Supplemental material

Jensen et al., https://doi.org/10.1084/jem.20171384

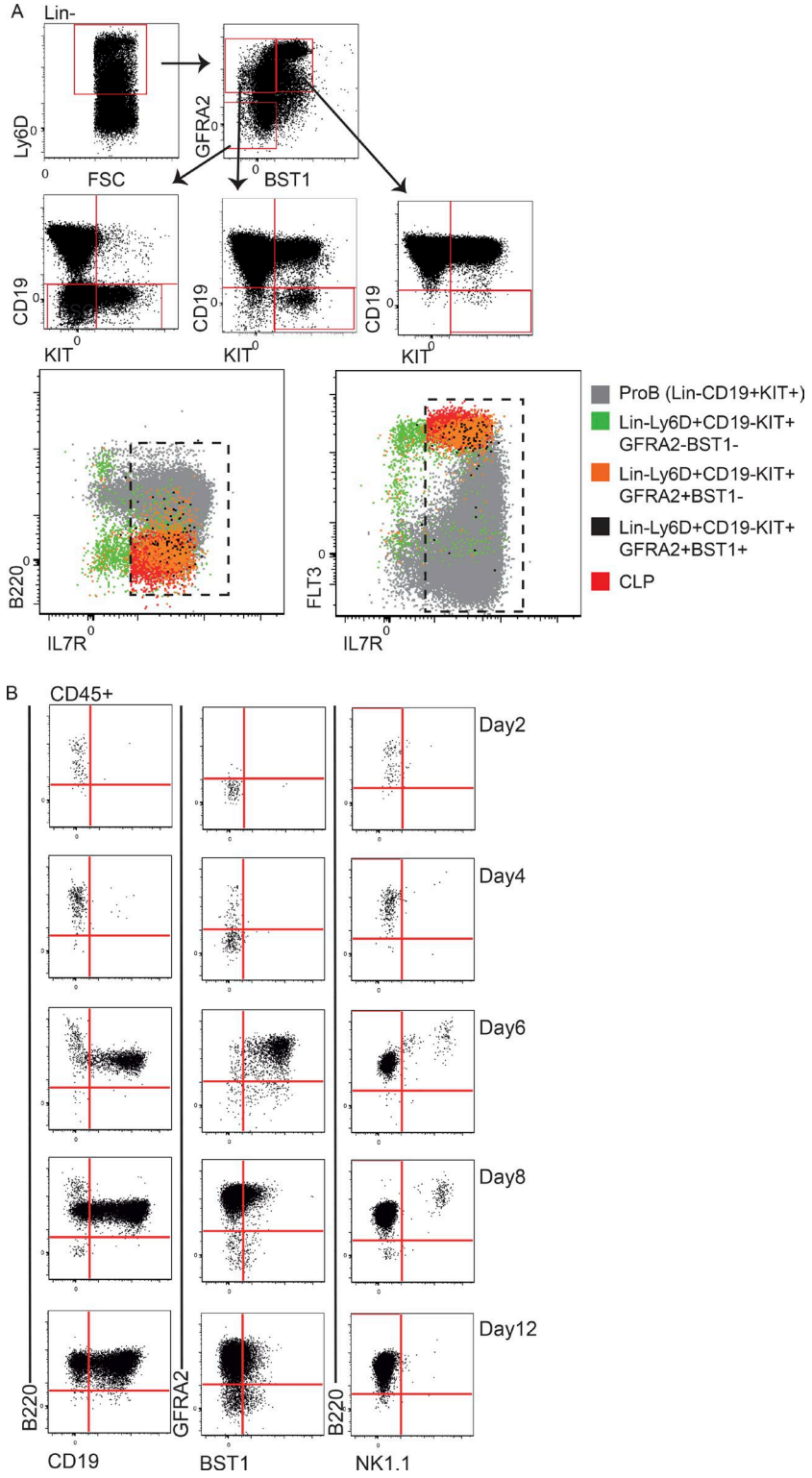

Figure S1. **Coexpression of surface markers identifies unique GFRA-2– and BST1-positive populations. (A)** FACS analysis of CD19, B220, KIT, IL7R, and FLT3 expression on Lin$^-$Ly6D$^+$GFRA2$^-$BST1$^-$, Lin-Ly6D$^+$GFRA2$^+$BST1$^-$, and Lin$^-$Ly6D$^+$GFRA2$^+$BST1$^+$ BM cells. Pro–B cells (Lin$^-$CD19$^+$KIT$^+$) and CLP (Lin$^-$CD19$^-$B220$^-$FLT3$^+$IL7R$^+$KIT^{int}SCA1^{int}) are shown as reference populations. FACS plots show a concatenation of three samples. **(B)** FACS plots show an example of the formation of cells on the indicated days after seeding of 50 CLPLy6D$^+$GFRA2$^-$ progenitors on OP9 stroma cells under B and NK cell differentiation conditions (Fig. 2, B–D).

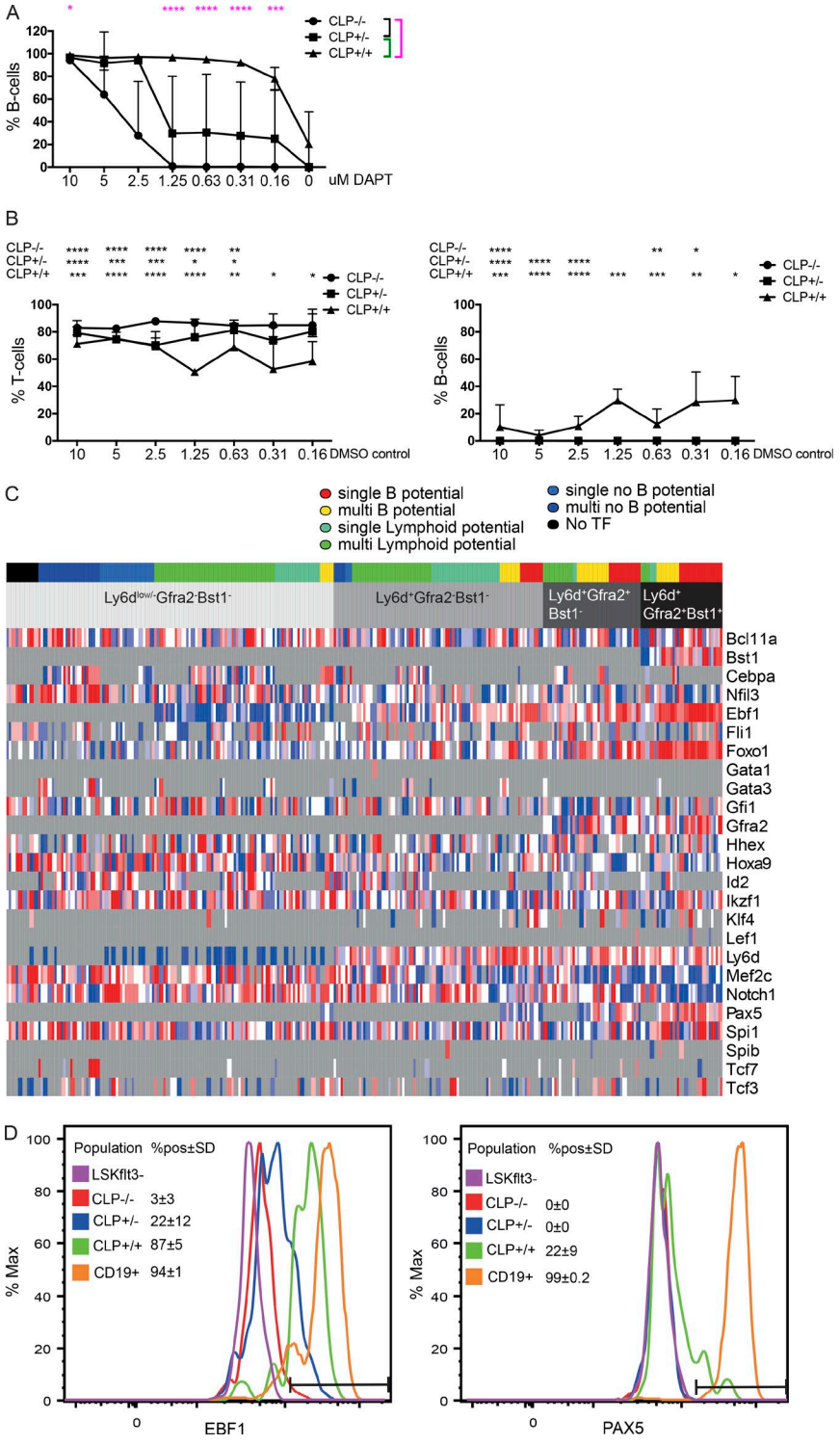

Figure S2. **Functional and molecular analyses reveal unique features of GFRA2⁺ progenitors. (A and B)** Diagrams presenting the mean cellular composition (SD) in cultures (n = 3) generated after seeding 20 CLPLy6D⁻GFRA2⁻ (CLP$^{-/-}$), CLPLy6D⁺GFRA2⁻ (CLP$^{+/-}$) or CLPLy6D⁺GFRA2⁺ (CLP$^{+/+}$) cells on OP9DL1 stroma under T cell differentiation conditions and supplemented with a twofold dilution series of DAPT or DMSO controls. **(A)** The percentage of B cells in culture after 14 d in cultures supplemented with DAPT. *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001 (Student's t test); from a total of three samples per condition from three experiments. **(B)** The percentage of T and B cells after 14 d in cultures supplemented with DMSO only. **(B)** P values above the graph refer to the percentage of T and B cells between DAPT and DMSO controls. *, P < 0.05; **, P < 0.01; ***, P < 0.001; ****, P < 0.0001 (Student's t test); from a total of three samples per condition from three experiments. **(C)** Heatmap showing SC-qPCR data from 338 CLP cells. Cells are grouped based on expression of *Ly6D*, *Gfra2*, and *Bst1*. Colors at the top show coexpression of the TFs *Ebf1, Pax5, Cebpa*, *Id2*, and *Gata3*. Single B potential, *Ebf1* and *Pax5* expression; multi B potential, *Ebf1*, *Pax5*, and other TFs); single lymphoid potential, *Ebf1*; multi lymphoid potential, *Ebf1* and TFs other than *Pax5*; single no B potential, TFs other than *Ebf1* and *Pax5*; and multi no B potential, two or more TFs other than *Ebf1* and *Pax5*. **(D)** Histograms based on FACS data displaying EBF1 and PAX5 expression in defined CLP subpopulations as indicated (n = 3, ±SD).

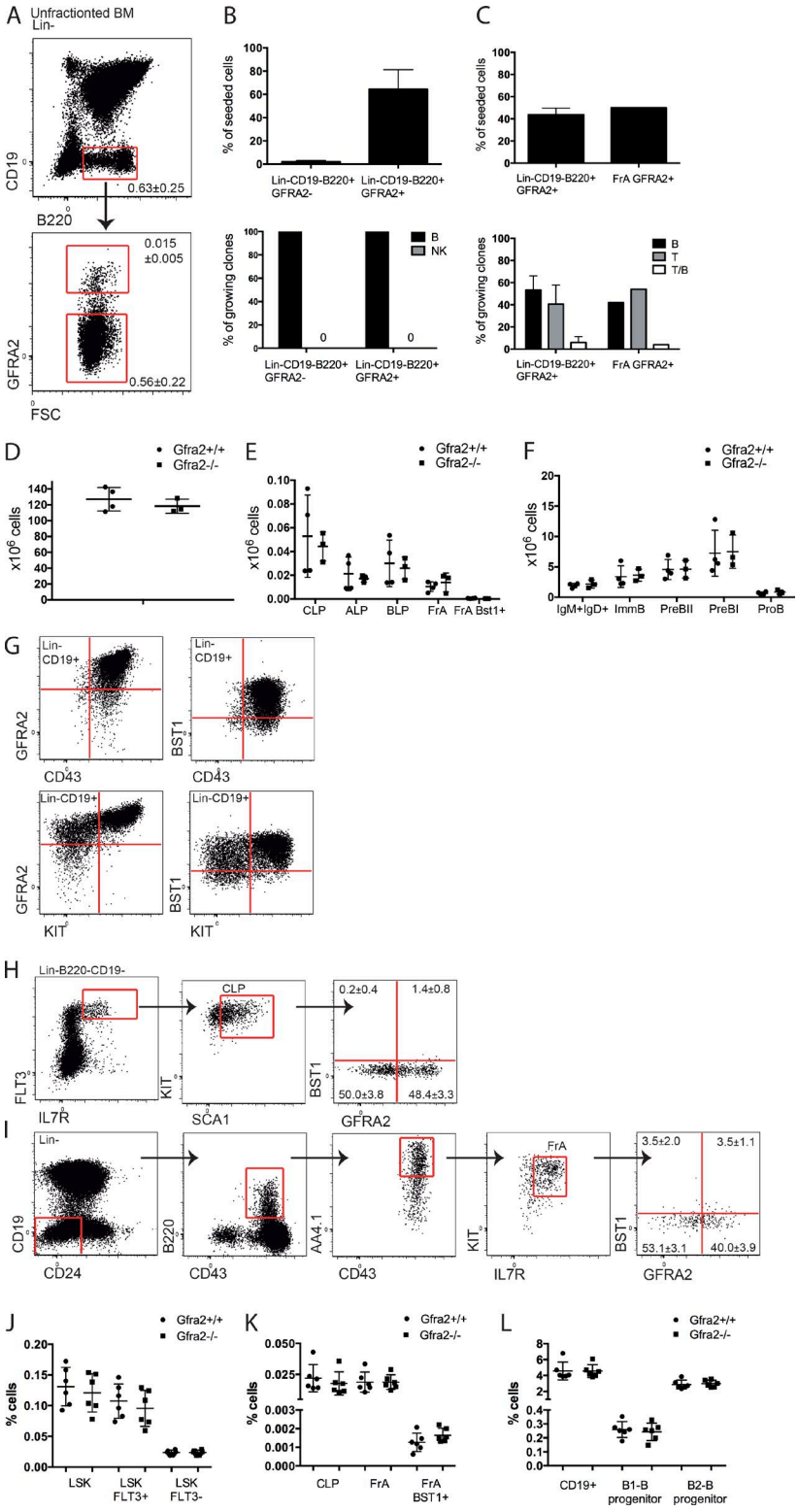

Figure S3. **GFRA2 defines progenitor compartments but is not critical for normal B cell development. (A)** Representative FACS profiles of Lin$^-$CD19$^-$B220$^+$ cells. Numbers in FACS plots shows the mean percentage of viable cells in unfractionated BM (SD, n = 6). **(B)** Cloning efficiency and lineage potentials of Lin$^-$CD19$^-$B220$^+$GFRA2$^{+/-}$ cells grown on OP9 stromal cell layers (286 seeded cells per population from three samples and three experiments). **(C)** Cloning efficiency and lineage potentials of Lin$^-$CD19$^-$B220$^+$GFRA2$^+$ (240 seeded cells from three samples and two experiments) cells grown on OP9DL1 stromal cell layers (with a GFRA2$^+$ FrA reference sample, 48 seeded cells). **(D–F)** BM cellularity and FACS-based phenotyping of the BM in adult Gfra2-deficient mice (Gfra2$^{+/+}$: n = 4; Gfra2$^{-/-}$: n = 3). **(G–I)** FACS profiles of GFRA2 and BST1 expression in FL (G) CD19$^+$ cells (H), FrA cells, and (I) CLP cells. Numbers in FACS plots show the mean percentage of indicated cell types in unfractionated FL (SD, n = 6). **(J–L)** FACS phenotyping of the FL at embryonic day 18 in Gfra2-deficient mice (Gfra2$^{+/+}$: n = 6; Gfra2$^{-/-}$: n = 6).

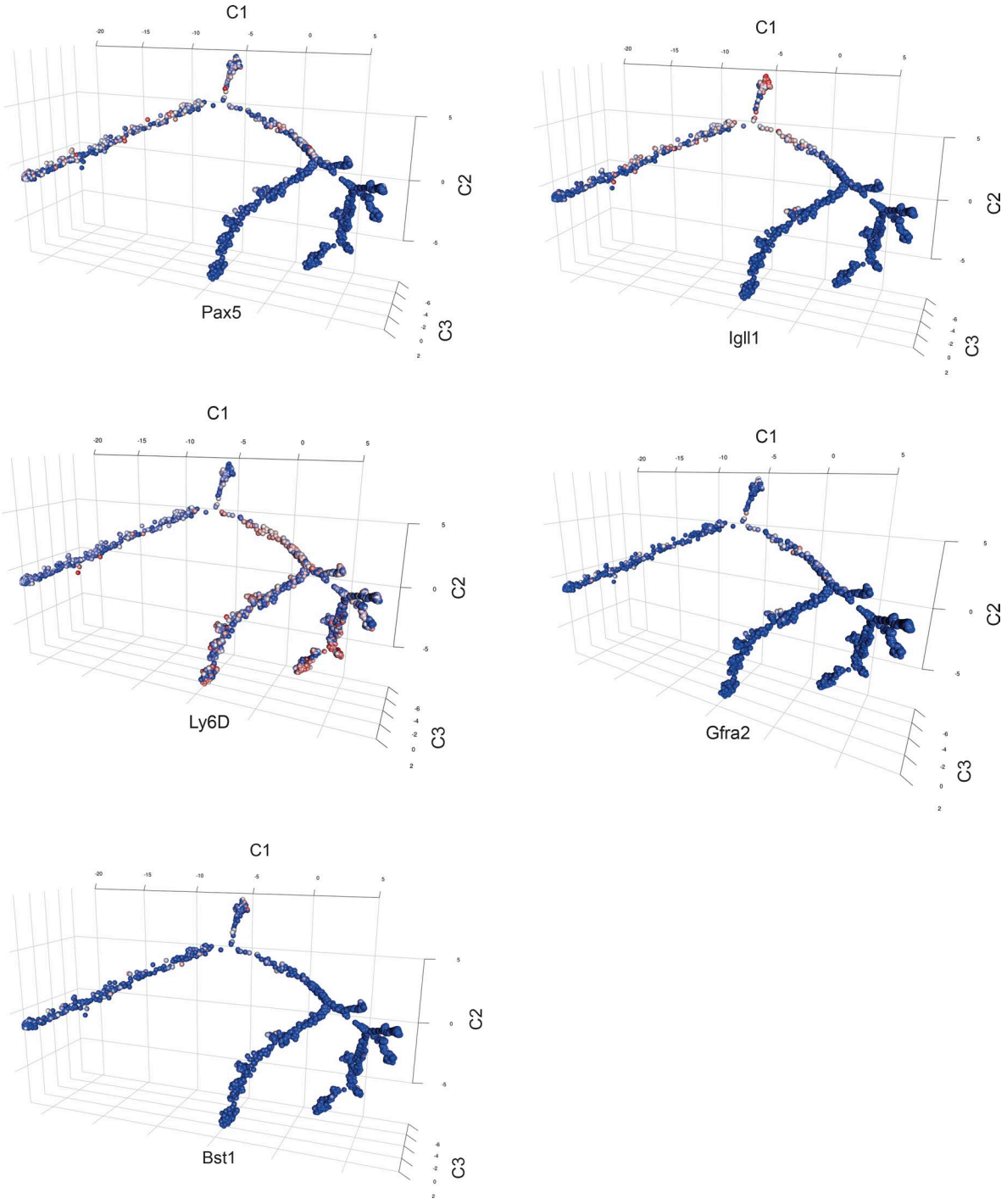

Figure S4. **Expression of lineage-restricted genes identifies committed B cell progenitors.** SC RNA-seq data from 1,753 CLPs (LIN$^-$FLT3$^+$IL-7R$^+$KIT$^+$SCA1$^+$), 279 FrA (LIN$^-$ B220$^+$IL7R$^+$KIT$^+$), and 472 CD19$^+$ B cell progenitors (LIN$^-$IgM$^-$IgD$^-$CD19$^+$). Pseudo time analysis of SC RNA-seq data by using DDRTree over three components. Expression levels of the indicated genes are displayed with blue to red color indicating low to high levels in SCs.

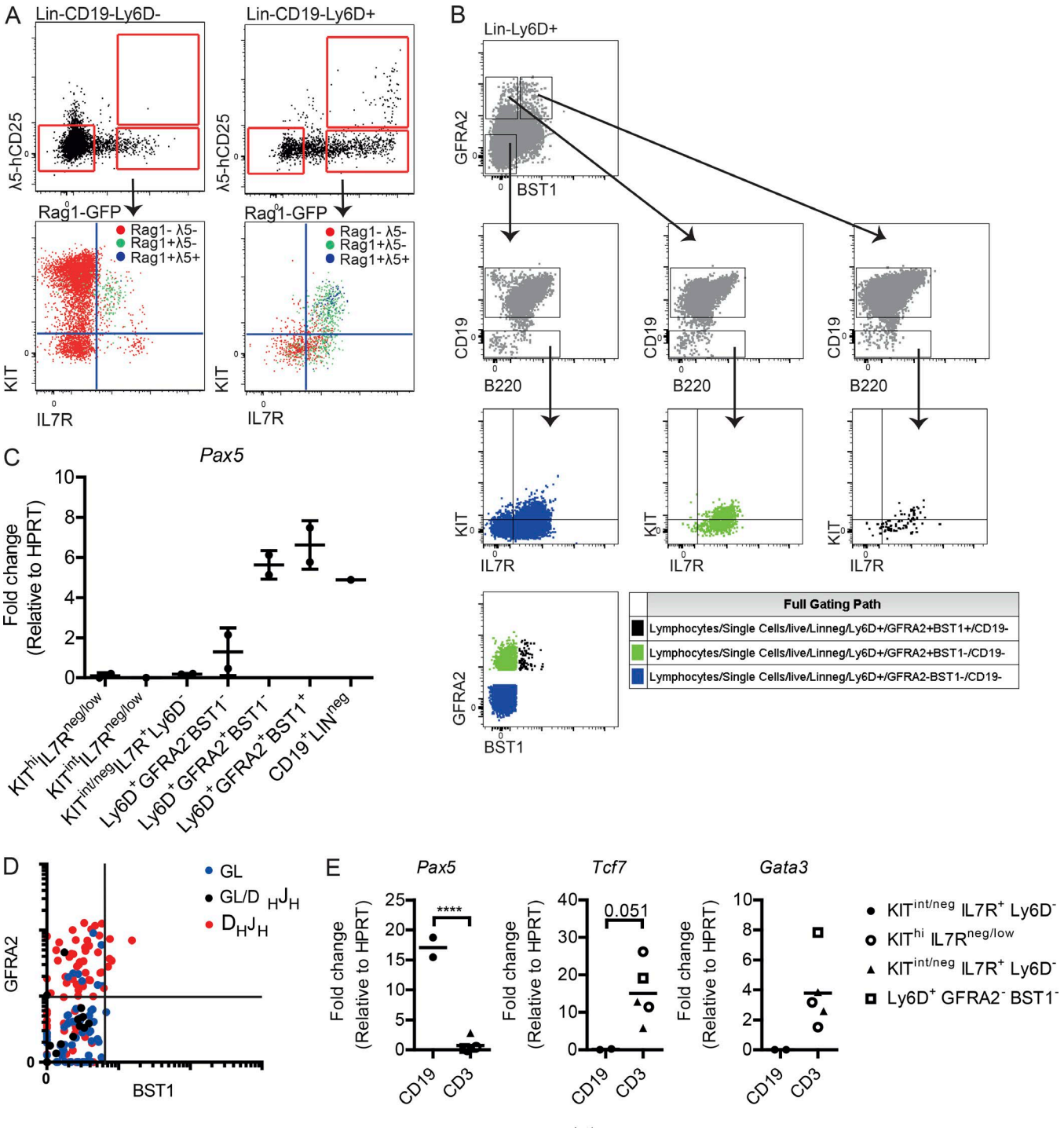

Figure S5. **GFRA2 and BST1 identify lineage-restricted cells in a complex progenitor compartment. (A)** Representative FACS plots demonstrating the expression of Rag1-GFP and Igll1 (l5-hCD25) reporter genes in LP cells. **(B)** Back gating of Lin$^-$Ly6D$^+$ GFRA2$^-$BST1$^-$, GFRA2$^+$BST1$^-$, and GFRA2$^+$BST1$^+$ cells. The CD19$^+$ population is indicated by gray. **(C)** Diagrams displaying qPCR data from LPs and CD19$^+$ cells sorted from BM. Figure indicates the expression of *Pax5* relative to *Hprt* in each population, and one dot represents an independent experiment analyzed by qPCR in triplicate. Samples are from two separate sorts. **(D)** A FACS plot generated from index-sorted LIN$^-$CD19$^-$IL7R$^+$Ly6D$^+$ cells (n = 105) after analysis of IgH DJ recombination events. Red indicates that only re-combined alleles were detected, black indicates that both germline and recombination were detected in the same cell, and blue indicates that only germline bands were detected. **(E)** Diagrams displaying qPCR data from donor-derived CD19$^+$ (n = 2) or CD3$^+$ (n = 5) cells sorted from the BM of transplanted mice. The four populations (Lin$^-$CD19$^-$KIT$^{int/neg}$IL7R$^+$Ly6D$^-$, Lin$^-$CD19$^-$KIT^{high}IL7R$^{low/neg}$, Lin$^-$CD19$^-$KIT$^{int/neg}$IL7R$^+$, Ly6D$^+$GFRA2$^-$BST1$^-$) indicate the identity of the transplanted progenitors. Plots show the relative expression to *Hprt* of *Pax5*, *Tcf7*, and *Gata3*. Each dot represents one mouse analyzed by qPCR in triplicate.

Tables S1, S2, S3, S4, and S5 are provided as Excel files.

Table S1 specifies the percentage of cells that stain positively with antibodies from the BD Lyoplate antibody screen with CLP LY6D$^-$, CLP LY6D$^+$, and CD19$^+$ cells.

Table S2 shows the statistics on differentially expressed genes in the SC-qPCR data sets belonging to Fig. 1. Differentially expressed genes were identified by using MAST for R, which builds a model that can account for the bimodality of gene expression in single-cell data by incorporating gene expression level (continuous) and binary expression (discrete), forming a unification in the hurdle score (Finak et al., 2015).

Table S3 shows the statistics on differentially expressed genes in the SC-qPCR data sets belonging to Fig. 4. Differentially expressed genes were identified by using MAST for R, which builds a model that can account for the bimodality of gene expression in single-cell data by incorporating gene expression level (continuous) and binary expression (discrete), forming a unification in the hurdle score (Finak et al., 2015).

Table S4 shows FACS antibodies used to identify LP populations in different figures in the manuscript.

Table S5 presents P values calculated for Fisher's exact test analysis determined for cloning frequencies and size based on seeded cells as indicated.

Table S6 presents P values calculated for Fisher's exact test analysis determined for cloning frequencies and cellular output based on seeded cells or generated clones as indicated.

Reference

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