(A)/p(A) lymphomas will be required, we speculate that the aberrant activation of the Notch signaling cascade leads to the induction of high levels of anti-sense

ThymoD transcription, ultimately leading to release of the Bcl11b intergenic region from the lamina. Since both sense and antisense ThymoD transcription are associated with changes in compartmentalization and chromatin folding, we suggest that ThymoD transcription, rather than its primary nucleotide sequence, is the critical determinant that dictates the specificity of Bcl11b enhancer-promoter communication.

Previous studies have demonstrated that somatic mutation of an intergenic region may generate an oncogenic superenhancer (Mansour et al., 2014). Our observations indicate that non-coding transcription may act as a tumor suppressor. Given that non-coding transcription is widespread, we suggest that, likewise, non-coding transcription is a common mechanism to activate the expression of nearby genes that encode for proteins with tumor suppressive function.

Conclusion

It is now well established that the genomes of animal and plant kingdoms are organized into euchromatic and heterochromatic regions. Numerous studies have also identified genomic regions that switch nuclear location during developmental progression. The underlying mechanism that directs changes in compartmentalization, however, remained to be revealed. Here, we present data indicating that non-coding transcription dictates remodeling of local chromatin structure, possibly involving loop extrusion, to reposition regulatory elements from the heterochromatic to the euchromatic compartment. We suggest that local remodeling of chromatin topology by non-coding transcription-induced loop extrusion is a universal mechanism that permits genomic regions to readily switch compartments.

Enhancers are often separated from promoter regions by vast genomic distances. How enhancer elements select their cognate promoter regions with high precision and specificity remains to be revealed (Farley et al., 2015). We suggest that developmental and lineage-specific patterns of non-coding transcription sequester, by loop extrusion, enhancer and promoter elements into a single-loop domain. Non-coding transcription-induced loop extrusion explains how enhancers and promoters, segregated from each other by vast genomic distances and positioned in separate compartments, find each other with great specificity during developmental progression.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - SCIDadh cell line culture
  - Mice
  - T cell culture
  - ThymoD p(A)/p(A) tumor culture
- METHOD DETAILS
  - Generating SCIDadh pA/pA cell line
  - sgRNA template sequence at upstream of ThymoD CR1
  - sgRNA template sequence at exon3 of thymoD (upstream of CR2)
  - Generating ThymoD pA/pA mice
  - Primer for genotyping
  - Flow cytometry
  - Wright Giemsa staining
  - Real-time PCR
  - Primer for qPCR
  - 3D-FISH
  - In Situ Hi-C
  - ATAC-Seq
  - ChIP-Seq
  - RNA-Seq
  - GRO-seq
  - Whole genome bisulfite sequencing (WGBS)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Filtering and Normalization of HiC Reads
  - Analysis for ATAC-Seq
  - ChIP-Seq
  - Analysis of CTCF direction and significant interaction
  - Analysis of SMC3 Occupancy in DN2 Cells Exposed to Actinomycin D
  - RNA-Seq
  - Analysis of GRO-seq
  - Analysis of WGBS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental information includes seven figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2017.09.001.

AUTHORS CONTRIBUTIONS

T.I. and C.M. designed experiments. T.I. performed the majority of the experiments. A.J.M. performed HiC. V.C. identified ThymoD. Z.H., M.D., K.M.F., and A.N.C. performed bioinformatics. J.P.v.H. characterized ThymoD isoforms. S.P.F. and D.L.W. provided cells. T.I. and C.M. wrote the manuscript. C.M. conceived and supervised the study.

ACKNOWLEDGMENTS

We thank Alex Bortnick for editing the manuscript. This study was supported by funding from the Center for Computational Biology & Bioinformatics (UL1TRR001442), the California Institute for Regenerative Medicine (RB5-07025), and the NIH (AI102853, AI00880, AI082850, and AI09599) to C.M. D.L.W. was supported by the NIH (AI102853). T.I. was supported by the Uehara Memorial Foundation.

Received: January 30, 2017
Revised: June 24, 2017
Accepted: August 30, 2017
Published: September 21, 2017

REFERENCES

Bain, G., and Murre, C. (1998). The role of E-proteins in B- and T-lymphocyte development. Semin. Immunol. *10*, 143–153.

Benner, C., Isoda, T., and Murre, C. (2015). New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing B cell progenitors. Proc. Natl. Acad. Sci. USA *112*, 12776–12781.

Bhardwaj, S., Schlackow, M., Rabajdova, M., and Gullerova, M. (2016). Transcription facilitates sister chromatid cohesion on chromosomal arms. Nucleic Acids Res. *44*, 6676–6692.

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.

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.

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.

Carleton, M., Ruetsch, N.R., Berger, M.A., Rhodes, M., Kaptik, S., and Wiest, D.L. (1999). Signals transduced by CD3epsilon, but not by surface pre-TCR complexes, are able to induce maturation of an early thymic lymphoma in vitro. J. Immunol. *163*, 2576–2585.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature *539*, 452–455.

Farley, E.K., Olson, K.M., and Levine, M.S. (2015). Regulatory principles governing tissue specificity of developmental enhancers. Cold Spring Harb. Symp. Quant. Biol. *80*, 27–32.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of chromosomal domains by loop extrusion. Cell Rep. *15*, 2038–2049.

García-Ojeda, M.E., Klein Wolterink, R.G., Lemaître, F., Richard-Le, Goff, O., Hasan, M., Hendriks, R.W., Cumano, A., and Di Santo, J.P. (2013). GATA-3 promotes T cell specification by repressing B cell potential in pro-T cells in mice. Blood *121*, 1749–1759.

Guo, Y., Maillard, I., Chakraborti, S., Rothenberg, E.V., and Speck, N.A. (2008). Core binding factors are necessary for natural killer cell development and cooperate with Notch signaling during T-cell specification. Blood *112*, 480–492.

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.Y., and Pellegrini, M. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics *14*, 774.

Gutierrez, A., Kentsis, A., Sanda, T., Holmfeldt, L., Chen, S.C., Zhang, J., Protopopov, A., Chin, L., Dahlberg, S.E., Neuberg, D.S., et al. (2011). The BCL11B tumor suppressor is mutated across the major molecular subtypes of T-cell acute lymphoblastic leukemia. Blood *118*, 4169–4173.

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.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A phase separation model for transcriptional control. Cell *169*, 13–23.

Ikawa, T., Kawamoto, H., Goldrath, A.W., and Murre, C. (2006). E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. J. Exp. Med. *203*, 1329–1342.

Ikawa, T., Hirose, S., Masuda, K., Kakugawa, K., Satoh, R., Shibano-Satoh, A., Kominami, R., Katsura, Y., and Kawamoto, H. (2010). An essential developmental checkpoint for production of the T cell lineage. Science *329*, 93–96.

Izumi, K., Nakato, R., Zhang, Z., Edmondson, A.C., Noon, S., Dulik, M.C., Rajagopalan, R., Venditti, C.P., Gripp, K., Samanich, J., et al. (2015). Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. Nat. Genet. *47*, 338–344.

Kamimura, K., Ohi, H., Kubota, T., Okazuka, K., Yoshikai, Y., Wakabayashi, Y., Aoyagi, Y., Mishima, Y., and Kominami, R. (2007). Haploinsufficiency of Bcl11b for suppression of lymphomagenesis and thymocyte development. Biochem. Biophys. Res. Commun. *355*, 538–542.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell *163*, 134–147.

Klein, L., Kyewski, B., Allen, P.M., and Hogquist, K.A. (2014). Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nat. Rev. Immunol. *14*, 377–391.

Kueh, H.Y., Yui, M.A., Ng, K.K., Pease, S.S., Zhang, J.A., Damle, S.S., Freedman, G., Siu, S., Bernstein, I.D., Elowitz, M.B., and Rothenberg, E.V. (2016). Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. Nat. Immunol. *17*, 956–965.

Li, L., Leid, M., and Rothenberg, E.V. (2010a). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science *329*, 89–93.

Li, P., Burke, S., Wang, J., Chen, X., Ortiz, M., Lee, S.C., Lu, D., Campos, L., Goulding, D., Ng, B.L., et al. (2010b). Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. Science *329*, 85–89.

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.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Liu, P., Li, P., and Burke, S. (2010). Critical roles of Bcl11b in T-cell development and maintenance of T-cell identity. Immunol. Rev. *238*, 138–149.

Longabaugh, W.J.R., Zeng, W., Zhang, J.A., Hosokawa, H., Jansen, C.S., Li, L., Romero-Wolf, M., Liu, P., Kueh, H.Y., Mortazavi, A., and Rothenberg, E.V. (2017). Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. Proc. Natl. Acad. Sci. USA *114*, 5800–5807.

Lucas, J.S., Zhang, Y., Dudko, O.K., and Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. Cell *158*, 339–352.

Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., et al. (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a non-coding intergenic element. Science *346*, 1373–1377.

Mansson, R., Welinder, E., Åhsberg, J., Lin, Y.C., Benner, C., Glass, C.K., Lucas, J.S., Sigvardsson, M., and Murre, C. (2012). Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. Proc. Natl. Acad. Sci. USA *109*, 21028–21033.

Melé, M., and Rinn, J.L. (2016). "Cat's Cradling" the 3D Genome by the Act of LncRNA Transcription. Mol. Cell *62*, 657–664.

Miyazaki, K., Miyazaki, M., and Murre, C. (2014). The establishment of B versus T cell identity. Trends Immunol. *35*, 205–210.

Miyazaki, M., Miyazaki, K., Chen, K., Jin, Y., Turner, J., Moore, A.J., Saito, R., Yoshida, K., Ogawa, S., Rodewald, H.R., et al. (2017). The E-Id protein axis specifies adaptive lymphoid cell identity and suppressed thymic innate lymphoid cell development. Immunity *46*, 818–834.e4.

Naito, T., Tanaka, H., Naoe, Y., and Taniuchi, I. (2011). Transcriptional control of T-cell development. Int. Immunol. *23*, 661–668.

Nasmyth, K. (2001). Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. Annu. Rev. Genet. *35*, 673–745.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rothenberg, E.V., Ungerbäck, J., and Champhekar, A. (2016). Forging T-lymphocyte identity: Intersecting networks of transcriptional control. Adv. Immunol. *129*, 109–174.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA *112*, 6456–6465.

Schmitt, T.M., and Zúñiga-Pflücker, J.C. (2002). Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. Immunity *17*, 749–756.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. *19*, 697–708.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctinos with RNA-Seq. Bioinformatics *25*, 1105–1111.

Tsagaratou, A., González-Avalos, E., Rautio, S., Scott-Browne, J.P., Togher, S., Pastor, W.A., Rothenberg, E.V., Chavez, L., Lähdesmäki, H., and Rao, A. (2017). TET proteins regulate the lineage specification and TCR-mediated expansion of iNKT cells. Nat. Immunol. *18*, 45–53.

van Steensel, B., and Belmont, A.S. (2017). Lamina-associated domains: Links with chromosome architecture, heterochromatin, and gene repression. Cell *169*, 780–791.

Weber, B.N., Chi, A.W., Chavez, A., Yashiro-Ohtani, Y., Yang, Q., Shestova, O., and Bhandoola, A. (2011). A critical role for TCF-1 in T-lineage specification and differentiation. Nature *476*, 63–68.

Weng, A.P., Ferrando, A.A., Lee, W., Morris, J.P., 4th, Silverman, L.B., Sanchez-Irizarry, C., Blacklow, S.C., Look, A.T., and Aster, J.C. (2004). Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science *306*, 269–271.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370–1379.

Yui, M.A., and Rothenberg, E.V. (2014). Developmental gene networks: a triathlon on the course to T cell identity. Nat. Rev. Immunol. *14*, 529–545.

Zhang, J.A., Mortazavi, A., Williams, B.A., Wold, B.J., and Rothenberg, E.V. (2012). Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell *149*, 467–482.

Cell

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-CTCF	EMD Millipore	Cat # 07-729; RRID: AB_441965
Anti-H3K4me1	Abcam	Cat # ab8895; RRID: AB_306847
Anti-H3K4me3	Abcam	Cat # ab8580; RRID: AB_306649
Anti-H3.3	Abcam	ab176840
Anti-RNA polymerase II CTD (8WG16)	Abcam	Cat # ab817; RRID: AB_306327
Anti-SMC3	Abcam	Cat # ab9263; RRID: AB_307122
Anti-BrdU (IIB5)	Santa Cruz	Cat # sc-32323 AC; RRID: AB_626766
Anti-Lamin B1	Santa Cruz	Cat # sc-6217; RRID: AB_648158
Anti-CD3e (145-2C11)	Biolegend	Cat # 100306; RRID: AB_312671
Anti-CD3e (145-2C11)	eBioscience	Cat # 13-0031-82; RRID: AB_466319
Anti-CD4 (GK1.5)	Biolegend	Cat # 100423; RRID: AB_389302
Anti-CD4 (GK1.5)	Biolegend	Cat # 100408; RRID: AB_312693
Anti-CD8a (53-6.7)	eBioscience	Cat # 11-0081-85; RRID: AB_464916
Anti-CD8a (53-6.7)	Bioleged	Cat # 100725; RRID: AB_493425
Anti-CD11b (M1/70)	eBiosciences	Cat # 11-0112-85; RRID: AB_464936
Anti-CD19 (MB19-1)	eBiosciences	Cat # 11-0191-82; RRID: AB_464965
Anti-B220 (RA3-6B2)	eBiosciences	Cat # 11-0452-82; RRID: AB_465054
Anti-B220 (RA3-6B2)	eBiosciences	Cat # 13-0452-85; RRID: AB_466450
Anti-Gr1 (RB6-8C5)	Biolegend	Cat # 108417; RRID: AB_389309
Anti-Gr1 (RB6-8C5)	eBiosciences	Cat # 13-5931-85; RRID: AB_466801
Anti-NK1.1(PK136)	eBiosciences	Cat # 11-5941-85; RRID: AB_465319
Anti-CD25 (PC61)	Biolegend	Cat # 102016; RRID: AB_312865
Anti-CD27 (LG3A)	Biolegend	Cat # 124211; RRID: AB_1236460
Anti-CD44 (IM7)	Biolegend	Cat # 103020; RRID: AB_493683
Anti-CD45.2 (104)	Biolegend	Cat # 109828; RRID: AB_893350
Anti-CD117 (ack45)	BD Pharmingen	Cat # 553869; RRID: AB_395103
Anti-CD117 (2B8)	eBiosciences	Cat # 13-1171-82; RRID: AB_466569
Anti-CD127 (A7R37)	eBiosciences	Cat # 13-1271-85; RRID: AB_466589
Anti-Flt-3 (A2F10)	eBiosciences	Cat # 13-1351-85; RRID: AB_466600
Anti-TCRb (H57-597)	Biolegend	Cat # 109211; RRID: AB_313434
Anti-TCRgd (eBioGL3)	Biolegend	Cat # 118105; RRID: AB_313829
Anti-Ter119 (TER119)	eBioscience	Cat # 11-5921-82; RRID: AB_465311
Anti-Ter119 (TER119)	eBiosciences	Cat # 13-5921-85; RRID: AB_466798
Anti-Thy1.2 (53-2.1)	eBiosciences	Cat #13-0902-85; RRID: AB_466534
Biological Samples		
Murine thymus	Jackson	Cat # 00056
Chemicals, Peptides, and Recombinant Proteins		
Actinomycin D	Thermo FIsher Scientific	Cat # 11805017
Puromycin dihydrochloride hydrate	Fisher Scientific	Cat # AC227420100
Unmethylated lambda DNA	Promega	Cat # D1521
MethoCult M3630	Stem Cell Technologies	Cat # 3630
Anti-biotin Microbeads	Miltenyi Biotech	Cat # 130-090-485

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Complete protease inhibitor cocktail, EDTA-free	Roche	Cat # 4693132001
Ghost dye red 780	Tonbo bissciences	Cat # 13-0865-T100
Ficol-paque plus	GE health care	Cat #17-1440-02
Formamide	Sigma	Cat # F9037-100ML
5-Bromouridine 5′-triphosphate	Sigma	Cat # B7166-10MG
MES sodium salt	Sigma	Cat # M3058-25G
SUPERase In Rnase inhibitor	Life Technologies	Cat # AM2694
Biotin-14-dATP	Life Technologies	Cat # 19524-016
Qdot 585 streptavidin conjugate	Invitrogen	Cat # Q10111MP
N-Lauroylsarcosine sodium salt solution	Sigma	Cat # L7414-10ml
10x T4DNA Ligation Buffer	NEB	Cat # B0202
T4 PNK	NEB	Cat # M0201
T4 RNA Ligase 2, truncated	NEB	Cat # M0242
T4 RNA Ligase 1 (ssRNA Ligase)	NEB	Cat # M0204
5′ DNA Adenylation Kit	NEB	Cat # E2610
DNA Polymerase I, Large fragment (Klenow)	NEB	Cat # M0210
Klenow 3′-5′ exo minus	NEB	Cat # M0212
MboI	NEB	Cat # R0147
Quick Ligation Kit	NEB	Cat # M2200
Paraformaldehyde Granular	Electron Microscopy Sciences	Cat #19208
Critical Commercial Assays		
Nextera DNA Sample Preparation Kit for ATAC-seq	Illumina	Cat # FC-121-1030
NEBNext Multiplex Oligos for Illumina Set1 and 2	NEB	Cat # E7335S
TruSeq DNA Sample Preparation Kits v2 for WGBS	Illumina	Cat # FC-121-2001
MethylCode Bisulfite Conversion Kit	Invitrogen	Cat # MECOV-50
RNeasy Mini Kit	QIAGEN	Cat # 74106
RNeasy Micro Kit	QIAGEN	Cat # 74004
RNase free Dnase (50 units)	QIAGEN	Cat # 79254
QIAGEN Miniprep Kit	QIAGEN	Cat # 27106
DNA Clean and Concentrator	Zymo Research	Cat # D4014
NEBuider HiFi DNA Assembly Master Mix	NEB	Cat # E2621S
dsDNA HS qubit kit	Life Technologies	Cat # Q32854
MEGAshortscript T7 Kit	Life Technologies	Cat # AM1354
MEGAclear Transcription Clean-Up Kit	Life Technologies	Cat # AM1908
SuperScriptIII First-Strand Synthesis System	Life Technologies	Cat # AM2010
RNA Fragmentation Reagents	Life Technologies	Cat # AM8740
Micro Bio-Spin 30 Columns #732-6250	Biorad	Cat # 732-6250
Ultrafree-MC Centrifugal Filter Units with Microporous Membrane	Millipore	Cat # UFC30HVNB
AMPureXP beads	Beckman Coulter	Cat # A63880
ProbeQuant G50 micro-columns	GE Health Care	Cat # 28-9034-08
Nick translation kit	Roche	Cat #11745808910
Deposited Data		
Hi-C, ATAC-seq, ChIP-seq, GRO-seq, RNA-seq	This manuscript	GEO: GSE90958
ATAC-seq	Bossen et al., 2015	GEO: GSE66978
Hi-C	N/A	GEO: GSE40173

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNA-seq	Zhang et al., 2012	GEO: GSE31235
RNA-seq	Longabaugh et al., 2017	GEO: GSE89198
Experimental Models: Cell Lines		
Human: 293T	ATCC	Cat # ATCC-CRL-3216
Mouse: OP9-DL1	Schmitt and Zúñiga-Pflücker, 2002	N/A
Mouse: SCIDadh cells	Carleton et al., 1999	N/A
Experimental Models: Organisms/Strains		
C57BL/6 Mice	Jackson	Cat # 00064
Recombinant DNA		
Fosmid: BCL11b enhancer	BacPac	Cat # WIBR1-1110C15
pGEM-T Easy Vector System I	Promega	Cat # A1360
Cas9 WT vector (WT Cas9 cut by NcoI/EcoRI from px330 was cloned into the pdcas9 vector by NcoI/EcorI)	This paper	Cat # 44246
pdCas9-humanized	Addgene	N/A
pX330-U6-Chimeric_BB-CBh-hSpCas9	Addgene	N/A
pGEM LHA pAS (ThymoD exon3) RHA	This paper	N/A
pGEM LHA GFP pAS (CR1-CR2) RHA	This paper	N/A
pGEM LHA GFP pAS (ThymoD exon3) RHA	This paper	N/A
Sequence-Based Reagents		
sgRNA at CR1-CR2 control region (cagctGAATTCTAATACGACTCACTATAGGGGGAAAACACAGACCCTAGTCgttttagagctagaaatagcaagttaaaataaggctagtccGttatcAacttgaaaaagtggcaccgagtcggtgc)	IDT	N/A
sgRNA at exon 3 of ThymoD (cagctGAATTCTAATACGACTCACTATAGGGTCTTCAAGGGTGCTATCACAgttttagagctagaaatagcaagttaaaataaggctagtccGttatcAacttgaaaaagtggcaccgagtcggtgc)	IDT	N/A
Software and Algorithms		
FlowJo software	Tree Star	N/A
GraphPad Prism 7	graphPad Software	N/A
Bowtie	N/A	http://bowtie-bio.sourceforge.net/index.shtml
HOMER	N/A	http://homer.ucsd.edu/homer
TopHat	Trapnell et al., 2009	https://ccb.jhu.edu/software/tophat/index.shtml
BSSeeker2	Guo et al., 2013	https://github.com/BSSeeker/BSseeker2
Other		
Poly-D-Lysine/Laminin; 12mm dish	Corning	Cat # 08-774-385
Fisherbrand Microscope cover glass	Fisher Scientific	Cat # 12-541AC

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to the Lead Contact, Cornelis Murre (cmurre@ucsd.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

SCIDadh cell line culture

SCIDadh cells were cultured in IMDM 10% FBS, PSG at 37°C in 5% CO2.

Mice

ThymoD pA/pA mice were generated using CRISPR-Cas9 engineering in the embryonic core facility at the University of California, San Diego. All mice were bred in specific pathogen-free condition in accordance with the Institutional Animal Care and Use Committee of the University of California, San Diego.

T cell culture

For OP9-DL1 culture, adult bone marrow single cell suspension were incubated with CD3e (145-2C11), B220 (RA3-6B2), CD11b (RB6-8C5), GR1 (RB6-8C5), TER119 (TER119), CD117 (2B8), CD127 (A7R37) and Flt3 (A2F10). Hematopoietic progenitor cells were collected by depleting lineage-positive cells using Automacs. Selected cells were cultured on OP9-DL1 in 6 well plates with αMEM containing 20% FCS, 2% PSG, FLT3L, recombinant IL-7 1-2 ng/ml at 37°C in 5% CO2.

ThymoD p(A)/p(A) tumor culture

Tumor 3 and tumor 6 were cultured on OP9-DLL1 in MEMa 20% FBS containing 5ng/ml Flt3-L, 5ng/ml IL7, 5ng/ml SCF, PSG at 37°C in 5% CO2. Tumor 5 was cultured without OP9-DLL1 in MEMa and 20% FBS containing 5ng/ml Flt3-L, 5ng/ml IL7, 5ng/ml SCF, PSG at 37°C in 5% CO2. All RNaseq data for tumors were obtained from primary cells. SMC3, RNA polII ChIP-seq was done for cultured tumors 3, 5 and primary tumor 6. DNA FISH was done for cultured tumor 3 and tmor 5.

METHOD DETAILS

Generating SCIDadh pA/pA cell line

WT Cas9 was digested using NcoI/EcoRI from pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene) and cloned into pdCas9-humanized vector (Addgene) digested using NcoI/EcoRI. WT Cas9 and PcI packaging vector were transfected into 293T cells using calcium phosphate precipitation. Virus supernatant was harvested at 48 hours. Spin infection was performed in 1x10^6 cells/ml concentration at 30°C for 90 min. Puromycin (final concentration 10ug/ml) was used for selection. sgRNA was generated by MEGAscript T7 transcription kit and MEGAclear transcription clean-up kit (ThermoFisher Scientific). In order to construct a repair template, gBlock gene fragments (Integrated DNA Technologies) carrying bGH poly(A) with 700bps homology arms that flank the gRNA target site were inserted in pGEM-T easy by NEBuilder HiFi DNA Assembly Master Mix (NEB). pAS knock in was conducted in a Neon transfection system (Invitrogen) using 10 μL tips. 1×10^6 of SCIDadh cells were harvested in a 1.5 mL tube and washed with PBS. Cells were suspended in 45 μL of buffer R. 3-4 μL of sgRNA (~20 ug) and 1-2 μL of repair template (10 μg/μl) were added in cell suspension to make a total 50 μL solution. 2×10^5 Cells were transfected at one time with run#3 (pulse voltage 1500, Pulse width 20, and Pulse no 1), then transferred to 1 well of 12 well plates with 1.5 mL of IMDM without PSG. The next day cells were transferred into Methocult M3630 (STEMCELL technology). Single colonies were picked and transferred into 96 well plates at day 3 and cultured for 5-6 days. Genotyping was performed for the cells isolated from each well.

sgRNA template sequence at upstream of ThymoD CR1

cagctGAATTCTAATACGACTCACTATAGGGGGAAAACACAGACCCTAGTCgttttagagctagaaatagcaagttaaaataaggctagtccGttatcAacttgaaaaagtggcaccgagtcggtgc

sgRNA template sequence at exon3 of thymoD (upstream of CR2)

cagctGAATTCTAATACGACTCACTATAGGGTCTTCAAGGGTGCTATCACAgttttagagctagaaatagcaagttaaaataaggctagtccGttatcAacttgaaaaagtggcaccgagtcggtgc

Primers for genotyping at upstream of ThymoD CR1	
IntCR1-2_F1	ATGGATGGAGAGGTGGACTG
IntCR1-2_R1	CCCTGAGAGAGCCCTAATCC
int5_out	CTCTGTCTGCCAACCCAACT

Primers for genotyping at exon3 (upstream of CR2)	
ThymoDex3F	GGGCAGACGAAACTGACTGT
ThymoDex3R2	AAGCCCTGCCTTGACTGTAA
CR2_ex3_F1	ACCAGGAGAAGAGTGCTGGA

Generating ThymoD pA/pA mice

A sgRNA target site (TCTTCAAGGGTGCTATCACA) as described for SCIDadh pA/pA CR2 mutants was computationally identified in ThymoD exon 3 and verified by transfection of sgRNAs into a pro-B cell line (IM3) that stably expressed Streptococcus pyogenes WT Cas9 using the same method for generating SCIDadh mutant cell line. In order to construct a repair template, homology arms that flank the sgRNA target site were isolated from amplified C57BL/6 genomic DNA. Homology arms flanking a bGH poly(A) addition site were inserted in pGEM-T easy. The inserted poly(A) addition site is located upstream of CR2 region as well as 12kb upstream of the CR1 region (Li et al., 2013). Next a mixture of sgRNA, Cas9 mRNA and repair template were injected in mouse zygotes using standard procedures. The inserted pAS sequence was detected by PCR.

Primer for genotyping

Set 1 F; 5′-GTGTCTCAAGGCGAGAAAGG-3′
Set 1 R; 5′-AAGCTCGCTCTGTTTTGAGG-3′
Set 2 F; 5′-ACTGAACAGCTCTCACCCTC-3′
Set 2 R; 5′-ACTGTAAAGCCCTGGGTCCT-3′

Flow cytometry

For flow cytometry and cell sorting, single cell suspensions of thymus were prepared and analyzed as follows. Dead cells were eliminated using ghost dye red 780. Cells were stained with CD3e (145-2C11), CD4 (GK1.5; RM4-5), CD8a (53-6.7), CD11b (M1/70), CD19 (MB19-1), B220 (RA3-6B2), Gr1 (RB6-8C5), NK1.1(PK136), CD25 (PC61), CD27 (LG3A), CD44 (IM7), CD45.2 (104), CD117 (ack45), TCRb (H57-597), TCRgd (eBioGL3), Ter119 (TER119), and Thy1.2 (53-2.1) antibodies (Becton Dickinson and e-Biosciences). Antibodies were conjugated to fluorescein isothiocyanate, phycoerythrin, peridinin chlorophyll protein-cyanine 5.5, allophycocyanin, allophycocyanin-indotricarbocyanate, Pacific Blue, eFluor 450, Brilliant Violet 421 or were biotinylated. Biotynylated antibodies were stained using QD588 quantum dots (Life Technologies). Data were collected on a LSRII (BD Biosciences) and analyzed with FlowJo software (TreeStar). Cell sorting was performed on a FACSAriaII (BD).

Wright Giemsa staining

Cytospin was done at 1000rpm for 3min. Coverslips were stained by Wright solution for 3 min. Coverslips were washed into each well in 6well plate 2 times. Giemsa staining (1:10) was done for 7 min and washed into each well in 6well plate 2 times. Images were captured by an Olympus BH2 microscope.

Real-time PCR

For quantitative PCR cells were stored in buffer RLT (QIAGEN) with 2-Mercaptoethanol at −80°C. Total RNA was extracted using an RNeasy Mini Kit (QIAGEN) with DNase digestion (QIAGEN) or alternatively with RNeasy Micro Kit (QIAGEN) for small numbers of cells. cDNA was synthesized using a first-strand synthesis kit (Life Technologies) and oligo-dT primers. Conventional PCR was performed for cDNA with SYBR Green Master Mix (Roche). Previously reported primer sets were used for BCL11b and ARP detection (Ikawa et al., 2010).

Primer for qPCR

ThymoD F	GGGCAGACGAAACTGACTGT
ThymoD R	AAGCCCTGCCTTGACTGTAA

3D-FISH

For 3D-DNA FISH fosmid probes were used. WIBR1-1110C15 was used for the Bcl11b enhancer region and obtained from the BACPAC Resource Center (BPRC) at Children's Hospital Oakland Research Institute. Fosmid DNA was labeled with fluorochrome by Nick translation. 20ul solution consisted of 1μg of DNA, 4 μl of 5x Alexa fluor mix, 5x nick translatioin mix, 2 μl of 10x buffer was incubated at 15°C for 4 hour, heat killed by adding 1 μl of 0.5M EDTA pH 8 at 65°C for 10 min, and purified through ProbeQuant G-50 micro column. 2×10^5 cells/40 μl were put on the center of the poly-L-Lysine coated coverslip and cultured in 37°C for 30 min in cell culture incubator. Cells were washed with PBS, fixed at room temperature (RT) for 10 min with 4% paraformaldehyde in 1X PBS, pH 7.2. Fixed cells were quenched at RT for 10 min with 0.1 M Tris-Cl, pH 7.4, washed with PBS and stored in 1xPBS at 4°C for up to 1 month. Cells were permeabilized in 0.1% saponin, 0.1% Triton X-100, 1x PBS for 10 min at RT, incubated 20 min at RT with 20% glycerol, 1X PBS, freeze-thawed in liquid nitrogen three times and rinsed in 1x PBS. For immunofluorescence, cells were blocked at 37°C for 30 min in 5% BSA, 0.1% Triton X-100, 1X PBS and stained using a primary Lamin B1 (M-20) antibody (sc-6217) obtained from Santa Cruz Biotechnology at 1/150 dilution in the blocking buffer at 37°C for 30 min. Primary stained cell were washed for 10 min at RT twice in 1X PBS, 0.1% Triton X-100 at RT with gentle agitation. For second staining, cells were blocked for 30 min at 37°C in 5% BSA, 0.1% Triton X-100, 5% donkey serum, 1X PBS, followed by secondary staining of donkey anti-goat IgG antibody conjugated to Alexa594 (A11058) (Invitrogen). Primary stained cell were washed for 10 min at RT twice in 1X PBS, 0.1% Triton X-100 at RT with gentle agitation, fixed again for 10 min at RT in 2% PFA, 1X PBS and quenched at RT for 10 min with 0.1 M Tris-Cl, pH 7.4, washed with PBS and stored in 1xPBS. For DNA-FISH, cells were denatured for 30 min in 0.1 M HCl at RT, blocked for 1hr at 37°C in 3% BSA + 100 μg/ml Rnase A in 1X PBS, and permeabilized for 30 min at RT in 0.5% saponin, 0.5% Triton X-100, 1X PBS. Cells were then washed once in PBS and stored in 2x SSC until hybridization. For hybridization, nuclear DNA was denatured by incubating coverslips for 2 min and 30 s at 73°C in 2x SSC, 70% formamide solution, followed by an incubation for 1 min in 2x SSC, 50% formamide solution. The hybridization solution contained 100 ng of labeled fosmid probe, 4 μg of mouse Cot-1 DNA, 1 μg of sheared salmon-sperm DNA dissoleved in 50% formamide, 4x SSC and 20% dextran sulfate. The probes were denatured at 73°C for 5 min. Denatured coverslips and probes were sealed and incubated at 37°C for over-night in hybridization oven. On the next day, coverslips were removed and washed once in 2x SSC, 50% formamide solution for 15 min and three times in 2x SSC for 5 min at 37°C with gentle agitation. Cells were washed once with PBS, excess PBS was removed and coverslips were mounted on slides with Prolong gold anti-fade reagent with DAPI (Invitrogen). Images were acquired on a Deltavision microscope using 100x objective lens. 0.2 μm optical sections were obtained in the DAPI, FITC and Red channels. The distances between the nuclear lamina and the BCL11b intergenic region were measured in 2D with ImageJ software.

In Situ Hi-C

We used *in situ* Hi-C with minor modifications to generate a genome-wide contact map essentially consisted of 70 steps as described in the supplementary material (Rao et al., 2014). Five million cells each of wild-type DN2 and ThymoD−/− DN2 cells were generated using OP9-DL1 stromal cells. For step 7, Complete Protease Inhibitor Cocktail, EDTA-free (Roche) was used. One tablet was dissolved in 1 mL dH_2O. 5 μl of 50x Protease Inhibitor was used to 250 mL Hi-C lysis buffer. For step 8 and 9, centrifugation was done at 200x G for 5 min. For step 10, nuclei were resuspended in 0.5% SDS diluted in a final concentration of 1X NEB Buffer 2. For step 11, 5% Triton X-100 was used instead of 10%. For step 12, digestion efficiency was determined by using 3 μl of sample with 87 μl of 1x tris and 10 μl of proteinase-K. Samples were incubated at 55°C for 30 min, at 65°C with shaking for at least 1 hour, and purified with DNA clean & concentrator (Zymo Research). Samples were run on 0.6% agarose gel. For step 27, sheared DNA size should be 300-600 bps. Samples were checked by taking 1μl of sample and run on 2% agarose gel. For step 54-57 each sequencing library was incubated with USER enzyme for 15 min at 37°C after step 56. Final libraries were submitted to paired-end sequencing of 100 bp length on an Illumina HiSeq 2500.

ATAC-Seq

ATAC-seq was performed as previously described (Buenrostro et al., 2013). 50000 cells were used for library preparation. Washed cells were resuspended in lysis buffer. After washing, cells were treated with transposition mix for 30 min at 37°C. DNA was purified by DNA clean & concentrator (Zymo Research). Library fragments were amplified using 1x NEBnext PCR master mix and 1.25 μM of custom Nextra PCR primers 1 and 2 with following PCR conditions: 72°C for 5 min, 98°C for 30 s, followed by thermocycling at 98°C for 10 s, 63°C for 30 s and 72°C for 1 min. Cycle of PCR amplification was determined by q-PCR with Sybr Green in order to stop amplification prior to saturation. Amplified 100-800 bps of Libraries were selected from 2% agarose gel. The quality of the library was checked with Agilent TapeStation and sequenced on an Illumina Hi-Seq4000.

ChIP-Seq

Chromatin was immunoprecipitated with following antibodies. CTCF antibody (07-729) was obtained from Millipore. H3.3 (ab176840), H3K4me1 (ab8895), H3K4me3 (ab8580), RNAPII 8WG16 (ab817) and SMC3 (ab9263) antibodies were purchased from Abcam. Cells were isolated from OP9-DL1 cultures and stained with FACS antibody and ghost dye to sort DN2 cells. Cells were fixed for 10 min in PBS containing 1% formaldehyde. Formaldehyde was quenched with 0.2 M glycine for 10 min. Lineage negative fixed DN2 cells were sorted by FACS AriaII and washed with PBS. Fixed cells were stored at −80°C until use. For preparation for cells treated with actinomycin D, lived cells were selected by ficol-paque plus (GE health care) at day 10. Selected live DN2 cells were suspended with MEM containing 20% FCS, 2% PSG, actinomycin D 10 ug/ml, IL7 and FLT3L and cultured without OP9-DLL1 for 1hr. Cells were harvested at 15, 30, and 60 min and immediately spun down and fixed as described above. Nuclei were isolated in cell buffer mix 10 mM of HEPES/KOH, 85 mM of KCL, 1 mM of EDTA, 1 mM of Benzamidine, 1x Protease inhibitor (Roche), 1% of NP-40 for 10-15 min on ice. Nuclei were spun down for 5 min at 3000 rpm and resuspended in lysis buffer with protease inhibitors. Nuclei were sonicated by Biorupter (Diagenode) with 20 cycles of 30 s on and 30 s off at high setting. Sonicated chromatin was immunoprecipitated with antibodies coated on Protein G-Sepharose for over-night at 4°C. Samples were washed 5 times with LiCl wash buffer, mixing 3 min for each wash on a rotator and washed with TE buffer, mixing 1 min on rotator. Samples were resuspended in 200 μL IP Elution Buffer, incubated with 65°C thermomixer for 3 hr, mixed with 2 μL of protenase-K at 50°C for 1hr. Samples were placed on magnet and transferred to new 1.5ml Eppendorf tube and incubated at 65°C for over-night in hybridization oven for reverse cross-linking. DNA was purified by ChIP DNA clean & concentrator (Zymo Research). Immunoprecipitated DNA was end-repaired, added with dATP by using Klenow exo-, ligated with adaptors from NEBNext Multiplex Oligos for Illumina (NEB) and purified with SPRI beads. Adaptor ligated DNA was amplified with NEBNext Multiplex Oligos for Illumina (NEB) and size-selected with SPRI beads. The quality of the library was checked with Agilent TapeStation. Libraries were run on Illumina Hi-Seq4000.

RNA-Seq

Total RNA was isolated from DN2 cells sorted by AriaII and treated with TURBO DNase (Ambion). mRNA was purified from total RNA by a Dynabeads mRNA purification kit (Life Technologies). cDNA was generated with a First-Strand Synthesis Kit (Life Technologies) and a random hexamers in presence of actinomycin D. Second-strand synthesis was performed with dUTP instead of dTTP. The double-stranded cDNA was sinicated to a length of 200-400 base pairs with S220 Focused-ultrasonicator (Covaris). Sonicated cDNA was ligated to adaptors. Libraries were prepared by TruSeq Stranded mRNA library prep kit (Illumina). Libraries for DN2 cells were sequenced for 50 cycles on Illumine HiSeq 2500.Total RNA from Tumor samples were isolated from whole enlarged thymus, prepared for library as same as DN2 RNA, and sequenced for 50 cycles on Illumina HiSeq 4000.

GRO-seq

Duplicates of GRO-seq experiments were performed. Nuclei from 5 million cells were isolated by hypotonic lysis. Nuclear run on (NRO) reaction was done for 5 min at 30°C in the presence of BrUTP 750 μM, ATP 750 μM, GTP 750 μM, reduced concentrations of CTP 4.5 μM and 1.65% sarkosyl. Total RNA was purified with Trizol and isopropanol precipitation, DNase treated (TURBO DNase, Ambion), fragmented with 2ul of fragmentation regent (Ambion AM1907). Fragmented RNA was re-buffered by P30 RNase-free spin column (Bio-Rad). RNA fragments were 3′ dephosphorylated with T4 polynucleotide kinase. BrUTP-labeled run-on RNA was immunopurified with anti-BrdUTP-coated agarose beads, washed, and EtOH-precipitated. Run-on RNA was de-capped with tobacco acid pyrophosphatase from FirstChoice RLM-RACE kit (Thermo Fisher Scientific), 5′ phosphorylated with polynucleotide kinase (NEB) and purified with Trizol LS/isopropanol precipitation.3′ adaptor ligation was done by ligationg a single-stranded, 3′-blocked, 5′′-adenylated 3′ oligonucleotide with mutant (K227Q) truncated RNA ligase 2 (NEB) to the 3′ end of the RNA fragments, followed by annealing a reverse transcription primer complementary to the 3′ adaptor to suppress adaptor dimer formation, and ligating a hybrid 5′ DNA-RNA oligonucleotide using RNA ligase I and reverse-transcribing with SuperScript III reverse transcriptase (ThermoFisher Scientific). The cDNA was purified with SPRI beads. The libraries were amplified with primers bearing primer landing site compatible with illumine sequencing. The libraries were size-selected with 2% agarose containing SYBR gold to 60-110 bp insert size, followed by the quality of the library was checked with Agilent TapeStation and then sequenced for 50 cycles on Illumina HiSeq 2500.

Whole genome bisulfite sequencing (WGBS)

Library preparation for WGBS was previously described (Benner et al., 2015). DNA was isolated from in-vitro cultured WT DN2 and ThymoD pA/pA DN2 cells. 1 μg of genomic DNA mixed with unmethylated lamda DNA at a concentration of 0.5% of total DNA was sonicated by Biorupter with 20 cycles (30 s on and 30 s off at low condition). Fragmented DNA was end-repaired using the End-It procedure (Epicenter), and incubated in the presence of Klenow (3′ to 5′ exo minus) for adding an A base to the 3′ end. Next TruSeq adapters were ligated to fragmented DNA and purified using 2% Agarose gel electrophoresis. Bisulfite conversion was performed as described by the manufacturer (MethylCode). Bisulfite-treated DNA was amplified by using a TruSeq PCR primer mixture and Pfu Turbo Cx Polymerase, agarose purified, and sequenced on an Illumina HiSEq 2500 sequencer.

QUANTIFICATION AND STATISTICAL ANALYSIS

Filtering and Normalization of HiC Reads

Reads were first trimmed from the 3′ end of sequences to GATC (MboI restriction enzyme site). Trimmed reads were aligned to mm9 with Bowtie software with parameters–chunkmbs 128–mm -m1–best–strata -p4 -S. The remainder of the analysis was performed using Homer (Heinz et al., 2010). Only unique reads and paired end tags with more than 1.5x fragment length on the same chromosome separating them were retained. Sequence reads were checked for GC and nucleotide bias. HiC reads derived from wild-type DN2 and ThymoDpA/pA DN2 cells were compared to pro-B and multipotent progenitors (Lin et al., 2012). Tag counts for all directories were normalized to the sample with lowest sequencing depth. For multipotent progenitors, pro-B and wild-type DN2 comparisons a map resolution of 50 kb was chosen to ensure that > 80% of loci were associated with at least 1000 genomic contacts. A superRes of 100 kb was used for comparison of wild-type DN2 and ThymoDpA/pA DN2 cells. A resolution of 40 kb was chosen to ensure that greater than 80% of loci were associated with at least 1000 contacts. Hi-C heatmaps were normalized to output the ratio of observed to expected interactions by assuming each region has an equal chance of interacting with other regions in the genome and that regions are expected to interact depending on their linear distance along the chromosome (Lin et al., 2012). Hi-C heatmaps were visualized using Java Tree View. Circos diagrams were generated using Circos software. Only interactions with p values $\leq$ 0.0001 are shown. Line thickness of Circos diagrams refers to p values. Darker lines depict lower p values. In depth explanations of normalization, generation of Hi-C correlation matrices, principal component analysis (PCA) and identifying significant interactions were performed as previously described (Lin et al., 2012).

Analysis for ATAC-Seq

ATAC-Seq data were mapped to reference genome mm9 by Bowtie (version 1.1.1), using following option:–best -m 1, and others as default. The bam file were then processed by Homer to create tag directory with the default setting of Homer makeTagDirectory, using option -tbp 1. For ATAC open region, following options were used: localSize 50000 -size 150-minDist 50 –fragLength 0, and we used the default setting for other peak/region calling.

ChIP-Seq

ChIP-seq data were mapped to reference genome mm9 by Bowtie (version 1.1.1), using following option:–best -m 1, and others as default. The bam file were then processed by Homer to create tag directory with the default setting of Homer makeTagDirectory, using option -tbp 1. Homer findPeak has been used to identify enriched regions.

Analysis of CTCF direction and significant interaction

p values associated with CTCF-CTCF interactions were computed as described by HOMER with the following modification. Bins were centered on CTCF bound site in wild-type or ThymoD p(A)/p(A) DN2 cells. CTCF directionality was identified by scanning using FIMO. The following parameter were used–motif CTCF_MOUSE.H10MO.A–text–thresh 0.01. MEME's mouse DNA-binding protein motif database was used (HOCOMOCOv10_MOUSE_mono_meme_format.meme). For multiple significant CTCF motifs that showed opposite directions within a CTCF binding site interactions originating from that site were identified as "unknown."

Analysis of SMC3 Occupancy in DN2 Cells Exposed to Actinomycin D

DN2 cells were incubated in the presence of actinomycin D, formaldehyde fixed and analyzed for SMC3 occupancy using ChIP-seq as described above. Active transcription start sites were identified using RNA polymerase II ChIP-seq. Active TSS sites were defined by RNAPII peaks spanning a 2 kb window centered on the TSSs. –log 10 p values of the Fisher test were calculated for the contingency table. SMC reads were determined across the ThymoD locus (chr12: 108.345 –108.405 Mb) in DN2 cells treated in the absence or presence of actinomycin D and normalized to the total number of reads (10^7 read). The mean and standard deviation for each experiment were derived from 20 independent samples using boot strapping for all reads across the genome.

RNA-Seq

RNaseq data were examined using Omics Pipe applying the RNA-seq count-based differential expression analysis pipeline. Quality control of the raw fastq files was performed using FastQC (Babraham Bioinformatics). Sequencing reads were aligned to the mouse genome (mm10) applying STAR aligner. Read quantification (exons) was performed using htseq-count with UCSC RefSeq annotation. The R BioConductor package DESeq2 was used to calculate size factors in order to normalize library sizes across replicates and calculate means and variances based on a negative binomial distribution model in order to detect differentially expressed genes based on adjusted p values of < 0.05. Functional enrichment of the differentially expressed genes was done using ToppGene Suite, WebGestalt and Metascape.

Analysis of GRO-seq

GRO-seq data were mapped to reference genome mm 9 by Bowtie (version 1.1.1), using following option:–best -m 1, and others as default. The bam file are then processed by Homer to create tag directory with the default setting of Homer makeTagDirectory, using option -tbp 1. UCSCfiles were generated with default setting with -strand separate.

Analysis of WGBS

Analysis of bisulfite sequencing was performed using BSSeeker2 (Guo et al., 2013), HOMER (Heinz et al., 2010), and custom awk scripts. Specifically, a bisulfite-sequencing amenable mm9 reference genome was built using the BSSeeker2 script *bs_seeker2-build.py* with default options. Paired-end sequencing data were aligned to this reference genome using the BSSeeker2 script *bs_seeker2-align.py* allowing 6 mismatches (-m 6) and a fragment length between 0 and 800 bases (-I 0 -X 800). PCR duplicates were removed from paired end data using picard tools (https://broadinstitute.github.io/picard/). Unpaired reads were saved, aligned individually with the same options, and merged with paired-end data. Methylation levels were called using the BSSeeker2 script *bs_seeker2-call_methylation.py* with default settings. BSSeeker2 methylation call data were reformatted to the allC format (Lister et al., 2009) and tag directories were created using HOMER's *makeTagDirectory* command with the -minCounts 5 option, thus including only those cytosines covered by at least 5 reads. Downstream analysis comparing DNA methylation levels between wild-type and *ThymoD* pA/pA cells included only those cytosines covered by 5 reads in both datasets.

DATA AND SOFTWARE AVAILABILITY

Genome-wide reads were deposited at GEO. They are accessible for analysis at GEO: GSE90958 and at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=avcrkuikxvsxxmr&acc=GSE90958.

Acknowledgments

Chapter 3, in full, is a reformatted reprint of the material as it appears as "Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate" in Cell, 2017. Isoda, T., Moore, A.J., He, Z., Chandra, V., Aida, M., Denholtz, M., van Hamburg, J.P., Fisch, K.M., Chang, A.N., Fahl, S.P., Wiest, D.L. and Murre, C,. The dissertation author was the primary investigator and author of this paper.

Chapter 4: Upon microbial challenge human neutrophils undergo rapid changes in nuclear architecture and chromatin folding to orchestrate an immediate inflammatory gene program

4.1 Abstract

Differentiating neutrophils undergo large-scale changes in nuclear morphology. How such alterations in structure are established and modulated upon exposure to microbial agents is largely unknown. Here, we found that prior to encounter with bacteria, an armamentarium of inflammatory genes was positioned in a transcriptionally passive environment suppressing premature transcriptional activation. Upon microbial exposure, however, human neutrophils rapidly (<3 hours) repositioned the ensemble of pro-inflammatory genes towards the transcriptionally permissive compartment. We show that the repositioning of genes was closely associated with the swift recruitment of cohesin across the inflammatory enhancer landscape permitting an immediate transcriptional response upon bacterial exposure. We found that activated enhancers, marked by increased deposition of H3K27Ac, were highly enriched for cistromic elements associated with PU.1, CEBPB, TFE3, JUN and FOSL2 occupancy. These data reveal how upon microbial challenge the cohesin machinery is recruited to an activated enhancer repertoire to instruct changes in chromatin folding, nuclear architecture and to activate an inflammatory gene program.

4.2 Introduction

The organization of the human genome within the nucleus is central to the control of gene expression and thus cell identity and function. At the largest scale, the genome is folded into chromosome territories that, with the exception of nucleoli, rarely intermingle. However, chromosomes are not randomly distributed across the nucleus. Large and gene poor chromosomes are predominantly positioned at the lamina whereas small and gene rich chromosomes concentrate in the nuclear interior (Fritz et al. 2016). Chromosomes themselves fold into loop domains that physically associate to establish the transcriptionally repressive or inert heterochromatic B compartment or transcriptionally permissive euchromatic A compartment (Dixon et al. 2012; Lieberman-Aiden et al. 2009). The heterochromatic compartment is highly enriched at the nuclear lamina whereas the euchromatic compartment is positioned in the nuclear interior (Kosak et al. 2002).

Loop domains are established in part by the CTCF protein (Dixon et al. 2012; Rao et al. 2014). Convergently oriented pairs of CTCF-bound loci can form CTCF-anchored loops, generated by recruitment of the cohesin complex (Rao et al. 2017; Nora et al. 2017). Cohesin is loaded onto transcribed regions located throughout loop bodies (Busslinger et al. 2017). Once sequestered, the cohesin complex extrudes chromatin in a progressive manner until a pair of convergent CTCF bound sites are reached, a process termed loop extrusion (Fudenberg et al. 2016). Gene promoters connected to transcriptional enhancers by CTCF-mediated loops tend to be highly expressed (Rao et al. 2014), and CTCF occupancy at nearby sites contributes to the maintenance of gene expression and stable chromatin structure (Rao et al. 2017; Nora et al. 2017; Schwarzer et al. 2017; Bintu et al. 2018).

Human neutrophils are abundant, short-lived circulating white blood cells that are critical first-responders to infection and tissue damage. Upon injury or infection, neutrophils exit the circulation via extravasation, migrate towards damaged tissues or infectious foci, phagocytose small pathogens, release reactive oxygen and nitrogen species, and extrude their chromatin as cytotoxic granule-laced extracellular traps (NETs). In addition to their direct role in killing invading pathogens, activated neutrophils rapidly induce the expression of a wide range of cytokines and chemokines to orchestrate an immediate inflammatory response (Ley et al. 2018).

The nuclei of polymorphonuclear (PMN) neutrophils are composed of multiple distinct but internally continuous lobes allowing them to swiftly migrate between (paracellular route) or through (transcellular route) endothelial cells that line blood vessels and interstitial spaces of tissues while maintaining their nuclear integrity (Rowat et al. 2013; Olins et al. 2009; Muller 2013). The Lamin B Receptor (LBR) is an important determinant for imposing a multi-lobular nuclear architecture on neutrophils (Shultz et al. 2003; Hoffmann et al. 2002). Neutrophils of mice deficient in the *Lbr* gene fail to adopt a multi-lobular nuclear shape (Shultz et al. 2003), and mouse neutrophilic cell lines lacking *Lbr* cannot form characteristic toroidal nuclei during differentiation (Zhu et al. 2017). Similarly, humans with *LBR* mutations manifest the Pelger-Huët anomaly, characterized by a reduction in nuclear lobe number in granulocytes (Hoffmann et al. 2002).

Chromatin folding in murine neutrophils is highly enriched for remote genomic interactions, primarily involving heterochromatic regions. These interactions span vast genomic distances resulting in large-scale chromosome condensation. Terminal differentiation of murine neutrophils is also associated with the relocation of centromeres, pericentromeres, telomeres, LINE elements, and ribosomal DNA from the nuclear interior to the nuclear lamina, a process that requires the *Lbr* gene (Zhu et al. 2017). As neutrophils differentiate, the LBR deforms the malleable nuclear envelope by wrapping it around the heterochromatic component of the neutrophil genome, resulting in its characteristic lobed shape.

Upon reaching a tissue site of infection, neutrophils neutralize bacteria in multiple ways: (i) engulfment through phagocytosis; (ii) degranulation to release microbicidal factors into the extracellular space; (iii) release of extracellular traps or NETs that are composed of extruded chromatin fibers and antimicrobial factors; and (iv) rapid induction of cytokine gene expression to coordinate a broader immune response (Ley et al. 2018; Brinkmann et al. 2004). To detect and respond appropriately to diverse invading pathogens, neutrophils express a variety of pattern recognition receptors including cell surface and endolysosomal Toll-like receptors (TLRs), C-type lectin receptors, and formyl peptide receptors, among others. Once activated a variety of downstream signaling pathways converge on the NF-κB and AP1 transcription factors to induce an inflammatory gene program including the cytokines and chemokines IL-8/CXCL8, TNFα, IL-1β, IL-17, and IFNγ (Thomas and Schroder 2013; Garcia-Romo et al. 2011; Tecchio et al. 2014).

The mechanisms by which pathogen sensing pathways interface with the neutrophil genome to induce a rapid and stimulant-appropriate inflammatory gene expression program remain unclear. Here we found that human neutrophil genomes display highly segmented compartments and contracted heterochromatin when compared to human embryonic stem cells. Upon microbe encounter, a specific subset of modestly euchromatic subdomains, spatially segregated from the highly euchromatic A compartment, displayed strengthening of their euchromatic character, and relocalized from a peri-nuclear envelope position towards the nuclear interior. Prominent among the regions that repositioned during human neutrophil activation were gene loci associated with an activated neutrophil-specific gene expression program. Microbial-induced changes in long-range chromatin interactions were globally associated with rapid loss of insulation at euchromatic subdomain boundaries, as well as the formation of *de novo* chromatin loops linking immune response genes to pre-existing and *de novo* formed transcriptional enhancers. The loop-mediated juxtaposition of inflammatory genes to transcriptional enhancers upon microbial exposure was closely associated with the deposition of histone 3 lysine 27 acetylation (H3K27ac), an enhancer-associated histone modification, and rapid loading (<3 hours) of the cohesin complex at the subset of enhancer elements that control an inflammatory gene program. Based on these observations, we propose that the microbe-induced transcriptional signature of activated neutrophils is driven by activated enhancer repertoires. Activated enhancers marked by elevated levels of H3K27Ac, in turn, rapidly recruit the cohesin machinery to dictate changes in chromatin folding and nuclear positioning of genes associated with an inflammatory gene program.

4.3 Results

Human neutrophil development is associated with segmented compartments and contracted genomes

Neutrophil nuclei undergo dramatic morphological changes during differentiation from multipotent progenitors, with terminally differentiated neutrophil nuclei having 3-5 internally continuous but spatially distinct lobes. To characterize the genomic interactions established during the development of PMN cells, neutrophils were isolated from human peripheral blood, formaldehyde-fixed, and analyzed using *in situ* HiC (Supplemental Table S1) (Rao et al. 2014). The genomes of human neutrophils were slightly enriched for inter-chromosomal interactions when compared to human embryonic stem cells (hESCs) (Fig. 4.1A). Chromosome territories remained intact and we found no evidence of individual chromosomes being split across multiple lobes. Notably, compared to hESCs, human neutrophils were depleted for genomic interactions that spanned less than 3Mb but were enriched for interactions that covered more than 3Mb (Fig. 4.1A).

We next constructed contact matrices for hESCs cells and human neutrophils (Fig. 4.1B). We found that a larger fraction of the neutrophil genome was sequestered in the B compartment when compared to hESCs. The stereotypic plaid pattern, resulting from the spatial segregation of the A and B compartments, was much more pronounced in human neutrophils compared to hESCs (Fig. 4.1B). Intra-chromosomal and inter-chromosomal interactions between A and B compartments were both less prevalent in neutrophils versus hESCs (Fig. 4.1C). Conversely, long-range genomic interactions across the B compartment were significantly more extensive in human neutrophils than hESCs (Fig. 4.1C). During differentiation, large genomic regions that exhibited a continuum of either positive or negative PC1 values in hESCs fragmented into smaller genomic regions that switched PC1 values in neutrophils (Fig. 4.1C,D). Many of the genomic regions that switched from negative to positive PC1 values during development were associated with a neutrophil-specific transcription signature, whereas those regions switching from positive to negative PC1 values were associated with silencing of lineage-inappropriate genes. Notably, the hyper-segmentation of compartment domains in the neutrophil genome established *de novo* loop domain and compartment boundaries (Fig. 4.1E). Specifically, although more than 75% of loop domain boundaries identified in hESCs were conserved in neutrophils, less than 40% of loop domain boundaries in neutrophils were present in hESCs (Fig. 4.1F). Overall compartment boundaries were poorly conserved between these two cell types (Fig. 4.1F). Genome-wide analysis of cell type-specific loop domain and compartment boundary element insulation strength confirmed this finding, indicating the existence of cell type-specific boundaries that were associated specifically with either hESCs or human neutrophils, in addition to shared boundaries (Fig. 4.1G). Taken together our data reveal that human neutrophils, when compared to hESCs, are characterized by a contracted genome with increased enforcement of compartmentalization and highly segmented A and B compartments.

PMA-induced activation of neutrophils rapidly modulates nuclear architecture

Upon detecting invading microbes, neutrophils rapidly activate an inflammatory-specific transcription signature. As a first approach to examine whether and how the nuclear architecture of neutrophils responds to inflammatory signals, HiC was performed on neutrophils cultured in both the absence and presence of the canonical neutrophil activator phorbol 12-myristate 13-acetate (PMA), a protein kinase C. PMA stimulation of human neutrophils resulted in a global decrease in short-range intra-chromosomal interactions and a global increase inter-chromosomal interactions (Fig. 4.2A), while exerting minimal effects on A-B compartmentalization and loop domain boundaries (Fig. 4.2B). Likewise, PMA-induced activation did not trigger large-scale switching of genes or regulatory elements between the A and B compartments (Fig. 4.2C). However, further scrutiny of chromatin folding across the A compartment revealed a small but significant number of discrete genomic regions that underwent significant PMA-dependent changes from low but positive PC1 values to highly positive PC1 values, indicating an increase in euchromatic character (Fig. 4.2C). We refer to these regions as PMA ΔPC1 domains (Fig. 4.2C, Methods). Notably, PMA ΔPC1 domains were strongly enriched for genes implicated in the neutrophil defense response, including genes downstream of key innate immune receptors such as the complement receptors, FCγ receptor, and dectin-2, as well as genes implicated in cell migration and regulation of lysosomal pH (Fig. 4.2D).

Analysis of intra-chromosomal HiC contact matrices revealed few significant changes across chromosome 18 following PMA-induced activation (Fig. 4.2E). PMA ΔPC1 domains, however, often displayed changes in interactions with euchromatin, both in their immediate vicinity, as well as across chromosome 18 (Fig. 4.2F). Direct measure of genome-wide changes in chromatin organization showed that in activated neutrophils PMA ΔPC1 domains showed large-scale changes in contact frequencies (Fig. 4.2G). Specifically, ~25% of PMA ΔPC1 domains fell within the top 10% most differentially interacting genomic regions (Fig. 4.2G). Likewise, PMA ΔPC1 domains on average displayed significantly lower chromatin interaction correlation with unstimulated neutrophils when compared to the genome as a whole (Fig. 4.2H). Taken together, these data indicate that in PMA-activated neutrophils genic and intergenic domains associated with innate immune genes increase their euchromatic character and undergo alterations in remote genomic interactions.

Upon microbial exposure a subset of neutrophil inflammatory genes increases their euchromatic character

To validate the alterations in neutrophil euchromatic character using a physiologically relevant stimulus, human neutrophils were cultured in the presence of live *Escherichia coli* bacteria for a period of three hours. *E. coli* co-cultured neutrophils were isolated, formaldehyde cross-linked and analyzed using HiC. Genomes of human neutrophils cultured in the presence of *E. coli* only displayed minor alterations in contact frequencies, maintained overall compartment and loop domain structures (Fig. 4.3A,B), and remained essentially free of detectable A-B compartment switching (Fig. 4.3C). However, similar to PMA-activated neutrophils, a distinct subset of genomic regions positioned in the A compartment displayed a substantial increase in euchromatic character upon *E. coli* encounter (Fig.4. 3C, *E. coli* ∆PC1 domains). Notably, the *E. coli* ∆PC1 domains included genes encoding for cytokines and chemokines, genes associated with neutrophil degranulation, and genes linked with the inflammatory response (Fig. 4.3D).

Similar to PMA-activated neutrophils, *E. coli* co-cultured neutrophils showed few large-scale changes in chromatin organization compared to unstimulated neutrophils (Fig. 4.3E). *E. coli* ∆PC1 domains, however, showed dramatic increases in genomic interactions involving neighboring euchromatic regions, as well as the remainder of the chromosome upon co-culture with *E. coli* (Fig. 4.3F). Similar to PMA ∆PC1 domains, *E. coli* ∆PC1 domains were among the most restructured genomic regions in response to *E. coli*, with 25% of *E. coli* ∆PC1 domains assigned to the top 15% of the most differentially interacting regions globally (Fig. 4.3G). *E. coli* ∆PC1 domains overall displayed significantly lower correlation with unstimulated neutrophil genome structure than the remainder of the genome (Fig. 4.3H). These data indicate that upon microbial exposure, a subset of genes associated with an inflammatory response increase their euchromatic character.

We next sought to ascertain to what degree ∆PC1 domains differed between stimuli. *E. coli* ∆PC1 domains only partially overlapped with PMA ∆PC1 domains. The identities of genes in ∆PC1 domains also depended on the stimulus that neutrophils encountered. *E. coli*-specific ∆PC1 domains were highly enriched for chemokine and cytokine genes as well as genes involved in chemotaxis (Supplemental Fig. S2B). In contrast, PMA-specific ∆PC1 regions were enriched for defensin gene clusters. These data suggest that the changes in euchromatic character regulate stimulant-appropriate inflammatory responses. Supporting this hypothesis, genes residing in stimulus-specific ∆PC1 domains underwent stimulus-specific changes in gene expression. Genes in *E. coli* ∆PC1 domains were more highly expressed upon *E. coli* encounter than upon PMA stimulation, whereas genes in PMA ∆PC1 domains were more highly expressed upon PMA stimulation than during *E. coli* co-culture. Taken together, these data indicate that neutrophil activation enhances the euchromatic character of a subset of inflammatory response gene loci in a stimulus-dependent manner.

Rapid assembly and relocalization of a CXCL transcriptional hub upon E. coli encounter

To determine how euchromatic character is strengthened upon microbial activation, we focused on an archetypal *E. coli* ∆PC1 domain containing inflammatory-specific genes encoded within the extended *CXCL* locus. The *CXCL* locus spans a cluster of genes encoding a class of chemokines that include *CXCL8* (IL8), *CXCL1*, and *CXCL2* (MIP2α), each of which is rapidly induced when exposed to microbial agents. We found that in unstimulated neutrophils the *CXCL* locus exists as a loop domain associated with a modestly positive PC1 score which is insulated from neighboring euchromatin (Fig. 4.4A). Notably, within three hours of exposure to *E. coli*, the euchromatic character of the *CXCL* locus was significantly strengthened (Fig. 4.4A), accompanied by large scale changes in chromatin folding, with genomic interactions and transcriptional activation spreading into neighboring regions (Fig. 4.4B,C).

To determine whether the alterations in genome folding were associated with gene expression, activated neutrophils were analyzed for transcript abundance as well as CTCF and SMC3 occupancy (Fig. 4.4C). As expected, CXCL8, CXCL1, and CXCL2 transcript abundance was significantly elevated upon *E. coli* encounter (Fig. 4.4C). Notably, a recently described non-coding genomic region located immediately upstream of *CXCL8* was also transcriptionally induced upon exposure to bacteria (Fig. 4.4C) (Fanucchi et al. 2019). While CTCF occupancy was elevated at a site closely linked with the *CXCL8* locus, other CTCF bound sites in the locus were not modulated upon activation (Fig. 4.4C). In contrast, we found that *E. coli* encounter substantially enriched cohesin occupancy across the locus (Fig. 4.4C). Cohesin occupancy was particularly prominent at sites closely associated with *de novo* loops that linked the *CXCL8*, *CXCL1,* and *CXCL2* gene bodies, promoter regions, and SMC3-enriched intergenic regions into a shared transcriptional hub (Fig. 4.4C).

To validate these findings in single cells, we performed fluorescence *in situ* hybridization (FISH) using a probe corresponding to the *E. coli*-specific *CXCL* ∆PC1 domain (Fig. 4.4D). In unstimulated neutrophils the *CXCL E. coli* ∆PC1 domain localized near the nuclear periphery (Fig. 4.4D). Upon *E. coli* encounter the *CXCL E. coli* ∆PC1 domain rapidly relocated away from the heterochromatic nuclear periphery towards the nuclear interior, concomitant with its change in euchromatic character and elevated transcript levels (Fig. 4.4D). Specifically, the *E. coli* ∆PC1 domain relocated from the DAPI-dense portion of the nucleus near the nuclear periphery to the DAPI-sparse nuclear interior (Fig. 4.4E,F). This change in nuclear positioning was not an indirect result of changes in nuclear morphology, nor activation-induced loss of nuclear integrity, as heterochromatic control probes remained tightly associated with the nuclear periphery during *E. coli* encounter. Collectively these observations indicate that upon microbial exposure human neutrophils rapidly remodel nuclear architecture to assemble a *CXCL* transcriptional hub in the nuclear interior.

Neutrophil activation is associated with global loss of insulation at inflammatory genes

The data described above reveal that when human neutrophils encounter bacteria, a subset of inflammatory genes undergo large-scale changes in chromatin folding that spread into neighboring loop domains. To quantitatively describe this loss of subdomain insulation, we computed the insulation scores for genomic regions that surrounded the boundaries of *E. coli* ∆PC1 domains, and upon microbial exposure gained euchromatic character to merge with surrounding euchromatin. We found that upon *E. coli* encounter the gain of euchromatic character across *E. coli* ∆PC1 domains was closely associated with decreased insulation strength at *E. coli* ∆PC1 domain boundaries (Fig. 4.5A). Although globally the genomic distances separating chromatin interaction anchor points were significantly decreased in activated versus unstimulated neutrophils, the distance separating anchor points of chromatin interactions with *E. coli* ∆PC1 domains increased (Fig. 4.5B). This loss of insulation at *E. coli* ∆PC1 domain boundaries suggests *de novo* formation of regulatory interactions with the surrounding area (Fig. 4.5B). Additionally, chromatin interactions contained entirely within *E. coli* ∆PC1 domains were found on average to be significantly stronger in unstimulated neutrophils as compared to *E. coli* co-cultured neutrophils, suggesting a loss of subdomain structure and self-association during microbial encounter (Fig. 4.5C).

Given the loss of insulation at *E. coli* ∆PC1 domain boundaries, we next sought to determine the relationship between ∆PC1 domains and gene regulatory chromatin interactions. Although chromatin interactions within *E. coli* ∆PC1 domains were on average weakened during *E. coli* encounter (Fig. 4.5C), *E. coli* encounter-specific chromatin loops within ∆PC1 domains were significantly stronger than chromatin loops found only in unstimulated neutrophils (Fig. 4.5D). This finding suggested a gene regulatory role for *E. coli* encounter-dependent loops, and a tight link between these loops and *E. coli* ∆PC1 domains. Supporting this finding, *E. coli*-dependent chromatin loops were generally closer to *E. coli* ∆PC1 domains than were unstimulated neutrophil-specific chromatin loops, and 11% of *E. coli* dependent chromatin loops were identified in *E. coli* ∆PC1 domains, which make up only 0.3% of the genome (Fig. 4.5E). Importantly, genes near *E. coli* co-culture-specific chromatin loop anchors were significantly more highly expressed than genes at chromatin loop anchors found only in unstimulated neutrophils (Fig. 4.5F).

Given the enrichment of neutrophil inflammatory response genes in *E. coli* ∆PC1 domains (Fig. 4.3D) and the link between expression levels and an increase in euchromatic character, we next determined the relationship between euchromatic character (PC1 score) and transcript levels during microbe encounter. Notably, we found a strong correlation between PC1 score dynamics and transcriptional dynamics, with the most highly induced genes also showing the largest increases in PC1 score, and the most repressed genes showing the largest decreases in PC1 score (Fig. 4.5G). These phenomena are readily visible at a number of inflammatory loci, wherein the tight self-association of *E. coli* ∆PC1 domains in unstimulated cells is lost in favor of distal regulatory interactions and transcriptional activation during *E. coli* encounter (Supplemental Fig. S4).

Taken together these data indicate that neutrophil transcriptional state, euchromatic character, and spatial localization of genes are closely linked.

Microbial exposure rapidly recruits cohesin to inflammatory enhancers

Examination of gene regulatory interactions associated with *E. coli* ∆PC1 domains (Fig. 4.4) hinted that a large number of *E. coli*-dependent interactions were associated with the recruitment of the cohesin complex to cis regulatory elements. To study this phenomenon and understand its role in *E. coli*-dependent changes in gene expression, we analyzed unstimulated and *E. coli*-exposed neutrophils for SMC3 and CTCF occupancy, as well as changes in H3K27Ac marked enhancer repertoires and transcription. We then focused our analysis on a specific subset of SMC3-amassed enhancers: those H3K27ac-defined enhancers present in *E. coli* co-cultured neutrophils that gained substantial SMC3 occupancy during *E. coli* encounter (Fig. 4.6A, Methods).

SMC3-amassed enhancers are characterized by modestly increased CTCF binding, and substantially increased H3K27ac deposition and polyadenylated RNA abundance (Fig. 4.6A). Supporting the importance of SMC3 and H3K27ac deposition at new regulatory interactions, *E. coli* co-culture-specific chromatin interaction anchors were found to be enriched for SMC3 occupancy and H3K27ac deposition, with only modest changes in CTCF occupancy compared to unstimulated neutrophil-specific interactions. Similarly, H3K27ac-defined enhancers enriched for polyadenylated RNA signal were likewise enriched for SMC3 occupancy and H3K27ac deposition.

Supporting the importance of SMC3-amassed enhancers in *E. coli* ∆PC1 domain behavior, SMC3-amassed enhancers were enriched at *E. coli* ∆PC1 domains (Fig. 4.6B). SMC-amassed enhancers were also, on average, localized closer to ∆PC1 domains when compared to the global enhancer repertoire, and were more enriched in ∆PC1 domains than the entire enhancer repertoire (Fig. 4.6C).

To understand the mechanism of cohesin targeting to SMC3-amassed enhancers, we identified transcription factor binding motifs enriched within both SMC3-amassed enhancers and non-SMC3-amassed enhancers found in *E. coli* co-cultured neutrophils. We then computed the enrichment of transcription factor motif density in SMC3-amassed enhancers compared to all enhancers found in *E. coli* co-cultured neutrophils. Validating this approach, we found that DNA binding motifs associated with known inflammatory regulating factor AP1 including JUND and FOSL2, as well as transcriptional regulators that orchestrate neutrophil differentiation and physiology including CEBPB, CEBP homologue CHOP, and PU.1, were significantly enriched across the bacterial-induced SMC3-amassed enhancer repertoire compared to the entire enhancer repertoire (Fig. 6D, top and Fig. 4.6E). Notably, transcript abundance associated with these factors was elevated in neutrophils exposed to *E. coli* (Fig. 4.6D, bottom). Apart from known inflammatory and myeloid regulatory transcription factors, we also found that DNA sequences associated with TFE3 occupancy were enriched at SMC3-amassed enhancers (Fig. 4.6D and 4.6E).

Recent studies revealed that in activated macrophages and microglial cells TFE3 orchestrate an inflammatory program of gene expression (Pastore et al., 2016). Notably, upon activation TFE3 relocated from the cytoplasm to the nuclear interior (Pastore et al., 2016). As a first approach to determine whether likewise TFE3 translocates from the cytoplasm to the nucleus we examined naïve and microbial-exposed neutrophils for TFE3 localization using immunofluorescence. We found that in naive neutrophils TFE3 was predominantly localized in the cytoplasm. Notably, however, we found that within three hours upon exposure to *E. coli* TFE3 repositioned from the cytoplasm to the nuclear interior in the vast majority of neutrophils.

Genes interacting with SMC3-amassed enhancers were next analyzed for functional group enrichment. Notably, bacterial-induced SMC3-amassed enhancers were closely associated with genes involved in neutrophil activation, including cytokine signaling and response, chemotaxis, and degranulation (Fig. 4.6F). Analysis of RNA-seq data revealed that genes interacting with enhancers in general showed little preference to be induced upon *E. coli* encounter as compared to any other gene in the genome. In contrast, genes linked to SMC3-amassed enhancers showed a significant increase in gene expression during *E. coli* encounter compared to genes globally, or to genes linked to enhancers in general (Fig. 4.6G). This phenomenon appears to depend on both SMC3 occupancy and H3K27ac deposition, as interactions with either SMC3-amassed *E. coli*-specific enhancers or SMC3-amassed pre-existing enhancers were both associated with increased gene expression, whereas interactions with enhancers only found in unstimulated cells were not associated with increased gene expression, regardless of SMC3 occupancy.

Taken together, these data indicate that upon bacterial exposure human neutrophils rapidly sequester the cohesin machinery at a specific subset of enhancers to modulate chromatin folding and activate an inflammatory gene transcription program.

4.4 Discussion

The unique morphology of neutrophils has been an enigma since its discovery more than a century ago (Cavaillon 2011). How neutrophil genomes are folded into three-dimensional space and how neutrophil nuclear architecture is altered upon microbial exposure has remained largely unknown. Here we used a genome-wide chromosome conformation capture approach (HiC) to address these questions. We found that human neutrophil nuclei, when compared to embryonic stem cells, displayed a distinct nuclear architecture: (i) a decline in genomic interactions across loop domains (< 3Mb); (ii) a segmentation of large, continuous A and B compartments into numerous small compartments, resulting in the establishment of new compartment and loop domain boundaries; and, (iii) an increase in remote chromosomal interactions across loop domains (> 3Mb). This increase in long-range genomic interactions primarily involved heterochromatic regions indicating a key role for heterochromatic interactions in influencing human neutrophil genome topology. Our data are consistent with previous studies involving murine neutrophils that also displayed a highly contracted genome when compared to progenitor cells and show that key features of neutrophil genome structure are conserved between the murine and human genomes (Zhu et al. 2017).

The neutrophil genome undergoes large-scale alterations in morphology upon bacterial encounter. Using genome-wide chromosome conformation capture studies, we found that such changes involve the repositioning of euchromatic *E. coli* ∆PC1 domains enriched for cytokine and other immune response genes. Upon encountering activating stimuli, these domains gained euchromatic character, repositioning themselves from the nuclear periphery to the more euchromatic nuclear interior. During this process, the boundaries of these domains lost insulation, allowing the domain to merge with neighboring highly euchromatic regions, and further allowing for new chromatin interactions to form and activate an inflammatory gene program. These subdomains resemble a previously identified euchromatic A2 spatial subcompartment positioned between the nuclear periphery and the nuclear interior (Rao et al. 2014; Chen et al. 2018). Based on our observations, we propose that the A2 subcompartment is associated with genes or regulatory elements that need to be transcriptionally repressed, but accessed quickly, precluding both their sequestration to the fully heterochromatic B compartment, as well as their presence in the transcriptionally active A1 compartment.

Our data further provide mechanistic insight as to how neutrophils instruct changes in nuclear positioning and domain insulation upon bacterial encounter. Alterations in chromatin topology both at ∆PC1 domains and across the genome are closely associated with the rapid recruitment of cohesin to a subset of H3K27ac-defined enhancers. While cohesin occupancy is substantially enriched at these enhancers, CTCF binding is only modestly elevated upon bacterial encounter. These observations imply that changes in nuclear architecture are predominantly activated by cohesin-dependent loop extrusion. This finding then raises the question as to how cohesin is being recruited to inflammatory genes upon bacterial encounter. We found that the increase in cohesin occupancy at SMC3-amassed enhancers was closely accompanied by substantial enrichment for the enhancer mark H3K27ac. Hence, we suggest that upon bacterial encounter, human neutrophils activate a signaling response that involves the Toll-like receptor pathway. Motif analysis suggests that Toll-like receptor mediated signaling modulates the expression and/or biochemical activities of key neutrophil-associated transcriptional regulators such as PU.1, CEBP/β, CEBP homolog CHOP, AP1 factors JUN and FOS, as well as TFE3. The activities of such regulators, in turn, would promote the assembly of an active enhancer repertoire as evidenced by the deposition of H3K27Ac, which then rapidly sequesters cohesin at inflammatory response enhancer-gene promoter clusters. Once recruited to SMC3-amassed enhancers, cohesin may act to extrude chromatin until convergent CTCF sites are reached, removing insulation at ΔPC1 domain boundaries by forming *de novo* loop domains in which activated enhancers are placed within close spatial proximity of gene promoters, altogether facilitating the rapid activation of an inflammatory response gene program (Fig. 7).

Why has such an elaborate mechanism of gene activation, including loop extrusion, evolved in human neutrophils? We suggest that segregating enhancers and promoters in spatially distinct loop domains ensures efficient silencing and prevents stochastic activation of an inflammatory specific gene program in unstimulated neutrophils. Only upon exposure to activating stimuli are unstimulated neutrophils instructed to juxtapose the inflammatory enhancer repertoire with their target gene promoters, thus facilitating enhancer-promoter communication and the induction of an inflammatory specific gene program. We hypothesize that the specificity of this response is likely governed by transcription factors downstream of activated receptors that bind their target enhancers, allowing cohesin and histone acetyl transferase recruitment, juxtaposition of target gene promoters, and stabilization of transcription units.

As documented here for human neutrophils during a microbial encounter, enhancers and promoters may be spatially segregated from each other in distinct loop domains until an appropriate environmental signal is received in order to prevent inappropriate or pathological activation of gene expression. Previous studies have documented a related mechanism that orchestrates the developmental progression of lymphoid cells. Specifically, regulatory regions associated with key developmental regulators such as EBF1 and Bcl11b are, in progenitor cells, positioned at the nuclear lamina to suppress premature activation during developmental progression. Upon reaching the appropriate developmental stage, alterations in chromatin folding readily reposition such enhancers away from the transcriptionally repressive environment at the lamina into the euchromatic nuclear interior, leading to assembly of transcriptionally productive enhancer-promoter interactions. The repositioning also directs the enhancer into a single loop domain to facilitate enhancer-promoter communication. Once placed within the euchromatic compartment and within spatial proximity to EBF1 and Bcl11b, enhancers and promoters then act to establish B or T cell identity, respectively (Lin et al. 2012; Isoda et al. 2017). Thus, the inflammatory gene response and activation of a developmental specific gene expression programs share a common mechanism that assures appropriate timing of gene expression.

In sum, here we demonstrate that in human neutrophils, prior to encounter with bacteria, an armamentarium of inflammatory genes was positioned in a transcriptionally passive environment suppressing premature transcriptional activation. Upon microbial exposure, however, human neutrophils rapidly (<3 hours) repositioned the ensemble of pro-inflammatory genes towards the transcriptionally permissive compartment. We found that the repositioning of genes was closely associated with the swift recruitment of cohesin across the inflammatory enhancer landscape permitting an immediate transcriptional response upon bacterial exposure. These data reveal at the mechanistic level how upon microbial challenge human neutrophils undergo rapid changes in nuclear architecture to orchestrate an immediate inflammatory gene program.

4.5 Methods

Human subject details

Blood for neutrophil isolation was obtained via venopuncture from healthy human volunteers under written informed consent approved by the UC San Diego Human Research Protection Program (#131002X).

Blood draws and neutrophil isolation

Whole blood was layered onto Polymorphprep reagent (Accurate Chemical and Scientific Corp., AN1114683) and was centrifuged for 45 minutes at 500g, 25°C, and allowed to stop without braking. The granulocyte layer was extracted and contaminating red blood cells were lysed as needed (generally 1-3 times) with brief resuspensions in sterile H2O followed by immediate flooding with 1x phosphate buffered saline (PBS) and centrifugation at 500g for 7 minutes at 25°C. Cells were checked for purity via Wright-Giemsa staining; the final granulocyte fraction was generally >95% neutrophils. For RNA-sequencing experiments, neutrophils were further purified to homogeneity using an EasySep Human Neutrophil Enrichment kit (Stemcell technologies, 19257) as per the manufacturer's protocol.

For all experiments neutrophils were cultured in HBSS +Ca/+Mg/-Phenol red (Thermo Fisher, 14025092) with the addition of 0.5% endotoxin-free BSA (Akron, AK8917-0100) at 37°C in a 5% CO2 humidified incubator.

Wright-Giemsa staining

Neutrophils (1x10^5) were spun onto cover slips using a Cytospin3 (Shandon, 74010121 GB) and flooded with Wright stain (Sigma, WS16-500ML) for 3 minutes. Cover slips were then washed with six consecutive dips in water baths. Cover slips were then allowed to air dry and were then flooded with Giemsa stain (Sigma, GS500-500ML) and allowed to incubate for 7 minutes before being washed as above and allowed to air dry.

Neutrophil activation

Neutrophils were plated at the desired cell numbers and treated with 25nM phorbol 12-myristate 13-acetate (PMA, Promega V1171) or co-cultured in the presence of *E. coli* strain K1 at a multiplicity of infection (MOI) of 5. Stimulations were performed for 3 hours and cells were harvested as detailed below.

Chromatin immunoprecipitation with sequencing (ChIP-seq)

Neutrophils were plated at 10-20x10^6 cells/10mL for each ChIP experiment. At the completion of each experiment cells were washed with fresh media, formaldehyde was added to the culture to a final concentration of 1%, and cells were cross linked with agitation for 10 minutes at room temperature. Fixation was then quenched for 5 minutes with glycine at a final concentration of 0.13M. Fixed cells were scraped from the plate and washed three times in ice cold 1x phosphate buffered saline (PBS) with 0.1mM EDTA and 1x EDTA-free complete protease inhibitors (Roche 05056489001). Cell pellets were snap frozen in liquid nitrogen and stored at -80°C until processing.

To bind antibody to ProteinG Dynabeads (Invitrogen 10004D), beads were washed 3x1mL with bead wash buffer (1x PBS, 5mg/mL BSA, Roche complete EDTA-free protease inhibitor, 0.22uM filtered) and resuspended in 500uL of the same. 1-5ug of antibody was added and allowed to bind beads overnight, rotating at 4°C. The following day beads were washed 3x1mL with bead wash buffer and resuspended in 100uL RIPA 150 (50mM Tris pH 8.0, 150mM NaCl, 0.1% SDS, 0.1% Sodium deoxycholate, 1% TritonX100, 1mM EDTA).

For each ChIP, cells were thawed and lysed on ice for 10 minutes with inversion in Farnham lysis buffer (5mM PIPES pH 8.0, 85mM KCl, 0.5% NP-40, 10mM EDTA, protease inhibitors) with or without 20 draws through an 18 gauge needle. Nuclei were spun down for 5 minutes, 2000rpm at 10°C in a bench top microfuge, supernatant was discarded, and nuclei were resuspended in 300uL RIPA 150. Chromatin was then sonicated in a Diagenode Bioruptor 300 chilled to 4°C for 3x8 cycles of 30” on and 30” off, set on high with 5 minutes of cooling time between each set of 8 cycles. The insoluble fraction was spun down for 20 minutes at maximum speed, 4°C, in a benchtop microfuge. Input and IP samples were split to separate new tubes, IP volume was adjusted to 900uL with RIPA 150, and 100uL of Protein G dynabeads bound to the antibody of interest in RIPA 150 was added to each IP. Chromatin was allowed to bind to antibody-bead conjugates overnight while rotating at 4°C. Following binding, beads were washed 2x5 minutes in RIPA 150, 2x5 minutes in RIPA 500 (50mM Tris pH 8.0, 500mM NaCl, 0.1% SDS, 01% Sodium deoxycholate, 1% TritonX100, 1mM EDTA), 2x3 minutes in LiCl wash (10mM Tris pH 8.0, 250mM LiCl, 1% NP-40, 1% Sodium deoxycholate, 1mM EDTA), and once in 1x TE. Beads were transferred to clean tubes at the start of each new wash buffer. DNA was eluted from beads with 200uL elution buffer (1mM Sodium carbonate, 1% SDS) for 1 hour, 65°C with shaking, at which point beads were removed and cross-links were reversed overnight at 65°C. Eluted DNA was purified using a ChIP DNA clean and concentrator kit (Zymo D5205).

DNA for ChIP and other high throughput sequencing approaches was processed as follows: End repair was performed using an Epicenter End-It kit (Lucigen ER0720), according to manufacturer’s instructions and column purified in a Zymo Minelute column (Zymo D4013). A-tails were added by incubating DNA in 1x NEB buffer 2 (New England Biolabs B7002S) with the addition of 200uM dATP and 7.5 units Klenow (exo-) (New England Biolabs M0212L) at 37 degrees for 45 minutes. NEB Next adaptors (New England Biolabs E7337A) were ligated using an NEB quick ligation kit (New England Biolabs M2200L) at benchtop temperature for 30 minutes followed by treatment with 2uL USER enzyme (New England Biolabs M5505L) for 15 minutes at 37°C. DNA was purified using an AmpureXP bead-analogous two-step SPRI bead protocol (Rohland and Reich 2012), resulting in purification of DNA fragments between ~200-800 bp.

PCR amplification of final libraries for sequencing was performed with Phusion hot start polymerase II system (ThermoFisher F549L) in conjunction with the NEB Next indexing system (New England Biolabs E7335L and E7500S). Final size selection for all high-throughput sequencing libraries was performed using a home-made two-step SPRI bead-based DNA purification system, resulting in final DNA fragment sizes of ~200-800bp.

RNA-sequencing

At specified time points neutrophils were washed once with PBS and lysed in the RLT buffer component of the Qiagen RNeasy mini kit (Qiagen 74106) with the addition of

10ul/mL 2-mercaptoethanol, homogenized via Qiashredder (Qiagen 79654), and snap frozen in liquid nitrogen. Total RNA was purified via RNeasy mini kit (Qiagen 74106) according to the manufactures instructions, including the RNase-free DNase (Qiagen 79254) treatment step. RNA was eluted in H2O, Turbo DNase kit buffer (ThermoFisher/Ambion AM1907) was added to a 1x concentration and RNA was treated with 4 units of Turbo DNase at 37degrees for 30 minutes. Turbo DNase was then treated with inactivation reagent per manufacturer’s specifications. mRNA was purified from total RNA using a Dynabead mRNA purification kit (Life Technologies 61006). First strand synthesis was performed using SuperScript III first strand synthesis system (ThermoFisher 18080051) as follows: 100-500ng RNA; 0.5uL Oligo(dT) primer; 0.8uL random hexamer; 1uL 10mM dNTP, H2O to 9.5uL. Mixture was incubated to 70 degrees for 10 minutes then snap frozen. First strand synthesis mix composed of 2uL 10x RT buffer, 4uL 25mM MgCl2, 2uL 0.1M DTT, 0.5 uL of 120ng/uL ActinomycinD, 40U RNaseOUT, 200U SuperScriptIII was added to the mixture which was then incubated at 25 degrees for 10 minutes, 42 degrees for 45 minutes, 50 degrees for 25 minutes, and 75 degrees for 15 minutes. Unincorporated nucleotides were removed from the mixture using a ProbeQuant G-50 column (Sigma GE28-9034-08). First strand synthesis reaction as then brought to 51uL with H2O and cooled on ice. 24uL of second strand mixture composed of 1uL 10x RT buffer, 2uL 25mM MgCl2, 1uL 0.1M DTT, 2uL of 10mM dATP, dGTP, dCTP, dUTP mix, 15uL of 5x second strand synthesis buffer (New England Biolabs B6117S), 0.5uL E. coli ligase (NEB M0205S), 2uL DNA polymerase I (NEB M0209S), and 0.5uL RNaseH was added and mixture was incubated at 16 degrees for 2 hours. DNA was purified using a DNA clean and concentrator kit (Zymo D4013) and sonicated on a Covaris E220 with the following settings; Duty cycle 10%; Intensity 5; Cycle per burst 200; Time (seconds) 180. Sonicated DNA was purified using a DNA clean and concentrator kit. DNA was prepared for high throughput sequencing using the methodology described above for ChIP-seq, with the addition of 1uL UNG (ThermoFisher/Applied Biosystems N8080096) during USER enzyme treatment.

Whole genome bisulfite sequencing

Neutrophils were washed twice with PBS and genomic DNA was isolated using a DNeasy Blood and Tissue kit (Qiagen 69504). 1 μg of genomic DNA mixed with unmethylated lamda DNA at a concentration of 0.5% of total DNA was sonicated by Biorupter 300 with 20 cycles (30 seconds on and 30 seconds off at low power). Fragmented DNA was end-repaired and A-tailed as described above. TruSeq adapters (Illumina FC-121-2001) were ligated to fragmented DNA which was then purified by running on a 2% agarose gel. Bisulfite conversion was performed using the MethylCode kit as described by the manufacturer (Invitrogen MECOV-50). Bisulfite-treated DNA was amplified by using a TruSeq PCR primer mixture and Pfu Turbo Cx Polymerase, agarose gel purified, and sequenced on an Illumina HiSeq 2500 sequencer with paired end 150bp reads.

E. coli culture and MOI determination

E. coli strain K1 was grown in LB at 37 degrees with shaking overnight, and diluted into a fresh culture and grown to exponential phase the day of each experiment. *E. coli* was then pelleted at 3000rpm for 10 minutes at 10 degrees on a benchtop centrifuge, washed in cell culture media, and added to neutrophil cultures at an MOI of ~5 in HBSS +Ca/+Mg/-Phenol red with 0.5% endotoxin free BSA. 9 1:10 serial dilutions of *E. coli* containing media were plated on LB agar and grown overnight at 37 degrees. The resulting colonies were counted in order to assess MOI for individual experiments.

In situ HiC

In situ HiC was performed as described (Rao et al. 2014), modifying only the MboI restriction enzyme digest time to assure proper digestion of chromatin. Generally, HiC libraries prepared from activated neutrophils were digested for 2-4 hours with 50 units MboI to avoid over-digesting the chromatin. The remainder of the library preparation adhered to the published protocol and reagents exactly. HiC library DNA was prepared for high throughput sequencing using the NEB Next platform according to manufacturer's instructions, and sequenced using paired end 100bp reads.

Fluorescence in situ hybridization (FISH)

Cover slips were incubated overnight in 1%HCl in 70% ethanol, washed 3x with H2O, once in 70% ethanol, and stored in 100% ethanol. Coverslips were allowed to air dry prior to adding cells. Cells were incubated on cover slips in 24 well plates as described above. At the completion of incubation times, cells were washed 3x3 minute in PBS and fixed for 30 minutes in 6% paraformaldehyde (Electron Microscopy Sciences 15710) in 1x PBS. PFA was flushed out with >5 volumes of PBS/0.05% Tween-20 (PBST), ensuring that cells never contact the air. Residual PFA was quenched via incubation with fresh 20mM glycine in PBS for 15 minutes at room temperature. Cells were permeabilized in

PBS+0.5% Triton x-100 for 20 minutes at room temperature, washed twice with PBST, and incubated in PBS + 100ug/mL RNase A (Qiagen 19101) for 1 hour at 37°C. Cells were then treated with 0.1N HCl for 5 minutes at room temperature, washed 2x3 minutes with 1xPBS, 2x5 minutes with 2xSSC, then incubated for >48 hours in 2xSSC/50% formamide at 4°C. Cover slips were then blotted dry and 5uL probe containing 75-200ng labeled DNA was added to each coverslip. Cover slips were then sealed on top of glass slides along with probe using rubber cement. Probes and genomic DNA were denatured together for 5 minutes at 78°C on a heat block and allowed to hybridize for 16-48 hours at 37°C. Following hybridization cover slips were washed 1x15 minutes in SSC/50% formamide pre-warmed to 37°C, 3x15 minutes in 2xSSC pre-warmed to 37°C, 3x7 minutes in 0.1xSSC pre-warmed to 60°C, 3x7 minutes in 4xSSC/0.02% Tween-20 pre-warmed to 42°C, 1x5 minutes with 2xSSC pre-warmed to 37°C, and 2x5 minutes in 1x PBS. Cells were then post-fixed in 4% PFA in 1x PBS for 10 minutes at room temperature, and PFA was flushed out as above. Cells were washed 1x10 minutes in PBST+DAPI, 4x5 minutes 1xPBS, and mounted in Prolong Gold mounting media (ThermoFisher P36930).

FISH probes were prepared from bacterial artificial chromosomes (BACs) using nick/translation (Roche 11745808910). 1ug BAC DNA was used in each 20uL nick/translation reaction along with the following fluorophores, as needed: ChromaTide Alexafluor 488-5-dUTP (ThermoFisher/Life Technologies C11397), Cy3-dUTP (VRW 42501), or AlexaFluor 647-aha-dUTP (ThermoFisher/Life Technologies A32763). Nick/Translation was performed for 5-16 hours at 15 degrees and terminated by addition of 1uL 0.5M EDTA. Unincorporated nucleotides were removed with ProbeQuant G-50 columns per manufacturer’s instructions. 100ng of labeled probe DNA was run on a 1.5%

agarose gel following each nick/translation reaction to ensure the majority of probe fragments were in the 300-800bp range. Up to 200ng total probe per cover slip was combined with 10ug salmon sperm DNA (ThermoFisher 15632011), 4ug human Cot1 DNA (ThermoFisher 15279011), 1/10 volume of 3M sodium acetate pH 5.2, and 2.5 volumes of 100% ethanol. Probes were allowed to precipitate for 30 minutes at -20°C, were centrifuged for 20 minutes at 4°C, maximum speed, washed twice with 70% ethanol and once with 100% ethanol, air dried, and resuspended in 6uL 100% formamide at 56°C. 6uL of 2x hybridization buffer (40% dextran sulfate in 8x SSC (20x SSC: 3M NaCl, 0.3M Sodium citrate)) was then added to each probe. Probes were denatured for 5 minutes at 80 degrees and snap-cooled on ice. Probes were then added to cover slips and denatured and hybridized to genomic DNA as noted above. The *CXCL* locus FISH probe utilized BAC RP11-243E9, the heterochromatic control probe utilized BAC RP11-134J16.

Imaging of FISH samples was performed at the Waitt Biophotonics Center at the Salk Institute. FISH samples were imaged on Zeiss Airyscan 880 microscopes using the Airyscan Fast mode (Huff 2016) at a resolution of 40nm in the x and y axes. Z sections were imaged every 160nm. Quantification of FISH data was performed using TANGO (Ollion et al. 2013) for FIJI (Schindelin et al. 2012). Nuclei and spot detection were performed with built in tools in TANGO. Image metrics analyzed in TANGO include: “Eroded Volume Fraction” and “Signal Quantification Layer” in Fig. 4, and “Distances” in Supplemental Fig. S3. Metrics were exported from TANGO as text files and statistical analysis and figure generation were performed in R using built-in tools (R Core Team).

HiC analysis

Raw HiC library read alignment to human genome build hg38, valid read pair filtering, matrix assembly at various resolutions, and ICE normalization of said matrices were performed using HiC-pro with default settings (Servant et al. 2015). Biological replicates were pooled following valid read pair filtering, and pooled data sets were used for analysis except where noted.

For all direct comparisons of HiC data (topological domain boundary location comparisons, insulation scores, plotted contact matrices, $\log_2$ differential matrices) ICE normalized sparse matrix files were created containing only the subset of interacting bins that recorded reads in all data sets being compared. Read numbers at these bins were then quantile normalized in R using the normalize.quantiles() function in the preprocessCore package (Bolstad), allowing direct comparison of chromatin interactions between libraries with different read distributions and sequencing depths (Hsu et al. 2017).

Topological domain boundaries were called on normalized HiC data at 40kb resolution using the domain calling software published in Dixon et al. (Dixon et al., 2012).

HiC-Pro defined valid read pairs were used in conjunction with HOMER (Heinz et al. 2010) to run principal component analysis (PCA, *runHiCpca.pl -res 10000*), generate distance vs. interaction frequency plots (*makeTagDirectory*), define compartment boundaries (*findHiCCompartments.pl*), determine interaction correlations (*getHiCcorrDiff.pl -res 40000 -superRes 40000*), define distance-normalized chromatin interactions (*analyzeHiC -res 20000 -superRes 40000 -minDist 100000*), and to generate whole chromosome pairing plots (*analyzeHiC -res 400000000*).

CTCF anchored-type loops were called using HICCUPS (Rao et al. 2014).

Insulation scores were determined as follows: The genome was divided into 40kb segments. Insulation scores for each segment were defined as the number of normalized (ICE and quantile, see above) valid read pairs within a 500kb window centered on the segment of interest whose ends map to opposite sides of the segment of interest divided by the total number of valid read pairs whose ends both map within the 500kb window.

∆PC1 domains were identified as follows: PCA was run at 10,000 base pair resolution on pooled HiC data using the *runHiCpca.pl* command in HOMER with the following settings: *-res 10000 -superRes 10000 -genome hg38*. Visual inspection showed that positive PC1 values corresponded to the gene-rich A compartment, and negative PC1 values corresponded to the gene-poor B compartment on all chromosomes and across all conditions. Genomic regions with PC1 score differentials between conditions greater than three standard deviations above the mean PC1 score differential between conditions were identified as potential ∆PC1 domains. PCA was then run on individual HiC biological replicates and only those potential ∆PC1 domains with a reproducible gain in PC1 value in each biological replicate were retained. Finally, reproducible ∆PC1 domains within 100kb of each other were merged into single continuous ∆PC1 domains which were used for downstream analysis.

ChIP-seq analysis

Raw fastq files were aligned to the human genome build hg38 using bowtie (Langmead et al.)_with the following parameters: *-m1 --best --strata*. Downstream processing of ChIP-seq data was performed using HOMER, except where noted. Uniquely mapped reads from high quality biological replicates were pooled for downstream analysis (Landt et al. 2012). Sequencing data was reorganized as a HOMER-formatted tag directory for each replicate and multiple reads mapping to the same base pair were collapsed to a single read using the *makeTagDirectory* command in HOMER with the following parameters: *-tbp 1*. ChIP peaks were called using the *findPeaks* command in HOMER with default parameters. Genes at ChIP peaks were identified using *annotatePeaks.pl* in HOMER, and the GenomicRanges package (Lawrence et al. 2013) in R.

SMC3-amassed enhancers were defined as follows: Enhancers were defined as H3K27ac peaks called as above. In order to identify enhancers with activation-dependent cohesin recruitment (SMC3-amassed enhancers), total unique SMC3 ChIP-seq reads mapping to enhancers were calculated using *annotatePeaks.pl* in HOMER. To directly compare binding strength between conditions read numbers at enhancers were quantile normalized across conditions using the preprocessCore R package. Those reads with a log2(normalized activated/normalized unstimulated read numbers) value greater than 1.5 were defined as SMC3-amassed. The GenomicRanges package in R was used to identify genes in contact with SMC3-amassed enhancers in conjunction with HOMER-defined chromatin interactions (detailed below). Enhancer-gene pairs were called as interacting if the center of one interaction anchor was within 10kb of an enhancer and the center of the other interaction anchor was within 50kb of a gene promoter.

RNA-seq analysis

RNA-seq data was analyzed using the Tuxedo tools, except where noted. Raw fastq files were aligned to the human genome build hg38 using tophat2 (Kim et al. 2013) with the following parameters: *--library-type fr-firststrand -a 15*. Duplicated reads were removed using Picard tools command MarkDuplicates REMOVE_DUPLICATES=T, and RNA-seq quality metrics were assessed using Picard tools command CollectRnaSeqMetrics (http://broadinstitute.github.io/picard). Gene expression values were computed for each replicate across each condition using cuffdiff with an hg38 refflat file as reference with the following parameters: *--library-type fr-firststrand*. Subsequent analysis of gene expression and integration of gene expression data with other data types was performed in R.

Metascape analysis (http://metascape.org)

Genes associated with various genomic features were identified using the GenomicRanges package in R and were analyzed for functional enrichment in the Metascape web portal using “Express Analysis” on default settings (Tripathi et al. 2015). Metascape gene set enrichment visualizations were performed in R.

Bisulfite-seq analysis

Bisulfite converted DNA sequencing data was processed using the BSseeker2 software suite (Guo et al. 2013). A bisulfite-sequencing amenable hg38 reference genome was built using the *bs_seeker2-build.py* command. DNA sequences were aligned to the hg38 bisulfite sequencing-amenable genome build using the *bs_seeker2-align.py* command with the following options: *-m 6 -I 0 -X 800*. Cytosine methylation levels were determined using *bs_seeker2-call_methylation.py* with default settings. Awk was used to convert CGmap files to HOMER-compatible allC formatted files. HOMER-formatted tag directories were built using HOMER’s *makeTagDirectories* command with the following options: *-format allC -minCounts 0 -genome hg38*. Due to sequencing coverage-induced biases in DNA methylation meta-analysis (data not shown), awk was used to create HOMER-formatted tag directories containing only those cytosine residues covered by both unstimulated and PMA-activated neutrophil data sets. HOMER’s *annotatePeaks.pl* command was used with the *-ratio* option to determine DNA methylation levels at particular genomic features.

Data visualization

Normalized HiC contact matrices presented in this paper were generated using HiCPlotter (Akdemir and Chin 2015). HiC interactions and ChIP-seq data in Fig. 4, Supplemental Figs. S4, and S6 were visualized using Sushi (Phanstiel et al. 2014). The remainder of linear genomic data was visualized using the Integrated Genomics Viewer (Robinson et al. 2011; Thorvaldsdóttir et al. 2013). FISH images were processed in FIJI. All other data was visualized using R.

Data availability

Data sets generated in this study are available as a series in the GEO database under accession number GSE126758.

4.6 Figures

Figure 4.1. Heterochromatic super-contraction and segmentation of the neutrophil genome during acquisition of the PMN shape. (A) Neutrophil (blue) and H1 hESC (teal) chromatin interaction frequencies as a function of linear genomic distance. Percent interchromosomal paired end tags (PETs) are indicated. (B) Normalized HiC contact matrices for H1 hESC (left) and neutrophil (right) chromosome 20. First principal component 1 eigenvector (PC1) for each HiC matrix is displayed above its respective matrix. Positive PC1 values correspond to the gene-rich A compartment, negative values to the gene poor B compartment, based on human genome build hg38. (C) HiC contact matrix showing the log2 fold-change in normalized interactions between the H1 hESC and neutrophil matrices in B, illustrating changes in the organization of chromosome 20 during terminal differentiation from the pluripotent state and acquisition of the PMN shape. PC1 values for H1 hESCs and neutrophils are shown above and to the right of the matrix, respectively. (D)Density plot showing the distribution of A (solid lines) and B (dotted lines) compartment domain sizes in H1 hESCs (teal) and neutrophils (blue). Total number of domains for each data set are listed at top. Domains smaller than 100,000 bp were not considered. (E) Example of a new TAD and compartment boundary formed during differentiation and hyper-compartmentalization of the neutrophil genome. Top (top to bottom): IGV tracks showing H1 hESC PC1 values, H1 hESC TADs, neutrophil PC1 values, and neutrophil TADs. Bottom: Normalized HiC contact matrix of H1 hESC and neutrophil HiC matrices and a log2 fold-change difference matrix at a new TAD/compartment boundary on chromosome 14. (F) Percent of TAD and compartment boundaries shared between H1 hESCs and neutrophils. Domain boundaries within 100kb (< 3 windows) were considered shared. (G) Insulation scores in H1 hESCs and neutrophils calculated for each 40kb bin genome-wide, at neutrophil-specific, H1 hESC-specific, and shared PC1 compartment/TAD boundaries for both cell types. Grouped pairs are all significantly different by the Wilcoxon rank sum test with log(p) values <1E-11.

Denholtz Figure 1

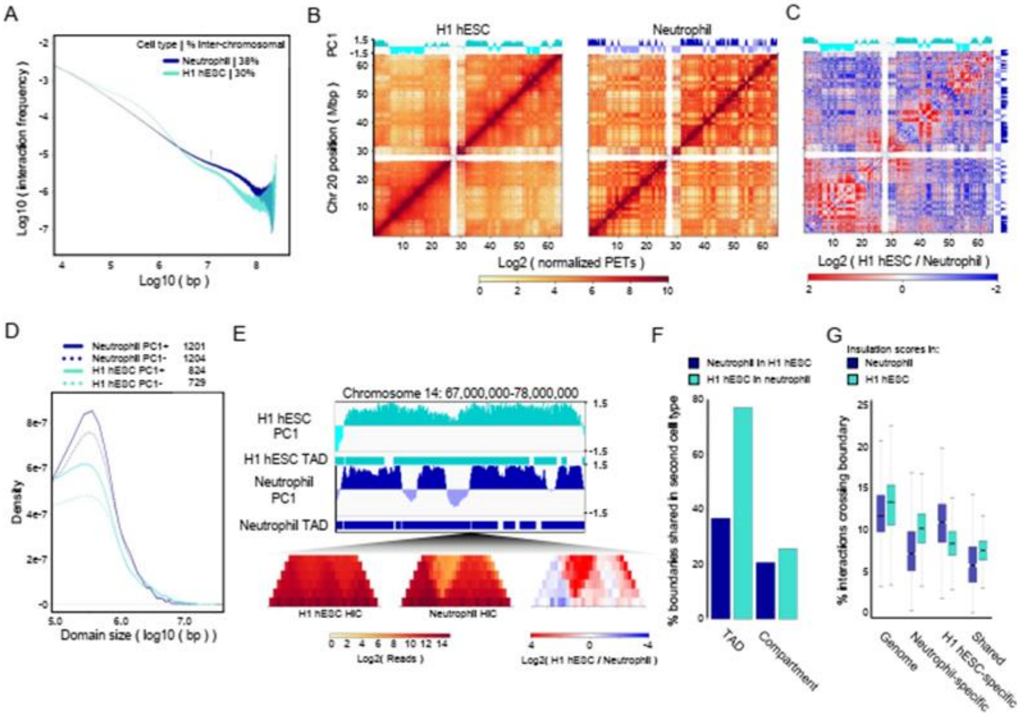


Figure 4.2. PMA-activation increases euchromatic character at distinct genomic loci encoding for neutrophil activation genes. (A) Unstimulated (blue) and PMA-activated neutrophil (red) (three hours) chromatin interaction frequencies as a function of linear genomic distance. Percent inter-chromosomal PETs are indicated. (B) Percentages of TAD and compartment boundaries shared between unstimulated and PMA-activated neutrophils. (C) Scatterplot comparing 10kb-windowed PC1 values between unstimulated and PMA-activated neutrophils. PMA ∆PC1 domains are marked in red. (D) Metascape-defined functional groups enriched for genes found within 100kb of PMA ∆PC1 domains. Complete metascape results can be found in Supplemental Table 2. (E) HiC contact matrices of chromosome 18 for unstimulated (left) and PMA-activated (right) neutrophils, 100kb resolution. Respective PC1 values are shown above each matrix. (F) Difference matrix showing the log2 fold-change in normalized interactions between unstimulated and PMA-activated neutrophils. PC1 differences are shown at top, PMA ∆PC1 domains are marked by red triangles. (E) Black and red points: Genome-wide interaction correlations for each 40kb bin in the genome, ranked from most to least differential, left to right. Bins containing PMA ∆PC1 domains are marked with red points. Red line with red shading: Proportion of total PMA ∆PC1 domains found at a given rank or lower, showing a preference for PMA ∆PC1 domains to fall in genomic regions with the most differential chromatin interactions upon PMA stimulation. (H) Boxplots of genome-wide interaction correlation values for PMA ∆PC1 domains and the remainder of the genome during PMA stimulation. Boxplot outliers are not shown. *=Wilcoxon rank-sum test p-value <2E-16.

Denholtz Figure 2

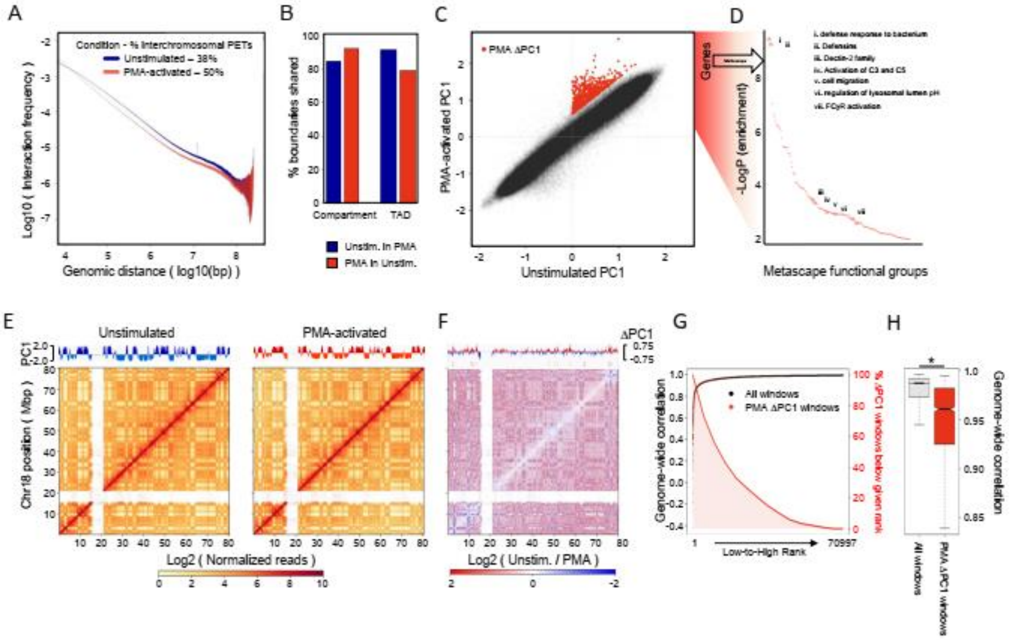


Figure 4.3. Neutrophil *E. coli* co-culture increases euchromatic character at distinct loci encoding for neutrophil pathogen response genes. (A) Unstimulated (blue) and E. coli co-cultured neutrophil (green) chromatin interaction frequencies as a function of linear genomic distance. Percent inter-chromosomal PETs are indicated. NOTE: Unstimulated neutrophil data is identical to that shown in Figure 2A and is shown here to illustrate differences between unstimulated and E. coli co-cultured neutrophil data. Neutrophils were cultured in the presence of E. coli for a three hours period. (B) Percent of TAD and compartment boundaries shared between unstimulated and E. coli co-cultured neutrophils. (C) Scatterplot comparing 10kb-windowed PC1 scores between unstimulated and E. coli co-cultured neutrophils. E. coli ΔPC1 domains are marked in green. (D) Metascape-defined functional groups enriched for genes found within 100kb of E. coli ΔPC1 domains. Complete metascape results are described in Supplemental Table 2. (E) HiC contact matrices of chromosome 20 for unstimulated (left) and E. coli co-cultured (right) neutrophils, 100kb resolution. Respective PC1 values are shown above each matrix. (F) Difference matrix showing the log2 fold-change in normalized interactions between unstimulated and E. coli co-cultured neutrophils. PC1 differences are shown at top with E. coli ΔPC1 domains marked with green triangles. (G) Black and green points: Genome-wide interaction correlations for each 40kb bin in the genome, ranked from most to least differential, left to right. Bins containing E. coli ΔPC1 domains are marked with green points. Green line with green shading: Proportion of total E. coli ΔPC1 domains found at a given rank or lower, showing a preference for E. coli ΔPC1 domains to fall within genomic regions with the most differential chromatin interactions upon E. coli encounter. (H) Boxplots of genome-wide interaction correlation values for E. coli ΔPC1 domains and the remainder of the genome during E. coli encounter. Boxplot outliers are not shown. *=Wilcoxon rank-sum test p-value <2E-16.

Denholtz Figure 3

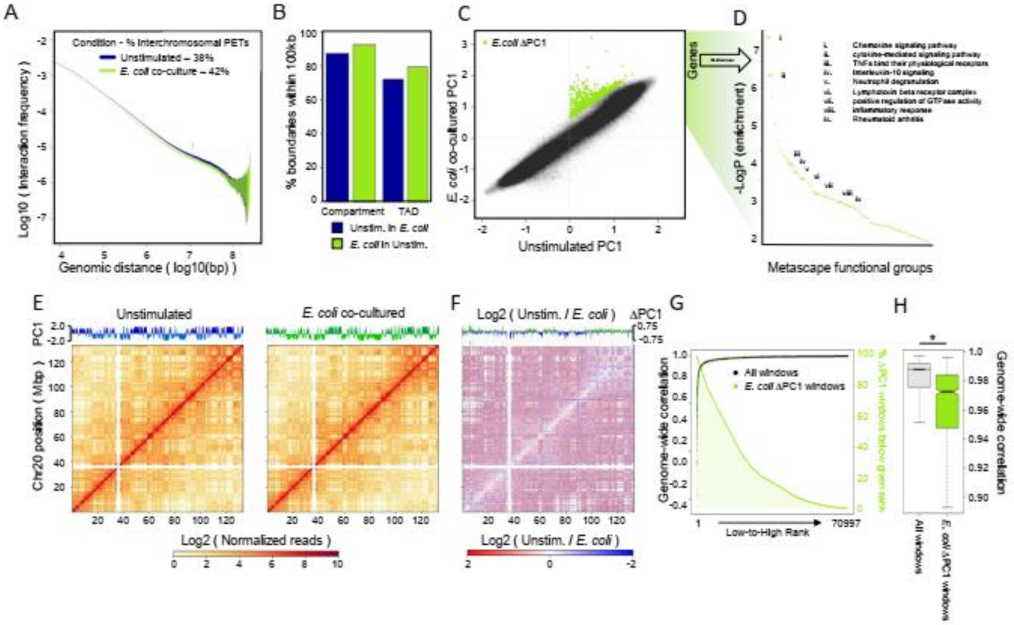


Figure 4.4*. E. coli* co-culture-induced topological changes at the *CXCL* sub-domain are associated with non-coding transcription, cohesin recruitment, and locus repositioning. (A) HiC contact maps of the extended CXCL gene cluster in unstimulated (top) and E. coli co-cultured (bottom) human neutrophils. PC1 scores are shown above their respective matrices, CXCL E. coli □PC1 domain position is noted. (B) Log2 difference matrix comparing HiC contacts between unstimulated and E. coli co-cultured neutrophils within the extended CXCL gene cluster. PC1 differential values (E. coli co-cultured – unstimulated PC1 values) are shown above matrix, protein-coding genes in the CXCL gene locus are shown below. (C) Top: Linear genomic features and significant chromatin interactions at the CXCL E. coli □PC1 domain. CTCF and SMC3 ChIP-seq, RNA-seq, and HOMER-defined chromatin interactions with -log(p) values less than -50 are shown for unstimulated (blue, top) and E. coli co-cultured neutrophils (green, bottom) with PC1 differential values (E. coli co-cultured – unstimulated PC1 values) shown between. Bottom panel displays SMC3 ChIP-seq and RNA-seq tracks at the CXCL8 (IL8) gene, demonstrating transcription-associated recruitment of SMC3 to the CXCL8 locus. Peaks of transcription and associated SMC3 recruitment are highlighted in yellow. (D) Representative FISH image (z-projection) showing the euchromatic CXCL8 locus (green) in unstimulated (top) and E. coli co-cultured (bottom) neutrophils. (E) Quantification of the proportion of nuclear volume between the FISH signal and the nuclear periphery (Eroded volume fraction) in unstimulated and E. coli co-cultured neutrophils. Number of alleles analyzed are listed below the boxplots. Wilcoxon rank sum test p-value for the data distributions: ** = <0.00005. (F) Quantification of DAPI signal intensity at the FISH spots identified in B. Wilcoxon rank sum test p-values: ** = <0.00005.

Denholtz Figure 4

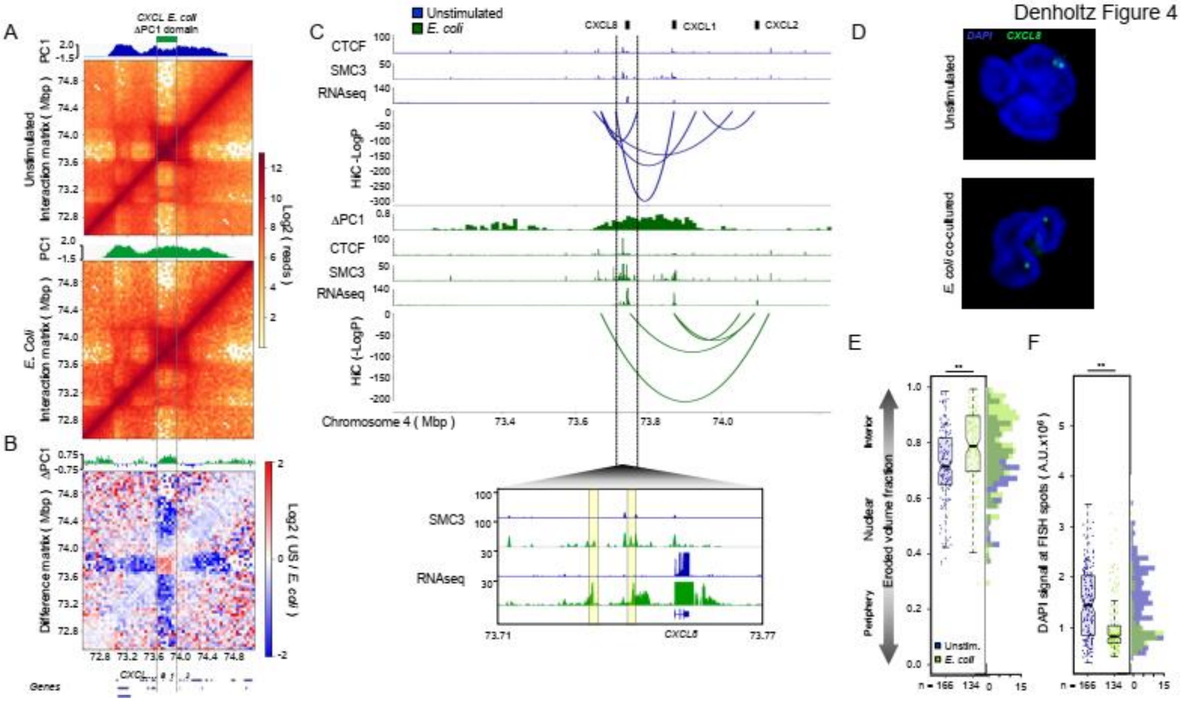


Figure 4.5. Analysis of subsystem functional logic. (A) Insulation score meta-analysis. Insulation scores were calculated for genomic regions surrounding *E. coli* ∆PC1 domain boundaries. Normalized insulation scores for unstimulated (blue) and E. coli co-cultured (green) neutrophils are shown. Vertical dotted lines demarcate meta-domain boundaries. Horizontal dotted lines show median normalized insulation scores at PC domain boundaries genome-wide for each condition. Calculations shown are for all E. coli□∆PC1 domains larger than 100kb. (B) Distribution of linear genomic distances between chromatin interaction anchor points genome-wide, for interactions anchored in ∆PC1 domains, and for interactions fully contained within E. coli ∆PC1 domains. Wilcoxon rank sum test: **p<2.2E-16; *p<1E-5; +not significant. (C) Interaction strength bias (see methods) of HOMER-defined chromatin interactions for unstimulated and E. coli co-cultured neutrophils genome-wide, for those interactions with a single anchor in an E. coli ∆PC1 domain, and for those interactions contained entirely within E. coli ∆PC1 domains. (D) HICCUPS defined loop strength with respect to E. coli ∆PC1 domains in unstimulated and E. coli co-cultured neutrophils. (E) Distance between HICCUPS loops and E. coli □PC1 domains. Differences are not significant by the Kolmogorov-Smirnov test. 5/45 (11%) E. coli co-culture specific loops are in E. coli ∆PC1 domains, 19/49 are within 1Mb of E. coli ∆PC1 domains (45%). E. coli ∆PC1 domains make up ~0.3% of the human neutrophil genome. (F)Log2 (normalized E. coli co-cultured / unstimulated) FPKM values for genes at chromatin interaction anchors shared between unstimulated and E. coli co-cultured neutrophils, or at chromatin interactions specific to one condition. Wicoxon rank sum test: **p<1E-4; *p<5E-3. (G) Change in PC1 values at genes with the given mRNA expression differential during E. coli encounter.

Denholtz Figure 5

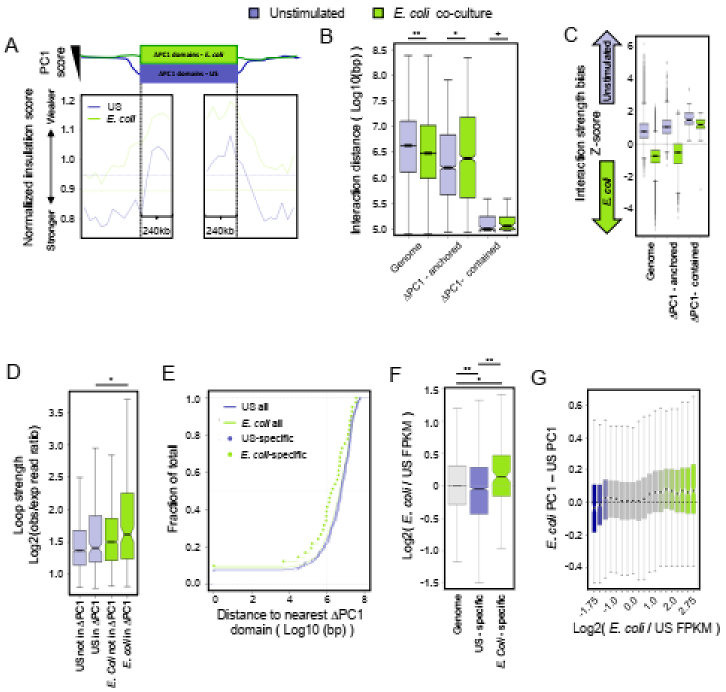


Figure 4.6. *E. coli* co-culture induces cohesin recruitment to a subset of H3K27ac-defined enhancers. (A) Log2 ratio (E. coli co-cultured / unstimulated) of the normalized ChIP-seq and RNA-seq signals at H3K27ac-defined enhancers that amass SMC3 as well as all enhancers genome-wide. (B) Histogram showing the percent of SMC3-amassed enhancers falling in E. coli ΔPC1 domains randomly positioned within the A compartment (grey), and the percent of SMC3-amassed enhancers falling within actual E. coli ΔPC1 domains (green arrow). 997/1000 of random permutations resulted in lower overlap between SMC3-amassed enhancers and ΔPC1 domains than were observed in the empirical data. (C) Top panel shows distance distribution between E. coli ΔPC1 domain boundaries and SMC3-amassed enhancers (green) and all enhancers (white). Bottom panel shows observed enrichment of SMC3-amassed enhancers or all enhancers in □PC1 domains divided by the expected enrichment of these enhancers in E. coli ΔPC1 domains based on 1000 random permutations of E. coli ΔPC1 domain positions within the A compartment. * = Wilcoxon rank sum test p-value < 0.005. (D) Top panel indicates ratio of mean transcription factor motif density (motifs per base pair per peak, SMC3-amassed enhancers / all enhancers) for representative transcription factors. Bottom panel shows gene expression values (FPKM) of representative transcription factors in unstimulated and E. coli co-cultured neutrophils. (E) Known transcription factor motifs identified in D. (F) Metascape gene functional analysis for genes interacting with SMC3-amassed enhancers. Full metascape analysis results is shown in Supplemental Table 2. (G) Empirical cumulative distribution of log2(E. coli co-cultured / unstimulated FPKM) values for all genes, genes interacting with any enhancer, and genes interacting with E. coli co-culture-dependent SMC3-amassed enhancers.

Denholtz Figure 6

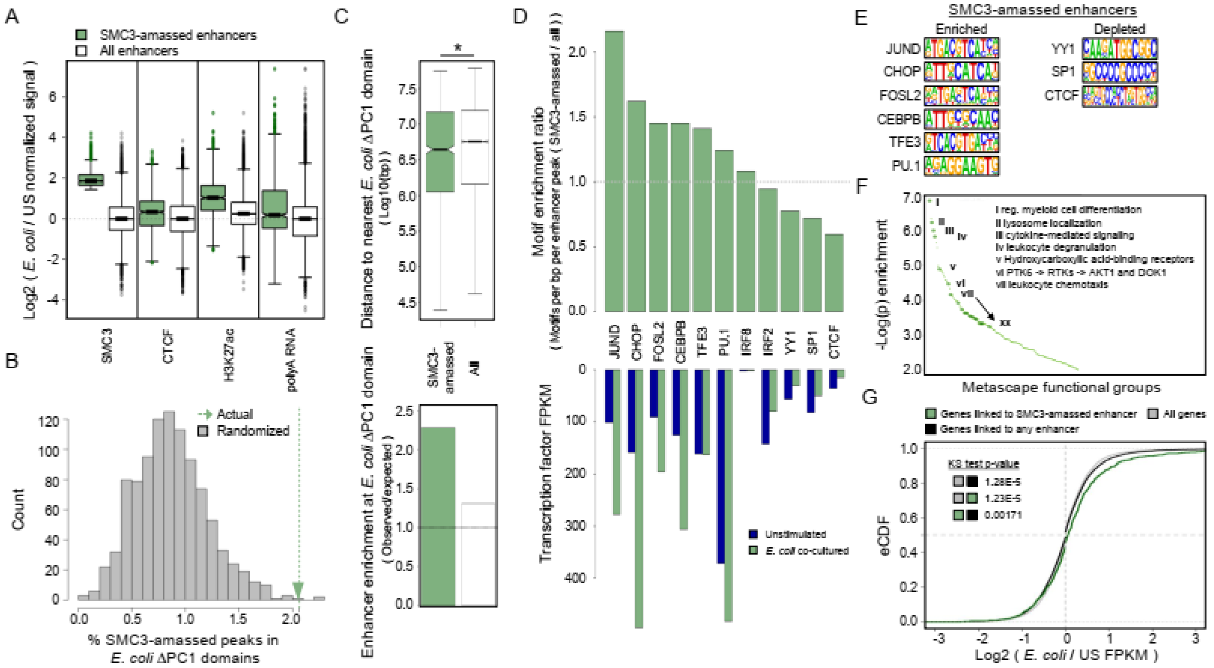
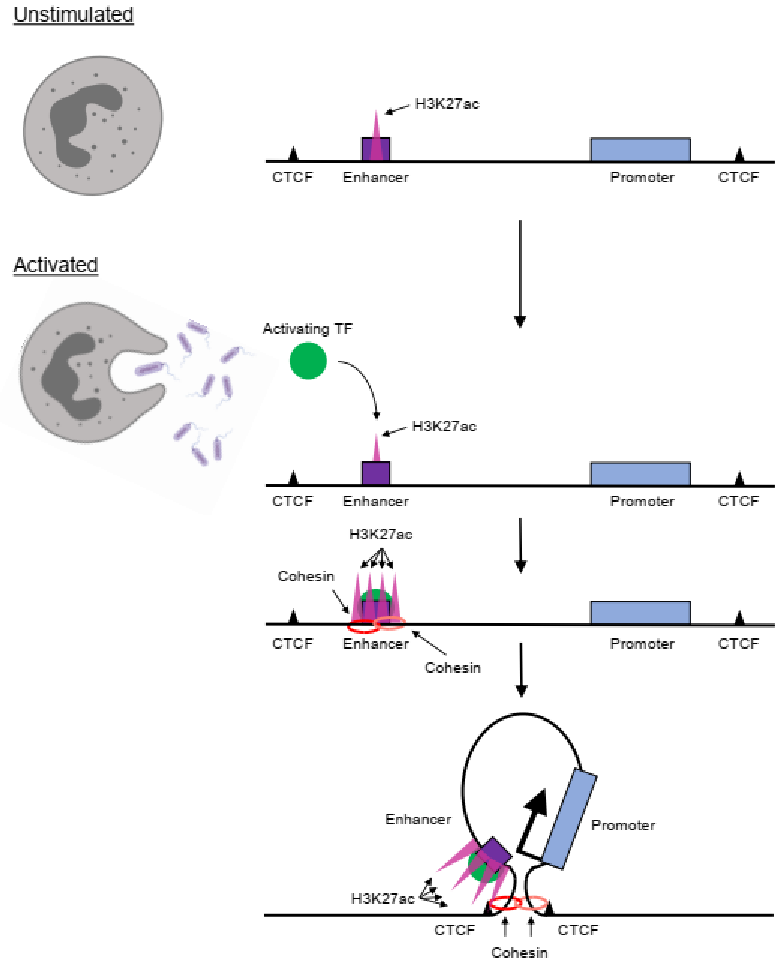


Figure 4.7. Microbial-induced human neutrophil activation instructs rapid changes in nuclear architecture to orchestrate an inflammatory gene program. Activation-induced transcription factor binding results in H3K27ac deposition, cohesin recruitment, and formation of de novo chromatin loops linking enhancers to inflammatory genes to orchestrate an inflammatory gene program.

Denholtz Figure 7



4.7 Acknowledgments

Chapter 4, in full, has been accepted for publication of the material as it may appear in *Genes & Development*, 2020. “Upon microbial challenge human neutrophils undergo rapid changes in nuclear architecture orchestrate an immediate inflammatory gene program”. Matthew Denholtz, Yina Zhu, Zhaoren He, Hanbin Lu, Takeshi Isoda, Simon Döhrmann, Victor Nizet and Cornelis Murre. The dissertation author was the primary investigator and author of this paper.

4.8 References

Akdemir K, Chin L. 2015. HiCPlotter integrates genomic data with interaction matrices. Genome Biol 16: 198.

Bintu B, Mateo LJ, Su J-HH, Sinnott-Armstrong NA, Parker M, Kinrot S, Yamaya K, Boettiger AN, Zhuang X. 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science (New York, NY) 362.

Bolstad B. preprocessCore: A collection of pre-processing functions. R package version 1.44.0. https://github.com/bmbolstad/preprocessCore.

Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, Weinrauch Y, Zychlinsky A. 2004. Neutrophil extracellular traps kill bacteria. Science (New York, NY) 303: 1532–5.

Busslinger GA, Stocsits RR, van der Lelij P, Axelsson E, Tedeschi A, Galjart N, Peters J-MM. 2017. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature 544: 503–507.

Cavaillon J-MM. 2011. The historical milestones in the understanding of leukocyte biology initiated by Elie Metchnikoff. J Leuk Biol 90: 413–24.

Chen Y, Zhang Y, Wang Y, Zhang L, Brinkman EK, Adam SA, Goldman R, van Steensel B, Ma J, Belmont AS. 2018. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J Cell Biol 217: 4025–4048.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485: 376–80.

Fanucchi S, Fok ET, Dalla E, Shibayama Y, Börner K, Chang EY, Stoychev S, Imakaev M, Grimm D, Wang KC, et al. 2019. Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. Nat Genet 51: 138–150.

Fritz AJ, Barutcu A, Martin-Buley L, van Wijnen AJ, Zaidi SK, Imbalzano AN, Lian JB, Stein JL, Stein GS. 2016. Chromosomes at Work: Organization of Chromosome Territories in the Interphase Nucleus. J Cell Biochem 117: 9–19.

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep 15: 2038–49.

Garcia-Romo GS, Caielli S, Vega B, Connolly J, Allantaz F, Xu Z, Punaro M, Baisch J, Guiducci C, Coffman RL, et al. 2011. Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. Sci Transl Med 3: 73ra20.

Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen PY, Pellegrini M. 2013. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics 14: 774.

Heinz S, Benner C, Spann N, Bertolino E, cell LY. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. MolCell 38: 576-589.

Hoffmann K, Dreger CK, Olins AL, Olins DE, Shultz LD, Lucke B, Karl H, Kaps R, Müller D, Vayá A, et al. 2002. Mutations in the gene encoding the lamin B receptor produce an altered nuclear morphology in granulocytes (Pelger-Huët anomaly). Nat Gen 31: 410–4.

Hsu SC, Gilgenast TG, Bartman CR, Edwards CR, Stonestrom AJ, Huang P, Emerson DJ, Evans P, Werner MT, Keller CA, et al. 2017. The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries. Mol Cell 66: 102-116.e7.

Huff J. 2016. The Fast mode for ZEISS LSM 880 with Airyscan: high-speed confocal imaging with super-resolution and improved signal-to-noise ratio. Nat Meth 13.

Isoda T, Moore AJ, He Z, Chandra V, Aida M, Denholtz M, van Hamburg J, Fisch KM, Chang AN, Fahl SP, et al. 2017. Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. Cell 171: 103-119.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14: R36.

Kosak ST, Skok JA, Medina KL, Riblet R, Beau MM, Fisher AG, Singh H. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science (New York, NY) 296: 158–62.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 22: 1813–31.

Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short reads to the human genome. Genome Biol 10:R25.

Lawrence M, Huber W, Pages P, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V. 2013. Software for computing and annotating genomic ranges. PLoS Comp 9(8):e10003118.

Ley K, Hoffman HM, Kubes P, Cassatella MA, Zychlinsky A, Hedrick CC, Catz SD. 2018. Neutrophils: New insights and open questions. Sci Immunol 3(30):eaat4579.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (New York, NY) 326: 289–93.

Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK, 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: 1196–204.

Muller W. 2013. Getting leukocytes to the site of inflammation. Vet Pathology 50: 7–22.

Nora EP, Goloborodko A, Valton A-LL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell 169: 930-944.e22.

Olins AL, Hoang TV, Zwerger M, Herrmann H, Zentgraf H, Noegel AA, Karakesisoglou I, Hodzic D, Olins DE. 2009. The LINC-less granulocyte nucleus. European Cell Biol 88: 203–14.

Ollion J, Cochennec J, Loll F, Escudé C, Boudier T. 2013. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. Bioinformatics (Oxford, England) 29:1840–1.

Pastore N, Brady OA, Diab HI, Martina JA, Sun L, Huynh T, Lim JA, Zare H, Raben N, Ballabio A, Puertollano R. 2016. TFEB and TFE3 cooperate in the regulation of the innate immune response in activated macrophages. Autophagy 12: 1240-1258.

Phanstiel D, Boyle A, Araya C, Snyder MP. 2014. Sushi. R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. Bioinformatics 19:2808-2810.

Rao SS, Huang S-CC, Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon K-RR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, et al. 2017. Cohesin Loss Eliminates All Loop Domains. Cell 171: 305-320.e24.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159: 1665–80.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.

Robinson J, Thorvaldsdóttir H, Winckler W, Guttman M, Lander E, Getz G, Mesirov J. 2011. Integrative genomics viewer. Nat Biotech 29: 24–26.

Rohland N, Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res22: 939–46.

Rowat AC, Jaalouk DE, Zwerger M, Ung W, Eydelnant IA, Olins DE, Olins AL, Herrmann H, Weitz DA, Lammerding J. 2013. Nuclear envelope composition determines the ability of neutrophil-type cells to passage through micron-scale constrictions. J Biol Chem 288: 8610–8.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source platform for biological-image analysis. Nat Methods 9: 676–82.

Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L, et al. 2017. Two independent modes of chromatin organization revealed by cohesin removal. Nature 551: 51–56.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-JJ, Vert J-PP, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 16: 259.

Shultz LD, Lyons BL, Burzenski LM, Gott B, Samuels R, Schweitzer PA, Dreger C, Herrmann H, Kalscheuer V, Olins AL, et al. 2003. Mutations at the mouse ichthyosis locus are within the lamin B receptor gene: a single gene model for human Pelger-Huët anomaly. Hum Mol Genet 12: 61–9.

Tecchio C, Micheletti A, Cassatella MA. 2014. Neutrophil-derived cytokines: facts beyond expression. Front Immunol 5: 508.

Thomas CJ, Schroder K. 2013. Pattern recognition receptor function in neutrophils. Trends Immunol 34: 317–28.

Thorvaldsdóttir H, Robinson JT, Mesirov J. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14:178-192.

Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, Moulton HM, DeJesus P, Che J, Mulder LC, et al. 2015. Meta- and Orthogonal Integration of Influenza “OMICs” Data Defines a Role for UBR4 in Virus Budding. Cell Host Microbe 18: 723–35.

Zhu Y, Gong K, Denholtz M, Chandra V, Kamps MP, Alber F, Murre C. 2017. Comprehensive characterization of neutrophil genome topology. Genes Devel31: 141–153.

Chapter 5: Insights and future directions

It is now well studied that upon lineage commitment the genomes of adaptive and innate immune cells went through large-scale alterations in genome topology including locus contraction, genome contraction and nuclear repositioning. Locus contraction involves the interaction of variable regions linked with antigen receptor loci in adaptive immune cells (Jhunjhunwala et al., 2009). Genome contraction involves remote genomic interactions that shape the genomes of neutrophils and spermatozoa (Zhu et al., 2017; Battulin et al., 2015).

In chapter 2 we showed that B cell activation is indeed accompanied by big changes in genome folding pattern. The plasma cell specific transcription signature is associated with changes in compartmentalization. Many of the key transcription factors experienced repositioning of their regulatory elements, including *Ebf1*, *Atf4*, *Prdm1*, etc. The Ebf1 locus repositioned from the euchromatic compartment to peri-centromeric heterochromatin compartment, and the PC1 value of *Ebf1*'s distal regulatory element flipped from positive to negative. Beside traditional folding changes, we also found that coordinated transcription during B cell activation was correlated with interchromosomal association. This may be because genome-wide transcriptional coordination is achieved through interchromosomal transcription and splicing hubs like speckles. Further experiments should be done to validate the exact scale of this phenomenon and the mechanism. For example, large-scale positional transcript labeling method like merFISH can be done to illustrate how the interchromsomal pre-mature RNA are positioned and how the coordinated transcriptional burst correlates with the coordinated positioning (Xia et al, 2019). By simultaneously identify positional information and quantity information at genome scale, new imaging technique will provide us more details to decipher the phenomenon.

In chapter 3 we studied the regulation of key T cell fate-controlling gene *Bcl11b*. To our surprise, we discovered a novo mechanism of controlling enhancer 3D positioning. The transcription of the long non-coding RNA ThymoD at *Bcl11b*'s distal enhancer was necessary and sufficient to re-localize the enhancer from nucleus periphery to euchromatic compartment away from the nuclear lamina. The transcription of ThymoD led to binding of two CTCF sites pointing toward the *Bcl11b* promoter, which allowed proper cohesin accumulation and loop formation between the enhancer region and the promoter. However, still much is unknown about how the transcription of ThymoD regulate the binding of CTCF. One possible mechanism is through DNA demethylase TET1 and TET2. Transcription could recruit TET protein to demethylate the CTCF binding site (Benner et al, 2015), and the demethylated CpG then allowed CTCF binding (Wiehle et al. 2019). Other possibility includes recruitment of CTCF co-binding regulators, and CTCF / cohesin / demethylase / co-binding factor recruitment through formation of specially DNA structure like G4. Other than then CTCF binding mystery, the transcription and splicing of ThymoD was also carefully regulated. There are two major promoters responsible for transcribing ThymoD and different stages of T cell development. From the ThymoD p(A)/p(A) tumor data, we found that there were at least two other promoter that were dormant during most stages of normal T cell development. Why so many promoters are necessary to regulate ThymoD transcription and what's the difference between the variance isoform of ThymoD are still largely unanswered. Knocking out the ThymoD promoters one-by-one may enable us to further understand to elaborate regulation process.

In chapter 4, neutrophils activation was studied through a combination of high-through sequencing techniques and Imaging techniques. Like the other immune cell types, we found that human neutrophil development was closely associated with segmented compartments and contracted genomes. We employed two kinds of activation stimuli, PMA and *E. coli*, and both activate neutrophils with rapid changes in nuclear structure. The rapid changes mostly involve genome-wide loss of sub-domain boundary between inflammatory response gene domains, which enable long range promoter co-activation network and enhancer-promoter interaction to form. However, it’s still no clear how the elaborate mechanism genome structure is regulated. We suggest that enhancers and promoters may be spatially segregated from each other in distinct loop domains until an appropriate environmental signal is received in order to prevent inappropriate or pathological activation of gene expression. More experiment, for example, live imaging of some of the key super-enhancer and key TF promoters, can be done to validate the hypothesis.

In sum, here we demonstrate that genome architecture alteration is closely correlated with transcriptional regulation in immune cell development and activation. Many of the key genes are under careful governance of the chromatin folding pattern.

References

Battulin, N., Fishman, V.S., Mazur, A.M., Pomaznoy, M., Khabarova, A.A., Afonnikov, D.A., Prokhortchouk, E.B., and Serov, O.L. (2015). Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. Genome Biol. 16, 77.

Benner, C., Isoda, T. and Murre, C., (2015). New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing B cell progenitors.Proc. Natl. Acad. Sci. USA 112, 12776–12781.

Jhunjhunwala, S., van Zelm, M., Peak, M. and Murre, C. (2009). Chromatin architecture and the generation of antigen receptor diversity. Cell 138, 435-448.

Xia, C., Babcock, H.P., Moffitt, J.R. and Zhuang, X., (2019). Multiplexed detection of RNA using MERFISH and branched DNA amplification. Sci. Rep., 9, 7721.

Wiehle, L., Thorn, G.J., Raddatz, G., Clarkson, C.T., Rippe, K., Lyko, F., Breiling, A. and Teif, V.B., (2019). DNA (de) methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. Genome Res., 29, 750-761.

Zhu, Y., Gong, K., Denholtz, M., Chandra, V., Kamps, M.P., Alber, F. and Murre, C., （2017）. Comprehensive characterization of neutrophil genome topology. Genes Dev. 31, 141-153.cells controls CTCF-dependent chromatin boundaries. Genome Res., 29, 750-761.