Single-cell analysis identifies dynamic gene expression networks that govern B cell development and transformation

Robin D. Lee1, Sarah A. Munro2, Todd P. Knutson2, Rebecca S. LaRue2, Lynn M. Heltemes-Harris1 & Michael A. Farrar1✉

Integration of external signals and B-lymphoid transcription factor activities organise B cell lineage commitment through alternating cycles of proliferation and differentiation, producing a diverse repertoire of mature B cells. We use single-cell transcriptomics/proteomics to identify differentially expressed gene networks across B cell development and correlate these networks with subtypes of B cell leukemia. Here we show unique transcriptional signatures that refine the pre-B cell expansion stages into pre-BCR-dependent and pre-BCR-independent proliferative phases. These changes correlate with reciprocal changes in expression of the transcription factor EBF1 and the RNA binding protein YBX3, that are defining features of the pre-BCR-dependent stage. Using pseudotime analysis, we further characterize the expression kinetics of different biological modalities across B cell development, including transcription factors, cytokines, chemokines, and their associated receptors. Our findings demonstrate the underlying heterogeneity of developing B cells and characterise developmental nodes linked to B cell transformation.

1Department of Laboratory Medicine and Pathology, Center for Immunology, Masonic Cancer Center, University of Minnesota, Minneapolis, United States.
2Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, United States. ✉email: farra005@umn.edu
Distinct stages of B-cell development have been delineated using flow cytometry and a variety of surface and intracellular markers. The use of such markers in conjunction with distinct gene knockout mice has greatly expanded our understanding of specific B-lymphoid transcription factors, cytokines, and signaling pathways that entrain B cell development. However, these markers are insufficient to fully demarcate distinct subsets, resulting in the analysis of mixed populations. These limitations have led to an incomplete understanding of B-lymphoid transcription-factor expression kinetics across the B-cell developmental trajectory, and the orchestration of transcriptional programs underlying the alternating cycles of proliferation and differentiation. Capturing transition states as B cells differentiate from one stage to the next is particularly difficult. Furthermore, perturbations during normal B-cell differentiation can lead to development of B-cell acute lymphoblastic leukemia (B-ALL)\textsuperscript{11–13}. However, exactly what stages are most permissive for transformation remains imprecisely defined. Recent characterization of B-ALL subtypes showed diverse transcriptional signatures, suggesting the existence of more than one point of origin or use of different signaling pathways to drive transformation\textsuperscript{12,14}. Therefore, understanding normal B-cell transcriptional programs becomes important.

To address the above questions, we used single-cell transcriptomics (scRNA-Seq) and proteomics (CITE-Seq, Cellular Indexing of Transcriptomes and Epitopes by Sequencing)\textsuperscript{15} to precisely characterize different subsets of B-cell development. Our analysis discovered several stages of pre-B-cell differentiation—including a pre-BCR-dependent and two pre-BCR-independent stages that exhibited distinct modalities of proliferation. This process of pre-B-cell differentiation was characterized by oscillatory regulation of the transcription factor EBF1, with reciprocal changes in the RNA-binding protein YBX3. In contrast, the pre-BCR-independent stages correlated with changes in chemokine and cytokine receptors and suggest that these stages may involve differential localization of pre-BCR-independent stage subsets within the bone marrow. Finally, comparisons of various human B-ALL transcriptomes to those of different stages of B-cell development highlight the developmental nodes that associate with varying subtypes of B-ALL and how they correlate with prognosis.

### Results

**Identifying B-cell development stages using scRNAseq and CITE-Seq.** To couple transcriptional information with B-cell-stage-defining surface marker expressions, we used combined single-cell RNA sequencing and CITE-Seq (also referred as antibody-derived-tags [ADT] hereafter) proteomics. Bone marrow from two wildtype C57BL/6 mice was harvested and stained with two distinct oligo-labeled antibodies that recognize CD45 and MHC class I, which allow identification of cells derived from each individual mouse (referred to as hashtag antibodies). We further stained cells with a panel of CITE-Seq antibodies (B220, CD19, CD93, CD25, IgM, and CD34) as well as fluorescently labeled B220 and CD43. Cells were sorted (representative gating; each individual mouse (referred to as hashtag antibodies). We and MHC class I, which allow identification of cells derived from each individual mouse (referred to as hashtag antibodies). We and MHC class I, which allow identification of cells derived from each individual mouse (referred to as hashtag antibodies). We and MHC class I, which allow identification of cells derived from each individual mouse (referred to as hashtag antibodies). We

**Transcriptional signatures of pre–pro B cells and committed pro B cells.** We first used antibody-derived CITE-Seq tags to define B220\textsuperscript{+}CD43\textsuperscript{+}CD19\textsuperscript{−} pre–pro B cells. These cells expressed early B-lineage-associated genes such as Fli3, Il7r, and Cd79a (Fig. 2a). Pre–pro B cells have also been shown to express genes associated with myeloid lineages, consistent with the observation that they can give rise to myeloid cells as well\textsuperscript{18,19}. Indeed, we found that the pre–pro B cells have high expression of myeloid lineage-associated transcription factors such as Runx2, Ifi6, and Tcf4, and plasmacytoid dendritic cell markers, such as Bst2; these genes were silenced upon commitment to the B-cell lineage at the pro-B-cell stage (Fig. 2b). We also assessed the expression of previously described EBF1-repressed target genes, including Tyrobp, Clec12a, Cd300a, Cd7, Odh2, and Myc\textsuperscript{20} (Fig. 2c). These target genes were highly expressed in the pre–pro B-cell cluster, while Ebf1-expressing pro-B cells had low or undetectable expression of these genes, further distinguishing the pre–pro B cells from pro-B cells (Fig. 2b). To identify the presence of B-cell-biased uncommitted progenitors in this cluster, we used CD93 ADT expression, previously shown to enrich for B cell progenitors\textsuperscript{21}. A median split of CD93 expression was performed in this cluster, and the CD93 above-median expressing cells had enrichment of B-cell commitment associated genes, such as Ebf1, Cd24a, and Vpreb1 (Fig. 2d). Pro-B cells, traditionally delineated as c-KIT\textsuperscript{+} cells\textsuperscript{22} had Kit gene expression (Fig. 2e). In addition, pro-B cells expressed the Bcl-2 family gene, Bok, and interferon-stimulated genes, such as Ifitm2 and Ifitm3 (Fig. 2e, Supplementary Data 1). Finally, EBF1 positively regulates the expression of pre-BCR-surrogate light-chain genes, Vpreb1 and IgGl\textsuperscript{23}. We found that cycling pro-B cells had high expression of both Vpreb1 and IgGl, compared with the pre-pro B cells that do not express Ebf1 (Fig. 2f). Thus, our analysis confirmed several genes expression patterns associated with the transition from pre–pro-B to pro-B cells, as well as identified highly specific markers of pro-B cells such as Bok.

**Pre-B-cell expansion comprises distinct pre-BCR-dependent and pre-BCR-independent proliferation stages.** Next, we assessed the transcriptional signature of the ADT-B220\textsuperscript{+}CD19\textsuperscript{+}CD43\textsuperscript{+}CD25\textsuperscript{−} cells. We refer to these cells as the pre-BCR-dependent proliferation cluster (also referred to pre-BCRd hereafter) (Fig. 3a). Pre-BCR signaling initiates silencing of the surrogate
light-chain locus\(^{24}\). In accordance with this, the pre-BCR-dependent cells have intermediate expression of both \(\text{Vpreb1}\) and \(\text{Igll1}\) (Fig. 2f). Genes that were uniquely upregulated in this cluster included \(\text{Nrgn}\) and \(\text{Ybx3}\) (Fig. 3a, Supplementary Figure 3a and Supplementary Data 2). \(\text{NRGN}\) (Neurogranin) is a calmodulin-binding protein that regulates the dynamics of calcium binding to calmodulin\(^{25}\). Neurogranin is upregulated in activated\(^{26}\) and anergic B cells\(^{37}\), which suggests that neurogranin expression is controlled by B-cell receptor signaling. \(\text{YBX3}\) is a DNA/RNA-binding protein that was recently shown to stabilize the amino acid transporter transcripts \(\text{Slc7a5}\) and \(\text{Slc3a2}\) in HELA cells, allowing for their robust translation\(^{28}\). \(\text{Slc7a5}\) and \(\text{Slc3a2}\) heterodimerize to form CD98, a large neutral amino acid transporter. Expression of CD98 in CD\(^8\) T cells has been shown to be tightly controlled by antigen-receptor signaling and is critical for MYC expression\(^{29}\). Our findings suggest that pre-BCR signaling may serve a similar function. Consistent with this idea, MYC was most highly expressed in the \(\text{Ybx3}\), \(\text{Slc7a5}\), and \(\text{Slc3a2}\) expressing pre-BCRd cluster (Fig. 3a). Surprisingly \(\text{Ebf1}\) expression is significantly reduced, while expression of the transcription factor \(\text{Pax5}\) was largely unchanged and \(\text{Ikef1}\) was modestly induced (Fig. 3b). Concordant with low \(\text{Ebf1}\) expression, EBF1-target genes, such as \(\text{Cd79a}\)\(^{30,31}\) and \(\text{Cd79b}\)\(^{32}\), are also significantly reduced in the pre-BCRd stage (Fig. 3c). In addition, \(\text{Il7r}\) expression, a negative regulator of pre-BCR-signaling components\(^{33}\), is also reduced (Fig. 3c). To identify potential transcriptional regulators that govern this pre-BCR-dependent expansion stage, we performed a Landscape In Silico deletion Analysis (LISA)\(^{34}\) using the top 100 differentially upregulated genes in the pre-BCRd cluster. After excluding factors not expressed in B cells (i.e., MYCN), we found MYC to be the top predicted regulator of this gene set in pre-BCRd cells. Consistent with its decreased expression, EBF1 was predicted to have minimal
transcriptional landscape during pre-BCR-dependent expansion. To further understand the importance of 

**contributions in pre-BCRd cells (Fig. 3d).** In addition, we performed a gene-set enrichment analysis (GSEA), which further supported a MYC expression signature and enrichment of metabolic reprogramming signatures (Fig. 3d). Together, this suggests that MYC is the critical transcription factor that governs the transcriptional landscape during pre-BCR-dependent expansion. To further understand the importance of Ebf1 downregulation during pre-BCR signaling, we examined differentially regulated genes that contained Ebf1 binding sites. Ebf1 has clear binding sites at the promoters for Nrgn and Myc (Fig. 3e). Likewise, Ebf1 binds to a
known superenhancer linked to the *Myc* locus. In addition, potential binding sites were observed within the promoters of the *Slc7a5* and *Slc3a2* genes (Fig. 3e). We performed RNA-seq in wildtype and *Pax5*<sup>−/−</sup> × *Ebf1*<sup>−/−</sup> leukemic progenitor B cells and evaluated expression changes of *Ybx3*, *Slc7a5*, and *Slc3a2*. We found that *Slc7a5* and *Slc3a2*, along with other pre-BCR-dependent stage module genes (*Ybx3*), were upregulated with decreased *Ebf1* gene dosage (Fig. 3f). This is consistent with our recent observation that *Myc* expression is increased in *Pax5*<sup>−/−</sup> × *Ebf1*<sup>−/−</sup> preleukemic and *Pax5*<sup>−/−</sup> × *Ebf1*<sup>−/−</sup> leukemic progenitor B cells. These findings suggest that EBF1 mediates repression of the pre-BCR gene expression module.
YBX3 binds to Jak1 transcripts and inhibits translation of Jak1 in HELA cells. To evaluate whether YBX3 levels could reduce JAK1 in pre-B cells, and to validate EBFI downregulation during pre-BCR signaling, we used flow cytometry to characterize B220+CD19+CD43+CD99high expressing cells. Ki67+JAK1low cells have lower EBFI expression (which correlates with high Ybx3) compared with the Ki67+JAK1high cells (Supplementary Fig. 4a). When evaluating the expression of Y-box family genes, Ybx1 was ubiquitously expressed throughout all developmental stages, with peak expression during the pre-BCRd stage, whereas Ybx3 expression was selectively induced during the pre-BCRd stage (Supplementary Fig. 3a). We assessed how YBX3 governs B-cell development by using Ybx3+/− mice with flow cytometry. Ybx3+/− B-lineage cells had no significant phenotypic defects during the early proliferative phases (Hardy fractions A–C, Supplementary Fig. 4b, d) and late stage differentiation of B cell development (Hardy Fractions D–F, Supplementary Fig. 4c, d). This is consistent with previous findings that YBX1 and YBX3 have redundancies in both function and target mRNA binding. Collectively, our data highlight the reciprocal regulation of Myc and Ebf1 during B-cell development, where MYC is critical for governing differentially expressed genes in the pre-BCRd cluster. Furthermore, pre-BCR signaling limits the IL7R-signaling axis by downregulating IL7R expression and JAK1 protein translation.

We next examined the identity and signature of the remaining cycling cells (Fig. 4a). These cells have minimal surrogate light-chain expression (Fig. 2i), indicating further silencing of the locus. Notably, these cells have intermediate expression of the ADT-CD43 (Fig. 4a), which suggests that they are transitioning toward the quiescent CD43lo/m− small pre-B-cell stage. Some cells within these clusters also express CD25 protein (Fig. 4a), a marker for pre-B cells. A subset of these cells still expressed high levels of Nrgn (Fig. 3a), which may suggest continued pre-BCR signaling. However, unlike the pre-BCR-dependent cluster, these cells have high expression of Bach2 (Fig. 4b), a transcription factor that represses antigen-receptor signaling and low expression of Ybx3 and Myc (Figs. 3a and 4c). This suggests that these pre-B cells are proliferating independently of pre-BCR signaling. When comparing the differentially expressed genes between the pre-BCR-dependent and pre-BCR-independent cells, the pre-BCR-independent cells from cluster I (Pre-BCRi I, Fig. 4a), we found that pre-BCR-independent cells reexpress high levels of Il7r and Ebf1 (Figs. 3b, c, and d), in accordance with restrained pre-BCR signaling. When comparing the two different pre-BCR-independent clusters (pre-BCRi I and pre-BCRi II), pre-BCRi I had heightened expression of histone genes (Hist1h2ac, Hist1h2ap, Hist1h1b, Hist1h1c, and Hist1h4d) (Fig. 4d), but no difference in Mki67 expression (Supplementary Fig. 3b). On the contrary, the pre-BCRi II cluster is enriched for cells expressing cell motility-associated actin (Actg1), dynein (Dyn1l1), and thymosin (Tmsb10 and Tmsb12) genes (Fig. 4e). Interestingly, expression of Cxcr4 is further induced as pre-B cells reach the pre-BCRi stages (Fig. 4b). Expression of CXCR4 would promote targeting of pre-BCRi cells to CXCL12-expressing cells, which also often express high levels of IL7 in the bone marrow. Since Il7r also increases in pre-BCRi cells, these changes in gene expression suggest that pre-BCRd cells rely on pre-BCR signals for survival/proliferation, while pre-BCRi cells rely on IL7 signals. Although the IL7R and CXCR4 likely promote pre-BCRi cell survival, we did observe that pre-BCRi cells also started to express Cd74 as well as Cd44 (Fig. 4e), which complex together to form a receptor for the chemokine ligand, MIF. This could promote chemotaxis of pre-BCRi cells to Mif-expressing Fbn1highIgf1high osteogenic cells, to also express IGF1, a paracrine growth factor that has been previously shown to be important for the generation of small pre-B cells. Thus, pre-BCRi cells likely require CXCR4/IL7R signals, with a possible contribution from the MIF/CD74/CD44 and IGFI/IGF1R signaling axes, to efficiently transit from a pre-BCR-dependent to a pre-BCR-independent state.

To better characterize the pre-BCR-dependent and pre-BCR-independent populations, we used flow cytometry using the markers identified from our single-cell study. We observed that surface CD74 expression was the highest in the early CD43+ progenitor B cells and had a stepwise decrease in surface expression as the cells matured into mature B cells (Supplementary Fig. 5a). In contrast, intracellular CD74 expression was low in the CD43+ progenitor B cells and small pre-B cells but was the highest in the immature and mature B cells (Supplementary Figure 5a), consistent with transcript expression (Fig. 4b). These results suggest spatiotemporal regulation of CD74 during B-cell development and may point to a differential function of CD74 -serving as a chemokine coreceptor early in development and in antigen processing, as the invariant chain, in more mature stages of development. Finally, to characterize the dynamic Ebf1 gene expression between the pre-BCR-dependent and pre-BCR-independent populations, we used Myc-GFP reporter mice to query for Ebf1 protein expression. MYC-GFP-expressing progenitor B cells (B220+CD19+CD43+IgM−/+) had a heterogeneous EBF1 protein expression (Supplementary Fig. 5b). Compared with the MYC-GFP+EBFIhigh cells, the MYC-GFP+EBFIlow expressing cells also had lower IL7R, CXCR4, and CD74 expression (Supplementary Fig. 5c), which is consistent with transcript expression in the pre-BCR-dependent cluster. To ensure that MYC-GFP-expressing pro-B cells were not the predominant population in the MYC-GFP+EBFIlow population, we examined cKit expression and observed that MYC-GFP+EBFIlow cells were primarily cKit-negative (Supplementary Fig. 5d).
Fig. 4 Pre-BCR-independent proliferation is distinct from pre-BCR-dependent proliferation. a Highlighted pre-BCRI I and pre-BCRI II populations (left). Violin plots of expression of ADT-CD43 and ADT-CD25 across the B-cell development stages (right). b Feature plot and violin plot for Cd74, Cd44, and Bach2. c Heatmap of differentially expressed genes between the pre-BCR-dependent (Pre-BCRd), pre-BCR-independent I (Pre-BCRI I), and pre-BCR-independent II (Pre-BCRI II) S, or G2/M phase. Scale represents normalized counts centered and scaled across cells. d Volcano plot showing differentially regulated genes between the pre-BCRd cluster and the pre-BCRI I cluster. e Differentially regulated genes between the pre-BCRI I cluster and the pre-BCRI II cluster. Color scales in a, b represent natural log-transformed SCTransform-corrected counts.
**B-cell differentiation and maturation.** Quiescent IgM− cells have high expression of Rag1 and Rag2, suggesting V(D)J recombination of pro-B and pre-B cells (Fig. 5a). While Igkc expression is detectable in all subsets, except for cycling pro-B cells and pre-BCR-dependent B cells, Igkc1, Igkc2, and Igkc3 are only expressed in a subset of cells (Fig. 5a). Furthermore, different B-cell stages exhibited differential expression of trace-element-associated genes such as selenoprotein genes. Selenom was expressed in a subset of pro-B VDJ cells (Supplementary Fig. 6a). Selenop was expressed in both pro-B and pre-B cells undergoing recombination, whereas Seleno was highly expressed in cycling pro-B and pre-B cells (Supplementary Fig. 6a). The significance of this differential gene expression program remains to be ascertained, although selenium has been associated with immune function and activation44. Finally, the IgM+ cells were broken down into three clusters corresponding to immature B cells, cycling immature B cells, and mature B cells. The cycling immature B cells have high expression of surface IgM. (Supplementary Fig. 6b). Immature B cells express Ms4a1 (CD20), whereas the mature B cells express Ms4a4c, H2-Aa, Sell (L-selectin), and Ltb (Fig. 5b). Notably, a subset of mature B cells expressed Apoe (Fig. 5b) and showed overlapping detection of both IgM and CD43 (Fig. 1d). This subset shares features with previously described B1 bone marrow B cells15, although the function of Apoe in these cells remains unknown.

**Pseudotime and module analysis of the B-cell development trajectory.** To understand the relationship between developmental stages and changes in gene expression over the B-cell developmental trajectory, we performed pseudotime analysis using Monocle16. To lessen the influence of cell cycle on UMAP positioning, we regressed out cell cycle genes within Monocle and performed UMAP dimension reduction. Using stage-defining markers, we identified 13 transcriptionally distinct B-cell developmental stages (Fig. 6a and Supplementary Fig. 7). Compared with the Seurat-based clustering (Fig. 1b), the immature B cell subset was split into two clusters and the preBCRii S and G2/M
Fig. 6 Pseudotime analysis illustrates the kinetics of transcription-factor expression and gene modules that are differentially expressed across the B-cell development trajectory. 

a Cell cycle-related genes were regressed within Monocle3 and UMAP dimension reduction was performed. Clustering was performed based on the Monocle3 clusters (left). Cells from Monocle-based clustering labeled with Seurat cluster labels (right). 

b Pseudotime values were calculated and plotted. 

c Module analysis to demonstrate gene modules that change across the B-cell developmental trajectory. A total of 33 modules and their expression intensity for each stage are shown. Color scale represents normalized module gene expression. 

d Gene ontology term analysis of selected modules 1, 6, and 8. 

e Expression of B-lymphoid transcription factors and epigenetic factors across B-cell development stages. 

f Expression of cytokine, chemokine and cytokine/chemokine receptors across the B-cell developmental trajectory.
subsets from Seurat were combined into one cluster in Monocle (Fig. 6a). Despite the differing clustering methods and regression of cell cycle genes, the Seurat and Monocle labeling of B-cell subsets demonstrated a high degree of overlap (Fig. 6a and Supplementary Data 3). Pseudotime values were then calculated to establish a developmental trajectory (Fig. 6b). This analysis let us identify modules of genes that are changing across the developmental trajectory (Fig. 6c). Gene ontology analysis of these gene modules suggested that the early B-cell stages, including pre–pro B and pro-B cells, are significantly enriched for cell-adhesion processes, whereas these signals are diminished in pre-B cells (Module 1 and Module 8, Fig. 6c, d). This is in accordance with previous findings that pro-B cells strongly promote adhesion to IL7–producing stromal cells, whereas pre-B-CR-signaling and its downstream target IKZF1, are important for downregulating stromal adhesion components. Furthermore, the cycling pro-B cells and pre-BCR-dependent cells were enriched for genes involved in metabolic processes (Module 6; Fig. 6d), suggesting metabolic reprogramming after major checkpoints (B-cell lineage commitment and pre-B-BCR selection, respectively) during B-cell development. Finally, evaluation of the expression of transcription factors, epigenetic modifiers (Fig. 6c), cytokines, chemokines and their respective receptors (Fig. 6f) indicates differential expression kinetics across the B-cell development trajectory. Thus, our findings confirm some previous observations, but also identify distinct gene programs that exhibit highly dynamic changes during B-cell development.

Stage-specific gene expression networks correlate with various human B-ALL subtypes and prognosis. Defects in B-cell differentiation and dysregulation of signaling pathways lead to B-cell transformation. To identify B cell developmental pathways that are associated with human B-ALL subtypes, we used differentially upregulated gene markers identified in both the Seurat and Monocle analyses. Using averaged z-scores of upregulated genes from each cluster, the transcriptional network of each stage was compared with various human B-ALL subtypes. This resulted in the hierarchical clustering of proliferation and differentiation stages (Fig. 7a).

Interestingly, the human B-ALL subtypes that were enriched for proliferating pro-B, pre-B-BCRd, and preBCRi clusters, included the BCL2/MYC, IKZF1 N159Y, and KMT2A subtypes, which are all associated with high risk poor prognosis. In contrast, the B-ALL subtypes that were enriched for genes characteristic of other B cell differentiation stages, such as ETV6–RUNX1 and ZNF384 rearrangements, were the ones with favorable outcomes. Performing hierarchical clustering using only the Monocle gene list gave identical clustering patterns. We then performed statistical testing for significance of the hierarchical clustering. Testing of the row-based hierarchical clustering of different developing B-cell stages showed statistically significant differences in the proliferating B cells versus the differentiation stages (Fig. 7b, left, P = 0.006, Monte Carlo test corrected for familywise error rate). Notably, the preBCRd stage was also distinct from the preBCRi and proliferating pro-B stage (Fig. 7b, left, P = 0.006, Monte Carlo test corrected for family-wise error rate). Likewise, column-based hierarchical clustering testing of the human B-ALL subtypes demonstrated that the high-risk leukemias that correlated with proliferative B cells were significantly distinct from the low-risk leukemias that correlated with differentiating B cells (Fig. 7b; right, P = 0.001; Monte Carlos test corrected for family-wise error rate). The pre-BCR module genes, including YBX3 and NRGN, are significantly upregulated in BCL2/MYC, IKZF1 N159Y, and MEF2D B-ALL subtypes (Fig. 7c). Despite the numerous pre-BCR module genes highly expressed in various B-ALL subtypes, not all pre-BCR genes share this pattern, indicating transcriptional heterogeneity of B-ALLs compared with normal-developing B cells (Fig. 7c, Supplementary Fig. 8a). Finally, we examined whether YBX3 expression correlated with outcome in B-ALL. We observed that pediatric B-ALLs with above-median YBX3 expression are associated with worse prognosis (Fig. 7d, hazard ratio = 2.03, P = 0.032, log-rank test), whereas adult B-ALL patients (that mainly comprise Ph+ B-ALLs) show no significant difference in survival (Fig. 7d). Overall, we identify B-cell gene expression networks that are modulated during B-cell development (Fig. 7e) and correlate with human B-ALL subtypes. In addition, we specifically demonstrate that a YBX3-related module is associated with a poor prognosis in human B-ALLs.

Discussion

The evaluation of different organs and niches at single-cell resolution has greatly expanded our understanding of the cellular diversity that is present within the bone marrow. However, these previous studies have not resulted in a detailed description of B-cell development. This is due to the paucity of developing B cells in these broad surveys of the bone marrow compartment. B-cell development has also been examined using cell surface markers in conjunction with flow cytometry. These studies, in conjunction with in vivo reconstitution experiments to ascertain precursor–progeny relationships, have provided a general outline of B-cell development. Furthermore, resources such as Immgen provide a wealth of transcriptional data about various sorted B-cell compartments. However, these sorted-cell populations are relatively heterogeneous and thus fail to provide detailed single-cell resolution of B-cell development. In this study, we coupled the conventional surface marker-based staging of B-cell development using CITE-Seq with single-cell transcriptomics to identify unappreciated transcriptional heterogeneity during B cell development and link them to various underlying biological processes. In addition, we identify the RNA-binding protein YBX3 as a marker of pre-B cell differentiation and correlate Ybx3 expression with outcomes in patients with B cell acute lymphoblastic leukemia.

Proliferation and differentiation during B-cell development are important biological modalities that underpin selection and BCR-repertoire diversification. Given our sorting strategy (Supplementary Fig. 1a), we did not capture all possible subsets of B cells, such as B220–CD19+ B1 progenitor cells or CD11c–expressing B cells. Nonetheless, our results highlight the diversity of transcriptional networks that are present within both proliferation and differentiation modalities during B-cell development. Specifically, the proliferating B cells were clustered tightly and labeled as G1PM, S, and G2/M–phase cells (Fig. 1c) based on their transcript status using a predetermined cell cycle gene list. The postmitotic G1 phase (G1PM) was recently described in a study demonstrating that stimulated splenic B cells undergo mitogen-independent proliferation in which genes associated with G2/M phase are not fully extinguished during the G1 phase. Thus, proliferating B cells can undergo an extremely short G1 stage that bears features of G2/M. Our study highlights that proliferating progenitor B cells also exhibit a G1 phase containing many G2/M–signature genes (Fig. 1c and Supplementary Fig. 2a), which is consistent with rapidly proliferating B cells with very short G1 stages. Our studies demonstrate unexpected dynamic changes in transcription factors that orchestrate B-cell development. We were largely able to confirm the expression kinetics for the critical B-lymphoid lineage transcription factors Pax5 and Ikzf1 across B cell development stages (Fig. 3b). In contrast, Ebf1, another key
B-cell transcription factor, has a strikingly large dynamic range of expression (Fig. 3b) that varies greatly over the course of B-cell development. *Ebf1* has the highest expression during both heavy and light chain recombination stages, intermediate expression during cytokine-mediated proliferative stages, and the lowest expression during pre-BCR and BCR-signaling stages (Fig. 3b). Likewise, the expression of EBF1-target genes exhibited the same patterns (Fig. 3c). This points to dynamic changes in *Ebf1* in regulating key gene expression networks throughout B-cell development. Furthermore, B-lymphoid transcription factors, including *Ebf1*, have been shown to serve as metabolic gatekeepers, where they repress genes encoding proteins for glucose uptake and utilization, and thereby prevent malignant transformation61. In light of this, we provide evidence that EBF1 represses *Myc* and that low gene expression of *Ebf1* correlates strongly with the activation of pre-BCR-dependent gene modules.
that promote metabolic reprogramming (such as amino acid transporters) and proliferation. Therefore, the expression kinetics of \( Ebf1 \) are tightly controlled to enable the unique aspects of alternating cycles of proliferation and differentiation throughout B-cell development.

Our single-cell RNA-Seq data pointed to an unappreciated expression of the RNA binding protein YBX3 in pre-B-cell differentiation and proliferation. Notably, Ybx3 expression was significantly upregulated during the pre-BCR-dependent stage (Fig. 3a). Likewise, the related YBX family member Ybx1 is also most highly expressed at the pre-BCRd stage (Fig. 3a). Analysis of Ybx3 expression in Ybx1−/− mice did not show major developmental defects (Supplementary Fig. 4b, c) and thus YBX1 likely serves a redundant function with YBX3 in developing B cells. Therefore, YBX1 and YBX3 may initiate the MYC-dependent transcriptional program that characterizes the pre-BCR-dependent expansion stage.

Elucidating the requirements and regulation of normal B-cell development has significantly improved our understanding of how developing B cells can undergo transformation and maintain leukemic states. B-ALL subtypes exhibit significant transcriptional diversity. While differences between mouse and human B cell populations exist, many aspects of mouse and human B-cell development are highly analogous. Our comparative analysis between stage-defining gene expression in wild-type progenitor B cells and different human B-ALL subtypes suggests that cycling pro-B, pre-BCR-dependent, and pre-BCR-independent subsets shared the highest similarities with high-risk human B-ALL subsets, including BCL2/MYC, IKZF1 N159Y, and KMT2A-rearranged leukemias. In contrast, the absence of these cycling-stage signatures was associated with low-risk human B-ALL subsets, including ETV6-RUNX1 and ZNF3840-rearranged leukemias (Fig. 7a). Importantly, while these cycling stage signatures do include some canonical cell cycle genes, they are also characterized by non-cell cycle related unique gene sets associated with these specific developmental stages, which may promote a more aggressive disease state (Supplementary Fig. 8). Therefore, our findings suggest that the presence or absence of stage-specific genes may associate with prognosis and could be targeted to further improve outcome.

When looking at individual genes from each cluster in different B-ALL subtypes, we found both overlapping and nonoverlapping gene modules, suggesting that B-cell leukemias partially associate with specific expression modules of multiple B-cell development stages. In particular, we found that pre-BCR module genes, including \( Nrgn \) and Ybx3, were significantly upregulated in the high-risk BCL2/MYC, IKZF1 N159Y, and MEF2D B-ALL subtypes (Fig. 7b). In addition to our findings on the importance of MYC during pre-BCR signaling, previous studies have demonstrated the upregulation of IKZF1 and MEF2D upon pre-BCR signaling. Therefore, aberrant dysregulation through mutations and rearrangements in IKZF1 and MEF2D, respectively, may lead to the heightened usage of pre-BCR-signaling genes for transformation. Importantly, we were able to demonstrate that high expression of Ybx3, a key component of the pre-BCR-signaling module, correlates with poor prognosis in pediatric B-ALL (Fig. 7c). Interestingly, high expression of Ybx3 also correlates with dramatic downregulation of \( Ebf1 \). The combination of elevated YBX3 and reduced \( Ebf1 \) expression likely maintains the MYC-driven transcriptional program that contributes to pre-B-cell transformation. As YBX3 has been shown to repress \( JAK1 \) translation, this may explain the requirement for ectopic activation of the JAK/STAT5 pathway to overcome a potential YBX3-driven negative feedback loop. Collectively, our data and previously published findings converge on multiple mechanisms that can activate different proliferation modules, leading to high-risk leukemia and a poor prognosis. Using strategies that target YBX3 or pre-BCR-related modules in pre-BCR-related leukemias may also further improve outcomes in patients with these subsets of high-risk B-ALL.

Methods

Animals. All animals used were bred and housed in specific pathogen-free facilities at the University of Minnesota and animal experiment protocols were approved by Institutional Animal Care and Use Committees (IACUC 2010-38515A and IACUC 1904-36975A). All of the animals used were 6-12-week-old C57BL/6J males and females with appropriate age- and sex-matched controls. Specifically, the two wild-type mice used for the single-cell RNA-seq experiment were C57BL/6J males and females, and C57/BL6-12-week-old. The \( JAK1 \) and \( EBF1 \) flow experiment was performed with C57BL/6J male and female mice, CD45.1 and CD45.2, that were 7-10 weeks old. The \( Ybx3 \)−/− phenotyping experiments were performed with C57BL/6J male and female mice, CD45.2, and 6-12-week-old. Msc-GFP mouse experiments were performed with C57Bl/6 male and female mice, CD45.2, that were 8-10 weeks old. The Ybx3−/− mice were graciously provided by Dr. Timothy Ley at Washington University in St. Louis and have been previously described.

Msc-GFP reporter mice were obtained from The Jackson Laboratory. All animals were housed in a 12/12 light cycle with light cycle was from 6 AM to 8 PM. Ambient temperature was at 72°F with a humidity ranging from 30 to 40%.

Tissue processing and cell preparation. For flow-cytometry and cell-sorting experiments, bilateral femurs and tibias were harvested from mice. Bones were flushed with 1X PBS with 0.2% fetal bovine serum (Sigma Aldrich, 12133 C), 0.1% sodium azide (Rica, 7144-8-16), and 0.05 m ethanolamine (EDTA, Fisher Scientific, S3113), pH 7.4. The cells were filtered through a 70 μm mesh, centrifuged at 350 × g for 5 min, and then incubated for 5 min with 5 μL of ACK lysis buffer (0.15 M ammonium chloride (Fisher, A661), 10 mM potassium chloride (Fisher, 56379), 1:100), B220-BUV395 (RA3-6B2, BDBiosciences, 103222, 1:100), CD11c-BUV395, (N418, BD Biosciences, 744180, 1:100), Ly6G-APC780 (RB6-8C5, ThermoFisher, 47-5931-82, 1:100), Ly6C-BV421 (Ly6C-BV421, 86-8CS, ThermoFisher, 47-931-82, 1:100), Ly6G-BV421 (Ly6G-BV421, 86-8CS, ThermoFisher, 47-931-82, 1:100), NK1.1-APCef780 (PK136, ThermoFisher, 47-0114-82, 1:100), CD11c-BUV395, (N418, BD Biosciences, 744180, 1:100), Ly6C-BV421 (Ly6C-BV421, 86-8CS, ThermoFisher, 47-931-82, 1:100), Ly6G-BV421 (Ly6G-BV421, 86-8CS, ThermoFisher, 47-931-82, 1:100), NK1.1-APC780 (PK136, ThermoFisher, 47-9841-82, 1:100), CD11c-BUV395, (N418, BD Biosciences, 744180, 1:100), GhostRed780 (Tonbo Biosciences, 13-0865, 1:100), Ter119-APC780 (TER-119, ThermoFisher, 47-0114-82, 1:100), CD11c-BUV395, (N418, BD Biosciences, 744180, 1:100), and staining.

Flow cytometry and antibodies. All flow cytometry was performed using the BD Fortessa cytometers (BD Biosciences) in the University of Minnesota Flow Cytometry Core. Bone marrow cells obtained using the method above were stained with various FACS antibodies, including B220-UV395 (RA3-6B2, BD Biosciences, 563793, 1:100), B220-Blue (RA3-6B2, BDBiosciences, 558108, 1:100) B220-Pe-Cy7 (RA3-6B2, BioLegend, 103222, 1:100), CD11c-AF670 (N418, Ther- moFisher, 07-0114-82, 1:100), CD11c-BUV395, (N418, BD Biosciences, 744180, 1:100), GhostRed780 (Tonbo Biosciences, 13-0865, 1:100), Ter119-APC780 (TER-119, ThermoFisher, 47-9841-82, 1:100), NK1.1-APC780 (PK136, ThermoFisher, 47-9841-82, 1:100), CD11c-BUV395, (N418, BD Biosciences, 744180, 1:100), and staining.
CD43-Biotin (S7, BD Biosciences, 553269, 1:100), CD43-BV786 (S7, BD Biosciences, 740857, 1:100), CD19-BV605 (6D5, BioLegend, 115901, 1:100), IL7R-BV421 (ATR34, BioLegend, 115352, 1:100), CXC4-R4-PEcy7.5 (L726L2, BioLegend, 146509, 1:100), CD47-BV439 (In-1, BioLegend, 470274, 1:100), CD98-PACe780 (N418, Thermo Fisher, 47-0014-82, Ter119-PACe780 (TER-119, Thermo Fisher, 47-5921-82, 1:100), NK1.1-PACe780 (PKE-10, Thermo Fisher, 47-5921-82, 1:100), PC4-PACe780 (R6-C5, Thermo Fisher, 47-5931-82, 1:100), CD48-PACe780 (GK1.5, Thermo Fisher, 47-0041-82, 1:100), CD8-PACe780 (53-67-6, Thermo Fisher, 47-0081-82, 1:100), GhostDye Red780 viability dye (Tonbo Biosciences, 13-0865, 1:1000) and 1 µg of CITE-Seq antibodies (BioLegend, A0301, A0103, RA3-6B2, 1:100, and TotalSeq A0093, 6D5, BioLegend, 115559), CD3 (TotalSeq A0113, AA4.1, BioLegend, 136513), CD25 (TotalSeq A0097, PC61, BioLegend, 102055) and IgM (TotalSeq A0450, R-M1, BioLegend, 406535). Cells were stained with the above-mentioned antibodies for 20 minutes on ice, washed, and resuspended in 1X PBS with 2% fetal bovine serum (FBS), 2 mM ethylenediaminetetraacetic acid, and pH 7.4 buffer containing 1 µg of streptavidin-PE (TotalSeq A0113, R&D Systems, IC4260G, 1:100), EBFI-PE (T62-818, BD Biosciences, 565494, 1:100), and anti- GFP-FITC (Rockland, 600-402-215, 1:400) surface stained cells were fixed/permeabilized using the eBioscience Transcription Factor staining kit (eBioscience, 00-5523-00) for 30 minutes at room temperature, washed, and then stained for 30 minutes in permeabilization buffer. Cells were then washed and resuspended in 1X PBS with 2% fetal bovine serum (FBS), 0.1% sodium azide and 0.5 mM ethylenediaminetetraacetic acid, pH 7.4, for flow cytometric analysis. Cell sorting was performed on a BD FACSAria sorter (BD Biosciences). All flow-cytometry data acquired were analyzed using flowJo software (Tree Star).

Local field force expression data (ADT) were filtered to include only the WT singlets and counts were normalized according to the centered-log-ratio method in Seurat. For each of the six markers measured (B220, CD19, CD43, CD25, IgM, and CD43), the normalized counts were centered (subtracting the mean expression value from each class) and scaled (dividing centered value by standard deviation) across all cells. The Seurat object with S and G2M phase scores and a resolution ~0.4 was converted into a cell_data_set object for use with the Monocle (v3) R package. The aligned genes were used for the residual_model_formula_str = "-S.Score + G2M.Score" to adjust for the cell cycle status. After adjustment for cell cycle status, UMAP dimensional reduction and clustering were performed in Monocle. The final resolution for Monocle clustering was 0.005. This resolution resulted in 2 separate partitions for the clusters. The first partition of cluster was in its own sub-dataset and was excluded from pseudotime trajectory analysis. The remaining two partitions were 1) pre-pro B cells and 2) all other cell populations. We relabeled the partitions so that the pre-pro B cells and the remaining cell populations of interest would be in the same partition for the initial population pseudotime analysis. We initiated the pseudotime trajectory in the pre-pro B cell population, using a custom function described in the Monocle documentation (get_earliest_principal_node) to automatically select the starting point for the pseudotime trajectory analysis. To compare the Monocle and Seurat clustering results, the number of cells and frequency of cluster membership overlap between Monocle and Seurat was calculated.

In order to compare the way Monocle 3 methods identify modules of coregulated genes, we used the function graph_test to identify variable genes in the data using the Moran’s I statistic. Then we identified the genes that had a significant q-value (<0.05) from the autocorrelation analysis and then grouped these genes into modules using UMAP and Louvain community analysis. We used the enrichGO function in the clusterProfiler package (v 3.14.3) to evaluate enrichment of the modules in GO terms across all three ontologies (BP, CC, and MF) [57,64].

Lightfield analysis of mRNA expression (LISA) and gene set enrichment analysis (GSEA).

The top 100 differentially upregulated genes obtained from Seurat FindAllMarkers function were used as input into the LISA Cistrome (lisa.cistrome.org). The transcription factor ChIP-seq database was used to determine putative enhancer modules for differentially regulated genes of the pre-BCR-dependent proliferation stage. We used the Gene set enrichment analysis (GSEA) software to identify gene sets from the Molecular Signature Databases [50,70], comparing the differentially expressed genes in the pre-BCR-dependent proliferating cells versus the pre-BCR-dependent cells. The MSigDB (v 7.0) annotations were extracted from the msigdb R package [7.0.1].
Enrichment analysis of single cell RNA-Seq cluster marker genes in human B-ALL gene expression. We used enrichment analysis to examine whether human B-ALL subtypes have similarity to particular B cell progenitor populations. We created robust signatures for each cell type using the Seurat and Monocle top cluster marker genes independently, and we also created combined cluster marker gene sets by taking the intersection of the independent Seurat and Monocle cluster gene signatures for each of the different cell types we identified in the single cell data. The top marker genes are the genes that are differentially expressed for a given cluster when compared with all other clusters. Default approaches for finding top marker genes were used for both Seurat and Monocle. The findAllMarkers() function from Seurat was used as described in the previous section, "Single-cell Bioinformatics Analysis", to identify which genes are differentially expressed in one cluster compared with all other clusters for the Seurat cluster. The top_markers() function from Monocle was used to identify which genes are differentially expressed in one cluster compared with all others for the Monocle cluster definitions. We used the Monocle default parameters, so 25 genes were returned per Monocle cluster with the top_markers() function. We did not apply any additional filters to the top gene markers for the Monocle analysis. For the Seurat cluster marker genes we only used genes with positive log fold changes to better match the Monocle cluster gene marker results. We did not include downregulated marker genes from the Seurat analysis. For both Monocle and Seurat cluster gene lists we had to convert the gene names to human gene names, so some genes were removed that did not have appropriate human homologs. Additionally, when there was not a direct 1:1 mapping between the Monocle and Seurat clusters, cluster gene lists from one method were merged before comparison to the other method’s cluster gene lists. We then used the intersection of the Monocle and the Seurat lists to get our final list of genes for each cluster for our individual B-ALL heatmaps for each group of cluster-marker genes.

We downloaded publicly available B-ALL bulk RNA-Seq count data. We summed expression values for each gene across biological replicates for each B-ALL subtype to create an average sample for each subtype. We normalized and log2-transformed the data to create log2cpm values for unsupervised hierarchical clustering of each combined cluster marker gene set against all B-ALL subtypes. To gain a bigger picture view of the association between gene networks and human B-ALL subtypes, we created a heatmap showing a summarized cluster gene set z-score for each B-ALL subtype. We calculated each cluster gene set z-score by averaging the z-scores of the cluster marker genes in each individual cluster gene set for each B-ALL subtype. In the heatmap of these data we allowed for both column (B-ALL subtype) and row (average cluster z-score) hierarchical clustering.

Statistical analysis. Differential expression analysis (DE) analysis was completed using the FindMarkers function, employing a Wilcoxon rank-sum test between all pairwise clusters or between a single cluster vs. all others. Genes were considered significant if the absolute value of log2-fold-change was $> 0.25$ and Bonferroni-adjusted $p$-value $< 0.01$. The statistical significance of hierarchical clustering for the human B-ALL heatmaps was done using a Monte-Carlo based method. Specifically, the sigclust2 package was used with the parameters of Euclidean distance and complete linkage for clustering methods. Data and statistical analyses were performed using Prism 8 (Graphpad). A Shapiro-Wilk test was performed to assess data normality, and unpaired data that passed normality were analyzed using an unpaired student $t$-test. Paired data derived from the same mouse were analyzed using a parametric paired student $t$-test.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The data supporting the findings of this study are available within the paper and supplementary information file. Single-cell RNA-Seq data were deposited at Gene Expression Omnibus, with the primary accession code: GSE168158. Mouse genome mm10 was used as reference sequence (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20). EBF1. We then used the intergene obtained from GSM2863146. Bulk RNA-seq data for wild type and Pax5$^{−/−}$ Ebf1$^{−/−}$ leukemia for Fig. 3f were obtained from GSE148680. Source data are provided with this paper.

Code availability. No custom code was used or generated in this paper, as publicly available software and code was used (Seurat, Monocle, DoubletFinder, GSEA, and GO-term analysis). Code can be provided upon request.
30. Sigvardsson, M. et al. Early B-Cell Factor, E2A, and Pax-5 cooperate to activate the early B-cell specific mb-1 promoter. Mol. Cell. Biol. 22, 8539–8551 (2002).

31. Äkerblad, P., Rosberg, M., Leanderson, T. & Sigvardsson, M. The B2 (interacting globulin β)-Chain gene is a genetic target for early B-cell factor. Mol. Cell. Biol. 19, 392–401 (1999).

32. Katendralh, C. D. S. et al. Antagonism of B cell enhancer networks by STAT5 drives leukemia and poor patient survival. Nat. Immunol. 18, 694–704 (2017).

33. Qin, Q. et al. Lisa: inferring transcriptional regulators through integrative modeling of public chromatin accessibility and ChIP-seq data. Genome Biol. 21, 32 (2020).

34. Roychoudhuri, R. et al. BACH2 regulates CD8+ T cell differentiation by controlling access of AP-1 factors to enhancers. Nat. Immunol. 17, 851–860 (2016).

35. Sidwell, T. et al. Attenuation of TCR-induced transcription by Bach2 controls regulatory T cell differentiation and homeostasis. Nat. Commun. 11, 1–17 (2020).

36. Cordeiro Gomes, A. et al. Hematopoietic stem cell niches produce lineage-instructive signals to control multipotent progenitor differentiation. Immunity 45, 1219–1231 (2016).

37. Tikhonova, A. N. et al. The bone marrow microenvironment at single-cell resolution. Nature 569, 222–228 (2019).

38. Klassen, C. et al. MIF promotes B cell chemotaxis through the receptors CXCR4 and CD74 and ZAP-70 signaling. J. Immunol. 192, 5273–5284 (2014).

39. Yu, V. W. C. et al. Distinctive mesenchymal cell pairings govern B cell differentiation in the bone marrow. Stem Cell Rep. 7, 220–235 (2016).

40. Hoffmann, P. R. & Berry, M. J. The influence of selenium on immune responses. Mol. Nutr. Food Res. 52, 1273–1280 (2008).

41. Choi, Y. S., Dieter, J. A., Rothaeusler, K., Luo, Z. & Baumgarth, N. B-1 cells in the bone marrow are a significant source of natural IgM. Eur. J. Immunol. 42, 120–129 (2012).

42. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386 (2014).

43. Froissard, M. et al. Cell circuits between B cell progenitors and IL-7+ mesenchymal progenitor cells control B cell development. J. Exp. Med. 215, 2586–2599 (2018).

44. Joshi, I. et al. Loss of Ikaros DNA-binding function confers integrin-dependent survival on pre-B cells and progression to acute lymphoblastic leukemia. Nat. Immunol. 15, 294–304 (2014).

45. Liu, W. et al. De Novo B1 double-hit B-cell precursor acute lymphoblastic leukemia (BCP-ALL) in pediatric and young adult patients associated with poor prognosis. Pediatr. Hematol. Oncol. 32, 535–547 (2015).

46. Mullighan, C. G. et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N. Engl. J. Med. 360, 470–480 (2009).

47. Hilden, J. M. et al. Analysis of prognostic factors of acute lymphoblastic leukemia in infants: report on CCG 1953 from the Children’s Oncology Group. Blood 108, 441–451 (2006).

48. Rubnitz, J. E. et al. Prospective analysis of TEL gene rearrangements in childhood acute lymphoblastic leukemia: a children’s oncology group study. J. Clin. Oncol. 26, 2186–2191 (2008).

49. Bhogwani, D. et al. ETV6-RUNX1-positive childhood acute lymphoblastic leukemia: Improved outcome with contemporary therapy. Leukemia 26, 265–270 (2012).

50. Qin, Y. Z. et al. The prognostic significance of ZNF384 fusions in adult Ph-negative B-cell precursor acute lymphoblastic leukemia: a comprehensive cohort study from a single chinese center. Front. Oncol. 11, 632532 (2021).

51. Hiraibashy, S. et al. Clinical characteristics and outcomes of B-ALL with ZNF384 rearrangements: a retrospective analysis by the Ponte di Legno Childhood ALL Working Group. Leukemia https://doi.org/10.1038/s41375-021-01199-0 (2021).

52. Kimes, P. K., Liu, Y., Hayes, D. N. & Marron, J. S. Statistical significance for hierarchical clustering. Biometrics 73, 811 (2017).

53. Schaum, N. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 562, 367–372 (2018).

54. Baccin, C. et al. Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. Nat. Cell Biol. 18, 3409–3417 (2016).

55. 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).

56. Heng, T. S. P. et al. The immunological genome project: networks of gene expression in immune cells. Nat. Immunol. 9, 1091–1094 (2008).

57. Schaum, N. et al. Metabolic gatekeeper function of B-lymphoid transcription factors. Nature 542, 479–483 (2017).

58. Herglotz, J. et al. Essential control of early B-cell development by Mef2 transcription factors. Blood 127, 572–581 (2016).

59. Harvey, R. C. et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. Blood 116, 4874–4884 (2010).

60. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst. 8, 339–357.e4 (2019).

61. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296 (2019).

62. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502 (2019).

63. Carbon, S. et al. The gene ontology resource: 20 years and still going strong. Nucleic Acids Res. 47, D330–D338 (2019).

64. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296 (2019).

65. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502 (2019).

66. Liberson, A. et al. The molecular signatures database Hallmark gene set collection. Cell Syst. 1, 417–425 (2015).

67. Gentles, A. J. et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med. 21, 938–945 (2015).

Acknowledgements
We thank G. Hubbard, A. Rost, and N. Keller, for mouse and technical assistance, J. Motl and P. Champoux for cell sorting and Flow Cytometry Core Facility maintenance at the University of Minnesota (5P01AI035296), D. George for providing Ybx3−/− bone marrow, E. Stanley, I. Daniels, and Dr. K. Beckman and the University of Minnesota Genomics Center for 10X genomics single-cell capture and sequencing, Dr. Meinrad Busslinger for Pax5−/− mice, Dr. Rudolf Grosschedl for Ebf1−/− mice, Dr. Tim Ley for Ybx3−/− mice, and Drs. J. Pereira, T. Leiben, and D. Owen for review and comments on the paper. The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results presented within this paper. This work was supported by an individual predoctoral F30 fellowship from the NIH (F30CA232399) and T32 training grant (T32 GM080244), R.D.L. and, NIH grants R01AI124512, R01AI147540, and R01CA232317, M.A.F.

Author contributions
Conceptualization, R.D.L. and M.A.F.; Methodology, R.D.L. and M.A.F.; Formal Analysis, R.D.L., S.A.M., T.P.K., R.S.L., L.H.H., and M.A.F.; Investigation, R.D.L., L.H.H.; Resources, M.A.F.; Data Curation, S.A.M., T.P.K., and R.S.L.; Writing—Original Draft, R.D.L. and M.A.F.; Writing—Review & Editing, R.D.L., S.A.M., T.P.K., R.S.L., M.A.F.; Visualization, R.D.L. and L.H.H.; Supervision, M.A.F.; Project Administration, M.A.F.; Funding Acquisition, R.D.L. and M.A.F.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-27232-5.

Correspondence and requests for materials should be addressed to Michael A. Farrar.

Peer review information Nature Communications thanks Rachel Gerstein and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer-reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
