A single-cell atlas and aging clock define biological age and risk-associated stem cell states in human hematopoiesis

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

Affiliations

[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@fudan.edu.cn

Key words: hematopoietic stem and progenitor cell, single cell transcriptomics, aging clock

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 six molecular programs (MPs) that define distinct functional states within the HSC/multipotent progenitor (MPP) compartment. Among these, MP1, associated with inflammaging, and MP5, related to RNA splicing and protein homeostasis, exhibited conserved age-associated dynamics across independent datasets, highlighting their central roles in stem cell aging. Leveraging these age-associated programs, we developed a machine learning-based stem cell aging clock from 84 donors to predict chronological age from HSC/MPP transcriptomes. Applying this aging clock to acute myeloid leukemia (AML) defines transcriptional age deviation (TAD), a measure of biological age divergence in leukemic HSC/MPPs. TAD captures disease-associated variation linked to genetic risk stratification and patient survival. Our study provides a high-resolution reference map of human HSPC aging, establishes a stem cell-specific aging clock and demonstrates its utility for uncovering aging-related dysregulation with prognostic relevance in hematological malignancies.

Introduction

Hematopoietic stem and progenitor cells (HSPCs) serve as the foundation of lifelong hematopoiesis, sustaining hematopoietic homeostasis and immune competence throughout the human lifespan. At the apex of this hierarchy, hematopoietic stem cells (HSCs) and multipotent progenitors (MPPs) are characterized by their self-renewal capability and multilineage differentiation potential, dynamically balancing quiescence, proliferation, and differentiation to meet physiological demands[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 HSCs in organismal aging. Multiple mechanistic studies in mouse have identified several key drivers and pathways involved in HSC aging, including increased inflammatory exposure (inflammaging) and disturbance of RNA splicing[8–13]. However, the extent to which these mechanisms operate in human HSC aging remains to be fully elucidated.

Understanding the transcriptional programs governing both physiological aging and pathological transformation in HSC/MPPs is crucial for identifying molecular vulnerabilities underlying age-related hematopoietic decline and malignant progression[14]. Studying human HSC aging remains challenging due to the inherent heterogeneity of HSC/MPPs and limited access to human samples across the lifespan. Traditional bulk RNA sequencing approaches fail to capture the complexity of individual HSC 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 HSC/MPP heterogeneity and functions remain incompletely understood.

Recent studies have introduced transcriptome-based 'aging clocks' at single cell resolution in peripheral blood mononuclear cells (PBMCs), testis, and endometrium, demonstrating that aging leaves measurable signatures that can be leveraged by machine-learning models to infer biological age[20–23]. Such approaches have proven valuable for detecting aging deviations in disease states and evaluating interventions targeting aging pathways. However, it remains unclear whether aging clocks can be robustly applied to HSC/MPPs across the full developmental continuum, from prenatal stages to late adulthood, and whether they can sensitively detect deviations in biological age associated with hematological disorders.

In this study, we constructed a comprehensive single-cell transcriptomic atlas of human $CD34^+$ HSPCs spanning developmental stages from early prenatal development (6 PCW) to late adulthood (74 years) across 28 healthy donors. Molecular characterization revealed six aging-associated molecular programs (MPs) representing distinct transcriptional signatures related to inflammatory signaling, differentiation, protein phosphorylation, aerobic respiration, RNA splicing and proteostasis, and cell cycle regulation. These molecular programs define age-associated functional states and are robust across independent datasets. Leveraging these programs, we developed a machine learning-based stem cell aging clock that robustly predicts donor chronological age from single-cell HSC transcriptomes and quantifies biological age from 84 donors. Applying this framework to acute myeloid leukemia (AML), we defined transcriptional age deviation (TAD) as the age deviation of leukemic HSC/MPPs relative to a healthy baseline. TAD reveals disease-associated remodeling of stem cell aging states that extends beyond intratumoral heterogeneity, capturing clinically meaningful variation linked to genetic risk stratification and patient outcome. Collectively, our work provides a reference map of human HSPC aging, establishes a stem cell aging clock that enables quantitative integration of age-associated stem cell states at single-cell resolution, and provides a framework for uncovering aging-related dysregulation linked to disease risk and outcome in hematological disorders.

Results

A single-cell transcriptomic atlas of human HSPCs across the lifespan

To systematically characterize age-related dynamics in human hematopoiesis, we re-analyzed raw FASTQ files from 30 samples obtained from six previously published datasets of CD34$^+$ hematopoietic stem and progenitor cells (HSPCs), comprising a total of 200,665 cells from 28 healthy donors aged from post-conception week (PCW) 6 to 74 years (**Fig. 1A-B**)[5,17,24–27]. After stringent quality control, doublet removal using scDblFinder[28], and exclusion of non-hematopoietic cells (**Fig. S1A-B**), 186,123 high-quality cells were retained for downstream analyses (**Fig. 1C**, **Supplementary Table 1**). To verify sample integrity, donor sex was confirmed based on expression of sex-chromosome markers *XIST* (female) and *UTY* (male) (**Fig. S1C**).

We applied Harmony[29] to integrate datasets across different sequencing datasets and experimental batches, followed by dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP) (**Fig. S1D-E, Supplementary Table 1**). Unsupervised Leiden clustering revealed transcriptionally distinct clusters, which we further annotated into eight major hematopoietic populations based on canonical marker genes (**Fig.1C-D, Fig. S1F**) and differentially expressed genes (DEGs) (**Fig. S1G**, **Supplementary Table 2**). These populations included hematopoietic stem and multipotent progenitors (HSC/MPP: *AVP*, *HOPX*, *MLLT3*), megakaryocyte-erythroid progenitors (MEP: *GATA1*, *GATA2*), erythroid progenitors (Ery: *APOC1*, *KLF1*), megakaryocyte progenitors (MK: *PF4, GP9*), eosinophil-basophil-mast progenitors (EoMBP: *TPSB2*, *LMO4*), lymphoid progenitors (Lym: *CD79A*, *CD79B*), myeloid progenitors (Mye: *MPO*, *ELANE*), and dendritic cell progenitors (DC: *SMPD3*, *SPIB*). Together, these data constitute a comprehensive atlas of healthy human HSPCs spanning the lifespan.

High-confidence transcriptional delineation of the HSC/MPP compartment

To investigate age-dependent heterogeneity within the most primitive hematopoietic compartment, we further refined the HSC/MPP population using a dual-filtering strategy. We intersected cells annotated as HSC/MPPs cluster from the main atlas

163 and then retained only cells with AUCell scores exceeding the average of 0.12 for an
164 established HSC/MPP gene signature[17], resulting in 59,723 high-confidence
165 HSC/MPP cells for downstream analysis (**Fig. 1E**).

166

167 We validated the stemness and differentiation potential of this refined population using
168 CytoTRACE2[30], which confirmed the highest differentiation potential among all
169 clusters (**Fig. S2A**). Furthermore, pseudotime-based trajectory analysis using
170 StaVIA[31] on the force-directed graph (FDG) embeddings independently supported the
171 undifferentiated status of this population (**Fig. S2B-C**).To facilitate systematic analysis
172 across the human lifespan, we grouped donors into five discrete developmental stages:
173 early fetal (eF, PCW 6-12), fetal (F, PCW 13 to birth), pediatric (P, age ≤18 years),
174 young adult (yA, 18 < age ≤40 years), and adult (A, age >40 years), comprising 2,232,
175 7,076, 6,706, 12,998 and 30,711 HSC/MPPs, respectively (**Fig. 1F**). This grouping
176 enabled comparisons across developmental and aging trajectories.

177

178 **Stage-specific transcriptional and regulatory programs in HSC/MPPs**

179 We applied the same preprocessing steps on the refined HSC/MPPs. Projection of
180 developmental stage groups onto the HSC/MPP UMAP embedding revealed distinct
181 enrichment patterns, indicating pronounced transcriptional heterogeneity across age
182 groups (**Fig. S2D)**. To characterize these age-associated transcriptional changes, we
183 performed pairwise differential expression analysis across the five defined
184 developmental stages using COSG[32] (**Fig. S2E, Supplementary Table 3**). Notably,
185 the largest number of differentially expressed genes (DEGs) was observed between
186 early fetal and adult groups, demonstrating substantial molecular remodeling
187 occurring over the human lifespan. The cell cycle signatures analysis revealed that the
188 proportion of cells in G2/M and S phases declined with age, while the proportion of
189 cells in G1 phase increased during aging (**Fig. S2F-G**), suggesting a shift toward
190 quiescence in aged HSC/MPPs.

191

192 To further elucidate upstream regulatory programs underlying these transcriptional
193 changes, we applied pySCENIC[33,34] to infer transcription factor (TF) regulon activity

194 across developmental groups. As expected, strong developmental stage-specific
195 shifts in gene regulatory networks were identified (**Fig. S2H**). Regulons including
196 MYC(+), E2F1(+), and ATF4(+) were predominantly active in early developmental
197 stages, consistent with roles in primitive hematopoiesis and stem cell maintenance. In
198 contrast, NFKB2(+), MYB(+), and ERG(+) were dominant in aged HSC/MPPs,
199 suggesting functional rewiring of stem cell identity and lineage priming with age. We
200 identified several TFs enriched in young HSC/MPPs, including MYC, which has been
201 previously reported to regulate HSC self-renewal[35]. Together, these findings reveal
202 progressive rewiring of transcriptional regulatory circuits that underpin age-associated
203 functional heterogeneity within the HSC/MPP compartment.

204

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

206 To delineate the latent transcriptional architecture of the refined HSC/MPP population,
207 we employed consensus non-negative matrix factorization (cNMF)[36] to identify
208 molecular programs (MPs) representing underlying functional states. From an initial
209 set of 17 age-associated programs identified by cNMF, we retained 6 robust MPs
210 whose activities shift significantly in both our refined HSC/MPP dataset and an
211 independent validation cohort[16] (**Fig. 2A-B, Fig. S3A-B, Supplementary Table 4,**
212 **Methods**). Pathway enrichment analysis revealed that each MP corresponds a distinct
213 transcriptional module and is associated with specific biological processes. MP1 was
214 enriched for genes involved in inflammatory NF-κB signaling. MP2 shows particular
215 enrichment for genes related to HSC identity (*MECOM, HLF, MLLT3*) that regulate
216 differentiation and proliferation. MP3 was enriched for transcription factor binding and
217 protein phosphorylation. MP4 captured signatures of aerobic respiration in
218 mitochondria. MP5 was characterized by genes associated with RNA splicing (*YBX1,*
219 *C1QBP, HNRNPA1*) and chaperone-associated proteostasis (*HSPA5, HERPUD1,*
220 *UBE2E2*) pathways. MP6 showed enrichment for cell cycle-related genes. (**Fig. 2C-D,**
221 **Supplementary Table 5**).

222

223 To delineate these transcriptional programs across developmental stages and
224 lineages, we quantified MP utilization across five age groups and major hematopoietic
225 lineages and projected program activities onto the full HSPC UMAP (**Fig. S3C**). MP1

226 displayed a monotonic increase along the aging trajectory, with progressive
227 enrichment from fetal to adult stages, particularly within the HSC/MPP compartment
228 and along megakaryocyte-erythroid differentiation branches. MP2 and MP3 showed
229 pronounced age-dependent upregulation in primitive HSC/MPPs, megakaryocyte and
230 erythroid differentiation trajectories, whereas MP3 was specifically upregulated in the
231 eosinophil-mast-basophil progenitor (EoMBP) lineage, highlighting lineage-specific
232 divergence of age-associated programs. In contrast, MP4 was consistently elevated
233 during early fetal and fetal stages across all hematopoietic lineages, suggesting a
234 developmentally restricted program. MP5 was enriched in early fetal/fetal
235 development and displayed a marked decline beginning at the pediatric stage in
236 HSC/MPP and myeloid lineage. MP6 exhibited gradual decrease with age and showed
237 modest variation across lineage contexts. Together, the lineage-biased deployment
238 and age-dependent remodeling of these MPs demonstrate that hematopoietic aging
239 is shaped by both globally coordinated transcriptional shifts and compartment-specific
240 regulatory pressures.

241

242 **Age-associated dynamics of molecular programs in HSC/MPPs**

243 To systematically explore age-associated changes in molecular programs in refined
244 HSC/MPPs, we assessed the distribution and activity of MPs across the five
245 developmental stages (**Fig. 3A**). MP1-3 were predominantly enriched in adult ages,
246 suggesting progressive activation of inflammaging pathways with age. In contrast,
247 MP4-6 showed pronounced enrichment during early development (eF and F stages)
248 but markedly decreased in postnatal and diminished in adult stages, indicating
249 developmental-specific programs associated with aerobic respiration, transcription
250 regulation, and proliferation that are attenuated after birth.

251

252 To validate the robustness of these patterns, we applied the MP signatures to an
253 independent publicly available HSC datasets from Li et al[16]. Among the six modules,
254 MP1 and MP5 exhibited the most dramatic age-related shifts in both cohorts (**Fig. 3B,**
255 **Fig. S3D, Supplementary Table 4**), confirming their role as core aging-associated
256 programs in human HSC/MPPs. The progressive increase in MP1 activity from fetal

257 to aged stages suggests accumulation of inflammation, while MP5 highlights a critical
258 window of RNA splicing and proteostasis during early hematopoiesis.

259

260 We next examined transcription factor (TF) activity within cells assigned to MP1 and
261 MP5 modules, stratified by developmental stage (**Fig. 3C**). This analysis revealed age-
262 and molecular program- dependent regulatory rewiring. Specifically, the MP5 program
263 in the fetal stage was driven by key proliferative transcriptional regulons such as
264 MYC(+), E2F1(+), and BRCA1(+)[37–39]. The adult MP5 signature was less identifiable
265 compared with MP1 and shift towards HSC maintenance regulators such as MYB(+)
266 and SPI1(+) and GFI1(+)[40–42]. In parallel, the adult MP1 program, linked to
267 inflammation, became dominated by stress- and inflammation-responsive TFs,
268 including members of the AP-1 (e.g., FOS, JUND) and NF-κB (e.g., NFKB2) families.
269 This regulatory shift mirrors the emergence of inflammatory and stress-adapted
270 transcriptional states during hematopoietic aging. These findings illustrated that MP1
271 and MP5 are governed by distinct and dynamically evolving TF networks that couple
272 developmental stage with functional state transitions in the HSC/MPP heterogeneity.

273

274 **Poor-prognosis transcriptional signature in AML defined by MP5**

275 Given MP5 program enrichment in early development, we next evaluated the clinical
276 relevance of this program in AML patients. In the TCGA-LAML cohort (n = 141), we
277 performed univariable Cox proportional hazards analysis on the top 200 MP5-
278 associated genes and identified 18 genes independently associated with shorter
279 overall survival after multiple-testing correction (HR > 1, FDR < 0.05; **Supplementary**
280 **Table 6)**[43]. This 18-gene set exhibited strong correlation with the original MP5
281 signature (Pearson r = 0.74; **Fig. S3E-F**), supporting it as a robust representation of
282 MP5 activity.

283

284 We next derived a cohort-specific MP5 score by summing the TPM expression values
285 of these 18 genes. Patients stratified into the MP5-high group showed significantly
286 poorer overall survival compared with MP5-low patients (median OS 7.2 vs 19.8
287 months; log-rank $P < 0.001$; **Fig. 3D**). Importantly, this prognostic association was
288 independently validated in the BeatAML cohort (n = 591), where MP5-high patients

289 exhibited shorter survival (n = 591; median OS 9.9 vs 15.0 months; log-rank *P* = 0.002;
290 **Fig. 3E**)[44]. These results demonstrate that MP5 activity defines a high-risk
291 transcriptional state in AML, linking age-associated MPs with clinical outcomes.

292

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

294 To develop a robust stem cell aging clock based on HSC/MPP transcriptomes, we
295 expanded the analysis with our refined dataset (54,424 HSC/MPPs from 28 donors)
296 with two independent published cohorts (17,211 HSC/MPPs from 24 donors; 4,004
297 HSC/MPPs from 32 donors, respectively)[16,45], forming the largest human stem cell
298 aging cohort comprising of 84 donors (80,938 cells). This expands cohort increased
299 donor diversity and statistical power across the human lifespan. To reduce technical
300 noise while preserving biological variability, we implemented a 'metacell' strategy that
301 20 randomly sampled HSC/MPP cells from each donor were aggregated to generate
302 one metacell. This resulted in a total of 4,088 metacells spanning ages from 4 PCW
303 to 77 years (**Fig. 4A**). Donor's chronological age was used as the ground truth label
304 for model training.

305

306 Given that MP1 and MP5 exhibited the strongest age association in our analyses
307 (**Supplementary Table 4**), we prioritized these modules for feature selection. From
308 MP1 and MP5, we extracted 55 informative genes used as input features for the
309 downstream model construction after excluding ribosomal, mitochondrial, HLA,
310 housekeeping[46,47], hemoglobin, sex chromosome genes, top 50 lineage-specific
311 marker and genes not robustly expressed across datasets (**Supplementary Table 2,**
312 **7**).

313

314 We next trained and benchmarked six supervised machine learning models, including
315 Linear Regression, Elastic Net, Lasso, Random Forest, XGBoost and MLPRegressor,
316 using a 10-fold cross-validation strategy with our dataset as discovery dataset and the
317 two independent cohorts as external validation datasets (**Methods**). While several
318 models achieved similarly high Pearson correlation coefficients (PCCs) on the
319 discovery and validation datasets (PCC = 0.99), MLPRegressor consistently

320 demonstrated superior generalization performance, maintaining the highest PCC on
321 the validation dataset (0.76). In contrast, XGBoost and Random Forest showed a
322 pronounced reduction in validation PCC (≈ 0.18-0.19), indicating limited
323 generalizability despite strong performance on the discovery set. Therefore, we
324 selected MLPRegressor as our final model (**Fig. 4B, Fig. S4A-B**). We retrained the
325 final model using all three cohorts, resulting in a global stem cell aging clock with an
326 MAE of 3.13 and a PCC of 0.99 (**Fig. 4C**). To further interpret the molecular drivers
327 underlying model predictions, we calculated permutation-importance scores for each
328 feature in the final model, identifying genes such as *AREG*, *CRHBP* and *GNG11* as
329 the most influential predictors, suggesting their relevance to age-related transcriptional
330 reprogramming (**Fig. S4C**). Moreover, to systematically assess the contribution of
331 each MP, we trained and compared separate models using genes derived from each
332 MP and MP1 + MP5 as we mentioned above. This analysis confirmed that [MP1 +
333 MP5] based models consistently outperformed those derived from other modules,
334 highlighting the key role of these modules in defining HSC/MPP age (**Fig. 4D**).

335

336 **Transcriptional age deviation (TAD) stratifies AML by risk and survival**

337 To investigate whether hematological malignancies are associated with altered
338 HSC/MPP biological aging, we applied the age prediction model to a large scRNA-seq
339 cohort of acute myeloid leukemia(AML), comprising 90 patients of 73,360 single
340 HSC/MPPs[48] (**Supplementary Table 8**). Cells were aggregated into 3,715 HSC/MPP
341 metacells and we calculated the transcriptional age deviation (TAD) for each metacell
342 after applying correction to the chronological age-dependent bias (**Methods**).
343 Metacells within the lowest and highest 10% of TAD were annotated as low-TAD and
344 high-TAD groups, respectively (low-TAD group: TAD < - 9.5, n = 334 metacells; high-
345 TAD group: TAD > 4.8, n = 272 metacells). Differential expression analysis identified
346 2483 genes up-regulated in low-TAD group and 366 genes up-regulated in high-TAD
347 group.

348

349 High-TAD group were dominated by pro-inflammatory and stress-response programs,
350 including TNF-α signalling via NF-κB, IL-2/STAT5 signaling and apoptosis. Low-TAD
351 cells were selectively enriched for aerobic respiration, ATP biosynthetic process and

352 primary cilium assembly (adjusted $P < 0.05$, **Fig. 4E**). The enrichment of aerobic-
353 respiration pathways in the low-TAD group mirrors the MP4 activity characteristic of
354 fetal and early-postnatal HSCs. Notably, upregulated mitochondrial oxidative
355 metabolism has been linked to the enhanced proliferative capacity of early HSCs and
356 to drug resistance in AML[49–51]. In addition, primary cilia are reported essential for
357 primitive hematopoiesis in both zebrafish and human and dysregulated in AML[52,53].

358

359 We next examined the relationship between TAD and established clinical risk
360 stratification in AML. Stratification of patients according to the European LeukemiaNet
361 (ELN) genetic risk groups[54] revealed lower TAD values enriched in the adverse-risk
362 category and higher TAD values in the favorable-risk category (Kruskal–Wallis $P = 1.4$
363 $\times 10^{-3}$, **Fig. 4F**). These results indicate that reduced TAD is associated with poorer
364 genetic risk as defined by the ELN classification.

365

366 To further evaluate the prognostic associations of TAD, we analyzed two independent
367 bulk RNA-seq AML cohorts, TCGA-LAML (n = 141) and BeatAML (n = 589). We
368 applied our model to predict the age of each patient and calculated TAD as described
369 in the **Methods**. Patients were stratified into low- and high-TAD groups based on
370 cohort-specific median TAD values. In both cohorts, the low-TAD group showed
371 significantly shorter overall survival compared with the high-TAD group (TCGA-LAML:
372 median OS 8.6 vs 13.1 months; BeatAML: median OS 9.0 vs 16.5 months; log-rank P
373 < 0.001 in both cohorts; **Fig. S4D-E**). Moreover, TAD declined stepwisely across ELN
374 risk groups in the BeatAML cohort, with adverse-risk cases exhibiting the lowest TAD
375 values (Kruskal–Wallis $P = 6.9 \times 10^{-5}$; **Fig. S4F**). Collectively, these results suggest
376 that TAD captures biologically and clinically meaningful heterogeneity within AML
377 group and serve as a potential biomarker for stratifying AML patients by genetic risk
378 and survival outcome.

379

380 **Discussion**

381 In this study, we constructed a comprehensive single-cell transcriptomic atlas
382 spanning human hematopoiesis from early prenatal development to adult age over 70.
383 By integrating and re-analyzing >186,000 high-quality cells across 28 healthy donors,

384 we identified 59,723 high-confidence HSC/MPPs with the maximal differentiation
385 potential and applied consensus non-negative matrix factorization to define six
386 molecular programs (MPs) governing developmental stages and functional states
387 across the human lifespan. We further developed a machine learning-based stem cell
388 aging clock trained on 80,938 cells HSC/MPPs from 84 donors ranging from 4 PCW
389 to 77 years old, constituting the largest human stem cell aging cohorts to date.

390

391 Age-associated transcriptional rewiring in HSC/MPPs is characterized by programs
392 related to stress responses, metabolic activity, RNA splicing and proteostasis. Notably,
393 Among the six molecular programs identified, MP1 (inflammaging) and MP5 (RNA
394 splicing and proteostasis) demonstrated the most prominent age-related changes and
395 were conserved across datasets, supporting age-specific transcriptional
396 reprogramming events. The postnatal attenuation of MP5 coupled with progressive
397 MP1 accumulation in aged cells suggests a developmental shift from active stress
398 response programs to stress-adaptive states, potentially underlying functional
399 exhaustion and altered differentiation capacity. These findings align with recent
400 mechanistic studies demonstrating that HSC inflammaging disrupts stem cell
401 quiescence and contributes to clonal hematopoiesis[8–10], while intact RNA splicing and
402 proteostasis maintains HSC self-renewal and confers a survival advantage to leukemic
403 stem cells[11–13,55–58]. Age-associated upregulation of MP1 is consistent with recent
404 study describing the expansion of inflammatory HSC subpopulations during aging[59].
405 We further identified an MP5-associated 18-gene signature that delineates a high-risk
406 disease state that patients with elevated expression exhibited significantly shorter
407 overall survival in both TCGA-LAML and BeatAML cohort. Together, these findings
408 link age-associated transcriptional rewiring to leukemic fitness and adverse clinical
409 outcome, suggesting that aging-related molecular programs may be co-opted or
410 dysregulated during malignant transformation.

411

412 By leveraging gene features from MP1 and MP5, our machine learning based stem
413 cell aging clock achieved high predictive accuracy (MAE = 3.13 years, PCC = 0.99),
414 demonstrating that 55 key features genes can reliably predict donor chronological age.
415 This represents, to our knowledge, the first transcriptomic aging clock specifically

designed for human HSC/MPPs spanning the entire developmental spectrum from prenatal through advanced age. The top-ranked features identified by permutation importance exhibited clear age dependent expression patterns offering candidate biomarkers for stem cell aging that warrant further functional validation.

Applying this aging clock to a cohort of 90 AML samples revealed a transcriptional age deviation (TAD) that delineates HSC/MPP states. Specifically, TAD separates pro-inflammatory, apoptosis-rich HSC subsets from aerobic respiration-dominant, developmentally biased counterparts, thereby revealing intra-tumoral heterogeneity. Importantly, TAD scaled with ELN genomic risk in single-cell data and independent AML patient cohorts, robustly stratified patients by overall survival. These findings establish TAD as a quantitative axis that linking stem-cell aging to leukemic fitness and clinical outcome, providing a broadly applicable framework for monitoring biological age and uncovering age-associated therapeutic vulnerabilities across hematologic malignancies.

Our study has several limitations. First, although we analyzed over 186,000 cells from 28 donors and expand the model to 84 donors, donor representation at extreme ages remains limited, which may affect the precision of age predictions at these boundaries. Second, our analysis is based on transcriptomic data alone, integration with epigenomic, proteomic, and metabolomic datasets at single cell level in the future would provide a more comprehensive view of HSC aging. Third, larger longitudinal AML cohorts with detailed clinical annotation will be required to determine the temporal stability and predictive utility of TAD during disease progression and treatment.

Despite these limitations, our work provides a high-resolution map of age-dependent transcriptional dynamics in human HSC/MPPs and establishes a predictive framework for estimating stem cell biological age. The aging clock framework enables quantitative assessment of biological age at single-cell resolution, with potential applications in evaluation of interventions that modulate age-associated stem cell states, and identification of aging-related vulnerabilities that predispose to hematological dysfunction and malignancy.

448 **References**

449 1. Seita, J. & Weissman, I. L. Hematopoietic Stem Cell: Self-renewal versus Differentiation.
450 *Wiley Interdiscip Rev Syst Biol Med* **2**, 640–653 (2010).
451 2. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation
452 landscapes. *Nature* **553**, 418–426 (2018).
453 3. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal
454 during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
455 4. Kasbekar, M., Mitchell, C. A., Proven, M. A. & Passegué, E. Hematopoietic stem cells
456 through the ages: A lifetime of adaptation to organismal demands. *Cell Stem Cell* **30**, 1403–
457 1420 (2023).
458 5. Ainciburu, M. *et al.* Uncovering perturbations in human hematopoiesis associated with
459 healthy aging and myeloid malignancies at single-cell resolution. *eLife* **12**, e79363 (2023).
460 6. Zhou, F. J. J. *et al.* Age-Related Changes in Hematopoietic Stem Cell Proteostasis Promote
461 the Emergence of Clonal Hematopoiesis and Leukemia in Older Adults. *Blood* **144**, 4047–
462 4047 (2024).
463 7. Conboy, I. M. *et al.* Rejuvenation of aged progenitor cells by exposure to a young systemic
464 environment. *Nature* **433**, 760–764 (2005).
465 8. Bogeska, R. *et al.* Inflammatory exposure drives long-lived impairment of hematopoietic
466 stem cell self-renewal activity and accelerated aging. *Cell Stem Cell* **29**, 1273-1284.e8
467 (2022).
468 9. Kovtonyuk, L. V. *et al.* IL-1 mediates microbiome-induced inflammaging of hematopoietic
469 stem cells in mice. *Blood* **139**, 44–58 (2022).
470 10. Esplin, B. L. *et al.* Chronic exposure to a TLR ligand injures hematopoietic stem cells. *J*
471 *Immunol* **186**, 5367–5375 (2011).
472 11. Cesana, M. *et al.* A CLK3-HMGA2 alternative splicing axis impacts human hematopoietic
473 stem cell molecular identity throughout development. *Cell Stem Cell* **22**, 575-588.e7 (2018).
474 12. Goff, D. J. *et al.* A Pan-BCL2 inhibitor renders bone-marrow-resident human leukemia
475 stem cells sensitive to tyrosine kinase inhibition. *Cell Stem Cell* **12**, 316–328 (2013).
476 13. van Galen, P. *et al.* The unfolded protein response governs integrity of the haematopoietic
477 stem-cell pool during stress. *Nature* **510**, 268–272 (2014).
478 14. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of
479 aging: An expanding universe. *Cell* **186**, 243–278 (2023).
480 15. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous
481 process. *Nat Cell Biol* **19**, 271–281 (2017).
482 16. Li, H. *et al.* The dynamics of hematopoiesis over the human lifespan. *Nat Methods* 1–13
483 (2024) doi:10.1038/s41592-024-02495-0.
484 17. Roy, A. *et al.* Transitions in lineage specification and gene regulatory networks in
485 hematopoietic stem/progenitor cells over human development. *Cell Reports* **36**, 109698
486 (2021).
487 18. Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and
488 differentiation programs upon aging of hematopoietic stem cells. *Genome Res* **25**, 1860–
489 1872 (2015).

19. Zhang, Y. *et al.* Temporal molecular program of human hematopoietic stem and progenitor cells after birth. *Developmental Cell* **57**, 2745-2760.e6 (2022).
20. Li, W. *et al.* Single-cell immune aging clocks reveal inter-individual heterogeneity during infection and vaccination. *Nat Aging* **5**, 607–621 (2025).
21. Cui, L. *et al.* Single-cell transcriptomic atlas of the human testis across the reproductive lifespan. *Nat Aging* **5**, 658–674 (2025).
22. Zhu, H. *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.* **9**, eabq7599 (2023).
23. Cao, D. *et al.* Time-series single-cell transcriptomic profiling of luteal-phase endometrium uncovers dynamic characteristics and its dysregulation in recurrent implantation failures. *Nat Commun* **16**, 137 (2025).
24. Hua, P. *et al.* Single-cell analysis of bone marrow–derived CD34+ cells from children with sickle cell disease and thalassemia. *Blood* **134**, 2111–2115 (2019).
25. Psaila, B. *et al.* Single-Cell Analyses Reveal Megakaryocyte-Biased Hematopoiesis in Myelofibrosis and Identify Mutant Clone-Specific Targets. *Molecular Cell* **78**, 477-492.e8 (2020).
26. Iskander, D. *et al.* Single-cell profiling of human bone marrow progenitors reveals mechanisms of failing erythropoiesis in Diamond-Blackfan anemia. *Sci. Transl. Med.* **13**, eabf0113 (2021).
27. Calvanese, V. *et al.* Mapping human haematopoietic stem cells from haemogenic endothelium to birth. *Nature* **604**, 534–540 (2022).
28. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* **10**, 979 (2021).
29. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).
30. Kang, M. *et al.* Improved reconstruction of single-cell developmental potential with CytoTRACE 2. *Nat Methods* 1–6 (2025) doi:10.1038/s41592-025-02857-2.
31. Stassen, S. V. *et al.* StaVia: spatially and temporally aware cartography with higher-order random walks for cell atlases. *Genome Biol* **25**, 224 (2024).
32. Dai, M., Pei, X. & Wang, X.-J. Accurate and fast cell marker gene identification with COSG. *Briefings in Bioinformatics* **23**, bbab579 (2022).
33. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).
34. Van De Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**, 2247–2276 (2020).
35. Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev* **18**, 2747–2763 (2004).
36. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* **8**, e43803 (2019).
37. Vincelette, N. D. *et al.* C-MYC Augments the Proliferation and Survival of Hematopoietic Stem Cells and Multipotent Progenitors to Drive Myeloproliferative Neoplasms. *Blood* **138**, 28 (2021).

533 38. Trikha, P. *et al.* E2f1–3 Are Critical for Myeloid Development. *J Biol Chem* **286**, 4783–
534 4795 (2011).
535 39. Mgbemena, V. E. *et al.* Distinct Brca1 Mutations Differentially Reduce Hematopoietic
536 Stem Cell Function. *Cell Rep* **18**, 947–960 (2017).
537 40. Baker, S. J. *et al.* B-*myb* is an essential regulator of hematopoietic stem cell and myeloid
538 progenitor cell development. *Proceedings of the National Academy of Sciences* **111**, 3122–
539 3127 (2014).
540 41. Delestré, L. *et al.* Senescence is a Spi1-induced anti-proliferative mechanism in primary
541 hematopoietic cells. *Haematologica* **102**, 1850–1860 (2017).
542 42. Hock, H. *et al.* Gfi-1 restricts proliferation and preserves functional integrity of
543 haematopoietic stem cells. *Nature* **431**, 1002–1007 (2004).
544 43. The Cancer Genome Atlas Research Network. Genomic and Epigenomic Landscapes of
545 Adult De Novo Acute Myeloid Leukemia. *N Engl J Med* **368**, 2059–2074 (2013).
546 44. Tyner, J. W. *et al.* Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**,
547 526–531 (2018).
548 45. Wang, Z. *et al.* An immune cell atlas reveals the dynamics of human macrophage
549 specification during prenatal development. *Cell* **186**, 4454-4471.e19 (2023).
550 46. Hsiao, L. L. *et al.* A compendium of gene expression in normal human tissues. *Physiol*
551 *Genomics* **7**, 97–104 (2001).
552 47. Hounkpe, B. W., Chenou, F., de Lima, F. & De Paula, E. V. HRT Atlas v1.0 database:
553 redefining human and mouse housekeeping genes and candidate reference transcripts by
554 mining massive RNA-seq datasets. *Nucleic Acids Res* **49**, D947–D955 (2021).
555 48. Whittle, J., Meyer, S., Lacaud, G., Baker, S. M. & Iqbal, M. Single-Cell Atlas of AML
556 Reveals Age-Related Gene Regulatory Networks in t(8;21) AML. *eLife* **14**, (2025).
557 49. Manesia, J. K. *et al.* Highly proliferative primitive fetal liver hematopoietic stem cells are
558 fueled by oxidative metabolic pathways. *Stem Cell Research* **15**, 715–721 (2015).
559 50. Chen, C. *et al.* TSC–mTOR maintains quiescence and function of hematopoietic stem cells
560 by repressing mitochondrial biogenesis and reactive oxygen species. *J Exp Med* **205**, 2397–
561 2408 (2008).
562 51. Hagen, J. T. *et al.* Acute myeloid leukemia mitochondria hydrolyze ATP to support
563 oxidative metabolism and resist chemotherapy. *Science Advances* **11**, eadu5511 (2025).
564 52. Liu, Z. *et al.* Primary cilia regulate hematopoietic stem and progenitor cell specification
565 through Notch signaling in zebrafish. *Nat Commun* **10**, 1839 (2019).
566 53. Singh, M., Chaudhry, P. & Merchant, A. A. Primary cilia are present on human blood and
567 bone marrow cells and mediate Hedgehog signaling. *Experimental Hematology* **44**, 1181-
568 1187.e2 (2016).
569 54. Döhner, H. *et al.* Diagnosis and management of AML in adults: 2022 recommendations
570 from an international expert panel on behalf of the ELN. *Blood* **140**, 1345–1377 (2022).
571 55. Gabut, M. *et al.* An alternative splicing switch regulates embryonic stem cell pluripotency
572 and reprogramming. *Cell* **147**, 132–146 (2011).
573 56. Wang, F. *et al.* Single-cell architecture and functional requirement of alternative splicing
574 during hematopoietic stem cell formation. *Sci. Adv.* **8**, eabg5369 (2022).
575 57. Galen, P. van *et al.* Integrated Stress Response Activity Marks Stem Cells in Normal
576 Hematopoiesis and Leukemia. *Cell Reports* **25**, 1109-1117.e5 (2018).

58. Kharabi Masouleh, B. *et al.* Mechanistic rationale for targeting the unfolded protein response in pre-B acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences* **111**, E2219–E2228 (2014).
59. Safina, K. R. *et al.* An inflammatory and quiescent HSC subpopulation expands with age in humans.
60. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).
61. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
62. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).
63. Buckley, M. T. *et al.* Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. *Nat Aging* **3**, 121–137 (2023).
64. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**, 1–22 (2010).
65. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Royal Statistical Society. Journal. Series B: Methodological* **58**, 267–288 (1996).
66. Zou, H. & Hastie, T. Regularization and Variable Selection Via the Elastic Net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67**, 301–320 (2005).
67. Ho, T. K. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition* vol. 1 278–282 vol.1 (1995).
68. 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* 785–794 (2016). doi:10.1145/2939672.2939785.
69. Hinton, G. E. Connectionist learning procedures. *Artificial Intelligence* **40**, 185–234 (1989).

Competing Interests

All authors declare no conflict of interest.

Author Contributions

Conceptualization: G.W. & H.C.; Methodology: G.W. & H.C.; Formal analysis H.C. & P.D.; Investigation: H.C. & J.X.; Data curation: H.C.; Writing: G.W. & H.C.; Visualization: H.C.; Supervision: G.W.; Funding acquisition: G.W. All authors have read and approved the final manuscript.

Figures

Figure 1

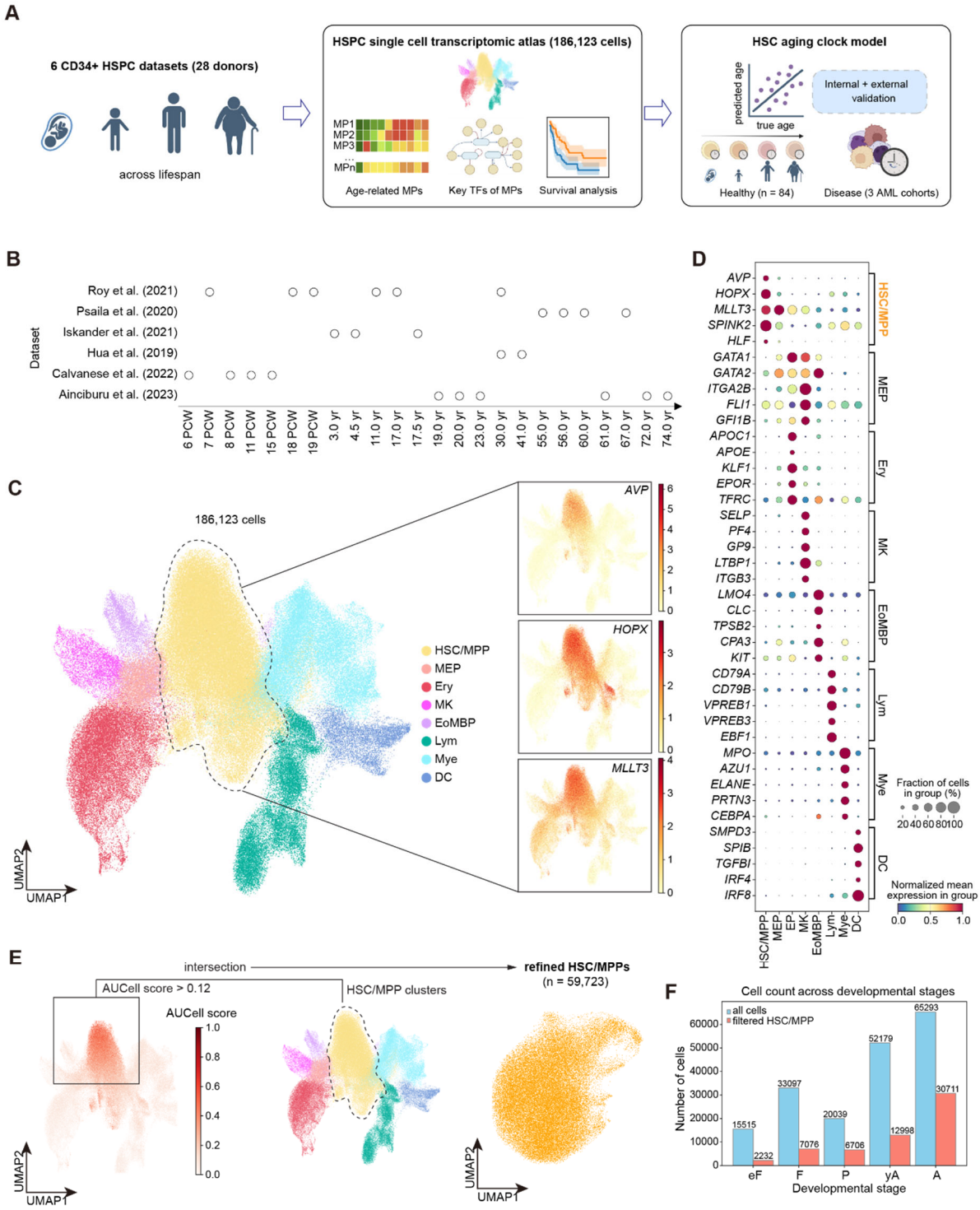

Fig. 1. Construction of a human lifespan HSPC single-cell transcriptome atlas. **A**. Schematic workflow of the study. **B**. Sample sources for this study. PCW, post-conception week; yr, age. **C**. UMAP dimensionality reduction plot of the human HSPC transcriptome atlas and highlighted expression of HSC marker genes (*AVP, HOPX, MLLT3*) on the UMAP projection. **D**. Normalized expression of lineage-specific marker genes in HSPCs. **E**. HSC/MPP filtration based on clustering

and AUCell enrichment. Distribution of HSPC AUCell scores on UMAP projection (left), clustering result (middle), and UMAP projection of refined HSC/MPPs (right) were shown. HSC/MPP, hematopoietic stem cells and multipotent progenitors; MEP, megakaryocyte-erythroid progenitors; EP, erythroid progenitors; MK, megakaryocytes; EoMBP, eosinophil-basophil-mast progenitors; Lym, lymphoid progenitors; Mye, myeloid progenitors; DC, dendritic cells. Fig. 1A were created using Biorender (www.biorender.com).

Figure 2

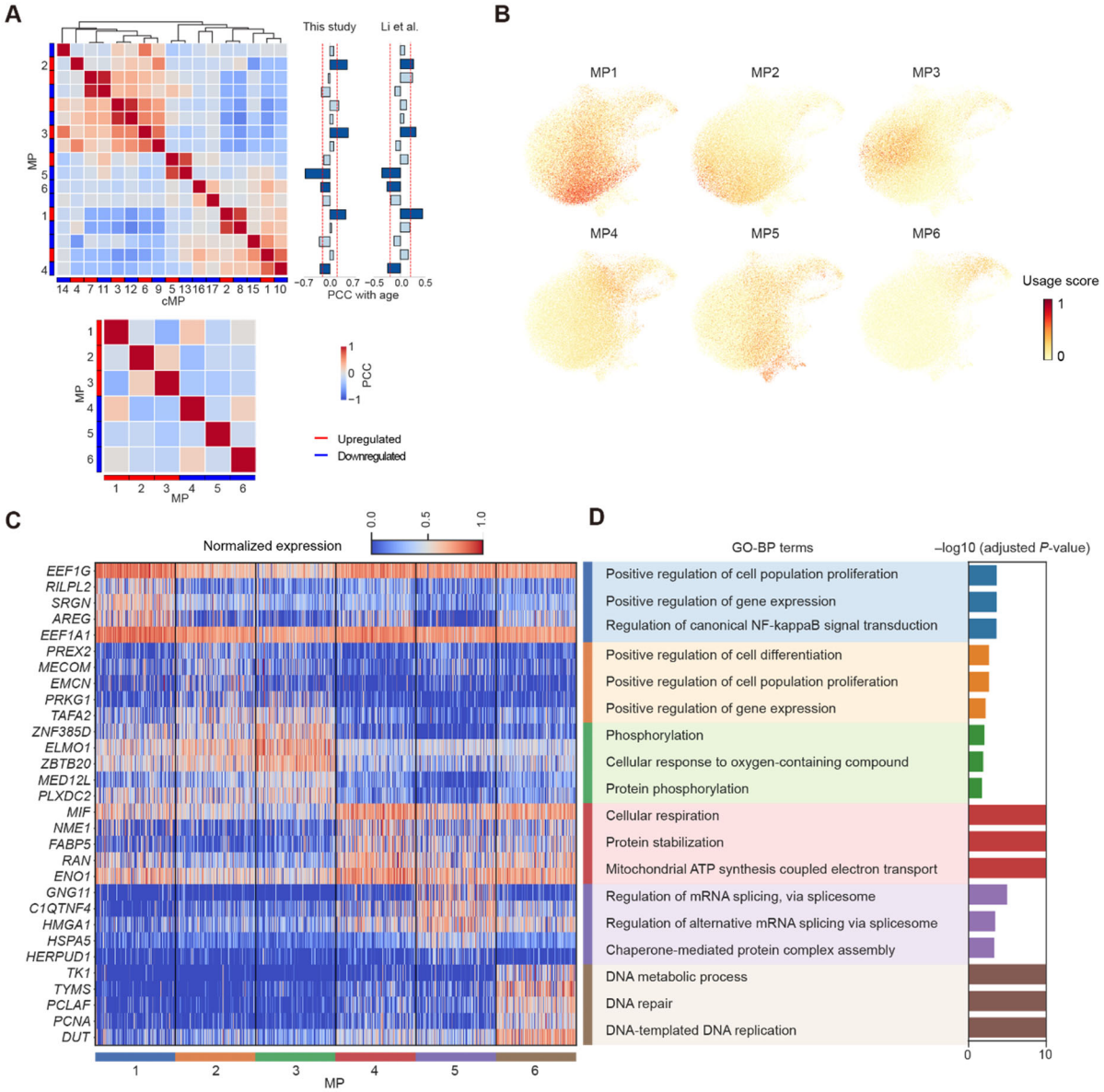

Fig. 2. Age-related molecular programs in lifespan HSC/MPPs. **A**. Correlation of 17 candidate molecular programs (cMPs, upper part) and 6 robust molecular programs (MPs, lower part). Correlation between cMP usage scores and age is shown on the right. 6 cMPs with Pearson's Correlation Coefficient (PCC)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

Figure 3

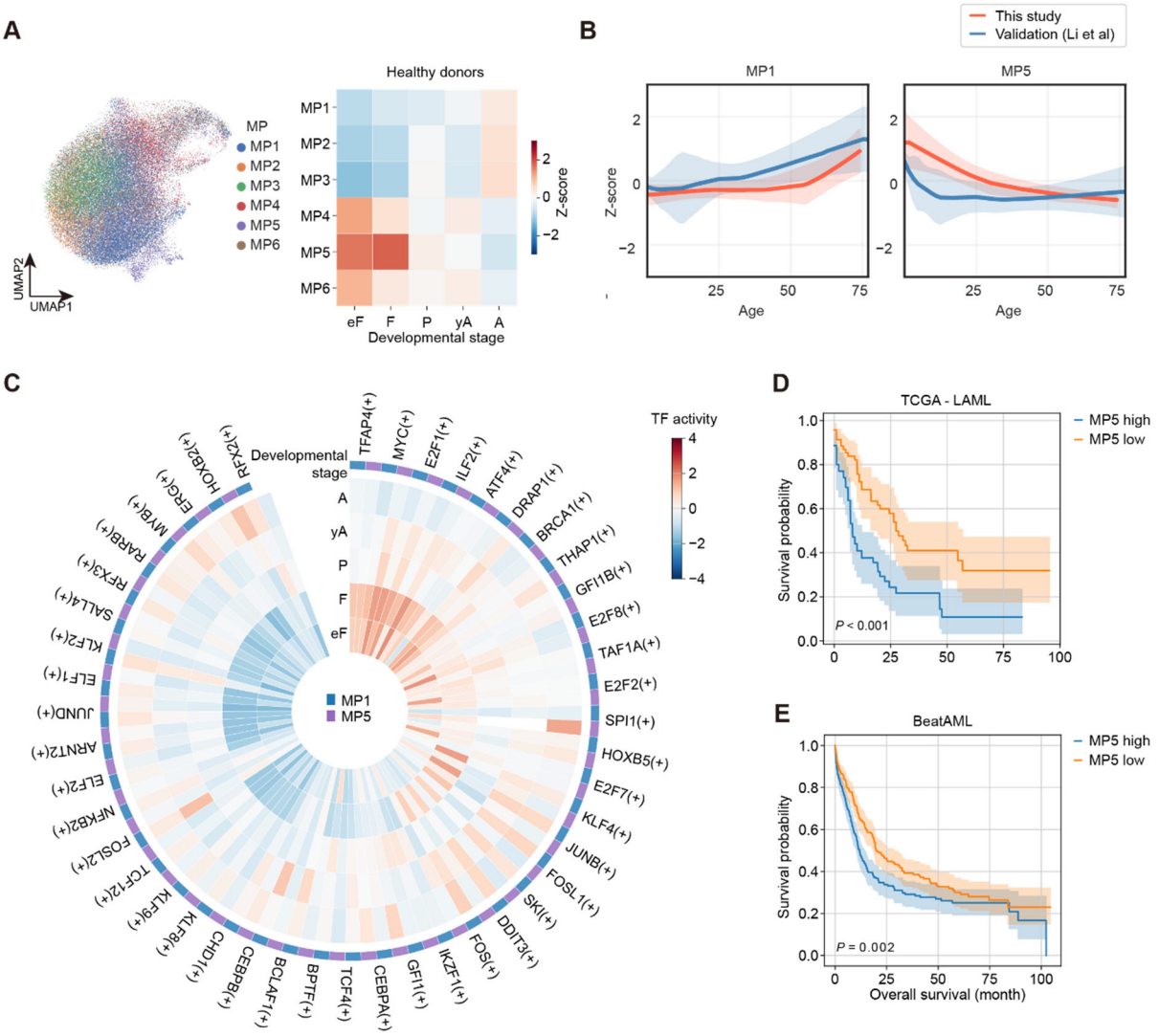

Fig. 3. Characterization of molecular programs across developmental stages. **A**. Distribution and activity of MPs across the five developmental stages. **B**. Temporal dynamics of the two age-related MPs. **C**. Activities of key transcription factors driving MP1 and MP5 across developmental stages.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

Figure 4

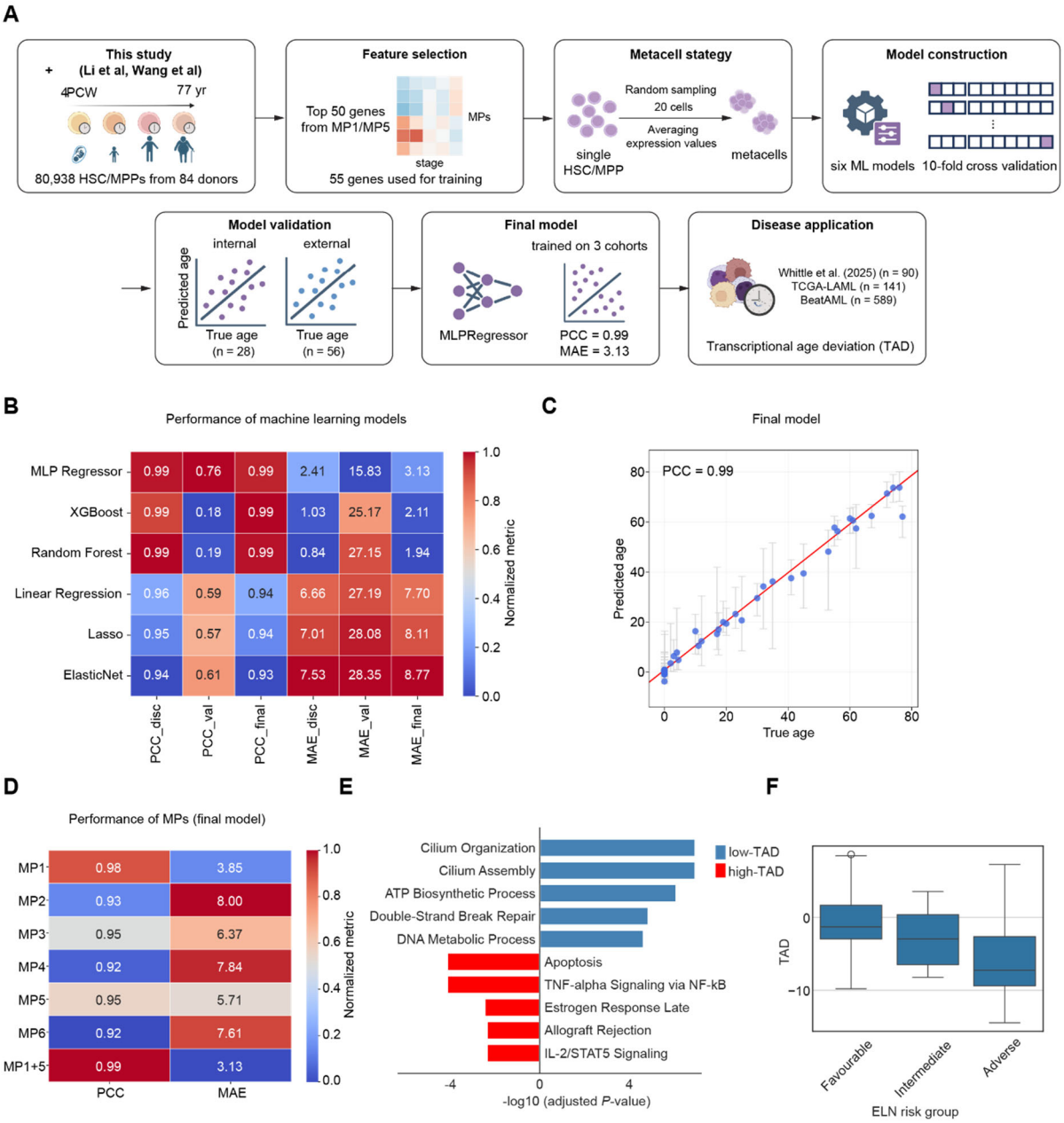

Fig. 4. A novel HSC/MPP aging clock model across the human lifespan. **A**. Workflow of aging clock model construction, evalutation and application. The main steps included processing of discovery and independent validation cohorts, feature selection, metacell calculation, machine learning (ML) model construction, internal and external validation, final model training on all 3 cohorts, and application of the model on AML datasets provided by Whittle et al[48]. **B**. Performance of six common machine learning models across the discovery, validation and combined cohorts. The models are ranked by their mean absolute error (MAE) on the validation cohort. **C**. Performance of the final model (trained on 3 cohorts). **D**. Performance of the final model trained on 3 cohorts using each MP and a combined [MP1 + MP5] as the model's features for training. **E**. Pathway enrichment results using differentially expressed genes between metacells belonging to

the low-TAD group and the high-TAD group. **F**. Distribution of TAD among patients stratified by ELN genetic risk groups. Disc, discovery cohort; val, validation cohort; final, combined cohorts; TAD, transcriptional age deviation. Fig. 4A were created using Biorender (www.biorender.com).