Supplementary Information Guide

Blood and immune development in human fetal bone marrow and Down syndrome

Laura Jardine[1,2†], Simone Webb[1†], Issac Goh[1], Mariana Quiroga Londoño[3], Gary Reynolds[1], Michael Mather[1], Bayanne Olabi[1], Emily Stephenson[1], Rachel A. Botting[1], Dave Horsfall[1], Justin Engelbert[1], Daniel Maunder[1], Nicole Mende[3], Caitlin Murnane[4], Emma Dann[6], Jim McGrath[1], Hamish King[5], Iwo Kucinski[3], Rachel Queen[1], Christopher D Carey[7], Caroline Shrubsole[2], Elizabeth Poyner[1], Meghan Acres[1], Claire Jones[8], Thomas Ness[8], Rowen Coulthard[8], Natalina Elliott[4], Sorcha O'Byrne[4], Myriam L. R. Haltalli[3], John E Lawrence[6], Steven Lisgo[1], Petra Balogh[6], Kerstin B Meyer[6], Elena Prigmore[6], Kirsty Ambridge[6], Mika Sarkin Jain[6], Mirjana Efremova[10], Keir Pickard[2], Thomas Creasey[2,9], Jaume Bacardit[11], Deborah Henderson[1], Jonathan Coxhead[1], Andrew Filby[1], Rafiqul Hussain[1], David Dixon[1], David McDonald[1], Dorin Mirel-Popescu[1], Monika S. Kowalczyk[12], Bo Li[12], Orr Ashenberg[12,13], Marcin Tabaka[12], Danielle Dionne[12], Timothy L. Tickle[12,14], Michal Slyper[12], Orit Rozenblatt-Rosen[12], Aviv Regev[12], Sam Behjati[6,15], Elisa Laurenti[3], Nicola K. Wilson[3,+], Anindita Roy[4,16,+], Berthold Göttgens[3,+], Irene Roberts[4,16,+], Sarah A. Teichmann[6,17,+], Muzlifah Haniffa[1,6,18,+*]

[1]Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK
[2]Haematology Department, Freeman Hospital, Newcastle-upon-Tyne Hospitals NHS Foundation Trust, Newcastle-upon-Tyne, NE7 7DN, UK
[3]Department of Haematology, Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, CB2 0AW, UK
[4]Department of Paediatrics, University of Oxford, Oxford OX3 9DS, UK
[5]Centre for Immunobiology, Blizard Institute, Queen Mary University of London, London E1 2AT, UK
[6]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK [7]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK [8]NovoPath, Department of Pathology, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE2 4LP, UK
[9]Wolfson Childhood Cancer Centre, Northern Institute for Cancer Research, Newcastle University, Newcastle-upon-Tyne, UK.
[10]Barts Cancer Institute, Queen Mary University of London, London, UK
[11]School of Computing, Newcastle University, NE4 5TG, UK
[12]Klarman Cell Observatory, Broad Institute of Harvard and MIT, Cambridge, MA, USA [13]Genentech, South San Francisco, CA, USA
[14]Data Sciences Platform, Broad Institute of Harvard and MIT, Cambridge, MA, USA
[15]Department of Paediatrics, University of Cambridge, Cambridge CB2 0SP, UK
[16]MRC Molecular Haematology Unit and Department of Paediatrics, Weatherall Institute of Molecular Medicine, University of Oxford, and BRC Blood Theme, NIHR Oxford Biomedical Centre, Oxford OX3 9DS, UK [17]Theory of Condensed Matter Group, Cavendish Laboratory/Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK
[18]Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE2 4LP, UK
[†]Equal contribution
[+]Authors who jointly supervised
[*]Corresponding author; Correspondence to: m.a.haniffa@ncl.ac.uk.

Table of contents

Supplementary Methods 3

Supplementary Tables 3

Supplementary Methods

Additional methodological information for both wet-lab experimental protocols and bioinformatic analysis are provided in this Supplementary material.

Supplementary Tables

Supplementary Table 1: Single cell dataset metadata manifest.

Lane manifest of all novel scRNA-seq and CITE-seq data in this study, with complimentary metadata provided (including: sampleID, age, sex, sequencing technology, FACs sort gate, and single cell frequency pre- and post- QC). Specific accession codes used for public data availability are also noted. For completeness, Smart-seq2 FBM data are included as a lane, although they technically represent a plate-based experiment (for further metadata for SS2 data, please see relevant accessions and Supplementary Table 15).

Supplementary Table 2: Adult BM scRNA-seq annotation key.

Annotation key for mapping any broad cell states used in analysis to the 46 refined cell states in the adult BM scRNA-seq data.

Supplementary Table 3: CB scRNA-seq annotation key.

Annotation key for mapping any broad cell states used in analysis to the 26 refined cell states in the CB scRNA-seq data.

Supplementary Table 4: CB scRNA-seq metadata.

CB scRNA-seq cell type metadata by barcode.

Supplementary Table 5: Adult BM scRNA-seq metadata.

Adult BM scRNA-seq cell type metadata by barcode.

Supplementary Table 6: FBM cell type annotation markers.

Core annotation markers for identifying cell states in FBM scRNA-seq data with references to published literature.

Supplementary Table 7: FBM scRNA-seq annotation key.

Annotation key for mapping any broad cell states used in analysis to the 64 refined cell states in the FBM scRNA-seq data.

Supplementary Table 8: FBM scRNA-seq metadata.

FBM scRNA-seq cell type metadata by barcode.

Supplementary Table 9: Dimensional reduction coordinates.

Dimensional reduction coordinates by barcode for all UMAP/FDG embeddings shown in manuscript.

Supplementary Table 10: FBM DS and non-DS cell type population numbers.
Raw numbers for annotated cell types by sample in FBM (DS and non-DS) 10x scRNA-seq datasets. Raw numbers for all novel single cell datasets are noted in the Supplementary Methods 'Statistics and reproducibility' section.

Supplementary Table 11: FBM (total) CITE-seq metadata.
FBM (total) CITE-seq cell type metadata by barcode. This cell type metadata in this table is derived from RNA-based annotation of dataset, with barcodes filtered for those present in both RNA and protein matrices after both protein and RNA-based QC.

Supplementary Table 12: FBM (total) CITE-seq DEGs by cell type.
DE analysis was performed (using Scanpy) on all vs all cell types in FBM (total) CITE-seq dataset, with a two-sided Wilcoxon rank sum test and top 100 genes by z-score selected. Top 100 genes were then filtered to genes expressed in at least 25% of cells in either of the two populations compared, and with a natural log fold change cut-off of 0.25. From this subset, the genes were sorted by logfc and the top 30 genes for each cell type are shown in this table along with their p-values and log fold change values as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 13: FBM (total) CITE-seq DE proteins by cell type.
DE analysis for protein markers was performed (using Scanpy) on all vs all cell states in FBM (total) CITE-seq data with a two-sided Wilcoxon rank sum test and markers sorted by z-score. Resultant p-values and log fold change values for all proteins in the dataset are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 14: FBM (total) CITE-seq decision tree statistics.
Confusion matrix statistics from continuous decision tree constructed using the Rpart package to distinguish between the lineage-committed immune cell types in FBM (total) CITE-seq dataset using the 198 proteins present (Extended Data Figs. 2a-b). Overall accuracy was computed with a 95% CI using a binomial test and checked with a one-sided test (see caret package documentation for confusionMatrix function).

Supplementary Table 15: FBM Smart-seq2 scRNA-seq metadata.
FBM Smart-seq2 scRNA-seq cell type metadata by barcode.

Supplementary Table 16: CD34+ (FBM, FL, CB) CITE-seq metadata.
CD34+ (FBM, FL, CB) CITE-seq cell type metadata by barcode. This cell type metadata is derived from mRNA-based annotation of the CITE-seq dataset, and contains complete annotations from RNA dataset; the number of cells in each cell type shown here and detailed in manuscript methods 'statistics and reproducibility' may differ from any manuscript figures based on protein/ADT matrices (as they represent smaller population of cells that intersect with the post-QC protein matrices). Further information on post-QC and post-intersect numbers for CITE-seq protein and RNA lanes are provided in Supplementary Table 1.

Supplementary Table 17: CD34+ FBM/FL/CB CITE-seq DE proteins by cell type.
DE analysis for protein markers was performed (using Scanpy) on all progenitor cell states in CD34+ FBM/FL/CB CITE-seq data with a two-sided Wilcoxon rank sum test and markers sorted by z-score. Resultant p-values and log fold change values for all proteins in the dataset are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 18: Statistics for FBM scRNA-seq cell type frequency change by gestational stage.
FBM scRNA-seq dataset statistical significance of cell frequency change by gestational stage shown in parentheses. P-values resulting from quasibinomial regression model (subject to one-sided ANOVA; with correction for sort gates; p-values computed at 95% CI and adjusted for multiple testing using Bonferroni correction; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). Significant differences in proportion of broad cell groupings across gestational stages were assessed using a likelihood-ratio test computed using the anova function in R (stats v4.0.4). Increasing or decreasing trends were denoted with 'up' and 'down' arrows respectively. Directionality of trend between proportions and gestational age were accessed by sign of the Spearman's rho and Pearson's r computed between proportions and stage.

Supplementary Table 19: Sort-corrected cell type proportions for FL and FBM scRNA-seq data.
FACs sort count information and sort-corrected cell type proportions derived from parent, and daughter gate contributions for FL and FBM scRNA-seq data. Sort proportions are represented as a fraction of sorted counts contributed by each daughter gate (CD45+/-) to the total number of events in the live/dead parent gate. Corrected scRNA-seq counts are represented as a corrected proportion of each cell type contribution by sort proportion for each scRNA-seq lane.

Supplementary Table 20: DS FBM scRNA-seq annotation key.
Annotation key for mapping any broad cell states used in analysis to the 40 refined cell states in the DS FBM scRNA-seq data.

Supplementary Table 21: DEGs between equivalent cell states in non-DS and DS FBM scRNA-seq datasets.
DEGs between non-DS and DS FBM scRNA-seq datasets were calculated for each cell type using the FindMarkers tool in Seurat which is based on two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method. Filtering was performed to exclude genes that were not expressed in at least 25% of cells in one condition and to exclude genes that did not have at least a 0.25 log fold change. The final list was filtered for significance for genes with an adjusted p-value<0.05.

Supplementary Table 22: Statistics for myeloid cell type frequency change by tissue.
Statistical significance of monocyte, neutrophil and DC lineage cell state frequencies across tissues from scRNA-seq datasets. P-values resulting from quasibinomial regression model (subject to one-sided ANOVA; with correction for sort gates; p-values computed at 95% CI and adjusted for multiple testing using Bonferroni correction; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). Significant differences in proportion of monocyte, neutrophil and DC lineage

cells across tissues were assessed using a likelihood-ratio test computed using the anova function in R (stats v4.0.4). Proportions of total monocyte, neutrophil and DC lineage cells out of total haematopoietic cells across tissue were tested using a one-way ANOVA with Tukey's multiple comparison correction in GraphPad (v.8.1.2, GraphPad Software).

Supplementary Table 23: Statistics for lymphoid cell type frequency change by tissue.
Statistical significance of B-lineage cell state frequency across tissues from scRNA-seq datasets. P-values resulting from quasibinomial regression model (subject to one-sided ANOVA; with correction for sort gates; p-values computed at 95% CI and adjusted for multiple testing using Bonferroni correction; *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). Significant differences in proportion of B-lineage cells across tissues were assessed using a likelihood-ratio test computed using the anova function in R (stats v4.0.4). Proportions of total B-lineage cells out of total haematopoietic cells across tissue were tested using a one-way ANOVA with Tukey's multiple comparison correction in GraphPad (v.8.1.2, GraphPad Software).

Supplementary Table 24: DEGs between sinusoidal and tip ECs in FBM scRNA-seq dataset.
DE analysis was performed (using Scanpy) on sinusoidal vs tip endothelial cells in FBM scRNA-seq dataset, with a two-sided Wilcoxon rank sum test and top 100 genes by z-score selected. Top 100 genes were then filtered to genes expressed in at least 25% of cells in either of the two populations compared, and with a natural log fold change cut-off of 0.25. Resultant p-values and log fold change values are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 25: DEGs between equivalent HSC/MPPs and endothelial cells in the DS and non-DS FBM scRNA-seq datasets.
Table containing DEG analysis result for analysis (using Scanpy) between HSC/MPP and endothelial cells across DS and non-DS FBM scRNA-seq datasets. Genes tested are HSC/MPP receptors and endothelial ligands predicted by CellPhoneDB to significantly interact in the niche of endothelial cells (in the arteriolar niche) in non-DS FBM (see Extended Data Fig. 10a-b). Resultant p-values and log fold change values are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 26: Antibodies for flow cytometry and immunofluorescence microscopy.
Lists antibodies used for cell sorting/analysis: i) enrichment sort of scRNA-seq, ii) progenitor sort for single cell cultures, iii) analysis of progenitor cultures and antibodies used for immunohistochemistry.

Supplementary Table 27: Antibodies in the CITE-seq panel
Lists antibodies used for CITE-seq panel.

Supplementary Table 28: Counts and coordinates for HSC/MPP and progenitors in immunofluorescence images
Counts and coordinates per ROI of HSC/MPP and progenitors derived from IF imaging of FBM. HSC/MPP and progenitors were manually annotated with QuPath v.0.2.3 within unbiased assignment of metaphyseal versus diaphyseal regions (assigned as metaphyseal if <1.3mm and

diaphyseal if >1.3mm from the epiphyseal cartilage) in four fetal bone marrow samples of similar gestational age (14-15 PCW). Relative cellular density within each ROI was accounted for by dividing manually counted HSC/MPP and progenitors by the total number of segmented cells in each ROI. We find that there is no significant preference for relative HSC/MPP and progenitor frequency between either metaphyseal or diaphyseal compartments by a one-way ANOVA (with the anova function in R using the stats package v4.0.4).

Supplementary Table 29: Sequencing information for novel single cell datasets in this study.
Sequencing type, software and reference genome used for alignment for all novel single cell datasets used in this study.

Supplementary Table 30: DEGs between promyelocytes and promonocytes in the FBM scRNA-seq dataset.
DE analysis was performed (using Scanpy) on promyelocyte vs promonocyte cells in FBM scRNA-seq dataset, with a two-sided Wilcoxon rank sum test and top 100 genes by z-score selected. Top 100 genes were then filtered to genes expressed in at least 25% of cells in either of the two populations compared, and with a natural log fold change cut-off of 0.25. Resultant p-values and log fold change values are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 31: Gene mutations implicated in severe congenital neutropenia
List of gene mutations implicated in severe congenital neutropenia, with maturation defect, AML/DS risk and literature reference. Extended Data Fig. 3e.

Supplementary Table 32: Monocle3 pseudotime values.
Pseudotime values by barcode for all Monocle3 analysis shown in the manuscript.

Supplementary Table 33: Monocle3 output for FBM neutrophil pseudotime analysis.
Monocle3 output table, displaying differentially expressed genes across FBM neutrophil pseudotime development trajectory. DE genes are calculated using a one-sided Moran's I statistical test, with test statistic and p-value for each gene shown in the table.

Supplementary Table 34: DEGs between the cycling and non-cycling Pre B progenitors in FBM scRNA-seq dataset.
DE analysis was performed (using Scanpy) on cycling vs non-cycling Pre B progenitors in FBM scRNA-seq dataset, with a two-sided Wilcoxon rank sum test and top 100 genes by z-score selected. Top 100 genes were then filtered to genes expressed in at least 25% of cells in either of the two populations compared, and with a natural log fold change cut-off of 0.25. Resultant p-values and log fold change values are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 35: Monocle3 output for FBM B-lineage pseudotime analysis.
Monocle3 output table, displaying differentially expressed genes across FBM B-cell development trajectory. DE genes are calculated using a one-sided Moran's I statistical test, with test statistic and p-value for each gene shown in the table.

Supplementary Table 36: Gene mutations implicated in B-ALL.
Lists details of genes implicated in B-ALL, their functions and provides a literature reference. Relates to analysis in Extended Data Fig. 4h.

Supplementary Table 37: Monocle3 output for adult BM B-lineage pseudotime analysis.
Monocle3 output table, displaying differentially expressed genes across adult BM B cell development trajectory. DE genes are calculated using a one-sided Moran's I statistical test, with test statistic and p value for each gene shown in the table.

Supplementary Table 38: Statistics for cell type frequency change in CD34+ (FBM, FL, CB) CITE-seq data by tissue.
Statistical significance of cell type proportion across tissue in CD34+ (FBM, FL, CB) CITE-seq data shown in parentheses. P-values resulting from quasibinomial regression model (subject to one-sided ANOVA; with correction for sort gates; p-values computed at 95% CI and adjusted for multiple testing using Bonferroni correction; *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). Significant differences in proportion of cells across tissue were assessed using a likelihood-ratio test computed using the anova function in R (stats v4.0.4). Includes information for FACs sort information and associated corrected proportions.

Supplementary Table 39: DE proteins between HSC/MPP in the CD34+ FBM/FL/CB CITE-seq data across tissue.
DE analysis for protein markers was performed by tissue (using Scanpy) on combined HSC/MPP (cell states HSC/MPP I-IV combined) in the CD34+ FBM/FL/CB CITE-seq data with a two-sided Wilcoxon rank sum test and markers sorted by z-score. Resultant p-values and log fold change values for all proteins in the dataset are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 40: DEGs between HSC/MPP 1 in the CD34+ FBM/FL/CB CITE-seq data across tissue.
DE analysis was performed by tissue (using Scanpy) on HSC/MPP 1 cells between FL, FBM and CB in the CD34+ CITE-seq dataset, with a two-sided Wilcoxon rank sum test and genes sorted by z-score. Resultant p-values and log fold change values for all genes in the dataset are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 41: Monocle3 output for non-DS FBM erythroid pseudotime analysis.
Monocle3 output table, displaying differentially expressed genes across non-DS FBM erythroid development trajectory. DE genes are calculated using a one-sided Moran's I statistical test, with test statistic and p-value for each gene shown in the table.

Supplementary Table 42: Monocle3 output for DS FBM erythroid pseudotime analysis.
Monocle3 output table, displaying differentially expressed genes across DS FBM erythroid development trajectory. DE genes are calculated using a one-sided Moran's I statistical test, with test statistic and p-value for each gene shown in the table.

Supplementary Table 43: Enrichment of inflammatory and cytokine production pathway genes across equivalent cell states in non-DS and DS FBM scRNA-seq datasets.

Table containing fold changes for genes associated with inflammatory and cytokine production pathways enriched in DE genes computed between analogous cell state compartments in non-DS and DS FBM scRNA-seq datasets (input DE genes provided in Supplementary Table 21; DEGs defined by a two-sided Wilcoxon rank-sum test with Benjamini-Hochberg procedure for multiple testing correction). See methods for 'TNF response gene annotation, TNF superfamily interactions and TNFα-signalling pathway enrichment'.

Supplementary Table 44: Lineage-wide CellPhoneDB-inferred putative receptor:ligand interactions in combined DS and non-DS FBM scRNA-seq datasets.

Putative receptor:ligand interactions as inferred by CellphoneDB - table of significant means. The non-DS FBM scRNA-seq dataset was subsampled to 10% of total cells and combined with the DS data. The statistical_method option in CellphoneDB was used with a p-value cut-off of 0.05 for significant receptor ligand pairs and a result precision of 3dp. Empty cells represent a value of '1'. Interactions were run for all vs all equivalent cell states between non-DS and DS scRNA-seq datasets, and results are displayed in Extended Data Fig. 7j.

Supplementary Table 45: Statistics for stromal cell type frequency change in FBM scRNA-seq data by gestational stage.

Statistical significance of stromal cell type proportion across gestational stage in FBM scRNA-seq data. P-values resulting from quasibinomial regression model (subject to one-sided ANOVA; with correction for sort gates; p-values computed at 95% CI and adjusted for multiple testing using Bonferroni correction; *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001). Significant differences in proportion of stromal cells across gestational age were assessed using a likelihood-ratio test computed using the anova function in R (stats v4.0.4). Increasing or decreasing trends were denoted with 'up' and 'down' arrows respectively. Directionality of trend between proportions and gestational age were accessed by sign of the Spearman's rho and Pearson's r computed between proportions and stage.

Supplementary Table 46: i) DEGs between sinusoidal ECs identified in FBM and FL scRNA-seq datasets, and ii) DE proteins between FBM vs FL sinusoidal ECs that were annotated in the CD34+ CITE-seq dataset.

Left: DE gene analysis was performed (using Scanpy) on sinusoidal ECs in FBM scRNA-seq dataset vs their equivalent that we re-annotated (see Supplementary Methods for details) from the FL dataset (Popescu et al, 2019). DEGs were run with a two-sided Wilcoxon rank sum test and for genes with known roles in adhesion, extracellular matrix and angiopoiesis. Resultant p-values and log fold change values are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method. Right: DE protein analysis was performed (using Scanpy) on FBM vs FL sinusoidal ECs that were annotated in the CD34+ CITE-seq dataset based on logistic regression projection with the FBM 10x dataset as reference. DEGs were run with a two-sided Wilcoxon rank sum test and for all proteins present in the dataset. Resultant p-values and log fold change values are shown in this table, as output from two-sided Wilcoxon rank-sum test adjusted for multiple testing using the Benjamini-Hochberg method.

Supplementary Table 47: CD34+ (FBM, FL, CB) CITE-seq sinusoidal endothelial cell metadata.
CD34+ (FBM, FL, CB) CITE-seq sinusoidal endothelial cell type metadata by barcode. Sinusoidal EC were captured in this dataset by virtue of also being CD34+. The cell type metadata in this table is derived from mRNA-based annotation of the CITE-seq dataset, and contains complete annotations from RNA dataset; the number of cells in each cell type shown here and detailed in Supplementary Methods 'Statistics and Reproducibility' may differ from manuscript figures based on protein/ADT matrices (as they represent smaller population of cells that intersect with the post-QC protein matrices). Further information on post-QC and post-intersect numbers for CITE-seq protein and RNA lanes are provided in Supplementary Table 1.

Supplementary Table 48: Population number for CD34, CD38, CD117 and CD71 -negative/-positive single cells in the FBM (total) CITE-seq dataset.
Population number for CD34, CD38, CD117 and CD71 -negative/-positive single cells in the FBM (total) CITE-seq dataset. To derive proportions, protein counts were ln(DSB-normalised and scaled), with scaling performed to a lower limit of 0. Positive and negative expression thresholds were defined as 2 and 1 for CD34/CD117/CD71 and CD38 protein expression, respectively.

Supplementary Table 49: CellPhoneDB output from FBM scRNA-seq stromal vs progenitor experiment.
CellPhoneDB output from FBM scRNA-seq stromal vs progenitor experiment (subsetted to cell type interactions displayed in Extended Data Fig. 10a). The statistical_method option in CellphoneDB was used with a p-value cut-off of 0.05 for significant receptor ligand pairs and a result precision of 3dp. The p-values results file ('pvalues.txt' file output from running CellPhoneDB with statistical method as described in Methods) reports statistical significance of predicted R:L interaction in a given interacting R:L pair in given cell types. The p-values are a result of the CellPhoneDB statistical test, which i) randomly permutes cell type label of all cells to create a null distribution of the mean of the average R:L expression, ii) tests proportion of mean R:L expression higher than actual mean R:L expression to derive p-value for likelihood of a R:Ls cell type specificity.

Supplementary Table 50: Genes implicated in inherited disorders of blood and immune cells.
A list of genes involved in inherited disorders of blood and immune cells that is integrated into our interactive web portal.

Supplementary Table 51: Fetal yolk sac scRNA-seq metadata.
Fetal yolk sac scRNA-seq cell type metadata by barcode. The metadata presented here are a result of sub-clustering the fetal yolk sac scRNA-seq dataset presented in Popescu et al, Nature, 2019.

Supplementary Table 52: Fetal liver scRNA-seq metadata.
Fetal liver scRNA-seq cell type metadata by barcode. The metadata presented here are a result of sub-clustering the fetal liver scRNA-seq dataset from Popescu et al, Nature, 2019.
Specific cell types sub-clustered are described in Supplementary Methods 'Statistics and Reproducibility'.

Supplementary Table 53: DS FBM scRNA-seq metadata.
DS FBM scRNA-seq cell type metadata by barcode.