Multiomics of human bone marrow decodes B cell ontogeny across the lifespan and guides in vitro B lymphopoiesis

Roser Vilarrasa-Blasi1, Seungjoon Kim1, Jakov Korzhenevich2, Franziska M. Schmidt2, J. Patrick Pett1, Batuhan Cakir1, Michael Svaton2, Chenqu Suo1,3,4, Oxana Nashchekina1, Frederick C.K. Wong1, Di Zhou1, Carmen Sancho-Serra1, Ryan Colligan1, Yale S. Michaels5,6, Carla Zimmerman7, Kevin Troule1, Yinshui Chang1, Rina Sakata1,8, Hana Aliee1, Valentina Lorenzi1, Marie Moullet1, Davide Carra9,10, Leonardo Morsut9,10, Elena Prigmore1, Martin Prete1, Moritz Mayr11, Hagen Schmal11, Khalid Shoumariyeh12, Antonia M. S. Müller13, Charlotte U. Zajc13, Iva Kelava1, Sarah A Teichmann3, Klaus Warnatz14, Biola M. Javierre15,16, Luz Garcia-Alonso1, Peter W. Zandstra7,17, Marta Rizzi2,14,18,19@, Roser Vento-Tormo1,3@

Affiliations:

1 Wellcome Sanger Institute, Cambridge, UK.

2 Division of Clinical and Experimental Immunology, Institute of Immunology, Medical University of Vienna, Vienna, Austria.

3 Cambridge Stem Cell Institute, Cambridge, UK.

4 Department of Paediatrics, University of Cambridge, Cambridge, UK.

5 Paul Albrechtsen Research Institute CancerCare Manitoba, Winnipeg, MB R3E 0V9, Canada.

6 Department of Biochemistry and Medical Genetics, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB R3E 0W2, Canada.

7 School of Biomedical Engineering, University of British Columbia, Vancouver, Canada.

8 MRC Laboratory of Molecular Biology, Cambridge, UK,

9 Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

10 Department of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA.

11 Department of Orthopaedic and Trauma Surgery, Medical Center - University ofMedical Center Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany.

12 Department of Medicine I, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany.

13 Department of Transfusion Medicine and Cell Therapy, Medical University of Vienna, Vienna, Austria.

14 Department of Rheumatology and Clinical Immunology, Medical Center - University ofMedical Center Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany.
15 Josep Carreras Leukaemia Research Institute, Barcelona, Spain.
16 Institute for Health Science Research Germans Trias i Pujol, Barcelona, Spain.
17 Michael Smith Laboratories, University of British Columbia, Vancouver, Canada.
18 Center for Chronic Immunodeficiency, Medical Center - University ofMedical Center Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany.
19 CIBSS – Centre for Integrative Biological Signalling Studies, University of Freiburg, Freiburg, Germany.
@ Marta Rizzi = marta.rizzi@uniklinik-freiburg.de and marta.rizzi@meduniwien.ac.at
@ Roser Vento-Tormo = rv4@sanger.ac.uk

Abstract

In humans, the bone marrow becomes the primary site for B lymphopoiesis during the second trimester of pregnancy and continues throughout life. Prenatal and adult B cell progenitors play distinct roles in the aetiology and pathology of pediatric and adult hematopoietic malignancies, though the molecular drivers of these differences remain unclear. Here, we created a comprehensive multiomics atlas of over 1.5M cells covering immune and stromal compartment from prenatal and adult bone marrow. By integrating our cohort with publicly available datasets, our cross-study single-cell atlas enables high-resolution analysis of the cell-intrinsic and cell-extrinsic processes that modulate prenatal and adult B lymphopoiesis. Even though B cells follow broadly similar developmental trajectories, we identify a postnatal ‘preProB-like’ cells subset with high similarity to prenatal ‘preProB’ cells, and uncover prenatal cells which carry several signatures characteristic of leukemias, including enhanced proliferation, higher *RAG1/RAG2* activity, extrinsic B cell signals such as *IL7*, and lower retention signals in the bone marrow. Our atlas of B cell development allowed us to develop and characterise, at both cellular and molecular levels, the first feeder-free protocol for generating B cell precursors from human induced pluripotent stem cells. Together, our single-cell multiomics atlas of B lymphopoiesis *in vivo* and *in vitro* offers detailed insights into the unique molecular features of prenatal and adult B cell lymphopoiesis, and serves as a powerful resource for investigating the early events that contribute to haematological disorders.

Introduction

B lymphopoiesis involves multiple coordinated steps that ensure the continuous generation of immature B cells, which populate the periphery. This process is tightly regulated through key mechanisms including coordinated expression of fate determining transcription factors (TFs)[1], stepwise rearrangement of the immunoglobulin receptor[2], stage specific proliferation[3], signals from the microenvironment[4,5] and negative selection against self-reactive specificities[6]. In humans, fetal hematopoietic stem cells (HSCs) initiate commitment to the B cell lineage in the fetal liver in the first trimester, but from the second trimester, the fetal bone marrow becomes the major site of B lymphopoiesis[7,8]. In contrast, in mice, the fetal liver remains the primary site of B lymphopoiesis until birth[9]. Further key differences distinguish mice and human B lymphopoiesis including markers[10], differentiation states[11] (e.g., early lymphoid progenitors, ELPs and preProB), required cytokines (e.g., IL-7[12]), and specific signaling pathways (e.g., BTK, BLNK[13,14]). These species-specific distinctions underscore the need for a comprehensive characterisation of human B cell ontogeny, as well as the development of accurate human *in vitro* models.

A developmental block in B cell progenitor differentiation[15], along with uncontrolled proliferation and increased susceptibility to chromosomal translocations that often promote interactions between B cell developmental genes and immunoglobulin heavy chain loci[16], can predispose to leukemia[17]. Most pediatric B cell acute lymphoblastic leukemias (B-ALL) are thought to originate *in utero*[18–22], with fetal ELPs identified as the cell of origin for the most common subtype of infant B-ALL (iB-ALL), driven by *KMT2A* rearrangements[23]. However, despite their central role in early detection, disease prevention, and therapeutic design, the fetal-specific gene programs that define early B lymphopoiesis, and those co-opted in pediatric leukemia, remain largely unexplored. Moreover, the TFs and regulatory networks that distinguish fetal from adult B lymphopoiesis, and potentially drive disease susceptibility, have yet to be systematically defined.

Here, we generated a comprehensive single-cell multiomics atlas of human bone marrow hematopoiesis across prenatal and adult life, integrating single-cell RNA sequencing (scRNA-seq) and single-cell open chromatin (scATAC-seq) to reveal fundamental differences in cell states, chromatin dynamics, microenvironment interactions and leukemia-associated programs in B lymphopoiesis. Unlike prior stage-specific atlases[24–33], our cross-lifespan dataset uncovers previously unrecognised features, including a novel ‘preProB-like’ population in adult bone marrow, reduced V(D)J diversity in the prenatal stage, and distinct TF regulatory networks driving prenatal versus adult B lymphopoiesis. Notably, fetal-specific gene programs and regulatory circuits identified in our atlas are recapitulated in pediatric B cell leukemias, suggesting mechanisms underlying the heightened susceptibility of fetal-

derived cells to malignant transformation. Leveraging this atlas, we established and benchmarked a novel feeder-free platform for generating B cell progenitors from human induced pluripotent stem cells (hiPSCs) that recapitulates *in vivo* developmental trajectories and overcomes the limitations of prior mouse feeder-based systems[34–37], which failed to produce definitive HSCs and offered limited experimental control. Thereby, we have laid the foundation for scalable B cell production for disease modelling and future immunotherapies.

Results

A single-cell multiomics atlas of prenatal and adult bone marrow in humans identifies differences in cell stages across lifespan

Following the second trimester of fetal development, the bone marrow becomes a specialised niche for early B cell differentiation, supporting the progression of HSCs into progenitors that progressively develop into precursor B cells, which undergo immunoglobulin (Ig) gene rearrangement and finally surface expression of the B cell receptor (BCR). To characterise this process, we generated a comprehensive single-cell multiomics atlas (scRNA-seq and scATAC-seq) by profiling 13 second-trimester fetal bone marrow samples (16-21 post conceptional weeks, pcw), when the bone marrow is the main hematopoietic organ producing B cells, and 14 adult bone marrow aspirates (24-82 years old, yo) (**Figure 1A, Supplementary Figure 1A-B**). ScATAC-seq was performed on 8 fetal and 12 adult bone marrow samples (**Figure 1B**).

After QC filtering, we retained 229,221 fetal and 217,633 adult cells from scRNA-seq with V(D)J data, as well as 61,965 fetal and 159,094 adult nuclei from scATAC (**Figure 1B**). We further integrated our transcriptomics data with publicly available scRNA-seq data from the human fetal bone marrow[24] (12-19 pcw) and postnatal bone marrow[25] (<2-84 yo), encompassing both immune and stromal compartments **(Figure 1C)**. This resulted in a unified dataset of 1,677,926 high quality cells - 366,576 fetal and 1,311,350 postnatal (**Figure 1C, Supplementary Figure 1C-E, Supplementary Figure 2A, Supplementary Table 1-2**). Postnatal bone marrow was further stratified (1-5 yo) to find out whether there are pediatric-specific changes (**Supplementary Figure 2B**). Cell types in scRNAseq datasets were annotated based on the expression of well-characterised lineage markers and we used label transfer to annotate scATAC-seq data (**Supplementary Figure 3A-B, Supplementary Table 3,** see **Supplementary Note 1-2**). The bone marrow atlas is accessible at https://www.hipimmuneatlas.org/vb2024/index.html (username: vb2024, password: vb2024).

We next focused on the scRNAseq dataset. To directly compare B cell differentiation processes between fetal and adult bone marrow while minimising batch effects, we performed dimensionality reduction and clustering analysis on cells spanning hematopoietic precursors to B cells (Vilarrasa-Blasi bone marrow dataset only) (**Figure 1D-E**). Plasma cells were identified in adults but absent in fetal bone marrow (**Figure 1E**). This is consistent with previous reports showing the appearance of serum IgM only from 24 pcw[38], suggesting that plasma cell development occurs later in gestation or in other anatomical sites. We identified a population of potential innate-like B1 cells in fetal and adult bone marrow (**Supplementary Figure 1E, Supplementary Figure 2A, Supplementary Note 3**). While B1

cells have been previously identified in various fetal organs[39], our single-cell transcriptomic analysis now reveals a population expressing canonical markers of B1 cells (*CD19*, *CD20*, *CD5*, *MZB1*[40] and *CCR10*[39]).

In addition, progenitors and B cell populations in fetal and adult human bone marrow were profiled using high-dimensional flow cytometry to validate the identified populations and determine their true proportions, which can be distorted in single-cell transcriptomics data, because of sorting enrichment of rare populations (**Supplementary Figure 4A-E**). Fetal and adult human bone marrow profiling by spectral flow cytometry confirmed the presence of cells expressing B1 markers (CD43, CD27, CD20 and CD5)[41] (**Supplementary Note 3**). In addition, compared to adult bone marrow, fetal bone marrow showed higher frequencies of HSCs (CD34+CD38-CD90+) and Myeloid Erythroid Progenitors; higher frequencies of B cell precursors; a similar distribution of differentiation stages, with the exception of increased immature B cells; and lower frequencies of mature B cell subpopulations (naive to antibody-secreting cells (ASCs)) (**Supplementary Figure 4A**).

PreProB-like cells, an intermediate phenotype between CLP and proB cells in adults

To compare transcriptomic differences between fetal and adult B cell development, we reconstructed inferred differentiation trajectories from HSC to preNaiveB cells (also known as transitional cells) in both fetal and adult samples (**Supplementary Figure 5A-F, Supplementary Table 4, Supplementary Note 4**). The overall progression was similar in both cases: HSCs transition to multipotent progenitors (MPPs), which then give rise to lympho-myeloid primed progenitors (LMPPs). LMPPs commit to either early lymphoid progenitors (ELPs) predominant in the fetal bone marrow or common lymphoid progenitors (CLPs) predominant in the adult bone marrow. ELPs and CLPs clustered together (**Figure 1E**), but showed differential expression of defined markers. For example, *CD7*[31,42] was upregulated in ELPs, while *MME* (CD10)[11] and *DNTT* (TdT) were exclusively expressed in CLPs (**Figure 1F**). In line with its use to identify lymphoid and B cell precursors, *MME* (CD10) extended to later developmental stages (earlyProB up to immatureB) in both fetal and adult samples (**Figure 1F, Supplementary Figure 5B**). Even though adult bone marrow was sourced both from hip replacement surgery and from needle aspirate, with the former cohort being older, no significant differences were observed among donors, supporting the robustness of our dataset **(Supplementary Note 4**).

In fetal bone marrow, differentiation from ELP to proB occurs predominantly via preProB cells (**Supplementary Figure 5C-D**), a population described as a hallmark of fetal B lymphopoiesis[11]. In adults, we defined a rare subset of adult cells clustering with preProB (**Figure 1E, Supplementary Figure 5G**) and positioned between ELP/CLP and proB in our trajectories (**Supplementary Figure**

5D) that we named 'PreProB-like' cells. From the transcriptional point of view, preProB-like and preProB shared the expression *IL7R* and *CD38*[10,43] (**Figure 1F**) that are expressed during B cell lineage specification[10]. PreProB expressed higher level of *MS4A1* (CD20), *BANK1*[44], *BLK*[45] and *CD24*[46], that are related to cell signaling, as well as *CD7*, which is expressed at the ELP stage[11] and correlates with the developmental decision between T-B and NK cell fate[47], and *LEF1,* known to control precursor proliferation[48]. Conversely, preProB-like showed enhanced expression of *ID2* that regulates activity of E2A and was shown to increase in adult age[49], as well as of genes related to control of proliferation and myeloid differentiation such as *KLF2*, *KLF4*, *KLF10*[50–52], and of a hematopoietic progenitor gene as *SPINK2*[53]. PreProB-like upregulated *DNTT* at levels similar to CLPs, but lower than earlyProB cells **(Figure 1F, Supplementary Figure 5H).** Both populations exhibit active cell cycling, with approximately 82% of preProB cells and 35% of preProB-like cells classified in S or G2/M phases (**Supplementary Figure 6A**). In addition, scRNAseq analysis of pediatric bone marrow revealed the presence of preProB-like, but not preProB cells, supporting the view that preProB cells as such are restricted to fetal development (**Supplementary Figure 6B).**

We developed a custom antibody panel to define preProB and preProB-like in the pre- and post-natal bone marrow, based on the expression of CD34 (HSC marker maintained in early B cell differentiation stages) and CD19, with low levels of CD10 and the absence of immunoglobulin expression (**Figure 1G-H, Supplementary Figure 6C**). While preProB cells made about 6% of CD34+ cells in fetal bone marrow, preProB-like, as defined by the same surface markers, were about 1% of CD34+ cells (**Figure 1H- I**). PreProB cells and preProB-like shared similar surface markers, as well as similar expression of CD38, CD24, CD45RA and CXCR4. In addition, CD79a transcript (**Figure 1F**) and protein (**Supplementary Figure 6D**) were expressed both in preProB-likeand preProB, with a discrepancy between higher RNA expression in fetal and higher protein expression in adult preProB-like. In line with global differences between fetal and adult B cell development, we found that fetal preProB cells had a higher expression of EBF1, PAX5, CD179a (*VPREB1*) (**Figure 1F-J, Supplementary Figure 6D**), confirming transcriptomics data, underlying intrinsic differences in the regulation of expression of fate-specific proteins between fetal and adult B lymphopoiesis (**Figure 1J, Supplementary Figure 6D**).

Hence, our comprehensive map of B lymphopoiesis across prenatal and adult samples enabled the discovery of preProB-like cells as an intermediate phenotype between CLP and proB cells postnatally, and a possible alternative developmental pathway. While sharing features with fetal preProB cells, preProB-like cells exhibit a distinct transcriptional profile, making a previously unrecognised stage in human B cell development.

Developmental divergence in V(D)J recombination reveals a stereotyped, public fetal B cell repertoire

The final phase of our inferred B cell trajectory involved the transition of lateProB cells to preB (large and small) followed by immatureB cells and preNaiveB cells (**Supplementary Figure 5G**). The transition from lateProB to preB was marked by the expression of the Ig heavy chain (IgH), part of the V(D)J recombination process. We reconstructed V(D)J recombination using Dandelion[54] in lateProB, largePreB, smallPreB, immatureB, preNaiveB, memoryB cells, and plasma cells (the last two being exclusive to adult bone marrow).

Our analysis of V(D)J rearrangements showed the expected progression of immunoglobulin gene rearrangements through the different developmental stages starting in lateProB cells (**Supplementary Figure 7A**), further supporting lateProB as an intermediate stage of early differentiation. As expected[55,56], the IgM heavy chain was productively rearranged from largePreB cells and light-chain IGK and IGL from smallPreB cells onwards in both fetal and adult bone marrow. We detected more IgD transcripts in preNaiveB cells in adults and the remaining isotypes (IgA and IgG) were exclusively present in antigen-experienced memoryB and plasma cells in adult bone marrow (**Supplementary Figure 7A)**. The higher frequency of nonproductive IGK rearrangements in fetal smallPreB cells and persistence of nonproductive IGL rearrangements in preNaiveB cells in the fetal bone marrow suggested ongoing receptor editing at these stages **(Figure 1K)**.

IGH chain diversity was similar between fetal and adult largePreB / smallPreB cells, while greater diversity was observed in adult bone marrow productive repertoire at the immatureB / preNaiveB cell stage compared to fetal bone marrow (**Figure 1L**). In line with previous reports[57], we observed a higher rate of N-nucleotide incorporation in the DJ junction in adults. Consistently with the earlier and higher expression of *DNTT* at the CLP, lateProB and earlyProB stage, when D-J gene recombination starts, in adults **(Figure 1F)**. In contrast to previous bulk analyses[57], we observed no significant difference in the NP-nucleotide length in the VD junction between fetal and adults (**Supplementary Figure 7B)**. This discrepancy is likely due to the higher number of cells we analysed and the specific analysis of V(D)J recombination events in distinct B cell precursor populations. The total junction length of IGH was longer in adult cells **(Supplementary Figure 7B).** Longer CDR3 (complementarity determining region) **(Supplementary Figure 7B)** results from the addition of NP nucleotides due to increased TdT activity at the D-J junction in adults, as well as preferential use of proximal D-segments, that are shorter in sequence, in fetal B cells. A similarly higher NP-nucleotide length and junction length were observed in both IGK and IGL light chain genes in adult B cell precursors **(Supplementary Figure 7B).** Taken

together, these data support the findings of the heavy chain and the reduced diversity of the overall fetal repertoire.

In light of the reduced diversity of fetal BCR repertoire, we calculated the Morisita index to assess the similarity of the repertoire across individuals, suggesting the presence of a public repertoire. The average value of the Morisita index is below 0.3 in fetal B cells, which does not suggest a large overlap in the repertoire across individuals **(Supplemental Figure 7C-D).** Nevertheless, in the preB and immatureB–preNaiveB groups analysed, fetal samples had a significantly higher Morisita index compared to adults (p-value= $3.95x10^{-5}$ and $8.31x10^{-11}$, respectively) (**Supplementary Figure 7D**). Considering sampling and the expected large interindividual variation, these data support a slightly higher similarity across fetal samples. In fact, analysis of the V-DJ gene usage showed that fetal repertoire was characterised by the predominant usage of the D_H-proximal IGHV1-2 and IGHV6-1, as well as the IGHD7-27 (**Supplementary Figure 7E**) as previously described[57–59]. Indeed, shared IGH clonotypes have been previously reported among different fetal samples[60]. Functionally, this reduced diversity fetal BCR repertoire, characterised by the use of conserved IGH genes, has been reported to bind apoptotic debris and commensal bacteria. It is also polyreactive but lacks the tolerance checkpoint between preNaive and naive B cells[61], suggesting that peripheral B cell tolerance mechanisms are not yet fully developed at this stage.

Together, our analysis indicates developmental differences in V(D)J recombination between fetal and adult B cell precursors, including distinct timing of receptor editing, isotype expression, and junctional diversity.

Fetal B lymphopoiesis transcriptional and microenvironmental programmes are co-opted in leukemias

B lymphopoiesis is a multistep process orchestrated by TFs and modulated by extracellular signals from the surrounding microenvironment[62]. To identify TFs dynamically activated along fetal and adult B lymphopoiesis, we retrieved chromatin-accessible regions (peaks) specific to the B cell lineage for prenatal and adult samples, and combined them to reconstruct enhancer-mediated gene regulatory networks (GRNs) with SCENIC+[63] (**Supplementary Figure 3B,** see **Methods**). Additionally, we linked these regulatory differences to signals from the microenvironment by identifying ligand-receptor interactions using CellPhoneDB v5[64] (see **Methods, Supplementary Table 5,** interactions visualisation accessible at https://www.hipimmuneatlas.org/).

We began by examining the TFs that are differentially expressed and activated during fetal development and the accessibility status of their chromatin compared to adult B lymphopoiesis. As expected from the shared developmental trajectories of fetal and adult B cell differentiation, we identified canonical regulators of B cell differentiation, PAX5, TCF3 (E2A) and FOXO1, in both contexts (**Supplementary Figure 8A-B**). However, fetal samples showed differential upregulation and activity of TFs linked to leukemia, including IKZF2[65], TCF4[66], EBF1[67], LEF1[68] and ZEB2[69] (**Figure 2A-B, Supplementary Figure 9A, Supplementary Figure 10, Supplementary Table 6**), compared to adult bone marrow. Expression of these TFs was increased compared to both adult and pediatric bone marrow (**Supplementary Figure 9B**). Dynamically associated peaks linked to the expression of these TFs were identified by SCENIC+ **(Supplementary Note 6)**. Interestingly, TFs upregulated during fetal B cell development were also linked to genetic risk variants for haematological disorders, suggesting that disease-associated GRNs resemble a fetal-like state during pathogenesis (**Supplementary Figure 9C**).

To investigate the downstream target of these fetal-enriched leukemia TFs, we performed regulon analysis. Regulon analysis showed that EBF1 regulates *RAG2* expression, ZEB2 regulates *RAG1* expression, and LEF1 and TCF4 regulate both *RAG1* and *RAG2* expression (**Figure 2D, Supplementary Figure 10**). *RAG1* and *RAG2* genes are essential for V(D)J recombination and have been implicated in leukemogenesis[70]. Transcriptomic data showed higher *RAG1* and *RAG2* expression along the fetal B cell trajectory compared to adults (**Figure 2E**).

To explore the potential link between TFs upregulated in fetal samples and leukemic gene expression programmes, we reanalysed pediatric KMT2A-rearranged iB-ALL, where rearrangement events occur very early in development[23] (**Supplementary Note 5**). Using stringent filtering to minimise false positives, we found that *EBF1* and *ZEB2* were also upregulated in iB-ALL, along with proliferation-associated genes such as *IGF2BP3*[71] and *SOCS2*[72] (**Figure 2F**). Analysis of the EBF1 regulon in fetal samples revealed genes co-upregulated in fetal samples and iB-ALL, including genes involved in B cell malignancies such as *ZCCHC7*[73,74] and genes involved in immune cell function such as *ATP1B3.* Similarly, ZEB2 was inferred to regulate expression of the B cell genes *BCL7A* and co-receptor *CD79B* (**Figure 2G-H).** This suggests that sustained upregulation of *EBF1* and *ZEB2* in iB-ALL may contribute to the more aggressive phenotype of leukemias, and this may reflect the roles of these genes in fetal B lymphopoiesis.

Finally, we examined the influence of stromal interactions on GRNs during B cell differentiation. *ITGA4-VCAM1* and *IL7-IL7R* interactions were upregulated in earlyProB and largePreB cells during fetal B lymphopoiesis (**Figure 2I-J, Supplementary Figure 10, Supplementary Table 5**). The receptors *ITGA4* and *IL7R* were significantly upregulated in the fetal compared to adult B

lymphopoiesis (**Figure 2I-J**). Their cognate ligands were expressed by sinusoidal endothelial cells (*VCAM1*) as well as osteoblasts and adipo-MSC (*VCAM1* and *IL7*) (**Figure 2I-J)**. Although IL-7 is dispensable for human B lymphopoiesis, it functions as an important feed-forward signal that enhances proliferation and supports B cell commitment[31]. *In vitro* studies further suggest that dependence on IL-7 signaling increases with age[75]. While our data do not allow us to directly assess whether fetal B cells are functionally dependent on IL-7, we observe increased *IL7R* expression along with *EBF1*, which is functionally associated with IL-7 signaling[31] **(Figure 2A-B)**. Moreover, in fetal samples, we detected expression of genes downstream of both *ITGA4* and *IL7R* signaling (**Supplementary Figure 9D**) that are critical for B cell commitment[76], survival[31,77] and proliferation of early precursors (*i.e.,* PI3K/AKT pathway). These genes promote B cell fate by activating key regulators such as BACH2, PAX5 and EBF1[31], all of which have been implicated in leukemia[78,79].

Together, our findings reveal that fetal B lymphopoiesis is characterised by specific transcriptional and microenvironmental programs, such as sustained IL-7 signaling, early RAG activity, and the activation of EBF1 and ZEB2 regulons, that are also active in pediatric leukemia.

Adult bone marrow rewires transcriptional networks to support B cell maturation and plasma cell niches

Using the same strategy on adult samples, we identified key TFs driving adult B cell differentiation (**Figure 3A**, **Supplementary Figure 8B, Supplementary Note 6)**. Several TFs regulating core cellular processes were upregulated during adult B lymphopoiesis (**Supplementary Figure 11A-C**). We previously defined *DNTT* (TdT) as a gene upregulated in the newly defined preProB-like state and maintained at high levels throughout subsequent B cell differentiation in adults **(Figure 1F, Figure 3B)**. Here, we found that several TFs active in the preProB-like state, including TCF3, TCF4 and NR3C1, target *DNTT* **(Figure 3C, Supplementary Figure 10**). Additionally, FOXO1, NR3C1, ZEB2, PAX5, TCF3 and TCF4 targeted *MME* (CD10), which was upregulated at the transcriptional level in CLP in respect to ELP **(Figure 3D)**.

Next, we quantified extrinsic regulators of B cell progenitors in adults and found upregulation of *CXCR4* (B cell progenitors), which binds to *CXCL12* in the stroma (upregulated in adult adipo-MSC, endothelial and fibroblasts) (**Figure 3E, Supplementary Figure 11D)**. This interaction supports HSC survival, maintenance[80], initial differentiation into B cells and retention of both HSCs and immatureB cells. In immatureB cells, the CXCR4/CXCL12 axis, together with PI3K signaling, regulates central tolerance by ensuring the selection of functional B cells[6,81]. Exit into circulation is mediated by *CXCR4*

downregulation with *CXCR4* higher in adult B cell progenitors compared to their fetal counterparts (**Supplementary Figure 11D**).

Plasma cells are terminally differentiated B cells responsible for long term antibody secreting memory. While they are absent in fetal development up to 21pcw (**Supplementary Figure 1E**), they are supported in adults by a specialized niche: cell-cell communication analysis revealed adult-specific stromal expression of *CCL2*[82,83] and *CCL28*[84,85] (**Figure 3E**), cytokines and chemokines known to promote plasma cell survival. These signals were predominantly produced by sinusoidal cells, fibroblasts and adipo-MSC in the adult bone marrow, with the corresponding receptors, *CCR2* and *CCR10*, expressed exclusively on plasma cells (**Figure 3E**).

In summary, our findings highlight that adult B lymphopoiesis is shaped by distinct transcriptional and microenvironmental programs that promote B cell retention, central tolerance, and the establishment of a plasma cell niche, which are absent during fetal development.

Atlas-benchmarked framework for *in vitro* generation of B cell progenitors

Differentiation of hiPSCs offers a scalable model to study early B cell development *in vitro*. However, existing hiPSC-based protocols[34–37] have shown low efficiency, and it remains unclear whether feeder cells contribute to lineage outcomes. A systemic evaluation of *in vitro* modelling is therefore essential to fully exploit its potential for studying malignant transformation and therapeutic targeting.

We benchmarked a published two-step protocol employing OP9 and MS-5 stromal cells[37,86], reported to generate CD19-positive cells, although molecular and functional characterization was not provided. Following the published conditions, we generated scRNA-seq data from the output cultures. We observed a strong myeloid bias and unreliable B cell differentiation across hiPSC lines (**Supplementary Figure 12A-B**), consistent with findings from other groups[36]. Progenitors from "Richardson protocol" clustered with stromal population rather than HSCs **(Supplementary Figure 12C)**.

To overcome the limitations of OP9 based systems in generating B cell progenitors (referred to as "Richardson protocol" in the figures), we modified the HSC differentiation step by adapting a previously described T cell protocol[87]. HiPSCs were cultured as aggregates for 8 days to generate HSCs [87], which were then seeded onto MS-5 stromal cells for 21 days supplementing the media with IL-3, IL-7, SCF, and Flt3L to promote B cell differentiation (**Figure 4A, Supplementary Figure 12D-E,** see **Methods**). We obtained CD19-positive cells, and collected *in vitro* cultured cells at multiple time points for scRNA-seq profiling (referred to as "B cell feeder protocol" in the figures) (**Figure 4A**). Notably, a

subset of HSCs cells displayed a definitive transcriptional profile at day 13 and day 15, with high expression of *CD34*, *SPINK2*, *ITGA4* and *CD74* (**Figure 4B**).

We next characterised CD19-positiveCD10-positive B cell precursors generated on MS-5 feeders at day 29 by scRNAseq, representing ~1% of all CD45-positive immune cells (**Supplementary Figure 12E**). Our protocol recapitulated major stages of B cell lymphopoiesis, including early progenitors (*CD34, SPINK2, SPN, PRSS57, HOPX, CD7*), ELP/CLP (*FLT3, LTB, IL7R, IL2RG, CD79A, VPREB1*), proB (*CD19, CD24E, MME, EBF1, PAX5*), and *bona fide* preB to immature B cells (*CD38, SPIB, CD40, MS4A1, PTPRC*) (**Figure 4C-D, Supplementary Table 8**). *FLT3* expression was detected in ELP/CLP, albeit at low levels. Leveraging *FLT3* expression, we added FLT3L to the culture to support expansion of early progenitors **(Figure 4A)**. Comparison with CD34-positive fetal bone marrow cultures revealed B cell differentiation efficiencies of 6-10% of immune cells at day 29 (**Supplementary Figure 12F-G**), with comparable representation of developmental stages (**Supplementary Figure 12F-G**).

To assess B cell differentiation dynamics, we compared *in vitro* trajectories with fetal and adult B lymphopoiesis. Canonical markers including *CD34*, *PTPRC* (CD45), *CD19, MME* (CD10), *EBF1* and *PAX5* were expressed in hiPSC-derived cells but with distinct expression dynamics (**Figure 4E**). *CD34* expression decreased sharply at the progenitor stage *in vitro*, while *in vivo* downregulation was more gradual, only nearing complete suppression at the preB to immatureB stage. Indeed, the loss of *CD34* expression is a common observation in protocols to differentiate cells in culture[88,89]. In parallel, *PTPRC* (CD45) upregulation began earlier *in vitro*, around the progenitor stage, compared to its initiation *in vivo* at the ELP/CLP stage, and its expression levels remained higher *in vitro* throughout the process. *MME* (CD10), *EBF1* and *PAX5* exhibited a slower and delayed increase in expression *in vitro* as compared to *in vivo* while *CD19* followed a comparable expression dynamic (**Figure 4E**). Altogether, these findings show that while the majority of B-cell markers were acquired with similar dynamics, the timing differed for some markers, which may reflect the fact that all cells were exposed to the same signals provided by the MS-5 stromal cells and the culture medium.

We next characterised the signals provided by MS-5 (see **Methods**). MS-5 cells expressed low levels of *IL7* and *TSLP*, as well as factors supporting more primitive progenitors (*KITLG)*, proB and committed B cell precursors (*CSF1)*, and niche localization cues (*VCAM1*, *CXCL12*, *FN1)* (**Figure 4F**, **Supplementary Figure 12H**). They also expressed *TGFbeta2* and *3*, negative regulators of B lymphopoiesis which may contribute to the low yield of immatureB cells in our culture[90]. Importantly, the cognate receptors for these molecules were expressed by the differentiating hiPSC-B cells. Unlike the spatially and temporally segregated niches encountered *in vivo*, MS-5 feeders provide overlapping

maintenance (e.g., *KITLG*) and differentiation (e.g., *CXCL12, CSF1*) cues throughout the culture period, potentially constraining stage-specific progression. In addition to B cells, hiPSC-derived cultures generated various myeloid lineages, including monocytes, macrophages, dendritic cells and mast cells, whose presence did not impact B cell differentiation (**Supplementary Note 7**).

A feeder-free hiPSC B lymphopoiesis system

Given the conflicting cues provided by MS-5 feeders, we next assessed whether feeder-free conditions could improve developmental progression. Supporting the notion that MS-5 inadequately sustain later stages, sorted fetal bone marrow preB cells remained developmentally arrested when cultured on MS-5 feeders (**Supplementary Figure 13A-B**). In contrast, under feeder-free conditions, sorted fetal preB cells efficiently matured into IgM-positive immatureB cells.

Building on our previous work[31,91], we established a fully feeder-free hiPSC B lymphopoiesis model ("B cell feeder-free protocol"). hiPSC were cultured for 8 days to generate HSCs, followed by 5 days of endothelial-to-hematopoietic transition [92,93], and subsequently differentiated for 28 days in medium supplemented with CXCL12, Flt3L, SCF and IL-7, with IL-7 withdrawn from day 21 onward (**Figure 5A-B**, **Supplementary Figure 13C**, see **Methods**). This system generated CD19-positive CD10-positive preB-stage cells (**Figure 5B-D, Supplementary Figure 13C**). ScRNA-seq analyses at day 41, annotated using a logistic regression classifier trained on *in vivo* data, revealed that most cells clustered as largePreB, with a smaller fraction classified as smallPreB (**Figure 5E**). *In vitro* cells from B cellfeeder-free protocol expressed key B lineage genes, including *CD19, SPIB* and *MS4A1* (**Figure 5F, Supplementary Figure 13D**).

Notably, immunoglobulin recombination genes were more highly expressed in feeder-free cultures compared with MS-5 conditions, and, while remaining below *in vivo* levels, more closely approximated the *in vivo* expression profile. Specifically, there was higher expression of *RAG1*, *RAG2, DNTT (TdT)* and *VPREB1* (**Figure 5F, Supplementary Figure 9, Supplementary Figure 13D**). *FOXO1* expression, however, remained low in both systems, potentially limiting recombination efficiency. BCR sequencing showed both nonproductive heavy chain rearrangements and productive and nonproductive light-chain rearrangements detected in these cells at the latest time point (**Figure 5G**), suggesting start of immunoglobulin gene rearrangement.

In summary, we developed a feeder-free system to develop B cell progenitors. Despite limited maturation, the system captures key precursor stages and can be potentially used for disease modelling, as it shows expression of genes implicated in blood malignancies (**Supplementary Figure 13E**).

Discussion

Understanding the regulation of human B lymphopoiesis is essential for addressing diseases ranging from leukemia to autoimmunity. Here, we present a high-resolution, multimodal atlas of fetal and adult bone marrow B cell development, revealing transcriptional, epigenetic, and niche-driven differences that shape immune function. Our findings reveal how fetal programs are co-opted in pediatric leukemia, highlighting potential drivers of disease aggressiveness. We develop and validate a new feeder-free protocol ("B cell feeder free protocol") that captures definitive HSC emergence and all major stages of early B cell differentiation *in vitro*. Together, these tools provide a valuable resource for dissecting the developmental origins of B cell function and transformation.

First, by comprehensively analysing all progenitor populations across prenatal and postnatal development, we identified 'PreProB-like' cells, a previously unrecognized population in adult bone marrow. Transcriptionally, preProB-like cells shares features with fetal preProB cells, including *IL7R* and *CD38* expression, but differs in reduced expression of early B cell markers like *MS4A1 (CD20)*, *BLK*, and *BANK1*. Notably, our lineage inference analyses shows that preProB-like cells and preProB cells bridge the ELP/CLP and proB stages and are transcriptionally distinct from both, suggesting a functionally unique intermediate in adult B cell development. Lineage tracing experiments are difficult in humans because of the need for longitudinal sampling, the intrinsic challenge of clone-sampling bias across aspirates, complicating identification of ancestor cells. We acknowledge that the use of human patient samples limits experimental tractability, particularly with respect to definitive lineage-tracing approaches, and that the relatively low proportion of this population in the bone marrow makes it difficult to isolate sufficient numbers to perform functional assays in independent samples. Further studies, including *in vivo* validation of the lineage potential of preProB-like cells will be needed to understand the contribution of preProB-like cells to adult B lymphopoiesis.

Our comparative confirms the presence of cells expressing B1 markers in fetal and adult bone marrow. Although B1 cells have been reported in human peripheral blood[41], pleural fluid[94] and other fetal organs[39], their existence remains debated[95,96]. Our single-cell data together with high dimensional flow cytometry profiling supports the presence of B1-like cells in fetal and adult bone marrow, consistent with murine studies[97–99]. Notably, we observed differences in the BCR repertoire compared with previously described B1 cells in other fetal organs[54], warranting further validation. The presence of innate-like B cells may contribute to the initial wave of IgM production in the fetus after 24 pcw[38]; as IgM does not cross the placenta, this endogenous response likely provides early immunity. Whether adult B1-like cells represent fetal remnants or arise continuously from adult HSCs remains an open question.

Third, by reconstructing the immunoglobulin repertoire at single-cell resolution, we revealed developmental differences in V(D)J recombination and BCR diversity, extending previous findings from bulk or stage-specific analyses[57,59,100]. We confirmed the biased V(D)J usage in fetal B cells, shorter CDR3 length, and fewer N-nucleotide additions compared to adults, in line with lower TdT *(DNTT)* expression during fetal development[57–59]. Interestingly, prenatally, *DNTT* peaked at the lateProB stage rather than preProB, in contrast to mice, where TdT is absent during liver-based prenatal development, resulting in impaired junctional diversity[101]. The later onset of *DNTT* (at lateProB) and increased *RAG* expression in fetal cells may explain the higher frequency of heavy-chain rearrangements in fetal lateProB cells. These features contribute to the reduced diversity and moderate overlap of fetal B cell repertoire across individuals. In fact, considering prenatal donors were unrelated, the observation of mild overlap across donors suggests convergence of the prenatal repertoire across individuals, hinting at a fetal B cell repertoire with a more limited antigen specificity than their adult counterpart. Repertoire similarity in preterm infants compared to term infants supports the idea of early clonal expansion against shared antigens[102]. A detailed epigenetic dissection of these developmental stages will be necessary to clarify the regulatory mechanisms underlying these differences.

Fourth, we identified TFs and microenvironmental signals active in fetal B lymphopoiesis that are also enriched in pediatric leukemia. Notably, TFs such as *EBF1* and *ZEB2* upregulated in fetal progenitors were also elevated in KMT2A-rearranged iB-ALL. These TFs regulate key components of BCR signaling (*CD79A/B*) and recombination (*RAG1, RAG2*). Increased RAG activity, coupled with high EBF1 and FOXO1 activity, has been associated with chromosomal translocations and leukemic transformation[16]. In parallel, fetal-associated signals, including *IL7R-IL7* and *ITGA4–VCAM1* interactions, were enriched and may contribute to enhanced EBF1 activity, a known driver of leukemogenic programs when dysregulated. Although our analyses reveal sustained fetal-like TF regulatory activity in leukemic blasts, they do not establish causality. Our findings align with the concept of onco-fetal reprogramming[103], whereby tumours reactivate fetal gene circuits associated with aggressive phenotypes, without resolving whether this reflects reactivation or persistence of fetal programmes, nor defining the precise cell of origin or regulatory elements driving transformation.

Finally, guided by our atlas, we developed and benchmarked improved protocols for B cell differentiation from hiPSC. In line with prior reports[36], OP9 and MS-5 based systems inefficiently generated definitive HSCs, resulting in lack of B cells. By promoting definitive-like HSCs emergence[87], we enabled more robust B cell differentiation dependent on MS-5, and subsequently established a fully feeder-free system with enhanced scalability and translation potential. Molecular profiling confirmed stage-faithful progression and closer recapitulation of physiological gene expression on the feeder-free

system, likely due to reduced confounding stromal signals. We could not determine whether the derived B cells more closely resemble fetal or the adult counterparts, potentially owing to culture-derived signals or limited representation of early precursors (e.g., low presence of PreProB). The system remains constrained by low yields of mature B cells and incomplete IgH recombination, indicating the need for further optimisation. Nevertheless, this molecularly defined hiPSC-B cell platform provides a foundation for modelling human B lymphopoiesis, investigating leukemogenic mechanisms, and advancing the development of off-the-shelf antibody-producing cells.

In summary, we present two complementary resources: a multimodal atlas and an *in vitro* differentiation protocol that together reveal the developmental logic of human B lymphopoiesis, shed light on the fetal origins of pediatric leukemia, and enable future advances in disease modelling, immunotherapy, and regenerative medicine.

Data and code availability

We provide user-friendly access to our scRNA-seq resource at https://www.hipimmuneatlas.org/. All the raw and processed sequencing data generated in this study are deposited to ArrayExpress (scRNA-seq, scATAC-seq). Transcriptomics scRNA-seq prenatal bone marrow (accession number E-MTAB-14706), V(D)J data from prenatal bone marrow (accession number: E-MTAB-14698). The code used to perform the analyses presented in the manuscript can be found at https://github.com/ventolab

Acknowledgements

This publication is part of the Human Cell Atlas– www.humancellatlas.org/publications/. We are thankful to the Sanger Core Sequencing pipeline for support with sequencing library preparation. We acknowledge Annie Moisan and Cara Buchanan, program director and coordinator from the Wellcome Leap HOPE program for their support. We thank Antonio García (https://www.bio-graphics.es/) for his invaluable help with conceptualising and making the illustrations that are part of this manuscript, and Aidan Maartens for proofreading and providing advice on the narrative of the manuscript. We extend our gratitude to Lisa Lindeboom for her assistance in setting up the project, and to Jessica Cox for her support with the CellPhoneDBViz portal. We also thank Bee Ling Ng and the Wellcome Sanger Institute Cytometry Core Facility for their support. The human embryonic and fetal material was provided by the Joint MRC/Wellcome Trust (grant no. MR/R006237/1) HDBR (http://www.hdbr.org). We thank the Immunology-Rheumatology Biobank (IR-Biobank) of the Department of Rheumatology and Clinical Immunology and the FREEZE-Biobank of the University Medical Center Freiburg for support in sample collection and storage. The illustration from 'Figure 4 and 5A' was partially created using BioRender (https://biorender.com).

Funding

This work was supported by funding from the Wellcome Trust Grant 220540/Z/20/A, the Wellcome Leap HOPE Program and BBSRC Flexible Talent Mobility Account (FTMA) BB/Z515073/1. R.V-B is funded by EMBO long-term ALTF 737-2021 and UKRI Postdoctoral Fellowship EP/X038068/1, awarded based on a selected proposal under the European Union's Horizon 2021 research and innovation programme, within the Marie Skłodowska-Curie Actions. B.M.J. is funded by the Spanish Health Institute Carlos III (CP22/00127). M.R. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project number 468499998).

Author information

R.V-B and R.V-T conceived and designed the experiments and analyses. R.V-B and S.K analysed the data with contributions from J.P.P, B.C, K.T, Y.C, M.P and L.G-A; V.L contributed to pseudotime analysis, H.A assisted with data integration, M.P contributed to integration analysis, and R.S contributed to leukemia analysis. R.V-B together with D.Z and C.S-S performed sample processing with contributions from E.P. R.V-B and J.K performed bone marrow flow cytometry profiling. R.V-B performed the hiPSC-B cell MS5 protocol with contributions from O.N, F.W, R.C, Y.S.M, C.Z, I.K, and the hiPSC-B cell feeder protocol with J.K and F.M.S. M.M, H.S, K.S, A.M.S.M. and C.Z contributed to adult bone marrow sample collection. R.V-B, R.V-T and M.R interpreted the data with contributions from J.K, M.S, L.G-A, C.S, D.C, L.M, S.T, K.W, B.M.J, P.Z. R.V-B, R.V-T and M.R supervised the work and wrote the manuscript with contributions from L.G-A. All authors read and approved the manuscript.

Competing interests

In the past 3 years, S.A.T. has received remuneration for scientific advisory board membership from Sanofi, GlaxoSmithKline, Foresite Labs and Qiagen. S.A.T. is a co-founder and holds equity in Transition Bio and Ensocell. From 8 January 2024, S.A.T. has been a part-time employee of GlaxoSmithKline. Y.S.M. and P.W.Z. are named inventors on patents for T cell differentiation technology. P.W.Z. is a co-founder of Notch Therapeutics. P.W.Z. and Y.S.M. are co-founders of Apiary Therapeutics. P.W.Z. consult for cell therapy companies. This work is currently under consideration for patent application.

Methods

Patient samples

The human embryonic and fetal material was provided by the Joint MRC / Wellcome Trust (Grant #MR/006237/1) Human Developmental Biology Resource (http://www.hdbr.org) following elective of pregnancy, with written informed consent and approved by the London - Fulham Research Ethics Committee (REC reference 23/LO/0312). HDBR is regulated by the UK Human Tissue Authority (HTA; https://www.hta.gov.uk) and operates in accordance with the relevant HTA Codes of Practice. Adult bone marrow tissues were obtained from healthy donors that underwent hip replacement surgery or donated for hematopoietic cell transplantation and were otherwise healthy. All participants provided written informed consent in accordance with approval by the Ethics Committee of the University Medical Center Freiburg (20-1109) or of the Medical University of Vienna (1087/2025 and 1198/2025).

Tissue processing

Adherent material was removed from the fetal bone marrow samples, and the ends of the bones were cut. Bone marrow cells were recovered by flushing from one end of the bone. Cells were filtered through a 70µm strainer and centrifuged at 500xg for 5 minutes. After centrifugation, red blood cells were lysed using RBC lysis buffer according to manufacturer's instructions (eBioscience #00-4300), and washed with wash buffer containing PBS (Gibco #20012027), 0.5% BSA (Miltenyi #130-091-376) and 2mM EDTA (Life Technologies #AM9260G) to reduce clumping[24]. Cells were then counted and processed in one or more of the following way: (1) sequencing the whole population; (2) enrichment of CD34-positive cells using the CD34-positive selection kit (Miltenyi Biotec #130-046-702), according to the manufacturer's instructions, followed by sequencing; (3) enrichment of the CD34-negative fraction for CD10-positive and CD19-positive cells using CD10 PE clone HI10A (BioLegend #312204) and CD19 PE clone HIB19 (BioLegend #302208) antibodies in combination with the EasySep PE selection kit II (StemCell Technologies #17684), followed by sequencing, (4) enrichment of CD34-positive cells followed by sorting of multipotent progenitors (MPP) and lympho-myeloid progenitors (LMPP) subpopulations[11]. Staining antibodies: Lin-FITC: CD2 clone RPA-2.10 (BioLegend #300206), CD3 clone OKT3 (BioLegend #317306), CD14 clone M5E2 (BioLegend #301804), CD16 clone 3G8 (BioLegend #302006), CD56 clone HCD56 (BioLegend #318339), CD235a clone Ga-R2/HiR2 (BD Biosciences #40174); CD34-BV786 clone 581 (BD Biosciences #743534); CD38 BV711 clone HIT2 (BioLegend #303527); CD90 APC-Cy7 clone 5E10 (BioLegend #328131); CD45RA APC clone HI100 (BioLegend #304111), and DAPI as live death dye. Sorted subpopulations were sequenced. (5) For one sample, CD45-positive cells were enriched using the CD45-positive selection kit (Miltenyi Biotec #130-045-801), according to the manufacturer's instructions, followed by sequencing. (6) Cells were cryopreserved at -80^{0}C for 24 hours before transferring to liquid nitrogen.

For a subset of samples, the extremities and shafts of the bones were separated, chopped into small pieces, and incubated separately in PBS (Gibco #20012027) containing 4mg/mL dispase II (Gibco #17105-041) and 2mg/mL collagenase I (Gibco #171000-17) for 1 hour at 37^0C[25]. Following incubation, samples were topped up with wash buffer and filtered through a 100µm cell strainer. Cells were centrifuged at 500xg for 5 minutes at 4^0C, washed with wash buffer, and filtered through a 70µm strainer, followed by an additional centrifugation step. Red blood cells were then lysed using RBC lysis buffer according to manufacturer's instructions (eBioscience #00-4300). Cells were counted and processed for sequencing.

Adult bone marrow samples were collected in PBS (Gibco #171000-17) supplemented with 0.1% EDTA (Life Technologies #AM9260G) and centrifuged at 150xg for 10 minutes to remove the remaining fat layer. Samples were then mashed and filtered using a 70µm strainer. Mononuclear cells were isolated by density gradient centrifugation (20 minutes at 1000xg) and subsequently washed with PBS containing 0.1% EDTA (7 minutes at 300xg). Cells were processed in one of the following ways: (1) cells were sorted to enrich for specific populations using Bigfoot spectral cell sorter (ThermoFisher scientific). Briefly, cells were incubated for 15 minutes on ice with antibodies: CD45RA BV480 clone HI100 (BD Biosciences #566114), CD33 PerCP-Cy5.5 clone WM53 (BioLegend #303414), CD34 PE-Cy7 clone 581 (BioLegend #343516), CD90 AF647 clone 5E10 (BioLegend #328116) together with Zombie NIR live/dead fluorescence dye (BioLegend #423106) in a total volume 150µl (PBS supplemented with 2% heat inactivated FBS and 1 mM EDTA). After washing, cells were sorted to obtain the CD90-positive and CD34-positive cells and to enrich for CD45-positive CD33-negative cells (50%), CD33-positive cells (25%) and CD45-negative CD33-negative cells (25%) for sequencing. Data were analysed using FlowJo Software (v10.10.0). (2) CD34-positive cells were enriched using the CD34-positive selection kit (Miltenyi Biotec #130-046-702), according to the manufacturer's instructions, and sequenced. The CD34-negative fraction was further enriched for CD10-positive and CD19-positive antibodies in combination with the EasySep PE selection kit II (StemCell Technologies #17684), followed by sequencing.

Flow cytometry of fetal and adult bone marrow samples

Phenotyping of cells was performed using the antibodies listed in **Supplementary Table 9**. For both surface and intracellular staining protocols, selected surface proteins (CXCR4, CD117, CCR10, CD135 and CD127) were stained first for 20 minutes at room temperature to improve resolution. After washing, the remaining surface markers were stained for 15 minutes at 4°C and washed once before acquisition. For intracellular protein detection, the eBioscience Foxp3/Transcription Factor Staining Buffer Set (Invitrogen) was used according to the manufacturer's protocol. Briefly, after washing, cells were incubated with 100µl of Foxp3 Fixation/Permeabilization working solution for 30 minutes at room temperature. After two washing steps with 1x Permeabilization Buffer, the intracellular proteins were

stained for 30 minutes at room temperature. Cells were then washed with 1x Permeabilization Buffer, incubated with secondary antibody for 15 minutes at 4°C, and washed twice with 1x Permeabilization Buffer prior to acquisition. Samples were acquired on Cytek Aurora (Cytek Biosciences) using SpectroFlo Software (version 2.2.0 and 3.1.0). Data were analysed using FlowJo Software (v10.7, TreeStar).

ELP/proB and preB B cell differentiation

For the development of fetal ELP/proB cells in B1 cells or pre-B cells into immature B cells, respectively: fetal bone marrow samples were sorted for CD33-negative, IgM-negative, CD34-positive cells that were either ELP/proB cells (CD19-negative, CD10-positive) or pre-B cells (CD19-positive, CD10-positive), respectively. For sorting following antibodies were used: IgM BV650 clone MHM-88 (BioLegend #314524), CD33 PerCP-Cy5.5 clone WM-53 (BioLegend #303414), CD10 PE clone LT10 (invitrogen #MA119659), CD34 PE-Cy7 clone 581 (BioLegend #343516), CD19 APC clone HIB19 (BioLegend #302212), zombie NIR (BioLegend #423106). For all cultures, sorted cells were cultured for 21 days under normoxic conditions (37°C, 6.5% CO_2) in a 96-well U-bottom plate (Nunc #9407471) in Iscove medium (Gibco #21980065) containing fetal bovine serum 10% (Sigma-Aldrich #S0615), L-glutamine 1X (Gibco #25030081), nonessential amino acids 1% (Gibco #11140035), reduced glutathione 1 µg/mL (Sigma-Aldrich #G4251), transferrin 2.5 µg/mL (Sigma-Aldrich #T5391), and insulin 1 µg/mL (Sigma-Aldrich #I5500), as previously described in[104]. On day 21, the development was monitored by spectral cytometry using CD19 BUV563 clone HIB19 (BD Bioscience #741361), CD38 Pacific Blue clone HIT2 (ExBio #PB-366-T100), CD45RA BV480 clone HI100 (BD Biosciences #566114), IgM BV650 clone MHM-88 (BioLegend #314524), CD10 BV711 clone HI10a (BioLegend #312226), CD27 BV786 clone L128 (BD Biosciences #563327), CD43 FITC clone 1G10 (BD Biosciences #555475), CD33 PerCP-Cy5.5 clone WM-33 (BioLegend #303414), CD20 PerCP clone 2H7 (BioLegend #302323), IgD PE-Dazzle 594 clone IA6-2 (BioLegend #348240), CD34 PE-Cy7 clone 581 (BioLegend #343516), zombie NIR (BioLegend #423106), CD5 APC-Cy7 clone L17F12 (BioLegend #364009).

Cell culture

iPS11 hiPSCs (ALSTEM Inc, kindly gifted by Zandstra's lab; ALSTEM Inc) were cultured in mTeSR-plus media (STEMCELL Tech #100-0276) on 10cm plates (Corning #430167) coated with hESC-qualified Matrigel (Corning #354277) at a final concentration of 180µg/mL, diluted in 6mL DMEM/F12 (Gibco #21331020). Plates were coated overnight in the fridge, wrapped with parafilm, and brought to room temperature before use. iPS11 hiPSCs were passaged as clumps when 60-70% confluency is reached by washing with PBS (Gibco #20012027) before treating with 4mL of TrypLE Express (Gibco #12605-010) at 37˚C for 5 minutes. TrypLE was gently aspirated and cells were scraped down into

mTeSR-plus media using a cell scraper (Sarstedt #83.3951) and plates were washed with fresh mTeSR-plus media. Gently pipette mix 1-3 times to break up remnant large colonies and seed on freshly Matrigel-coated plates. Cells were maintained at 37°C, 5% CO_2. Media was changed daily.

Kolf_2 (HipSci) were cultured in E8-flex (Life Technologies #A2858501) on 10 cm plates (Corning #430167) coated with Vitronectin XF matrix (VTN, STEMCELL Technologies #07180) at a 1:25 dilution with PBS (Gibco #10010015). Plates were coated for an hour at 37^0C. Kolf_2 were passaged using ReLesR (STEMCELL Tech #100-0483) at 37^0C for 3 minutes. An appropriate volume of media was added, and colonies were suspended to achieve a uniform size before seeding. Cells were maintained at 37°C, 5% CO_2. Media was changed daily.

MIFF3 hiPSC (UOSi001-B, University of Sheffield) were culture in E8 (Life Technologies #A1517001) on 10cm plates (Corning #430167) coated with Vitonectin-N recombinant human protein (VTN-N, LifeTechnologies #A14700) in a dilution 1:100 with DPBS Ca^{2+}/Mg^{2+} free (Gibco #14190094). Plates were coated for an hour at room temperature. MIFF3 cells were passaged using ReLeSR (STEMCELL Technologies #100-0483). Cells were washed with ReLeSR, the solution was then aspirated, and colonies were left for 5 minutes to detach. An appropriate volume of media was added, and colonies were suspended to achieve a uniform size before seeding. Cells were maintained at 37°C, 5% CO_2. Media was changed daily.

A18945, gibco human episomal iPSC line (ThermoFisher) were cultured in E8 (Life Technologies #A1517001) on 10cm plates (Corning #430167), coated with Vitronectin XF matrix (VTN, STEMCELL Technologies #07180) in a dilution 1:50 with DPBS (Gibco #14190094). Plates were coated for an hour at 37°C. A18966 cells were passaged using ReLesR (STEMCELL Technologies #100-0483) at 37^0C for 4-5 minutes. An appropriate volume of media was then added, and colonies were suspended and seeded to achieve a uniform colony size. Cells were maintained at 37°C, 5% CO_2. Media was changed daily.

OP9 (LGC #CRL-2749) stromal cells were cultured following Richardson et al., protocol[37] with minor changes. Briefly, OP9 cells were cultured in OP9-M (maintenance) media composed of aMEM (MEMa, no nucleosides, powder (Gibco #10442312) prepared with 970mL distilled Water (Gibco #15230089) and 29.3mL sodium bicarbonate solution 7.5% (Merck Life Science #S8761))[37] together with 20% FBS non heat-inactivated (Gibco #26140079) and 1x Pen/Strep (Gibco #15140122). Cells were passaged using freshly thawed 0.05% trypsin (Gibco #15090046), diluted in DPBS Ca^{2+}/Mg^{2+} free (Gibco #14190094) with 5mM EDTA (Life Technologies #AM9260G), and incubated for 5 minutes at 37^0C. After incubation, trypsin was quenched with 5mL OP9-M and cells were centrifuged 300xg for 5 minutes. Cells were then counted, and split 1/4 dilution (approximately 100,000 cells) were seeded into a pregelatinized 10cm tissue culture dish (Falcon #353003) 0.1% bovine gelatin (Sigma-Aldrich #G1393, prepared with: DPBS Ca^{2+}/Mg^{2+} free (Gibco #14190094) and autoclaved) at 4°C overnight

wrapped with parafilm. Cells were maintained at 37°C with 5% CO_2, monitored daily, and split every 2–3 days when they reached 80% confluency.

MS-5 (DSMZ #ACC 441) stromal cells were cultured following Richardson et al., protocol[37] with minor changes. Briefly, MS-5 cells were cultured in MS5-M (maintenance) media composed of aMEM (MEMa, no nucleosides, powder (Gibco #10442312) prepared with 970mL distilled water (Gibco #15230089) and 29.3 mL sodium bicarbonate solution 7.5% (Merck Life Science #S8761))[37] together with 10% FBS non heat-inactivated (Gibco #26140079) and 1x Pen/Strep (Gibco #15140122). Cells were passaged using accutase (STEMCELL Technologies #07920) and seeded into a pregelatinized flask with 0.1% porcine gelatin (Sigma-Aldrich #G1890, prepared with DPBS Ca^{2+}/Mg^{2+} free (Gibco #14190094) and autoclaved) for 2 hours at 37^0C. 100,000 cells were seeded into a 75 cm^2 flask (Corning #430641U). Cells were maintained at 37°C with 5% CO_2, monitored daily, and split every 2–3 days when they reached 80% confluency.

Derivation of CEBPa and CEBPb double knock-out hiPSCs

Stock gRNAs for *CEBPa* and *CEBPb* (Synthego Gene Knockout Kit V2 1.5nmol comprising 3x equimolar gRNAs per gene) were prepared in 1X TE pH 7.5 solution (IDTE, IDT #11-01-02-02) at 200pmol/µL (200µM). 22.5pmol of each gRNA mix (a total of 45pmol) were assembled into ribonucleoproteins (RNPs) complexes with 20pmol of eSpCas9NLS (NEB #M0606T) and left at room temperature for 20 minutes inside the tissue culture flow hood. Cultured iPS11 iPSCs of early passage (p7) on a 10cm plate were isolated as single cells by washing with PBS and treating with 4mL of TrypLE Express (Gibco #12604013) at 37˚C for 2.5 minutes. Gently aspirate TrypLE without dislodging colonies before resuspending cells and washing the plate once in a total of 10mL of mTeSR-plus media. Single-cell density was checked and counted by diluting equivolume with 0.4% trypan blue solution (Gibco #T10282) using a haemocytometer (VWR #631-1098). 200K cells were centrifuged at 300xg for 3 minutes and supernatant was aspirated. Cell pellet was resuspended in 16.4µL P3 buffer + 3.6µL P3 supplement + 2µL RNP mix and transferred to a 16-well strip cuvette (Lonza #V4XP-3032) for nucleofection using the CA-137 program in a 4D-Nucleofector-X unit (Lonza). Cells were immediately transferred to room temperature-equilibrated mTeSR-plus before seeding in a serial dilution on a Matrigel-coated 6-well plate in mTeSR-plus media supplemented with 10µM Y-27632 (Merck #Y0503-1MG). The cells were cultured for approximately 10 days to allow the formation of single-cell colonies. Colonies were picked by treating each well with 1mL of gentle cell dissociation reagent for 2-3 minutes at room temperature on a microscope until cell-cell adhesions were weakened without dislodging. Bent filtered P10 tips were used to pick colonies into a Matrigel-coated well of a 96-well plate containing mTeSR-plus supplemented with 10µM Y-27632.

gRNA target sequence (PAM)	Target gene
CGGCTCCGCCTCGTAGAAGT(CGG)	CEBPa
TGCTCGCAGATGCCGCCCAG(CGG)	CEBPa
CGCGCCCCGGGGAAAGCCGA(AGG)	CEBPa
GCCGCCTGCCTTTAAATCCA(TGG)	CEBPb
TGCTTGGCTGCTGCGTACGG(CGG)	CEBPb
TCATGCAACGCCTGGTGGCC(TGG)	CEBPb

Genotyping of amplified clonal lines were done by extracting genomic DNA

For genotyping, genomic DNA was extracted from 96-well plates using the DNeasy kit (Qiagen #69504) according to the manufacturer's protocol. Screening was conducted via colony PCR using the Taq PCR Kit (Qiagen #203203). The presence of *CEBPa/CEBPb* knockouts was assessed using forward and reverse primers designed to anneal approximately 200–400 bp upstream and downstream of the targeted sequence. PCR products were resolved on 1% agarose gel, and KO clones exhibiting either the absence of a PCR product or an altered fragment size were selected for sequencing.

Primers used for KO Screening:

Target Gene	Screening Primers	Expected WT size	Primer Sequence
CEBPa	F4-sF91	490bp	F4: 5'-GCGGCCTGCCGGGTATAAA-3' FS91: 5'-GCCCGGGTAGTCAAAGTCGCC-3'
CEBPb	F6-R5	620bp	F6: 5'-CCGGGCTCAGGAGAAACTTTA-3' R5: 5'-TTCTTGCAGTTCTTGCCCCC-3'
CEBPb	F10-R9	900bp	F10: 5'-TGACGCAGCGGTTGCTA-3' R9: 5'-GCCTGGTAGCCGAGGTAAG-3'

Feeder hiPSC-B cell differentiation protocol

To generate the developmental precursors of hematopoietic stem cells (HSCs), we mimicked the hemogenic endothelium process. Briefly, hiPSCs were aggregated by centrifugation into spheroid cultures using microwell plates. hiPSCs in aggregate were specified into mesoderm and definitive hemato-endothelial identity following a stepwise protocol from Michaels et al.,[87]. One hundred ninety-two hours after initiating the differentiation, day 8, hiPSC-CD34 positive cells were enriched using the CD34-positive selection kit (Miltenyi Biotec #130-046-702), following manufacturer's instructions.

hiPSC-CD34 positives-enriched cells were cryopreserved using CryoStor CS10 (STEMCELL Technologies #07930) or used directly for B-cell differentiation. At day 7 of the protocol, 300,000 MS-5 stromal cells were seeded with 2mL MS5-M media (see 'cell culture' section) into a 0.1% porcine gelatin pregelatinized well of a 6-well plates (Corning #3516) for 2 hours at 37^0C. At day 8, 50,000 CD34-positive cells were seeded onto the MS-5 stromal cells monolayer with 2mL OP9-D differentiation media composed of aMEM (MEMa, no nucleosides, powder (Gibco #10442312) prepared with 970mL distilled water (Gibco #15230089) and 29.3mL sodium bicarbonate solution 7.5% (Merck Life Science #S8761))[37] together with 10% FBS non heat-inactivated (Gibco #26140079), 1x Pen/Strep (Gibco #15140122), 100µm Monothioglycerol (MTG) 1000X (100mM) (Sigma Aldrich #M6145), and the following cytokines: hIL-3 10 ng/mL (Miltenyi Biotec #130-093-909), hIL-7 20 ng/mL (Miltenyi Biotec #130-095-367), hFlt3L 50 ng/mL (Miltenyi Biotec #130-096-479) and hSCF 50 ng/mL (Miltenyi Biotec #130-096-695). After 4 days of culturing CD34-positive cells on MS-5 stromal cells, at day 12 of the protocol, the media was topped-up with 2mL of OP9-D media together with hIL-3, hIL-7, hFlt3L and hSCF. After 8 days on MS-5 stromal cells, at day 16 of the protocol, media was replaced for 2mL of OP9-D media together with hIL-7, hFlt3L and hSCF (excluding hIL-3). At day 12 on MS-5 stromal cells, corresponding to day 20 of the protocol, the media was topped-up with 2mL of OP9-D media together with hIL-7, hFlt3L and hSCF (excluding hIL-3). At day 14 on MS-5 stromal cells, corresponding to day 22 of the protocol, cells were transferred into a new MS-5 monolayer with 2mL of OP9-D media together with hIL-7, hFlt3L and hSCF (excluding hIL-3), and some cells were analysed at day 22 time point. At day 18 on MS-5 stromal cells, corresponding to day 26 of the protocol, the media was topped-up with 2mL of OP9-D media together with hIL-7, hFlt3L and hSCF (excluding hIL-3). Cells were analysed at the day 29 time point. For analysis, cells were sorted using the following antibodies: CD45-BV785/APC clone HI30 (BioLegend #304048/#304037), CD14-BV605/PE-Cy7 clone M5E2/HCD14 (BD Biosciences #564054/BioLegend #325618); CD19-PE/FITC clone HIB19 (BD Bioscience #555413/BioLegend #302206); CD10-APC/BV711 clone HI10A (BD Biosciences #332777/BioLegend #312226); CD20-FITC/BV510 clone 2H7 (BioLegend #302304/BD Biosciences #563067); IgM-BV510/BV785 clone G20-127/MHM-88 (BD Biosciences #563113/BioLegend #314543); CD29-APC-Cy7/PE clone HMB1-1 (BioLegend #102226/#102208) and DAPI as live death dye. Cell hashing was added in some conditions (BioLegend #TotalSeq-C0251 to C0258). After washing, cells were sorted using mainly Bigfoot spectral cell sorter (ThermoFisher scientific). Occasionally, SONY MA900 cell sorter was used. To gate the populations, controls of MS-5 staining and peripheral blood mononuclear cells (PBMC, STEMCELL Technologies #70025.1) were used. The gating strategy (**Supplementary Figure 9B**) involved selecting singlets from all cells, then CD45-positive cells and CD14-negative cells, and from these we gated: (1) CD19/CD10 (2) CD19/CD20 and, where possible, (3) CD19/IgM. Samples were sorted in four-way mode as follows: (1) CD19-positive cells, subdivided into CD10-positive and CD10-negative populations, (2) CD20-

positive cells, including both CD19-positive and CD19-negative cells, (3) CD14-positive cells and (4) remaining immune-positive (CD45-positive) cells gated from CD20-negative and CD19-negative populations. Analysed time points included day 8, day 13 (day 5 on MS-5 stromal cells), day 15 (day 7 on MS-5 stromal cells), day 22 (day 14 on MS-5 stromal cells) and day 29 (day 21 on MS-5 stromal cells). Fluorescence-activated cell sorting data were analysed using FlowJo Software (v10.10.0).

Pictures of the cell culture were taken using: EVOS XL (Invitrogen #AMEX1200) and EVOS FL (Invitrogen #AMF4300).

For Richardson et al.,[37] and French et al.,[35] protocols, steps were followed as published. For Richardson et al., media was topped-up on day 17 (without IL3), and media was replaced on day 24. For French et al., media was topped-up on day 17 (without IL3), CD45-positive cell enrichment was performed on day 21 using a selection kit (Miltenyi Biotec #130-045-801) and cultured with media without cytokines, and media without cytokines was replaced on day 28. Analysed time points for the Richardson et al., protocol were day 10 (CD34-positive cells), day 24 (day 14 on MS-5 stromal cells), day 31 (day 21 on MS-5 stromal cells). For the French et al., protocol, analysed time points were day 10 (CD34-positive cells), day 21 (day 11 on MS-5 stromal cells or the day of the CD45-enrichment), day 29 (day 19 on MS-5 stromal cells). Fluorescence-activated cell sorting data was analysed by FlowJo Software (v10.10.0).

Feeder-free hiPSC-B cell differentiation protocol

To generate the developmental precursors of hematopoietic stem cells (HSCs), we mimicked the hemogenic endothelium process. Briefly, hiPSCs were aggregated by centrifugation into spheroid cultures using using AggreWell™400 plates (STEMCELL Technologies #34425). hiPSCs in aggregate were specified into mesoderm and definitive hemato-endothelial identity following a stepwise protocol from Michaels et al., [105]. One hundred ninety-two hours after initiation of differentiation (day 8), hiPSC-CD34 positive cells were enriched using the CD34-positive MicroBead kit (Miltenyi Biotec #130-046-702), following manufacturer's instructions. The enriched hiPSC-derived CD34-positive population was either cryopreserved in CryoStor® CS10 (STEMCELL Technologies #07930) or used directly for subsequent B-cell differentiation.

hiPSC-derived CD34-positive cells were used as the input population to induce the endothelial-to-hematopoietic transition. Cells, (10,000 cells per well), were seeded in a 96-well flat bottom TC-treated microplate (Corning #3599) pre-coated with 50uL of Matrigel® matrix (Corning #354277) at final concentration of 180ug/ml diluted with DMEM/F12 with GlutaMAX (Gibco #31331-028). Plates were coated overnight at 4^0C wrapped in parafilm. Immediately before use, coating solution was aspirated, and plates were washed once with PBS (Gibco #10010015). Cells were seeded with completed basal media containing: StemPro™-34 SFM (Gibco #10639011), 1x GlutaMAX (Gibco #35050038), 0.5x Penicillin/Streptomycin (Gibco #15140-22), 150ug/mL apo-Transferrin (Sigma-Aldrich #T5391),

0.039 uL/mL 1-Thioglycerol (Sigma-Aldrich #M6145) and 50ug/mL L-Ascorbic acid (Sigma-Aldrich #A8960), supplemented with cytokines: 10ng/mL hBMP4 (R&D Systems #314-BP-010), 5ng/mL hFGF2 (PreproTech #100-18B-10ug), 5ng/mL hVEGF165 (R&D Systems #293-VE-010), 10ng/mL hIL-6 (Miltenyi Biotec #130-093-929), 5ng/mL hIL-11 (PreproTech #200-11-10ug), 30ng/mL hTPO (Miltenyi Biotec #130-095-745), 25ng/mL hIGF-1 (R&D Systems #291-G1-200), 50ng/mL hSCF (Miltenyi Biotec #130-096-695), 10ng/mL hFlt3L (Miltenyi Biotec #130-096-479), 10ng/ml hIL-3 (Miltenyi Biotec #130-093-909). If cells were thawed, 5 µM ROCK inhibitor Y-27632 dihydrochloride (Sigma-Aldrich #Y0503) was added. Cells were incubated under normoxia conditions (37°C, 5% CO_2) for 5 days (corresponding to day 13 of the protocol), with medium topped-up at day 3 (corresponding to day 11 of the protocol)[92,105].

Finally, hematopoietic progenitor cells (50K) were input to induce B-cell differentiation. Cells were incubated under normoxic conditions (37°C, 6.5% CO_2) in a 96-well U-bottom plate (Nunc #9407471) in Iscove medium (Gibco #21980065) containing fetal bovine serum 10% (Sigma-Aldrich #S0615), L-glutamine 1X (Gibco #25030081), nonessential amino acids 1% (Gibco #11140035), reduced glutathione 1 µg/mL (Sigma-Aldrich #G4251), transferrin 2.5 µg/mL (Sigma-Aldrich #T5391), and insulin 1 µg/mL (Sigma-Aldrich #I5500), as previously described in[104]. Media was refreshed twice a week and supplemented with hCXCL12a (20 ng/mL, Immunotools #11343363), hFlt3L (25 ng/mL, Immunotools #11343305), hIL-7 (20 ng/mL, Immunotools #11340075) for 21 more days (corresponding to day 34 of the protocol) and hSCF (25 ng/mL, Immunotools #11343325) for 28 days (corresponding to day 41 of the protocol).

For sorting of lymphoid/B cells of the feeder-free hiPSC culture, cells were incubated for 15 minutes at 4°C with the following antibodies: CD10 PE clone LT10 (Invitrogen #MA119659), CD19 APC clone HIB19 (BioLegend #302212), CD33 PerCP-Cy5.5 clone WM-33 (BioLegend #303414), IgM BV650 clone MHM-88 (BioLegend #314524), CD34 PE-Cy7 clone 581 (BioLegend #343516), zombie NIR (BioLegend #423106). After washing, cells were sorted live CD19-positive and/or CD10-positive cells were used for single-cell RNA sequencing on days 20 and 27 of culture (corresponding to day 33 and 40 of the protocol). Flow cytometric analysis was performed on days 14, 21, 28 of culture (corresponding to days 27, 34, 41 of culture).

10x Genomics Chromium GEX (gene expression) and ATAC assay for transposase-accessible chromatin) library preparation and sequencing

For the scRNA-seq experiments, cells were loaded according to the manufacturer's protocol for the Chromium Next GEM Single Cell 5' v2 (DUAL) Kit (10x Genomics) to attain between 500 and 10,000 cells per reaction. Libraries were sequenced, aiming at a minimum coverage of 50,000 raw reads per cell, on the Novaseq 6000 system, using the sequencing format: read 1: 28 cycles; i7 index: 8 cycles, i5 index: 0 cycles, Read 2: 98 cycles. For cell hashing workflow, Chromium Next GEM Single Cell 5′ v2

with Feature Barcoding technology for Cell Surface Protein was used (5' Feature Barcode kit) according to the manufacturer's protocol (10x Genomics). The BCR and TCR were sequenced at a target depth of 5,000 reads per cell and following the same sequencing format as the scRNA-seq.

For the scATAC-seq and multimodal snRNA-seq/scATAC-seq experiments, cells were loaded according to manufacturer's protocol for the Chromium Single Cell ATAC v2 and Chromium Next GEM Single Cell Multiome ATAC+Gene Expression (10x Genomics) to attain between 1,000 and 10,000 nuclei. Libraries were sequencing, aiming at a minimum coverage of 50,000 raw reads per nuclei on the Novaseq 6000 systems, using the sequencing format ATAC v2: read 1: 50 cycles; i7 index: 8 cycles, i5 index: 16 cycles, read 2: 50 cycles. ATAC from multiome: read 1: 50 cycles; i7 index: 8 cycles, i5 index: 24 cycles, read 2: 49 cycles.

External human bone marrow scRNA-seq and CITE-seq

We collected raw sequencing data from previously published human fetal and adult bone marrow scRNA-seq and CITE-seq datasets. Specifically, we downloaded publicly available .fastq files either from the Gene Expression Omnibus (GEO), ArrayExpress or the European Genome-phenome Archive (EGA). These datasets included: (1) Jardine et al., (E-MTAB-9389, GEO: GSE166895)[24], (2) Bandyopadhyay et al., (GEO: GSE253355)[25], (3) Oetjen et al., (GSE120446)[26], (4) Granja et al., (GSE139369)[27], (5) Setty et al., (S-SUBS8)[28], (6) Caron et al., (GEO: GSE132509)[29], (7) Ainciburu et al., (GSE180298)[30], (8) Kaiser et al., (GSE214693)[31], (9) Zhang et al., (GSE245108)[32], (10) Human Cell Atlas (HCA, S-SUBS12)[33].We also collected scRNA-seq data of ALL from: (1) Khabirova et al., (EGA: EGAD00001007854)[23].

External human fetal YS, AGM, LV, *in vitro* scRNA-seq and CITE-seq

We collected raw sequencing data from previously published human fetal yolk sac(YS), aorta-gonad mesonephros (AGM), liver (LV) scRNA-seq and CITE-seq datasets. Specifically, we downloaded publicly available .fastq files either from the Gene Expression Omnibus (GEO) or ArrayExpress. These datasets included: (1) Goh et al., (E-MTAB-10552, E-MTAB-11549, E-MTAB-11613, E-MTAB-11618)[106–108], (2) Wang et al., (GEO: GSE144024)[109], (3) Calvanese et al., (GEO: GSE162950)[53], (4) Popescu et al., (E-MTAB-7407) [110], (5) Suo et al., (E-MTAB-11343)[39], (6) Alsinet et al., (E-MTAB-11623)[111].

Alignment and quantification of scRNA-seq data

Reads from both the newly generated scRNA-seq libraries and external datasets were alignment to the 10x Genomics' human reference genome GRCh38 (v.3.0.0), followed by cell calling, transcript quantification and QC using the Cell Ranger Software (v.3.0.2; 10x Genomics) with default parameters.

In vitro datasets cultured with MS-5 cells were aligned to the both GRCh38 (v.3.0.0) and mm10 (v.3.1.0). Cell Ranger filtered count matrices were used for downstream analysis.

Downstream scRNA-seq and scATAC-seq analysis

Donor demultiplexing of in vivo libraries

We used 10x sequenced libraries from the public datasets (E-MTAB-11549, E-MTAB-11613, E-MTAB-11618, GSE166895) which were generated with multiplexed donors. Similarly, for the newly generated libraries, we also pooled cell suspensions from multiple donors. To ensure accurate assignment of cells to their respective donor, we genotyped some donors and then pooled sample combinations in a way that each scRNA-seq library contained at least one genotyped donor. To assign each cell in the scRNA-seq libraries back to its donor-of-origin, we first called the single nucleotide polymorphisms (SNPs) in the reads of each barcode with Souporcell[112] (v.2.4). scRNA-seq reads were genotyped from the Cell Ranger BAM files, aligned to the GRCh38 human reference genome, and compared against the filtered_2p_1kgenomes_GRCh38.vcf reference. Once the cells in scRNA-seq libraries were genotyped, we linked them back to their donor-of-origin genotype using shared_samples.py script from Souporcell across all samples. In yolk sac (YS) and liver (LV), we used the genotype information provided by the original study[106]. Genotype doublets and unassigned cells were discarded in downstream analysis.

Demultiplexing of pooled hiPSC-differentiation libraries

First, we used HashSolo[113] to demultiplex human and mouse cells from the MS-5 culture with a default prior distribution. To distinguish each condition in the *in vitro* experiment, we used HashSolo with a modified prior distribution per each sample. In general, we set higher priors for both negative classification and doublets. Cells assigned as ‘doublet’ or ‘negative’ were excluded from downstream analysis. To verify the above classification, we derived a Gaussian Mixture Model (GMM) with scikit-learn (v0.24) and identified two clusters per each hashtag oligo. From a cluster that has higher hashtag oligo counts, we set a minimum count based on the lower 2nd percentile, and cells with counts below this threshold were also excluded.

Doublet detection based on transcriptional mixtures

Doublet detection was performed with the Single-Cell Toolkit (sctk) package. Per-cell doublet scores were computed with Scrublet (v0.2.3) on a per-library basis. Cells were clustered using a two-stage Leiden clustering (initial clustering followed by size-dependent subclustering) and Scrublet scores were propagated to transcriptionally similar barcodes by assigning the median Scrublet score to the cells in the cluster. Clusters with unusually high median Scrublet scores were identified using a one-sided robust

z-score test, with Benjamini–Hochberg FDR correction across clusters. Doublets were removed using precomputed annotations from the *stringent_doublets* column, combining Scrublet scores >0.3 and Benjamini–Hochberg–adjusted *P* values <0.05.

Quality filters, batch correction and clustering

We used the filtered count matrices from Cell Ranger 3.0.2 for downstream analysis of the scRNA-seq libraries with Scanpy (v.1.11.5), following their recommended standard practices[114]. We further filtered the cells to retain only high-quality cells that either (1) expressed more than 500 genes or (2) with a mitochondrial content lower than 20% or (3) expressed less than 100k UMI counts.

To quantify cell cycle effect, we flagged cell cycle genes using a data-driven approach as described in[110,115]. We used the scanpy function 'score_genes_cell_cycle' to infer the cell cycle stage of each cell (that is, G1, G2/M or S) that was later used to interpret the clusters.

Next, we generated an integrated manifold for human bone marrow using all 10x scRNA-seq libraries listed in **Supplementary Table 1**. First, we discard genes expressed in three or fewer cells in the concatenated dataset. Next, we identified the top 5,000 most highly variable genes in each dataset and used scVI[116] (v.1.4.0.post1) to estimate a low dimensional space. To minimize cell cycle bias, the previously flagged cell cycle genes were excluded before estimating the highly variable genes. *Donor_ID* was used as batch variable, with cell cycle *phase* as categorical covariate and *log1p_ncounts* and *percent_mito* included as continuous covariate. All the remaining parameters were kept as default, with *n_latent:*20, *n_layers:*2. With the resulting scVI-corrected latent representation of each cell, we estimated the neighbor graph, generated a uniform manifold approximation and projection (UMAP) visualization and performed Leiden clustering. The resolution of the leiden clustering was adjusted manually to capture well-known, low-abundance cell types.

Next, we subsetted the cells in (i) clusters assigned to the B-cell lineage and (ii) from the in-house Vilarrasa-Blasi bone marrow dataset (fetal and adult) to generate a lineage-specific manifold. This refined manifold was used for fine-grained annotation of B-cell differentiation states: HSC, LMPP, MPP, CMP, ELP/CLP, PreProB/LateCLP, EarlyProB, LateProB, LargePreB, SmallPreB, ImmatureB, PreNaiveB, MemoryB and Plasma cells. To do so, we followed the same approach described above (i.e. for all the bone marrow cells from all datasets), with cell cycle genes excluded from the 5,000 highly variable genes and scVI set to *n_latent:* 10 and *n_layer*:1. The final representation of B-cell differentiation was generated excluding the myeloid-primed CMPs, with scVI using *n_latent*: 10 and *n_layer*: 2.

To characterise the hiPSC-differentiation protocols, we generated a manifold that integrated in vivo tissue scRNA-seq data from human fetal yolk sac, aorta-gonad mesonephros, liver and fetal and adult bone marrow (see section "External human bone marrow scRNA-seq and CITE-seq" and "External human bone marrow scRNA-seq and CITE-seq") together with the hiPSC-differentiation cells (Alsinet et al.,[111] and our cultured and MS-5 and feeder-free hiPSC-B cells differentiation) with scVI, using n_latent: 20, n_layer: 2. We followed the steps described above, and defined batch as the combination of study id, experimental model (*in vitro, in vivo* or cultured fetal bone marrow), the origin (fetal or adult), the organ (yolk sac, aorta-gonad mesonephros, liver and bone marrow), the sequencing type (5' GEX or 3'GEX) and sequencing platform (scRNA-seq, CITE-seq or cell hashing. Mouse cells and genes were discarded prior to the integration from the MS-5 hiPSC-B cell differentiation.

Annotation of cell types

We performed a full re-annotation of the cell clusters in each integrated scRNA-seq manifold generated. When assigning lineage/coarse- and cell type/fine-grained annotations to the Leiden clusters, we first performed a QC round to exclude clusters likely driven by technical artefacts (that is, low QC cells or doublets). Briefly, we flagged as low QC those clusters that (1) express an overall lower number of genes, or (2) express an overall lower number of counts, or (3) display a higher than average mitochondrial RNA content and, importantly, and (4) do not express any distinctive marker gene (and thus are not representing any independent biological entity). Next, we flagged as doublets those clusters that met the following criteria: (1) exhibit higher scrublet doublet score (2) express marker genes from multiple lineages (for example, display both stromal and immune markers) and (3) lacked unique gene markers beyond the doublet's combined gene expression profile. Distinctive marker genes were identified using the Term Frequency–Inverse Document Frequency approach (TF-IDF), as implemented in the SoupX package[117] (v.1.5.0).

Next, we assigned cell type labels to remaining high-quality clusters. General lineage/coarse annotation of the in vivo main datasets was done on the main bone marrow manifold. Cell state/fine grained annotation was defined on the per-lineage manifold (that is, from reanalyzing the cells in each lineage, as described in the previous section). Annotations were defined based on expression of distinctive genes (identified with TF-IDF) with bona fide markers from the literature. From Jardine and colleagues' dataset, we also relied on previous annotations from the author. To aid cell type annotation in external or in vitro datasets, we also used CellTypist[118] trained either with cells (i) from our whole in vivo bone marrow annotations or (ii) from our in vivo B-cell differentiation fine annotations from our in-house dataset.

For downstream analyses, only finely annotated cell states with more than 50 cells per origin (fetal or adult) were considered. We further discarded fetal cells that co-clustered with adult memoryB cells because these fetal cells lacked expression of the markers characteristic of memoryB cells.

hiPSC-derived B cells were automatically annotated using CellTypist trained on our human bone marrow (fetal and adult) in vivo annotations. Cells lacking expression of key cell-type markers (B1, plasma, memoryB) or unhealthy cells downregulating housekeeping genes were subsequently filtered out. The remaining hiPSC- derived B cells were integrated with the in vivo dataset using scVI following the same strategy, with *n_latent* = 1 and *n_layers* = 10 (for MS5 only) and *n_latent* = 2 and *n_layers* = 10 (for MS5 and feeder free), followed by leiden clustering and manual annotation based on CellTypist predictions and bona-fide marker genes.

Similarly, the publically available Leukemia dataset was automatically annotated using CellTypist trained on our human bone marrow (fetal and adult) coarse annotation of cell types. Doublets were removed using the stringent_score threshold. Progenitors and B cells were selected and subsequently refined through automatic annotation using CellTypist trained on our in vivo B-cell differentiation fine annotations. The remaining cells were integrated with the in vivo B-cell differentiation dataset using scVI following the approach described above, removing cell-cycle–related genes from the set of 5,000 highly variable genes, and running scVI using *n_latent* = 2 and *n_layers* = 20

BCR/TCR

All single-cell V(D)J data from the 5′ Chromium 10X kit were initially processed with cellranger vdj pipeline (v.6.1.2) with Cell Ranger V(D)J algorithm using GRCh38 reference (v.5.0.0). BCR contigs contained in 'all_contigs.fasta' and 'all_contig_annotations.csv' were then processed further using dandelion singularity container[54] (v.0.3.7). Specifically, V(D)J output data were re-annotated with igblastn against IMGT (international ImMunoGeneTics) reference sequences by Dandelion preprocessing. For *in vitro* BCR data, we skipped tigger which reassigned heavy chain V gene alleles to avoid errors occurring from low quality contigs. From Dandelion output, we used 'all_contig_dandelion.tsv' output, while we used 'all_contig_igblast_db-all.tsv' to analyze productive and nonproductive repertoires. To define productive and nonproductive, it was defined by True and False from the productive_ratio function for IGH, IGK, and IGL respectively. For most repertoire characteristics including isotype, mutation frequency, junction length, np length and gene usage, we used Dandelion output and were analysed per donor. In gene usage, we performed multiple testing corrections with the FDR approach using p.adjust function from stats (v.4.0.4). In gene usage, we ignored polymorphic variants from IMGT gene names to reduce the data groups and downsampled per donor to match the same number of contigs. For the visualization, we removed genes that were less than

1% in average and re-ordered based on genomic position from IMGT. In isotype, we remove multi_contig from the analysis. To analyze repertoire diversity metrics, we used Immunarch[119] (v.0.9.1). For the input of the data, we first subsetted per cell type and per donor and downsampled to have the same number of contigs per donor from Dandelion output. To identify overlapping repertoire, we measured Morisita index using the repOverlap function with method = 'morisita' excluding HSC, MPP, LMPP, ELP/CLP, PreProB/LateCLP, EarlyProB and B1 populations. To investigate repertoire diversity, we measured Inverse Simpson index using the repDiversity function with method = 'inv.simp'. To analyze gene usage overlap, we measured jensen-shannon using geneUsage function, following geneUsageAnalysis function with method = 'js'. For all statistical difference tests, the glm function from stats (v.4.0.4) was used.

Differential gene expression fetal vs adult bone marrow

We evaluated the magnitude and significance of gene expression differences between subpopulations across the B cell lineage or the stromal cells in fetal and adult bone marrow using limma[120] (v.3.46.0). First, in case of the B cell lineage, we downsample to 500 cells per subpopulation. Secondly, to account for within-sample correlations (i.e. cells coming from the same donor), pseudobulking with sum aggregation was performed prior to applying limma. Specifically, we generated three pseudobulks per donor and per cell type by aggregating the cells of each cell type and taking the mean gene expression within the cell type. Finally, we tested for differential expression between conditions (fetal vs adult) using the "limma-voom" approach. We reported as differentially expressed genes with logFC > 0 and an adjusted p-value < 0.05.

Cell–cell communication analysis with CellPhoneDB

We used CellPhoneDB v5.0.0[64] to study cell-cell interactions between the bone marrow microenvironment and B cell lineage cells, identifying those differentially expressed between fetal and adult. Using CellPhoneDB method-3 based on differential gene expression, we retrieved pairs of interacting ligands and receptors meeting the following requirements: (1) all the interacting partners were expressed by at least 5% of the cell type under consideration; (2i) the receptors were significantly overexpressed by the B cells; or (2ii) the ligands were significantly overexpressed by the bone marrow microenvironment. Differential expression analysis was performed on a per-cell type basis comparing matched fetal and adult cell types along the B-cell lineage trajectory (from HSC to preNaiveB) or in the bone marrow microenvironment using the approach described above. These interactions between immune cells and stromal cells can be iteratively queried via the CellPhoneDBViz browser at: https://www.cellphonedb.org/viz/viz.html?projectid=toBcells_StromalLigands_DEGs&auth=OyPZeD5NRRirUUfH1m1VdA (to query the ligands differentially secreted by the BM microenvironment in fetal vs adult) or:

https://www.cellphonedb.org/viz/viz.html?projectid=toBcells_BcellReceptors_DEGs&auth=dfhHkHuWzCzndeiDXs5XMA (to query receptor changes in B cells between fetal and adult samples).

Trajectory inference and differential expression along trajectories

Trajectory modelling was restricted to the in-house Vilarrasa-Blasi dataset to avoid batch effects. To account for differences in cell abundance, populations were downsampled to 500 cells per developmental origin, and an scVI-corrected manifold was generated as described above. Next, we used the manifold for trajectory inference with Slingshot[121] (v1.8.0), designating hematopoietic stem cells (HSC) as the starting point of the trajectory and preNaiveB as the endpoint. Each pseudotime was then scaled between 0 and 1. The scaled pseudotime ordering of the cells along with the weighted assignment was then used as input for TradeSeq[122] (v1.18.0) to extract genes that are differentially expressed along the B-cell lineage or between fetal/adult with the *associationTest()* function.

The Slingshot-inferred trajectory derived from in-house cells was independently validated using PAGA on bone marrow cells across all studies. Here, we performed scVI integration datasets as described above (*n_layers* = 1, *n_latent* = 20), followed by PAGA-based reconstruction of the lineage structure.

Differential gene expression leukemia vs adult bone marrow

To identify the genes differentially expressed in the ELP/CLP subpopulation between each leukemia patient compared to adult bone marrow, we used the *sc.tl.rank_genes_groups* function from scanpy (v1.8.2) using the two-sided Wilcoxon rank-sum test. Genes with log fold-change (logFC) > 1 and adjusted p-value < 0.01 were considered upregulated in leukemia.

Subsequently, we compared these leukemia upregulated genes to those upregulated in fetal versus adult bone marrow. This allowed us to extract the genes that were consistently upregulated in both leukemia and fetal bone marrow cells compared to their adult counterparts.

Chromatin accessibility data preprocessing

For single-cell ATAC-seq analysis, fragment files and associated outputs generated by CellRanger-arc (v.2.0.1) for 10X multiome libraries or CellRanger-atac (v2.1.0) for scATAC-seq libraries were imported into ArchR[123] (v1.0.2) for downstream processing. Initial per-droplet quality control was performed considering the number of unique nuclear fragments, signal-to-background ratio and the fragment size distribution. Moreover, droplets with transcription start site enrichment score < 4 and number of fragments < 1,000 were removed. Doublets were identified using ArchR's doublet scoring (addDoubletScores) with default setting, and subsequently removed with filterDoublets. Dimensionality reduction on the tile matrix was performed with Latent Semantic Indexing using default values, and the low-dimensional variables were then used to compute the neighbourhood graph, UMAP visualisation and clustering. Initial clustering was performed at a resolution of 0.8. Then, pseudo-bulk

replicates were made for each broad cluster per region from the initial clustering results. Peak calling (501-bp fixed-width peaks) was performed based on pseudo-bulk coverages by MACS2 (v2.2.7.1). Then, a cell-by-peak count matrix was obtained and exported. Peaks were extracted and integrated with RNA count matrix data using GLUE[124]. Cell type labels were transferred from the RNA to the ATAC data via the shared knn-graph using majority voting. To minimise potential batch effects on the label transfer, we restricted the RNA-seq dataset to only in-house Vilarrasa-Blasi bone marrow cells, and downsampled cell populations to 110 (corresponding to the smallest annotation size) per cell fine-grained type and origin (fetal or adult) for HSC, MPP, LMPP, CMP, ELP/CLP, PreProB/LateCLP, EarlyProB, LateProB, LargePreB, SmallPreB, ImmatureB, and PreNaiveB populations. Memory B, plasma, and B1 cells (coming only from adult origin) were downsampled to 220 cells. Peak calling was then repeated per cell annotated type and new cell-by-peak count matrices were created for either all cells, and repeated later for B cell lineage only cells, starting from hematopoietic stem cells to preNaiveB cells. The peak locations reported by ArchR were used to summarise the peak types (promoter/exonic/intronic/distal/). In addition, the peak matrices for fetal and adult bone marrow were used to analyse the conserved and specific peaks. Peaks were ranked according to the proportion of cells exhibiting accessibility. Peaks were defined as conserved if they were in the top 0.1 quantile of this distribution whereas peaks were defined as specific if they were in the bottom 0.1 quantile.

Gene regulatory network (GRN) analysis

The SCENIC+[63] (v1.0.0) pipeline was used to link transcription factors to their putative target genes and regulatory genomic regions through binding sites. First, we used ArchR to build a matrix with fetal-adult consensus peaks using the iterative overlap peak merging procedure[125]. We then subsetted the matrix to include only cell populations from the B cell differentiation trajectory (from hematopoietic stem cells (HSC) to preNaiveB cells) that we used for downstream analyses. Pseudo-multiome meta-cells were created from separate RNA and ATAC by clustering cell barcodes into groups of around 10–15 cells based on a bipartite knn-graph constructed from the shared RNA and ATAC embedding, and then aggregating counts and fragments subsequently. The pipeline was applied twice: first across all bone marrow cells, and second within the B-cell lineage to identify cell state specific transitions. First, CisTopic (pycistopic v1.0.2) was used to identify region topics and differentially accessible regions from the fragment counts, defining candidate cis-regulatory elements (i.e. peak regions). Next, CisTarget (pycistarget v1.0.2) was then applied to scan these peak regions for transcription factor-binding motifs, linking transcription factors to candidate regulatory peak regions. Next, GRNBoost2 (arboreto v0.1.6)[126] was subsequently used to infer transcription factor-target gene relationships based on peak-to-gene co-expression and co-accessibility. Together, these steps allowed us to reconstruct regulons for individual transcription factors. Finally, transcription factor activity was estimated using AUCell, which quantifies regulon activity based on the expression of target genes and derives regulon

specificity scores. Enhancer-driven gene regulatory networks were constructed by integrating transcription factors, regulatory regions and target genes across individual regulons. TF-to-gene importance (TF2G) reflects the strength of association between a transcription factor and its target gene, derived from a gradient boosted model that integrates region-to-gene co-accessibility links with TF motif enrichment within accessible regulatory regions. The number of target regulatory regions corresponds to the count of regulatory elements containing binding motifs for a given TF that are linked to a specific gene.

Coverage ATAC plots for SCENIC+ detected peaks around genes of interest (± 150,000 bp of the start/end of the nearest gene or promoter) across fetal and adult bone marrow cell types were computed from the original fragment files using bedtools (v2.31.0) and pyGenomeTracks package.

SNP2Cell

The SNP2Cell tool[127] was applied to compute SNP scores for the genes and distal regulatory elements forming the gene regulatory networks of B cell development. Briefly, SNP scores were computed from published full GWAS summary statistics and weighted by linkage disequilibrium. SNP scores were then mapped to regulatory elements within the gene regulatory networks, and propagated to nearby TFs and genes in the network[128]. Statistical significance was determined based on random simulations of perturbations. Analogously, marker gene scores for each cell cluster were mapped to the network and propagated to detect cell cluster specific network nodes. Finally, both types of scores were combined to identify cell-type-specific disease-associated sub-networks.

Human orthologue genes for MS-5 mouse stromal

To study the expression of relevant genes in mouse MS-5 cells from the in vitro experiment, we filtered the dual-aligned count matrix (GRCh38 v3.0.0 and mm10 v3.1.0) to include only mm10 genes and mouse-derived cells. Genes expressed in MS-5 stromal cells were matched to their human orthologues using the ortholog_one2one function in Ensembl. The resulting expression matrix of MS-5 cells with human orthologous genes was then combined with that of the human cells in the hiPSC-differentiated dataset to identify ligands specifically produced by the MS-5 cells (with cognate receptors present in our hiPSC-differentiated subpopulations, according to CellPhoneDB database).

Figures

Figure 1. Single-cell multiomics atlas and B lymphopoiesis of the human fetal and adult bone marrow. A, Schematic of study design. Multiomics profiling (scRNA-seq, scVDJ-seq, and scATAC-seq) from fetal and adult bone marrow and integrated with publicly available scRNA-seq datasets[24,25] to define (1) B cell states, (2) antigen receptor repertoire, (3) comparison fetal and adult B lymphopoiesis trajectories (4) reconstruction of gene regulatory networks, (5) mapping of B-cell and bone marrow niche interactions, and (6) development of protocols for *in vitro* B cell modelling. **B,** List of datasets analysed, with donor and cell counts per dataset. **C,** Batch-corrected Uniform Manifold Approximation and Projection (UMAP) of the integrated bone marrow scRNA-seq data from a total of 41 donors and 490,563 cells coloured by cell state. **D,** Schematic illustration of the B-cell differentiation process in the fetal and adult bone marrow. Early lymphoid progenitors (ELPs) and preProB are only defined in the fetal bone marrow. Common lymphoid progenitors (CLPs) and LateCLP are only defined in the adult bone marrow. **E,** Batch-corrected Uniform Manifold Approximation and Projection (UMAP) embedding of the HSC differentiation to B cells containing exclusively cells sequenced for this study coloured by cell type (left) and their origin (right). **F,** Dot plot showing log-transformed, min-max normalised expression of marker genes for adult lateCLP cells. Populations highlighted in grey were included in the differential gene expression analysis between fetal and adult bone marrow (log-fold change >1, p-adjusted <0.05). **G,** Batch-corrected Uniform Manifold Approximation and Projection (UMAP) showing expression of *CD34, CD19*, and *MME* (CD10). **H,** Flow cytometric analysis of the lateCLP population, shown for one representative adult and fetal bone marrow donor. **I,** Frequency of lateCLP cells in adult and PreProB cells in fetal bone marrow as percentage of CD34-positive cyIgM-negative cells from 7 fetal and 7 adult bone marrow donors. Data were analysed using Mann-Whitney test and depicted as ***p<0.001. **J,** Median Fluorescence Intensity (MFI) for CD38, CD24 and PAX5 measured in indicated subpopulations of fetal and adult bone marrow. **K,** Boxplot of the proportion of cells with productive or nonproductive heavy chain (IGH) and light chain (IGK or IGL). Orange colour represents fetal samples, blue colour represents adult samples. N, indicated number of cells. **L,** Box plot representing inverse simpson index for a group of early (largePreB and smallPreB) and late (immatureB and preNaiveB) B cell progenitors. Orange colour represents fetal samples, blue colour represents adult samples. BM, bone marrow; scRNA-seq, single cell RNA sequencing; sc-ATAC-seq, single cell sequencing assay for transposase-accessible chromatin; BCR, B cell receptor; TCR, T cell receptor; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; MEMP, megakaryocyte-erythorid-mast progenitor; GMP, granulocyte-monocyte progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor; Eo, eosinophil; Baso, basophil; MK, megakaryocyte; MOP,

myeloid progenitor; PDC, plasmacytoid dendritic cell; CDP, common dendritic cell; DC1, dendritic cell type 1; DC2, dendritic cell type 2; ILC3, innate lymphoid cell type 3; NK, natural killer; NKT, natural killer-T; T-DP, T cell double positive; T-reg, regulatory T cell; EC/Endo, endothelial; MSC, mesenchymal stromal cell; VSMC, vascular smooth muscle cell; Pre, precursor; Ig, immunoglobulin; IGH, immunoglobulin heavy locus; IGK, immunoglobulin kappa locus; IGL, immunoglobulin lambda locus.

Figure 2. Regulatory programs upregulated in fetal B lymphopoiesis. A, Heatmap showing transcription factor (TF) activity (x-axis) during fetal B lymphopoiesis (y-axis), based on scRNA-seq and scATAC-seq co-modelling with SCENIC+. Cell colour indicates TF expression, dot size reflects chromatin accessibility of TF eRegulon regions, and dot transparency represents overall TF regulon activity. TFs shown are upregulated and differentially active in fetal vs. adult cell types. **B,** Smoothed spline curves showing gene expression dynamics of key TFs upregulated in fetal B lymphopoiesis. **C-D,** Subset of the SCENIC+ inferred gene regulatory networks highlighting the regulatory relationships between the TFs shown in 'A' and 'B' (C; above) and the TFs that regulate the B cell recombination genes *RAG1* and *RAG2* (D; below), which are overexpressed in the fetal bone marrow. **E,** Smoothed spline curves showing log-transformed expression dynamics of B cell recombination genes *RAG1* and *RAG2* upregulated in fetal B lymphopoiesis. **F,** Bar plot showing the expression levels of the targets of EBF1 and ZEB2 upregulated in iB-ALL; error bars represent 95% confidence interval (CI) of the mean. **G,** Subset of the SCENIC+ inferred gene regulatory network showing the targets of EBF1 and ZEB2 that are differentially expressed between iB-ALL with KMTA-MLL1, KMTA-MLL2 and KMTA-MLL4 rearrangement compared to adult bone marrow. **H,** Bar plot showing the expression levels of genes involved in proliferation and B cell differentiation that meet two criteria: (1) are upregulated in iB-ALL with KMTA-MLL1, KMTA-MLL2 and KMTA-MLL4 rearrangements when compared to adult CLPs (log-fold change >1, adjusted p-value<0.01) and (2) are upregulated in fetal ELP when compared to adult CLP (log-fold change >0, adjusted p-value <0.05) (related to Supplementary Note 5). **I,** Smoothed spline curves showing the log-transformed expression of *IL7R* and *ITGA4* in fetal B lymphopoiesis. **J,** Dot plot showing log-transformed, min-max normalised expression of two ligands (*IL7* and *VCAM1*) expressed in the stromal compartment of the fetal bone marrow, cognate ligands of IL7R and ITGA4, interactions identified with CellPhoneDB. BM, bone marrow; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor.

Figure 3. Regulatory programs in adult B lymphopoiesis. A, Heatmap showing transcription factor (TF) activity (x-axis) during adult B lymphopoiesis (y-axis), based on scRNA-seq and scATAC-seq co-modelling with SCENIC+). Cell colour indicates TF expression, dot size reflects chromatin accessibility of TF eRegulon regions, and dot transparency represents overall TF regulon activity. Cell types are ordered by B cell differentiation. TFs shown are upregulated and differentially active in adults vs. fetal cell types. **B,** Smoothed spline curves showing *DNTT* log-transformed gene expression dynamics during fetal and adult B lymphopoiesis. **C-D,** Subsets of the SCENIC+ inferred gene regulatory network showing the TFs (from panel 'A') regulating DNTT involved in B cell V(D)J recombination (C; left) and the TFs (from panel 'A') regulating the marker CD10 (*MME*; D right). **E,** Dot plot showing log-transformed, min-max normalised expression of genes encoding cell-cell interaction molecules (y-axis) involved in B and plasma cell retention and function. Expression is shown across distinct B cell populations in adults (x-axis, left) and stroma/endothelium (x-axis, right). All ligands are upregulated in adult samples (log-fold change >0, adjusted p-value <0.05). Cell-cell interactions were identified with CellPhoneDB. BM, bone marrow; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor.

Figure 4. Characterisation of B cells differentiated from human induced pluripotent stem cells (hiPSC) under feede-based conditions. A, Schematic illustration of 'B cell feeder protocol' to produce hiPSC-derived B cells under feeder conditions. **B,** Dot plot showing log-transformed, min-max normalised expression of primitive and definitive HSCs markersin hiPSC-derived endothelial and progenitor cells. **C,** UMAP embedding of hiPSC-differentiated cells together with cultured fetal bone marrow cells and fetal and adult bone marrow. Coloured by cell lineage (left) and experimental model (right, hiPSC-differentiated: orange, cultured fetal bone marrow: dark grey tone and fetal and adult bone marrow: light grey tone). Total number of 291,332 cells. **D**, Dot plot showing log-transformed, min-max normalised expression of gene markers per cell type of hiPSC-derived B cells (left). Bar plot showing the percentage of cells for each cell state at each time point of the experiment (right). **E,** Gene expression (log-normalised) of B cell related markers and genes (*CD19*, *CD34*, *PTPRC* (CD45), *EBF1*, *MME* (CD10) and *PAX5*) across four B cell differentiation groups in fetal and adult bone marrow and hiPSC-differentiated cells. The line indicated the mean expression across differentiation in vivo (light grey) and in vitro (orange). **F,** Dot plot showing log-transformed, min-max normalised expression of ligands (human orthologous) expressed by MS-5 stromal cells and receptors in hiPSC-B cell differentiated subpopulations. Arrows represent the ligand-receptor communications. The illustration from 'Figure 4A' was partially created using BioRender (https://biorender.com). hiPSC, human induced

pluripotent stem cell; HSC, hematopoietic stem cell; MPP, multipotent progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor.

Figure 5. Generation of hiPSC derived B cells under feeder-free conditions. A, Schematic illustration of 'B cell feeder-free protocol' to produce hiPSC-derived B cells under feeder-free conditions. **B,** Flow cytometry plots for day 34 of culture showing an example of an independent experiment. Indicated is the frequency of CD33-negative cells. **C**, Frequency of CD10-positive/CD19-positive cells across days of culture. **D**, Intracellular staining of PAX5, cytosolic CD79a and cytosolic CD179a (VpreB1) at different days of culture. Negative control is shown in grey. **E**, UMAP embedding of hiPSC-differentiated cells under feeder and feeder-free conditions together with cultured fetal bone marrow cells and fetal and adult bone marrow. Coloured by cell lineage (left) and experimental model (right; hiPSC-differentiated under feeder-free conditions: dark green, hiPSC-differentiated under feeder conditions: orange, cultured fetal bone marrow: dark grey tone and fetal and adult bone marrow: light grey tone). Total number of 291,742 cells. **F**, Dot plot showing log-transformed, min-max normalised expression of B-cell related genes (left) and seven genes (x-axis) involved in the rearrangement of the B-cell receptor and immunoglobulin recombination (right) across early progenitors, ELP/CLP, proB cells (preProB, earlyProB and lateProB) and preB cells (largePreB and smallPreB) from merged fetal and adult bone marrow (y-axis), and across hiPSC-differentiated cells preB cells (largePreB and smallPreB) under feeder and feeder-free conditions. **G**, Percentage of preB cells with productive and non-productive contigs at the immunoglobulin heavy chain (IGH), lambda (IGL) and kappa (IGK) loci from a total of 7,218 cells in feeder conditions and 363 cells in feeder-free conditions. The illustration from 'Figure 5A' was partially created using BioRender (https://biorender.com). HSC, hematopoietic stem cell; MPP, multipotent progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor; hiPSC, human induced pluripotent stem cell; diff., differentiated.

Supplementary Figures

Supplementary Fig. 1 Single-cell RNA-sequencing datasets to build the integrated atlas of the human bone marrow. A, Gating strategies to analyse immune cells in the bone marrow by 10x-genomics in the fetal bone marrow: CD34 enrichment (*top*) and CD19-positive CD10-positive enrichment (*bottom*). **B,** Gating strategies to analyse immune cells in the bone marrow by 10x-genomics in the fetal bone marrow: enrichment of specific subpopulations (*top*), CD34 enrichment (*bottom left*) and CD19-positive CD10-positive (*bottom right*). **C,** Batch-corrected Uniform Manifold Approximation and Projection (UMAP) projections of scRNA-seq data coloured by cell lineage (*left*), cell source (*middle*) and cell cycle phase (*right*). **D**, Dot plot showing log-transformed, min-max normalised expression of gene markers per lineage. **E,** Bar plot showing the percentage of cells coming from the fetal and adult dataset. Cell types with fewer than 50 cells in either group were excluded. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; MEMP, megakaryocyte-erythorid-mast progenitor; GMP, granulocyte-monocyte progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor; Eo, eosinophil; Baso, basophil; MK, megakaryocyte; MOP, myeloid progenitor; PDC, plasmacytoid dendritic cell; CDP, common dendritic cell; DC1, dendritic cell type 1; DC2, dendritic cell type 2; ILC, innate lymphoid cell; ILC3, innate lymphoid cell type 3; NK, natural killer; NKT, natural killer-T; T-DP, T cell double positive; T-reg, regulatory T cell; EC/Endo, endothelial; MSC, mesenchymal stromal cell; VSMC, vascular smooth muscle cell; Pre, precursor.

Supplementary Fig. 2 Markers of cell populations by lineage. A, Dot plot showing log-transformed, min-max normalised expression of gene markers per cell type. **B,** Batch-corrected Uniform Manifold Approximation and Projection (UMAP) projections of integrated human bone marrow highlighting fetal (orange), pediatric (green) and adult (blue) bone marrow dataset. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; MEMP, megakaryocyte-erythorid-mast progenitor; GMP, granulocyte-monocyte progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor; Eo, eosinophil; Baso, basophil; MK, megakaryocyte; MOP, myeloid progenitor; PDC, plasmacytoid dendritic cell; CDP, common dendritic cell; DC1, dendritic cell type 1; DC2, dendritic cell type 2; ILC3, innate lymphoid cell type 3; NK,

natural killer; NKT, natural killer-T; T-DP, T cell double positive; T-reg, regulatory T cell; EC, endothelial; MSC, mesenchymal stromal cell; VSMC, vascular smooth muscle cell; Pre, precursor.

Supplementary Fig. 3 Chromatin accessibility in fetal and adult bone marrow. A, ATAC annotation for all BM subpopulations fetal bone marrow (up) and adult bone marrow (down). Batch-corrected Uniform Manifold Approximation and Projection (UMAP) projections of scRNA-seq data coloured by cell lineage and integrated omics. Populations with fewer than 100 cells and nuclei combined were filtered out. **B**, ATAC annotation for all B-cell lineage fetal BM (left) and adult BM (right). Batch-corrected Uniform Manifold Approximation and Projection (UMAP) projections of scRNA-seq data coloured by cell lineage and integrated omics. BM, bone marrow; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; MEMP, megakaryocyte-erythorid-mast progenitor; GMP, granulocyte-monocyte progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor; Eo, eosinophil; Baso, basophil; MK, megakaryocyte; MOP, myeloid progenitor; PDC, plasmacytoid dendritic cell; CDP, common dendritic cell; DC1, dendritic cell type 1; DC2, dendritic cell type 2; ILC3, innate lymphoid cell type 3; NK, natural killer; NKT, natural killer-T; T-DP, T cell double positive; T-reg, regulatory T cell; EC, endothelial; MSC, mesenchymal stromal cell; VSMC, vascular smooth muscle cell; Pre, precursor.

Supplementary Fig. 4 Characterisation of B cell differentiation in fetal and adult bone marrow by flow cytometry. A, Gating strategies to identify early hematopoietic progenitors. Exemplary gating is shown for a fetal (*top*) and adult bone marrow (*bottom*). Numbers indicate the frequency of population in the parent gate. **B,** Frequencies of early progenitors as percentage from CD34-positive cells in fetal (n=7) and adult (n=6) bone marrow. Data analysed using multiple Mann-Whitney tests and corrected for multiple comparisons by Holm-Šidák correction and depicted as *p<0.05. **C,** Gating strategies to identify early B cell progenitors and mature B cell subsets. Exemplary gating is shown for a fetal (*top*) and adult bone marrow (*bottom*). Numbers indicate the frequency of population in the parent gate. **D,** Frequency of CD10-positive/CD19-positive as percentage of CD33-negative cells. **E,** Frequency of early B cell progenitors (*left*) and of mature B cell subsets (*right*) as percentage of CD10-positive/CD19-positive cells in fetal (n=7) and adult (n=6) bone marrow . Data analysed using (multiple) Mann-Whitney tests and corrected (panel E) for multiple comparisons by Holm-Šidák correction and depicted as *p<0.05, **p<0.01. HSC, hematopoietic stem cell; MPP, multipotent progenitor; LMPP, lympho-myeloid primed progenitor; MLP, multi-lymphoid progenitor, CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, Megakaryocyte erythroid progenitor; CLP, common lymphoid progenitor; DN, double-negative; MZ, marginal zone; Mem, memory; ASC, antibody-secreting cells; Pre, precursor.

Supplementary Fig. 5 Single-cell RNA sequencing of B cell differentiation in prenatal and adult bone marrow samples. A, Dot plot showing log-transformed, min-max normalised expression of gene markers per cell type of B-cell lineage. **B,** Relative proportion of CD10 and CD19 cells. **C,** Density plots of the distinct B cell populations over the pseudotime differentiation in the fetal (*left*) and adult (*right*) bone marrow from Vilarrasa-Blasi bone marrow dataset. **D,** UMAP visualisation of fetal (*left*) and adult (*right*) HSC to B cell differentiation. Trajectories reconstructed with *Slingshot* are overlaid on the embedding from Vilarrasa-Blasi bone marrow dataset. **E,** UMAP visualisation of fetal (*left*) and adult (*right*) HSC to B cell differentiation. Trajectories reconstructed with *Paga* are overlaid on the embedding across all studies. **F,** Density plots of the distinct B cell populations over the pseudotime differentiation in the fetal (*left*) and adult (*right*) bone marrow across all studies. **G,** Batch-corrected Uniform Manifold Approximation and Projection (UMAP) embedding of the HSC differentiation to B cells containing exclusively cells sequenced for this study coloured by donor highlighting preProB in FBM and lateCLP in ABM. **H,** Dot plot showing log-transformed, min-max normalised expression of gene markers expressed for adult lateCLP cells from HSC to late B-cell differentiation. BM, bone marrow; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor.

Supplementary Fig. 6 Characterisation of LateCLP. A, Percentage of cell cycle phases throughout B cell lineage. **B,** Dot plot showing log-transformed, min-max normalised expression of marker genes for lateCLP cells with data stratified to highlight pediatric bone marrow samples across all analysed bone marrow studies. **C,** Flow cytometry gating strategy to detect lateCLP in adult human bone marrow and PreProB in fetal human bone marrow. **D**, Median Fluorescence Intensity (MFI) for indicated markers measured in fetal (n=7 donors) and adult (n=7 donors) bone marrow subpopulations. BM, bone marrow; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor.

Supplementary Fig. 7 B-cell repertoire in fetal and adult populations. A, Bar plot showing the percentage of cells with no contig or contig with specific constant region (IgA, IgD, IgG, IgM, IgM/IgD) per subpopulation. **B,** From left to right, box plots showing the number of N/P nucleotides in VD and DJ junctions (*top*) and CDR3 junction length (*bottom*) of the heavy chain (IGH) (*left*) and light chain (IGK and IGL) (*middle, right*). **C,** Heatmap showing Morisita overlap index between samples. **D,** Box

plot showing Morisita overlap index between samples compared within early B cell progenitors (Large and small PreB) and late B cell progenitors (immatureB and preNaiveB). **E,** Bar plot showing proportion of cells expressing distinct genes for IGHV, IGHD, IGKV, IGLV. Ordered by genomic position (from 5' to 3') and removed genes expressed less than 1% (for both fetal and adult bone marrow). Significant differences are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). BM, bone marrow; FBM, fetal bone marrow; ABM, adult bone marrow; IGH, immunoglobulin heavy locus; IGK, immunoglobulin kappa locus; IGL, immunoglobulin lambda locus; IGHV, immunoglobulin heavy chain variable region; IGHD, immunoglobulin heavy chain diversity region; IGLV, immunoglobulin lambda light chain variable region; IGKL, immunoglobulin kappa light chain variable region.

Supplementary Fig. 8 Gene regulatory programs. A, Fetal activator regulons. **B,** Adult activator regulons. TFs marked with an arrow: PAX5, TCF3 (E2A), FOXO1.

Supplementary Fig. 9 Gene regulatory programs differentially expressed in B cells from the fetal bone marrow. A, Dot plot showing log-transformed, min-max normalised expression of transcription factors differentially expressed and active in fetal B cells, across both fetal and adult B cell populations. **B,** Dot plot showing log-transformed, min-max normalised expression of marker transcription factors with data stratified to highlight pediatric bone marrow samples across all analysed bone marrow studies. **C,** Heatmap showing enrichment scores for the top four genes per four different hematological diseases across B cell differentiation in fetal and adult bone marrow. **D,** Dot plot showing log-transformed, min-max normalised expression of IL-7 target pathways in B cells. BM, bone marrow; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor.

Supplementary Fig. 10 Gene expression inter-variability of key B cell developmental regulators. Gene expression variability across B cell differentiation in fetal and adult bone marrow donors for key B cell transcription factors and developmental markers, including *IKZF2, TCF4, EBF1, LEF1, ZEB2, RAG1, RAG2, IL7R, ITGA4, DNTT, CD79A*, and *VPREB1*.

Supplementary Fig. 11 Gene regulatory programs differentially expressed in B-cells from the adult bone-marrow. A, Smoothed spline curves showing gene expression dynamics of the TFs FOS and FOSB upregulated in adult B lymphopoiesis. **B**, Subset of the SCENIC+ inferred gene regulatory

network showing the regulatory relationships between FOS, FOSB, and the TFs in panel 'Figure 3A'. **C,** Dot plot showing log-transformed, min-max normalised expression of transcription factors differentially expressed and active in adult B cells, in both fetal and adult B cell populations. **D,** Dot plot showing log-transformed, min-max normalised expression of genes encoding cell-cell interaction molecules (y-axis) involved in B and plasma cell retention and function across B cell differentiation in fetal and adult bone marrow. BM, bone marrow; HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor.

Supplementary Fig. 12 Feeder hiPSC-B cell differentiation. **A,** UMAP embedding of the integrated hematopoiesis atlas and hiPSC-differentiated coloured by lineage (*left*) and specific protocols (*right*). **B,** Flow cytometry plots for Richardson *et al.,* protocol in iPSC11-derived cells showing the myeloid bias. The percentage corresponds to the number of events in the gated population relative to the total number of events in the selected upstream population, S, indicating that the gate was used for sorting. **C**, UMAP embedding of the integrated hematopoiesis atlas and hiPSC-differentiated highlighting hiPSC-derived progenitors and coloured by differentiation protocol. **D**, Cell culture images (D8, 10X; D22, 10X; D29, 10X and 20X). **E,** Flow cytometry plots and sorting gates of hiPSC-differentiation at Day 29 (*top*) and peripheral blood mononuclear cell sample used as gating control (*bottom*). The percentage corresponds to the number of events in the gated population relative to the total number of events in the selected upstream population. S, indicates the sorted population or the gate used for sorting. **F**, Flow cytometry plots of cultured fetal bone marrow at Day 29: cultured CD34-positive fetal bone marrow cells (*top*) and cultured CD34-positive CD19-negative fetal bone marrow cells (*bottom*). The percentage corresponds to the number of events in the gated population relative to the total number of events in the selected upstream population. S, indicates the sorted population or the gate used for sorting. **G**, Dot plot showing log-transformed, min-max normalised expression of markers shown in 'Figure 4D'. Indicated number (n) of cells per cell state. **H,** Dot plot showing log-transformed, min-max normalised expression of TSLPR in fetal and adult bone marrow. PBMC, peripheral blood mononuclear cells; AGM, aorta-gonad-mesonephros; KO, knockout; HSC, hematopoietic stem cell; MPP, multipotent progenitor; LMPP, lympho-myeloid primed progenitor; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor; Eo, eosinophil; Baso, basophil.

Supplementary Fig. 13 Feeder-free hiPSC-B cell differentiation. A, Sorted fetal bone marrow pre-B (CD19-positive, CD10-positive) cells cultured for 21 days in either feeder-free or feeder system. Shown are exemplary plots. Numbers indicate frequency of populations in the parent gate. **B,** Frequency

of immature B cells (CD19-positive IgM-positive) as percentage of CD10-positive CD38-positive cells in feeder-free and feedersystems. Each dot represents biological replicate. **C,** Flow cytometry plots for days 27 and 41 of culture showing two independent experiments. Indicated is the frequency of CD33-negative cells. **D,** Dot plot showing log-transformed, min-max normalised expression of B cell related genes (*left*) and seven genes involved in the rearrangement of the B-cell receptor and immunoglobulin recombination (*right*) across early progenitors, ELP/CLP, proB cells (preProB, earlyProB and lateProB) and preB cells (largePreB and smallPreB) from fetal and adult bone marrow, and across hiPSC-differentiated preB cells (largePreB and smallPreB) under feeder and feeder-free conditions. **E,** Dot plot showing log-transformed, min-max normalised expression of selected genes drivers of leukemia and altered in immunodeficiency during fetal and adult bone marrow B-cell differentiation and hiPSC-differentiation. BM, bone marrow; hiPSC, human induced pluripotent stem cell; BM, bone marrow; diff., differentiated; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor.

Supplementary Tables

Supplementary Table 1. Metadata of samples analysed. A, 10x scRNAseq libraries. **B,** 10x scATAC-seq and 10x cell-coupled snRNA/ATAC-seq multiomic libraries.

Supplementary Table 2. Quality control of samples. A, Summary statistics from Cell Ranger for each 10x scRNA-seq library that we sequenced. **B,** Summary statistics from 10x Cell Ranger ATAC 2.0.0 (scATAC-seq) and 10x Cell Ranger ARC 2.0.0 (multiomics) for each 10x ATAC library that we sequenced.

Supplementary Table 3. Absolute cell counts per donor and dataset. A. Cell counts 'Figure 1C'. **B** Cell counts 'Figure 1E'.

Supplementary Table 4. Differentially expressed genes during fetal and adult B lymphopoiesis.

Supplementary Table 5. CellPhoneDB relevant interaction lists. A, Relevant interaction list from receptor differential gene expression analysis in fetal and adult pairwise subpopulations during B cell differentiation. **B,** Relevant interactions from ligand differential gene expression analysis in fetal and adult pairwise stromal cells.

Supplementary Table 6. Regulome for IKZF2, TCF4, AIRE, EBF1, LEF1, ZEB2 in fetal B cells.

Supplementary Table 7. Differentially commonly expressed genes in iB-ALL vs adult.

Supplementary Table 8. Absolute cell counts hiPSC B-cell differentiation per time point. Cell numbers from 'Figure 4D'.

Supplementary Table 9. List of antibodies for adult bone marrow.

References

1. Lin, Y. C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* **11**, 635–643 (2010).
2. Young, F. *et al.* Influence of immunoglobulin heavy- and light-chain expression on B-cell differentiation. *Genes Dev.* **8**, 1043–1057 (1994).
3. Clark, M. R., Mandal, M., Ochiai, K. & Singh, H. Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. *Nat. Rev. Immunol.* **14**, 69–80 (2014).
4. Tokoyoda, K., Egawa, T., Sugiyama, T., Choi, B.-I. & Nagasawa, T. Cellular niches controlling B lymphocyte behavior within bone marrow during development. *Immunity* **20**, 707–718 (2004).
5. Nakatani, T. *et al.* Ebf3+ niche-derived CXCL12 is required for the localization and maintenance of hematopoietic stem cells. *Nat. Commun.* **14**, 6402 (2023).
6. Pelanda, R. *et al.* B-cell intrinsic and extrinsic signals that regulate central tolerance of mouse and human B cells. *Immunol. Rev.* **307**, 12–26 (2022).
7. Zheng, Z. *et al.* Uncovering the emergence of HSCs in the human fetal bone marrow by single-cell RNA-seq analysis. *Cell Stem Cell* **29**, 1562–1579.e7 (2022).
8. Prashad, S. L. *et al.* GPI-80 defines self-renewal ability in hematopoietic stem cells during human development. *Cell Stem Cell* **16**, 80–87 (2015).
9. Calvanese, V. & Mikkola, H. K. A. The genesis of human hematopoietic stem cells. *Blood* **142**, 519–532 (2023).
10. Korzhenevich, J., Janowska, I., van der Burg, M. & Rizzi, M. Human and mouse early B cell development: So similar but so different. *Immunology Letters* **261**, 1–12 (2023).
11. 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).
12. Prieyl, J. A. & LeBien, T. W. Interleukin 7 independent development of human B cells. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10348–10353 (1996).
13. Minegishi, Y. *et al.* An essential role for BLNK in human B cell development. *Science* **286**, 1954–1957 (1999).
14. Pappu, R. *et al.* Requirement for B cell linker protein (BLNK) in B cell development. *Science* **286**, 1949–1954 (1999).
15. Okuyama, K. *et al.* PAX5 is part of a functional transcription factor network targeted in lymphoid leukemia. *PLoS Genet.* **15**, e1008280 (2019).
16. Mielczarek, O. *et al.* Intra- and interchromosomal contact mapping reveals the Igh locus has extensive conformational heterogeneity and interacts with B-lineage genes. *Cell Rep.* **42**, 113074 (2023).
17. Sigvardsson, M. Transcription factor networks link B-lymphocyte development and malignant transformation in leukemia. *Genes Dev.* **37**, 703–723 (2023).
18. Jackson, T. R., Ling, R. E. & Roy, A. The origin of B-cells: Human fetal B cell development and implications for the pathogenesis of childhood acute lymphoblastic leukemia. *Front. Immunol.* **12**, 637975 (2021).
19. de Smith, A. J. *et al.* Backtracking to the future: unraveling the origins of childhood leukemia. *Leukemia* **38**, 416–419 (2024).
20. 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).
21. Greaves, M. F., Maia, A. T., Wiemels, J. L. & Ford, A. M. Leukemia in twins: lessons in natural history. *Blood* **102**, 2321–2333 (2003).
22. Greaves, M. F. & Wiemels, J. Origins of chromosome translocations in childhood leukaemia. *Nat. Rev. Cancer* **3**, 639–649 (2003).
23. Khabirova, E. *et al.* Single-cell transcriptomics reveals a distinct developmental state of KMT2A-rearranged infant B-cell acute lymphoblastic leukemia. *Nat. Med.* **28**, 743–751 (2022).

24.Jardine, L. *et al.* Blood and immune development in human fetal bone marrow and Down syndrome. *Nature* **598**, 327–331 (2021).
25.Bandyopadhyay, S. *et al.* Mapping the cellular biogeography of human bone marrow niches using single-cell transcriptomics and proteomic imaging. *Cell* **187**, 3120–3140.e29 (2024).
26.Oetjen, K. A. *et al.* Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight* **3**, (2018).
27.Granja, J. M. *et al.* Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37**, 1458–1465 (2019).
28.Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).
29.Caron, M. *et al.* Single-cell analysis of childhood leukemia reveals a link between developmental states and ribosomal protein expression as a source of intra-individual heterogeneity. *Scientific reports* **10**, (2020).
30.Ainciburu, M. *et al.* Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single-cell resolution. *Elife* **12**, (2023).
31.Kaiser, F. M. P. *et al.* IL-7 receptor signaling drives human B-cell progenitor differentiation and expansion. *Blood* **142**, 1113–1130 (2023).
32.Zhang, X. *et al.* An immunophenotype-coupled transcriptomic atlas of human hematopoietic progenitors. *Nat. Immunol.* **25**, 703–715 (2024).
33.A single cell immune cell atlas of human hematopoietic system. https://explore.data.humancellatlas.org/projects/cc95ff89-2e68-4a08-a234-480eca21ce79.
34.Carpenter, L. *et al.* Human induced pluripotent stem cells are capable of B-cell lymphopoiesis. *Blood* **117**, 4008–4011 (2011).
35.French, A., Yang, C.-T., Taylor, S., Watt, S. M. & Carpenter, L. Human induced pluripotent stem cell-derived B lymphocytes express sIgM and can be generated via a hemogenic endothelium intermediate. *Stem Cells Dev.* **24**, 1082–1095 (2015).
36.Dubois, F. *et al.* Toward a better definition of hematopoietic progenitors suitable for B cell differentiation. *PLoS One* **15**, e0243769 (2020).
37.Richardson, S. E., Ghazanfari, R., Chhetri, J., Enver, T. & Böiers, C. In vitro differentiation of human pluripotent stem cells into the B lineage using OP9-MS5 co-culture. *STAR Protoc.* **2**, 100420 (2021).
38.Vick, D. J., Hogge, W. A., Normansell, D. E., Burkett, B. J. & Harbert, G. M., Jr. Determination of normal human fetal immunoglobulin M levels. *Clin. Diagn. Lab. Immunol.* **2**, 115–117 (1995).
39.Suo, C. *et al.* Mapping the developing human immune system across organs. *Science* **376**, eabo0510 (2022).
40.Flach, H. *et al.* Mzb1 protein regulates calcium homeostasis, antibody secretion, and integrin activation in innate-like B cells. *Immunity* **33**, 723–735 (2010).
41.Griffin, D. O., Holodick, N. E. & Rothstein, T. L. Human B1 cells in umbilical cord and adult peripheral blood express the novel phenotype CD20+ CD27+ CD43+ CD70-. *J. Exp. Med.* **208**, 67–80 (2011).
42.Keita, S. *et al.* Distinct subsets of multi-lymphoid progenitors support ontogeny-related changes in human lymphopoiesis. *Cell Rep.* **42**, 112618 (2023).
43.Lee, R. D. *et al.* Single-cell analysis identifies dynamic gene expression networks that govern B cell development and transformation. *Nat. Commun.* **12**, 6843 (2021).
44.Gómez Hernández, G., Domínguez, T., Galicia, G., Morell, M. & Alarcón-Riquelme, M. E. Bank1 modulates the differentiation and molecular profile of key B cell populations in autoimmunity. *JCI Insight* **9**, (2024).
45.Saijo, K. *et al.* Essential role of Src-family protein tyrosine kinases in NF-kappaB activation during B cell development. *Nat. Immunol.* **4**, 274–279 (2003).
46.Israel, E. *et al.* Expression of CD24 on CD19- CD79a+ early B-cell progenitors in human bone marrow. *Cell. Immunol.* **236**, 171–178 (2005).
47.Hao, Q. L. *et al.* Human intrathymic lineage commitment is marked by differential CD7 expression: identification of CD7- lympho-myeloid thymic progenitors. *Blood* **111**, (2008).

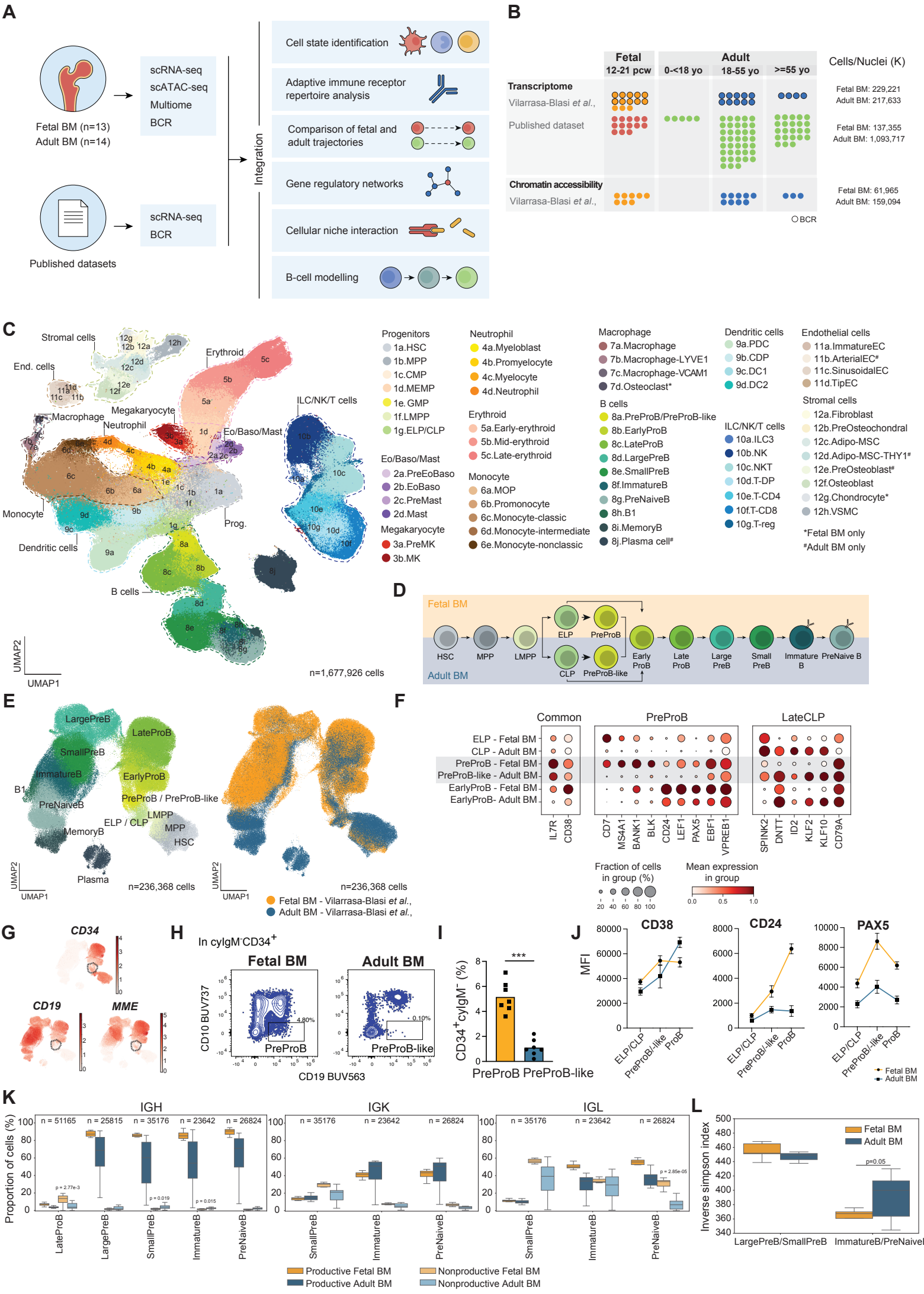
48.Reya, T. *et al.* Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. *Immunity* **13**, 15–24 (2000).
49.Jensen, K. *et al.* Increased ID2 levels in adult precursor B cells as compared with children is associated with impaired Ig locus contraction and decreased bone marrow output. *J. Immunol.* **191**, 1210–1219 (2013).
50.van Zelm, M. C. *et al.* Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. *J. Immunol.* **175**, 5912–5922 (2005).
51.Kharas, M. G. *et al.* KLF4 suppresses transformation of pre-B cells by ABL oncogenes. *Blood* **109**, 747–755 (2007).
52.Yusuf, I. *et al.* KLF4 is a FOXO target gene that suppresses B cell proliferation. *Int. Immunol.* **20**, 671–681 (2008).
53.Calvanese, V. *et al.* Mapping human haematopoietic stem cells from haemogenic endothelium to birth. *Nature* **604**, 534–540 (2022).
54.Suo, C. *et al.* Dandelion uses the single-cell adaptive immune receptor repertoire to explore lymphocyte developmental origins. *Nat. Biotechnol.* **42**, 40–51 (2024).
55.Alt, F. W. *et al.* Ordered rearrangement of immunoglobulin heavy chain variable region segments. *The EMBO journal* **3**, (1984).
56.Nishimoto, N. *et al.* Normal pre-B cells express a receptor complex of mu heavy chains and surrogate light-chain proteins. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 6284–6288 (1991).
57.Rother, M. B. *et al.* Decreased IL7Rα and TdT expression underlie the skewed immunoglobulin repertoire of human B-cell precursors from fetal origin. *Sci. Rep.* **6**, 33924 (2016).
58.Schroeder, H. W., Jr & Wang, J. Y. Preferential utilization of conserved immunoglobulin heavy chain variable gene segments during human fetal life. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6146–6150 (1990).
59.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).
60.Rechavi, E. *et al.* Timely and spatially regulated maturation of B and T cell repertoire during human fetal development. *Sci. Transl. Med.* **7**, 276ra25 (2015).
61.Chen, J. W. *et al.* Autoreactivity in naïve human fetal B cells is associated with commensal bacteria recognition. *Science* **369**, 320–325 (2020).
62.Busslinger, M. Transcriptional control of early B cell development. *Annu. Rev. Immunol.* **22**, 55–79 (2004).
63.Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* 1–13 (2023).
64.Troulé, K. *et al.* CellPhoneDB v5: inferring cell-cell communication from single-cell multiomics data. *arxiv* (2023) doi: https://doi.org/10.48550/arXiv.2311.04567.
65.Park, S.-M. *et al.* IKZF2 drives leukemia stem cell self-renewal and inhibits myeloid differentiation. *Cell Stem Cell* **24**, 153–165.e7 (2019).
66.in 't Hout, F. E. M., van der Reijden, B. A., Monteferrario, D., Jansen, J. H. & Huls, G. High expression of transcription factor 4 (TCF4) is an independent adverse prognostic factor in acute myeloid leukemia that could guide treatment decisions. *Haematologica* **99**, e257–9 (2014).
67.Ramamoorthy, S. *et al.* EBF1 and Pax5 safeguard leukemic transformation by limiting IL-7 signaling, Myc expression, and folate metabolism. *Genes Dev.* **34**, 1503–1519 (2020).
68.Gutierrez, A., Jr *et al.* LEF-1 is a prosurvival factor in chronic lymphocytic leukemia and is expressed in the preleukemic state of monoclonal B-cell lymphocytosis. *Blood* **116**, 2975–2983 (2010).
69.Li, H. *et al.* The EMT regulator ZEB2 is a novel dependency of human and murine acute myeloid leukemia. *Blood* **129**, 497–508 (2017).
70.Yu, X. *et al.* RAG1 and RAG2 non-core regions are implicated in leukemogenesis and off-target V(D)J recombination in BCR-ABL1-driven B-cell lineage lymphoblastic leukemia. *Elife* **12**, (2024).
71.Tran, T. M. *et al.* The RNA-binding protein IGF2BP3 is critical for MLL-AF4-mediated leukemogenesis. *Leukemia* **36**, 68–79 (2022).

72.Vitali, C. *et al.* SOCS2 controls proliferation and stemness of hematopoietic cells under stress conditions and its deregulation marks unfavorable acute leukemias. *Cancer Res.* **75**, 2387–2399 (2015).
73.Kwon, A. *et al.* Rare circulating lymphoblasts with striking eosinophilia: A rare case of B-lymphoblastic leukemia with PAX5::ZCCHC7. *Am. J. Hematol.* **98**, 989–990 (2023).
74.Li, Y., Zhang, Q. & Shao, H. Recurrent PAX5::ZCCHC7 rearrangement in B-cell acute lymphoblastic leukemia. *Ann. Hematol.* **103**, 5599–5605 (2024).
75.Parrish, Y. K. *et al.* IL-7 Dependence in human B lymphopoiesis increases during progression of ontogeny from cord blood to bone marrow. *J. Immunol.* **182**, 4255–4266 (2009).
76.Tsapogas, P. *et al.* IL-7 mediates Ebf-1-dependent lineage restriction in early lymphoid progenitors. *Blood* **118**, 1283–1290 (2011).
77.Corfe, S. A. & Paige, C. J. The many roles of IL-7 in B cell development; mediator of survival, proliferation and differentiation. *Semin. Immunol.* **24**, 198–208 (2012).
78.Geron, I. *et al.* An instructive role for Interleukin-7 receptor α in the development of human B-cell precursor leukemia. *Nat. Commun.* **13**, 659 (2022).
79.Attia, H. R. *et al.* ITGA4 gene methylation status in chronic lymphocytic leukemia. *Future Sci. OA* **6**, FSO583 (2020).
80.Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25**, 977–988 (2006).
81.Alves da Costa, T. *et al.* Central human B cell tolerance manifests with a distinctive cell phenotype and is enforced via CXCR4 signaling in hu-mice. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2021570118 (2021).
82.Rafei, M. *et al.* Mesenchymal stromal cell-derived CCL2 suppresses plasma cell immunoglobulin production via STAT3 inactivation and PAX5 induction. *Blood* **112**, 4991–4998 (2008).
83.Dasoveanu, D. C. *et al.* Lymph node stromal CCL2 limits antibody responses. *Sci. Immunol.* **5**, eaaw0693 (2020).
84.Wang, W. *et al.* Identification of a novel chemokine (CCL28), which binds CCR10 (GPR2). *J. Biol. Chem.* **275**, 22313–22323 (2000).
85.Mei, H. E. *et al.* Blood-borne human plasma cells in steady state are derived from mucosal immune responses. *Blood* **113**, 2461–2469 (2009).
86.Böiers, C. *et al.* A Human IPS Model Implicates Embryonic B-Myeloid Fate Restriction as Developmental Susceptibility to B Acute Lymphoblastic Leukemia-Associated ETV6-RUNX1. *Dev. Cell* **44**, 362–377.e7 (2018).
87.Michaels, Y. S. *et al.* DLL4 and VCAM1 enhance the emergence of T cell-competent hematopoietic progenitors from human pluripotent stem cells. *Sci. Adv.* **8**, eabn5522 (2022).
88.Arakelian, L. *et al.* Endothelial CD34 expression and regulation of immune cell response in-vitro. *Sci. Rep.* **13**, 13512 (2023).
89.Dooley, D. C., Oppenlander, B. K. & Xiao, M. Analysis of primitive CD34- and CD34+ hematopoietic cells from adults: gain and loss of CD34 antigen by undifferentiated cells are closely linked to proliferative status in culture. *Stem Cells* **22**, 556–569 (2004).
90.Lee, G., Ellingsworth, L. R., Gillis, S., Wall, R. & Kincade, P. W. Beta transforming growth factors are potential regulators of B lymphopoiesis. *J. Exp. Med.* **166**, 1290–1299 (1987).
91.Thiel, J. *et al.* Defects in B-lymphopoiesis and B-cell maturation underlie prolonged B-cell depletion in ANCA-associated vasculitis. *Ann. Rheum. Dis.* **83**, 1536–1548 (2024).
92.Ditadi, A. *et al.* Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat. Cell Biol.* **17**, 580–591 (2015).
93.Michaels, Y. S. *et al.* DLL4 and VCAM1 enhance the emergence of T cell-competent hematopoietic progenitors from human pluripotent stem cells. *Sci. Adv.* **8**, eabn5522 (2022).
94.Shen, Q. *et al.* TCF1 and LEF1 promote B-1a cell homeostasis and regulatory function. *Nature* **646**, 442–451 (2025).
95.Tangye, S. G. To B1 or not to B1: that really is still the question! *Blood* **121**, 5109–5110 (2013).

96.Reynaud, C.-A. & Weill, J.-C. Gene profiling of $CD11b^+$ and $CD11b^-$ B1 cell subsets reveals potential cell sorting artifacts. *J. Exp. Med.* **209**, 433–434 (2012).
97.Montecino-Rodriguez, E., Leathers, H. & Dorshkind, K. Identification of a B-1 B cell-specified progenitor. *Nat. Immunol.* **7**, 293–301 (2006).
98.Ghosn, E. E. B., Sadate-Ngatchou, P., Yang, Y., Herzenberg, L. A. & Herzenberg, L. A. Distinct progenitors for B-1 and B-2 cells are present in adult mouse spleen. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2879–2884 (2011).
99.Smith, F. L. & Baumgarth, N. B-1 cell responses to infections. *Curr. Opin. Immunol.* **57**, 23–31 (2019).
100. Rechavi, E. & Somech, R. Survival of the fetus: fetal B and T cell receptor repertoire development. *Semin. Immunopathol.* **39**, 577–583 (2017).
101. Melchers, F. B cell development and its deregulation to transformed states at the pre-B cell receptor-expressing pre-BII cell stage. *Curr. Top. Microbiol. Immunol.* **294**, 1–17 (2005).
102. Le, B. L. *et al.* Maternal and infant immune repertoire sequencing analysis identifies distinct ig and TCR development in term and preterm infants. *J. Immunol.* **207**, 2445–2455 (2021).
103. Li, Z. *et al.* Presence of onco-fetal neighborhoods in hepatocellular carcinoma is associated with relapse and response to immunotherapy. *Nat. Cancer* **5**, 167–186 (2024).
104. Kraus, H. *et al.* A feeder-free differentiation system identifies autonomously proliferating B cell precursors in human bone marrow. *J. Immunol.* **192**, 1044–1054 (2014).
105. Michaels, Y. S. *et al.* DLL4 and VCAM1 enhance the emergence of T cell-competent hematopoietic progenitors from human pluripotent stem cells. *bioRxiv* 2021.11.26.470145 (2021) doi:10.1101/2021.11.26.470145.
106. Goh, I. *et al.* Yolk sac cell atlas reveals multiorgan functions during human early development. *Science* **381**, eadd7564 (2023).
107. Webb, S., Stephenson, E. & Haniffa, M. Human embryonic liver CITE-seq data. *BioStudies, E-MTAB-11618* **https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-11618**, (2022).
108. Stephenson, E., Webb, S. & Haniffa, M. Human fetal liver CITE-seq data. *BioStudies, E-MTAB-11613* **https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-11613**, (2022).
109. Wang, H. *et al.* Decoding Human Megakaryocyte Development. *Cell Stem Cell* **28**, 535–549.e8 (2021).
110. Popescu, D.-M. *et al.* Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
111. Alsinet, C. *et al.* Robust temporal map of human in vitro myelopoiesis using single-cell genomics. *bioRxiv* 2021.11.17.469005 (2021) doi:10.1101/2021.11.17.469005.
112. Heaton, H. *et al.* Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).
113. Bernstein, N. J. *et al.* Solo: Doublet identification in single-cell RNA-seq via semi-supervised deep learning. *Cell Syst.* **11**, 95–101.e5 (2020).
114. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
115. Park, J.-E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, eaay3224 (2020).
116. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
117. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020).
118. Domínguez Conde, C. *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).
119. ImmunoMind Team. Immunarch: An R Package for Painless Analysis of Large-Scale Immune Repertoire Data. *Zenodo* (2019).
120. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

121. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
122. Van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1–13 (2020).
123. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
124. Cao, Z.-J. & Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458–1466 (2022).
125. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
126. Moerman, T. *et al.* GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics (Oxford, England)* **35**, (2019).
127. To, K. *et al.* A multi-omic atlas of human embryonic skeletal development. *Nature* **635**, 657–667 (2024).
128. Cowen, L., Ideker, T., Raphael, B. J. & Sharan, R. Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* **18**, 551–562 (2017).

Figure 1


A
Fetal BM (n=13)
Adult BM (n=14)
scRNA-seq
scATAC-seq
Multiome
BCR
Published datasets
scRNA-seq
BCR
Integration
Cell state identification
Adaptive immune receptor repertoire analysis
Comparison of fetal and adult trajectories
Gene regulatory networks
Cellular niche interaction
B-cell modelling
B
Fetal
12-21 pcw
Adult
0-<18 yo
18-55 yo
>=55 yo
Cells/Nuclei (K)
Transcriptome
Vilarrasa-Blasi et al.,
Published dataset
Fetal BM: 229,221
Adult BM: 217,633
Fetal BM: 137,355
Adult BM: 1,093,717
Chromatin accessibility
Vilarrasa-Blasi et al.,
Fetal BM: 61,965
Adult BM: 159,094
BCR
C
Stromal cells
End. cells
Erythroid
Megakaryocyte
Macrophage
Neutrophil
Eo/Baso/Mast
ILC/NK/T cells
Monocyte
Prog.
Dendritic cells
B cells
UMAP2
UMAP1
n=1,677,926 cells
Progenitors
1a.HSC
1b.MPP
1c.CMP
1d.MEMP
1e.GMP
1f.LMPP
1g.ELP/CLP
Eo/Baso/Mast
2a.PreEoBaso
2b.EoBaso
2c.PreMast
2d.Mast
Megakaryocyte
3a.PreMK
3b.MK
Neutrophil
4a.Myeloblast
4b.Promyelocyte
4c.Myelocyte
4d.Neutrophil
Erythroid
5a.Early-erythroid
5b.Mid-erythroid
5c.Late-erythroid
Monocyte
6a.MOP
6b.Promonocyte
6c.Monocyte-classic
6d.Monocyte-intermediate
6e.Monocyte-nonclassic
Macrophage
7a.Macrophage
7b.Macrophage-LYVE1
7c.Macrophage-VCAM1
7d.Osteoclast*
B cells
8a.PreProB/PreProB-like
8b.EarlyProB
8c.LateProB
8d.LargePreB
8e.SmallPreB
8f.ImmatureB
8g.PreNaiveB
8h.B1
8i.MemoryB
8j.Plasma cell#
Dendritic cells
9a.PDC
9b.CDP
9c.DC1
9d.DC2
ILC/NK/T cells
10a.ILC3
10b.NK
10c.NKT
10d.T-DP
10e.T-CD4
10f.T-CD8
10g.T-reg
Endothelial cells
11a.ImmatureEC
11b.ArterialEC#
11c.SinusoidalEC
11d.TipEC
Stromal cells
12a.Fibroblast
12b.PreOsteochondral
12c.Adipo-MSC
12d.Adipo-MSC-THY1#
12e.PreOsteoblast#
12f.Osteoblast
12g.Chondrocyte*
12h.VSMC
*Fetal BM only
#Adult BM only
D
Fetal BM
Adult BM
HSC
MPP
LMPP
ELP
CLP
PreProB
PreProB-like
Early ProB
Late ProB
Large PreB
Small PreB
Immature B
PreNaive B
E
LargePreB
LateProB
SmallPreB
ImmatureB
EarlyProB
B1
PreProB / PreProB-like
PreNaiveB
LMPP
ELP / CLP
MPP
MemoryB
HSC
Plasma
UMAP2
UMAP1
n=236,368 cells
n=236,368 cells
Fetal BM - Vilarrasa-Blasi et al.,
Adult BM - Vilarrasa-Blasi et al.,
F
Common
PreProB
LateCLP
ELP - Fetal BM
CLP - Adult BM
PreProB - Fetal BM
PreProB-like - Adult BM
EarlyProB - Fetal BM
EarlyProB- Adult BM
IL7R
CD38
CD7
MS4A1
BANK1
BLK
CD24
LEF1
PAX5
EBF1
VPREB1
SPINK2
DNTT
ID2
KLF2
KLF10
CD79A
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
G
CD34
CD19
MME
H
In cyIgM-CD34+
Fetal BM
Adult BM
CD10 BUV737
CD19 BUV563
4.80%
0.10%
PreProB
PreProB-like
I
CD34+cyIgM- (%)
***
PreProB PreProB-like
J
MFI
CD38
CD24
PAX5
ELP/CLP
PreProB/-like
ProB
Fetal BM
Adult BM
K
IGH
IGK
IGL
Proportion of cells (%)
n = 51165
n = 25815
n = 35176
n = 23642
n = 26824
p = 2.77e-3
p = 0.019
p = 0.015
p = 2.85e-05
LateProB
LargePreB
SmallPreB
ImmatureB
PreNaiveB
Productive Fetal BM
Productive Adult BM
Nonproductive Fetal BM
Nonproductive Adult BM
L
Inverse simpson index
Fetal BM
Adult BM
p=0.05
LargePreB/SmallPreB
ImmatureB/PreNaiveB

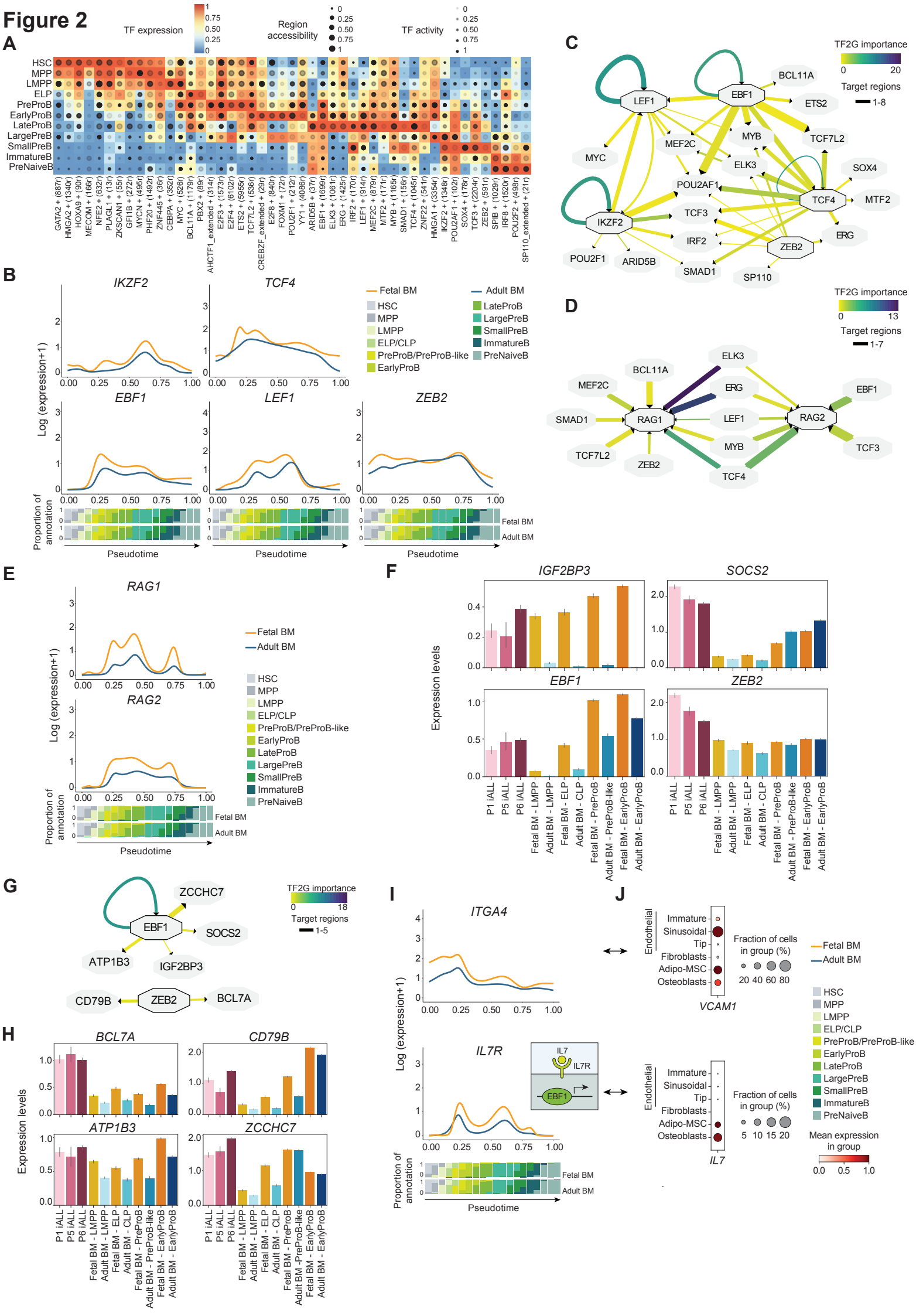


Figure 2
A
TF expression
Region accessibility
TF activity
HSC
MPP
LMPP
ELP
PreProB
EarlyProB
LateProB
LargePreB
SmallPreB
ImmatureB
PreNaiveB
B
IKZF2
TCF4
EBF1
LEF1
ZEB2
Fetal BM
Adult BM
HSC
MPP
LMPP
ELP/CLP
PreProB/PreProB-like
EarlyProB
LateProB
LargePreB
SmallPreB
ImmatureB
PreNaiveB
Log (expression+1)
Proportion of annotation
Pseudotime
C
TF2G importance
Target regions
LEF1
EBF1
BCL11A
ETS2
MYB
TCF7L2
MEF2C
MYC
ELK3
SOX4
POU2AF1
TCF4
MTF2
TCF3
IKZF2
IRF2
ERG
ZEB2
POU2F1
ARID5B
SMAD1
SP110
D
TF2G importance
Target regions
ELK3
BCL11A
MEF2C
ERG
EBF1
SMAD1
RAG1
LEF1
RAG2
MYB
TCF3
TCF7L2
ZEB2
TCF4
E
RAG1
RAG2
Fetal BM
Adult BM
HSC
MPP
LMPP
ELP/CLP
PreProB/PreProB-like
EarlyProB
LateProB
LargePreB
SmallPreB
ImmatureB
PreNaiveB
Log (expression+1)
Proportion of annotation
Pseudotime
F
IGF2BP3
SOCS2
EBF1
ZEB2
Expression levels
P1 iALL
P5 iALL
P6 iALL
Fetal BM - LMPP
Adult BM - LMPP
Fetal BM - ELP
Adult BM - CLP
Fetal BM - PreProB
Adult BM - PreProB-like
Fetal BM - EarlyProB
Adult BM - EarlyProB
G
TF2G importance
Target regions
ZCCHC7
EBF1
SOCS2
ATP1B3
IGF2BP3
CD79B
ZEB2
BCL7A
H
BCL7A
CD79B
ATP1B3
ZCCHC7
Expression levels
I
ITGA4
IL7R
IL7
IL7R
EBF1
Log (expression+1)
Proportion of annotation
Pseudotime
Fetal BM
Adult BM
J
Endothelial
Immature
Sinusoidal
Tip
Fibroblasts
Adipo-MSC
Osteoblasts
Fraction of cells in group (%)
VCAM1
IL7
Mean expression in group

Figure 3

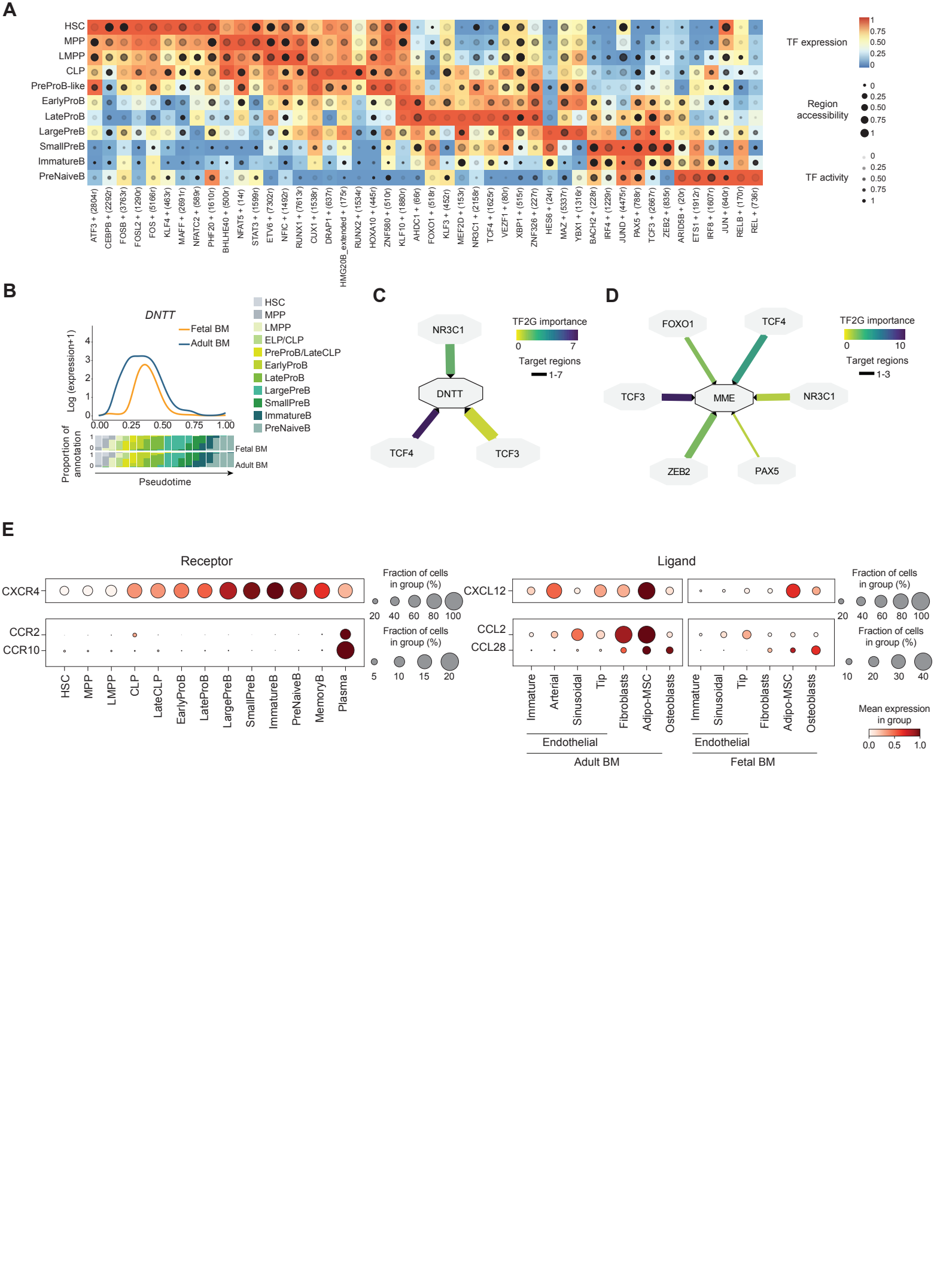

A
HSC
MPP
LMPP
CLP
PreProB-like
EarlyProB
LateProB
LargePreB
SmallPreB
ImmatureB
PreNaiveB
TF expression
Region accessibility
TF activity
ATF3 + (2804r)
CEBPB + (2292r)
FOSB + (3763r)
FOSL2 + (1290r)
FOS + (5166r)
KLF4 + (463r)
MAFF + (2691r)
NFATC2 + (589r)
PHF20 + (1610r)
BHLHE40 + (500r)
NFATS + (14r)
STAT3 + (1599r)
ETV6 + (7302r)
NFIC + (1492r)
RUNX1 + (7613r)
CUX1 + (1538r)
DRAP1 + (637r)
HMG20B_extended + (175r)
RUNX2 + (1534r)
HOXA10 + (445r)
ZNF580 + (510r)
KLF10 + (1880r)
AHDC1 + (66r)
FOXO1 + (518r)
KLF3 + (452r)
MEF2D + (153r)
NR3C1 + (2158r)
TCF4 + (1625r)
VEZF1 + (80r)
XBP1 + (515r)
ZNF326 + (227r)
HES6 + (24r)
MAZ + (5337r)
YBX1 + (1316r)
BACH2 + (228r)
IRF4 + (1229r)
JUND + (4475r)
PAX5 + (788r)
TCF3 + (2667r)
ZEB2 + (835r)
ARID5B + (20r)
ETS1 + (1912r)
IRF8 + (1607r)
JUN + (640r)
RELB + (170r)
REL + (736r)
B
DNTT
Fetal BM
Adult BM
Log (expression+1)
Proportion of annotation
Pseudotime
HSC
MPP
LMPP
ELP/CLP
PreProB/LateCLP
EarlyProB
LateProB
LargePreB
SmallPreB
ImmatureB
PreNaiveB
C
NR3C1
DNTT
TCF4
TCF3
TF2G importance
Target regions
1-7
D
FOXO1
TCF4
TCF3
MME
NR3C1
ZEB2
PAX5
TF2G importance
Target regions
1-3
E
Receptor
Ligand
CXCR4
CCR2
CCR10
CXCL12
CCL2
CCL28
Fraction of cells in group (%)
HSC
MPP
LMPP
CLP
LateCLP
EarlyProB
LateProB
LargePreB
SmallPreB
ImmatureB
PreNaiveB
MemoryB
Plasma
Immature
Arterial
Sinusoidal
Tip
Fibroblasts
Adipo-MSC
Osteoblasts
Endothelial
Adult BM
Fetal BM
Mean expression in group

Figure 4

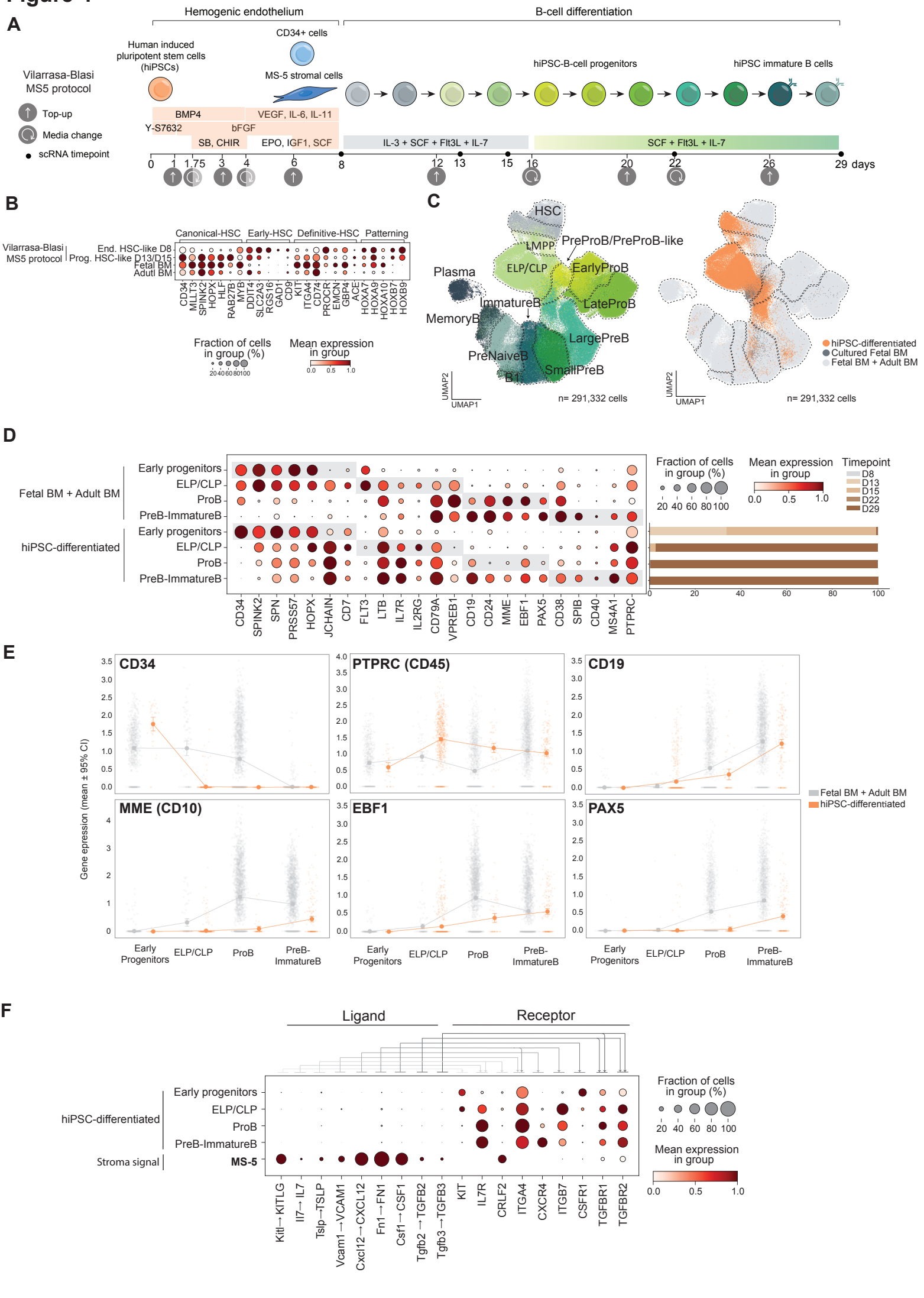
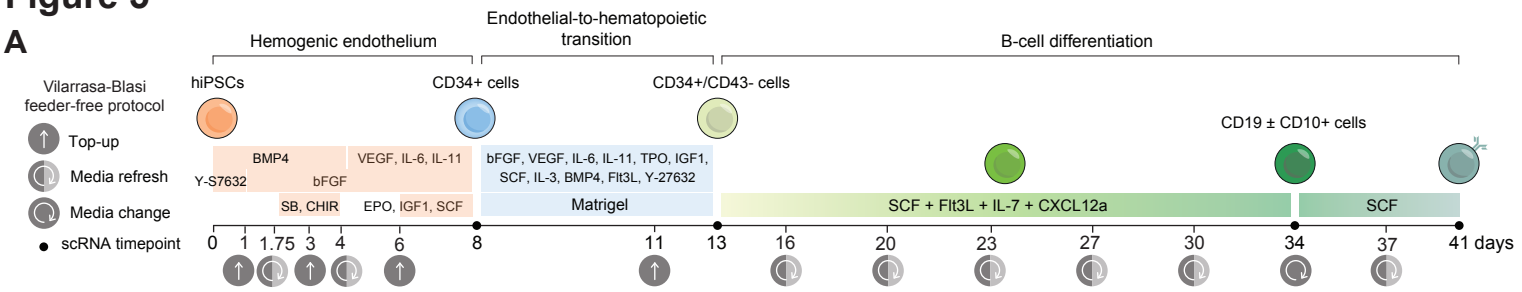
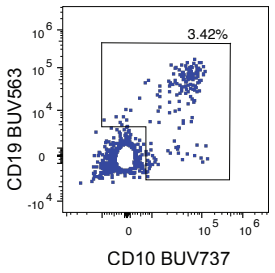
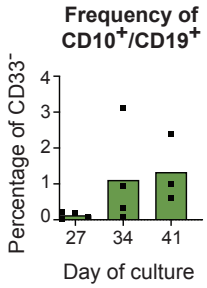
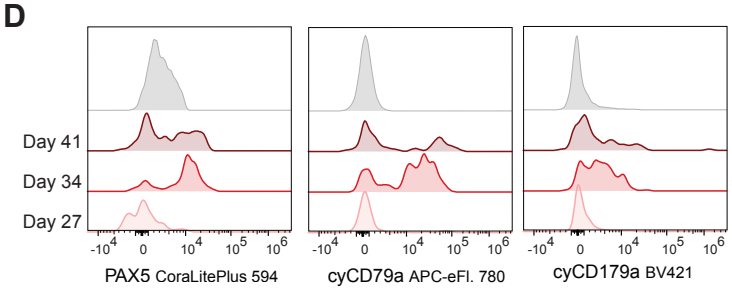

A
Hemogenic endothelium
B-cell differentiation
Human induced pluripotent stem cells (hiPSCs)
CD34+ cells
MS-5 stromal cells
hiPSC-B-cell progenitors
hiPSC immature B cells
Vilarrasa-Blasi MS5 protocol
Top-up
Media change
scRNA timepoint
BMP4
VEGF, IL-6, IL-11
Y-S7632
bFGF
SB, CHIR
EPO, IGF1, SCF
IL-3 + SCF + Flt3L + IL-7
SCF + Flt3L + IL-7
0 1 1.75 3 4 6 8 12 13 15 16 20 22 26 29 days
B
Canonical-HSC
Early-HSC
Definitive-HSC
Patterning
Vilarrasa-Blasi MS5 protocol
End. HSC-like D8
Prog. HSC-like D13/D15
Fetal BM
Adult BM
CD34 MLLT3 SPINK2 HOPX HLF RAB27B MYB DDIT4 SLC2A3 RGS16 GAD1 CD9 KIT ITGA4 CD74 PROCR EMCN GBP4 ACE HOXA7 HOXA9 HOXA10 HOXB7 HOXB9
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
C
HSC
LMPP
PreProB/PreProB-like
Plasma
ELP/CLP
EarlyProB
ImmatureB
LateProB
MemoryB
LargePreB
PreNaiveB
SmallPreB
B1
UMAP2
UMAP1
n= 291,332 cells
hiPSC-differentiated
Cultured Fetal BM
Fetal BM + Adult BM
n= 291,332 cells
D
Fetal BM + Adult BM
hiPSC-differentiated
Early progenitors
ELP/CLP
ProB
PreB-ImmatureB
CD34 SPINK2 SPN PRSS57 HOPX JCHAIN CD7 FLT3 LTB IL7R IL2RG CD79A VPREB1 CD19 CD24 MME EBF1 PAX5 CD38 SPIB CD40 MS4A1 PTPRC
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
Timepoint
D8
D13
D15
D22
D29
0 20 40 60 80 100
E
CD34
PTPRC (CD45)
CD19
MME (CD10)
EBF1
PAX5
Gene epression (mean ± 95% CI)
Early Progenitors
ELP/CLP
ProB
PreB-ImmatureB
Fetal BM + Adult BM
hiPSC-differentiated
F
Ligand
Receptor
hiPSC-differentiated
Early progenitors
ELP/CLP
ProB
PreB-ImmatureB
Stroma signal
MS-5
Kitl→KITLG
Il7→IL7
Tslp→TSLP
Vcam1→VCAM1
Cxcl12→CXCL12
Fn1→FN1
Csf1→CSF1
Tgfb2→TGFB2
Tgfb3→TGFB3
KIT
IL7R
CRLF2
ITGA4
CXCR4
ITGB7
CSFR1
TGFBR1
TGFBR2
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0

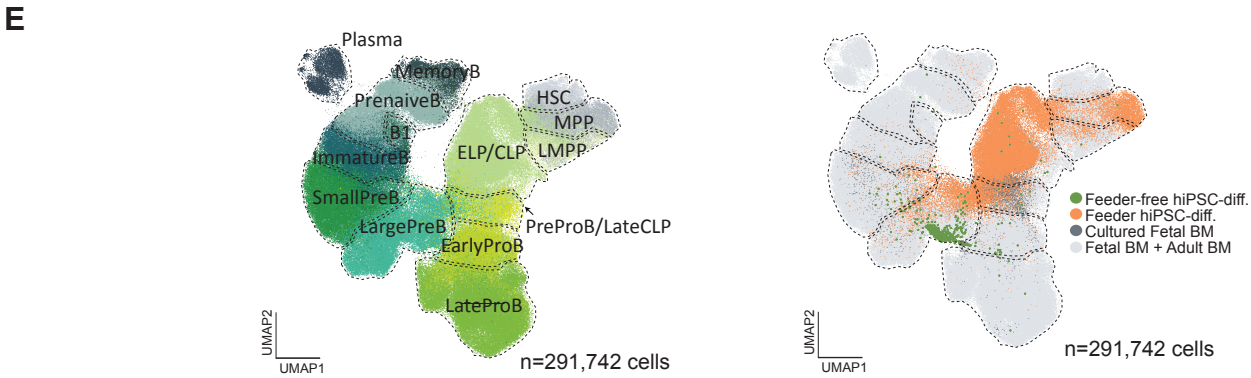
Figure 5

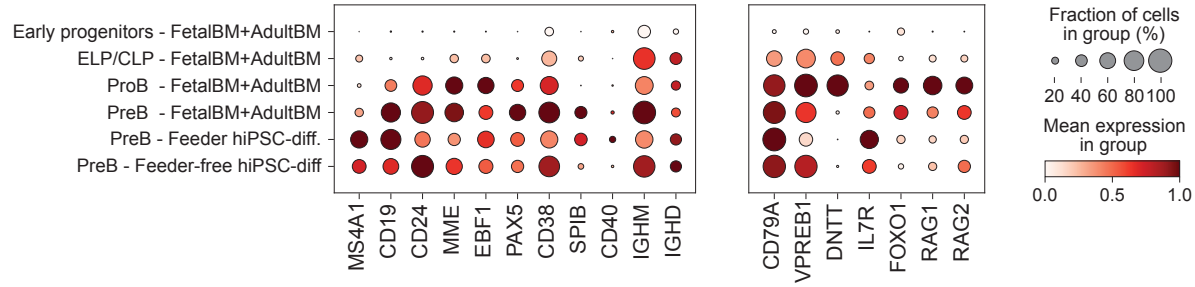

A
Hemogenic endothelium
Endothelial-to-hematopoietic transition
B-cell differentiation
Vilarrasa-Blasi feeder-free protocol
Top-up
Media refresh
Media change
scRNA timepoint
hiPSCs
CD34+ cells
CD34+/CD43- cells
CD19 ± CD10+ cells
BMP4
Y-S7632
bFGF
VEGF, IL-6, IL-11
SB, CHIR
EPO, IGF1, SCF
bFGF, VEGF, IL-6, IL-11, TPO, IGF1, SCF, IL-3, BMP4, Flt3L, Y-27632
Matrigel
SCF + Flt3L + IL-7 + CXCL12a
SCF
0 1 1.75 3 4 6 8 11 13 16 20 23 27 30 34 37 41 days

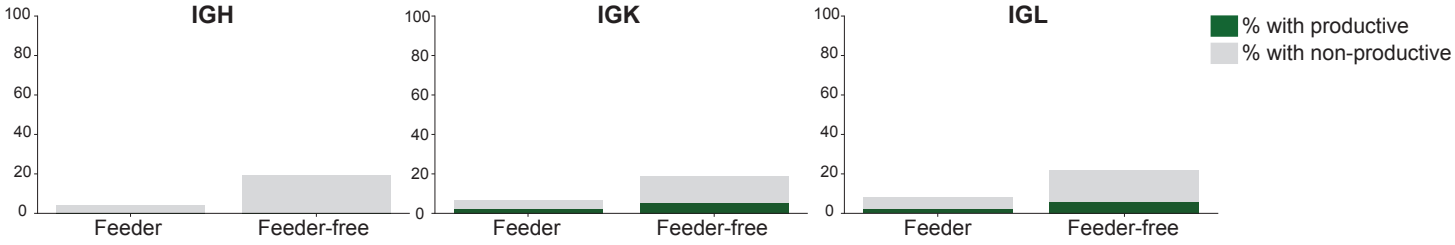

B
Day 34
3.42%
CD19 BUV563
CD10 BUV737


C
Frequency of CD10+/CD19+
Percentage of CD33-
Day of culture
27 34 41


D
Day 41
Day 34
Day 27
PAX5 CoraLitePlus 594
cyCD79a APC-eFl. 780
cyCD179a BV421


E
Plasma
MemoryB
PrenaiveB
B1
ImmatureB
SmallPreB
LargePreB
EarlyProB
LateProB
HSC
MPP
LMPP
ELP/CLP
PreProB/LateCLP
UMAP2
UMAP1
n=291,742 cells
Feeder-free hiPSC-diff.
Feeder hiPSC-diff.
Cultured Fetal BM
Fetal BM + Adult BM
n=291,742 cells


F
Early progenitors - FetalBM+AdultBM
ELP/CLP - FetalBM+AdultBM
ProB - FetalBM+AdultBM
PreB - FetalBM+AdultBM
PreB - Feeder hiPSC-diff.
PreB - Feeder-free hiPSC-diff
MS4A1 CD19 CD24 MME EBF1 PAX5 CD38 SPIB CD40 IGHM IGHD
CD79A VPREB1 DNTT IL7R FOXO1 RAG1 RAG2
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0


G
IGH
IGK
IGL
% with productive
% with non-productive
Feeder
Feeder-free