

# Single-cell atlas of AML reveals age-related gene regulatory networks in t(8;21) AML

Jessica Whittle<sup>1,2</sup>, Stefan Meyer<sup>3,4,5</sup>, Georges Lacaud<sup>2\*</sup>, Syed Murtuza-Baker<sup>1\*</sup>, Mudassar Iqbal<sup>1\*</sup>

<sup>1</sup>Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom; <sup>2</sup>Stem Cell Biology Group, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, United Kingdom; <sup>3</sup>Manchester Cancer Research Centre (MCRC), Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom; <sup>4</sup>Department of Paediatric and Adolescent Oncology, Royal Manchester Children's Hospital, Manchester, United Kingdom; <sup>5</sup>Department of Adolescent Oncology, The Christie NHS Foundation Trust, Manchester, United Kingdom

**\*For correspondence:**

georges.lacaud@manchester.ac.uk (GL);  
syed.murtuzabaker@manchester.ac.uk (SM-B);  
mudassar.iqbal@manchester.ac.uk (MI)

**Competing interest:** The authors declare that no competing interests exist.

**Funding:** See page 17

**Sent for Review**

15 January 2025

**Preprint posted**

17 January 2025

**Reviewed preprint posted**

15 May 2025

**Reviewed preprint revised**

12 November 2025

**Version of Record published**

11 February 2026

**Reviewing Editor:** Murim

Choi, Seoul National University, Republic of Korea

© Copyright Whittle *et al.* This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## eLife Assessment

This manuscript provides a single-cell transcriptomic atlas for AML (222 samples comprising 748,679 cells) integrating data from multiple studies. They use this dataset to investigate t(8;21) AML, and they reconstruct the Gene Regulatory Network and enhancer Gene Regulatory Network, which allowed identification of interesting targets. This aggregation is **important** and can help infer differences in genetic regulatory modules based on the age of disease onset. Their **compelling** effort may help explain age-related variations in prognosis and disease development in subtype-specific manner.

**Abstract** Acute myeloid leukemia (AML) is characterized by cellular and genetic heterogeneity, which correlates with clinical course. Although single-cell RNA sequencing (scRNA-seq) reflects this diversity to some extent, the low sample numbers in individual studies limit the analytic potential when comparing specific patient groups. We performed large-scale integration of published scRNA-seq datasets to create a unique single-cell transcriptomic atlas for AML (AML scAtlas), totaling 748,679 cells, from 159 AML patients and 51 healthy donors from 20 different studies. This is the largest single-cell data resource for human AML to our knowledge, publicly available at <https://cellxgene.cziscience.com/collections/071b706a-7ea7-47a4-bddf-6457725839fc>. This AML scAtlas allowed investigations into 20 patients with t(8;21) AML, where we explored the clinical importance of age, given the in-utero origin of pediatric disease. We uncovered age-associated gene regulatory network (GRN) signatures, which we validated using bulk RNA sequencing data to delineate distinct groups with divergent biological characteristics. Furthermore, using an additional multiomic dataset (scRNA-seq and scATAC-seq), we validated our initial findings and created a de-noised enhancer-driven GRN reflecting the previously defined age-related signatures. Applying integrated data analysis of the AML scAtlas, we reveal age-dependent gene regulation in t(8;21) AML, potentially reflecting immature/fetal HSC origin in prenatal origin disease vs postnatal origin. Our analysis revealed that BCLAF1, which is particularly enriched in pediatric AML with t(8;21) of inferred in-utero origin, is a promising prognostic indicator. The AML scAtlas

provides a powerful resource to investigate molecular mechanisms underlying different AML subtypes.

---

## Introduction

Acute myeloid leukemia (AML) is an aggressive blood cancer driven by non-random genomic rearrangements in hematopoietic stem/progenitor cells (HSPCs). Recurrent AML-associated genomic aberrations, which often involve transcriptional or epigenetic regulators, give rise to distinct patterns of gene expression strongly associated with clinical course and chemotherapy response (Tenen, 2003; Bolouri et al., 2018). Single-cell RNA sequencing (scRNA-seq) studies have demonstrated that HSPCs acquire lineage priming at an early stage when still phenotypically immature and disperse down an erythromyeloid or lymphomyeloid differentiation trajectory (Velten et al., 2017). In the context of AML, diverse clonal hierarchies include the co-existence of normal hematopoietic clones. Leukemic clones can partially recapitulate myeloid differentiation and have been shown to display functional differences even when defined by the same genotype (Velten et al., 2021; van Galen et al., 2019; Beneyto-Calabuig et al., 2023; Zeng et al., 2022). Indeed, analysis of AML using scRNA-seq has revealed key clonal hierarchies, defining subtype-associated cell types, and dynamic changes following therapy, and have been critical in characterizing leukemic stem cells (LSCs), which propagate the disease and drive relapse (van Galen et al., 2019; Beneyto-Calabuig et al., 2023; Zeng et al., 2022; Stetson et al., 2021).

Most AML scRNA-seq studies are limited by small sample numbers and include a mixture of different AML subtypes which may not be directly comparable to one another. Therefore, it is difficult to make biological conclusions with sufficient robustness to be clinically translatable in these individual datasets. To overcome this, we performed large-scale integration of public scRNA-seq datasets to create a single-cell transcriptomic atlas for AML (AML scAtlas). Due to the range of data sources spanning time, locations, and experimental designs, complex batch effects often arise between scRNA-seq datasets, which requires a tailored data integration approach (Luecken et al., 2022; Heumos et al., 2023). Thus, we benchmarked some widely used batch correction tools (Korsunsky et al., 2019; Lopez et al., 2018; Xu et al., 2021) for our specific data use case.

Given the broad representation of age groups in AML scAtlas, we sought to investigate a developmental aspect of AML biology. Pediatric AML has substantially better clinical outcomes compared to adult AML (Balgobind et al., 2011; Wiggers et al., 2019; Chaudhury et al., 2018). The molecular landscape of AML differs between children and adults (Bolouri et al., 2018; Balgobind et al., 2011; Wiggers et al., 2019; Chaudhury et al., 2018) this may, in part, reflect differences in the developmental origins of the disease. Chromosomal changes in pediatric leukemia are acquired in-utero, as evidenced by leukemia-specific genomic aberrations detected in the Guthrie spots of children who later developed leukemia, sometimes several years after birth (Wiemels et al., 2002). Adult leukemia, in contrast, is thought to develop later in life through acquisition of pre-leukaemic changes and clonal evolution of adult HSPCs (Welch et al., 2012; Jaiswal et al., 2014). The impact of these developmental stages on leukemia biology remains incompletely understood, and no current methods exist to quantify and characterize differences in the origin of the disease. However, as childhood AML with presumed in-utero origin has a better outcome, for teenagers and young adults, determination of the pre- or postnatal origin might be important for better treatment stratification and prognostication.

AML with t(8;21) (AML-ETO/RUNX1-RUNX1T1) is one of the most frequent AML subtypes in young people, although it affects all ages (Bolouri et al., 2018). The prenatal origins of the t(8;21) rearrangement has been confirmed even in older children presenting with AML (Wiemels et al., 2002). The prognosis of AML with t(8;21) is better in children than in teenagers and even more so than in young adults (National Cancer Registration and Analysis Service, Northern Ireland Cancer Registry, Scottish Cancer Registry, Unit WClas, 2021). This outcome difference is not fully explainable by co-morbidities and may instead be related to the developmental origins of the disease. In the intermediate teenage group, t(8;21) AML may comprise both late childhood and early adult disease entities, a distinction that could have prognostic implications and could help to explain disease biology and clinical course.

We leveraged our AML scAtlas resource to characterize age and developmental stage-specific signatures in t(8;21) AML by applying single-cell gene regulatory network (GRN) inference (Aibar

*et al., 2017; Van de Sande et al., 2020*), as a means of revealing cell state heterogeneity across age groups. We then validated and refined our findings in a larger cohort using bulk RNA sequencing (RNA-seq) data from the TARGET (*Bolouri et al., 2018*) and BeatAML (*Burd et al., 2020*) studies, defining age-associated GRN signatures and key regulators of t(8;21) AML, that may reflect the developmental origins of the leukemia.

Profiling both gene expression and chromatin accessibility together can decipher the enhancer-driven GRN (eGRN) and enriched transcriptional regulators. Significant heterogeneity across different patients and time points (*Lambo et al., 2023*) was recently described by analyzing combined scRNA-seq and single-cell Assay for Transposase Accessible Chromatin sequencing (scATAC-seq). We used the t(8;21) AML data from this study to validate our initial findings, by applying cutting edge GRN inference methodology (*Bravo González-Blas et al., 2023*). This encompasses both modalities to provide a denoised eGRN which we could correlate with our age-associated signatures.

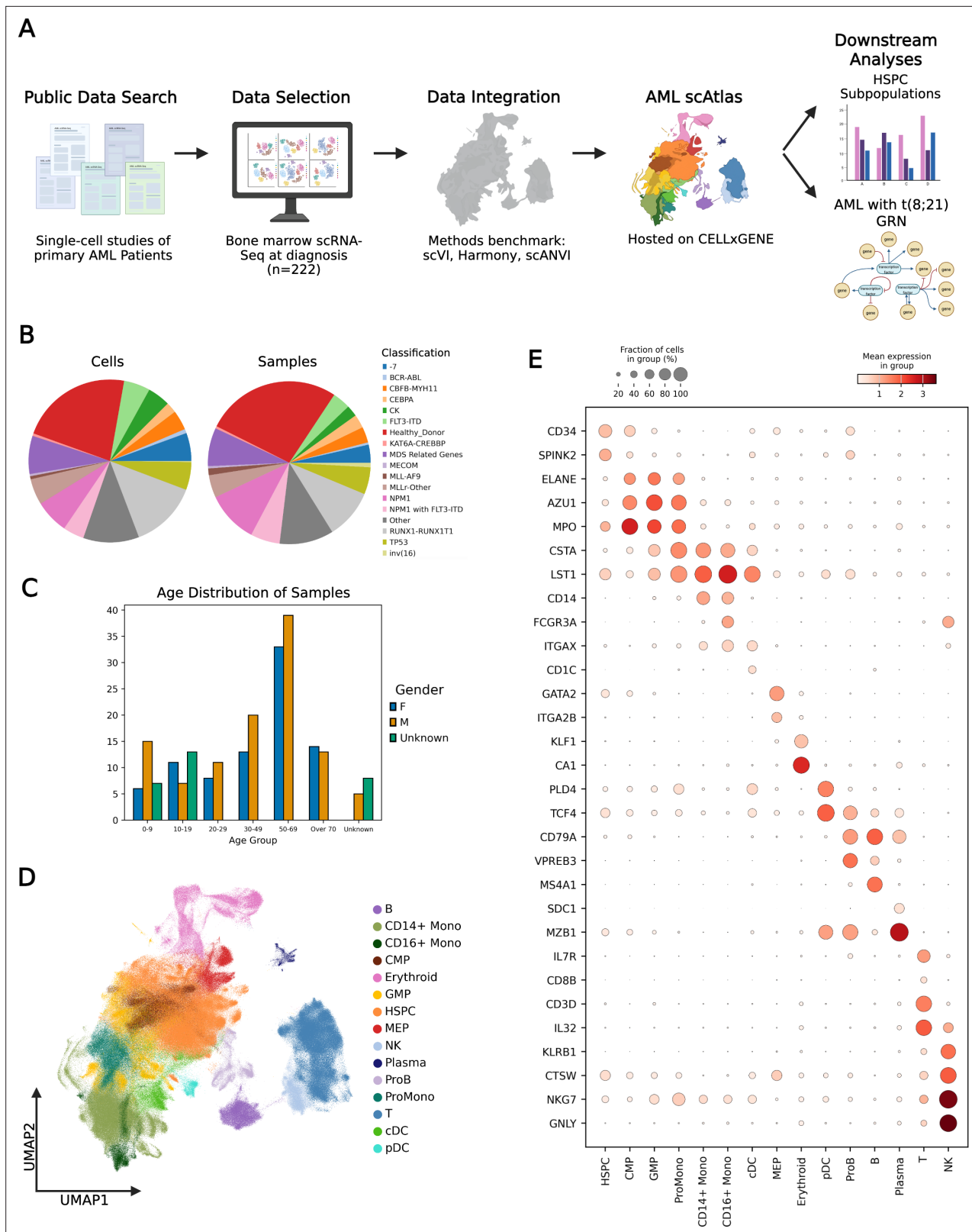
## Results

### Large scale data integration to construct a single-cell transcriptomic atlas of AML (AML scAtlas)

To create the AML scAtlas, we integrated published scRNA-seq data of primary AML bone marrow samples, from 16 suitable high-quality studies (see Materials and methods), comprising 159 AML samples (*Figure 1A; Source data 1*). Where on-treatment time points were available, we selected only diagnostic samples to establish a reference atlas of primary AML at diagnosis. If studies had healthy donor bone marrow samples, these were included, alongside data from healthy bone marrow samples from four additional scRNA-seq studies (*Source data 1*) to enable comparisons between malignant and healthy bone marrow populations. After cell filtering and quality control, the AML scAtlas contains data from 748,679 high quality cells derived from a total of 20 different scRNA-seq studies (*Velten et al., 2021; van Galen et al., 2019; Beneyto-Calabuig et al., 2023; Stetson et al., 2021; Zheng et al., 2017; Petti et al., 2019; Jiang et al., 2020; Johnston et al., 2020; Pei et al., 2020; Li et al., 2023; Lasry et al., 2023; Fiskus et al., 2023; Naldini et al., 2023; Mumme et al., 2023; Zhang et al., 2023; Li et al., 2022; Oetjen et al., 2018; Setty et al., 2019; Caron et al., 2020; Figure 1—figure supplement 1A*). Each sample was assigned to an AML clinical subtype, based on the recent European Leukemia Net (ELN) clinical guidelines (*Döhner et al., 2022*), and classified into the corresponding prognostic risk group. This resource captures a broad range of molecular subtypes of AML and spans different age groups, including both pediatric and adult AML cases (*Figure 1B–C*). Overall, this is the largest dataset to date for exploring AML biology at single-cell resolution.

In the initial analysis of the combined dataset, batch effects were noted with study-specific clustering, which was quantified using several benchmarking metrics (*Figure 1—source data 2; Figure 1—figure supplement 1A and B*). Even within samples of the same study, sample-wise clustering was noted (*Figure 1—figure supplement 1C*). To address this, we benchmarked several widely used batch correction methods (*Figure 1—source data 2; Figure 1—figure supplement 2A–C*), identifying scVI as the best method for this dataset (*Figure 1—source data 2*). We, therefore, employed scVI to correct for batch effects, before clustering and cell type annotation in the AML scAtlas, by using the consensus of multiple annotation tool results (*Figure 1—source data 2*), verified using cluster-wise marker gene expression (*Figure 1D–E*).

Cell type proportions analyses across the clinically relevant subtypes in the dataset show that the AML subtypes were significantly biased towards myeloid cell types (CMP, MEP, GMP, ProMono, CD14+ Mono, CD16+ Mono, cDC, Erythroid) with each subtype exhibiting a clear predominant cell type consistent with AML clonal expansion (*Figure 2A; Figure 2—figure supplement 1A–B*). In contrast, healthy donor samples had more balanced lineage proportions, with lymphoid cells (T, B, NK, ProB, pDC, Plasma) well represented (*Figure 2A; Figure 2—figure supplement 1A–B*). Given the established critical role of HSPCs and LSCs in propagating AML, and their importance as therapeutic targets (*Montefiori et al., 2021*), we focused on HSPC clusters for further analysis (*Figure 2B*). To identify LSCs, we applied a curated reference profile of leukaemic stem and progenitor cells (LSPCs) (*Zeng et al., 2022; Figure 2B–C*) and correlated this with calculated LSC6 (*Elsayed et al., 2020*) and LSC17 (*Ng et al., 2016*) scores for each cell (*Figure 2D*). We then compared the proportions of HSPC/LSPCs across different AML subtypes and risk groups, as defined by the ELN clinical guidelines (*Döhner*



**Figure 1.** Large scale data integration creates a single-cell atlas of acute myeloid leukemia (AML). **(A)** Overview of the analysis steps in creating AML scAtlas. **(B)** Proportion of cells (left panel) and samples (right panel) belonging to each AML subtype as defined by the European Leukemia Net (ELN) clinical guideline. **(C)** Age group and gender distribution of AML single-cell Atlas (scAtlas) cohort samples. **(D)** scVI harmonized UMAP colored by

Figure 1 continued on next page

Figure 1 continued

annotated cell types. (E) The expression of key hematopoietic marker genes across annotated cell types shown on a dotplot. Color scale shows mean gene expression, dot size represents the fraction of cells expressing the given gene.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

**Source data 1.** Batch correction benchmarking metrics.

**Source data 2.** Automated cell type annotation results.

**Figure supplement 1.** Initial analysis establishes presence of batch effects.

**Figure supplement 2.** Benchmarking batch correction methods.

---

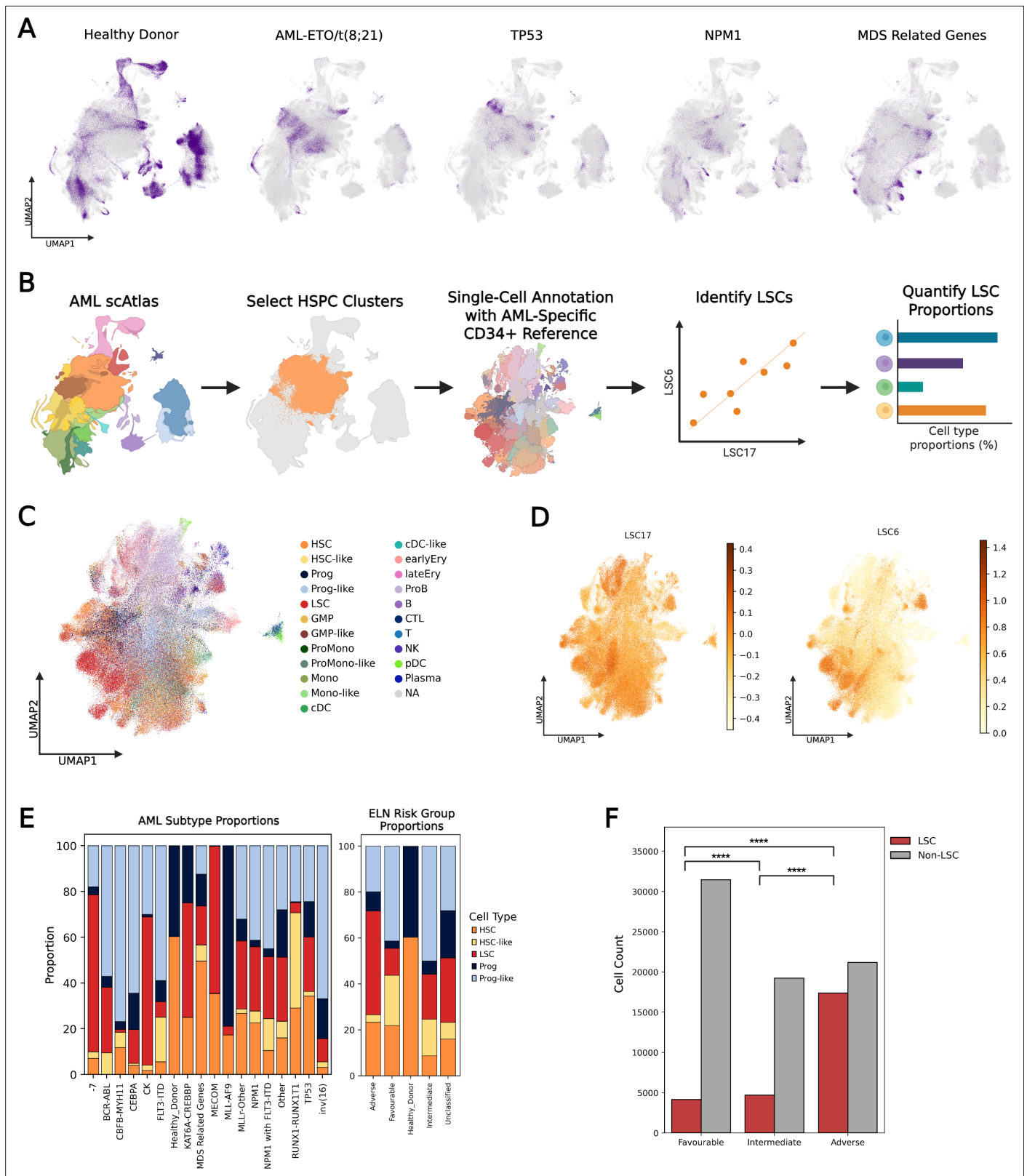
*et al., 2022; Figure 2E*). Higher-risk subtypes displayed a higher proportion of LSCs compared to favorable risk disease (*Figure 2E–F*).

## Application of AML scAtlas to identifying age-associated gene regulatory networks in t(8;21) AML

The AML scAtlas enables robust comparison of adult and pediatric AML. We hypothesized that in adolescents and young adults with t(8;21) AML, the potential for either in-utero or postnatal HSPC origin disease might affect disease biology and prognosis. Thus, we sought to explore biological differences between pediatric and adult cases of t(8;21) AML, aiming to explain and potentially improve prognostication in adolescents and young adults. We selected samples with t(8;21) AML from the AML scAtlas, resulting in 105,663 cells from 13 adult cases (aged 20–67), seven adolescent cases (aged 12–17), and three pediatric cases (aged 6–8) (*Figure 3A–C*). Where gender information was not available, this was inferred from ChrY/XIST gene expression (*Figure 3—figure supplement 1A*). Several adult samples underwent CD34 selection in original studies, excluding more differentiated cell types (mature lymphoid populations, monocytes, granulocytes) in these samples. Thus, these cell types were excluded from comparative analysis, focusing only on HSPCs and myeloid progenitors (CMP, GMP, MEP), which were well represented in all studies (*Figure 3C*).

We reconstructed the GRN for the t(8;21) subset using the pySCENIC (*Van de Sande et al., 2020*) pipeline, which is a Python-based efficient implementation of original SCENIC method (*Aibar et al., 2017*). It is a state-of-the-art method for network inference from scRNA data, popular in the community (*Hamed et al., 2022; Barnett et al., 2024; Zhang et al., 2024*) and has shown strong performance in a recent benchmarking study (*Nguyen et al., 2021*). SCENIC's three major steps are: First, it identifies groups of co-expressed genes as potential targets of a transcription factor (TF). Second, it refines these groups of genes based on the enrichment of the corresponding TF binding motif, forming 'regulons.' Third, it uses the AUCell method (embedded within SCENIC) to quantify the activity of each regulon in every cell. AUCell calculates the Area Under the Curve (AUC) for the regulon's genes set in a ranking of all genes by expression for each cell. The top 20 regulons for each age group were selected based on the regulon specificity score (RSS) (*Figure 3—figure supplement 1B*). Unsupervised clustering on the Zscore normalized regulon activity score matrix revealed clear differences in the GRN across different age groups (*Figure 3—figure supplement 1C*). We hypothesize that the differences in the GRN might reflect differences in the pre- or postnatal developmental origins of the disease. Additional testing of GRN inference from individual studies shows that the high number of cells refines the overall GRN (see Methods; *Figure 3—figure supplement 1D–E*).

To define gene regulatory programs (co-occurring gene modules, defined by a transcription factor and its targets) which are specific to different age groups (termed 'regulon signature'), we used the clustered dendrogram to select the regulon clusters most associated with the pediatric (below 10 years-old) and adult samples (over 18 years-old) (*Figure 3D; Figure 3—figure supplement 1C*). The pediatric regulon signature, proposed to represent in-utero origin t(8;21) AML (henceforth termed 'inferred-prenatal'), includes 16 regulons defined by a distinct group of hematopoietic transcription factors (TFs) (TRIM28, CTCF, RAD21, SOX4, TAL1, MYB, FOXN3, JUN, BCLAF1, ZBTB7A, IKZF1, MAZ, REST, YY1, CUX1, KDM5A), many of which have clearly defined roles in HSPCs and AML (*Lu et al., 2018; Fisher et al., 2017; Kumar et al., 2023*). The adult regulon signature, presumed representative of the postnatally acquired t(8;21) AML (henceforth termed 'inferred-postnatal'), combines three discrete clusters of regulons (YBX1, ENO1, and HDAC2; GATA1, POLE3, TFDP1, MYBL2, E2F4, and KLF1; IRF1, STAT1, IRF7, MAFF, ATF4, TAGLN2, SPI1, and KLF2), defined by TFs previously



**Figure 2.** Characterizing cell type distributions in acute myeloid leukemia (AML) subtypes. **(A)** UMAP highlighting the distribution of cells from different AML subtypes in AML single-cell Atlas (scAtlas). **(B)** Schematic showing the workflow used to identify leukemic stem cells (LSCs) from the AML scAtlas hematopoietic stem and progenitor cell (HSPC) clusters. **(C)** Using the AML scAtlas HSPC clusters only, UMAP was regenerated and annotated with an AML-specific reference of leukemia stem and progenitor cells (LSPCs). **(D)** UMAPs showing the leukemic stem cell scores of each cell, for the LSC17

Figure 2 continued on next page

Figure 2 continued

(left) and LSC6 (right). (E) Proportions of HSPC/LSPC populations in different AML subtypes (left) and AML risk groups (right), as defined by European Leukemia Net (ELN) clinical guidelines. (F) Comparison of LSC abundance in favourable and adverse ELN risk groups. Chi-Square test statistic: 8658.98, degrees of freedom: 1, p-value: 0.0.

The online version of this article includes the following figure supplement(s) for figure 2:

**Figure supplement 1.** Cell type proportions vary by acute myeloid leukemia (AML) subtype.

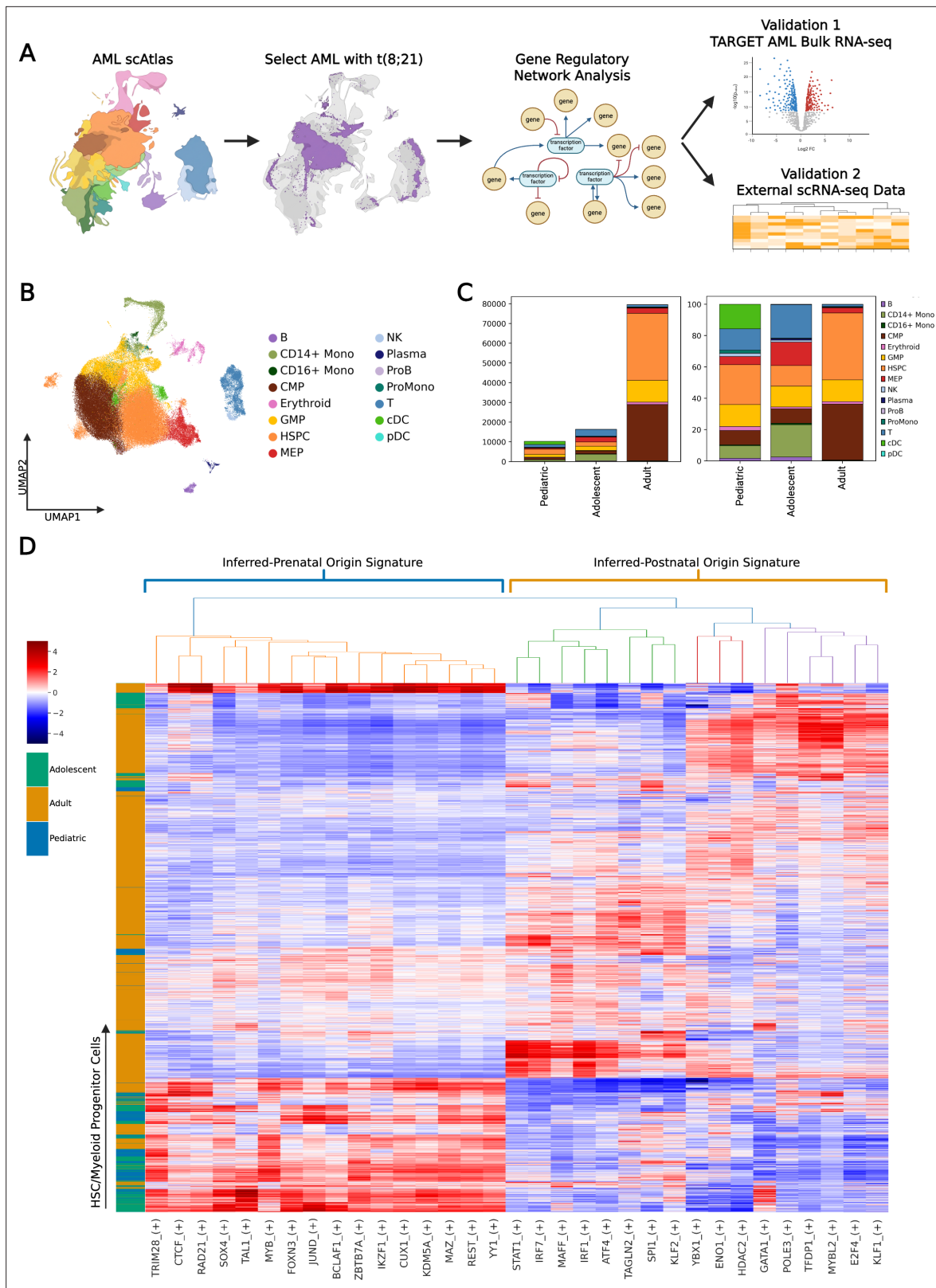
implicated in various hematopoietic, leukemic and inflammatory processes (Ning et al., 2011; Fischer et al., 2019; Fang et al., 2018). Importantly, both signatures contain key components of the AP-1 complex, which is heavily implicated in the biology of t(8;21) AML (Ptasinska et al., 2019; Martinez-Soria et al., 2018) and undergoes dynamic changes during aging (Patrick et al., 2024). Samples of 6 individuals aged 12–17 clustered with the pediatric samples and showed enrichment for the inferred-prenatal signature (Figure 3D), suggesting that older adolescents (up to age 17 in our cohort) more closely resemble pediatric AML with t(8;21) and remain biologically distinct from adult-onset disease. This implies that the inferred in-utero origin of t(8;21) AML can also be present in AML diagnosed in older children.

### Validation of age-associated regulons in bulk-RNA-seq cohorts of t(8;21) AML

We next sought to externally validate our age-associated regulon signatures in a larger cohort of patients. Bulk RNA-seq samples were obtained from the TARGET (Bolouri et al., 2018) and BeatAML (Burd et al., 2020) cohorts, selecting bone marrow samples taken at diagnosis in line with AML scAtlas data (n=83; Source data 2). We applied the AUCell algorithm from pySCENIC (Van de Sande et al., 2020) to calculate the activity of our pediatric inferred-prenatal and adult inferred-postnatal regulons in each sample. Unsupervised clustering of the bulk RNA-seq AUCell results revealed discrete clusters of samples that were highly enriched for our inferred-prenatal and inferred-postnatal origin-associated regulons (Figure 4A).

Given the limitations of most scRNA-seq platforms in detecting lowly expressed genes, notably TFs, we leveraged bulk RNA-seq samples to refine our identified gene regulatory networks by detecting differentially expressed regulon-associated TFs. We used our inferred-prenatal and inferred-postnatal signature clusters and performed differential gene expression analysis between these samples, using two widely used tools (DESeq2 Robinson et al., 2010 and edgeR Love et al., 2014) to ensure robustness of the results (Figure 4B; Figure 4—figure supplement 1A). We then compared differentially expressed regulon-associated TFs between the two groups and intersected this with the differential genes detected by each method. Although changes in TF expression are subtle (Figure 4B; Figure 4—figure supplement 1A), we identify significantly differentially expressed TFs which reflect the observed differences in regulon activity and indicate the most critical regulons in our age-related GRN signatures (Figure 4A–B; Figure 4—figure supplement 1B). This further delineated the inferred-prenatal signature to five key TFs (KDM5A, REST, BCLAF1, YY1, and RAD21), and the inferred-postnatal signature to eight TFs (ENO1, TFDP1, MYBL2, TAGLN2, KLF2, IRF7, SPI1, and YBX1).

We next performed gene set enrichment analysis (GSEA) on significantly differentially expressed genes as determined by edgeR (Love et al., 2014), to investigate pathways enriched in the inferred-prenatal samples compared to the inferred-postnatal ones (Figure 4C). Notably, inferred-prenatal samples showed increased expression of stemness-associated genes, and SMARCA2 target genes, a key player in HSC gene expression regulation and chromatin remodeling (Holmfeldt et al., 2016). SMARCA2 is also known to be upregulated during the fetal-to-adult HSC transition (Chen et al., 2019), implying that the observed SMARCA2 enrichment may indeed reflect the inferred fetal HSC cell-of-origin. Genes impacted by YY1 depletion were also downregulated compared to the samples of inferred-postnatal leukemia origin, which supports the identification of YY1 as an inferred-prenatal regulon (Figure 4C). To explore therapeutic implications, we performed GSEA using drug response signatures from published studies (Unnikrishnan et al., 2017; Williams et al., 2020; Xu et al., 2019; Zhang et al., 2020; Figure 4D). This analysis revealed that inferred-prenatal origin t(8;21)



**Figure 3.** Acute myeloid leukemia (AML) single-cell Atlas (scAtlas) reveals age-associated heterogeneity in t(8;21) AML. **(A)** Depiction of the workflow to generate and validate the t(8;21) AML gene regulatory network (GRN) from AML scAtlas. **(B)** Using the AML scAtlas t(8;21) sample cells, UMAP is re-computed and shows the different cell types. **(C)** Bar plots of the absolute cell type numbers (left panel) and the cell type proportions (right panel) stratified by age group. The CD34 enrichment performed on several adult samples is reflected. **(D)** Using HSPCs and CMPs only, the pySCENIC gene

*Figure 3 continued on next page*

Figure 3 continued

regulatory network (GRN) and regulon AUC scores were calculated. Z-score normalized scores underwent hierarchical clustering to create a clustered heatmap and identify age-associated regulons. Regulons were prioritized using their regulon specificity scores (RSS).

The online version of this article includes the following figure supplement(s) for figure 3:

**Figure supplement 1.** Acute myeloid leukemia (AML) with t(8;21) pySCENIC analysis.

AML is enriched for genes associated with increased chemosensitivity to cytarabine, venetoclax, and daunorubicin.

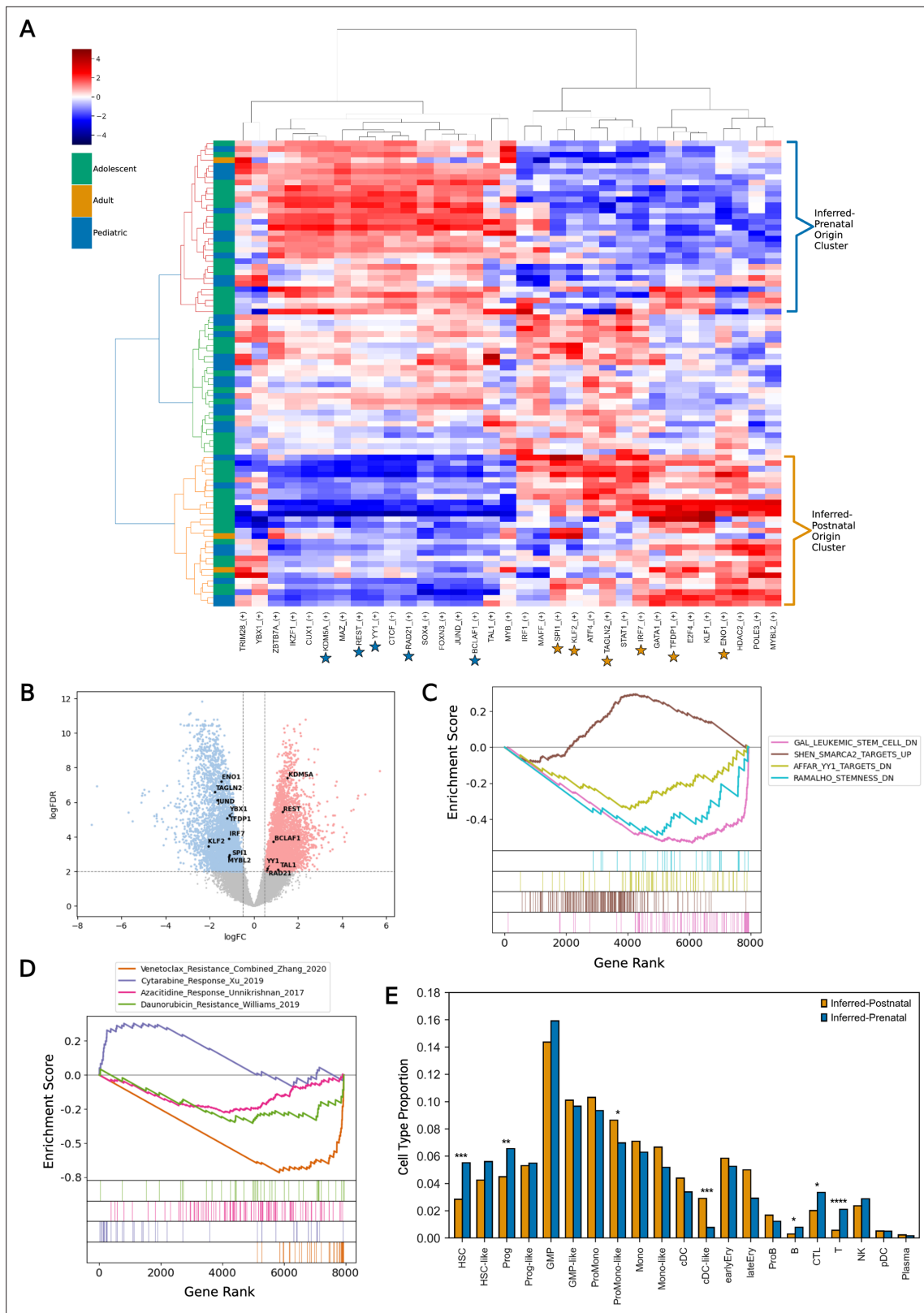
We hypothesized that the increase in stemness-associated genes in the leukemia with the inferred-prenatal origin could be reflective of potential differences in the leukemic cell-of-origin and its impact on myeloid differentiation. We therefore performed cell type deconvolution using AutoGeneS (Aliee and Theis, 2021), with a curated LSPC reference profile (Zeng et al., 2022), to compare the cellular heterogeneity between prenatal and postnatal origin bulk RNA-seq samples (Figure 4E). This revealed a higher proportion of HSPC cell types (HSC, Prog), with a reduction in some differentiated myeloid cell types (ProMono-like, cDC-like) in the samples of inferred-prenatal origin (Figure 4E). To corroborate this finding, we examined cell type proportions in the original t(8;21) subset of AML scAtlas, confirming that cells with the inferred-prenatal signature comprise more HSCs than inferred-postnatal signature cells (Figure 4—figure supplement 1D–E). However, comparison of cell type proportions in this dataset is confounded by differences in sample processing as some studies performed CD34 selection, hence, there is more cell type diversity observed in the pediatric samples (Figure 4—figure supplement 1D–E).

## Multomics single-cell data reveals a denoised GRN and identifies candidate perturbations in prenatal origin t(8;21) AML

We next used the scRNA-seq and scATAC-seq data from a recent cohort of pediatric t(8;21) AML patients (Lambo et al., 2023) at multiple clinical time points to uncover the enhancer-driven GRN (eGRN) in inferred-prenatal and inferred-postnatal origin t(8;21) AML (Source data 2). Initially, we identified two representative samples of our inferred-prenatal and inferred-postnatal signatures by using pySCENIC AUCCell (Van de Sande et al., 2020) to measure the activity of our previously defined regulons. Unsupervised clustering of the AUC scores was used to infer whether each sample matched the regulon signatures, identifying one inferred-prenatal sample and one inferred-postnatal sample for downstream analysis (Figure 5—figure supplement 1A).

We then applied SCENIC+ (Bravo González-Blas et al., 2023), which integrates scRNA-seq and scATAC-seq to identify candidate enhancer regions and TF-binding motifs, linking TFs to target genes and identified enhancers. This creates enhancer-driven regulons (eRegulons), forming an eGRN. We applied SCENIC+ (Bravo González-Blas et al., 2023) to the leukemia samples with the inferred-prenatal and inferred-postnatal origin at diagnosis and relapse, keeping only regulons that showed a correlation between both modalities to retain only the most robust regulons (Figure 5—figure supplement 1B). This revealed several eRegulons across both patients (Figure 5—figure supplement 1D), many of which were patient-specific, particularly when comparing HSPC populations (Figure 5A). The inferred-prenatal sample displayed a specific HSC eRegulon profile. In contrast, the inferred-postnatal sample more closely resembled the corresponding Granulocyte-Monocyte Progenitor (GMP) (Figure 5A). Interestingly, at relapse the inferred-prenatal origin patient undergoes a chemotherapy-driven lineage switch to a lymphoid phenotype, which may suggest that the leukemia originated from a less committed progenitor (Figure 5A).

To identify clusters of closely related eRegulons, we computed the correlations between eRegulons enrichment. We identified two main clusters of eRegulons which correspond to different inferred-signature samples (Figure 5B; Figure 5—figure supplement 1C). For each eRegulon cluster, we used the associated target genes as input for gene ontology over representation analysis (ORA), to assess functional differences in the eGRN. This revealed fundamental differences in the underlying biological processes (Figure 5C; Figure 5—figure supplement 2A). The AML sample with inferred-prenatal origin was enriched for many processes associated with development. In contrast, inferred-postnatal samples appeared more metabolism focused (Figure 5C; Figure 5—figure supplement



**Figure 4.** Validation of age-associated regulons in large bulk RNA-seq cohorts. **(A)** Using previously defined age-associated regulons, pySCENIC AUC scores (Z-score normalized) were clustered to identify bulk RNA-seq samples (n=83) most enriched for inferred-prenatal and inferred-postnatal origin signatures. **(B)** Volcano plot of differentially expressed genes when comparing the inferred-prenatal origin (n=31) and inferred-postnatal origin (n=27) samples. Adjusted p-value threshold 0.01; log2 fold change threshold 0.5. Regulon signature associated transcription factors (TFs) are

Figure 4 continued on next page

Figure 4 continued

indicated. (C) Enrichment plot of significant gene sets enriched in the inferred-prenatal origin samples. GSEA was performed on the DEGs using MSigDB databases. FDR q-value threshold <0.05. (D) Enrichment plot of drug sensitivity gene sets enriched in the inferred-prenatal samples. GSEA was performed on the DEGs, using drug response signatures from published studies of four widely used acute myeloid leukemia (AML) drugs. FDR q-value threshold <0.05. (E) The predicted cell type proportions estimated using AutoGeneS deconvolution, of the inferred-prenatal (n=31) and inferred-postnatal origin samples (n=27) were compared using t-tests. Significant p-values <0.05 (\*), <0.01 (\*\*), <0.001 (\*\*\*), and <0.0001 (\*\*\*\*) are indicated.

The online version of this article includes the following figure supplement(s) for figure 4:

**Figure supplement 1.** Validation of pySCENIC regulons.

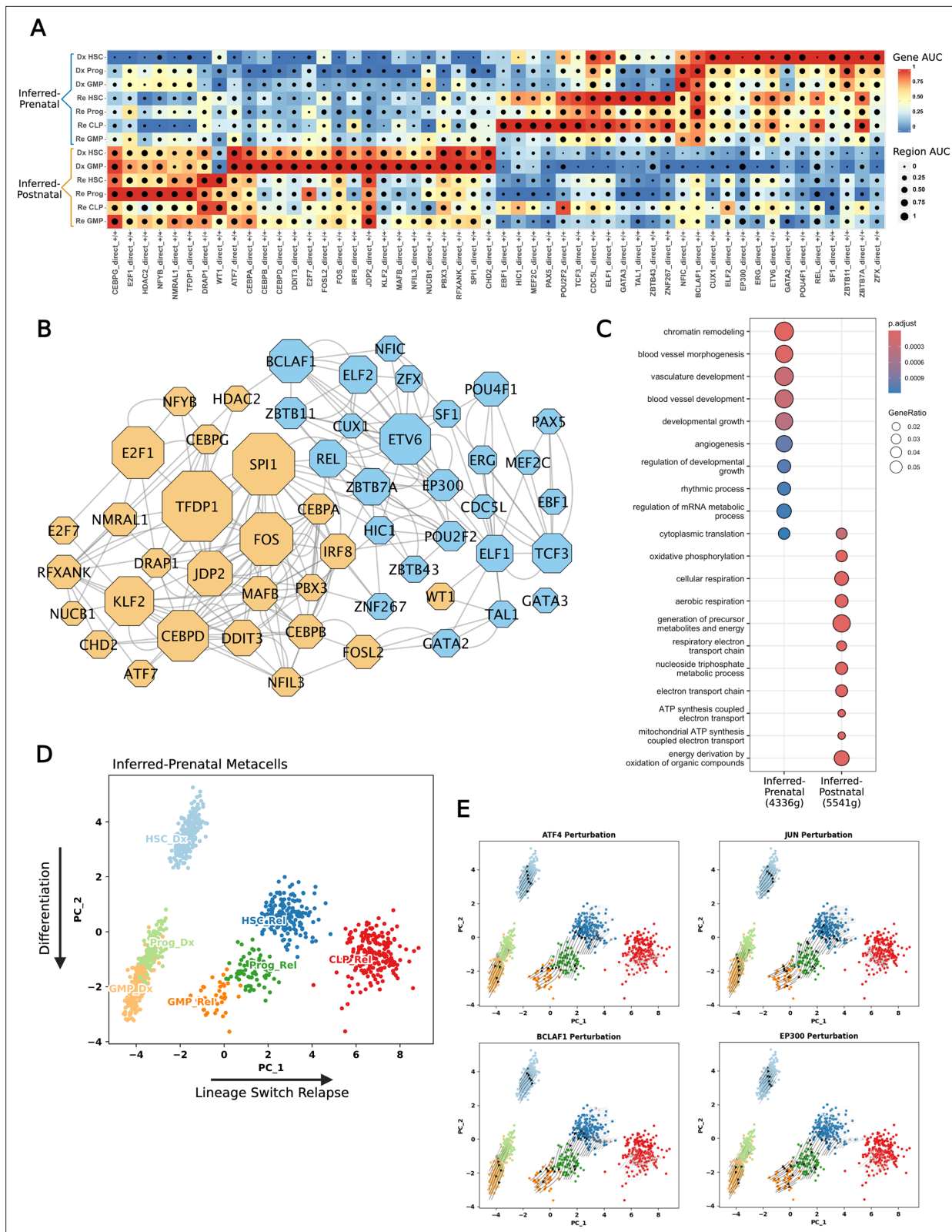
**2A).** This further supports the association of these eRegulons with presumed prenatal origin t(8;21) AML, compared to postnatal origin disease.

Previous analysis using the TARGET (Bolouri et al., 2018) and BeatAML (Burd et al., 2020) datasets indicated that inferred-prenatal and inferred-postnatal origin t(8;21) AML may harbor different levels of chemosensitivity based on published drug response signatures (Figure 4D). Therefore, we performed in silico perturbations of eRegulon-associated TFs. PCA of the diagnosis and relapse samples recapitulated the expected differentiation trajectories along PC2, while separating diagnosis from relapse along PC1 (Figure 5D). Using the SCENIC+ (Bravo González-Blas et al., 2023) perturbation simulation workflow, we identified TFs estimated to induce differentiation, as defined by a negative shift in PC2 (Figure 5—figure supplement 2B). We prioritized TFs predicted to impact the HSC compartment and identified 18 TFs with predicted significant effects on HSC differentiation (Figure 5—figure supplement 2C). Several of these are components of the AP-1 complex (JUN, ATF4, FOSL2), which are established downstream targets of the t(8;21) fusion protein and are known to propagate t(8;21) AML (Ptasinska et al., 2019; Eferl and Wagner, 2003; Figure 5E; Figure 5—figure supplement 2C).

Using AP-1 complex members as a comparative baseline, we identified EP300 as one of the most impactful hits. EP300 has recently been shown to drive t(8;21) AML self-renewal through acetylation dependent mechanism (Wang et al., 2011). This suggests that presumed prenatal origin pediatric t(8;21) AML may be particularly sensitive to EP300 inhibition. One of the most striking predictions, for both diagnostic and relapse HSC populations, was BCLAF1 (Figure 5E; Figure 5—figure supplement 2C). BCLAF1 is a regulator of normal HSPCs (Crowley et al., 2022a), and its expression level declines during hematopoietic differentiation. While recent studies have identified a role for BCLAF1 in AML (Dell'Aversana et al., 2017), this has not been explored in detail in the context of pediatric AML or t(8;21) AML and may present a therapeutic opportunity.

We also performed SCENIC+ (Bravo González-Blas et al., 2023) perturbation modelling on the postnatal origin sample (AML12). In this case, the PCA was less straightforward to interpret, as branching differentiation trajectories towards a lymphoid or myeloid fate appear along the PC2 axis, while PC1 distinguishes diagnosis and relapse samples (Figure 5—figure supplement 2D). Therefore, we prioritized TFs based on a predicted effect similar to AP-1 complex components, as it is known that this complex is a critical regulator in t(8;21) AML. We identified several TFs from our original postnatal origin signature were predicted to have an effect (Figure 5—figure supplement 2D–F), supporting the relevance of the GRNs identified in our previous analyses.

To further investigate EP300 and BCLAF1, we queried the DepMap (Tsherniak et al., 2017) database to assess the dependency of t(8;21) AML cell lines to these genes (Figure 5—figure supplement 2G). We found that the two widely used cell lines of t(8;21) AML, KASUMI-1 (7-year-old donor) (Asou et al., 1991) and SKNO-1 (22-year-old donor) (Matozaki et al., 1995), were among the most sensitive to these perturbations based on their DepMap effect scores (Figure 5—figure supplement 2G). Several other cell lines sensitive to BCLAF1 were derived from pediatric cancers, most notably neuroblastomas, which also arise in-utero (Körber et al., 2023; Figure 5—figure supplement 2H). Together, these findings suggest that EP300 inhibition may be particularly effective in t(8;21) AML, and that BCLAF1 may present a new therapeutic target for t(8;21) AML, particularly in pediatric cases with inferred prenatal origin of the driver translocation.



**Figure 5.** Combining multiomics data interrogates age-associated regulons. **(A)** SCENIC+ eRegulon dot plot of showing correlation between single-cell RNA sequencing (scRNA-seq) target gene activity (indicated by the color scale) and scATAC-seq target region accessibility (depicted by spot size). Regulon specificity score (RSS) identified the key activating eRegulons (+/+) between inferred-prenatal and inferred-postnatal origin disease and allows comparison of diagnosis (Dx) and relapse (Rel) time points. **(B)** Network showing the inferred-prenatal (blue) and inferred-postnatal (orange) associated *Figure 5 continued on next page*

Figure 5 continued

eRegulons. Node size represents the number of target genes in each regulon. Edges represent interactions between nodes. (C) Over-representation analysis of age-associated eRegulon target genes using Gene Ontology (GO) Biological Processes curated gene sets. Adjusted p-value threshold 0.05. (D) Principal components analysis (PCA) of the gene based eRegulon enrichment scores for the inferred-prenatal origin disease at diagnosis and relapse. PC1 axis explains variance occurring between diagnosis and relapse, where this patient underwent a lineage switch. PC2 captures variance related to hematopoietic differentiation. (E) SCENIC+ perturbation simulation shows the predicted effect of knockout of selected transcription factors (TFs) on the previously computed PCA embedding. Arrows indicate the predicted shift in cell states relative to the initial PCA embedding.

The online version of this article includes the following source data and figure supplement(s) for figure 5:

**Source data 1.** SCENIC+ eRegulons for *Lambo et al., 2023* t(8;21) acute myeloid leukemia (AML) samples.

**Figure supplement 1.** Multiomics data of t(8;21) acute myeloid leukemia (AML) Rrefines gene regulatory network (GRN).

**Figure supplement 2.** SCENIC+ analysis identifies age-associated candidate perturbations in t(8;21) acute myeloid leukemia (AML).

## Discussion

Here, we have generated a new data resource, AML scAtlas, to investigate AML biology across a broad range of subtypes at single-cell resolution. By including 222 samples comprising 748,679 cells of patients with a wide range of clinical characteristics, AML scAtlas overcomes the limitations of many standalone single-cell studies enabling AML subtype-focused analysis with enough data for robust statistical comparisons. This dataset is publicly available (<https://cellxgene.cziscience.com/collections/071b706a-7ea7-47a4-bddf-6457725839fc>) providing the AML research community with a resource to address diverse biological questions and generate new hypotheses.

To further address a clinically relevant question using this data source, we compared differences between pediatric and adult-onset disease based on the potential biological effect of the in-utero origin of pediatric leukemia. Data of our AML scAtlas was used to explore the GRNs in adult and pediatric t(8;21) AML and revealed a strong age-associated GRN signature. This suggests that while pediatric and adult t(8;21) AMLs are propagated by the same driver translocation, they exhibit clear biological differences correlated with age. This may be due to differences in the cell-of-origin, with mouse models showing that t(8;21) AML can arise from a HSC or a more lineage-restricted GMP (*Cabezas-Wallscheid et al., 2013*). As pediatric t(8;21) can arise in-utero, as evidenced by previous studies (*Wiemels et al., 2002*), and adult t(8;21) is acquired postnatally (*Welch et al., 2012; Jaiswal et al., 2014*), we propose that the observed age-related differences in AML with t(8;21) reflect these differences in the developmental origins of the disease. We identified two distinct groups of regulons corresponding to either inferred-prenatal origin and inferred-postnatal origin disease. These regulons constitute the GRN underlying the cellular state, which can be informative when identifying molecular vulnerabilities to target leukemia.

Our cohort is the largest scRNA-seq dataset to explore t(8;21) AML biology to date, however, the number of patients included remains low (n=22), and many of the studies containing the adult samples used CD34 selection in their experimental protocol creating a bias towards HSPCs in these samples. To overcome some of these limitations, we used bulk RNA-seq samples from the TARGET (*Bolouri et al., 2018*) and BeatAML (*Burd et al., 2020*) studies with t(8;21) AML (n=83, **Source data 2**) to validate our regulon signatures. This identifies two clusters of samples which closely match these signatures, showing that the regulon patterns identified from our AML scAtlas are recapitulated with bulk RNA-seq data enabling exploration of larger patient cohorts. Comparisons between inferred-prenatal and inferred-postnatal origin transcriptomes prioritized TFs which were differentially expressed and highlighted differences in underlying biology and drug response. We identified five signature TFs (KDM5A, REST, BCLAF1, YY1, RAD21) for inferred-prenatal origin disease, several of which have roles in embryonic stem cells (*Singh et al., 2008; Dahl et al., 2016*), and critical functions in the maintenance of HSCs (*Lu et al., 2018; Fisher et al., 2017; Kumar et al., 2023*). In contrast, TFs identified in inferred-postnatal origin samples, such as interferon regulatory factors (IRFs), HDAC2, and SPI1, reflect inflammatory and immune processes, many of which have been implicated in leukemia (*Ning et al., 2011; Fischer et al., 2019; Fang et al., 2018*). We also found that inferred-prenatal origin samples had a higher proportion of HSC/Prog cell types compared to inferred-postnatal origin samples, a more primitive state than postnatal onset t(8;21) AML cases, supporting the hypothesis that age-associated differences in the cell-of-origin influence disease biology.

Given these biological differences, we used bulk RNA-seq to predict chemosensitivity using published drug response signatures (*Unnikrishnan et al., 2017; Williams et al., 2020; Xu et al., 2019; Zhang et al., 2020*). Inferred-prenatal samples were enriched for genes indicative of cytarabine sensitivity and depleted of genes suggestive of daunorubicin and venetoclax resistance. These findings suggest that the developmental origins of the disease may influence drug responses, with potential implications in the design of novel therapeutic strategies and providing further biological evidence that pediatric AML might benefit from different clinical management compared with adult-onset AML. Importantly, venetoclax is currently in the AML23 trial (NCT05955261); our results support further evaluation of venetoclax treatment in pediatric t(8;21) AML.

Using an additional single-cell multiomic dataset, using SCENIC+, we reconstructed the eGRN in samples matching our inferred-prenatal and inferred-postnatal regulon signatures. Upon comparing eRegulons for each patient at diagnosis and relapse, we identified clusters of highly correlated eRegulons defined by different biological processes. Inferred-prenatal origin samples are characterized by developmental and transcriptional dysregulation, whereas inferred-postnatal origin samples are largely driven by fundamental cellular processes linked to inflammation. We used SCENIC+ to model the predicted impact of TF perturbations on our prenatal origin sample at diagnosis and relapse identified several key components of the AP-1 complex, which are critical in t(8;21) AML biology and are also associated with dynamic age-related transcriptional changes (*Ptasinska et al., 2019; Martinez-Soria et al., 2018; Patrick et al., 2024*).

Through our analysis, we identified EP300 as a candidate target, which has been shown to be critical for t(8;21) AML biology (*Wang et al., 2011*) with demonstrable effects in KASUMI-1 and SKNO1 cell lines. EP300 has been identified as a promising therapeutic target in AML with several molecules in development (*Nicosia et al., 2023*) our data indicate potential specific therapeutic benefit in prenatal origin t(8;21) AML. One of the most impactful perturbation predictions for the HSC compartment at diagnosis and relapse was BCLAF1. This is consistent with previous evidence of its importance in HSCs (*Crowley et al., 2022a; Crowley et al., 2022b*) and AML (*Dell'Aversana et al., 2017*), but has not been studied specifically in the context of pediatric AML or t(8;21) AML previously. The DepMap data shows that KASUMI-1 is the most sensitive myeloid cell line to BCLAF1 perturbation, and our GRN analyses suggest it is particularly active in pediatric t(8;21) AML of inferred in-utero origin, thus this may represent an additional prognostic indicator.

Further investigations are required to characterize the roles of both EP300 and BCLAF1 in prenatal origin t(8;21) AML before any clinical realization. EP300 has already been investigated as a target in AML, so future work should focus on the pediatric AML setting with in vitro and in vivo studies using EP300/CBP inhibitors such as inobrodib (*Nicosia et al., 2023*). In contrast, BCLAF1 is relatively unexplored, and additional work is required to elucidate its molecular function and assess its potential as a therapeutic target. BCLAF1 may ultimately prove most valuable as a biomarker of in-utero t(8;21) AML, enabling distinction between late-onset in-utero and postnatal disease. This would require molecular validation in a large cohort of pediatric patients with Guthrie spots to confirm whether they had acquired t(8;21) in-utero.

## Conclusions

Overall, our study demonstrates that large-scale single-cell data integration is a powerful approach to dissect specific patient groups in detail and enables robust comparative analyses. We present the AML scAtlas as a publicly available resource for the research community to address diverse biological questions. By applying AML scAtlas to t(8;21) AML, we identified age-associated gene regulatory networks that likely reflect differences in the developmental origins, biology, and outcome of the disease. These findings also highlight novel candidate therapeutic targets which may be more relevant in pediatric t(8;21) AML compared to adult-onset disease, offering opportunities for more tailored treatment strategies.

## Methods

For the complete analysis code, including the conda environments used for analysis, see GitHub Repo (<https://github.com/jesswhitts/AML-scAtlas>, copy archived at *Whittle, 2026*).

## Data collection

A literature search was performed for published AML scRNA-seq datasets (*Velten et al., 2021; van Galen et al., 2019; Beneyto-Calabuig et al., 2023; Stetson et al., 2021; Zheng et al., 2017; Petti et al., 2019; Jiang et al., 2020; Johnston et al., 2020; Pei et al., 2020; Li et al., 2023; Lasry et al., 2023; Fiskus et al., 2023; Naldini et al., 2023; Mumme et al., 2023; Zhang et al., 2023; Li et al., 2022; Oetjen et al., 2018; Setty et al., 2019; Caron et al., 2020*). Suitable studies were selected based on the data quality (over 1000 counts and 500 genes detected per cell for most of the data). Diagnostic, primary AML samples were selected from each AML study. Where healthy donor samples were present, these were also included, along with an additional 4 studies with healthy bone marrow samples.

## Initial data processing

Each scRNA-seq dataset underwent initial quality control individually using Scanpy (v1.9.3) (*Wolf et al., 2018*) as some studies provided raw data and others provided pre-filtered data. Where raw data was provided, doublets were removed using Scrublet (v0.2.3) (*Wolock et al., 2019*) and cells were filtered using the median absolute deviation as described in this single-cell best practices handbook (*Heumos et al., 2023; Heumos and Schaar, 2023*).

Once filtered, datasets were combined, and quality control was performed using Scanpy (v1.9.3) (*Wolf et al., 2018*). The full dataset had quality thresholds applied (percentage mitochondrial counts <10, read counts >1000, gene counts >500), removing any samples which had fewer than 50 cells remaining after filtering. Genes present in <50 cells were removed. MALAT1 was removed as this was highly abundant in many cells and considered artefactual.

## Batch correction

The presence of batch effects was determined through dimensionality reduction and clustering using Scanpy (v1.9.3) (*Wolf et al., 2018*) and using the kBET algorithm (v0.99.6) (*Büttner et al., 2019*). This was repeated on individual studies, to assess whether there were sample-wise batch effects. Batch correction benchmarking was implemented using Harmony (Scanpy v1.9.3 implementation) (*Korsunsky et al., 2019*), scVI (v1.0.3) (*Lopez et al., 2018*), and scANVI (v1.0.3) (*Xu et al., 2021*) and quantified using scIB (1.1.4) (*Luecken et al., 2022*). Different numbers of highly variable genes were used to select the optimal number for integration. Batch correction was performed using scVI (v1.0.3) (*Lopez et al., 2018*) with the top 2000 highly variable genes, using sample as the model covariate.

## AML scAtlas cell type annotation

The scVI corrected embedding was used to run UMAP and Leiden clustering using Scanpy functions (v1.9.3) (*Wolf et al., 2018*). Cell type annotation was performed using CellTypist (v1.6.0) (*Domínguez Conde et al., 2022*) using the 'Immune\_All\_Low.pkl' model, SingleR (v2.0.0) (*Aran et al., 2019*) using the Novershtern hematopoietic reference (*Novershtern et al., 2011*), and scType (v1.0) (*Ianevski et al., 2022*) with the tissue defined as 'Immune system.' Full automated tool outputs are detailed in *Figure 1—source data 2*; overall, we found that the results given varied significantly between different tools. We postulate that this is, in part, due to differences in the reference profiles used. Thus, we opted to use the best consensus of these different tools for our cluster identity assignments.

## AML scAtlas LSC annotation

HSPC clusters were selected from AML scAtlas, and the scVI corrected embedding was used to recompute UMAP using Scanpy functions (v1.9.3) (*Wolf et al., 2018*). As our previous cell type annotations used generic reference profiles and were not AML-specific, we generated a custom cell type annotation reference to identify LSCs. We created a custom SingleR (v2.0.0) (*Aran et al., 2019*) reference using the *Zeng et al., 2022* revised annotations of the *van Galen et al., 2019* dataset (*Figure 1—source data 2*). This was also correlated with the LSC6 (*Elsayed et al., 2020*) and LSC17 (*Ng et al., 2016*) scores for each cell. To compare LSC abundance between ELN risk groups, chi2\_contingency was implemented from SciPy (v1.12.0).

## AML with t(8;21) analysis

Samples with the t(8;21) translocation were selected from the full AML scAtlas. The UMAP was re-computed, and genes were filtered to remove those detected in fewer than 50 cells for the revised dataset, leaving 24,866 genes remaining. Gene regulatory network analysis was performed using pySCENIC (Aibar et al., 2017; Van de Sande et al., 2020) (v0.12.1) as per the recommended workflow. To facilitate comparisons between age groups, cell types were focused on HSPCs, as many adult samples were originally enriched for CD34. The RSS was calculated for the adult and pediatric samples to select the top 20 differential regulons per age group. Using SciPy hierarchical clustering (v1.12.0), regulons were filtered to identify regulon signature groups used for downstream analysis.

## Bulk RNA-seq analysis

Bulk RNA-Seq data was downloaded for the TARGET (Bolouri et al., 2018) and BeatAML (Burd et al., 2020) cohorts and samples with t(8;21) were selected. Only bone marrow samples taken at diagnosis were used for downstream analyses (Source data 2). Using the previously defined signature regulons, the AUCell algorithm (Aibar et al., 2017) (v0.12.1) was implemented to measure regulon activity. Hierarchical clustering was performed using SciPy v1.12.0 to identify samples most enriched for each age-related signature. Differential gene expression analysis was implemented using edgeR (Robinson et al., 2010) (v3.42.4) and DESeq2 (Love et al., 2014) (v1.40.2), using a log<sub>2</sub> fold change threshold of 0.5 and an adjusted p-value cutoff of 0.01. Candidate differential genes, ranked on log<sub>2</sub> fold change, underwent GSEA with GSEAPy (v0.10.8) using a significance threshold of 0.05. Cell type deconvolution was performed using AutoGeneS (Aliee and Theis, 2021) (v1.0.4) using the recommended workflow. Significance when comparing groups was ascertained using a Student's t-test on the predicted cell type proportion values for each sample.

## Single-cell multi-omics analysis

The Lambo et al., 2023 scRNA-seq and scATAC-seq data from pediatric AML bone marrow samples was downloaded, and the t(8;21) samples selected (Source data 2). Using our previously defined signature regulons, the AUCell algorithm (Aibar et al., 2017) (v0.12.1) was implemented to measure regulon activity. This identified the samples most enriched for each age-related signature as AML16 and AML12.

The SCENIC+ pipeline (Bravo González-Blas et al., 2023) (v1.0a1) was implemented as per the recommended Snakemake workflow for creating pseudo-multiome data. Regulons were filtered by correlation between modalities, using a threshold of 0.2 for non-multiome data. The most robust regulons were prioritized based on the SCENIC+ recommendations (direct +/+) (Bravo González-Blas et al., 2023). To facilitate comparisons, the eRegulon RSS was calculated for each patient and the top 30 eRegulons selected. The correlation between the gene sets underpinning eRegulons was calculated and sample-associated clusters were selected for over-representation analysis with clusterProfiler (Wu et al., 2021) (v4.8.3).

To predict the impact of specific TF perturbations on key cell types, SCENIC+ (Bravo González-Blas et al., 2023) perturbation modelling was implemented using the recommended parameters. TFs were then prioritized on their predicted impact on HSC differentiation and visualized using the PCA embedding. Candidate targets EP300 and BCLAF1 were queried in the DepMap (Tsherniak et al., 2017) databases to infer their potential importance.

## Acknowledgements

JW was funded by MRC DTP award (MR/W007428/1), SM by Blood Cancer UK (15038), and CCLG (2016 09), while GL by Cancer Research UK (C5759/A20971 & C5759/A27412) and MI by MRC (MR/X014088/1). We would like to thank all authors of the public data used in this study for their contributions to scientific community. We also acknowledge useful discussions around this work with Magnus Rattray.

## Additional information

### Funding

Funder	Grant reference number	Author
Medical Research Council	MR/W007428/1	Jessica Whittle
Blood Cancer UK	15038	Stefan Meyer
Children's Cancer and Leukaemia Group	2016 09	Stefan Meyer
Cancer Research UK	C5759/A20971	Georges Lacaud
Cancer Research UK	C5759/A27412	Georges Lacaud
Blood Cancer UK	23002	Georges Lacaud
Medical Research Council	MR/X014088/1	Mudassar Iqbal

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Jessica Whittle, Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review and editing; Stefan Meyer, Conceptualization, Formal analysis, Writing – original draft, Writing – review and editing; Georges Lacaud, Syed Murtuza-Baker, Mudassar Iqbal, Conceptualization, Formal analysis, Supervision, Funding acquisition, Writing – original draft, Project administration, Writing – review and editing

### Author ORCIDs

Jessica Whittle  <https://orcid.org/0000-0002-1139-1671>

Georges Lacaud  <https://orcid.org/0000-0002-5630-2417>

Mudassar Iqbal  <https://orcid.org/0000-0002-5006-4331>

### Peer review material

Reviewer #1 (Public review): <https://doi.org/10.7554/eLife.104978.3.sa1>

Author response <https://doi.org/10.7554/eLife.104978.3.sa2>

## Additional files

### Supplementary files

MDAR checklist

Source data 1. Datasets included in acute myeloid leukemia (AML) single-cell Atlas (scAtlas).

Source data 2. External validation datasets used in this study.

### Data availability

The AML scAtlas is hosted online for public use (<https://cellxgene.cziscience.com/collections/071b706a-7ea7-47a4-bddf-6457725839fc>). The processed AnnData object is also available to download from figshare (<https://doi.org/10.48420/27269946.v2>). Details of all data used in this study can be found in **Source data 1**, along with associated links to the original data. All code used to perform the analyses presented here can be accessed in the GitHub repository: (<https://github.com/jesswhitts/AML-scAtlas>, copy archived at **Whittle, 2026**). The samples used for validation analyses are publicly available and are detailed in **Source data 2**. The SCENIC+ eGRN files are provided as **Figure 5—source data 1**.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Whittle J, Meyer S, Lacaud G, Baker SM, Iqbal M	2025	A Single-Cell Transcriptomic Atlas of Acute Myeloid Leukemia	<a href="https://doi.org/10.48420/27269946">https://doi.org/10.48420/27269946</a>	figshare, 10.48420/27269946

## References

- Aibar S**, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nature Methods* **14**:1083–1086. DOI: <https://doi.org/10.1038/nmeth.4463>, PMID: 28991892
- Aliee H**, Theis FJ. 2021. AutoGeneS: Automatic gene selection using multi-objective optimization for RNA-seq deconvolution. *Cell Systems* **12**:706–715. DOI: <https://doi.org/10.1016/j.cels.2021.05.006>, PMID: 34293324
- Aran D**, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology* **20**:163–172. DOI: <https://doi.org/10.1038/s41590-018-0276-y>, PMID: 30643263
- Asou H**, Tashiro S, Hamamoto K, Otsuji A, Kita K, Kamada N. 1991. Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood* **77**:2031–2036 PMID: 2018839.
- Bal gobind BV**, Van den Heuvel-Eibrink MM, De Menezes RX, Reinhardt D, Hollink I, Arentsen-Peters S, van Wering ER, Kaspers GJL, Cloos J, de Bont E, Cayuela JM, Baruchel A, Meyer C, Marschalek R, Trka J, Stary J, Beverloo HB, Pieters R, Zwaan CM, den Boer ML. 2011. Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. *Haematologica* **96**:221–230. DOI: <https://doi.org/10.3324/haematol.2010.029660>, PMID: 20971820
- Barnett SN**, Cujba AM, Yang L, Maceiras AR, Li S, Kedlian VR, Pett JP, Polanski K, Miranda AMA, Xu C, Cranley J, Kanemaru K, Lee M, Mach L, Perera S, Tudor C, Joseph PD, Pritchard S, Toscano-Rivalta R, Tuong ZK, et al. 2024. An organotypic atlas of human vascular cells. *Nature Medicine* **30**:3468–3481. DOI: <https://doi.org/10.1038/s41591-024-03376-x>, PMID: 39566559
- Beneyto-Calabuig S**, Merbach AK, Kniffka JA, Antes M, Szu-Tu C, Rohde C, Waclawiczek A, Stelmach P, Gräßle S, Pervan P, Janssen M, Landry JJM, Benes V, Jauch A, Brough M, Bauer M, Besenbeck B, Felden J, Bäumer S, Hundemer M, et al. 2023. Clonally resolved single-cell multi-omics identifies routes of cellular differentiation in acute myeloid leukemia. *Cell Stem Cell* **30**:706–721. DOI: <https://doi.org/10.1016/j.stem.2023.04.001>, PMID: 37098346
- Bolouri H**, Farrar JE, Triche T Jr, Ries RE, Lim EL, Alonzo TA, Ma Y, Moore R, Mungall AJ, Marra MA, Zhang J, Ma X, Liu Y, Liu Y, Guidry Auvil JM, Davidsen TM, Gesuwan P, Hermida LC, Salhia B, Capone S, et al. 2018. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nature Medicine* **24**:103–112. DOI: <https://doi.org/10.1038/nm.4439>, PMID: 29227476
- Bravo González-Blas C**, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V, Poovathingal S, Wouters J, Aibar S, Aerts S. 2023. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature Methods* **20**:1355–1367. DOI: <https://doi.org/10.1038/s41592-023-01938-4>, PMID: 37443338
- Burd A**, Levine RL, Ruppert AS, Mims AS, Borate U, Stein EM, Patel P, Baer MR, Stock W, Deininger M, Blum W, Schiller G, Olin R, Litzow M, Foran J, Lin TL, Ball B, Boyiadzis M, Traer E, Odenike O, et al. 2020. Precision medicine treatment in acute myeloid leukemia using prospective genomic profiling: feasibility and preliminary efficacy of the Beat AML Master Trial. *Nature Medicine* **26**:1852–1858. DOI: <https://doi.org/10.1038/s41591-020-1089-8>, PMID: 33106665
- Büttner M**, Miao Z, Wolf FA, Teichmann SA, Theis FJ. 2019. A test metric for assessing single-cell RNA-seq batch correction. *Nature Methods* **16**:43–49. DOI: <https://doi.org/10.1038/s41592-018-0254-1>, PMID: 30573817
- Cabezas-Wallscheid N**, Eichwald V, de Graaf J, Löwer M, Lehr HA, Kreft A, Eshkind L, Hildebrandt A, Abassi Y, Heck R, Dehof AK, Ohngemach S, Sprengel R, Wörtge S, Schmitt S, Lotz J, Meyer C, Kindler T, Zhang DE, Kaina B, et al. 2013. Instruction of haematopoietic lineage choices, evolution of transcriptional landscapes and cancer stem cell hierarchies derived from an AML1-ETO mouse model. *EMBO Molecular Medicine* **5**:1804–1820. DOI: <https://doi.org/10.1002/emmm.201302661>, PMID: 24124051
- Caron M**, St-Onge P, Sonntag T, Wang YC, Richer C, Ragoussis I, Sinnott D, Bourque G. 2020. 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**:8079. DOI: <https://doi.org/10.1038/s41598-020-64929-x>, PMID: 32415257
- Chaudhury S**, O'Connor C, Cañete A, Bittencourt-Silvestre J, Sarrou E, Prendergast Á, Choi J, Johnston P, Wells CA, Gibson B, Keeshan K. 2018. Age-specific biological and molecular profiling distinguishes paediatric from adult acute myeloid leukaemias. *Nature Communications* **9**:5280. DOI: <https://doi.org/10.1038/s41467-018-07584-1>, PMID: 30538250

- Chen C**, Yu W, Tober J, Gao P, He B, Lee K, Trieu T, Blobel GA, Speck NA, Tan K. 2019. Spatial genome reorganization between fetal and adult hematopoietic stem cells. *Cell Reports* **29**:4200–4211. DOI: <https://doi.org/10.1016/j.celrep.2019.11.065>, PMID: 31851943
- Crowley SJ**, Bednarski JJ, Magee JA, Li Y, White LS, Yang W. 2022a. BCLAF1 regulates expression of AP-1 genes and fetal hematopoietic stem cell repopulation activity. *Blood* **140**:2852–2853. DOI: <https://doi.org/10.1182/blood-2022-162867>
- Crowley S**, White LS, Li Y, Yang W, Magee JA, Bednarski JJ. 2022b. *Bclaf1* promotes hematopoietic stem cell repopulating capacity and self-renewal. *The Journal of Immunology* **208**:47.01. DOI: <https://doi.org/10.4049/jimmunol.208.Supp.47.01>
- Dahl JA**, Jung I, Aanes H, Greggains GD, Manaf A, Lerdrup M, Li G, Kuan S, Li B, Lee AY, Preissl S, Jermstad I, Haugen MH, Suganthan R, Björås M, Hansen K, Dalen KT, Fedorcsak P, Ren B, Klungland A. 2016. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**:548–552. DOI: <https://doi.org/10.1038/nature19360>, PMID: 27626377
- Dell’Aversana C**, Giorgio C, D’Amato L, Lania G, Matarese F, Saeed S, Costanzo A, Belsito Petrizzi V, Ingenito C, Martens JHA, Pallavicini I, Minucci S, Carissimo A, Stunnenberg HG, Altucci L. 2017. deregulation in AML tumorigenesis. *Leukemia* **31**:2315–2325. DOI: <https://doi.org/10.1038/leu.2017.64>
- Döhner H**, Wei AH, Appelbaum FR, Craddock C, DiNardo CD, Dombret H, Ebert BL, Fenaux P, Godley LA, Hasslerjian RP, Larson RA, Levine RL, Miyazaki Y, Niederwieser D, Ossenkoppele G, Röllig C, Sierra J, Stein EM, Tallman MS, Tien HF, et al. 2022. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood* **140**:1345–1377. DOI: <https://doi.org/10.1182/blood.2022016867>, PMID: 35797463
- Domínguez Conde C**, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, Howlett SK, Suchanek O, Polanski K, King HW, Mamanova L, Huang N, Szabo PA, Richardson L, Bolt L, Fasouli ES, Mahbubani KT, Prete M, Tuck L, Richoz N, et al. 2022. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**:eabl5197. DOI: <https://doi.org/10.1126/science.abl5197>, PMID: 35549406
- Eferl R**, Wagner EF. 2003. AP-1: a double-edged sword in tumorigenesis. *Nature Reviews. Cancer* **3**:859–868. DOI: <https://doi.org/10.1038/nrc1209>, PMID: 14668816
- Elsayed AH**, Rafiee R, Cao X, Raimondi S, Downing JR, Ribeiro R, Fan Y, Gruber TA, Baker S, Klco J, Rubnitz JE, Pounds S, Lamba JK. 2020. A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid leukemia. *Leukemia* **34**:735–745. DOI: <https://doi.org/10.1038/s41375-019-0604-8>, PMID: 31645648
- Fang WF**, Chen YM, Lin CY, Huang HL, Yeh H, Chang YT, Huang KT, Lin MC. 2018. Histone deacetylase 2 (HDAC2) attenuates lipopolysaccharide (LPS)-induced inflammation by regulating PAI-1 expression. *Journal of Inflammation* **15**:3. DOI: <https://doi.org/10.1186/s12950-018-0179-6>
- Fischer J**, Walter C, Tönges A, Aleth H, Jordão MJC, Leddin M, Gröning V, Erdmann T, Lenz G, Roth J, Vogl T, Prinz M, Dugas M, Jacobsen ID, Rosenbauer F. 2019. Safeguard function of PU.1 shapes the inflammatory epigenome of neutrophils. *Nature Immunology* **20**:546–558. DOI: <https://doi.org/10.1038/s41590-019-0343-z>, PMID: 30911105
- Fisher JB**, Peterson J, Reimer M, Stelloh C, Pulakanti K, Gerbec ZJ, Abel AM, Strouse JM, Strouse C, McNulty M, Malarkannan S, Crispino JD, Milanovich S, Rao S. 2017. The cohesin subunit Rad21 is a negative regulator of hematopoietic self-renewal through epigenetic repression of Hoxa7 and Hoxa9. *Leukemia* **31**:712–719. DOI: <https://doi.org/10.1038/leu.2016.240>, PMID: 27554164
- Fiskus W**, Mill CP, Birdwell C, Davis JA, Das K, Boettcher S, Kadia TM, DiNardo CD, Takahashi K, Loghavi S, Soth MJ, Heffernan T, McGeehan GM, Ruan X, Su X, Vakoc CR, Daver N, Bhalla KN. 2023. Targeting of epigenetic co-dependencies enhances anti-AML efficacy of Menin inhibitor in AML with MLL1-r or mutant NPM1. *Blood Cancer Journal* **13**:53. DOI: <https://doi.org/10.1038/s41408-023-00826-6>, PMID: 37055414
- Hamed AA**, Kunz DJ, El-Hamamy I, Trinh QM, Subedar OD, Richards LM, Foltz W, Bullivant G, Ware M, Vladoiu MC, Zhang J, Raj AM, Pugh TJ, Taylor MD, Teichmann SA, Stein LD, Simons BD, Dirks PB. 2022. A brain precursor atlas reveals the acquisition of developmental-like states in adult cerebral tumours. *Nature Communications* **13**:4178. DOI: <https://doi.org/10.1038/s41467-022-31408-y>, PMID: 35853870
- Heumos L**, Schaar A. 2023. Consortium S-CBP. Single-cell best practices. <https://www.sc-best-practices.org/preamble.html> [Accessed March 29, 2024].
- Heumos L**, Schaar AC, Lance C, Litinetskaya A, Drost F, Zappia L, Lücken MD, Strobl DC, Henao J, Curion F, Schiller HB, Theis FJ, Single-cell Best Practices Consortium. 2023. Best practices for single-cell analysis across modalities. *Nature Reviews. Genetics* **24**:550–572. DOI: <https://doi.org/10.1038/s41576-023-00586-w>, PMID: 37002403
- Holmfeldt P**, Ganuza M, Marathe H, He B, Hall T, Kang G, Moen J, Pardieck J, Saulsberry AC, Cico A, Gaut L, McGoldrick D, Finkelstein D, Tan K, McKinney-Freeman S. 2016. Functional screen identifies regulators of murine hematopoietic stem cell repopulation. *The Journal of Experimental Medicine* **213**:433–449. DOI: <https://doi.org/10.1084/jem.20150806>, PMID: 26880577
- Ianevski A**, Giri AK, Aittokallio T. 2022. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nature Communications* **13**:1246. DOI: <https://doi.org/10.1038/s41467-022-28803-w>, PMID: 35273156
- Jaiswal S**, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen HA, Ladenvall C, Getz G, Correa A, et al. 2014. Age-related clonal hematopoiesis associated with adverse outcomes. *The New England Journal of Medicine* **371**:2488–2498. DOI: <https://doi.org/10.1056/NEJMoa1408617>, PMID: 25426837

- Jiang L**, Li XP, Dai YT, Chen B, Weng XQ, Xiong SM, Zhang M, Huang JY, Chen Z, Chen SJ. 2020. Multidimensional study of the heterogeneity of leukemia cells in t(8;21) acute myelogenous leukemia identifies the subtype with poor outcome. *PNAS* **117**:20117–20126. DOI: <https://doi.org/10.1073/pnas.2003900117>, PMID: 32747558
- Johnston G**, Ramsey HE, Liu Q, Wang J, Stengel KR, Sampathi S, Acharya P, Arrate M, Stubbs MC, Burn T, Savona MR, Hiebert SW. 2020. Nascent transcript and single-cell RNA-seq analysis defines the mechanism of action of the LSD1 inhibitor INCB059872 in myeloid leukemia. *Gene* **752**:144758. DOI: <https://doi.org/10.1016/j.gene.2020.144758>, PMID: 32422235
- Körber V**, Stainczyk SA, Kurilov R, Henrich KO, Hero B, Brors B, Westermann F, Höfer T. 2023. Neuroblastoma arises in early fetal development and its evolutionary duration predicts outcome. *Nature Genetics* **55**:619–630. DOI: <https://doi.org/10.1038/s41588-023-01332-y>, PMID: 36973454
- Korsunsky I**, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P-R, Raychaudhuri S. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**:1289–1296. DOI: <https://doi.org/10.1038/s41592-019-0619-0>, PMID: 31740819
- Kumar P**, Zhang N, Lee J, Cheng H, Kurtz K, Conneely SE, Sasidharan R, Rau RE, Pati D. 2023. Cohesin subunit RAD21 regulates the differentiation and self-renewal of hematopoietic stem and progenitor cells. *Stem Cells* **41**:971–985. DOI: <https://doi.org/10.1093/stmcls/sxad058>, PMID: 37534584
- Lambo S**, Trinh DL, Ries RE, Jin D, Setiadi A, Ng M, Leblanc VG, Loken MR, Brodersen LE, Dai F, Pardo LM, Ma X, Vercauteren SM, Meshinchi S, Marra MA. 2023. A longitudinal single-cell atlas of treatment response in pediatric AML. *Cancer Cell* **41**:2117–2135. DOI: <https://doi.org/10.1016/j.ccell.2023.10.008>, PMID: 37977148
- Lasry A**, Nadorp B, Fornerod M, Nicolet D, Wu H, Walker CJ, Sun Z, Witkowski MT, Tikhonova AN, Guillamot-Ruano M, Cayanan G, Yeaton A, Robbins G, Obeng EA, Tsigos A, Stone RM, Byrd JC, Pounds S, Carroll WL, Gruber TA, et al. 2023. An inflammatory state remodels the immune microenvironment and improves risk stratification in acute myeloid leukemia. *Nature Cancer* **4**:27–42. DOI: <https://doi.org/10.1038/s43018-022-00480-0>, PMID: 36581735
- Li B**, Kowalczyk MS, Slyper M, Jellert G, Tabaka M, Ashenberg O, Waldman J, Dionne D, Abigail K, Hui M, Yang Y, Rozenblatt-Rosen O, Regev A. 2022. A single cell immune cell atlas of human hematopoietic system. <https://explore.data.humancellatlas.org/projects/cc95ff89-2e68-4a08-a234-480eca21ce79> [Accessed February 14, 2025].
- Li K**, Du Y, Cai Y, Liu W, Lv Y, Huang B, Zhang L, Wang Z, Liu P, Sun Q, Li N, Zhu M, Bosco B, Li L, Wu W, Wu L, Li J, Wang Q, Hong M, Qian S. 2023. Single-cell analysis reveals the chemotherapy-induced cellular reprogramming and novel therapeutic targets in relapsed/refractory acute myeloid leukemia. *Leukemia* **37**:308–325. DOI: <https://doi.org/10.1038/s41375-022-01789-6>, PMID: 36543880
- Lopez R**, Regier J, Cole MB, Jordan MI, Yosef N. 2018. Deep generative modeling for single-cell transcriptomics. *Nature Methods* **15**:1053–1058. DOI: <https://doi.org/10.1038/s41592-018-0229-2>
- Love MI**, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550. DOI: <https://doi.org/10.1186/s13059-014-0550-8>, PMID: 25516281
- Lu Z**, Hong CC, Kong G, Assumpção A, Ong IM, Bresnick EH, Zhang J, Pan X. 2018. Polycomb group protein YY1 is an essential regulator of hematopoietic stem cell quiescence. *Cell Reports* **22**:1545–1559. DOI: <https://doi.org/10.1016/j.celrep.2018.01.026>, PMID: 29425509
- Luecken MD**, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, Strobl DC, Zappia L, Dugas M, Colomé-Tatché M, Theis FJ. 2022. Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods* **19**:41–50. DOI: <https://doi.org/10.1038/s41592-021-01336-8>, PMID: 34949812
- Martinez-Soria N**, McKenzie L, Draper J, Ptasinska A, Issa H, Potluri S, Blair HJ, Pickin A, Isa A, Chin PS, Tirtakusuma R, Coleman D, Nakjang S, Assi S, Forster V, Reza M, Law E, Berry P, Mueller D, Osborne C, et al. 2018. The oncogenic transcription factor RUNX1/ETO corrupts cell cycle regulation to drive leukemic transformation. *Cancer Cell* **34**:626–642. DOI: <https://doi.org/10.1016/j.ccell.2018.08.015>, PMID: 30300583
- Matozaki S**, Nakagawa T, Kawaguchi R, Aozaki R, Tsutsumi M, Murayama T, Koizumi T, Nishimura R, Isobe T, Chihara K. 1995. Establishment of a myeloid leukaemic cell line (SKNO-1) from a patient with t(8;21) who acquired monosomy 17 during disease progression. *British Journal of Haematology* **89**:805–811. DOI: <https://doi.org/10.1111/j.1365-2141.1995.tb08418.x>, PMID: 7772516
- Montefiori LE**, Bendig S, Gu Z, Chen X, Pölönen P, Ma X, Murison A, Zeng A, Garcia-Prat L, Dickerson K, Iacobucci I, Abdelhamed S, Hiltenbrand R, Mead PE, Mehr CM, Xu B, Cheng Z, Chang TC, Westover T, Ma J, et al. 2021. Enhancer hijacking drives oncogenic *BCL11B* expression in lineage-ambiguous stem cell leukemia. *Cancer Discovery* **11**:2846–2867. DOI: <https://doi.org/10.1158/2159-8290.CD-21-0145>, PMID: 34103329
- Mumme H**, Thomas BE, Bhasin SS, Krishnan U, Dwivedi B, Perumalla P, Sarkar D, Ulukaya GB, Sabnis HS, Park SI, DeRyckere D, Raikar SS, Pauly M, Summers RJ, Castellino SM, Wechsler DS, Porter CC, Graham DK, Bhasin M. 2023. Single-cell analysis reveals altered tumor microenvironments of relapse- and remission-associated pediatric acute myeloid leukemia. *Nature Communications* **14**:6209. DOI: <https://doi.org/10.1038/s41467-023-41994-0>, PMID: 37798266
- Naldini MM**, Casirati G, Barcella M, Rancoita PMV, Cosentino A, Caserta C, Pavesi F, Zonari E, Desantis G, Gilioli D, Carrabba MG, Vago L, Bernardi M, Di Micco R, Di Serio C, Merelli I, Volpin M, Montini E, Ciceri F, Gentner B. 2023. Longitudinal single-cell profiling of chemotherapy response in acute myeloid leukemia. *Nature Communications* **14**:1285. DOI: <https://doi.org/10.1038/s41467-023-36969-0>, PMID: 36890137
- National Cancer Registration and Analysis Service, Northern Ireland Cancer Registry, Scottish Cancer Registry, Unit WClas**. 2021. Children, teenagers and young adults UK cancer statistics report. WClas. <http://>

- [www.ncin.org.uk/cancer\\_type\\_and\\_topic\\_specific\\_work/cancer\\_type\\_specific\\_work/cancer\\_in\\_children\\_teenagers\\_and\\_young\\_adults/](http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/cancer_in_children_teenagers_and_young_adults/)
- Ng SWK**, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, Arruda A, Popescu A, Gupta V, Schimmer AD, Schuh AC, Yee KW, Bullinger L, Herold T, Görlich D, Büchner T, Hiddemann W, Berdel WE, Wörmann B, Cheok M, et al. 2016. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**:433–437. DOI: <https://doi.org/10.1038/nature20598>, PMID: 27926740
- Nguyen H**, Tran D, Tran B, Pehlivan B, Nguyen T. 2021. A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data. *Briefings in Bioinformatics* **22**:bbaa190. DOI: <https://doi.org/10.1093/bib/bbaa190>, PMID: 34020546
- Nicosia L**, Spencer GJ, Brooks N, Amaral FMR, Basma NJ, Chadwick JA, Revell B, Wingelhofer B, Maiques-Diaz A, Sinclair O, Camera F, Ciceri F, Wiseman DH, Pegg N, West W, Knurowski T, Frese K, Clegg K, Campbell VL, Cavet J, et al. 2023. Therapeutic targeting of EP300/CBP by bromodomain inhibition in hematologic malignancies. *Cancer Cell* **41**:2136–2153. DOI: <https://doi.org/10.1016/j.ccell.2023.11.001>, PMID: 37995682
- Ning S**, Pagano JS, Barber GN. 2011. IRF7: activation, regulation, modification and function. *Genes & Immunity* **12**:399–414. DOI: <https://doi.org/10.1038/gene.2011.21>
- Novershtern N**, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, Habib N, Yosef N, Chang CY, Shay T, Frampton GM, Drake ACB, Leskov I, Nilsson B, Preffer F, Dombkowski D, Evans JW, Liefeld T, Smutko JS, Chen J, et al. 2011. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**:296–309. DOI: <https://doi.org/10.1016/j.cell.2011.01.004>, PMID: 21241896
- Oetjen KA**, Lindblad KE, Goswami M, Gui G, Dagur PK, Lai C, Dillon LW, McCoy JP, Hourigan CS. 2018. Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight* **3**:e124928. DOI: <https://doi.org/10.1172/jci.insight.124928>, PMID: 30518681
- Patrick R**, Naval-Sanchez M, Deshpande N, Huang Y, Zhang J, Chen X, Yang Y, Tiwari K, Esmaili M, Tran M, Mohamed AR, Wang B, Xia D, Ma J, Bayliss J, Wong K, Hun ML, Sun X, Cao B, Cottle DL, et al. 2024. The activity of early-life gene regulatory elements is hijacked in aging through pervasive AP-1-linked chromatin opening. *Cell Metabolism* **36**:1858–1881. DOI: <https://doi.org/10.1016/j.cmet.2024.06.006>, PMID: 38959897
- Pei S**, Pollyea DA, Gustafson A, Stevens BM, Minhajuddin M, Fu R, Riemony KA, Gillen AE, Sheridan RM, Kim J, Costello JC, Amaya ML, Inguva A, Winters A, Ye H, Krug A, Jones CL, Adane B, Khan N, Ponder J, et al. 2020. Monocytic subclones confer resistance to venetoclax-based therapy in patients with acute myeloid leukemia. *Cancer Discovery* **10**:536–551. DOI: <https://doi.org/10.1158/2159-8290.CD-19-0710>
- Petti AA**, Williams SR, Miller CA, Fiddes IT, Srivatsan SN, Chen DY, Fronick CC, Fulton RS, Church DM, Ley TJ. 2019. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. *Nature Communications* **10**:3660. DOI: <https://doi.org/10.1038/s41467-019-11591-1>, PMID: 31413257
- Ptasinska A**, Pickin A, Assi SA, Chin PS, Ames L, Avellino R, Gröschel S, Delwel R, Cockerill PN, Osborne CS, Bonifer C. 2019. RUNX1-ETO depletion in t(8;21) AML leads to C/EBP $\alpha$ - and AP-1-mediated alterations in enhancer-promoter interaction. *Cell Reports* **28**:3022–3031. DOI: <https://doi.org/10.1016/j.celrep.2019.08.040>, PMID: 31533028
- Robinson MD**, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140. DOI: <https://doi.org/10.1093/bioinformatics/btp616>, PMID: 19910308
- Setty M**, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. 2019. Characterization of cell fate probabilities in single-cell data with Palantir. *Nature Biotechnology* **37**:451–460. DOI: <https://doi.org/10.1038/s41587-019-0068-4>, PMID: 30899105
- Singh SK**, Kagalwala MN, Parker-Thornburg J, Adams H, Majumder S. 2008. REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature* **453**:223–227. DOI: <https://doi.org/10.1038/nature06863>, PMID: 18362916
- Stetson LC**, Balasubramanian D, Ribeiro SP, Stefan T, Gupta K, Xu X, Fourati S, Roe A, Jackson Z, Schauer R, Sharma A, Tamilselvan B, Li S, de Lima M, Hwang TH, Balderas R, Sauntharajah Y, Maciejewski J, LaFramboise T, Barnholtz-Sloan JS, et al. 2021. Single cell RNA sequencing of AML initiating cells reveals RNA-based evolution during disease progression. *Leukemia* **35**:2799–2812. DOI: <https://doi.org/10.1038/s41375-021-01338-7>, PMID: 34244611
- Tenen DG**. 2003. Disruption of differentiation in human cancer: AML shows the way. *Nature Reviews. Cancer* **3**:89–101. DOI: <https://doi.org/10.1038/nrc989>, PMID: 12563308
- Tsherniak A**, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, Gill S, Harrington WF, Pantel S, Krill-Burger JM, Meyers RM, Ali L, Goodale A, Lee Y, Jiang G, Hsiao J, Gerath WFJ, Howell S, Merkel E, Ghandi M, et al. 2017. Defining a cancer dependency map. *Cell* **170**:564–576. DOI: <https://doi.org/10.1016/j.cell.2017.06.010>, PMID: 28753430
- Unnikrishnan A**, Papaemmanuil E, Beck D, Deshpande NP, Verma A, Kumari A, Woll PS, Richards LA, Knezevic K, Chandrakanthan V, Thoms JAI, Tursky ML, Huang Y, Ali Z, Olivier J, Galbraith S, Kulasekararaj AG, Tobiasson M, Karimi M, Pellagatti A, et al. 2017. Integrative genomics identifies the molecular basis of resistance to azacitidine therapy in myelodysplastic syndromes. *Cell Reports* **20**:572–585. DOI: <https://doi.org/10.1016/j.celrep.2017.06.067>, PMID: 28723562
- Van de Sande B**, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, Seurinck R, Saelens W, Cannoodt R, Rouchon Q, Verbeiren T, De Maeyer D, Reumers J, Saeys Y, Aerts S. 2020. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature Protocols* **15**:2247–2276. DOI: <https://doi.org/10.1038/s41596-020-0336-2>, PMID: 32561888

- van Galen P**, Hovestadt V, Wadsworth IJ, Hughes TK, Griffin GK, Battaglia S, Verga JA, Stephansky J, Pastika TJ, Lombardi Story J, Pinkus GS, Pozdnyakova O, Galinsky I, Stone RM, Graubert TA, Shalek AK, Aster JC, Lane AA, Bernstein BE. 2019. Single-cell RNA-Seq reveals AML hierarchies relevant to disease progression and immunity. *Cell* **176**:1265–1281. DOI: <https://doi.org/10.1016/j.cell.2019.01.031>, PMID: 30827681
- Velten Lars**, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D, Boch T, Hofmann W-K, Ho AD, Huber W, Trumpp A, Essers MAG, Steinmetz LM. 2017. Human haematopoietic stem cell lineage commitment is a continuous process. *Nature Cell Biology* **19**:271–281. DOI: <https://doi.org/10.1038/ncb3493>, PMID: 28319093
- Velten L**, Story BA, Hernández-Malmierca P, Raffel S, Leonce DR, Milbank J, Paulsen M, Demir A, Szu-Tu C, Frömel R, Lutz C, Nowak D, Jann JC, Pabst C, Boch T, Hofmann WK, Müller-Tidow C, Trumpp A, Haas S, Steinmetz LM. 2021. Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. *Nature Communications* **12**:1366. DOI: <https://doi.org/10.1038/s41467-021-21650-1>, PMID: 33649320
- Wang L**, Gural A, Sun XJ, Zhao X, Perna F, Huang G, Hatlen MA, Vu L, Liu F, Xu H, Asai T, Xu H, Deblasio T, Menendez S, Voza F, Jiang Y, Cole PA, Zhang J, Melnick A, Roeder RG, et al. 2011. The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. *Science* **333**:765–769. DOI: <https://doi.org/10.1126/science.1201662>, PMID: 21764752
- Welch JS**, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J, Kandoth C, Fulton RS, McLellan MD, Dooling DJ, Wallis JW, Chen K, Harris CC, Schmidt HK, Kalicki-veizer JM, Lu C, et al. 2012. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**:264–278. DOI: <https://doi.org/10.1016/j.cell.2012.06.023>, PMID: 22817890
- Whittle J**. 2026. AML-scatlas. swl:1:rev:79ee35bd13a168470ed2d9271217d73e135beb1e. Software Heritage. <https://archive.softwareheritage.org/swl:1:dir:07d2e95fc5cb18e1e4f254da0742cac889e81112;origin=https://github.com/jesswhitts/AML-scAtlas;visit=swl:1:snp:3427ea754164022460f156f97410b3f3bf9e0c7;anchor=swl:1:rev:79ee35bd13a168470ed2d9271217d73e135beb1e>
- Wiemels JL**, Xiao Z, Buffler PA, Maia AT, Ma X, Dicks BM, Smith MT, Zhang L, Feusner J, Wiencke J, Pritchard-Jones K, Kempski H, Greaves M. 2002. In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood* **99**:3801–3805. DOI: <https://doi.org/10.1182/blood.v99.10.3801>, PMID: 11986239
- Wiggers CRM**, Baak ML, Sonneveld E, Nieuwenhuis EES, Bartels M, Creyghton MP. 2019. AML subtype is a major determinant of the association between prognostic gene expression signatures and their clinical significance. *Cell Reports* **28**:2866–2877. DOI: <https://doi.org/10.1016/j.celrep.2019.08.012>, PMID: 31509748
- Williams MS**, Amaral FMR, Simeoni F, Somerville TCP. 2020. A stress-responsive enhancer induces dynamic drug resistance in acute myeloid leukemia. *The Journal of Clinical Investigation* **130**:1217–1232. DOI: <https://doi.org/10.1172/JCI130809>, PMID: 31770110
- Wolf FA**, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**:15. DOI: <https://doi.org/10.1186/s13059-017-1382-0>, PMID: 29409532
- Wolock SL**, Lopez R, Klein AM. 2019. Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Systems* **8**:281–291. DOI: <https://doi.org/10.1016/j.cels.2018.11.005>, PMID: 30954476
- Wu T**, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**:e100141. DOI: <https://doi.org/10.1016/j.xinn.2021.100141>, PMID: 34557778
- Xu H**, Muise ES, Javadi S, Chen L, Cristescu R, Mansueto MS, Follmer N, Cho J, Kerr K, Altura R, Machacek M, Nicholson B, Addona G, Kariv I, Chen H. 2019. Identification of predictive genetic signatures of Cytarabine responsiveness using a 3D acute myeloid leukaemia model. *Journal of Cellular and Molecular Medicine* **23**:7063–7077. DOI: <https://doi.org/10.1111/jcmm.14608>, PMID: 31449347
- Xu C**, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. 2021. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Molecular Systems Biology* **17**:e9620. DOI: <https://doi.org/10.15252/msb.20209620>, PMID: 33491336
- Zeng AGX**, Bansal S, Jin L, Mitchell A, Chen WC, Abbas HA, Chan-Seng-Yue M, Voisin V, van Galen P, Tierens A, Cheok M, Preudhomme C, Dombret H, Daver N, Futreal PA, Minden MD, Kennedy JA, Wang JCY, Dick JE. 2022. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. *Nature Medicine* **28**:1212–1223. DOI: <https://doi.org/10.1038/s41591-022-01819-x>, PMID: 35618837
- Zhang H**, Nakauchi Y, Köhnke T, Stafford M, Bottomly D, Thomas R, Wilmot B, McWeeney SK, Majeti R, Tyner JW. 2020. Integrated analysis of patient samples identifies biomarkers for venetoclax efficacy and combination strategies in acute myeloid leukemia. *Nature Cancer* **1**:826–839. DOI: <https://doi.org/10.1038/s43018-020-0103-x>, PMID: 33123685
- Zhang Y**, Jiang S, He F, Tian Y, Hu H, Gao L, Zhang L, Chen A, Hu Y, Fan L, Yang C, Zhou B, Liu D, Zhou Z, Su Y, Qin L, Wang Y, He H, Lu J, Xiao P, et al. 2023. Single-cell transcriptomics reveals multiple chemoresistant properties in leukemic stem and progenitor cells in pediatric AML. *Genome Biology* **24**:199. DOI: <https://doi.org/10.1186/s13059-023-03031-7>, PMID: 37653425
- Zhang B**, He P, Lawrence JEG, Wang S, Tuck E, Williams BA, Roberts K, Kleshchevnikov V, Mamanova L, Bolt L, Polanski K, Li T, Elmentaite R, Fasouli ES, Prete M, He X, Yayon N, Fu Y, Yang H, Liang C, et al. 2024. A human embryonic limb cell atlas resolved in space and time. *Nature* **635**:668–678. DOI: <https://doi.org/10.1038/s41586-023-06806-x>, PMID: 38057666

**Zheng GXY**, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**:14049. DOI: <https://doi.org/10.1038/ncomms14049>, PMID: 28091601