IL-7R signaling activates widespread VH and DH gene usage to drive antibody diversity in bone marrow B cells

Graphical abstract

Highlights
- The IL-7R drives recombination of VH and DH genes in the immunoglobulin heavy chain
- Deletion of the IL-7R impairs usage of all except 3’ VH and flanking DH genes
- IL-7R loss diminishes large-scale VH and DH antisense transcription in the Igh
- IL-7R loss causes reduced expression of B lineage transcription factors and targets

Authors
Amanda Baizan-Edge, Bryony A. Stubbs, Michael J.T. Stubbington, Daniel J. Bolland, Kristina Tabbada, Simon Andrews, Anne E. Corcoran

Correspondence
anne.corcoran@babraham.ac.uk

In brief
Baizan-Edge et al. show that the interleukin-7 receptor drives antibody repertoire diversity. Deletion of the IL-7R impairs usage of most VH genes and DH genes in V(D)J recombination of the Igh, causing a severely restricted repertoire. Defects include reduced Igh antisense transcription and diminished expression of B lineage transcription factors.
IL-7R signaling activates widespread \(V_H\) and \(D_H\) gene usage to drive antibody diversity in bone marrow B cells

Amanda Baizan-Edge,1,4 Bryony A. Stubbs,1 Michael J.T. Stubbington,1,5 Daniel J. Bolland,1,2 Kristina Tabbada,1,2,6 Simon Andrews,3 and Anne E. Corcoran1,2,7,*

1Nuclear Dynamics Programme, Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK
2Lymphocyte Signaling and Development Programme, Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK
3Bioinformatics Group, Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK
4Present address: Centre for Gene Regulation and Expression, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK
5Present address: 10x Genomics, 6230 Stoneridge Mall Road, Pleasanton, CA 94588, USA
6Present address: Palaeogenomics, PalaeoBARN, School of Archaeology, University of Oxford, Oxford OX1 3TG, UK
7Lead contact
*Correspondence: anne.corcoran@babraham.ac.uk
https://doi.org/10.1016/j.celrep.2021.109349

SUMMARY

Generation of the primary antibody repertoire requires V(D)J recombination of hundreds of gene segments in the immunoglobulin heavy chain (Igh) locus. The role of interleukin-7 receptor (IL-7R) signaling in Igh recombination has been difficult to partition from its role in B cell survival and proliferation. With a detailed description of the Igh repertoire in murine IL-7R\(^{−/−}\) bone marrow B cells, we demonstrate that IL-7R signaling profoundly influences \(V_H\) gene selection during \(V_H\)-to-\(DJ_H\) recombination. We find skewing toward \(3′\) \(V_H\) genes during de novo \(V_H\)-to-\(DJ_H\) recombination more severe than the fetal liver (FL) repertoire and uncover a role for IL-7R signaling in \(D_H\)-to-\(J_H\) recombination. Transcriptome and accessibility analyses suggest reduced expression of B lineage transcription factors (TFs) and targets and loss of \(D_H\) and \(V_H\) antisense transcription in IL-7R\(^{−/−}\) B cells. Thus, in addition to its roles in survival and proliferation, IL-7R signaling shapes the Igh repertoire by activating underpinning mechanisms.

INTRODUCTION

Interleukin-7 (IL-7) is a critical cytokine for B and T lymphocyte development. It is bound by the IL-7 receptor (IL-7R), composed of the common gamma chain (\(\gamma_c\)), shared with the IL-2R, and the IL-7-specific IL-7R\(\alpha\) chain (CD127—encoded by \(I\/i7\)), which also pairs with the thymic stromal lymphopoietin receptor (TSLPR), important in fetal liver (FL) B cell survival (Vosshenrich et al., 2003; Rochman et al., 2009). Binding of IL-7 to the IL-7R in pro-B cells activates several signaling pathways, including Janus kinase/signal transducers and activators of transcription (JAK/STAT), phosphoinositide-3 kinase (PI3K)/Akt (Protein kinase B), mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK), and PLC\(\gamma\)/PKC/mammalian target of Rapamycin (mTOR), and has been associated with proliferation and survival, B lineage commitment, and Igh recombination (Corcoran et al., 1996, 1998; reviewed by Corfe and Paige, 2012; Yu et al., 2017).

IL-7R signaling is essential for B lineage commitment. Its absence in IL-7R\(^{−/−}\) mice impairs early B cell development from the common lymphoid progenitor (CLP) stage, resulting in reduced progenitors and impaired B-lineage potential (Peschon et al., 1994; Miller et al., 2002; Erlandsson et al., 2004; Dias et al., 2005, Medina et al., 2013). This is due in part to failure to activate early B-Cell Factor 1 (EBF1), a crucial transcription factor (TF) for B lineage specification (Kikuchi et al., 2005; Tsapogas et al., 2011; Pongubala et al., 2008; Roessler et al., 2007; Boller and Grosschedl, 2014). Paired box 5 (PAX5), a key TF for B cell commitment (Nutt et al., 1999; Rumfelt et al., 2006), is also reduced in IL-7R\(^{−/−}\) pro-B cells (Corcoran et al., 1998), but this may be due to reduced EBF1 expression (O’Riordan and Grosschedl 1999; Decker et al., 2009; Lin et al., 2010). However, some cells progress to the pre-pro-B stage (Kikuchi et al., 2005; Peschon et al., 1994; Miller et al., 2002) and fewer to the CD19\(^+\) pro-B compartment (Corcoran et al., 1998; Miller et al., 2002).

Although IL-7R\(^{−/−}\) pro-B cells in the bone marrow (BM) undergo Igh \(V_H\)-to-\(DJ_H\) recombination (Corcoran et al., 1998), their usage of \(V_H\) genes is severely restricted, indicating a role of IL-7R signaling in this process. Importantly, this is independent of proliferation: IL-7R\(^{−/−}\) cells expressing an IL-7R\(\alpha\) chain lacking Tyr449, required for PI3K signaling, express a recombinated Ig\(\mu\) polypeptide but do not proliferate in vitro (Corcoran et al., 1996). Conversely, a chimeric receptor comprising the IL-7R\(\alpha\) extracellular domain and the IL-2R\(\beta\) intracellular domain restored proliferation in IL-7R\(^{−/−}\) cells, but not Ig\(\mu\) expression, indicating a non-redundant role for the IL-7R in Igh V(D)J recombination.
A diverse antibody repertoire requires inclusion of all available V\textsubscript{H} and D\textsubscript{JH} genes. Large-scale processes, including non-coding transcription and Ig locus contraction, are thought to facilitate accessibility of distal 5\' V\textsubscript{H} genes to the recombination center over the DJ region, where the recombination activating gene (RAG) 1 and 2 bind (Ji et al., 2010, Corcoran et al., 1998; Bolland et al., 2004; Yancopolous and Alt, 1985, Chowdhury and Sen, 2003; Fuxa et al., 2004; Jhunjhunwala et al., 2008; Sayegh et al., 2005; Stubbington and Corcoran, 2013). Nevertheless, V\textsubscript{H} genes recombine at widely different frequencies; frequently recombining V\textsubscript{H} genes also have one of two local active chromatin states (Bolland et al., 2016).

IL-7R\textsuperscript{−/−} pro-B cells in vivo displayed decreased non-coding transcription and recombination of 5\' V\textsubscript{H} genes (Corcoran et al., 1998), prompting the hypothesis that the IL-7R influences Igh recombination through increasing accessibility of 5\' V\textsubscript{H} genes, supported by studies linking IL-7R signaling and active histone modifications in the Igh locus (Chowdhury and Sen, 2001; Johnson et al., 2003; Xu et al., 2008), and suggesting that IL-7R activation of the Igh locus is mediated by STAT5 (Bertolino et al., 2005). However, conditional deletion of STAT5 was rescued by the pro-survival factor B-Cell Lymphoma-2 (BCL-2) with no deficiency in 5\' V\textsubscript{H} recombination, suggesting that the dominant role of STAT5 was pro-B cell survival (Malin et al., 2010). IL-7R\textsuperscript{−/−} B cells were only partially rescued and cannot be rescued by Eu-BCL-2 expression (Maraskovsky et al., 1998), indicating that the IL-7R has additional downstream signaling roles. Heterogeneous expression of the IL-7R during the cell cycle both inhibits Rag recombinase expression to prevent DNA breaks during replication, while maintaining Igh locus accessibility throughout the cell cycle (Johnson et al., 2012).

Other studies have suggested that loss of IL-7R\textsuperscript{α} prevents B cell progression beyond the pro-B stage and that B cells in the BM originate from the FL (Kikuchi et al., 2005; Peschon et al., 1994; Miller et al., 2002; Carvalho et al., 2001), supported by similarity in V\textsubscript{H} repertoire between IL-7R\textsuperscript{−/−} and FL B cells (Corcoran et al., 1998; Jeong and Teale, 1988; Yancopolous et al., 1988). Definitive conclusions have been hampered by incomplete knowledge of the Igh locus, provided later (Johnston et al., 2006), and low-resolution Igh repertoire assays.

With next-generation sequencing (NGS) enabling more comprehensive analysis of Igh repertoires, we have revisited the IL-7R\textsuperscript{−/−} mouse (Peschon et al., 1994) to address the uncertainties above, which confound a complete picture of the role of the IL-7R in B cell development. Using VDJ sequencing (VDJ-seq) (Bolland et al., 2016), a DNA-based NGS technique, we have fully characterized the Igh DJ\textsubscript{JH} and VDJ\textsubscript{JH} repertoires in IL-7R\textsuperscript{−/−} BM B cells. Widespread use of gene segments in both D\textsubscript{JH} and V\textsubscript{H}DJ\textsubscript{JH} recombination was severely impaired by loss of IL-7R signaling.

Importantly, IL-7R\textsuperscript{−/−} BM B cells are not derived from the FL. Junctions between V\textsubscript{H}, D\textsubscript{JH} and J\textsubscript{H} gene segments in IL-7R\textsuperscript{−/−} BM B cells are much more variable than FL B cell sequences, indicating that IL-7R\textsuperscript{−/−} BM pro-B cells undergo de novo V(D)J recombination. Furthermore, they have a much more severe reduction in repertoire diversity than FL. Transcriptome analysis reveals loss of large-scale antisense intergenic transcripts in both D\textsubscript{H} and V\textsubscript{H} regions and reduced expression of key transcription factors required for Igh recombination, including EBF1 and Pax5. Thus, IL-7R signaling promotes Igh repertoire diversity in BM pro-B cells by activating mechanisms that enable large-scale access to V\textsubscript{H} and D\textsubscript{JH} genes.

**RESULTS**

**Usage of V and D genes in Igh recombination is severely skewed in pro-B cells lacking the IL-7R\textsuperscript{α} chain**

VDJ-seq was performed on two IL-7R\textsuperscript{−/−} BM and two wild-type (WT) FL pro-B cell samples and compared with two WT BM data-sets we previously generated (GEO: GSE80155; Bolland et al., 2016). Replicate libraries for all genotypes were highly correlated, indicating VDJ-seq is highly reproducible (Figure S1). Although IL-7R\textsuperscript{−/−} libraries were generated with approximately 6-fold fewer reads (Figure S2), the ratio of VDJ\textsubscript{JH} to DJ\textsubscript{JH} recombinated products was similar to WT, indicating that dynamic progression through first DJ\textsubscript{JH} to VDJ\textsubscript{JH} recombination, was not slowed (Figure S3). A binomial test was applied to determine genes with significantly greater read counts than expected by chance, considered to be actively recombining (Bolland et al., 2016). Consistent with a requirement for IL-7R signaling to enable usage of V\textsubscript{H} genes in recombination, fewer V\textsubscript{H} genes passed the binomial test in IL-7R\textsuperscript{−/−} (84 genes) relative to WT pro-B cells (128 genes). All but three were within the WT group (Table S1).

To visualize V\textsubscript{H} gene recombination frequencies and compare between genotypes, V\textsubscript{H} gene quantifications were normalized to the total number of reads over all V\textsubscript{H} genes for each genotype (Figure 1A). A much higher proportion of sequences mapped to the most 3\' V\textsubscript{H} genes in IL-7R\textsuperscript{−/−} than in the WT repertoire (Figure 1B). Notably, the first five V\textsubscript{H} genes comprised 45% of the total repertoire, compared with 20% in WT. Of the 44 V\textsubscript{H} genes that recombine in WT, but not in IL-7R\textsuperscript{−/−} BM, the vast majority were at the 5' end of the V region, including several that normally recombine at high frequency (J558.16.106, J558.26.116, and J558.67.166).

The 195 V\textsubscript{H} genes segregate into 16 families in 3 clans (Johnston et al., 2006). Their diverged evolutionary origins correlate with differential TF binding and chromatin state around V\textsubscript{H} genes (Bolland et al., 2016). Because accessibility, TF expression, and chromatin state have been linked to IL-7R signaling, we investigated the relationship between recombination frequency and V\textsubscript{H} gene clan or chromatin state in IL-7R\textsuperscript{−/−} pro-B cells. V\textsubscript{H} genes in clan 1 recombine at lower frequency in the IL-7R\textsuperscript{−/−}, although clan 2 and 3 V\textsubscript{H} genes recombine at similar frequencies to WT (Figure 1C). When the data are broken down to V\textsubscript{H} gene families, a more nuanced picture emerges. Within clan 2 and 3, 7158 and Q52, the two largest and most 3\' V\textsubscript{H} families, recombine more frequently in IL-7R\textsuperscript{−/−}. However, several of the smaller families in the middle region recombine less frequently (Figure S3). The enhancer state V\textsubscript{H} genes (including clan 1 and the distal 3609 family from clan 2) were significantly less frequently recombined relative to the architectural state V\textsubscript{H} genes (clan 2, except 3609, and clan 3; Figure 1D). However, some architectural state V\textsubscript{H} families also recombined less frequently. Thus, loss of IL-7R impacts on V\textsubscript{H} genes in the enhancer state (distal and middle...
genes) and on several middle region families in the architectural state. This distribution also applies to the clans: loss of IL-7R reduces recombination of clan 1 (mostly distal V genes) as well as the middle genes from clans 2 and 3. Importantly, this suggests that the IL-7R does not influence either clans or local chromatin states selectively but rather linear positioning in the IgH V region, i.e., loss of IL-7R impairs recombination of middle and 5' V genes in the IgH locus.

Figure 1. IL-7R<sup>−/−</sup> pro-B cells have impaired V<sub>H</sub> and D<sub>H</sub> repertoires

(A) Recombination frequencies of 195 V<sub>H</sub> genes measured by VDJ-seq. WT BM pro-B cells (blue) and IL-7R<sup>−/−</sup> (red) are shown. Two biological replicates are shown as open circles (bar height represents average). Reads for each V<sub>H</sub> gene are shown as percentage of total reads quantified. V<sub>H</sub> gene number legend is shown in Table S1.

(B) The mean of each V<sub>H</sub> gene was divided with the WT mean followed by log2 transformation. Only genes recombining in either genotype are shown. * represents V<sub>H</sub> genes with value 0 in IL-7R<sup>−/−</sup> (only in WT). V<sub>H</sub> gene raw read counts and recombining/non-recombining classification are shown in Table S1.

(C and D) Log2 values for each gene in graph B (excluding those marked by *) were grouped by (C) evolutionary origin: clan 1 (n = 78), clan 2 (n = 27), and clan 3 (n = 26); ANOVA (degrees freedom [Df] = 2; F-value = 5.39; p = 0.009) and (D) chromatin state: enhancer (n = 68), architectural (n = 30), and background (n = 33); ANOVA (Df = 2; F-value = 4.54; p = 0.012).

(E) Reads for each D<sub>H</sub> gene as percentage of total reads quantified for two biological replicates of WT (blue) and IL-7R<sup>−/−</sup> pro-B cells.
Active recombining D$_i$ genes were also identified by binomial test. VDJ-seq revealed profound changes in individual D$_i$ usage in IL-7R$^{-/-}$ pro-B cells. Several centrally positioned DSP gene segments (DSP2$\times$2', DSP2$\times$3', DSP2.3, and DSP2.5) do not recombine in IL-7R$^{-/-}$ pro-B cells (Figure 1E). Conversely, relative recombination frequencies of the most 3' D$_{i}$ gene, DQ52; the most 5', DFL16.1; and its adjacent DSP2.9 were increased in IL-7R$^{-/-}$ cells.

**IL-7R$^{-/-}$ pro-B cells do not originate from the FL**

We investigated whether recombination events in IL-7R$^{-/-}$ BM were comparable to B cells derived from FL, rather than de novo in the BM. VDJ-seq data from WT embryonic day 15.5 pro-B cells were compared with IL-7R$^{-/-}$ and WT BM. Notably, the ratio of DJ to VDJ recombining in FL was 12:1, in contrast to WT and IL-7R$^{-/-}$ BM ratios of approximately 2:1 and 1:1, respectively (Figure S3). Consistent with previous reports, the FL V$_i$ gene repertoire exhibited a 3' bias relative to WT, including more frequent use of the most recombed V$_i$ gene, V$_{\mu}$-81X (11% compared with 7% for WT BM; Figures 2A and S2C). However, IL-7R$^{-/-}$ cells had a more pronounced phenotype, with V$_{\mu}$-81X comprising 14% of VDJ recombined sequences. Additionally, many 5' V$_i$ genes recombined less frequently than in FL pro-B cells (Figure 2A). V$_{\mu}$ usage within Igk V$_i$ gene family sub-domains (Figure 2B) is distributed evenly across the locus in WT but is somewhat biased toward the D-proximal 3' gene families in FL B cells. However, this shift is mild for all but the central and distal J558 genes (Figure 2C). The IL-7R$^{-/-}$ repertoire is also biased toward the 7183/O52 V$_i$ family but much more severely than FL B cells. In contrast to FL, this increase in 7183/O52 V$_{\mu}$ gene usage was mirrored by a decrease in usage for every other family except the small J606 family. Thus, recombination in IL-7R$^{-/-}$ cells is markedly more biased toward the 3' V$_i$ genes than FL, suggesting IL-7R$^{-/-}$ BM pro-B cells do not simply originate from FL precursors.

To distinguish whether VDJ$_{\mu}$ sequences in IL-7R$^{-/-}$ BM cells are derived from FL or adult BM, we analyzed VDJ-seq libraries with IMGT to compare non-templated (N-) incorporations within V$_{\mu}$ and DJ$_{\mu}$ junctions. Terminal deoxynucleotide transferase (TdT), which inserts N-nucleotides, is absent in FL and first expressed in pro-B cells in the BM (Feeney, 1990; Li et al., 1993). Accordingly, we identified very few N- additions in FL junctions; only 25% and 15% had more than one incorporation in the VD and DJ junction, respectively. In contrast, around 80% of VD and DJ junctions in WT BM had N- additions (Figures 2D and 2E). 60% of VD and DJ joins from IL-7R$^{-/-}$ cells also had N- additions. IL-7R$^{-/-}$ and WT sequences also had a similar distribution, including up to 10 additions (Figures 2D and 2E). Together, these data demonstrate that IL-7R$^{-/-}$ pro-B cells undergo V(D)J recombination de novo in the BM, but loss of the IL-7R$^{-/-}$ severely restricts usage of V$_{\mu}$ and DJ$_{\mu}$ genes in the formation of the primary Igk repertoire.

**IL-7R signaling does not influence local V gene chromatin state**

We next investigated how IL-7R signaling may regulate V(D)J recombination. Reduction in recombination of 5' V$_i$ genes in IL-7R$^{-/-}$ BM, together with normal DJ/VDJ ratios, suggests that signaling through IL-7R is specifically required for 5' V$_{\mu}$ gene recombination. We first asked whether altered Recombination Signal Sequence (RSS) accessibility could account for this reduced recombination, because the local enhancer chromatin state is predominantly associated with 5' V$_{\mu}$ genes (Bolland et al., 2016). We performed ATAC-seq to identify accessible DNA in a chromatin context (Pulvarthy et al., 2016). We performed these experiments in Rag recombinase-deficient models, which cannot recombine Ig loci, thereby enabling analysis of the intact, poised Igk locus and avoiding interference from loss of sequence, or elevated V$_{\mu}$ gene expression, due to recombination. We compared Rag2$^{-/-}$ pro-B cells (which express the endogenous IL-7R) with IL-7R$^{-/-}$ x Rag2$^{-/-}$ (referred to as IL-7R/ Rag2$^{-/-}$) BM. Duplicate libraries for both genotypes were highly correlated, indicating the data are highly reproducible (Figure S1).

In Rag2$^{-/-}$ pro-B cells, V$_{\mu}$ RSSs coincided with a peak of accessibility (Figure 3A). V$_{\mu}$ RSSs in IL-7R$^{-/-}$/Rag2$^{-/-}$ cells had a similarly highly accessible profile, suggesting IL-7R signaling does not activate local accessibility over V$_{\mu}$ RSSs. The Igk light chain V gene (Vc) RSSs were used as a negative control because the Igk locus does not become fully activated until the pre-B stage (Matheson et al., 2017). Accordingly, Vc RSSs were less accessible than V$_{\mu}$ RSSs (Figures 3A and 3B). Notably, this pattern was similar in Rag2$^{-/-}$ and IL-7R$^{-/-}$/Rag2$^{-/-}$ pro-B cells, suggesting that Igk V$_{\mu}$ genes are not activated prematurely in the absence of the IL-7R$_{\mu}$.

We divided the RSSs into clans and chromatin states to test whether these groups showed different accessibility patterns in IL-7R$^{-/-}$/Rag2$^{-/-}$. There was no significant difference between clans (Figure 3C). RSS accessibility in both enhancer and architectural groups was significantly reduced compared to the background state V$_{\mu}$ genes, which showed a small increase in RSS accessibility relative to WT (Figure 3D). Together, these data indicate little difference in local accessibility at recombining V$_{\mu}$ genes in the absence of the IL-7R$_{\mu}$.

**Igh antisense intergenic transcription, but not V$_{\mu}$ genic transcription, is impaired in IL-7R$^{-/-}$ pro-B cells**

Non-coding transcription has been proposed to promote chromatin accessibility to facilitate Igk V$_{\mu}$ gene recombination. The Igk V$_{\mu}$ region has short non-coding genic sense transcripts at V$_{\mu}$ promoters (Yancopoulos and Alt, 1985; Corcoran et al., 1998) and long intergenic antisense transcripts (Bolland et al., 2004, 2016; Verma-Gaur et al., 2012). To investigate changes in the absence of IL-7R signaling, we performed RNA-seq (Figures S1G–S1I). This revealed that there is generally little V$_{\mu}$ genic transcription over the 3' V genes; 31 of the 39 most D-proximal V genes showed no transcription in Rag2$^{-/-}$ pro-B cells. Very few non-coding V$_{\mu}$ genic transcripts were differentially expressed between IL-7R$^{-/-}$/Rag2$^{-/-}$ and Rag2$^{-/-}$ cells (Table S2). Notable exceptions included all four members of the middle region SM7 family, highly abundant in Rag2$^{-/-}$ but almost completely absent in IL-7R$^{-/-}$/Rag2$^{-/-}$ cells. Conversely, two of the most 5' V$_{\mu}$ genes, J558.88.194 and J558.87.193, not transcribed in Rag2$^{-/-}$ pro-B cells, are highly expressed in IL-7R$^{-/-}$/Rag2$^{-/-}$ pro-B cells (Figure 3E). Long intergenic antisense non-coding transcripts in the V$_{\mu}$ (Pax5-activated intergenic repeat 4 [PAIR4], PAIR6, J558, and J606) and DJ$_{\mu}$ regions (J4, J6, and J8), and...
D<sub>H</sub>-antisense) were also analyzed. Although the RNA-seq libraries were not strand specific, these known transcripts are easily distinguished from the much less frequently transcribed genic transcripts in Rag2<sup>−/−</sup> pro-B cells (Figure 3E). Strikingly, we next examined genome-wide alterations in gene expression in the absence of the IL-7R (Figure S4). Gene set enrichment analysis (GSEA) confirmed the essential role of IL-7R signaling in cell cycle and clonal expansion as genes related to E2F, G2M antisense transcription throughout the V<sub>H</sub> region was almost completely absent in IL-7R<sup>−/−</sup>/Rag2<sup>−/−</sup> pro-B cells. DESeq2 revealed significant reduction in all V<sub>H</sub> antisense transcripts tested (Table S2). Furthermore, although sense transcription over the J<sub>H</sub> (μ<sub>H</sub>) or the C<sub>H</sub> (λ<sub>H</sub>) transcript regions was relatively unchanged (Table S2), D<sub>H</sub> antisense transcription was almost completely lost over the entire DJ<sub>H</sub> region in IL-7R<sup>−/−</sup>/Rag2<sup>−/−</sup> cells (Figure 3F).

**EBF1, PAX5, and other key B-cell-lineage-specifying genes are mis-regulated in IL-7R<sup>−/−</sup> pro-B cells**
checkpoint, and MYC were downregulated in the IL-7R+/Rag2−/− cells (Figures S4B–S4D). Consistent with previous reports, expression of both Ebf1 and Pax5 was substantially reduced in IL-7R+/Rag2−/− pro-B cells (Table S3). Importantly, several key B-lineage genes regulated by Ebf1 and Pax5 had reduced transcription levels in IL-7R+/Rag2−/− cells, including Foxo1, Rag1, and Cd79a (coding for Igα, pre-B cell receptor complex), downregulated in EBF1-deficient cells (Györy et al., 2012). PAX5-activated genes, including Smarca4 (encoding BRG1) and Lef1, were also decreased. Conversely, FLT3R (Flt3), downregulated in pro-B cells by PAX5 (Pridans et al., 2008), was upregulated in IL-7R+/Rag2−/− pro-B cells. GSEA demonstrated that, overall, genes with PAX5 binding sites at their promoters were depleted in IL-7R+/Rag2−/− pro-B cells (Figure S4E). However, some B-cell-specific genes regulated by PAX5 and EBF1 were expressed normally: Irf4 (direct target of both), Myb, and Pdcd1 (activated and repressed by EBF1) showed no significant transcriptional changes in IL-7R+/Rag2−/− cells. Indeed, Irf8 and Ika/3 (Aiolos), activated by both, were more highly expressed in IL-7R+/Rag2−/− cells. Together, these results suggest that EBF1 and PAX5 function is mis-regulated in IL-7R−/− cells.

To determine whether IL-7R signaling additionally influences the binding pattern of these and other TFs, we compared accessible hypersensitivity sites with ATAC-seq, using the Model-based Analysis of ChiP-Seq (MACS) caller (Zhang et al., 2008) function in Seqmonk. We divided the sites into two groups: those that had fewer reads (less accessible) or more reads (more accessible) in IL-7R+/Rag2−/− than Rag2−/− cells. We used hypergeometric optimization of motif enrichment (HOMER) to identify TF motifs within these sites. This allowed us to infer TFs that bind less often (less accessible sites) in IL-7R+/Rag2−/− relative to Rag2−/− and vice versa. EBF1 bound less often in IL-7R+/Rag2−/− cells, correlating with reduction in its expression and that of its target genes. The PAX5 motif was not found in HOMER. We infer that the PAX8 motif, with a similar binding pattern, is PAX5, because Pax8 is not expressed in B cells. Again, we found reduced representation of this motif in ATAC-seq-accessible sites in IL-7R+/Rag2−/− cells (Figure 4A).

To further interrogate the differences in PAX5 and EBF1 binding in the Ig locus, we examined MACS peaks in published PAX5 (GSM932924) and EBF1 (GSM876622 and GSM876623) chromatin immunoprecipitation (ChiP) data in Rag−/− pro-B cells. We quantified ATAC-seq reads over these sites to infer TF binding differences between IL-7R+/Rag2−/− and Rag2−/− cells. Of 24 accessible sites in the Igh locus (not counting RSSs), all overlapped with one of the 34 PAX5 binding sites. Five showed a marked decrease in accessibility in IL-7R+/Rag2−/− cells (Table S3). Importantly, two of these sites correspond to the PAIR4 and PAIR6 PAX5 binding sites (Figure S4I), providing a mechanism for the loss of non-coding transcription above. We only detected three EBF1 binding sites in the Igh locus, with no difference in accessibility. However, quantification of ATAC-seq reads over 2,896 EBF1 binding sites genome-wide revealed 643 significantly differentially enriched sites. 630 had lower accessibility in IL-7R+/Rag2−/− cells (Table S4). In the HOMER analysis, we also detected reduced accessibility at motifs of other important TFs involved in B cell specification, including E2A. Reduced specification was reflected in TF motifs at more accessible sites in IL-7R+/Rag2−/− pro-B cells, which included T cell development motifs (GATA family TFs) and early B cell priming TFs, including Pu.1, Myb, and Runx (Figure 4B). Furthermore, GSEA analysis showed that genes with binding sites for GATA, LIM domain only 2 (LMO2), and NFE2 (TFs in T- and erythroid development) were enriched in IL-7R+/Rag2−/− pro-B cells (Figures S4F–S4H). Overall, the pattern of TF motif accessibility and target gene alteration suggests that the IL-7R−/− enforces commitment to the B cell lineage.

**DISCUSSION**

Here, we asked whether and how the IL-7R plays a role in de novo immunoglobulin gene recombination in BM B cells. As previously reported (Jeong and Teale, 1988; Yancopoulos et al., 1988), we found that FL pro-B cells had a bias toward usage of 3’ Vh genes. However, the FL B cell antibody repertoire was far less restricted than previously thought, suggesting the current model of B cell ontogeny, in which complex antibody repertoires do not develop until after birth warrants revisiting (Siegrist and Aspinall 2009). It will be important to investigate which mechanisms that underpin adult Igh repertoire formation are already in place in FL, including long-range looping and local V gene activation, dependent on CCCTC-binding factor (CTCF) and PAX5 (Gerasimova et al., 2015; Bolland et al., 2016; Jain et al., 2018). Because PAX5 is essential for FL Igh recombination (Nutt et al., 1997), it may play similar roles in FL B cells.

Our unprecedented depth of analysis of IL-7R±/− VDJh and DJh sequences demonstrates that BM pro-B cells lacking the

![Figure 3. IL-7R+/Rag2−/− pro-B cells show no significant changes in RSS accessibility but striking loss of non-coding transcription over the Igh locus](image-url)
IL-7R<sup>−/−</sup> display widespread defects at both stages of *Igh* recombination. Most importantly, the *V<sub>hi</sub>* repertoire was highly biased toward 3′ *V<sub>hi</sub>* genes. Reduced use of 5′ *V<sub>hi</sub>* genes was much more pronounced than in FL pro-B cells, indicating that IL-7R signaling is specifically needed in the BM to make all *V<sub>hi</sub>* genes available for the primary antibody repertoire. The presence of N-additions within IL-7R<sup>−/−</sup> VDJ<sub>hi</sub> sequences demonstrates that they are derived from BM, not FL, progenitors. Our findings concur with a study in neonatal IL-7R<sup>−/−</sup> BM (Hesslein et al., 2006). Detection of N-additions at D<sub>hi</sub>-J<sub>hi</sub> junctions indicated that D<sub>hi</sub> to J<sub>hi</sub> recombination also took place de novo in the BM. Thus, V(D)J recombination progressed with normal dynamics but severely restricted participation of both V<sub>hi</sub> and D<sub>hi</sub> genes. Previous models of a block in B cell development in IL-7R<sup>−/−</sup> BM (Kikuchi et al., 2005; Peschon et al., 1994; Miller et al., 2002; Carvalho et al., 2001) inferred that IL-7R<sup>−/−</sup> BM B cells had originated in the FL, where the IL-7R is not essential (Erlandsson et al., 2004), due to their restricted *V<sub>h</sub>* gene usage. Here, our demonstration that V(D)J recombination occurs de novo in the BM, albeit in the very few remaining B cells, has enabled us to uncover specific roles of the IL-7R in regulating DH and VH gene usage in the BM, albeit in the very few remaining B cells, has enabled us to uncover specific roles of the IL-7R in regulating DH and VH gene usage.

A previous study showing that IL-7R<sup>−/−</sup> B cell development could be partially rescued by a vav-cre bcl2 transgene, indicating a crucial role in CLP survival (Malin et al., 2010), suggested that IL-7R signaling is not required for BM B cell recombination. However, *Bcl2* driven by the *Igh E<sub>μ</sub>* enhancer did not rescue accessibility, has not occurred. Nevertheless, several marks of V region accessibility are acquired, against which we measured the effect of IL-7R loss. There were no significant differences in local chromatin accessibility over *V<sub>h</sub>* gene RSSs in IL-7R<sup>−/−</sup>/Rag2<sup>−/−</sup> cells. Importantly, low accessibility at *V<sub>h</sub>* gene RSSs suggests that surviving IL-7R<sup>−/−</sup>/Rag2<sup>−/−</sup> B cells have not “rushed through” to the IL-7R-independent pre-B cell stage where increased *V<sub>h</sub>* access occurs. IL-7R signaling must be downregulated to enable Igk recombination (Johnson et al., 2008; Mandal et al., 2019). Overall, we find no evidence that defects in local accessibility at 5′ V genes account for the preference for recombination of 3′ *V<sub>hi</sub>* genes in IL-7R<sup>−/−</sup>/Rag2<sup>−/−</sup> cells. We also found that *V<sub>hi</sub>* specific non-coding transcription rarely changed in IL-7R<sup>−/−</sup>/Rag2<sup>−/−</sup> pro-B cells, supporting multomics studies that found no correlation with *V<sub>hi</sub>* usage (Choi et al., 2013; Bolland et al., 2016).

Non-coding intergenic transcription activates T cell receptor α (TCRA) locus recombination in vivo (Abarrategui and Krangel, 2006), although de novo antisense transcription over *Igh* 3′ *V<sub>hi</sub>* genes increases their recombination (Guo et al., 2011). These and other findings support a model in which intergenic transcription drives recombination (Corcoran, 2010). Here, widespread loss of all PAX5-dependent (PAIRs 4 and 6) and PAX5-independent (J558 and J606) antisense intergenic transcripts (Bolland et al., 2004, 2016; Ebert et al., 2011; Verma-Gaur et al., 2012; Choi et al., 2013) suggests that the IL-7R regulates all *Igh*...
antisense transcription and supports a role in promoting long-range mechanisms underpinning V_{H} to D_{H} recombination. We did not observe de novo antisense transcription over 3’ V_{H} genes in IL-7Rα−/− pro-B cells, indicating the relative increase in 3’ V_{H} gene recombination is secondary to the defect in 5’ recombination rather than a bona fide increase in 3’ recombination (Guo et al., 2021).

PAX5, downregulated in IL-7Rα−/− cells, is essential for Igh locus contraction (Fuxa et al., 2004; Nutt et al., 1997; Hesslein et al., 2003; Medvedovic et al., 2013; Montefiori et al., 2016). PAX5 has pleiotropic functions, but here, reduced accessibility at PAX5 sites on PAIR promoters and downregulation of PAX5-dependent PAIR transcription in the Igh locus suggests that PAX5 binding and function at these regulatory regions is directly impaired by loss of the IL-7R (Ebert et al., 2011).

A key finding was that D_{H} to J_{H} recombination is also impaired in IL-7Rα−/− BM B cells. Representation of the central DSP family was severely reduced. Strikingly, antisense non-coding transcription over the D_{H} region was also ablated. These findings support our model that antisense transcription over the DSP genes activates their recombination (Bolland et al., 2007) and reveal a role for the IL-7R in activating this transcription to drive D_{H} to J_{H} recombination.

Reduced accessibility at hundreds of EBF1 binding sites and reduced expression of multiple PAX5 targets by GSEA suggests specific functional consequences of reduced EBF1 and PAX5 expression. Conversely, increased accessibility at putative T cell TF motifs suggests that, although IL-7Rα−/− B cells are committed to the B cell lineage, they nevertheless remain plastic, similarly to PAX5−/− B cells (Nutt et al., 1999).

Our findings have important implications for human B cell development. Pediatric studies suggested early human B cell development did not require IL-7 (LeBien, 2000), but recent studies have shown that adult B cell development is dependent on IL-7R signaling, thereby aligning the dynamics of mouse and human IL-7R dependency (Parrish et al., 2009; Milford et al., 2016). It will be important to determine whether the IL-7R regulates immunoglobulin recombination in human B cells. Human immunodeficiency diseases and aging both have restricted antibody repertoire and poor response to infection (Siegrist and Aspinall, 2009; Martin et al., 2015). Notably, both Igh recombination and IL-7R signaling are impaired in aging mice (Stephan et al., 1997; Szabo et al., 1999) and humans. The therapeutic potential of the IL-7R in human aging is an emerging area of interest (Pastoors et al., 2015), and our findings suggest therapeutic potential of IL-7 for boosting naive antibody repertoires.

In conclusion, we reveal that, in addition to its roles in pro-B cell survival and proliferation, IL-7R signaling shapes the Igh repertoire at both the D_{H}−to-J_{H} and V_{H}−to-D_{H} stages of recombination in mouse BM and identify several mechanisms by which it can activate the Igh locus. IL-7R signaling is therefore essential for expanding antibody diversity to ensure robust activation of the adaptive immune system.

STAR METHODS

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Mice
- **METHOD DETAILS**
  - Primary cells
  - DNA extraction
  - RNA-seq
  - VDJ-seq
  - ATAC-seq
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109349.

ACKNOWLEDGMENTS

We thank Geoff Butcher, Martin Turner, and members of the Corcoran lab for helpful discussions and critical reading of the manuscript. We thank Mark Veldhoen for provision of IL-7Rα/Rag2−/− mice. We thank Flow Cytometry and Biological Services Unit for excellent support. Work in our laboratory is supported by grants from the BBSRC (BBS/E/B/000C0404, BBS/E/B/000C0405, BBS/E/B/000C0427, BBS/E/B/000C0426, and Core Capability Grant), A.B.-E. was supported by an MRC PhD studentship (1236141), B.A.S. was supported by an MRC PhD studentship (1129229), and M.J.T.S. was supported by a Babraham Institute PhD studentship.

AUTHOR CONTRIBUTIONS


DECLARATION OF INTERESTS


INCLUSION AND DIVERSITY

We worked to ensure sex balance in the selection of non-human subjects. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: December 9, 2020
Revised: April 5, 2021
Accepted: June 15, 2021
Published: July 13, 2021
REFERENCES


STAR★METHODS

KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibodies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD11b Monoclonal Antibody, Biotin</td>
<td>eBioscience Clone M1/70; Cat# 13-0112-82; RRID:AB_466359</td>
<td></td>
</tr>
<tr>
<td>Ly-6G/Gr-1 Monoclonal Antibody, Biotin</td>
<td>eBioscience Clone RB6-8C5; Cat# 13-5931-82; RRID:AB_466800</td>
<td></td>
</tr>
<tr>
<td>RAT ANTI MOUSE Ly-6C:Biotin</td>
<td>AbD Serotec Clone ER-MP20 Cat# MCA2389B RRID:AB_844550</td>
<td></td>
</tr>
<tr>
<td>CD3e Monoclonal Antibody, Biotin</td>
<td>eBioscience Clone TER-119 Cat# 13-5921-82 RRID:AB_466797</td>
<td></td>
</tr>
<tr>
<td>BV421 Rat Anti-Mouse CD45R/B220</td>
<td>BD bioscience Clone RA3-6B2 Cat# 562922 RRID:AB_2737894</td>
<td></td>
</tr>
<tr>
<td>PerCP-Cy5.5 Rat Anti-Mouse CD19</td>
<td>BD bioscience Clone 1D3 Cat# 561113 RRID:AB_10563071</td>
<td></td>
</tr>
<tr>
<td>FITC Rat Anti-Mouse CD43</td>
<td>BD bioscience Clone S7 Cat# 553270 RRID:AB_394747</td>
<td></td>
</tr>
<tr>
<td>Chemicals, peptides and recombinant proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agencourt AMPure XP beads</td>
<td>Beckman Cat# A63880</td>
<td></td>
</tr>
<tr>
<td>Ovation RNA-seq SystemV2 kit</td>
<td>NuGen Cat# 7102-08</td>
<td></td>
</tr>
<tr>
<td>TruSeq RNA Library Prep Kit v2</td>
<td>Illumina Cat# RS-122-2001</td>
<td></td>
</tr>
<tr>
<td>Nextera DNA Sample Preparation Kit</td>
<td>Illumina Cat# 15028211</td>
<td></td>
</tr>
<tr>
<td>Deposited data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw and analyzed data</td>
<td>This paper GEO: GSE157603</td>
<td></td>
</tr>
<tr>
<td>PAX5 ChIP Rag&lt;sup&gt;+/−&lt;/sup&gt; pro-B</td>
<td>Revilla-I-Domingo et al., 2012 GEO: GSM932924</td>
<td></td>
</tr>
<tr>
<td>EBF1 ChIP Rag&lt;sup&gt;+/−&lt;/sup&gt; pro-B</td>
<td>Vilagos et al., 2012 GEO: GSM876622, GSM876623</td>
<td></td>
</tr>
<tr>
<td>Experimental models: organisms/strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse: IL7Rx&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>J.J. Peschon Peschon et al., 1994</td>
<td></td>
</tr>
<tr>
<td>Mouse: RAG2&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>Frederick Alt Shinkai et al., 1992</td>
<td></td>
</tr>
<tr>
<td>Software and algorithms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seqmonk</td>
<td>The Babraham Institute</td>
<td><a href="https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/">https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/</a></td>
</tr>
<tr>
<td>LinkON</td>
<td>The Babraham Institute Chovanec et al., 2018</td>
<td><a href="https://github.com/peterch405/BabrahamLinkON/blob/master/README.md">https://github.com/peterch405/BabrahamLinkON/blob/master/README.md</a></td>
</tr>
<tr>
<td>HOMER</td>
<td>Heinz et al., 2010</td>
<td><a href="http://homer.ucsd.edu/homer/ngs/peaks.html">http://homer.ucsd.edu/homer/ngs/peaks.html</a></td>
</tr>
<tr>
<td>IMGT</td>
<td>Lefranc et al., 2015</td>
<td><a href="http://www.imgt.org/">http://www.imgt.org/</a></td>
</tr>
<tr>
<td>GSEA4.1</td>
<td>Subramanian et al., 2005</td>
<td><a href="http://www.gsea-msigdb.org/gsea/index.jsp">http://www.gsea-msigdb.org/gsea/index.jsp</a></td>
</tr>
</tbody>
</table>

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Anne Corcoran (anne.corcoran@babraham.ac.uk).

Materials availability
This study did not generate new unique reagents.
Data and code availability

- The VDJ-seq, ATAC-seq and RNA-seq raw sequencing files generated in this study, as well as processed files have been deposited at GEO, and are publicly available as of the date of publication. Accession numbers are listed in the Key resources table. This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the Key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice
Wild-type, RAG2−/− (Shinkai et al., 1992), IL-7Rα−/− (Peschon et al., 1994) and IL-7Rα−/− crossed with RAG2−/− (IL-7Rα−/− x RAG2−/−) C57BL/6 mice were maintained in accordance with Babraham Institute Animal Welfare and Ethical Review Body and Home Office rules under Project License 80/2529. Recommended ARRIVE reporting guidelines were followed. Mice were bred and maintained in the Babraham Institute Biological Services Unit under Specific Opportunistic Pathogen Free (SOPF) conditions. After weaning, mice were maintained in individually ventilated cages (2–5 mice per cage). Mice were fed CRM (P) VP diet (Special Diet Services) ad libitum, and millet, sunflower or poppy seeds at cage-cleaning as environmental enrichment. Health status was monitored closely and any mouse with clinical signs of ill-health or distress persisting for more than three days was culled. Treatment with antibiotics was not permitted to avoid interference with immune function. Thus, all mice remained ‘sub-threshold’ under UK Home Office severity categorization. 6-8-week-old IL-7Rα−/− and IL-7Rα−/− x RAG2−/− mice (all mixed sex), and 10-12 week old RAG2−/− mice (one female replicate and one male replicate) were used. Although wild-type (WT) comparison data were from 12 week old mice, IL-7Rα−/− animals were taken before 10 weeks because they produce fewer BM B cells as they age, with very few produced after 10 weeks (Peschon et al., 1994; Erlandsson et al., 2004). To maximize cell numbers and considering IL-7Rα−/− mice as young as 3 weeks have adult B cell populations (Hesslein et al., 2006), pro-B cells from 6-8-week old mice were taken for sorting. Fetal livers (FL) were harvested from day 15.5 mouse embryos.

METHOD DETAILS

Primary cells
Following CO2 asphyxiation and cervical dislocation, mouse BM was flushed from femurs and tibias, resuspended at 25 × 10^6 cells/ml, and depleted of macrophages, granulocytes, erythroid lineage and T cells using biotinylated antibodies against CD11b (MAC-1; ebioscience; 1:800), incubated on ice for 30 mins, followed by streptavidin MACs beads (10 μl/10^7 cells in 100 μl) (Miltenyi) at 4°C for 15 mins. MACS LS columns were equilibrated and the flow through collected for flow sorting. MACS depletion for FL was carried out as for BM, using TER119-biotin at a higher concentration (1:200). Cells were stained for 45 mins on ice in the dark with the following sorting antibodies from BD Bioscience: BV421-anti B220, 1:200; PerCP-Cy5.5-antiCD19, 1:400; FITC-anti CD43, 1:200) following sorting antibodies from BD Bioscience: BV421-anti B220, 1:200; PerCP-Cy5.5-antiCD19, 1:400; FITC-anti CD43, 1:200) following sorting.

DNA extraction
Genomic DNA was isolated from mouse B cells using the DNeasy kit (QIAGEN) according to the manufacturer’s instructions, except for the incubation step, which was changed to 30 mins. DNA was eluted in nuclease-free water and quantified by Nanodrop.

RNA-seq
Total RNA was extracted from ~200,000 cells for each replicate using the RNeasy Plus kit (QIAGEN). cDNA preparation was performed using the Ovation RNaseq System V2 kit (NuGen) protocol, and 200 ng of cDNA (made up to 130 μl with nuclease-free water) was carried through to generate 50bp paired-end RNA-seq libraries for Illumina sequencing. cDNA was sonicated using a Covaris M220 to fragment lengths between 200-700 bp (10% duty cycle, 140W peak incident power, 200 cycles per burst, 80 s processing time). End repair was carried out by adding 16 μl of 10x T4 Ligase Buffer (NEB), 4 μl of 10 mM dNTP mix, 3 μl T4 DNA polymerase (5 U/μl, Invitrogen), 1 μl Klenow (2U/μl, Invitrogen), 1 μl T4 polynucleotide kinase (10 U/μl, NEB) and 3 μl nuclease-free water, and incubating the reaction for 30 min at 20°C. Samples were purified again using QIAquick columns and eluted in 43 μl of nuclease-free water. A-tailing was then performed by adding 5 μl of 10x Klenow Buffer (NEB), 1 μl of dATP (10 mM) and 1 μl of exo minus Klenow (SU/μl, Fermentas); the reaction was incubated at 37°C for 30 min and purified using MinElute PCR purification columns (QIAGEN) and cDNA eluted in 10 μl. To sequence multiple libraries in one lane, Illumina TruSeq adaptors (6 bp index) (TruSeq RNA Library Prep Kit v2) were ligated by adding 15 μl 2x Rapid Ligation Buffer (Enzymatics), 4 μl of Rapid T4 DNA Ligase (6U/μl, Enzymatics) and 1 μl of adaptor mix (indexed adaptor and universal adaptor at 1.5 uM each) to the 10 μl of eluted cDNA. This reaction was incubated at
et al., 2013, 2015) on 70,000-100,000 cells. Sorted cells were washed, centrifuged, and resuspended in 50 μL of each Illumina paired-end primer (25 μM, Sigma) to the sonicated DNA (161 μL reaction), and incubating it at room temperature for 30 min. The sample was purified following the QIAquick PCR purification column protocol (QIAGEN). Repaired DNA fragments were eluted in 50 μL of nuclease-free water, and A-tailing was then carried out by adding 6 μL of 10x buffer 2 (NEB), 1 μL dATP (10 mM) and 3 μL Klenow exo- (5U/μL, NEB), and incubating the mix at 37°C for 30 min. DNA was purified using QIAquick PCR columns again. Two altered PE1 adaptor mixes were used, both including 6 random nucleotide barcodes. For the latter reaction, the A-tailed samples were split in two, each ligating to one of the two adaptor mixes. Adaptor ligation was carried out by adding 6 μL 10x T4 DNA ligase buffer (NEB), 4 μL adaptor (50 μM) and 5 μL T4 DNA ligase (400,000 U/ml, NEB) to each sample, making a 60 μL reaction. Reactions were incubated at 16°C overnight. The split reactions were pooled after incubation, and QIAquick columns were again used for purification. Depletion of unrecombined sequences was achieved by the use of 4 pairs of biotinylated primers (Bolland et al., 2016) which target the intergenic region upstream of each JH gene. Each sample was split so that each aliquot was < 1 μg of DNA in 50 μL containing 5 μL 10x Thermopol reaction buffer (Roche), 2 μL dNTP (10 mM each) and 1 μL Vent exo- (200U/ml, NEB). Incubation: 95°C-4 min, 55°C-5 min, 72°C-15 min. Reactions for each sample were pooled and purified using QIAquick columns. Unrecombined JH sequences were removed from the samples using streptavidin magnetic beads (Dynabeads MyOne Streptavidin C1, Invitrogen) and samples were purified using QIAquick columns. To enrich for unrecombined sequences, primer extension was carried out as above (annealing temperature 59°C rather than 55°C) using biotinylated reverse primers for each of the four JH genes. Recombined sequences were recovered using streptavidin beads as above. Beads were washed twice using 100 μL of washing buffer and once with EB buffer, and resuspended in 42 μL of EB buffer. To amplify the library, a PE1 primer (Illumina) and a mix of four JH primers which included a PE2 adaptor sequence (Bolland et al., 2016), were used. The resuspended beads were split into four aliquots (10.5 μL each) and 12.5 μL Pwo master mix (Sigma-Aldrich), 1 μL PE1 primer (10 μM) and 1 μL JH reverse primer mix (10 μM) were added to make a 25 μL reaction. Incubation: 94°C-2 min, 15 cycles of 94°C-15 s, 61°C-30 s, 72°C-45 s, followed by a final 72°C-5 min. The sample was pooled again, and beads were washed using 30 μL EB (QIAGEN), which was added to the supernatant containing the amplified library. These were put through a double-sided selection using Ampure XP beads (Bekman Coulter) for 200-700 bp. The library was further amplified to incorporate flowcell-binding indexes. PE1 and PE2 (including the index) primers (Illumina) were added to a PCR reaction as described above but with only 5 cycles and an annealing temperature of 55°C rather than 61°C. PCR products were combined and purified using SPRI beads, eluting in 20 μL.

Libraries were sequenced on an Illumina HiSeq (8-12 libraries/lane; 250 BP paired-end). Due to the low number of cells in IL-7Rx−/− BM, VDJ-seq libraries were generated with approximately 6-fold less starting material, resulting in reduced numbers of sequences relative to the WT BM VDJ-seq libraries (Table S7). Nevertheless, this amount of starting material does not compromise detection of the wide dynamic range of frequency of VDJ and DJ recombinated sequences (Chovanec et al., 2018).

**ATAC-seq**

ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) was performed as previously described (Buenrostro et al., 2013, 2015) on 70,000-100,000 cells. Sorted cells were washed, centrifuged, and resuspended in 50 μL of lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP40) on ice for 15 min. Nuclei were centrifuged at 600 rcf. for 10 min at 4°C, resuspended in 50 μL 1xTD buffer containing 2.5 μL TDE1 transposase (Illumina Nextera DNA Sample Preparation Kit), and incubated for 30 min at 37°C. Samples were purified using MiniElute columns (QIAGEN) according to the manufacturer's instructions and eluted in 21 μL RSB buffer (10mM Tris HCl pH7.6, 10mM NaCl, 1.5mM MgCl2, 0.1% NP40). PCR amplification and index incorporation were performed in a 50 μl reaction containing 5 μl of forward and reverse index primers (Illumina Nexter Index Kit), 15 μl NPM, 5 μl PPC (Illumina Nexter DNA Sample Preparation Kit) and 20 μl DNA. Libraries were purified using QIAquick PCR clean-up columns (QIAGEN) and sequenced on an Illumina HiSeq (6 libraries/lane).

**QUANTIFICATION AND STATISTICAL ANALYSIS**

RNA-seq reads were mapped to the mouse genome build NCBI37/mm9 using Bowtie2 and quantified using Seqmonk (Babraham Bioinformatics; https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Differential expression analysis was performed using DESeq2 (Love et al., 2014), using all annotated genes, V, Δ, and intergenic transcripts and intergenic transcripts in a single analysis. Gene Set Enrichment Analysis (GSEA) was performed using GSEA 4.1 (Subramanian et al., 2005). Mouse gene names were
converted to human gene symbols, and ran with default parameters for genes with BaseMean $\geq 5$. The Molecular Signature Database (MSigDB) hallmark gene sets and the transcription factor targets/regulatory target were used to perform pathway enrichment analysis limiting the output to the top 1000 gene sets.

For VDJ-seq, reads were mapped to the mouse genome build NCBI37/mm9 using Bowtie2. To quantify individual V and D genes, probes were created over each gene segment and correctly orientated reads were quantified over each probe using Seqmonk. Libraries were also analyzed using IMGT (International ImMunoGeneTics information system – http://www.imgt.org/ Lefranc et al., 2015). Analysis of VDJ-recombined sequences was carried out as described previously (Bolland et al., 2016); however, to analyze DJ-recombined sequences using IMGT, it was necessary to artificially add a V$\mu$ gene 5’ of the D$\mu$ sequence, as IMGT/HighVQUEST can only process VDJ sequences. The J558.78.182 V$\mu$ gene was appended, as it is functional and in-frame. These data were kept separate, and only D$\mu$ genes and the DJ$\mu$ junction were used for the analysis.

Due to the expectation for less variable VDJ events in IL-7R$\alpha$ cells, measures were taken to distinguish and discount technical duplicates in VDJ-seq libraries. VDJ-seq libraries were de-duplicated based on the sequence of read 2 (containing both V$\mu$-DJ$\mu$ and D$\mu$-J$\mu$ junctions) and the sequence and position of the V$\mu$ gene (LinkON pipeline described in Bolland et al. (2016). This method relies on the variability of gene usage, junction diversity, and the sonication step; therefore, reduced variability in the junctions and in V$\mu$ usage would lead to sequences being more likely to be identical, particularly in partially (D$\mu$-J$\mu$) recombined alleles. To overcome this issue, modified PE1 adaptors containing random barcodes were used to generate the IL-7R$\alpha$ and FL libraries, allowing PCR duplicates, biological duplicates and Illumina sequencing errors to be distinguished (Chovanec et al., 2018).

ATAC-seq reads were mapped to the mouse genome build NCBI37/mm9 using Bowtie2 and quantified using the MACS peak caller within Seqmonk. DESeq2 was used to identify genomic locations exhibiting significant differences in ATAC-seq reads in IL-7R$\alpha$ relative to Rag2$^{+/+}$, and these sites were tested for TF motifs using the HOMER analysis tool (http://homer.ucsd.edu/homer/ngs/peaks.html - Heinz et al., 2010).

ChIP reads were mapped using Bowtie and peaks called using MACS2 (in the narrow peak mode): PAX5 Rag$^{+/+}$ pro-B (GSM932924) and EBF1 Rag$^{+/+}$ pro-B (GSM876622, GSM876623).

For calculating statistical significance between groups for the above datasets, a two-tailed ANOVA (type III) together with a pairwise t test (adjusted by Benjamini and Hochberg method) were used to calculate significance and p values (significant when $p < 0.05$) when data were normal. When data failed the normality test a Kruskal-Wallis test followed by pairwise Wilcoxon test (adjusted by Benjamini and Hochberg method) was performed. All results for specific tests are explained in the figure legends, including the statistical test used, value and meaning of n, and confidence intervals.