JEM Journal of Experimental Medicine

ARTICLE

Dissection of progenitor compartments resolves developmental trajectories in B-lymphopoiesis

Christina T. Jensen[1]*, Josefine Åhsberg[1]*, Mikael N.E. Sommarin[1], Tobias Strid[1,2], Rajesh Somasundaram[2], Kazuki Okuyama[2], Jonas Ungerbäck[2], Jussi Kupari[3], Matti S. Airaksinen[3], Stefan Lang[1], David Bryder[1], Shamit Soneji[1], Göran Karlsson[1], and Mikael Sigvardsson[1,2]

To understand the developmental trajectories in early lymphocyte differentiation, we identified differentially expressed surface markers on lineage-negative lymphoid progenitors (LPs). Single-cell polymerase chain reaction experiments allowed us to link surface marker expression to that of lineage-associated transcription factors (TFs) and identify GFRA2 and BST1 as markers of early B cells. Functional analyses in vitro and in vivo as well as single-cell gene expression analyses supported that surface expression of these proteins defined distinct subpopulations that include cells from both the classical common LPs (CLPs) and Fraction A compartments. The formation of the GFRA2-expressing stages of development depended on the TF EBF1, critical both for the activation of stage-specific target genes and modulation of the epigenetic landscape. Our data show that consecutive expression of Ly6D, GFRA2, and BST1 defines a developmental trajectory linking the CLP to the CD19+ progenitor compartment.

Introduction

Despite the fact that B cell development represents one of the most carefully characterized developmental pathways in the hematopoietic system, the identity of the earliest committed B cell progenitor remains obscure. Even though the B220+CD43+ pre-pro–B or Fraction A (FrA) cells (Hardy et al., 1991; Li et al., 1993) contain a substantial fraction of early B-lineage progenitors, it constitutes a heterogeneous population of cells with varying lineage potentials. Despite the development of more advanced isolation strategies (Sen et al., 1990; Rolink et al., 1994; Li et al., 1996; Tudor et al., 2000), a large portion of the cells in the B220+CD19− FrA subpopulations retain T-lineage potential (Martin et al., 2003; Rumfelt et al., 2006; Mansson et al., 2010) as well as the ability to generate myeloid cells (Alberti-Servera et al., 2017). The difficulty in identifying CD19-negative lineage committed B cell progenitors indicated that B-lineage cell fate is associated with CD19 expression (Rumfelt et al., 2006). This would be in line with the fact that CD19 is a direct target gene for the transcription factor (TF) PAX5, which forms a regulatory network with EBF1, FOXO1, and TCF3 to establish stable B-lineage commitment (Nutt et al., 1999; Rolink et al., 1999; Mikkola et al., 2002; Ikawa et al., 2004; Pongubala et al., 2008; Welinder et al., 2011; Mansson et al., 2012; Nechanitzky et al., 2013). However, by using mice carrying an *Igll1* (λ5) reporter gene (human CD25 [hCD25]) (Mårtensson et al., 1997), it was possible to prospectively isolate B cell–committed

progenitors among CD19-negative cells (Mansson et al., 2008, 2010). These B-lineage–committed population phenotypically belongs to a lineage-negative (Lin−) B220−SCA1intKITintIL7R+FLT3+ common lymphoid progenitor (LP [CLP]) compartment (Kondo et al., 1997; Karsunky et al., 2008) originally thought to consist of multipotent cells with potential to differentiate to all lymphoid lineages. Further exploration of the CLP compartment revealed that functionally distinct subpopulations could be identified based on the expression of a Rag1 reporter gene (Igarashi et al., 2002; Mansson et al., 2010) or the surface marker Ly6D (Inlay et al., 2009; Mansson et al., 2010). Despite the fact that Ly6D+ LP cells generated mainly B-lineage cells after transplantation (Inlay et al., 2009), single-cell (SC) analysis indicated that a substantial fraction of these progenitors still possessed a T-lineage potential that could be evoked by a strong Notch signal (Mansson et al., 2010). Hence, there exists an unresolved heterogeneity in the CD19− progenitor population, which obscures our current understanding of B cell commitment.

To gain insight into the earliest events in B-lymphopoiesis, we applied a strategy where we combine an antibody library screen with genome-wide expression analyses to identify heterogeneously expressed cell-surface proteins on LPs. SC gene expression analyses allowed us to link the surface markers GDNF Family Receptor Alpha 2 (GFRA2) and bone marrow (BM)

[1]Division of Molecular Hematology, Lund University, Lund, Sweden;   [2]Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden;   [3]Faculty of Medicine, University of Helsinki, Helsinki, Finland.

*C.T. Jensen and J. Ahsberg contributed equally to this paper;   Correspondence to Mikael Sigvardsson: mikael.sigvardsson@med.lu.se.

stroma cell antigen 1 (BST1) to the combined expression of the B-lineage TFs *Ebf1* and *Pax5*. While up-regulation of GFRA2 on Ly6D$^+$ LPs correlated with reduction of natural killer (NK) cell but retained T and B cell potential, BST1-positive cells were largely B cell–restricted in vitro as well as in vivo. The formation of the GFRA2$^+$ compartments and activation of B-lineage genes in the progenitors were directly dependent on EBF1. Our data reveal that consecutive expression of Ly6D, GFRA2, and BST1 defines a developmental trajectory linking the CLP compartment with the CD19$^+$ pro–B cells in the mouse BM.

Results

Surface expression of GFRA2 and BST1 correlates with distinct stages of B cell development in the lineage negative LP compartment

To identify surface markers that can be used to isolate subpopulations of LP cells, we explored the expression of 176 surface antigens on LP populations (CLP/LP) (Lin$^-$B220$^-$IL7R$^+$FLT3$^+$SCA1^{int}KIT^{int}Ly6D$^{+/-}$; Karsunky et al., 2008; Inlay et al., 2009) from mouse BM using an arrayed antibody library and FACS analysis (Fig. 1 A and Table S1). Although several antigens, including CD102, CD44, CD47, and CD48, could be detected on all the progenitors (Table S1), CD11C, CD120B, CD5, CD162, and CD4 were expressed on a fraction of the CLP/LPs but not on CD19$^+$ cells (Jensen et al., 2016). CD24 and CD157 (BST1) were expressed on subpopulations of CLPs and in a substantial fraction of the CD19$^+$ cells, indicating a positive correlation with B-lineage cell fate (Fig. 1 A and Table S1). The antibody screen was complemented by RNA-sequencing (RNA-seq) of Ly6D$^{-/+}$ LPs (Inlay et al., 2009) depleted of LPAM1 (α4β7)-expressing cells (Possot et al., 2011) and a CD19$^-$ B-lymphoid–restricted progenitor population expressing an *Igll1* (λ5) hCD25 reporter gene (Mårtensson et al., 1997; Mansson et al., 2008; Fig. 1 B). This identified several differentially expressed surface markers that could be linked to B-lineage development, including BST1 and GFRA2.

To link the expression of differentially expressed surface markers to that of lineage-associated TFs, we used SC qRT-PCR (SC-qRT-PCR) to analyze gene expression patterns in 338 Lin$^-$B220$^-$IL7R$^+$FLT3$^+$SCA1^{int}KIT^{int} CLP/LP cells (Fig. 1 C). This identified 37 genes that were differentially expressed between subpopulations, including the surface markers *Ly6d*, *Bst1*, *Gfra2*, *Cd24a* (CD24), and *Selplg* (CD162), as well as the TFs *Ebf1*, *Pax5*, and *Foxo1* (Fig. 1, C and D, Table S2). The expression of *Bst1*, *Gfra2*, and *Cd24a* displayed a positive correlation with *Pax5* and *Ebf1* although the opposite patterns were observed for *Selplg*, *Itgb7*, and *Csf3r* (CD114; Fig. 1 C). These experiments suggested that the expression of *Gfra2* and *Bst1* could be linked to B-lineage priming in LP compartments.

GFRA2 and BST1 are both glycosylphosphatidylinositol-anchored proteins, but whereas GFRA2 is a receptor for the neurotrophic proteins neurotropic factor (GDNF) and neurturin (Sanicola et al., 1997), BST1 is an ADP-ribosyl cyclase. Although no function of GFRA2 has been reported within the hematopoietic system, BST1 has been shown to be expressed on B cell progenitors (McNagny et al., 1988; Ishihara et al., 1996) and to be important for normal immune response (Itoh et al., 1998) as

well as for neutrophil function (Funaro et al., 2004). To investigate their expression on BM progenitor populations, we performed FACS analysis for GFRA2 and BST1 on multipotent as well as B-lineage–restricted progenitors (Fig. 1 E). Although the expression of both markers was low or absent in the multipotent progenitors (LSK), they could be detected on a fraction of the classical CLP population. FrA cells were heterogeneous regarding surface levels of these proteins, whereas CD19$^+$ pro–B cells were largely positive for expression of both GFRA2 and BST1. The heterogeneous expression of both GFRA2 and BST1 in progenitor compartments and rather robust expression on CD19$^+$ pro–B cells indicated that these markers could be used to define early B cell progenitor populations.

GFRA2 expression defines a B-lineage primed progenitor population within the Ly6D$^+$ compartment

Assessing the expression of GFRA2 in the CLP/LP compartment (Lin$^-$B220$^-$IL7R$^+$FLT3$^+$SCA1^{int}KIT^{int}Ly6D$^{+/-}$; Karsunky et al., 2008; Inlay et al., 2009) revealed that it was largely restricted to the Ly6D$^+$ population (Fig. 2 A). Expression of BST1 was even more restricted and found mainly on GFRA2$^+$ cells (Fig. 2 A). To further explore whether these markers identify unique populations within the LP compartments, we examined the expression of CD19 and KIT on Lin$^-$ Ly6D$^+$GFRA2$^-$BST1$^-$, Ly6D$^+$GFRA2$^+$BST1$^-$, and Ly6D$^+$GFRA2$^+$BST1$^-$ cells (Fig. S1 A). All the populations contained CD19$^-$ and KIT$^+$ putative LP cells, in addition to CD19 expressing B cells. In line with this, CD19$^-$KIT$^+$ cells expressed lower levels of B220 and higher levels of FLT3 compared with Lin$^-$CD19$^+$KIT$^+$ pro–B cells. These data support the idea that GFRA2 and BST1 can be used to identify early CD19$^-$ progenitors in the mouse BM. To unravel the functional properties of the identified populations, we seeded Ly6D and GFRA2 double-negative (CLP$^{-/-}$), Ly6D single-positive (CLP$^{+/-}$), and Ly6D$^+$GFRA2$^+$ (CLP$^{+/+}$) progenitors on OP9 stroma cells in cultures supplemented with Kit ligand, Fms-like tyrosine kinase 3 ligand, IL-7, IL-15, and IL-2. The progressive development into B (CD19$^+$) or NK (NK1.1$^+$) cells was then followed by FACS analysis (Fig. 2 B). Ly6D$^+$GFRA2$^+$ cells underwent a rapid expansion and more efficient generation of CD19$^+$ cells compared with the GFRA2$^-$ populations although the generation of NK cells from GFRA2$^+$ cells was minimal. Expression of CD19 in cultures of CLP$^{-/-}$ cells was preceded by GFRA2 expression (Fig. 2 C), suggesting that B cell development proceeded through a GFRA2$^+$CD19$^-$ stage. To resolve the developmental hierarchy among the populations, we followed the formation of GFRA2- and BST1-positive cells after seeding Ly6D$^+$GFRA2$^-$ CLPs on OP9 stroma cells. Although substantial numbers of GFRA2$^+$ cells were detected 2 d after seeding (Figs. 2 D and S1 B), BST1$^+$ cells were not present until after 6 d of culture. This suggests that GFRA2 and BST1 can be used to define a developmental hierarchy in the early lymphoid compartment in the BM.

To identify residual T-lineage potential in the progenitor subpopulations, single progenitors were seeded on OP9-Delta1 (OP9DL1) cells, which promote the formation of T-lineage cells from multipotent hematopoietic progenitors (Schmitt and Zúñiga-Pflücker, 2002). The cellular content of the generated colonies was analyzed for expression of early T-lineage
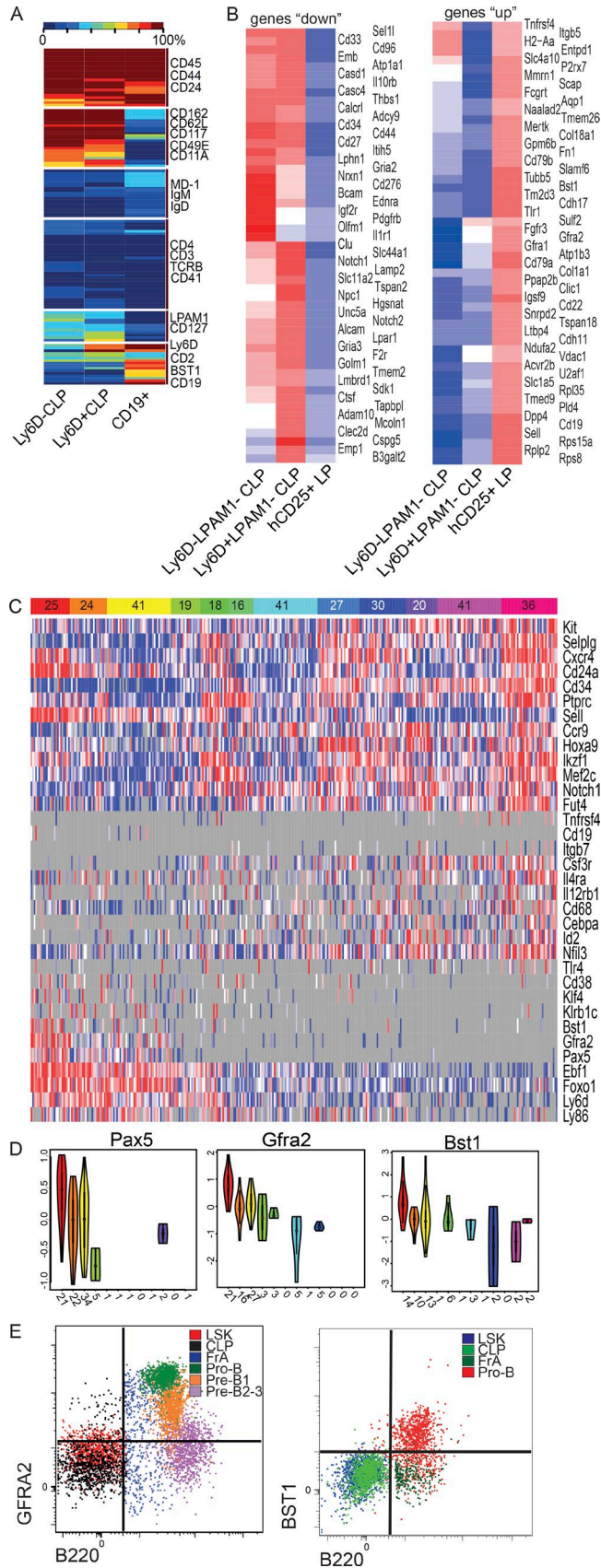
Jensen et al.
Heterogeneity in early lymphoid compartments

Journal of Experimental Medicine 1948
https://doi.org/10.1084/jem.20171384

Figure 1. **Heterogenic surface marker expression allows for the identification of LP subpopulations in the mouse BM. (A)** Heatmap showing data from a BD Lyoplate antibody screen with CLP Ly6D⁻, CLP Ly6D⁺, and CD19⁺

(CD19⁻NK1.1⁻Thy1.2^hi, or CD19⁻NK1.1⁻Thy1.2^hiCD25⁺) or B-lymphoid (CD19) markers. Although all the growing Ly6D⁺GFRA2⁻ cells displayed T-lineage potential, the Ly6D⁺GFRA2⁺ population contained cells insensitive to the NOTCH signal (Fig. 2 E). Hence, even though about half of the GFRA2⁺ cells retained some T cell potential, all B-lineage committed cells were found within the GFRA2⁺ population (Fig. 2 E). This suggests a cellular heterogeneity that potentially could be explained by changes in relative sensitivity to NOTCH signaling. To examine this possibility, we incubated Ly6D⁻ (CLP^(−/−)), Ly6D⁺GFRA2⁻ (CLP^(+/−)), and Ly6D⁺G-FRA2⁺ (CLP^(+/+)) CLP cells on OP9DL1 cells with increasing concentrations of the γ secretase inhibitor DAPT modulating the strength of the NOTCH signal. Although we observed a profound loss of T cell formation from Ly6D⁺GFRA2⁺ progenitors already at the lowest concentrations of DAPT, formation of T-lineage cells from Ly6D⁺GFRA2⁻ cells were less sensitive to the NOTCH signal inhibitor (Figs. 2 F and S2, A and B). At higher concentrations of DAPT, we observed a significant difference in the ability of Ly6D⁺GFRA2⁻ cells to generate T-lineage progenitors compared with Ly6D⁻GFRA2⁻ cells. Thus, GFRA2 expression marks a functionally distinct population within the classical CLP compartment.

BST1 expression was highly restricted within the CLP compartment. When assessing the content of cultures generated from single BST1⁺ LP cells, we found that 90% of the clones contained B-lineage cells when cultured on OP9DL1 cells (Fig. 2 G), revealing a reduced T-lineage potential. Reanalysis of the CLP SC-qPCR data in Fig. 1 by clustering cells based on expression of *Bst1, Gfra2*, and *Ly6d* (Fig. S2 C) suggested that although the expression of *Gfra2* is associated with increased expression levels of B-lineage genes including *Ebf1, Foxo1*, and *Pax5*, the levels are further increased in *Bst1*-expressing cells. Hence, the expression of Ly6D, GFRA2, and BST1 resolves a functional hierarchy among LP populations in the early CD19⁻ LP compartment.

cells; data from CD19⁺ cells are originally published by Jensen et al. (2016). Data are shown as percentage of cells that stained positive with the library antibodies (>6% in at least one of the investigated populations). Selected markers are indicated. For full information, see Table S1. **(B)** RNA-seq data from Ly6D⁻LPAM1⁻CLP (n = 2), Ly6D⁺LPAM1⁻CLP (n = 3), and hCD25⁺Lin⁻IL7R⁺FLT3⁺SCA1^IntKIT^Int LP (n = 5) cells. The heatmap shows relative expression of differentially expressed surface markers. Differentially expressed genes were called by using DESeq2 (FDR < 0.1, blue to red color denoting low to high expression, replicates averaged). **(C)** Heatmap of SC-qRT-PCR Fluidigm data showing differentially expressed surface markers and TFs in the CLP compartment (n = 338). Differentially expressed genes were called by using the MAST hurdle model (P < 0.01). Each cluster is indicated in colors at the top of the map. Numbers in the colored cluster indicate the number of cells in each cluster. **(D)** Violin plots of SC-qRT-PCR data in C displaying the expression level and frequency of expression of key surface markers and *Pax5* (n = 338). Numbers at the bottom of each plot indicate the number of cells expressing each gene in each given cluster. Colors in the violin plots match the cluster colors in the heatmap in C. Full information about differentially expressed genes can be found in Table S2. **(E)** Representative FACS plots showing the expression pattern of GFRA2 and BST1 in BM LPs. Early progenitors were defined as Lin⁻KIT^HighSCA1^High (LSK), CLP as Lin⁻IL7R⁺FLT3⁺KIT^Int, and FrA as Lin⁻B220⁺AA4.1⁺IL7R⁺KIT⁺. Pro–B cells are defined as Lin⁻CD19⁺B220⁺Ly6D^low/-CD24^lowKIT⁺IL7R⁺ (GFRA2 panel) or Lin⁻CD19⁺B220⁺KIT⁺IL7R⁺ (BST1 panel), pre–B1 cells as Lin⁻CD19⁺B220⁺Ly6D^low/-CD24^hi, and PreB2-3 as Lin⁻CD19⁺B220⁺Ly6D^hiCD24^hi.
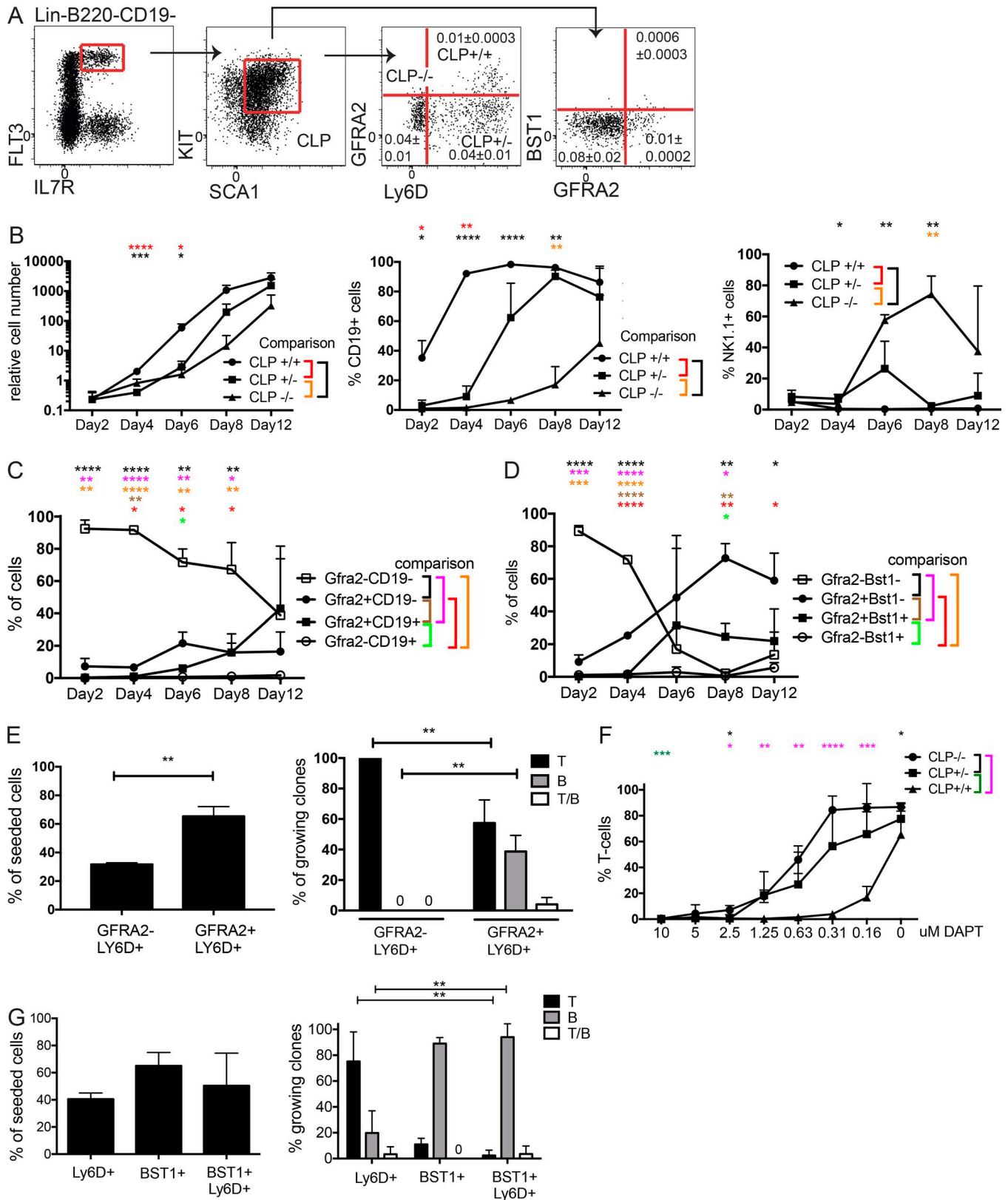
Figure 2. **Surface expression of Ly6D, GFRA2, and BST1 resolves a developmental hierarchy in early LPs. (A)** Representative FACS profiles of Ly6D, GFRA2, and BST1 expression on CLPs. Numbers in FACS plots show mean ± SD percentage of viable cells in unfractionated BM (n = 3). **(B)** Diagrams presenting the number of cells relative to that of control beads (left panel) or the mean cellular composition in cultures (n = 3) generated in the indicated time after seeding 50 CLPLy6D⁻GFRA2⁻ (CLP$^{-/-}$), CLPLy6D⁺GFRA2⁻ (CLP$^{+/-}$), or CLPLy6D⁺GFRA2⁺ (CLP$^{+/+}$) cells on OP9 stromal cell layers. **(C)** Diagram displaying the generation of GFRA2⁺ and/or CD19⁺ cells at the indicated time after seeding 50 CLPLy6D⁻GFRA2⁻ (CLP$^{-/-}$) on OP9 stromal cell layers. **(D)** Diagrams displaying the formation of GFRA2- and/or BST1-positive cells at the indicated time after seeding 50 Ly6D⁺GFRA2⁻ CLPs (CLP$^{+/-}$) on OP9 stromal cell layers. **(B–D)** *, P < 0.05; **,

Jensen et al.
Heterogeneity in early lymphoid compartments

Journal of Experimental Medicine 1950

To study whether the functional differences between GFRA2+ and negative progenitors were reflected in unique molecular signatures, we performed RNA-seq analysis from Ly6D−, Ly6D+GFRA2− and Ly6D+GFRA2+ CLP/LP cells (Fig. 3 A). Among the genes up-regulated in the GFRA2+ cells were *Ebf1* and *Pax5* encoding two transcriptional regulators essential for early B cell development. To investigate if the protein levels of EBF1 and PAX5 recapitulated that of the mRNA, we used FACS to explore their presence within defined fractions of the CLP/LP compartment (Fig. 3 B and S2 D). Although low amounts of EBF1 could be detected in a minority of the Ly6D+GFRA2− progenitors, 87% of the GFRA2+ cells expressed EBF1 at levels approaching that of CD19+ cells. PAX5 could only be detected at low levels in 22% of the GFRA2+ cells, whereas the protein was expressed in 99% of the CD19+ cells.
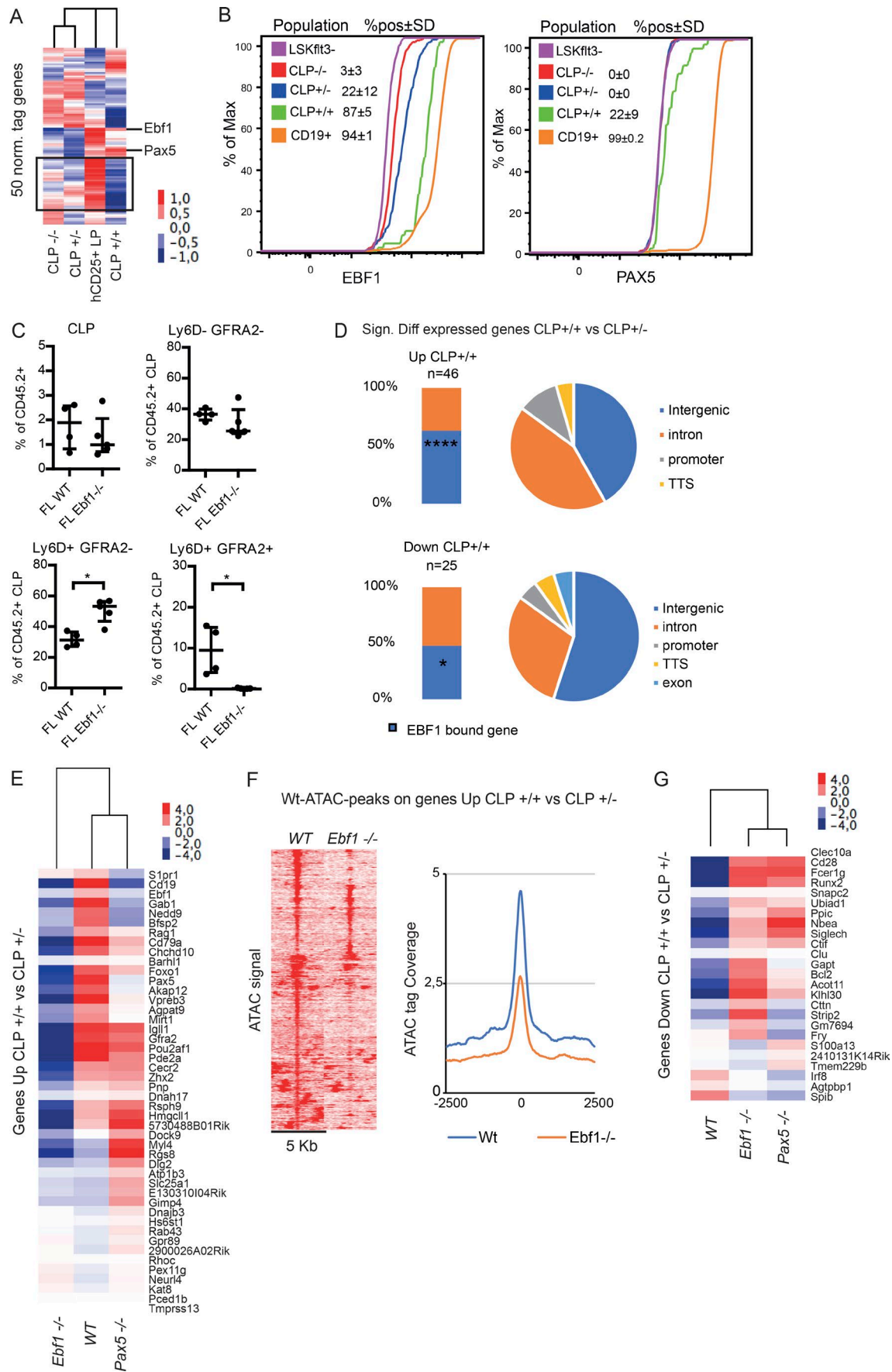
The increased expression of EBF1 in GFRA2+ cells prompted us to explore the importance of this TF for the formation of LP populations. Although the fraction of total CLPs generated after transplantation of *Ebf1−/−* fetal liver (FL) LSK cells was comparable to *Wt* cells, the GFRA2+ fraction was selectively reduced (Fig. 3 C). Hence, the formation of GFRA2+ cells is critically dependent on EBF1. Because it has been reported that *Gfra2* is a direct target for EBF1 (Treiber et al., 2010), we investigated whether the loss of GFRA2+ cells reflected a developmental block or primarily targeted the expression of the surface marker per se. To this end, we analyzed the association of genes differentially expressed in Ly6D+GFRA2− compared with Ly6D+GFRA2+ CLPs with EBF1 binding sites as defined by chromatin immunoprecipitation sequencing (ChIP-seq) data from in vitro expanded pro–B cells. This suggested that almost two thirds of the genes up-regulated and about half of the genes repressed could be linked to EBF1 binding sites (Fig. 3 D). Analyzing the mRNA levels of the same set of differentially expressed genes in in vitro expanded *Wt*, *Ebf1−/−*, or *Pax5−/−* FL pro–B cells revealed that the majority of the up-regulated genes were expressed at high levels in *Wt* pro–B cells, whereas the expression was decreased in *Ebf1−/−* progenitors (Fig. 3 E). *Pax5−/−* FL cells expressed the majority of the genes at levels similar to *Wt* cells. Examining the chromatin accessibility in regions linked to the differentially expressed genes by assay for transposase accessible chromatin by using sequencing (ATAC-seq) analysis from *Wt* and *Ebf1−/−* progenitor cells revealed that several of these sites were devoid of accessibility in the absence of EBF1 (Fig. 3 F). Analyzing the expression of genes expressed at lower levels in GFRA2-positive LPs revealed that a majority of these were overexpressed in the *Ebf1−/−* cells (Fig. 3 G). These data strongly support the idea that GFRA2 expression defines a unique EBF1-dependent population of B-lineage primed cells within the classical Ly6D+ BLP compartment.

GFRA2 and BST1 identify functionally distinct subpopulations in the B220+ FrA/pre-pro–B cell compartment

Having established GFRA2 and BST1 expression identified unique progenitor populations in the CLP compartment, we investigated the possibility of using these markers for identification of lineage-restricted progenitors in the B220+CD19− FrA/pre-pro–B cell compartment. The protocols for isolation of these progenitors has been gradually improved (Hardy et al., 1991; Li et al., 1996; Rumfelt et al., 2006), and the highest enrichment for committed B cell progenitors has been reported for Lin−CD19−B220+CD24low/−CD43+AA4.1highKITlowIL7R+ cells (Rumfelt et al., 2006). Exploring surface expression of GFRA2 and BST1 on FrA cells (Fig. 4 A), we observed a similar pattern as for CLP/LP cells, with GFRA2−BST1− (FrA−/−), GFRA2+BST1− (FrA+/−) and GFRA2+BST1+ (FrA−/−) populations. When seeded on OP9 stroma cells, both the GFRA2+ populations displayed a similar proliferative capacity; however, we detected a significant difference in the fraction of CD19+ cells generated from BST1+ progenitors after 2 d of culture (Fig. 4 B). In contrast, the GFRA2−BST1− cells did not expand as well as the other populations, and the generation of CD19+ cells was slower and preceded by formation of GFRA2+ cells. At day 12 after seeding, we noted a reduced expression of GFRA2 on a fraction of the CD19+ cells generated by GFRA2+ progenitors, a finding likely resulting from the heterogeneous expression of this marker in CD19+ progenitor compartments (Fig. 1 E). Thus, as for the CLP/LP compartment, expression of GFRA2 could be linked to a high B-lineage potential. To detail this, we sorted the three populations and seeded SCs on OP9 cells. Although most (∼60%) GFRA2+ progenitors generated pure B cell colonies, only ∼10% of the GFRA2− cells generated colonies under these conditions, and approximately one third of these was composed of NK cells (Fig. 4 C). When the progenitors were seeded on OP9DL1 stroma cells, the GFRA2− cells generated almost exclusively T-lineage colonies, GFRA2+BST1− cells generated similar numbers of B- and T-lineage clones (Fig. 4, D and E), and the BST1+ cells generated ∼90% B-lineage clones. Similar cloning frequencies and lineage potentials were observed after seeding single Lin−CD19−B220+GFRA2+ cells (Fig. S3, A–C), revealing reduced need for additional markers for identification of B-lineage progenitors within the FrA compartment. Transplantation of GFRA2-positive or -negative B220+ progenitors to immune-deficient NOD *scid*γc−/− (NSG; CD45.1) hosts revealed that, even though both populations were able to generate B-lineage cells, the GFRA2 negative cells generated more CD19− cells, including a substantial fraction of NK1.1+ cells

P < 0.01; ***, P < 0.001; ****, P < 0.0001 (Student's *t* test); n = 3. **(E)** Cloning efficiency defined as percentage of seeded cells and lineage potentials of single GFRA2+Ly6D+ (n = 190) and GFRA2−Ly6D+ (n = 144) CLPs grown on OP9DL1 stromal cell layers. T cells were defined as CD19−NK1.1−Thy.1.2hiCD25+. Values are reported as mean ± SD; *, P < 0.05; **, P < 0.01 (Student's two-tailed *t* test); from three independent samples and two to three experiments. **(F)** Diagrams presenting the mean cellular composition in cultures (n = 3) generated after seeding 20 CLPLy6D−GFRA2− (CLP−/−), CLPLy6D+GFRA2− (CLP+/−) or CLPLy6D+GFRA2+ (CLP+/+) cells on OP9DL1 stroma under T cell differentiation conditions supplemented with a twofold dilution series of DAPT. Panels show the percentage of T cells in culture after 14-d incubation with media supplemented with DAPT. *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001 (Student's *t* test). DMSO controls are shown in Fig. S2B. **(G)** Cloning efficiency and lineage potentials of single Ly6D+ (n = 122 from three experiments), CD24hiBST1+ (BST1+; n = 47) and CD24hiBST1+Ly6D+ (BST1+Ly6D+; n = 58). CLPs grown on OP9DL1 stromal cell layers for 14 d. T cells were defined as CD19−NK1.1−Thy.1.2hi. Values are reported as mean ± SD; *, P < 0.05; **, P < 0.01 (Student's two-tailed *t* test); from three independent samples and two to three experiments.

A — heatmap, 50 norm. tag genes; labeled Ebf1, Pax5; columns CLP−/−, CLP+/−, hCD25+ LP, CLP+/+; color scale 1,0 / 0,5 / 0,0 / −0,5 / −1,0

B —

Left panel (EBF1):

Population	%pos±SD
LSKflt3−	
CLP−/−	3±3
CLP+/−	22±12
CLP+/+	87±5
CD19+	94±1

Right panel (PAX5):

Population	%pos±SD
LSKflt3−	
CLP−/−	0±0
CLP+/−	0±0
CLP+/+	22±9
CD19+	99±0.2

C — Plots: CLP (% of CD45.2+), Ly6D− GFRA2− (% of CD45.2+ CLP), Ly6D+ GFRA2− (% of CD45.2+ CLP), Ly6D+ GFRA2+ (% of CD45.2+ CLP); groups FL WT and FL Ebf1−/−; significance marks *

D — Sign. Diff expressed genes CLP+/+ vs CLP+/−

Up CLP+/+ n=46 (****); Down CLP+/+ n=25 (*); EBF1 bound gene; pie categories: Intergenic, intron, promoter, TTS, exon

E — heatmap, Genes Up CLP +/+ vs CLP +/−; columns Ebf1−/−, WT, Pax5−/−; scale 4,0 / 2,0 / 0,0 / −2,0 / −4,0

Gene labels (E): S1pr1, Cd19, Ebf1, Gab1, Nedd9, Bfsp2, Rag1, Cd79a, Chchd10, Barhl1, Foxo1, Pax5, Akap12, Vpreb3, Agpat9, Mirt1, Igll1, Gfra2, Pou2af1, Pde2a, Cecr2, Zhx2, Pnp, Dnah17, Rsph9, Hmgcll1, 5730488B01Rik, Dock9, Myl4, Rgs8, Dlg2, Atp1b3, Slc25a1, E130310I04Rik, Gimp4, Dnajb3, Hs6st1, Rab43, Gpr89, 2900026A02Rik, Rhoc, Pex11g, Neurl4, Kat8, Pced1b, Tmprss13

F — Wt-ATAC-peaks on genes Up CLP +/+ vs CLP +/−; ATAC signal WT and Ebf1−/−; 5 Kb; ATAC tag Coverage plot; Wt (blue) and Ebf1−/− (orange); x-axis −2500 to 2500

G — heatmap, Genes Down CLP +/+ vs CLP +/−; columns WT, Ebf1−/−, Pax5−/−; scale 4,0 / 2,0 / −2,0 / −4,0

Gene labels (G): Clec10a, Cd28, Fcer1g, Runx2, Snapc2, Ubiad1, Ppic, Nbea, Siglech, Ctif, Clu, Gapt, Bcl2, Acot11, Klhl30, Cttn, Strip2, Gm7694, Fry, S100a13, 2410131K14Rik, Tmem229b, Irf8, Agtpbp1, Spib

(Fig. 4 F). Similar cellular compositions were observed in blood at 2 and 4 wk after transplantation (Fig. 4 F). In the spleen, we detected NK cells generated from the GFRA2⁻ cells, and in three of the transplanted animals, we detected CD3⁺ cells (Fig. 4 G). The BM of mice transplanted with GFRA2⁺ cells (Fig. 4 H) contained a higher fraction of donor cells, and the finding that 20% of these were CD19⁺IgM⁻ cells indicates that the GFRA2⁺ progenitors were able to remain in an immature stage for at least 4 wk. Hence, the lineage potential of GFRA2- and BST1-positive FrA cells is highly similar to that of the CLP/LP compartment.

GFRA2 is not essential for B cell development

As the functional relevance of GFRA2 in early B cell development is largely unexplored, we analyzed the presence of LP populations as well as different B-lineage compartments in the BM from GFRA2-deficient mice. We detected a comparable number of cells as well as highly similar proportions of CLPs and B cell populations in the absence of GFRA2 (Fig. S3, D–F), suggesting that this protein is not essential for B cell development in the BM.

GFRA2 signals via the tyrosine kinase receptor RET (Cik et al., 2000) has been reported to be of importance for normal hematopoiesis in the fetus (Fonseca-Pereira et al., 2014). To explore whether GFRA2 was expressed in the FL, we dissected FL from *Wt* mice and explored the expression of surface markers by FACS. Similar to what we observed in the BM, CD19⁺ cells in the FL expressed high levels of both GFRA2 and BST1 (Fig. S3 G). GFRA2 expression and to some extent BST1 was also detected on FL FrA (Fig. S3 I) as well as CLPs (Fig. S3 H). Despite robust expression of GFRA2 on FL cells, we noted no differences in the presence of LSK cells, CLPs, total CD19, or B1 progenitors in *Wt* compared with *Gfra2⁻/⁻* FL (Fig. S3, J–L). Hence, although expression of GFRA2 in the LP compartments in the BM is linked to B-lineage progenitors, we were unable to detect evidence for a critical function of this protein in B cell development.

GFRA2⁺ CLP and FrA cells display similar gene expression patterns

To identify potential molecular differences between GFRA2- and BST1-expressing cells in the FrA or CLP compartments, we analyzed the expression of lineage-associated genes in the defined populations by SC-qPCR (Fig. 5 A and Table S3). This revealed an association of GFRA2⁻ CLP/LP and GFRA2⁻ FrA cells with a relative absence of B-lineage priming. In contrast, B-lineage genes, including *Pax5*, *Pou2af1*, and *Igll1*, were expressed in both GFRA2-positive CLP/LP and FrA cells. Component analysis (C; Fig. 5 B) revealed that the GFRA2⁻ FrA cells clustered with GFRA2⁻ CLPs whereas all the GFRA2⁺ progenitor populations displayed a high degree of similarity in gene expression. These data suggest that GFRA2 expression defines similar populations in both the classical CLP compartment and FrA, arguing against these cells representing distinct progenitor populations.

To further investigate the molecular heterogeneity in B-lineage progenitor subpopulations we performed SC RNA sequencing of classic CLPs (Karsunky et al., 2008), FrA cells as well as CD19⁺IgM⁻ pro–B cells. Analyzing the data by using dimension reduction (DDRTree; Fig. 5 C) suggested a clear discrimination between the CLPs (red) and pro–B cells (blue) with FrA cells (green) and some CLPs bridging the two populations. In line with the observation that FrA is rather heterogeneous (Rumfelt et al., 2006), we detected cells from this population clustering with both the more mature B cells and the CLP compartments. The impact of cell cycle on the clustering of cells shows that the C3 dimension is largely dependent on cell cycle status and that a large portion of the examined cells from all populations was in either S or G2/M phase (Fig. 5 D). Expression levels of *Ebf1* (Fig. 5 E), *Pax5, or Igll1* (Fig. S4) were generally increased in the C1 and C2 dimensions of the DDRTree plot, suggesting progressive differentiation. Although we detected mRNA encoding Ly6D in several populations, the expression of *Gfra2* or *Bst1* was more restricted to a few cells of more differentiated phenotype. These data suggest that, even though FrA as well as the Ly6D-expressing BLP compartment is enriched for B cell–primed progenitors, both populations display a high molecular heterogeneity.

GFRA2 and BST1 expression allows for high-resolution identification of B cell development in the BM

Having established that GFRA2 and BST1 expression identified committed B-lineage progenitors in both the classical CLP and FrA populations, we wanted to examine whether these markers could be used to create a unified model for early B cell

---

Figure 3. **EBF1 is a critical regulator of genes associated with lineage restriction in early B cell development. (A)** Heatmap displaying log2 mean RNA-seq gene expression data in Ly6D⁻GFRA2⁻CLP (CLP⁻/⁻; n = 6), Ly6D⁺GFRA2⁻CLP (CLP⁺/⁻; n = 4), Ly6D⁺GFRA2⁺CLP (CLP⁺/⁺; n = 3), and hCD25⁺ LP (n = 5) cells (Fig. 1 B) of genes filtered on ≥50 normalized reads in one or more samples and normalized to 10M reads. **(B)** Cumulative distribution function plots of FACS data showing the expression of EBF1 and PAX5 protein in CLP populations. Numbers in graphs show the mean ± SD percentage of positive cells relative to Lin⁻SCA1⁺KIT⁺FLT3⁻ (LSKFlt3⁻) cells (n = 3). **(C)** Graph displaying the percentage of the indicated cell populations (% of PI⁻CD45.2⁺Lin⁻B220⁻CD19⁻ events) in the BM of transplanted NSG mice. Each dot indicates one transplanted mouse (*Wt*, n = 4; *Ebf1⁻/⁻*, n = 5; from two experiments). Statistical analysis was performed by using an unpaired *t* test. *, P < 0.05. Error bars indicate SD of mean. **(D)** Graphs displaying the fraction of genes with differential expression in CLP⁺/⁺ (n = 3) compared with CLP⁺/⁻ (n = 4; FDR: 0.1; log2fold: 1) associated with EBF1 binding as estimated by proximity analysis of ChIP-seq data (GEO accession no. GSE69227). Statistical analysis was based on comparison with data generated by proximity analysis of EBF1 binding to the total number of annotated genes in the genome (24,072) by using χ² test with Yates' correction in Prism. (****, P < 0.0001; *, P = 0.0266). Functional annotation of EBF1 ChIP-seq peaks annotated (by proximity) on genes up- or down-regulated in CLP⁺/⁺ vs CLP⁺/⁻ is presented as pie charts. **(E)** Heatmap displaying log2 mean RNA-seq gene expression data in cultured FL pro–B *Wt* (n = 3), *Ebf1⁻/⁻* (n = 3) and *Pax5⁻/⁻* (n = 3) cells (GEO accession no. GSE92434) of genes significantly up-regulated in CLP⁺/⁺ compared with CLP⁺/⁻ as defined in D (FDR: 0.1; log2fold: 1). **(F)** Heatmap displaying mean ATAC-seq signal in *Wt* (n = 3) and *Ebf1⁻/⁻* (n = 3) ATAC-seq data (GSE92434) on *Wt* ATAC-seq peaks (with at least five normalized tags in one of the samples) associated with genes significantly up-regulated in CLP⁺/⁺ compared with CLP⁺/⁻ (FDR: 0.1; log2fold: 1) in a 5-Kb window. ATAC coverage in a 5-Kb window centered on these sites was quantitated in a histogram. **(G)** Heatmap displaying log2 mean RNA-seq gene expression data in *Wt* (n = 3), *Ebf1⁻/⁻* (n = 3) and *Pax5⁻/⁻* (n = 3) in in vitro expanded FL pro–B cells (accession no. GSE92434) on genes significantly down-regulated in CLP⁺/⁺ compared with CLP⁺/⁻.
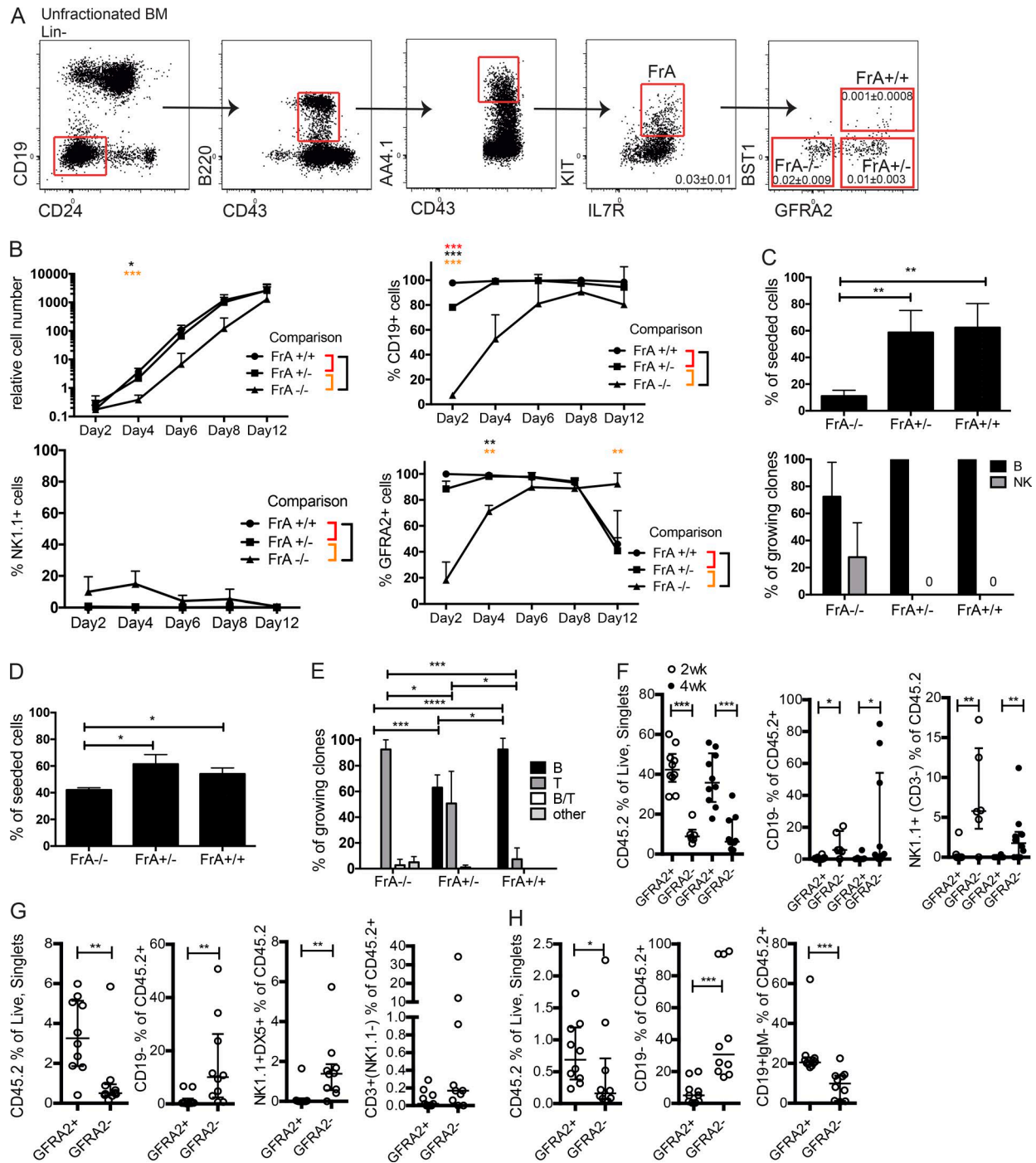

Figure 4. **The classic FrA compartment can be functionally resolved based on expression of GFRA2 and BST1. (A)** Representative FACS profiles for GFRA2 and BST1 expressing FrA cells. Numbers in FACS plots shows the mean ± SD percentage of viable cells in unfractionated BM (n = 7). **(B)** Diagrams presenting the number of cells relative to that of control beads (left upper panel) or the average cellular composition in cultures (n = 3) generated the given number of days after seeding of 50 FrA GFRA2⁻BST1⁻ (FrA⁻/⁻), FrA GFRA2⁺BST1⁻ (FrA⁺/⁻), and FrA GFRA2⁺BST1⁺ (FrA⁺/⁺) cells on OP9 stromal cell layers. Values are reported as mean (SD), with *, P < 0.05, **, P < 0.01, and ***, P < 0.001 (Student's t test), from a total of three samples from three experiments. **(C–E)** Cloning efficiency and lineage potentials on OP9 stroma and (D) cloning efficiency and (E) lineage potentials on OP9DL1 stroma of single FrA⁻/⁻ (OP9: n = 72; OP9DL1: n = 108), FrA⁺/⁻ (OP9: n = 56; OP9DL1, n = 112), and FrA⁺/⁺ (OP9: n = 52; OP9DL1: n = 88) cells 14 d after seeding. Values are reported as mean ± SD; *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001 (Student's two-tailed t test); from three independent samples and two to three experiments. **(F)** Graphs showing level of engraftment and lineage contribution of CD45.2 donor cells in peripheral blood 2 and 4 wk after intravenous transplantation of 2,000 GFRA2⁺ (n = 10) or 2,000 GFRA2⁻ Lin⁻CD19⁻B220⁺ cells (n = 10) to NSG recipients in two experiments. **(G)** Graphs showing level of engraftment and lineage contribution of CD45.2 donor cells in spleen 4 wk after intravenous transplantation. **(H)** Graphs showing level of engraftment and lineage contribution of CD45.2 donor cells in BM 4 wk after intravenous transplantation (as in F). Each dot represents one mouse, and statistical analysis in F–H was performed in GraphPad Prism by using the Mann-Whitney test; *, P < 0.05; **, P < 0.01; ***, P < 0.001.

Jensen et al.
Heterogeneity in early lymphoid compartments

Journal of Experimental Medicine
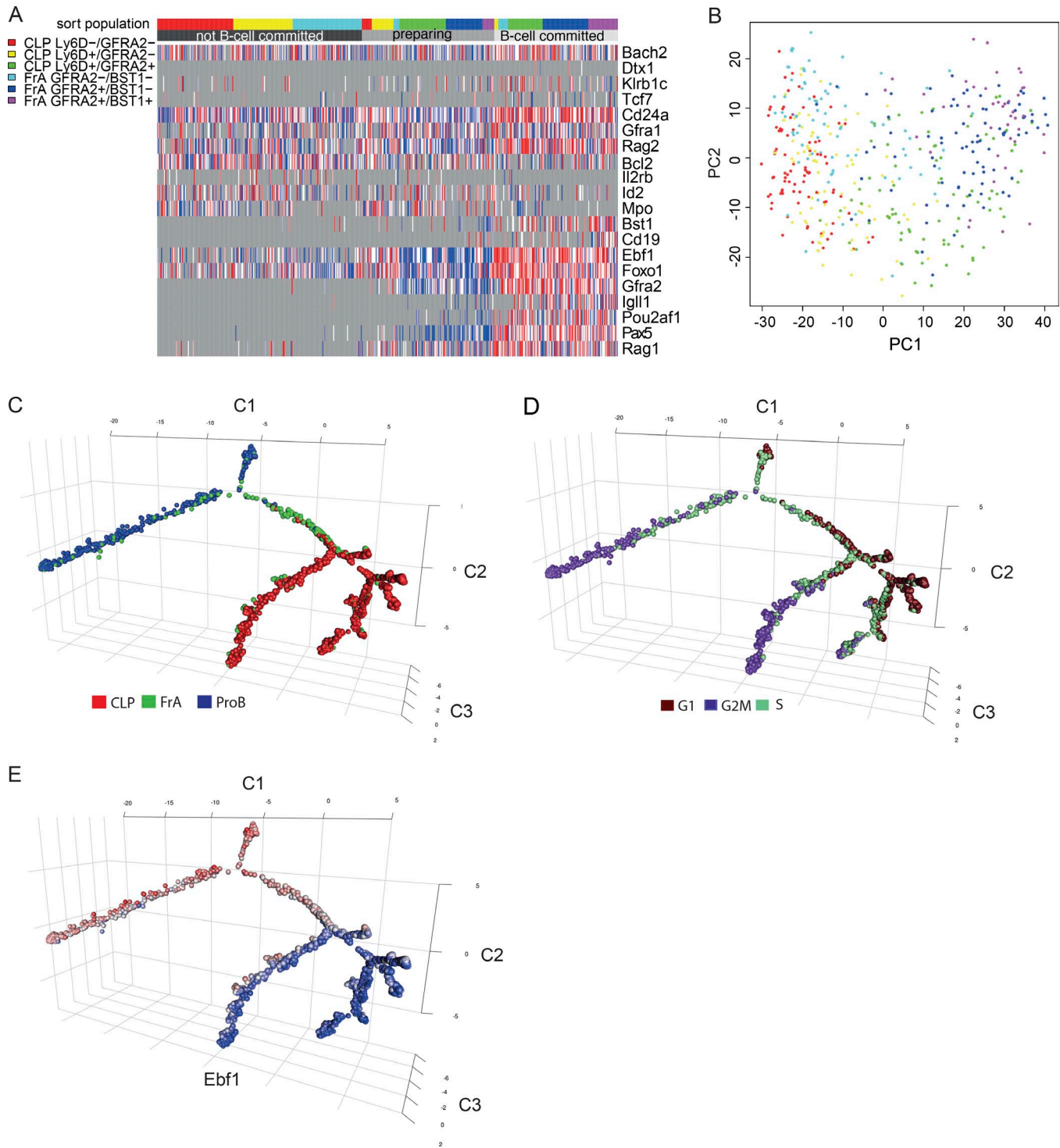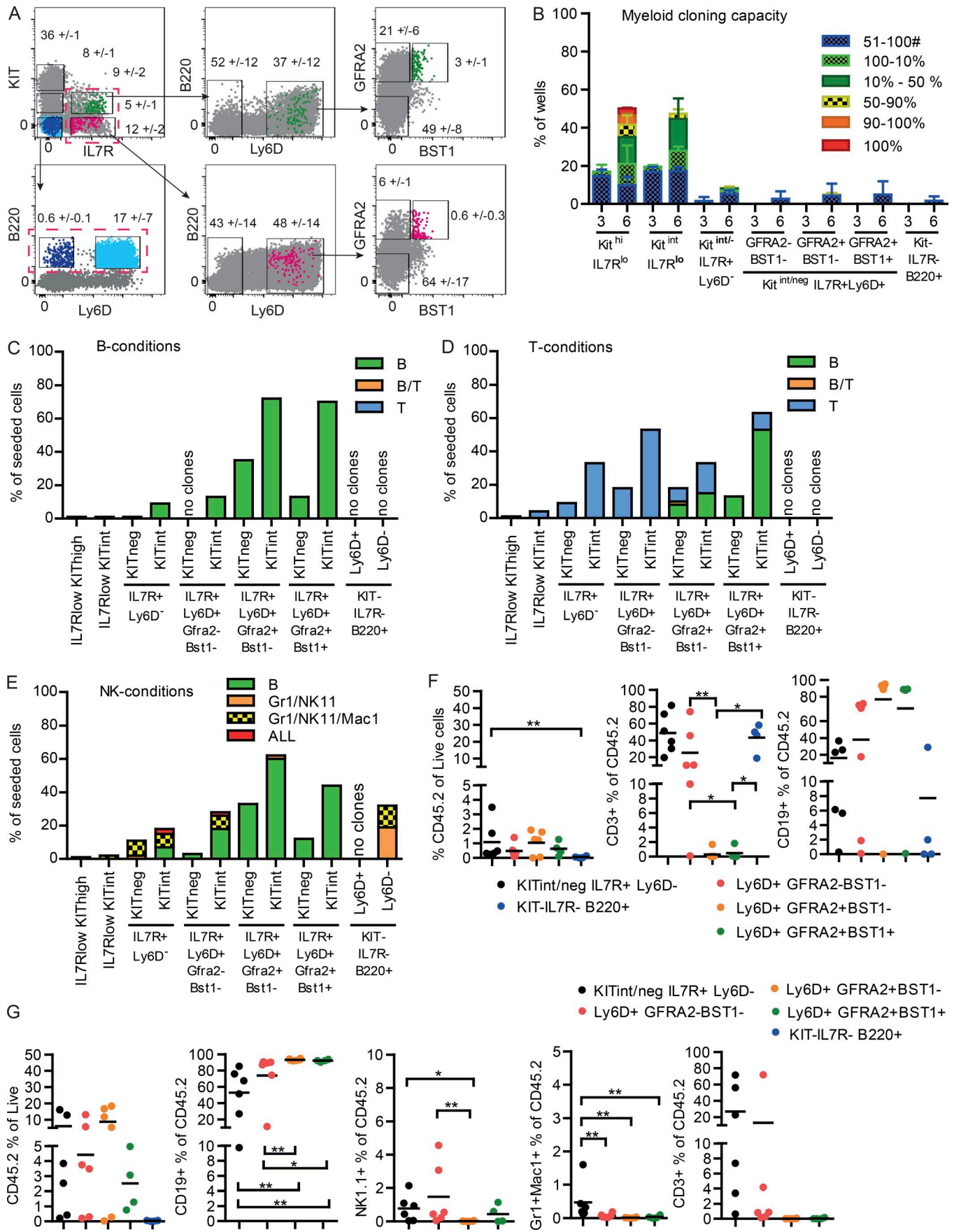1954

https://doi.org/10.1084/jem.20171384

Figure 5.   **GFRA2+ CLPs and FrA cells display similarities in gene expression patterns. (A)** Heatmap displaying SC-qRT-PCR data from sorted single CLPLy6D−GFRA2− (CLP−/−; n = 87), CLPLy6D+GFRA2− (CLP+/−; n = 86), CLPLy6D+GFRA2+ (CLP+/+; n = 82), FrAGFRA2−BST1− (FrA−/−; n = 86), FrAGFRA2+BST1− (FrA+/−; n = 83), and FrAGFRA2+BST1+ (FrA+/+; n = 42) cells. The heatmap shows the differentially expressed genes (MAST hurdle model, P < 0.05). Blue to red color denotes low to high expression. The grayscale color bar above the heatmap indicates the clustering of the different populations based on gene expression patterns. Full information about differentially expressed genes can be found in Table S3. **(B)** Component analysis of the SC-qRT-PCR data in A. **(C–E)** Pseudo time analysis of SC RNA-seq data. **(C)** DDRTree showing the sorted subpopulations of B cell progenitors. **(D)** DDRTree displaying cell-cycle status as determined by gene expression patterns. **(E)** DDRTree showing expression levels of *Ebf1*. Blue indicates low or no expression, and red indicates high levels in SCs. Data are from CLPs (LIN−FLT3+IL7R+KIT+SCA1+; n = 1,753), FrA (LIN− B220+IL7R+KIT+; n = 279), and CD19+ (LIN−IgM−IgD−CD19+; n = 472) B cell progenitors.

Jensen et al.
Heterogeneity in early lymphoid compartments

Journal of Experimental Medicine
https://doi.org/10.1084/jem.20171384

1956

development in the BM. To identify committed and primed progenitors in the lineage negative compartment, we used reporter transgenic mice carrying *Rag1*-GFP and *Igll1*-hCD25 reporter constructs (Mansson et al., 2010; Fig. S5 A). Even though the Lin⁻IL7R⁺KIT^{Int} CLP population was highly enriched for reporter-expressing cells, a substantial fraction of the Rag-expressing cells could be detected in the IL7R⁺KIT⁻ fraction, indicating that B-lineage progenitors might be identified outside of the classical CLP (Kondo et al., 1997) or FrA (Rumfelt et al., 2006) compartment.

As would be predicted from the analysis of reporter mice, we detected expression of Ly6D, GFRA2, and BST1 in both the KIT^{Int} and the KIT⁻ populations (Figs. 6 A and S5 B). qPCR analysis suggested that the expression of GFRA2 was linked to increased expression of *Pax5* mRNA (Fig. S5 C), and exploring Ig heavy chain (IgH) DJ recombination in SCs (Rumfelt et al., 2006) identified fewer cells with remaining germline alleles in the GFRA2⁺ compartment of the Ly6D⁺ cells (Fig. S5 D).

Investigation of the myeloid potential of SCs in vitro suggested that high expression of IL7R was sufficient to identify progenitors with reduced myeloid potential (Fig. 6 B). Hence, even though a few myeloid clones with more than 100 cells were generated from IL7R⁺Ly6D⁻ cells, the inclusion of additional markers would be of limited value. Exploring B-lineage potential of SCs suggested that expression of IL7R and intermediate expression of KIT was associated with higher cloning frequencies (Fig. 6 C). However, we detected cells with potential to generate B-lineage clones in both IL7R⁺KIT^{int} and IL7R⁺KIT⁻GFRA2⁺ cells. Hence, GFRA2 expression identifies cells with B-lineage potential outside the IL7R⁺KIT^{Int} compartment harboring both the CLP (Kondo et al., 1997) and FrA (Rumfelt et al., 2006) populations. Cultivation of the defined progenitors under conditions stimulating T cell development (Fig. 6 D) generated data highly similar to what we observed for purified CLP and FrA cells (Figs. 2 and 4). GFRA2 expression was associated with formation of both B and T cell clones, whereas the expression of BST1 was linked to a clear dominance of B cell clones (Fig. 6 D). Although the expression of KIT was associated with increased cloning frequency, this had no impact on lineage potential. Exploring the ability of the populations to generate NK cells, we noted NK cell production by IL7R⁺Ly6D⁻ as well as the IL7R⁻KIT⁻B220⁺ progenitors (Fig. 6 E). Although Ly6D expression generally associated with a reduction in the ability to generate NK cells (Inlay et al., 2009), we detected a virtual loss of NK cell potential among all

the GFRA2-expressing populations. Within the IL7R⁺ compartment, the expression of KIT could be linked to higher cloning frequency, although this had most limited impact on the lineage potential in the cells. Hence, SC in vitro analysis suggests that the Ly6D⁺ population can be divided into GFRA2-negative cells (BLP1), GFRA2⁺BST1⁻ cells, (BLP2), and GFRA2⁺BST1⁺ progenitors (BLP3).

To resolve the in vivo potential of the progenitors, we transplanted 2,000 Ly6D⁻ KIT$^{int/-}$IL7R⁺, BLP1, BLP2, and BLP3 cells as well as IL7R⁻KIT⁻B220⁺ progenitors to CD45.1⁺ NSG mice. We examined the presence of CD45.2⁺ cells in the spleen and BM at 4 wk after transplantation (Fig. 6, F and G). Although the Ly6D⁻KIT$^{int/-}$IL7R⁺ as well as the BLP1-2 cells all generated comparable levels of CD45.2⁺ cells in the spleen (Fig. 6 G), BLP3 cells were less efficient. B220⁺IL7R⁻ cells were largely unable to generate CD45.2⁺ cells in the spleen. All the reconstituting populations generated mainly CD19⁺ B-lineage cells, and none of the transplanted LP populations generated a macroscopically identifiable thymus at 4 wk after transplantation. However, we detected some formation of CD3⁺ cells in the spleen and BM from the Ly6D⁻KIT$^{int/-}$IL7R⁺ progenitors and BLP1 cells. The generation of NK and myeloid cells in the spleen was low from all the tested populations. The BM generally contained lower levels of CD45.2⁺ cells (Fig. 6 F). Although all the populations generated CD19⁺ cells, the Ly6D⁻ KIT$^{int/-}$IL7R⁺ and BLP1 as well as the IL7R⁻B220⁺ cells generated a substantial fraction of CD3⁺ cells. These cells displayed low expression of *Pax5* and high mRNA levels encoding the TFs of *Gata3* and *Tcf7* (Fig. S5 E), verifying that they represent T-lineage cells. Hence, expression of GFRA2 on LP cells is associated with functional commitment to B-lymphoid development.

Discussion

Although seminal work in the 1990s (Sen et al., 1990; Hardy et al., 1991; Li et al., 1993, 1996; Rolink et al., 1994; Kondo et al., 1997; Tudor et al., 2000) created a foundation for our understanding of B-lineage development, the identity of the earliest committed B cell progenitor has been rather elusive. Despite the fact that expression of B220 enriches for B cell progenitors and appears to proceed the expression of GFRA2 in in vitro differentiation assays (Fig. S1 B), it has been reported to have a limited value as a marker for early committed cells (Martin et al., 2003; Rumfelt et al., 2006; Mansson et al., 2010; Alberti-Servera et al., 2017).

Figure 6. **Surface expression of Ly6D, GFRA2, and BST1 in lineage negative cells defines a developmental trajectory reaching from the CLP to the pro–B cell stage. (A)** Representative FACS plots demonstrating gating strategy for sorting of LPs from lineage-depleted BM. The percentage indicates the fraction of the parental gate (SD) collected from four sorting experiments. **(B)** Diagrams displaying the colony size and cloning frequency 3 or 6 d after seeding single progenitor cells (*n* = 144) from each indicated population into Terasaki plates. The cells were cultured in the presence of IL-3, GM-CSF, M-CSF, G-CSF, KIT ligand, and Flt3 ligand in two experiments. KIT$^{Int/neg}$ cells were defined based on the use of the dashed pink gate in A. Complete statistical analysis (Fisher's exact test) is shown in Table S5. **(C–E)** Diagrams displaying the cloning frequency as well as the cellular content of colonies as determined by FACS analysis 14 d after seeding single progenitor cells on OP9 (C and E) or OP9DL1 (D) stroma cells under conditions stimulating B-, NK-, or T-lineage differentiation as indicated. Cells were sorted based on the gates indicated in A. The data are based on seeding of (*n* = 144) IL7R⁻KIT^{high}, (*n* = 108) IL7R⁻KIT^{int}, (*n* = 144) IL7R⁺Ly6D⁻KIT⁻ and KIT^{int}, (*n* = 60) Ly6D⁺GFRA2⁻BST1⁻KIT⁻, (*n* = 60) KIT^{int}, (*n* = 60) Ly6D⁺GFRA2⁺BST1⁻KIT⁻, (*n* = 60) KIT^{int}, (*n* = 30) Ly6D⁺GFRA2⁺BST1⁻KIT⁻, (*n* = 30) KIT^{int} cells, (*n* = 80) KIT⁻IL7R⁻B220⁺Ly6D⁺ cells, and (*n* = 16) KIT-IL7R⁻B220⁺Ly6D⁻ cells for each condition. Complete statistical analysis for B–E (Fisher's exact test) is shown in Table S6. **(F and G)** Graphs showing level of engraftment and lineage contribution of CD45.2 donor cells in BM (F) and spleen (G) 4 wk after intravenous transplantation of 2,000 sorted cells with the indicated surface receptor phenotype. KIT$^{int/neg}$ IL7R⁺Ly6D⁻ (*n* = 6), Ly6D⁺GFRA2⁻BST1⁻ (*n* = 6), Ly6D⁺GFRA2⁺BST1⁻ (*n* = 6), Ly6D⁺GFRA2⁺BST1⁺ (*n* = 4), and KIT- IL7R⁻B220⁺ (*n* = 4) from two transplantation experiments. Each dot represents one mouse, and statistical analysis was performed in GraphPad Prism by using Mann-Whitney test; *, P < 0.05; **, P < 0.01; ***, P < 0.001.

Jensen et al.
Heterogeneity in early lymphoid compartments

Journal of Experimental Medicine
https://doi.org/10.1084/jem.20171384

1957

Furthermore, even though surface expression of Ly6D identifies cells largely B-lineage restricted in vivo (Inlay et al., 2009), a small population of T-lineage cells was detected in the thymus after transplantation (Inlay et al., 2009). The high-resolution analysis presented in this study unravels both functional and molecular heterogeneity within the Ly6D+ BLP population.

The data presented in this study show that the expression of B-lineage–restricted genes is largely confined to the GFRA2+ BLP2-3 compartments and that GFRA2-negative progenitors retain the ability to adopt T-lineage fate when exposed to a strong NOTCH signal or when transplanted to lymphopenic mice. Even though we were able to detect CD3+ cells generated from Ly6D+ progenitors, we failed to identify a macroscopically detectable thymus 4 wk after transplantation. This is likely a consequence of the limited ability of BM-derived LPs to sustain long-term generation of T-lineage cells (Allman et al., 2003) in combination with the use of adult NSG mice as hosts for transplantation experiments.

GFRA2 expression is tightly linked to the expression of EBF1, and a substantial part of the genes associated selectively with the BLP2-3 stage appears to be direct targets for EBF1. This is well in line with the idea that EBF1 is essential for the activation of B-lineage restricted transcriptional programs (Zandi et al., 2008; Treiber et al., 2010) in early progenitor cells. EBF1 expression has also been linked to lineage restriction because ectopic expression of EBF1 has been reported to result in reduced development of T- and NK-lineage cells in vivo (Zhang et al., 2003), and inactivation of the *Ebf1* gene in B-lineage cells results in lineage plasticity (Mikkola et al., 2002; Nechanitzky et al., 2013; Ungerbäck et al., 2015). This is likely explained by the ability of EBF1 to suppress expression of *Gata3* (Banerjee et al., 2013) and *Id2* (Thal et al., 2009), which are essential for development of T and NK cells, respectively. Interestingly, the expression of *Bst1* has been associated with functional PAX5 (Pridans et al., 2008; Revilla-I-Domingo et al., 2012) in line with that, and expression of this marker is associated with commitment to the B cell lineage. Although the generation of GFRA2+ cells is dependent on EBF1, the expression of Ly6D would appear to be under the control of alternative regulatory circuits (Fig. 3 C; Tsapogas et al., 2011). One possibility could be that this reflects the action of E-proteins because of the importance of these TFs for the formation of Ly6D+ progenitors (Inlay et al., 2009; Welinder et al., 2011). The same regulatory network may be involved in the regulation of Ig recombination because, even though we detected an increased frequency of recombination events in the GFRA2 expressing cells, Rag1 expression as well as recombination events are detected outside this compartment.

Blood cell generation is often depicted in a linear model in which multipotent progenitors gradually lose functional potentials as they develop along branches representing the different blood lineages. However, emerging evidence suggests that the developmental trajectories may not follow a strict tree-like structure but rather resemble a network of paths where the goal can be reached via different routes (Paul et al., 2015; Perié et al., 2015). This concept is rather well established in fetal development where progenitor populations with combined B-macrophage or T-macrophage potential (Katsura, 2002) as well as B/T and B/NK bipotent progenitors are found in the FL (Possot et al., 2011; Pereira de Sousa et al., 2012). Our analysis as well as previous
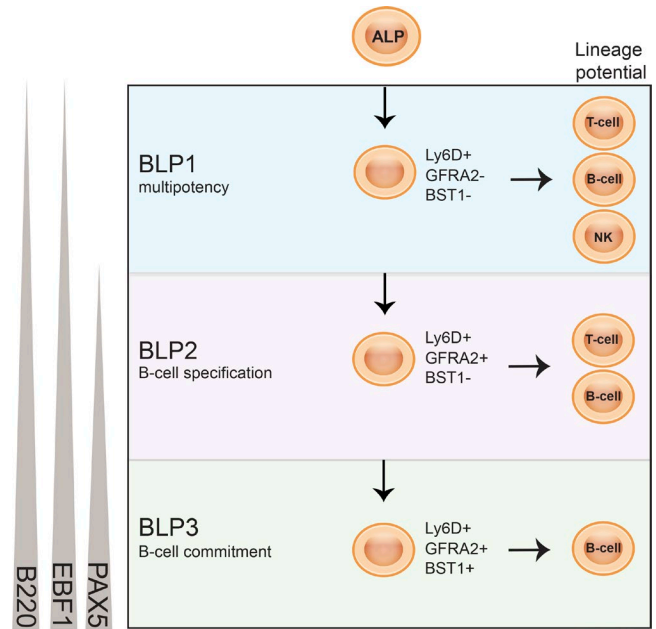

Figure 7. **Ly6D, GFRA2, and BST1 expression defines a developmental trajectory in B cell development.** A schematic drawing depicting how GFRA2 and BST1 expression can be used to capture different stages of the B cell commitment process. The expression levels of B220, EBF1, and PAX5 are indicated.

work (Martin et al., 2003; Rumfelt et al., 2006; Inlay et al., 2009; Mansson et al., 2010) suggest that B-lineage–committed cells can be found both in the CLP/LP compartment and in FrA. These data could be interpreted as if B cell development proceeds along two parallel developmental trajectories. However, both functional and molecular analysis suggested that GFRA2-expressing cells from the B220− (CLP) and B220+ (FrA) populations represented a highly similar or even identical developmental stage. Even though it should not be excluded that development may proceed directly from a GFRA2+ stage into the CD19+ compartment, our data support that the BST1+ population indeed represents a more mature compartment of CD19− cells. Our inability to identify committed B cells outside of the GFRA2+ progenitor compartments strongly suggests that the developmental trajectory defined by this marker represents the major pathway for B-lineage development from a functional CLP into a CD19+ pro–B cell. Even though the expression of KIT did not impact the lineage potential of the progenitor cells, it should be noted that the frequency of functional progenitor cells was reduced in the KIT-negative populations. Based on our findings, we would propose that the pro–B cell (BLP; Inlay et al., 2009) compartment can be divided into Ly6D+GFRA2− (BLP1), GFRA2+BST1− (BLP2), and GFRA2+BST1+ (BLP3) cells (Fig. 7). We believe this subdivision of early progenitor cells will increase our understanding of molecular events regulating normal as well as malignant B cell development.

Materials and methods
Animal models
Wt, Gfra2−/− (Rossi et al., 1999), *Ebf1−/−* (Lin and Grosschedl, 1995), an *hCD25* reporter under the regulatory elements of the

Jensen et al.
Heterogeneity in early lymphoid compartments

Journal of Experimental Medicine
https://doi.org/10.1084/jem.20171384

1958

Lambda5 (λ5, *Igll1*) promoter (Mårtensson et al., 1997), and Rag1-GFP- (Igarashi et al., 2002) mice were on a C57BL/6 background and backcrossed for more than seven generations. BM samples were harvested from *Wt* at 6–11 wk and from *Gfra2*[−/−] mice at 8 wk or embryonic day 18 FL. Comparative analyses were based on littermate controls. Animal procedures were performed with consent from the local ethics committees at Lund University and Linköping University.

FACS staining and cell sorting
For isolation of BM LP cells, bones were crushed in a mortar, and the cell suspension passed through a 50-µm filter. For progenitor cell isolation by FACS sorting, BM cells were either sorted unenriched, enriched for KIT[+] cells by magnet-activated cell sorting column for using anti-CD117 immunomagnetic beads (Miltenyi Biotec), or lineage depleted by using sheep anti–rat IgG beads (Dynabeads; Invitrogen), recognizing purified lineage antibodies against CD4 (GK1.5; BioLegend), CD8A (53–6.7; BioLegend), CD11B (M1/70; BioLegend), Ter119 (TER-119; BD Biosciences), GR1 (RB6-8C5; BioLegend), and CD19 (6D5; BioLegend) before antibody staining. Cells were thereafter stained with antibodies (full information of antibody combinations, clones, and conjugates is shown in Table S4). Cell sorting was performed on BD FACSAriaII or FACSAriaIII cell sorters (BD Biosciences) and analyses on a BD FACSLSRII, FACSAriaIII, FACSCantoII, or FACSFortessa.

BD Lyoplate mouse cell surface marker screening
The Lyoplate analysis was performed on 5 million unfractionated BM cells per sample according to the manufacturer's instructions (BD Biosciences). In addition to the Lyoplate antibodies, cells were costained with surface markers identifying CD19[+] and CLP/LP cells (see Table S4 for full antibody information).

In vitro evaluation of myeloid B–, NK-, and T-lineage potentials
For evaluation of B, NK, and T cell potential, SCs were deposited directly into 96-well plates containing preplated (2,000 cells/well) stroma cells. T-cultures (on OP9-DL1 stroma layers) were supplemented with 10 ng/ml KIT ligand, 10 ng/ml Fms-like tyrosine kinase 3 ligand (FLT3L), and 10 ng/ml IL-7. B/NK cultures (on OP9 stroma layers) were supplemented with the same cytokines as the OP9-DL1 cultures, with the addition of 20 ng/ml IL-15 and 40 ng/ml IL-2. All cytokines were acquired from PeproTech. Cultures were substituted with fresh cytokines after 7 d. OptiMEM supplemented with 10% heat-inactivated FCS, 50 µg/ml gentamicin, and 50 µM β-mercaptoethanol was used for maintaining the OP9/OP9-DL1 stromal cell lines as well as for the co-cultures. Co-cultures were evaluated by FACS with CD19 (ID3), NK1.1 (PK136), CD11c (N418), CD45 (30-F11), CD11b (M1/70), GR1 (RB6-8C5), CD25 (7D4), and CD90.2/Thy1.2 (53–2.1) antibodies. Cells were considered B cells if CD45[+]CD19[+], NK cells if CD45[+]NK1.1[+], and T cells if CD45[+]CD90.2[+]CD19[−]NK1.1[−]CD11c[−]CD11b[−]GR1[−] or CD45[+]CD90.2[+]CD25[+]CD19[−]NK1.1[−]CD11c[−]CD11b[−]GR1[−]. Co-cultures were analyzed at days 12–16 with a BD FACSLSRII analyzer (BD Biosciences).

Myeloid potential of SCs in vitro was determined by cultivation of cells in OptiMEM supplemented with 10% heat-inactivated FCS, 50 µg/ml gentamicin, and 50 µM β-mercaptoethanol.

The medium was complemented with 5 ng/ml IL-3, 5 ng/ml GM-CSF, 5 ng/ml M-CSF, 10 ng/ml G-CSF, 25 ng/ml Kit-ligand, and 25 ng/ml FLT3L (PeproTech). SC dispersion was based on limiting dilution (170 sorted cells diluted into 3.4 ml of medium) to generate on average one cell per 20 µl per well.

SC-qPCR analysis
SCs, with 10- and 20-cell controls, were sorted into 96-well plates with 4 µl lysis buffer (10% NP-40, 10 mM deoxynucleotide triphosphate, 0.1 M 1,4-dithiothreitol, RNAsout [Invitrogen], and nuclease-free water) and immediately frozen at –80°C. Upon thawing, Taq/SSIII enzyme reaction mix (CellsDirect One-Step RT-qPCR kit; Invitrogen) and 96 TaqMan assays at a final dilution of 0.025× were added to each well, excluding four wells where no RT control of Taq enzyme (Invitrogen) reaction mix was added. Preamplification cycling was preformed according to previously described protocols (Pina et al., 2012) with minor alterations; the first cycle, 50°C at 15 min, was extended to 60 min. The preamplified SCs were loaded on to a 48.48 or 96.96 Dynamic Array Chip for Gene Expression (Fluidigm) with 48 or 96 TaqMan assays, respectively. The unfractionated CLP Fluidigm data were analyzed by using the SCExV webtool (Lang et al., 2015). Gene expression was normalized by using the median expression and positive and negative control wells as well as cells showing the same expression patterns as those that were removed. Random forest clustering was applied as described in (Shi, 2006), and the resulting grouping was reordered to follow the expression changes of PAX5, EBF1, and CD114. Statistics on differentially expressed genes were calculated using the MAST R package according to the hurdle model, and differentials were called at a cut-off of P < 0.01 (Finak et al., 2015). The subfractionated CLP and FrA Fluidigm data were read into a Rscexv object, and all cells showing XRNA expression >25 CT (cycle of threshold) as well as the multicell controls were dropped. The data were normalized to the median expression per cell and median absolute deviation z-scored. Unsupervised clustering was performed by using the random forest clustering implemented in Rscexv. Subsequently, the expression of the B cell marker genes (*Pax5*, *Rag1*, *Ebf1*, and *Igll1*) was used to manually group the FACS sort and random forest clustered samples into the B cell–specific groups "not B cell committed," "preparing," and "B cell committed." Statistics on differentially expressed genes were calculated using the MAST R package according to the hurdle model, and differentials were called at a cut-off of P < 0.05 (Finak et al., 2015).

SC IgH DJ recombination assay
Single Lin[−]CD19[−]Kit[int/neg]IL7R[+]Ly6D[+] cells were index sorted, based on the mentioned markers and additionally GFRA2, BST1, and B220, directly into lysis buffer in 96-well plates. $D_H J_H$ rearrangement status was investigated as previously described (Rumfelt et al., 2006), with the exceptions that the Advantage 2 Polymerase Mix and PCR buffer (Takara Bio) were used, and PCR products were visualized on 2% agarose gels (E-Gel96; Invitrogen) by using an E-Base integrated power supply device (Thermo Fisher Scientific). The validity of these PCR products was not verified by sequence analysis. SC-qPCR data can be found in GEO accession no. GSE92540.

Next-generation sequencing and data analysis

Total RNA for RNA-seq was isolated from sorted CLP/LPs by using RNAeasy Micro Kit (Qiagen) according to manufacturer's recommendations. Libraries were constructed by using NuGEN's Ovation Ultralow Library systems (NuGEN Technologies) and were subsequently subjected to 50–76 cycles of sequencing on a HiSeq 2000 or NextSeq500 (Illumina). For analysis of LPAM1⁻Ly6D⁻ CLPs, LPAM1⁻Ly6D⁺ CLPs, and hCD25 (Igll1)⁺ LPs, reads were mapped by using HISAT (Kim et al., 2015) against the mm10 version of the mouse genome and counted against the Gencode M4 annotation. Differentially expressed genes were identified by using DESeq2 (Love et al., 2014) for *R* using a false discovery rate (FDR) <10% and reduced to surface marker proteins only by using a surface protein atlas (Bausch-Fluck et al., 2015). Replicate samples were averaged before generating heatmaps. For analysis of Ly6D⁻ GFRA2⁻, Ly6D⁺ GFRA2⁻, and Ly6D⁺ GFRA2⁺ CLPs and hCD25⁺ LPs, the reads were aligned to mouse reference genome (mm10/GRCm38) using TopHat (Kim et al., 2013). If not indicated, further analyses were performed using the HOMER platform (Heinz et al., 2010). After conversion of Bam-files to Tag directories with the command makeTagdirectory.pl, samples were analyzed by using normalization to 10M mapped reads by the analyzeRepeats.pl command with the options –count exons -condeseGenes. For analysis of statistical significance among differently expressed genes between the three different CLP populations, the data were analyzed using analyzeRepeats.pl with the –noadj option followed by the getDiffExpression.pl command by using EdgeR (Robinson et al., 2010) analyzing -AvsA with parameters -fdr 0.10 -log2fold 1. Heatmaps were created by hierarchical clustering in Cluster3 (de Hoon et al., 2004) followed by visualization in Java Treeview (Saldanha, 2004).

RNA-seq data from *Wt*, *EBF⁻/⁻*, and *Pax5⁻/⁻* pro–B cells, ATAC-seq data from *Wt* and *EBF⁻/⁻* pro–B cells, and pro–B EBF1 ChIP-seq data were mapped against reference genome (mm10/GRCm38) by using TopHat (Kim et al., 2013) and Bowtie 2 (Langmead and Salzberg, 2012), respectively. Further analyses were performed using HOMER (Heinz et al., 2010) with an initial conversion of Bam files to tag directories by using makeTagdirectory.pl. Expression matrices of RNA-seq data normalized to 10M mapped reads were created by the analyzeRepeats.pl command with the options –count exons -condeseGenes. ATAC-seq peaks were identified by using findPeaks.pl in HOMER with the settings -style histone -size 75 -minDist 75 -minTagThreshold 6 -L 8 -F 8. ChIP-seq peaks were identified by using findPeaks.pl in HOMER with the setting -style factor. Files for fishbone heatmaps were created by using annotatePeaks.pl -size 5000 -hist 25 -ghist, and files for histograms were created using annotatPeaks.pl with the options -size 5000 -hist 25 in HOMER.

Heatmaps were finalized by hierarchical clustering in Cluster3 (de Hoon et al., 2004) followed by visualization in Java Treeview (Saldanha, 2004). Gene expression heatmaps were clustered by using log transformation center genes on means, and average linkage (uncentered correlation), hierarchical clustering of genes and arrays, and fishbone heatmaps by clustering of genes only by using average linkage (uncentered correlation). Publicly available data retrieved from GEO and used in this report was pro–B EBF1 ChIP (accession no. GSE69227; Ungerbäck et al., 2015). RNA-seq and ATAC-seq data from Wt and EBF1-deficient cells as well as CLP subfractions were deposited in GEO (accession no. GSE92434 and GSE92540).

SC RNA-seq data were generated on the 10X emulsion platform (10X Genomics) according to the manufacturer's instructions. The SC RNA-seq from the Chromium platform was processed using CellRanger v1.3 and default settings, and subsequent normalization, quality control, filtering, differential gene expression analysis, and cell-cycle assignment were performed in R by using Seurat v2.2 (Butler et al., 2018). Dimension reduction was performed using DDRTree over three components. SC RNA-seq data was deposited in GEO (accession no. GSE92540).

In vitro analysis of population dynamics

Subfractions of CLP and FrA cells were plated at a concentration of 50 cells per well in 12-well plates containing preplated (27,000 cells per well) OP9 stroma cells. Cultures were kept in OptiMEM supplemented with 10% heat-inactivated FCS, 50 µg/ml gentamicin, and 50 µM β-mercaptoethanol and were supplemented with 10 ng/ml KIT ligand, 10 ng/ml FLT3L, 10 ng/ml IL-7, 20 ng/ml IL-15, and 40 ng/ml IL-2. All cytokines were acquired from PeproTech. Cultures were substituted with fresh cytokines after 7 d. Cell cultures were analyzed by FACS at days 2, 4, 6, 8, and 12 and stained against NK1.1 (PK136; BioLegend), CD19 (1D3; eBioscience), CD45.2 (104; BioLegend), BST1 (BP3; BioLegend), B220 (RA3-6B2; BioLegend), and GFRA2 (R&D Systems). Cell expansion in cultures was estimated by FACS through the addition of a standardized number of beads (Trucount Control; BD Biosciences) to all FACS samples.

Transplantation assays of NSG mice for evaluation of in vivo lineage potential

Bones (femur, tibiae, crista iliac) from 10–12-wk-old CD45.2 mice were taken, and hematopoietic BM cells were extracted by crushing the bones in mortar. Red cells were removed by adding ice-cold Erythrocyte Lysis Buffer (1.5 M NH₄Cl, 100 mM NaHCO₃, and 10 mM EDTA, pH 7.4). Lineage-positive cells were removed by staining with purified antibodies CD19(6D5), CD8a(53–6.7), CD4(GK1.5), Gr1(RB6-8C5), Mac1(M1/70), and Ter119(TER-119), followed by negative selection by using Dynabeads Sheep Anti–Rat IgG (11035; Invitrogen). Remaining cells were next incubated with biotinylated mGFRα-2 (BAF429; R&D Systems) and further stained with lineage-PECy5 Gr1(RB6-8C5), Mac1(M1/70), Ter119(TER-119), CD11c(N418), NK1.1(PK136), CD3 ε(145–2c11), CD19-PECy7(ID3; eBioscience), B220-Alexa Flour 700 (RA3-6B2), and Streptavidin-Brilliant Violet (BV) 605. Lin⁻CD19⁻B220⁺GFRA2⁺ and Lin⁻CD19⁻B220⁺GFRA2⁻ cells were sorted on BD FACSAriaIII, and 2,000 cells per 150 µl 1% PBS/FCS were transplanted by intravenous injection via tail vein to 9–11-wk-old preconditioned (2.25 Gray) NOD.Cg-*Prkdc^scid Il2rg^tm1Wjl*/SzJ (NSG) mice. Peripheral blood was analyzed 2 wk after injection, and at 4 wk peripheral blood, spleen, and BM were taken for flow cytometric analysis by using IgM-APCCy7 (RMM-1) and CD4-APC (GK1.5) or CD49b-APC (DX5), CD19-PECy7 (ID3; eBiosciences), CD3ε-BV605 (17A2), CD8-PacificBlue (53–6.7), NK1.1-FITC (PK136), CD45.1-PE (A20; eBiosciences), and CD45.2-Alexa Fluor 700 (104). All antibodies are from BioLegend, unless otherwise indicated.

For analysis of EBF1-deficient cells, 2,000 sorted LSK cells from Wt or Ebf1$^{-/-}$ FL were transplanted into sublethally irradiated NSG mice by tail vein injection. The BM samples of transplanted mice were subjected to FACS analysis 6 wk after transplantation to identify LP populations.

Modulation of NOTCH signal strength by DAPT addition
Subfractions of CLP cells were plated at a concentration of 20 cells per well in 24-well plates containing preplated (13,000 cells per well) OP9DL1 stroma cells. Cultures were kept in OptiMEM supplemented with 10% heat-inactivated FCS, 50 µg/ml gentamicin, and 50 µM β-mercaptoethanol and were supplemented with 10 ng/ml KIT ligand, 10 ng/ml FLT3L, and 10 ng/ml IL-7 (PeproTech). Cultures were substituted with fresh DAPT and cytokines after 7 d. Cultures were analyzed by FACS at day 14 and stained against CD19 (ID3), NK1.1 (PK136), CD11c (N418), CD45 (30-F11), CD11b (M1/70), GR1 (RB6-8C5), and CD90.2/Thy1.2 (53–2.1).

Intracellular FACS staining of PAX5 and EBF1
WT mice were individually lineage-depleted by using sheep anti–rat IgG beads (Dynabeads; Invitrogen) recognizing purified lineage antibodies against CD4 (GK1.5; BioLegend), CD8A (53–6.7; BioLegend), CD11B (M1/70; BioLegend), Ter119 (TER-119; BD Biosciences), and GR1 (RB6-8C5; BioLegend) and subsequently stained against surface markers (Table S4). Fixation, permeabilization, and staining of TFs were done by using the Transcription Factor Buffer Set (BD Pharmingen) according to the manufacturer's instructions.

Online supplemental material
Table S1 includes detailed data from the antibody screening process. Tables S2 and S3 show differentially expressed genes in the SC-qPCR experiments. Table S4 shows the FACS antibodies used to identify LP populations. Tables S5 and S6 show statistical analysis of SC in vitro culture data. Fig. S1 shows back gating of defined populations and FACS data from kinetics experiments. Fig. S2 shows SC PCR analysis of CLP cells and data from DMSO control experiments for the Notch inhibition experiments in Fig. 2 as well as intracellular FACS analysis of EBF1 and PAX5. Fig. S3 shows in vitro analysis of B220$^+$ progenitor populations and FACS data from defined progenitors in BM and FL in Wt and Gfra2$^{-/-}$ mice. Fig. S4 shows DDRTree projections highlighting gene expression patterns obtained from SC-RNA-seq analysis, and Fig. S5 shows FACS plots from reporter transgenic mice, gating strategies, qPCR data from sorted populations, and SC IgH DJ analysis.

References
Alberti-Servera, L., L. von Muenchow, P. Tsapogas, G. Capoferri, K. Eschbach, C. Beisel, R. Ceredig, R. Ivanek, and A. Rolink. 2017. Single-cell RNA sequencing reveals developmental heterogeneity among early lymphoid progenitors. *EMBO J.* 36:3619–3633. https://doi.org/10.15252/embj .201797105

Allman, D., A. Sambandam, S. Kim, J.P. Miller, A. Pagan, D. Well, A. Meraz, and A. Bhandoola. 2003. Thymopoiesis independent of common lymphoid progenitors. *Nat. Immunol.* 4:168–174. https://doi.org/10.1038/ni878

Banerjee, A., D. Northrup, H. Boukarabila, S.E. Jacobsen, and D. Allman. 2013. Transcriptional repression of Gata3 is essential for early B cell commitment. *Immunity.* 38:930–942. https://doi.org/10.1016/j .immuni.2013.01.014

Bausch-Fluck, D., A. Hofmann, T. Bock, A.P. Frei, F. Cerciello, A. Jacobs, H. Moest, U. Omasits, R.L. Gundry, C. Yoon, et al. 2015. A mass spectrometric-derived cell surface protein atlas. *PLoS One.* 10:e0121314. https://doi .org/10.1371/journal.pone.0121314

Butler, A., P. Hoffman, P. Smibert, E. Papalexi, and R. Satija. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36:411–420. https://doi.org/10.1038/nbt.4096

Cik, M., S. Masure, A.S. Lesage, I. Van Der Linden, P. Van Gompel, M.N. Pangalos, R.D. Gordon, and J.E. Leysen. 2000. Binding of GDNF and neurturin to human GDNF family receptor alpha 1 and 2. Influence of cRET and cooperative interactions. *J. Biol. Chem.* 275:27505–27512.

de Hoon, M.J., S. Imoto, J. Nolan, and S. Miyano. 2004. Open source clustering software. *Bioinformatics.* 20:1453–1454. https://doi.org/10.1093/ bioinformatics/bth078

Finak, G., A. McDavid, M. Yajima, J. Deng, V. Gersuk, A.K. Shalek, C.K. Slichter, H.W. Miller, M.J. McElrath, M. Prlic, et al. 2015. MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 16:278. https://doi.org/10.1186/s13059-015-0844-5

Fonseca-Pereira, D., S. Arroz-Madeira, M. Rodrigues-Campos, I.A. Barbosa, R.G. Domingues, T. Bento, A.R. Almeida, H. Ribeiro, A.J. Potocnik, H. Enomoto, and H. Veiga-Fernandes. 2014. The neurotrophic factor receptor RET drives haematopoietic stem cell survival and function. *Nature.* 514:98–101. https://doi.org/10.1038/nature13498

Funaro, A., E. Ortolan, B. Ferranti, L. Gargiulo, R. Notaro, L. Luzzatto, and F. Malavasi. 2004. CD157 is an important mediator of neutrophil adhesion and migration. *Blood.* 104:4269–4278. https://doi.org/10.1182/blood -2004-06-2129

Hardy, R.R., C.E. Carmack, S.A. Shinton, J.D. Kemp, and K. Hayakawa. 1991. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* 173:1213–1225. https://doi.org/ 10.1084/jem.173.5.1213

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, and C.K. Glass. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell.* 38:576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Igarashi, H., S.C. Gregory, T. Yokota, N. Sakaguchi, and P.W. Kincade. 2002. Transcription from the RAG1 locus marks the earliest lymphocyte progenitors in bone marrow. *Immunity.* 17:117–130. https://doi.org/10.1016/ S1074-7613(02)00366-7

Ikawa, T., H. Kawamoto, L.Y. Wright, and C. Murre. 2004. Long-term cultured E2A-deficient hematopoietic progenitor cells are pluripotent. *Immunity*. 20:349–360. https://doi.org/10.1016/S1074-7613(04)00049-4

Inlay, M.A., D. Bhattacharya, D. Sahoo, T. Serwold, J. Seita, H. Karsunky, S.K. Plevritis, D.L. Dill, and I.L. Weissman. 2009. Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes Dev*. 23:2376–2381. https://doi.org/10.1101/gad.1836009

Ishihara, K., Y. Kobune, Y. Okuyama, M. Itoh, B.O. Lee, O. Muraoka, and T. Hirano. 1996. Stage-specific expression of mouse BST-1/BP-3 on the early B and T cell progenitors prior to gene rearrangement of antigen receptor. *Int. Immunol*. 8:1395–1404. https://doi.org/10.1093/intimm/8.9.1395

Itoh, M., K. Ishihara, T. Hiroi, B.O. Lee, H. Maeda, H. Iijima, M. Yanagita, H. Kiyono, and T. Hirano. 1998. Deletion of bone marrow stromal cell antigen-1 (CD157) gene impaired systemic thymus independent-2 antigen-induced IgG3 and mucosal TD antigen-elicited IgA responses. *J. Immunol*. 161:3974–3983.

Jensen, C.T., S. Lang, R. Somasundaram, S. Soneji, and M. Sigvardsson. 2016. Identification of stage-specific surface markers in early B cell development provides novel tools for identification of progenitor populations. *J. Immunol*. 197:1937–1944. https://doi.org/10.4049/jimmunol.1600297

Karsunky, H., M.A. Inlay, T. Serwold, D. Bhattacharya, and I.L. Weissman. 2008. Flk2+ common lymphoid progenitors possess equivalent differentiation potential for the B and T lineages. *Blood*. 111:5562–5570. https://doi.org/10.1182/blood-2007-11-126219

Katsura, Y. 2002. Redefinition of lymphoid progenitors. *Nat. Rev. Immunol*. 2:127–132. https://doi.org/10.1038/nri721

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S.L. Salzberg. 2013. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 14:R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kim, D., B. Langmead, and S.L. Salzberg. 2015. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods*. 12:357–360. https://doi.org/10.1038/nmeth.3317

Kondo, M., I.L. Weissman, and K. Akashi. 1997. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*. 91:661–672. https://doi.org/10.1016/S0092-8674(00)80453-5

Lang, S., A. Ugale, E. Erlandsson, G. Karlsson, D. Bryder, and S. Soneji. 2015. SCExV: A webtool for the analysis and visualisation of single cell qRT-PCR data. *BMC Bioinformatics*. 16:320. https://doi.org/10.1186/s12859-015-0757-z

Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. 9:357–359. https://doi.org/10.1186/s12859-015-0757-z

Li, Y.S., K. Hayakawa, and R.R. Hardy. 1993. The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver. *J. Exp. Med*. 178:951–960. https://doi.org/10.1084/jem.178.3.951

Li, Y.S., R. Wasserman, K. Hayakawa, and R.R. Hardy. 1996. Identification of the earliest B lineage stage in mouse bone marrow. *Immunity*. 5:527–535. https://doi.org/10.1016/S1074-7613(00)80268-X

Lin, H., and R. Grosschedl. 1995. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature*. 376:263–267. https://doi.org/10.1038/376263a0

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550. https://doi.org/10.1186/s13059-014-0550-8

Mansson, R., S. Zandi, K. Anderson, I.L. Martensson, S.E. Jacobsen, D. Bryder, and M. Sigvardsson. 2008. B-lineage commitment prior to surface expression of B220 and CD19 on hematopoietic progenitor cells. *Blood*. 112:1048–1055. https://doi.org/10.1182/blood-2007-11-125385

Mansson, R., S. Zandi, E. Welinder, P. Tsapogas, N. Sakaguchi, D. Bryder, and M. Sigvardsson. 2010. Single-cell analysis of the common lymphoid progenitor compartment reveals functional and molecular heterogeneity. *Blood*. 115:2601–2609. https://doi.org/10.1182/blood-2009-08-236398

Mansson, R., E. Welinder, J. Åhsberg, Y.C. Lin, C. Benner, C.K. Glass, J.S. Lucas, M. Sigvardsson, and C. Murre. 2012. Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. *Proc. Natl. Acad. Sci. USA*. 109:21028–21033. https://doi.org/10.1073/pnas.1211427109

Mårtensson, I.L., F. Melchers, and T.H. Winkler. 1997. A transgenic marker for mouse B lymphoid precursors. *J. Exp. Med*. 185:653–661. https://doi.org/10.1084/jem.185.4.653

Martin, C.H., I. Aifantis, M.L. Scimone, U.H. von Andrian, B. Reizis, H. von Boehmer, and F. Gounari. 2003. Efficient thymic immigration of B220+ lymphoid-restricted bone marrow cells with T precursor potential. *Nat. Immunol*. 4:866–873. https://doi.org/10.1038/ni965

McNagny, K.M., P.A. Cazenave, and M.D. Cooper. 1988. BP-3 alloantigen. A cell surface glycoprotein that marks early B lineage cells and mature myeloid lineage cells in mice. *J. Immunol*. 141:2551–2556.

Mikkola, I., B. Heavey, M. Horcher, and M. Busslinger. 2002. Reversion of B cell commitment upon loss of Pax5 expression. *Science*. 297:110–113. https://doi.org/10.1126/science.1067518

Nechanitzky, R., D. Akbas, S. Scherer, I. Györy, T. Hoyler, S. Ramamoorthy, A. Diefenbach, and R. Grosschedl. 2013. Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nat. Immunol*. 14:867–875. https://doi.org/10.1038/ni.2641

Nutt, S.L., B. Heavey, A.G. Rolink, and M. Busslinger. 1999. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature*. 401:556–562. https://doi.org/10.1038/44076

Paul, F., Y. Arkin, A. Giladi, D.A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner, et al. 2015. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell*. 163:1663–1677. https://doi.org/10.1016/j.cell.2015.11.013

Pereira de Sousa, A., C. Berthault, A. Granato, S. Dias, C. Ramond, B.L. Kee, A. Cumano, and P. Vieira. 2012. Inhibitors of DNA binding proteins restrict T cell potential by repressing Notch1 expression in Flt3-negative common lymphoid progenitors. *J. Immunol*. 189:3822–3830. https://doi.org/10.4049/jimmunol.1103723

Perié, L., K.R. Duffy, L. Kok, R.J. de Boer, and T.N. Schumacher. 2015. The branching point in erythro-myeloid differentiation. *Cell*. 163:1655–1662. https://doi.org/10.1016/j.cell.2015.11.059

Pina, C., C. Fugazza, A.J. Tipping, J. Brown, S. Soneji, J. Teles, C. Peterson, and T. Enver. 2012. Inferring rules of lineage commitment in haematopoiesis. *Nat. Cell Biol*. 14:287–294. https://doi.org/10.1038/ncb2442

Pongubala, J.M., D.L. Northrup, D.W. Lancki, K.L. Medina, T. Treiber, E. Bertolino, M. Thomas, R. Grosschedl, D. Allman, and H. Singh. 2008. Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. *Nat. Immunol*. 9:203–215. https://doi.org/10.1038/ni1555

Possot, C., S. Schmutz, S. Chea, L. Boucontet, A. Louise, A. Cumano, and R. Golub. 2011. Notch signaling is necessary for adult, but not fetal, development of RORγt(+) innate lymphoid cells. *Nat. Immunol*. 12:949–958. https://doi.org/10.1038/ni.2105

Pridans, C., M.L. Holmes, M. Polli, J.M. Wettenhall, A. Dakic, L.M. Corcoran, G.K. Smyth, and S.L. Nutt. 2008. Identification of Pax5 target genes in early B cell differentiation. *J. Immunol*. 180:1719–1728. https://doi.org/10.4049/jimmunol.180.3.1719

Revilla-I-Domingo, R., I. Bilic, B. Vilagos, H. Tagoh, A. Ebert, I.M. Tamir, L. Smeenk, J. Trupke, A. Sommer, M. Jaritz, and M. Busslinger. 2012. The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO J*. 31:3130–3146. https://doi.org/10.1038/emboj.2012.155

Robinson, M.D., D.J. McCarthy, and G.K. Smyth. 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26:139–140. https://doi.org/10.1093/bioinformatics/btp616

Rolink, A., U. Grawunder, T.H. Winkler, H. Karasuyama, and F. Melchers. 1994. IL-2 receptor alpha chain (CD25, TAC) expression defines a crucial stage in pre-B cell development. *Int. Immunol*. 6:1257–1264. https://doi.org/10.1093/intimm/6.8.1257

Rolink, A.G., S.L. Nutt, F. Melchers, and M. Busslinger. 1999. Long-term in vivo reconstitution of T-cell development by Pax5-deficient B-cell progenitors. *Nature*. 401:603–606. https://doi.org/10.1038/44164

Rossi, J., K. Luukko, D. Poteryaev, A. Laurikainen, Y.F. Sun, T. Laakso, S. Eerikäinen, R. Tuominen, M. Lakso, H. Rauvala, et al. 1999. Retarded growth and deficits in the enteric and parasympathetic nervous system in mice lacking GFR alpha2, a functional neurturin receptor. *Neuron*. 22:243–252. https://doi.org/10.1016/S0896-6273(00)81086-7

Rumfelt, L.L., Y. Zhou, B.M. Rowley, S.A. Shinton, and R.R. Hardy. 2006. Lineage specification and plasticity in CD19- early B cell precursors. *J. Exp. Med*. 203:675–687. https://doi.org/10.1084/jem.20052444

Saldanha, A.J. 2004. Java Treeview–extensible visualization of microarray data. *Bioinformatics*. 20:3246–3248. https://doi.org/10.1093/bioinformatics/bth349

Sanicola, M., C. Hession, D. Worley, P. Carmillo, C. Ehrenfels, L. Walus, S. Robinson, G. Jaworski, H. Wei, R. Tizard, et al. 1997. Glial cell line-derived

neurotrophic factor-dependent RET activation can be mediated by two different cell-surface accessory proteins. *Proc. Natl. Acad. Sci. USA.* 94:6238–6243. https://doi.org/10.1073/pnas.94.12.6238

Schmitt, T.M., and J.C. Zúñiga-Pflücker. 2002. Induction of T cell development from hematopoietic progenitor cells by delta-like-1 in vitro. *Immunity.* 17:749–756. https://doi.org/10.1016/S1074-7613(02)00474-0

Sen, J., N. Rosenberg, and S.J. Burakoff. 1990. Expression and ontogeny of CD2 on murine B cells. *J. Immunol.* 144:2925–2930.

Shi, T.H.S. 2006. Unsupervised learning with random forest predictors. *J. Comput. Graph. Stat.* 15:118–138. https://doi.org/10.1198/106186006X94072

Thal, M.A., T.L. Carvalho, T. He, H.G. Kim, H. Gao, J. Hagman, and C.A. Klug. 2009. Ebf1-mediated down-regulation of Id2 and Id3 is essential for specification of the B cell lineage. *Proc. Natl. Acad. Sci. USA.* 106:552–557. https://doi.org/10.1073/pnas.0802550106

Treiber, T., E.M. Mandel, S. Pott, I. Györy, S. Firner, E.T. Liu, and R. Grosschedl. 2010. Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. *Immunity.* 32:714–725. https://doi.org/10.1016/j.immuni.2010.04.013

Tsapogas, P., S. Zandi, J. Åhsberg, J. Zetterblad, E. Welinder, J.I. Jönsson, R. Månsson, H. Qian, and M. Sigvardsson. 2011. IL-7 mediates Ebf-1-dependent lineage restriction in early lymphoid progenitors. *Blood.* 118:1283–1290. https://doi.org/10.1182/blood-2011-01-332189

Tudor, K.S., K.J. Payne, Y. Yamashita, and P.W. Kincade. 2000. Functional assessment of precursors from murine bone marrow suggests a sequence of early B lineage differentiation events. *Immunity.* 12:335–345. https://doi.org/10.1016/S1074-7613(00)80186-7

Ungerbäck, J., J. Åhsberg, T. Strid, R. Somasundaram, and M. Sigvardsson. 2015. Combined heterozygous loss of Ebf1 and Pax5 allows for T-lineage conversion of B cell progenitors. *J. Exp. Med.* 212:1109–1123. https://doi.org/10.1084/jem.20132100

Welinder, E., R. Mansson, E.M. Mercer, D. Bryder, M. Sigvardsson, and C. Murre. 2011. The transcription factors E2A and HEB act in concert to induce the expression of FOXO1 in the common lymphoid progenitor. *Proc. Natl. Acad. Sci. USA.* 108:17402–17407. https://doi.org/10.1073/pnas.1111766108

Zandi, S., R. Mansson, P. Tsapogas, J. Zetterblad, D. Bryder, and M. Sigvardsson. 2008. EBF1 is essential for B-lineage priming and establishment of a transcription factor network in common lymphoid progenitors. *J. Immunol.* 181:3364–3372. https://doi.org/10.4049/jimmunol.181.5.3364

Zhang, Z., C.V. Cotta, R.P. Stephan, C.G. deGuzman, and C.A. Klug. 2003. Enforced expression of EBF in hematopoietic stem cells restricts lymphopoiesis to the B cell lineage. *EMBO J.* 22:4759–4769. https://doi.org/10.1093/emboj/cdg464