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) reflects this diversity to some extent, the low sample numbers in individual studies limit the analytic potential when comparing 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 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.

Conclusions

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.

31 **Introduction**

32 Acute myeloid leukemia (AML) is an aggressive blood cancer driven by non-random genomic
33 rearrangements in hematopoietic stem/progenitor cells (HSPCs). Recurrent AML-associated
34 genomic aberrations, which often involve transcriptional or epigenetic regulators, give rise to
35 distinct patterns of gene expression strongly associated with clinical course and chemotherapy
36 response [1, 2]. Single-cell RNA sequencing (scRNA-seq) studies have demonstrated that
37 HSPCs acquire lineage priming at an early stage when still phenotypically immature and
38 disperse down an erythromyeloid or lymphomyeloid differentiation trajectory [3]. In the context
39 of AML, diverse clonal hierarchies include the co-existence of normal hematopoietic clones.
40 Leukemic clones can partially recapitulate myeloid differentiation and have been shown to
41 display functional differences even when defined by the same genotype [4-7]. Indeed, analysis
42 of AML using scRNA-seq has revealed key clonal hierarchies, defining subtype-associated
43 cell types, and dynamic changes following therapy, and have been critical in characterizing
44 leukemic stem cells (LSCs), which propagate the disease and drive relapse [5-8].

45 Most AML scRNA-seq studies are limited by small sample numbers and include a mixture of
46 different AML subtypes which may not be directly comparable to one another. Therefore, it is
47 difficult to make biological conclusions with sufficient robustness to be clinically translatable in
48 these individual datasets. To overcome this, we performed large-scale integration of public
49 scRNA-seq datasets to create a single-cell transcriptomic atlas for AML (AML scAtlas). Due
50 to the range of data sources spanning time, locations, and experimental designs, complex
51 batch effects often arise between scRNA-seq datasets which requires a tailored data
52 integration approach [9, 10]. Thus, we benchmarked some widely used batch correction tools [11-
53 13] for our specific data use case.

54 Given the broad representation of age groups in AML scAtlas, we sought to investigate a
55 developmental aspect of AML biology. Pediatric AMLs have substantially better clinical
56 outcomes compared to adult AMLs[14-16]. The molecular landscape of AML differs between
57 children and adults[2, 14-16]; this may, in part, reflect differences in the developmental origins of
58 the disease. Chromosomal changes in pediatric leukemia are acquired *in-utero*, as evidenced
59 by leukemia-specific genomic aberrations detected in the Gunthrie spots of children who later
60 developed leukemia, sometimes several years after birth [17]. Adult leukemia, in contrast, is
61 thought to develop later in life through acquisition of pre-leukaemic changes and clonal
62 evolution of adult HSPCs [18, 19]. The impact of these developmental stages on leukemia biology
63 remains incompletely understood, and no current methods exist to quantify and characterize
64 differences in the origin of the disease. However, as childhood AML with presumed *in-utero*

65 origin has a better outcome, for teenagers and young adults, determination of the pre- or
66 postnatal origin might be important for better treatment stratification and prognostication.

67 AML with t(8;21) (AML-ETO/RUNX1-RUNX1T1) is one of the most frequent AML subtypes in
68 young people, although it affects all ages [2]. The prenatal origins of the t(8;21) rearrangement,
69 has been confirmed even in older children presenting with AML [17]. The prognosis of AML with
70 t(8;21) is better in children than in teenagers and even more so than in young adults [20]. This
71 outcome difference is not fully explainable by co-morbidities and may instead be related to the
72 developmental origins of the disease. In the intermediate teenage group, t(8;21) AML may
73 comprise both late childhood and early adult disease entities, a distinction that could have
74 prognostic implications and could help to explain disease biology and clinical course.

75 We leveraged our AML scAtlas resource to characterize age and developmental stage specific
76 signatures in t(8;21) AML by applying single-cell gene regulatory network (GRN) inference[21,
77 22], as a means of revealing cell state heterogeneity across age groups. We then validated and
78 refined our findings in a larger cohort using bulk RNA sequencing (RNA-seq) data from the
79 TARGET[2] and BeatAML[23] studies, defining age-associated GRN signatures and key
80 regulators of t(8;21) AML, that may reflect the developmental origins of the leukemia.

81 Profiling both gene expression and chromatin accessibility together can decipher the
82 enhancer-driven GRN (eGRN) and enriched transcriptional regulators. Significant
83 heterogeneity across different patients and time points [24] was recently described by analyzing
84 combined scRNA-seq and single-cell Assay for Transposase Accessible Chromatin
85 sequencing (scATAC-seq). We used the t(8;21) AML data from this study to validate our initial
86 findings, by applying cutting edge GRN inference methodology [25]. This encompasses both
87 modalities to provide a denoised eGRN which we could correlate with our age-associated
88 signatures.

89 **Results**

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

92 To create the AML scAtlas, we integrated published scRNA-seq data of primary AML bone
93 marrow samples, from 16 suitable high-quality studies (**see Materials and Methods**),
94 comprising 159 AML samples (**Figure 1A; Supplementary Table 1**). Where on-treatment
95 time points were available, we selected only diagnostic samples to establish a reference atlas
96 of primary AML at diagnosis. If studies had healthy donor bone marrow samples, these were
97 included, alongside data from healthy bone marrow samples from four additional scRNA-seq

studies (**Supplementary Table 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 [4-6, 8, 26-40] (**Supplementary Figure 1A**). Each sample was assigned to an AML clinical subtype, based on the recent European Leukemia Net (ELN) clinical guidelines [41], 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;1C**). 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 (**Supplementary Table 2; Supplementary Figure 1A;1B**). Even within samples of the same study, sample-wise clustering was noted (**Supplementary Figure 1C**). To address this, we benchmarked several widely used batch correction methods (**Supplementary Table 2**; **Supplementary Figure 2A;2C**), identifying scVI as the best method for this dataset (**Supplementary Table 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 (**Supplementary Table 3**), verified using cluster-wise marker gene expression (**Figure 1D;1E**).

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; Supplementary Figure 3A**). In contrast, healthy donor samples had more balanced lineage proportions, with lymphoid cells (T, B, NK, ProB, pDC, Plasma) well represented (**Figure 2A; Supplementary Figure 3A**). Given the established critical role of HSPCs and LSCs in propagating AML, and their importance as therapeutic targets [42], 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) [7] (**Figure 2B;2C**) and correlated this with calculated LSC6 [43] and LSC17[44] 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

129 guidelines [41] (**Figure 2E**). Higher-risk subtypes displayed a higher proportion of LSCs
130 compared to favorable risk disease (**Figure 2E;2F**).

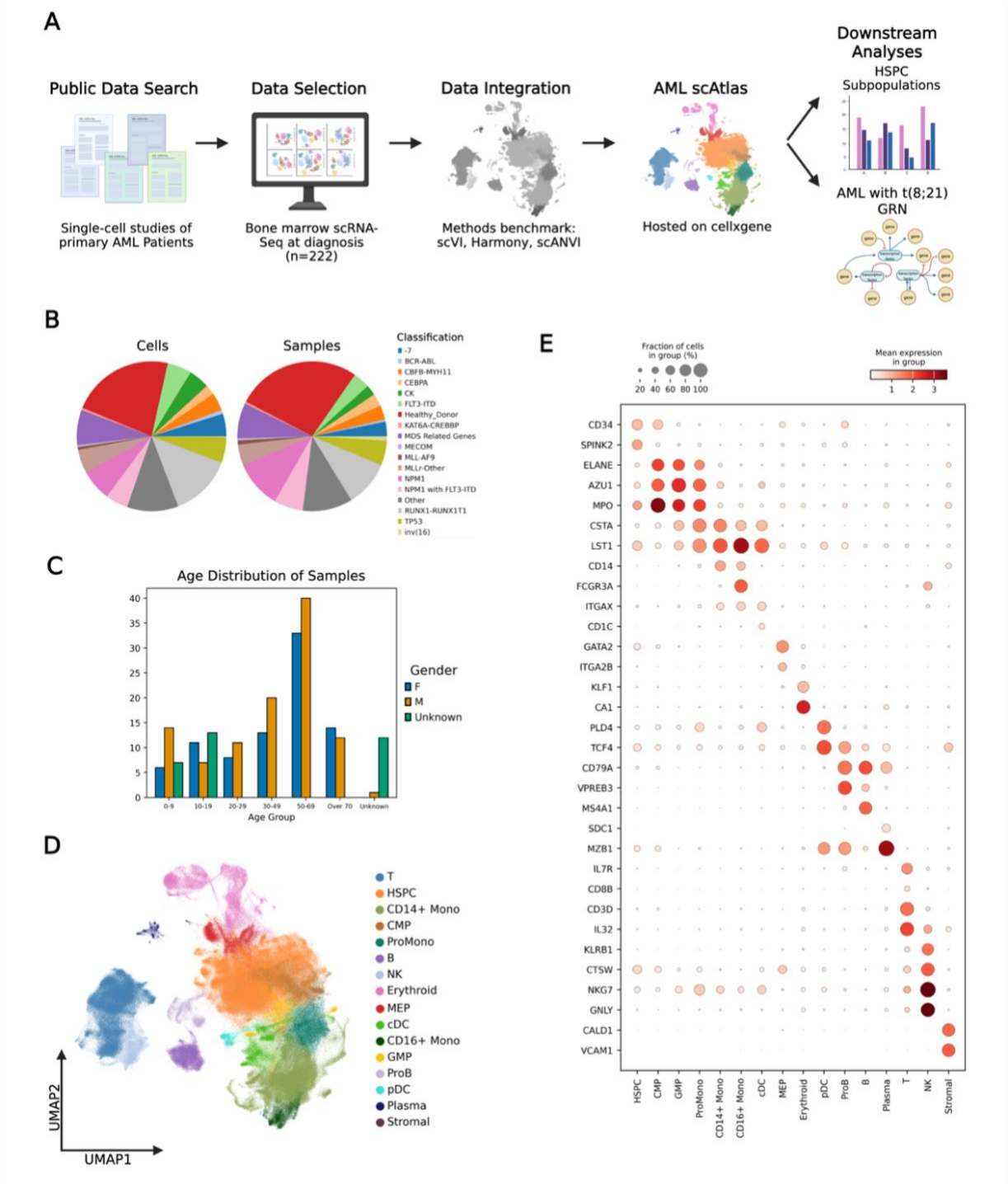


Figure 1 – Large Scale Data Integration Creates a Single-Cell Atlas of AML

*(**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 ELN clinical guideline. (**C**) Age group and gender distribution of AML scAtlas cohort samples. (**D**) scVI harmonized UMAP colored by 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.*

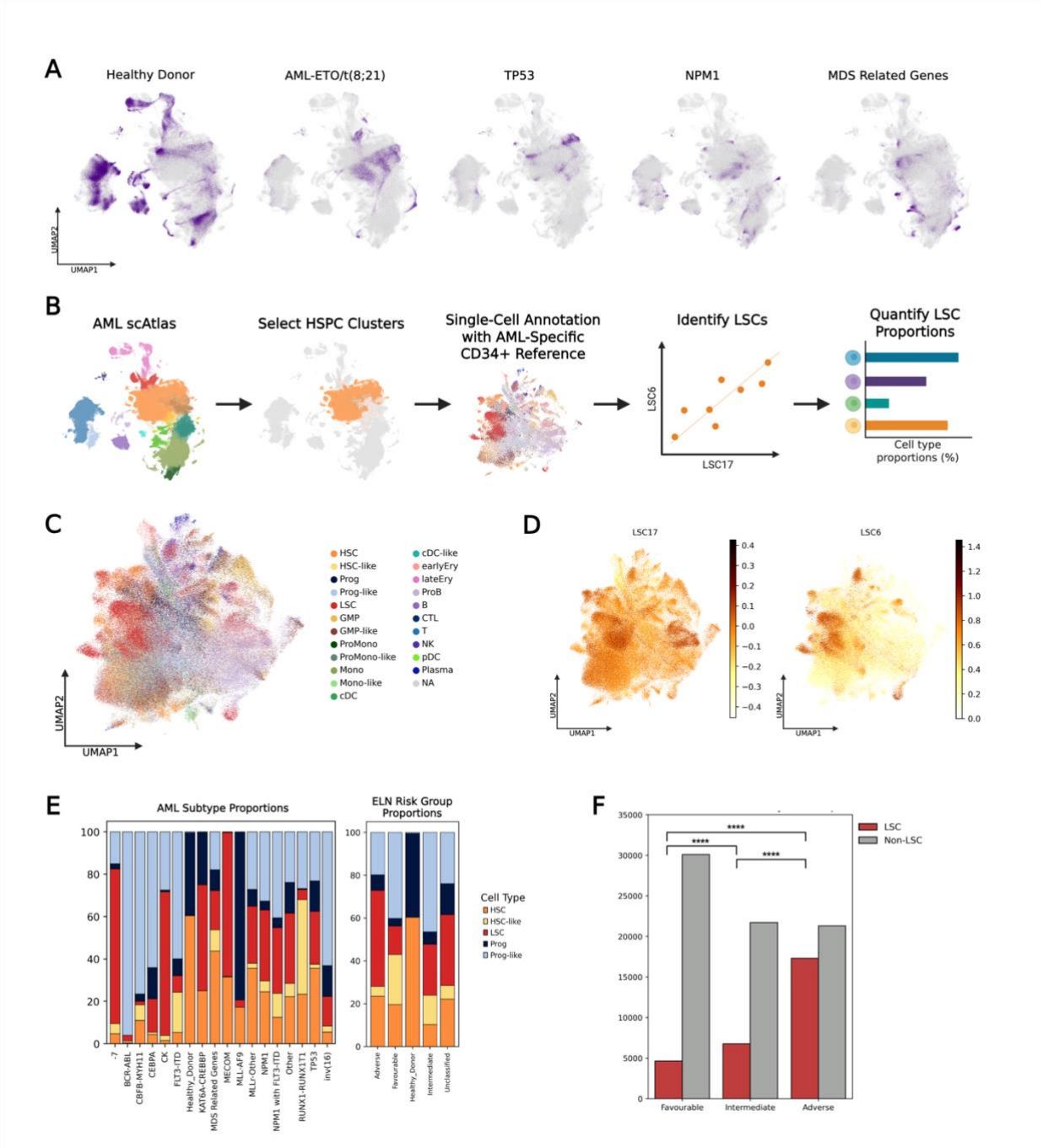


Figure 2 – Characterizing Cell Type Distributions in AML Subtypes

*(**A**) UMAP highlighting the distribution of cells from different AML subtypes in AML 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 (left) and LSC6 (right). (**E**) Proportions of HSPC/LSPC populations in different AML subtypes (left) and AML risk groups (right), as defined by 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.*

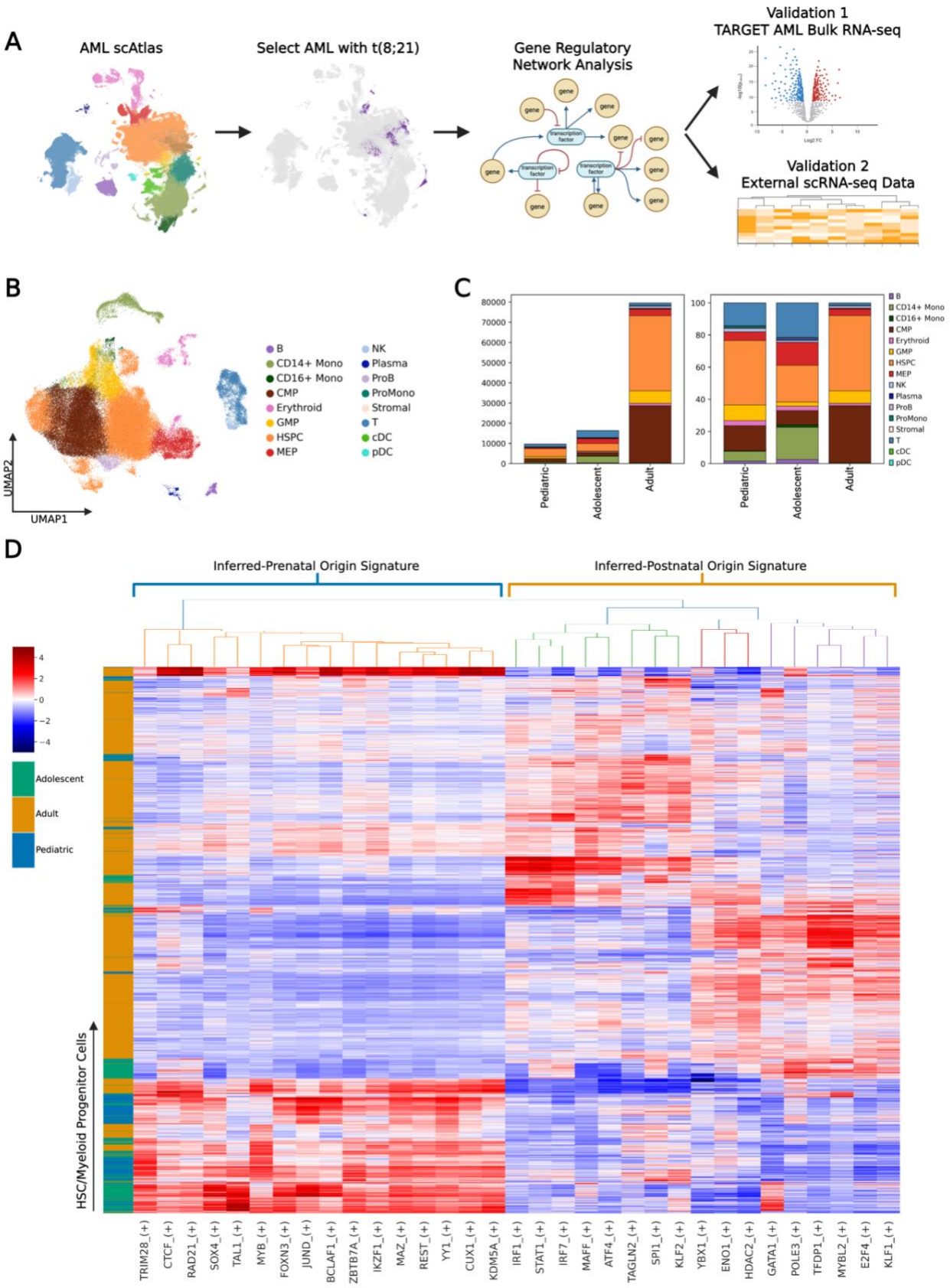
131

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

134 The AML scAtlas enables robust comparison of adult and pediatric AML. We hypothesized
135 that in adolescents and young adults with t(8;21) AML, the potential for either *in-utero* or
136 postnatal HSPC origin disease might affect disease biology and prognosis. Thus, we sought
137 to explore biological differences between pediatric and adult cases of t(8;21) AML, aiming to
138 explain and potentially improve prognostication in adolescents and young adults. We selected
139 samples with t(8;21) AML from the AML scAtlas, resulting in 105,663 cells from 13 adult cases
140 (aged 20-67), 7 adolescent cases (aged 12-17), and 3 pediatric cases (aged 6-8) (**Figure 3A-**
141 **3C**). Where gender information was not available, this was inferred from ChrY/XIST gene
142 expression (**Supplementary Figure 4A**). Several adult samples underwent CD34 selection in
143 original studies, excluding more differentiated cell types (mature lymphoid populations,
144 monocytes, granulocytes) in these samples. Thus, these cell types were excluded from
145 comparative analysis, focusing only on HSPCs and myeloid progenitors (CMP, GMP, MEP),
146 which were well represented in all studies (**Figure 3C**).

147 We reconstructed the GRN for the t(8;21) subset using the pySCENIC [22] pipeline, which is a
148 python-based efficient implementation of original SCENIC method [21]. It is a state-of-the-art
149 method for network inference from scRNA data, popular in the community [45-47] and has shown
150 strong performance in a recent benchmarking study [48]. SCENIC's three major steps are: First,
151 it identifies groups of co-expressed genes as potential targets of a transcription factor (TF).
152 Second, it filters these groups of genes to retain only TF targets with the corresponding binding
153 motif, forming "regulons." Third, it uses the AUCell method (embedded within SCENIC) to
154 quantify the activity of each regulon in every cell. AUCell calculates the Area Under the Curve
155 (AUC) for the regulon's genes set in a ranking of all genes by expression for each cell. The
156 top 20 regulons for each age groups were selected based on the regulon specificity score
157 (RSS) (**Supplementary Figure 4B**). Unsupervised clustering on the Z score normalized
158 regulon activity score matrix revealed clear differences in the GRN across different age groups
159 (**Supplementary Figure 4C**). We hypothesize that the differences in the GRN might reflect
160 differences in the pre- or postnatal developmental origins of the disease. Additional testing of
161 GRN inference from individual studies shows that the high number of cells refines the overall
162 GRN (**see Methods; Supplementary Figure 4D-E**).

163


A
AML scAtlas
Select AML with t(8;21)
Gene Regulatory Network Analysis
Validation 1
TARGET AML Bulk RNA-seq
Validation 2
External scRNA-seq Data
B
UMAP2
UMAP1
B
CD14+ Mono
CD16+ Mono
CMP
Erythroid
GMP
HSPC
MEP
NK
Plasma
ProB
ProMono
Stromal
T
cDC
pDC
C
Pediatric
Adolescent
Adult
D
Inferred-Prenatal Origin Signature
Inferred-Postnatal Origin Signature
Adolescent
Adult
Pediatric
HSC/Myeloid Progenitor Cells
TRIM28_(+)
CTCF_(+)
RAD21_(+)
SOX4_(+)
TAL1_(+)
MYB_(+)
FOXN3_(+)
JUND_(+)
BCLAF1_(+)
ZBTB7A_(+)
IKZF1_(+)
MAZ_(+)
REST_(+)
YY1_(+)
CUX1_(+)
KDM5A_(+)
IRF1_(+)
STAT1_(+)
IRF7_(+)
MAFF_(+)
ATF4_(+)
TAGLN2_(+)
SPI1_(+)
KLF2_(+)
YBX1_(+)
ENO1_(+)
HDAC2_(+)
GATA1_(+)
POLE3_(+)
TFDP1_(+)
MYBL2_(+)
E2F4_(+)
KLF1_(+)

164

Figure 3 – AML 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 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).*

165 To define gene regulatory programs (co-occurring gene modules, defined by a transcription
166 factor and its targets) which are specific to different age groups (termed 'regulon signature'),
167 we used the clustered dendrogram to select the regulon clusters most associated with the
168 pediatric (below 10 years-old) and adult samples (over 18 years-old) (**Figure3D;**
169 **Supplementary Figure 4C**). The pediatric regulon signature, proposed to represent *in-utero*
170 origin t(8;21) AML (henceforth termed 'inferred-prenatal'), includes 16 regulons defined by a
171 distinct group of hematopoietic transcription factors (TFs) (TRIM28, CTCF, RAD21, SOX4,
172 TAL1, MYB, FOXN3, JUND, BCLAF1, ZBTB7A, IKZF1, MAZ, REST, YY1, CUX1, KDM5A),
173 many of which have clearly defined roles in HSPCs and AML [49-51].The adult regulon signature,
174 presumed representative of the postnatally acquired t(8;21) AML (henceforth termed 'inferred-
175 postnatal), combines 3 discrete clusters of regulons (YBX1, ENO1, and HDAC2; GATA1,
176 POLE3, TFDP1, MYBL2, E2F4, and KLF1; IRF1, STAT1, IRF7, MAFF, ATF4, TAGLN2, SPI1,
177 and KLF2), defined by TFs previously implicated in various hematopoietic, leukemic and
178 inflammatory processes [52-54]. Importantly, both signatures contain key components of the AP-
179 1 complex, which is heavily implicated in the biology of t(8;21) AML [55, 56] and undergoes
180 dynamic changes during aging [57]. Samples of 6 individuals aged 12-17 clustered with the
181 pediatric samples and showed enrichment for the inferred-prenatal signature (**Figure 3D**),
182 suggesting that older adolescents (up to aged 17 in our cohort) more closely resemble
183 pediatric AML with t(8;21) and remain biologically distinct from adult-onset disease. This
184 implies that the inferred *in-utero* origin of t(8;21) AML can also be present in AML diagnosed
185 in older children.

186 **Validation of Age-Associated Regulons in Bulk-RNA-Seq Cohorts of t(8;21) AML**

187 We next sought to externally validate our age-associated regulon signatures in a larger cohort
188 of patients. Bulk RNA-seq samples were obtained from the TARGET [2] and BeatAML [23]
189 cohorts, selecting bone marrow samples taken at diagnosis in line with AML scAtlas data
190 (n=83; **Supplementary Table 4**). We applied the AUCell algorithm from pySCENIC [22] to
191 calculate the activity of our pediatric inferred-prenatal and adult inferred-postnatal regulons in
192 each sample. Unsupervised clustering of the bulk RNA-seq AUCell results revealed discrete

193 clusters of samples that were highly enriched for our inferred-prenatal and inferred-postnatal
194 origin-associated regulons (**Figure 4A**).

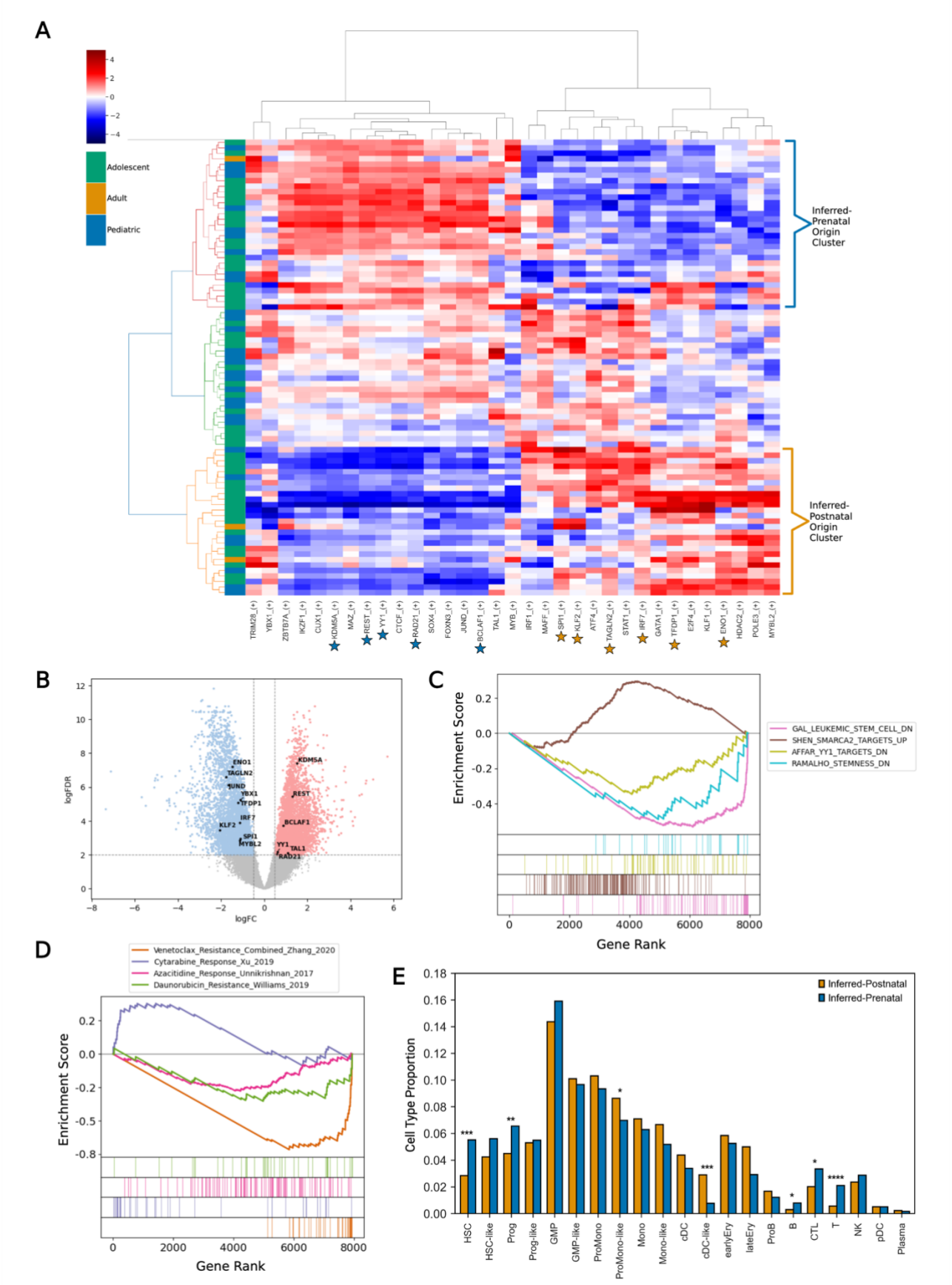


Figure 4 – Validation of Age-Associated Regulons in Large Bulk RNA-Seq Cohorts

*(**A**) Using previously defined age-associated regulons, pySCENIC AUC scores (Z-score normalized) were clustered to identify 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.*

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[59] and edgeR[58]) to ensure robustness of the results (**Figure 4B; Supplementary Figure 5A**). 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; Supplementary Figure 5A**), 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; Supplementary Figure 5B**). This further delineated the inferred-prenatal signature to 5 key TFs (KDM5A, REST, BCLAF1, YY1, and RAD21), and the inferred-postnatal signature to 8 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[58], 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 [60]. SMARCA2 is also known to be upregulated during the fetal-to-adult HSC transition [61], 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

220 implications, we performed GSEA using drug response signatures from published studies [62-
221 65] (**Figure 4D**). This analysis revealed that inferred-prenatal origin t(8;21) AML is enriched for
222 genes associated with increased chemosensitivity to cytarabine, venetoclax, and daunorubicin
223 (**Figure 4D**).

224 We hypothesized that the increase in stemness-associated genes in the leukemia with the
225 inferred-prenatal origin could be reflective of potential differences in the leukemic cell-of-origin
226 and its impact on myeloid differentiation. We therefore performed cell type deconvolution using
227 AutoGeneS [66], with a curated LSPC reference profile [7], to compare the cellular heterogeneity
228 between prenatal and postnatal origin bulk RNA-seq samples (**Figure 4E**). This revealed a
229 higher proportion of HSPC cell types (HSC, Prog), with a reduction in some differentiated
230 myeloid cell types (ProMono-like, cDC-like) in the samples of inferred-prenatal origin (**Figure**
231 **4E**). To corroborate this finding, we examined cell type proportions in the original t(8;21)
232 subset of AML scAtlas, confirming that cells with the inferred-prenatal signature comprise
233 more HSCs than inferred-postnatal signature cells (**Supplementary Figure 5D-E**). However,
234 comparison of cell type proportions in this dataset is confounded by differences in sample
235 processing as some studies performed CD34 selection, hence there is more cell type diversity
236 observed in the pediatric samples (**Supplementary Figure 5D-E**).

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

239 We next used the scRNA-seq and scATAC-seq data from a recent cohort of pediatric t(8;21)
240 AML patients [24] at multiple clinical time points to uncover the enhancer-driven GRN (eGRN)
241 in inferred-prenatal and inferred-postnatal origin t(8;21) AML (**Supplementary Table 4**).
242 Initially, we identified two representative samples of our inferred-prenatal and inferred-
243 postnatal signatures by using pySCENIC AUCell [22] to measure the activity of our previously
244 defined regulons. Unsupervised clustering of the AUC scores was used to infer whether each
245 sample matched the regulon signatures, identifying one inferred-prenatal sample and one
246 inferred-postnatal sample for downstream analysis (**Supplementary Figure 6A**).

247 We then applied SCENIC+ [25], which integrates scRNA-seq and scATAC-seq to identify
248 candidate enhancer regions and TF-binding motifs, linking TFs to target genes and identified
249 enhancers. This creates enhancer-driven regulons (eRegulons), forming an eGRN. We
250 applied SCENIC+ [25] to the leukemia samples with the inferred-prenatal and inferred-postnatal
251 origin at diagnosis and relapse, keeping only regulons that showed a correlation between both
252 modalities to retain only the most robust regulons (**Supplementary Figure 6B**). This revealed
253 several eRegulons across both patients (**Supplementary Figure 6D**), many of which were

254 patient specific, particularly when comparing HSPC populations (**Figure 4A**). The inferred-
255 prenatal sample displayed a specific HSC eRegulon profile. In contrast, the inferred-postnatal
256 sample more closely resembled the corresponding Granulocyte-Monocyte Progenitor (GMP)
257 (**Figure 5A**). Interestingly, at relapse the inferred-prenatal origin patient undergoes a
258 chemotherapy-driven lineage switch to a lymphoid phenotype, which may suggest that the
259 leukemia originated from a less committed progenitor (**Figure 5A**).

260 To identify clusters of closely related eRegulons, we computed the correlations between
261 eRegulons enrichment. We identified 2 main clusters of eRegulons which correspond to
262 different inferred-signature samples (**Figure 4B; Supplementary Figure 6C**). For each
263 eRegulon cluster, we used the associated target genes as input for gene ontology over
264 representation analysis (ORA), to assess functional differences in the eGRN. This revealed
265 fundamental differences in the underlying biological processes (**Figure 5C**; **Supplementary**
266 **Figure 7A**). The AML sample with inferred-prenatal origin was enriched for many processes
267 associated with development. In contrast, inferred-postnatal samples appeared more
268 metabolism focused (**Figure 5C**; **Supplementary Figure 7A**). This further supports the
269 association of these eRegulons with presumed prenatal origin t(8;21) AML, compared to
270 postnatal origin disease.

271 Previous analysis using the TARGET [2] and BeatAML [23] datasets indicated that inferred-
272 prenatal and inferred-postnatal origin t(8;21) AML may harbor different levels of
273 chemosensitivity based on published drug response signatures (**Figure 4D**). Therefore, we
274 performed *in silico* perturbations of eRegulon-associated TFs. PCA of the diagnosis and
275 relapse samples recapitulated the expected differentiation trajectories along PC2, while
276 separating diagnosis from relapse along PC1 (**Figure 5D**). Using the SCENIC+ [25] perturbation
277 simulation workflow, we identified TFs estimated to induce differentiation, as defined by a
278 negative shift in PC2 (**Supplementary Figure 7B**). We prioritized TFs predicted to impact the
279 HSC compartment and identified 18 TFs with predicted significant effects on HSC
280 differentiation (**Supplementary Figure 7C**). Several of these are components of the AP-1
281 complex (JUN, ATF4, FOSL2), which are established downstream targets of the t(8;21) fusion
282 protein and are known to propagate t(8;21) AML [55, 67] (**Figure 5E; Supplementary Figure**
283 **7C**).

284 Using AP-1 complex members as a comparative baseline, we identified EP300 as one of the
285 most impactful hits. EP300 has recently been shown to drive t(8;21) AML self-renewal through
286 acetylation dependent mechanism [68]. This suggests that presumed prenatal origin pediatric
287 t(8;21) AML may be particularly sensitive to EP300 inhibition. One of the most striking
288 predictions, for both diagnostic and relapse HSC populations, was BCLAF1 (**Figure 5E;**

289 **Supplementary Figure 7C**). BCLAF1 is a regulator of normal HSPCs [69], and its expression
290 level declines during hematopoietic differentiation. While recent studies have identified a role
291 for BCLAF1 in AML [70], this has not been explored in detail in the context of pediatric AML or
292 t(8;21) AML and may present a therapeutic opportunity.

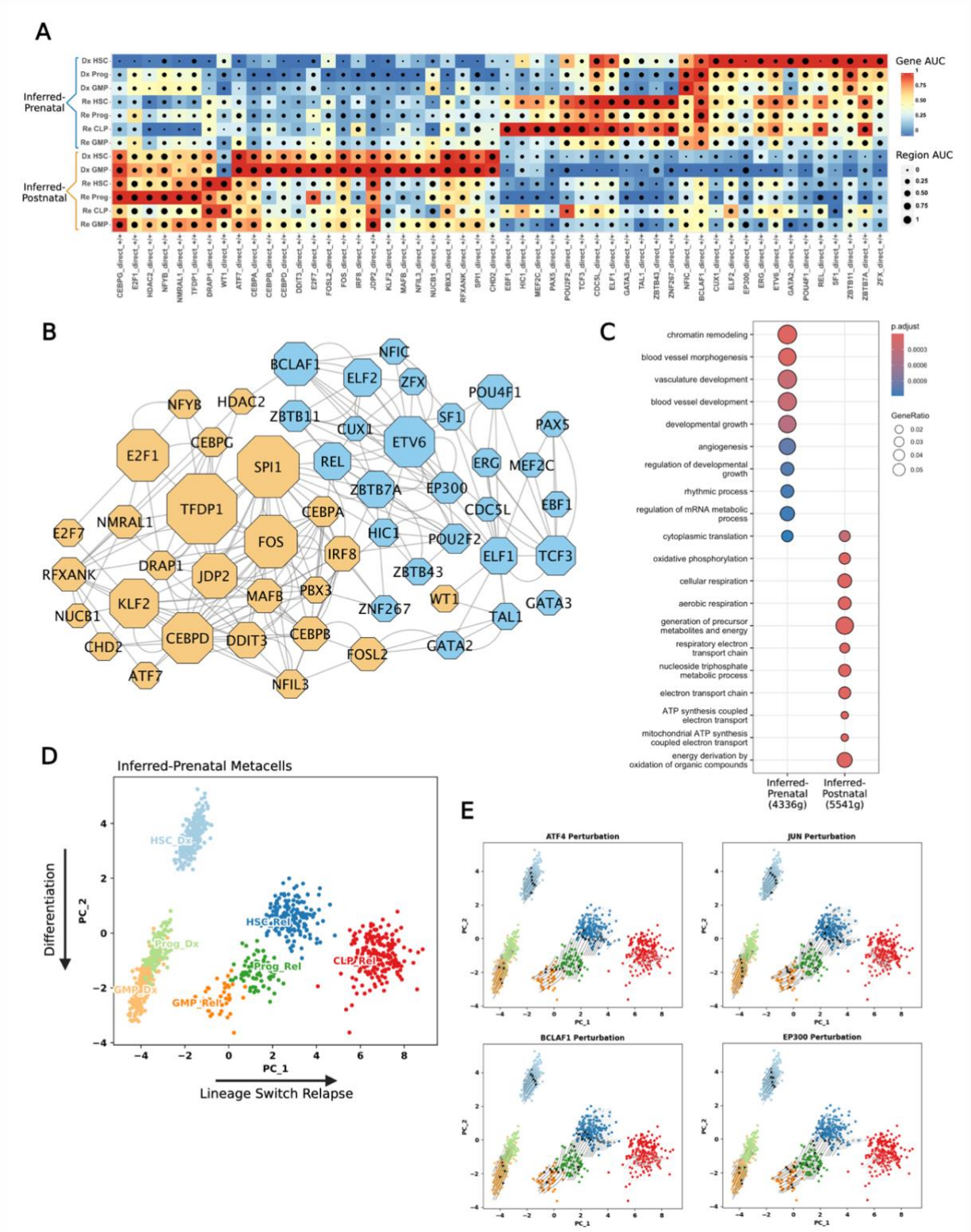


Figure 5 – Combining Multiomics Data Interrogates Age-Associated Regulons

*(**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**) Network showing the inferred-prenatal (blue) and inferred-postnatal (orange) associated 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 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.*

We also performed SCENIC+ [25] 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 (**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 from our original postnatal origin signature were predicted to have an effect (**Supplementary Figures 7D-F**), supporting the relevance of the GRNs identified in our previous analyses.

To further investigate EP300 and BCLAF1, we queried the DepMap [71] 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) [72] and SKNO-1 (22-year-old donor) [73], were among the most sensitive to these perturbations based on their DepMap effect scores (**Supplementary Figure 7G**). Several other cell lines sensitive to BCLAF1 were derived from pediatric cancers, most notably neuroblastomas, which also arise *in-utero* [74] (**Supplementary Figure 7H**). 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 pre-natal origin of the driver translocation.

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

318 (https://cellxgene.bmh.manchester.ac.uk/AML/) providing the AML research community with
319 a resource to address diverse biological questions and generate new hypotheses.

320 To further address a clinically relevant question using this data source, we compared
321 differences between pediatric and adult-onset disease based on the potential biological effect
322 of the *in-utero* origin of pediatric leukemia. Data of our AML scAtlas was used to explore the
323 GRNs in adult and pediatric t(8;21) AML and revealed a strong age-associated GRN signature.
324 This suggests that while pediatric and adult t(8;21) AMLs are propagated by the same driver
325 translocation, they exhibit clear biological differences correlated with age. This may be due to
326 differences in the cell-of-origin, with mouse models showing that t(8;21) AML can arise from
327 a HSC or a more lineage restricted GMP [75]. As pediatric t(8;21) can arise *in-utero*, as
328 evidenced by previous studies [17], and adult t(8;21) is acquired postnatally [18, 19], we propose
329 that the observed age-related differences in AML with t(8;21) reflect these differences in the
330 developmental origins of the disease. We identified two distinct groups of regulons
331 corresponding to either inferred-prenatal origin and inferred-postnatal origin disease. These
332 regulons constitute the GRN underlying the cellular state, which can be informative when
333 identifying molecular vulnerabilities to target leukemia.

334 Our cohort is the largest scRNA-seq dataset to explore t(8;21) AML biology to date, however,
335 the number of patients included remains low (n=22), and many of the studies containing the
336 adult samples used CD34 selection in their experimental protocol creating a bias towards
337 HSPCs in these samples. To overcome some of these limitations, we used bulk RNA-seq
338 samples from the TARGET [2] and BeatAML [23] studies with t(8;21) AML (n=83) to validate our
339 regulon signatures. This identifies two clusters of samples which closely match these
340 signatures, showing that the regulon patterns identified from our AML scAtlas are recapitulated
341 with bulk RNA-seq data enabling exploration of larger patient cohorts. Comparisons between
342 inferred-prenatal and inferred-postnatal origin transcriptomes prioritized TFs which were
343 differentially expressed and highlighted differences in underlying biology and drug response.
344 We identified 5 signature TFs (KDM5A, REST, BCLAF1, YY1, RAD21) for inferred-prenatal
345 origin disease, several of which have roles in embryonic stem cells [76, 77], and critical functions
346 in the maintenance of HSCs [49-51]. In contrast, TFs identified in inferred-postnatal origin
347 samples, such as interferon regulatory factors (IRFs), HDAC2, and SPI1, reflect inflammatory
348 and immune processes, many of which have been implicated in leukemia [52-54]. We also found
349 that inferred-prenatal origin samples had a higher proportion of HSC/Prog cell types compared
350 to inferred-postnatal origin samples, a more primitive state than postnatal onset t(8;21) AML
351 cases, supporting the hypothesis that age-associated differences in the cell-of-origin influence
352 disease biology.

353 Given these biological differences, we used bulk RNA-seq to predict chemosensitivity using
354 published drug response signatures [62-65]. Inferred-prenatal samples were enriched for genes
355 indicative of cytarabine sensitivity and depleted of genes suggestive of daunorubicin and
356 venetoclax resistance. These findings suggest that the developmental origins of the disease
357 may influence drug responses, with potential implications in the design of novel therapeutic
358 strategies and providing further biological evidence that pediatric AML might benefit from
359 different clinical management compared with adult-onset AML. Importantly, venetoclax is
360 currently in the AML23 trial (NCT05955261); our results support further evaluation of
361 venetoclax treatment in pediatric t(8;21) AML.

362 Using an additional single-cell multiomic dataset, using SCENIC+, we reconstructed the eGRN
363 in samples matching our inferred-prenatal and inferred-postnatal regulon signatures. Upon
364 comparing eRegulons for each patient at diagnosis and relapse, we identified clusters of highly
365 correlated eRegulons defined by different biological processes. Inferred-prenatal origin
366 samples are characterized by developmental and transcriptional dysregulation, whereas
367 inferred-postnatal origin samples are largely driven by fundamental cellular processes linked
368 to inflammation. We used SCENIC+ to model the predicted impact of TF perturbations on our
369 prenatal origin sample at diagnosis and relapse identified several key components of the AP-
370 1 complex, which are critical in t(8;21) AML biology and are also associated with dynamic age-
371 related transcriptional changes [55-57].

372 Through our analysis, we identified EP300 as a candidate target, which has been shown to
373 be critical for t(8;21) AML biology [68] with demonstrable effects in KASUMI-1 and SKNO1 cell
374 lines. EP300 has been identified as a promising therapeutic target in AML with several
375 molecules in development [78]; our data indicate potential specific therapeutic benefit in
376 prenatal origin t(8;21) AML. One of the most impactful perturbation predictions for the HSC
377 compartment at diagnosis and relapse was BCLAF1. This is consistent with previous evidence
378 of its importance in HSCs [69, 79] and AML [70], but has not been studied specifically in the context
379 of pediatric AML or t(8;21) AML previously. The DepMap data shows that KASUMI-1 is the
380 most sensitive myeloid cell line to BCLAF1 perturbation, and our GRN analyses suggest it is
381 particularly active in pediatric t(8;21) AML of inferred *in-utero* origin, thus this may represent
382 an additional prognostic indicator.

383 Further investigations are required to characterize the roles of both EP300 and BCLAF1 in
384 prenatal origin t(8;21) AML before any clinical realisation. EP300 has already been
385 investigated as a target in AML, so future work should focus on the pediatric AML setting with
386 *in-vitro* and *in-vivo* studies using EP300/CBP inhibitors such as inobrodib [78]. In contrast

387 BCLAF1 is relatively unexplored, and additional work is required to elucidate its molecular
388 function and assess its potential as a therapeutic target. BCLAF1 may ultimately prove most
389 valuable as a biomarker of *in-utero* t(8;21) AML, enabling distinction between late-onset *in-*
390 *utero* and postnatal disease. This would require molecular validation in a large cohort of
391 pediatric patients with Gunthrie spots to confirm whether they had acquired t(8;21) *in-utero*.
392
393 **Conclusions**
394 Overall, our study demonstrates that large-scale single-cell data integration is a powerful
395 approach to dissect specific patient groups in detail, and enabling robust comparative
396 analyses. We present the AML scAtlas as a publicly available resource for the research
397 community to address diverse biological questions. By applying AML scAtlas to t(8;21) AML,
398 we identified age-associated gene regulatory networks that likely reflect differences in the
399 developmental origins, biology and outcome of the disease. These findings also highlight novel
400 candidate therapeutic targets which may be more relevant in pediatric t(8;21) AML compared
401 to adult-onset disease, offering opportunities for more tailored treatment strategies.
402
403
404 **Methods**
405 For the complete analysis code, including the conda environments used for analysis, see
406 GitHub Repo (https://github.com/jesswhitts/AML-scAtlas).
407
408 Data Collection
409 A literature search was performed for published AML scRNA-seq datasets [4-6, 8, 26-40]. Suitable
410 studies were selected based on the data quality (over 1000 counts and 500 genes detected
411 per cell for most of the data). Diagnostic, primary AML samples were selected from each AML
412 study. Where healthy donor samples were present, these were also included, along with an
413 additional 4 studies with healthy bone marrow samples.
414
415 Initial Data Processing
416 Each scRNA-seq dataset underwent initial quality control individually using Scanpy (v1.9.3)
417 [80] as some studies provided raw data and others provided pre-filtered data. Where raw data
418 was provided, doublets were removed using Scrublet (v0.2.3) [81] and cells were filtered using
419 the median absolute deviation as described in this single-cell best practices handbook [10, 82].
420
421 Once filtered, datasets were combined, and quality control was performed using Scanpy
422 (v1.9.3) [80]. The full dataset had quality thresholds applied (percentage mitochondrial counts

423 <10, read counts >1000, gene counts >500), removing any samples which had fewer than 50
424 cells remaining after filtering. Genes present in <50 cells were removed. MATAL1 was
425 removed as this was highly abundant in many cells and considered artefactual.
426
427 Batch Correction
428 The presence of batch effects was determined through dimensionality reduction and clustering
429 using Scanpy (v1.9.3) [80] and using the kBET algorithm (v0.99.6) [83]. This was repeated on
430 individual studies, to assess whether there were sample-wise batch effects. Batch correction
431 benchmarking was implemented using Harmony (Scanpy v1.9.3 implementation) [11], scVI
432 (v1.0.3) [12], and scANVI (v1.0.3) [13] and quantified using scIB (1.1.4) [9]. Different numbers of
433 highly variable genes were used to select the optimal number for integration. Batch correction
434 was performed using scVI (v1.0.3) [12] with the top 2000 highly variable genes, using sample
435 as the model covariate.
436
437 AML scAtlas Cell Type Annotation
438 The scVI corrected embedding was used to run UMAP and Leiden clustering using Scanpy
439 functions (v1.9.3) [80]. Cell type annotation was performed using CellTypist (v1.6.0) [84] using
440 the 'Immune_All_Low.pkl' model, SingleR (v2.0.0) [85] using the Novershtern hematopoietic
441 refreference[86], and scType (v1.0) [87] with the tissue defined as 'Immune system'. Full
442 automated tool outputs are detailed in **Supplementary Table 3**; overall we found that the
443 results varied significantly between different tools. We postulate that this is, in part, due to
444 differences in the reference profiles used. Thus, we opted to use the best consensus of these
445 different tools for our cluster identity assignments.

446 AML scAtlas LSC Annotation
447 HSPC clusters were selected from AML scAtlas, and the scVI corrected embedding was used
448 to re-compute UMAP using Scanpy functions (v1.9.3) [80]. As our previous cell type annotations
449 used generic reference profiles and were not AML specific, we generated a custom cell type
450 annotation reference to identify LSCs. We created a custom SingleR (v2.0.0) [85] reference
451 using the Zeng et al [7] revised annotations of the Van Galen et al [5] dataset (**Supplementary**
452 **Table 3**). This was also correlated with the LSC6 [43] and LSC17 [44] scores for each cell. To
453 compare LSC abundance between ELN risk groups, chi2_contingency was implemented from
454 SciPy (v1.12.0).
455 P
456 AML with t(8;21) Analysis
457 Samples with the t(8;21) translocation were selected from the full AML scAtlas. The UMAP
458 was re-computed, and genes were filtered to remove those detected in fewer than 50 cells for

459 the revised dataset, leaving 24,866 genes remaining. Gene regulatory network analysis was
460 performed using pySCENIC [21, 22] (v0.12.1) as per the recommended workflow. To facilitate
461 comparisons between age groups, cell types were focused on HSPCs, as many adult samples
462 were originally enriched for CD34. The RSS was calculated for the adult and pediatric samples
463 to select the top 20 differential regulons per age group. Using SciPy hierarchical clustering
464 (v1.12.0), regulons were filtered to identify regulon signature groups used for downstream
465 analysis.
466
467 Bulk RNA-Seq Analysis
468 Bulk RNA-Seq data was downloaded for the TARGET [2] and BeatAML [23] cohorts and samples
469 with t(8;21) were selected. Only bone marrow samples taken at diagnosis were used for
470 downstream analyses (**Supplementary Table 4**). Using the previously defined signature
471 regulons, the AUCell algorithm [21] (v0.12.1) was implemented to measure regulon activity.
472 Hierarchical clustering was performed using SciPy v1.12.0) to identify samples most enriched
473 for each age-related signature. Differential gene expression analysis was implemented using
474 edgeR [58] (v3.42.4) and DESeq2 [59] (v1.40.2), using a log2 fold change threshold of 0.5 and
475 an adjusted p value cutoff of 0.01. Candidate differential genes, ranked on log2 fold change,
476 underwent GSEA with GSEApy (v0.10.8) using a significance threshold of 0.05. Cell type
477 deconvolution was performed using AutoGeneS [66] (v1.0.4) using the recommended workflow.
478 Significance when comparing groups was ascertained using a student's T-test on the
479 predicted cell type proportion values for each sample.
480
481 Single-Cell Multi-Omics Analysis
482 The Lambo et al [24] scRNA-seq and scATAC-seq data from pediatric AML bone marrow
483 samples was downloaded, and the t(8;21) samples selected (**Supplementary Table 4**). Using
484 our previously defined signature regulons, the AUCell algorithm [21] (v0.12.1) was implemented
485 to measure regulon activity. This identified the samples most enriched for each age-related
486 signature as AML16 and AML12.
487
488 The SCENIC+ pipeline [25] (v1.0a1) was implemented as per the recommended Snakemake
489 workflow for creating pseudo-multiome data. Regulons were filtered by correlation between
490 modalities, using a threshold of 0.2 for non-multiome data. The most robust regulons were
491 prioritized based on the SCENIC+ recommendations (direct +/+) [25]. To facilitate comparisons,
492 the eRegulon RSS was calculated for each patient and the top 30 eRegulons selected. The
493 correlation between the gene sets underpinning eRegulons was calculated and sample-
494 associated clusters were selected for over-representation analysis with clusterProfiler [88]
495 (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 [71] 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). The samples used for validation analyses are publicly available and are detailed in Supplementary Table 4. The SCENIC+ eGRN files are provided as Supplementary Table 5.

Author Contributions

MI, SMB, and GL conceived the study and oversaw the research. JW did the data collection and implemented the analyses. SM guided sub-analyses and assisted with interpretation. 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

533 landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational
534 interactions. Nat Med. 2018;24(1):103-12.
535 3. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D,
536 Boch T, Hofmann WK, Ho AD, Huber W, Trumpp A, Essers MA, Steinmetz LM. Human haematopoietic stem cell
537 lineage commitment is a continuous process. Nat Cell Biol. 2017;19(4):271-81.
538 4. Velten L, Story BA, Hernández-Malmierca P, Raffel S, Leonce DR, Milbank J, Paulsen M, Demir A, Szu-
539 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,
540 Steinmetz LM. Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell
541 transcriptomics. Nature Communications. 2021;12(1):1366.
542 5. van Galen P, Hovestadt V, Wadsworth Ii MH, Hughes TK, Griffin GK, Battaglia S, Verga JA, Stephansky
543 J, Pastika TJ, Lombardi Story J, Pinkus GS, Pozdnyakova O, Galinsky I, Stone RM, Graubert TA, Shalek AK, Aster
544 JC, Lane AA, Bernstein BE. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and
545 Immunity. Cell. 2019;176(6):1265-81.e24.
546 6. Beneyto-Calabuig S, Merbach AK, Kniffka JA, Antes M, Szu-Tu C, Rohde C, Waclawiczek A, Stelmach
547 P, Gräßle S, Pervan P, Janssen M, Landry JJM, Benes V, Jauch A, Brough M, Bauer M, Besenbeck B, Felden J,
548 Bäumer S, Hundemer M, Sauer T, Pabst C, Wickenhauser C, Angenendt L, Schliemann C, Trumpp A, Haas S,
549 Scherer M, Raffel S, Müller-Tidow C, Velten L. Clonally resolved single-cell multi-omics identifies routes of cellular
550 differentiation in acute myeloid leukemia. Cell Stem Cell. 2023;30(5):706-21.e8.
551 7. Zeng AGX, Bansal S, Jin L, Mitchell A, Chen WC, Abbas HA, Chan-Seng-Yue M, Voisin V, van Galen P,
552 Tierens A, Cheok M, Preudhomme C, Dombret H, Daver N, Futreal PA, Minden MD, Kennedy JA, Wang JCY, Dick
553 JE. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid
554 leukemia. Nature Medicine. 2022;28(6):1212-23.
555 8. Stetson LC, Balasubramanian D, Ribeiro SP, Stefan T, Gupta K, Xu X, Fourati S, Roe A, Jackson Z,
556 Schauner R, Sharma A, Tamilselvan B, Li S, de Lima M, Hwang TH, Balderas R, Saunthararajah Y, Maciejewski
557 J, LaFramboise T, Barnholtz-Sloan JS, Sekaly RP, Wald DN. Single cell RNA sequencing of AML initiating cells
558 reveals RNA-based evolution during disease progression. Leukemia. 2021;35(10):2799-812.
559 9. Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, Strobl DC, Zappia L, Dugas
560 M, Colomé-Tatché M, Theis FJ. Benchmarking atlas-level data integration in single-cell genomics. Nature Methods.
561 2022;19(1):41-50.
562 10. Heumos L, Schaar AC, Lance C, Litinetskaya A, Drost F, Zappia L, Lücken MD, Strobl DC, Henao J,
563 Curion F, Aliee H, Ansari M, Badia-i-Mompel P, Büttner M, Dann E, Dimitrov D, Dony L, Frishberg A, He D, Hediyeh-
564 zadeh S, Hetzel L, Ibarra IL, Jones MG, Lotfollahi M, Martens LD, Müller CL, Nitzan M, Ostner J, Palla G, Patro R,
565 Piran Z, Ramírez-Suástegui C, Saez-Rodriguez J, Sarkar H, Schubert B, Sikkema L, Srivastava A, Tanevski J,
566 Virshup I, Weiler P, Schiller HB, Theis FJ, Single-cell Best Practices C. Best practices for single-cell analysis across
567 modalities. Nature Reviews Genetics. 2023.
568 11. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P-r,
569 Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. Nature Methods.
570 2019;16(12):1289-96.
571 12. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell
572 transcriptomics. Nature Methods. 2018;15(12):1053-8.
573 13. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of
574 single-cell transcriptomics data with deep generative models. Molecular Systems Biology. 2021;17(1):e9620.
575 14. Balgobind BV, Van den Heuvel-Eibrink MM, De Menezes RX, Reinhardt D, Hollink IH, Arentsen-Peters
576 ST, van Wering ER, Kaspers GJ, Cloos J, de Bont ES, Cayuela JM, Baruchel A, Meyer C, Marschalek R, Trka J,

577 Stary J, Beverloo HB, Pieters R, Zwaan CM, den Boer ML. Evaluation of gene expression signatures predictive of
578 cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. Haematologica. 2011;96(2):221-30.
579 15. Wiggers CRM, Baak ML, Sonneveld E, Nieuwenhuis EES, Bartels M, Creyghton MP. AML Subtype Is a
580 Major Determinant of the Association between Prognostic Gene Expression Signatures and Their Clinical
581 Significance. Cell Reports. 2019;28(11):2866-77.e5.
582 16. Chaudhury S, O'Connor C, Cañete A, Bittencourt-Silvestre J, Sarrou E, Prendergast Á, Choi J, Johnston
583 P, Wells CA, Gibson B, Keeshan K. Age-specific biological and molecular profiling distinguishes paediatric from
584 adult acute myeloid leukaemias. Nat Commun. 2018;9(1):5280.
585 17. Wiemels JL, Xiao Z, Buffler PA, Maia AT, Ma X, Dicks BM, Smith MT, Zhang L, Feusner J, Wiencke J,
586 Pritchard-Jones K, Kempski H, Greaves M. In utero origin of t(8;21) AML1-ETO translocations in childhood acute
587 myeloid leukemia. Blood. 2002;99(10):3801-5.
588 18. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J,
589 Kandoth C, Fulton RS, McLellan MD, Dooling DJ, Wallis JW, Chen K, Harris CC, Schmidt HK, Kalicki-Veizer JM,
590 Lu C, Zhang Q, Lin L, O'Laughlin MD, McMichael JF, Delehaunty KD, Fulton LA, Magrini VJ, McGrath SD, Demeter
591 RT, Vickery TL, Hundal J, Cook LL, Swift GW, Reed JP, Alldredge PA, Wylie TN, Walker JR, Watson MA, Heath
592 SE, Shannon WD, Varghese N, Nagarajan R, Payton JE, Baty JD, Kulkarni S, Klco JM, Tomasson MH, Westervelt
593 P, Walter MJ, Graubert TA, DiPersio JF, Ding L, Mardis ER, Wilson RK. The origin and evolution of mutations in
594 acute myeloid leukemia. Cell. 2012;150(2):264-78.
595 19. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burtt
596 N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen HA, Ladenvall C,
597 Getz G, Correa A, Banahan BF, Gabriel S, Kathiresan S, Stringham HM, McCarthy MI, Boehnke M, Tuomilehto J,
598 Haiman C, Groop L, Atzmon G, Wilson JG, Neuberg D, Altshuler D, Ebert BL. Age-related clonal hematopoiesis
599 associated with adverse outcomes. N Engl J Med. 2014;371(26):2488-98.
600 20. National Cancer Registration and Analysis Service, Northern Ireland Cancer Registry, Scottish Cancer
601 Registry, Unit WCIaS. Children, teenagers and young adults UK cancer statistics report 2021 2021 [Available from:
602 http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/cancer_in_children_tee
603 nagers_and_young_adults/.
604 21. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine
605 J-C, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S. SCENIC: single-cell regulatory network
606 inference and clustering. Nature Methods. 2017;14(11):1083-6.
607 22. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, Seurinck R, Saelens W,
608 Cannoodt R, Rouchon Q, Verbeiren T, De Maeyer D, Reumers J, Saeys Y, Aerts S. A scalable SCENIC workflow
609 for single-cell gene regulatory network analysis. Nature Protocols. 2020;15(7):2247-76.
610 23. Burd A, Levine RL, Ruppert AS, Mims AS, Borate U, Stein EM, Patel P, Baer MR, Stock W, Deininger M,
611 Blum W, Schiller G, Olin R, Litzow M, Foran J, Lin TL, Ball B, Boyiadzis M, Traer E, Odenike O, Arellano M, Walker
612 A, Duong VH, Kovacsovics T, Collins R, Shoben AB, Heerema NA, Foster MC, Vergilio J-A, Brennan T, Vietz C,
613 Severson E, Miller M, Rosenberg L, Marcus S, Yocum A, Chen T, Stefanos M, Druker B, Byrd JC. Precision
614 medicine treatment in acute myeloid leukemia using prospective genomic profiling: feasibility and preliminary
615 efficacy of the Beat AML Master Trial. Nature Medicine. 2020;26(12):1852-8.
616 24. Lambo S, Trinh DL, Ries RE, Jin D, Setiadi A, Ng M, Leblanc VG, Loken MR, Brodersen LE, Dai F, Pardo
617 LM, Ma X, Vercauteren SM, Meshinchi S, Marra MA. A longitudinal single-cell atlas of treatment response in
618 pediatric AML. Cancer Cell. 2023.
619 25. Bravo González-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V, Poovathingal
620 S, Wouters J, Aibar S, Aerts S. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory
621 networks. Nature Methods. 2023;20(9):1355-67.

622 26. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott
623 GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin
624 M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente
625 WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH. Massively parallel digital
626 transcriptional profiling of single cells. Nature Communications. 2017;8(1):14049.
627 27. Petti AA, Williams SR, Miller CA, Fiddes IT, Srivatsan SN, Chen DY, Fronick CC, Fulton RS, Church DM,
628 Ley TJ. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. Nat
629 Commun. 2019;10(1):3660.
630 28. Jiang L, Li XP, Dai YT, Chen B, Weng XQ, Xiong SM, Zhang M, Huang JY, Chen Z, Chen SJ.
631 Multidimensional study of the heterogeneity of leukemia cells in t(8;21) acute myelogenous leukemia identifies the
632 subtype with poor outcome. Proc Natl Acad Sci U S A. 2020;117(33):20117-26.
633 29. Johnston G, Ramsey HE, Liu Q, Wang J, Stengel KR, Sampathi S, Acharya P, Arrate M, Stubbs MC, Burn
634 T, Savona MR, Hiebert SW. Nascent transcript and single-cell RNA-seq analysis defines the mechanism of action
635 of the LSD1 inhibitor INCB059872 in myeloid leukemia. Gene. 2020;752:144758.
636 30. Pei S, Pollyea DA, Gustafson A, Stevens BM, Minhajuddin M, Fu R, Riemondy KA, Gillen AE, Sheridan
637 RM, Kim J, Costello JC, Amaya ML, Inguva A, Winters A, Ye H, Krug A, Jones CL, Adane B, Khan N, Ponder J,
638 Schowinsky J, Abbott D, Hammes A, Myers JR, Ashton JM, Nemkov T, D'Alessandro A, Gutman JA, Ramsey HE,
639 Savona MR, Smith CA, Jordan CT. Monocytic Subclones Confer Resistance to Venetoclax-Based Therapy in
640 Patients with Acute Myeloid Leukemia. Cancer Discov. 2020;10(4):536-51.
641 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
642 W, Wu L, Li J, Wang Q, Hong M, Qian S. Single-cell analysis reveals the chemotherapy-induced cellular
643 reprogramming and novel therapeutic targets in relapsed/refractory acute myeloid leukemia. Leukemia.
644 2023;37(2):308-25.
645 32. Lasry A, Nadorp B, Fornerod M, Nicolet D, Wu H, Walker CJ, Sun Z, Witkowski MT, Tikhonova AN,
646 Guillamot-Ruano M, Cayanan G, Yeaton A, Robbins G, Obeng EA, Tsirigos A, Stone RM, Byrd JC, Pounds S,
647 Carroll WL, Gruber TA, Eisfeld A-K, Aifantis I. An inflammatory state remodels the immune microenvironment and
648 improves risk stratification in acute myeloid leukemia. Nature Cancer. 2023;4(1):27-42.
649 33. Fiskus W, Mill CP, Birdwell C, Davis JA, Das K, Boettcher S, Kadia TM, DiNardo CD, Takahashi K,
650 Loghavi S, Soth MJ, Heffernan T, McGeehan GM, Ruan X, Su X, Vakoc CR, Daver N, Bhalla KN. Targeting of
651 epigenetic co-dependencies enhances anti-AML efficacy of Menin inhibitor in AML with MLL1-r or mutant NPM1.
652 Blood Cancer Journal. 2023;13(1):53.
653 34. Naldini MM, Casirati G, Barcella M, Rancoita PMV, Cosentino A, Caserta C, Pavesi F, Zonari E, Desantis
654 G, Gilioli D, Carrabba MG, Vago L, Bernardi M, Di Micco R, Di Serio C, Merelli I, Volpin M, Montini E, Ciceri F,
655 Gentner B. Longitudinal single-cell profiling of chemotherapy response in acute myeloid leukemia. Nature
656 Communications. 2023;14(1):1285.
657 35. Mumme H, Thomas BE, Bhasin SS, Krishnan U, Dwivedi B, Perumalla P, Sarkar D, Ulukaya GB, Sabnis
658 HS, Park SI, DeRyckere D, Raikar SS, Pauly M, Summers RJ, Castellino SM, Wechsler DS, Porter CC, Graham
659 DK, Bhasin M. Single-cell analysis reveals altered tumor microenvironments of relapse- and remission-associated
660 pediatric acute myeloid leukemia. Nature Communications. 2023;14(1):6209.
661 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,
662 Zhou Z, Su Y, Qin L, Wang Y, He H, Lu J, Xiao P, Hu S, Wang Q-F. Single-cell transcriptomics reveals multiple
663 chemoresistant properties in leukemic stem and progenitor cells in pediatric AML. Genome Biology.
664 2023;24(1):199.
665 37. Li B, Kowalczyk MS, Slyper M, Jellert G, Tabaka M, Ashenberg O, Waldman J, Dionne D, Abigail K, Hui
666 M, Yang Y, Rozenblatt-Rosen O, Regev A. A single cell immune cell atlas of human hematopoietic system Human

667 Cell Atlas Data Portal: Human Cell Atlas; 2022 [Available from:
668 https://explore.data.humancellatlas.org/projects/cc95ff89-2e68-4a08-a234-480eca21ce79.
669 38. Oetjen KA, Lindblad KE, Goswami M, Gui G, Dagur PK, Lai C, Dillon LW, McCoy JP, Hourigan CS. Human
670 bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. JCI Insight.
671 2018;3(23).
672 39. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. Characterization of cell fate probabilities
673 in single-cell data with Palantir. Nat Biotechnol. 2019;37(4):451-60.
674 40. Caron M, St-Onge P, Sontag T, Wang YC, Richer C, Ragoussis I, Sinnett D, Bourque G. Single-cell
675 analysis of childhood leukemia reveals a link between developmental states and ribosomal protein expression as
676 a source of intra-individual heterogeneity. Scientific Reports. 2020;10(1):8079.
677 41. Döhner H, Wei AH, Appelbaum FR, Craddock C, DiNardo CD, Dombret H, Ebert BL, Fenaux P, Godley
678 LA, Hasserjian RP, Larson RA, Levine RL, Miyazaki Y, Niederwieser D, Ossenkoppele G, Röllig C, Sierra J, Stein
679 EM, Tallman MS, Tien H-F, Wang J, Wierzbowska A, Löwenberg B. Diagnosis and management of AML in adults:
680 2022 recommendations from an international expert panel on behalf of the ELN. Blood. 2022;140(12):1345-77.
681 42. Montefiori LE, Bendig S, Gu Z, Chen X, Pölönen P, Ma X, Murison A, Zeng A, Garcia-Prat L, Dickerson
682 K, Iacobucci I, Abdelhamed S, Hiltenbrand R, Mead PE, Mehr CM, Xu B, Cheng Z, Chang TC, Westover T, Ma J,
683 Stengel A, Kimura S, Qu C, Valentine MB, Rashkovan M, Luger S, Litzow MR, Rowe JM, den Boer ML, Wang V,
684 Yin J, Kornblau SM, Hunger SP, Loh ML, Pui CH, Yang W, Crews KR, Roberts KG, Yang JJ, Relling MV, Evans
685 WE, Stock W, Paietta EM, Ferrando AA, Zhang J, Kern W, Haferlach T, Wu G, Dick JE, Klco JM, Haferlach C,
686 Mullighan CG. Enhancer Hijacking Drives Oncogenic BCL11B Expression in Lineage-Ambiguous Stem Cell
687 Leukemia. Cancer Discov. 2021;11(11):2846-67.
688 43. Elsayed AH, Rafiee R, Cao X, Raimondi S, Downing JR, Ribeiro R, Fan Y, Gruber TA, Baker S, Klco J,
689 Rubnitz JE, Pounds S, Lamba JK. A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid
690 leukemia. Leukemia. 2020;34(3):735-45.
691 44. Ng SWK, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, Arruda A, Popescu A, Gupta V,
692 Schimmer AD, Schuh AC, Yee KW, Bullinger L, Herold T, Görlich D, Büchner T, Hiddemann W, Berdel WE,
693 Wörmann B, Cheok M, Preudhomme C, Dombret H, Metzeler K, Buske C, Löwenberg B, Valk PJM, Zandstra PW,
694 Minden MD, Dick JE, Wang JCY. A 17-gene stemness score for rapid determination of risk in acute leukaemia.
695 Nature. 2016;540(7633):433-7.
696 45. Hamed AA, Kunz DJ, El-Hamamy I, Trinh QM, Subedar OD, Richards LM, Foltz W, Bullivant G, Ware M,
697 Vladoiu MC, Zhang J, Raj AM, Pugh TJ, Taylor MD, Teichmann SA, Stein LD, Simons BD, Dirks PB. A brain
698 precursor atlas reveals the acquisition of developmental-like states in adult cerebral tumours. Nat Commun. 2022
699 Jul 19;13(1):4178.
700 46. Barnett SN, Cujba AM, Yang L, Maceiras AR, Li S, Kedlian VR, Pett JP, Polanski K, Miranda AMA, Xu C,
701 Cranley J, Kanemaru K, Lee M, Mach L, Perera S, Tudor C, Joseph PD, Pritchard S, Toscano-Rivalta R, Tuong
702 ZK, Bolt L, Petryszak R, Prete M, Cakir B, Huseynov A, Sarropoulos I, Chowdhury RA, Elmentaite R, Madissoon
703 E, Oliver AJ, Campos L, Brazovskaja A, Gomes T, Treutlein B, Kim CN, Nowakowski TJ, Meyer KB, Randi AM,
704 Noseda M, Teichmann SA. An organotypic atlas of human vascular cells. Nat Med. 2024 Dec;30(12):3468-3481.
705 47. Zhang B, He P, Lawrence JEG, Wang S, Tuck E, Williams BA, Roberts K, Kleshchevnikov V, Mamanova
706 L, Bolt L, Polanski K, Li T, Elmentaite R, Fasouli ES, Prete M, He X, Yayon N, Fu Y, Yang H, Liang C, Zhang H,
707 Blain R, Chedotal A, FitzPatrick DR, Firth H, Dean A, Bayraktar OA, Marioni JC, Barker RA, Storer MA, Wold BJ,
708 Zhang H, Teichmann SA. A human embryonic limb cell atlas resolved in space and time. Nature. 2024
709 Nov;635(8039):668-678.
710 48. Nguyen H, Tran D, Tran B, Pehlivan B, Nguyen T. A comprehensive survey of regulatory network
711 inference methods using single cell RNA sequencing data. Brief Bioinform. 2021 May 20;22(3):bbaa190.

712 49. Lu Z, Hong CC, Kong G, Assumpção A, Ong IM, Bresnick EH, Zhang J, Pan X. Polycomb Group Protein
713 YY1 Is an Essential Regulator of Hematopoietic Stem Cell Quiescence. Cell Rep. 2018;22(6):1545-59.
714 50. Fisher JB, Peterson J, Reimer M, Stelloh C, Pulakanti K, Gerbec ZJ, Abel AM, Strouse JM, Strouse C,
715 McNulty M, Malarkannan S, Crispino JD, Milanovich S, Rao S. The cohesin subunit Rad21 is a negative regulator
716 of hematopoietic self-renewal through epigenetic repression of Hoxa7 and Hoxa9. Leukemia. 2017;31(3):712-9.
717 51. Kumar P, Zhang N, Lee J, Cheng H, Kurtz K, Conneely SE, Sasidharan R, Rau RE, Pati D. Cohesin
718 Subunit RAD21 Regulates the Differentiation and Self-Renewal of Hematopoietic Stem and Progenitor Cells. Stem
719 Cells. 2023;41(10):971-85.
720 52. Ning S, Pagano JS, Barber GN. IRF7: activation, regulation, modification and function. Genes & Immunity.
721 2011;12(6):399-414.
722 53. Fischer J, Walter C, Tönges A, Aleth H, Jordão MJC, Leddin M, Gröning V, Erdmann T, Lenz G, Roth J,
723 Vogl T, Prinz M, Dugas M, Jacobsen ID, Rosenbauer F. Safeguard function of PU.1 shapes the inflammatory
724 epigenome of neutrophils. Nature Immunology. 2019;20(5):546-58.
725 54. 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
726 2 (HDAC2) attenuates lipopolysaccharide (LPS)-induced inflammation by regulating PAI-1 expression. Journal of
727 Inflammation. 2018;15(1):3.
728 55. Ptasinska A, Pickin A, Assi SA, Chin PS, Ames L, Avellino R, Gröschel S, Delwel R, Cockerill PN, Osborne
729 CS, Bonifer C. RUNX1-ETO Depletion in t(8;21) AML Leads to C/EBPα- and AP-1-Mediated Alterations in
730 Enhancer-Promoter Interaction. Cell Reports. 2019;28(12):3022-31.e7.
731 56. Martinez-Soria N, McKenzie L, Draper J, Ptasinska A, Issa H, Potluri S, Blair HJ, Pickin A, Isa A, Chin PS,
732 Tirtakusuma R, Coleman D, Nakjang S, Assi S, Forster V, Reza M, Law E, Berry P, Mueller D, Osborne C, Elder
733 A, Bomken SN, Pal D, Allan JM, Veal GJ, Cockerill PN, Wichmann C, Vormoor J, Lacaud G, Bonifer C, Heidenreich
734 O. The Oncogenic Transcription Factor RUNX1/ETO Corrupts Cell Cycle Regulation to Drive Leukemic
735 Transformation. Cancer Cell. 2018;34(4):626-42.e8.
736 57. Patrick R, Naval-Sanchez M, Deshpande N, Huang Y, Zhang J, Chen X, Yang Y, Tiwari K, Esmaeili M,
737 Tran M, Mohamed AR, Wang B, Xia D, Ma J, Bayliss J, Wong K, Hun ML, Sun X, Cao B, Cottle DL, Catterall T,
738 Barzilai-Tutsch H, Troskie RL, Chen Z, Wise AF, Saini S, Soe YM, Kumari S, Sweet MJ, Thomas HE, Smyth IM,
739 Fletcher AL, Knoblich K, Watt MJ, Alhomrani M, Alsanie W, Quinn KM, Merson TD, Chidgey AP, Ricardo SD, Yu
740 D, Jardé T, Cheetham SW, Marcelle C, Nilsson SK, Nguyen Q, White MD, Nefzger CM. The activity of early-life
741 gene regulatory elements is hijacked in aging through pervasive AP-1-linked chromatin opening. Cell Metab.
742 2024;36(8):1858-81.e23.
743 58. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression
744 analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139-40.
745 59. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with
746 DESeq2. Genome Biology. 2014;15(12):550.
747 60. Holmfeldt P, Ganuza M, Marathe H, He B, Hall T, Kang G, Moen J, Pardieck J, Saulsberry AC, Cico A,
748 Gaut L, McGoldrick D, Finkelstein D, Tan K, McKinney-Freeman S. Functional screen identifies regulators of
749 murine hematopoietic stem cell repopulation. J Exp Med. 2016;213(3):433-49.
750 61. Chen C, Yu W, Tober J, Gao P, He B, Lee K, Trieu T, Blobel GA, Speck NA, Tan K. Spatial Genome Re-
751 organization between Fetal and Adult Hematopoietic Stem Cells. Cell Rep. 2019;29(12):4200-11.e7.
752 62. Unnikrishnan A, Papaemmanuil E, Beck D, Deshpande NP, Verma A, Kumari A, Woll PS, Richards LA,
753 Knezevic K, Chandrakanthan V, Thoms JAI, Tursky ML, Huang Y, Ali Z, Olivier J, Galbraith S, Kulasekararaj AG,
754 Tobiasson M, Karimi M, Pellagatti A, Wilson SR, Lindeman R, Young B, Ramakrishna R, Arthur C, Stark R, Crispin
755 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.

63. 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.

64. 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.

65. 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.

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

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

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. Körber V, Stainczyk SA, Kurilov R, Henrich KO, Hero B, Brors B, Westermann F, Höfer T. Neuroblastoma arises in early fetal development and its evolutionary duration predicts outcome. Nat Genet. 2023 Apr;55(4):619-630.

75. 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.

76. 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.

77. 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

801 histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature.
802 2016;537(7621):548-52.
803 78. Nicosia L, Spencer GJ, Brooks N, Amaral FMR, Basma NJ, Chadwick JA, Revell B, Wingelhofer B,
804 Maiques-Diaz A, Sinclair O, Camera F, Ciceri F, Wiseman DH, Pegg N, West W, Knurowski T, Frese K, Clegg K,
805 Campbell VL, Cavet J, Copland M, Searle E, Somervaille TCP. Therapeutic targeting of EP300/CBP by
806 bromodomain inhibition in hematologic malignancies. Cancer Cell. 2023;41(12):2136-53.e13.
807 79. Crowley S, White LS, Li Y, Yang W, Magee JA, Bednarski JJ. Bclaf1 promotes hematopoietic stem cell
808 repopulating capacity and self-renewal. The Journal of Immunology. 2022;208(1_Supplement):47.01-47.01.
809 80. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome
810 Biology. 2018;19(1):15.
811 81. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell
812 Transcriptomic Data. Cell Systems. 2019;8(4):281-91.e9.
813 82. Heumos L, Schaar A, Consortium S-CBP. Single-cell best practices 2023. Available from: https://www.sc-
814 best-practices.org/preamble.html#.
815 83. Büttner M, Miao Z, Wolf FA, Teichmann SA, Theis FJ. A test metric for assessing single-cell RNA-seq
816 batch correction. Nature Methods. 2019;16(1):43-9.
817 84. Domínguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, Howlett SK, Suchanek O,
818 Polanski K, King HW, Mamanova L, Huang N, Szabo PA, Richardson L, Bolt L, Fasouli ES, Mahbubani KT, Prete
819 M, Tuck L, Richoz N, Tuong ZK, Campos L, Mousa HS, Needham EJ, Pritchard S, Li T, Elmentaite R, Park J,
820 Rahmani E, Chen D, Menon DK, Bayraktar OA, James LK, Meyer KB, Yosef N, Clatworthy MR, Sims PA, Farber
821 DL, Saeb-Parsy K, Jones JL, Teichmann SA. Cross-tissue immune cell analysis reveals tissue-specific features in
822 humans. Science. 2022;376(6594):eabl5197.
823 85. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte
824 AJ, Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic
825 macrophage. Nat Immunol. 2019;20(2):163-72.
826 86. Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, Habib N, Yosef N,
827 Chang CY, Shay T, Frampton GM, Drake AC, Leskov I, Nilsson B, Preffer F, Dombkowski D, Evans JW, Liefeld T,
828 Smutko JS, Chen J, Friedman N, Young RA, Golub TR, Regev A, Ebert BL. Densely interconnected transcriptional
829 circuits control cell states in human hematopoiesis. Cell. 2011 Jan 21;144(2):296-309.
830 87. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific
831 marker combinations from single-cell transcriptomic data. Nature Communications. 2022;13(1):1246.
832 88. 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.
833 clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3).