A stem cell aging clock links biological age deviation to clinical outcome in acute myeloid leukemia

Hongcheng Chen[1], Pengwei Dong[1], Jianing Xu[1], and Guanlin Wang[1, *]

[1]Shanghai Key Laboratory of Metabolic Remodelling and Health, Institute of Metabolism and Integrative Biology, School of Basic Medical Sciences, Centre for Evolutionary Biology, Fudan University, Shanghai, 200438, China

*Corresponding author: Guanlin Wang, e-mail: guanlin_wang@fudan.edu.cn

Abstract

Aging of hematopoietic stem and progenitor cells (HSPCs) impairs regenerative capacity and predisposes to hematological diseases. Here, we constructed a comprehensive single-cell transcriptomic atlas comprising 186,123 $CD34^+$ HSPCs spanning early prenatal development (6 post-conception weeks) to late adulthood (74 years). We identified two conserved core molecular programs (MPs) of inflammaging and RNA splicing / protein homeostasis. Leveraging these programs, we developed a machine learning-based stem cell aging clock from 84 donors. Applying this clock to acute myeloid leukemia (AML), we define Transcriptional Age Deviation (TAD), a novel metric of biological age divergence. We found that a biologically "younger" state (low TAD), reflecting oncofetal reprogramming, is a powerful independent predictor of poor survival in two large AML cohorts. Low TAD was associated with high-risk genetics and therapy resistance, and critically, it re-stratified patient outcomes within each ELN 2022 risk category. Our work establishes a quantitative link between stem cell aging biology and AML prognosis, offering a robust tool to refine clinical risk assessment.

Introduction

Hematopoietic stem and progenitor cells (HSPCs) serve as the foundation of lifelong hematopoiesis, sustaining hematopoietic homeostasis and immune competence throughout the human lifespan[1–4]. However, aging imposes a progressive decline in HSPC function, leading to reduced regenerative capacity, lineage skewing toward myeloid differentiation, accumulation of DNA damage and increased susceptibility to hematological disorders such as acute leukemia[5,6]. Notably, the aged hematopoietic system contributes to systemic aging across multiple organs, as demonstrated by heterochronic parabiosis experiments showing that young blood can rejuvenate aged tissues[7], highlighting the central role of HSPCs in organismal aging. Multiple mechanistic studies in mouse have identified several key drivers and pathways involved in HSPC aging, including increased inflammatory exposure (inflammaging) and disturbance of RNA splicing[8–13]. However, the extent to which these mechanisms operate in human HSPC aging remains to be fully elucidated.

Understanding the transcriptional programs governing both physiological aging and pathological transformation in HSPCs is crucial for identifying molecular vulnerabilities underlying age-related hematopoietic decline and malignant progression[14]. Studying human HSPC aging remains challenging due to the inherent heterogeneity of HSPCs and limited access to human samples across the lifespan. Traditional bulk RNA sequencing approaches fail to capture the complexity of individual HSPC gene expression profiles, missing subtle but biologically meaningful age-related transcriptional shifts. The advent of single-cell RNA sequencing (scRNA-seq) technologies has revolutionized this field by enabling high-resolution profiling of HSPC subpopulations across human development and aging, revealing a higher proportion of lymphoid-biased cells during early development, followed by a progressive shift toward myeloid-predominant hematopoiesis in the bone marrow niche during aging[15–19]. Nevertheless, the regulatory programs driving age-associated HSPC heterogeneity and functions remain incompletely understood.

Recent studies have introduced 'aging clocks' in various tissues[20–23], but it remains unclear whether a transcriptomics-based stem-cell-specific clock can be robustly applied across the human lifespan, from prenatal to elderly age. More importantly, while clinical risk stratification in acute myeloid leukemia (AML), such as the ELN 2022 guidelines, has improved, significant heterogeneity remains within each risk group[24–27]. There is a critical need for novel, biologically-grounded biomarkers that can capture this heterogeneity and provide additional

62 prognostic value. It is unknown whether a metric derived from the fundamental process of stem
63 cell aging could fulfill this clinical need.
64
65 In this study, we constructed a comprehensive single-cell transcriptomic atlas of human $CD34^+$
66 HSPCs spanning developmental stages from early prenatal development (6 PCW) to late
67 adulthood (74 years) across 28 healthy donors. Molecular characterization revealed six
68 conserved aging-associated molecular programs (MPs) representing distinct transcriptional
69 signatures related to inflammatory signaling, HSC identity, transcription signaling, aerobic
70 respiration, RNA splicing and proteostasis, and cell cycle regulation. Leveraging these
71 programs, we developed a machine learning-based stem cell aging clock that robustly predicts
72 donor chronological age from single-cell HSC transcriptomes and quantifies biological age
73 from 84 donors. Applying this framework to AML, we defined transcriptional age deviation
74 (TAD) as the age deviation of leukemic stem cells relative to a healthy baseline. TAD reveals
75 disease-associated remodeling of stem cell aging states that extends beyond intratumoral
76 heterogeneity, capturing clinically meaningful variation linked to genetic risk stratification and
77 patient outcome. Collectively, our work provides a reference map of human HSPC aging,
78 establishes a stem cell aging clock that enables quantitative integration of age-associated stem
79 cell states at single-cell resolution, and provides a framework for uncovering aging-related
80 dysregulation linked to disease outcome in hematological disorders.
81
82
83 **Results**
84 **A single-cell transcriptomic atlas of HSPCs across human lifespan**
85 To systematically characterize age-related dynamics in human hematopoiesis, we integrated
86 six previously published datasets of $CD34^+$ hematopoietic stem and progenitor cells (HSPCs),
87 comprising a total of 200,665 cells from 28 healthy donors aged from post-conception week
88 (PCW) 6 to 74 years (**Figure 1A-B**)[5,17,44–47]. To verify sample integrity, donor sex was
89 confirmed based on expression of sex-chromosome markers *XIST* (female) and *UTY* (male)
90 (**Supplemental Figure 1C**). After stringent quality control (**Supplemental Figure 1A-B**,
91 **Supplemental Methods**), 186,123 high-quality cells were retained for downstream analyses
92 (**Figure 1C-D**, **Supplemental Figure 1D-G**, **Supplemental Tables 1 and 2**). From this atlas,
93 we isolated 59,723 high-confidence hematopoietic stem cells and multipotent progenitors
94 (HSC/MPPs) based on unsupervised clustering and a validated gene signature score (**Figure**
95 **1E**)[17]. The primitive, undifferentiated state of this population was confirmed by differentiation

potential (CytoTRACE2) and trajectory analyses (StaVIA, **Supplemental Figure 2A-C**, **Methods**)[31,32].

Six age-related molecular programs define HSC/MPP functional states

To delineate the age-related transcriptional reprogramming of the refined HSC/MPP population, we conducted UMAP projection, differential expression, and pySCENIC regulon analysis (**Supplemental Figure 2D-H**, **Supplemental Methods**)[30,48]. The analysis revealed a sharp age-dependent transcriptional and regulatory rewiring of HSC/MPP. Proliferative, MYC/E2F-dominated transcription dominates early fetal cells, whereas adult HSC/MPPs exhibit a quiescent, myeloid-primed state marked by heightened NF-κB inflammatory regulons (e.g. NFKB2, ERG), capturing the core transition from development to inflammaging in human hematopoiesis.

To delineate the latent transcriptional architecture of HSC/MPP aging beyond single gene alteration, we employed consensus non-negative matrix factorization (cNMF)[49] to identify molecular programs (MPs) representing underlying functional states. We calculated six robust age-related MPs from our HSC/MPP discovery cohort and an independent validation cohort[16] (**Figure 2A-B, Supplemental Figure 3A-B, Supplemental Table 4, Supplemental Methods**). Pathway analysis showed that the six MPs map to distinct functions, with MP1 (most significantly upregulated) reflecting pro-inflammatory NF-κB–driven "inflammaging," MP5 (most significantly downregulated) characterizing RNA-splicing/proteostasis, and the remaining modules corresponding to HSC-identity, transcription-signal, mitochondrial respiration, and cell-cycle processes (**Figure 2C-D, Supplemental Table 5**).

Age-associated dynamics of molecular programs in HSC/MPPs

To systematically explore age-associated changes in molecular programs in refined HSC/MPPs, we assessed the distribution and activity of MPs across five developmental stages and major hematopoietic lineages (**Figure 3A-B**, **Supplemental Figure 3C-D**). MP1-3 were predominantly enriched in adult ages, particularly within the uncommitted HSC/MPP and the myeloid compartment, suggesting progressive activation of myeloid-biased inflammaging pathways with age. In contrast, MP4-6 showed pronounced enrichment during early development (early fetal and fetal stages) but markedly decreased in postnatal and diminished in adult stages with modest variation across lineage contexts, indicating developmental-

specific programs associated with aerobic respiration, transcription regulation, and proliferation that are attenuated after birth.

We next examined transcription factor (TF) activity within cells assigned to MP1 and MP5 modules, stratified by developmental stage (**Figure 3C**). fetal MP5 is driven by proliferative regulators (MYC, E2F1, BRCA1) [50–52], whereas in adults it switches to HSC-maintenance TFs (MYB, SPI1, GFI1) [53–55]; simultaneously, adult MP1 becomes dominated by inflammatory AP-1 and NF-κB regulons (e.g., FOS, JUND, NFKB2), underscoring distinct and dynamically evolving regulatory networks of MP1 and MP5.

To identify the clinical relevance of MPs, we conducted survival analysis on the developmental program MP5. We found that an 18-gene MP5 signature was associated with poor survival in two large AML cohorts (log-rank $P < 0.005$ for both cohorts, **Figure 3D-E**, **Supplemental Figure 3E-F**, **Supplemental Table 6**, **Supplemental Methods**). This motivated us to create a more sophisticated model integrating multiple aging programs to quantify biological age.

A machine learning-derived transcriptomic clock for human HSC/MPP aging

To develop a robust stem cell aging clock based on HSC/MPP transcriptomes, we expanded the analysis with our refined dataset (59,723 HSC/MPPs from 28 donors) with two independent published cohorts (17,211 HSC/MPPs from 24 donors; 4,004 HSC/MPPs from 32 donors, respectively)[16,33], forming the largest human stem cell aging cohort comprising of 84 donors (80,938 cells). This expands cohort increased donor diversity and statistical power across the human lifespan. To reduce technical noise while preserving biological variability, we implemented a "metacell" strategy that 20 randomly sampled HSC/MPP cells from each donor were aggregated to generate one metacell. This resulted in a total of 4,088 metacells spanning ages from 4 PCW to 77 years (**Figure 4A**). Donor's chronological age was used as the ground truth label for model training.

We prioritized MP1 and MP5 for feature selection given their strongest associations with age, and extracted 55 informative genes used as input features for model construction (**Supplemental Tables 2 and 7**). We next trained and benchmarked six supervised machine learning models using a 10-fold cross-validation strategy, with our dataset as discovery dataset and the two independent cohorts as external validation datasets (**Methods**). While several models achieved similarly high Pearson correlation coefficients (PCCs) on the discovery and

163 validation datasets (PCC = 0.99), MLPRegressor consistently demonstrated superior
164 generalization performance, maintaining the highest PCC on the validation dataset (0.76). In
165 contrast, XGBoost and Random Forest showed a pronounced reduction in validation PCC
166 (≈0.18-0.19), indicating limited generalizability. Therefore, we selected MLPRegressor as our
167 final model (**Figure 4B, Supplemental Figure 4A-B**). We retrained the final model using all
168 three cohorts, resulting in a global stem cell aging clock with an MAE of 3.13 and a PCC of
169 0.99 (**Figure 4C**). To further interpret the molecular drivers underlying model predictions, we
170 calculated permutation-importance scores for each feature in the final model, identifying genes
171 such as *AREG*, *CRHBP* and *GNG11* as the most influential predictors, suggesting their
172 relevance to age-related transcriptional reprogramming (**Supplemental Figure 4C**). Moreover,
173 to systematically assess the contribution of each MP, we trained and compared separate models
174 using genes derived from each MP and MP1 + MP5. This analysis confirmed that [MP1 + MP5]
175 based models consistently outperformed those derived from other modules, highlighting the
176 key role of these modules in defining HSC/MPP age (**Figure 4D**).
177
178 **TAD reveals biologically distinct and clinically relevant states in AML stem cells**
179 To investigate whether hematological malignancies are associated with altered HSC/MPP
180 biological aging, we applied the age prediction model to a large scRNA-seq cohort of acute
181 myeloid leukemia (AML), comprising 90 patients of 73,360 single HSC/MPPs[43]
182 (**Supplemental Table 8**). Cells were aggregated into 3,715 HSC/MPP metacells and we
183 calculated the transcriptional age deviation (TAD) for each metacell after applying correction
184 to the chronological age-dependent bias (**Methods**). Metacells within the lowest and highest
185 10% of TAD were annotated as low-TAD and high-TAD groups, respectively (low-TAD group:
186 TAD < -9.5, n = 334 metacells; high-TAD group: TAD > 4.8, n = 272 metacells). Differential
187 expression analysis identified 2483 genes up-regulated in low-TAD group and 366 genes up-
188 regulated in high-TAD group.
189
190 High-TAD group were dominated by pro-inflammatory and stress-response programs,
191 including TNF-α signalling via NF-κB, IL-2/STAT5 signaling and apoptosis. Low-TAD cells
192 were selectively enriched for aerobic respiration and primary cilium assembly (adjusted $P <$
193 0.05, **Figure 4E**). The enrichment of aerobic-respiration pathways in the low-TAD group
194 mirrors the MP4 activity characteristic of fetal and early-postnatal HSCs. Notably, upregulated
195 mitochondrial oxidative metabolism has been linked to the enhanced proliferative capacity of

196 early HSCs and to drug resistance in AML[56–58]. In addition, primary cilia are reported essential
197 for primitive hematopoiesis in both zebrafish and human and dysregulated in AML[59,60].
198
199 Strikingly, this biologically defined TAD showed a strong correlation with established clinical
200 risk. We observed a stepwise decrease in TAD values across European LeukemiaNet (ELN)
201 2022 risk groups, with adverse-risk patients exhibiting significantly lower TAD than favorable-
202 risk patients (Kruskal–Wallis $P = 1.4 \times 10^{-3}$, **Supplemental Figure 4D**). This initial finding
203 suggested that a biologically "younger" stem cell state is associated with higher-risk disease
204 genetics.
205
206 **TAD provides independent and additive prognostic value in bulk AML cohorts**
207 To assess the clinical utility and broad applicability of TAD, we extended our analysis to two
208 large-scale bulk RNA-seq cohorts, TCGA-LAML (n=141) and BeatAML (n=451). Across both
209 cohorts, TAD powerfully stratified patient survival. Patients with low TAD exhibited
210 dramatically shorter overall survival (OS) compared to those with high TAD (TCGA-LAML:
211 median OS 8.1 vs 15.2 months; BeatAML: median OS 7.0 vs 16.4 months; log-rank $P < 0.001$
212 in both cohorts; **Figure 5A-B**). Importantly, in a multivariate Cox regression model on the
213 BeatAML cohort, after adjusting for established risk factors including age (⩾60 vs <60 years)
214 and ELN 2022 risk groups, TAD remained a highly significant, independent predictor of OS
215 (HR = 0.95, 95% CI: 0.91-0.98, $P < 0.005$). These results firmly establish TAD as a prognostic
216 marker with power extending beyond current clinical standards.
217
218 We next investigated whether TAD could resolve the known heterogeneity within ELN 2022
219 risk groups in the BeatAML cohort. Kaplan-Meier analysis revealed that TAD robustly re-
220 stratified patients into distinct survival groups within each ELN category (**Figure 5C-E**). This
221 effect was most pronounced in the clinically challenging intermediate-risk group, where
222 patients with low TAD had a dismal median OS of just 8.6 months, compared to 18.4 months
223 for their high-TAD counterparts (log-rank $P < 0.001$). This powerful re-stratification ability
224 was also observed in the favorable-risk (median OS 19.6 vs 30.7 months, $P = 0.001$) and
225 adverse-risk groups (median OS 5.4 vs 10.2 months, $P = 0.003$).
226
227 Furthermore, TAD demonstrated a significant ability to predict treatment response. While
228 complete remission (CR) or CR with incomplete hematologic recovery (CRi) rates expectedly
229 decreased across ELN favorable (73%), intermediate (49%), and adverse (41%) risk groups

230 (**Figure 5F**), TAD provided additional predictive resolution. Within each ELN category,
231 patients with low TAD consistently exhibited lower CR/CRi rates compared to their high-TAD
232 counterparts (**Figure 5G**). This re-stratification was statistically significant in both the
233 favorable-risk (high-TAD: 85%, low-TAD: 59%; chi-squared test, $P = 0.009$) and the clinically
234 challenging intermediate-risk groups (high-TAD: 59%, low-TAD: 34%; $P = 0.002$). A similar
235 trend was observed in the adverse-risk group (high-TAD: 51%, low-TAD: 35%; $P = 0.063$),
236 collectively highlighting the potential of TAD to identify individuals unlikely to respond to
237 standard therapy across the entire risk spectrum.
238
239 Collectively, these results establish TAD not just as a correlative biomarker, but as a
240 quantitative metric reflecting fundamental aging-related cellular states that are co-opted in
241 malignancy, drive clinical outcomes, and offer a valuable new dimension for risk assessment
242 in AML.
243
244
245 **Discussion**
246 In this study, we leveraged a comprehensive lifespan atlas of human hematopoiesis to develop
247 a stem-cell-specific aging clock, leading to the definition of transcriptional age deviation
248 (TAD), a quantitative biomarker with profound clinical implications for acute myeloid
249 leukemia (AML). Our central finding is that TAD, which measures the degree of oncofetal
250 reprogramming in cancer cells, serves as a powerful and independent predictor of patient
251 survival and therapy response. We demonstrated that TAD not only stands on its own as a
252 prognostic marker but also provides significant additive value to the current gold-standard ELN
253 2022 risk classification. Specifically, TAD effectively re-stratifies the clinically ambiguous
254 intermediate-risk group, identifying a high-risk subset with dismal survival outcomes. This
255 work bridges the fundamental biology of HSPC aging with the urgent clinical need for
256 improved risk assessment in AML.
257
258 Our analysis provides a compelling biological explanation for the paradoxical "younger-is-
259 worse" paradigm observed with TAD. The low-TAD state is not a sign of cellular health, but
260 rather a hallmark of developmental reprogramming, where leukemic cells reactivate
261 transcriptional programs characteristic of fetal hematopoiesis. These programs, particularly
262 those related to aerobic respiration, RNA splicing, and proteostasis (as captured by our MP4
263 and MP5 modules), confer a highly proliferative, chemo-resistant phenotype optimized for

264 malignant expansion. This finding aligns with the established concept that reactivation of
265 developmental pathways is a common strategy for tumorigenesis[61–63]. Unlike traditional LSC
266 markers (e.g., CD34, CD123) that define cellular identity[64–66], TAD acts as a functional state
267 descriptor, quantifying a cell's position on a continuous aging-developmental axis and thereby
268 offering an orthogonal layer of prognostic information.
269
270 The foundation of our clinically relevant biomarker is a meticulously constructed atlas of over
271 186,000 human HSPCs, spanning from early fetal development to late adulthood. This resource
272 allowed us to delineate six robust molecular programs (MPs) governing HSPC function across
273 the human lifespan. We identified two core, conserved aging programs—MP1 (inflammaging)
274 and MP5 (RNA splicing and proteostasis)—whose opposing dynamics from birth to old age
275 encapsulate the key trade-offs in HSC biology. The progressive accumulation of MP1 activity
276 reflects the well-documented pro-inflammatory drift in the aged bone marrow niche, a known
277 driver of myeloid skewing and clonal hematopoiesis[8–10,67]. Conversely, the sharp decline of the
278 MP5 program after birth highlights a critical developmental window for RNA splicing and
279 proteostasis, pathways essential for maintaining HSC fitness but which are co-opted in
280 malignancy[11–13,68–71]. Our ability to build a highly accurate aging clock (MAE = 3.13 years)
281 from just 55 genes derived from these two core programs underscores their central role in
282 defining stem cell age.
283
284 Our study has several limitations. First, while our cohort of 84 donors is the largest of its kind,
285 representation at extreme ages remains limited. Second, our analysis is based on transcriptomic
286 data alone; future integration with epigenomic and proteomic data will provide a more
287 comprehensive view of HSC aging and its dysregulation in disease. Third, while we validated
288 TAD in large retrospective cohorts, its predictive utility should be prospectively evaluated in
289 clinical trials to determine its stability and role in guiding treatment decisions.
290
291 In conclusion, our work provides a foundational map of hematopoietic aging and establishes
292 TAD as a robust, clinically translatable biomarker. By quantifying the deviation from a healthy
293 aging trajectory, we have created a powerful new tool for precision oncology that converts a
294 complex biological signature into a single, interpretable metric. The strong performance of
295 TAD in bulk RNA-seq cohorts highlights its potential for immediate implementation in clinical
296 settings. This study demonstrates that a deep understanding of fundamental aging biology can

297 unlock novel strategies for stratifying disease risk and personalizing therapy in AML and
298 potentially other hematologic malignancies.
299
300
301 **Authorship**
302 Contribution: Conceptualization: G.W. & H.C.; Methodology: G.W. & H.C.; Formal analysis:
303 H.C. & P.D.; Investigation: H.C. & J.X.; Data curation: H.C.; Writing: G.W. & H.C.;
304 Visualization: H.C.; Supervision: G.W.; Funding acquisition: G.W.
305 Conflict of interest disclosure: All authors declare no conflict of interest.
306
307
308 **References**

309 1. Seita J, Weissman IL. Hematopoietic Stem Cell: Self-renewal versus Differentiation.
310 *Wiley Interdiscip Rev Syst Biol Med*. 2010;2(6):640-653. doi:10.1002/wsbm.86
311 2. Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation
312 landscapes. *Nature*. 2018;553(7689):418-426. doi:10.1038/nature25022
313 3. Wilson A, Laurenti E, Oser G, et al. Hematopoietic stem cells reversibly switch from
314 dormancy to self-renewal during homeostasis and repair. *Cell*. 2008;135(6):1118-1129.
315 doi:10.1016/j.cell.2008.10.048
316 4. Kasbekar M, Mitchell CA, Proven MA, Passegué E. Hematopoietic stem cells through
317 the ages: A lifetime of adaptation to organismal demands. *Cell Stem Cell*.
318 2023;30(11):11. doi:10.1016/j.stem.2023.09.013
319 5. Ainciburu M, Ezponda T, Berastegui N, et al. Uncovering perturbations in human
320 hematopoiesis associated with healthy aging and myeloid malignancies at single-cell
321 resolution. Shim J, Banerjee U, Kim JK, eds. *eLife*. 2023;12:e79363.
322 doi:10.7554/eLife.79363
323 6. Zhou FJJ, Wang HC, Le MK, et al. Age-Related Changes in Hematopoietic Stem Cell
324 Proteostasis Promote the Emergence of Clonal Hematopoiesis and Leukemia in Older
325 Adults. *Blood*. 2024;144(Supplement 1):4047-4047. doi:10.1182/blood-2024-202652
326 7. Conboy IM, Conboy MJ, Wagers AJ, Girma ER, Weissman IL, Rando TA. Rejuvenation
327 of aged progenitor cells by exposure to a young systemic environment. *Nature*.
328 2005;433(7027):760-764. doi:10.1038/nature03260

329 8. Bogeska R, Mikecin AM, Kaschutnig P, et al. Inflammatory exposure drives long-lived
330 impairment of hematopoietic stem cell self-renewal activity and accelerated aging. *Cell*
331 *Stem Cell*. 2022;29(8):1273-1284.e8. doi:10.1016/j.stem.2022.06.012
332 9. Kovtonyuk LV, Caiado F, Garcia-Martin S, et al. IL-1 mediates microbiome-induced
333 inflammaging of hematopoietic stem cells in mice. *Blood*. 2022;139(1):44-58.
334 doi:10.1182/blood.2021011570
335 10. Esplin BL, Shimazu T, Welner RS, et al. Chronic exposure to a TLR ligand injures
336 hematopoietic stem cells. *J Immunol*. 2011;186(9):5367-5375.
337 doi:10.4049/jimmunol.1003438
338 11. Cesana M, Guo MH, Cacchiarelli D, et al. A CLK3-HMGA2 alternative splicing axis
339 impacts human hematopoietic stem cell molecular identity throughout development. *Cell*
340 *Stem Cell*. 2018;22(4):575-588.e7. doi:10.1016/j.stem.2018.03.012
341 12. Goff DJ, Court Recart A, Sadarangani A, et al. A Pan-BCL2 inhibitor renders bone-
342 marrow-resident human leukemia stem cells sensitive to tyrosine kinase inhibition. *Cell*
343 *Stem Cell*. 2013;12(3):316-328. doi:10.1016/j.stem.2012.12.011
344 13. van Galen P, Kreso A, Mbong N, et al. The unfolded protein response governs integrity
345 of the haematopoietic stem-cell pool during stress. *Nature*. 2014;510(7504):268-272.
346 doi:10.1038/nature13228
347 14. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: An
348 expanding universe. *Cell*. 2023;186(2):243-278. doi:10.1016/j.cell.2022.11.001
349 15. Velten L, Haas SF, Raffel S, et al. Human haematopoietic stem cell lineage commitment
350 is a continuous process. *Nat Cell Biol*. 2017;19(4):271-281. doi:10.1038/ncb3493
351 16. Li H, Côté P, Kuoch M, et al. The dynamics of hematopoiesis over the human lifespan.
352 *Nat Methods*. Published online December 5, 2024:1-13. doi:10.1038/s41592-024-02495-
353 0
354 17. Roy A, Wang G, Iskander D, et al. Transitions in lineage specification and gene
355 regulatory networks in hematopoietic stem/progenitor cells over human development.
356 *Cell Reports*. 2021;36(11):11. doi:10.1016/j.celrep.2021.109698
357 18. Kowalczyk MS, Tirosh I, Heckl D, et al. Single-cell RNA-seq reveals changes in cell
358 cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res*.
359 2015;25(12):1860-1872. doi:10.1101/gr.192237.115
360 19. Zhang Y, Xie X, Huang Y, et al. Temporal molecular program of human hematopoietic
361 stem and progenitor cells after birth. *Developmental Cell*. 2022;57(24):24.
362 doi:10.1016/j.devcel.2022.11.013

20. Li W, Zhang Z, Kumar S, et al. Single-cell immune aging clocks reveal inter-individual heterogeneity during infection and vaccination. *Nat Aging*. 2025;5(4):607-621. doi:10.1038/s43587-025-00819-z
21. Cui L, Nie X, Guo Y, et al. Single-cell transcriptomic atlas of the human testis across the reproductive lifespan. *Nat Aging*. 2025;5(4):658-674. doi:10.1038/s43587-025-00824-2
22. Zhu H, Chen J, Liu K, et al. Human PBMC scRNA-seq–based aging clocks reveal ribosome to inflammation balance as a single-cell aging hallmark and super longevity. *Sci Adv*. 2023;9(26):26. doi:10.1126/sciadv.abq7599
23. Cao D, Liu Y, Cheng Y, et al. Time-series single-cell transcriptomic profiling of luteal-phase endometrium uncovers dynamic characteristics and its dysregulation in recurrent implantation failures. *Nat Commun*. 2025;16(1):1. doi:10.1038/s41467-024-55419-z
24. Mrózek K, Kohlschmidt J, Blachly JS, et al. Outcome prediction by the 2022 European LeukemiaNet genetic-risk classification for adults with acute myeloid leukemia: an Alliance study. *Leukemia*. 2023;37(4):788-798. doi:10.1038/s41375-023-01846-8
25. Song GY, Kim HJ, Kim T, et al. Validation of the 2022 European LeukemiaNet risk stratification for acute myeloid leukemia. *Sci Rep*. 2024;14(1):8517. doi:10.1038/s41598-024-57295-5
26. Mudra SE, Sackstein P, Williams LS, et al. Real-World Validation of the European Leukemia Network 2022 Risk Stratification in Acute Myelogenous Leukemia. *Blood*. 2023;142(Supplement 1):4297. doi:10.1182/blood-2023-189309
27. Zhao EJ, Narayanan S, Kuchenbauer F, et al. Performance of the ELN 2022 risk stratification in a real-world Canadian AML patient population. *Blood*. 2025;146(Supplement 1):5178. doi:10.1182/blood-2025-5178
28. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16(12):12. doi:10.1038/s41592-019-0619-0
29. Dai M, Pei X, Wang XJ. Accurate and fast cell marker gene identification with COSG. *Briefings in Bioinformatics*. 2022;23(2):2. doi:10.1093/bib/bbab579
30. Aibar S, González-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods*. 2017;14(11):11. doi:10.1038/nmeth.4463
31. Kang M, Gulati GS, Brown EL, et al. Improved reconstruction of single-cell developmental potential with CytoTRACE 2. *Nat Methods*. Published online October 27, 2025:1-6. doi:10.1038/s41592-025-02857-2

32. Stassen SV, Kobashi M, Lam EY, Huang Y, Ho JWK, Tsia KK. StaVia: spatially and temporally aware cartography with higher-order random walks for cell atlases. *Genome Biol*. 2024;25:224. doi:10.1186/s13059-024-03347-y
33. Wang Z, Wu Z, Wang H, et al. An immune cell atlas reveals the dynamics of human macrophage specification during prenatal development. *Cell*. 2023;186(20):20. doi:10.1016/j.cell.2023.08.019
34. Hsiao LL, Dangond F, Yoshida T, et al. A compendium of gene expression in normal human tissues. *Physiol Genomics*. 2001;7(2):97-104. doi:10.1152/physiolgenomics.00040.2001
35. Hounkpe BW, Chenou F, de Lima F, De Paula EV. HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Res*. 2021;49(D1):D947-D955. doi:10.1093/nar/gkaa609
36. Buckley MT, Sun ED, George BM, et al. Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. *Nat Aging*. 2023;3(1):1. doi:10.1038/s43587-022-00335-4
37. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*. 2010;33:1-22. doi:10.18637/jss.v033.i01
38. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *Royal Statistical Society Journal Series B: Methodological*. 1996;58(1):267-288. doi:10.1111/j.2517-6161.1996.tb02080.x
39. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. *J R Stat Soc Ser B Stat Methodol*. 2005;67(2):301-320. doi:10.1111/j.1467-9868.2005.00503.x
40. Ho TK. Random decision forests. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol 1. 1995:278-282 vol.1. doi:10.1109/ICDAR.1995.598994
41. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016:785-794. doi:10.1145/2939672.2939785
42. Hinton GE. Connectionist learning procedures. *Artificial Intelligence*. 1989;40(1):185-234. doi:10.1016/0004-3702(89)90049-0

427 43. Whittle J, Meyer S, Lacaud G, Baker SM, Iqbal M. Single-Cell Atlas of AML Reveals
428 Age-Related Gene Regulatory Networks in t(8;21) AML. *eLife*. 2025;14.
429 doi:10.7554/eLife.104978.1
430 44. Hua P, Roy N, de la Fuente J, et al. Single-cell analysis of bone marrow–derived CD34+
431 cells from children with sickle cell disease and thalassemia. *Blood*. 2019;134(23):23.
432 doi:10.1182/blood.2019002301
433 45. Psaila B, Wang G, Rodriguez-Meira A, et al. Single-Cell Analyses Reveal
434 Megakaryocyte-Biased Hematopoiesis in Myelofibrosis and Identify Mutant Clone-
435 Specific Targets. *Molecular Cell*. 2020;78(3):3. doi:10.1016/j.molcel.2020.04.008
436 46. Iskander D, Wang G, Heuston EF, et al. Single-cell profiling of human bone marrow
437 progenitors reveals mechanisms of failing erythropoiesis in Diamond-Blackfan anemia.
438 *Sci Transl Med*. 2021;13(610):610. doi:10.1126/scitranslmed.abf0113
439 47. Calvanese V, Capellera-Garcia S, Ma F, et al. Mapping human haematopoietic stem cells
440 from haemogenic endothelium to birth. *Nature*. 2022;604(7906):7906.
441 doi:10.1038/s41586-022-04571-x
442 48. Van De Sande B, Flerin C, Davie K, et al. A scalable SCENIC workflow for single-cell
443 gene regulatory network analysis. *Nat Protoc*. 2020;15(7):7. doi:10.1038/s41596-020-
444 0336-2
445 49. Kotliar D, Veres A, Nagy MA, et al. Identifying gene expression programs of cell-type
446 identity and cellular activity with single-cell RNA-Seq. Valencia A, Barkai N, Mereu E,
447 Göttgens B, eds. *eLife*. 2019;8:e43803. doi:10.7554/eLife.43803
448 50. Vincelette ND, Moon J, Kuykendall AT, et al. C-MYC Augments the Proliferation and
449 Survival of Hematopoietic Stem Cells and Multipotent Progenitors to Drive
450 Myeloproliferative Neoplasms. *Blood*. 2021;138(Supplement 1):28. doi:10.1182/blood-
451 2021-146399
452 51. Trikha P, Sharma N, Opavsky R, et al. E2f1–3 Are Critical for Myeloid Development. *J*
453 *Biol Chem*. 2011;286(6):4783-4795. doi:10.1074/jbc.M110.182733
454 52. Mgbemena VE, Signer RAJ, Wijayatunge R, Laxson T, Morrison SJ, Ross TS. Distinct
455 Brca1 Mutations Differentially Reduce Hematopoietic Stem Cell Function. *Cell Rep*.
456 2017;18(4):947-960. doi:10.1016/j.celrep.2016.12.075
457 53. Baker SJ, Ma'ayan A, Lieu YK, et al. B-*myb* is an essential regulator of hematopoietic
458 stem cell and myeloid progenitor cell development. *Proceedings of the National Academy*
459 *of Sciences*. 2014;111(8):8. doi:10.1073/pnas.1315464111

460 54. Delestré L, Cui H, Esposito M, et al. Senescence is a Spi1-induced anti-proliferative
461 mechanism in primary hematopoietic cells. *Haematologica*. 2017;102(11):11.
462 doi:10.3324/haematol.2016.157636
463 55. Hock H, Hamblen MJ, Rooke HM, et al. Gfi-1 restricts proliferation and preserves
464 functional integrity of haematopoietic stem cells. *Nature*. 2004;431(7011):1002-1007.
465 doi:10.1038/nature02994
466 56. Manesia JK, Xu Z, Broekaert D, et al. Highly proliferative primitive fetal liver
467 hematopoietic stem cells are fueled by oxidative metabolic pathways. *Stem Cell*
468 *Research*. 2015;15(3):715-721. doi:10.1016/j.scr.2015.11.001
469 57. Chen C, Liu Y, Liu R, et al. TSC–mTOR maintains quiescence and function of
470 hematopoietic stem cells by repressing mitochondrial biogenesis and reactive oxygen
471 species. *J Exp Med*. 2008;205(10):2397-2408. doi:10.1084/jem.20081297
472 58. Hagen JT, Montgomery MM, Aruleba RT, et al. Acute myeloid leukemia mitochondria
473 hydrolyze ATP to support oxidative metabolism and resist chemotherapy. *Science*
474 *Advances*. 2025;11(15):eadu5511. doi:10.1126/sciadv.adu5511
475 59. Liu Z, Tu H, Kang Y, et al. Primary cilia regulate hematopoietic stem and progenitor cell
476 specification through Notch signaling in zebrafish. *Nat Commun*. 2019;10(1):1839.
477 doi:10.1038/s41467-019-09403-7
478 60. Singh M, Chaudhry P, Merchant AA. Primary cilia are present on human blood and bone
479 marrow cells and mediate Hedgehog signaling. *Experimental Hematology*.
480 2016;44(12):1181-1187.e2. doi:10.1016/j.exphem.2016.08.009
481 61. Somervaille TCP, Matheny CJ, Spencer GJ, et al. Hierarchical maintenance of MLL
482 myeloid leukemia stem cells employs a transcriptional program shared with embryonic
483 rather than adult stem cells. *Cell Stem Cell*. 2009;4(2):129-140.
484 doi:10.1016/j.stem.2008.11.015
485 62. Giambra V, Gusscott S, Gracias D, et al. Epigenetic Restoration of Fetal-like IGF1
486 Signaling Inhibits Leukemia Stem Cell Activity. *Cell Stem Cell*. 2018;23(5):714-726.e7.
487 doi:10.1016/j.stem.2018.08.018
488 63. Chen C, Yu W, Alikarami F, et al. Single-cell multiomics reveals increased plasticity,
489 resistant populations, and stem-cell–like blasts in KMT2A-rearranged leukemia. *Blood*.
490 2022;139(14):14. doi:10.1182/blood.2021013442
491 64. Majeti R, Becker MW, Tian Q, et al. Dysregulated gene expression networks in human
492 acute myelogenous leukemia stem cells. *Proceedings of the National Academy of*
493 *Sciences*. 2009;106(9):3396-3401. doi:10.1073/pnas.0900089106

494 65. Jordan CT, Upchurch D, Szilvassy SJ, et al. The interleukin-3 receptor alpha chain is a
495 unique marker for human acute myelogenous leukemia stem cells. *Leukemia*.
496 2000;14(10):1777-1784. doi:10.1038/sj.leu.2401903
497 66. Hosen N, Park CY, Tatsumi N, et al. CD96 is a leukemic stem cell-specific marker in
498 human acute myeloid leukemia. *Proceedings of the National Academy of Sciences*.
499 2007;104(26):11008-11013. doi:10.1073/pnas.0704271104
500 67. Safina KR, Salik B, Kotliar DA, et al. An inflammatory and quiescent HSC
501 subpopulation expands with age in humans.
502 68. Gabut M, Samavarchi-Tehrani P, Wang X, et al. An alternative splicing switch regulates
503 embryonic stem cell pluripotency and reprogramming. *Cell*. 2011;147(1):132-146.
504 doi:10.1016/j.cell.2011.08.023
505 69. Wang F, Tan P, Zhang P, et al. Single-cell architecture and functional requirement of
506 alternative splicing during hematopoietic stem cell formation. *Sci Adv*. 2022;8(1):1.
507 doi:10.1126/sciadv.abg5369
508 70. Galen P van, Mbong N, Kreso A, et al. Integrated Stress Response Activity Marks Stem
509 Cells in Normal Hematopoiesis and Leukemia. *Cell Reports*. 2018;25(5):1109-1117.e5.
510 doi:10.1016/j.celrep.2018.10.021
511 71. Kharabi Masouleh B, Geng H, Hurtz C, et al. Mechanistic rationale for targeting the
512 unfolded protein response in pre-B acute lymphoblastic leukemia. *Proceedings of the*
513 *National Academy of Sciences*. 2014;111(21):E2219-E2228.
514 doi:10.1073/pnas.1400958111

515 **Figure legends**

516 **Figure 1. Construction of a human lifespan HSPC single-cell transcriptome atlas.** (**A**)
517 Schematic workflow of the study. (**B**) Sample sources for this study. PCW, post-conception
518 week; yr, age. (**C**) UMAP dimensionality reduction plot of the human HSPC transcriptome
519 atlas and highlighted expression of HSC marker genes (*AVP, HOPX, MLLT3*) on the UMAP
520 projection. (**D**) Normalized expression of lineage-specific marker genes in HSPCs. (**E**)
521 HSC/MPP filtration based on clustering and AUCell enrichment. Distribution of HSPC AUCell
522 scores on UMAP projection (left), clustering result (middle), and UMAP projection of refined
523 HSC/MPPs (right) were shown. HSC/MPP, hematopoietic stem cells and multipotent
524 progenitors; MEP, megakaryocyte-erythroid progenitors; EP, erythroid progenitors; MK,
525 megakaryocytes; EoMBP, eosinophil-basophil-mast progenitors; Lym, lymphoid progenitors;
526 Mye, myeloid progenitors; DC, dendritic cells; eF, early fetal (PCW 6-12); F, fetal (PCW 13
527 to birth); P, pediatric (age ≤18 years); yA, young adult (18 < age ≤40 years); A, adult (age >40
528 years). Figure 1A were created using Biorender (www.biorender.com).

529

530 **Figure 2. Age-related molecular programs in lifespan HSC/MPPs.** (**A**) Correlation of 17
531 candidate molecular programs (cMPs, upper part) and 6 robust molecular programs (MPs,
532 lower part). Correlation between cMP usage scores and age is shown on the right. 6 cMPs with
533 Pearson's Correlation Coefficient (PCC) > 0.20 and $P < 0.05$ in both discovery and validation
534 cohorts were chosed. (**B**) Distribution of 6 MPs on UMAP embeddings. (**C** and **D**) Expression
535 of top 5 genes from each MP on program-specific HSC/MPPs (**C**) and pathway enrichment
536 results (**D**) of each MP. A cell was considered specific to the MP with the highest usage score.
537 Pathway enrichment was performed using the GO Biological Process database. All pathways
538 shown reached significance of adjusted $P < 0.05$.

539

540 **Figure 3. Characterization of molecular programs across developmental stages.** (**A**)
541 Distribution and activity of MPs across the five developmental stages. (**B**) Temporal dynamics
542 of the two age-related MPs. (**C**) Activities of key transcription factors driving MP1 and MP5
543 across developmental stages. (**D** and **E**) Survival analysis results of MP5-related 18-gene set
544 on TCGA-LAML (**D**) and BeatAML (**E**) cohorts.

545

546 **Figure 4. A novel HSC/MPP aging clock model across the human lifespan.** (**A**) Workflow
547 of aging clock model construction, evalutation and application. The main steps included
548 processing of discovery and independent validation cohorts, feature selection, metacell

549 calculation, machine learning (ML) model construction, internal and external validation, final
550 model training on all 3 cohorts, and application of the model on AML datasets provided by
551 Whittle et al[43]. (**B**) Performance of six common machine learning models across the discovery,
552 validation and combined cohorts. The models are ranked by their mean absolute error (MAE)
553 on the validation cohort. (**C**) Performance of the final model (trained on 3 cohorts). (**D**)
554 Performance of the final model trained on 3 cohorts using each MP and a combined [MP1 +
555 MP5] as the model's features for training. (**E**) Pathway enrichment results using differentially
556 expressed genes between metacells belonging to the low-TAD group and the high-TAD group.
557
558 **Figure 5. TAD stratifies risk and predicts outcome in bulk RNA-seq AML cohorts.** (**A-B**)
559 Survival analysis of AML patients divided by TAD in TCGA-LAML (**A**) and BeatAML (**B**)
560 cohorts. (**C-E**) Survival analysis of AML patients re-stratified by TAD within the "Favorable"
561 (**C**), "Intermediate" (**D**), and "Adverse" (**E**) ELN risk group. (**F**) Complete remission (CR/CRi)
562 rates among patients in the three ELN risk groups. (**G**) CR/CRi rates among patients in the
563 ELN risk groups re-stratified by TAD.
564

Figures

Figure 1

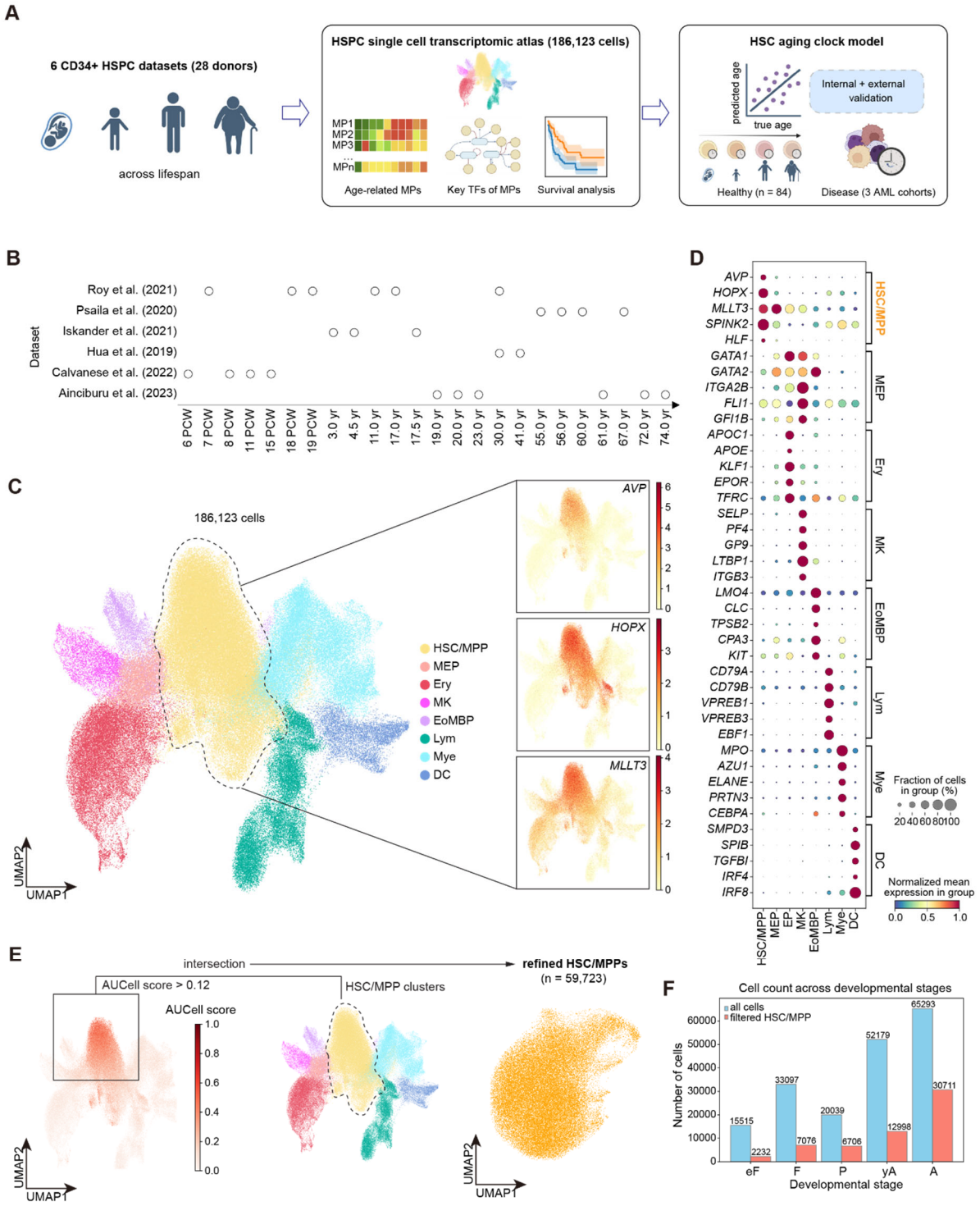


Figure 2

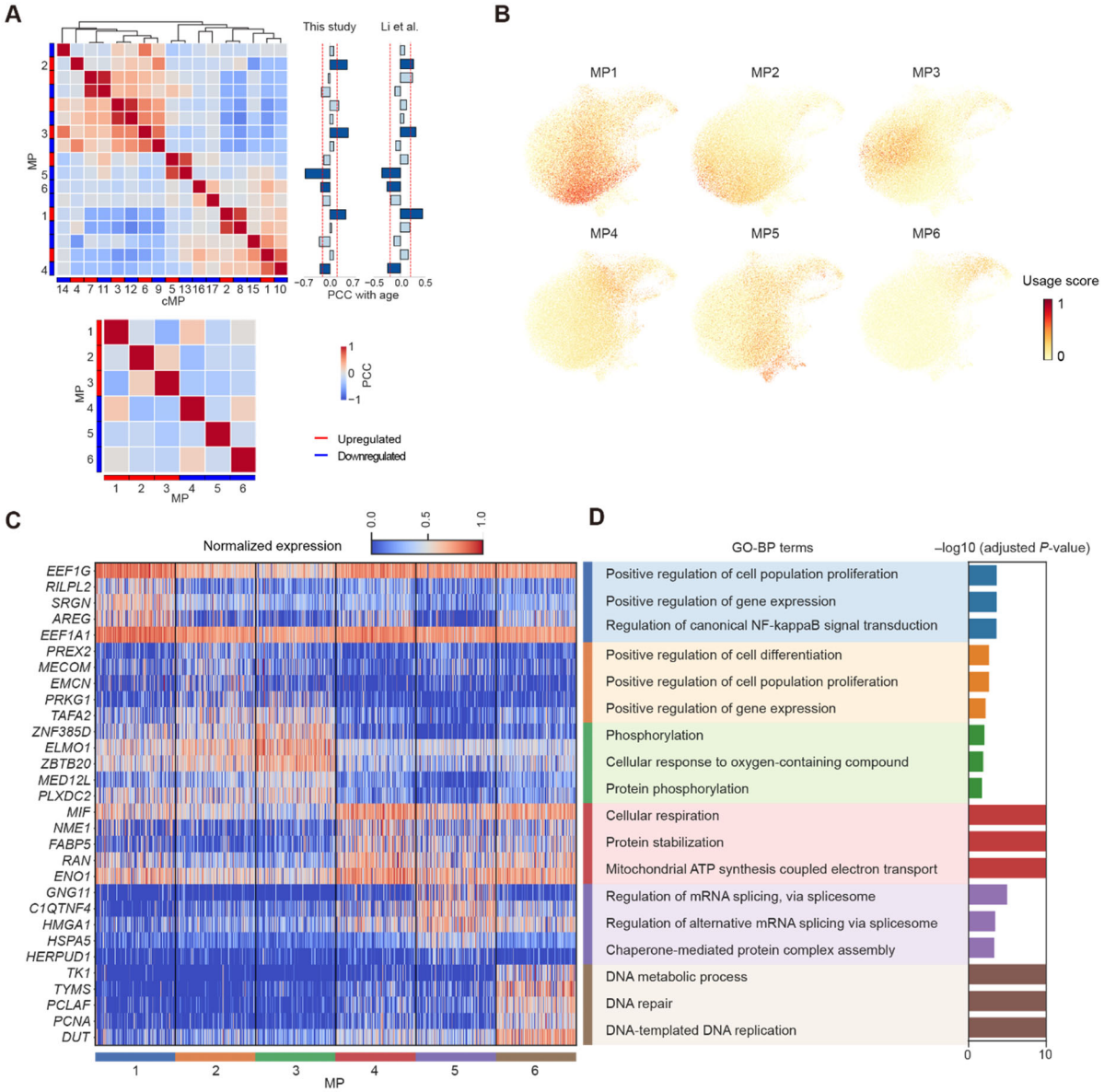

A
This study
Li et al.
MP
cMP
PCC with age
PCC
Upregulated
Downregulated
B
MP1
MP2
MP3
MP4
MP5
MP6
Usage score
C
Normalized expression
MP
D
GO-BP terms
–log10 (adjusted P-value)
Positive regulation of cell population proliferation
Positive regulation of gene expression
Regulation of canonical NF-kappaB signal transduction
Positive regulation of cell differentiation
Positive regulation of cell population proliferation
Positive regulation of gene expression
Phosphorylation
Cellular response to oxygen-containing compound
Protein phosphorylation
Cellular respiration
Protein stabilization
Mitochondrial ATP synthesis coupled electron transport
Regulation of mRNA splicing, via splicesome
Regulation of alternative mRNA splicing via splicesome
Chaperone-mediated protein complex assembly
DNA metabolic process
DNA repair
DNA-templated DNA replication

Figure 3

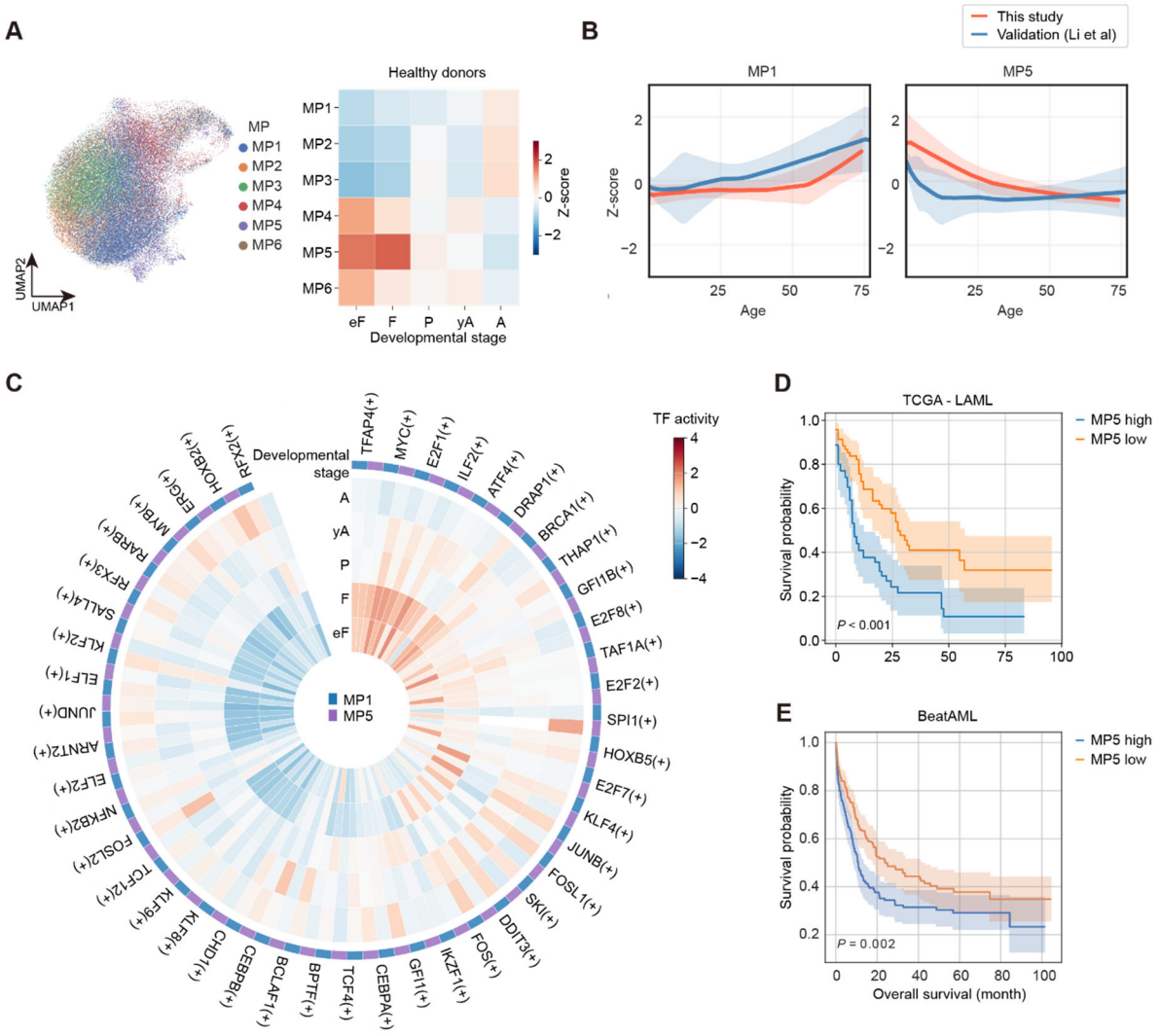
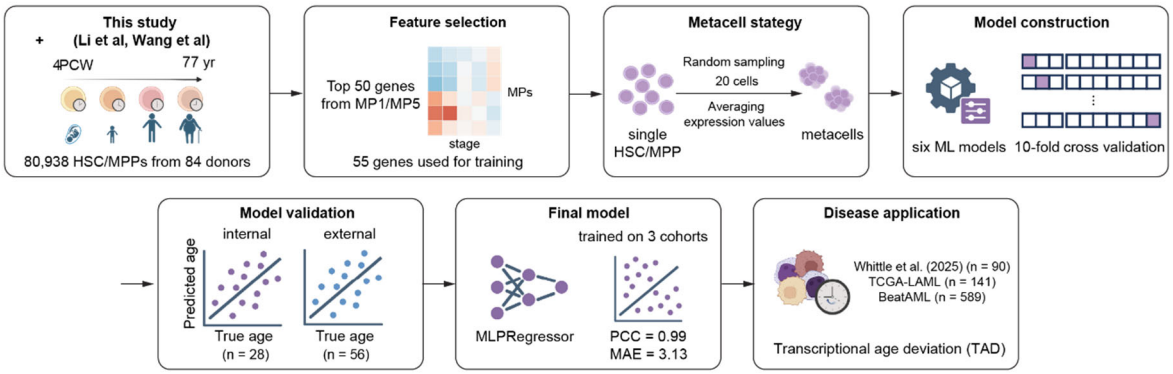
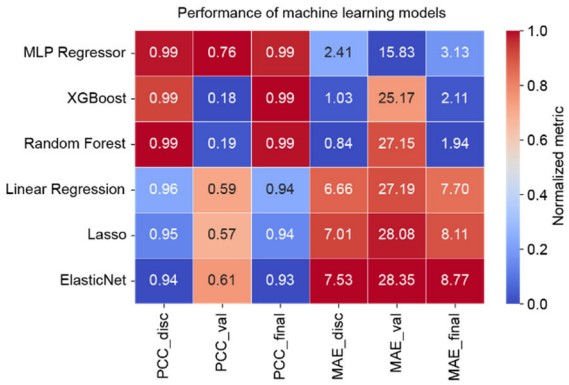

A
MP
MP1
MP2
MP3
MP4
MP5
MP6
UMAP2
UMAP1
Healthy donors
MP1
MP2
MP3
MP4
MP5
MP6
eF F P yA A
Developmental stage
Z-score
B
This study
Validation (Li et al)
MP1
MP5
Z-score
Age
C
Developmental stage
A
yA
P
F
eF
MP1
MP5
TF activity
TFAP4(+)
MYC(+)
E2F1(+)
ILF2(+)
ATF4(+)
DRAP1(+)
BRCA1(+)
THAP1(+)
GFI1B(+)
E2F8(+)
TAF1A(+)
E2F2(+)
SPI1(+)
HOXB5(+)
E2F7(+)
KLF4(+)
JUNB(+)
FOSL1(+)
SKI(+)
DDIT3(+)
FOS(+)
IKZF1(+)
GFI1(+)
CEBPA(+)
TCF4(+)
BPTF(+)
BCLAF1(+)
CEBPB(+)
CHD1(+)
KLF8(+)
KLF9(+)
TCF12(+)
FOSL2(+)
NFKB2(+)
ELF2(+)
ARNT2(+)
JUND(+)
ELF1(+)
KLF2(+)
SALL4(+)
RFX3(+)
RARB(+)
MYB(+)
ERG(+)
HOXB2(+)
RFX2(+)
D
TCGA - LAML
MP5 high
MP5 low
Survival probability
P < 0.001
E
BeatAML
MP5 high
MP5 low
Survival probability
P = 0.002
Overall survival (month)

Figure 4

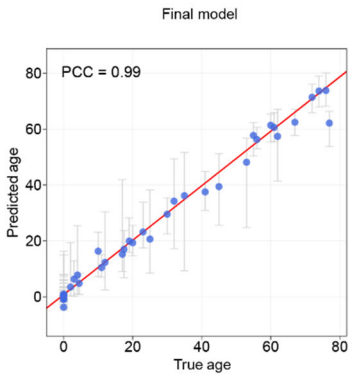
A



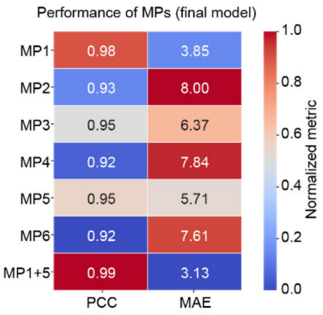
B



C



D



E

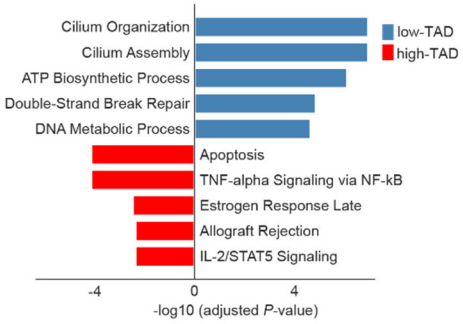
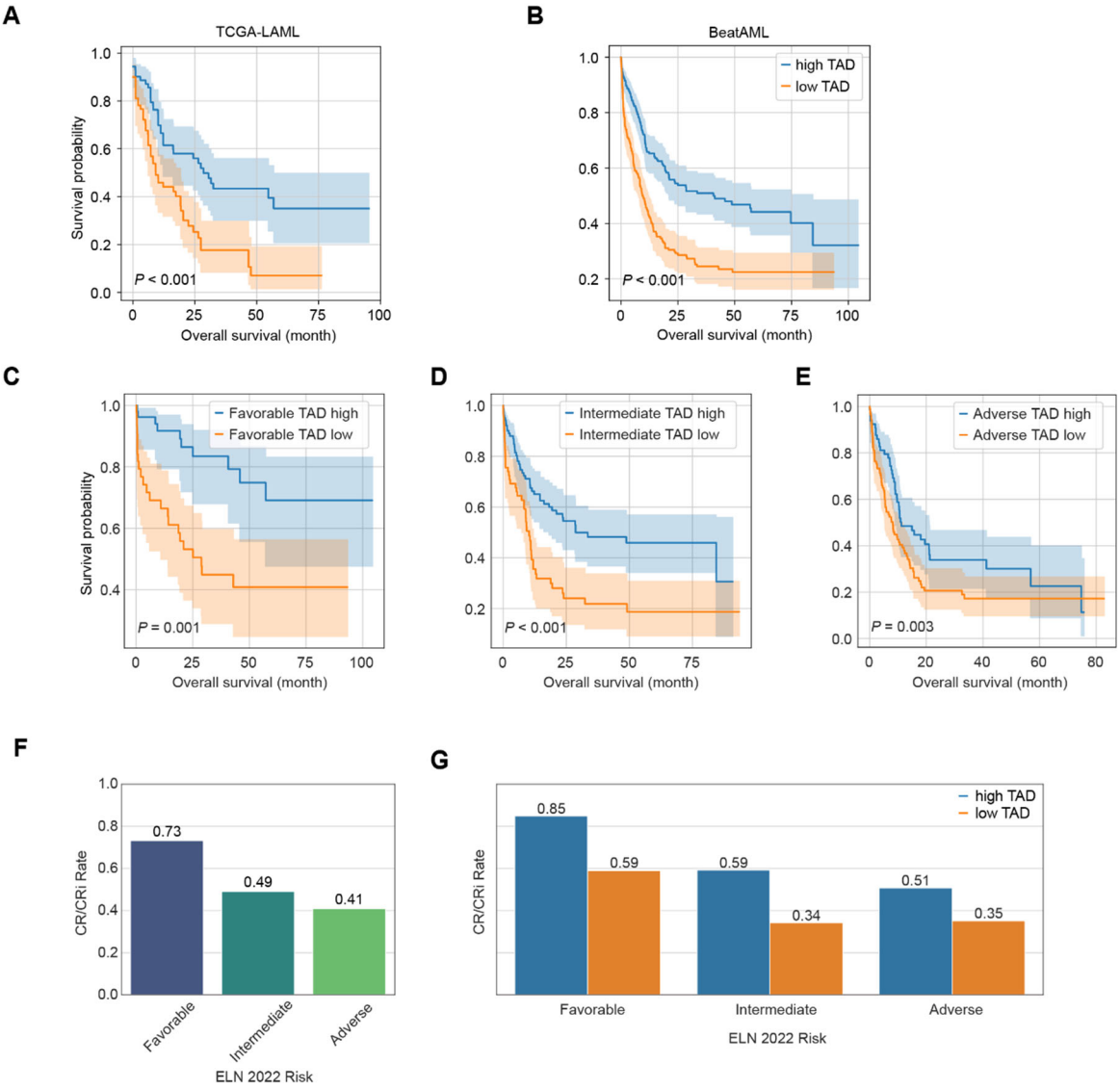


Figure 5


A
TCGA-LAML
Survival probability
P < 0.001
Overall survival (month)
B
BeatAML
high TAD
low TAD
P < 0.001
Overall survival (month)
C
Favorable TAD high
Favorable TAD low
Survival probability
P = 0.001
Overall survival (month)
D
Intermediate TAD high
Intermediate TAD low
P < 0.001
Overall survival (month)
E
Adverse TAD high
Adverse TAD low
P = 0.003
Overall survival (month)
F
CR/CRi Rate
0.73
0.49
0.41
Favorable
Intermediate
Adverse
ELN 2022 Risk
G
high TAD
low TAD
CR/CRi Rate
0.85
0.59
0.59
0.34
0.51
0.35
Favorable
Intermediate
Adverse
ELN 2022 Risk