nature immunology

ARTICLES

https://doi.org/10.1038/s41590-019-0468-0

CXCR4 signaling directs *Igk* recombination and the molecular mechanisms of late B lymphopoiesis

Malay Mandal[1,2*], Michael K. Okoreeh[1,2,7], Domenick E. Kennedy[1,2,7], Mark Maienschein-Cline[3], Junting Ai[1,2], Kaitlin C. McLean[1,2], Natalya Kaverina[4], Margaret Veselits[1,2], Iannis Aifantis[5,6], Fotini Gounari[1,2] and Marcus R. Clark[1,2*]

In B lymphopoiesis, activation of the pre-B cell antigen receptor (pre-BCR) is associated with both cell cycle exit and *Igk* recombination. Yet how the pre-BCR mediates these functions remains unclear. Here, we demonstrate that the pre-BCR initiates a feed-forward amplification loop mediated by the transcription factor interferon regulatory factor 4 and the chemokine receptor C-X-C motif chemokine receptor 4 (CXCR4). CXCR4 ligation by C-X-C motif chemokine ligand 12 activates the mitogen-activated protein kinase extracellular-signal-regulated kinase, which then directs the development of small pre- and immature B cells, including orchestrating cell cycle exit, pre-BCR repression, *Igk* recombination and BCR expression. In contrast, pre-BCR expression and escape from interleukin-7 have only modest effects on B cell developmental transcriptional and epigenetic programs. These data show a direct and central role for CXCR4 in orchestrating late B cell lymphopoiesis. Furthermore, in the context of previous findings, our data provide a three-receptor system sufficient to recapitulate the essential features of B lymphopoiesis in vitro.

B lymphopoiesis consists of alternating and mutually exclusive states of either stochastic immunoglobulin gene recombination or cell proliferation with selection[1]. Pro-B cells initiate immunoglobulin gene heavy chain (*Igh*) recombination. Cells that successfully recombine *Igh* express Igμ and assemble a pre-B cell antigen receptor (pre-BCR) containing surrogate light chain (λ5 and VpreB). These cells first proliferate as large pre-B cells and then transit to become small pre-B cells where they exit the cell cycle and recombine the immunoglobulin gene light chain loci (*Igk* then *Igl*)[2,3].

Proliferation in pro and large pre-B cells is driven by interleukin-7 (IL-7) receptor (IL-7R)-dependent activation of signal transducer and activator of transcription 5 (STAT5) and phosphatidylinositol-3-OH kinase (PI_3K)[1]. Active STAT5 induces expression of *Ccnd3* (encoding cyclin D3), while PI_3K activation induces *Myc* and aerobic glycolysis[4–8]. These signaling pathways also directly repress *Igk* recombination[1,9,10]. In parallel, PI_3K activation represses expression of the transcription factors forkhead box O1 (FOXO1) and forkhead box O3 (FOXO3), which induce *Rag1* and *Rag2* (refs. [11,12]).

In contrast, canonically, the pre-BCR transmits signals that initiate cell cycle exit and *Igk* recombination. Downstream activation of the protein kinase extracellular-signal-regulated kinase (ERK) induces expression of the transcription factors Aiolos and Ikaros, which repress *Ccnd3* and *Myc*[8,13]. ERK also induces the transcription factor E2A, which binds and activates the *Igk* intronic enhancer (Eκi)[9,14,15]. Furthermore, escape from IL-7R signaling allows upregulation of FOXO1 and FOXO3. Interestingly, these differentiation mechanisms occur in small pre-B cells where there is concurrent repression of pre-BCR expression[1,16]. Therefore, it is unclear whether initial transient pre-BCR signaling is sufficient to execute the entire developmental program in small pre-B cells or if other signals are required.

The pre-BCR also upregulates C-X-C motif chemokine receptor 4 (CXCR4), which senses C-X-C motif chemokine ligand 12 (CXCL12) gradients and has been proposed to mediate movement of pre-B cells out of IL-7-rich bone marrow (BM) niches[17–19]. By controling exposure to IL-7, CXCR4 is thought to control the balance between IL-7R and pre-BCR signaling[1]. However, CXCR4 transmits signals, including activating Ras-ERK, and in cancer has been implicated in multiple processes including invasion, epithelial–mesenchymal transition and proliferation[20–24]. Furthermore, in T cell development, CXCR4 synergizes with the pre-TCR to augment proliferation[25]. These data suggest that CXCR4 can mediate more than chemotaxis.

Results

Small pre-B cells contact CXCL12$^+$ stroma. To understand the spatial relationships between proliferating and differentiating B cell progenitors and IL-7, we isolated intact BM cores from wild-type (WT) C57BL/6 mice and stained them with antibodies specific for IL-7, B220, Ki67 and immunoglobulin M (IgM) (Fig. 1a). Visualization of stained BM by multicolor confocal microscopy revealed that cycling pro- and large pre-B cells (B220$^+$IgM$^-$Ki67$^+$) localized in IL-7^{hi} BM niches, while B220$^+$IgM$^+$ B cell progenitors, which mostly include immature B cells, resided in IL-7$^{-/lo}$ BM regions (Fig. 1a,b and Supplementary Fig. 1a).

[1]Department of Medicine, Section of Rheumatology, University of Chicago, Chicago, IL, USA. [2]Gwen Knapp Center for Lupus and Immunology Research, University of Chicago, Chicago, IL, USA. [3]Core for Research Informatics, University of Illinois at Chicago, Chicago, IL, USA. [4]Division of Nephrology, University of Washington, Seattle, WA, USA. [5]Department of Pathology, NYU School of Medicine, New York, NY, USA. [6]Laura and Isaac Perlmutter Cancer Center, NYU School of Medicine, New York, NY, USA. [7]These authors contributed equally: Michael K. Okoreeh, Domenick E. Kennedy.
*e-mail: mmandal@medicine.bsd.uchicago.edu; mclark@uchicago.edu

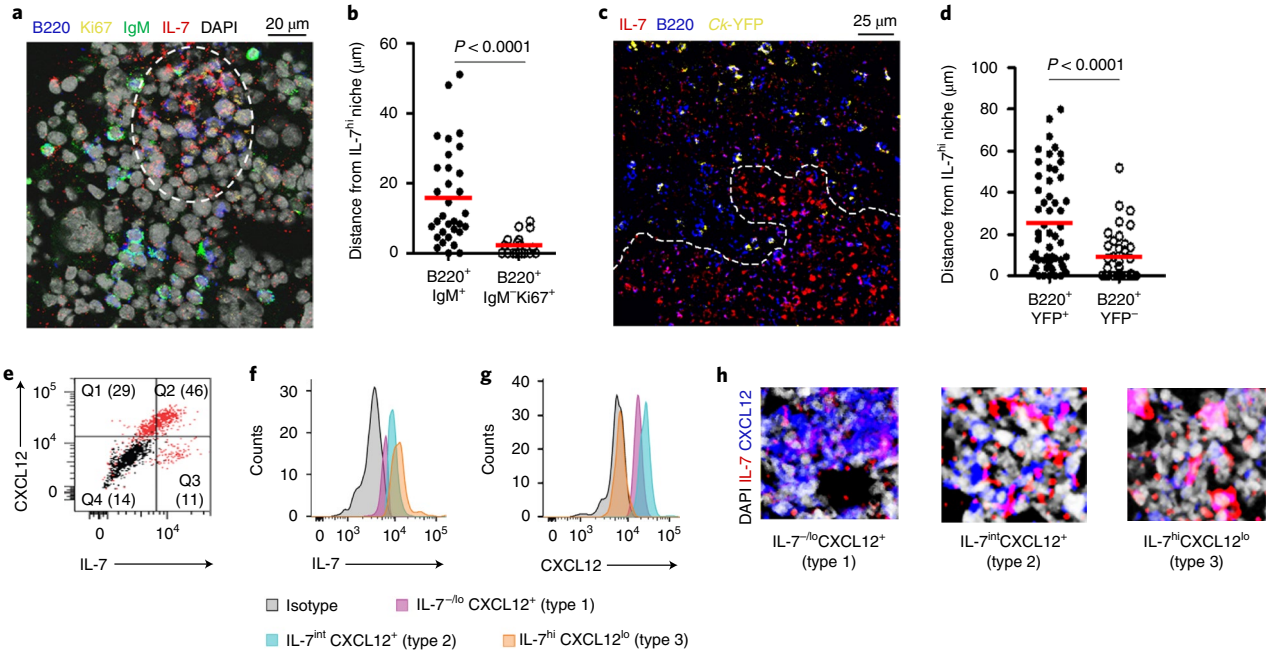

Fig. 1 | Location of proliferating and differentiating pre-B cells in the BM. a, Confocal microscopy of WT BM section (8-μm-thick femur) stained with antibodies to IL-7 (red), B220 (blue), Ki67 (yellow), IgM (green) and DAPI (gray) to visualize the locations of proliferating and IgM$^+$ B cell progenitors. Single-color panels are presented in Supplementary Fig. 1a. The image is representative of four independent images from three WT mice. **b**, Distance of IgM$^+$ and IgM$^-$Ki67$^+$ B cell progenitors from IL-7^{hi} niches (white dashed lines in **a** and **c**). Data were pooled from three independent experiments. Horizontal bars represent the means of the distances of all cells counted. The *P* value was calculated by unpaired *t*-test. **c**, Visualization by confocal microscopy of B cell progenitors in BM sections (8-μm-thick femur) of *Ck*-YFP mice with antibodies to IL-7 (red) and B220 (blue). The image is representative of three independent images from three WT mice. **d**, Distances of YFP$^+$ B cell progenitors from IL-7^{hi} niches in the BM of *Ck*-YFP mice. Data were pooled from two independent experiments. Horizontal bars represent the means of the distances of all cells counted. The *P* value was calculated by unpaired *t*-test. **e**, Flow cytometric analysis of IL-7 and CXCL12 expression by cultured BM stromal cells ($n = 3$), with staining broken into four quadrants, Q1–Q4. **f**,**g**, Identification of stromal cells expressing IL-7 (**f**) and CXCL12 (**g**) by flow cytometry ($n = 3$). **h**, Different areas of BM showing varying degrees of IL-7 (red) and CXCL12 (blue) expression by type 1, type 2 and type 3 stromal cells. Images are representative of five independent areas from two WT BM samples.

To identify the relative positions of small pre-B cells in the BM, we used an *Igk*-YFP reporter mouse in which cells expressing the *Igk* germline (*Ck*) and rearranged transcripts are marked with yellow fluorescent protein (YFP)[26]. YFP expression was low in pro-B cells, with strong upregulation in small pre-B cells (Supplementary Fig. 1b). There was intermediate expression of YFP in large pre-B cells. However, IgM$^-$YFP$^+$Ki67$^+$ cells were rare in the BM (data not shown), indicating that YFP expression in large pre-B cells is insufficient to mark these cells in confocal micrographs. Examination of the position of B220$^+$YFP$^+$ B cell progenitors relative to IL-7 revealed they were excluded from IL-7^{hi} niches (Fig. 1c,d). These studies suggest that small pre-B cells migrate away from IL-7-rich BM regions.

Flow cytometry of Ter119$^-$CD45$^-$ BM stromal cells revealed that IL-7 and CXCL12 were usually co-expressed, with IL-7^{int}CXCR12$^+$ more common than single-positive populations (Fig. 1e)[17]. Linear visualization of IL-7 and CXCL12 expression suggested the presence of three populations consisting of IL-7^{lo}CXCL12$^+$ (type 1 cells), IL-7^{int}CXCL12$^+$ (type 2) or IL-7^{hi}CXCL12^{lo} (type 3) (Fig. 1f,g). Examination of whole BM single planes revealed widespread distributions of each cell type (Supplementary Fig. 1c). However, individual high-power fields revealed distinct areas that were relatively enriched for type 1, type 2 or type 3 cells (Fig. 1h and Supplementary Fig. 1d).

BM from *Ck*-YFP mice was then stained with antibodies specific for IgM, IL-7 and CXCL12 (Fig. 2a,b). Small pre-B (YFP$^+$IgM$^-$) and immature B (YFP$^+$IgM$^+$) cells resided in niches enriched for type 1 (IL-7$^{-/lo}$CXCL12$^+$) stromal cells (Fig. 2a–c).

Examination of B cell progenitors for IL-7R and CXCR4 expression, by both messenger RNA (mRNA) and surface staining (Fig. 2d–g), revealed that IL-7R and CXCR4 were reciprocally expressed with small pre-B cells expressing relatively high surface densities of CXCR4 and low densities of IL-7R. Both IL-7R and CXCR4 were downregulated in immature B cells. Although pro-B and large pre-B cells can migrate along an IL-7 gradient, direct comparison of in vitro chemotaxis revealed that both large pre-B and small pre-B cells responded strongly to CXCR4 (Fig. 2h). Interestingly, large pre-B cells showed the strongest chemotaxis, even though CXCR4 surface densities were higher on small pre-B cells. However, most small pre-B cells (YFP$^+$IgM$^-$) were in intimate contact with CXCL12$^+$ stroma while IgM$^+$ immature B cells resided in the same area but were not contacting CXCL12$^+$ stroma (Fig. 2i). Additional high-power-field confocal microscopy with three-dimensional reconstruction showed that small pre-B cells were in tight contact with both CXCL12$^+$ cells and high local accumulations of extracellular CXCL12 (Fig. 2j and Supplementary Fig, 1e). Furthermore, these small pre-B cells clearly had CXCL12 in their cytoplasm, suggesting recent internalization of this ligand. This tight association suggests that CXCR4 might be doing more than positioning small pre-B cells away from IL-7.

CXCR4 directly regulates pre-B cell differentiation. Next, we crossed *Cxcr4*$^{fl/fl}$ mice with mice expressing the Cre recombinase under the *mb-1* promoter (*mb1*-Cre)[27,28]. In these mice, deletion of *Cxcr4* in large pre-B cells appeared complete with no detectable undeleted alleles in *Cxcr4*$^{fl/fl}$*mb1*-Cre$^{+/-}$ (hereafter referred to as *Cxcr4*$^{-/-}$) mice (Fig. 3a). BM was harvested from WT, *mb1*-Cre$^{+/-}$, *Cxcr4*$^{fl/fl}$ and *Cxcr4*$^{-/-}$ mice and analyzed by flow cytometry. While the pre-pro to large pre-B cell compartments were normal in *Cxcr4*$^{-/-}$ mice (Fig. 3b,c), the number of small pre-B cells was decreased

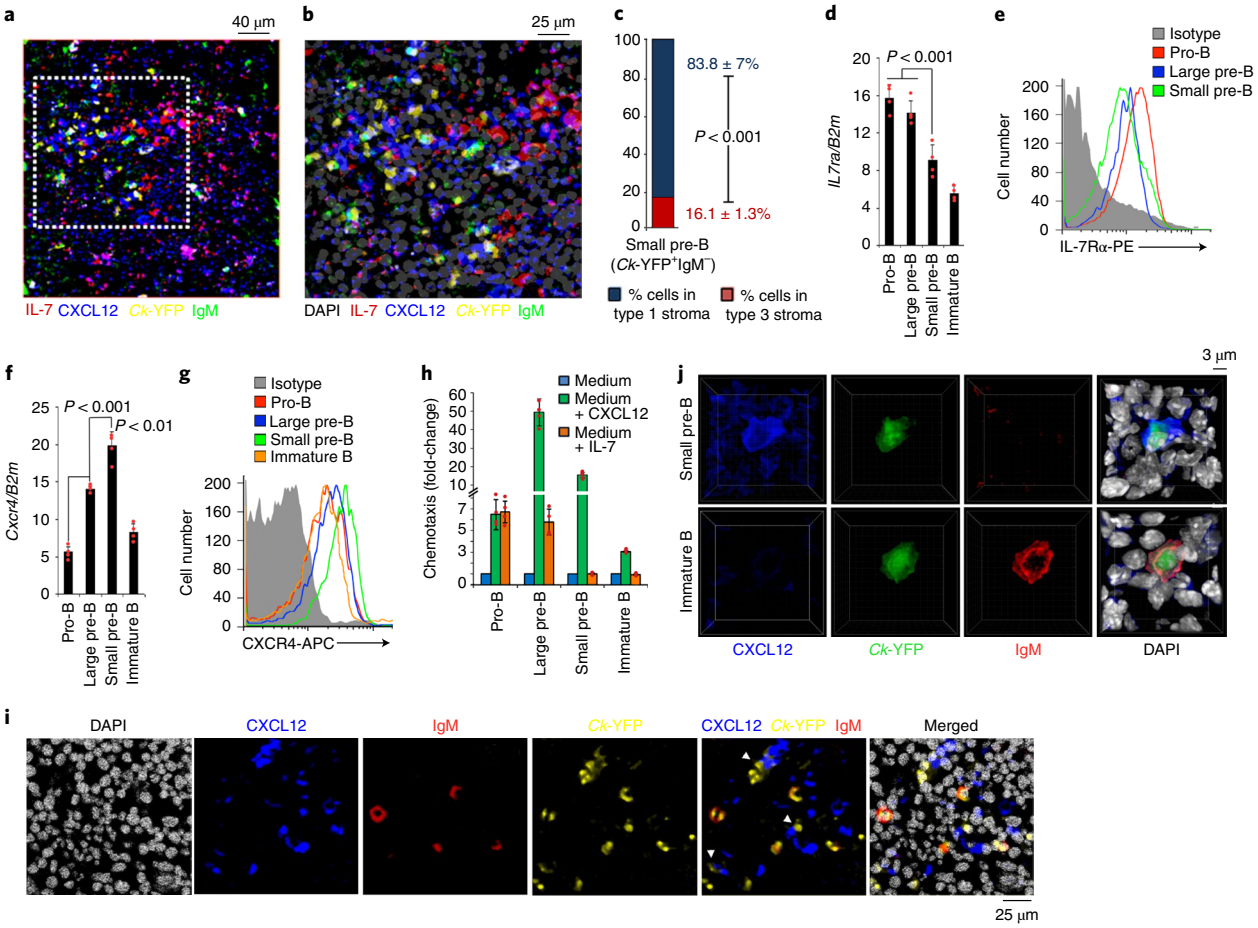

Fig. 2 | Small pre-B cells are in intimate contact with CXCL12$^+$ stroma. a,**b**, Distributions of small pre-B (YFP$^+$IgM$^-$) and immature (YFP$^+$IgM$^+$) B cell progenitors with respect to IL-7^{hi} and CXCL12$^+$ stromal cells in *Ck*-YFP BM sections (8-μm-thick femur) stained with antibodies to IL-7 (red), CXCL12 (blue), IgM (green) and DAPI (gray). The area indicated by a white dashed square in **a** is magnified in **b**. The image is representative of four independent images. **c**, Frequency of small pre-B (YFP$^+$IgM$^-$) cell progenitors in contact with type 1 (IL-7$^{-/lo}$CXCL12$^+$) and type 3 (IL-7^{hi}CXCL12^{lo}) stromal cells. The *P* value was calculated by unpaired *t*-test. **d**–**g**, Quantitative real-time PCR analysis of the expression of *Il7rα* (**d**) and *Cxcr4* (**f**), and flow cytometric analysis of the corresponding cell-surface expression of IL-7Rα (**e**) and CXCR4 (**g**) on indicated B cell progenitor populations ($n = 4$). Data are presented as means ± s.d. *P* values were determined by unpaired *t*-test. **h**, Chemotaxis of different B cell progenitors to IL-7 (10 ng ml^{-1}) and CXCL12 (100 ng ml^{-1}) by transwell migration assay. Data are normalized to migration with medium alone ($n = 4$). Data are presented as means ± s.d. **i**, Distribution of small pre-B (YFP$^+$IgM$^-$) and immature B (YFP$^+$IgM$^+$) cells in the CXCL12$^+$ niches in the BM (8-μm-thick femur) of *Ck*-YFP mice stained with antibodies to CXCL12 (blue), IgM (red) and DAPI (gray). Each image is representative of three independent images. **j**, Three-dimensional reconstructions of small pre-B (YFP$^+$IgM$^-$) and immature B (YFP$^+$IgM$^+$) cells in the BM of *Ck*-YFP mice stained with antibodies to CXCL12 (blue), IgM (red) and DAPI (gray) using Imaris software (representative of $n = 9$ images).

approximately threefold while the number of immature and mature B cells was decreased fivefold. Multicolor confocal microscopy of BM from *Cxcr4*$^{fl/fl}$ and *Cxcr4*$^{-/-}$ mice revealed that almost all *Cxcr4*$^{-/-}$ IgM$^-$ B cell progenitors were near IL-7^{hi} BM stroma, with few IgM$^+$ cells (Fig. 3d–f and Supplementary Fig. 2a,b). These data indicate that CXCR4 is required for normal development of small pre-B cells and their positioning away from IL-7^{hi} BM niches.

Previously, it has been reported that, in vitro, withdrawal of IL-7 is sufficient to induce pre-BCR activation, cell cycle exit and *Igk* recombination. However, these experiments were done using the stroma feeder cell line OP9, which express CXCL12 (refs. [8,9,19,29,30]). Furthermore, there was no other positive control. Therefore, in a stromal cell-free system, B220$^+$IgM$^-$ progenitors from WT BM were grown in the presence of high IL-7 (16 ng ml^{-1}) for 5 d, followed by 2 d of culture with IL-7 (16 ng ml^{-1}; +IL-7) or low IL-7 (0.2 ng ml^{-1}; −IL-7) without or with CXCL12 (100 ng ml^{-1}; +CXCL12). Cells were harvested and subjected to RNA sequencing (RNA-Seq). Overall, the combination of withdrawing IL-7 and adding CXCR4 regulated 4,274 genes (Fig. 4a). Plotting the fold-change versus significance (volcano plots) revealed that CXCL12 regulated many more genes by twofold or more compared with IL-7 withdrawal alone (Fig. 4b,c). To confirm the direct role of CXCR4 in signaling, we cultured B220$^+$IgM$^-$ progenitors from *Cxcr4*$^{-/-}$ BM and assayed for gene expression in cells withdrawn from IL-7 without or with CXCL12. As expected, the addition of CXCL12 did not substantially regulate genes in *Cxcr4*$^{-/-}$ B cell progenitors (Fig. 4d).

There were four distinct gene clusters regulated by CXCL12 and IL-7 withdrawal (Fig. 4a). In cluster 1 (Fig. 4e), downregulation of genes was primarily dependent on CXCL12. Gene ontology analysis revealed that these genes were enriched for cell cycle, metabolism and RNA processing programs. Among the cell cycle genes, *Ccnd2*, *Ccnd3* (Supplementary Fig. 3a), *Myc* (Supplementary Fig. 3b) and MYC-dependent genes (gene set enrichment analysis (GSEA); Supplementary Fig. 3c) were strongly repressed by CXCL12. In cluster 2 (Fig. 4f), gene repression was exclusively dependent on CXCL12, and these genes were enriched for mitosis and DNA

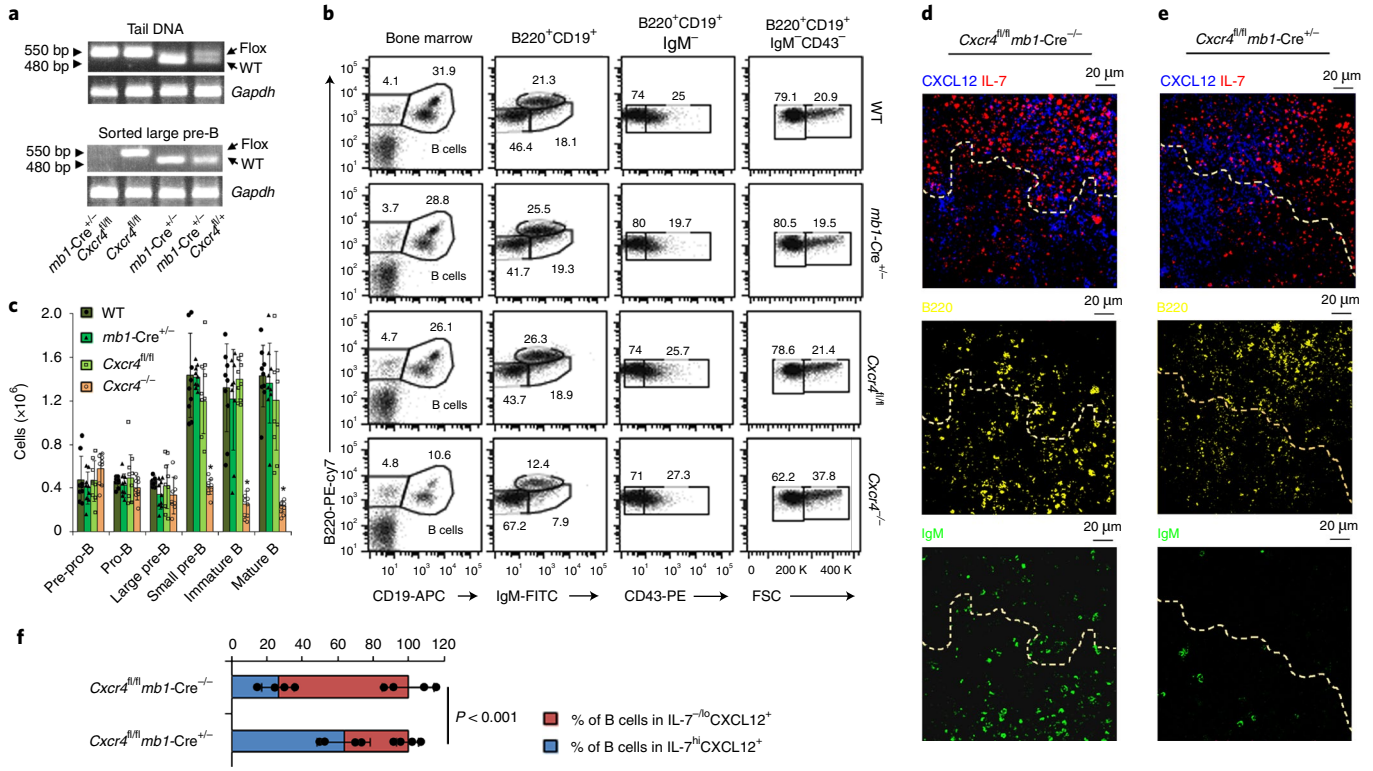

Fig. 3 | *Cxcr4* is required for the development of small pre-B cells. a, Genomic PCR of WT and floxed alleles for *Cxcr4* deletion in tail DNA and flow-sorted large pre-B cells from the indicated mice, with *Gapdh* as a control ($n=3$). **b**, Flow cytometric analysis of different developmental stages of B lymphopoiesis in the BM of WT, *mb1*-Cre$^{+/-}$, *Cxcr4*$^{fl/fl}$ and *Cxcr4*$^{-/-}$ littermate control mice ($n=9$). B cell progenitors are defined as B220$^+$CD19$^+$, pro-B cells are defined as B220$^+$CD19$^+$CD43$^+$IgM$^-$, large and small pre-B cells are defined as B220$^+$CD19$^+$CD43$^-$IgM$^-$FSC^{hi} and B220$^+$CD19$^+$CD43$^-$IgM$^-$FSC^{lo}, respectively, and immature B cells are defined as B220$^+$CD19$^+$CD43$^-$IgM$^+$. FSC, forward scatter. **c**, Absolute number of cells per mouse at different stages of B cell development in the BM of WT, *mb1*-Cre$^{+/-}$, *Cxcr4*$^{fl/fl}$ and *Cxcr4*$^{-/-}$ mice ($n=9$). $^*P<0.001$ compared with all of the controls (that is, WT, *mb1*-Cre$^{+/-}$ and *Cxcr4*$^{fl/fl}$). Data are presented as means ± s.d. **d**,**e**, Distribution of B cell progenitors in the BM of *Cxcr4*$^{fl/fl}$ (CXCR4 sufficient; **d**) and *Cxcr4*$^{fl/fl}$-*mb1*-Cre$^{+/-}$ (CXCR4 deficient; **e**) mice by confocal microscopy of the corresponding BM sections (8-μm-thick femur) stained with antibodies to IL-7 (red), CXCL12 (blue), B220 (yellow), IgM (green) and DAPI (gray) ($n=4$). **f**, Percentage of B cell progenitors (B220$^+$) in proximity to IL-7$^{-/lo}$CXCL12$^+$ and IL-7^{hi}CXCL12$^{+/-}$ stroma (from $n=4$ independent images). Data are presented as means ± s.d. Each dot represents the average distance obtained from each image of the indicated genotype. The *P* value was calculated by unpaired *t*-test.

replication and repair pathways. In cluster 3 (Fig. 4g), gene repression was mostly dependent on IL-7 withdrawal with an enrichment for RNA metabolism. CXCL12 primarily induced genes in cluster 4 (Fig. 4h) implicated in cell migration, cell adhesion and signaling, including the nuclear factor kappa light chain enhancer of activated B cells (NF-κB) (Supplementary Fig. 3d,e) and mitogen-activated protein kinase pathways. Transcription factors critical for late B lymphopoiesis were also induced, including *Irf4*, *Irf8*, *Spib* and *Ikzf3* (encoding Aiolos) (Supplementary Fig. 3f). CXCL12 did not induce *Tcf3* (encodes E2A), *Pax5*, *Foxo1*, *Foxo3* or *Ikzf1* (encodes Ikaros) (Supplementary Fig. 3g). Other genes induced in cluster 4 were *Rag1*, *Rag2* and *Brwd1* (refs. [10,31]) (Supplementary Fig. 3h,i). Overall, withdrawal of IL-7 and addition of CXCL12 were associated with repression of cell cycle programs, including E2F (Supplementary Fig. 3j), and induction of cell differentiation programs (Supplementary Fig. 3k).

Next, we isolated WT and *Cxcr4*$^{-/-}$ small pre-B cells and subjected them to RNA-Seq (Fig. 4i). There were 4,887 differentially expressed genes ($q<0.05$) in *Cxcr4*$^{-/-}$ small pre-B cells, with increased expression being more common than decreased expression. (The *q*-value is an adjusted *P*-value, taking into account the false discovery rate (FDR).) Gene ontology analysis revealed that CXCR4 was required for the repression of cell cycle, metabolic pathways, DNA replication and repair pathways (Supplementary Fig. 3l). In contrast, CXCR4 was required to upregulate signaling pathways including Ras, NF-κB, cell motility and cell adhesion (Supplementary Fig. 3m). GSEA confirmed that CXCR4 was necessary for repressing cell cycle genes, including E2F targets, and inducing differentiation programs (Fig. 4j,k). More *Cxcr4*$^{-/-}$ large and small pre-B cells were progressing through the cell cycle than WT cells (Fig. 4l). Volcano plots confirmed that CXCR4 was required for broadly repressing cell cycle genes and inducing differentiation genes (Fig. 4m). Many of these genes were confirmed by quantitative PCR (Fig. 4n). The expression of some B cell lineage and maintenance transcription factors, including *Tcf3*, *Ikzf1*, *Foxo1* and *Foxo3*, was not different in *Cxcr4*$^{-/-}$ small pre-B cells[10,29,32,33].

Repression of the surrogate light chain is critical for subsequent antigen selection and tolerance in immature B cells[16]. The presence or absence of IL-7 had no significant effect on *Vpreb1*, *Vpreb2* or *Igll1* (λ5) transcription (Supplementary Fig. 4a–c). In contrast, the addition of CXCL12 repressed *Vpreb1* and *Vpreb2* transcription and essentially silenced *Igll1* transcription. CXCL12-mediated repression was dependent on CXCR4. In vivo, *Vpreb1*, *Vpreb2* and *Igll1* expression were highest in pro-B and large pre-B cells and low in small pre-B cells (Supplementary Fig. 4d–f). Furthermore, *Vpreb1*, *Vpreb2* and *Igll1* were not repressed in *Cxcr4*$^{-/-}$ small pre-B cells. Consistent with these data, in in vitro-cultured pre-B cells, CXCR4 but not IL-7 withdrawal strongly repressed pre-BCR expression (Supplementary Fig. 4g). Overall, these data indicate

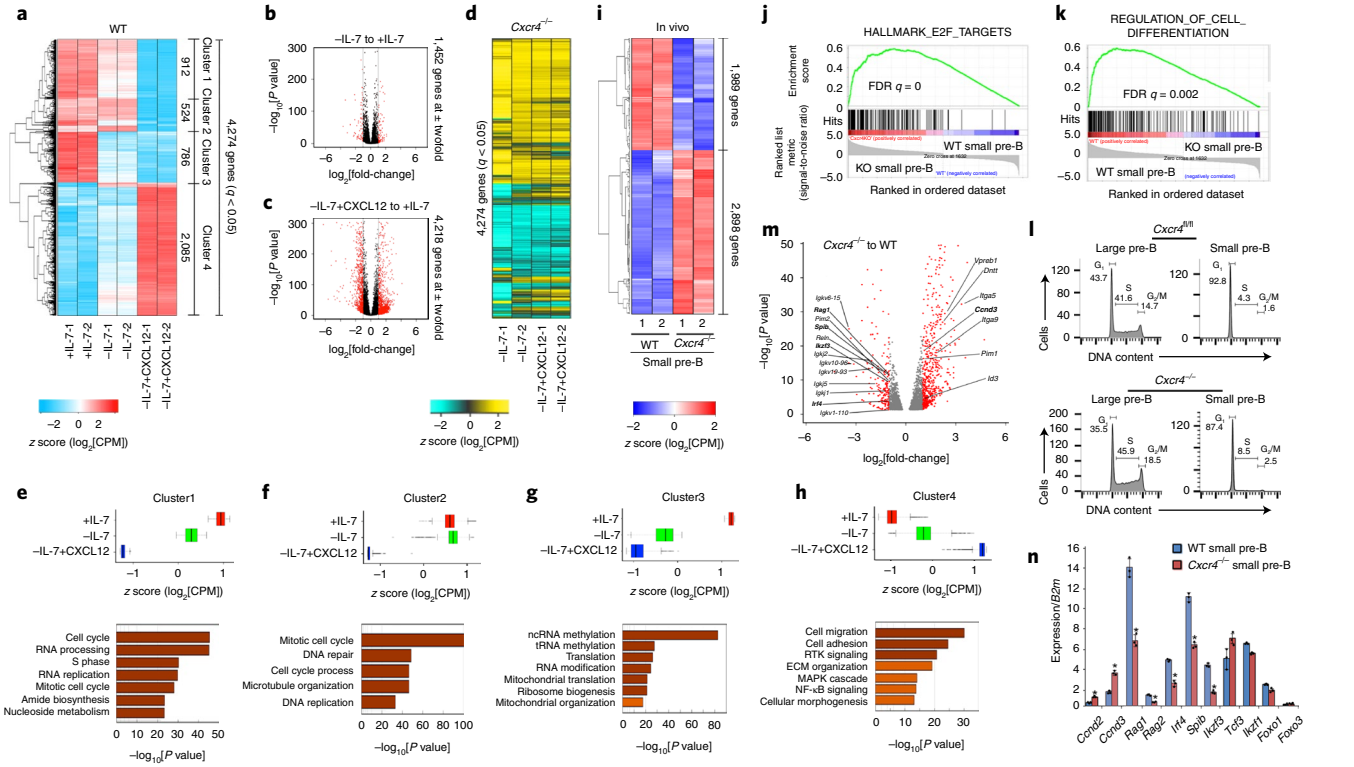

Fig. 4 | Pre-B cell differentiation is directly regulated by CXCR4 signaling. a, Hierarchical clustering of differentially regulated genes (RNA-Seq) identified at $q < 0.05$ in WT pre-B cells (replicates shown) cultured for 48 h with 16 ng ml^{-1} of IL-7 (+IL-7), 0.2 ng ml^{-1} of IL-7 (−IL-7) or 0.2 ng ml^{-1} of IL-7 with 100 ng ml^{-1} of CXCL12 (−IL-7+CXCL12) without stromal cells. CPM, counts per million. **b**,**c**, Volcano plots of the differential expression of genes in WT pre-B cells cultured with or without IL-7 (**b**), or for −IL-7+CXCL12 versus −IL-7 (**c**) ($n = 2$ RNA-Seq; FDR < 0.01). Red dots are genes with significantly increased (right side) or decreased (left side) expression. −log$_{10}$[*P* values] were determined by Kal's statistical test. **d**. Heat map of the same genes in the same order shown in **a**, from RNA-Seq of *Cxcr4*$^{-/-}$ pre-B cells (replicates shown) cultured as indicated. **e**–**h**, Average expression of all of the differentially expressed genes in cluster 1 (**e**), cluster 2 (**f**), cluster 3 (**g**) and cluster 4 (**h**) of WT pre-B cells under +IL-7, −IL-7 or −IL-7+CXCL12 culture conditions (top), and the corresponding gene ontology analyses (bottom). Only terms with FDR < 5%, $P < 0.05$ and greater than twofold log$_2$ enrichment are reported. Gene expression levels for individual RNA-Seq replicates were averaged, and the *z* scores are plotted. Boxes represent interquartile ranges (IQRs; Q1–Q3 percentile) and black vertical lines represent median values. Maximum and minimum values (ends of whiskers) are defined as Q3 + 1.5× the IQR and Q1 − 1.5× the IQR, respectively. Outliers are indicated as black dots along the whiskers. −log$_{10}$[*P* values] were determined by Kal's statistical test. ECM, extracellular matrix; MAPK, mitogen-activated protein kinase; ncRNA, non-coding RNA; RTK, receptor tyrosine kinase; tRNA, transfer RNA. **i**, Heat map of RNA-Seq with clustering of upregulated and downregulated genes in flow-purified *Cxcr4*$^{-/-}$ and WT small pre-B cells ($q < 0.05$; replicates shown). **j**,**k**, GSEA of 'HALLMARK_E2F_TARGETS' (**j**) and 'REGULATION_OF_CELL_DIFFERENTIATION' (**k**). GSEA reports *q* values based on the median of the *P* value distribution. In **k**, pathways were enriched in *Cxcr4*$^{-/-}$ (knockout (KO)) small pre-B cells compared with WT small pre-B cells. **l**, Cell cycle analysis of flow-purified large and small pre-B cells from *Cxcr4*$^{fl/fl}$ and *Cxcr4*$^{-/-}$ mice ($n = 2$). **m**, Volcano plot of differentially expressed genes in *Cxcr4*$^{-/-}$ versus WT small pre-B cells. Red dots are genes with significantly increased (right side) or decreased (left side) expression in *Cxcr4*$^{-/-}$ small pre-B cells compared with WT cells (FDR < 0.01). −log$_{10}$ [*P* values] were determined by Kal's statistical test. Expression of genes whose names are shown in bold was confirmed by quantitative real-time PCR in **n**. **n**, Quantitative real-time PCR analysis of the expression of *Ccnd2*, *Ccnd3*, *Rag1*, *Rag2*, *Irf4*, *Spib*, *Ikzf3* (encodes Aiolos), *Tcf3* (encodes E2A), *Ikzf1* (encodes Ikaros), *Foxo1* and *Foxo3* in flow-purified WT and *Cxcr4*$^{-/-}$ small pre-B cells ($n = 3$). Data are presented as means ± s.d. *P* values were determined by unpaired *t*-test (*$P < 0.001$ compared with expression in WT small pre-B cells). Primers and probes are presented in Supplementary Table 1.

that CXCR4 directly regulates specific developmental programs of late B lymphopoiesis.

CXCR4 signaling determines small pre-B cell identity. Next, we compared the transcriptional programs of WT in vitro-cultured B cell progenitors under different conditions from those of genes differentially expressed by *Cxcr4*$^{-/-}$ and WT small pre-B cells. Approximately, 85% of genes differentially regulated in vitro (3,637 of 4,274 genes) were also differentially expressed in vivo. Clustering of these genes revealed that the direction of regulation by the CXCL12–CXCR4 axis was similar in vitro and in vivo (Fig. 5a).

We then focused on those genes that were differentially regulated in vitro by at least twofold and highly expressed (at least one-tenth *B2m* expression), and in which differential regulation was statistically robust ($P < 10^{-5}$). With these parameters, the withdrawal of IL-7 and addition of CXCL12 induced 219 genes and repressed 388 genes (Fig. 5b). As a control, 125 highly expressed genes were randomly chosen that showed no significant change in transcription. The transcription of those genes induced by CXCL12 was dependent on the expression of CXCR4.

Plotting the expression of each in vitro-regulated gene group as a function of normal B cell development showed that those genes strongly induced by CXCL12 in vitro were also strongly upregulated in vivo (Fig. 5c). The converse was true (that is, those genes repressed by CXCL12 decreased in expression) on transition to the small pre-B cell stage. Those genes that were not regulated in vitro did not change in expression during B lymphopoiesis.

Next, we captured all genes differentially regulated under any in vitro condition or in vivo in WT versus *Cxcr4*$^{-/-}$ small pre-B cells. These genes were then used to cluster the indicated different in vitro

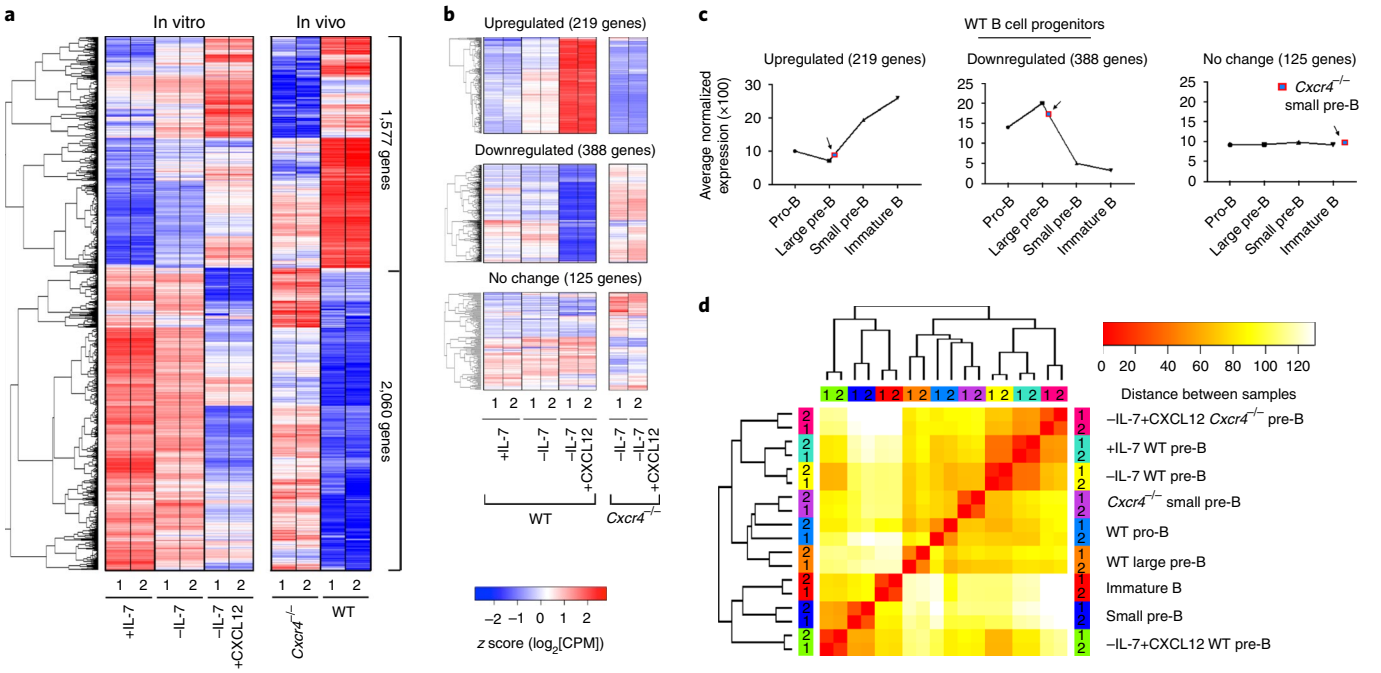

Fig. 5 | CXCR4 signaling determines small pre-B cell identity. a, Hierarchical clustering of differentially regulated genes (RNA-Seq) identified at $q < 0.05$ in WT pre-B cells cultured as indicated compared with flow-purified *Cxcr4*$^{-/-}$ and WT small pre-B cells ($q < 0.05$). Common differentially expressed genes (both in vitro and in vivo) were plotted in the same order. **b**, Heat map of differentially regulated genes in vitro that changed by at least twofold and were highly expressed (at least one-tenth of *B2m* expression), and in which differential regulation was statistically robust ($P < 10^{-5}$). Upregulated, downregulated and representative unchanged genes from the indicated culture conditions are presented in the top, middle and bottom panels, respectively. **c**, Average expression of the same upregulated (left), downregulated (middle) and unchanged genes (right) in different in vivo WT B cell progenitors, as indicated. **d**, Comparison of transcriptional programs regulated by CXCR4, both in vivo and in vitro, based on the expression levels of the 3,637 genes that were differentially expressed under any condition. Hierarchical clustering was run on *z*-scored $\log_2$-normalized expression levels using a Euclidian correlation distance metric. The clustering dendrogram was plotted against the distance matrix for all indicated samples.

and in vivo populations (Fig. 5d). Interestingly, in vitro cultures in which IL-7 was present or absent, or cultures in which CXCL12 was added to *Cxcr4*$^{-/-}$ cells, largely clustered together. These in vitro conditions were most closely related to *Cxcr4*$^{-/-}$ small pre-B cells and then WT pro-B cells. In marked contrast, those in vitro-cultured cells in which IL-7 had been withdrawn and CXCL12 had been added strongly clustered with ex vivo small pre-B cells. These data indicate that the withdrawal of IL-7 and addition of CXCL12 in vitro largely reconstitute the in vivo developmental progression from proliferating pro-B cells to small pre-B cells.

CXCR4 dictates small pre-B cell chromatin accessibility. Examination of chromatin accessibility (by assay for transposase-accessible chromatin using sequencing (ATAC-Seq)) as a function of normal B cell development revealed a progressive loss of accessibility from pro-B cells through the small pre-B cell stage (31,913 peaks; Fig. 6a)[31]. In contrast, in *Cxcr4*$^{-/-}$ small pre-B cells, there were 71,366 open peaks, 81% of which were common with those of proliferative large pre-B cells (Fig. 6b). These data show that, in the absence of CXCR4, small pre-B cells adopt a chromatin landscape with features of large pre-B cells.

Comparison of accessible regions in *Cxcr4*$^{-/-}$ small pre-B cells versus WT small pre-B cells revealed a preferential closing of intragenic and intergenic chromatin areas in WT small pre-B cells (Fig. 6c,d). In total, 56% of intergenic and 73% of intragenic accessibility peaks were in chromatin regions bearing epigenetic enhancer marks (H3K4me1$^+$H3K27Ac$^+$; data not shown). Of the 2,829 accessibility peaks that did not open in *Cxcr4*$^{-/-}$ small pre-B cells, only 16% were at enhancers (data not shown). These data suggest that CXCR4 preferentially silences large pre-B cell-specific enhancers.

Analysis of differentially regulated accessibility peaks for transcription factor binding motifs (Supplementary Fig. 5a) revealed enrichment for those predicted to be recognized by interferon regulatory factor (IRF), Spi-B transcription factor (SPIB), NF-κB and E2A in all accessible regions. However, CXCR4 appeared to specifically repress a subset of RUNX, STAT5 and MYC predicted binding sites, all of which are transcription factors important for earlier stages of B cell development[9,31,34,35]. CXCR4 also repressed accessibility at sites known to bind STAT5 and Myc during B lymphopoiesis (Fig. 6e)[9,35]. Conversely, CXCR4 was required to open the binding motifs for the transcription factors E2A, SPIB and IRF4 (Fig. 6e). CXCR4 also repressed accessibility at sites low or moderately enriched for IRF4, PAX5 and FOXO1 binding sites. It also opened sites poorly enriched for PAX5 and FOXO1 binding sites and highly enriched for IRF4 binding sites (Fig. 6f,g)[29,31]. These findings suggest that, in vivo, CXCR4 preferentially represses STAT5 and Myc binding while opening the genome at sites of IRF4 binding.

Next, we asked how signaling through CXCR4 modulates chromatin accessibility in vitro. In contrast with the in vivo studies, the addition of CXCL12 enhanced overall chromatin accessibility (Fig. 6h). The genomic sites specifically opened by CXCR4 signaling were enriched in SPIB, E2A and SPIB:IRF composite motifs (Fig. 6i and Supplementary Fig. 5b). This was similar to the transcription factor motif sites enriched in WT small pre-B cells compared with *Cxcr4*$^{-/-}$ small pre-B cells. Increased accessibility at sites that can bind IRF4 was observed (Fig. 6j). In addition, CXCR4 signaling was associated with enrichment at sites that can bind PAX5 and FOXO1 (Fig. 6k). As an example, IRF4 binds strongly to the *Jk* proximal promoter and Eκi, which become accessible in both WT small pre-B

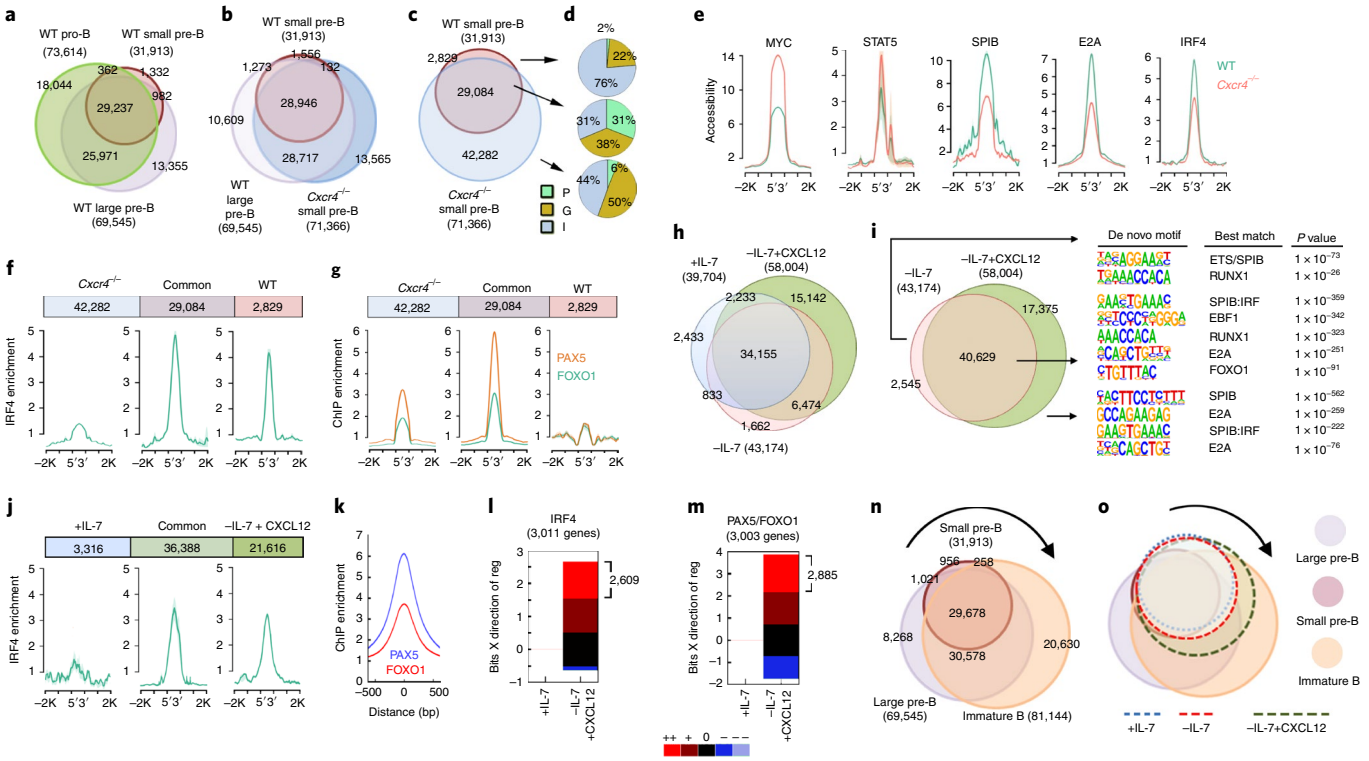

Fig. 6 | CXCR4 signaling sets the small pre-B cell epigenetic landscape. a, Total and overlapping open chromatin regions (ATAC-Seq) in flow-purified WT pro-B, large pre-B and small pre-B cell progenitor populations. **b**, Total and overlapping open chromatin regions (ATAC-Seq) in flow-purified WT large pre-B, small pre-B and *Cxcr4*$^{-/-}$ small pre-B cells. **c**, Overlap of accessible regions from flow-purified WT and *Cxcr4*$^{-/-}$ small pre-B cells. In **a**–**c**, the total number of peaks for each population is shown in parentheses, with the numbers in each Venn region indicated as well. **d**, Distribution of accessible regions across the genome in promoters (P), gene bodies (G) and intergenic regions (I). **e**, Average accessibility (ATAC-Seq) at MYC-, STAT5-, SPIB-, E2A- and IRF4-bound sites (determined from ChIP-Seq data) in WT and *Cxcr4*$^{-/-}$ small pre-B cells. **f**,**g**, Average enrichment at IRF4 binding (**f**) and PAX5 and FOXO1 binding (**g**) (both ChIP-Seq) in unique and common accessible regions of WT and *Cxcr4*$^{-/-}$ small pre-B cells. In **e**–**g**, plots represent the average of two independent ChIP-Seq and ATAC-Seq datasets. **h**, Total and overlapping open chromatin regions (ATAC-Seq) in WT pre-B cells cultured for 48 h, as indicated before ($n = 2$). **i**, Total and overlapping accessible regions in −IL-7- and −IL-7+CXCL12-cultured WT pre-B cells (left) and the de novo transcription factor binding motifs associated with unique and common accessible regions. EBF1, early B cell factor 1; RUNX1, runt-related transcription factor 1; ETS, E26 transformation-specific. **j**, Average enrichment of IRF4 binding (ChIP-Seq) in unique and common accessible regions of +IL-7- and −IL-7+CXCL12-cultured WT pre-B cells. **k**, Average enrichment of PAX5 and FOXO1 binding (ChIP-Seq) in accessible regions of −IL-7+CXCL12-cultured WT pre-B cells. In **j**, and **k**, the plots represent the average of three independent ChIP-Seq datasets for IRF4 and two independent ChIP-Seq datasets for FOXO1 and PAX5, respectively. **l**,**m**, EMBER analysis combining the ChIP-Seq results for IRF4 (**l**) and PAX5/FOXO1 (**m**), with assessment (RNA-Seq) of the predominant expression patterns of genes within 100 kb of the IRF4- (**l**) and PAX5/FOXO1-bound (**m**) sites in WT pre-B cells cultured as indicated. Changes in expression were calculated relative to the expression in +IL-7-cultured WT pre-B cells. In some instances, more than one gene was within 100 kb of a peak, giving more genes than the number of accessible peaks. Changes in mean expression were categorized as follows: −−, <−3 s.d.; −, −1 s.d. to −3 s.d.; 0, −1 s.d. to +1 s.d.; +, 1 s.d. to 3 s.d.; ++, >3 s.d. (where s.d. is the sum of the s.d. values calculated for experimental replicates). **n**, Total and overlapping open chromatin regions (ATAC-Seq) in flow-purified WT large pre-B, small pre-B and immature B cell progenitors. The total number of peaks for each population is shown in parentheses, with numbers in each Venn region indicated as well. **o**, Total and overlapping open chromatin regions in WT pre-B cells cultured as indicated (+IL-7, −IL-7 or −IL-7+CXCL12), overlaid on accessibility levels observed in flow-sorted WT progenitors.

cells and in vitro pre-B cells cultured with CXCL12 (Supplementary Fig. 5c). *Brwd1* activation was associated with enhanced accessibility at sites that can bind IRF4 in the promoter and proximal intronic enhancer, as well as enhanced accessibility at sites that can bind both PAX5 and FOXO1 in the proximal enhancer. FOXO1 alone binds the open 3′ intronic enhancer[29]. Overall, there was increased accessibility at important transcription factor binding sites, including those for IRF, PAX5 and FOXO, in both promoters and enhancers (H3K4me1^{+}H3K27Ac^{+}) of WT small pre-B cells and of –IL-7+CXCL12-cultured pre-B cells (Supplementary Fig. 5d). These in vitro data suggest that CXCR4 signaling plays an additional role in opening PAX5 and FOXO sites in late B lymphopoiesis.

Analysis of gene expression within 100 kilobases (kb) of IRF4-bound peaks revealed 3,011 genes. Of these, 2,609 (86.7%) were strongly induced by IL-7 withdrawal and the addition of CXCL12 (Fig. 6l). Similarly, there were 3,003 genes within 100 kb of a PAX5- and/or FOXO1-bound site. Of these, 2,885 (96.1%) were induced in −IL-7+CXCL12-treated small pre-B cells (Fig. 6m). These data suggest that withdrawal of IL-7 and addition of CXCL12 induce both IRF4 and FOXO1/PAX5 binding to regulate transcriptional programs of late B lymphopoiesis.

In vivo, CXCR4 appeared to repress overall chromatin accessibility, while in vitro it enhanced accessibility. To understand this apparent paradox, we compared those accessibility peaks induced by CXCR4 in vitro with those of later stages of in vivo B cell development (Fig. 6n,o and Supplementary Fig. 5e–g). During normal B lymphopoiesis, there was a radical increase in genomic accessibility from small pre-B cells to immature B cells. Most accessibility peaks opened by CXCR4 in vitro represent peaks found in immature B cells. These sites included both de novo accessibility peaks and peaks that

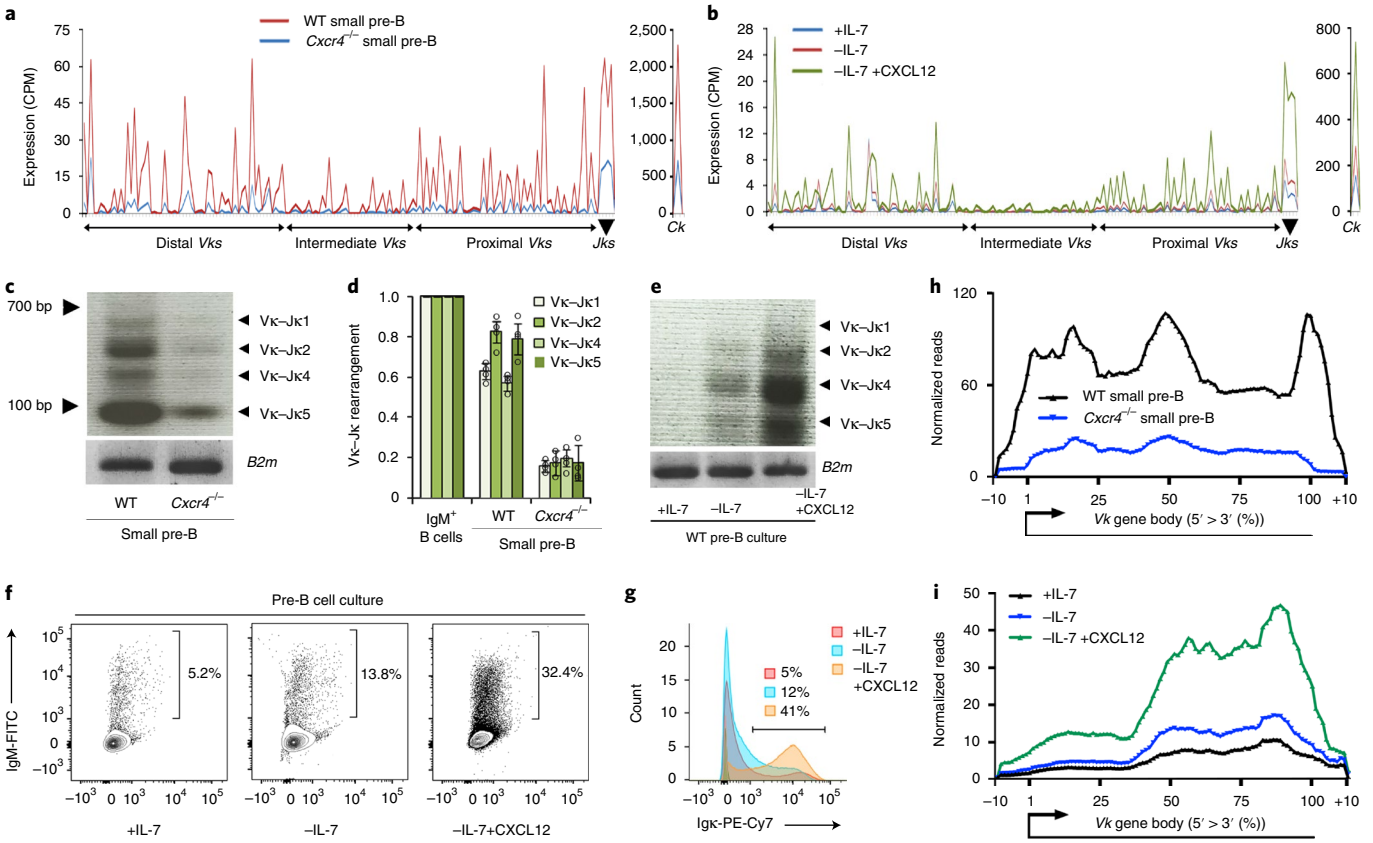

Fig. 7 | CXCR4 signaling is necessary for *Igk* recombination. a,**b**, Expression (RNA-Seq) of all *Vks*, *Jks* and *Ck* genes in WT and *Cxcr4*$^{-/-}$ small pre-B cells (**a**; $n=2$) and pre-B cells cultured as indicated (**b**; $n=2$). **c**, Semi-quantitative PCR analysis of *Igk* rearrangements in WT and *Cxcr4*$^{-/-}$ small pre-B cells. Data are representative of three experiments. **d**, Quantitative analysis of *Vk*–*Jk* rearrangements in WT and *Cxcr4*$^{-/-}$ small pre-B cells compared with WT immature B cells ($n=4$). Data are presented as means ± s.d. *P* values were determined by unpaired *t*-test ($P<0.001$ versus WT small pre-B cells, all measurements). **e**, Semi-quantitative PCR analysis of *Igk* rearrangements in cultured WT pre-B cells. Data are representative of three experiments. **f**, Surface IgM$^+$ B cells after culture of WT pre-B cells as indicated. Data are representative of four experiments. **g**, Histogram of Igκ expression in pre-B cells cultured as indicated ($n=3$). **h**,**i**, Quantified and integrated transcription across all *Vk* gene segments in in vivo B cell progenitors (**h**) and in vitro cultured pre-B cells (**i**), including immediately before and after *Vk* gene bodies.

were previously open in large pre-B cells. Our in vivo data reveal that CXCR4 signaling, and not IL-7 withdrawal, is largely responsible for repressing accessibility in large pre-B cells. Our in vitro data reveal an additional role for CXCR4, at the next stage of B cell development, in contributing to the open chromatin state of immature B cells.

CXCR4 signaling is essential for *Igk* recombination. Initiation of *Igk* rearrangement is the hallmark of pre-B cell differentiation. Therefore, we next examined transcription at *Igk* in WT and *Cxcr4*$^{-/-}$ small pre-B cells. Transcription across the whole *Igk* locus was dependent on CXCR4, both at the *Vk* and *Jk* gene segments (Fig. 7a). To determine the direct role of CXCR4, we examined the corresponding in vitro data (Fig. 7b). Withdrawal of IL-7 only induced modest transcription across the *Igk* locus. In contrast, addition of CXCL12 induced a pattern of transcription broadly similar to that observed in vivo in WT small pre-B cells.

Next, we examined Vκ–Jκ recombination. Semi-quantitative and quantitative PCR of Vκ–Jκ genomic recombination in *Cxcr4*$^{-/-}$ small pre-B cells revealed severely diminished recombination compared with either WT immature or small pre-B cells (Fig. 7c,d). Likewise, Vκ–Jκ recombination was dependent on CXCL12 in vitro (Fig. 7e). Furthermore, staining in in vitro differentiated B cell progenitor cultures for Igμ or Igκ showed that the addition of CXCL12 enhanced the frequency of cells expressing a BCR (Fig. 7f,g). These data show that CXCR4 is required for efficient Vκ–Jκ recombination.

Diminished *Vk* transcription could be a consequence of reduced recombination and/or could reflect diminished accessibility before recombination. Therefore, we quantified and integrated transcription across all *Vk* gene segments including immediately before and after the *Vk* gene bodies (Fig. 7h,i). Transcription of the *Vk* gene bodies occurs both before and after recombination. In contrast, sequences immediately downstream of *Vk* are lost following recombination, and therefore reflect pre-recombination transcription. As can be seen both in vitro and in vivo, *Vk* transcription is globally diminished, including transcription in the immediately downstream flanking regions containing the nine nucleotide *Vk* recombination signal sequences (the beginning of the 100–110% interval). The difference in transcription through the recombination signal sequences is more pronounced in ex vivo cells. This difference in magnitude might reflect the higher rate of ongoing *Igk* recombination observed in vitro in the absence of CXCL12. Therefore, transcription pre-recombination, and therefore accessibility, is dependent on CXCR4.

CXCR4-mediated ERK activation drives B cell development. One of the principal signaling mediators of CXCR4 is ERK[20,22,36]. In vitro stimulation of differentiated B cell progenitors with CXCL12 in the absence of IL-7 increased intracellular phospho-ERK, which was blocked by ERK inhibitor (ERK-i; Fig. 8a). Therefore, B cell progenitors were then cultured in the absence of IL-7, with and without

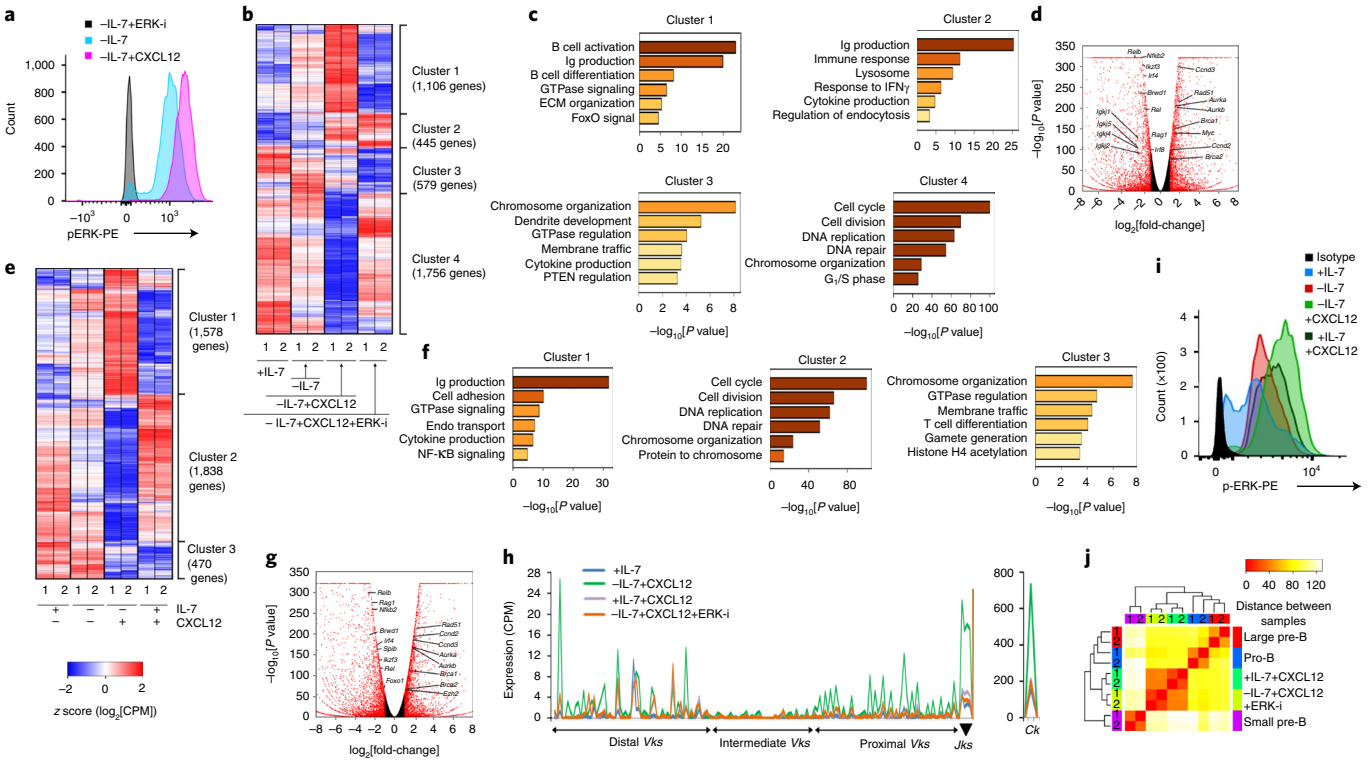

Fig. 8 | ERK signaling downstream of CXCR4 enables pre-B cell differentiation. a, Flow cytometry for intracellular p-ERK in WT pre-B cells cultured as indicated. The negative controls were −IL-7-cultured cells in the presence of ERK-i. Data are representative of three independent experiments. **b**, Hierarchical clustering of differentially regulated genes (RNA-Seq) identified both in vitro and in vivo at $q < 0.05$ in WT pre-B cells (replicates shown) cultured as indicated. **c**, Gene ontology analysis of the clusters identified in **b**. Ig, immunoglobulin; IFN, interferon. **d**, Volcano plot of differentially expressed genes cultured in −IL-7+CXCL12+ERK-i compared with −IL-7+CXCL12. Red points mark genes with significantly increased (right side) or decreased (left side) expression in WT pre-B cells ($n = 2$ RNA-Seq; FDR < 0.01). $-\log_{10}[P\text{ values}]$ were calculated by Kal's statistical test. **e**, Hierarchical clustering of differentially regulated genes (RNA-Seq) identified both in vitro and in vivo at $q < 0.05$ in WT pre-B cells (replicates shown) cultured as indicated. **f**, Gene ontology analysis of the clusters identified in **e**. Endo, endoplasmic. **g**, Volcano plot of differentially expressed genes in WT pre-B cells cultured in +IL-7+CXCL12 versus −IL-7+CXCL12 ($n = 2$ RNA-Seq; FDR < 0.01). Red dots are genes with significantly increased (right side) or decreased (left side) expression. $-\log_{10}[P\text{ values}]$ were calculated by Kal's statistical test. **h**, Expression (RNA-Seq) of all *Vks*, *Jks* and *Ck* genes in WT pre-B cells cultured as indicated ($n = 2$). **i**, Flow cytometry for intracellular p-ERK in WT pre-B cells cultured under +IL-7, −IL-7, +IL-7+CXCL12 and −IL-7+CXCL12 conditions, as indicated. Data are representative of three independent experiments. **j**, Comparison of differential transcriptional programs regulated by CXCR4 signals in WT pro-B, large pre-B and small pre-B cell progenitors and in vitro pre-B cells cultured in the presence or absence of IL-7 with CXCl12, and with or without ERK-i. Hierarchical clustering was run on *z*-scored $\log_2$-normalized expression levels using a Euclidian correlation distance metric. The clustering dendrogram was plotted against the distance matrix for all of the indicated samples.

CXCL12, as well as with CXCL12 plus ERK-i. Under these conditions, four transcriptional clusters were apparent (Fig. 8b,c). In the largest clusters (clusters 1 and 4), treatment with ERK-i largely reversed the transcriptional program induced by CXCL12. Cluster 2 transcription was mostly unaffected by ERK-i, while cluster 3 was repressed by ERK inhibition. Gene ontology analysis of each cluster revealed that ERK both induced immunoglobulin gene production and B cell differentiation programs (cluster 1) and repressed cell cycle programs (cluster 4) (Fig. 8c). The magnitude of ERK-dependent transcription was readily apparent in a volcano plot of differentially regulated genes, which confirmed that ERK was required to repress cell cycle genes and induce transcription of *Igk* and other differentiation genes (Fig. 8d). These data indicate that ERK mediates most of the transcriptional program downstream of CXCR4.

As sensing of CXCL12 is required for moving out of IL-7-rich niches, some B cell progenitors will be subjected to signals through both receptors. Furthermore, pro-B cells tightly associate with IL-7^{int}CXCL12^{+} (type 2) stroma[17]. To understand the consequences of this dual-receptor signaling state, B cell progenitors cultured in IL-7 were stimulated with CXCL12 and subjected to RNA-Seq. Different combinations of IL-7 and CXCL12 stimulation yielded three gene clusters (Fig. 8e,f). In clusters 1 and 2, stimulation with IL-7 and CXCL12 most closely resembled the gene transcription profile induced by IL-7 alone. Remarkably, in these clusters, CXCL12 appeared to reinforce the IL-7-dependent transcriptional state. In contrast, in cluster 3, IL-7 did not modulate normal CXCL12-mediated repression. Genes in cluster 1 included transcribed *Igk* gene segments, while cluster 2 was enriched for cell cycle genes (Fig. 8f). Therefore, the main transcriptional programs regulated by CXCL12 were strongly repressed in the presence of IL-7. Only cluster 3, which includes genes involved in chromatin organization and GTPase regulation, were regulated by CXCL12 in the presence of IL-7 (Fig. 8f). A volcano plot of differentially regulated genes confirms that IL-7 repressed most CXCL12-induced genes (Fig. 8g). These data indicate that for most gene targets, IL-7 is dominant over CXCR4.

Next, we examined how IL-7, CXCL12 and ERK signals integrate to regulate *Igk* transcription (Fig. 8h). The combination of IL-7 withdrawal and CXCL12 strongly induced broad transcription across the *Igk* locus. In contrast, either continual IL-7 signaling in the presence and absence of CXCL12, or inhibition of ERK, failed

to induce *Igk* locus transcription (Fig. 8h). These data indicate that only a precise combination of developmental cues, including escape from IL-7R signaling and ERK-dependent induction through CXCR4, will open the *Igk* locus to recombination.

Next, we examined how signaling through both IL-7R and CXCR4 regulated ERK phosphorylation (Fig. 8i). Interestingly, withdrawal of IL-7 modestly increased phosphorylated ERK (p-ERK). In contrast, in the presence of IL-7 and CXCL12, p-ERK levels were intermediate between +IL-7 and −IL-7+CXCL12. Analysis of RNA-Seq data from cells treated with +IL-7+CXCL12 versus −IL-7+CXCL12+ERK-i revealed strong co-clustering together (Fig. 8j). These data suggest that one major function of IL-7 and IL-7R signaling is inhibition of ERK activation.

IRF4-mediated amplification of CXCR4 expression. Pre-BCR signaling induces IRF4 expression, which has been postulated to drive CXCR4 upregulation[19,37]. Indeed, *Irf4* was induced in large pre-B cells expressing the pre-BCR, and was further upregulated in small pre-B cells (Supplementary Fig. 6a). Increased expression was associated with enhanced accessibility at both *Irf4* enhancers and the promoter (Supplementary Fig. 6b). During development from the pro-B to large pre-B to small pre-B stage, there was a progressive increase in accessibility at the *Cxcr4* locus (Supplementary Fig. 6c), and high *Cxcr4* expression was associated with binding of IRF4, FOXO1 and PAX5 at both the promoter and enhancers in small pre-B cells. These data are consistent with IRF4 driving *Cxcr4* expression.

To determine whether IRF4 could induce *Cxcr4* expression, we infected B cell progenitors cultured with IL-7 with control retrovirus or retrovirus encoding IRF4. Infected cells were isolated by sorting for green fluorescent protein (GFP) and cultured. In the presence or absence of IL-7, cells expressing IRF4 upregulated *Cxcr4* mRNA (Supplementary Fig. 6d), and CXCR4 surface expression in vivo was dependent on IRF4 and IRF8 (Supplementary Fig. 6e). Furthermore, in vitro transwell migration assays revealed that IRF4-transfected cells migrated more robustly along a CXCL12 gradient than mock-transfected control cells (Supplementary Fig. 6f). Finally, we isolated pro-B and large pre-B cells from *Irf4*$^{-/-}$*Irf8*$^{-/-}$ mice and measured chemotaxis in response to CXCL12 (Supplementary Fig. 6g). As demonstrated, chemotaxis was significantly diminished in large pre-B cells in response to CXCL12. These data show that IRF4 induces *Cxcr4* and enhances chemotaxis along CXCL12 gradients. As CXCR4 also induces IRF4, overall, our data suggest a IRF4–CXCR4 feed-forward loop that enhances migration into CXCL12-rich BM niches.

Discussion

Previous models of late B cell development have evoked a two-receptor system in which IL-7R drives proliferation in pro- and large pre-B cells while expression of the pre-BCR directs cell cycle exit and *Igk* recombination[1]. Canonically, the pre-BCR does so both directly and by inducing CXCR4 expression, which allows escape from IL-7-rich BM niches[19]. In this model, the balance between IL-7R and pre-BCR signaling determines whether a cell proliferates or recombines *Igk*. By controlling cell positioning in the BM, it was thought that CXCR4 indirectly regulates the balance between these two receptor determiners of cell state. Here, we demonstrate that CXCR4 plays a direct and fundamental role in inducing the small pre-B cell state, including upregulating the recombination-activating gene (RAG) proteins, repressing pre-BCR expression, and opening *Igk* to recombination. Indeed, many of the biological functions ascribed to the pre-BCR are directly mediated by CXCR4.

Until now, the essential functions of CXCR4 in B lymphopoiesis have not been appreciated. This is probably because previous in vitro systems used OP9 cells, which secrete CXCL12 (refs. [8,9,19,29,30]). Without a defined in vitro system, one cannot dissect changes due to direct CXCR4 signaling versus indirect effects due to positioning relative to IL-7-rich niches. In such a defined system, IL-7 escape and CXCL12 are the essential (and sufficient) extrinsic cues needed to dictate the development of small pre-B cells.

In the BM, small pre-B cells occupied a unique niche closely associated with both CXCL12^{+} stromal cells and extracellular deposits of CXCL12. Furthermore, small pre-B cells in these niches were rich in intracellular CXCL12, suggesting recent endocytosis. This proximity is not simply a consequence of chemotaxis, as large pre-B cells, which migrate more strongly towards CXCL12 in vitro, are rarely in direct contact with CXCL12^{+} stroma. Spatial proximity in the BM is predicted to ensure the strong delivery of CXCL12, specifically to small pre-B cells.

Late B lymphopoiesis requires coordinated induction of both transcription factor networks and chromatin remodeling complexes that dictate which regulatory sites are open to transcription factor binding[1,31]. CXCR4 does both. It induces the expression of IRF4 and NF-κB, which are both critical for late B lymphopoiesis[19,28]. In addition, CXCR4 signaling enhances accessibility at sites bound by multiple mediators of late lymphopoiesis, including IRF, E2A, SPIB, PAX5 and FOXO1. Conversely, sites bound by early mediators of B cell development, such as MYC and STAT5, are closed by CXCR4 signaling. It is likely that at least some CXCR4-dependent changes in chromatin accessibility are mediated by bromodomain and WD repeat-containing protein 1, which both opens enhancers of late lymphopoiesis and represses those targeted by early transcription factor developmental programs[31].

Our data suggest a feed-forward mechanism that both amplifies CXCR4 signaling and enforces the small pre-B cell transcriptional state. Successful rearrangement of *Igk*, and expression of the pre-BCR, induces expression of IRF4. IRF4, in turn, induces expression of CXCR4, which both mitigates the effects of IL-7R signaling and induces a transcriptional program that includes IRF4. IRF4 then feeds forward to further increase CXCR4 expression and signaling. In addition, CXCR4 signaling strongly feeds back to silence pre-BCR expression. We propose that these feed-forward and feed-back loops function to separate proliferative and differentiative states, thereby ensuring genomic integrity[1] and enforcing ordered developmental progression.

The primacy of CXCR4 in directing late B lymphopoiesis begs a reconsideration of pre-BCR function. Clearly, a principal function of the pre-BCR is to induce the IRF4/CXCR4 feed-forward loop. It is also likely that the pre-BCR initiates molecular pathways that complement those activated by CXCR4. Most notably, the pre-BCR induces expression of the anti-apoptotic molecule myeloid cell leukemia 1 (refs. [35,38]). In contrast, CXCR4 does not appear to regulate any component of the apoptotic pathway except for mildly repressing *Bcl2l11* (encodes Bim) and *Bid* (data not shown). The pre-BCR also induces Ikaros and Aiolos, which silence *Ccnd3* (refs. [8,13,39]). *Ikaros* is not regulated by CXCR4. Therefore, the pre-BCR provides both complementary and permissive signals that enable CXCR4-mediated differentiation.

From these studies, and previous work, a new model of late B cell development emerges in which coordinated signals through three receptors dictate differentiation from the pro-B to immature B cell stages[1]. Proliferation in pro-B and large pre-B cells is driven by IL-7R, which also represses *Igk* accessibility. Expression of the pre-BCR, and escape from IL-7R signaling, initiates a differentiation program. However, it is the induction of CXCR4 by IRF4, and the delivery of strong CXCL12-dependent signals, that completes differentiation into small pre-B cells robustly undergoing *Igk* recombination. Pre-BCR signaling, rather than providing transit across a discrete 'checkpoint', initiates a complex CXCR4-dependent program that guides B cell progenitors through a precise developmental program. Therefore, it is the constant interplay and integration of environmental cues, in the context of pre-BCR expression, that determines cell fate and orchestrates B lymphopoiesis.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41590-019-0468-0.

Received: 4 April 2019; Accepted: 16 July 2019;
Published online: 2 September 2019

References

1. Clark, M. R., Mandal, M., Ochiai, K. & Singh, H. Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. *Nat. Rev. Immunol.* **14**, 69–80 (2014).
2. Herzog, S., Reth, M. & Jumaa, H. Regulation of B-cell proliferation and differentiation by pre-B-cell receptor signalling. *Nat. Rev. Immunol.* **9**, 195–205 (2009).
3. Clark, M. R., Cooper, A. B., Wang, L. & Aifantis, I. The pre-B cell receptor in B cell development: recent advances, persistent questions and conserved mechanisms. *Curr. Top. Microbiol. Immunol.* **290**, 87–104 (2005).
4. Cooper, A. B. et al. A unique function for cyclin D3 in early B cell development. *Nat. Immunol.* **7**, 489–497 (2006).
5. Lazorchak, A. S. et al. Sin1-mTORC2 suppresses *rag* and *il7r* gene expression through Akt2 in B cells. *Mol. Cell* **39**, 433–443 (2010).
6. Rathmell, J. C. et al. Akt-directed glucose matabolism can prevent Bax conformation change and promote growth factor-independent survival. *Mol. Cell Biol.* **23**, 7315–7328 (2003).
7. Gottlob, K. et al. Inhibition of early apoptotic events by Akt/PKB is dependent on the first committed step of glycolysis and mitochondrial hexokinase. *Genes Dev.* **15**, 1406–1418 (2001).
8. Mandal, M. et al. Ras orchestrates exit from the cell cycle and light-chain recombination during early B cell development. *Nat. Immunol.* **10**, 1110–1117 (2009).
9. Mandal, M. et al. Epigenetic repression of the *Igk* locus by STAT5-mediated recruitment of the histone methyltransferase Ezh2. *Nat. Immunol.* **12**, 1212–1220 (2011).
10. Mandal, M. et al. Histone reader BRWD1 targets and restricts recombination to the *Igk* locus. *Nat. Immunol.* **16**, 1094–1103 (2015).
11. Herzog, S. et al. SLP-65 regulates immunoglobulin light chain gene recombination through the PI(3)K-PKB-Foxo pathway. *Nat. Immunol.* **9**, 623–631 (2008).
12. 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).
13. Heizmann, B., Kastner, P. & Chan, S. Ikaros is absolutely required for pre-B cell differentiation by attenuating IL-7 signals. *J. Exp. Med.* **210**, 2823–2832 (2013).
14. 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).
15. Sakamoto, S. et al. E2A and CBP/p300 act in synergy to promote chromatin accessibility of the immunoglobulin κ locus. *J. Immunol.* **188**, 5547–5560 (2012).
16. Van Loo, P. F., Dingjan, G. M., Maas, A. & Hendriks, R. W. Surrogate-light-chain silencing is not critical for the limitation of pre-B cell expansion but is for the termination of constitutive signaling. *Immunity* **27**, 468–480 (2007).
17. Fistonich, C. et al. Cell circuits between B cell progenitors and the IL-7$^+$ mesenchymal progenitor cells control B cell development. *J. Exp. Med.* **215**, 2596–2599 (2018).
18. Tokoyoda, K., Egawa, T., Sugiyama, T., Choi, B. I. & Nagasawa, T. Cellular niches controlling B lymphocyte behavior within bone marrow during development. *Immunity* **20**, 335–344 (2004).
19. Johnson, K. et al. Regulation of immunoglobulin light-chain recombination by the transcription factor IRF-4 and the attenuation of interleukin-7 signaling. *Immunity* **28**, 335–345 (2008).
20. Tian, Y. et al. CXCL12 induces migration of oligodendrocyte precursor cells through the CXCR4-activated MEK/ERK and PI3K/AKT pathways. *Mol. Med. Rep.* **18**, 4374–4380 (2018).
21. Song, Z. Y., Wang, F., Cui, S. X., Gao, Z. H. & Qu, X. J. CXCR7/CXCR4 heterodimer-induced histone demethylation: a new mechanism of colorectal tumorigenesis. *Oncogene* **38**, 1560–1575 (2019).
22. Pozzobon, T., Goldoni, G., Viola, A. & Molon, B. CXCR4 signaling in health and disease. *Immunol. Lett.* **177**, 6–15 (2016).
23. Zheng, N. et al. CXCR7 is not obligatory for CXCL12-CXCR4-induced epithelial-mesenchymal transition in human ovarian cancer. *Mol. Carcinog.* **58**, 144–155 (2019).
24. Chatterjee, S., Behnam Azad, B. & Nimmagadda, S. The intricate role of CXCR4 in cancer. *Adv. Cancer Res.* **124**, 31–82 (2014).
25. Trampont, P. C. et al. CXCR4 acts as a costimulator during thymic β-selection. *Nat. Immunol.* **11**, 162–170 (2010).
26. Amin, R. H. et al. Biallelic, ubiquitous transcription from the distal germline Igκ locus promoter during B cell development. *Proc. Natl Acad. Sci. USA* **106**, 522–527 (2009).
27. Hobeika, E. et al. Testing gene function early in the B cell lineage in *mb1*-Cre mice. *Proc. Natl Acad. Sci. USA* **103**, 13789–13794 (2006).
28. Derudder, E. et al. Development of immunoglobulin λ-chain-positive B cells, but not editing of immunoglobulin κ-chain, depends on NF-κB signals. *Nat. Immunol.* **10**, 647–454 (2009).
29. 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).
30. Lagergren, A. et al. The *Cxcl12*, *periostin*, and *Ccl9* genes are direct targets for early B-cell factor in OP-9 stroma cells. *J. Biol. Chem.* **282**, 14454–14462 (2007).
31. Mandal, M. et al. BRWD1 orchestrates epigenetic landscape of late B lymphopoiesis. *Nat. Commun.* **9**, 3888–3902 (2018).
32. Essafi, A. et al. Direct transcriptional regulation of Bim by FoxO3a mediates STI571-induced apoptosis in Bcr-Abl-expressing cells. *Oncogene* **24**, 2317–2329 (2005).
33. Nodland, S. E. et al. IL-7R expression and IL-7 signaling confer a distinct phenotype on developing human B-lineage cells. *Blood* **118**, 2116–2127 (2011).
34. Malin, S. et al. Role of STAT5 in controlling cell survival and immunoglobulin gene recombination during pro-B cell development. *Nat. Immunol.* **11**, 171–179 (2010).
35. Bossen, C. et al. The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth. *Nat. Immunol.* **16**, 775–784 (2015).
36. Montresor, A. et al. CXCR4- and BCR-triggered integrin activation in B-cell chronic lymphocytic leukemia cells depends on JAK2-activated Bruton's tyrosine kinase. *Oncotarget* **9**, 35123–35140 (2018).
37. Thompson, E. C. et al. Ikaros DNA-binding proteins as integral components of B cell developmental-stage-specific regulatory circuits. *Immunity* **26**, 335–344 (2007).
38. Vikstrom, I. B. et al. MCL-1 is required throughout B-cell development and its loss sensitizes specific B-cell subsets to inhibition of BCL-2 or BCL-XL. *Cell Death Dis.* **7**, e2345 (2016).
39. Ma, S. et al. Ikaros and Aiolos inhibit pre-B cell proliferation by directly suppressing c-Myc expression. *Mol. Cell Biol.* **30**, 4149–4158 (2010).

Acknowledgements

We thank M. Olson, R. Ladd and D. Leclerc for cell-sorting services, and the ImmGen Consortium for data assembly. We thank H. Singh (University of Pittsburgh) for providing the PAX5 and FOXO1 ChIP-Seq data, and M. Schlissel (University of Michigan) for providing the *Cκ*-YFP reporter mice. This work is supported by US National Institutes of Health grants AI120715, AI128785 and AI143778 (to M.R.C.), AI120715-02 and AI128785-01A1 (to M.M.), F32AI143120 (to D.E.K.), T32GM007281 (to K.C.M.), UL1TR002003 (to M.M.-C.) and T32HD007009 (to M.K.O.). Part of the bioinformatics analysis was performed by the UIC Research Informatics Core, supported in part by NCATS through grant UL1TR002003.

Author contributions

M.M. and M.R.C. designed the experiments. M.M. carried out and analyzed most of the experiments. M.M. and M.M.-C. analyzed the high-throughput sequencing data. M.K.O and D.E.K. assisted with the ATAC-Seq, RNA-Seq and ChIP-Seq experiments, data analyses and flow cytometry. M.V. and K.C.M. performed some sorting and real-time PCR. J.I. and N.K. assisted with the confocal microscopy. I.A. helped to design some of the experiments. F.G. provided the RaDR-GFP mice. M.M. and M.R.C. oversaw the entire project and wrote the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41590-019-0468-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.M. or M.R.C.

Peer review information: Laurie Dempsey was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Mice. WT, Cκ-YFP (WT), RaDR-GFP (WT), *Cxcr4*$^{-/-}$ and *Cxcr4*$^{fl/fl}$*mb1*-Cre^{+} mice were housed in the animal facilities of the University of Chicago. Male and female mice were used at 6–12 weeks of age, and experiments were carried out in accordance with the guidelines of the Institutional Animal Care and Use Committee of the University of Chicago.

Isolation, culture and flow cytometry of BM B cell progenitors. BM was collected from WT, *Ck*-YFP, *mb1*-Cre$^{+/-}$, *Cxcr4*$^{fl/fl}$ or *Cxcr4*$^{fl/fl}$*mb1*-Cre^{+} mice and cells were resuspended in staining buffer (3% (*v*/*v*) fetal bovine serum (FBS) in phosphate buffered saline). Erythrocytes were lysed and cells were stained with anti-IL-7Rα (CD127; SB/199), anti-CXCR4 (2B11), anti-CD43 (S7), IgM (R6–60.2), IgD (11–36), anti-CD19 (1D3) and anti-B220 (RA3-6B2l) (all from BD Biosciences) as described previously[9,10]. Pre-pro-B cells (Lin^{-}CD19^{-}B220^{+}IgM^{-}), pro-B cells (Lin^{-}CD19^{+}B220^{+}CD43^{+}IgM^{-}), large pre-B cells (Lin^{-}B220^{+}CD43^{-}IgM^{-}FSC^{hi}), small pre-B cells (Lin^{-}B220^{+}CD43^{-}IgM^{-}FSC^{low}) and immature B cells (Lin^{-}B220^{+}CD43^{-}IgM^{+}) were isolated by cell sorting with a FACSAria II (BD Biosciences).

B cell progenitors (B220^{+}IgM^{-}) from WT, RaDR-GFP or *Cxcr4*$^{fl/fl}$*mb1*-Cre^{+} mice were isolated from BM with a MACS separation column (Miltenyi Biotec) and cultured in complete Opti-MEM containing 10% (*v*/*v*) FBS and IL-7 (16 ng ml^{-1}) for 5 d. Further culture for 48 h with 16 ng ml^{-1} IL-7 (+IL-7), 0.2 ng ml^{-1} IL-7 (−IL-7) or 0.2 ng ml^{-1} IL-7 with 100 ng CXCL12 (−IL-7+CXCL12) without any stromal cells was performed before analysis.

For intracellular staining with antibodies to phosphorylated ERK (20A; catalog number: 612593; BD Biosciences), cells were fixed with BD Cytofix buffer and were made permeable with BD Phosflow Perm Buffer II (catalog number: 558030; BD Biosciences). The level of phosphorylation was determined by flow cytometry. Some 5 μM Erk inhibitor II (FR 180204; CAS 865362-74-9; Millipore) was used for last 4 h of the 48-h culture before the assay, as a negative control.

BM microscopy. Both femurs of WT, *Ck*-YFP, *Cxcr4*$^{fl/fl}$ or *Cxcr4*$^{fl/fl}$*mb1*-Cre^{+} mice were isolated and put in DMEM+ 10% FBS medium. After removing ends, the body of the femur was connected to a syringe filled with media. Intact BM was flushed out gently, transferred into a mold containing OCT, and frozen in 2-methylbutane in dry ice immediately. Frozen BM was sectioned (8–10 μm) serially. For tissue staining, the sections were fixed in 4% paraformaldehyde, blocked with 10% normal donkey serum and stained with primary antibodies to B220 (clone RA3-6B2 (eBioscience) and ab64100 (Abcam)), IgM (clone ll/41 (eBioscience) and FITC-conjugated Goat anti-mouse (115-097-020/115-096-075; Jackson)), Ki67 (clone SP6; Ab16667; Lot#GR59808-1; Abcam), IL-7 (M-19 polyclonal; sc1268; Lot#C2408; Santa Cruz Biotechnology) or CXCL12 (FL-93 polyclonal (Santa Cruz Biotechnology) and ab18919 (Lot#GR116-13; Abcam) in various combinations and thereafter incubated with fluorochrome-conjugated secondary antibodies specific to the primary species and isotypes (Invitrogen). Finally, 4′,6-diamidino-2-phenylindole (DAPI; Invitrogen) was applied to stain the nucleus. Images were acquired at 12-bit depth and 1,024 × 1,024 pixel size, utilizing either the SP5 or SP8 laser scanning confocal microscope (Leica). Raw images were stored in manufacturer-specified .lif format and converted to multichannel .tif images for further use. *z*-stack images were obtained using the Leica SP8 laser scanning confocal microscope in 0.5-μm increments and processed by Imaris software for three-dimensional reconstruction.

Quantitative PCR, RNA-Seq and analysis. Total cellular RNA was isolated with an RNeasy kit (Qiagen) and RNA was reverse-transcribed with SuperScript III reverse transcriptase (Invitrogen). Then, quantitative PCR in quadruplicate was performed using SYBR Green PCR Master Mix (Applied Biosystems). Gene expression was analyzed with an ABI PRISM 7300 Sequence Detector and ABI PRISM Sequence Detection Software version 1.9.1 (Applied Biosystems). The results were normalized by division of the value for the unknown gene by that obtained for *B2m*. Sequences of primers are presented in Supplementary Table 1.

For RNA-Seq, mRNA was isolated by oligo-dT beads and a library was prepared using the standard Illumina library protocol (kit: RS-122-2101 TruSeq Stranded mRNA LT-SetA). Libraries were sequenced on the Illumina HiSeq 2500. Quality control-passed reads then aligned to the mm9 reference genome in a splice-aware manner using the STAR aligner. Gene expression was quantified using cuffquant, and normalized expression levels and differential expression levels were generated with cuffnorm and cuffdiff, respectively (version 2.2.1), as described previously[31]. Differential expression statistics (fold-change and *P* value) were computed, using DESeq2 and edgeR, on raw expression counts obtained from quantification. *P* values were adjusted for multiple testing using the false discovery rate (FDR) correction of Benjamini and Hochberg. All heat maps use hierarchical clustering based on the *z* score of $\log_2$-transformed normalized expression between various experimental conditions.

IL-7 and CXCL12 expression by BM stromal cells. BM was isolated from WT mice. Red blood cells were lysed, and the remaining cells were plated in six-well plates in complete medium to allow stromal cells to adhere to the bottom of the well. Non-adherent cells were aspirated, and flow cytometry was performed on the remaining adherent cells. Contaminating hematopoietic lineage cells and red blood cells were first excluded from our analysis using the CD45 and Ter119 markers, respectively. Then, CXCL12 and IL-7 expression was assessed on the remaining BM stromal cells by flow cytometry using the same antibodies used for microscopy (IL-7 (M-19; sc1268; Lot#C2408; Santa Cruz goat anti-mouse) and CXCL12 (ab25117; Lot#GR116-13) (both from Abcam)).

Transwell migration assay. Migration assays were performed as described previously[19]. MACS column-purified WT B220^{+} cells (0.25–0.5 × 10^{6} cells 100 μl^{-1}) were placed in the upper compartment of a transwell chamber (5 μm pore size; Corning 3421) with 600 μl medium containing 20 ng ml^{-1} recombinant IL-7 (catalog number: 407-ML; R&D Systems) or 100 ng ml^{-1} recombinant CXCL12 (catalog number: 460-SD-010; R&D Systems). The numbers and proportions of cells before (as input in the upper chamber) and after 3 h (migrated into the lower chamber) were measured by flow cytometry. The chemotaxis was expressed as either a percentage or a fold-change relative to the number of input cells.

Cell cycle analysis. Cells were incubated in a solution containing propidium iodide and then analyzed by flow cytometry as described previously[8]. The proportions of cells in the G_1, S and G_2-M phases of the cell cycle were analyzed with FlowJo.

PCR analysis of *Igk* rearrangements. Genomic DNA (or complementary DNA) was isolated from small pre-B cells of WT, *Cxcr4*$^{fl/fl}$ and *Cxcr4*$^{fl/fl}$*mb1*-Cre^{+} mice, and from WT pre-B cells cultured for 48 h without stoma, as described above[8–10] (primers are shown in Supplementary Table 1). *B2m* expression was used to control for the amount of genomic DNA. The intensity of the band for each rearrangement product was normalized whenever necessary to values obtained from IgM^{+} B cells, given a value of 1.

Retroviral transduction. The complementary DNA encoding mouse IRF4 was subcloned into the retroviral vector MIGR1-GFP[8]. Retroviruses containing constructs were produced by transient transfection of PLAT-E packaging cell lines. Infection of B cell progenitors from *Irf4*$^{-/-}$*Irf8*$^{-/-}$ was done as described previously[8,9]. After 48 h, GFP^{+} cells were isolated by cell sorting and B cell progenitors were cultured in complete medium with IL-7 at a high (16 ng ml^{-1}) or low concentration (0.2 ng ml^{-1}). Then, the expression of *Cxcr4* and chemotaxis to CXCL12 were measured by quantitative real-time PCR and transwell migration assay, respectively, as described above.

GSEA. Normalized $\log_2$[counts per million] values from RNA-Seq data and the HALLMARK database, which has 50 pathways, were used as done previously[31]. GSEA was performed using the program offered at http://software.broadinstitute.org/gsea/index.jsp, using default options. Enrichment Map was used for visualization of the GSEA results.

ATAC-Seq. ATAC-Seq was performed as described previously[10,31]. All raw sequence data were quality trimmed to a minimum phred score of 20 using Trimmomatic[40]. Alignment to the reference genome mm9 was done with BWA. For the ATAC-Seq data, read pairs for which one pair passed quality trimming but the other did not were aligned separately and merged with the paired-end alignments. PCR duplicates were removed using Picard's MarkDuplicates, and alignments with an edit distance greater than 2 to the reference, or that were mapped multiple times to the reference, were removed.

ATAC-Seq analysis followed the procedure described previously[10,31]. We used the peak calling results as a guide to identify regulatory elements, then quantified enrichment in these regulatory elements and ran differential analysis to compare the samples. Reproducibility was factored into the differential statistics that we calculated using estimates of dispersion.

ATAC-Seq and chromatin immunoprecipitation sequencing (ChIP-Seq) peak calling and motif analysis. Peaks for ChIP-Seq samples were called using MACS2, as done previously[10,31]. HOMER software (hypergeometric optimization of motif enrichment) for de novo motif discovery and next-generation sequencing analysis was used for new prediction of motifs in the peaks.

Comparison of ChIP-Seq data and mRNA expression. The EMBER program[9] was used for the identification of genes targeted by IRF4, PAX5 and FOXO1 binding and how those genes acted over the in vitro culture in the presence (+IL-7) or absence (−IL-7) of −IL-7, and in the presence of CXCL12 under attenuated IL-7 conditions (−IL-7+CXCL12). EMBER integrates transcription factor binding data with RNA-Seq expression data and uses an unsupervised learning algorithm to identify genes targeted by the transcription factor. This is done by defining a set of pairwise comparisons, making the changes in expression mathematically discrete and searching for over-represented patterns among these data for the genes within 100 kb of the transcription factor binding sites. Only genes that matched an expression pattern were selected; therefore, not all transcription factor binding sites were assigned to a target gene.

Statistical analysis. Data were analyzed by unpaired *t*-test and analysis of variance, followed by the test of least-significant difference for comparisons within and between groups. All categories in each analyzed experimental panel were compared and $P < 0.05$ was considered significant. All *P* values <0.001 were rounded to facilitate comparisons of the results.

Accession codes. Gene Expression Omnibus accession codes for our publicly available datasets are as follows: GSE103057 (ATAC-Seq and RNA-Seq for WT pro-B, WT large pre-B, WT small pre-B and WT immature B cells); GSE103057 (IRF4 ChIP-Seq for WT small pre-B cells); GSE69478 (RAG1, RAG2 and H3K4me3 ChIP-Seq for small pre-B cells); GSE40173 (MYC ChIP-Seq); GSE31039 (H3K27ac ChIP-Seq in CH12 cells); and GSE31039 (BM H3K4me1, H3K4me3 and H3K27ac ChIP-Seq). The dataset for H3K4me1 ChIP-Seq in CH12 cells is available from ENCODE (with the accession code ENCSR000DHQ), and the STAT5 ChIP-Seq dataset is available from ref. [9].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The sequences, ATAC-Seq data for WT small pre-B cells, *mb1*-Cre^{+}*Cxcr4*$^{fl/fl}$ (*Cxcr4* knockout) small pre-B cells, +IL-7, −IL-7 and −IL-7+CXCL12 cultured WT pre-B cells, and −IL-7 and −IL-7+CXCL12 cultured *Cxcr4*$^{-/-}$ small pre-B cells, as well as all of the corresponding RNA-Seq data, including for +IL-7+CXCL12 and −IL-7+CXCL12+ERK-i cultured pre-B cells, are deposited in the GenBank database (accession number GSE129311).

References

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

nature research

Corresponding author(s): Malay Mandal, Marcus Clark

Last updated by author(s): Jul 2, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	commercial Softwares 1. FACSDiva V8.0. l 2. ABI Prism Sequence Detection v 1.9.1, 3. Imaris 3D
Data analysis	commercial Softwares 1. FlowJo 10.0.8r.1 for analyses Free Softwares 1. BWA MEM v0.7.Sa. 2. Picard MarkDuplicates vl.107 3. Macs2 (v2.l.0) 4. --SPM R flag 5. UCSC tool bedGraphToBigWig 6. HOMER 7. NGS plots 8. cuffnorm, cuffdiff v 2.2.1, 9. DESeq2, 10. edgeR, 11. GSEA, 12. trimmomatic, 13. EMBER No custom code was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Accession no GSE129311 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129311).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed as studies were performed on inbred, genetically identical mice.
Data exclusions	No data was excluded.
Replication	All attempts at replication were successful. Replication was repeated 2-9 times to confirm the results. Sometimes the experiment was done by an another individual of the lab.
Randomization	Experiments were all done with independent replicates with indicated controls. As experiments were done in genetically identical mice, no randomization was necessary.
Blinding	Investigators were not blinded. However, much of the bioinformatic analysis was done by an independent person who did not perform the actual experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

All antibodies were used at a dilution of 1:350 except otherwise mentioned.
anti-CDllc (HL3), BD Biosciences, 550283,
anti-NKl.1 (PK136), BD Biosciences , 560515,
anti-CD71 (C2), BD Biosciences, 562716,
anti-Ter119 (TER-119), BD Biosciences, 553673,
anti-Mac-1 (Ml/70), BD Biosciences, 557960,
anti-Gr-1 (RB6-8C5), BD Biosciences, 550291,
anti-CD34 (RAM34), BD Biosciences, 560238, (1:200)
anti-Seal (Ly-6A/E, D7), BD Biosciences, 553108,
anti-cKit {CDll 7, 2B8), BD Biosciences, 553352, (1:250)
anti-Flt3 (CD135, A2F10.1), BD Biosciences, 553842, (1:250)
Validation

anti-CD4 (H129.19), BD Biosciences, 553647,
anti-CDS (53-6. 7), BD Biosciences, 550281,
anti-CD25 (IL2Ra, 7D4), BD Biosciences, 553075,
anti-CD44 (I M7), BD Biosciences, 550538,
anti-CCR9 (CWl.2), BD Biosciences, 565412,
anti-CXCR4 (2Bll), BD Biosciences, 555976, (1:50)
anti-CD43 (57), BD Biosciences, 553271,
lgM (R6-60.2), BD Biosciences, 553406,
lgD (11-36), BD Biosciences, 560868,
anti-CD19 (103), BD Biosciences, 557655,
anti-B220 (RA3-6B21), BD Biosciences,
anti-CD93 (AA4.1), BD Biosciences, 563805,
CD21 (7G6), BD Biosciences, 558658,
CD23 (B3B4); BD Biosciences, 553139,
anti-lRF4 (M-17, Santa Cruz Biotechnogy, sc-6059X, lot#J2015), (1:200)
anti-lRF4-AF647 (IRF4.3E4), Biolegend, 646408 , lot#B240433,(1:200)
phosphorylated-Erk (20A; Cat#612593, BD Biosciences Pharmingen), (1:100)
B220 (ab64100, Lot#GR29057-4, Abeam and RA3-6B2, BD Biosciences, (1:100)
lgM (FITC; Goat anti-mouse Jackson 115-097-020/115-096-075; APC anti-mouse Jackson 17-5790-82
Nat lmmunol. 2009 Oct; 10(10): 1110-1117; (1:100)
Ki-67 (SP6, Ab16667, Abeam, Lot#GR59808-1; rabbit mAb anti-mouse;
IL-7 (M-19, sc1268, Lot#C2408, Santa Cruz goat anti-mouse), (1:50)
CXCL12 (ab18919, rabbit anti-mouse and ab25117, Lot#GR116-13 both from Abeam; (1:50)
DAPI (lnvitrogen) .
Alexa-Fluor 647 Chicken/Donkey anti-rat lgG (H+L; A-21472, ThermoFisher/ab150155, Abeam),(1:500)
Alexa-Fluor 594/546/647 Donkey anti-goat lgG (A-11058, A-11056, A-215447 ThermoFisher),(1:500)
Alexa-Fluor 546 Donkey anti-Rabbit/Chicken lgG (A-10040/A-21441, ThermoFisher).(1:500)
AlexaFluor 488 Donkey anti-rabbit lgG (A21206 ThermoFisher). (1:500)

Validation

AlexaFluor 488 Donkey anti-rabbit lgG (A21206 ThermoFisher).
anti-CDllc (HL3), J Leukoc Biol. 2008; 84(4) :1039-1046
anti-NKl.1 (PK136), J Exp Med. 1997; 186(12):1957-1963
anti-CD71 (C2), J Hist Cytochem. 1996; 44(8) :929-941.
anti-Ter119 (TER-119), Br J Haematol. 2000; 109(2):280-7.
anti-Mac-1 (Ml/70), Eur J lmmunol. 1979; 9(4) :301-306.
anti-Gr-1 (RB6-8C5), J Virol. 1996; 70(2):898-904.
anti-CD34 (RAM34), Nature. 2000; 404(6774) :193-197
anti-Seal (Ly-6A/E, D7), Blood. 1997; 89(6) :1915-1921
anti-cKit (CDll 7, 2B8), Cell. 1990; 63(1):235-243.
anti-Flt3 (CD135, A2F10.l), Nature. 1994; 368(2) :643-8.
anti-CD4 (H129.19), Nature. 1991; 349(6304) :71-74
anti-CDS (53-6.7), Nature. 1989; 342(6247) :278-281.
anti-CD25 (IL2Ra, 7D4), Cell. 1993; 73(1) :5-8
anti-CD44 (IM7), Science. 1996; 271(5248):509-512
anti-CCR9 (CWl.2), Immunity. 2011; 34(4) :602-615.
anti-CXCR4 (2Bll), Cell. 1996; 87(4):745-756.
anti-CD43 (57), Nature. 1995; 377(6549):535-538
lgM (R6-60.2), J lmmunol. 2011; 187(8):3942-3952.
lgD (11-36), J lmmunol. 1993; 150(6):2253-2262.
anti-CD19 (103), Nature. 1995; 376(6538):352-355.
anti-B220 (RA3-6B21), Nat lmmunol. 2009 Oct; 10(10): 1110-1117.
anti-CD93 (AA4.1), J Exp Med. 1999; 189(4):735-740.
CD21 (7G6), Science. 1998; 280(5363):582-585.
CD23 (B3B4); Nature. 1994; 369(6483):753-756.
anti-lRF4 (M-17), Immunity. 2013 May 23; 38(5): 918-929.
anti-lRF4-AF647 (IRF4.3E4), Nature.2009.;458:351-356
phosphorylated-Erk (20A); Nat lmmunol. 2009 Oct; 10(10): 1110-1117.
B220 (ab64100, and RA3-6B2), rat anti-mouse CD45R; Suitable for: IP, IHC-P, IHC-Fr, Flow Cyt, ICC/IF), Leukemia 32:364-375 (2018); Mol Med Rep 17:7701-7707 (2018)
lgM (FITC and APC); Goat anti-mouse, Nat lmmunol. 2009 Oct; 10(10): 1110-1117; Nat Commun. 2018 Sep 28;9(1):3973
Ki-67 (SP6, Ab16667)Lot#GR59808-1; rabbit mAb anti-mouse; IHC-FoFr, ICC/IF, Flow Cyt, IHC-Fr, WB, IHC-P)
Exp Neurol 311:15-32 (2019); Cell Biochem Funct 37:11-20 (2019)
IL7 (M-19, sc1268, Lot#C2408, goat anti-mouse), Science 257: 379-382; Immunity. 2008 Mar;28(3):335-45;
also tested by flow cytometry with negative control (IL-7 non expressing cells) and isotype control .
CXCL12 (ab18919 and ab25117 rabbit anti-mouse, Lot#GR116-13) suitable for IHC-Fr, IHC-P, IP, WB)
Nat Cell Biol 15:284-94 (2013); Nat Commun 4:1795 (2013);
also tested by flow cytometry.
DAPI (lnvitrogen) . Science (2002) 296:1860-1864; J Biol Chem (1999) 274:38083-38090
Alexa-Fluor 647 Chicken/Donkey anti-rat lgG (H+L; A-21472), Nat Commun. 2018 Jun 19;9(1):2409
Alexa-Fluor 594/546/647 Donkey anti-goat lgG (A-11058, A-11056, A-215447), Neuron. 2017 Apr 5;94(1) :138-152.e5
Alexa-Fluor 546 Donkey anti-Rabbit/Chicken lgG (A-10040/A-21441). Nat Commun. 2017 Feb 3;8:14172.
AlexaFluor 488 Donkey anti-rabbit lgG (A21206). Nat Methods. 2011 Nov 6;8(12) :1083-8

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	rf4-/-lrf8-/- B cell line were used. B Cell progenitors were isolated from IRF4 and IRF8 double deficient mice and grown in IL-7.
Authentication	The cells were collected from directly from bone marrow of lrf4-/-lrf8-/- mice and grown in recombinant IL7 as described previously in 1. Immunity. 2008 Mar;28(3):335-45. 2. Nat lmmunol. 2009 Oct;l0(l0):1110-7 3. Nat lmmunol. 2011 Oct 30;12(12) :1212-20 4. Nat lmmunol. 2012 Jan 22;13(3):300-7 the cultured cells were genotyped for to confirm IRF4 and IRF8 deletion by PCR.
Mycoplasma contamination	All cell line tested are negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Wild type (WT), Ck-YFP (WT), mbl-Cre+/-, Cxcr4fl/fl and mbl-Cre+Cxcr4fl/fl mice (both sexes) in B6 background of 6-12 weeks of age
Wild animals	study did not involve wild animals.
Field-collected samples	study did not involve samples collected from the field.
Ethics oversight	Experiments were carried out in accordance with the guidelines of the Institutional Animal Care and Use Committee of the University of Chicago.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Accession no GSE129311 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129311). and https://www.ncbi .nlm.nih.gov/geo/query/acc.cgi ?acc=GSE10305 7
Files in database submission	GSM3704394 mm_atac_spreb_cxcr4ko.l GSM3704395 mm_atac_spreb_cxcr4ko.2 GSM3704396 mm_atac_preb_il7plus.l GSM3704397 mm_atac_preb_il7plus.2 GSM3704398 mm_atac_preb_il7minus.1 GSM3704399 mm_atac_preb_il7minus.2 GSM3704400 mm_atac_preb_il7minus-cxcl12plus.1 GSM3704401 mm_atac_preb_il7minus-cxcl12plus.2 GSM3704402 mm_rna_spreb_wt.l GSM3704403 mm_rna_spreb_wt.2 GSM3704404 mm_rna_spreb_cxcr4ko.l GSM3704405 mm_rna_spreb_cxcr4ko.2 GSM3704406 mm_rna_preb_il7plus.l GSM3704407 mm_rna_preb_il7plus.2 GSM3704408 mm_rna_preb_il7minus.1 GSM3704409 mm_rna_spreb_il7minus.2 GSM3704410 mm_rna_preb_il7minus-cxcl12plus.1 GSM3704411 mm_rna_preb_il7minus.cxcl12plus.2 GSM3704412 mm_rna_spreb_cxcr4ko_il7minus.l GSM3704413 mm_rna_spreb_cxcr4ko_il7minus.2 GSM3704414 mm_rna_spreb_cxcr4ko_il7minus.cxcl12plus.l GSM3704415 mm_rna_spreb_cxcr4ko_il7minus.cxcl12plus.2 GSM3704416 mm_rna_preb_il7plus-cxcl12plus.1 GSM3704417 mm_rna_preb_il7plus-cxcl12plus.2 GSM3704418 mm_rna_preb il7m inus-cxcl 12plus-ERKi nhi bitorl lplus. l

	GSM3704419 mm_rna_preb_il7minus-cxcl12plus-ERKinhibitorllplus.2
Genome browser session (e.g. UCSC)	Accession no GSE129311 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129311). and https://www.ncbi .nlm.nih.gov/geo/query/acc.cgi ?acc=GSE103057

Methodology

Replicates	3 for each sample, RAW, BigWig (BW) and BED
Sequencing depth	(aligned reads after removal of PCR duplication) mm_chip_spreb_irf4.1: 23704156 mm_chip_spreb_irf4_lowdna.l: 60641987 mm_chip_spreb_irf4_lowdna.2 : 39825495
Antibodies	anti-lRF4 (M-17, Santa Cruz Biotechnogy, sc-6059X, lot#J2015)
Peak calling parameters	Raw reads were aligned to the reference genome using BWA MEM v0.7.Sa. Apparent PCR duplicates were removed using Picard MarkDuplicates vl.107. Peaks were called against specified inputs using Macs2 (v2.l.0); normalized bedgraph tracks were generated using the -SPMR flag, and converted to bigWig using UCSC tool bedGraphToBigWig. Peaks with a score >5 were retained.
Data quality	All raw sequence data was quality trimmed to a minimum phred score of 20 using trimmomatic. Apparent PCR duplicates were removed using Picard MarkDuplicates vl.107. Peaks with a score >5 were retained. Details are in "Methods".
Software	BWA MEM v0.7.Sa. Picard MarkDuplicates vl.107 Macs2 (v2.l.0) --SPMR flag UCSC tool bedGraphToBigWig

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Bone marrow (BM) was collected from WT, Ccxr4-fl/fl or Cxcr4-fl/fl X mbl-cre+/- mice. Cells resuspended in staining buffer {3% (vol/vol) FBS in PBS). Erythrocytes were lysed and cells were stained with the antibodies described in the reporting summary and methods section of the manuscript. For cultured pre-B cell with or without IL-7 and CXCL12 were directly washed and resuspended in staining buffer {3% (vol/vol) FBS in PBS). Then the cells were stained with the antibodies described in the reporting summary and methods section of the manuscript.
Instrument	FACS Aria II (BD) for sorting. LSR Fortessa 4-15 (BD) for data collection
Software	FACSDiva V8.0. l for data collection FlowJo 10.0.8r.1 for analyses
Cell population abundance	Sorted samples were re-run on flow cytometer to ensure purity.
Gating strategy	To analyze bone marrow cells, first progenitors/lymphocytes were discriminated from granulocytes by FSC and SSC when necessary. Doublets were excluded using FSC-H and FSC-A gate. All downstream analyses were performed on singlets.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.