



ERG Controls B Cell Development by Promoting Igh V-to-DJ Recombination

Søndergaard, Elisabeth; Rauch, Alexander; Michaut, Magali; Rapin, Nicolas; Rehn, Matilda; Wilhelmson, Anna S.; Camponeschi, Alessandro; Hasemann, Marie S.; Bagger, Frederik O.; Jendholm, Johan; Knudsen, Kasper J.; Mandrup, Susanne; Mårtensson, Inga Lill; Porse, Bo T.

Published in:
Cell Reports

DOI:
[10.1016/j.celrep.2019.10.098](https://doi.org/10.1016/j.celrep.2019.10.098)

Publication date:
2019

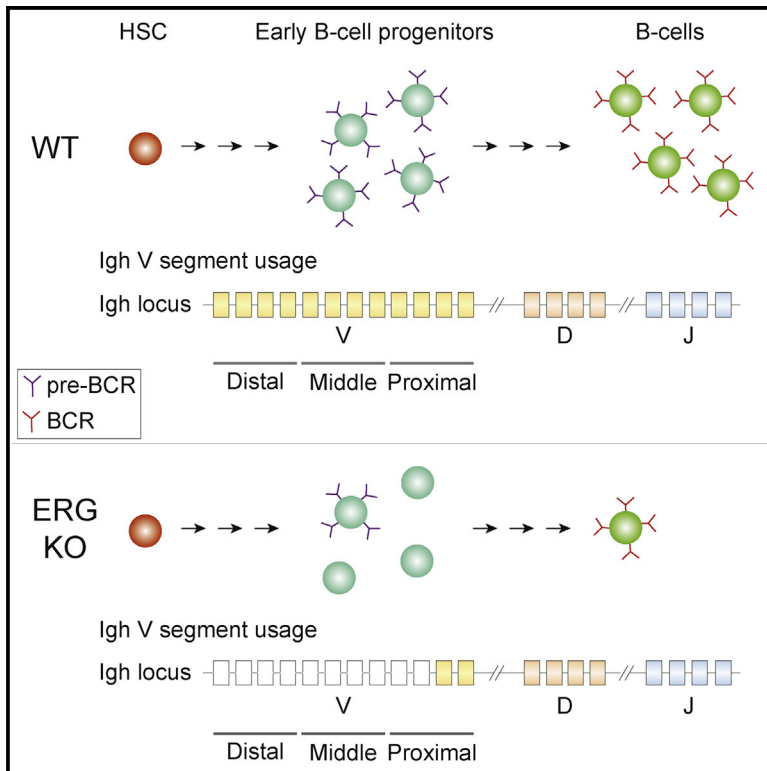
Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Citation for published version (APA):
Søndergaard, E., Rauch, A., Michaut, M., Rapin, N., Rehn, M., Wilhelmson, A. S., ... Porse, B. T. (2019). ERG Controls B Cell Development by Promoting *Igh* V-to-DJ Recombination. *Cell Reports*, 29(9), 2756-2769.e6. <https://doi.org/10.1016/j.celrep.2019.10.098>

ERG Controls B Cell Development by Promoting *Igh* V-to-DJ Recombination

Graphical Abstract



Authors

Elisabeth Søndergaard,
Alexander Rauch, Magali Michaut, ...,
Susanne Mandrup, Inga-Lill Mårtensson,
Bo T. Porse

Correspondence

bo.porse@finsenlab.dk

In Brief

Søndergaard et al. demonstrate that ERG is a critical transcriptional regulator essential for B cell development. Loss of ERG leads to a marked reduction in V-to-DJ recombination at the *Igh* locus, precluding B cell progenitors from expressing the pre-B cell receptor, which is required for entering the final stages of B cell development.

Highlights

- Loss of the transcription factor ERG leads to an early block in B cell development
- ERG is required for *Igh* V-to-DJ recombination involving distal V segments
- ERG is an important transcriptional regulator of B cell development



ERG Controls B Cell Development by Promoting *Igh* V-to-DJ Recombination

Elisabeth Søndergaard,^{1,2,3} Alexander Rauch,⁴ Magali Michaut,^{1,2,3,5} Nicolas Rapin,^{1,2,3,5} Matilda Rehn,^{1,2,3} Anna S. Wilhelmson,^{1,2,3} Alessandro Camponeschi,⁶ Marie S. Hasemann,^{1,2,3} Frederik O. Bagger,^{1,2,3,5} Johan Jendholm,^{1,2,3} Kasper J. Knudsen,^{1,2,3} Susanne Mandrup,⁴ Inga-Lill Mårtensson,⁶ and Bo T. Porse^{1,2,3,7,*}

¹The Finsen Laboratory, Rigshospitalet, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen N, Denmark

²Biotech Research and Innovation Centre (BRIC), University of Copenhagen, 2200 Copenhagen N, Denmark

³Danish Stem Cell Centre (DanStem) Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen N, Denmark

⁴Department of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense, Denmark

⁵The Bioinformatics Centre, Department of Biology, Faculty of Natural Sciences, University of Copenhagen, 2200 Copenhagen N, Denmark

⁶Department of Rheumatology and Inflammation Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 405 30 Gothenburg, Sweden

⁷Lead Contact

*Correspondence: bo.porse@finsenlab.dk

<https://doi.org/10.1016/j.celrep.2019.10.098>

SUMMARY

B cell development depends on the coordinated expression and cooperation of several transcription factors. Here we show that the transcription factor ETS-related gene (ERG) is crucial for normal B cell development and that its deletion results in a substantial loss of bone marrow B cell progenitors and peripheral B cells, as well as a skewing of splenic B cell populations. We find that ERG-deficient B lineage cells exhibit an early developmental block at the pre-B cell stage and proliferate less. The cells fail to express the immunoglobulin heavy chain due to inefficient V-to-DJ recombination, and cells that undergo recombination display a strong bias against incorporation of distal V gene segments. Furthermore, antisense transcription at PAX5-activated intergenic repeat (PAIR) elements, located in the distal region of the *Igh* locus, depends on ERG. These findings show that ERG serves as a critical regulator of B cell development by ensuring efficient and balanced V-to-DJ recombination.

INTRODUCTION

B cells are central players in the adaptive immune system, in which they fight various foreign antigens, which is made possible by their ability to synthesize a vastly diverse B cell antigen receptor (BCR) repertoire. The BCRs are generated during B cell development, which in the adult takes place in the bone marrow (BM), before the resulting immature B cells migrate to the spleen, where they undergo further maturation to naive B cells. In the BM, hematopoietic progenitors develop into B cells by passing through several progenitor stages in a highly regulated process. Key to this process is transcriptional control, and several transcription factors have been identified as essential regulators of B cell development (Sigvardsson, 2018).

Among these are E2A, FOXO-1, EBF-1, and PAX5, which work together to facilitate restriction and commitment to the B cell lineage during early B cell development (Lin et al., 2010; Nutt et al., 1999). Apart from transcriptional regulators, ordered expression of key cell surface receptors such as the interleukin 7 receptor (IL7R), pre-BCR, and BCR, is also required to faithfully execute the B cell developmental program (Kitamura et al., 1991; Pelanda et al., 2002; Peschon et al., 1994; Zou et al., 2003).

The main mechanism underlying the ability of B cells to obtain a large BCR repertoire is the process of V(D)J recombination. During this process, immunoglobulin gene segments are recombined sequentially; the immunoglobulin heavy-chain (*Igh*) locus is recombined before the immunoglobulin light-chain (*Igl*) loci. Furthermore, recombination of the *Igh* locus is a two-step process in which D-to-J precedes V-to-DJ recombination. Once in-frame recombination of the *Igh* locus has been obtained, synthesis of the immunoglobulin μ heavy chain (μ HC) commences. The μ HC assembles with the surrogate light chain (SLC) to form a pre-BCR and, subsequently, with the immunoglobulin light chain to form a BCR (Perlot and Alt, 2008).

The murine *Igh* locus comprises a 3-Mb large region located on chromosome 12 and encompasses 8 constant (C) regions, 4 J, \sim 10 functional D, and just over 100 functional V gene segments. The latter are divided into 16 V gene families of varying size that are dispersed over a \sim 2.5-Mb region, with the individual families and members located different distances from the D gene segments (Johnston et al., 2006). Recombination of the *Igh* locus is both complex and highly regulated, and many changes occur at the locus as a prelude to recombination. These changes include nuclear relocalization, changes in DNA accessibility, initiation of non-coding RNA transcription, and changes in the 3D conformation of the locus (Kumari and Sen, 2015). The latter is particularly important to obtain fairly even usage of V gene segments, despite their different localizations relative to the D genes. The resulting conformation of the *Igh* locus is characterized by local loop domains and large-scale locus contraction that are facilitated by the binding of several factors (Degner et al., 2011; Fuxa et al., 2004; Guo et al., 2011; Jhunjhunwala



et al., 2008; Kosak et al., 2002; Liu et al., 2007; Medvedovic et al., 2013).

ETS-related gene (ERG), a member of the ETS domain family of transcription factors, has been shown to be essential for definitive hematopoiesis and to promote leukemic development (Baldus et al., 2006; Loughran et al., 2008; Martens, 2011; Rainis et al., 2005). Moreover, our group has identified ERG as a critical regulator of hematopoietic stem cell maintenance (Knudsen et al., 2015). Previous work has also hinted at a role for ERG in B cell development. Specifically, overexpression of *Erg* was shown to confer a survival advantage to progenitor B cells, and mice harboring a heterozygous mutation in the ERG transactivation domain exhibited a trend toward fewer peripheral B cells (Loughran et al., 2008; Tsuzuki et al., 2011). ERG has also been linked to B cell acute lymphoblastic leukemia (B-ALL), in which heterozygous deletion of *ERG*, in combination with an *IGH-DUX4* fusion, was recently shown to define a novel B-ALL subtype (Lilljebjörn et al., 2016; Zhang et al., 2016). Collectively, these observations suggest that ERG may play a role during normal B cell development.

In the present work, we set out to test the potential role of ERG during B cell development. By using a lymphoid-restricted *Erg* knockout mouse line, we showed that the ablation of *Erg* resulted in a substantial reduction in peripheral B cell numbers and a skewing in the ratio of splenic B cell subsets. We mapped the loss of peripheral B cells to an early block in BM B cell development and found that ERG-deficient progenitor B cells displayed decreased levels of the IL7R and the pre-BCR, thus resulting in reduced proliferation. Mechanistically, we showed that ERG-deficient progenitor B cells were unable to undergo efficient *Igh* V-to-DJ recombination, as evidenced by fewer rearrangements and a strong bias against the use of distal V gene segments. Altogether, these results identify ERG as a factor of critical importance during early B cell development, with a specific role during *Igh* locus recombination.

RESULTS

ERG Is Required for Normal B Cell Development

To assess the role of ERG in lymphoid development, we introduced the *hCD2-iCre* allele into our previously generated conditional *Erg* mouse line, which facilitates deletion of the DNA binding domain of ERG (de Boer et al., 2003; Knudsen et al., 2015). Because the *hCD2* promoter is activated at the common lymphoid progenitor (CLP) stage, the resulting *Erg^{fl/fl};CD2iCre* (from here on, *Erg^{Δ/Δ}*) mice lack ERG activity in all lymphoid cells (Figure 1A) (Schindler et al., 2009). Normally, *Erg* is expressed at substantial levels in CLPs, and the expression is sustained during the early stages of B and T cell development, as assessed via publicly available gene expression repositories (Figures S1A and S1B) (Bagger et al., 2016).

We first examined the cell lineage distribution in peripheral blood from control and *Erg^{Δ/Δ}* mice by flow cytometry. Loss of ERG resulted in a marked reduction of peripheral blood B cells, while the frequencies of T, myeloid, and natural killer (NK) cells increased concomitantly (Figures 1B and 1C). This phenotype was accompanied by a 50% drop in spleen and inguinal lymph node masses, presumably due to the strong reduction of B cells in these organs. (Figures 1D and 1E).

Given the substantial decrease in peripheral B cell numbers, we analyzed the splenic B cell compartment and noted a pronounced skewing of B cell subsets in *Erg^{Δ/Δ}* mice (Figures 1F and 1G). Specifically, follicular (FO) B cells were dramatically decreased, both in frequency and in number, while the opposite was true for marginal zone (MZ) B cells. The frequency of B1 B cells was unaffected in *Erg^{Δ/Δ}* mice, but we observed a mild skewing in the B1 B cell subsets, resulting in an increased B1a-to-B1b ratio (Figure S1C). Furthermore, the transitional B cells were almost absent. Thus, the loss of ERG affected the earliest splenic B cell populations, likely reflecting that fewer immature B cells were seeding this organ.

The strong reduction in peripheral B cell frequencies following lymphoid-specific loss of ERG could in principle mask a potential T cell phenotype in *Erg^{Δ/Δ}* mice (Figures 1B, 1C, and S1D). We therefore assessed the thymic T cell compartment and found no effects on relative frequencies of mature T cells, although their overall numbers were somewhat decreased (Figure S1E). Furthermore, only minor changes were detected in T cell progenitor numbers and frequencies (Figures S1E and S1F). Given that the effect on B cells was more pronounced, we focused our efforts on this compartment.

In summary, loss of ERG in the lymphoid compartment markedly affects B cell numbers in peripheral blood and lymphoid organs, which may be caused by lower efflux of immature B cells from the BM.

ERG-Deficient Progenitors Encounter an Early B Cell Developmental Block

The data presented so far suggest that ERG may play a role in early B cell development, which is a continuous process taking place in the BM (Figure 2A). To address the earliest stages of B cell development, we first verified that deletion of *Erg*, in our model, was initiated at the CLP stage and was efficient in all tested downstream populations (Figure S2A). Lymphoid-specific loss of ERG was only associated with a mild reduction in overall BM cellularity, but this translated into a 6-fold reduction in the number of B220⁺CD19⁺ cells, thus demonstrating a strong impact on the B cell lineage (Figures 2B, 2C, and S2B).

We next combined panels of B lineage markers to pinpoint the main stages of B cell development affected in *Erg^{Δ/Δ}* animals. Normally, upregulation of CD25 occurs at the pre-BI to pre-BII transition, at which stage c-Kit has been downregulated and pre-BCR signaling commences (Rolink et al., 1994). In *Erg^{Δ/Δ}* mice, we detected a tendency toward a reduction in the number of c-Kit⁺ B cell progenitors and observed a clear population of CD25⁺ cells, of which a small percentage expressed both c-Kit and CD25, markers that are normally mutually exclusive (Figures 2D, 2F, and S2C). Furthermore, within the normal B cell progenitor compartment, most cells that are CD43⁺ do not yet express CD25 (herein P1; CD43⁺CD25⁻), and vice versa (P3; CD43⁻CD25⁺) (Figures 2D–2F). In sharp contrast to control cells, *Erg^{Δ/Δ}* progenitors failed to downregulate CD43 and became arrested at an unconventional CD43⁺CD25⁺ (P2) stage. The accumulation of cells in P1 and P2 and the decrease in P3 cells strongly suggest that pre-BCR signaling was derailed following loss of ERG, and as a consequence, we decided to use the CD43/CD25 marker combination for subsequent analyses.

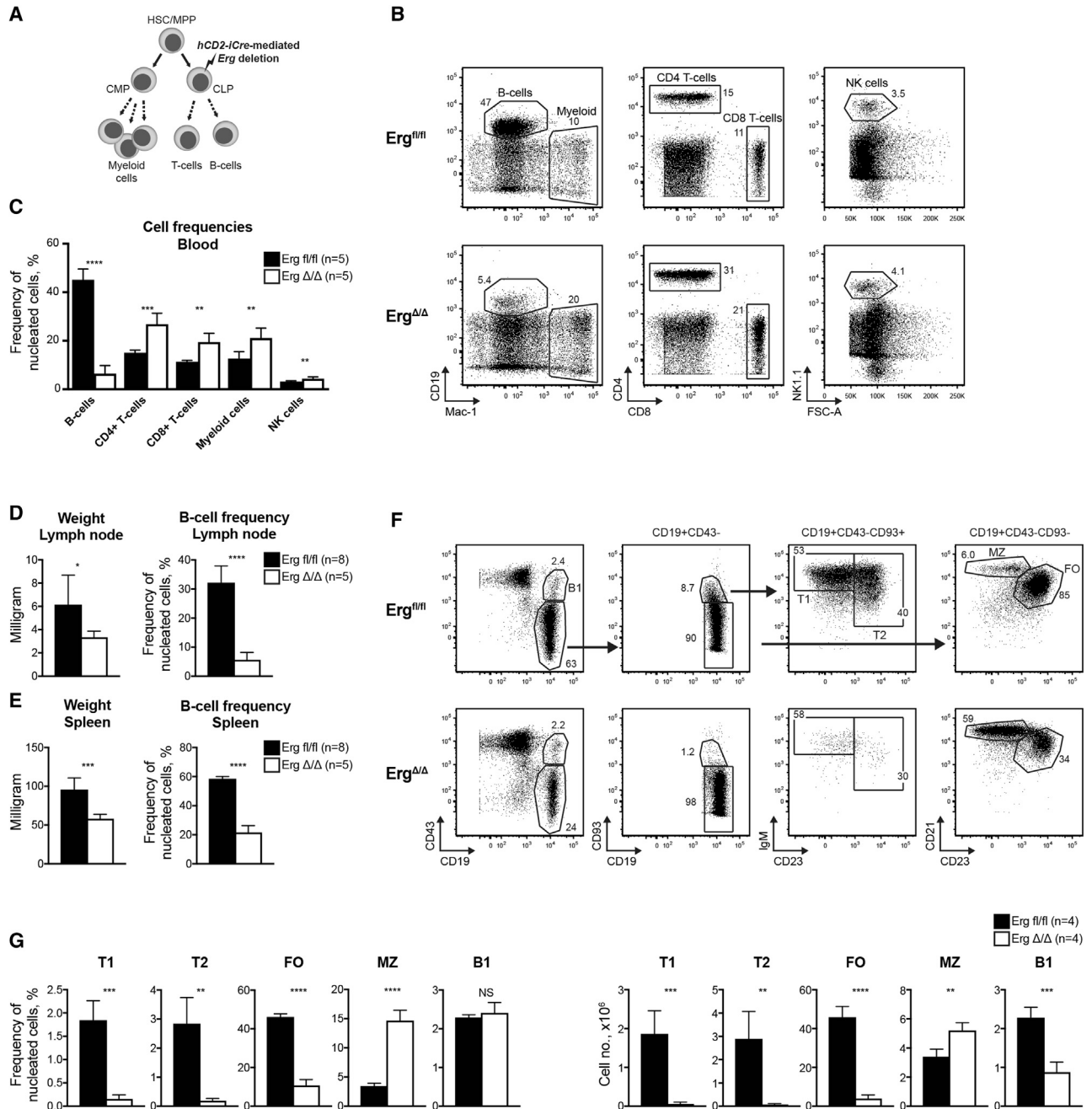
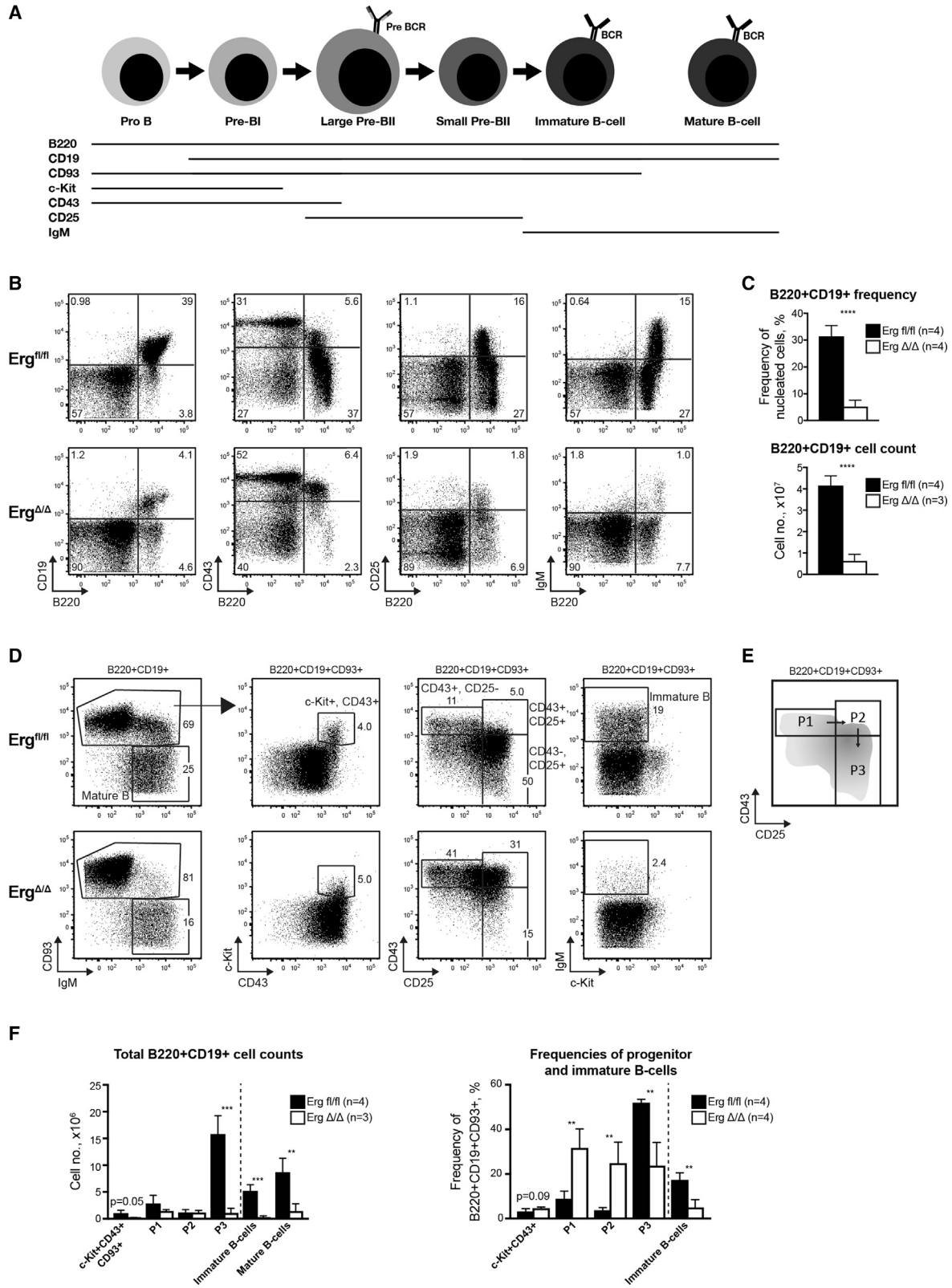


Figure 1. ERG Is Required for Normal B Cell Development

(A) Deletion of the DNA-binding domain of *Erg* takes place at the CLP stage, leading to a pan-lymphoid loss of ERG. (B) Flow cytometry analysis of peripheral blood from *Erg*^{fl/fl} and *Erg*^{fl/fl}; *CD2iCre* (*Erg*^{Δ/Δ}) mice. (C) Quantification of the flow analysis in (B), showing lineage frequencies in peripheral blood. (D) Weights and B cell frequencies (B220+ cells, as determined by flow cytometry) of inguinal lymph nodes from *Erg*^{fl/fl} and *Erg*^{Δ/Δ} mice. (E) Weights and B cell frequencies (B220+ cells, as determined by flow cytometry) of spleens from *Erg*^{fl/fl} and *Erg*^{Δ/Δ} mice. (F) Flow cytometry analysis assessing the splenic B cell subsets of *Erg*^{fl/fl} and *Erg*^{Δ/Δ} mice. Cell populations were defined as shown by the gating strategy. (G) Quantification of the flow cytometry analysis in (F). Left, the frequencies of the various B cell populations of the spleen. Right, total number of cells within each population. For the flow cytometry analyses, representative flow cytometry plots for the two genotypes are shown. For quantification, the data are shown as mean + SD. Statistical analyses were performed using unpaired, two-tailed Student's t test. **p* ≤ 0.05, ***p* ≤ 0.01, ****p* ≤ 0.001, *****p* ≤ 0.0001. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; T, transitional; FO, follicular; MZ, marginal zone. See also Figure S1.



(legend on next page)

Heterozygous loss of ERG is associated with the development of B-ALL, and we therefore tested whether *Erg*^{Δ/+} animals display B cell abnormalities that could set the stage for cancer development. However, we only detected minor changes in the B cell lineage of these mice (Figures S2D and S2E). Moreover, to assess whether the B cell phenotype reported earlier for young *Erg*^{Δ/Δ} mice progressed during aging, we phenotyped aged (~1.5 years) *Erg*^{Δ/Δ} mice but found no evidence of further phenotypic progression (Figures S2F–S2H). These data demonstrate that the sole loss of ERG is not sufficient to drive development of B cell malignancies.

Collectively, these findings suggest that the main function of ERG during B cell development is to efficiently push cells through stages in which the *Igh* locus undergoes recombination and pre-BCR expression is initiated.

ERG Is Required for the Expression of IL7R and μ HC to Drive B Cell Progenitor Proliferation

To directly examine whether loss of ERG was associated with failure to initiate pre-BCR expression, we next quantified the intracellular levels of the immunoglobulin μ HC (Figures 3A and 3B). Strikingly, whereas μ HC expression in control mice was initiated in P1 and reached maximum levels in P2, these two progenitor populations were nearly devoid of μ HC in *Erg*^{Δ/Δ} mice. Most *Erg*^{Δ/Δ} cells that did reach P3 displayed relatively normal μ HC levels, suggesting that the few cells that succeed in downregulating CD43 are progenitors that express a pre-BCR. We also measured the expression of IL7R, another essential receptor for B cell development (Peschon et al., 1994), and found it to be expressed throughout the P1–P3 progenitor populations, albeit at lower levels in *Erg*^{Δ/Δ} cells (Figures 3C and 3D). Because both IL7R and pre-BCR have been associated with B cell progenitor proliferation (Corcoran et al., 1996; Hess et al., 2001), we next quantified the distribution of cells across the cell cycle in the P1–P3 populations. Consistent with a proliferative phenotype, we found that loss of ERG led to a marked decrease of cells in the S and G2/M phases of the cell cycle, accompanied by a compensatory increase of cells in G0/G1 (Figures 3E and 3F). These data are supported by *in vivo* bromodeoxyuridine (BrdU)-incorporation experiments, which demonstrated not only a marked decrease in the frequencies of replicating progenitors in *Erg*^{Δ/Δ} mice but also a reduction in the amount of label incorporated in individual cells (Figures 3G and 3H). The latter could reflect a lower DNA synthesis rate that would be expected

to lead to an accumulation of cells in S phase. However, as we observe accumulation of cells in G0/G1, we favor an interpretation in which loss of ERG leads to accumulation of cells at the G1-S boundary and delayed entry into S phase.

In summary, loss of ERG is associated with major cell cycle deficits during B cell development, likely caused by moderately reduced levels of IL7R and failure to express a pre-BCR.

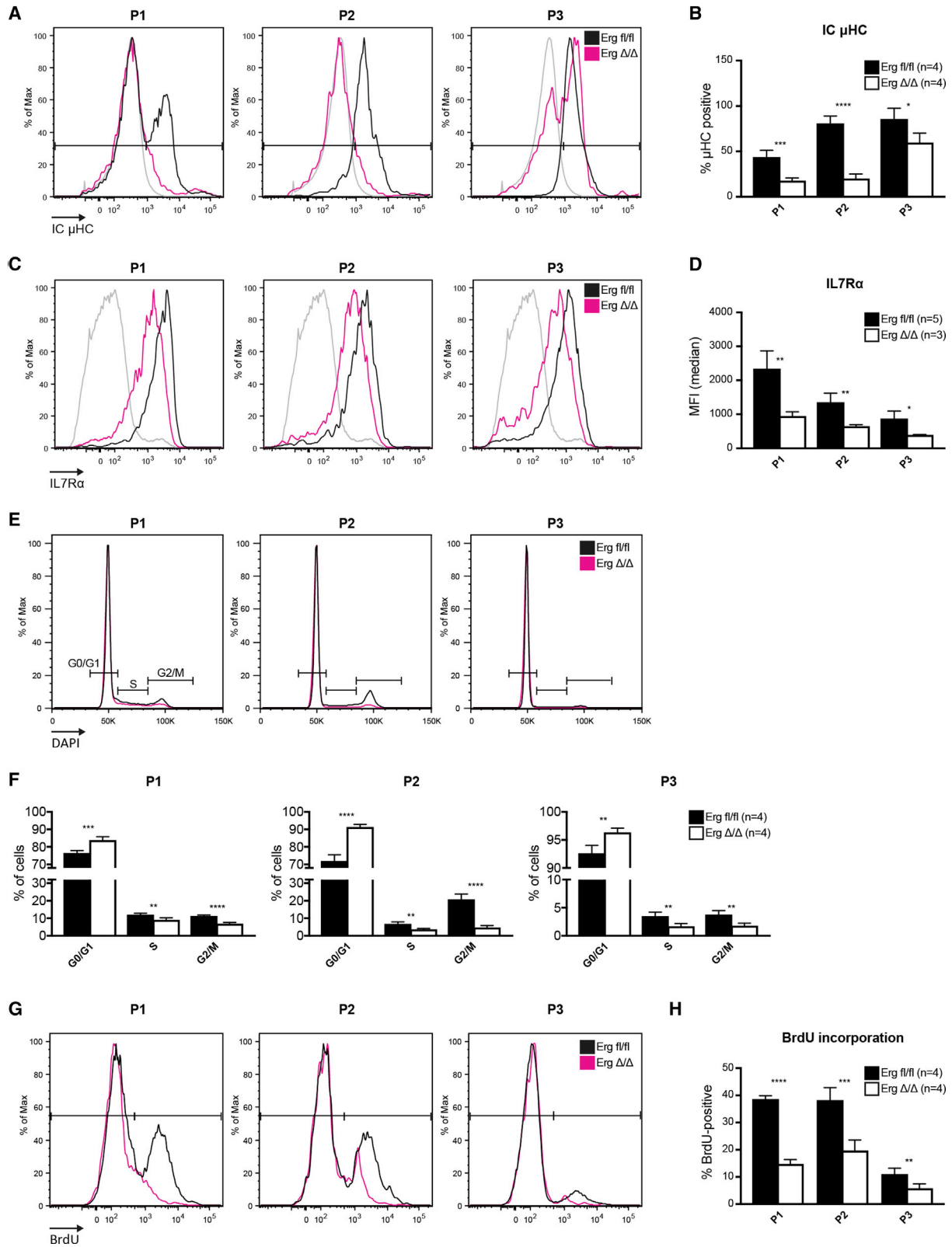
ERG Is Part of a B Cell Transcriptional Network that Drives B Cell Development

To gain insights into the molecular mechanisms underlying the block in B cell development in ERG-deficient mice, we next subjected the P1–P3 populations to RNA sequencing (RNA-seq) (Table S1). On the global level, we detected a substantial number of deregulated genes in the ERG-deficient P1 and P2 progenitor populations (P1: 115 up, 118 down; P2: 91 up, 131 down; adjusted p value ≤ 0.05 ; log2 fold change ≥ 1 or ≤ -1). In contrast, fewer genes were differentially expressed at the P3 cell stage (P3: 37 up, 9 down; adjusted p value ≤ 0.05 ; log2 fold change ≥ 1 or ≤ -1), supporting the notion that the main functional deficits following loss of ERG occur before pre-BCR expression (Figure 4A). Consistent with the observed cell cycle deficit, gene set enrichment analysis (GSEA) demonstrated an overall depletion of gene signatures associated with active cell proliferation (exemplified by E2F targets, G2M checkpoint, and chromosome segregation) in all three ERG-deficient progenitor subsets (Figure 4B; Table S2). We also noted a depletion of a MYC target gene signature, as well as depletion of several translation- and ribosome-related signatures in *Erg*^{Δ/Δ} P1 progenitors (Figure 4B; Table S2). This appears to be driven by reduced *Myc* expression and is consistent with the well-known impact of MYC on protein synthesis (Figure 4A) (van Riggelen et al., 2010).

To understand the impact of ERG on chromatin accessibility and how ERG integrates with established B cell transcription factors, we performed Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) on control and *Erg*^{Δ/Δ} progenitor B cells. Consistent with the GSEA, we found that genes associated with sites exhibiting decreased accessibility following loss of ERG were downregulated and enriched for Gene Ontology (GO) categories associated with proliferation (Figures S3A–S3C). We next used the Integrated analysis of Motif Activity and Gene Expression changes of transcription factors (IMAGE) tool (Madsen et al., 2018) to determine changes

Figure 2. ERG-Deficient Progenitors Encounter an Early B Cell Developmental Block

- (A) Schematic figure of B cell development in the BM from the pro-B cell stage to mature B cells. Surface marker expression is annotated.
- (B) Flow cytometry analysis of BM cells from *Erg*^{fl/fl} and *Erg*^{Δ/Δ} mice.
- (C) Quantification of the analysis in (B). Top, frequency of the BM cells that are double-positive for B220 and CD19. Bottom, total quantification of the number of B220+CD19+ cells in the BM (2 \times femur, tibia, ilium) of *Erg*^{fl/fl} and *Erg*^{Δ/Δ} mice.
- (D) Flow cytometry analysis of the B220+CD19+ compartment in *Erg*^{fl/fl} and *Erg*^{Δ/Δ} mice.
- (E) Schematic flow cytometry plot depicting the BM B cell progenitor populations called P1–P3.
- (F) Quantification of the flow cytometry analysis seen in (D), according to the gating strategy shown in the plots. Left, quantification of the total number of cells within the indicated cell populations in the BM (2 \times femur, tibia, ilium) of *Erg*^{fl/fl} and *Erg*^{Δ/Δ} mice. Right, frequency of the indicated cell populations within the B220+CD19+CD93+ progenitor and immature B cell gate. All populations are B220+CD19+. Dashed lines separate progenitor B cells from more downstream populations.
- For the flow cytometry analyses, representative plots are shown. For quantification, the data are shown as mean + SD. Statistical analyses were performed using unpaired, two-tailed Student's t test. *p ≤ 0.05 , **p ≤ 0.01 , ***p ≤ 0.001 , ****p ≤ 0.0001 . See also Figure S2.



(legend on next page)

in transcription factor motif activities following loss of ERG. IMAGE models the contribution of transcription factor motifs to the regulatory activity of ATAC regions based on the ATAC-seq signal and the presence of transcription factor motifs in ATAC regions ($n = 52,311$). For the contribution of motifs to gene expression, motifs within the ATAC regions are weighted based on distance to the respective transcription start sites. Thus, changes in motif activity between conditions, e.g., control and *Erg* $^{\Delta/\Delta}$ progenitor B cells, indicate that a particular transcription factor contributes differentially to the overall regulatory activity of ATAC regions in the context of all analyzed motifs (Rauch et al., 2019). Gratifyingly, we found that the ERG motif, together with the motifs of the two key B cell transcription factors, EBF1 and PAX5, exhibited a marked reduction in regulatory activity following loss of ERG (Figure S3D). Using the IMAGE tool-based prediction of transcription factor target genes, we could show that transcription factors with decreased motif activity in *Erg* $^{\Delta/\Delta}$ progenitor B cells were highly interconnected, e.g., regulated the expression of each other, compared with those with increased motif activity or randomized transcriptional networks (Figures S3E and S3F). These findings suggest that ERG, together with PAX5 and EBF1, is a key component of a large transcription factor network responsible for driving B cell development.

Having established the more global impact of ERG on B cell transcriptional control, we turned our attention to the expression of individual genes important for B cell development (Figure 4C). Among B lineage factors, we detected a minor reduction in *Ebf-1* expression (\log_2 fold change = -0.78 ; adjusted p value = 0.048). However, because heterozygous loss of EBF-1 has a relatively mild impact on B cell development, the observed reduction in *Ebf-1* expression is unlikely to explain the observed *Erg* $^{\Delta/\Delta}$ phenotype (Lukin et al., 2010). Loss of ERG is associated with a dramatic reduction in μ HC, and we therefore assessed the expression of other pre-BCR components at the P1–P3 progenitor stages. At the P1 stage, there were no significant differences in the expression of any of these receptor components. At the following cell stages, the expression levels of the SLC genes (i.e., *Igll1* and *Vpreb1*) were downregulated in control cells, which is expected subsequent to signaling from the pre-BCR (Parker et al., 2005). An equivalent decrease in SLC gene expression was not observed in the *Erg* $^{\Delta/\Delta}$ cells, likely due to a lack of pre-BCR signaling. Strikingly, *Erg* displayed a similar expression profile, perhaps suggesting that it is also downregulated in response to

pre-BCR signaling (Figure 4C). Finally, we noted a marked reduction in the expression of a substantial number of *Igh* locus V gene segments in *Erg* $^{\Delta/\Delta}$ cells, especially in the P2 population, which could explain the near absence of μ HC expression (Figure 4A, yellow dots). In contrast to the changes observed for the expression of gene segments of the *Igh* locus, we did not observe significant transcriptional changes at the immunoglobulin light-chain loci, as assessed in the most mature P3 population (Table S1).

The reduction in *Igh* V gene segment expression does not appear to be associated with changes in the expression of vital genes encoding members of the recombination machinery or regulators of the 3D structure of the locus (Figure 4C). In *Erg* $^{\Delta/\Delta}$ P1 progenitors, we did observe a reduction in the expression of *Dntt*, which encodes TdT, the protein that is responsible for n-nucleotide incorporation during V(D)J recombination. However, this cannot account for the *Erg* $^{\Delta/\Delta}$ phenotype, because *Dntt*-deficient mice do not display a B cell developmental phenotype (Schelonka et al., 2011).

Collectively, our chromatin and transcriptional analyses demonstrate an integral role of ERG in a transcriptional network driving B cell development. Moreover, they identify transcriptional deregulation of *Igh* V gene segment expression as a potential cause of the B cell developmental phenotype in *Erg* $^{\Delta/\Delta}$ mice.

Loss of ERG Abrogates Efficient V-to-DJ Recombination

Given that our functional and molecular analyses demonstrate that the main consequence of ERG deficiency is centered on the *Igh* locus, we were intrigued by the transcriptional downregulation of several *Igh* V gene segments. We therefore analyzed the RNA-seq data of all V gene segments, from proximal to distal, and found a marked overall reduction in their expression in *Erg* $^{\Delta/\Delta}$ P1 and P2 progenitors (Figures 5A–5C). This reduction was less pronounced in *Erg* $^{\Delta/\Delta}$ P3 progenitors, in which *Igh* recombination has already occurred. Instead, we detected a clear preference for the expression of proximal V gene segments in *Erg* $^{\Delta/\Delta}$ P3 progenitors compared with their wild-type counterparts.

To determine whether the reduction in V gene segment expression was due to reduced V-to-DJ recombination, we next used the high-throughput sequencing-based immunoSEQ platform to assess the VDJ junctions of the *Igh* loci. This analysis revealed a strong reduction in the total numbers of rearrangements in *Erg* $^{\Delta/\Delta}$ P1 and P2 progenitors, whereas P3 and the mature FO and MZ B cell populations were unperturbed (Figure 6A).

Figure 3. ERG Is Required for the Expression of IL7R and μ HC to Drive B Cell Progenitor Proliferation

(A) Histograms showing the levels of intracellular immunoglobulin μ heavy chain (IC μ HC) in P1–P3 cells from *Erg* $^{fl/fl}$ and *Erg* $^{\Delta/\Delta}$ mice as assessed by flow cytometry. Cells from four mice of each genotype were examined. The gray line illustrates the level of IC μ HC in B220-CD19 $^-$ cells.
 (B) Quantification of the flow cytometry analysis in (A). The percentages of μ HC-positive cells were determined using the gating shown in the histogram plots.
 (C) Histograms showing the surface expression of IL7R α on P1–P3 cells as assessed by flow cytometry. The gray line illustrates the level of IL7R α on B220-CD19 $^-$ cells.
 (D) Quantification of the flow cytometry analysis in (C). MFI, median fluorescence intensity.
 (E) Cell cycle analysis of P1–P3 cells from *Erg* $^{fl/fl}$ and *Erg* $^{\Delta/\Delta}$ mice as determined by DAPI staining using flow cytometry.
 (F) Quantification of the flow cytometry analysis in (E). The frequencies of cells found in the different cell cycle stages were determined using the gates shown in the histogram plots.
 (G) Mice were injected with BrdU, and the bones were harvested 3 h later. The histograms depict the levels of incorporated BrdU in P1–P3 cells of *Erg* $^{fl/fl}$ and *Erg* $^{\Delta/\Delta}$ mice as measured by flow cytometry.
 (H) Quantification of the flow cytometry analysis in (G). The percentages of BrdU-positive cells were determined using the gating shown in the histogram plots. For the flow cytometry analyses, representative plots are shown. For quantification, the data are shown as mean + SD. Statistical analyses were performed using unpaired, two-tailed Student's t test. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

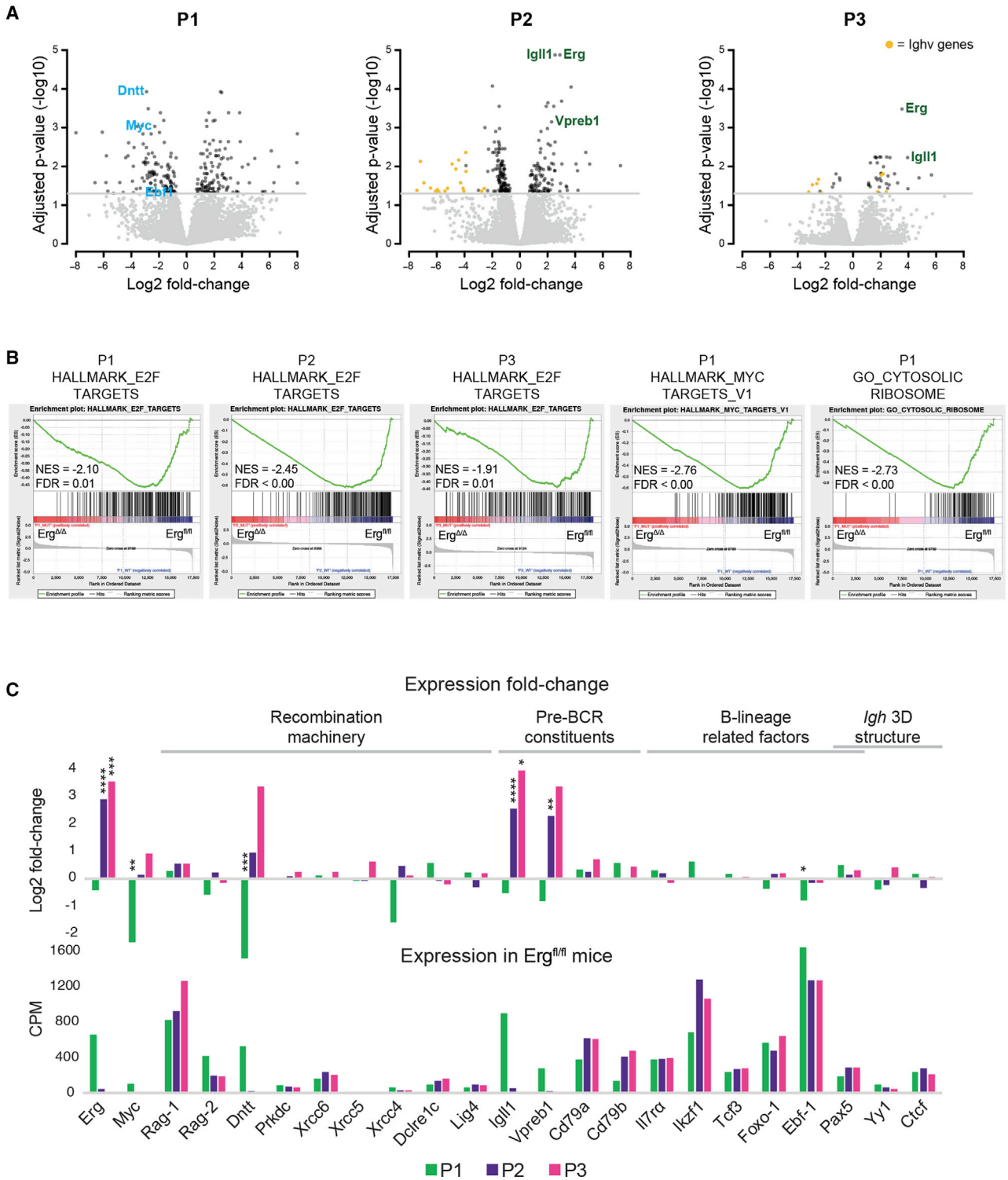


Figure 4. Gene Expression Analysis of ERG-Deficient B Cell Progenitors

(A) RNA sequencing (RNA-seq) was performed on RNA in four biological replicates isolated from P1–P3 cells from *Erg*^{fl/fl} and *Erg*^{Δ/Δ} mice (see also Table S1). Volcano plots depict differentially expressed genes between *Erg*^{fl/fl} and *Erg*^{Δ/Δ} cells in the P1 (left), P2 (middle), and P3 (right) populations. Selected genes are highlighted by gene names and color coded in blue (down in *Erg*^{Δ/Δ}) or green (up in *Erg*^{Δ/Δ}). *IghV* gene segments that are significantly deregulated are illustrated by yellow dots (P1, 0 gene segments; P2, 20 gene segments; P3, 9 gene segments). Genes with a fold change of < -8 or > 8 are located at -8 or 8 on the x axis, respectively.

(legend continued on next page)

Moreover, the frequency of non-productive rearrangements was increased in the P1 and P2 populations (Figure 6B). We also found that loss of ERG was associated with a reduction in complementary-determining region 3 (CDR3) length, which was most pronounced at the early progenitor stages (Figure 6C). Finally, when we quantified the frequencies of productive rearrangements for distinct V gene segment families, there was a clear preference for the use of proximal families in ERG-deficient cells (Figures 6D, 6E, and S4A). Collectively, these data demonstrate that ERG is required for efficient and balanced V-to-DJ recombination.

V-to-DJ recombination is governed by several distinct mechanisms. Because the reduction in the use of distal V gene segments in ERG-deficient progenitors could be due to decreased chromatin accessibility at these regions, we assessed the ATAC-seq data, but we found no differences in the overall accessibility of the locus (Figure 6F). Another feature that has been tightly linked to V-to-DJ recombination is antisense transcription that is initiated at regulatory PAX5-activated intergenic repeat (PAIR) elements in the distal end of the *Igh* locus in progenitor B cells and that ceases immediately after recombination (Ebert et al., 2011; Liu et al., 2007; Verma-Gaur et al., 2012). PAIR elements have been shown to be involved in contacts between the 5' and the 3' ends of the locus, and in particular, PAIR4, PAIR6, and PAIR11 display high levels of transcription. In accordance with these previous studies, we observed high levels of antisense transcription at these PAIR elements in control P1 progenitors, but not in the subsequent P2 and P3 cell subsets (Figures 6G and S4B). In contrast, loss of ERG nearly ablated expression of PAIR-associated antisense RNA, suggesting a role for ERG in this process. In an attempt to connect ERG binding in B cell progenitors to events at the *Igh* locus, we performed numerous chromatin immunoprecipitation (ChIP) experiments but were unable to obtain high-quality data. Instead, we took advantage of an existing dataset derived from the immature murine hematopoietic HPC-7 cell line (Wilson et al., 2010). Consistent with a role for ERG in mediating recombination involving distal V gene segments, ERG exhibits extensive binding at these regions (Figure S4C).

In summary, we find that loss of ERG impairs V-to-DJ recombination, in particular to distal V gene regions, and ablates PAIR-associated antisense transcription.

DISCUSSION

Transcriptional control is key to B cell development, and in the present work, we identify ERG as an important player in this process. Using a lymphoid-specific *Erg* knockout mouse model, we showed that loss of ERG resulted in an early block in B cell development accompanied by loss of peripheral B cells, with the intriguing exception of MZ B cells. The inability of *Erg*^{Δ/Δ} B cell progenitors to efficiently synthesize μ HC, and thus a pre-BCR, was shown to be caused by derailed V-to-DJ recombination, characterized by reduced recombination frequency and a proximal V

gene segment bias. Interestingly, loss of ERG resulted in a developmental block at an unconventional CD43⁺CD25⁺ P2 cell stage immediately before pre-BCR assembly. Normally CD25 is expressed concomitantly with the pre-BCR; however, this relation is clearly short-circuited in the absence of ERG, indicating that CD25 expression is driven by signals unrelated to V-to-DJ recombination and μ HC expression (Rolink et al., 1994). Consistent with their inability to elicit proper pre-BCR signaling, *Erg*^{Δ/Δ} B cell progenitors display reduced proliferation rates, which is most likely exacerbated by their inability to express normal levels of known proliferative drivers such as IL7R and MYC (Corcoran et al., 1996; Hess et al., 2001; Vallespinós et al., 2011). Finally, transcriptional and chromatin accessibility analyses place ERG as a key constituent of a transcription factor network important for B cell development.

Apart from a marked general depletion of peripheral B cells, our data demonstrate that ERG affects the relative distribution of mature splenic B cell subsets, and specifically, we find that the numbers of MZ B cells were augmented in *Erg*^{Δ/Δ} mice. VDJ sequencing showed that among the tested cell populations, MZ B cells had the shortest average CDR3 length in control mice. Furthermore, this was the only population in which we did not observe any difference in the average CDR3 length between *Erg*^{Δ/Δ} and control cells. This suggests that the shorter CDR3s in *Erg*^{Δ/Δ} B lineage cells could influence the mature B cell pool. In support of this, previous work has demonstrated that MZ B cells have shorter CDR3s compared with FO B cells and that a larger proportion of the MZ B cells harbors *Igh* rearrangements lacking n-nucleotides (Carey et al., 2008; Prohaska et al., 2018). Furthermore, loss of *Dnmt* is associated with increased MZ B cell potential (Carey et al., 2008). Thus, we propose that the increase in MZ B cells in *Erg*^{Δ/Δ} mice is due to a preference of ERG-deficient progenitors to develop into the MZ lineage. This priming takes place in the *Erg*^{Δ/Δ} BM and is likely caused by an overall truncation of the CDR3 or by reduced n-nucleotide incorporation as a result of decreased *Dnmt* expression. A direct role of ERG in peripheral B cells is less probable given its low expression in these cells (Figure S1A).

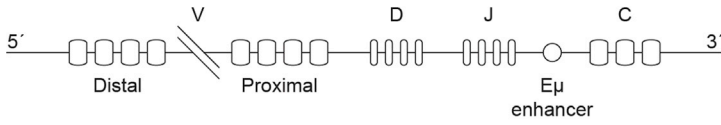
We demonstrated that the failure of the *Erg*^{Δ/Δ} progenitor B cells to efficiently express the μ HC was associated with a strong V(D)J recombination deficiency. Not only were overall recombination frequencies reduced, but productively rearranged alleles were markedly depleted for distal V gene segments. Part of the recombination deficiency could be caused by reduced D-to-J recombination rates, which would result in fewer V-to-DJ recombination events. In this scenario, recombination to both proximal (*Igh*V2+*Igh*V5) and distal (*Igh*V1+*Igh*V8) V gene segments would be expected to be decreased. However, we only observe a significant decrease in the number of rearrangements involving distal, but not proximal, V gene segments when assessed in the immature P1 progenitor population. Although we cannot rule out an effect on DJ recombination, this finding indicates that the main V(D)J

(B) Gene set enrichment analysis (GSEA) (see also Table S2) of the RNA-seq data of the three progenitor populations, *Erg*^{Δ/Δ} versus *Erg*^{fl/fl}.

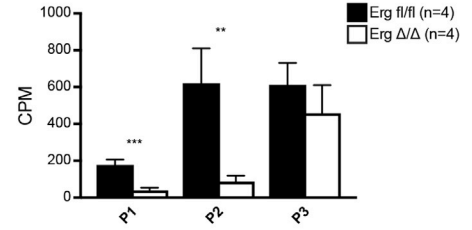
(C) Gene expression changes of selected genes as examined by RNA-seq. P1, P2, and P3 populations are represented with green, purple, and pink bars, respectively. Top, log2 fold change in gene expression, *Erg*^{Δ/Δ} versus *Erg*^{fl/fl}. Bottom, average gene expression values in *Erg*^{fl/fl} cells, depicted as average CPM (counts per million). *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001.

See also Figure S3.

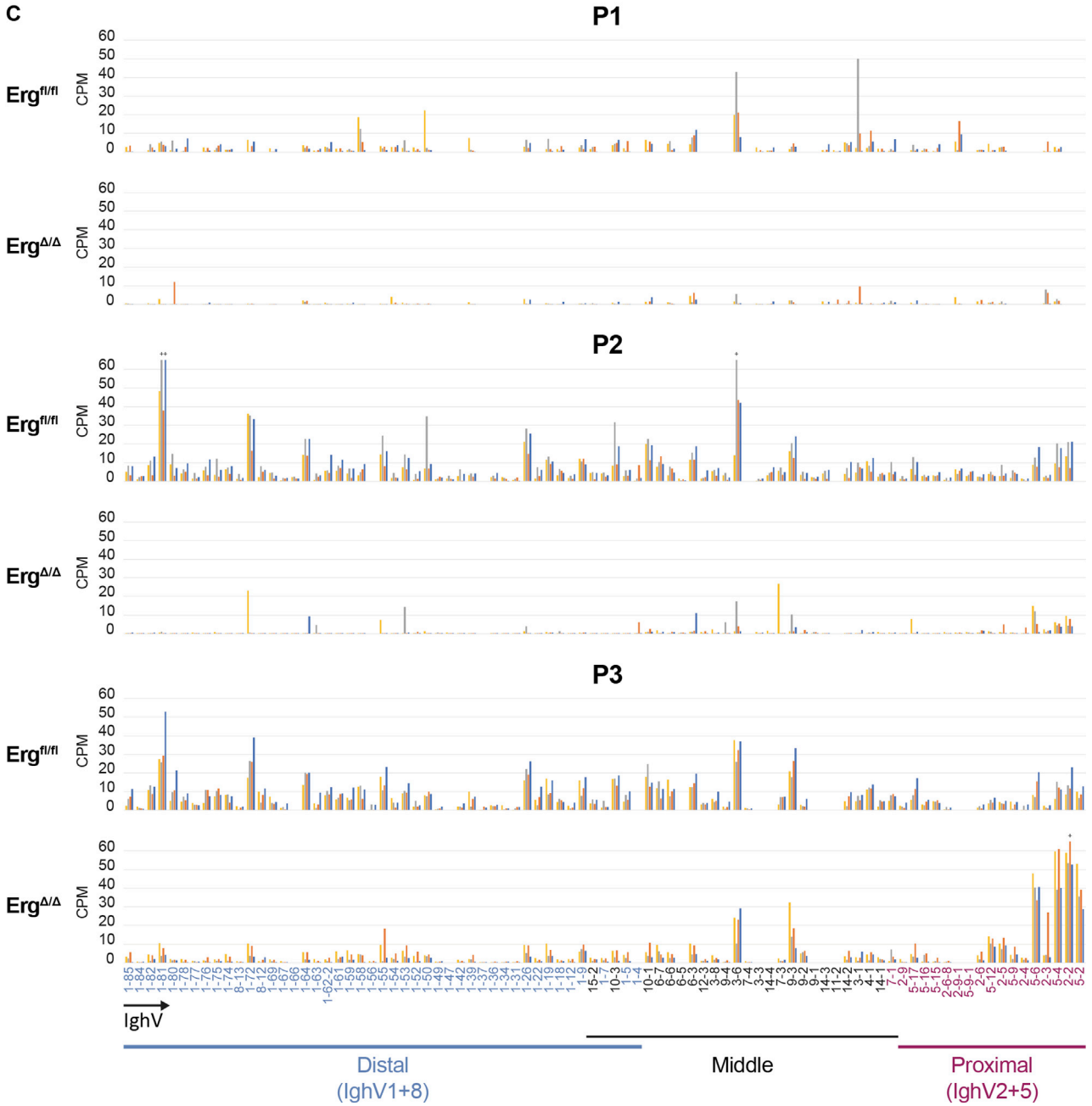
A
Igh locus



B V gene segment reads



C



(legend on next page)

recombination deficiency observed in *Erg*^{ΔΔ} progenitor B cells is caused by the inefficient use of distal V gene segments.

Efficient V-to-DJ recombination involving distal V gene segments is associated with major structural changes at the *Igh* locus to bring the 3' and the 5' ends into proximity in a process termed locus contraction (Kumari and Sen, 2015). Previous work has demonstrated that the structural proteins CTCF and YY1, along with the lineage-restricted transcription factor PAX5, play key roles in this process. Specifically, loss of either YY1 or PAX5 was associated with inefficient and biased V-to-DJ recombination, characterized by reduced locus contraction and loss of PAIR-associated antisense transcription, i.e., phenotypes that bear resemblance to what we observe in the *Erg*^{ΔΔ} mice (Ebert et al., 2011; Fuxa et al., 2004; Liu et al., 2007; Medvedovic et al., 2013; Verma-Gaur et al., 2012). However, in contrast to *Erg*^{ΔΔ} mice, PAX5-deficient animals lack CD25⁺ B cells and are unable to recombine to distal V gene segments, while *Mb1-Cre Yy1*^{fl/fl} mice lack CD25⁺ B cells and exhibit a bias against the use of distal V gene segments (Hesslein et al., 2003; Liu et al., 2007; Nutt et al., 1997). It is therefore conceivable that ERG collaborates with YY1 and/or PAX5 during the latter stages of *Igh* locus recombination. Locus contraction and efficient usage of distal V gene segments have been proposed to be associated with the establishment of so-called transcription factories. According to this model, PAX5-mediated PAIR-associated antisense transcription, along with transcription at the 3' end of the *Igh* locus, facilitates interactions between the distal V gene region and the E_μ enhancer region (Verma-Gaur et al., 2012). Although the exact order of events is not fully resolved, we note that ERG is able to bind to both the distal region of the *Igh* locus in hematopoietic progenitor cells and an E_μ enhancer element in cell-free experiments (Rivera et al., 1993; Wilson et al., 2010). These findings are consistent with a model in which ERG is involved in the establishment of higher-order structures at the *Igh* locus, either by promoting physical contacts through space or by driving transcriptional events, e.g., PAIR-associated antisense transcription, connected with the formation of these structures.

In conclusion, we have identified ERG as an essential player in B cell development, with particular importance for proper and efficient rearrangement of the *Igh* locus.

STAR★METHODS

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

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS

- Animals
- METHOD DETAILS
 - Genotyping
 - Mouse experiments
 - Flow cytometry
 - Gene expression analysis
 - ATAC-seq analyses
 - ImmunoSEQ
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Gene expression analysis
 - ATAC-seq analyses
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2019.10.098>.

ACKNOWLEDGMENTS

We thank Anna Fossum for advice and practical help with flow cytometry. Sequencing of ATAC-seq libraries was carried out at the Villum Center for Bio-analytical Sciences, Functional Genomics and Metabolism Research Unit, University of Southern Denmark. We thank Tenna P. Mortensen and Ronni Nielsen for sequencing assistance. In addition, we thank Erwin Schoof for conducting experiments that were not included in the paper and Bertie Göttgens for providing us with an ERG chromatin immunoprecipitation sequencing (ChIP-seq) bed file of mapped reads. This study was supported by the Lundbeck Foundation, the Independent Research Fund Denmark, and through a center grant from the Novo Nordisk Foundation (Novo Nordisk Foundation Center for Stem Cell Biology, DanStem; grant NNF17CC0027852). Work in the Mårtensson lab was supported by grants from the Swedish Cancer Fund (CAN 2016/481) and the Swedish Childhood Cancer Fund (PR2018-0170), whereas work in the Mandrup group was supported by the Novo Nordisk Foundation and the Independent Research Fund Denmark.

AUTHOR CONTRIBUTIONS

E.S., M.R., M.S.H., J.J., K.J.K., A.S.W., and A.R. carried out the experiments. E.S. carried out analysis of flow cytometry and immunoSEQ data. M.M. and E.S. performed analysis of the RNA-seq data. N.R. processed the microarray data. A.R. carried out the bioinformatical analysis of the ATAC-seq and microarray data. A.C. and F.O.B. assisted in data analysis. E.S. and B.T.P. drafted the manuscript. S.M., I.-L.M., and B.T.P. directed the research.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 7, 2019
Revised: August 28, 2019
Accepted: October 24, 2019
Published: November 26, 2019

Figure 5. ERG Is Required for Proper Expression of *IghV* Gene Segments

(A) Schematic figure of the immunoglobulin heavy-chain (*Igh*) locus.

(B) Accumulated expression values for *IghV* gene segments as assessed by RNA-seq. Cells from four *Erg*^{fl/fl} and *Erg*^{ΔΔ} mice were sequenced, and expression values are shown as CPM (counts per million). The data are shown as mean + SD. Statistical analyses were performed using unpaired, two-tailed Student's t test. **p ≤ 0.01, ***p ≤ 0.001.

(C) Expression values, shown as CPM, for individual *IghV* gene segments as assessed by RNA-seq. Expression values for four individual mice from each genotype are shown. + designates that the height of the indicated bar exceeds 65 CPM. The *IghV* gene segments are depicted on the x axis, with the most distal V gene segments (blue) to the left and the most proximal V gene segments (purple) to the right.

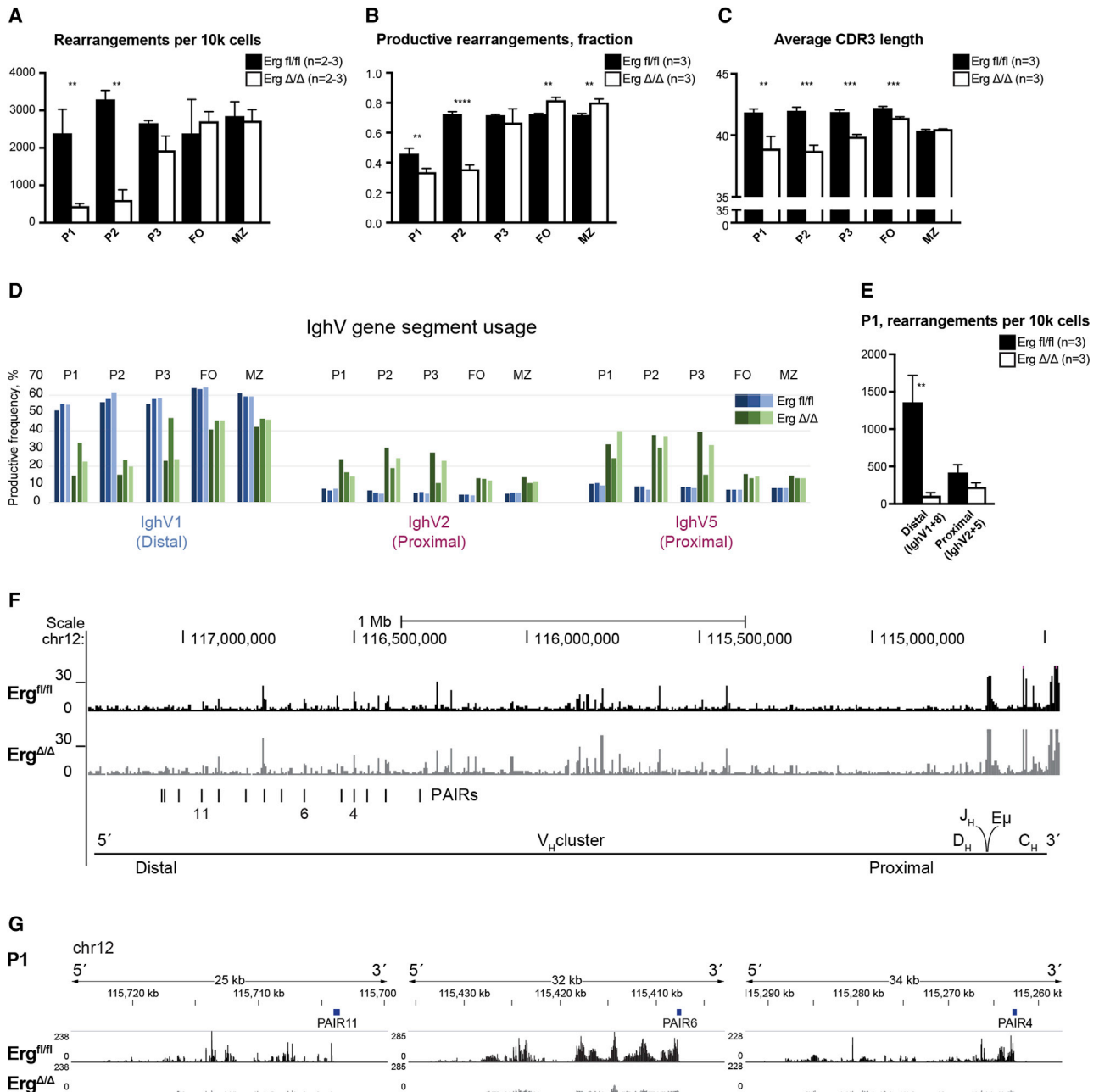


Figure 6. Inefficient and Irregular V-to-DJ Recombination in the Absence of ERG

(A) ImmunoSEQ analysis of P1–P3 progenitor cells and splenic FO and MZ B cells from *Erg^{fl/fl}* and *Erg^{Δ/Δ}* mice. The bar plot shows the quantification of the total number of detected rearrangements (templates) within each population as calculated per 10,000 sorted cells. For all cell populations, three mice of each genotype were analyzed, but only samples with an equal number of sorted cells were compared in the quantification of total rearrangements. Thus, only two replicates were included for *Erg^{fl/fl}* P2, *Erg^{fl/fl}* P3, and *Erg^{Δ/Δ}* P3 in this plot.

(B) Fraction of the rearrangements (templates) that are productively (in-frame, no-stop codon) recombined in the P1–P3, FO, and MZ populations as assessed by immunoSEQ.

(C) Average length of the CDR3 region in the P1–P3, FO, and MZ populations as determined using immunoSEQ.

(D) Results from the immunoSEQ analysis showing the usage of distal (blue) and proximal (purple) IghV gene families. The V gene segment usage is depicted as the frequency of the productive rearrangements within each sample. Bars of the same color represent cells from the same mouse for the P1–P3 populations and for the FO and MZ populations.

(E) ImmunoSEQ analysis of P1 progenitor cells showing the quantification of the total number of detected rearrangements (templates) involving distal (IghV1 and IghV8) and proximal (IghV2 and IghV5) gene segments as calculated per 10,000 sorted cells.

(legend continued on next page)

REFERENCES

- Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* 9, 9354.
- Bagger, F.O., Sasivarevic, D., Sohi, S.H., Laursen, L.G., Pundhir, S., Sønderby, C.K., Winther, O., Rapin, N., and Porse, B.T. (2016). BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. *Nucleic Acids Res.* 44 (D1), D917–D924.
- Baldus, C.D., Burmeister, T., Martus, P., Schwartz, S., Gökbuget, N., Bloomfield, C.D., Hoelzer, D., Thiel, E., and Hofmann, W.K. (2006). High expression of the ETS transcription factor ERG predicts adverse outcome in acute T-lymphoblastic leukemia in adults. *J. Clin. Oncol.* 24, 4714–4720.
- Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* 109, 21.29.1–21.29.9.
- Carey, J.B., Moffatt-Blue, C.S., Watson, L.C., Gavin, A.L., and Feeney, A.J. (2008). Repertoire-based selection into the marginal zone compartment during B cell development. *J. Exp. Med.* 205, 2043–2052.
- Corcoran, A.E., Smart, F.M., Cowling, R.J., Crompton, T., Owen, M.J., and Venkitaraman, A.R. (1996). The interleukin-7 receptor alpha chain transmits distinct signals for proliferation and differentiation during B lymphopoiesis. *EMBO J.* 15, 1924–1932.
- de Boer, J., Williams, A., Skavdis, G., Harker, N., Coles, M., Tolaini, M., Norton, T., Williams, K., Roderick, K., Potocnik, A.J., and Kioussis, D. (2003). Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur. J. Immunol.* 33, 314–325.
- Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., et al. (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc. Natl. Acad. Sci. USA* 108, 9566–9571.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Ebert, A., McManus, S., Tagoh, H., Medvedovic, J., Salvagiotto, G., Novatchkova, M., Tamir, I., Sommer, A., Jaritz, M., and Busslinger, M. (2011). The distal V(H) gene cluster of the Igh locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. *Immunity* 34, 175–187.
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048.
- Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E., and Busslinger, M. (2004). Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev.* 18, 411–422.
- Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20, 307–315.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80.
- Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., Oltz, E.M., and Sen, R. (2011). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* 147, 332–343.
- 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.
- Hess, J., Werner, A., Wirth, T., Melchers, F., Jäck, H.M., and Winkler, T.H. (2001). Induction of pre-B cell proliferation after *de novo* synthesis of the pre-B cell receptor. *Proc. Natl. Acad. Sci. USA* 98, 1745–1750.
- Hesslein, D.G., Pflugh, D.L., Chowdhury, D., Bothwell, A.L., Sen, R., and Schatz, D.G. (2003). Pax5 is required for recombination of transcribed, acetylated, 5' Igh V gene segments. *Genes Dev.* 17, 37–42.
- Jhunjunwala, S., van Zelm, M.C., Peak, M.M., Cutchin, S., Riblet, R., van Dongen, J.J., Grosveld, F.G., Knoch, T.A., and Murre, C. (2008). The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* 133, 265–279.
- Johnston, C.M., Wood, A.L., Bolland, D.J., and Corcoran, A.E. (2006). Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. *J. Immunol.* 176, 4221–4234.
- Kitamura, D., Roes, J., Kühn, R., and Rajewsky, K. (1991). A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin mu chain gene. *Nature* 350, 423–426.
- Knudsen, K.J., Rehn, M., Hasemann, M.S., Rapin, N., Bagger, F.O., Ohlsson, E., Willer, A., Frank, A.K., Søndergaard, E., Jendholm, J., et al. (2015). ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. *Genes Dev.* 29, 1915–1929.
- Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., and Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296, 158–162.
- Kumari, G., and Sen, R. (2015). Chromatin Interactions in the Control of Immunoglobulin Heavy Chain Gene Assembly. *Adv. Immunol.* 128, 41–92.
- Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring Reproducibility of High-Throughput Experiments. *Ann. Appl. Stat.* 5, 1752–1779.
- Lilljebjörn, H., Henningsson, R., Hyrenius-Wittsten, A., Olsson, L., Orsmark-Pietras, C., von Palffy, S., Askmyr, M., Rissler, M., Schrappe, M., Cario, G., et al. (2016). Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. *Nat. Commun.* 7, 11790.
- Lin, Y.C., Jhunjunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowski, J., et al. (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* 11, 635–643.
- Liu, H., Schmidt-Suppran, M., Shi, Y., Hobeika, E., Barteneva, N., Jumaa, H., Pelanda, R., Reth, M., Skok, J., Rajewsky, K., and Shi, Y. (2007). Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev.* 21, 1179–1189.
- Loughran, S.J., Kruse, E.A., Hacking, D.F., de Graaf, C.A., Hyland, C.D., Willson, T.A., Henley, K.J., Ellis, S., Voss, A.K., Metcalf, D., et al. (2008). The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat. Immunol.* 9, 810–819.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Lukin, K., Fields, S., Lopez, D., Cherrier, M., Ternyak, K., Ramirez, J., Feeney, A.J., and Hagman, J. (2010). Compound haploinsufficiencies of Ebf1 and

(F) ATAC-seq tracks visualized for mm9 in the University of California Santa Cruz Genome Browser (<http://genome.ucsc.edu>) for the *Igh* locus in *Erg^{fl/fl}* and *Erg^{Δ/Δ}* progenitor B cells. Cells from four mice of each genotype were analyzed. The cells underwent fluorescence-activated cell sorting (FACS) using the following marker combination: Lin-B220+CD19+CD93+CD43+^{fl/fl}.

(G) RNA-seq tracks visualized for mm10 in the Integrative Genomics Viewer (IGV) browser (<https://software.broadinstitute.org/software/igv/>) showing transcription (raw RNA-seq reads) immediately next to PAIR11 (left), PAIR6 (middle), and PAIR4 (right) in *Erg^{fl/fl}* and *Erg^{Δ/Δ}* P1 cells. Four replicates were made for each genotype. There was no significant difference in the total number of mapped reads between the *Erg^{fl/fl}* and *Erg^{Δ/Δ}* P1 samples (data not shown).

For the bar plots in (A)–(C), the data are shown as mean + SD. Statistical analyses were performed using unpaired, two-tailed Student's t test. **p* ≤ 0.05, ***p* ≤ 0.01, ****p* ≤ 0.001, *****p* ≤ 0.0001. For tracks showing ATAC-seq and RNA-seq data, representative samples are shown. *Erg^{fl/fl}* and *Erg^{Δ/Δ}* tracks are black and gray, respectively. See also Figure S4.

- Runx1 genes impede B cell lineage progression. *Proc. Natl. Acad. Sci. USA* **107**, 7869–7874.
- Madsen, J.G.S., Rauch, A., Van Hauwaert, E.L., Schmidt, S.F., Winnefeld, M., and Mandrup, S. (2018). Integrated analysis of motif activity and gene expression changes of transcription factors. *Genome Res.* **28**, 243–255.
- Martens, J.H. (2011). Acute myeloid leukemia: a central role for the ETS factor ERG. *Int. J. Biochem. Cell Biol.* **43**, 1413–1416.
- Medvedovic, J., Ebert, A., Tagoh, H., Tamir, I.M., Schwickert, T.A., Novatchkova, M., Sun, Q., Huis In 't Veld, P.J., Guo, C., Yoon, H.S., et al. (2013). Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. *Immunity* **39**, 229–244.
- Nutt, S.L., Urbánek, P., Rolink, A., and Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the Igh locus. *Genes Dev.* **11**, 476–491.
- Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**, 556–562.
- Parker, M.J., Licence, S., Erlandsson, L., Galler, G.R., Chakalova, L., Osborne, C.S., Morgan, G., Fraser, P., Jumaa, H., Winkler, T.H., et al. (2005). The pre-B-cell receptor induces silencing of VpreB and lambda5 transcription. *EMBO J.* **24**, 3895–3905.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419.
- Pelanda, R., Braun, U., Hobeika, E., Nussenzweig, M.C., and Reth, M. (2002). B cell progenitors are arrested in maturation but have intact VDJ recombination in the absence of Ig-alpha and Ig-beta. *J. Immunol.* **169**, 865–872.
- Perlot, T., and Alt, F.W. (2008). Cis-regulatory elements and epigenetic changes control genomic rearrangements of the Igh locus. *Adv. Immunol.* **99**, 1–32.
- Peschon, J.J., Morrissey, P.J., Grabstein, K.H., Ramsdell, F.J., Maraskovsky, E., Gliniak, B.C., Park, L.S., Ziegler, S.F., Williams, D.E., Ware, C.B., et al. (1994). Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. *J. Exp. Med.* **180**, 1955–1960.
- Prohaska, T.A., Que, X., Diehl, C.J., Hendriks, S., Chang, M.W., Jepsen, K., Glass, C.K., Benner, C., and Witzum, J.L. (2018). Massively Parallel Sequencing of Peritoneal and Splenic B Cell Repertoires Highlights Unique Properties of B-1 Cell Antibodies. *J. Immunol.* **200**, 1702–1717.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
- Rainis, L., Toki, T., Pimanda, J.E., Rosenthal, E., Machol, K., Strehl, S., Göttgens, B., Ito, E., and Izraeli, S. (2005). The proto-oncogene ERG in megakaryoblastic leukemias. *Cancer Res.* **65**, 7596–7602.
- Rauch, A., Haakonsson, A.K., Madsen, J.G.S., Larsen, M., Forss, I., Madsen, M.R., Van Hauwaert, E.L., Wiwie, C., Jespersen, N.Z., Tencerova, M., et al. (2019). Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as repressors of adipogenesis. *Nat. Genet.* **51**, 716–727.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47.
- Rivera, R.R., Stuiver, M.H., Steenbergen, R., and Murre, C. (1993). Ets proteins: new factors that regulate immunoglobulin heavy-chain gene expression. *Mol. Cell. Biol.* **13**, 7163–7169.
- Rolink, A., Grawunder, U., Winkler, T.H., Karasuyama, H., and Melchers, F. (1994). IL-2 receptor alpha chain (CD25, TAC) expression defines a crucial stage in pre-B cell development. *Int. Immunol.* **6**, 1257–1264.
- Schelonka, R.L., Ivanov, I.I., Vale, A.M., Dimmitt, R.A., Khaled, M., and Schroeder, H.W., Jr. (2011). Absence of N addition facilitates B cell development, but impairs immune responses. *Immunogenetics* **63**, 599–609.
- Schindler, J.W., Van Buren, D., Foudi, A., Krejci, O., Qin, J., Orkin, S.H., and Hock, H. (2009). TEL-AML1 corrupts hematopoietic stem cells to persist in the bone marrow and initiate leukemia. *Cell Stem Cell* **5**, 43–53.
- Sigvardsson, M. (2018). Molecular Regulation of Differentiation in Early B-Lymphocyte Development. *Int. J. Mol. Sci.* **19**, E1928.
- Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550.
- Tsuzuki, S., Taguchi, O., and Seto, M. (2011). Promotion and maintenance of leukemia by ERG. *Blood* **117**, 3858–3868.
- Vallespinós, M., Fernández, D., Rodríguez, L., Alvaro-Blanco, J., Baena, E., Ortiz, M., Dukovska, D., Martínez, D., Rojas, A., Campanero, M.R., and Moreno de Alborán, I. (2011). B Lymphocyte commitment program is driven by the proto-oncogene c-Myc. *J. Immunol.* **186**, 6726–6736.
- van Riggelen, J., Yetil, A., and Felsher, D.W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat. Rev. Cancer* **10**, 301–309.
- Verma-Gaur, J., Torkamani, A., Schaffer, L., Head, S.R., Schork, N.J., and Feeney, A.J. (2012). Noncoding transcription within the Igh distal V(H) region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells. *Proc. Natl. Acad. Sci. USA* **109**, 17004–17009.
- Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E., et al. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532–544.
- Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.
- Zhang, J., McCastlain, K., Yoshihara, H., Xu, B., Chang, Y., Churchman, M.L., Wu, G., Li, Y., Wei, L., Iacobucci, I., et al.; St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome Project (2016). Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat. Genet.* **48**, 1481–1489.
- Zou, X., Piper, T.A., Smith, J.A., Allen, N.D., Xian, J., and Brüggemann, M. (2003). Block in development at the pre-B-II to immature B cell stage in mice without Ig kappa and Ig lambda light chain. *J. Immunol.* **170**, 1354–1361.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
B220 APC Anti-Human/Mouse CD45R (B220) APC	Thermo Fisher Scientific	Cat# 17-0452-82; RRID:AB_469395
B220 A700 Anti-Human/Mouse CD45R (B220) Alexa Fluor 700	Thermo Fisher Scientific	Cat# 56-0452-82; RRID:AB_891458
B220 APC eFluor 780 Anti-Human/Mouse CD45R (B220) APC-eFluor 780	Thermo Fisher Scientific	Cat# 47-0452-82; RRID:AB_1518810
B220 PE-Cy5 Anti-Human/Mouse CD45R (B220) PE-Cy5	Thermo Fisher Scientific	Cat# 15-0452-83; RRID:AB_468756
B220 PerCP Anti-Mouse CD45R (B220) PerCP	BD Biosciences	Cat# 553093; RRID:AB_394622
CD3e APC Anti-Mouse CD3e APC	Thermo Fisher Scientific	Cat# 17-0031-83; RRID:AB_469316)
CD3e PECy5 Anti-Mouse CD3e PE-Cy5	Thermo Fisher Scientific	Cat# 15-0031-82; RRID:AB_468690
CD4 FITC Anti-Mouse CD4 FITC	Thermo Fisher Scientific	Cat# 11-0041-82; RRID:AB_464892
CD4 PECy5 Anti-Mouse CD4 PE-Cy5	Thermo Fisher Scientific	Cat# 15-0041-81; RRID:AB_468694
CD5 BV421 Anti-Mouse CD5 BV421	BD Biosciences	Cat# 562739; RRID:AB_2737758
CD8a PECy5 Anti-Mouse CD8a PE-Cy5	Thermo Fisher Scientific	Cat# 15-0081-81; RRID:AB_468705
CD16/CD32 purified Anti-Mouse CD16/CD32 (Mouse BD Fc Block) Purified	BD Biosciences	Cat# 553142; RRID:AB_394657
CD19 APC Anti-Mouse CD19 APC	BD Biosciences	Cat# 550992; RRID:AB_398483
CD19 eFluor 450 Anti-Mouse CD19 eFluor 450	Thermo Fisher Scientific	Cat# 48-0193-82; RRID:AB_2734905
CD19 eFluor 605 Anti-Mouse CD19 eFluor 605NC	Thermo Fisher Scientific	Cat# 93-0193-41; RRID:AB_1603212
CD21 APC anti-Mouse CD21/CD35 APC	BD Biosciences	Cat# 561770; RRID:AB_10892818
CD23 PE Anti-Mouse CD23 PE	BD Biosciences	Cat# 561773; RRID:AB_10895122
CD25 PE Anti-Mouse CD25 PE	BD Biosciences	Cat# 553866; RRID:AB_395101
CD43 APC Anti-Mouse CD43 APC	BD Biosciences	Cat# 560663; RRID:AB_1727479
CD43 biotin Anti-Mouse CD43 Biotin	BD Biosciences	Cat# 553269; RRID:AB_2255226
CD44 APC eFluor 780 Anti-Human/Mouse CD44 APC-eFluor 780	Thermo Fisher Scientific	47-0441-82; RRID:AB_1272244

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CD93 FITC	Thermo Fisher Scientific	Cat# 11-5892-82; RRID:AB_465298
Anti-Mouse CD93 (AA4-1) FITC		
CD93 PE-Cy7	Thermo Fisher Scientific	Cat# 25-5892-81; RRID:AB_469658
Anti-Mouse CD93 (AA4.1) PE-Cy7		
CD135 PE	Thermo Fisher Scientific	Cat# 12-1351-83; RRID:AB_465860
Anti-Mouse CD135 (Flt3) PE		
c-kit Alexa APC 780	Thermo Fisher Scientific	Cat# 47-1171-82; RRID:AB_1272177
Anti-Mouse CD117 (c-Kit) APC-eFluor 780		
c-kit APC	Thermo Fisher Scientific	Cat# 17-1171-82; RRID:AB_469430
Anti-Mouse CD117 (c-Kit) APC		
Gr1 PE-Cy5	Thermo Fisher Scientific	Cat# 15-5931-82; RRID:AB_468813
Anti-Mouse Ly-6G (Gr-1) PE-Cy5		
IgM FITC (μ chain-specific)	Sigma-Aldrich	Cat# F9259; RRID:AB_259799
Anti-Mouse IgM (mu-chain specific) FITC		
IL7R biotin	Thermo Fisher Scientific	Cat# 13-1271-82; RRID:AB_466588
Anti-Mouse CD127 Biotin		
Mac-1 APC	Thermo Fisher Scientific	17-0112-82; RRID:AB_469343
Anti-Mouse CD11b APC		
Mac-1 PE-Cy5	BioLegend	Cat# 101210, RRID:AB_312793
Anti-mouse/human CD11b PE-Cy5		
Nk1.1 PE-Cy7	Thermo Fisher Scientific	Cat# 25-5941-82, RRID:AB_469665
Anti-Mouse NK1.1 PE-Cy7		
SA-PECF594	BD Biosciences	Cat# 562318, RRID:AB_11154218
Streptavidin PE-CF594		
SA-QD655	Life Technologies	CAT#Q10121MP, RRID: N/A
Sca1 PB	BioLegend	Cat# 108120, RRID:AB_493273
Anti-mouse Ly-6A/E (Sca-1) Pacific blue		
Ter119 PE-Cy5	eBioscience	15-5921-81, RRID:AB_468809
Anti-Mouse TER-119 PE-Cy5		
Chemicals, Peptides, and Recombinant Proteins		
Fetal bovine serum	GE healthcare	CAT#SV30160.03
Pharmlyse	BD Biosciences	CAT#555899
Paraformaldehyde	Sigma-Aldrich	CAT#P6148
Saponin	Fluka	CAT#47036
DNase I	Sigma-Aldrich	CAT#D4527
DAPI	Invitrogen	CAT#D3571
2-mercaptoethanol	Sigma-Aldrich	CAT#M3148
Critical Commercial Assays		
FITC BrdU flow kit	BD Biosciences	CAT# 559619, RRID:AB_2617060
AllPrep DNA/RNA micro kit	QIAGEN	CAT#80284
Ovation RNA-seq System V2	NuGEN	CAT# 7102-32
Ovation Ultralow System V2	NuGEN	CAT#0344-32
RNeasy micro kit	QIAGEN	CAT#74004
Nextera DNA Library Preparation Kit	Illumina	CAT#FC-121-1030
Nextera Index Kit	Illumina	CAT#FC-121-1011
ImmunoSEQ, mmIGH, survey resolution	Adaptive biotechnologies	N/A
Affymetrix's Mouse Gene 1.0 ST GeneChip array	Affymetrix	N/A

(Continued on next page)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
RNA-seq	This paper	GEO: GSE128978
ATAC-seq	This paper	GEO: GSE128978
Microarray	This paper	GEO: GSE128978
ImmunoSEQ	This paper	immuneACCESS database (https://clients.adaptivebiotech.com/login) DOI: 10.21417/ES2019CR
Mouse reference genome mm9/NCBI37	Genome Reference Consortium	N/A
Mouse reference genome mm10/GRCm38	Genome Reference Consortium	N/A
Experimental Models: Organisms/Strains		
B6- <i>Erg</i> ^{fllox/fllox} ; <i>CD2iCre</i>	Knudsen et al., 2015	N/A
Oligonucleotides		
Primers for <i>Erg</i> genotyping PCR: ATCATGACAATA AGCCGGGT; AACCAGAACGGTGGTAGTCT; GTG GTGGCTCCTTAAGGGTC	Knudsen et al., 2015	N/A
Primers for <i>CD2iCre</i> genotyping PCR: GACAGGCAG GCCTTCTCTGAA; CTTCTCCACACCAGCTGTGGA	Knudsen et al., 2015	N/A
Software and Algorithms		
FlowJo	BD	https://www.flowjo.com/solutions/flowjo RRID:SCR_008520
Prism	GraphPad	https://www.graphpad.com/ RRID:SCR_002798
UCSC Genome Browser	University of California at Santa Cruz; California; USA	http://genome.ucsc.edu RRID:SCR_005780
Integrative Genomics Viewer (IGV)	Broad Institute	http://software.broadinstitute.org/software/igv/ RRID:SCR_011793
bcbio-nextgen		
STAR aligner	Dobin et al., 2013	https://github.com/alexdobin/STAR
Salmon	Patro et al., 2017	https://combine-lab.github.io/salmon/
FastQC		http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
multiqc	Ewels et al., 2016	https://multiqc.info/
R/Bioconductor	Gentleman et al., 2004	https://www.R-project.org/
Differential gene expression recommendations	Soneson et al., 2015	
limma package	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html
GSEA	Subramanian et al., 2005	http://software.broadinstitute.org/gsea/index.jsp
MSigDB		http://software.broadinstitute.org/gsea/msigdb
Bioconductor (Affy package, RMA)	Gautier et al., 2004	http://bioconductor.org/packages/release/bioc/html/affy.html
ENCODE IDR pipeline	Li et al., 2011	https://github.com/nboley/idr
MACS2	Zhang et al., 2008	https://github.com/taoliu/MACS
HOMER	Heinz et al., 2010	http://homer.ucsd.edu/homer/
DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
goseq	Young et al., 2010	https://bioconductor.org/packages/release/bioc/html/goseq.html
IMAGE	Madsen et al., 2018	https://www.sdu.dk/en/om_sdu/institutter_centre/bmb_biokemi_og_molekylaer_biologi/forskning/forskningsgrupper/functionalgenomics/bioinformatics+tools
Bedtools	Quinlan and Hall, 2010	http://bedtools.readthedocs.io/en/latest/
Bloodspot	Bagger et al., 2016	http://servers.binf.ku.dk/bloodspot/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Bo T. Porse (bo.porse@finsenlab.dk). This study did not generate any new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

Erg^{fl/fl} and *Erg^{fl/fl};CD2iCre (Erg^{Δ/Δ})* mice were bred as described in Knudsen et al. (2015). In short, mice, in which the ETS domain-containing exon of *Erg* is flanked by LoxP sites, were crossed to mice harboring *CD2iCre*, which drives a lymphoid-restricted expression of *iCre*. The mice were backcrossed onto a C57BL/6J background for at least 6 generations. The mice were bred and housed in individually ventilated cages (IVC) at the Department of Experimental Medicine at the University of Copenhagen according to institutional guidelines. Experiments were carried out with permissions from the Danish Animal Research Ethical Committee. 10-14 week old female mice were used for all analyses unless specifically stated.

METHOD DETAILS

Genotyping

A three-primer system (ATCATGACAATAAGCCGGGT; AACGAGAACGGTGGTAGTCT; GTGGTGGCTCCTTAAGGGTC) was used for PCR genotyping of *Erg* resulting in bands of the following sizes: 397bp (*Erg^{fl}*), 302bp (*Erg^Δ*) and 340bp (*Erg⁺*). *CD2iCre* genotyping was done using two primers (GACAGGCAGGCTTCTCTGAA; CTTCTCCACACCAGCTGTGGA) giving rise to a 522bp PCR product.

Mouse experiments

For BrdU incorporation analysis, mice were subjected to intraperitoneal injection of 200μl PBS containing 2mg BrdU (Sigma-Aldrich, B5002), and BM was collected three h later.

Flow cytometry

Bones were harvested by dissection of the hindlegs (2x tibia, femur, ilium), and crushed in PBS+3% fetal bovine serum (FBS, GE healthcare, Little Chalfont, UK). Thymus and spleen were dissected and the cells were homogenized in PBS+3% FBS and filtered through a 70μm cell strainer.

Cells in single cell suspension were blocked with CD16/CD32 and incubated with antibodies for 20-25min. The cells were subsequently stained with a secondary antibody when applicable. For FACS, cells were treated with PharmLyse (BD, Franklin Lakes, New Jersey, USA) for 5min to reduce the amount of red blood cells.

For intracellular staining, cells (stained for cell-surface antigens) were fixed for 5min in PBS+1% paraformaldehyde (PFA) (Sigma-Aldrich, St. Louis, Missouri, USA) and permeabilized with PBS+0.1% saponin (Fluka) at RT for 10min. The cells were blocked with CD16/32 before antibody incubation. For the analysis of BrdU incorporation, cells (stained for cell-surface antigens) were fixed, permeabilized, treated with DNase I (D4527, Sigma-Aldrich) and stained with anti-BrdU antibody using a FITC BrdU Flow kit (BD) according to manufacturer's instructions.

Finally, for cell cycle analyses, cells (stained for cell-surface antigens) were fixed for 10min in PBS+2% PFA at room temperature (RT) and permeabilized with PBS+0.1% saponin at RT for 45min before being stained for 30min with DAPI (0.5μg/mL) (Invitrogen, Carlsbad, California, USA).

Flow cytometry analyses were carried out using a BD Accuri C6 or LSR II instrument, and cells were sorted on BD Aria I or III instruments. FlowJo software (BD) was used for subsequent analyses.

Cells were analyzed and sorted according to the following surface markers:

BM: P1 (B220+, CD19+, CD93+, CD43+, CD25-), P2 (B220+, CD19+, CD93+, CD43+, CD25+), P3 (B220+, CD19+, CD93+, CD43-, CD25+), c-Kit, CD43+ (B220+, CD19+, CD93+, CD43+, c-Kit+), immature B cells (B220+, CD19+, CD93+, IgM+), mature B cells (B220+, CD19+, CD93-, IgM+).

Spleen: T1 B cells (CD19+, CD43-, CD93+, IgM+, CD23-), T2 B cells (CD19+, CD43-, CD93+, CD23+), B1 B cells (CD19+, CD43+), marginal zone B cells (CD19+, CD43-, CD93-, CD21^{hi}, CD23-), follicular B cells (CD19+, CD43-, CD93-, CD21^{int}, CD23+), B1a B cells (B220+, CD19+, CD43+, IgM+, CD5+), B1b B cells (B220+, CD19+, CD43+, IgM+, CD5-), CD4+ T cells (CD3e+, CD4+), CD8+ T cells (CD3e+, CD8a+), all B cells (B220+).

Thymus: B cells (B220+), CD4+ T cells (CD4+, CD8a-), CD8+ T cells (CD4-, CD8a+), DP (CD4+, CD8a+), DN (CD4-, CD8a-), DN1 (CD4-, CD8a-, CD44+, CD25-), DN2 (CD4-, CD8a-, CD44+, CD25+), DN3 (CD4-, CD8a-, CD44-, CD25+), DN4 (CD4-, CD8a-, CD44-, CD25-). DP and DN indicate double positive and double negative, respectively.

Peripheral blood: B cells (CD19+), CD4+ T cells (CD4+), CD8+ T cells (CD8a+), myeloid cells (Mac-1+), NK cells (Nk1.1+).

Lymph nodes: B cells (B220+).

For ATAC-seq and gene expression microarrays, progenitor B cells (Lin⁻, B220⁺, CD19⁺, CD93⁺ CD43^{+/-}) were sorted. For *Erg* PCR recombination analysis, the following cells were sorted: MPPs (Lin⁻, Sca1⁺, c-Kit⁺, CD150⁻), CLPs (Lin⁻, Sca1^{lo}, c-Kit^{lo}, Flt3L⁺, IL7R α ⁺), prepro-B (Lin⁻, B220⁺, CD19⁻, CD93⁺ CD43^{+/-}), pro-B (Lin⁻, B220⁺, CD19⁺, CD93⁺ CD43^{+/-}), pre-B+B (Lin⁻, B220⁺, CD19⁺, CD43⁻), splenic B cells (B220⁺).

Gene expression analysis

RNA-seq-based

Progenitor B cells were sorted into RLT plus buffer (QIAGEN, Hilden, Germany) with 2-mercaptoethanol (Sigma-Aldrich, 10 μ l/mL RLT buffer). The RNA was extracted with QIAGEN's AllPrep micro kit. cDNA was obtained using the Ovation RNA-seq System V2 (Nugen, Redwood City, California, USA), and libraries were prepared with the Ovation Ultralow System V2 (Nugen), all according to manufacturer's instructions. The samples were sequenced on an Illumina NextSeq 500 instrument (San Diego, California, USA). RNA-seq reads were aligned to the mm10 genome using STAR (Dobin et al., 2013) and the RNA-seq tracks were visualized in the IGV browser (<http://software.broadinstitute.org/software/igv/>).

Microarray-based

Progenitor B cells were sorted into RLT plus buffer with 2-mercaptoethanol (10 μ l/ml RLT buffer). Total RNA was extracted using the RNeasy micro kit (QIAGEN), and libraries were obtained using the Ovation Pico WTA system V2 (NuGEN) following manufacturer's instructions. The libraries were hybridized to the Mouse Gene 1.0 ST GeneChip array (Affymetrix, Santa Clara, California, USA).

ERG gene expression

ERG expression levels during the early stages of B- and T cell development were assessed via publicly available gene expression repositories using the Bloodspot database (Bagger et al., 2016).

ATAC-seq analyses

Genome-wide analyses of chromatin accessibility

ATAC-seq was performed on 10,000 sorted progenitor B cells as previously described (Buenrostro et al., 2015). Briefly, sorted cells were permeabilized and exposed to 0.5 μ l transposase. Library amplification was done using the Nextera DNA library Prep kit (Illumina) according to the instructions by the manufacturer, and sequencing was performed on a HiSeq 1500 Illumina machine. Sequencing reads were mapped to the mouse genome (mm9) using STAR (Dobin et al., 2013). The ATAC-seq data were visualized using the UCSC browser (<http://genome.ucsc.edu>).

ImmunoSEQ

Sorted progenitors and mature B cells were subjected to immunoSEQ assessing the VDJ junctions of the *Igh* locus. The immunoSEQ assay utilizes a bias-controlled multiplex PCR (using gene segment-specific primers) followed by next-generation sequencing in order to allow identification of the CDR3 sequences of the recombined antigen receptor loci.

60,000 mature B cells or 10,000-30,000 progenitor B cells were sorted, pelleted, and frozen at -80°C . The pellets were sent to the facilities of Adaptive Biotechnologies (Seattle, Washington, USA) according to their instructions for further processing. The cells were assessed for the murine *Igh* locus (gDNA) at survey resolution (used for samples with low cell numbers). The online immunoSEQ analyzer software was used for extracting data.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical details of the experiments (e.g., statistical tests used, value of *n*, definition of center, and dispersion and precision measures) are stated in the figures, the figure legends and/or under the relevant section of the STAR Methods. Generally, unless stated otherwise in the figure legends or STAR Methods, we have used unpaired, two-tailed Student's *t* test and visualized the results as barplots showing mean + SD, when two conditions were compared. *n* represents the number of biological replicates. Statistical significance is defined as *p* value ≤ 0.05 . * = *p* ≤ 0.05 , ** = *p* ≤ 0.01 , *** = *p* ≤ 0.001 , **** = *p* ≤ 0.0001 .

Gene expression analysis

RNA-seq-based

RNA-seq data were analyzed with bcbio-nextgen (<https://github.com/bcbio/bcbio-nextgen>). Fastq files were aligned to the mm10 genome using STAR (Dobin et al., 2013). Transcript expression levels were estimated with Salmon (Patro et al., 2017). FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used for QC metrics, and multiqc (Ewels et al., 2016) for reporting. Data analysis was then performed with R/Bioconductor (<https://www.R-project.org/>; Gentleman et al. (2004)). For the differential gene expression, we followed the recommendations from Sonesson et al. (2015). Gene expression levels were assessed by adding all the transcript levels for a given gene and normalization of the counts was done with the lengthScaledTPM function of the tximport package followed by a limma-voom procedure: for each population (P1-3), genes with CPM > 1 in at least 2 samples were selected, the TMM normalization was run using the calcNormFactors function of the limma package (Ritchie et al., 2015), and a linear model with *Erg* $^{\Delta\Delta}$ versus *Erg*^{fl/fl} was fitted (voom / lmFit / eBayes). *P* values were adjusted using the Benjamini-Hochberg procedure.

Gene set enrichment analysis was carried out using GSEA (<http://software.broadinstitute.org/gsea/index.jsp>) (Subramanian et al., 2005). Gene sets originated from the MSigDB (<http://software.broadinstitute.org/gsea/msigdb>).

Microarray-based

Microarray data were normalized using RMA (Bioconductor) (Gautier et al., 2004). Differential expression analysis was performed using the limma package (Ritchie et al., 2015) with uncorrected and corrected p values (modified smyth t test) being reported for each probeset.

ATAC-seq analyses

Sequencing reads were mapped to the mouse genome (mm9) using STAR (Dobin et al., 2013). The ATAC-seq data were visualized using the UCSC browser (<http://genome.ucsc.edu>). ATAC-seq sites were identified according to the ENCODE IDR pipeline (Li et al., 2011) using MACS2 (Zhang et al., 2008) peak calling from each replicate. IDR p value was calculated for pairwise combinations of replicates (6 combinations for 4 replicates a to d; a:b, a:c, a:d, b:c, b:d, c:d). For all downstream analyses, ATAC-seq sites scoring a p value < 0.05 in at least 4 out of 6 combinations were merged across both genotypes followed by filtering for the mouse ENCODE blacklist (Amemiya et al., 2019) using bedtools (Quinlan and Hall, 2010). Sequencing tags in peaks were quantified using HOMER (Heinz et al., 2010) allowing only one read per position per length and differential accessibility was determined using DEseq2 (Love et al., 2014). Dynamic ATAC-sites were defined as sites with a significant change in ATAC-seq signal (Benjamini-Hochberg correction FDR < 0.01) between *Erg^{fl/fl}* and *Erg^{Δ/Δ}* mice.

Gene ontology analysis of genes in the vicinity of dynamic ATAC-seq sites

Genes with transcription start site ± 50 kb from ATAC-seq sites with differential accessibility between *Erg^{fl/fl}* and *Erg^{Δ/Δ}* progenitor B cells were subjected to gene ontology analysis using goseq (Young et al., 2010). Biological processes reaching a p value < 0.01 after Benjamini-Hochberg correction were reported.

IMAGE based modeling of motif activity and network analysis

We used the IMAGE tool (Madsen et al., 2018) to compute the contribution of motifs first to chromatin accessibility as defined by the ATAC-seq signal in all ATAC-seq regions (n = 52311) and second to gene expression as defined by the microarray signal for each biological replicate (ATAC-seq n = 4 and microarray n = 3 per genotype). Dynamic motifs were defined by a significant change in motif activity (p value < 0.01) when comparing progenitor B cells from *Erg^{fl/fl}* and *Erg^{Δ/Δ}* mice. IMAGE output further contains predicted target regions and target genes for each motif which was used to build a transcriptional network. Network visualization was done with the R package igraph. To estimate the connectivity within the group of transcription factors that either lost or gained motif activity in *Erg^{Δ/Δ}* progenitor B cells compared to *Erg^{fl/fl}* cells, we determined how many group members were targeted by each individual transcription factor of the group. The values were normalized by group size and averaged over all group members. In order to control for a possible network bias, we performed 1000 randomizations in which edges of the network were shuffled while keeping the specific number of target transcription factors among all targets constant. For each permutation we determined the connectivity, and permutation results are represented as mean including standard deviation over 1000 permutations. A p value was generated by asking how often the shuffled network reached at least or exceeded the values of the real distribution.

Statistics and reproducibility

ATAC-seq and microarray data were generated from isolated progenitor B cells of independent *Erg^{fl/fl}* and *Erg^{Δ/Δ}* mice (not paired). All boxplots depict the first and third quartiles as the lower and upper bounds of the box, with a thicker band inside the box showing the median value and whiskers representing 1.5 × the interquartile range. FDR tests for differential chromatin accessibility and gene ontology enrichment were done using Benjamini-Hochberg procedure integrated in R packages DEseq2 (Love et al., 2014) and goseq (Young et al., 2010). P values derived from randomization tests indicate the number of trials yielding similar or higher values than the true distribution divided by 1000.

DATA AND CODE AVAILABILITY

The accession number for the RNA-seq data, microarray gene expression data, and ATAC-seq data <http://www.ncbi.nlm.nih.gov/gds> reported in this paper is GEO: GSE128978. ImmunoSEQ data are available in the immuneACCESS database (<https://clients.adaptivebiotech.com/login>), <https://doi.org/10.21417/ES2019CR>.