Intrinsic and extrinsic regulation of human fetal bone marrow haematopoiesis and perturbations in Down syndrome

Laura Jardine[1,2†], Simone Webb[1†], Issac Goh[1], Gary Reynolds[1], Michael Mather[1], Bayanne Olabi[1], Mariana Quiroga Londoño[3], Daniel Maunder[1], Emily Stephenson[1], Justin Engelbert[1], Rachel A. Botting[1], Nicole Mende[3], Caitlin Murnane[4], Dave Horsfall[1], Jim McGrath[1], Hamish King[5], Emma Dann[6], Iwo Kucinski[3], Rachel Queen[1], Christopher D Carey[7], Caroline Shrubsole[2], Meghan Acres[1], Claire Jones[8], Thomas Ness[8], Natalina Elliott[4], Sorcha O'Byrne[4], Steven Lisgo[1], Petra Balogh[6], Kerstin 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], Marcin Tabaka[12], Danielle Dionne[12], Timothy L. Tickle[12,13], Michal Slyper[12], Orit Rozenblatt-Rosen[12], Aviv Regev[12], Sam Behjati[6,14], Elisa Laurenti[3], Nicola K. Wilson[3,+], Anindita Roy[4,15,+], Berthold Göttgens[3,+*], Irene Roberts[4,15,+*], Sarah A. Teichmann[6,16,+*], Muzlifah Haniffa[1,6,17,+*]

Affiliations

[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]Data Sciences Platform, Broad Institute of Harvard and MIT, Cambridge, MA, USA

[14]Department of Paediatrics, University of Cambridge, Cambridge CB2 0SP, UK

[15]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

[16]Theory of Condensed Matter Group, Cavendish Laboratory/Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK

[17]Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE2 4LP, UK

[†]Equal contribution

[+]Co-senior

*Corresponding authors

Keywords: human development, haematopoiesis, immunology, single cell RNA-sequencing, bone marrow, Down syndrome, trisomy 21

Summary

Throughout postnatal life, haematopoiesis in the bone marrow (BM) maintains blood and immune cell production. Haematopoiesis first emerges in human BM at 11-12 post conception weeks while fetal liver (FL) haematopoiesis is still expanding. Yet, almost nothing is known about how fetal BM evolves to meet the highly specialised needs of the fetus and newborn infant. Here, we detail the development of fetal BM, including stroma, using multi-omic assessment of mRNA and multiplexed protein epitope expression. We find that the full blood and immune cell repertoire is established in fetal BM in a short time window of 6-7 weeks early in the second trimester. Fetal BM promotes rapid and extensive diversification of myeloid cells, with granulocytes, eosinophils and dendritic cell subsets emerging for the first time. B-lymphocyte expansion occurs, in contrast with erythroid predominance in FL at the same gestational age. We identify transcriptional and functional differences that underlie tissue-specific identity and cellular diversification in fetal BM and FL. Finally, we reveal selective disruption of B-lymphocyte, erythroid and myeloid development due to cell intrinsic differentiation bias as well as extrinsic regulation through an altered microenvironment in the fetal BM from constitutional chromosome anomaly Down syndrome during this crucial developmental time window.

Introduction

Following waves of haematopoiesis in the yolk sac (YS) and fetal liver (FL), the bone marrow (BM) is established as the site of lifelong blood and immune cell production. In humans, haematopoietic cells appear in the fetal BM from 11-12 post conception weeks (PCW)[1,2]. By this time, development of the immune repertoire has been initiated in the liver, with contributions from spleen and thymus[3,4]. The contribution of the BM, both in terms of fetal haematopoiesis and in laying foundations for BM haematopoiesis in the longer term remains to be established.

Haematopoiesis must supply the fetus with erythrocytes for oxygen transport, platelets for haemostasis, macrophages for tissue remodelling and an immune system that is poised to respond to insult without risking tissue damage. In fetal life, the priorities are to establish an innate immune system primed to respond to any pathogen encountered in perinatal life, and to provide the pool of B- and T- lymphocyte progenitors needed to prescribe the adaptive immune response to pathogens after birth, particularly in the first few weeks of life when the newborn infant is first exposed to the extra-uterine environment.

Perturbations in haematopoiesis *in utero* can have far reaching implications. Oncogenic events acquired in haematopoietic cells during fetal life, such as *ETV6-RUNX1* or *KMT2A* fusion genes, often progress to childhood leukaemia[5–8]. Somatic, fetally-acquired *GATA1* mutations leading to neonatal pre-leukaemia and increased risk of myeloid leukaemia in early childhood are specific to Down syndrome (DS)[9,10]. Extensive abnormalities in haematopoiesis have been documented in DS FL preceding *GATA1* mutations[11] as well as in DS newborns without *GATA1* mutations[12]. The prevalence of immune deficiency, autoimmune disease and childhood leukaemias are increased in children with DS[13–15]. This makes understanding the

development of the immune system of clinical importance in DS and yet almost nothing is known of how this occurs in fetal life.

Longer term haematopoiesis depends on a finite pool of haematopoietic stem cells (HSCs), supported by their niche. In mice, BM development begins once vascularization allows HSCs to enter collagenous bone[16]. Stromal cells that support HSC development in humans have been identified but no systematic examination of BM development has been achieved to date. The niche is accepted to be critical to HSC function and niche abnormalities are implicated in both malignant and nonmalignant disorders[17].

In this study, we use single-cell multi-omics to dissect the composition of disomic and trisomy 21 human fetal BM, as haematopoiesis emerges and develops during the early second trimester. For disomic fetal BM, FL and cord blood (CB) $CD34^+$ progenitors, we generate combined surface epitope and transcriptome profiles (including a panel of 198 antibodies), to allow reliable comparison of tissue-specific differentiation landscapes. We predict receptor-ligand interactions to comprehend the non-haematopoietic support required for differentiation. We validate: i) newly emerging cell states in fetal BM by FACS-based prospective isolation for single cell RNA sequencing (scRNA-seq) and morphology assessment; ii) BM niche by multiplex immunofluorescence imaging, and; iii) HSC differentiation potential using single cell clonogenic differentiation assays. Drawing upon existing scRNA-seq data from YS, FL, CB and adult BM, we show for the first time in humans how life-long definitive haematopoiesis is established in the BM early in fetal life and the unique characteristics that contribute towards the assembly of a complex multilineage blood and immune system within a matter of weeks.

Results

A single cell atlas of human fetal bone marrow

To characterize both fetal BM haematopoiesis and the environment that supports it, we obtained single cell suspensions from mechanical disruption of fetal femur. Suspensions were; i) labeled with CD45-fluorochrome antibody, FACS-enriched for $CD45^{+}$/- cells and processed with either 3-prime or 5-prime reagents to yield mRNA and TCR/BCR data, or; ii) labelled with CD34-fluorochrome antibody plus ADT cocktail and FACS-enriched for live cells or $CD34^{+}$ to generate CITE-seq (total) or CITE-seq ($CD34^{+}$) data (10x Genomics) (**Fig. 1A; Supplementary Table 1**). FL and CB $CD34^{+}$ CITE-seq data were generated in parallel. Reference scRNA-seq data sets on YS and FL[4], CB and adult BM (https://data.humancellatlas.org/) were assembled to allow comparisons across haematopoietic development (**Extended Data Figs. 1A-D).**

Following removal of low quality cells and doublets, we used Harmony to integrate data across batches (see Methods). Data were iteratively partitioned by graph-based leiden clustering. Differentially expressed genes (DEGs) were calculated between clusters, and those with unique transcriptional signatures assigned as cell states by reference to expression profiles reported in the literature **Supplementary Table 2**). To ensure recognition of unique cell states in different tissues, datasets were annotated independently, facilitated by label-transfer, and logistic regression was used to confirm consistency of annotations. An exception was the $CD34^{+}$ CITE-seq data in FL, fetal BM and CB, in which cross-tissue data were integrated for clustering and annotation.

Our fetal BM scRNA-seq dataset sampled 9 karyotypically normal fetuses from the time of macroscopically detectable haematopoiesis (12 PCW) until established second trimester

haematopoiesis (18-19 PCW). From 122,584 cells, of which 103,228 passed quality control (QC), 64 transcriptionally distinct cell states were identified and manually grouped into 10 broad cellular compartments (**Figs. 1B-D, Extended Data Fig. 1E; Supplementary Tables 3-6**). During this time window, the ratio of haematopoietic to stromal cells expanded rapidly, from 3.4:1 to 12.8:1 (**Figs. 1B, D**). All 10 broad cellular compartments were present by 12 PCW, which could be accounted for by influx of committed FL progenitors, alongside FL HSCs, and rapid differentiation in their new environment. B-lymphopoiesis progressively expanded over gestation while proportion of myeloid cells remained similar (**Fig. 1D, Supplementary Table 7**). Fewer broad cellular compartments were observed in CB and adult BM, with eosinophils, basophils and mast cells absent in both, and stromal cells absent in CB (**Extended Data Figs. 1B, C; Supplementary Tables 8-11**). Our fetal BM CITE-seq (total) data were drawn from 3 further BM suspensions (16-17 PCW, k=8,978) (**Supplementary Table 12, Extended Data Fig. 2A**). In CITE-seq data, we observed a strong correlation between gene expression and equivalent surface protein expression across the broad cell categories. Our fetal BM 10x dataset can be explored using an interactive web portal: (https://developmentcellatlas.ncl.ac.uk/fbm_index; username: fbm_portal_access; password: JxsaCj3zHZu38oqh), which integrates a database for genes implicated in inherited disorders of blood and immune cells.

A panel of population-defining DEGs expressed as surface antigens in fetal BM 10x were selected to prospectively isolate cell states for validation (**Extended Data Figs. 2B-D**). Emphasis was placed on validating the myeloid innate immune cells not detected in age-matched FL (neutrophils and their precursors, eosinophils, basophils)[4]. Cells were FACS-isolated into plates for Smart-seq2 (SS2) scRNA-seq and cytospins were prepared for morphological assessment. There was good correlation between 10x and SS2 transcriptional

states (**Fig. 1E, Extended Data Fig. 2D; Supplementary Tables 5, 13**) and morphology-based identity assignment (**Fig. 1F, Supplementary Table 2**). As the sorting strategy did not effectively discriminate promonocytes from monocytes, we calculated differentially expressed proteins between mRNA-defined clusters in our CITE-seq (total) data to generate candidate proteins markers for developing monocyte populations (**Supplementary Table 14)**. In addition to expected candidates (HLA-DR, CD14, CD11b), differences in expression of some novel monocyte antigens (CD35, CD123, CD49d) were seen (**Fig. 1G**). For developing neutrophils, CD112, CD99 and CD1d emerged as new discriminatory markers.

Diversification of innate myeloid and lymphoid cells

Clustering and DEG analysis of fetal BM 10x myeloid cells identified 18 monocyte, dendritic cell (DC), neutrophil and macrophage cell states spanning from committed precursors to terminally-differentiated cells (**Fig. 2A, Extended Data Figs. 3A, B, Supplementary Table 5**). Compared with YS and FL, these lineages were significantly expanded in BM (**Fig. 2B**). Neutrophils and their precursors, and additional DC cell states (pDC, tDC and DC3) were detected in BM, but not in earlier sites of haematopoiesis (**Figs. 1F, 2A, B, Supplementary Table 15**). Fetal BM and adult BM were comparable in terms of monocyte, DC and neutrophil diversity, however $CD16^{+}$ monocytes (expressing *FCGR3A*, *HES4*, *CSF1R*) and monocyte-DC populations were only observed after birth.

To understand how myeloid development is differentially programmed in FL and fetal BM, we sought to explore the origins of neutrophil commitment. By force-directed graph (FDG), a granulocyte monocyte progenitor (GMP)-like population bridged uncommitted BM progenitor states and the earliest neutrophilic cells (promyelocytes) (**Extended Data Fig.**

3C). Between GMPs and the earliest monocytes (promonocytes), we identified cells fitting the transcriptional profile of committed monocyte progenitors reported in CB and adult BM[18] (MOP). This prompted us to re-analyse our previously published FL scRNA-seq data, upon which we identified the GMPs in the HSC/MPP group and MOPs and a small number of promyelocytes within the 'neutrophil-myeloid progenitor' group (**Supplementary Table 16**). Profiling single cells along the continuum of BM myeloid progenitors, divergence of monocyte and neutrophil signatures was observed at the GMP stage, consistent with understanding derived from mouse BM studies[19,20] (**Extended Data Fig. 3D, Supplementary Table 17**). As the balance of transcription factors CEPBα (*CEBPA*) and PU.1 (*SPI1*) is critical for neutrophil versus monocyte specification[21], we examined the relative expression of *CEBPA* and *SPI1* in FL versus fetal BM GMPs (**Fig. 2C**). This revealed that GMPs often expressed either *CEBPA* or *SPI1*, which is (i) consistent with the notion that fate choices are often made upstream of the classically defined GMP compartment [19,20], and (ii) that high, unopposed expression of *CEBPA* in fetal BM may account for its propensity to make neutrophils.

To leverage temporal information captured by fetal BM samples in the 12-19 PCW window, we used the Waddington-OT[22] algorithm to visualise the probability of neutrophil-monocyte-lineage cells transitioning into: CD14 monocyte, neutrophil, or other (upstream neutrophil-monocyte-lineage) fate at each week of development. The fate matrix revealed a high probability of terminal differentiation into CD14 monocyte fate at 12 PCW and an increasing proportion of cells approximating neutrophil fate later in the developmental window (**Extended Data Fig. 3E**). We then inferred the neutrophil lineage trajectory using Monocle3[23] and identified genes dynamically regulated over pseudotime (**Supplementary Tables 18, 19**). We used this to explore the temporal expression of genes proven critical to

neutrophil development by their implication in congenital neutropenias- inherited disorders of low neutrophil number leading to infection susceptibility, recombinant growth-factor-dependence and, variably, BM failure, myelodysplastic syndrome (MDS) and acute myeloid leukaemia (AML). Genes implicated in congenital neutropenia showed expression peaks spanning pseudotime (**Fig. 2D, Supplementary Table 20**). Genes associated with MDS/AML-risk neutropenias (*SBDS, HAX1, G6PC3*) tended to be expressed in HSC/MPPs, whereas genes causing congenital neutropenias without recognized MDS/AML-risk (*AP3B1, CXCR4*) were expressed in the terminal stages of differentiation (**Fig. 2D**), underscoring the importance of progenitor dysregulation in leukaemia risk.

We compared the transcriptome profile of DCs in fetal BM, YS, FL and adult BM with the well-characterised DC subsets in adult blood[24] (**Extended Data Fig. 3F**). The transcriptional signatures of DC1 and pDCs (DC6) were highly conserved across haematopoietic development and tissue, whereas DC2 and DC3 signatures were less conserved. To explore the origins of fetal BM DC diversity, we visualized terminal DC states and their putative precursors by FDG (**Fig. 2E, Supplementary Table 5**). This revealed tDC (DC5) as a transcriptional state intermediate between DC2 and pDC. FDG suggested a dual origin of pDCs, from myeloid and lymphoid precursors, as previously reported in FL[4]. Using iRegulon we found similar TFs were predicted to drive terminal differentiation of fetal BM pDC and tDC (**Extended Data Fig. 3G**).

We also identified innate lymphocytes, NK cells, NKT cells and ILC precursors in fetal BM. NK progenitors expressed lymphoid lineage progenitor genes (*SOX4* and *TOX*) (**Extended Data Fig. 3H**) and emerged from HSC via the lymphoid primed multipotent progenitor (LMPP), as inferred by FDG analysis (**Fig. 2F, Supplementary Table 5**). NK cells and ILCs

have been identified by scRNA-seq in FL and fetal skin prior to 12 PCW, suggesting that peripheral tissues are seeded prior to BM NK and ILC precursor production[4]. Fetal BM NK cells were enriched for NK cytotoxicity genes, relative to YS and FL. This may indicate that fetal BM permits more extensive maturation of NK cells, or that early fetal development is less dependent on effectors of cytotoxicity (**Fig. 2G**).

Establishment of the adaptive immune repertoire

The post-natal and adult mammalian BM is recognized as a critical site for B cell differentiation, positive and negative selection[25]. However, the extent to which fetal BM contributes to the prenatal B cell repertoire is not fully understood[26]. We delineated 5 distinct cell states in the fetal BM B cell lineage and observed two bursts of proliferative activity, at the Pre pro-B progenitor and the Pre-B precursor stages (**Figs. 3A, B, Extended Data Fig. 4A, Supplementary Table 5**). Leveraging corresponding BCR-enriched VDJ data from 2 fetal BM samples at 15 PCW, we detected productive heavy chain rearrangement from the Pre-B precursor stage and productive heavy+light chain from the Immature B cell stage (**Fig. 3B**). The emerging B cell repertoire was diverse, with the small number of shared clonotypes detected predominantly between Pre B progenitors (25 clonotypes shared) and Pro B progenitors (10 clonotypes shared) (**Extended Data Fig. 4B**). The proportion of B-lineage cells per mononuclear cell in fetal BM was 10-fold higher than in adult BM, and fetal BM B lineage composition was markedly skewed towards the earlier cell states (**Fig. 3C, Supplementary Table 21**).

Differentiation trajectories reconstructed using Monocle3 revealed linear differentiation from HSC to Pro B progenitors with a subsequent branch-point at the Pre-B progenitor stage (**Extended Data Fig. 4C**). One path of Pre-B progenitor ('Cell cycling') significantly

upregulated proliferative genes including *MKI67*, *TOP2A* and *CDK1*. The other path of Pre-B progenitors ('B cell differentiation') differentiated into immature and naive B cells, significantly expressing TFs and proteins characteristic of B cell maturation e.g. *IGHM*, *CXCR4* and *SOX4* **(Fig. 3D, Supplementary Table 22)**. Apoptosis gene expression was relatively minimal in the 'Cell cycling' path, suggesting that cells in this state were capable of returning to the path of progressive differentiation (**Extended Data Fig. 4D**). Apoptosis genes were most enriched in the non-cycling Pro-B and Pre-B cell stages (**Extended Data Fig. 4D**), in keeping with programmed death of cells failing successful heavy chain recombination and heavy chain integration into the Pre-B receptor, respectively.

Genes differentially regulated during B lineage pseudotemporal development were calculated with Monocle3 and used as a foundation for exploring the normal expression patterns of genes implicated in childhood B-ALL (**Fig. 3E, Extended Data Fig. 4C, Supplementary Tables 18, 23, 24**). B-ALL most commonly presents in the first five years of life and it is known that leukaemia-initiating events can occur *in utero*[8,27,28]. Small deletions and translocations in a limited set of genes are well-characterized in B-ALL[29]. These genes were highly expressed in fetal BM B cell states, especially in the Pre-B progenitor and earlier stages (**Fig. 3E**), while more limited expression was seen in adult BM B cell states (**Extended Data Fig. 4E, Supplementary Table 25**). As proliferation status and chromatin accessibility contribute to mutagenesis risk[30], it is plausible that the dependence of proliferating B progenitors on these genes exposes a substrate for mutagenesis that is unique to haematopoiesis during early life.

While the adult BM is a considerable reservoir of both naive and memory T lymphocytes (**Extended Data Fig. 1C**), the fetal BM contains few T lymphocytes (**Fig. 1B**). Thymic

lymphopoiesis is established several weeks before the fetal BM is colonized[3], and in keeping with this, we found CD4, CD8 and T regulatory cells expressing genes characteristic of the post thymic state (**Fig. 3F**). TCR-enriched VDJ data of 2 fetal BM samples at 15 PCW confirmed that these single positive T cell subsets contained productive TRA and TRB transcripts (**Extended Data Fig. 4F**). In FL, early lymphocytes were transcriptionally similar to the early thymocyte precursor (ETP) in the thymus, supporting the hypothesis that FL lymphoid precursors seed the thymus[4]. We combined early lymphoid precursor states from the FL, thymus and fetal BM and identified a proportion of fetal BM early lymphocyte precursors (ELP) with a similar role (**Fig. 3G, Supplementary Table 5**) suggesting that, in keeping with previous reports, fetal BM contributes to thymopoiesis[31].

Intrinsic features of haematopoietic progenitors

Having dissected the committed immune populations of fetal BM, we sought to characterise the progenitors driving haematopoiesis using our fetal BM 10x data (**Extended Data Fig. 5A, B**). A spectrum of haematopoietic progenitors was observed in these 3,741 cells, ranging from uncommitted HSCs expressing *MLLT3, HLF* and *HOPX,* to lineage-committed precursors of lymphoid, myeloid, and erythroid/MK cells. Our $CD34^+$-selected CITE-seq data provided better resolution of the rarer cell states at both mRNA and protein level and, through integration with fetal BM and CB CITE-seq data, allowed us to explore the unique features of fetal BM haematopoiesis (**Fig. 4A, B; Extended Data Fig. 5C, D, Supplementary Table 5, 26**).

All progenitor states were represented in FL, although erythroid precursors dominated (**Fig. 4C, Extended Data Fig. 5E; Supplementary Table 27**). Lymphoid precursors were most prevalent in fetal BM, particularly the later B lineage subsets (LyP II, III, IV). In contrast,

CB was enriched in HSC/MPPs, CLPs, and the earliest erythroid precursors (EryP I). As expected, CB HSPCs were less proliferative than fetal tissues (**Fig. 4D**). Fetal BM HSCs were at least as proliferative as FL HSCs and cycling progenitors were seen across the BM landscape, rendering it unlikely that the BM repertoire reflects passive colonisation by FL cells.

To assess whether the contrasting profiles of FL, fetal BM and CB were forged by differentiation bias of the earliest $CD34^+$ progenitors, we performed Direction of Transition (DoT) analysis[32] with adult BM 10x data as a reference landscape. We calculated 'cell state change vectors' using DEGs between HSCs (HSC/MPP I cluster) in FBM, FL and CB (**Fig. 4E, Supplementary Table 28**) and fetal BM and FL (**Extended Data Fig. 6A, Supplementary Table 29**). FL DEGs were consistent with a bias towards erythroid fate (red region on dotscore UMAP) and away from myeloid and lymphoid fate (blue regions). Fetal BM progenitors appeared to be significantly biased towards neutrophil and B lineages (**Extended Data Fig. 6A)**. This further supports the notion that fetal BM haematopoiesis is not simply a mirror of FL haematopoiesis, as differential programming is observed from the earliest progenitors. CB progenitors appeared to favour myeloid and lymphoid cell production over erythroid, in keeping with erythroid specification as particularly important in early fetal development (**Fig. 4E**).

To establish whether transcriptional differences and predicted biases in FL and fetal BM HSC/MPPs corresponded to functionally relevant differentiation potential, we FACS-isolated single HSCs from paired FL and fetal BM suspensions and performed single cell clonal cultures over MS5 mouse stroma with cytokines and growth factors supporting erythroid, megakaryocyte, NK lymphoid, monocyte and neutrophil output (**Fig. 4F-H, Extended Data**

Figs. 6B-C). Notably, the ability to produce B cells was not tested in this assay. More committed populations ($CD34^{+}CD38^{-/mid}$ and MLP) were also FACS-isolated and cultured for comparison. Relative to FL HSC/MPPss, fewer fetal BM HSC/MPPs produced colonies (**Fig. 4F)**. Fetal BM colonies showed greater lineage restriction (**Fig. 4G, H)**. Myeloid colonies arose frequently from both FL and fetal BM HSC/MPPs but myeloid-restricted colonies (i.e. no other potential) were typical of fetal BM (**Fig. 4H)**, supporting the greater bias of fetal BM HSC/MPPs to the myeloid fate, as predicted by DoT analysis, and the more established myeloid cell repertoire observed in BM (**Fig. 2B, Extended Data Figs. 6A**).

Intrinsic and extrinsic perturbation of haematopoiesis in Down syndrome

Children with trisomy 21 (Down syndrome; DS) have an approximately 50-fold increased risk of developing acute leukaemia in the first four years of life[15]. Leukaemias of both myeloid and lymphoid lineages occur, although myeloid leukaemia risk is more substantially elevated compared with the disomic population (150-fold for AML and 600-fold for AMKL)[15]. Myeloid leukaemia of DS is known to arise *in utero,* when fetal cells acquire somatic mutations in the *GATA1* TF gene and drive a highly proliferative neonatal pre-leukaemic syndrome[33]. Lymphoid leukaemias in DS are exclusively B-lineage, typically present after 12 months of age, and are less clearly attributable to *in utero* haematopoiesis. Children with DS frequently experience immune pathology, including increased susceptibility to infection[34], higher infection-related mortality[14], B-lymphopenia[35,36] and a markedly increased prevalence of autoimmunity[13,37]. While haematopoietic perturbations have been reported in DS FL, little is known about how haematopoiesis and immune cell development may be affected in fetal BM.

We generated scRNA-seq data from 4 DS fetal BM samples (12-13 PCW), in which *GATA1* mutations had been excluded, with 16,743 of 17,956 cells passing QC **(Fig. 5A; Supplementary Table 30, 31**). The average gene count per single cell in DS fetal BM 10x was comparable with disomic fetal BM 10x, and good consensus was observed between annotations (**Extended Data Fig. 7A, B**). The relative abundance of broad cell states in DS fetal BM was compared with age-matched disomic (non-DS) fetal BM using Milo[38] (**Fig. 5B**). The erythroid lineage was significantly more abundant, and the MK and B lineages significantly reduced in DS. Relative suppression of the MK lineage in DS BM was not expected as DS FL HSCs show intrinsic erythroid/MK bias and increased megakaryocytes[11]. There was greater representation of mid and late erythroid cells in the DS data, accompanied by increased enrichment in cell cycle gene expression in early and mid erythroid cells (**Extended Data Fig. 7C, D**).

Erythroid differentiation was reconstructed using Monocle3 in both DS and non-DS fetal BM (**Extended Data Fig. 7E, Supplementary Tables 18, 32, 33**). Distinct patterns of gene expression across DS pseudotime included increased expression of cell cycle genes (*CCND3, MKI67*), early recruitment of haem synthesis genes (*ALAS2, FECH*) and elevated, sustained expression of glycolysis gene *PKLR*. These data suggest that more rapid proliferation of emerging erythroid cells in DS, supported by metabolic and biosynthetic adaptations, underpin the DS BM erythroid dominance. Reconstruction of MK and B lineage trajectories were not performed due to limited cell numbers in DS.

We performed in vitro differentiation assays to test our hypothesis that the cellular topography of DS fetal BM reflects differentiation bias of DS BM stem/progenitors. Single HSC/MPPs from DS and age-matched non-DS BM were sorted onto methylcellulose and

MS-5- supported systems (sorting as per **Extended Data Fig. 6B**). On methylcellulose, DS HSC/MPPs produced significantly more erythroid colonies and fewer myeloid colonies (GEMM/GM) (**Fig. 5C**). In keeping with our observation that early- and mid-erythroid cells are more actively cycling in DS, erythroid colonies were consistently larger and produced a greater number of sub-colonies (**Extended Data Fig. 7F**). Colonies could not be generated from DS HSC/MPPs on MS5 with either 1 or 20 HSC/MPPs per well, perhaps reflecting the need of DS fetal BM HSC/MPP for specific additional cytokines, given that matched FL samples from the same fetus efficiently generated colonies with more erythroid and fewer myeloid cells than non-DS FL.

We sought to explore the lineage biases imposed by trisomy 21 on fetal BM HSC/MPPs. DEGs were calculated between paired cell states in DS and non-DS 10x fetal BM datasets, and all showed increased expression of chromosome 21 genes in DS (**Supplementary Table 34**), in addition to genome-wide transcriptional differences (**Fig. 5D**). Marked expression differences were noted for populations affected in DS, including MEMP, MK and B-lineage cells, suggesting these genes are likely to be important in driving lineage biases. DEG analysis further revealed overexpression of chromosome 21 TFs with documented roles in haematopoiesis in DS, including *U2AF1* (MEMP), *U2AF1* and *ETS2* (MK) and *ETS2* (B lineage) (**Extended Data Fig. 7G**).

To elucidate the molecular regulation of gene expression differences, we used PySCENIC to infer TF activity for DS and non-DS cell states (**Fig. 5E**). Inferences proposed that non-DS HSCs were regulated by TFs with well-defined roles in early haematopoietic programming (TAL1, SPI1, FLI1[39,40]), but DS HSCs lacked the expected regulatory programmes. Instead, DS regulons included 3 ETS-domain containing TFs- ELK1, SRF and GABPA. These TFs

activate basal splicing, transcription and translation, but exhibit overlap in their promoter binding regions[41]. GABPA, encoded on chromosome 21 is implicated in differentiation and maintenance of HSC/MPPs[42]. Analysis for non-DS MKs and their progenitors (MEMP) implicated TFs critical to erythroid versus MK fate specification (FLI1, KLF1, GATA1)[43], but analysis for DS-MKs suggested disruption of these programmes, and are in keeping with recent data showing marked silencing of the FLI1 promoter in DS[44]. While attempts have been made to distill the dysregulated TFs underpinning DS haematopoiesis in model systems[45,46], our data suggest that a complex blend of overlapping TFs and regulatory mechanisms underpin the lineage biases we observe in the human fetal BM.

Prompted by the frequency of TNF family proteins in DEGs between DS and non-DS fetal BM cell states (**Supplementary Table 34**), and the prediction that NFkB subunit RELA regulates multiple DS cells states, we sought to investigate whether inflammatory signaling may contribute to haematopoietic dysregulation in DS. *TNF* was overexpressed in most myeloid lineages in DS fetal BM relative to non-DS, in keeping with reports of raised circulating TNF**α** in DS[47] (**Fig. 5F)**. TNF**α** signalling pathway genes were over-represented in DS myeloid, erythroid, NK and stromal cells (**Fig. 5F)**. CellPhone DB analysis predicted that DS BM HSCs integrate TNF-family signals, primarily from BM myeloid cells, indicating that HSC-extrinsic factors as well as intrinsic lineage bias shape haematopoiesis in DS fetal BM (**Fig. 5G; Supplementary Table 35**).

Haematopoietic niche in fetal bone marrow

The non-haematopoietic fraction of the fetal BM 10x data comprised 6,287 cells. Fine-resolution clustering revealed 19 unique cell states (**Fig. 6A, Extended Data Fig. 8A-D, Supplementary Table 36**). Annotation was harmonised with stromal cell states recently

described in mouse BM[48] (**Extended Data Fig. 8E**). While osteoclasts were not captured in the mouse data due to experimental factors, there was otherwise strong correlation between fetal BM and mouse counterparts (**Extended Data Fig. 8E**). The initial stage of BM had a unique balance of cell states, with reduced proportions of endothelial cells (EC), osteoblasts and adipo-CAR cells and a greater proportion of osteoclasts compared to later stages, probably reflecting the requirement for bone remodelling cells early in BM niche establishment (**Extended Data Fig. 8A**).

We focused our investigation on osteochondral and EC, which are reported to form the endosteal and endothelial HSC niches[49]. We annotated one cluster of 191 cells as osteochondral precursors based on their expression of genes characteristic of osteoblasts (*CPE*, *IBSP*), Adipo-CARs (*VCAN, GGT5*), chondrocytes (*CSPG4, EDIL3*), genes involved in musculoskeletal development (*AEBP1*) and position within FDG visualisation (**Extended Data Fig. 8D**). Sinusoidal ECs expressing *STAB1, STAB2* and *TSPAN7* were the most abundant EC type in fetal BM (**Fig. 6A, Extended Data Figs. 8A, B**). The sinusoids form a portal of entry and exit for migratory progenitors and differentiated cell types in BM as well as other organs capable of haematopoiesis (liver and spleen). Refined clustering and re-annotation of ECs from FL allowed us to identify an analogous EC population **(Supplementary Table 16**). Fetal BM sinusoidal ECs had significantly higher expression of *SELE*, *VCAM1* and *ICAM2* than FL sinusoidal ECs, permitting SELPLG/PSGL1 and ITGA4/ITGB1 interactions but blocking LFA-1-dependent cell adhesion (**Fig. 6B, Supplementary Table 16)**. Fetal BM sinusoidal ECs expressed more extracellular matrix genes, including *THBS1*, which contributes to HSC/MPP retention[50] and matrix metalloproteinases which facilitate mature cell egress[51] (**Fig. 6B)**. Fetal BM sinusoidal ECs expressed more *CCL14*, promoting proliferation of myeloid progenitors, while FL expressed

more *CCL23* which has the opposite effect[52], supporting a role for ECs in shaping tissue-specific myelopoiesis.

We used CellphoneDB to predict significant receptor-ligand interactions between fetal BM HSCs and non-haematopoietic stromal cells (**Fig. 6C, Extended Data Fig. 8G, Supplementary Table 37**). HSC-stromal interactions with significant association and receptor-ligand expression were grouped by (1) directionality of signalling and (2) niche environment, whether arteriolar, endosteal or stromal; see Methods for niche groupings. EC and adipo-CAR cells were predicted to provide critical signals to support haematopoiesis via CD44, KIT, FLT3, NOTCH1/2, NRP2, MERTK, FZD6 and FGFR1 and to support HSC homing and retention via SELL, JAM3 and ESAM (**Fig. 6C; Extended Data Fig. 8G, left**). Osteochondral cells and stromal cells appeared capable of providing similar support, though NOTCH signalling interactions were fewer (**Fig. 6C; Extended Data Fig. 8G, left**). Protein-level expression of HSC receptors was confirmed in the $CD34^+$ fetal BM CITE-seq data (where corresponding antibodies were present in the panel) (**Extended Data Fig. 8G, right**). In the other direction, HSCs were predicted to signal to, and support development of their niches, in particular to the tip and proliferating ECs and the osteochondral precursor via ANGPT2, DLK1, EFNA1 and FGF7 **(Extended Data Fig. 9A, B).**

To identify predicted interacting cell types *in situ*, we used standard histochemistry and multiplex immunofluorescence microscopy on sections of fetal bone **(Fig. 6D).** $CD34^+VEGFR2^+$ sinusoids permeated the entire haematopoietic cavity (**Extended Data Figs. 9C, E, F**) and megakaryocytes, erythroblasts and mature myeloid cells were easily seen (**Extended Data Fig. 9D**). $CXCL12^+$ CAR cells were numerous and abutted the sinusoidal endothelium (**Extended Data Figs. 9E, F)**. $CD163^+$ macrophages were not in direct contact

with ECs but followed the sinusoidal architecture (**Extended Data Figs. 9E, F**). $CD34^{+}VEGFR2^{-}CD117^{lo/+}$ HSCs were rare, but were frequently located adjacent to sinusoids **(Fig. 6E).** Quantitative analysis demonstrated significant association of HSCs and ECs relative to background **(Fig. 6E, F; Supplementary Table 38**), validating our predictions that the arteriolar niche provides important support for BM HSCs.

Non-haematopoietic cells isolated from DS fetal BM were few in number, precluding the detailed annotation of cell states we performed for non-DS fetal BM. However, three broad categories could be ascertained (ECs, macrophages/osteoclasts and osteochondral cells). The key interactions predicted between EC and HSC in non-DS fetal BM were examined in DS fetal BM (**Fig. 6G, Supplementary Table 39**). NOTCH ligands *NOV*, *DLL4* and *DLK1* were more abundant in DS endothelium. NOTCH signalling has a critical role in HSC emergence, maintenance and response to proinflammatory signals[53,54]. Having identified activation of TNFα signaling pathways in DS haematopoietic cells, we probed for inflammatory programmes in DS stromal cell groups, relative to non-DS counterparts. Expression of genes implicated in TNFα pathways were increased in macrophages/osteoclasts, in line with myeloid cells shown in **Fig. 5F,** but genes implicated in type I interferon, IFNɣ and other inflammatory cytokine (IL1, 6, 7, 12) response pathways were overexpressed in DS ECs and osteochondral cells, indicating that broad dysregulation of immune homeostasis alters the BM niche in DS (**Fig. 6H, Supplementary Table 40**).

Discussion

It is well-established that survival of the fetus depends on successful initiation of haematopoiesis in waves across early developmental time and space. Understanding how the BM contributes to fetal blood and immune cell production has been limited to date. Through multi-omic dissection of the fetal BM and its stroma, we identify the BM as a key site of neutrophil emergence, myeloid diversification and B lymphoid selection.

Using these data, we explore vulnerabilities of early life haematopoiesis, including susceptibility to B-ALL and leukaemia associated with congenital neutropenia. In DS, we discover haematopoietic abnormalities that are associated with intrinsic bias of stem/progenitors, underpinned by genome-wide transcriptional changes that reflect the combinational effects of gene regulatory network disturbance and the influence of immune dysregulation. By comparison with scRNA-seq datasets across haematopoietic time and location, we reveal unique intrinsic properties of fetal BM HSCs, and we describe how BM stromal populations exert extrinsic influence on HSC development. Our data constitute a novel resource spanning human developmental haematopoiesis that has potential to inform regenerative and transplantation therapies, for example, through co-opting developmental programmes to accelerate reconstitution of haematopoietic stem cells transplants, and manipulating the lineage bias of differentiating progenitors to address specific deficiencies or for cellular therapy.

References

1. O'Byrne, S. *et al.* Discovery of a CD10-negative B-progenitor in human fetal life identifies unique ontogeny-related developmental programs. *Blood* **134**, 1059–1071 (2019).
2. Charbord, P., Tavian, M., Humeau, L. & Péault, B. Early ontogeny of the human marrow from long bones: an immunohistochemical study of hematopoiesis and its microenvironment. *Blood* **87**, 4109–4119 (1996).
3. Park, J.-E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, (2020).
4. Popescu, D.-M. *et al.* Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
5. Wiemels, J. L. *et al.* Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet* **354**, 1499–1503 (1999).
6. Mahmoud, H. H. *et al.* Intrauterine monoclonal origin of neonatal concordant acute lymphoblastic leukemia in monozygotic twins. *Med. Pediatr. Oncol.* **24**, 77–81 (1995).
7. Hong, D. *et al.* Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science* **319**, 336–339 (2008).
8. Gale, K. B. *et al.* Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13950–13954 (1997).
9. Muntean, A. G., Ge, Y., Taub, J. W. & Crispino, J. D. Transcription factor GATA-1 and Down syndrome leukemogenesis. *Leuk. Lymphoma* **47**, 986–997 (2006).
10. Tamblyn, J. A. *et al.* Prenatal therapy in transient abnormal myelopoiesis: a systematic review. *Arch. Dis. Child. Fetal Neonatal Ed.* **101**, F67–71 (2016).
11. Roy, A. *et al.* Perturbation of fetal liver hematopoietic stem and progenitor cell

development by trisomy 21. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17579–17584 (2012).

12. Roberts, I., O'Connor, D., Roy, A., Cowan, G. & Vyas, P. The impact of trisomy 21 on foetal haematopoiesis. *Blood Cells, Molecules, and Diseases* vol. 51 277–281 (2013).
13. Jørgensen, I. F. *et al.* Comorbidity landscape of the Danish patient population affected by chromosome abnormalities. *Genet. Med.* **21**, 2485–2495 (2019).
14. Garrison, M. M., Jeffries, H. & Christakis, D. A. Risk of death for children with down syndrome and sepsis. *J. Pediatr.* **147**, 748–752 (2005).
15. Hasle, H. Pattern of malignant disorders in individuals with Down's syndrome. *Lancet Oncol.* **2**, 429–436 (2001).
16. Coşkun, S. *et al.* Development of the fetal bone marrow niche and regulation of HSC quiescence and homing ability by emerging osteolineage cells. *Cell Rep.* **9**, 581–590 (2014).
17. Gao, X., Xu, C., Asada, N. & Frenette, P. S. The hematopoietic stem cell niche: from embryo to adult. *Development* **145**, (2018).
18. Kawamura, S. *et al.* Identification of a Human Clonogenic Progenitor with Strict Monocyte Differentiation Potential: A Counterpart of Mouse cMoPs. *Immunity* vol. 46 835–848.e4 (2017).
19. Kwok, I. *et al.* Combinatorial Single-Cell Analyses of Granulocyte-Monocyte Progenitor Heterogeneity Reveals an Early Uni-potent Neutrophil Progenitor. *Immunity* (2020) doi:10.1016/j.immuni.2020.06.005.
20. Olsson, A. *et al.* Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Experimental Hematology* vol. 44 S24 (2016).
21. Dahl, R. *et al.* Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPalpha ratio and granulocyte colony-stimulating factor. *Nat. Immunol.* **4**, 1029–1036 (2003).

22. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 1517 (2019).
23. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
24. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, (2017).
25. Rolink, A. G., Schaniel, C., Andersson, J. & Melchers, F. Selection events operating at various stages in B cell development. *Curr. Opin. Immunol.* **13**, 202–207 (2001).
26. Roy, A. *et al.* High resolution IgH repertoire analysis reveals fetal liver as the likely origin of life-long, innate B lymphopoiesis in humans. *Clin. Immunol.* **183**, 8–16 (2017).
27. Greaves, M. Author Correction: A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat. Rev. Cancer* **18**, 526 (2018).
28. Stieglitz, E. *et al.* The genomic landscape of juvenile myelomonocytic leukemia. *Nat. Genet.* **47**, 1326–1333 (2015).
29. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* vol. 446 758–764 (2007).
30. Shi, Y. *et al.* Chromatin accessibility contributes to simultaneous mutations of cancer genes. *Scientific Reports* vol. 6 (2016).
31. Haddad, R. *et al.* Dynamics of thymus-colonizing cells during human development. *Immunity* **24**, 217–230 (2006).
32. Kucinski, I. *et al.* Interactions between lineage-associated transcription factors govern haematopoietic progenitor states. *EMBO J.* e104983 (2020).
33. Roberts, I. & Izraeli, S. Haematopoietic development and leukaemia in Down syndrome. *Br. J. Haematol.* **167**, 587–599 (2014).
34. Sung, L., Lange, B. J., Gerbing, R. B., Alonzo, T. A. & Feusner, J. Microbiologically

documented infections and infection-related mortality in children with acute myeloid leukemia. *Blood* **110**, 3532–3539 (2007).

35. Farroni, C. *et al.* Dysregulated miR-155 and miR-125b Are Related to Impaired B-cell Responses in Down Syndrome. *Front. Immunol.* **9**, 2683 (2018).
36. Immune Dysfunction in Down's Syndrome: Primary Immune Deficiency or Early Senescence of the Immune System? *Clin. Immunol. Immunopathol.* **78**, 209–214 (1996).
37. Goldacre, M. J., Wotton, C. J., Seagroatt, V. & Yeates, D. Cancers and immune related diseases associated with Down's syndrome: a record linkage study. *Arch. Dis. Child.* **89**, 1014–1017 (2004).
38. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Milo: differential abundance testing on single-cell data using k-NN graphs. doi:10.1101/2020.11.23.393769.
39. Pimanda, J. E. *et al.* Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 17692–17697 (2007).
40. Iwasaki, H. *et al.* Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood* **106**, 1590–1600 (2005).
41. Boros, J. *et al.* Elucidation of the ELK1 target gene network reveals a role in the coordinate regulation of core components of the gene regulation machinery. *Genome Res.* **19**, 1963–1973 (2009).
42. Yu, S. *et al.* GABP controls a critical transcription regulatory module that is essential for maintenance and differentiation of hematopoietic stem/progenitor cells. *Blood* **117**, 2166–2178 (2011).
43. Palii, C. G. *et al.* Single-Cell Proteomics Reveal that Quantitative Changes in Co-expressed Lineage-Specific Transcription Factors Determine Cell Fate. *Cell Stem*

Cell **24**, 812–820.e5 (2019).

44. Muskens, I. S. *et al.* The genome-wide impact of trisomy 21 on DNA methylation and its implications for hematopoiesis. *Nat. Commun.* **12**, 821 (2021).
45. Birger, Y. *et al.* Perturbation of fetal hematopoiesis in a mouse model of Down syndrome's transient myeloproliferative disorder. *Blood* **122**, 988–998 (2013).
46. Banno, K. *et al.* Systematic Cellular Disease Models Reveal Synergistic Interaction of Trisomy 21 and GATA1 Mutations in Hematopoietic Abnormalities. *Cell Rep.* **15**, 1228–1241 (2016).
47. Sullivan, K. D. *et al.* Trisomy 21 causes changes in the circulating proteome indicative of chronic autoinflammation. *Sci. Rep.* **7**, 1–11 (2017).
48. Baccin, C. *et al.* Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nat. Cell Biol.* **22**, 38–48 (2020).
49. Boulais, P. E. & Frenette, P. S. Making sense of hematopoietic stem cell niches. *Blood* **125**, 2621–2629 (2015).
50. Long, M. W., Briddell, R., Walter, A. W., Bruno, E. & Hoffman, R. Human hematopoietic stem cell adherence to cytokines and matrix molecules. *Journal of Clinical Investigation* vol. 90 251–255 (1992).
51. Lane, W. J. *et al.* Stromal-derived factor 1–induced megakaryocyte migration and platelet production is dependent on matrix metalloproteinases. *Blood* vol. 96 4152–4159 (2000).
52. Mukaida, N., Tanabe, Y. & Baba, T. Chemokines as a Conductor of Bone Marrow Microenvironment in Chronic Myeloid Leukemia. *Int. J. Mol. Sci.* **18**, (2017).
53. Butko, E., Pouget, C. & Traver, D. Complex regulation of HSC emergence by the Notch signaling pathway. *Dev. Biol.* **409**, 129–138 (2016).
54. Espín-Palazón, R. *et al.* Proinflammatory signaling regulates hematopoietic stem cell

emergence. *Cell* **159**, 1070–1085 (2014).

55. Matsushima, H. *et al.* Neutrophil differentiation into a unique hybrid population exhibiting dual phenotype and functionality of neutrophils and dendritic cells. *Blood* **121**, 1677–1689 (2013).
56. Pandey, V. D., Singh, V., Nigam, G. L., Usmani, Y. & Yadav, Y. Fetal foot length for assessment of gestational age: A comprehensive study in north India. *Journal of the Anatomical Society of India* vol. 65 S19 (2016).
57. Tunstall-Pedoe, O. *et al.* Abnormalities in the myeloid progenitor compartment in Down syndrome fetal liver precede acquisition of GATA1 mutations. *Blood* **112**, 4507–4511 (2008).
58. Mulè, M. P., Martins, A. J. & Tsang, J. S. Normalizing and denoising protein expression data from droplet-based single cell profiling. doi:10.1101/2020.02.24.963603.
59. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

Figure Legends

Figure 1: A single cell transcriptome map of human fetal BM

1A) Experimental overview depicting i) generation of scRNA-seq and CITE-seq data from dissociated fetal BM cells, ii) validating findings with FACS-isolated cells for scRNA-seq (SS2) and assessment of morphology, and iii) use of comparison data sets.

1B) UMAP visualisation of the discovery (10x) fetal BM dataset with broad annotation of cell states (n=9, k = 103,228). Cell type is represented by colour, as shown in panel D legend.

1C) Dot plot showing gene expression of cell state-defining genes. Log-transformed, normalised and scaled gene expression value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

1D) Scatter plot showing age-distribution of fetal BM samples and grouping into 4 stages. Bar plot showing frequency of cell states across gestational age, with colour of bar representing each cell type (colour key in legend). Statistical significance of cell frequency change by stage shown in parentheses (Quasibinomial regression with bootstrap correction for sort gates; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, details in **Supplementary Table 7)**.

1E) UMAP visualisation of SS2 data (n=2, k=486) with a 50-cell per subset sample of predicted 10x counterparts (n=9, k=600). Cell states in 10x data are represented by coloured

areas and cell states in SS2 data are represented by dots of the equivalent colour. 10x and SS2 plots can be viewed without overlay in **Extended Data Fig. 2D**.

1F) Cytospin images of selected populations sorted according to gating strategy shown in **Extended Data Fig. 2B** and stained with Giemsa. 100x images were concatenated as shown by dotted lines.

1G) Dot plot showing differentially expressed proteins in developing monocyte and neutrophil subsets, as defined by mRNA-annotation, in fetal BM total CITE-seq data. DSB-normalised protein expression value is represented by the colour of the dot. Percentage of cells in each subset expressing the marker is shown by the size of the dot.

Extended Data Figure 1: A single cell transcriptome map of human fetal BM

1A) Summary of scRNA-seq datasets used for comparison: published YS and FL data[4] and publicly available CB and adult BM data from the Human Cell Atlas Data Coordination Portal.

1B) UMAP visualisation of CB scRNA-seq dataset with our broad annotations of cell state applied (n=4, k=148,442). Cell type is represented by colour, as shown in legend.

1C) UMAP visualisation of adult BM scRNA-seq dataset with our broad annotations of cell state applied (n = 4, k = 142,026). Cell type is represented by colour, as shown in legend.

1D) Logistic regression of probability of consensus plotted for intersecting cell states annotated between fetal BM 10x dataset and the equivalent compartment in adult BM and CB 10x datasets.

1E) UMAP visualisation of fetal BM 10x gene expression (GEX) data (n=9, k=103,228), pre and post batch correction. Sequencing type and sample is represented by colour, as shown in legend.

Extended Data Figure 2: A single cell transcriptome map of human fetal BM

2A) Dot plots showing gene and corresponding surface protein expression of cell-state defining markers in fetal BM CITE-seq dataset. Log-transformed, normalised and scaled GEX values and DSB-normalised protein expression values are represented by the colour of the dots. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

2B) Sorting strategy used to isolate cell types for validation based on subset defining markers from scRNA-seq (10x) data. From live, $CD45^+$ single cells, $CD123^+HLA\text{-}DR^-$ basophils and $CD123^+HLA\text{-}DR^+$ pDCs were gated. From the remaining cells, $CD34^+CD117^{mid\text{-}hi}$ progenitors and $CD117^{hi}$ mast cells were gated. Next, $CD125^+FSC^{hi}$ eosinophils were gated. Subsequently, $HLA\text{-}DR^+CD79a^+$ B cells were separated. As CD79 is weakly expressed on the cell surface, a significant number of B cells fall in the $HLA\text{-}DR^+CD79^-$ gate, forming the $CD14^-CD204^-CLEC9A^-CD1c^-$ population. From $HLA\text{-}DR^+$ cells, $CD14^+CD204^-$ monocytes were gated. Within the $CD14^-CD204^-$ population, $CLEC9A^+$ DC1 and $CD1c^+$ DC2 were identified. From $HLA\text{-}DR^-$ cells, $CD11b^-CD52^+$ T and NK cells were excluded. From the remaining cells $CD11b^-CD66b^-$ promyelocytes, $CD11b^-CD66b^+$ metamyelocytes/myelocytes and $CD11b^+CD66b^+$ neutrophils were selected. Note, the CD11b gating was as per published reports mature neutrophils being $CD11b^+$ and immature neutrophils being $CD11b^-$[55].

2C) Dot plot showing expression of cell-state defining genes in 10x data used to design sorting strategy. Equivalent surface antigen names are shown in parentheses. Log-transformed, normalised and scaled GEX value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

Dashed boxes indicate the gating strategy for cell-sorting, for example pDCs were isolated as $CD45^{+}HLA\text{-}DR^{+}CD123^{+}$ cells.

2D) UMAP visualisation of validation SS2 data (n=2, n=486) with a 50-cell per subset sample of predicted 10x counterparts (n=9, k=600). Top plot- coloured dots show SS2 data and grey dots 10x data. Bottom plot- coloured dots show 10x data and grey dots SS2 data. Overlay can be viewed in **Fig. 1E**.

Figure 2: Diversification of innate myeloid and lymphoid cells

2A) UMAP visualisation of fetal BM myeloid cells (n = 34,318). Cell type is represented by colour, as shown in legend.

2B) Scatterplot detailing the frequency of monocyte, neutrophil and DC lineage cells as a proportion of total haematopoietic cells for YS (n=3), FL (n=14), fetal BM (FBM; n=9) and adult BM (ABM; n=4) within the 10x GEX datasets (statistical significance tested by one-way ANOVA with Tukey's multiple comparison tests). Bar plot showing the relative proportions of cell states within monocyte, neutrophil and DC lineages for YS, FL, fetal BM and adult BM 10x GEX datasets, with colour of bar representing each cell type (colour key in legend). Macrophages were not compared across tissue as they were not captured in adult BM samples. Statistical significance of myeloid cell frequency change by tissue shown in parentheses (Quasi binomial regression with bootstrap correction for sort gates; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, NT = not tested; details in **Supplementary Table 15**).

2C) Left: Illustration displaying interaction of *SPI1* and *CEBPA* in monocyte and neutrophil differentiation from GMPs. Right: log, normalised and scaled expression of *CEBPA* and *SPI1* in GMP cells from FL and fetal BM shown as a scatter plot.

2D) Heat map showing expression of genes implicated in severe congenital neutropenia (**Supplementary Table 20**) across fetal BM Monocle3-inferred neutrophil pseudotime. Expression values are log-transformed, normalized and scaled. Genes differentially expressed

across fetal BM neutrophil pseudotime are marked with an asterisk (and full list provided in **Supplementary Table 19**).

2E) FDG visualisation of fetal BM DC cell states (n=5,702). Cell state is represented by colour, as shown in legend.

2F) FDG visualisation of fetal BM NK and ILC cell states (n=915). Cell state is represented by colour, as shown in legend. Grey ellipse highlights proliferating cells.

2G) Heat map showing cytotoxicity of NK cell states in YS, FL and fetal BM, as defined by enrichment of genes in the KEGG NK cytotoxicity pathway (See Methods). Relative enrichment is indicated by colour scale.

Extended Data Figure 3: Diversification of innate myeloid and lymphoid cells

3A) Dot plot showing expression of cell state-defining marker genes in fetal BM DC and monocyte lineage cells. Log-transformed, normalised and scaled GEX value is represented by the colour of dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

3B) Dot plot showing expression of cell state-defining marker genes in fetal BM myeloid precursors and neutrophil lineage cells. Log-transformed, normalised and scaled GEX value is represented by the colour of dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

3C) FDG visualisation of fetal BM progenitor, neutrophil and monocyte cell states (n=32,448). Cell state is represented by colour, as shown in legend.

3D) Heatmap showing gene expression of markers for early monocyte and neutrophil commitment (derived from promonocyte vs. promyelocyte DEGs) in fetal BM progenitor cell states (full DEG list shown in **Supplementary Table 17**). GEX shown is log-transformed, normalised and scaled.

3E) Waddington-OT fate map for monocyte and neutrophil lineage fetal BM cells at each week of gestation (19PCW is not shown as the algorithm does not generate a map for the terminal time point). Vertices represent maximum probability for pure fates of: CD14 monocyte, neutrophil, or other (upstream monocyte/neutrophil-lineage cell types as depicted in **Extended Data Fig 3C**).

3F) Heat map showing transcriptional similarity between DC/monocyte cell states reported in blood[24] with those identified in developing and mature haematopoietic tissues (YS, FL[4] fetal BM, adult BM). Gene enrichment scores are based on gene signatures pooled from top 100 statistically significant monocyte/DC DEGs in typical blood[24].

3G) Heatmap of predicted activity of TFs across pseudotime. TF activity was inferred using PySCENIC and pseudotime was calculated in Scanpy (*sc.tl.dpt*). For each TF the expression data was normalised to between 0-1 prior to plotting.

3H) Dot plot showing expression of genes in fetal BM NK and ILC cells. Log-transformed, normalised and scaled gene expression value is represented by the colour of dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

Figure 3: Establishment of the adaptive immune repertoire

3A) FDG visualisation of fetal BM B lineage cell states (n=28,583). Cell state is represented by colour, as shown in legend. Grey ellipses highlight cycling cells. Pre pro B = Pre Pro B progenitor, Pro B = Pro B progenitor, Pre B = Pre B precursor, Imm. B = Immature B cell. Differentiation paths are noted with dashed arrows for i) Cycling Pre B cells and ii) B cell differentiation.

3B) Bar plot showing frequency of productive chains and proportion of heavy and light chains in fetal BM B lineage cells (n=8,298), as defined by BCR-enriched VDJ data. Accompanying pie charts show percentage of cycling cells (defined as > mean cell cycle score) displayed per B-lineage cell-type.

3C) Scatter plot detailing the frequency of B lineage cells as a proportion of haematopoietic mononuclear cells present in YS (n=3), FL (n=14), FBM (n=9) and ABM (n=4) 10x GEX datasets (statistical significance tested by one-way ANOVA with Tukey's multiple comparison tests). The denominator is haematopoietic mononuclear cells to minimize the effect of variable granulocyte depletion in sample preparation. Bar plot showing relative proportions of cell states within the B lineage in YS, FL, FBM and ABM 10x GEX datasets, with colour of bar representing each cell type (colour key in legend; the YS bar is shown in white to illustrate that no B lineage cells are present). Statistical significance of cell state frequency change by tissue is shown in parentheses (Quasibinomial regression with bootstrap correction for sort gates; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, details in **Supplementary Table 21**).

3D) Violin-plot of DEGs between Pre-B progenitors in paths (i - Cycling) and (ii - B cell differentiation) as per **Fig. 3A**. Pre-B progenitor cells were assigned to a path based on cell cycle enrichment score, as described in Methods. Expression values shown on y axis are log-transformed, normalised and scaled. All genes are differentially expressed between both conditions; **** $p<0.0001$ (**Supplementary Table 22**).

3E) Heat map showing expression of genes implicated in B-ALL across fetal BM B lineage pseudotime inferred by Monocle3 (details in **Extended Data Fig. 4C**, and **Supplementary Tables 23, 24**). Expression values are log-transformed, normalized and scaled. Genes differentially expressed across pseudotime are marked with an asterisk.

3F) Dot plot comparing expression of cell state- defining markers genes in thymus and fetal BM (FBM). Log-transformed, normalised and scaled GEX value is represented by the colour of dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

3G) UMAP visualisation of fetal BM, FL, thymus progenitors and T precursors (n=4,766). Cell type is represented by colour, as shown in legend. ELP = early lymphocyte precursor, DN T cell = double negative T cell, ETP = early thymocyte precursor.

Extended Data Figure 4: Establishment of the adaptive immune repertoire

4A) Dot plot showing expression of cell state-defining marker genes in fetal BM B lineage cells. Log-transformed, normalised and scaled GEX value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

4B) Heatmap showing number of shared clonotypes between B lineage cell types, as defined by CellRanger. Number of shared clonotypes is shown by colour of square, with colour scale in legend.

4C) FDG visualisation of fetal BM B lineage cell states (n=30,066) and their upstream progenitors used in Monocle3 pseudotime analysis. Left: cell states are coloured, as shown in legend. Right: cells are coloured by pseudotime value. Two paths inferred from Monocle3 pseudotime trajectory are depicted by dashed arrows (i - Cycling Pre B cells and ii - B cell differentiation, as shown in **Fig. 3D**).

4D) FDG visualisation of fetal BM B lineage cell states (n=28,583). Cells are coloured by apoptotic gene enrichment score, as defined by their expression of genes in the KEGG apoptotic pathway (details in Methods).

4E) Heat map showing expression of genes implicated in B-ALL across adult BM B lineage pseudotime inferred by Monocle3 (**Supplementary Table 24, 25**). Expression values are log-transformed, normalized and scaled. Genes differentially expressed across pseudotime are marked with an asterisk.

4F) Barplot showing proportion of TRA/B/G/D chains per productively rearranged TCRs by fetal BM T lineage cell state, according to CellRanger VDJ output. Bars (mean) and error bars (SD) of n=2 15 PCW fetal BM samples are shown. Mean±SD percent productivity of TCRs was 93±9%, 81±16% and 92±11% for CD4 T cells, CD8 T cells and Treg, respectively.

Figure 4: Intrinsic features of haematopoietic progenitors

4A) FDG visualisation of $CD34^+$ fetal BM/FL/CB CITE-seq cells on GEX landscape (n=35,273). Cell type is represented by colour, as shown in legend. The HSC/MPP cluster labels refer to unsupervised sub-clustering of the most immature compartment rather than previously described functional MPP subpopulations.

4B) Dot plot showing expression of proteins used for progenitor characterization in $CD34^+$ CITE-seq data. DSB-normalised protein expression value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

4C) FDG visualisation of $CD34^+$ fetal BM (FBM)/FL/CB CITE-seq cells on GEX landscape (n=35,273). Smoothed cell abundance in each tissue is approximated by kernel density estimation (KDE) in the FDG space.

4D) Heat map showing cell cycle gene enrichment score for progenitor cell types in $CD34^+$ fetal BM (FBM)/FL/CB CITE-seq. Relative enrichment is indicated by colour scale.

4E) DoT-scores computed for genes differentially expressed between indicated comparisons among HSC/MPP1 cluster cells using adult bone marrow scRNA-Seq landscape as the reference. DoT score takes the origin point (here a matching subset of the HSC/MPP population matching the respective fetal tissue) and visually represents the direction of cell state changes introduced by the differential expression. Red-coloured cells indicate a shift

towards their state and blue-coloured cells a shift away from their state. DoT-score values correspond to z-scores estimated against simulated data.

4F) Outputs from single cell culture on MS5 stromal layer for paired FL and fetal BM HSC/MPPs (conducted as per **Extended Data Figs. 6B, C**). Proportion of sorted wells producing colonies, separated by progenitor type and tissue of origin (n=3 paired fetal BM and FL, n=4 fetal BM only). ***p=0.006 by Mann Whitney test of HSC/MPP vs HSC/MPP and ***p=0.006 committed progenitor vs committed progenitor.

4G) Well contents analysed by flow cytometry and number of lineage outputs per well compared between HSC and committed progenitors from FL vs. fetal BM. U=colony present but lineage undefinable by this assay. Statistical comparison is of unipotential vs. multipotential colonies: HSC/MPP FL vs fetal BM ** p=0.0012; committed progenitor FL vs BM 'ns' p=0.36 by binomial test.

4H) Proportion of FL vs fetal BM HSC/MPPs producing myeloid-containing colonies. Statistical comparison is of 'myeloid-only' vs. 'myeloid plus other' *** p=0.0002 by binomial test.

Extended Data Figure 5: Intrinsic features of haematopoietic progenitors

5A) FDG visualisation of fetal BM progenitor cells (n=3,741) from 10x data. Cell type is represented by colour, as shown in legend.

5B) Violin plots showing expression of MEM-, myeloid- and lymphoid- lineage genes in fetal BM progenitor cells from 10x data. Log-transformed, normalised and scaled GEX values are displayed on the y-axis.

5C) Dot plot showing expression of genes used for progenitor characterization in $CD34^+$ CITE-seq data. Log-transformed, normalised and scaled GEX value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

5D) Logistic Regression of probability of consensus plotted for intersecting cell states annotated between $CD34^+$ CITE-seq data and the equivalent compartment in fetal BM 10x data. Using the logistic regression technique outlined in Methods, we found good consensus between annotations.

5E) Bar graph showing proportion of each progenitor subset within progenitor cells captured in FL (n=4), FBM (n=3) and CB (n=4). Proportions are normalised across donors. Coloured bars represent mean and error bars SD. Quasibinomial regression with bootstrap correction for sort gates; *$p< 0.05$, **$p< 0.01$, ***$p< 0.001$, ****$p< 0.0001$ (**Supplementary Table 27**).

Extended Data Figure 6: Intrinsic features of haematopoietic progenitors

6A) DoT-scores computed for genes differentially expressed between HSC/MPP1 cluster cells from FL and fetal BM using the adult BM scRNA-seq landscape as the reference. DoT score takes the origin point (here a matching subset of the HSC/MPP population matching the respective fetal tissue) and visually represents the directions of cell state changes introduced by the differential expression. Red-coloured cells indicate a shift towards their state and blue-coloured cells a shift away from their state. DoT-score values correspond to z-scores estimated against simulated data.

6B) Sort gates for HSC culture experiments. Single, live, lineage negative $CD34^+$ cells were divided into $CD34^+CD38^{hi}$ (top 20%), $CD34^+CD38^{mid}$ (middle 60%) and $CD34^+CD38^-$ (bottom 20%) fractions. $CD34^+CD38^-$ cells were gated further into $CD45RA^-$ HSC/MPPs and $CD45RA^+$ MLP. HSC/MPP, MLP and $CD34^+CD38^{mid}$ cells were index sorted for single cell culture on an MS5 stromal layer. MLP and $CD34^+CD38^{mid}$ cells were analysed as "committed progenitor".

6C) Examples of single cell HSC culture outputs. Example 1 showing MK ($CD41^+$), erythroid ($CD235a^+$) and myeloid ($CD14^+$ monocyte and $CD15^+$ neutrophil) outputs. Example 2 showing NK ($CD56^+$) and myeloid ($CD14^+$ monocyte and $CD15^+$ neutrophil) outputs.

Figure 5: Intrinsic and extrinsic perturbation of haematopoiesis in Down syndrome

5A) UMAP visualisation of fetal BM scRNA-seq data (10x) generated from 12-13 PCW samples with confirmed trisomy 21 and no *GATA1* mutation (n=4, k=6,743). Cell type is represented by colour, as shown in legend.

5B) Beeswarm plot showing log-fold change in abundance between DS and aged-matched non-DS fetal BM cells in neighbourhoods from different broad cell state clusters (DS n=16,743, non-DS n=9,717). Coloured dots indicate significant difference in abundance (p-value adjusted for multiple testing over 0.1 i.e. FDR at 10%).

5C) Colony outputs from DS and aged-matched non-DS HSC/MPPs. Left: single fetal BM HSC/MPPs sorted onto methylcellulose (DS, n=2, k=365 and non-DS, n=2, k=246). Stacked bar graph show the proportion of colonies designated MK (megakaryocyte), E (erythroid), GEMM (granulocyte, erythroid, macrophage, megakaryocyte) and GM (granulocyte, macrophage). Statistical difference between DS and expected distribution based on non-DS tested by Chi square with **** $p<0.0001$. Right: 20 FL HSC/MPPs per well sorted onto MS5 stromal layer. Scatterplots show the proportion of differentiated cells identified as erythroid and myeloid ($CD11b^+CD14^+$and/or $CD15^+$) by flow cytometry (as in **Extended Data Fig. 6C-D**). Coloured bars show mean and error bars SEM. Statistical difference between DS and non-DS tested by Mann-Whitney with **** $p<0.0001$ and *** $p<0.001$.

5D) Heatmap showing number of DEGs between paired cell states in DS vs. non-DS fetal BM for each chromosome, including correction for number of genes per chromosome (full DEG list provided in **Supplementary Table 34**).

5E) Predicted most highly active TFs in DS vs non-DS HSCs, MEMPs and MK cells (for cell groupings see **Supplementary Tables 3, 31**). PySCENIC was used to predict TF activity in each cell state. This was used to produce median TF activity in each cluster and DS vs non-DS were compared by t-test for each TF to calculate p-values. Top 30 most significantly differentially active TFs by p-value shown.

5F) Top panel: matrix plot of TNF GEX across non-DS and DS fetal BM cell states. Bottom panel: TNFα signalling pathway enrichment was calculated for each cell fraction based on DEGs between non-DS and DS fetal BM (full list shown in **Supplementary Table 34**). Gene set enrichment analysis performed for MSigDB hallmark pathways using the fgsea package in R. Size of dot represents normalised enrichment score (NES) for TNFα signalling pathway genes.

5G) Sankey plot of putative TNF superfamily interactions in DS fetal BM. For each cell category, DEGs between DS and non-DS fetal BM (full list shown in **Supplementary Table 34**) were filtered for ligands and their receptors were identified using CellPhoneDB. Fold change in expression in DS over non-DS is shown in the boxes above (red scale). Below is the predicted receptor coloured by its gene expression level in combined DS and non-DS HSCs (blue scale).

Extended Data Figure 7: Intrinsic and extrinsic perturbation of haematopoiesis in Down syndrome

7A) Violin plot showing consistency of gene capture between DS and non-DS fetal BM 10x scRNA-seq lanes.

7B) Logistic Regression of probability of consensus plotted for intersecting cell states annotated between fetal BM (DS and non-DS) scRNA-seq datasets, using the logistic regression technique outlined in Methods.

7C) Pie charts showing proportions of detailed erythroid cell states in DS vs. non-DS fetal BM. ****$p < 0.0001$ by Chi-square test.

7D) Heat map showing cell cycle gene enrichment score in DS and non-DS fetal BM erythroid cell states. Relative enrichment is indicated by colour scale.

7E) Heat maps showing expression of genes implicated in erythropoiesis across non-DS and DS fetal BM Monocle3-inferred erythroid pseudotime. All genes selected are differentially expressed across both the DS and non-DS erythroid pseudotime (full gene lists shown in **Supplementary Tables 32, 33)**. Expression values are log-transformed, normalized and scaled.

7F) Representative images of erythrocyte colony size and structure from DS (top) and non-DS (bottom) fetal BM HSC/MPP methylcellulose cultures.

7G) Dot plot of chromosome 21 transcription factors significantly differentially expressed (adjusted p value <0.05) in DS vs. non-DS in the different subsets. Size of dot represents average log2 fold change expression in DS vs. non-DS fetal BM, and colour of dot represents log(log(adjusted p value)) for differential expression.

Figure 6: Haematopoietic niche in fetal bone marrow

6A) UMAP visualisation of the non-haematopoietic fetal BM cells (n=6,287). Cell type is represented by colour, as shown in legend. Broad groupings of macrophage/osteoclast, endothelial and osteochondral cell types are shown using dotted lines.

6B) Violin plot of GEX in FL and fetal BM sinusoidal ECs. Genes with documented role in adhesion, extracellular matrix formation and angiopoiesis are shown. Genes that are differentially expressed between tissues are highlighted with an asterisk (subject to Wilcoxon rank-sum test with Benjamini-Hochberg correction for multiple testing; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$).

6C) Summary of receptor-ligand interactions predicted between fetal BM HSC and non-haematopoietic cells by CellPhoneDB (**Supplementary Table 37**). Cells are grouped into arteriolar niche (EC-sinusoidal, EC-proliferating, EC-tip, EC-arteriolar, Fb-arteriolar, adipo-CAR), endosteal niche (Fb-endosteal, Osteochondral precursor, Early osteoblast) and stromal niche (Mac-stromal, Fb-fibroblast). Significant putative receptor-ligands across fetal BM niches are indicated in Venn diagram overlapping regions.

6D) Cross section of fetal bone stained with hematoxylin and eosin, demonstrating development of haematopoietic tissue at both poles.

6E) Multiplex immunofluorescence showing two $CD34^{+}CD117^{lo/+}$ HSC/progenitor cells (red arrow) in proximity to $VEGFR2^{+}$ endothelium, considered sinusoidal as it is a linear structure one EC thick (green arrow).

6F) Quantitation of mean distance per region of interest (ROI) between CD34+CD117$^{lo/+}$ HSC/progenitor cells and VEGFR2+ sinusoidal ECs (green circle) or CD163+ macrophages (blue square). CD34-CD117++ mast cells were discerned by the panel of markers and are included for comparison. Bars show mean ∓ SEM euclidean distance between cells in 20x tiled images in k=177 ROIs. (**Supplementary Table 38)**. Statistically significant (Wilcoxon signed rank test) distribution of HSC proximity to each cell state compared to background is denoted by a *. See Immunofluorescence microscopy section in Methods.

6G) Violin plot of GEX in DS and non-DS fetal BM HSC and ECs. Genes shown have a significant receptor-ligand interaction in non-DS fetal BM, as predicted by CellPhoneDB analysis and detailed in **Fig. 6C**. Genes that are differentially expressed between DS and non-DS fetal BM are highlighted with an asterisk (subject to Wilcoxon rank-sum test with Benjamini-hochberg correction for multiple testing; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). Full DEG lists provided in **Supplementary Table 39**.

6H) Heatmap of differentially enriched inflammatory and cytokine production pathways in DS vs non-DS non-haematopoietic cell groupings. Computed from DEGs between DS and non-DS non-haematopoietic cell groupings weighted by fold change. See Methods for further details and **Supplementary Table 40** for the full list of genes and pathways.

Extended Data Figure 8: Haematopoietic niche in fetal bone marrow

8A) Left panel: Dot plot showing age-distribution of fetal BM samples and grouping into 4 stages (n=9). Right panel: Stacked bar plot showing frequency of non-haematopoietic cell states across stages. Statistical significance of cell frequency change shown in parentheses (Quasi binomial regression with bootstrap correction for sort gates; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, details in **Supplementary Table 36**).

8B) Identification of 4 distinct EC states in fetal BM shown by dot plot of defining marker genes. Log-transformed, normalised and scaled GEX value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

8C) Identification of distinct osteoclast and macrophage cell states in fetal BM shown by dot plot of defining marker genes. Log-transformed, normalised and scaled GEX value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

8D) Identification of osteochondral lineage states in fetal BM shown by dot plot of defining marker genes. Log-transformed, normalised and scaled GEX value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

8E) Heat map showing transcriptional similarity between post-natal mouse BM stromal cells[48] and non-haematopoietic cells in fetal BM. Gene enrichment scores are based on gene

signatures pooled from top 100 statistically significant DEGs for stromal cell populations in mouse BM[48].

8F) FDG visualisation of osteochondral lineage fetal BM cells (n=1,723). Cell type is represented by colour, as shown in legend.

8G) Left panel: stromal ligands signalling to HSC receptors in fetal BM 10x data (as per Venn diagram in **Fig. 6C, Supplementary Table 37**). Violin plots showing expression of receptor-ligand pairs predicted by CellPhoneDB to have significant interactions in fetal BM 10x data. Log-transformed, normalised and scaled GEX value is represented on the y-axis. Colours represent grouping of stromal cell types, as shown in **Fig. 6C**. Right panel: Protein level expression of HSC receptors in $CD34^+$ CITE-seq fetal BM HSCs shown by dot plot. Log-transformed, normalised and scaled GEX value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

Extended Data Figure 9: Haematopoietic niche in fetal bone marrow

9A) Summary of receptor-ligand interactions predicted by CellPhoneDB where HSC ligands signal to stromal receptors (**Supplementary Table 37**). Cells are grouped into arteriolar niche (EC-sinusoidal, EC-proliferating, EC-tip, EC-arteriolar, Fb-arteriolar, adipo-CAR) and endosteal niche (Fb-endosteal, Osteochondral precursor, Early osteoblast). Significant putative receptor-ligands across fetal BM niches are indicated in Venn diagram overlapping regions.

9B) Dot plots showing expression of receptor-ligand pair genes predicted by CellPhoneDB, as listed in **Extended Data Fig. 9A** (**Supplementary Table 37**). Log-transformed, normalised and scaled GEX value is represented by the colour of the dot. Percentage of cells in each cell type expressing the marker is shown by the size of the dot.

9C) Cross section of fetal bone stained with hematoxylin and eosin and imaged at 10x magnification to demonstrate haematopoietic architecture.

9D) Cross section of fetal bone stained with hematoxylin and eosin and imaged at 50x magnification to demonstrate the more abundant, morphologically distinct cell types predicted by 10x scRNA-seq annotations.

9E) Representative single stains from multiplex IF showing $CD163^{+}$ macrophages; $CD117^{hi}$ mast cells and $CD117^{dim}$ HSCs; $VEGFR2^{+}$ sinusoidal ECs; $CD34^{+}$ ECs and HSCs; and $CXCL12^{+}$ CAR cells.

9F) Representative images from two multiplex IF panels imaged at 10x magnification, with cells coloured according to the legend.

Methods

Sample preparation

Fetal bone marrow tissue acquisition

Human developmental tissues were obtained from the MRC–Wellcome Trust-funded Human Developmental Biology Resource (HDBR; http://www.hdbr.org) with written consent and approval from the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee (08/H0906/21+5). HDBR is regulated by the UK Human Tissue Authority (HTA; www.hta.gov.uk) and operates in accordance with the relevant HTA Codes of Practice.

Fetal developmental stage assignment and chromosomal assessment

Developmental age was estimated from standardized measurements of foot length and heel-to-knee length[56]. Quantitative Fluorescence PCR of chromosomes X, Y, 13, 15, 16, 18, 21 and 22 was performed on fetal skin or chorionic villi to assign gender and exclude common chromosomal abnormalities. In DS samples, *GATA1* mutation was excluded as previously described[57].

Dissociation of fetal bone marrow tissue

Adherent material was removed from fetal femur and bone was cut into small pieces before grinding with a pestle and mortar. Flow buffer (PBS containing 5% (v/v) FBS and 2 mM EDTA) was added to reduce clumping. The suspension was filtered with a 70μm filter then centrifuged for 5 min at 500g. The supernatant was removed before cells were treated with 1x RBC lysis buffer (eBioscience) for 5 min at room temperature and washed once with Flow Buffer before counting.

Flow cytometry and FACS for scRNA-seq

Up to 1 million cells were stained with antibody cocktail, incubated for 30 minutes on ice, washed with flow buffer and resuspended at 10 million cells per ml, with DAPI (Sigma-Aldrich) added to a final concentration of 3μM immediately before FACS (**Supplementary Table 41**). FACS was performed on a BD FACSAria Fusion instrument running DIVA v.8 to formulate and execute sort decisions, and data were analysed post-sorting using FlowJo (v.10.6.2, BD Biosciences). For 10x sequencing, cells were sorted into 500μl PBS in pre-chilled FACS tubes coated with FBS (Thermo Scientific). For Smart-seq2, index sorting was used to isolate single cells into 96-well LoBind plates (Eppendorf) containing 10μl lysis buffer (TCL (Qiagen) + 1% (v/v) β-mercaptoethanol) per well. Plates were centrifuged at 300g for 10 seconds and snap-frozen on dry ice for storage at -80° until further processing.

10x scRNA-seq

Single cell suspensions of fetal BM (n=9; 12, 13+6, 14+3, 15, 15, 16+2, 16, 17 and 19 PCW) were prepared for FACS-isolation, as above (**Supplementary Table 41**). Live, single, $CD45^+$ and $CD45^-$ fractions from each sample were sorted and manually counted, then 7,000 cells were added to each channel of a Single Cell Chip before loading onto the 10x Chromium Controller (10x Genomics). Reverse transcription, cDNA amplification and sequencing libraries were generated using either the Single Cell 3' v2 or Single Cell 5' with V(D)J Reagent kits (10x Genomics) as per the manufacturer's protocol. Libraries were sequenced using an Illumina HiSeq 4000 with v.4 SBS chemistry. For the gene expression libraries the following parameters were used: Read 1: 26 cycles, i7 index: 8 cycles, i5 index: 0 cycles, Read 2: 98 cycles. For the V(D)J libraries the following parameters were used: Read

1: 150 cycles, Read 2: 150 cycles. All libraries were sequenced to achieve a minimum of 50,000 reads per cell.

Plate-based scRNA-seq

Two fetal BM suspensions (both 17 PCW) were prepared for FACS sorting as described above (antibody details in **Supplementary Table 41**). Target populations were gated as shown in **Extended Data Fig. 2B**. As CD79 is weakly expressed on the cell surface, a significant number of B cells fall in the HLA-DR^{+}CD79^{-} gate, forming the CD14^{-}CD204^{-}CLEC9A^{-}CD1c^{-} population, thus do not enter subsequent sort gates. The number of sorted cells per subset were: 68 neutrophils, 68 myelocytes/metamyelocytes, 56 promyelocytes, 60 pDCs, 72 HSC/MPPs, 60 pDCs, 72 B cells, 70 eosinophils, 70 mast cells, 72 monocytes, 70 DC1 and 70 DC2. Plates containing lysed single cells were processed using a modified Smart-seq2 protocol[24]. Libraries were generated using the Nextera XT kit (Illumina) with 384 cells per library. Cells were barcoded using Index v.2 sets A, B, C and D (Illumina). Libraries were sequenced using an Illumina NextSeq 550 on High-output mode to achieve a minimum of 1 million reads per cell.

Cytospins

From the 2 fetal BM suspensions prepared for plate-based RNA-seq, target populations were sorted into FACS tubes containing chilled PBS. Slides were prepared using a Thermo Cytospin 4 cytocentrifuge and ShandonTM coated slides (Thermo, 5991059), dried at room temperature, then fixed with ice-cold methanol and stained using Giemsa (Sigma-Aldrich), according to manufacturer's instructions. Slides were viewed using a Zeiss AxioImager microscope, images taken of 4 fields from n=3 samples using the 100X objective, and viewed using Zen (v.2.3) as previously described[4].

CITE-seq

Cryopreserved fetal BM (n=3, 14-17 PCW), FL (n=4, 14-17 PCW) and CB cells (n=4, 40-42 PCW) were thawed on the day of experiment and added to pre-warmed RF-10 (RPMI (Sigma-Aldrich) supplemented with 10% (v/v) heat-inactivated fetal bovine serum, FBS (Gibco), 100 U ml−1 penicillin (Sigma-Aldrich), 0.1 mg ml−1 streptomycin (Sigma-Aldrich), and 2 mM L-glutamine (Sigma-Aldrich)) dropwise whilst swirling. Cells were centrifuged at 500 g for 5 mins before counting. Pools of cells were made if cell numbers were low. Cells were then stained with Fc receptor blocking reagent (Biolegend) and CD34 APC/Cy-7 (Biolegend) for 10 mins in the dark and on ice. During the incubation, the CITE-seq antibody vial was centrifuged at 14,000 g for 1 min then reconstituted with Flow buffer. The vial was incubated for 5 mins at room temperature then centrifuged at 14,000 g for 10 mins at 4 °C. The CITE-seq antibody cocktail (**Supplementary Table 42**) was then added to the cells along with a competition antibody mix (**Supplementary Table 41**) and incubated for 30 mins in the dark and on ice. The stained cells were then washed with Flow buffer, before resuspension in Flow buffer supplemented with 3 mM DAPI (Sigma Aldrich).

As many as possible live, single, $CD34^+$ and $CD34^-$ cells were sorted by FACS into 500µl PBS in pre-chilled FACS tubes coated with FBS. Sorted cells were then centrifuged at 500g for 5 mins before manually counting. Cells were then submitted to the CRUK CI Genomics Core Facility for 10x Chromium loading, library preparation and sequencing. Single cell 3' version 3 (10x Genomics) kits were used and gene expression and cell surface protein libraries were generated as per manufacturer's protocols. Libraries were sequenced using a NovaSeq (Illumina) to achieve a minimum of 20,000 reads per cell for gene expression and 5,000 reads per cell for cell surface protein.

Culture experiments

Single cell culture experiments on MS5 were performed on paired fetal BM and FL samples (n = 3; 15, 17, 17 PCW), and four additional fetal BM samples (n = 4; 14, 15, 16, 17 PCW). Cryopreserved single cell suspensions were thawed and sorted into HSC/MPP, MLP and $CD34^+CD38^{mid}$ fractions as previously described[4] (**Supplementary Table 41**, **Extended Data Figs. 6B, C**). Single cells were index-sorted into 96 well plates containing MS5 in log phase growth (DSMZ, passage 6–10), using culture conditions as previously described. The proportion of wells producing colonies was not correlated with gestational age. Single cell colonies were isolated after 14 days, and prepared for flow cytometry as above. Erythroid colonies were identified as $CD45^-GYPA^+$ ≥ 30 cells, megakaryocyte colonies as $CD41^+$ ≥ 30 cells, myeloid colonies as [($CD45^+CD14^+$) + ($CD45^+CD15^+$)] ≥ 30 cells, NK colonies as $CD45^+CD56^+$ ≥ 30 cells. Two-tailed Fisher's exact tests, performed in Prism (v.8.1.2, GraphPad Software), were applied to the numbers of colonies of each type by stage to determine statistical significance in lineage differentiation potential with development.

Comparisons of DS and non-DS HSC/MPP cultured on an MS5 stromal layer, from paired fetal BM and FL samples all at 12-13 PCW (DS n=2 each; k=740, FL, k=280 BM; non-DS n=2 each; k=840 FL, k=500, BM), were performed with 20 cells per well, and analysed after 21 days. This number of cells per well was scaled to 20 and the duration of culture increased to 21 days to allow maximum opportunity for colonies to arise from DS fetal BM. To avoid nutrient and growth factor depletion, 50% of the initial culture media volume was added to wells each week.

Single cell methylcellulose cultures were performed on HSC/MPP from DS (n=2; PCW=17, 19; k=246) and non-DS fetal BM (n=3; PCW=17, 19, 21; k=365) . Methocult (Stem Cell Technologies, H4230) was supplemented with 20% IMDM (Gibco) including 1ng/ml Pen/Strep (Gibco), and human cytokines: 10ng/mL IL-6, IL-11, SCF, Flt3L, 50ng/mL GM-CSF and TPO, 20ng/mL IL-3 (all PeproTech), and 4 U/ml EPO (EPREX/Janssen) as described in Roy et al, PNAS 2012[11]. Human fetal BM HSPC subpopulations were index-sorted as viable single cells via FACS directly into 96 well TC plates (Sigma) containing supplemented methocult (50ul/well). Sorting was performed on a BD FACSAria Fusion Cell Sorter (BD Biosciences). Plates were then incubated at 37ºC and 5% CO2 for 14 days, colony readout was performed at D14 and imaged for morphological characterisation using an EVOS XL Core Imaging System microscope.

Immunofluorescence microscopy

Specimens of FFPE tissue were decalcified in EDTA. 4μm sections were applied to adhesive slides (Trajan-3). Automated staining was performed on a Vectra Automated Multispectral Imaging System (Antibodies listed in **Supplementary Table 41**). Images at 20X magnification were spectrally unmixed and analysed using supervised machine learning algorithms within Inform v.2.4.8 (Akoya). HSCs were inadequately segmented by this process, so were manually annotated in QuPath v.0.2.3. Using cartesian coordinates for each phenotyped cell, mean Euclidean distances of HSCs to $CD163^{+}$ macrophages and $VEGFR2^{+}CD34^{+}$ CAR were calculated. Mean euclidean distances were calculated between centroid coordinates using the scipy (v1.6) *cdist* function from within 158 sampled regions (ROI) acquired from two separate experiments (88 and 85 ROI respectively). Mean euclidean distances between HSCs and cell states of interest per ROI were compared to distances between the same HSCs and all other cells in the ROI. The scipy implementation of

the Wilcoxon signed rank test was applied in this comparison to determine if the proximity distribution of HSCs to cells of interest varied significantly above random.

Data analysis

Alignment, quantification and quality control of scRNA-seq

The fetal BM scRNA-seq datasets described in this study (non-DS and DS) underwent pre-processing as detailed below. 10x droplet-based sequencing data was quantified with the CellRanger Single Cell Software Suite (10x Genomics, Inc) and aligned to a GRCh38 human reference genome (see details in **Supplementary Table 43**). Smart-seq2 sequencing data was aligned with STAR (v2.7.3a) using the STAR index and aligned to the GRCh38 human reference genome. Gene-specific read counts for Smart-seq2 data were calculated using HTSeq-count (v0.10.0). Cells with fewer than 200 detected genes, genes expressed in fewer than 3 cells, and total mitochondrial gene expression exceeding 20% were removed from downstream analysis. The methodology for incorporation of external datasets (including: YS, FL, adult BM, CB, blood, thymus, mouse BM) can be found in *Statistics and Reproducibility*, with methods used for any re-annotation described in *Dimensional reduction, visualisation and clustering.*

Alignment, quantification and quality control of CITE-seq datasets

The fetal BM CITE-seq (including: CD34+, total) and FL/CB CD34+ CITE-seq datasets described in this study underwent pre-processing as detailed below. CITE-seq transcriptomic data was quantified with the CellRanger Single Cell Software Suite (10x Genomics, Inc) and aligned to a GRCh38 human reference genome, with cell surface protein data quantified using CITE-seq-Count (see details in **Supplementary Table 43**). CITE-seq cells from pooled

lanes were demultiplexed using Souporcell, using the singularity container provided by the authors of Souporcell on GitHub for reproducibility and with CellRanger human reference genome GRCh38 v3.0.0 for alignment. For the fetal BM CD34+ samples, the common variant option was also called (using the 'common_variants_grch38.vcf' file provided by Souporcell) in order to solve complex demultiplexing for these samples.

Cells in CITE-seq transcriptome data with fewer than 200 detected genes, genes expressed in fewer than 3 cells, and total mitochondrial gene expression exceeding 20% were removed from downstream analysis. Cells in CITE-seq cell surface protein data with total counts higher than 5,000, or fewer than 30 antibodies, or antibodies expressed in fewer than 3 cells were removed from downstream analysis.

Doublet exclusion and transformation of GEX and antibody matrices

We ran Scrublet (v0.2.1) on each 10x and CITE-seq RNA lane independently, obtaining per-cell scrublet scores. A doublet exclusion threshold of the median plus three times the median absolute deviation scrublet score was applied, as described previously[4].

Raw gene expression matrices from RNA lanes (including those from both 10x and CITE-seq experiments) underwent correction for cell-to-cell variation via normalisation, using the *normalize_per_cell* function in Scanpy (v1.4.4). Data were then transformed using the *log1p* function in Scanpy to alleviate skewness of data and mean-variance relationship. Expression values of each gene were then scaled and centred using the *scale* function in Scanpy. Highly variable genes (HVGs) were detected using the *highly_variable_genes* function in Scanpy, with minimum mean variance, maximum mean variance and minimum dispersion set as 0.0125, 3 and 0.5, respectively.

Raw protein count matrices (from CITE-seq experiments) underwent correction for cell-to-cell variation via normalisation using the DSB (Denonised and Scaled by Background) algorithm[58]. DSB normalisation quantified protein counts above background levels in individual cells, and allowed for improved interpretability of protein expression.

Dimensional reduction, visualisation, batch correction and clustering

Principal components were calculated using the *pca* function in Scanpy. Dependent upon the plateau observed in the elbow curve from the *pca_variance_ratio* function in Scanpy, an informative number of PCs were selected for downstream analysis.

Principal components were adjusted for sequencing type variation (i.e., 3' and 5' sequencing platforms), biological replicate, or tissue type using the Harmony package for batch correction (v1.0). Harmony parameter optimisation was performed using our novel fetal BM scRNA-seq as validation dataset and iterating through theta values 1-10. For each value of theta, we took the data substructure and technical variations (such as those introduced by FACs sorting) into consideration by calculating the mean kBET rejection rate and silhouette scores for all cell types post-Harmony correction (kBET package, v0.99.6, sklearn package v0.22). Optimum theta selection was then determined by the observation of lowest kBET rejection score vs lowest silhouette score; theta was then set to 3 for Harmony batch correction.

The *neighbours* function in Scanpy was used to calculate the neighbourhood graph. Uniform manifold approximation and projection (UMAP) embedding was calculated using the *umap* function in Scanpy. The neighbourhood graph was then clustered using the *leiden* function in

Scanpy. Gene expression dot plots were visualised using the Scanpy package in Python with dot colour indicating the mean normalised, logged and scaled expression values and dot size representing the proportion of cells annotated to each category that express the given gene. When upper limits were placed on either expression or population in gene expression dot plots, this is indicated by a '≥' in the given figure legend.

Annotation of clusters

Cluster cell identity was assigned through analysis of DEGs and their alignment with marker genes identified through literature search (**Supplementary Table 2**). DEGs were calculated in Scanpy using the *rank_genes_groups* function, which performed a Wilcoxon rank sum test restricted 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. All p-values were adjusted for multiple testing using the Benjamini-Hochberg method. Following initial broad rounds of annotations, clusters of broad similarity (e.g., lymphoid cells) were subset for further rounds of feature selection, visualisation, clustering and annotation as described above. Clusters whose gene signatures indicated additional diversity were further investigated in an iterative manner, and those with unique signatures were selected for downstream analysis.

Logistic regression for label transfer of annotations

Label transfers and probability of label correspondence between datasets were carried out using a Logistic Regression (LR) model trained against the fetal BM data. The LR model was built utilising the 'sklearn.linear_model.LogisticRegression' module in the sklearn package (v0.22). The LR model was trained on normalized gene expression data (x_var =33,712) and annotations (x_sample=64) of the whole fetal BM dataset (103,228 cells), and used to predict the probability of correspondence between labels in the target dataset. The model used was

L2 Ridge Regression regularised with penalty strength of 0.2 using the 'lbfgs' solver. Predicted label probability distribution within pre-computed clusters in the target data were used to assign cluster identity by majority vote. We further assessed the predicted labels by computing Adjusted Rand index and Mutual information scores from the modules 'sklearn.metrics.adjusted_rand_score' and 'sklearn.metrics.mutual_info_score' between original cell labels and predicted comparative cluster labels in overlapping cell states in each dataset.

Comparisons between scRNA-seq datasets

To compare gene expression programs across tissue compartments (and scRNA-seq datasets), we combined raw GEX matrices and performed joint matrix transformation and dimension reduction as previously described in Methods. Datasets were integrated using Harmony with tissue source as a covariate. DEG analysis was performed as described in 'Annotation of clusters', unless otherwise stated in tool-specific description in Methods (using Wilcoxon rank sum testing and Benjamini-Hochberg for multiple testing adjustment).

Kernel density estimation of cellular abundance

Gaussian kernel density functions (scipy v1.4.1) were estimated in the FDG space (bandwidth of 0.1) separately using 7,500 sampled cells from each tissue-specific dataset. To visualise cell densities on the common landscape, values were computed for all cells with respective kernel density functions.

Calculating differences in cell type proportions across gestational stages and organ

GraphPad Prism (v8.1.0) was used for plotting and statistical comparison. Statistically significant differences in cell type proportions across tissue were conducted using a one-way

ANOVA with Tukey's multiple comparison tests. Significance was noted on corresponding scatter plots using asterisks, where scatterplots display proportion per biological replicate. Cell-type proportions per sample were obtained by adjusting observed proportions by $CD45^{+}/CD45^{-}$ sort gate. For cell type proportion statistical analysis across gestational stage and tissue, proportions were modelled as a quasibinomial distribution. For both analyses the condition (gestational stage or tissue) was provided as a covariate for the proportion of the cell type being assessed. The quasibinomial model was fit using glm from the MASS R package. The p-value for the significance of the change in proportion between conditions was assessed using a likelihood-ratio test, computed using the anova function. Significant changes in cell type proportion were highlighted on bar plots using asterisks. The Spearman's rho test was used to assess monotonic trends of flux between analogous cell states of different developmental stages. Increasing or decreasing trends were denoted with 'up' and 'down' arrows respectively.

Differential abundance testing using Milo

To identify cell subpopulations enriched or depleted in DS vs disomic fetal BM samples, we used Milo (development version) for differential abundance testing on KNN graph neighbourhoods. This was implemented in the R package miloR (https://github.com/MarioniLab/miloR). Briefly, we performed KNN graph embedding on DS and disomic fetal BM samples at matched age (12-13 PCW) using the reduced dimension space derived from Harmony integration (using the same number of dimensions and value of K used for UMAP embedding). We use a refined sampling algorithm to select a subset of cells spanning the KNN graph (defined as index cells) and we count the number of cells from each sample in the neighbourhoods of index cells, where a neighbourhood is defined as the group of cells connected by an edge to an index cell. We test for differences in abundance

between the cells from DS and non-DS samples in each neighbourhood using the Quasi-likelihood test implemented by edgeR, controlling the FDR across the graph neighbourhoods. To assign cell type annotations to neighbourhoods, we take the most frequent annotation between cells in each neighbourhood. Neighbourhoods are generally homogeneous, retaining neighbourhoods with at least 60% of cells belonging to the most abundant cell type label.

Gene enrichment scores

To conduct gene enrichment scores against a reference published blood dataset[24], the top 100 DEGs (log2 fold change) of blood DC and monocyte cell types were input into *sc.tl.score_genes* function in Scanpy. Gene enrichment value for the blood reference cell type was then calculated as the average expression of the top DEG from the reference dataset, minus the average expression of another reference set of genes (randomly sampled from each binned expression value). Gene enrichment scores were visualised using a heatmap in the seaborn (v.0.9.0) package.

Cell cycle gene enrichment scores were calculated through use of a publicly available curated list of genes implicated in the human cell cycle[59] as the input for *sc.tl.score_genes_cell* function in Scanpy. The G2/M and S phase score for each cell thus represented low to high enrichment for a particular phase's genes. In order to serve as a proxy for a 'proliferative phase' score, the mean of the G2/M and S phase scores were calculated and plotted in UMAP space. Cells with G2/M/S cell cycle score greater than the mean were assigned as 'cycling' cells, else assigned as 'not cycling'.

Gene enrichment scores calculated using *sc.tl.score_genes* function in Scanpy were used to ascertain (1) apoptotic gene enrichment through use of genes implicated in the KEGG apoptotic pathway (GSEA: M8492), (2) NK cytotoxic gene enrichment through use of genes implicated in the KEGG NK cytotoxicity pathway (GSEA: M5669), and (3) TNF response gene enrichments using genes from the GO biological process database (GO:0034612).

Direction of transmission (DoT)-score analysis

To define the suitable origin for the DoT-score method, we projected HSC/MPP I cells from each tissue (fetal BM, fetal liver or cord blood) onto the adult BM (ABM) scRNA-seq landscape. Both fetal and adult data were scaled together and fetal cells from each tissue were projected onto 50 PC vectors computed using the ABM data. For each fetal cell, 15 nearest neighbours in the ABM data were identified. The normalised sum of neighbours identified per cell in the ABM served as a similarity score. We used all cells with similarity scores >0.05 (around 60 cells for each tissue) to compute the average gene expression, subsequently used as the origin points for the DoT-score analysis.

DoT-score was computed as previously described[32], using the dotscore package publicly available at https://github.com/Iwo-K/dotscore. Weights were derived from genes with significant differential expression between: HSC/MPP I cells from FL compared to fetal BM (FDR < 0.05, abs($\log_2$(Fold Change) > 0.5); fetal BM compared to combined CB and FL cells (FDR < 0.05, abs($\log_2$(Fold Change) > 1); CB compared to combined fetal BM and FL cells (FDR < 0.05, abs($\log_2$(Fold Change) > 1. For clarity, colour-scales in DoT-score plots are clipped at 0.1% and 99.9% percentiles to avoid plotting extreme outliers.

VDJ analysis

Using the pyVDJ Python package (v0.1.2), lanes of BCR-enriched and TCR-enriched 10x data were integrated with their corresponding 10x GEX lane data in Scanpy. Filtered CellRanger output files were then imported into the Scanpy workflow to investigate productivity of chains, presence of heavy and light chains and clonal assignment. VDJ metadata by cell type was then exported from Scanpy and plotted in GraphPad Prism.

Trajectory inference using Monocle3

GEX matrices for cell-types of interest (filtered by highly variable genes, as defined in Scanpy) were loaded into the Monocle workflow as CellDataSet objects using Monocle3 (v0.2.1). GEX values were then normalised by log and size factor to address depth differences using *preprocess_cds* function. For known lineages, cells were clustered with a resolution parameter of 1e-07 in order that one partition was returned (to ensure pseudotime with incorporated all cells). Cells were then ordered along pseudotime and with root state provided using the *order_cells* function. DEGs across pseudotime were calculated using a Moran's I statistical test (*graph_test* function) and DEGs grouped into 'modules' by their Moran's-derived correlation across pseudotime using the *find_gene_modules* function. Dynamically expressed genes across a given pseudotime were then plotted as heatmaps, with normalised logged and scaled gene expression values. Paired heat maps across conditions were the product of combined processing (log, normalising, scaling) of GEX counts and plotting gene expression over independently derived pseudotime trajectories.

Trajectory inference using Waddington-OT

The Waddington-OT 'wot' package (v.1.0.8) was used to calculate differentiation trajectories, or, transitions between cell states in the fetal BM scRNA-seq data (code closely following the author's GitHub documentation). Transport matrices were calculated for the selected

monocyte-neutrophil lineage cells between 12-19 PCW. Using transport matrices as input, the probability that a cell at each PCW interval would transition into either a monocyte or neutrophil fate was then saved as fate matrices. Fate matrices were visualised in triangles, with two vertices representing the pure cell fates of a bifurcating differentiation trajectory, and the third fate/vertex representing 'other'.

Prediction of cell-cell communication using CellPhoneDB

To assign putative cell-cell interactions within our fetal BM 10x GEX scRNA-seq dataset, we used CellPhoneDB (v2.1.2). Log, normalised and scaled expression values for cell types of interest were exported from Scanpy along with their respective cell type metadata. Using the receptor-ligand database (v2.0.0), CellPhoneDB was run using the statistical method, with *p*-value cut-off of 0.05 for significant receptor ligand pairs and a result precision of 3dp. To visualise spatiality of non-DS fetal BM niche interactions via Venn diagram, non-haematopoietic stromal cells were grouped into: arteriolar niche (EC-sinusoidal, EC-proliferating, EC-tip, EC-arteriolar), endosteal niche (Fb-endosteal, Osteochondral precursor, Early osteoblast) and stromal niche (Mac-stromal, Fb-fibroblast). Total cell numbers of interest were included in analysis unless down-sampling stated in *Statistics and reproducibility.*

Inference of transcription factors and their gene regulatory networks using PySCENIC

The PySCENIC package (v0.9.19) and pipeline was used to identify transcription factors and their target genes in the combined non-DS and DS fetal BM datasets. The ranking database (hg38__refseq-r80__500bp_up_and_100bp_down_tss.mc9nr.feather), motif annotation database (motifs-v9-nr.hgnc-m0.001-o0.0.tbl) and list of transcription factors (lambert2018.txt) were downloaded from the Aert's laboratory github page. An adjacency

matrix of transcription factors and their targets was generated and pruned using the Aert's group suggested parameters. Comparisons of TF activity between non-DS and DS fetal BM were made using t-tests of the AUCell output of predicted transcription factor activity in each cell for each cluster. The regulons generated were used to predict which genes controlled by each transcription factor in downstream analysis.

Network analysis and clustered pathway annotation

The FindConservedMarkers function in Seurat (v3.1) (Bonferroni-corrected FDR adjusted *p*-values < 0.05) and a Benjamini-Hochberg-corrected Wilcoxon rank-sum test (log fold change > 0.25 and *p*-values < 0.05) were used to identify conserved and differentially expressed genes between analogous cellstates in each dataset. Genes were submitted for over-representation analysis (ORA) using the G-profiler2 package in R to query two databases simultaneously (Reactome, Gene Ontology (GO) Biological Process). We derived statistical significance (q < 0.05) for each gene set enrichment and performed Markov clustering (MCL) using the MCL package in R to derive network neighbourhoods based on shared genes between the gene sets. The gene set clusters were annotated using the AutoAnnotate Cytoscape package. Clusters were ranked by the mean enrichment score of all gene sets within each cluster and manually curated based on biological significance. We used Cytoscape to visualise clusters of enriched gene sets. Transcription factor regulation of each markov cluster was identified by hypergeometric modelling of genes in the clustered genesets to a pre-compiled TF association matrix acquired from the Enrichr database of ENCODE and ChEA consensus TF associations using the Hypergeometric package in R, over-represented TFs were then ranked and filtered by *p*-value (<0.05). To characterise inflammatory response pathways associated with the stroma in DS fetal BM, we derived DEGs between analogous stromal compartments in DS and non-DS and submitted the genes for network analysis as

outlined prior. Fold changes for genes associated with the top five Markov clusters ranked by cluster size were visualised with the ggplot2 (v3.3.2) package in R.

TNF response gene annotation

We derived DEGs between analogous cellstate compartments in disomic and trisomy 21 fetal BM, statistically significant DEGs (p-value < 0.05) were compared for intersect against TNF response associated genes acquired from the GO biological process database (GO:0034612). Intersecting TNF response genes in DEGs were ranked by log fold change between cellstates of DS and non-DS. The full list of intersecting TNF response genes are shown in (**Supplementary Table 40**).

Statistics and reproducibility

Fetal liver, fetal yolk sac and thymus 10x data

For all re-analysis of FL and YS 10x data (n=113,063 and 10,071), n=14 and 3 biologically independent samples were used and total population of annotated cell types were shown in figures unless otherwise stated. Original cell numbers can be found in the original publication[4].

For **Fig. 2B**, proportions of myeloid cell states arising from BM haematopoiesis and their counterparts in other tissues were compared. The YS cell state originally assigned as DC progenitor in Popescu et al[4] was renamed as macrophage in line with further GEX exploration and re-annotation, and therefore included in **Fig. 2B** but not in **Extended Data Fig. 3F**. Further YS progenitor (HSC, ELP, CMP, MEMP, MEP) and myeloid (GMP, promonocyte, MOP, macrophage) cell states were identified upon re-clustering and

re-annotating the YS Lymphoid progenitor, MEMP, Myeloid progenitor and YS progenitor/MPP (metadata for reannotation available in **Supplementary Table 44**).

For **Fig. 2B**, the original FL Monocyte precursor and Neutrophil-myeloid progenitor were subclustered to resolve heterogeneity and revealed further myeloid states including MOP, monocyte, promonocyte and promyelocyte cells (metadata for reannotation available in **Supplementary Table 16**). For **Fig. 2B** and **Extended Data Fig. 3F,** FL myeloid nomenclature was updated such that monocyte precursor became promonocyte, and monocyte became $CD14^{+}$ monocyte. Due to the presence of a distinctive pDC precursor in FL, pDC and pDC precursor were merged into one 'pDC' grouping for purposes of the cross-tissue bar plot.

In **Fig. 6B,** FL endothelial populations were re-clustered to annotate sinusoidal endothelium, with metadata for reannotation available in **Supplementary Table 16**. For all analysis of thymus 10x data (n=259,265), n=24 biologically independent samples were used and total population of annotated cell types were shown in figures, with exception of down-sampling in **Fig. 3G**, where 1,000 thymus DN cells were used for analysis. Cell numbers can be found in the original publication[3,4].

Fetal bone marrow 10x data

For analysis of combined fetal BM 10x 5GEX and 3GEX data (n=103,228), n=9 biologically independent samples were used. Total population of annotated cell types were shown in figures unless otherwise noted.

The following refined cell number annotations were displayed in each of the figures (annotations available in **Supplementary Table 4**): CD4 T cell - 327, CD8 T cell - 171, CD14 monocyte - 8763, CD56 bright NK - 449, CMP- 425, DC1 - 50, DC2 - 598, DC3 - 705, DC precursor - 201, erythroid macrophage - 92, ELP - 1357, GMP - 1281, HSC - 92, ILC precursor - 67, LMPP - 34, MEMP - 16, MEP - 269, MK - 1000, MOP - 3838, MPP myeloid - 92, NK T cell - 111, NK progenitor - 26, Treg - 62, adipo-CAR - 353, arteriolar fibroblast - 83, basophil - 139, chondrocyte - 80, early MK - 1624, early erythroid - 7474, early osteoblast - 280, endosteal fibroblast - 54, eo/baso/mast precursor - 175, eosinophil - 321, erythroid macrophage - 92, immature B cell - 1998, immature EC - 42, late erythroid - 4636, mast cell - 648, mature NK - 136, mid erythroid - 14297, monocytoid macrophage - 290, muscle - 131, muscle stem cell - 254, myelocyte - 3794, myeloid DC progenitor - 31, myofibroblast - 78, naive B cell - 1411, neutrophil - 4501, osteoblast - 363, osteoblast precursor - 456, osteochondral precursor - 191, osteoclast - 1221, pDC - 712, pDC progenitor - 23, pre B progenitor - 14229, pre pro B progenitor - 5427, proliferating EC - 26, promonocyte - 7437, promyelocyte - 2191, schwann cells - 9, sinusoidal EC - 550, stromal macrophage -1464, tDC - 193, tip EC - 362, pro B progenitor - 5528.

When cell types were grouped into broad lineages (e.g., **Figs. 1B, C**), cell numbers were as follows: HSC_MPP - 3795, erythroid - 26407, MK - 2624, B_lineage - 28583, DC - 2459, eo/baso/mast - 1108, neutrophil - 10486, monocyte - 20038, T_NK - 1349, stroma - 6379. Other broad groupings used are detailed in **Supplementary Table 3**.

In **Fig. 3B**, only B cells with corresponding VDJ data (5,052/28,583) were shown in bar plot. In **Extended Data Fig. 4F**, only single positive T cells with corresponding VDJ data (138/560) were shown in barplots. Refined annotations were merged in the case of **Extended**

Data Fig. 7F, where non-DS fetal BM MEP and MEMP were merged and non-DS fetal BM cells were downsampled to only 12-13 PCW (at matched age with DS fetal BM). For inference of cellular interactions in **Fig. 5G**, significant DEGs between DS and non-DS BM cells (as shown in **Supplementary Table 34**) were filtered for TNF superfamily genes which were overrepresented, then CellphoneDB was used to predict interactions. In **Fig. 6A**, muscle stem cell and muscle were merged into muscle lineage. In **Fig. 6G**, fetal BM tip, sinusoidal and proliferating EC were merged into 'endothelial' cells.

Fetal bone marrow Smart-seq2 data

For analysis of fetal BM Smart-seq2 validation data (n=486), n=2 biologically independent samples were used and the following cell numbers were shown: $CD34^+$ = 32, B cell = 52, DC1 = 34, DC2 = 15, monocyte = 32, PMN = 65, basophil = 20, eosinophil = 54, mast cell = 47, myelocyte = 61, pDC = 30, promyelocyte = 44. Annotations are available in **Supplementary Table 13**.

Fetal bone marrow Down syndrome 10x data

For all analysis of fetal DS BM 10x 5GEX data (n=16,743), n=4 biologically independent samples and total cell populations were used, unless stated otherwise.

The following refined cell number annotations were displayed in each of the figures (annotations are available in **Supplementary Table 30**): CAR - 4, CD14 monocyte - 320, CD56 bright NK - 79, CD8 T cell - 181, CMP - 50, DC1 - 45, DC2 - 228, DC3 - 108, HSC - 105, ILC precursor -13, MEMP - 130, MK - 83, MOP - 422, MSC -53, Treg - 8, chondrocyte - 4, early B cell - 42, early MK - 34, early erythroid - 1,348, endothelium - 111, eo/baso/mast precursor - 53, eosinophil - 63, late erythroid - 6,336, macrophage - 113, mast cell - 66,

mature B cell - 31, mature NK - 147, mid erythroid - 5,230, myelocyte - 243, neutrophil - 273, osteoblast -11, osteoclast - 57, pDC - 14, pre B cell - 115, promonocyte - 395, pre pDC - 110, promyelocyte - 107, transitional NK cell -11.

When cell types were grouped into broad lineages (e.g., **Fig. 5A, B**), cell numbers were as follows: HSC/MPP = 338, Erythroid = 12,914, MK = 117, B lineage = 188, DC = 505, Neutrophil = 623, Eo/baso/mast = 129, Monocyte = 1,137, TNK = 439, Stroma = 353. Other broad groupings are detailed in **Supplementary Table 31**.

Adult bone marrow 10x data

For all analysis of adult BM 10x (n=142,026) data, n=4 biologically independent samples were used, unless stated otherwise. Lanes from four donors (BM1, BM2, BM5, BM6) were downloaded from the Immune Cell Atlas public repository (https://data.humancellatlas.org/).

The following refined cell number annotations were displayed in each of the figures (annotations are available in **Supplementary Table 11)**: CD14 monocyte - 3670, CD16 monocyte - 1938, CD56 bright NK - 1228, CLP - 882, CMP - 288, DC1 - 135, DC2 - 481, DC3 - 550, DC precursor - 462, HSC - 497, LMPP - 80, MEMP - 785, MK - 577, MOP - 1440, MPP - 365, Treg - 6327, early MK - 136, early erythroid - 5441, erythroid macrophage - 77, immature B cell - 2728, late erythroid -1150, mature CD8 T cell - 15725, mature NK - 6074, memory B cell - 4106, memory CD4 T cell - 22197, mid erythroid - 2192, monocyte-DC - 515, myelocyte - 6675, myeloid DC progenitor - 110, naive B cell - 19265, naive CD4 T cell - 5873, naive CD8 T cell - 8965, neutrophil - 2482, pDC - 1134, pDC progenitor - 63, plasma cell - 2074, pre B cell - 971, pro B progenitor - 1390, promonocyte -

7448, promyelocyte - 2197, stroma - 161, tDC - 75, transitional B cell - 2151, transitional NK - 946.

When cell types were grouped into broad lineages (e.g., **Extended Data Fig. 1C**), cell numbers were as follows: HSC/MPP = 3,007, Erythroid = 8,783, MK = 713, B lineage = 32,685, DC = 3,415, Neutrophil = 11,354, Monocyte = 14,496, TNK = 67,335, Stroma = 238. Other broad annotations are available in **Supplementary Table 8.**

Cord blood 10x data

For all analysis of CB 10x (n=148,442) data, n=4 biologically independent samples were used, unless stated otherwise. Lanes from four donors (CB1, CB2, CB5, CB6 respectively) were downloaded from the Immune Cell Atlas public repository (https://data.humancellatlas.org/).

The following refined cell number annotations were displayed in each of the figures (annotations are available in **Supplementary Table 10**): CD8 T cell - 16345, CD14 monocyte - 13324, CD16 monocyte - 888, CD56 bright NK - 4066, CMP - 272, DC1 - 67, DC2 -155, DC precursor - 169, GMP - 203, HSC - 194, ILC precursor -1519, MEMP - 338, MK - 1262, early MK - 496, early erythroid - 532, late erythroid - 878, mature NK - 7860, mid erythroid - 2627, myelocyte - 3726, naive B cell - 19516, naive CD4 T cell - 69338, neutrophil - 3458, pDC - 242, preDC - 269, promonocyte - 607, tDC - 91.

When cell types were grouped into broad lineages (e.g., **Extended Data Fig. 1B**), cell numbers were as follows: HSC/MPP - 1007, erythroid - 4037, MK - 1758, B cells - 19516,

DC - 993, neutrophil - 7184, monocyte - 14819, T/NK - 99128. Broad annotations are available in **Supplementary Table 9.**

CD34+ CITE-seq data

For analysis of CD34+ CITE-seq data, 10x lanes from fetal BM (n=3, k=8,829), fetal liver (n=4, k=18,904) and cord blood (n=4, k=7,540) were run using both 3GEX and ADT technology. The total population of annotated cell types were shown in figures unless otherwise noted. The following refined cell number annotations were displayed in each of the figures (annotations available in **Supplementary Table 26**): DC progenitor I_CB - 54, DC progenitor I_FBM - 247, DC progenitor I_FL - 50, DC progenitor II_CB - 76, DC progenitor II_FBM - 298, DC progenitor II_FL - 344, Early LyP_CB - 183, Early LyP_FBM - 301, Early LyP_FL - 440, EoBasoMC_CB - 153, EoBasoMC_FBM - 153, EoBasoMC_FL - 568, EryP I_CB - 224, EryP I_FBM - 172, EryP I_FL - 1,927, EryP II_CB - 10, EryP II_FBM - 70, EryP II_FL - 849, EryP III_CB - 86, EryP III_FBM - 133, EryP III_FL - 1,020, EryP IV_CB - 223, EryP IV_FBM - 319, EryP IV_FL - 1,864, HSC/MPP I_CB - 1,455, HSC/MPP I_FBM - 284, HSC/MPP I_FL - 1,699, HSC/MPP II_CB - 1,086, HSC/MPP II_FBM - 378, HSC/MPP II_FL - 1,298, HSC/MPP III_CB - 83, HSC/MPP III_FBM - 159, HSC/MPP III_FL - 544, HSC/MPP IV_CB - 1,020, HSC/MPP IV_FBM - 307, HSC/MPP IV_FL - 269, Late EryP I (Pro-erythroblast)_CB - 174, Late EryP I (Pro-erythroblast)_FBM - 229, Late EryP I (Pro-erythroblast)_FL - 1,598, Late EryP II (Erythroblast)_CB - 26, Late EryP II (Erythroblast)_FBM - 31, Late EryP II (Erythroblast)_FL - 868, LyP I (CLP)_CB - 787, LyP I (CLP)_FBM - 520, LyP I (CLP)_FL - 891, LyP II (pre pro-B)_CB - 460, LyP II (pre pro-B)_FBM - 1,499, LyP II (pre pro-B)_FL - 554, LyP III (pro-B)_CB - 3, LyP III (pro-B)_FBM - 450, LyP III (pro-B)_FL - 36, LyP IV (pre-B)_CB - 479, LyP IV (pre-B)_FBM - 777, LyP IV (pre-B)_FL - 619, MEP/MkP_CB - 348, MEP/MkP_FBM - 207,

MEP/MkP_FL - 1,184, MEP_CB - 39, MEP_FBM - 324, MEP_FL - 384, MyP_CB - 427, MyP_FBM - 951, MyP_FL - 845, Cycling LyP_CB - 144, Cycling LyP_FBM - 1,020, Cycling LyP_FL - 1,053.

Fetal bone marrow CITE-seq data

For analysis of fetal BM CITE-seq data (n=3, k=8,978), the total population of annotated cell types were shown in figures unless otherwise noted. The following refined cell number annotations were displayed in each of the figures (annotations available in **Supplementary Table 12**): basophil - 15, CD14 monocyte - 1,384, CD4 T cell - 39, CD56 bright NK - 66, CMP - 78, DC1 - 13, DC2 - 87, DC3 - 20, early erythroid - 517, early MK - 91, ELP - 177, eosinophil - 22, GMP - 108, HSC - 36, immature B cell - 403, late erythroid - 670, mast cell - 57, mid erythroid - 466, MK - 31, MOP - 280, naive B cell - 249, neutrophil - 294, osteoclast - 58, pDC - 139, pre B progenitor - 2,241, pre pro B progenitor - 248, pro B progenitor - 366, promonocyte - 620, promyelocyte - 103, sinusoidal EC - 42, stromal macrophage - 47, tip EC - 11.

Blood monocyte and DC 10x data

Monocyte-DC blood SS2 data were downloaded from a published study[24]. The available RPKM counts for 1140 monocytes (n=768) and DCs (n=372) were logged and scaled (in line with 10x analysis), in preparation for DEG analysis conducted as described below. Refined celltype population frequency can be found in original study.

Differentially expressed gene statistics

Differential gene expression analysis referenced in text and shown in violin plots were run using the Wilcoxon rank-sum statistical test with Benjamini-Hochberg procedure for multiple testing correction. *p*-values are shown in the relevant supplementary tables.

Statistics from barplots

For cell type proportion analysis across gestational stages and different tissues, proportions were modelled as a quasibinomial distribution. *p*-values for the significance of change in proportion between conditions were assessed and values subsequently detailed in **Supplementary Tables 7, 15, 21, 27, 36**. For all barplot statistics, asterisks were used to indicate significant changes in proportion, with *, **, *** and **** representing *p*-values of <0.05, <0.01, <0.001 and 0.0001 respectively.

Statistics from colony experiments

Statistical analysis of culture wells producing colonies between fetal BM and FL HSCs by Mann Whitney test yielded p=0.006 (***, n=7 experiments). Comparison between fetal BM and FL HSC committed progenitors by the same method yielded *p*=0.006 (***, n=1536 wells). Comparison of number of colony types per well for paired progenitor types was performed by binomial test, comparing 1 colony type with >1 colony type. *p*-values were 0.0012 for fetal BM and FL HSCs (**, n=164) and 0.3566 for fetal BM and FL HSC committed progenitors (ns, n=152). Comparison of number of myeloid-only colonies for paired fetal BM and FL HSCs was performed by binomial test, comparing 'myeloid only' with 'myeloid+other'. *p*-values were 0.0002 (***, n=79).

Disease lists for interactive web portal

The interactive web portal offers the capability to search expression profiles by genes associated with haematological disease. These gene lists were intended to compass inherited disorders with haematological phenotype and are derived from Genomics England clinical testing panels (https://panelapp.genomicsengland.co.uk/) (**Supplementary Table 45**).

Data and materials availability

The raw and processed DS/non-DS fetal bone marrow 10x GEX/BCR-/TCR-enriched scRNA-seq data for this study are deposited at EMBL-EBI ArrayExpress, EMBL-EBI ENA, and NCBI GEO, with accession codes as follows: E-MTAB-9389, E-MTAB-10042 and ERP125305. Related accession codes for this study (including Smart-seq2 and CITE-seq data) are linked to the main accession at E-MTAB-9389 and include: E-MTAB-9801 for fetal BM Smart-seq2; GSE166895 for CD34+ fetal BM, fetal liver and cord blood CITE-seq; GSE166895 for fetal BM CITE-seq.

Code availability

Single-cell sequencing data were processed and analysed using publicly available software packages. Python/R code and notebooks for reproducing single-cell analyses are available at https://github.com/haniffalab/FCA_bone_marrow

Acknowledgements

We thank the Newcastle University Flow Cytometry Core Facility, Bioimaging Core Facility, NovoPath, Genomics Facility, NUIT for technical assistance, School of Computing for access to the High-Performance Computing Cluster, CellGenIT, and Alison Farnworth for clinical liaison. We thank CRUK CI Genomics core for processing all Cambridge libraries/sequencing, Kristopher 'Kit' Nazor, Bertrand Yeung and Tse Shun Huang (Biolegend) for helpful discussions to optimise the TotalSeq™ panel and protocol, and Elizabeth Poyner for assistance in segmenting microscopy images. The human embryonic and fetal material was provided by the Joint MRC / Wellcome (MR/R006237/1) Human Developmental Biology Resource (www.hdbr.org).

Author Contributions

M.H.; S.A.T and I.R. conceived and directed the study. M.H.; S.A.T.; I.R.; L.J.; and E.L. designed the experiments. Samples were isolated by S.L.; R.A.B.; I.G.; J.E.; P.B.; K.A.; S.O.B.; N.E.; libraries prepared by E.P.; and E.S; and sequencing by J.C.; R.Q.; R.H.; and WSI core facility. Flow cytometry and FACS experiments were performed by R.A.B.; L.J. and D.M., supported by D.McD.; and A.F. Cytospins were performed by L.J. and D.D.; and *in vitro* culture differentiation experiments were performed by L.J.; C.M and D.M. Immunofluorescence microscopy was performed by C.J.; T.N.; C.C.;. C.S.;. M.A., with analysis performed by M.M., B.O., C.S. and I.G. M.S.K.; B.L.; O.A.; M.T.; D.D.; T.L.T.; M.S.; O.R-R. and A.R. generated adult and cord blood scRNA-seq datasets. CITE-seq datasets were generated by E.S.; N.M.; and N.K.W. Computational analysis was performed by S.W.; G.R.; I.G.; M.Q.L.; E.D.; I.K.; M.M.; J.B.; M.S.J.; M.E.; and web portals were

constructed by I.G.; D.H.; and J.McG., with disease information assembled by K.P. and T.C. M.H.; L.J.; S.W.; I.G. G.R.; B.O.; H.K.; K.M.; T.C.; N.M.; N.K.W.; K.M.; D.H.; D.M.P.; S.B.; A.R.; E.L.; B.G.; I.R.; and I.G.; and S.A.T. interpreted the data. M.H.; L.J.; S.W.; I.G.; G.R.; B.G.; I.R. and S.A.T. wrote the manuscript. All authors read and accepted the manuscript.

Competing interests

None declared

Funding

We acknowledge funding from the Wellcome Human Cell Atlas Strategic Science Support (WT211276/Z/18/Z); M.H. is funded by Wellcome (WT107931/Z/15/Z), The Lister Institute for Preventive Medicine and NIHR and Newcastle-Biomedical Research Centre; S.A.T. is funded by Wellcome (WT206194), ERC Consolidator Grant ThDEFINE and EU FET-OPEN MRG-GRAMMAR awards; relevant research in the B.G. group was funded by Wellcome (206328/Z/17/Z) and MRC (MR/M008975/1 and MR/S036113/1); I.R. is funded by Blood Cancer UK and by the NIHR Oxford Biomedical Centre Research Fund; A.R is funded by Wellcome Trust Clinical Research Career Development Fellowship (216632/Z/19/Z) and supported by the NIHR Oxford Biomedical Centre Research Fund; L.J is funded by NIHR Academic Clinical Lectureship and S.W is funded by a Barbour Foundation PhD studentship. This research was funded in part by the Wellcome Trust [see above for grant numbers]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.