Single-Cell Atlas of AML Reveals Age-Related Gene Regulatory Networks in t(8;21) AML

Jessica Whittle[1,2], Stefan Meyer[3,4,5], Georges Lacaud[2]*, Syed Murtuza Baker[1]*, Mudassar Iqbal[1]*

[1]Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK

[2]Stem Cell Biology Group, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK

[3]Manchester Cancer Research Centre (MCRC), Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, UK

[4]Department of Paediatric and Adolescent Oncology, Royal Manchester Children's Hospital, Manchester, UK

[5]Department of Adolescent Oncology, The Christie NHS Foundation Trust, Manchester, UK

*Corresponding authors

Abstract

Background

Acute myeloid leukemia (AML) is characterized by cellular and genetic heterogeneity, which correlates with clinical course. Although single-cell RNA sequencing (scRNA-seq) reflect this diversity to some extent, the low sample numbers in individual studies limit the analytic potential of comparisons of specific patient groups.

Results

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 44 healthy donors from 20 different studies. This is the largest single-cell data resource for AML to our knowledge, publicly available at https://cellxgene.bmh.manchester.ac.uk/AML/. This AML scAtlas, allowed exploration of the clinical importance of age in t(8;21) AML to an unprecedented level, 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 created a de-noised GRN reflecting the previously defined age-related signatures.

Conclusions

Applying integrated data analysis of the AML scAtlas, we reveal age-dependent gene regulation in t(8;21), perhaps reflecting immature/fetal HSC origin in prenatal origin disease vs postnatal origin. Our analysis revealed that BCLAF1, which is particularly enriched in t(8;21) pediatric AML 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 [1, 2]. 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 [3]. 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 [4-7]. 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 [5-8].

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, geographical locations, and experimental designs, complex batch effects often arise between scRNA-seq datasets which requires a tailored data integration approach [9, 10]. Thus, we benchmarked some widely used batch correction tools [11-13] for our specific data use case.

To begin interrogating the data in the AML scAtlas and demonstrate its utility, given the vast representation of different age groups, we set out to address a developmental question in the biology of AML. The molecular landscape of AML in children is different from adults [2, 14-16]. Correspondingly, the clinical course is different with substantially better outcomes in children. The observed differences in molecular profiles in adult and pediatric AML may, in part, be attributed to the developmental origins of the disease. Acquired chromosomal changes in pediatric leukemia are acquired *in utero*, as leukemia-specific genomic aberrations have been detected in the Gunthrie spots of children who subsequently developed leukemia, sometimes several years later in life [17]. Adult leukemia, in contrast, is believed to develop later in life through acquisition of pre-leukaemic changes and clonal evolution of adult HSPCs [18, 19]. The impact of these developmental stages on leukemia biology are not fully understood, and there is currently no means of quantifying and characterizing differences in the origin of leukemia.

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 [2]. The prenatal origins of the t(8;21) rearrangement, has been confirmed even in older children presenting with AML [17]. The prognosis of AML with t(8;21) is better in children than in teenagers and even more so than in young adults [20]. This outcome difference is not fully explainable by co-morbidities and may 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, which may have prognostic implications and could help to explain disease biology and clinical course.

We used our AML scAtlas data resource to characterize age and developmental stage specific signatures in t(8;21) AML by applying single-cell gene regulatory network (GRN) inference methods [21, 22], as a means of revealing cell state heterogeneity across age groups. We then validate and refine our findings in a larger cohort using bulk RNA sequencing (RNA-seq) data from the TARGET [2] and BeatAML [23] studies, defining age-associated GRN signatures and key regulators of t(8;21) AML, which may reflect the developmental origins of the leukemia.

Profiling together both gene expression and chromatin accessibility can be used to decipher the enhancer-driven GRN (eGRN) and enriched transcriptional regulators. Significant heterogeneity across different patients and time points [24] was recently described by analyzing combined scRNA-seq and single-cell Assay for Transposase Accessible Chromatin sequencing (scATAC-seq). This also identified longitudinal transcriptional changes across key clinical milestones, revealing significant heterogeneity across different patients and timepoints. We used the t(8;21) AML data from this study to apply cutting edge GRN inference methods [25], which encompasses both modalities to provide a cleaner GRN which we could correlate with our age-related signatures.

Results

Large Scale Data Integration to Construct a Single-Cell Transcriptomic Atlas of AML (AML scAtlas)

To create AML scAtlas, we combined published scRNA-seq data of primary AML bone marrow samples, from 16 suitable high-quality studies (**see Materials and Methods**), which included 159 AML samples (**Supplementary Table 1**). Where on-treatment time points were available, we selected only diagnostic samples to form a reference atlas of primary AML at diagnosis. If

studies had healthy donor bone marrow samples these were included, in addition to data from healthy bone marrow samples from four scRNA-seq studies (**Supplementary Table 1**) to enable comparisons with healthy bone marrow populations. After cell filtering and quality control the AML scAtlas contains data from 748,679 high quality cells from a total of 20 different scRNA-seq studies [4-6, 8, 26-40] (**Supplementary Figure 1A**). Each sample was allocated to an AML clinical subtype, based on the recent European Leukemia Net (ELN) clinical guidelines [41], and a corresponding prognostic risk group. This resource includes a range of molecular subtypes of AML and spans different age groups, including both pediatric and adult AML cases (**Figure 1A; 1B**). Overall, this dataset is the largest data resource to date for exploring AML biology at single-cell resolution.

From initial analysis of the combined dataset, batch effects were noted, as study-specific clusters were formed. Even within samples of the same study, sample-wise clustering was noted (**Supplementary Figure 1A-1C**). To address this problem, we performed a benchmark of some widely used batch correction methods (**Supplementary Table 2**; **Supplementary Figure 2A;2C**). This identified scVI as the best method for this dataset (**Supplementary Table 2**), which we employed to correct for batch effects. Next, we performed clustering and cell type annotation of AML scAtlas, based on manually curating automated cell type annotation tools results and cluster-wise marker gene expression (**Figure 1C;1D**). Cell type proportions analyses across the clinically relevant subtypes in the dataset (**Supplementary Figure 3A**) show that the AML subtypes were significantly biased towards myeloid cell types (CMP, MEP, GMP, ProMono, CD14+ Mono, CD16+ Mono, cDC, Erythroid) and had a clear predominant cell type, in keeping with AML clonal expansion. On the other hand, healthy donor samples had more balanced cell type proportions and lymphoid cells (T, B, NK, ProB, pDC, Plasma) were well represented (**Supplementary Figure 3A**).

Due to the established critical role of HSPCs and LSCs in propagating AML and their importance as therapeutic targets [42], we selected HSPC clusters for further analysis, as defined by CD34 expression (**Figure 1E**; **Supplementary Figure 3B;3C**). To identify LSCs, we used a curated reference profile of leukaemic stem and progenitor cells (LSPCs) [7] (**Figure 1E**) and correlated this with calculated LSC6 [43] and LSC17 [44] scores for each cell (**Supplementary Figure 3D**). We compared the proportions of HSPC/LSPCs across different AML subtypes and risk groups, as defined by the ELN clinical guidelines [41] (**Figure 1F**). This shows that higher-risk subtypes have a higher proportion of LSCs compared to favorable risk disease.

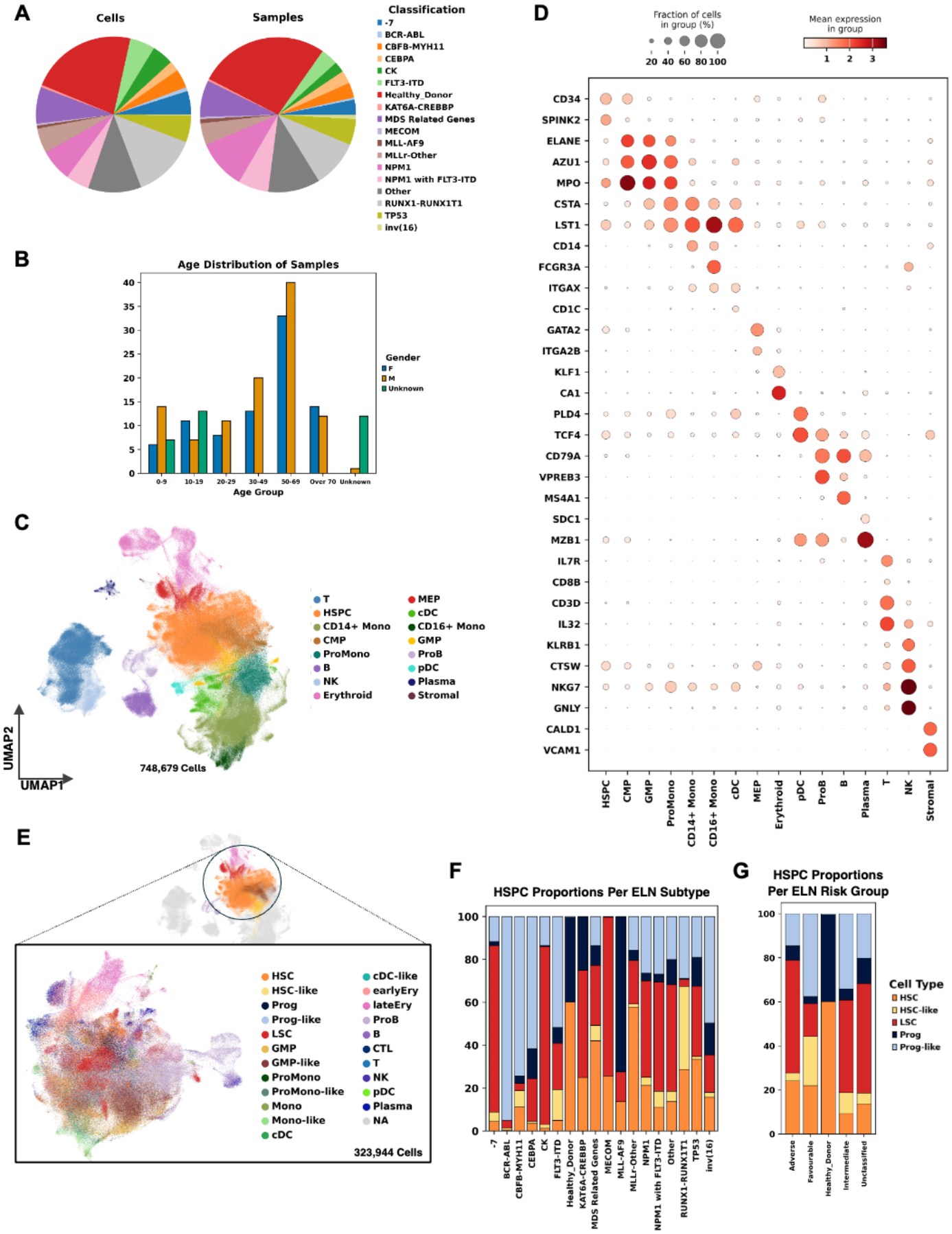
A
Cells
Samples
Classification
-7
BCR-ABL
CBFB-MYH11
CEBPA
CK
FLT3-ITD
Healthy_Donor
KAT6A-CREBBP
MDS Related Genes
MECOM
MLL-AF9
MLLr-Other
NPM1
NPM1 with FLT3-ITD
Other
RUNX1-RUNX1T1
TP53
inv(16)
B
Age Distribution of Samples
Gender
F
M
Unknown
0-9
10-19
20-29
30-49
50-69
Over 70
Unknown
Age Group
C
T
HSPC
CD14+ Mono
CMP
ProMono
B
NK
Erythroid
MEP
cDC
CD16+ Mono
GMP
ProB
pDC
Plasma
Stromal
UMAP2
UMAP1
748,679 Cells
D
Fraction of cells in group (%)
Mean expression in group
CD34
SPINK2
ELANE
AZU1
MPO
CSTA
LST1
CD14
FCGR3A
ITGAX
CD1C
GATA2
ITGA2B
KLF1
CA1
PLD4
TCF4
CD79A
VPREB3
MS4A1
SDC1
MZB1
IL7R
CD8B
CD3D
IL32
KLRB1
CTSW
NKG7
GNLY
CALD1
VCAM1
HSPC
CMP
GMP
ProMono
CD14+ Mono
CD16+ Mono
cDC
MEP
Erythroid
pDC
ProB
B
Plasma
T
NK
Stromal
E
HSC
HSC-like
Prog
Prog-like
LSC
GMP
GMP-like
ProMono
ProMono-like
Mono
Mono-like
cDC
cDC-like
earlyEry
lateEry
ProB
B
CTL
T
NK
pDC
Plasma
NA
323,944 Cells
F
HSPC Proportions Per ELN Subtype
G
HSPC Proportions Per ELN Risk Group
Cell Type
HSC
HSC-like
LSC
Prog
Prog-like
Adverse
Favourable
Healthy_Donor
Intermediate
Unclassified

Figure 1 – Large Scale Data Integration Creates an Atlas of AML

*(**A**) Proportion of cells (left panel) and samples (right panel) belonging to each AML subtype as defined by the ELN clinical guideline. (**B**) Age group and gender distribution of AML scAtlas cohort samples. (**C**) scVI harmonized UMAP colored by annotated cell types. (**D**) 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. (**E**) Selected hematopoietic stem and progenitor cell (HSPC) clusters from the AML scAtlas, annotated with a curated reference of leukemia stem and progenitor cells (LSPCs). (**F**) Proportions of HSPC/LSC populations in each AML subtype as defined by the ELN clinical guideline. (**G**) Proportions of HSPC/LSC populations in each AML risk group as defined by the ELN clinical guideline.*

Application of AML scAtlas to Identifying Age-Associated Gene Regulatory Networks in t(8;21)

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. This included 105,663 cells from 13 adult cases (aged 20-67), 7 adolescent cases (aged 12-17), and 3 pediatric cases (aged 6-8) (**Figure 2A-2C**). Where gender was not available, this was inferred from ChrY/XIST gene expression (**Supplementary Figure 4A**). Several adult samples underwent CD34 selection in original studies, excluding more differentiated cell types in these samples. Thus, these cell types were excluded from comparative analysis, focusing only on HSPCs (**Figure 2B**).

We reconstructed the GRN for the t(8;21) subset of our AML scAtlas using the pySCENIC [22] pipeline and defined critical regulons for this AML subtype. To compare pediatric and adult t(8;21) AML, the top 20 regulons for each of these groups were selected based on the regulon specificity score (RSS) (**Supplementary Figure 4B**). Unsupervised clustering was performed on the Z score normalized regulon activity score matrix, which highlighted clear differences in the GRN across different age groups (**Supplementary Figure 4C**). We presume that the differences in the GRN might reflect differences in the pre- or postnatal developmental origins of the disease.

157

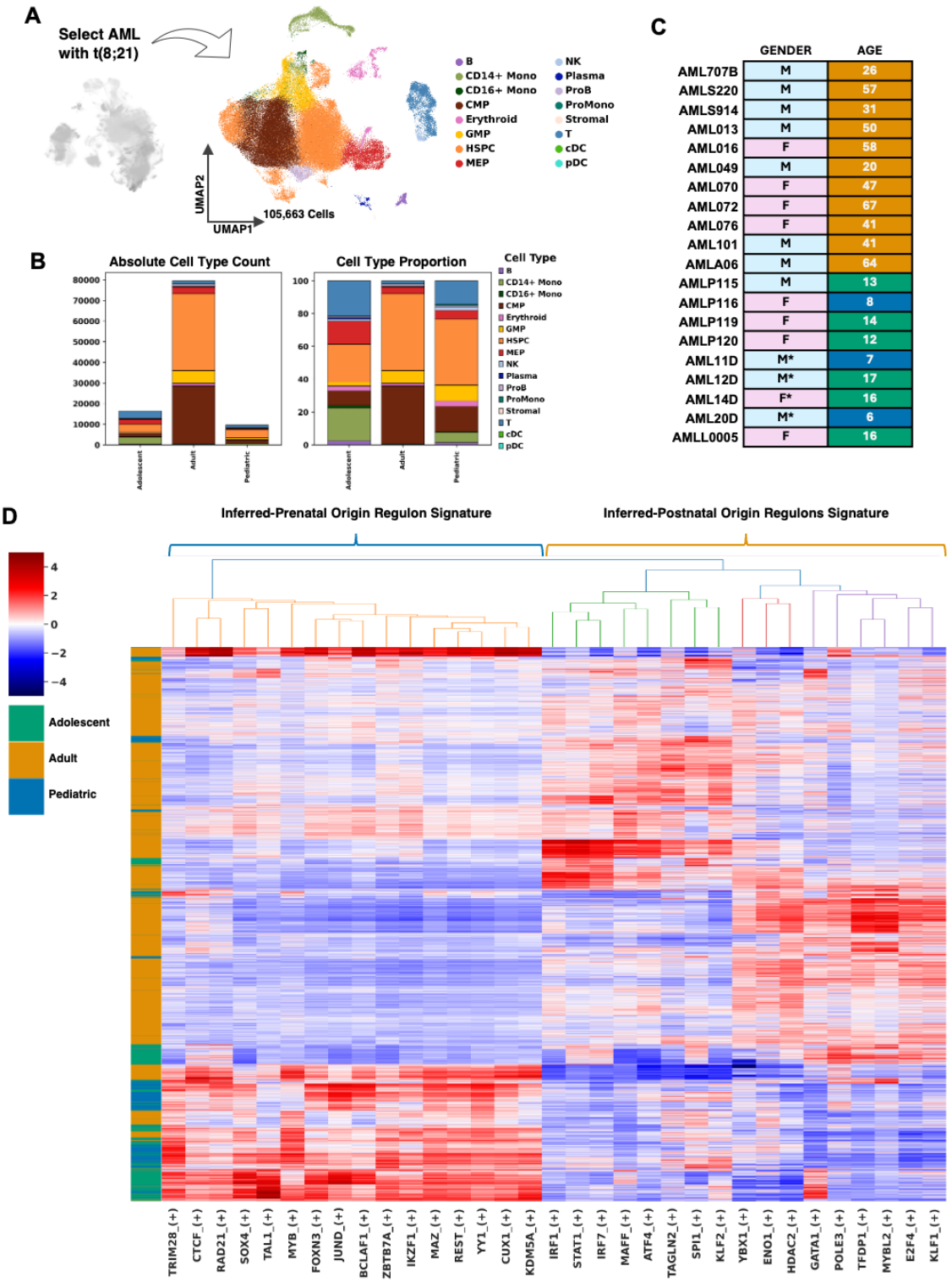

Figure 2 – AML scAtlas Reveals Age-Associated Heterogeneity in t(8;21) AML

*(**A**) Clinical characteristics of the AML scAtlas t(8;21) samples, which were divided into three age groups. Gender is indicated where available. *Sample sex inferred from XIST/ChrY expression. (**B**) Using the AML scAtlas t(8;21) sample cells, UMAP was 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 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).*

To define signature regulons, 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) (**Supplementary Figure 4C**; **Figure 2D**). The pediatric regulon signature is proposed to represent *in-utero* origin t(8;21) AML (henceforth termed 'inferred-prenatal'), and includes 16 regulons defined by a distinct group of hematopoietic transcription factors (TFs) (TRIM28, CTCF, RAD21, SOX4, TAL1, MYB, FOXN3, JUND, BCLAF1, ZBTB7A, IKZF1, MAZ, REST, YY1, CUX1, KDM5A), many of which have clearly defined roles in HSPCs and AML [45-47].The adult regulon signature, presumed representative of the postnatally acquired t(8;21) AML (henceforth termed 'inferred-postnatal), combines 3 discrete clusters of regulons (YBX1, ENO1, and HDAC2; GATA1, POLE3, TFDP1, MYBL2, E2F4, and KLF1; IRF1, STAT1, IRF7, MAFF, AFT4, TAGLN2, SPI1, and KLF2), defined by TFs previously implicated in various hematopoietic, leukemic and inflammatory processes [48-50]. Importantly, both signatures contain key components of the AP-1 complex, which is heavily implicated in the biology of t(8;21) AML [51, 52] and undergoes dynamic changes during aging [53].

Samples of 6 individuals aged 12-17 clustered with the pediatric samples and showed enrichment for the inferred-prenatal signature (**Figure 2D**), which suggests that older adolescents (up to aged 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 perform external validation of our age-associated regulon signatures in a larger cohort of patients. We applied the AUCell algorithm from pySCENIC [22], to the t(8;21) AML bulk RNA-seq from the TARGET [2] and BeatAML [23] cohorts (n=103) to calculate the activity of our pediatric inferred-prenatal and adult inferred-postnatal regulons. Using unsupervised clustering of the bulk RNA-seq AUCell results, we identified discrete clusters of samples that were highly enriched for our inferred-prenatal and inferred-postnatal origin-associated regulons (**Figure 3A**).

Given the limitations of most scRNA-seq platforms in detecting lowly expressed genes, notably TFs, we used these 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, which are well separated by principal components analysis (PCA) (**Supplementary Figure 5A**). We then looked for differentially expressed regulon-associated TFs between the two groups; although changes in TF expression are subtle (**Supplementary**

Figure 5B; **Figure 3B**), 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 3B**). This further delineated the inferred-prenatal signature to 5 key TFs (KDM5A, REST, BCLAF1, YY1, and RAD21), and the inferred-postnatal signature to 9 TFs (ENO1, TFDP1, MYBL2, KLF1, TAGLN2, KLF2, IRF7, SPI1, and YXB1).

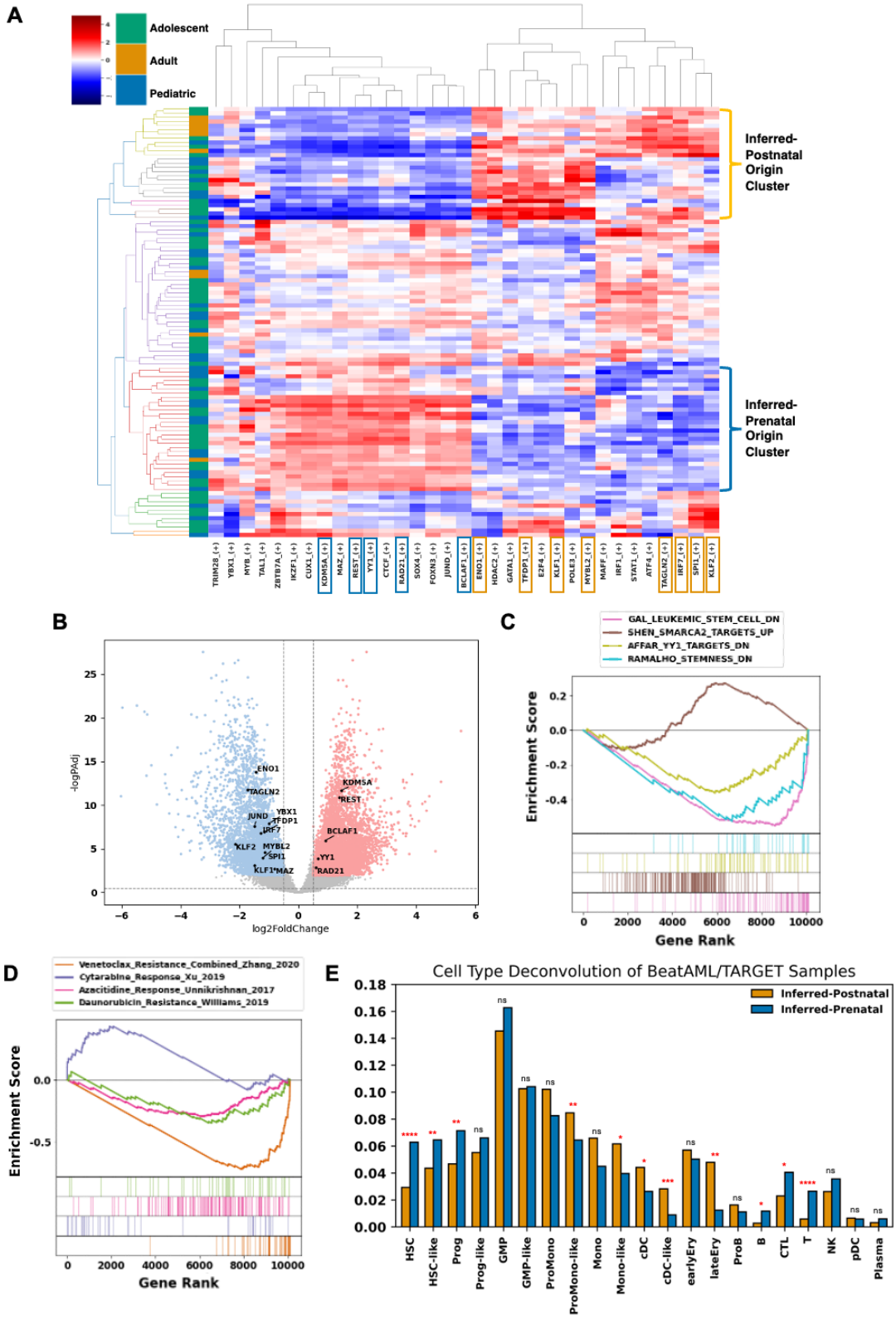

Figure 3 – Validation of Age-Associated Regulons in Large Bulk RNA-Seq Cohorts

*(**A**) Using previously defined age-associated regulons, pySCENIC AUC scores (Z-score normalized) were clustered to identify samples most enriched for inferred-prenatal and inferred-postnatal origin signatures. (**B**) Volcano plot of differentially expressed genes when comparing the inferred-prenatal origin and inferred-postnatal origin samples. Adjusted P value threshold 0.01; log2 fold change threshold 0.5. Regulon signature associated TFs are 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 4 widely used AML drugs. FDR q-value threshold <0.05. (**E**) The predicted cell type proportions estimated using AutoGeneS deconvolution, of the inferred-prenatal and inferred-postnatal origin samples were compared using T-Tests. Significant P values <0.05 (*), <0.01 (**), <0.001 (***) and <0.0001 (****) are indicated.*

We next performed gene set enrichment analysis (GSEA) using all significant differentially expressed genes, to investigate pathways which were enriched in the inferred-prenatal samples compared to the inferred-postnatal ones (**Figure 3C**). Notably, this identified an increase in stemness-associated genes, and in the SMARCA2 target genes, which is a key player in HSC gene expression regulation and chromatin remodeling [54]. SMARCA2 is also known to be upregulated during the fetal-to-adult HSC transition [55], 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 3C**). We also performed GSEA using drug response signatures from published studies [56-59], to deduce whether the inferred-prenatal or inferred-postnatal origin of the leukemia affects chemosensitivity related genes (**Figure 3D**). This analysis revealed that inferred-prenatal origin t(8;21) AML is enriched for genes associated with increased chemosensitivity to cytarabine, venetoclax, and daunorubicin (**Figure 3D**).

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. To compare the cellular heterogeneity between inferred-prenatal and inferred-postnatal samples, we performed cell type deconvolution using AutoGeneS [60], with a curated LSPC reference profile [7], to compare the cellular heterogeneity between prenatal and postnatal origin bulk RNA-seq samples (**Figure 3E**). This revealed a higher proportion of most HSPC cell types (HSC, HSC-like, LSPC-Cycle, LSPC-Primed, LSPC-Quiescent), with a reduction in differentiated myeloid cell types (CMP, MEP, GMP, ProMono, CD14+ Mono, CD16+ Mono, cDC, Erythroid) in the leukemia data with inferred-prenatal origin (**Figure 3E**). To corroborate this finding, we compared cell type proportions in the original t(8;21) subset of AML scAtlas, confirming that cells with the inferred-prenatal

signature comprise more HSC/HSC-like cells than inferred-postnatal signature cells (**Supplementary Figure 5C;5D**). 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 (**Supplementary Figure 5C;5D**).

Multiomics Single-Cell Data Reveals a Denoised GRN and Identifies Candidate Perturbations in Prenatal Origin t(8;21) AML

We used the scRNA-seq and scATAC-seq data from a recent cohort of pediatric t(8;21) AML patients [24] at multiple clinical time points to uncover the eGRN (enhancer-driven GRN) in inferred-prenatal and inferred-postnatal origin t(8;21) AML. Initially, we identified the two most representative samples of our inferred-prenatal and inferred-postnatal signatures by using pySCENIC AUCell [22] 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 (**Supplementary Figure 6A**).

SCENIC+ [25] analysis uses both scRNA-seq and scATAC-seq to identify candidate enhancer regions and TF-binding motifs, which are then used to link TFs to target genes and identified enhancers. This creates enhancer-driven regulons (eRegulons), forming an eGRN. We applied SCENIC+ [25] 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 (**Supplementary Figure 6B**). This revealed several eRegulons across both patients (**Supplementary Figure 6C**), many of which were patient specific, particularly when comparing HSPC populations (**Figure 4A**). The sample of inferred-prenatal origin shows a specific HSC eRegulon profile, whereas the inferred-postnatal sample more closely resembles the corresponding Granulocyte-Monocyte Progenitor (GMP) (**Figure 4A**). 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 4A**).

To identify clusters of closely related eRegulons, we computed the correlations between eRegulons enrichment. We identified 5 main clusters of eRegulons which correspond to different inferred-signature samples (**Figure 4B**). For each eRegulon cluster, we used the associated target genes as input for gene ontology over representation analysis (ORA), to compare functional differences in the eGRN. This revealed fundamental differences in the underlying biological processes (**Figure 4C**; **Supplementary Figure 7A**). The AML sample

with inferred-prenatal origin shows critical transcriptional components as the key molecular processes, such as RNA splicing, histone modification, and TF binding. Whereas AML sample with inferred-postnatal origin appears more immune-mediated, with leucocyte proliferation and leucocyte receptor activation (**Figure 4C**; **Supplementary Figure 7A**). This further supports the stemness signature in presumed prenatal origin t(8;21) AML, compared to postnatal origin disease.

Previous analysis using the TARGET [2] and BeatAML [23] 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 3D**). Therefore, we performed in silico perturbations of eRegulon-associated TFs. PCA of the diagnosis and relapse samples recapitulates expected differentiation trajectories along PC2, while capturing differences between diagnosis and relapse along PC1 (**Figure 4D**). Using the SCENIC+ [25] perturbation simulation workflow to model the predicted effects of TF perturbation on cell state, we identified TFs estimated to induce differentiation, as defined by a negative shift in PC2 (**Supplementary Figure 7B**). We prioritized TFs predicted to induce an effect on the HSC compartment and identified 18 TFs predicted to induce a significant effect on HSC differentiation (**Supplementary Figure 7C**). Several of these TFs are components of the AP-1 complex (JUN, ATF4, FOSL2), which are established targets of the t(8;21) fusion protein and are known to propagate t(8;21) AML [51, 61] (**Supplementary Figure 7C**; **Figure 4E**).

EP300 was one of the most impactful hits, which has recently had a role established in driving t(8;21) AML biology through acetylation permitting self-renewal [62]. 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 (**Supplementary Figure 7C**; **Figure 4E**). BCLAF1 is a regulator of normal HSPCs [63], and its expression level declines during hematopoietic differentiation. While recent studies have identified a role for BCLAF1 in AML [64], 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+ [25] perturbation modelling on the postnatal origin sample (AML12). PCA was more difficult to interpret in this case, as branching differentiation trajectories towards a lymphoid or myeloid fate appear along the PC2 axis, while PC1 distinguishes diagnosis and relapse samples (**Supplementary Figure 7D**). 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 which were identified in our original postnatal origin signature as likely to have an effect (**Supplementary Figures 7D-F**), indicating our previous analyses identified biologically relevant GRNs.

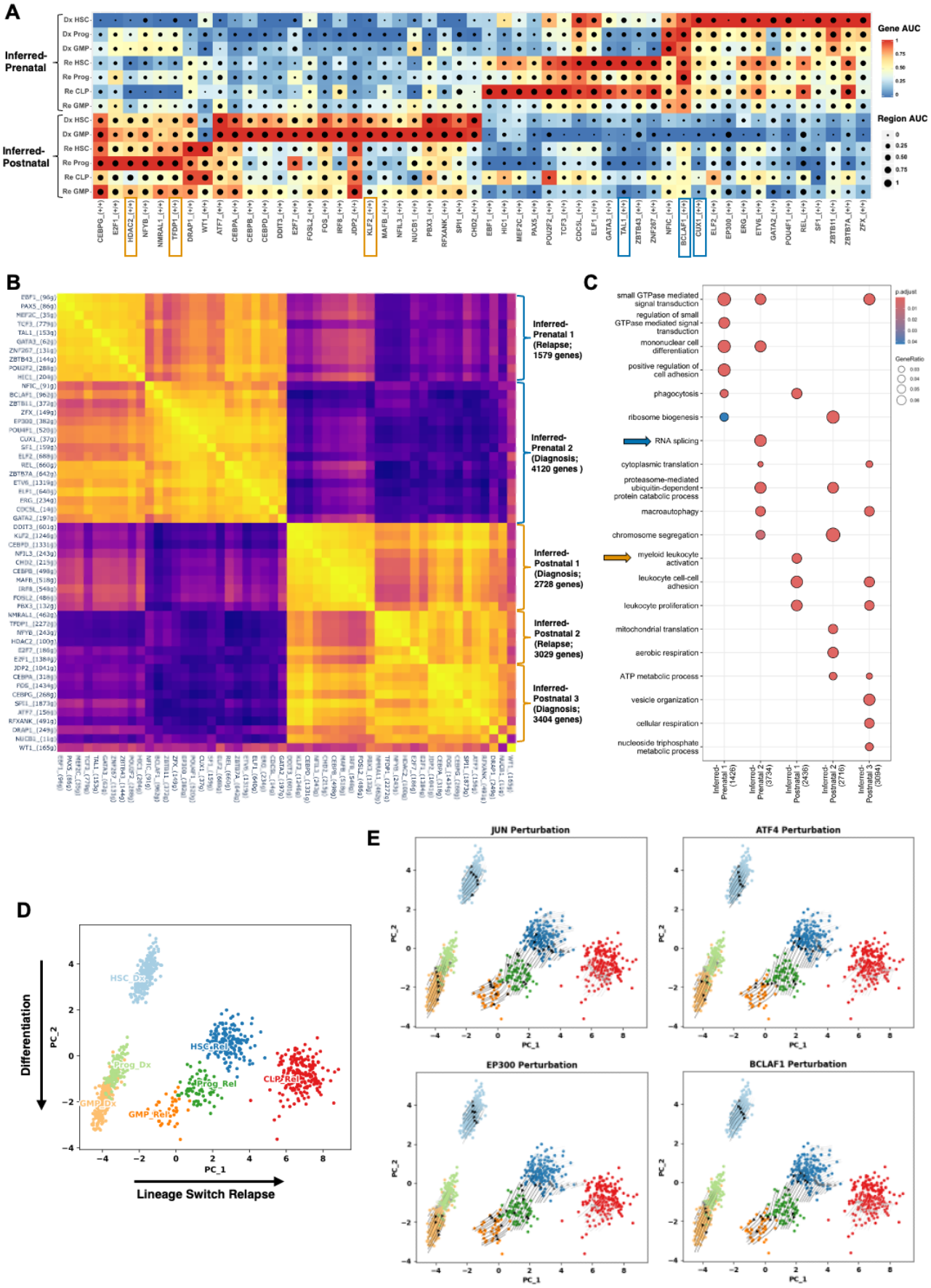
A
Inferred-Prenatal
Inferred-Postnatal
Gene AUC
Region AUC
B
Inferred-Prenatal 1 (Relapse; 1579 genes)
Inferred-Prenatal 2 (Diagnosis; 4120 genes)
Inferred-Postnatal 1 (Diagnosis; 2728 genes)
Inferred-Postnatal 2 (Relapse; 3029 genes)
Inferred-Postnatal 3 (Diagnosis; 3404 genes)
C
small GTPase mediated signal transduction
regulation of small GTPase mediated signal transduction
mononuclear cell differentiation
positive regulation of cell adhesion
phagocytosis
ribosome biogenesis
RNA splicing
cytoplasmic translation
proteasome-mediated ubiquitin-dependent protein catabolic process
macroautophagy
chromosome segregation
myeloid leukocyte activation
leukocyte cell-cell adhesion
leukocyte proliferation
mitochondrial translation
aerobic respiration
ATP metabolic process
vesicle organization
cellular respiration
nucleoside triphosphate metabolic process
Inferred-Prenatal 1 (1426)
Inferred-Prenatal 2 (3734)
Inferred-Postnatal 1 (2436)
Inferred-Postnatal 2 (2716)
Inferred-Postnatal 3 (3094)
D
Differentiation
Lineage Switch Relapse
HSC_Dx
HSC_Rel
Prog_Dx
Prog_Rel
GMP_Dx
GMP_Rel
CLP_Rel
PC_1
PC_2
E
JUN Perturbation
ATF4 Perturbation
EP300 Perturbation
BCLAF1 Perturbation

Figure 4 – Combining Multiomics Data Interrogates Age-Associated Regulons

*(**A**) SCENIC+ eRegulon dotplot of showing correlation between scRNA-seq target gene activity (indicated by the color scale) and scATAC-seq target region accessibility (depicted by spot size). 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**) Correlation plot of the eRegulon target genes. Clusters reflecting inferred-prenatal and inferred-postnatal origin disease at diagnosis (Dx) and relapse (Rel) reflect co-binding of eRegulon TFs. (**C**) Over-representation analysis of age-associated eRegulon target genes using 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 TFs on the previously computed PCA embedding. Arrows indicate the predicted shift in cell states relative to the initial PCA embedding.*

To investigate the impact EP300 and BCLAF1 in more detail, we used the DepMap [65] database to assess the dependency of t(8;21) AML cell lines to these genes (**Supplementary Figure 7G**). We found that the two widely-used cell lines of t(8;21) AML, KASUMI-1 (7-year-old donor) [66] and SKNO-1 (22-year-old donor) [67], were among the most sensitive to these perturbations based on their DepMap effect scores. Together this indicates 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 children with inferred pre-natal origin of the driver translocation.

Discussion

Here we have generated a new data resource, AML scAtlas, to investigate single-cell biology across a range of subtypes of AML. By including 222 samples comprising 748,679 cells of patients with a wide range of clinical characteristics, we overcome the limitations of many standalone single-cell studies by allowing AML subtype-focused analysis with enough data for robust statistical comparisons. The AML scAtlas dataset is publicly available (https://cellxgene.bmh.manchester.ac.uk/AML/) to allow researchers from all areas of AML, to utilize this dataset to answer their own biological questions and generate new hypotheses.

To further address a clinically relevant question using this data source, we compare 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) AML is propagated by the same driver translocation, there are clear biological differences which correlate with age. This may be due to differences in the cell-of-origin, which has been shown to transform from a HSC or

a more lineage restricted GMP in t(8;21) AML mouse models [68]. As pediatric t(8;21) can arise in-utero, as evidenced by previous studies [17], and adult t(8;21) is acquired postnatally [18, 19], we propose that the observed age-related differences in AML with t(8;21) reflect these differences in the developmental origins of the disease. Here, we find two distinct groups of regulons which best reflect inferred-prenatal origin and inferred-postnatal origin disease. 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 [2] and BeatAML [23] studies with t(8;21) AML (n=103) 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 also applicable to bulk RNA-seq data which allows exploration of larger patient cohorts. We compared the transcriptomes of the groups with inferred-prenatal and inferred-postnatal origin, to prioritize TFs which were differentially expressed and to compare differences in underlying biology and drug response. We identified 5 signature TFs (KDM5A, REST, BCLAF1, YY1, RAD21) for inferred-prenatal origin disease, some of which have roles in embryonic stem cells [69, 70], and critical functions in the maintenance of HSCs [45-47]. Whereas 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 [48-50]. We also found that inferred-prenatal origin samples had a higher proportion of HSC/HSC-like cell types compared to inferred-postnatal origin samples, which suggests that prenatal origin t(8;21) AML confers a more primitive state compared to postnatal onset t(8;21) AML cases, supporting age-associated differences in the cell-of-origin playing a role in disease biology.

Given these biological differences, we used bulk RNA-seq to predict chemosensitivity using published drug response signatures [56-59], finding that the inferred-prenatal samples were enriched for genes indicative of cytarabine sensitivity, and depleted of genes suggestive of daunorubicin and venetoclax resistance. Our analysis further suggests that the developmental origins of the disease may be important for drug response signatures, which might inform the design of novel therapeutic strategies and provides 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); this data supports further evaluation of venetoclax treatment in pediatric t(8;21) AML.

Using an additional single-cell multiomic dataset, using SCENIC+ we reconstructed the enhancer-driven regulons (eRegulons) 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 which were defined by different biological processes. This indicated that the samples with inferred-prenatal origin are propagated by transcriptional dysregulation, whereas the samples of inferred-postnatal origin are largely driven by inflammatory signaling. We used SCENIC+ to model the predicted impact of TF perturbations on our prenatal origin sample at diagnosis and relapse. Importantly, we identified several key components of the AP-1 complex, which is critical in t(8;21) AML biology and is also associated with dynamic age-related transcriptional changes [51-53].

Through our analysis, we identified EP300 as a candidate target, which has been shown to be critical for t(8;21) AML biology [62]. This has been shown to cause an effect in KASUMI-1 and SKNO1 cell lines. EP300 has been identified as a promising therapeutic target in AML with several molecules in development [71]; our data indicates 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 [63, 72] and AML [64], 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.

Conclusions

Overall, we demonstrate that large-scale single-cell data integration provides a powerful tool to explore specific patient groups in detail, allowing for robust comparisons. We present the AML scAtlas as a data resource for researchers to answer their own biological questions. Applying our AML scAtlas to t(8;21) AML, we uncover age-associated gene regulatory networks which we propose relates to differences in the developmental origins, biology and oucome of the disease, and might point to novel candidate therapeutic targets which may be more effective in pediatric t(8;21) AML compared to adult-onset disease.

Methods

For the complete analysis code, including the conda environments used for analysis, see GitHub Repo (https://github.com/jesswhitts/AML-scAtlas).

Data Collection

A literature search was performed for published AML scRNA-seq datasets [4-6, 8, 26-40]. 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) [73] 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) [74] and cells were filtered using the median absolute deviation as described in this single-cell best practices handbook [10, 75].

Once filtered, datasets were combined, and quality control was performed using Scanpy (v1.9.3) [73]. 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. MATAL1 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) [73] and using the kBET algorithm (v0.99.6) [76]. 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) [11], scVI (v1.0.3) [12], and scANVI (v1.0.3) [13] and quantified using scIB (1.1.4) [9]. Different numbers of highly variable genes were used to select the optimal number for integration. Batch correction was performed using scVI (v1.0.3) [12] with the top 2000 highly variable genes, using sample as the model covariate.

AML scAtlas Annotation

The scVI corrected embedding was used to run UMAP and Leiden clustering using Scanpy functions (v1.9.3) [73]. Cell type annotation was performed using CellTypist (v1.6.0) [77], SingleR (v2.0.0) [78], and scType (v1.0) [79], taking the consensus of the different tools and the expression of key marker genes to manually curate the cell types. To annotate LSCs, a custom SingleR (v2.0.0) [78] reference was created using the Zeng et al [7] revised annotations of the Van Galen et al [5] dataset. This was also correlated with the LSC6 [43] and LSC17 [44] scores for each cell.

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 [21, 22] (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 [2] and BeatAML [23] cohorts and samples with t(8;21) were selected. Using the previously defined signature regulons, the AUCell algorithm [21] (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 DESeq2 [80] (v1.40.2), using a log2 fold change threshold of 0.5 and an adjusted p value cutoff of 0.01. Candidate differential genes, ranked on log2 fold change, underwent GSEA with GSEApy (v0.10.8) using a significance threshold of 0.05. Cell type deconvolution was performed using AutoGeneS [60] (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 [24] scRNA-seq and scATAC-seq data was downloaded, and the t(8;21) samples selected. Using our previously defined signature regulons, the AUCell algorithm [21] (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 [25] (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 +/+) [25]. 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 [81] (v4.8.3).

To predict the impact of specific TF perturbations on key cell types, SCENIC+ [25] perturbation modelling was implemented using the recommended parameters. TFs were then prioritized on their predicted impact on HSC differentiation and visualised using the PCA embedding. Candidate targets EP300 and BCLAF1 were queried in the DepMap [65] databases to infer their potential importance.

Availability of Data and Materials

The AML scAtlas is hosted online for public use (https://cellxgene.bmh.manchester.ac.uk/AML/). The processed AnnData object is also available to download from figshare (DOI: 10.48420/27269946). Details of all data used in this study can be found in Supplementary Table 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).

Author Contributions

MI, SMB, and GL conceived the study and oversaw the research. SM guided sub-analyses and assisted with interpretation. JW did the data collection and implemented downstream analyses. JW wrote the first draft of the manuscript. All authors interpreted the data and edited the manuscript. All authors approved the final manuscript.

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.

References

1. Tenen DG. Disruption of differentiation in human cancer: AML shows the way. Nature Reviews Cancer. 2003;3(2):89-101.

2. 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, Auvil JMG, Davidsen TM, Gesuwan P, Hermida LC, Salhia B, Capone S, Ramsingh G, Zwaan CM, Noort S, Piccolo SR, Kolb EA, Gamis AS, Smith MA, Gerhard DS, Meshinchi S. The molecular

landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med. 2018;24(1):103-12.

3. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D, Boch T, Hofmann WK, Ho AD, Huber W, Trumpp A, Essers MA, Steinmetz LM. Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol. 2017;19(4):271-81.

4. 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 J-C, Pabst C, Boch T, Hofmann W-K, Müller-Tidow C, Trumpp A, Haas S, Steinmetz LM. Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. Nature Communications. 2021;12(1):1366.

5. van Galen P, Hovestadt V, Wadsworth Ii MH, 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. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell. 2019;176(6):1265-81.e24.

6. 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, Sauer T, Pabst C, Wickenhauser C, Angenendt L, Schliemann C, Trumpp A, Haas S, Scherer M, Raffel S, Müller-Tidow C, Velten L. Clonally resolved single-cell multi-omics identifies routes of cellular differentiation in acute myeloid leukemia. Cell Stem Cell. 2023;30(5):706-21.e8.

7. 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. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. Nature Medicine. 2022;28(6):1212-23.

8. Stetson LC, Balasubramanian D, Ribeiro SP, Stefan T, Gupta K, Xu X, Fourati S, Roe A, Jackson Z, Schauner R, Sharma A, Tamilselvan B, Li S, de Lima M, Hwang TH, Balderas R, Sauntharajah Y, Maciejewski J, LaFramboise T, Barnholtz-Sloan JS, Sekaly RP, Wald DN. Single cell RNA sequencing of AML initiating cells reveals RNA-based evolution during disease progression. Leukemia. 2021;35(10):2799-812.

9. Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, Strobl DC, Zappia L, Dugas M, Colomé-Tatché M, Theis FJ. Benchmarking atlas-level data integration in single-cell genomics. Nature Methods. 2022;19(1):41-50.

10. Heumos L, Schaar AC, Lance C, Litinetskaya A, Drost F, Zappia L, Lücken MD, Strobl DC, Henao J, Curion F, Aliee H, Ansari M, Badia-i-Mompel P, Büttner M, Dann E, Dimitrov D, Dony L, Frishberg A, He D, Hediyeh-zadeh S, Hetzel L, Ibarra IL, Jones MG, Lotfollahi M, Martens LD, Müller CL, Nitzan M, Ostner J, Palla G, Patro R, Piran Z, Ramírez-Suástegui C, Saez-Rodriguez J, Sarkar H, Schubert B, Sikkema L, Srivastava A, Tanevski J, Virshup I, Weiler P, Schiller HB, Theis FJ, Single-cell Best Practices C. Best practices for single-cell analysis across modalities. Nature Reviews Genetics. 2023.

11. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P-r, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. Nature Methods. 2019;16(12):1289-96.

12. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nature Methods. 2018;15(12):1053-8.

13. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Molecular Systems Biology. 2021;17(1):e9620.

14. Balgobind BV, Van den Heuvel-Eibrink MM, De Menezes RX, Reinhardt D, Hollink IH, Arentsen-Peters ST, van Wering ER, Kaspers GJ, Cloos J, de Bont ES, Cayuela JM, Baruchel A, Meyer C, Marschalek R, Trka J,

Stary J, Beverloo HB, Pieters R, Zwaan CM, den Boer ML. Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. Haematologica. 2011;96(2):221-30.

15. Wiggers CRM, Baak ML, Sonneveld E, Nieuwenhuis EES, Bartels M, Creyghton MP. AML Subtype Is a Major Determinant of the Association between Prognostic Gene Expression Signatures and Their Clinical Significance. Cell Reports. 2019;28(11):2866-77.e5.

16. Chaudhury S, O'Connor C, Cañete A, Bittencourt-Silvestre J, Sarrou E, Prendergast Á, Choi J, Johnston P, Wells CA, Gibson B, Keeshan K. Age-specific biological and molecular profiling distinguishes paediatric from adult acute myeloid leukaemias. Nat Commun. 2018;9(1):5280.

17. 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. In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. Blood. 2002;99(10):3801-5.

18. 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, Zhang Q, Lin L, O'Laughlin MD, McMichael JF, Delehaunty KD, Fulton LA, Magrini VJ, McGrath SD, Demeter RT, Vickery TL, Hundal J, Cook LL, Swift GW, Reed JP, Alldredge PA, Wylie TN, Walker JR, Watson MA, Heath SE, Shannon WD, Varghese N, Nagarajan R, Payton JE, Baty JD, Kulkarni S, Klco JM, Tomasson MH, Westervelt P, Walter MJ, Graubert TA, DiPersio JF, Ding L, Mardis ER, Wilson RK. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012;150(2):264-78.

19. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burtt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen HA, Ladenvall C, Getz G, Correa A, Banahan BF, Gabriel S, Kathiresan S, Stringham HM, McCarthy MI, Boehnke M, Tuomilehto J, Haiman C, Groop L, Atzmon G, Wilson JG, Neuberg D, Altshuler D, Ebert BL. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371(26):2488-98.

20. National Cancer Registration and Analysis Service, Northern Ireland Cancer Registry, Scottish Cancer Registry, Unit WCIaS. Children, teenagers and young adults UK cancer statistics report 2021 2021 [Available from: http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/cancer_in_children_teenagers_and_young_adults/.

21. 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. SCENIC: single-cell regulatory network inference and clustering. Nature Methods. 2017;14(11):1083-6.

22. 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. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nature Protocols. 2020;15(7):2247-76.

23. 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, Arellano M, Walker A, Duong VH, Kovacsovics T, Collins R, Shoben AB, Heerema NA, Foster MC, Vergilio J-A, Brennan T, Vietz C, Severson E, Miller M, Rosenberg L, Marcus S, Yocum A, Chen T, Stefanos M, Druker B, Byrd JC. Precision medicine treatment in acute myeloid leukemia using prospective genomic profiling: feasibility and preliminary efficacy of the Beat AML Master Trial. Nature Medicine. 2020;26(12):1852-8.

24. 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. A longitudinal single-cell atlas of treatment response in pediatric AML. Cancer Cell. 2023.

25. Bravo González-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V, Poovathingal S, Wouters J, Aibar S, Aerts S. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nature Methods. 2023;20(9):1355-67.

26. 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, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH. Massively parallel digital transcriptional profiling of single cells. Nature Communications. 2017;8(1):14049.

27. Petti AA, Williams SR, Miller CA, Fiddes IT, Srivatsan SN, Chen DY, Fronick CC, Fulton RS, Church DM, Ley TJ. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. Nat Commun. 2019;10(1):3660.

28. Jiang L, Li XP, Dai YT, Chen B, Weng XQ, Xiong SM, Zhang M, Huang JY, Chen Z, Chen SJ. Multidimensional study of the heterogeneity of leukemia cells in t(8;21) acute myelogenous leukemia identifies the subtype with poor outcome. Proc Natl Acad Sci U S A. 2020;117(33):20117-26.

29. Johnston G, Ramsey HE, Liu Q, Wang J, Stengel KR, Sampathi S, Acharya P, Arrate M, Stubbs MC, Burn T, Savona MR, Hiebert SW. Nascent transcript and single-cell RNA-seq analysis defines the mechanism of action of the LSD1 inhibitor INCB059872 in myeloid leukemia. Gene. 2020;752:144758.

30. Pei S, Pollyea DA, Gustafson A, Stevens BM, Minhajuddin M, Fu R, Riemondy 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, Schowinsky J, Abbott D, Hammes A, Myers JR, Ashton JM, Nemkov T, D'Alessandro A, Gutman JA, Ramsey HE, Savona MR, Smith CA, Jordan CT. Monocytic Subclones Confer Resistance to Venetoclax-Based Therapy in Patients with Acute Myeloid Leukemia. Cancer Discov. 2020;10(4):536-51.

31. 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. Single-cell analysis reveals the chemotherapy-induced cellular reprogramming and novel therapeutic targets in relapsed/refractory acute myeloid leukemia. Leukemia. 2023;37(2):308-25.

32. 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, Tsirigos A, Stone RM, Byrd JC, Pounds S, Carroll WL, Gruber TA, Eisfeld A-K, Aifantis I. An inflammatory state remodels the immune microenvironment and improves risk stratification in acute myeloid leukemia. Nature Cancer. 2023;4(1):27-42.

33. 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. Targeting of epigenetic co-dependencies enhances anti-AML efficacy of Menin inhibitor in AML with MLL1-r or mutant NPM1. Blood Cancer Journal. 2023;13(1):53.

34. 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. Longitudinal single-cell profiling of chemotherapy response in acute myeloid leukemia. Nature Communications. 2023;14(1):1285.

35. 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. Single-cell analysis reveals altered tumor microenvironments of relapse- and remission-associated pediatric acute myeloid leukemia. Nature Communications. 2023;14(1):6209.

36. 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, Hu S, Wang Q-F. Single-cell transcriptomics reveals multiple chemoresistant properties in leukemic stem and progenitor cells in pediatric AML. Genome Biology. 2023;24(1):199.

37. 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. A single cell immune cell atlas of human hematopoietic system Human

Cell Atlas Data Portal: Human Cell Atlas; 2022 [Available from: https://explore.data.humancellatlas.org/projects/cc95ff89-2e68-4a08-a234-480eca21ce79.

38. Oetjen KA, Lindblad KE, Goswami M, Gui G, Dagur PK, Lai C, Dillon LW, McCoy JP, Hourigan CS. Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. JCI Insight. 2018;3(23).

39. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. Characterization of cell fate probabilities in single-cell data with Palantir. Nat Biotechnol. 2019;37(4):451-60.

40. Caron M, St-Onge P, Sontag T, Wang YC, Richer C, Ragoussis I, Sinnett D, Bourque G. 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. 2020;10(1):8079.

41. Döhner H, Wei AH, Appelbaum FR, Craddock C, DiNardo CD, Dombret H, Ebert BL, Fenaux P, Godley LA, Hasserjian RP, Larson RA, Levine RL, Miyazaki Y, Niederwieser D, Ossenkoppele G, Röllig C, Sierra J, Stein EM, Tallman MS, Tien H-F, Wang J, Wierzbowska A, Löwenberg B. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood. 2022;140(12):1345-77.

42. 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, Stengel A, Kimura S, Qu C, Valentine MB, Rashkovan M, Luger S, Litzow MR, Rowe JM, den Boer ML, Wang V, Yin J, Kornblau SM, Hunger SP, Loh ML, Pui CH, Yang W, Crews KR, Roberts KG, Yang JJ, Relling MV, Evans WE, Stock W, Paietta EM, Ferrando AA, Zhang J, Kern W, Haferlach T, Wu G, Dick JE, Klco JM, Haferlach C, Mullighan CG. Enhancer Hijacking Drives Oncogenic BCL11B Expression in Lineage-Ambiguous Stem Cell Leukemia. Cancer Discov. 2021;11(11):2846-67.

43. 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. A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid leukemia. Leukemia. 2020;34(3):735-45.

44. 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, Preudhomme C, Dombret H, Metzeler K, Buske C, Löwenberg B, Valk PJM, Zandstra PW, Minden MD, Dick JE, Wang JCY. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature. 2016;540(7633):433-7.

45. Lu Z, Hong CC, Kong G, Assumpção A, Ong IM, Bresnick EH, Zhang J, Pan X. Polycomb Group Protein YY1 Is an Essential Regulator of Hematopoietic Stem Cell Quiescence. Cell Rep. 2018;22(6):1545-59.

46. 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. The cohesin subunit Rad21 is a negative regulator of hematopoietic self-renewal through epigenetic repression of Hoxa7 and Hoxa9. Leukemia. 2017;31(3):712-9.

47. Kumar P, Zhang N, Lee J, Cheng H, Kurtz K, Conneely SE, Sasidharan R, Rau RE, Pati D. Cohesin Subunit RAD21 Regulates the Differentiation and Self-Renewal of Hematopoietic Stem and Progenitor Cells. Stem Cells. 2023;41(10):971-85.

48. Ning S, Pagano JS, Barber GN. IRF7: activation, regulation, modification and function. Genes & Immunity. 2011;12(6):399-414.

49. 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. Safeguard function of PU.1 shapes the inflammatory epigenome of neutrophils. Nature Immunology. 2019;20(5):546-58.

50. Fang W-F, Chen Y-M, Lin C-Y, Huang H-L, Yeh H, Chang Y-T, Huang K-T, Lin M-C. Histone deacetylase 2 (HDAC2) attenuates lipopolysaccharide (LPS)-induced inflammation by regulating PAI-1 expression. Journal of Inflammation. 2018;15(1):3.

51. Ptasinska A, Pickin A, Assi SA, Chin PS, Ames L, Avellino R, Gröschel S, Delwel R, Cockerill PN, Osborne CS, Bonifer C. RUNX1-ETO Depletion in t(8;21) AML Leads to C/EBPα- and AP-1-Mediated Alterations in Enhancer-Promoter Interaction. Cell Reports. 2019;28(12):3022-31.e7.

52. 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, Elder A, Bomken SN, Pal D, Allan JM, Veal GJ, Cockerill PN, Wichmann C, Vormoor J, Lacaud G, Bonifer C, Heidenreich O. The Oncogenic Transcription Factor RUNX1/ETO Corrupts Cell Cycle Regulation to Drive Leukemic Transformation. Cancer Cell. 2018;34(4):626-42.e8.

53. Patrick R, Naval-Sanchez M, Deshpande N, Huang Y, Zhang J, Chen X, Yang Y, Tiwari K, Esmaeili M, Tran M, Mohamed AR, Wang B, Xia D, Ma J, Bayliss J, Wong K, Hun ML, Sun X, Cao B, Cottle DL, Catterall T, Barzilai-Tutsch H, Troskie RL, Chen Z, Wise AF, Saini S, Soe YM, Kumari S, Sweet MJ, Thomas HE, Smyth IM, Fletcher AL, Knoblich K, Watt MJ, Alhomrani M, Alsanie W, Quinn KM, Merson TD, Chidgey AP, Ricardo SD, Yu D, Jardé T, Cheetham SW, Marcelle C, Nilsson SK, Nguyen Q, White MD, Nefzger CM. The activity of early-life gene regulatory elements is hijacked in aging through pervasive AP-1-linked chromatin opening. Cell Metab. 2024;36(8):1858-81.e23.

54. 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. Functional screen identifies regulators of murine hematopoietic stem cell repopulation. J Exp Med. 2016;213(3):433-49.

55. Chen C, Yu W, Tober J, Gao P, He B, Lee K, Trieu T, Blobel GA, Speck NA, Tan K. Spatial Genome Re-organization between Fetal and Adult Hematopoietic Stem Cells. Cell Rep. 2019;29(12):4200-11.e7.

56. 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, Wilson SR, Lindeman R, Young B, Ramakrishna R, Arthur C, Stark R, Crispin P, Curnow J, Warburton P, Roncolato F, Boultwood J, Lynch K, Jacobsen SEW, Mufti GJ, Hellstrom-Lindberg E, Wilkins MR, MacKenzie KL, Wong JWH, Campbell PJ, Pimanda JE. Integrative Genomics Identifies the Molecular Basis of Resistance to Azacitidine Therapy in Myelodysplastic Syndromes. Cell Reports. 2017;20(3):572-85.

57. Williams MS, Amaral FMR, Simeoni F, Somervaille TCP. A stress-responsive enhancer induces dynamic drug resistance in acute myeloid leukemia. The Journal of Clinical Investigation. 2020;130(3):1217-32.

58. Xu H, Muise ES, Javaid 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. Identification of predictive genetic signatures of Cytarabine responsiveness using a 3D acute myeloid leukaemia model. Journal of Cellular and Molecular Medicine. 2019;23(10):7063-77.

59. Zhang H, Nakauchi Y, Köhnke T, Stafford M, Bottomly D, Thomas R, Wilmot B, McWeeney SK, Majeti R, Tyner JW. Integrated analysis of patient samples identifies biomarkers for venetoclax efficacy and combination strategies in acute myeloid leukemia. Nature Cancer. 2020;1(8):826-39.

60. Aliee H, Theis FJ. AutoGeneS: Automatic gene selection using multi-objective optimization for RNA-seq deconvolution. Cell Syst. 2021;12(7):706-15.e4.

61. Eferl R, Wagner EF. AP-1: a double-edged sword in tumorigenesis. Nature Reviews Cancer. 2003;3(11):859-68.

62. Wang L, Gural A, Sun X-J, 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, Nimer SD. The Leukemogenicity of AML1-ETO Is Dependent on Site-Specific Lysine Acetylation. Science. 2011;333(6043):765-9.

63. Crowley SJ, Bednarski JJ, Magee JA, LI Y, White LS, Yang W. BCLAF1 Regulates Expression of AP-1 Genes and Fetal Hematopoietic Stem Cell Repopulation Activity. Blood. 2022;140(Supplement 1):2852-3.

64. Dell'Aversana C, Giorgio C, D'Amato L, Lania G, Matarese F, Saeed S, Di Costanzo A, Belsito Petrizzi V, Ingenito C, Martens JHA, Pallavicini I, Minucci S, Carissimo A, Stunnenberg HG, Altucci L. miR-194-5p/BCLAF1 deregulation in AML tumorigenesis. Leukemia. 2017;31(11):2315-25.

65. 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, Garraway LA, Root DE, Golub TR, Boehm JS, Hahn WC. Defining a Cancer Dependency Map. Cell. 2017;170(3):564-76.e16.

66. Asou H, Tashiro S, Hamamoto K, Otsuji A, Kita K, Kamada N. Establishment of a Human Acute Myeloid Leukemia Cell Line (Kasumi-1) With 8;21 Chromosome Translocation. Blood. 1991;77(9):2031-6.

67. Matozaki S, Nakagawa T, Kawaguchi R, Aozaki R, Tsutsumi M, Murayama T, Koizumi T, Nishimura R, Isobe T, Chihara K. Establishment of a myeloid leukaemic cell line (SKNO-1) from a patient with t(8;21) who acquired monosomy 17 during disease progression. Br J Haematol. 1995;89(4):805-11.

68. 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, Castle JC, Trumpp A, Sahin U, Bockamp E. Instruction of haematopoietic lineage choices, evolution of transcriptional landscapes and cancer stem cell hierarchies derived from an AML1-ETO mouse model. EMBO Mol Med. 2013;5(12):1804-20.

69. Singh SK, Kagalwala MN, Parker-Thornburg J, Adams H, Majumder S. REST maintains self-renewal and pluripotency of embryonic stem cells. Nature. 2008;453(7192):223-7.

70. 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, Bjørås M, Hansen K, Dalen KT, Fedorcsak P, Ren B, Klungland A. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature. 2016;537(7621):548-52.

71. 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, Copland M, Searle E, Somervaille TCP. Therapeutic targeting of EP300/CBP by bromodomain inhibition in hematologic malignancies. Cancer Cell. 2023;41(12):2136-53.e13.

72. Crowley S, White LS, Li Y, Yang W, Magee JA, Bednarski JJ. Bclaf1 promotes hematopoietic stem cell repopulating capacity and self-renewal. The Journal of Immunology. 2022;208(1_Supplement):47.01-47.01.

73. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biology. 2018;19(1):15.

74. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Systems. 2019;8(4):281-91.e9.

75. Heumos L, Schaar A, Consortium S-CBP. Single-cell best practices 2023. Available from: https://www.sc-best-practices.org/preamble.html#.

76. Büttner M, Miao Z, Wolf FA, Teichmann SA, Theis FJ. A test metric for assessing single-cell RNA-seq batch correction. Nature Methods. 2019;16(1):43-9.

77. 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, Tuong ZK, Campos L, Mousa HS, Needham EJ, Pritchard S, Li T, Elmentaite R, Park J, Rahmani E, Chen D, Menon DK, Bayraktar OA, James LK, Meyer KB, Yosef N, Clatworthy MR, Sims PA, Farber DL, Saeb-Parsy K, Jones JL, Teichmann SA. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science. 2022;376(6594):eabl5197.

78. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163-72.

79. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nature Communications. 2022;13(1):1246.

80. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.

81. 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. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3).