An Erg driven transcriptional program controls B-lymphopoiesis

Ashley P. Ng[1,8,*], Hannah D. Coughlan[2,8], Soroor Hediyeh-zadeh[2], Kira Behrens[1], Timothy M. Johanson[3,8], Michael Sze Yuan Low[3,5,8], Charles C. Bell[7,9], Omer Gilan[7,9], Yih-Chih Chan[7,9], Andrew J. Kueh[1,8], Thomas Boudier[4,8], Ladina DiRago[1], Craig D. Hyland[1], Helen Ierino[1], Sandra Mifsud[1], Elizabeth Viney[1], Tracy Willson[1], Mark A. Dawson[7,9,10], Rhys S. Allan[3,8], Marco J. Herold[1,8] , Kelly Rogers[4,8], David M Tarlinton[6], Gordon K. Smyth[2,8], Melissa J. Davis[2,8], Stephen L. Nutt[3,8] and Warren S. Alexander[1,8]

[1] Blood Cells and Blood Cancer Division, [2] Bioinformatics Division, [3] Immunology Division, [4] Advanced Technology and Biology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia, 3010. [5] Monash Haematology, Monash Hospital, Clayton, Victoria, Australia, 3004. [6] Department of Immunology and Pathology, Monash University, Melbourne, Victoria, Australia, 3004. [7] Translational Haematology Program, Peter MacCallum Cancer Centre, Parkville, Australia, 3000, [8] Department of Medical Biology, The University of Melbourne, Parkville, Australia, 3010. [9] Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Australia, 3010. [10] Centre for Cancer Research, The University of Melbourne, Parkville, Australia, 3010.

***Correspondence/Lead contact**

Dr Ashley P. Ng
The Walter and Eliza Hall Institute of Medical Research
1G Royal Parade
Parkville, Victoria, 3052
AUSTRALIA
Tel: +61-3-93435-2555
Fax: +61-3-9347-0852
Email: ang@wehi.edu.au

Running title: An *Erg*-mediated *Ebf1* and *Pax5* transcriptional program controls early B-cell development

Word count: 3,890

Conflict of Interest: None

Summary/Abstract

B-cell development is initiated by the stepwise differentiation of hematopoietic stem cells into lineage committed progenitors, ultimately generating the mature B-cells that mediate protective immunity. This highly regulated process also generates clonal immunological diversity via recombination of immunoglobulin genes. While several transcription factors that control B-cell development and V(D)J recombination have been defined, how these processes are initiated and coordinated into a precise regulatory network remains poorly understood. Here, we show that the transcription factor ETS Related Gene (*Erg*) is essential for the earliest steps in B-cell differentiation. Erg initiates a transcriptional network involving the B-cell lineage defining genes, *Ebf1* and *Pax5*, that directly promotes the expression of key genes involved in V(D)J recombination and formation of the B-cell receptor. Complementation of the Erg-deficiency with a productively rearranged immunoglobulin gene rescued B-cell development, demonstrating that Erg is an essential and exquisitely stage specific regulator of the gene regulatory network controlling B-lymphopoiesis.

Keywords

Transcription Factors; Erg; Ebf1; Pax5; Transcriptional Network; Gene Regulatory Network, VDJ recombination; Immunoglobulin Heavy Chain; Lymphopoiesis; pre-BCR complex; Gene regulation in immune cells

INTRODUCTION

Transcription factors are critical for controlling the expression of genes that regulate B-cell development. The importance of specific B-cell transcription factors is highlighted by the phenotype of gene knockout models. Failure of B-cell lineage specification from multi-potential progenitors occurs with deletion of *Ikzf1* [1] and *Spi1* (*Pu.1)* [2], while deletion of *Tcf3 (E2A)* [3] and *Foxo1* [4] results in failure of B-cell development from common lymphoid progenitors (CLPs). Developmental arrest at later B-cell stages is observed with deletion of *Ebf1* and *Pax5* at the pre-proB and proB stages respectively [5] [6]. This sequential pattern of developmental arrest associated with loss of gene function, along with ectopic gene complementation studies [2], gene expression profiling [7] and analysis of transcription factor binding to target genes, support models in which transcription factors are organised into hierarchical gene regulatory networks that specify B-cell lineage fate, commitment and function [8].

Two transcription factors that have multiple roles during B-cell development are *Ebf1*, a member of the COE family, and *Pax5,* a member of the PAX family. While Ebf1 and Pax5 have been shown to bind to gene regulatory elements of a common set of target genes in a co-dependent manner during later stages of B-lineage commitment [9], both manifest distinct roles during different B-cell developmental stages. Ebf1 forms an early B-cell transcriptional network with E2A and Foxo1 in CLPs that appears important in early B-cell fate determination [10], while during later stages of B-cell development, Ebf1 acts as a pioneer transcription factor that regulates chromatin accessibility at a subset of genes co-bound by Pax5 [11] as well as at the Pax5 promoter itself [12]. Pax5 in contrast, regulates B-cell genomic organisation [13] including the *Immunoglobulin heavy chain (Igh)* locus during V(D)J recombination, co-operating with factors such as CTCF [14], as well as transactivating [15] and facilitating the activity of the recombinase activating gene complex [16].

It is unclear, however, how these various functions of Ebf1 and Pax5 are co-ordinated during different stages of B-cell development. In particular, it would be important to ensure co-ordinated *Ebf1* and *Pax5* co-expression before the pre-BCR checkpoint, such that *Ebf1* and *Pax5* co-regulated target genes required for V(D)J recombination and pre-B-cell receptor complex formation are optimally expressed [9].

Here we show that the ETS related gene (Erg), a member of the ETS family of transcription factors, plays this vital role in B-lymphopoiesis. Deletion of Erg from early lymphoid progenitors resulted in B-cell developmental arrest at the early pre-proB cell stage and loss of V_H-to-DJ_H recombination. Gene expression profiling, DNA binding analysis and complementation studies demonstrated Erg to be a stage-specific master transcriptional regulator that lies at the apex of an Erg-dependent Ebf1 and Pax5 gene regulatory network in pre-proB cells. This co-dependent transcriptional network directly controls expression of the *Rag1/Rag2* recombinase activating genes and the *Lig4* and *Xrcc6* DNA repair genes required for V(D)J recombination, as well as expression of components of the pre-BCR complex such as *CD19, Igll1, Vpreb1* and *Vpreb2*. Taken together, we define an essential Erg-mediated transcription factor network required for regulation of *Ebf1* and *Pax5* expression that is exquisitely stage specific during B-cell development.

RESULTS

Erg is required for B-cell development

To build on prior work defining the role of the hematopoietic transcription factor *Erg* in regulation of haematopoietic stem cells (HSCs) [17] and megakaryocyte-erythroid specification [18], we sought to identify whether *Erg* played roles in other haemopoietic lineages. *Erg* expression in adult hematopoiesis was first examined by generating mice carrying the $Erg^{tm1a(KOMP)wtsi}$ knock-in first reporter allele (Erg^{KI}) (***Figure 1A***). Consistent with the known

role for *Erg* in hematopoiesis [17, 18, 19, 20, 21], significant *LacZ* expression driven by the endogenous *Erg* promoter was observed in hematopoietic stem cells (HSCs) and multi-potential progenitor cells, as well as in granulocyte-macrophage and megakaryocyte-erythroid progenitor populations, with declining activity accompanying erythroid maturation (***Figure 1B*** with definitions of cells examined provided in ***Table S1*** and representative flow cytometry plots in ***Figure S1***). In other lineages, transcription from the *Erg* locus was evident in common lymphoid (CLP), all lymphoid (ALP) and B-cell-biased lymphoid (BLP) progenitor cells, as well as in B-lineage committed pre-proB, proB and preB cells and double-negative thymic T-lymphoid cell subsets, with a reduction in transcription with later B-cell and T-cell maturation (***Figure 1B,C***). We confirmed these findings with RNA sequencing (RNA-seq) analysis that showed significant *Erg* RNA in pre-proB, proB and preB cells (***Figure 1D***). This detailed characterisation of *Erg* expression raised the possibility that *Erg* plays a stage-specific function at early developmental stages of the lymphoid lineages.

To determine whether *Erg* had a role in lymphoid development, mice carrying floxed *Erg* alleles ($Erg^{fl/fl}$, ***Figure 1A***) were interbred with *Rag1Cre* transgenic mice that efficiently delete floxed alleles in common lymphoid (CLPs) and T- and B- committed progenitor cells [22], but have normal lymphoid development (***Figure S2A***). The resulting $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ mice specifically lack *Erg* throughout lymphopoiesis (***Figure 1E, Figure S2B***). While numbers of red blood cells, platelets and other white cells were normal, $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ mice displayed a deficit in circulating lymphocytes (***Table S2***). This was due to a specific absence of B-cells; the numbers of circulating T-cells and thymic progenitors were not decreased (***Figure 1F, Figure S2C***).

B-cells are produced from bone marrow progenitor cells that progress through regulated developmental stages. B-cell development was markedly compromised in $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ mice, with proB, preB, immature B and mature recirculating B cells (Hardy fractions C-F,

defined in ***Table S1***) markedly reduced in number or virtually absent (***Figure 1F***). A B-lymphoid developmental block was clearly evident at the pre-proB (Hardy fraction A-to-B) stage, with excess numbers of these cells present in the bone marrow.

Erg deficient pre-proB cells have perturbed V_H-to-DJ_H recombination

To further characterise the developmental B-cell block in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice, B220$^+$ bone marrow progenitors were examined for *Igh* somatic recombination. Unlike cells from control *Erg*$^{fl/fl}$ mice, B220$^+$ cells from *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice had not undergone significant V_H-to-DJ_H immunoglobulin heavy chain gene rearrangement, although D_H-to-J_H recombination was relatively preserved (***Figure 2A***).

We next investigated the abnormalities underlying *Igh* recombination in greater detail. We first undertook fluorescence in situ hybridization (FISH) at the *Igh* locus to measure the intra-chromosomal distance between distal V_HJ558 and proximal V_H7183 V_H family genes, as cell stage specific contraction of the *Igh* locus is essential for efficient V(D)J recombination [23]. This revealed that B-cell progenitors from *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice had reduced locus contraction compared to *Erg*$^{fl/fl}$ controls (***Figure 2B***). To assess whether other structural perturbations across the *Igh* locus were also present, chromatin conformation capture and high throughput sequencing (Hi-C) was performed. This analysis revealed a reduction of long-range interactions across the *Igh* locus in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ B-cell progenitors when compared to *Erg*$^{fl/fl}$ and C57BL/6 controls (***Figure 2C***). As these findings were also observed in *Pax5* deficient B-cell progenitors [23] [13] reflecting a direct role for Pax5 in co-ordinating the structure of the IgH locus [14], we mapped *Erg* binding sites across the *Igh* locus by ChIP-seq. Unlike well-defined Pax5 binding to Pax5- and CTCF-associated intergenic regions (PAIR domains) [14] [16], Erg binding to V_H families was not identified across the locus (***Figure 2C***, ***Figure S4A***), suggesting that Erg was unlikely to be required structurally to maintain the multiple long-range

interactions and V_H-to-DJ_H recombination lacking in the *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ B-cell progenitors. Analysis of *Igh* locus accessibility by ATAC-seq did not reveal any significant difference between *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ pre-proB cells and control cells (***Figure S4A***) suggesting that loss of locus accessibility either by chromatin regulation [24] or peripheral nuclear positioning with lamina-associated domain silencing [25], were not mechanisms that could adequately explain reduced *Igh* locus contraction, reduction of long range interactions, and loss of V_H-to-DJ_H recombination in the absence of *Erg*.

A potential role for ETS family of transcription factors in regulation of immunoglobulin gene rearrangement was proposed from experiments investigating the iEμ enhancer: a complex cis-activating element located in the intronic region between the *Igh* joining region (J_H) and constant region (Cμ) implicated in efficient V_H-to-DJ_H recombination and *Igh* chain transcription [26]. The iEμ enhancer is proposed to nucleate a three-loop domain at the 3' end of *Igh* interacting with the V_H region to juxtapose 5' and 3' ends of the heavy chain locus [27]. *Erg* and its closest related ETS family member, *Fli1*, were shown to bind to the μA element and trans-activate iEμ co-operatively with a bHLH transcription factor *in vitro* [28]. We therefore sought to determine whether the lack of Erg, and Erg binding in particular to the μA site of iEμ, could account for loss of V_H-DJ_H recombination observed in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice *in vivo*. While ChIP-PCR demonstrated Erg binding to the iEμ enhancer containing the μA element (***Figure S3A***), mice in which the μA region (μA$^{\Delta/\Delta}$) was deleted had preserved numbers of circulating mature B-cells compared to cEμ$^{\Delta/+}$ controls (***Figure 2D***) and intact V_H-DJ_H recombination (***Figure S3C***). This was in contrast to cEμ$^{\Delta/\Delta}$ mice, in which a core 220bp element of iEμ was deleted, that demonstrated a marked reduction of circulating mature IgM$^+$IgD$^+$ B-cells in peripheral blood in keeping with previous models [29] (***Figure 2D***). Together these data show that while Erg can bind to the μA region of the iEμ *in vivo*, deletion of this region did not result in significant perturbation of the B-cell development. It is therefore

unlikely that Erg binding to μA element of iEμ could account for the loss of V_H-to-DJ_H recombination in particular, or the *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ phenotype in general.

The VH10tar *IgH* knock-in allele permits B-lymphoid development in the absence of *Erg*

Given the loss of V_H-DJ_H recombination associated with structural perturbation of the *Igh* locus in Erg-deficient pre-proB cells, we sought to complement the loss of formation of a functional *Igh* μ transcript and in doing so, determine whether failure to form a pre-BCR complex was a principal reason for the developmental block in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice [30]. Complementation with a functionally re-arranged *Igh* allele in models of defective V(D)J recombination such as deletion of *Rag1, Rag2,* or components of DNA-dependent protein kinase (DNA-PK) that mediate V(D)J recombination, can overcome the pre-BCR developmental block [31 32 33 34].

The *IgH*^{VH10tar} knock-in allele that expresses productive *Igh*^{HEL} transcripts under endogenous *Igh* locus regulation [32] was therefore used to generate mice that lacked *Erg* in B-cell progenitors but would undergo stage-appropriate expression of the rearranged *Igh*^{HEL} chain (*Rag1Cre*$^{T/+}$;*Erg* $^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$). The presence of the *IgH*^{VH10tar} allele permits B-cell development in the absence of *Erg*. The bone marrow of *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$ mice contained significant numbers of B220$^+$IgM$^+$ B-cells and, notably, CD25$^+$CD19$^+$IgM$^-$ PreB cells, a population coincident with successful pre-BCR formation [35], that were virtually absent in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice (***Figure 3A***). Similarly, in the spleens of *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$ mice, near normal numbers of all B-lymphoid populations were observed, in contrast to the marked reduction in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice (***Figure 3B***). Notably, IgκL chain recombination had proceeded in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$ cells (***Figure 3C***).

We next tested whether the rescued *Rag1Cre*$^{T/+}$;*Erg* $^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$ splenic B-cells were functional in the absence of *Erg*. *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$ splenocytes were indistinguishable from wild-type controls in *in vitro* proliferative assays using anti-μ

stimulation, T-cell dependent stimulation with CD40 ligand, IL4 and IL5, or T-cell independent stimulation using lipopolysaccharide (***Figure 3D***). *Rag1Cre*$^{T/+}$*;Erg* $^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$splenic B-cells were also able to differentiate normally as measured by formation of plasma cells and IgG1 class switch recombination (***Figure 3E***). Circulating *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$ B-cells also expressed IgD, unlike their *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ counterparts (***Figure 3F***). These experiments demonstrated that loss of a functional *Igh* μ transcript and failure to form a pre-BCR complex was a principal reason for lack of B-cell development in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ mice.

Erg-deficient pre-proB cells do not express Ebf1 and Pax5 transcription factors

To define the mechanism by which *Erg* regulates V_H-to-DJ_H recombination and pre-BCR formation, we undertook gene expression profiling of *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells. Differential gene expression and gene-ontogeny analysis of differentially expressed genes in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB compared to *Erg*$^{fl/fl}$ pre-proB cells demonstrated deregulated expression of multiple B-cell genes (***Figure 4A***). These included genes encoding cell surface or adhesion receptors and core components of the pre-BCR complex *CD19, CD22, Igll1, Vpreb1, Vpreb2, CD79a and CD79b*, genes required for IgH recombination such as *Rag1* and *Rag2* and components of non-homologous end-joining repair complex associated with V(D)J recombination: *Xrcc6* (Ku70) and *Lig4*, and importantly, transcription factors implicated in B-cell development (*Ebf1, Pax5, Tcf3, Bach2, Irf4, Myc, Pou2af1, Lef1, Myb*) (***Figure 4B***).

Ebf1 and *Pax5* are critical for B-lineage specification [5] and maintenance [36, 37] and act co-operatively to regulate a gene network in early B-cell fates [9]. Because we observed with loss of *Erg* reduced expression of several critical B-cell genes previously identified to be controlled by *Ebf1* and/or *Pax5*, for example *CD19*, *Vpreb1*, and *Igll1* (***Figure 4A***), we speculated that Erg may play an important role in regulating the expression of these two essential transcription

factors and their targets. To determine if *Erg* bound *Ebf1* and/or *Pax5* gene regulatory regions and directly regulated their expression, we undertook ChIP-seq analysis in wild-type B-cell progenitors and ATAC-seq to assess locus accessibility at the *Ebf1* and *Pax5* loci in the absence of Erg in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ B-cell progenitors. This demonstrated direct Erg binding to the proximal (β) promoter region of *Ebf1* [38] as well as to the *Pax5* promoter and *Pax5* lymphoid specific intron 5 enhancer [12] (***Figure 4C***), which together with the absence of *Ebf1*and *Pax5* expression in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells and the loss of Ebf1 and Pax5 protein in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ B-cell progenitors by Western Blot, demonstrated that Erg was a direct transcriptional regulator of *Ebf1* and *Pax5* (***Figure 4D***). Importantly, the loss of *Ebf1* and *Pax5* expression occurred while expression of other known regulators of *Ebf1* expression, namely, *Foxo1*, *Spi1, Tcf3* and *Ikzf1* were maintained (***Figure 4C and Figure S4B***), and both *Ebf1* and *Pax5* loci remained accessible by ATAC-seq in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ B-cell progenitors (***Figure 4C***).

A co-dependent gene regulatory network dependent on Erg, Ebf1 and Pax5

Because expression of multiple B-cell genes were deregulated in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells, including those to which Ebf1 and Pax5 had been shown to directly bind and regulate, we investigated the possibility that Erg co-bound common target genes to reinforce the Ebf1 and Pax5 gene network using a genome wide motif analysis of Erg DNA binding sites in B-cell progenitors. As expected, the most highly enriched motif underlying Erg binding was the ETS motif. However, significant enrichment of *Ebf1, E2A, Pax5* and *Foxo1* binding motifs were also identified within 50bp of Erg binding sites (***Figure 4E***), suggesting that Erg may indeed act co-operatively with other transcription factors to regulate target gene expression in a co-dependent gene network. Analysis of the binding of each of Erg, Ebf1 and Pax5 to regulatory regions of genes that were differentially expressed in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB

cells was then undertaken. This analysis identified significant overlap of Erg, Ebf1 and Pax5 binding sites within 5kb of the transcriptional start site (TSS) of genes differentially expressed in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells compared with control pre-proB cells (***Figure 4F***). Taken together, these data provided compelling evidence for a gene regulatory network, in which Erg is required for maintaining expression of Ebf1 and Pax5 at the pre-proB cell stage of development, as well as reinforcing expression of target genes within the network by co-operative binding and co-regulation of target genes with Ebf1 and Pax5.

To further explore our finding that Erg, Ebf1 and Pax5 form the core of a gene regulatory network in pre-proB cells, examination of Ebf1 and Pax5 binding to the *Erg* locus was undertaken. Ebf1 and Pax5 binding within intron 1 of the *Erg* locus associated with the H3K27ac mark was found, as was Pax5 binding at the *Erg* promoter (***Figure S4B***). To determine if Ebf1 and Pax5 directly regulate *Erg* expression, gene expression changes in B-cell progenitors from a publicly available dataset in which *Ebf1* (*Ebf1*$^{\Delta/\Delta}$) or *Pax5* (*Pax5*$^{\Delta/\Delta}$) had been deleted were examined (***Figure 5A***). Deletion of either *Ebf1* or *Pax5* resulted in reduced *Erg* expression (***Figure 5B***), with Ebf1 appearing to be the stronger influence. We next compared gene expression changes in *Ebf1*$^{\Delta/\Delta}$ and *Pax5*$^{\Delta/\Delta}$ B-cell progenitors to those genes regulated by Erg in pre-proB cells. As would be predicted if Erg, Ebf1 and Pax5 were components of a co-dependent gene regulatory network, this analysis showed a highly significant correlation in gene expression changes observed with *Ebf1* or *Pax5* deletion in B-cell progenitors and those observed with *Erg* deletion in pre-proB cells. This was noted for down-regulated genes in Erg, Ebf1 and Pax5 deficient B-cell progenitors in particular (***Figure 5C***).

Finally, to confirm that *Ebf1* and *Pax5* were transcriptional regulators down-stream of Erg in pre-proB cells, transfection of *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ progenitor cells with MSCV-driven constructs for constitutive expression of *Ebf1* and *Pax5* was performed. This experiment

demonstrated rescue of B220 expression with *Ebf1* or *Pax5* over-expression in Erg deficient progenitors (***Figure 5D***). Notably, only partial rescue of CD19 expression and V_H-to-DJ_H recombination was observed with *Ebf1* over-expression while no rescue was observed with *Pax5* over-expression (***Figure 5D,E***). These observations suggest that while Ebf1 over-expression could partially compensate for several aspects of B-cell development in the absence of Erg, Pax5 over-expression alone could not. This is in keeping with a hierarchical model highlighting the importance of Erg as a key mediator of the network.

Taken together, these experiments demonstrated the existence of a co-dependent transcriptional network between Erg, Ebf1 and Pax5, that co-regulate critical target genes at the pre-proB cells stage of B-cell development.

To further the delineate the directly regulated target genes in an Erg-dependent Ebf1 and Pax5 transcriptional network, we undertook mapping of ChIP-seq binding of Erg, Ebf1 and Pax5 to differentially expressed genes at the pre-proB cell stage of development in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ cells. We identified that the majority of these target genes demonstrated direct combinatorial binding of Erg, Ebf1 and/or Pax5 to annotated promoter regions, gene body enhancer/putative enhancer regions or putative distal enhancer regions of these genes (***Figure 6A***). Detailed examination of several key target genes for which expression was completely dependent on Erg in pre-proB cells identified direct binding of Erg to the promoter and enhancer regions for several pre-BCR components, including *CD19*, *Igll1, Vpreb1* and *CD79a.* This occurred with co-ordinate binding of Ebf1 and Pax5 to the regulatory regions of these genes [15] (***Figure 6B***). In addition, indirect regulation by Erg at the *Rag1/Rag2* locus was also identified, with down-regulation of expression of transcription factors that bind and regulate the *Rag2* promoter such *Pax5*, *Lef1* and *c-Myb* in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells (***Figure 4B***) [39], as well as direct binding of Erg to the conserved B-cell specific *Erag* enhancer [40] (***Figure S4C***). Importantly, the loss of *Rag1* and *Rag2* expression in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells occurred while

expression of *Foxo1*, a positive regulator of the locus [41] was relatively maintained (***Figure S4B***).

An Erg-Ebf1-Pax5 mediated gene regulatory network was then mapped using each target gene, expression of which was perturbed in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells, and that was directly bound by Erg, Ebf1 and/or Pax5 at promoter, proximal or distal gene regions, to provide a comprehensive representation of this gene network (***Figure 6C***).

A key observation arising from our data was that the B-cell developmental block arising in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells could be overcome with the provision of a rearranged functional IgH VH10tar allele. This suggested that once the pre-BCR checkpoint was bypassed, *Erg* was no longer critical for further B-cell development and function, including V_LJ_L recombination of the *Igl* and BCR formation (***Figure 3C,D***). Indeed, beyond the pre-BCR checkpoint, re-emergence of *Ebf1* and *Pax5* expression occurred (***Figure 4C***) as well as expression of target genes of the *Ebf1* and *Pax5* network (***Figure 6B, Figure S4B,C***) in Erg-deficient *Rag1Cre*$^{T/+}$*;Erg* $^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$ proB and preB cells rescued with a VH10tar allele. This defines the role of Erg as an exquisitely stage specific regulator of early B cell development.

Discussion

In this study we explored the role of the transcription factor *Erg* in B-lymphopoiesis. Our studies suggest two regions controlling Erg expression during B-cell development: the *Erg* promoter region and the H3K27ac-marked putative enhancer region in the first intron, to which the B-cell transcription factors Ebf1 and Pax5 directly bind. Complete loss of *Erg* expression in Ebf1$^{\Delta/\Delta}$ B-cell progenitors in which *Pax5* and *Foxo1* expression was also lost, place the initiation of Erg expression in the B-lymphoid lineage downstream of the E2A, Ebf1, Foxo1 transcriptional network at the CLP stage of lymphoid development [10]. The importance of Erg

in B-cell development was demonstrated in mice in which *Erg* had been deleted throughout lymphopoiesis, which exhibited a developmental block at the pre-proB cell stage that was associated with profound defects in V_H-to-DJ_H recombination, *Igh* locus organization and transcriptional changes in multiple B-cell genes, including loss of expression of *Ebf1*, and *Pax5*. Combining RNA-seq, ChIP-seq and gene complementation studies, we were able to define a co-dependent transcriptional network between Erg, Ebf1 and Pax5, with direct Erg binding to the proximal (β) *Ebf1* promoter, to which Pax5, Ets1 and Pu.1 also co-operatively bind [38], as well as Erg binding to the *Pax5* promoter and potent intron 5 enhancer region, two critical *Pax5* regulatory elements required for correct transcriptional initiation of *Pax5* in early B-cell development [12]. These data support a model (***Figure 6D***) in which Ebf1 expression, initially Erg-independent in CLPs, requires Erg in pre-proB cells to promote and maintain its expression. Erg is also required for simultaneous *Pax5* expression at this stage of development for the establishment of an inter-dependent B-lymphoid gene regulatory network.

Together Erg, Ebf1 and Pax5 directly co-regulated the expression multiple genes that had previously been identified as direct transcriptional targets of Ebf1 and Pax5 (***Figure 6D***). Direct Erg binding to promoters of the pre-BCR signalling complex genes such as *Igll1*, *VpreB* and *CD79a,* establish Erg as a transcriptional regulator of target genes in this network. In addition to *Rag1* and *Rag2*, we also identified network regulation of expression of *Xrcc6*, the gene encoding the Ku70 subunit of DNA-dependent protein kinase holoenzyme (DNA-PK) that binds DNA double strand breaks during V(D)J recombination [42], and *Lig4,* encoding the XRCC4 associated DNA-ligase that is required for DNA-end joining during V(D)J recombination [43] (***Figure 6C, S4C***). Along with direct Erg promotion of expression of *Pax5* as a structural regulator of the *Igh* locus, these findings are sufficient to explain the *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ phenotype in which V_H-toDJ_H recombination was lost. Together with loss of expression of components of the pre-BCR complex, we can conclude B-cell development

was blocked as a consequence of *Erg* deletion due to the collapse of the Erg-mediated transcriptional network.

Importantly, re-emergence of *Ebf1* and *Pax5* expression beyond the pre-BCR checkpoint in *IgH*-rescued *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$ cells was observed, along with expression of target genes of Ebf1 and Pax5. This demonstrates that Erg is a stage-specific regulator of B-cell development, with emergence of an Erg-independent Ebf1 and Pax5 gene network during later stages of B-cell development, once clones have transitioned through the pre-BCR checkpoint. This would allow IgL chain V_L to J_L recombination and BCR formation to proceed in preB cells in which endogenous *Erg* expression is also reduced (***Figure 1B,C***). *Erg* however, is critical for promoting *Ebf1* and *Pax5* expression in pre-pro-B cells, orchestrating a transcriptional network required for V_H-to-DJ_H recombination, pre-BCR formation, and early B cell development. In this role, Erg not only co-ordinates the transcriptional functions of Ebf1 and Pax5, but reinforces the *Erg*-mediated transcriptional network by directly binding and activating critical target genes required for transition through the pre-BCR checkpoint.

Acknowledgements

We thank Janelle Lochland, Jason Corbin, Jasmine McManus, Melanie Salzone, Carolina Alvarado, Keti Stoev, Nicole Lynch and Shauna Ross for skilled assistance. We thank Professor Robert Brink for the V_H10tar knock-in mouse line. This work was supported by Program Grants (1113577, 1016647, 1054618, 1054925), Project Grant (APN 1060179, 1122783), Fellowship (DMT 1060675, SLN 1155342, WSA 1058344, TMJ 1124081), C.R.B. Blackburn Scholarship (MSYL, Australian National Health and Medical Research Council jointly with Royal Australasian College of Physicians) and Independent Research Institutes Infrastructure Support Scheme Grant (361646) from the Australian National Health and Medical Research Council, the Australian Cancer Research Fund and Victorian State Government Operational Infrastructure Support. YCC was supported through Maddie Riewoldt's Vision. The MAGEC laboratory was supported by the Australian Phenomics Network and the Australian Government through the National Collaborative Research Infrastructure Strategy Program.

Author Contribution

Conceptualization, A.P.N., M.S.Y.L, A.J.K., T.M.J., M.A.D., R.S.A., K.R., D.M.T., G.K.S., M.J.D., S.L.N. and W.S.A.; Methodology, A.P.N., M.S.Y.L, T.M.J., T.B., M.A.D., R.S.A., K.R., D.M.T., G.K.S., M.J.D., S.L.N. and W.S.A.; Investigation, A.P.N., H.C., S.H., K.B., T.M.J., M.S.Y.L., C.C.B., O.G., Y.C.C., T.B., L.D., C.D.H., H.I., S.M., E.V., T.W., K.R., G.K.S., M.J.D.; Formal analysis, A.P.N., H.C., S.H., M.S.Y.L., O.G., C.C.B., Y.C.C. T.B., K.R., M.J.D., S.L.N; Writing – Original Draft, A.P.N.; Writing – Review & Editing, A.P.N., H.C., S.H., G.K.S., S.L.N., and W.S.A.; Funding Acquisition, A.P.N. and W.S.A.; Supervision, A.P.N., M.A.D., D.M.T., G.K.S., M.J.D., S.L.N. and W.S.A.

COMPETING FINANCIAL INTERESTS

The authors declare that there are no competing financial interests.

Figure Legends

Figure 1. Expression of the *Erg* locus and targeted disruption of *Erg* in lymphopoiesis.

Figure 2. The immunoglobulin heavy chain locus in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice.

Figure 3. A rearranged V_H10_{tar} IgH allele rescues *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ B-lymphoid development.

Figure 4. Gene expression in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ pre-proB cells and Erg DNA binding.

Figure 5. Gene expression in Ebf1- and Pax5-deficient B-cell progenitors and rescue of Erg-deficient B-cell progenitors.

Figure 6. The Erg mediated Ebf1 and Pax5 gene regulatory network in pre-proB cells.

Figure 1. Expression and targeted disruption of *Erg* in lymphopoiesis. **A.** Wild-type (*Erg)*, $Erg^{tm1a(KOMP)wtsi}$, *lacZ* reporter (Erg^{KI}), conditional (Erg^{fl}), and Cre recombinase-deleted (Erg^{Δ}) alleles with exons, Cre (loxP) and Flp (frt) recombinase recognition sites. IRES, internal ribosome entry site; Neo, neomycin-resistance cassette. **B.** *Erg* transcriptional activity by *lacZ* expression in Erg^{KI} bone marrow (BM) and thymus cell populations (see Materials and Methods and ***Figure S1***, ***Table S1***). Mean fluorescent intensity (MFI) ratio ± S.D of Erg^{KI} (n=4) to C57BL/6 (n=4). $P_{adj} < 0.027$ corrected using Holm's modification for multiple testing for each population except BM $Ter119^+$ and $NK1.1^+$, and thymic DP, $CD4^+CD8^-$ and $CD8^+CD4^-$ populations ($P_{adj} > 0.05$). **C.** Representative flow cytometry plots: BM pre-proB (blue), proB (green) and preB (orange) and control $B220^+IgM^-IgD^-$ (black) (left) with *lacZ* MFI (right). **D.** *Erg* expression by RNA-seq (mean±S.D, Fragments Per Kilobase of transcript per Million mapped reads, FPKM) in $Erg^{fl/fl}$ pre-proB, proB and preB cells (n = 2) **E.** *Erg* RNA-seq (FPKM) in $Erg^{fl/fl}$ and $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ pre-proB cells (n=2) (left; *, P = 1.41e-5, ***Table S4***). *Erg* locus RNA-seq in $Erg^{fl/fl}$ (WT) and $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ (Erg KO, with pink highlighting absent expression) in pre-proB cells, H3K4me3 and H3k27ac ChIP-seq and chromatin accessibility (ATAC-seq, blue). **F.** $Erg^{fl/fl}$ (n=4) and $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ (n=7) $B220^+$B-cell, $Gr1^+Mac1^+$myeloid-cell, and $CD3^+$T-cell blood counts, mean±S.E.M; *, P = 6e-8 (top left). B-lymphoid populations in $Erg^{fl/fl}$ (n=9) and $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ (n=10) BM as ratio of cell number to $Erg^{fl/fl}$ (bottom left, see ***Table S1***). *, $P_{adj} < 0.04$ corrected using Holm's modification for multiple testing. Representative flow cytometry plots (right).

Figure 2. The immunoglobulin heavy chain locus in $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ mice**. A.** Genomic PCR using degenerate primers to IgH locus V_H558, V_H7183, V_HQ52 segments for detection of V_H to DJ_H (top panel) and D_H to J_H (middle panel) recombination with Mu0 loading controls (bottom panel) in $B220^+$ BM cells. **B.** Intra-chromosomal distance between distal VhJ558 and

proximal Vh7183 V_H families by Fluorescent In Situ Hybridisation from (N=129) *Igh* alleles from $Erg^{fl/fl}$ and $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$ B-cell progenitors. *P* value by Student's two-tailed unpaired t-test. **C.** Long-range chromatin interaction by chromatin conformation and capture analysis (Hi-C) of the immunoglobulin heavy chain locus of C57BL6 (wildtype) and $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$ B-cell progenitors. Reduced long-range interactions in $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$ B-cell progenitors indicated by blue arcs. Erg binding by ChIP (black bars) across the heavy chain locus (pink bar) as indicated (see also ***Figure S4A***). Location of 3'regulatory region (red bars), iEμ enhancer (purple bar) and PAIR domains (green bars) are indicated. **D.** Schematic representation of iEμ enhancer with the core 220bp $cE\mu^{\Delta}$ deletion and μA^{Δ} deletion shown (top). Peripheral blood counts of $B220^+$, $B220^+CD19^+$ and IgM^+IgD^+ B-cells in $cE\mu^{\Delta/+}$ (n=8), $cE\mu^{\Delta/\Delta}$ (n=3) and $\mu A^{\Delta/\Delta}$ (n=7) mice (bottom). * *P* value < 0.0001 by Benjamini Hochberg correction for multiple testing. See also ***Figure S3C.***

Figure 3. Rearranged V_H10_{tar} IgH allele permits $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$ B-lymphoid development. **A.** Representative flow cytometry plots of BM B-lymphoid populations (n = 9 $Erg^{fl/fl}$, n = 8 $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$, n = 8 $Rag1Cre^{T/+;}Erg^{\Delta/\Delta};IgH^{VH10tar/+}$) with mean percentage of viable cells indicated. B220/IgM profile (whole BM); CD25/CD19 profile ($B220^+IgM^-$ BM cells). Δ, $P<10^{-3}$ comparing $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$ to $Erg^{fl/fl}$; *, $P<0.01$ comparing $Rag1Cre^{T/+;}Erg^{\Delta/\Delta};IgH^{VH10tar/+}$ to $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$. **B.** Proportions of viable splenic B-lymphoid populations (n=14 $Erg^{fl/fl}$, n=10 $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$, n=9 $Rag1Cre^{T/+;}Erg^{\Delta/\Delta};IgH^{VH10tar/+}$). Δ, $P<3x10^{-4}$ comparing $Erg^{fl/fl}$ to $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$; *, $P<2x10^{-3}$ comparing $Rag1Cre^{T/+;}Erg^{\Delta/\Delta}$ to $Rag1Cre^{T/+;}Erg^{\Delta/\Delta};IgH^{VH10tar/+}$, mean±S.D. Fol=follicular, MZ=marginal zone (see ***Figure S1***, ***Table S1***). **C.** PCR of genomic DNA from $B220^+$ splenocytes for *Erg* (top panel; fl, floxed allele, Δ, cre-deleted allele), V_H-to-DJ_H recombination of V_H558, V_H7183, V_HQ52 families (second panel), V_H10tar allele (third panel) and $V_κ$ light chain recombination (bottom panel). **D.** Proliferation by cell

trace violet assay of wild-type (C57BL/6), *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$*;IgH*$^{VH10tar/+}$, *Rag1Cre*$^{+/+}$*;Erg*$^{fl/fl}$ and *Rag1Cre*$^{T/+}$ B220^{+} splenocytes to anti-IgM, CD40L+IL4+IL5 (T-cell-dependent) and LPS (T-cell-independent) stimulation. Mean percentage of viable cells for each cell division shown. No significant differences between genotypes were observed (P>0.90, 2-way ANOVA). **E.** Percentage of B220^{+} splenocytes differentiating to CD138^{+} plasma cells and undergoing IgG1 class switch recombination in response to CD40L+IL4+IL5 stimulation by flow cytometry. No significant differences were observed (P>0.40 for CD138 plasma cell differentiation, P>0.07 for IgG1 switch, corrected by Sidak's multiple comparison test). For **D** and **E**, n=2-4 C57BL/6, 3 *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$*;IgH*$^{VH10tar/+}$, 2-4 *Rag1Cre*$^{+/+}$*;Erg*$^{fl/fl}$ and 3 *Rag1Cre*$^{T/+}$ mice. **F.** Percentage circulating IgM^{+}IgD^{+} B220^{+}B-cells in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ (n=31), *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$*;IgH*$^{VH10tar/+}$ (n=17), and *Erg*$^{fl/fl}$ (n=9), $P<10^{-27}$ comparing *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ to *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$*;IgH*$^{VH10tar/+}$ by unpaired t-test.

Figure 4. Gene expression in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells and Erg DNA binding. **A.** Differentially expressed genes in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells compared to *Erg*$^{fl/fl}$ controls, manually curated according to function based on GO term analysis (see ***Table S4***) with the number of genes for each functional category shown by the horizontal axis and selected genes highlighted in boxes. **B.** Heatmap of gene expression differences between *Erg*$^{fl/fl}$ and *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells curated to transcription factors [44] and ordered by logFC. Selected B-cell transcription factors highlighted in red. **C.** RNA-seq for *Ebf1*, *Pax5* and *Tcf3* loci, with ChIP-seq for Erg binding in C57BL/6 B-cell progenitors and thymic *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ Erg knockout cells (Erg KO) to control for sites of non-Erg ChIP binding to DNA, H3K4me3 promoter mark, H3K27ac promoter and enhancer mark, and ATAC-seq, in *Erg*$^{fl/fl}$ pre-proB cells (pre-proB), *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB (Erg KO pre-proB), and Erg deficient proB and preB cells in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$*;IgH*$^{VH10tar/+}$mice that develop with a

functionally rearranged immunoglobulin heavy chain allele. * indicates Erg binding to the promoter region (blue bar) of *Ebf1* and *Pax5*. Solid pink bars: Erg binding to intragenic enhancer regions, with intron number as indicated. Open blue bar: promoter region with no Erg binding, open pink bar: putative enhancer region with no Erg binding. **D.** Western blot for Erg, Ebf1, Pax5 and β-actin in B-cell progenitors of genotypes indicated. **E.** Whole genome HOMER motif discovery underlying Erg bound regions in B-cell progenitors **F.** Heatmap of Erg, Ebf1 and Pax5 binding to differentially expressed genes in $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ pre-proB cells centred around the transcriptional start site (TSS) ± 5.0kB (see ***Table S5*** for all annotated ChIP binding sites).

Figure 5. Gene expression in Ebf1- and Pax5-deficient B-cell progenitors and rescue of Erg-deficient B-cell progenitors**. A.** Heatmap of top 100 most variable genes in wild-type (n=3), $Ebf^{\Delta/\Delta}$ (n=3), $Pax5^{\Delta/\Delta}$ (n=3) B-cell progenitors with hierarchical clustering applied. **B.** Expression of *Erg*, *Pax5*, *Foxo1* and *Tcf3* in wild-type, $Ebf^{\Delta/\Delta}$ and $Pax5^{\Delta/\Delta}$ B-cell progenitors (RPKM). **C.** Barcode enrichment plots depicting strongly associated gene expression signatures of down (vertical blue bars) and up (vertical red bars) regulated genes in $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ pre-proB cells compared to $Ebf1^{\Delta/\Delta}$ (top) and $Pax5^{\Delta/\Delta}$ (bottom) B-cell progenitors. Genes are ordered (from left to right) as most downregulated to most upregulated in $Ebf1^{\Delta/\Delta}$ or $Pax5^{\Delta/\Delta}$ B-cell progenitors compared to wild-type. The x-axis shows the moderated t-statistic in $Ebf1^{\Delta/\Delta}$ or $Pax5^{\Delta/\Delta}$ versus wild-type cells. A camera gene set test [45] confirms the correlation between $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ pre-proB cell and $Ebf1^{\Delta/\Delta}$ or $Pax5^{\Delta/\Delta}$ B-cell progenitor expression signatures with *P* values as shown for up- and down- regulated genes. **D.** Percentage of $B220^{+}$ and $CD19^{+}$ expressing GFP^{+} B-cell progenitors derived from lineage negative $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ BM transfected with MSCV control (n=3), Ebf1- (n=3) or Pax5-expressing (n=3) retroviruses and cultured on OP9 stromal cells with IL-7, SCF and Flt-

ligand. *, $P < 0.005$ by Student's unpaired t-test compared to MSCV control. **E.** V_H-to-DJ_H recombination of V_H558, V_H7183, V_HQ52 segments (top panel) recombination with Mu0 loading controls (bottom panel) in $B220^+$ enriched B-cell progenitors derived from lineage negative C57BL/6 (n=2), *Rag1Cre$^{T/+}$;Erg$^{\Delta/\Delta}$* (n=2), and *Rag1Cre$^{T/+}$;Erg$^{\Delta/\Delta}$* BM transfected with Ebf1 (n=2) and Pax5 (n=2) retroviruses.

Figure 6. The Erg mediated Ebf1 and Pax5 gene regulatory network in pre-proB cells. **A.** Erg, Ebf1 and Pax5 binding to annotated regions of differentially expressed genes in *Rag1Cre$^{T/+}$;Erg$^{\Delta/\Delta}$* pre-proB cells. **B.** RNA-seq gene expression at *CD19, Igll1, VpreB1* and *CD79a* loci, with ChIP-seq of Erg, Ebf1, and Pax5 binding, H3K4me3 promoter mark, H3K27ac promoter and enhancer mark, and ATAC-seq of *Erg$^{fl/fl}$* pre-proB cells, *Rag1Cre$^{T/+}$;Erg$^{\Delta/\Delta}$* pre-proB (Erg KO pre-proB), and Erg deficient proB and preB cells in *Rag1Cre$^{T/+}$;Erg$^{\Delta/\Delta}$;IgH$^{VH10tar/+}$*mice rescued with a functionally rearranged immunoglobulin heavy chain allele. Solid blue bar: Erg, Ebf1 and/or Pax5 binding promoter. Solid pink bar: Erg, Ebf1 and/or Pax5 binding to enhancer regions. Open blue bar: promoter region with no binding of Erg, Ebf1 or Pax5. **C.** The Erg dependent Ebf1 and Pax5 transcriptional network in pre-proB cells with binding of each transcription factor shown to annotated promoter, proximal and distal gene regions of differentially expressed genes in *Rag1Cre$^{T/+}$;Erg$^{\Delta/\Delta}$* pre-proB cells. See ***Table S5***. **D.** Summary of the Erg dependent Ebf1 and Pax5 transcriptional network in V_H-to-DJ_H recombination and pre-BCR formation.

Figure 1

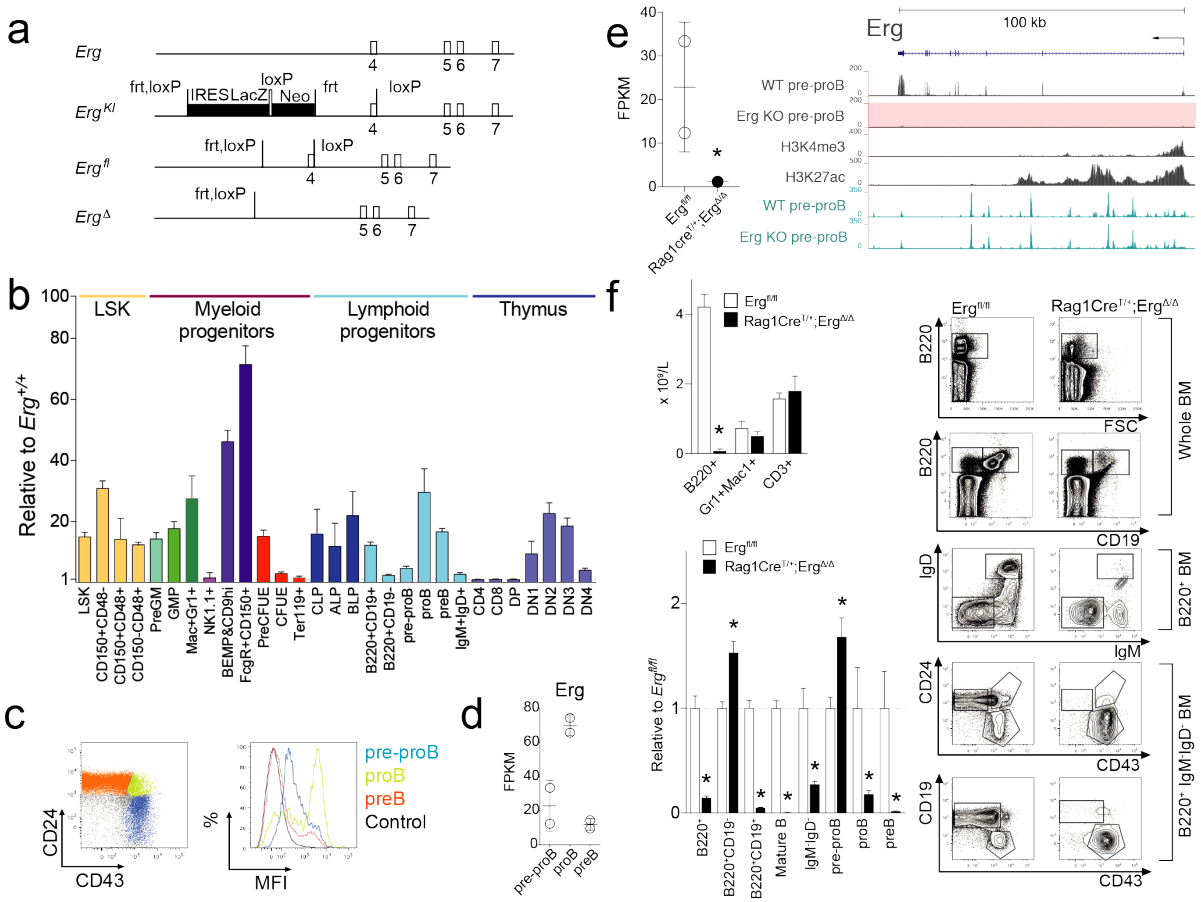


Figure 2

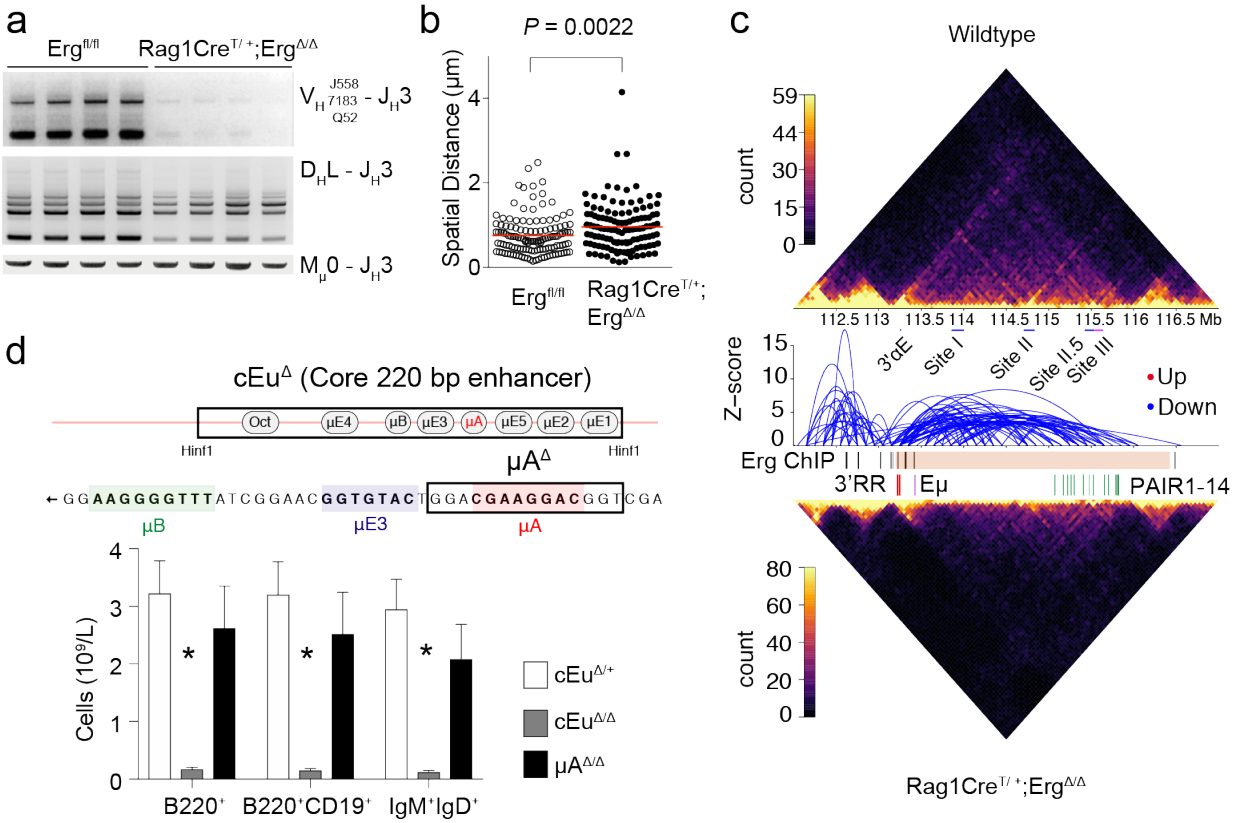


Figure 3

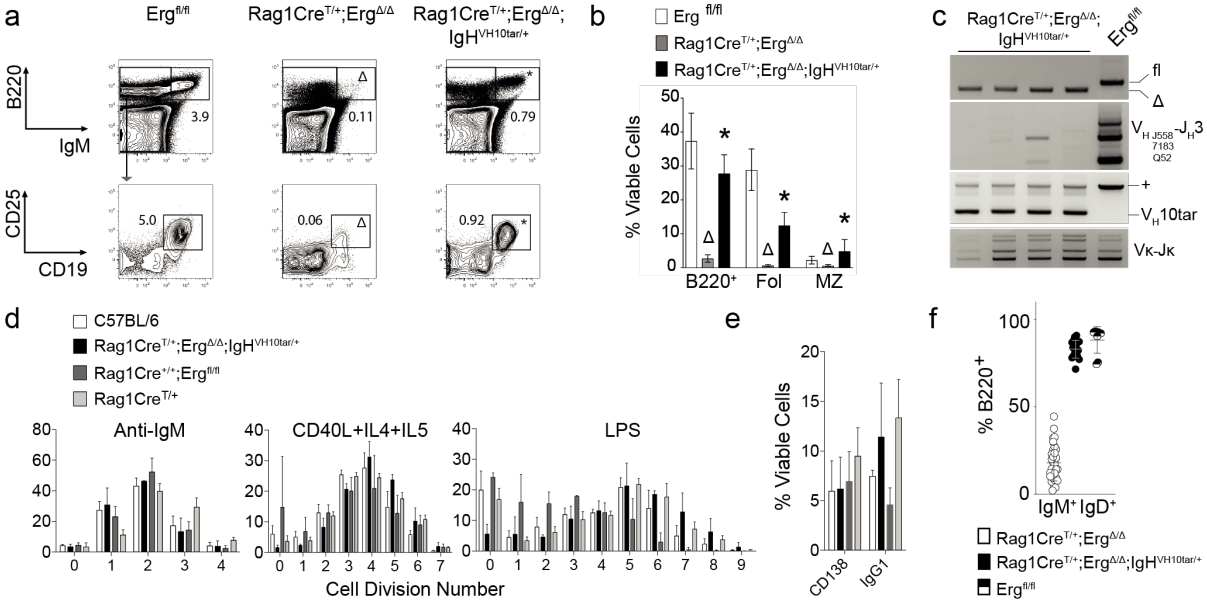


Figure 4

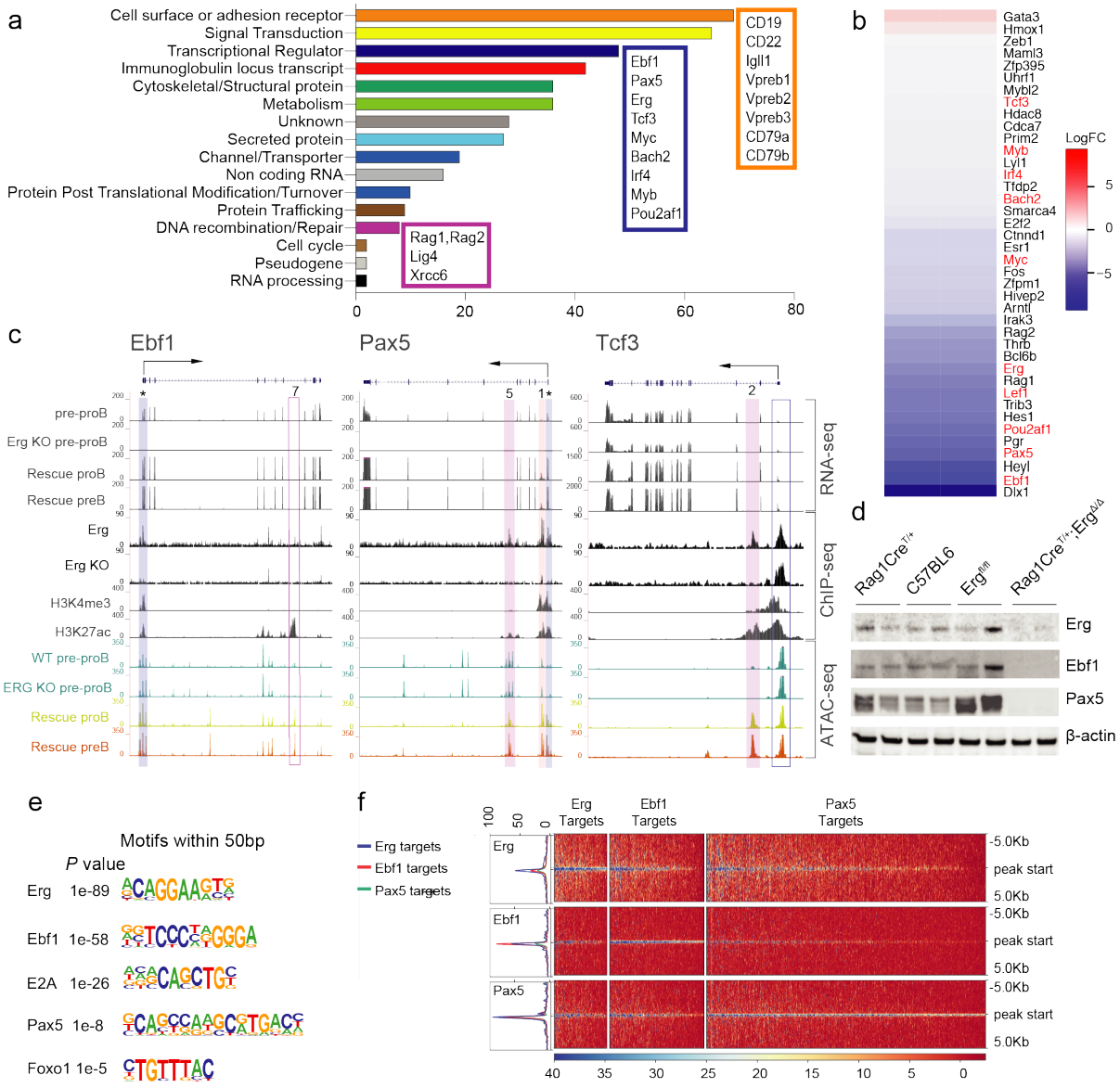


Figure 5

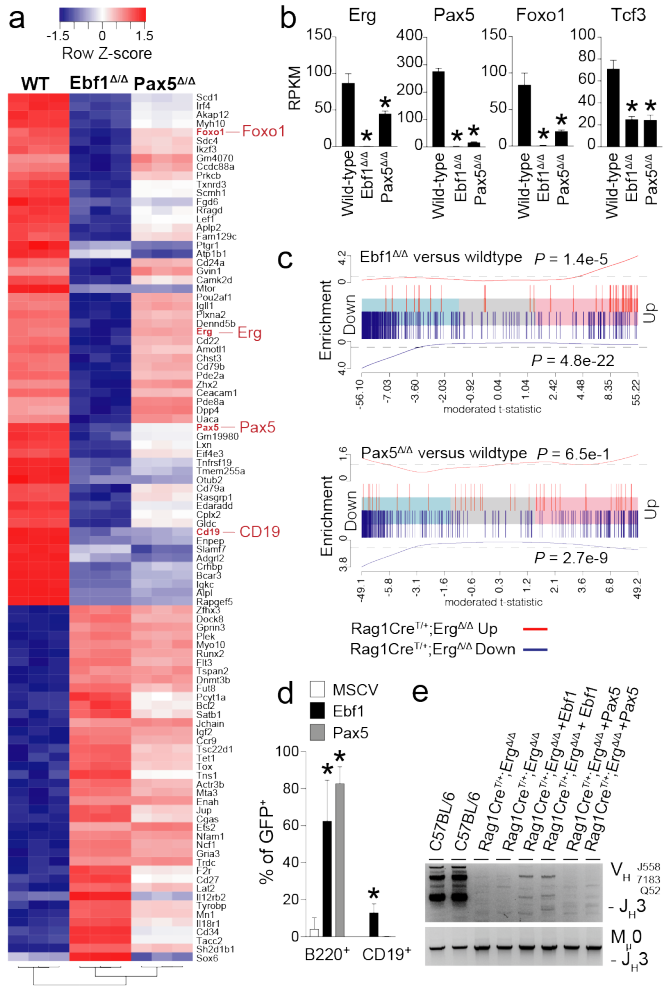
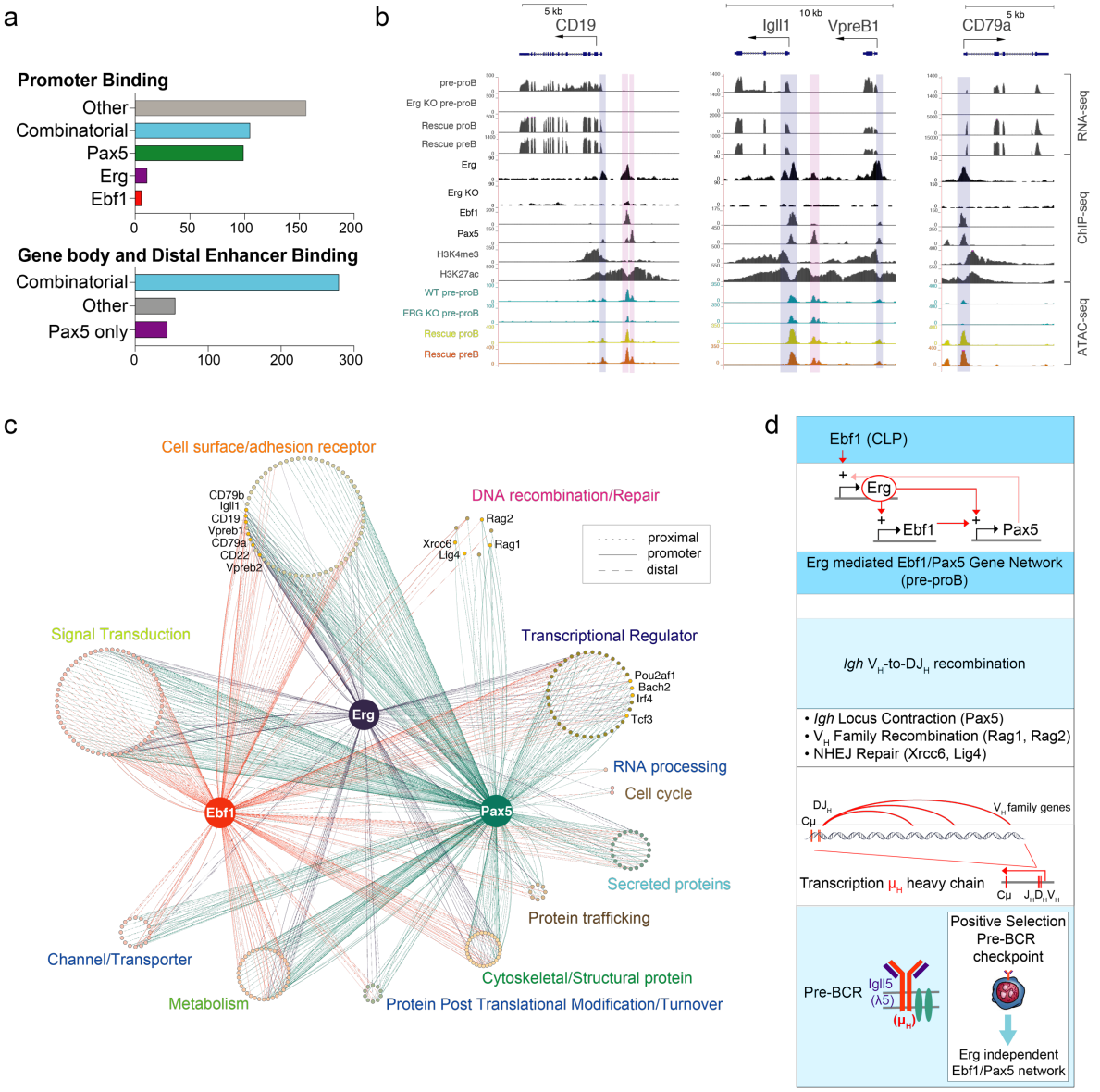


Figure 6



CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ashley Ng (ang@wehi.edu.au) subject to Material Transfer Agreements.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice. Mice carrying the *Erg*$^{tm1a(KOMP)wtsi}$ knock-first reporter allele [46] (*Erg*^{KI}, KOMP Knockout Mouse Project) were generated by gene targeting in ES cells. Mice with a conditional *Erg* knockout allele (*Erg*^{fl}) from which the IRES-LacZ cassette was excised was generated by interbreeding *Erg*^{KI} mice with *Flpe* transgenic mice [47]. *Rag1Cre* mice [48], in which Cre recombinase is expressed during lymphopoiesis from the CLP stage [22], were interbred with *Erg*^{fl} mice to generate mice lacking *Erg* in lymphopoiesis (*Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$) and *Rag1Cre*$^{+/+}$;*Erg*$^{fl/fl}$ (*Erg*$^{fl/fl}$) controls. Mice carrying the rearranged immunoglobulin heavy chain *IgH*^{VH10tar} allele [49] were a gift from Professor Robert Brink. *Rag1*$^{-/-}$ mice were obtained from the Jackson Laboratory. The cEμ$^{\Delta/\Delta}$ and μA$^{\Delta/\Delta}$ mice were generated by the MAGEC laboratory (Walter and Eliza Hall Institute of Medical Research) as previously described [50] on a C57BL/6J background. To generate cEμ$^{\Delta}$ mice, 20 ng/μl of Cas9 mRNA, 10 ng/μl of sgRNA (GTTGAGGATTCAGCCGAAAC and ATGTTGAGTTGGAGTCAAGA) and 40 ng/μl of oligo donor (CAAGCTAAAATTAAAAGGTTGAACTCAATAAGTTAAAAGAGGACCTCTCCAGTTTCGGCTCAACTCAACATTGCTCAATTCATTTAAAAATATTTGAAACTTAATTTATTATTGTTAAAA) were injected into the cytoplasm of fertilized one-cell stage embryos. To generate μA$^{\Delta}$ mice, 20 ng/μl of Cas9 mRNA, 10 ng/μl of sgRNA

(GAACACCTGCAGCAGCTGGC) and 40 ng/µl of oligo donor (GCTACAAGTTTACCTAGTGGTTTTATTTTCCCTTCCCCAAATAGCCTTGCCACAT GACCTGCCAGCTGCTGCAGGTGTTCTGGTTCTGATCGGCCATCTTGACTCCAACT CAACATTGCT) were injected into the cytoplasm of fertilized one-cell stage embryos. Twenty-four hours later, two-cell stage embryos were transferred into the oviducts of pseudo-pregnant female mice. Viable offspring were genotyped by next-generation sequencing. Mice were analysed from 5 to 14 weeks of age. Male and female mice were used. Experimental procedures were approved by the Walter and Eliza Hall Institute of Medical Research Animal Ethics Committee.

Primary cell culture. B-cell progenitors were obtained from bone marrow that was lineage depleted using biotinylated Ter119, Mac1, Gr1, CD3, CD4, and CD8 antibodies, anti-biotin microbeads and LS columns (Miltenyi Biotec) and cultured on OP9 stromal cells in Iscove's Modified Dulbecco's Medium (Gibco, Invitrogen) supplemented with 10% (v/v) foetal calf serum (Gibco, Invitrogen), 50µM β-mercaptoethanol as well as murine interleukin-7 (10ng/mL) at 37°C in 10% CO_2 for 7 days. Splenic B-cells were purified by negative selection using a B-cell isolation kit (Miltenyi Biotec) as described [51] and purity was confirmed by flow cytometry prior to labelling with Cell Trace Violet (CTV; Life technologies) as per manufacturer instructions. Labelled cells were seeded at $5x10^4$ cells per well and cultured for 90 hours.

METHOD DETAILS

Haematology. Blood was collected into tubes containing EDTA (Sarstedt) and analysed on an Advia 2120 analyser (Bayer).

Flow Cytometry. Single-cell suspensions from bone marrow, lymph node or spleen were prepared in balanced salt solution (BSS-CS: 0.15M NaCl, 4mM KCl, 2mM $CaCl_2$, 1mM $MgSO_4$, 1mM KH_2PO_4, 0.8mM K_2HPO_4, and 15mM HEPES supplemented with 2% [vol/vol] bovine calf serum). Analysis of blood was performed after erythrocyte lysis in buffered 156mM NH_4Cl. Staining was performed using biotinylated or fluorochrome-conjugated antibodies specific for murine antigens Ter119 (Ly-76), CD41 (MWReg30), Gr1 (Ly6G and Ly6C), Mac1 (CD11b), NK1.1, CD11c (N418), CD45R/B220 (RA3-6B2), CD19 (1D3), CD3 (17A2), CD4 (GK1.5), CD8a (53.6.7), Sca1 (Ly6A/E, D7), cKit (CD117, ACK4 or 2B8), CD150 (TC15-12F12.2), CD105 (MJ7/18), CD16/32 (24G2), CD127 (A7R34), CD135 (A2F10), Ly6D (49-H4), CD21/CD35 (7G6), CD23 (B3B4), CD93 (AA4.1), CD24 (M1/69), CD43 (S7), CD45.2 (S450-15-2), CD45.1 (A20), IgM^b (AF6-78), IgD (11-26c.2a), CD138 (281.2), IgG1 (X56), CD25 (3C7), CD44 (IM7). Secondary staining used streptavidin PE-Texas-Red (Invitrogen). FACS-Gal analysis was performed using warm hypotonic loading of fluorescein di β-D-galactopyranoside (Molecular Probes) on single cells as described [52] followed by immunophenotyping using relevant surface antigens as defined in ***Table S1***. Cells were analyzed using a LSR II or FACS Canto flow cytometer (Becton Dickinson) or sorted using a FACSAria II (Becton Dickinson) flow cytometer after antibody staining and lineage selection or depletion using anti-biotin beads and LS columns (Miltenyi Biotec). Data was analysed using FlowJo software (Version 8.8.7, Tree Star).

Genomic PCR. Genomic DNA was extracted using DirectPCR lysis reagent (Viagen) with proteinase K (Sigma-Aldrich) or the DNeasy minikit (Qiagen). 1μL of supernatant from murine tail samples lysed in 200μL or 50-100ng of genomic DNA were used for each reaction. Conditional *Erg* genomic deletion was detected using primers designed to detect the wild-type,

floxed or deleted *Erg* alleles (***Table S3***). Degenerate PCR primers for detection of genomic recombination across distal V_H558 or proximal V_H7183, V_HQ52 regions, the D_H region or Mu0 regions to J3 segments [53], and Vκ [54] were used as described, as were primers to detect TCR VβJ recombination [55, 56] and the *IgH*^{VH10tar} allele [57] ***(Table S3)*** [58] [59, 60, 61]. PCR products were separated by agarose gel electrophoresis and visualized with ethidium bromide staining. For quantitative genomic PCR using SYBR green (Life technologies), primers spanning individual *Erg* exons were used as described [19] and relative quantitation was performed using the $2^{-\Delta\Delta CT}$ method [62].

Splenic B-cell culture. Splenic B-cells were purified and purity was confirmed by flow cytometry prior to labelling with Cell Trace Violet (CTV; Life technologies) as per manufacturer instructions. Labelled cells were seeded at $5x10^4$ cells per well and cultured for 90 hours with either AffiniPure F(ab')$_2$ Fragment Goat Anti-Mouse IgM μ Chain Specific (20mg/ml; Jackson Immunoresearch), CD40L (produced in-house as described [63]) supplemented with IL4 (10ng/ml; R&D systems) and IL5 (5ng/ml; R&D systems) to assess T-cell dependent responses, or LPS (25mg/ml; Difco) to assess T-cell independent responses, and analysed by flow cytometry.

RNA-seq of primary B-cell progenitor samples. Total RNA was extracted using the RNeasy Plus minikit (Qiagen) from bone marrow B-lymphoid populations sorted independently from two *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ and *Rag1Cre*$^{+/+}$*;Erg*$^{fl/fl}$ mice at 7-10 weeks of age. Sequencing was performed on an Illumina Hi-Seq 2500 to generate 100bp paired-end reads. Two biological replicates were sequenced for each mouse strain and B-cell development stage. Adapter sequences were removed using Trimgalore (https://github.com/FelixKrueger/TrimGalore). Reads were aligned to the mm10 mouse genome using STAR [64]. Genewise counts were

obtained using featureCounts [65] with Rsubread's built-in Entrez Gene annotation [66]. Downstream analysis as conducted using edgeR 3.22.5 [67]. For each B-cell stage, genes were filtered as non-expressed if they were assigned 0.5 counts per million mapped reads (CPM) in fewer than two libraries. Library sizes were TMM normalized and differential expression was assessed using quasi-likelihood F-tests [68]. Genes were called differentially expressed if they achieved a false discovery rate of 0.05 (***Table S4***). For plotting purposes, counts were converted to Fragments Per Kilobase of transcript per Million mapped reads (FPKM) using edgeR's rpkm function. These data have been deposited in Gene Expression Omnibus database (accession number GSE132854).

Analysis of publicly available RNA-seq datasets. FASTQ files containing RNA-seq profiles of B-cell progenitor cells were downloaded from GEO for Ebf1$^{\Delta/\Delta}$ (GSM2879293, GSM2879294, GSM2879295), Pax5$^{\Delta/\Delta}$ (GSM2879296, GSM2879297, GSM2879298) and wildtype mice (GSM2879299, GSM2879300, GSM2879301). Reads were aligned to the mm10 genome using Rsubread's align function and read counts were summarized at the gene level as for the primary B-cell samples [66]. Genes were filtered from downstream analysis using edgeR's filterByExpr function and library sizes were TMM normalized. Counts were transformed to log2-CPM and the mean-variance relationship estimated using the *voom* function in limma [69]. Heatmaps were generated using heatmap.2 function in gplots. Genes were tested for differential expression using linear modelling in limma 3.38.2 [70]. Gene set testing was performed using *camera* [45] and barcode plots were generated with limma.

Chromatin Immunoprecipitation (ChIP). Chromatin immunoprecipitation was performed on 2x10^7 cultured ProB cells or primary *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ thymocytes as a negative control

for Erg binding. Cells were cross-linked with 1% formaldehyde for 15 min at room temperature, terminated by the addition of 0.125M glycine. Cells were then lysed in 1% SDS, 10mM EDTA, 50mM Tris-HCl, pH8.0, and protease inhibitors. Lysates were sonicated in a Covaris ultrasonicator to achieve a mean DNA fragment size of 500 bp. Immunoprecipitation was performed using 10µg of antibodies for a minimum of 12h at 4°C in modified RIPA buffer (1% Triton X-100, 0.1% deoxycholate, 90mM NaCl, 10mM Tris-HCl, pH8.0 and protease inhibitors). An equal volume of protein A and G magnetic beads (Life Technologies) were used to bind the antibody and associated chromatin. Reverse crosslinking of DNA was performed at 65°C overnight with RNaseA digestion followed by DNA purification using QIAquick PCR purification kits (Qiagen). Immunoprecipitated DNA was analysed on an Applied Biosystems StepOnePlus System with SYBR green reagents for iEµ µA and intergenic negative control regions using specific primers as detailed in ***Table S2***. Relative ChIP PCR enrichment of the iEµ µA containing region in ProB cells compared to *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ thymocytes was performed and normalized to the intergenic negative control region using the $2^{-\Delta\Delta CT}$ method [62].

ChIP-seq. For sequencing analysis of immunoprecipitated DNA, DNA was quantified using the Qubit dsDNA HS Assay (Life Technologies). Library preparations were performed using the standard ThruPLEXTM-FD Prep Kit protocol (Rubicon Genomics) and size selected for 200–400bp fragments using Pippen Prep (Sage Science Inc.). Fragment sizes were confirmed using either the High Sensitivity DNA assay or the DNA 1000 kit and 2100 bioanalyzer (Agilent Technologies). Libraries were quantified with qPCR, normalized and pooled to 2nM before sequencing with single-end 75bp reads using standard protocols on the NextSeq (Illumina). DNA reads were adapter trimmed using Trimmomatic [71] and aligned to the GRCm38/mm10 build of the Mus musculus genome using the BWA aligner [72]. Peaks were called using MACS2 [73] with default parameters to identify peaks using C17 antibody for Erg

binding with *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ thymocytes as a negative control to filter peaks not due to Erg binding, and were annotated to closest (peak start within 10kb from TSS) and overlapping genes using the R/Bioconductor package ChIPpeakAnno [74] (***Table S5***). These data have been deposited in Gene Expression Omnibus database (accession number GSE132853). Publicly available FASTQ files for Ebf1 (GSM1296532, GSM1296537), Pax5 (GSM932924), H3K4me3 (GSM2255547) and H3K27Ac (GSM2255552) ChIP-seq experiments were aligned to the mm10 mouse reference genome (GRCm38, December 2011) using Rsubread [75]. Peak-calling was performed using MACS2 [73] against input FASTQ files (GSM1296537, GSM1145867). Coordinates for Pu.1, Pax5, Irf4, YY1, Rad21 and CTCF binding were as published [76], while coordinates from annotated immunoglobulin heavy chains were obtained from Ensemble/Biomart (accessed 6th March 2017) and coordinates for 3' regulatory region (3'RR) hypersensitivity regions, 3'αE, iEμ, were as published [77] [78] [79] [14].

ATAC-seq analysis. ATAC-seq [80] was performed on sorted pre-proB, proB and preB populations. Briefly, 5x10^4 nuclei were fragmented by sonication for 30 minutes at 37^{o}C and the DNA purified prior to amplification with indexing primers (HiFi Ready Mix, Kapa Biosciences) for 13 PCR cycles followed by quality assessment by Bioanalyser. High quality libraries were size selected (150 – 700 base pairs) and sequenced using a high output paired end 75 base pair kit on the Nextseq 500 (Illumina) to a minimum of 50 million reads. ATAC-seq reads were aligned to mm10 genome using Bowtie2 [81] (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml accessed 6th March 2017). Peak calling was performed using MACS2 [73]. Intersections of genetic coordinates were performed using Bedtools (http://bedtools.readthedocs.io/en/latest/ accessed 6th March 2017). Heatmaps of unique peaks were generated using pHeatmap in R. These data have been deposited in Gene Expression Omnibus database (accession number GSE132852).

Visualisation of RNA-seq, ChIP-seq and ATASeq data

RNA-seq, ChIP-seq and ATAC-seq files were converted to BigWig files using deepTools (version 2) [82] and uploaded to Cyverse (www.cyverse.org) for visualization in UCSC Genome Browser [83] (genome.ucsc.edu).

Gene Network Analysis. All Ebf1, Pax5 and Erg ChIP-seq peaks mapping to differentially expressed genes in *Rag1Cre*$^{T/+}$*;Erg*$^{\Delta/\Delta}$ pre-proB cells within 10Kb of the transcriptional start sites (TSS) were identified. Peaks inside the gene body were annotated as "proximal targets", peaks overlapping the TSS were labelled as promoter regulated targets, peaks less than 3kb upstream or downstream of the TSS were labelled as putative promoter regulated targets, peaks more than 3kb upstream or downstream TSS were labelled as putative distal targets. Gene Ontogeny (GO) annotation of differentially expressed genes was performed and underwent expert manual curation. The network was constructed using [84] CRAN package, and exported to Cytoscape [85] for customization using RCy3 [86] R/Bioconductor package.

Hi-C Analysis. *In situ* Hi-C was performed as previously described [87]. The data preprocessing and analysis was performed as previously described with changes in parameters [13]. In brief, primary immune cell libraries were generated in biological duplicates for each genotype. An Illumina NextSeq 500 was used to sequence libraries with 80bp paired-end reads to produce libraries of sizes between 42 million and 100 million valid read pairs. Each sample was aligned to the mm10 genome using the *diffHic* package v1.14.0 [88] which utilizes cutadapt v0.9.5 [89] and bowtie2 v2.2.5 [81] for alignment. The resultant BAM file was sorted by read name, the FixMateInformation command from the Picard suite v1.117 (https://broadinstitute.github.io/picard/) was applied, duplicate reads were marked and then re-

sorted by name. Read pairs were determined to be dangling ends and removed if the pairs of inward-facing reads or outward-facing reads on the same chromosome were separated by less than 1000 bp for inward-facing reads and 6000 bp for outward-facing reads. Read pairs with fragment sizes above 1000 bp were removed. An estimate of alignment error was obtained by comparing the mapping location of the 3' segment of each chimeric read with that of the 5' segment of its mate. A mapping error was determined to be present if the two segments were not inward-facing and separated by less than 1000 bp, and around 1-2% were estimated to have errors. Differential interactions (DIs) between the three different groups were detected using the *diffHic* package [88]. Read pairs were counted into 100 kbp bin pairs. Bins were discarded if found on sex chromosomes, contained a count of less than 10, contained blacklisted genomic regions as defined by ENCODE for mm10 [90] or were within a centromeric or telomeric region. Filtering of bin-pairs was performed using the filterDirect function, where bin pairs were only retained if they had average interaction intensities more than 5-fold higher than the background ligation frequency. The ligation frequency was estimated from the inter-chromosomal bin pairs from a 500 kbp bin-pair count matrix. The counts were normalized between libraries using a loess-based approach. Tests for DIs were performed using the quasi-likelihood (QL) framework [68] of the edgeR package. The design matrix was constructed using a one-way layout that specified the cell group to which each library belonged and the mouse sex. A mean-dependent trend was fitted to the negative binomial dispersions with the estimateDisp function. A generalized linear model (GLM) was fitted to the counts for each bin pair [91], and the QL dispersion was estimated from the GLM deviance with the glmQLFit function. The QL dispersions were then squeezed toward a second mean-dependent trend, using a robust empirical Bayes strategy [92]. A p-value was computed against the null hypothesis for each bin pair using the QL F-test. P-values were adjusted for multiple testing using the Benjamini-Hochberg method. A DI was defined as a bin pair with a false discovery rate (FDR) below 5%.

DIs adjacent in the interaction space were aggregated into clusters using the diClusters function to produce clustered DIs. DIs were merged into a cluster if they overlapped in the interaction space, to a maximum cluster size of 1 Mbp. The significance threshold for each bin pair was defined such that the cluster-level FDR was controlled at 5%. Cluster statistics were computed using the *csaw* package v1.16.0 [93]. Overlaps between unclustered bin pairs and genomic intervals were performed using the InteractionSet package [94]. Plaid plots were constructed using the contact matrices and the plotHic function from the *Sushi R* package [95]. The color palette was inferno from the *viridis* package [96] and the range of color intensities in each plot was scaled according to the library size of the sample. The plotBedpe function of the *Sushi* package was used to plot the unclustered DIs as arcs where the z-score shown on the vertical access was calculated as $-\log_{10}(p\text{-value})$. These data have been deposited in Gene Expression Omnibus database (accession number GSE133246).

Fluorescence In Situ Hybridisation. Cultured B-cell progenitors were resuspended in hypotonic 0.075M KCl solution and warmed to 37ºC for 20 minutes. Cells were pelleted and resuspended in 3:1 (vol/vol) methanol:glacial acetic acid fixative. Fixed cells were dropped onto coated Shandon™ polysine slides (ThermoFisher Scientific) and air dried. The cells were hybridized with FISH probes (Creative Bioarray) at 37ºC for 16 hours beneath a coverslip sealed with Fixogum (Marabu) after denaturation at 73ºC for 5 minutes. Cells were washed at 73ºC in 0.4x SSC/0.3%NP_{40} for 2 minutes followed by 2x SSC/0.1%NP_{40} for less than 1 minute at room temperature and air dried in the dark and cover slipped. Images of nuclei were captured on an inverted Zeiss LSM 880 confocal using a 63x/1.4 NA oil immersion objective. Z-stacks of images were then captured using the lambda scan mode, a 405 and a multi-band pass beam splitter (488/561/633). The following laser lines were used: 405, 488, 561 and 633 nm. Spectral data was captured at 8 nm intervals. In all cases, images were set up with a pixel size of 70 nm

and an interval of 150 nm for z-stacks. Single dye controls using the same configuration were captured and spectra imported for spectral unmixing using the Zen software (Zen 2.3, Zeiss Microscopy). Unmixed data was then deconvolved using the batch express tool in Huygens professional software (Scientific Volume Imaging). Images were analysed using TANGO software [97] after linear deconvolution. Nuclear boundaries were extracted in TANGO using the background nuclear signal in the Aqua channel. A 3D median filter was applied and the 3D image projected with maximum 2D image projection for nuclei detection using the Triangle method for automated thresholding in ImageJ [98]. Binary image holes were filled and a 2D procedure implemented to separate touching nuclei using ImageJ 2D watershed implementation. The 2D boundaries of the detected nuclei were expanded in 3D and inside each 3D delimited region, Triangle thresholding was applied to detect the nuclear boundary in the 3D space. Acquired images from immunofluorescent probes were first filtered using 3D median and 3D tophat filter to enhance spot-like structures followed by application of the "spotSegmenter" TANGO plugin with only the best 4 spots having the brightest intensity kept for analysis. The spots identified by TANGO were manually verified against the original immunofluorescent image to identify and record the correct distance computed by TANGO between the aqua and 5-Rox immunofluorescent probes for both *Igh* alleles within a nucleus.

Statistical Analysis. Student's unpaired two-tailed t-tests were used using GraphPad Prism (GraphPad Software), unless otherwise specified. Unless otherwise stated, a P value of < 0.05 was considered significant.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Erg	Santa Cruz Biotechnology	sc-354
Anti-Ebf1	Abcam	ab108369
Anti-Pax5	In-house	Clone: 1H9
Ter119	BD Biosciences	Clone : Ly-76
CD41	BD Biosciences	Clone : MWReg30
Gr1	BD Biosciences	Clone :Ly6G &Ly6C
Mac1	BD Biosciences	Clone : CD11b
NK1.1	BD Biosciences	Clone : N418
CD45R/B220	BD Biosciences	Clone : RA3-6B2
CD19	BD Biosciences	Clone : 1D3
CD3	BD Biosciences	Clone : 17A2
CD4	BD Biosciences	Clone : CK1.5
CD8a	BD Biosciences	Clone : 53.6.7
Sca1 (Ly6A/E)	BD Biosciences	Clone : D7
cKit (CD117)	BD Biosciences	Clone : 2B8 or ACK4
CD150	Biolegend	Clone : TC15-12F12.2
CD105	BD Biosciences	Clone : MJ7/18
CD16/32	BD Biosciences	Clone ; 24G2
CD127	eBioscience	Clone : A7R34
CD135	Biolegend	Clone : A2F10
Ly6D	BD Biosciences	Clone : 49-H4
CD21/35	BD Biosciences	Clone : 7G6
CD23	BD Biosciences	Clone : B3B4
CD93	BD Biosciences	Clone : AA4.1
CD24	BD Biosciences	Clone : M1/69
CD43	BD Biosciences	Clone : S7
CD45.2	BD Biosciences	Clone : S450-15-2
CD45.1	BD Biosciences	Clone : A20
IgM^b	BD Biosciences	Clone : AF6-78
IgD	BD Biosciences	Clone : 11-26c.2a
CD138	BD Biosciences	Clone : 281.2
IgG1	BD Biosciences	Clone : X56
CD25	BD Biosciences	Clone : 3C7
CD44	BD Biosciences	Clone : IM7
Bacterial and Virus Strains		
Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
Fluorescein di β galactopyranoside	Molecular Probes/Invitrogen	Catalog # : F1179
Recombinant Murine Interleukin-7	Peprotech	Catalog # : 217-17
Recombinant Murine Interleukin-4	R&D Systems	Catalog #: 404-ML-010
Recombinant Murine Interleukin-5	R&D Systems	Catalog # : 405-ML-005
Cell Trace Violet	Life Technologies	Catalog # : C34557
Lipopolysacchride	Difco	
AffiniPure $F(ab')_2$ Fragment Goat Anti-Mouse IgM μ chain specific	Jackson Immunoresearch	Catalog # : 715-006-020

CD40L	In-house	N/A
Critical Commercial Assays		
Deposited Data		
Sequence data and analysis related to this paper	This paper	https://www.ncbi.nlm.nih.gov/geo/
RNA-seq	This paper	GSE132854
ChIP-seq	This paper	GSE132853
ATAC-seq	This paper	GSE132852
Hi-C PENDING	This paper	GSE133246
ChIP-seq Ebf1	[99]	GSM1296532, GSM1296537
ChIP-seq Prob-Rag2_Input	[99]	GSM1296537
ChIP-seq Pax5	[100]	GSM932924
ChIP-seq Prob_Rag2_Input_2	[100]	GSM1145867
ChIP-seq H3K4me3	[101]	GSM2255547
ChIP-seq H3K27ac	[101]	GSM2255552
RNA-seq Ebf1 knockout	[102]	GSM2879293, GSM2879294, GSM2879295
RNA-seq Pax5 knockout	[102]	GSM2879296, GSM2879297, GSM2879298
RNA-seq wild-type	[102]	GSM2879299, GSM2879300, GSM2879301
Experimental Models: Cell Lines		
Experimental Models: Organisms/Strains		
Mouse: $Flpe^{T/+}$	[47] Susan Dymecki, Harvard University	N/A
Mouse: $Rag1Cre^{T/+}$	[48] Terry Robberts, University of Leicester	N/A
Mouse: $IgH^{VH10tar}$	[49] Robert Brink, Garvan Institute of Medical Research	N/A
Oligonucleotides		
WA972	Geneworks	WA972 5'-GGTGAGGTCTCTTCCTGAACC-3' common forward
WA974	Geneworks	WA974 5'-TTGGATCCTCAGAATCTACCG-3' exon 4 reverse
WA1092	Geneworks	WA1092 5'-TTATCCTACCTGCCCCTGGT-3' 3' exon 4 reverse
VH558	Geneworks	VH558 5'-CGAGCTCTCCARCACAGCCTWCATGCARCTCARC-3'

VQ52	Geneworks	VQ52 5'-CGGTACCAGACTGARCATCASCAAGGACAAYTCC-3'
VH7183	Geneworks	VH7183 5'-CGGTACCAAGAASAMCCTGTWCCTGCAAATGASC-3'
J3		5'-GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG-3'
DHL	Geneworks	DHL 5'-GGAATTCGMTTTTTGTSAAGGGATCTACTACTGTG-3'
Mu0	Geneworks	Mu0 5'-CCGCATGCCAAGGCTAGCCTGAAAGATTACC-3'
VκD	Geneworks	VκD 5'-GGCTGCAGSTTCAGTGGCAGTGGRTCWGGRAC-3'
Mar35	Geneworks	Mar35 5'-AACACTGGATAAAGCAGTTTATGCCCTTTC-3'
VH10tarU	Geneworks	VH10tarU 5'-GTCTCTGCAGGTGAGTCCTAACTTCT-3'
VH10tarL	Geneworks	VH10tarL 5'-CAACTATCCCTCCAGCCATAGGAT-3'
iEμ forward	Geneworks	iEμ 5'-TTTCGG/CTGAATCCTCAACT-3' forward
iEμ reverse	Geneworks	iEμ 5'-GGTCATGTGGCAAGGCTATT-3' reverse
Erg negative forward	Geneworks	Erg negative 5'-GGGAAACAACACCCTTCTCA-3' forward
Erg negative reverse	Geneworks	Erg negative 5'-AATGTTGATCCTGCCAATCC-3' reverse
Recombinant DNA		
$Erg^{tm1a(KOMP)Wtsi}$ targeting vector	KOMP Knockout Mouse Project	Project ID: CSD48771
MSCV-mEbf1	In-house	N/A
MSCV-mPax5	In-house	N/A
FISH probe: VhJ558 Distal Region, IgHv1-72, BAC RP23-230L2	Creative Bioarray	N/A
FISH probe: Vh7183 Proximal Region, IgH5-2, BAC RP23-404D8	Creative Bioarray	N/A

Software and Algorithms		
FlowJo	Tree Star	RRID:SCR_008520
R Studio	RStudio, Inc	RRID:SCR_000432
R	R Project for Statistical Computing	RRID:SCR_001905
R packages Rsubread, limma, edgeR, csaw, diffHic, InterationSet, Sushi	Bioconductor	RRID:SCR_001905
MACS2	Liu Lab, Harvard University	RRID:SCR_013291
Bowtie2	[81]	RRID:SCR_005476
deepTools	[82]	RRID:SCR_016366
R packages pheatmap viridis	CRAN	RRID:SCR_003005S
ImageJ	[98]	RRID:SCR_003070
Zen	Zeiss Microscopy	RRID:SCR_013672
Huygens Software	Scientific Volume Imaging	RRID:SCR_014237
Other		

Inventory of Supplemental Information

Supplemental Figures

Figure S1. Representative flow cytometry plots indicating gating strategies for analysis of hematopoietic cell populations. Related to ***Figure 1***

Figure S2. B-lymphopoiesis in *Rag1Cre*$^{T/+}$ mice and T-lymphopoiesis in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice. Related to ***Figure 1.***

Figure S3. In vivo Erg binding to the iEμ enhancer in pre-proB cells. Related to ***Figure 2.***

Figure S4. RNA-seq and Erg, Ebf1 and Pax5 binding and chromatin accessibility at selected gene loci. Related to ***Figure 2, 4, 6.***

Supplemental Tables

Table S1. Immunophenotype of hematopoietic cell populations. Related to ***Figure 1.***

Table S2. Peripheral blood counts of *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice. Related to ***Figure 1.***

Table S3. Primers and PCR reactions. Related to ***Figure 2, 3.***

Table S4. RNA-seq. Differentially expressed genes in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ pre-proB cells, and Ebf1$^{\Delta/\Delta}$ Pax5$^{\Delta/\Delta}$ B-cell progenitors (EXCEL FILE). Related to ***Figure 4.***

Table S5. Erg, Ebf1 and Pax5 ChIP binding coordinates to differentially expressed genes in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ pre-proB cells (EXCEL FILE). Related to ***Figure 6.***

Figure S1. Representative flow cytometry plots indicating gating strategies for analysis of hematopoietic cell populations. **A**. bone marrow LSK cells, **B.** thymus sub-populations, **C.** bone marrow B-lineage cells and **D.** bone marrow myeloid cell populations in $Erg^{KI/+}$ mice. The cell surface markers and definitions of cell populations used are provided in ***Table S1***.

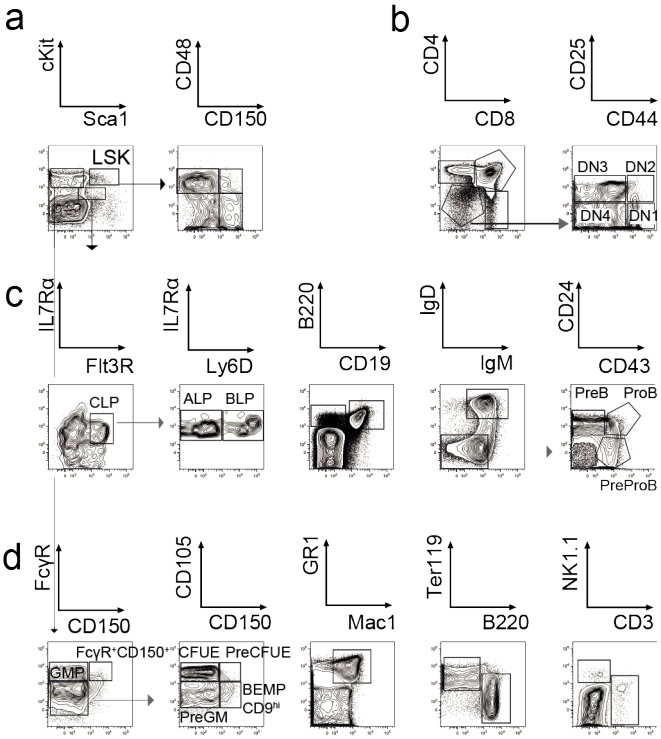
Figure S2. B-lymphopoiesis in $Rag1Cre^{T/+}$ mice and T-lymphopoiesis in $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ mice. **A.** Representative flow cytometry plots (left panels) of $Rag1Cre^{T/+}$ bone marrow and spleen cells. The IgM/IgD profile is from $B220^{+}$ bone marrow cells (top panel), the CD24/CD43 profiles from $B220^{+}IgM^{-}IgD^{-}$ bone marrow cells (middle panel) and the CD21/CD23 profiles from $B220^{+}CD93^{lo}$ spleen cells (bottom panel). Ratio of $Rag1Cre^{+/+}$ and $Rag1Cre^{T/+}$ B-lymphoid cells shown relative to $Rag1Cre^{+/+}$ controls in bone marrow and spleen with standard error of means (right panels). No significant differences were observed between the two genotypes for any population by Student's two-tailed unpaired t-test corrected using Holm's modification for multiple testing ($P_{adj} > 0.19$, n=6 mice per genotype). **B.** Quantitative genomic PCR on DNA from $Erg^{fl/fl}$ and $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ thymocytes using primers spanning individual *Erg* exons [19] demonstrating efficient exon 4 deletion in $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ thymocytes by $2^{-\Delta\Delta CT}$ method normalised to IL2 receptor and Erg exon 1. **C**. Representative flow cytometry plots (left panels) from $Erg^{fl/fl}$ and $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ thymi identifying the specific cell populations indicated with the mean with standard error of the mean of $Erg^{fl/fl}$ (n=7) and $Rag1Cre^{T/+};Erg^{\Delta/\Delta}$ mice (n=8) shown relative to the mean of $Erg^{fl/fl}$ controls (right panel). No significant differences were observed other that in the DN2 population ($P_{adj} < 6.3e\text{-}4$ by Student's two-tailed unpaired t-test corrected using Holm's modification for multiple testing).

Figure S3. In vivo Erg binding to the iEμ enhancer. **A.** Region of the iEu enhancer as

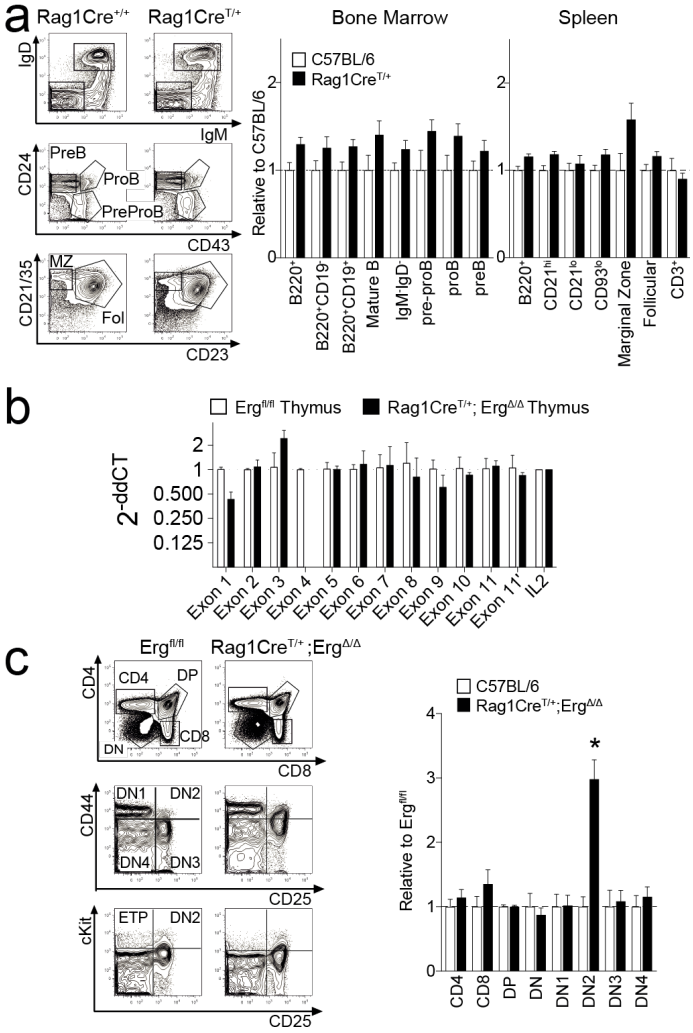
indicated (red bars) with the ATAC-seq and ChIP-seq for Erg, Pu.1, Pax5, Irf4, YY1, Rad21 and CTCF shown. **B.** Erg binding to iEμ containing the μA element by ChIP-PCR showing fold-enrichment in wild type B-cell progenitors relative to *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ thymocytes and normalised to a negative intergenic region control using the $2^{-\Delta\Delta CT}$ method (n=2 biological replicates, * $P<10^{-4}$ by Student's two-tailed unpaired t-test). **C.** Genomic PCR using degenerate primers to IgH locus V_H558, V_H7183, V_HQ52 segments for detection of $V_{H\ to}$ DJ_H (top panel) and D_H to J_H (middle panel) recombination with Mu0 loading controls (bottom panel) in B220$^+$ splenocytes.

Figure S4. RNA-seq and Erg, Ebf1 and Pax5 binding and chromatin accessibility at selected gene loci. **A.** *Igh* locus with representative RNA-seq tracks shown for *Erg*$^{fl/fl}$ and *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ (Erg KO) pre-proB cells, and Erg deficient proB and preB cells in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$mice rescued with a functionally rearranged immunoglobulin heavy chain allele (Rescue proB, Rescue preB). ChIP-seq tracks for Erg (highlighted in pink), *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ thymus cells (Erg KO) to control for sites of non-Erg ChIP binding to DNA, H3K4me3 and H3k27ac shown. Chromatin accessibility by ATAC-seq (blue) in WT and Erg KO pre-proB cells, Erg deficient Rescue proB (yellow) and Rescue preB (orange) cells. **B and C.** RNA-seq gene expression at gene loci, with Erg, Ebf1, and Pax5 binding, H3K4me3 promoter mark, H3K27ac promoter and enhancer mark, and ATAC-seq in *Erg*$^{fl/fl}$ pre-proB cells (PreProB), *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ pre-proB (Erg KO pre-proB), and Erg deficient proB and preB cells in *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$;*IgH*$^{VH10tar/+}$mice rescued with a functionally rearranged immunoglobulin heavy chain allele (Rescue proB, Rescue preB). Solid blue bar: Erg, Ebf1 and/or Pax5 binding promoter. Solid pink bar: Erg, Ebf1 and/or Pax5 binding to enhancer regions. Open blue bar: promoter region with no binding of Erg, Ebf1 or Pax5. **B.** *Erg, Foxo1, Ikzf1* and *Spi1* B-lineage transcription factor loci. **C.** *Rag1* and *Rag2*, *Lig4, Xrcc6* and *Poll* loci.

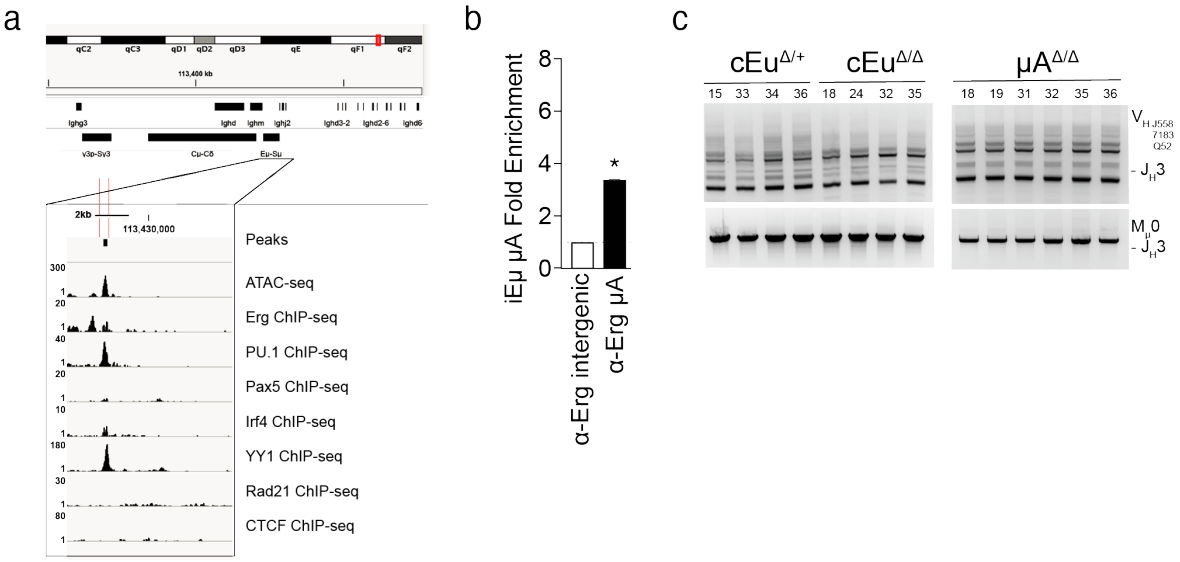
Supplementary Figure 1



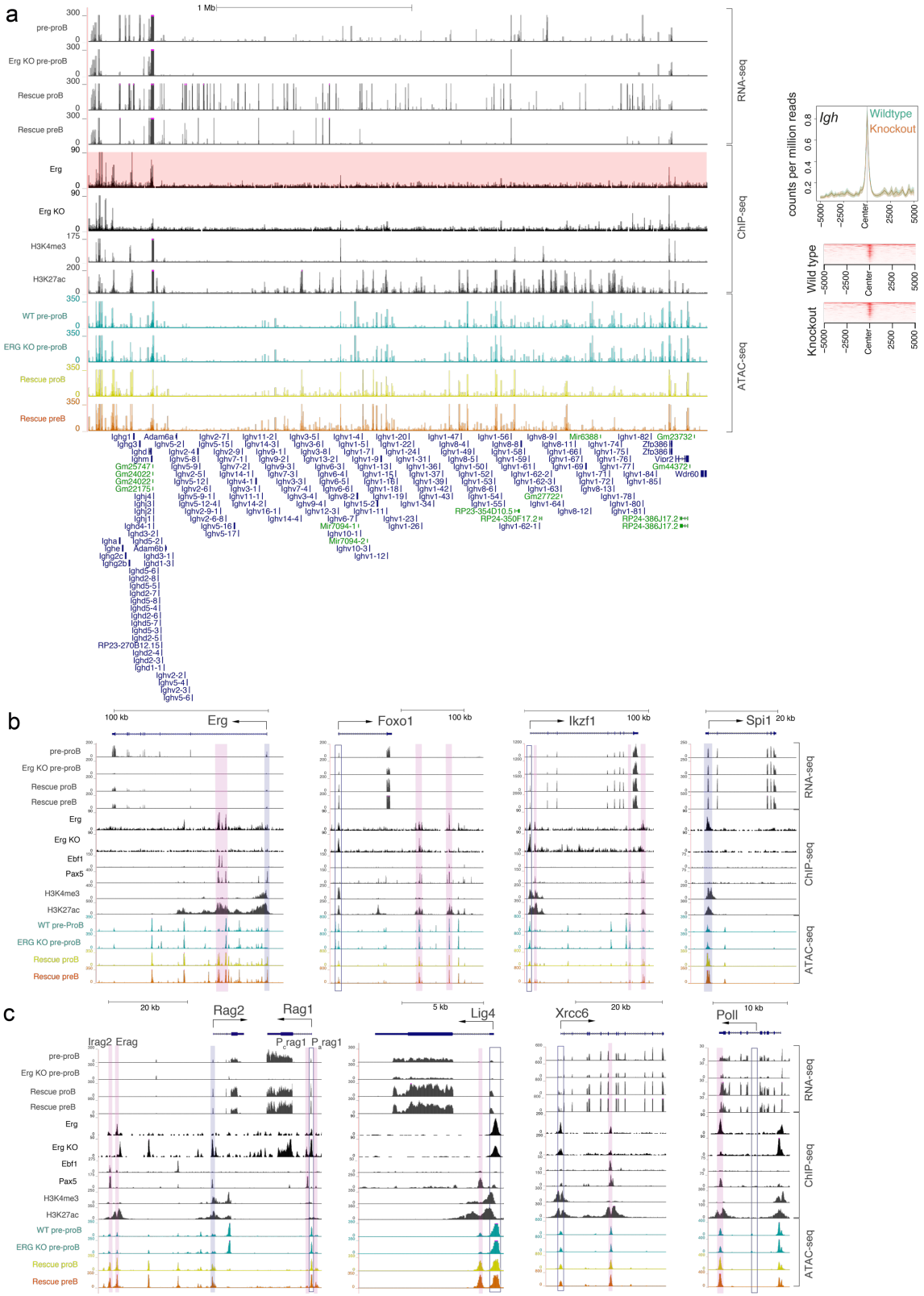
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Tables

Table S1. Immunophenotype of hematopoietic cell populations. See **Figure 1**

Table S2. Peripheral blood counts of *Rag1Cre*$^{T/+}$;*Erg*$^{\Delta/\Delta}$ mice. See **Figure 1**

Table S3. Primers and PCR reactions. See **Figure 2, 3**

Table S4. RNA-seq. Differentially expressed genes in *Rag1Cre*$^{T/+;}$*Erg*$^{\Delta/\Delta}$ pre-proB cells, and Ebf1$^{\Delta/\Delta}$ Pax5$^{\Delta/\Delta}$ B-cell progenitors (EXCEL FILE). See **Figure 4.**

Table S5. Erg, Ebf1 and Pax5 ChIP binding coordinates to differentially expressed genes in *Rag1Cre*$^{T/+;}$*Erg*$^{\Delta/\Delta}$ pre-proB cells (EXCEL FILE). See **Figure 6**

Supplementary Table 1. Immunophenotype of hematopoietic cell populations

	Immunophenotype	Reference
Bone marrow		
LSK	Lineage$^-$ Sca-1$^+$Kit$^+$	103, 104
LT-HSC, CD150$^+$CD48$^-$	Lineage$^-$ Sca-1$^+$Kit$^+$CD150$^+$CD48$^-$	105
ST-HSC, CD150$^+$CD48$^+$	Lineage$^-$ Sca-1$^+$Kit$^+$CD150$^+$CD48$^+$	
MPP, CD150$^-$CD48$^+$	Lineage$^-$ Sca-1$^+$Kit$^+$CD150$^-$CD48$^+$	
PreGM	Lineage$^-$ IL7Rα$^-$cKit$^+$Sca1$^-$CD150$^-$ Endoglin$^-$FcγRII/III$^-$	106, 107, 108
GMP	Lineage$^-$ IL7Rα$^-$cKit$^+$Sca1$^-$CD150$^-$ Endoglin$^-$FcγRII/III$^-$	109
BEMP & CD9^{hi}	Lineage$^-$ IL7Rα$^-$cKit$^+$Sca1$^-$ CD150$^+$Endoglin$^-$FcγRII/III$^-$	107
PreCFUE	Lineage$^-$ IL7Rα$^-$cKit$^+$Sca1$^-$ CD150$^+$Endoglin$^+$FcγRII/III$^-$	108
CFUE	Lineage$^-$ IL7Rα$^-$cKit$^+$Sca1$^-$CD150$^-$ Endoglin$^+$FcγRII/III$^-$	108
CLP	Lineage$^-$ IL7Rα$^+$Flt3$^+$cKit^{lo}Sca1^{lo}	110
ALP	Lineage$^-$ IL7Rα$^+$Flt3$^+$cKit^{lo}Sca1^{lo}Ly6D$^-$	111
BLP	Lineage$^-$ IL7Rα$^+$Flt3$^+$cKit^{lo}Sca1^{lo}Ly6D$^+$	
pre-proB (Hardy Fraction A-to-B)	B220$^+$IgM$^-$IgD$^-$NK1.1$^-$CD11c$^-$CD19$^-$ CD43$^+$CD24$^-$	112, 113, 114, 115, 116, 117
proB (Hardy Fraction C)	B220$^+$IgM$^-$IgD$^-$ NK1.1$^-$CD11c$^-$ CD19$^+$CD43$^+$CD24$^+$	
preB (Hardy Fraction D)	B220$^+$IgM$^-$IgD$^-$ NK1.1$^-$CD11c$^-$ CD19$^+$CD43$^-$CD24$^+$	
Immature B (Hardy Fraction E)	B220$^+$CD19$^+$IgM$^+$IgD$^-$	112
Mature recirculating B (Hardy Fraction F)	B220$^+$CD19$^+$IgM$^+$IgD$^+$	
Thymus		
DP	CD4$^+$CD8$^+$	
DN1	CD4$^-$CD8$^-$CD25$^-$CD44$^+$	118
DN2	CD4$^-$CD8$^-$CD25$^+$CD44$^+$	
DN3	CD4$^-$CD8$^-$CD25$^+$CD44$^-$	
DN4	CD4$^-$CD8$^-$CD25$^-$CD44$^-$	
Spleen		
Marginal Zone	B220$^+$CD19$^+$CD93^{lo}CD21/35^{hi}CD23^{lo}	119, 120
Follicular	B220$^+$ CD19$^+$CD93^{lo}CD21/35^{med}CD23^{hi}	

Supplementary Table 2. Peripheral blood counts in *Rag1Cre$^{T/+}$;Erg$^{\Delta/\Delta}$* mice.

Genotype	RBC x10^{12}/L	Platelets x10^{9}/L	WBC x10^{9}/L	Neutrophil x10^{9}/L	Lymphocyte x10^{9}/L	Monocyte x10^{9}/L	Eosinophil x10^{9}/L
Erg$^{fl/fl}$ (n= 32)	11.03 ± 0.48	1038 ± 192	9.07 ± 1.64	1.00 ± 0.90	7.52 ± 1.66	0.18 ± 0.15	0.21 ± 0.11
Rag1Cre$^{T/+}$;Erg$^{\Delta/\Delta}$ (n=25)	10.73 ± 1.85	1256 ± 204	4.82 ± 1.72 *	1.09 ± 0.82	3.22 ± 1.12 *	0.22 ± 0.17	0.24 ± 0.06

Blood was collected into EDTA and differential cell counts performed using an ADVIA 120 Hematology System. RBC, red blood cells; WBC, white blood cells. * *Padj* $< 10^{-11}$ by Student's two-tailed unpaired t-test corrected for multiple testing by Benjamini-Hochberg procedure to control for false discovery rate.

Supplementary Table 3. Primers and PCR reactions

	Primer 1	Primer 2	Expected sizes
Erg[i]	WA972 5'-GGTGAGGTCTCTTCCTGAACC-3' common forward	WA974 5'-TTGGATCCTCAGAATCTACCG-3' exon 4 reverse WA1092 5'- TTATCCTACCTGCCCCTGGT-3' 3' exon 4 reverse	Wildtype 206bp Targeted 231bp Floxed 355bp
V_H-to-DJ_H[ii]	VH558 5'-CGAGCTCTCCARCACAGCCTWCATGCARCTCARC-3' VQ52 5'-CGGTACCAGACTGARCATCASCAAGGACAAYTCC-3' VH7183 5'-CGGTACCAAGAASAMCCTGTWCCTGCAAATGASC-3' [53]	J3 5'-GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG-3'	See ***Figure 2.***
D_H-to-J_H[iii]	DHL 5'-GGAATTCGMTTTTTGTSAAGGGATCTACTACTGTG-3' [53]	J3 5'-GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG-3'	See ***Figure 2.***
Mu0[ii]	Mu0 5'-CCGCATGCCAAGGCTAGCCTGAAAGATTACC-3' [53]	J3 5'-GTCTAGATTCTCACAAGAGTCCGATAGACCCTGG-3'	Germline 1,259bp
VκD[iv]	VκD 5'-GGCTGCAGSTTCAGTGGCAGTGGRTCWGGRAC-3' [54]	Mar35 5'-AACACTGGATAAAGCAGTTTATGCCCTTTC-3' [121]	See ***Figure 3.***
HEL-IgH[i]	VH10tarU 5'-GTCTCTGCAGGTGAGTCCTAACTTCT-3' [57]	VH10tarL 5'-CAACTATCCCTCCAGCCATAGGAT-3'	Wildtype 865bp Knockin 302bp
IL2[i]	WA735 5'-CTAGGCCACAGAATTGAAAGATCT-3'	WA736 5'-GTAGGTGGAAATTCTAGCATCATCC-3'	
iEμ π/μA[vii]	iEμ 5'-TTTCGG/CTGAATCCTCAACT-3' forward Chr12;113427394 to 113427413	iEμ 5'-GGTCATGTGGCAAGGCTATT-3' reverse Chr12; 113427561 to 113427581	374bp
Erg –ve[vii]	Erg negative 5'-GGGAAACAACACCCTTCTCA-3' forward	Erg negative 5'-AATGTTGATCCTGCCAATCC-3' reverse	749bp

i. 94°C denaturation for 3 minutes, followed by 35 cycles at 94°C for 1 minute, 60°C for 30 seconds, 72°C for 1 minutes, and a final 5 minutes extension at 72°C.
ii. 94°C denaturation for 3 minutes, followed by 35 cycles at 94°C for 1 minute, 60°C for 30 seconds, 72°C for 2 minutes, and a final 5 minutes extension at 72°C.
iii. 94°C denaturation for 3 minutes, followed by 40 cycles of 1 minute at 94°C, 1 minute at 60°C, and 1 minute at 72°C, and a final extension step at 72°C for 10 min [59].
iv. 94°C denaturation for 3 minutes, followed by 27 cycles at 94°C for 1 minute, 60°C for 1 minute, 68°C for 3 minute, and a final 5 minute extension at 68°C [60].
v. 95°C denaturation for 3 minutes, followed by 35 cycles at 95°C for 30 seconds, 58°C for 40 seconds, 72°C for 45 seconds [122].
vi. 94°C denaturation for 5 min, followed by 30 cycles at 94°C for 30 seconds, annealing temperature 30 seconds, 72°C for 1 min, and a final 10 min extension at 72°C.
The annealing temperature was held at 68°C, 65°C and 62°C for five cycles each and 58°C for 15 cycles [61].
vii. 95°C denaturation for 20 seconds, followed by 40 cycles at 95°C for 3 seconds, 60°C for 30 seconds.

References

1. Reynaud, D. *et al.* Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat Immunol* **9**, 927-936 (2008).

2. Medina, K.L. *et al.* Assembling a gene regulatory network for specification of the B cell fate. *Dev Cell* **7**, 607-617 (2004).

3. Bain, G. *et al.* E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* **79**, 885-892 (1994).

4. Mansson, R. *et al.* Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. *Proc Natl Acad Sci U S A* **109**, 21028-21033 (2012).

5. Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376**, 263-267 (1995).

6. Nutt, S.L., Urbanek, P., Rolink, A. & Busslinger, M. 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 (1997).

7. Miyai, T. *et al.* Three-step transcriptional priming that drives the commitment of multipotent progenitors toward B cells. *Genes Dev* **32**, 112-126 (2018).

8. Nutt, S.L. & Kee, B.L. The transcriptional regulation of B cell lineage commitment. *Immunity* **26**, 715-725 (2007).

9. Li, R. *et al.* Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. *Genes Dev* **32**, 96-111 (2018).

10. Lin, Y.C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol* **11**, 635-643 (2010).

11. Boller, S. *et al.* Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. *Immunity* **44**, 527-541 (2016).

12. Decker, T. *et al.* Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis. *Immunity* **30**, 508-520 (2009).

13. Johanson, T.M. *et al.* Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. *Nat Immunol* **19**, 1257-1264 (2018).

14. Ebert, A. *et al.* 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 (2011).

15. Ochiai, K. *et al.* A self-reinforcing regulatory network triggered by limiting IL-7 activates pre-BCR signaling and differentiation. *Nat Immunol* **13**, 300-307 (2012).

16. Zhang, Z. *et al.* Transcription factor Pax5 (BSAP) transactivates the RAG-mediated V(H)-to-DJ(H) rearrangement of immunoglobulin genes. *Nat Immunol* **7**, 616-624 (2006).

17. Ng, A.P. *et al.* Erg is required for self-renewal of hematopoietic stem cells during stress hematopoiesis in mice. *Blood* **118**, 2454-2461 (2011).

18. Ng, A.P. *et al.* Early lineage priming by trisomy of erg leads to myeloproliferation in a down syndrome model. *PLoS Genet* **11**, e1005211 (2015).

19. Loughran, S.J. *et al.* The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat Immunol* **9**, 810-819 (2008).

20. Ng, A.P. *et al.* Trisomy of Erg is required for myeloproliferation in a mouse model of Down syndrome. *Blood* **115**, 3966-3969 (2010).

21. Knudsen, K.J. *et al.* ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. *Genes Dev* **29**, 1915-1929 (2015).

22. Malin, S. *et al.* Role of STAT5 in controlling cell survival and immunoglobulin gene recombination during pro-B cell development. *Nat Immunol* **11**, 171-U197 (2010).

23. Fuxa, M. *et al.* Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* **18**, 411-422 (2004).

24. Kumari, G. & Sen, R. Chromatin Interactions in the Control of Immunoglobulin Heavy Chain Gene Assembly. *Adv Immunol* **128**, 41-92 (2015).

25. Zullo, J.M. *et al.* DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149**, 1474-1487 (2012).

26. Perlot, T. & Alt, F.W. Cis-regulatory elements and epigenetic changes control genomic rearrangements of the IgH locus. *Adv Immunol* **99**, 1-32 (2008).

27. Guo, C. *et al.* Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* **147**, 332-343 (2011).

28. Rivera, R.R., Stuiver, M.H., Steenbergen, R. & Murre, C. Ets proteins: new factors that regulate immunoglobulin heavy-chain gene expression. *Mol Cell Biol* **13**, 7163-7169 (1993).

29. Marquet, M. *et al.* The Emu enhancer region influences H chain expression and B cell fate without impacting IgVH repertoire and immune response in vivo. *J Immunol* **193**, 1171-1183 (2014).

30. Rajewsky, K. Clonal selection and learning in the antibody system. *Nature* **381**, 751-758 (1996).

31. Chang, Y., Bosma, G.C. & Bosma, M.J. Development of B cells in scid mice with immunoglobulin transgenes: implications for the control of V(D)J recombination. *Immunity* **2**, 607-616 (1995).

32. Cook, A.J. *et al.* Reduced switching in SCID B cells is associated with altered somatic mutation of recombined S regions. *J Immunol* **171**, 6556-6564 (2003).

33. Spanopoulou, E. *et al.* Functional immunoglobulin transgenes guide ordered B-cell differentiation in Rag-1-deficient mice. *Genes Dev* **8**, 1030-1042 (1994).

34. Young, F. *et al.* Influence of immunoglobulin heavy- and light-chain expression on B-cell differentiation. *Genes Dev* **8**, 1043-1057 (1994).

35. Rolink, A., Grawunder, U., Winkler, T.H., Karasuyama, H. & Melchers, F. IL-2 receptor alpha chain (CD25, TAC) expression defines a crucial stage in pre-B cell development. *Int Immunol* **6**, 1257-1264 (1994).

36. Nutt, S.L., Heavey, B., Rolink, A.G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401**, 556-562 (1999).

37. Nechanitzky, R. *et al.* Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nat Immunol* **14**, 867-875 (2013).

38. Roessler, S. *et al.* Distinct promoters mediate the regulation of Ebf1 gene expression by interleukin-7 and Pax5. *Mol Cell Biol* **27**, 579-594 (2007).

39. Jin, Z.X. *et al.* Lymphoid enhancer-binding factor-1 binds and activates the recombination-activating gene-2 promoter together with c-Myb and Pax-5 in immature B cells. *J Immunol* **169**, 3783-3792 (2002).

40. Hsu, L.Y. *et al.* A conserved transcriptional enhancer regulates RAG gene expression in developing B cells. *Immunity* **19**, 105-117 (2003).

41. Amin, R.H. & Schlissel, M.S. Foxo1 directly regulates the transcription of recombination-activating genes during B cell development. *Nat Immunol* **9**, 613-622 (2008).

42. Gu, Y. *et al.* Growth retardation and leaky SCID phenotype of Ku70-deficient mice. *Immunity* **7**, 653-665 (1997).

43. Frank, K.M. *et al.* Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* **396**, 173-177 (1998).

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

45. Wu, D. & Smyth, G.K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res* **40**, e133 (2012).

46. Skarnes, W.C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337-342 (2011).

47. Farley, F.W., Soriano, P., Steffen, L.S. & Dymecki, S.M. Widespread recombinase expression using FLPeR (flipper) mice. *Genesis* **28**, 106-110 (2000).

48. McCormack, M.P., Forster, A., Drynan, L., Pannell, R. & Rabbitts, T.H. The LMO2 T-cell oncogene is activated via chromosomal translocations or retroviral insertion during gene therapy but has no mandatory role in normal T-cell development. *Mol Cell Biol* **23**, 9003-9013 (2003).

49. Phan, T.G. *et al.* B cell receptor-independent stimuli trigger immunoglobulin (Ig) class switch recombination and production of IgG autoantibodies by anergic self-reactive B cells. *J Exp Med* **197**, 845-860 (2003).

50. Kueh, A.J. *et al.* An update on using CRISPR/Cas9 in the one-cell stage mouse embryo for generating complex mutant alleles. *Cell Death Differ* **24**, 1821-1822 (2017).

51. Hasbold, J., Corcoran, L.M., Tarlinton, D.M., Tangye, S.G. & Hodgkin, P.D. Evidence from the generation of immunoglobulin G-secreting cells that stochastic mechanisms regulate lymphocyte differentiation. *Nat Immunol* **5**, 55-63 (2004).

52. Berger, C.N., Tan, S.S. & Sturm, K.S. Simultaneous detection of beta-galactosidase activity and surface antigen expression in viable haematopoietic cells. *Cytometry* **17**, 216-223 (1994).

53. Schlissel, M.S., Corcoran, L.M. & Baltimore, D. Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. *J Exp Med* **173**, 711-720 (1991).

54. Schlissel, M.S. & Baltimore, D. Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. *Cell* **58**, 1001-1007 (1989).

55. Anderson, S.J., Abraham, K.M., Nakayama, T., Singer, A. & Perlmutter, R.M. Inhibition of T-cell receptor beta-chain gene rearrangement by overexpression of the non-receptor protein tyrosine kinase p56lck. *EMBO J* **11**, 4877-4886 (1992).

56. Wojciechowski, J., Lai, A., Kondo, M. & Zhuang, Y. E2A and HEB are required to block thymocyte proliferation prior to pre-TCR expression. *J Immunol* **178**, 5717-5726 (2007).

57. Jurado, S. *et al.* The Zinc-finger protein ASCIZ regulates B cell development via DYNLL1 and Bim. *J Exp Med* **209**, 1629-1639 (2012).

58. Abe, K. *et al.* Novel lymphocyte-independent mechanisms to initiate inflammatory arthritis via bone marrow-derived cells of Ali18 mutant mice. *Rheumatology (Oxford)* **47**, 292-300 (2008).

59. Angelin-Duclos, C. & Calame, K. Evidence that immunoglobulin VH-DJ recombination does not require germ line transcription of the recombining variable gene segment. *Mol Cell Biol* **18**, 6253-6264 (1998).

60. Xiang, Y. & Garrard, W.T. The Downstream Transcriptional Enhancer, Ed, positively regulates mouse Ig kappa gene expression and somatic hypermutation. *J Immunol* **180**, 6725-6732 (2008).

61. Jackson, A., Kondilis, H.D., Khor, B., Sleckman, B.P. & Krangel, M.S. Regulation of T cell receptor beta allelic exclusion at a level beyond accessibility. *Nat Immunol* **6**, 189-197 (2005).

62. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408 (2001).

63. Kehry, M.R. & Castle, B.E. Regulation of CD40 ligand expression and use of recombinant CD40 ligand for studying B cell growth and differentiation. *Semin Immunol* **6**, 287-294 (1994).

64. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

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

66. Liao, Y., Smyth, G.K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* (2019).

67. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

68. Chen, Y., Lun, A.T. & Smyth, G.K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5**, 1438 (2016).

69. Law, C.W., Chen, Y., Shi, W. & Smyth, G.K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29 (2014).

70. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

71. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

72. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2; 2013.

73. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

74. Zhu, L.J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).

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

76. Bolland, D.J. *et al.* Two Mutually Exclusive Local Chromatin States Drive Efficient V(D)J Recombination. *Cell Rep* **15**, 2475-2487 (2016).

77. Thomas-Claudepierre, A.S. *et al.* Mediator facilitates transcriptional activation and dynamic long-range contacts at the IgH locus during class switch recombination. *J Exp Med* **213**, 303-312 (2016).

78. D'Addabbo, P., Scascitelli, M., Giambra, V., Rocchi, M. & Frezza, D. Position and sequence conservation in Amniota of polymorphic enhancer HS1.2 within the palindrome of IgH 3'Regulatory Region. *BMC Evol Biol* **11**, 71 (2011).

79. Chakraborty, T. *et al.* A 220-nucleotide deletion of the intronic enhancer reveals an epigenetic hierarchy in immunoglobulin heavy chain locus activation. *J Exp Med* **206**, 1019-1027 (2009).

80. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).

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

82. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014).

83. Raney, B.J. *et al.* Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* **30**, 1003-1005 (2014).

84. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Systems*, 1695 (2006).

85. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).

86. Ono, K., Muetze, T., Kolishovski, G., Shannon, P. & Demchak, B. CyREST: Turbocharging Cytoscape Access for External Tools via a RESTful API. *F1000Res* **4**, 478 (2015).

87. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).

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

89. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).

90. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

91. McCarthy, D.J., Chen, Y. & Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297 (2012).

92. Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S. & Smyth, G.K. Robust Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect Differential Expression. *Ann Appl Stat* **10**, 946-963 (2016).

93. Lun, A.T. & Smyth, G.K. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res* **44**, e45 (2016).

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

95. Phanstiel, D.H., Boyle, A.P., Araya, C.L. & Snyder, M.P. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* **30**, 2808-2810 (2014).

96. Garnier, S. viridis: Default Color Maps from 'matplotlib'. (2018).

97. Ollion, J., Cochennec, J., Loll, F., Escude, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840-1841 (2013).

98. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).

99. Schwickert, T.A. *et al.* Stage-specific control of early B cell development by the transcription factor Ikaros. *Nat Immunol* **15**, 283-293 (2014).

100. Revilla, I.D.R. *et al.* The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO J* **31**, 3130-3146 (2012).

101. Smeenk, L. *et al.* Molecular role of the PAX5-ETV6 oncoprotein in promoting B-cell acute lymphoblastic leukemia. *EMBO J* **36**, 718-735 (2017).

102. Jensen, C.T. *et al.* Dissection of progenitor compartments resolves developmental trajectories in B-lymphopoiesis. *J Exp Med* **215**, 1947-1963 (2018).

103. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242-245 (1996).

104. Okada, S. *et al.* In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. *Blood* **80**, 3044-3050 (1992).

105. Kiel, M.J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109-1121 (2005).

106. Boiers, C. *et al.* Expression and role of FLT3 in regulation of the earliest stage of normal granulocyte-monocyte progenitor development. *Blood* **115**, 5061-5068 (2010).

107. Ng, A.P. *et al.* Characterization of thrombopoietin (TPO)-responsive progenitor cells in adult mouse bone marrow with in vivo megakaryocyte and erythroid potential. *Proc Natl Acad Sci U S A* **109**, 2364-2369 (2012).

108. Pronk, C.J. *et al.* Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1**, 428-442 (2007).

109. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I.L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193-197 (2000).

110. Kondo, M., Weissman, I.L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661-672 (1997).

111. Inlay, M.A.a.B.D.a.S.D.a.S.T.a.S.J.a.K.H.a.P.S.K.a.D.D.L.a.W.I.L. Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes \& Development* **23**, 2376--2381 (2009).

112. Hardy, R.R., Carmack, C.E., Shinton, S.A., Kemp, J.D. & Hayakawa, K. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J Exp Med* **173**, 1213-1225 (1991).

113. Blasius, A.L., Barchet, W., Cella, M. & Colonna, M. Development and function of murine B220+CD11c+NK1.1+ cells identify them as a subset of NK cells. *J Exp Med* **204**, 2561-2568 (2007).

114. Asselin-Paturel, C. *et al.* Mouse type I IFN-producing cells are immature APCs with plasmacytoid morphology. *Nat Immunol* **2**, 1144-1150 (2001).

115. Chan, C.W. *et al.* Interferon-producing killer dendritic cells provide a link between innate and adaptive immunity. *Nat Med* **12**, 207-213 (2006).

116. Taieb, J. *et al.* A novel dendritic cell subset involved in tumor immunosurveillance. *Nat Med* **12**, 214-219 (2006).

117. Rumfelt, L.L., Zhou, Y., Rowley, B.M., Shinton, S.A. & Hardy, R.R. Lineage specification and plasticity in CD19- early B cell precursors. *J Exp Med* **203**, 675-687 (2006).

118. Godfrey, D.I., Kennedy, J., Suda, T. & Zlotnik, A. A developmental pathway involving four phenotypically and functionally distinct subsets of CD3-CD4-CD8- triple-negative adult mouse thymocytes defined by CD44 and CD25 expression. *J Immunol* **150**, 4244-4252 (1993).

119. Allman, D. *et al.* Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. *J Immunol* **167**, 6834-6840 (2001).

120. Loder, F. *et al.* B cell development in the spleen takes place in discrete steps and is determined by the quality of B cell receptor-derived signals. *J Exp Med* **190**, 75-89 (1999).

121. Inlay, M., Alt, F.W., Baltimore, D. & Xu, Y. Essential roles of the kappa light chain intronic enhancer and 3' enhancer in kappa rearrangement and demethylation. *Nat Immunol* **3**, 463-468 (2002).

122. LeBleu, V. *et al.* Stem cell therapies benefit Alport syndrome. *J Am Soc Nephrol* **20**, 2359-2370 (2009).