Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single cell resolution

Marina Ainciburu[1,2*], Teresa Ezponda[1,2*], Nerea Berastegui[1], Ana Alfonso-Pierola[2,3], Amaia Vilas-Zornoza[1,2], Patxi San Martin-Uriz[1,2], Diego Alignani[4], Jose Lamo de Espinosa[3] Mikel San Julian[3], Tamara Jimenez[5], Félix López[5], Sandra Muntion[5,9], Fermin Sanchez-Guijo[5,9], Antonieta Molero[6], Julia Montoro[6], Guillermo Serrano[8], Aintzane Diaz-Mazkiaran[2,8], Miren Lasaga[7], David Gomez-Cabrero[7], María Díez-Campelo[5], David Valcarcel[6], Mikel Hernaez[8], Juan P. Romero[1,2, §], Felipe Prosper[1,2,3,9, §]

[1]Area de Hemato-Oncología, Centro de Investigación Médica Aplicada (CIMA), Universidad de Navarra, IDISNA, Pamplona, Spain.
[2]Centro de Investigación Biomédica en Red de Cáncer, CIBERONC.
[3]Clínica Universidad de Navarra, Pamplona, Spain
[4]Flow Cytometry Core, CIMA, Universidad de Navarra, Pamplona, Spain.
[5]Hospital Universitario de Salamanca-IBSAL, Salamanca, Spain
[6]Department of Hematology, University Hospital Vall d'Hebron, Barcelona, Spain
[7]Translational Bioinformatics Unit, NavarraBiomed, Pamplona, Spain
[8]Computational Biology Program, CIMA, Universidad de Navarra, Pamplona, Spain
[9]Red de Investigación Cooperativa en Terapia Celular TerCel, ISCIII.

* contributed equally and should be considered joint first authors
§ contributed equally and should be considered joint senior authors.

Email addresses

Marina Ainciburu	mainciburu@alumni.unav.es
Teresa Ezponda	tezponda@unav.es
Nerea Berastegui	nberastegui@ alumni.unav.es
Ana Alfonso-Pierola	aalfonso@unav.es
Amaia Vilas-Zornoza	avilaszo@unav.es
Patxi San Martin-Uriz	psanmart@unav.es
Diego Alignani	dalignani@unav.es
Jose Lamo de Espinosa	jlamodeespi@unav.es
Mikel San Julian	msjulian@unav.es
Tamara Jimenez	tamara_js@usal.es
Félix López	flcadenas@saludcastillayleon.es
Sandra Muntion	smuntion@usal.es
Fermín Sanchez-Guijo	ferminsg@usal.es
Antonieta Molero	amolero@vhio.net
Julia Montoro	jmontoro@vhio.net
Guillermo Serrano	gserranos@unav.es
Aintzane Diaz-Mazkiaran	adiazma@unav.es
Miren Lasaga	mlagasag@unav.es
David Gomez-Cabrero	david.gomez.cabrero@navarra.es
María Díez-Campelo	mdiezcampelo@usal.es
David Valcarcel	dvalcarcel@vhio.net
Mikel Hernaez	mhernaez@unav.es
Juan P. Romero	jromeror@unav.es
Felipe Prosper	fprosper@unav.es

Institutional addresses

CIMA	Av. de Pío XII, 55, 31008 Pamplona, Navarra, Spain
Clínica Universidad de Navarra	Av. de Pío XII, 36, 31008 Pamplona, Navarra, Spain
University of Navarra	Campus Universitario, 31009 Pamplona, Navarra,Spain
Hospital Universitario de Salamanca-IBSAL, Paseo de San Vicente 58-182, 37007, Salamanca, Spain.	
University Hospital Vall d'Hebron	Passeig de la Vall d'Hebron, 119, 08035 Barcelona, Spain
NavarraBiomed	Calle de Irunlarrea, 3, 31008 Pamplona, Navarra, Spain

ABSTRACT

Background

Early hematopoiesis is a continuous process in which hematopoietic stem and progenitor cells (HSPCs) gradually differentiate and are primed toward specific lineages. Aging and myeloid malignant transformation are characterized by changes in the composition and regulation of HSPCs. In this study, we evaluated human HSPCs obtained from young and elderly healthy donors using single-cell RNA sequencing to identify the transcriptional and regulatory perturbations associated with healthy aging at single cell resolution. We then applied this knowledge to identify specific changes associated with the development of myeloid malignancies.

Results

Based on the transcriptional profile obtained, we identified changes in the proportions of progenitor compartments during aging, and differences in their functionality, as evidenced by gene set enrichment analysis. Trajectory inference revealed that altered gene expression dynamics accompanied cell differentiation, which could explain age-associated aberrant hematopoiesis. Next, we focused on key regulators of transcription by constructing gene regulatory networks and detected regulons that were specifically active in elderly individuals. Using the previous findings as a reference, we analyzed scRNA-seq data obtained from patients with myelodysplastic syndrome and acute myeloid leukemia and detected an alteration of the expression dynamics of genes involved in erythroid differentiation and identified specific transcription factors deregulated in acute myeloid leukemia.

Conclusions

We demonstrate that the combination of single cell technologies and computational tools enables the study of a variety of cellular mechanisms involved in early hematopoiesis and can be used to dissect perturbed differentiation trajectories associated with aging and malignant transformation. Furthermore, the identification of abnormal regulatory mechanisms associated with myeloid malignancies could be exploited for personalized therapeutic approaches in individual patients.

Keywords

Single Cell, Aging, Hematopoiesis, Genomics, Myelodysplastic Syndromes, Acute Myeloid Leukemia

BACKGROUND

Mature blood and immune cells are generated by hematopoiesis, which is a well-characterized process that has been studied for more than a century (1). Classically, hematopoiesis has been modeled as a stepwise differentiation process, at the core of which reside hematopoietic stem cells (HSCs), which are common precursors with self-renewal capacity. A tree-like hierarchy arises from HSCs, in which lineage commitment occurs at binary branching points, giving rise to functionally and phenotypically homogeneous progenitor populations (2,3). However, recent single cell studies have questioned the validity of the classical model of hematopoiesis and have revealed heterogeneity within the HSC compartment and within progenitor populations that were previously considered homogeneous (4). Furthermore, it is difficult to stablish boundaries between populations, which has led to the replacement of this concept by the idea of smooth transitions. Thus, early hematopoiesis is now viewed as a continuous landscape composed of undifferentiated hematopoietic stem and progenitor cells (HSPCs) with a variable degree of priming toward specific lineages (lymphoid, myeloid, or erythroid) (5–8).

During aging, the hematopoietic system undergoes various changes. There is evidence of an age-related increase in the relative number of HSCs, that presumably aims to compensate for a loss of repopulation ability (9,10). Despite this increase in the number of HSCs, the total bone marrow cellularity decreases (11). Additionally, a loss of lymphocyte production and a skewing toward myeloid differentiation have been demonstrated (12). Moreover, recent studies have revealed an increase in the number of platelet primed HSCs (13). A general loss of immune function that affects both innate and adaptive immunity has also been associated with aging (14). The molecular basis of these phenotypic changes includes an increased rate of random mutations in hematopoietic progenitors. Clonal hematopoiesis is a common trait of elderly individuals, in which blood cell subpopulations are clonally derived from a single HSC or progenitor with acquired mutations. Notably, some of these mutations are also associated with hematopoietic malignancies (15–18). Additionally, aging is also accompanied by transcriptional dysregulation. Gene expression analyses have revealed changes in cell cycle regulators (19), higher expression of myeloid signatures (13,19) and genes associated with leukemia (20), stronger response to inflammatory stimuli (21,22), and upregulation of pathways such as nuclear factor kappa beta (NF-κB) or tumor necrosis factor alpha (TNF-α) and downregulation of DNA repair (23). Lastly, epigenetic modifications that are in line with the transcriptional lesions detected have been observed in HSPCs obtained from elderly individuals (24).

Aging is associated with a higher risk of developing myeloid malignancies (25,26), suggesting a strong predisposition of hematopoietic cells from elderly individuals to lead to further alterations. Myelodysplastic syndromes (MDS) are among the main age-related hematological disorders. MDS are characterized by ineffective hematopoiesis and predisposition to transformation into acute myeloid leukemia (AML) (27), a highly aggressive neoplasm that is characterized by defective maturation and uncontrolled proliferation of myeloid progenitor cells.

Collectively, previous evidence suggests the existence of a progressive decline in the hematopoietic system with age, which can result in disease if certain pathological events occur. HSC damage is the primary consequence of this process and its outcomes range from lineage bias to differentiation blockade and leukemic transformation (28). In this study, we examined these successive alterations in detail by considering hematopoiesis as a continuous process and analyzing the single cell transcriptional profile of early hematopoietic progenitors. We aimed to characterize key regulators of hematopoiesis in young and aged hematopoiesis and to apply this knowledge to identify specific perturbations occurring in patients with myeloid malignancies. These regulators have the potential to predict, in a personalized manner, what abnormalities to expect and to identify potential therapeutic targets even at the early stages of transformation.

RESULTS

Transcriptional profiling of human young and elderly hematopoietic progenitor systems

To investigate the age-dependent changes in the hematopoietic system, we performed single-cell RNA sequencing (scRNA-seq) of bone marrow $CD34^+$ cells obtained from five young (18-20-year-old) and three elderly (>65-year-old) healthy donors **(Figure 1a)**. Briefly, single-cell libraries were prepared using a Chromium controller instrument (10x Genomics) and sequenced up to an average depth of 30,000 reads per cell. A total of 34,590 and 40,641 cells were profiled from young and elderly donors, respectively. We first constructed an integrated dataset with cells from both young and elderly donors using an integration procedure implemented in the Seurat R package (29). Subsequently, we subjected the integrated dataset to quality control filtering and dimensional reduction to obtain a reference map for visualization (**Supplemental Table 1**). Secondary analyses including unsupervised clustering, identification of cluster markers, trajectory inference and gene regulatory network (GRN) reconstruction were performed independently for each dataset.

After integration, we extracted the cells obtained from young donors and subjected them to unsupervised clustering and differential expression analysis to identify cluster markers for manual cell type annotation and labelling (**Supplemental Table 2)**. This analysis identified 14 cellular subpopulations comprising the landscape of HSCs, early progenitors (MEP, LMPP, CLP, and GMP cells), and cells already committed to specific lineages (Pro-B cells, monocytic, erythroid, megakaryocytic, basophil and dendritic cell progenitors) (**Figure 1b).** Early progenitors were characterized by the expression of genes such as *CRHBP*, *HOPX* (initial lympho-myeloid differentiation), and *PBX1* (initial megakaryocyte and erythroid lineage). More mature progenitors presented myeloid (*MPO*, *CTSG*, *PRTN3*, *LYZ*, and *IRF8*), lymphoid (*DNTT*, *VPREB1*, and *EBF1*), erythroid (*HBD*, and *CA1*) and basophil (*HDC*, and *MS4A2*)-specific markers **(Figure 1c).** Once cellular subpopulations from the young donors were identified, we labeled cells of elderly individuals using a classification method based on a logistic regression with elastic-net regularization (see Materials and Methods). Regularized logistic regression has been previously applied to scRNA-seq data, due to its high interpretability and good performance with sparse input (30,31). We created binary classifiers that showed the probability that each cell belongs to a specific cluster **(Supplemental Figure 2a)** and assigned labels based on the highest probability. We noted a significant increase in the proportion of HSCs and a reduction in the number of both committed lymphoid and myeloid lineages (Adj p.value < 0.05, **Supplemental Table 3**) as well as an increase in the proportion of erythroid-committed cells in the overall proportion, both in the integrated and individual datasets **(Figure 1d and Supplemental Figure 1).** The performance of our classifier was determined by using a publicly available dataset containing 8,176 human $CD34^+$ progenitors with known identities (32) **(Supplemental Figure 2b)**. We calculated the Jaccard Index to assess the similarity of the cellular populations and observed that clusters within the same lineage displayed the highest values **(Supplemental Figure 2c)**. As both datasets include early progenitors, it is not straight forward to state a clear point on which a specific immature cell type becomes part of a more mature population. This can be observed as our monocytic progenitor population was part of the GMP population from the reference study.

Next, we checked for functional differences between the young and elderly populations using gene set enrichment analysis (GSEA) after performing differential expression per cluster **(Supplemental Table 4)**. Cells from elderly donors exhibited a generalized enrichment of pathways activated in response to external molecules and insults, such as TNF-α, transforming growth factor beta (TGF-β), hypoxia or inflammation. Furthermore, the p53 pathway and apoptosis programs, which are activated by stressor stimuli, were also enriched in these cells **(Figure 1e)**. In particular, the TGF-β response and apoptosis genes, together with genes upregulated in response to ultraviolet radiation were specially enriched in age-associated HSCs, indicating an enhanced accumulation of transcriptional lesions in this compartment. These results suggested the presence of an age-related response to an inflammatory microenvironment in elderly individuals. Shifts in the abovementioned stress-related pathways were primarily driven by genes such as *JUNB*, *FOSB*, *FOS* ID2 or *DDIT4* (**Supplemental Figure 3**). On the other hand, we found that cell cycle progression (E2F targets) and proliferation (MYC targets) pathways were enriched in multiple cell populations from young donors **(Figure 1e)**, whereas

MYC expression was generally downregulated in cells from elderly individuals (**Supplemental Figure 3**). Hematopoietic progenitors from young donors were characterized by an increase in metabolism, including signatures related to glycolysis, fatty acid metabolism, and oxidative phosphorylation, in components of the respiratory chain (*UQCR*, *NDUF*, and *COX*), and the glycolytic enzymes *PGK1* and *TPI1*. These findings indicate an increased metabolic activity in progenitors from young donors (**Supplemental Figure 3**). DNA repair was also found to be more active in cells isolated from young individuals, suggesting a greater predisposition of age-associated progenitors to accumulate genetic lesions.

Overall, we found that the proportion of cell types changed with age, which indicated an inefficient maturation of the lymphoid and myeloid lineages. Furthermore, GSEA revealed age-related transcriptional alterations in all compartments, which suggested a general decline in the fitness of HSPC, which is probably related to their response to the microenvironmental stimuli, and to proliferation and differentiation cues.

Trajectory inference revealed age-specific drivers along lineages

To study the distinct differentiation trajectories observed during hematopoiesis and, specifically, alterations during aging, we subjected the single-cell expression datasets to trajectory inference using STREAM (33). We first inferred the differentiation trajectories of the young donors to obtain a reference that could serve as a basis for comparison. These trajectories revealed a common starting point that comprised early progenitors (HSCs, LMPPs, and MEPs). This initial root reached the branching point that divides cellular differentiation into the three main branches that resemble the main hematopoietic lineages: the myeloid, erythroid, and lymphoid lineages (**Figure 2a).** The identity of these branches was confirmed by plotting the expression profiles of previously described genes, such as *IRF8*, *GATA1*, and *EBF1* (**Figure 2b).** Then, in order to identify age-dependent alterations, the elderly dataset was projected onto the reference trajectories using the approach described in STREAM, (**Figure 2c**). We observed an increased proportion of HSCs in the root node, a reduced number of cells biased to both the myeloid and lymphoid lineages, and an increased number of progenitor cells committed to the erythroid compartment **(Figure 2d**).

Next, we aimed to identify shifts in the gene expression dynamics along differentiation trajectories. We applied the Palantir algorithm (34), with which we recovered a general pseudotime (**Supplemental Figure 4a**) and differentiation potential (**Supplemental Figure 4b**) for each cell. A greater differentiation potential in age-associated HSPCs and a lower pseudotime suggested an immature cell state, which was consistent with our previous analysis. We could also define six trajectories corresponding to six committed compartments: i.e., erythroid, lymphoid, dendritic cells, monocytes, basophils, and megakaryocytes, and assigned each cell a probability of belonging to a specific branch **(Supplemental Figure 4c)**.

We next focused on monocytic differentiation, which appeared to be impaired in the elderly donors, based on our STREAM results. We plotted, in a per-cell bases, the associated pseudotime along the monocytic trajectory with the Palantir probability of reaching the final stage of the studied route (**Figure 2e**). This analysis, revealed that HSCs and LMPPs from elderly donors had a higher probability of attaining the monocyte progenitor state, suggesting a stronger bias towards the monocytic compartment. However, GMPs displayed the opposite behavior which can indicate that although age-associated progenitors appeared to have a stronger initial bias toward this lineage, a large number of more advanced progenitor cells lost the capacity for monocytic differentiation. This result is in line with the aged-associated decrease in monocytic progenitors described above.

To determine the transcriptional lesions that potentially alter monocytic differentiation during aging, we computed gene expression trends and clustered them based on their dynamics along the monocytic differentiation trajectory, following the Palantir pipeline. We characterized young gene clusters by over-representation analysis. Terms related to the immune system and monocytic function were enriched as the trajectory progressed, such as antimicrobial response or regulation of the actin cytoskeleton in intermediate stages, and antigen processing and

presentation, or interferon gamma signaling, in gene clusters expressed near the terminal state. On computing trends for the cells of elderly donors, we observed that the overall behavior was similar. Nevertheless, genes with altered expression patterns in the elderly donors were visible (**Figure 2f**). Several myeloid differentiation markers (*CST7, PRTN3, MPO, CD74, CALR*, and *GNAS*) were expressed at lower levels in the elderly donors, indicating a less differentiated state. Accordingly, genes characteristic of stem or very early progenitor cells (*MYCT1, MLLT3,* and *ALDH1A1*) showed higher levels of expression in cells from the elderly donors across differentiation. On the other hand, genes related to stress and inflammation response (*FOS, JUNB, TSC22D3, DUSP1,* and *DDIT4*) or monocyte chemotaxis and extravasation (*ANXA1, LGALS3*, and *JAML*), exhibited higher expression levels at the start of differentiation in the elderly, with several genes showing decreased expression levels at the terminal state (**Figure 2g**), thereby corroborating a loss of capacity for monocytic differentiation in cells from elderly donors.

These analyses of the differentiation trajectory provided data regarding the genes that may be involved in the aberrant differentiation of age-associated HSCs. However, the regulation of these transcriptional programs cannot be inferred from this analysis.

GRNs guiding young and elderly hematopoiesis

To elucidate the regulatory mechanisms underlying healthy hematopoietic differentiation, we constructed lineage-specific GRNs using SCENIC (35) for each of the datasets independently. Briefly, we obtained the set of activation regulons (transcription factor and their associated targets) for each of the cellular subpopulations and binarized its activity (on/off) in a per-cell basis using AUC values provided by SCENIC (through the AUCell algorithm). This approach enabled quantification of the proportion of cells that displayed an activated state for each regulon in each cluster. We then filtered the regulons based on the regulon specificity score (RSS) **(Supplemental Figure 5a and 5b)** to identify specific regulatory mechanisms per cell type. For the young donors, we obtained a series of GRNs guided by transcription factors that were previously considered relevant for specific cell populations, such as *HOXA9* in early progenitor cells, or *CEBPG*, *PAX5*, and *GATA1* in the myeloid, lymphoid and erythroid compartment, respectively (**Figure 3a)**. The AUC values of some of these regulons displayed a behavior similar to that of the corresponding expression profiles, whereas others such as CEBPG or TCF3 displayed a more specific dynamic of GRN activity (AUC). For each of the hematopoietic lineages, we observed sets of regulons that regulated the transcriptional state of specific cellular differentiation programs and also the presence of common targets among different transcription factors (**Figure 3b**). This observation indicates that a single transcription factor not only regulates cellular differentiation toward a specific route, but also contributes to other regulatory elements in other differentiation trajectories.

We performed the same analysis using the elderly dataset and observed that progenitor populations displayed a reduced proportion of cells with activated regulons. Specifically, HSCs showed a less active state of transcriptions factors, such as *MECOM* and *GATA3*, that were deemed specific to young donors (**Figure 3c)**. In already committed cellular subpopulations, we observed similar proportions of activated cells between young and elderly donors, with the erythroid lineage showing the most similar profiles. However, progenitors (GMP, MEP, and CLP cells) from elderly donors showed a lower number of cells with specific regulons activated. Overall, we observed reduced numbers of common targets **(Figure 3d)** between transcription factors, specifically for HSCs, which could result in the loss of co-regulatory mechanisms in the elderly donors.

We then performed gene ontology enrichment analyses using the highest ranked regulons in early progenitors (HSC, GMP, CLP, and LMPP cells). HSCs from the young donors displayed enriched terms related to the differentiation of hematopoietic lineages, such as myeloid cell differentiation, lymphocyte cell differentiation and regulation of hemopoiesis, whereas HSCs from the elderly donors did not exhibit such enrichment, (**Figure 3e)**, suggesting a loss of differentiation potential among aged HSCs. Although no increase in the percentage of LMPPs was observed in the elderly donors, this analysis indicated a clear enrichment in terms related to

DNA replication in the elderly but not in the young donors. These results suggest an aged-mediated increase in the proliferation potential of LMPPs, which could represent a compensatory mechanism for the less efficient differentiation process present in elderly individuals. More mature progenitors also showed a youth-specific enrichment of differentiation processes such as lymphocyte differentiation, B-cell receptor signaling pathway, or T- and B-cell activation for CLPs, and erythrocyte differentiation and homeostasis for MEPs suggesting that the main drivers of regulation in age-associated progenitor cells undergo changes that results in the loss of expression of genes that are required for the proper hematopoietic differentiation of different lineages. These results indicate the existence of complex regulatory mechanisms in healthy hematopoiesis as different regulons participate in a coordinated manner throughout cellular differentiation processes. There was a loss of differentiation capabilities in early progenitors, which may contribute to the understanding of how pathological conditions introduce alterations in the biological systems in homeostasis.

scRNA-seq analysis of MDS and AML specimens reveals molecular lesions affecting normal hematopoietic differentiation

To further demonstrate the potential of the used computational methods in the prediction of the mechanisms underlying aberrant hematopoietic differentiation, we explored the transcriptional alterations that distinguish normal hematopoiesis during aging from abnormal hematopoiesis associated with the development of myeloid malignancies. To this end, we performed scRNA-seq for bone marrow $CD34^+$ cells from two samples obtained from one patient with MDS and one patient with AML.

We recovered 10,765 cells that passed the quality control from a patient with high-risk MDS with multilineage dysplasia (**Supplemental Table 5)**. The cell types were predicted using the GLMnet classification algorithm, as described previously (**Figure 4a**). We noted the absence of lymphoid compartments, considerable reduction in the number of committed myeloid progenitors (monocytic and dendritic cell progenitors), and accumulation of immature stages (LMPP and GMP cells). The erythroid compartments also appeared to be diminished, especially the early erythroid progenitor cluster (**Figure 4b**). These changes were also apparent when cells were mapped to the reference trajectory created using STREAM, as well as the cell distribution along the three principal branches (**Figure 4c**). We found that erythroid and megakaryocytic cells aligned correctly to the reference, despite the known reduction in the number of their progeny in peripheral blood.

Based on the cytopenias present in the patient with MDS characterized by anemia, we explored gene dynamics along the HSC-erythroid branch. We reasoned that transcriptional alterations already present in $CD34^+$ cells could contribute to the observed decrease in the early erythroid progenitor cluster, or to the aberrant differentiation occurring at more mature stages. We reconstructed the erythroid trajectory and computed gene trends along pseudotime using Palantir (**Supplemental Figure 6a**). Genes that were specifically expressed by any of the progenitors belonging to this branch (HSCs, MEPs, and erythroid progenitors) were selected and clustered according to their expression pattern (**Figure 4d**). Although, we did not detect broad changes in the transcriptional profiles among the conditions, we identified specific genes participating in erythroid differentiation that displayed different dynamics from both young and elderly healthy samples (**Figure 4d**). HSPCs from the patient with MDS showed an increased levels of transcription factors involved in the prevention of erythropoiesis such as *JUN*, *YBX1*, *LMO2*, and *LYL1* (36,37). On the other hand, the expression of genes such as *SLC25A6*, *EPSTI1* and *PHF6*, which play a prominent role in the promotion of erythroid differentiation, failed to raise in HSPCs from this patient (38,39). Despite these changes, we observed that essential drivers of erythropoiesis (*GATA1*, and *KLF1*) as well as erythroid specific markers (*AHSP*, *CA1*, and *HBB*) followed trends similar to those of healthy HSPCs, which is in agreement with the apparently normal progression of differentiation at the progenitor stages. These results indicate that, although at very early progenitor stages some drivers of erythroid differentiation exhibit normal trends of expression, others show clear abnormalities that may manifest in more mature stages as erythroid differentiation defects and may be responsible for the observed cytopenia.

In the case of the patient with AML, 14,049 $CD34^+$ cells were analyzed using scRNA-seq. Cell identity prediction revealed an increase in the number of HSCs and early progenitors, both LMPP and MEP cells, together with a reduction in the proportion of the remainder populations (**Figure 4a**). Using SCENIC we analyzed the transcriptional programs regulating HSPCs and reconstructed GRNs by extracting those regulons that were specifically active in each of the cellular compartments (**Supplemental Figure 6b**). We identified active GRNs that were guided by transcription factors characteristic of the different populations (**Figure 4e**), such as GATA3 and MECOM in HSCs or GATA1 and KLF1 in erythroid compartments. Nevertheless, we characterized regulons that were not present in HSPCs from young or elderly samples and were only identified during AML development. Among them, we detected two factors that were specifically active in HSCs: *ZSCAN18*, which is highly expressed in leukemic stem cells, and *IRF2* (40), a regulator of cell growth and survival. In addition, we found that *GFI1* was prominently active at the myeloid progenitor compartments, whereas *SIRT6* was active across all compartments. Both genes have been implicated in the development of AML (41,42). Collectively, these data demonstrate the power of these analyses to identify deregulated GRN in patients with myeloid malignancies, which could be exploited therapeutically in individual patients.

DISCUSSION

In this study we report the generation of high-resolution scRNA maps of human HSPCs and the changes associated with aging. Although previous studies focused on other species (22,43) and other layers of information such as mutations (44), proteomics (45), and proteo-genomics (46), our study represents one of the first analyses that describes early human hematopoiesis based on its dynamic gene expression and transcriptional regulation by identifying age related changes that may be responsible for some of the phenotypic modifications observed during healthy aging. Applying this knowledge to HSPCs obtained from patients with AML and MDS, we establish as a proof of concept of how this analysis has the potential to help identifying transcriptional alterations that play potential roles during disease development, which may eventually lead to the identification of potential therapeutic targets.

Single-cell experiments are providing data at an unprecedented scale, which has allowed the identification of novel cell types and alterations that occur in biological systems. This increase in resolution has proceeded concurrently with the development of multiple computational tools that aspire to solve common problems arising with these technologies. An essential dilemma of this approach is how to establish reliable identities of the cells that will be used for subsequent analysis (47). In this study, we generated a reference system based on the transcriptional profile of HSPCs from healthy young donors and used it to assign labels to cells from elderly healthy and pathological donors. We generated an in-house cell classifier that allowed us to reduce the effect of technical artifacts that can contribute to erroneous prediction of cell identity. This classifier was applied to cells that underwent the same sorting, library preparation and sequencing procedures; thus, we expected to identify similar cell populations in each dataset. Additionally, when applied to external data, we found minor differences from the originally established labels, with the three lineages as well as the most immature states being precisely identified.

In accordance with previous studies, our analyses revealed an age-driven expansion of HSCs, which may represent a compensatory mechanism for less efficient hematopoiesis (9,10). We also noted a decrease in the proportion of lymphoid precursors, which is in agreement with the described aged-associated decrease in lymphoid differentiation capacity, as described previously (12). Using functional analyses, we also detected an enrichment of pathways related to apoptosis and inflammatory conditions in early progenitors among elderly donors, which can be explained by the inflammatory microenvironment known to be present in elderly individuals (48). Conversely, cells from young donors showed highly active differentiation and proliferation profiles, reinforcing the idea of the higher differentiation capacity among young individuals (28). In addition to these analyses, our study incorporated two other approaches: the characterization of transcriptional alterations across differentiation trajectories and the analysis

of GRNs that guide different hematopoietic stages. As an example, these analyses allowed for the detection of an abnormal transcriptional rewiring across monocytic development during aging. We observed that although the most immature clusters (HSCs and LMPPs) showed an increased monocytic potential of differentiation in the elderly, more mature precursor cells presented reduced potential. These results would suggest that early immature cells are more primed toward the monocytic lineage during aging, but alterations during later stages of differentiation result in a loss of monocytic differentiation capacity and, accordingly, a delay in the expression of genes that characterize the monocytic lineage. In agreement with our results, an age-associated skewing toward myeloid-biased HSCs and multipotent progenitor compartments has been described in mice (49). By focusing on GRNs, we observed that aged progenitor cell types showed a decrease in the activity of specific regulons, a lower degree of interaction between transcription factors and targets, and no enrichment of pathways involved in the cellular differentiation of specific linages. Therefore, we were able to point toward regulatory factors guiding differentiation defects in the elderly.

We also shed light into the molecular pathogenesis of hematopoietic malignancies associated with aberrant hematopoietic differentiation, such as MDS and AML. Myeloid malignancies are characterized by a significant phenotypic and genomic heterogeneity; in that context, our discoveries could have a direct clinical application by promoting the identification of patient-specific transcriptional lesions with a potential involvement in the aberrant hematopoietic differentiation. Although previous studies focused on the transcriptional lesions of MDS exist (50–56), they were performed using bulk populations of $CD34^+$ cells and, thus, provide a limited perspective of the lesions that promote aberrant differentiation in these patients. As an example, the detailed analysis of erythroid differentiation, which was altered in our patient with MDS, led to the identification of genes with altered expression dynamics, which may play a key role in promoting dyserythropoiesis. In the case of AML, GRN analysis revealed key regulons, whose activity could contribute to the phenotype of these cells. These results highlight the relevance of the methods used for understanding the mechanisms underlying aberrant hematopoiesis in patients, and provide new opportunities for the application of single-cell technologies in the clinic, which may help infer the transcriptional alterations involved in the aberrant differentiation of specific cell lineages in a patient-specific manner.

CONCLUSIONS

Our study described the transcriptional alterations that occur along the main hematopoietic lineages using multiple approaches, such as cellular proportions, differentiation trajectories and GRNs. This study is the first single-cell study of the aging of the human hematopoietic system to focus on the transcriptomic landscape using novel computational methods. The inclusion of pathological samples provided insights into the potential use of single-cell omics technologies for understanding hematological malignancies.

MATERIAL AND METHODS

Sample collection

The samples and data from the patients included in the study were provided by the Biobank of the University of Navarra and were processed according to standard operating procedures. Patients and healthy donors provided informed consent, and the study was approved by the Clinical Research Ethics Committee of the Clinica Universidad de Navarra. Bone marrow aspirates were obtained from healthy controls [young individuals (n= 5), median age, 20 years, range, 19-23 years] or patients undergoing orthopedic surgery [elderly donors (n=3), median age, 72 years, range, 61-84 years]. Samples from newly diagnosed patient with MDS and AML samples were obtained from the Clinica Universidad de Navarra. The patient´s clinical characteristics are shown in **Supplemental Table 5**.

Fluorescence-activated cell sorting

For the CD34+ cells purification, bone marrow mononuclear cells (BMMCs) were isolated by Ficoll-Paque Plus (GE healthcare) density gradient centrifugation and stained using CD34 (clone 8G12; BD bioscience) CD64 (clone 10.1; Biolegend) CD19 (clone SJ25C1; Biolegend) CD10 (clone HI10A; Biolegend) CD3 (clone OKT3; BIolegend) CD36 (clone CLB-IVC7; Sanquin Plesmanlaan) CD61 (clone RUU-PL7F12; BD bioscience) for 15 min at RT. CD34+ CD64- CD19- CD10- CD3-CD36+CD61+ cells were then sorted in a BD FACSAria II (BD Biosciencs). Purified CD34+ cells were directly used for scRNA seq analysis.

scRNA library preparation

The transcriptome of the bone marrow $CD34^+$ cells were examined using NEXTGEM Single Cell 3' Reagent Kits v3.1 (10× Genomics) according to the manufacturer's instructions. Between 20,000 and 17,000 cells were loaded at a concentration of 700–1200 cells/µL onto a Chromium Controller instrument (10× Genomics) to generate single-cell gel bead-in-emulsions (GEMs). In this step, each cell was encapsulated with primers containing a fixed Illumina Read 1 sequence, a cell-identifying 16-bp 10× barcode, a 10-bp unique molecular identifier (UMI), and a poly-dT sequence. Upon cell lysis, reverse transcription yielded full-length, barcoded cDNA, which was then released from the GEMs, amplified using polymerase chain reaction, and purified using magnetic beads (SPRIselect, Beckman Coulter). Enzymatic fragmentation and size selection were used to optimize the cDNA size prior to library construction. Fragmented cDNA was then end-repaired, A-tailed, and ligated to Illumina adaptors. A final polymerase chain reaction amplification using barcoded primers was performed for sample indexing. Library quality control and quantification were performed using a Qubit 3.0 Fluorometer (Life Technologies) and an Agilent 4200 TapeStation System (Agilent), respectively. Sequencing was performed on a NextSeq500 instrument (Illumina) (Read1: 28 cycles; Read 55 cycles; i7 index: 8 cycles) at an average depth of 20,000 reads/cell.

scRNA preprocessing

The demultiplexing of raw base call files (BCL) was performed using the 10× software cellranger mkfastq. The generated FASTQ files were aligned to the GRCh38 version of the human genome, and count matrices were constructed using cellranger count. The default barcode filtering was performed at this step. To remove doublets, we plotted the distribution of total genes and UMIs detected per cell and established a customized superior threshold for each sample (**Supplemental table 1**). Additionally, we filtered out cells with >5% or 10% of counts landing in mitochondrial genes, as this is an indicator of dying cells. In the elderly and pathological samples, we detected a small number of cells in which <1% of the mitochondrial genes clustered together; thus, we excluded them from the analysis.

Data integration, clustering and visualization

Samples from the same condition (young or elderly) were integrated using the Seurat pipeline. Gene counts were divided by the total expression per cell, multiplied by a scaling factor of 10,000, and log transformed. Normalized counts were scaled across cells. The 2,000 genes with highest variance were selected using the variance stabilizing transformation method. Next, we performed integration as described previously (29) using 50 dimensions. After integration, we rescaled the data, regressed the cell cycle effect, and conducted principal component analysis (PCA). Based on the visual exploration of the scree plot, we selected the appropriate number of components to continue the analysis (**Supplemental table 1**).

Next, we performed unsupervised clustering using the algorithm implemented in Seurat. We constructed a shared nearest neighbor graph based on Euclidean distances in the PCA space with the chosen dimensionality and clustered the cells using the default Louvain algorithm. We tested several resolutions and assessed the results by calculating the average silhouette for each cluster. We determined the cluster markers using the Seurat function *FindAllMarkers* with the MAST method. Next, we annotated the clusters by manually inspecting the most specific markers and searching for curated markers in the literature.

To ease visualization, we embedded the cells from the young and elderly samples under the same coordinates. Thus, we integrated the young and elderly datasets using the pipeline

described above. Using the first 30 principal components, we calculated the UMAP coordinates, as implemented in Seurat, which were then used in all the figures.

Classification model

We constructed a classification model to predict the cell identity of elderly, MDS, and AML single cells using the previously annotated young dataset as a reference. To this end, we performed logistic regression with elastic-net regularization in the glmnet R package. This approach is appropriate for sparse input data, and the elastic-net penalty allows flexible selection of variables.

We constructed an individual binary prediction model for each of the 14 identities established. To train and validate each model, we selected all the cells from the appropriate cluster and an equal number of random cells from the remaining clusters. This set was then randomly divided in two: 75% was destined to training the model and 25% was dedicated to validation. We also performed an initial feature selection by selecting the subset of highly variable genes present in both the training–validation data (young) and the data to be classified (elderly, MDS, or AML). Initially, we trained models with different α values (1, 0.75, 0.5, 0.25, and 0.1). For each model, we used 10-fold cross validation, and the algorithm selected the optimal value for λ. Subsequently, we selected the most appropriate α value by evaluating the performance of the models on the validation dataset. If possible, we only retained the models that used between 20 and 150 variables. Among them, we selected the model with the maximum AUC and, in case of a tie, we successively searched for the minimum false-positive rate, minimum false-negative rate, and maximum number of variables. After the optimal model was selected, we used it to classify new data as positive or negative for a particular identity.

We repeated the steps from the creation of the training and validation sets to binary prediction, 10 times and stored the resulting predictions: both the class (positive or negative) and its associated probability of being positive. Finally, we averaged the 10 resulting probabilities for each binary model and assigned the identity that corresponded to the highest probability. In cases in which none of the identities were associated with a probability of >0.5, the cell was labeled as *not assigned*.

Proportion test

We tested for differences in cell population proportions, both among individuals from the same condition and between conditions. We carried out pairwise comparison between pairs of proportions and adjusted p values using the Bonferroni-Holm correction method.

GSEA

We performed differential expression analysis of cells belonging to the same identity to identify differences between conditions. We applied the Seurat *FindMarkers* function to the log-normalized counts using the MAST method and the options logfc.threshold = 0 and min.pct = 0. Only cell types with at least 25 cells per condition were used. We ranked the genes according to their average log fold change. We performed GSEA using the fgsea R package and tested for hallmark gene sets.

Trajectory analysis using STREAM

We extracted the differentiation branches in our data using STREAM. We then used the young batch-corrected matrix to establish the reference structure, onto which we then mapped the elderly and MDS cells. Using the STREAM pipeline, we computed the 2,000 most variable genes and 20 principal components. Then, we performed nonlinear dimensionality reduction using modified locally linear embedding. We projected cells onto a three-dimensional space and set the algorithm to use the 30 nearest neighbors. An elastic principal graph was constructed by computing 20 clusters and was refined using the following parameters: epg_alpha = 0.02, epg_mu = 0.05, and epg_lambda = 0.01. Subsequently, mapping was performed using the default STREAM parameters. For this step, we used the elderly batch-corrected matrix and the MDS log-normalized data.

Trajectory analysis using Palantir

Palantir was used to reconstruct hematopoietic lineages and recover gene expression dynamics along the lineages. This algorithm models differentiation and fate choice as a continuous probabilistic process. It orders cells along a global pseudotime, determines final states, and assigns each cell a probability to reach each terminal state.

We used the batch-corrected matrices to infer trajectories per condition. PCA was performed, and 30 diffusion components were calculated. For the young dataset, we selected 10 HSCs according to their position in the UMAP coordinates. We sequentially applied the Palantir algorithm using each of these cells as the initial state. Terminal states were stored for each run, and we retained those that appeared in more than five cases. If more than one terminal state corresponded to the same cell identity, we randomly selected one of them. We ran Palantir a final time adding the selected terminal states as prior information. For the elderly dataset, we selected the HSC with the highest probability, as determined by the GLMnet model, as the initial state. We then searched for cells with the highest similarity to the terminal states found in the young dataset. To this end, we calculated the nearest neighbor in the UMAP space.

In the case of the MDS sample, we used the normalized counts matrix, calculated 30 diffusion components, and set 6 eigen vectors to determine the multiscale space. For each dataset, gene expression trends along pseudotime were calculated as described in Palantir using generalized additive models, with the addition of the branch probabilities as prior weights for the model. Gene trends were finally clustered using phonograph in Palantir.

GRNs

Python implementation of SCENIC was used for GRN inference. All required inputs were downloaded from https://resources.aertslab.org/cistarget/ and https://pyscenic.readthedocs.io/en/latest/installation.html. Raw UMI matrices were provided as input for the pyscenic tutorial, as described in https://pyscenic.readthedocs.io/en/latest/tutorial.html. Regulon specificity scores were obtained using the *regulon_specificity_scores* command.

AUC matrices were imported into R and batch-corrected using the *removeBatchEffect* implemented in the limma R package. Batch-corrected AUC matrices were binarized using the BinarizeAUC function implemented in the AUCell R package. The percentage of active regulons per cell type was determined by counting the number of cells with a value of 1 in the binarized AUC matrix and dividing it by the number of cells per cell type. The top five regulons per cell type were selected according to the RSS scores.

Networks were drawn in Cytoscape, and only edges with an importance value larger than the 3rd quartile (per regulon) were retained.

Declarations

List of abbreviations used

AML Acute myeloid leukemia
AUC Area under the curve
GRN Gene regulatory network
GSEA Gene set enrichment analysis
HSC Hematopoietic stem cell
HSPC Hematopoietic stem and progenitor cell
PCA Principal component analysis
RSS Regulon specificity score

scRNA-seq Single Cell RNA Sequencing
UMI Unique molecular identifier

Ethics approval and consent to participate
Samples and data from patients included in the study were provided by the Biobank of the University of Navarra and were processed following standard operating procedures approved by the Ethical and Scientific Committees.

Consent for publication
Not applicable.

Availability of data and material
All the single cell RNA sequencing data is available at Gene Expression Omnibus under accession number GSE180298

Competing interests
JPR has been employed by 10x Genomics since February 2021; this employment had no bearing on this work

Funding
This work was supported by the Instituto de Salud Carlos III and co-finance by FEDER funds (PI17/00701, PI19/00726 and PI20/01308), CIBERONC (CB16/12/00489 and CB16/12/00225); Gobierno de Navarra (ERAPerMed MEET-AML 0011-2750-2019-000001; AGATA 0011-1411-2020-000010/0011-1411-2020-000011 and DIANA 0011-1411-2017-000028/0011-1411-2017-000029/0011-1411-2017-000030); Fundación La Caixa (GR-NET NORMAL-HIT HR20-00871); and Cancer Research UK [C355/A26819] and FC AECC and AIRC under the Accelerator Award Program. NB was supported by a PhD fellowship from Gobierno de Navarra (0011-0537-2019-000001); MA was supported by a PhD fellowship from Ministerio de Ciencia, Innovación y Universidades (FPU18/05488); TE was supported by an Investigador AECC award from the Fundación AECC. MH was supported by H2020 Marie S. Curie IF Action, European Commission, Grant Agreement No. 898356

Authors' contributions
Conception and design:
JPR, FP, MA, TE
Acquisition of data:
NB, AAP, AVZ, PSMU, DA, JLE, MSJ, TJ, FL, SM, FSG, AM, JM
Computational analysis:
MA, JPR, GS, ADM, ML, DGC, DV, MH
Writing of the manuscript:
MA, TE, JPR,FP
Study supervision:
JPR, FP

Acknowledgements
We particularly acknowledge the patients for their participation and the Biobank of the University of Navarra for its collaboration.

REFERENCES

1. Jagannathan-Bogdan M, Zon LI. Hematopoiesis. Dev. 2013;140(12):2463–7.

2. Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. Nature. 2018;553(7689):418–26.
3. Notta F, Zandi S, Takayama N, Dobson S, Gan OI, Wilson G, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science. 2016;351(6269)2116.
4. Haas S, Trumpp A, Milsom MD. Causes and consequences of hematopoietic stem cell heterogeneity. Cell Stem Cell. 2018;22(5):627–38.
5. Watcham S, Kucinski I, Gottgens B. New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. Blood. 2019;133(13):1415–26.
6. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Hennig BP, Hirche C, et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol. 2017;19(4):271–81.
7. Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells article. Nat Immunol. 2018;19(1):85–97.
8. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell. 2018;173(6):1535-1548.
9. Dykstra B, Olthof S, Schreuder J, Ritsema M, Haan G De. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. J Exp Med 2011; 208(13):2691–703.
10. Pang WW, Price EA, Sahoo D, Beerman I, Maloney WJ, Rossi DJ, et al. Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. Proc Natl Acad Sci U S A. 2011;108(50):20012–7.
11. Ogawa T, Kitagawa M, Hirokawa K. Age-related changes of human bone marrow: A histometric estimation of proliferative cells, apoptotic cells, T cells, B cells and macrophages. Mech Ageing Dev. 2000;117(1–3):57–68.
12. Young K, Borikar S, Bell R, Kuffler L, Philip V, Trowbridge JJ. Progressive alterations in multipotent hematopoietic progenitors underlie lymphoid cell loss in aging. J Exp Med. 2016;213(11):2259–67.
13. Grover A, Sanjuan-Pla A, Thongjuea S, Carrelha J, Giustacchini A, Gambardella A, et al. Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. Nat Commun. 2016;7,11075.
14. Weiskopf D, Weinberger B, Grubeck-Loebenstein B. The aging of the immune system. Transplant International. 2009;22(11):1041–50.
15. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med. 2014;20(12):1472–8.
16. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman P V., Mar BG, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. N Engl J Med. 2014;371(26):2488–98.
17. McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J, et al. Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis. Cell Rep. 2015;10(8):1239–45.
18. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. N Engl J Med. 2014;371(26):2477–87.
19. Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res 2015;25(12):1860-72.
20. Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ, et al. Cell intrinsic alterations underlie hematopoietic stem cell aging. Proc Natl Acad Sci U S A. 2005;102(26):9194–9.
21. Martinez-Jimenez CP, Eling N, Chen HC, Vallejos CA, Kolodziejczyk AA, Connor F, et al. Aging increases cell-to-cell transcriptional variability upon immune stimulation. Science. 2017;355(6332):1433–6.
22. Mann M, Mehta A, De Boer CG, Farouq D, Regev A, Baltimore D, et al. Heterogeneous

responses of hematopoietic stem cells to inflammatory stimuli are altered with age. Cell Rep. 2018;25(11):2992-3005.
23. Chambers SM, Shaw CA, Gatza C, Fisk CJ, Donehower LA, Goodell MA. Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. PLoS Biol. 2007;5(8):e201.
24. Sun D, Luo M, Jeong M, Rodriguez B, Xia Z, Hannah R, et al. Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. Cell Stem Cell. 2014;14(5):673–88.
25. Zeidan AM, Shallis RM, Wang R, Davidoff A, Ma X. Epidemiology of myelodysplastic syndromes: Why characterizing the beast is a prerequisite to taming it. Blood Reviews. 2019;34:1–15.
26. Deschler B, Lübbert M. Acute myeloid leukemia: Epidemiology and etiology. Cancer. 2006;107(9):2099–107.
27. Shallis RM, Ahmad R, Zeidan AM. The genetic and molecular pathogenesis of myelodysplastic syndromes. Eur J Haematol. 2018;101(3):260–71.
28. Chung SS, Park CY. Aging, hematopoiesis, and the myelodysplastic syndromes. Hematology. 2017;2017(1):73–8.
29. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888-1902.e21.
30. Torang A, Gupta P, Klinke DJ. An elastic-net logistic regression approach to generate classifiers and gene signatures for types of immune cells and T helper cell subsets. BMC Bioinformatics. 2019;20(1):1–15.
31. Nguyen QH, Lukowski SW, Chiu HS, Senabouth A, Bruxner TJC, Christ AN, et al. Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. Genome Res. 2018;28(7):1053–66.
32. Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat Biotechnol. 2019;37(12):1458–65.
33. Chen H, Albergante L, Hsu JY, Lareau CA, Lo Bosco G, Guan J, et al. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat Commun. 2019;10(1):1-14.
34. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. Characterization of cell fate probabilities in single-cell data with Palantir. Nat Biotechnol. 2019;37(4):451–60.
35. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020;15(7):2247–76.
36. Bhullar J, Sollars VE. YBX1 expression and function in early hematopoiesis and leukemic cells. Immunogenetics. 2011;63(6):337–50.
37. Lee WH, Chung MH, Tsai YH, Chang JL, Huang HM. Interferon-γ suppresses activin A/NF-E2 induction of erythroid gene expression through the NF-κB/c-Jun pathway. Am J Physiol - Cell Physiol. 2014;306(4):407-14.
38. Loontiens S, Dolens AC, Strubbe S, Van de Walle I, Moore FE, Depestel L, et al. PHF6 expression levels impact human hematopoietic stem cell differentiation. Front Cell Dev Biol. 2020;8.
39. Cho J, Seo J, Lim CH, Yang L, Shiratsuchi T, Lee MH, et al. Mitochondrial ATP transporter Ant2 depletion impairs erythropoiesis and B lymphopoiesis. Cell Death Differ. 2015;22(9):1437–50.
40. Zhang F, Zhu K, Liu L, Zhu J, Li J, Zhang P, et al. IRF2-INPP4B axis inhibits apoptosis of acute myeloid leukaemia cells via regulating T helper 1/2 cell differentiation. Cell Biochem Funct. 2020;38(5):582–90.
41. Cagnetta A, Soncini D, Orecchioni S, Talarico G, Minetto P, Guolo F, et al. Depletion of SIRT6 enzymatic activity increases acute myeloid leukemia cells' vulnerability to DNA-damaging agents. Haematologica. 2018;103(1):80–90.
42. Möröy T, Khandanpour C. Role of GFI1 in epigenetic regulation of MDS and AML pathogenesis: Mechanisms and therapeutic implications. Front Oncol. 2019;9:824.

43. Svendsen AF, Yang D, Kim KM, Lazare SS, Skinder N, Zwart E, et al. A comprehensive transcriptome signature of murine hematopoietic stem cell aging. Blood. 2021.
44. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. Science. 2019;366(6465):e4673.
45. Hennrich ML, Romanov N, Horn P, Jaeger S, Eckstein V, Steeples V, et al. Cell-specific proteome analyses of human bone marrow reveal molecular features of age-dependent functional decline. Nat Commun. 2018;9(1).
46. Triana SH, Vonficht D, Jopp-saile L, Raffel S, Lutz R, Leonce D, et al. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. biorxiv. 2021;28:69120.
47. Abdelaal T, Michielsen L, Cats D, Hoogduin D, Mei H, Reinders M, et al. A comparison of automatic cell identification methods for single-cell RNA-sequencing data. Genome Biol. 2019;20(1):1-19.
48. Leimkühler NB, Schneider RK. Inflammatory bone marrow microenvironment. Hematol. 2019;2019(1):294–302.
49. Elias HK, Bryder D, Park CY. Molecular mechanisms underlying lineage bias in aging hematopoiesis. Semin Hematol. 2017;54(1): 4–11.
50. Hofmann WK, De Vos S, Komor M, Hoelzer D, Wachsman W, Phillip Koeffler H. Characterization of gene expression of CD34 + cells from normal and myelodysplastic bone marrow. Blood. 2002;100(10):3553–60.
51. Pellagatti A, Cazzola M, Giagounidis AAN, Malcovati L, Della Porta MG, Killick S, et al. Gene expression profiles of CD34+ cells in myelodysplastic syndromes: Involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. Blood. 2006;108(1):337–45.
52. Ueda M, Ota J, Yamashita Y, Choi YL, Ohki R, Wada T, et al. DNA microarray analysis of stage progression mechanism in myelodysplastic syndrome. Br J Haematol. 2003;123(2):288–96.
53. Miyazato A, Ueno S, Ohmine K, Ueda M, Yoshida K, Yamashita Y, et al. Identification of myelodysplastic syndrome-specific genes by DNA microarray analysis with purified hematopoietic stem cell fraction. Blood. 2001;98(2):422–7.
54. Im H, Rao V, Sridhar K, Bentley J, Mishra T, Chen R, et al. Distinct transcriptomic and exomic abnormalities within myelodysplastic syndrome marrow cells. Leuk Lymphoma. 2018;59(12):2952–62.
55. Montalban-Bravo G, Class CA, Ganan-Gomez I, Kanagal-Shamanna R, Sasaki K, Richard-Carpentier G, et al. Transcriptomic analysis implicates necroptosis in disease progression and prognosis in myelodysplastic syndromes. Leukemia. 2020;34(3):872–81.
56. Pellagatti A, Cazzola M, Giagounidis A, Perry J, Malcovati L, Della Porta MG, et al. Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. Leukemia. 2010;24(4):756–64.

FIGURE LEGENDS

Figure 1. Transcriptional profiling of CD34$^+$ cells from young and elderly healthy donors. a) CD34$^+$ cells were obtained from bone marrow aspirates of young (n = 5) and elderly (n = 3) donors and subjected to single-cell RNA sequencing. b) UMAP plot with young cells colored according to unsupervised clustering results (left) and elderly cells labeled using an in-house cell classifier (right). c) Dot plot of cluster markers (adjusted *P*-value < 0.05) for the different cellular subpopulations identified. Dot size represents the percentage of cells that express each marker, and color represents scaled expression values. d) Bar plots showing the proportion of cells assigned to each cellular subpopulation for each donor independently. e) Dot plot of enriched terms after performing GSEA for each identified cluster. Dot color represents the enriched group, size indicates the NES absolute value, and transparency indicates the statistical significance.

Figure 2. Trajectory inference of the hematopoietic linages at single-cell resolution. a) STREAM plot obtained using cells from young healthy donors. Color denotes cellular subpopulations. The X-axis displays inferred pseudotime values. b) Expression of known cell-type markers for the different hematopoietic lineages projected in the STREAM plot. Color represents scaled expression values. c) STREAM plot of elderly differentiation trajectories projected in the young reference. Color indicates the proportion of cells belonging to each condition under study. d) STREAM plot of elderly differentiation trajectories projected in the young reference. Color represents the cell-type identity (gray color represents the proportion of young cells). e) Scatter plot of pseudotime vs. branch probabilities for the monocytic trajectory obtained using Palantir for young and elderly donors. Color represents the cellular subpopulation. f) Heatmap of gene expression trends for dynamic genes along the monocytic trajectory in young and elderly donors. The columns are arranged according to pseudotime values, and the rows are grouped according to gene clustering results. A summary of enriched terms for the gene clusters in young donors is shown. g) Expression trends in the comparison of young and elderly cells regarding the different genes involved in the monocytic trajectory.

Figure 3. Gene regulatory network reconstruction of hematopoietic cellular populations. a) (Left) Heatmap showing the proportion of cells per cluster that have an activated state for different regulons in young cells. (Right) UMAP plots with normalized expression and AUC values for specific transcription factors. b) Gene regulatory network of the identified regulons for the hematopoietic system in young donors. Regulons were trimmed to include only targets with an importance score higher than the 3rd quartile in each regulon. Node shape denotes gene-type identity, and color denotes cell population. Any target that can be assigned to multiple transcription factors is colored in gray. c) (Left) Heatmap showing the proportion of cells per cluster that have an activated state for different regulons in elderly cells. (Right) UMAP plots with normalized expression and AUC values for specific transcription factors. d) Gene regulatory network of the identified regulons for the hematopoietic system in elderly donors. Regulons were trimmed to include only the targets with an importance score higher than the 3rd quantile in each regulon. Node shape denotes gene-type identity, and color denotes cell population. Any target that can be assigned to multiple transcription factors is colored in gray. e) Bar plot with enriched gene ontology categories after over-representation analysis. Categories are grouped per cell type, and color denotes the enriched group. Bar length represents statistical significance.

Figure 4. Computational analysis of pathological conditions, including myelodysplastic syndromes and acute myeloid leukemia. a) UMAP plot of CD34$^+$ cells from MDS (n = 1) and AML (n = 1) samples. Cells are colored according to identity, as assessed using a previously described cell-type classification method. b) Bar plots showing the proportion of cells assigned to each cellular subpopulation for each donor independently. Color denotes the cellular subpopulation. c) STREAM plot for trajectory inference of young, elderly, and MDS samples. Both elderly and MDS data were projected over the young dataset. d) (Left) Heatmap of scaled gene expression values for dynamic genes along the erythroid trajectory in young, elderly, and

MDS samples. The columns are arranged according to pseudotime values, and the rows are grouped according to gene clustering results. (Right) Expression trends of genes with lower (top), higher (medium), and similar (bottom) levels in the MDS sample compared with those in the healthy samples. e) (Left) Heatmap showing the proportion of cells per cluster that had an activated state for different regulons among AML cells. (Right) Reconstructed gene regulatory network of the AML sample. Node shape denotes gene-type identity, and color denotes cell population. Any target that can be assigned to multiple transcription factors is colored in gray.

Additional Files

Supplemental file 1
Supplemental figures S1 to S6

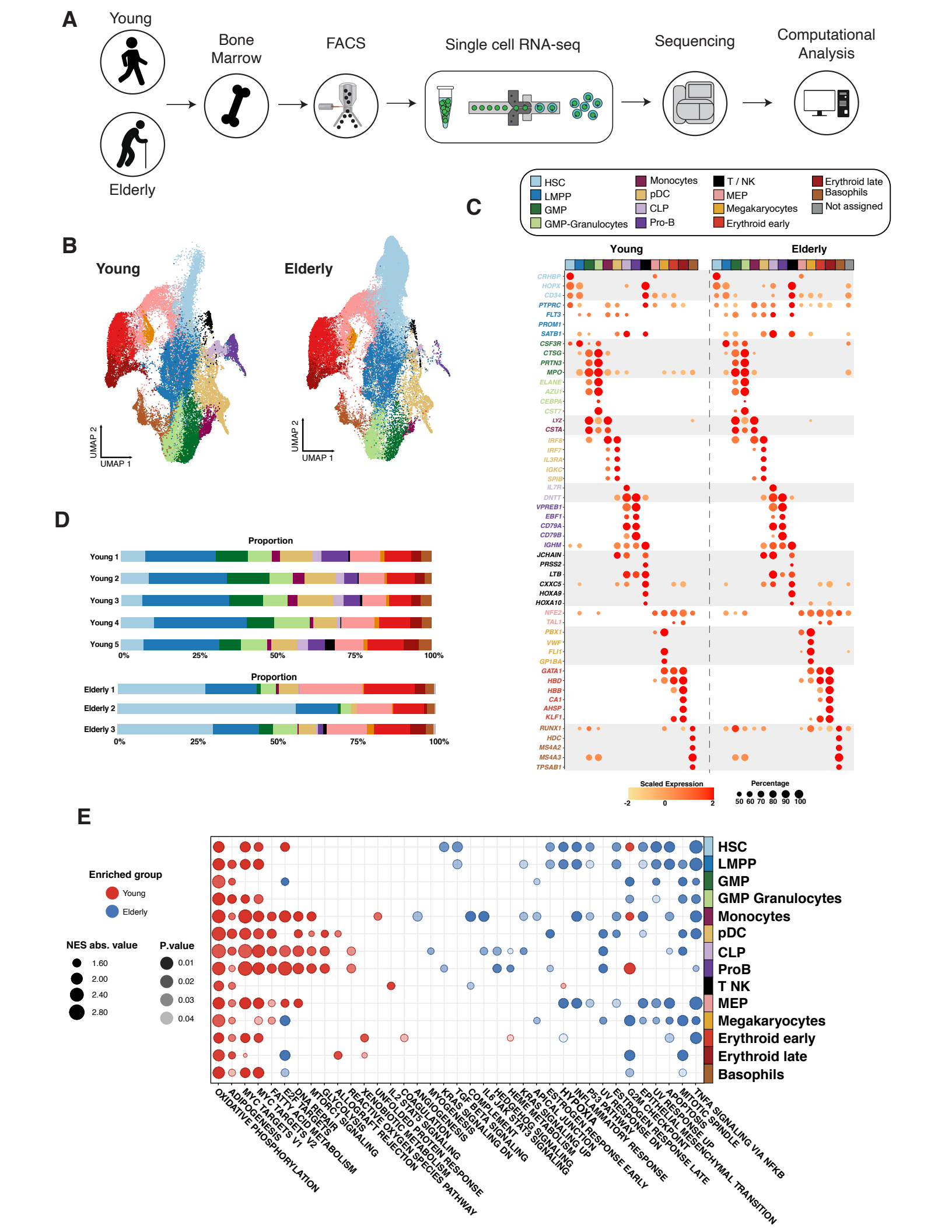
Supplemental Table S1
Parameters used for scRNAseq analysis

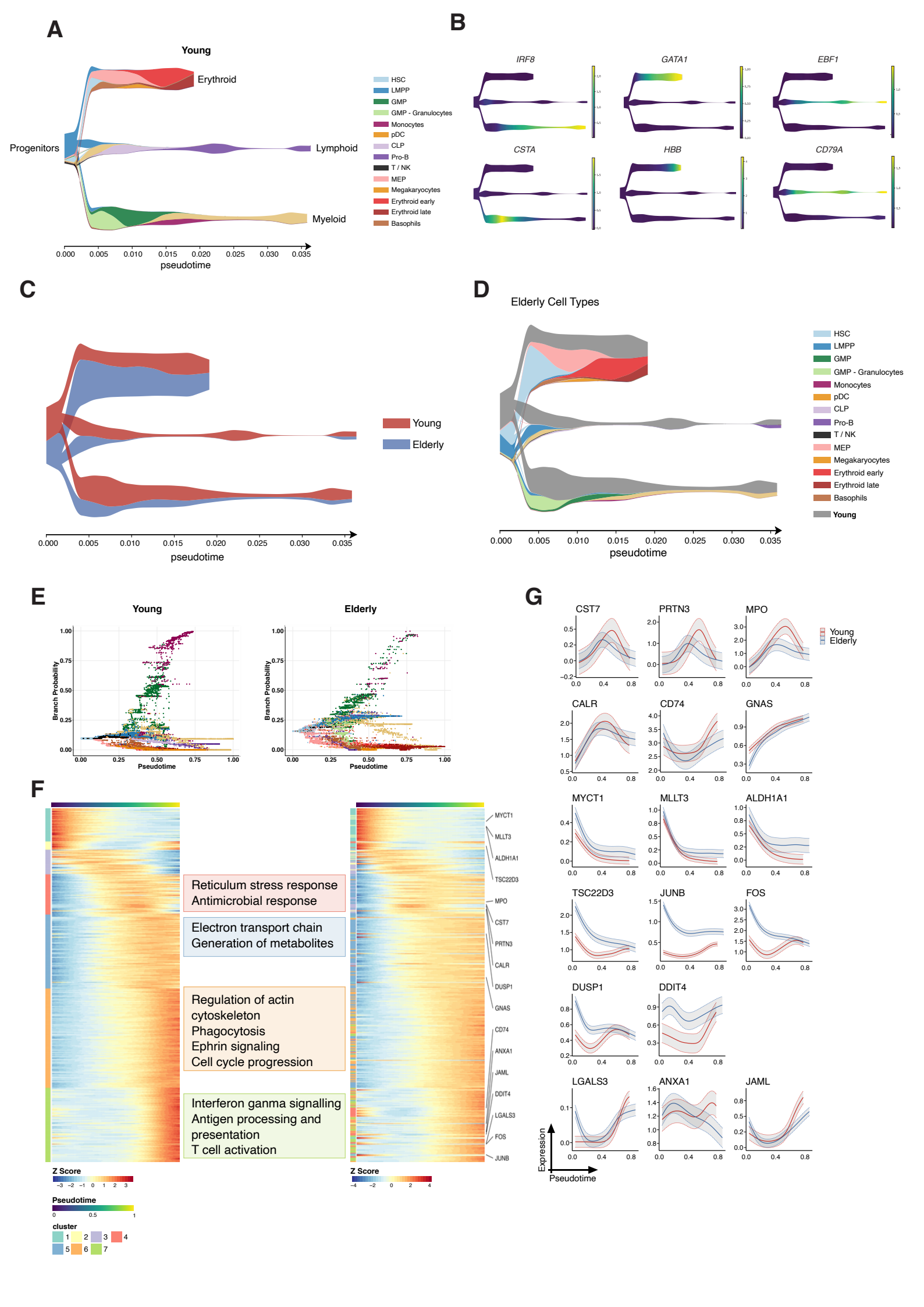
Supplemental Table S2
Cell type specific markers for each of the studied conditions (Adj p.value < 0.01 and logFC > 0.1)

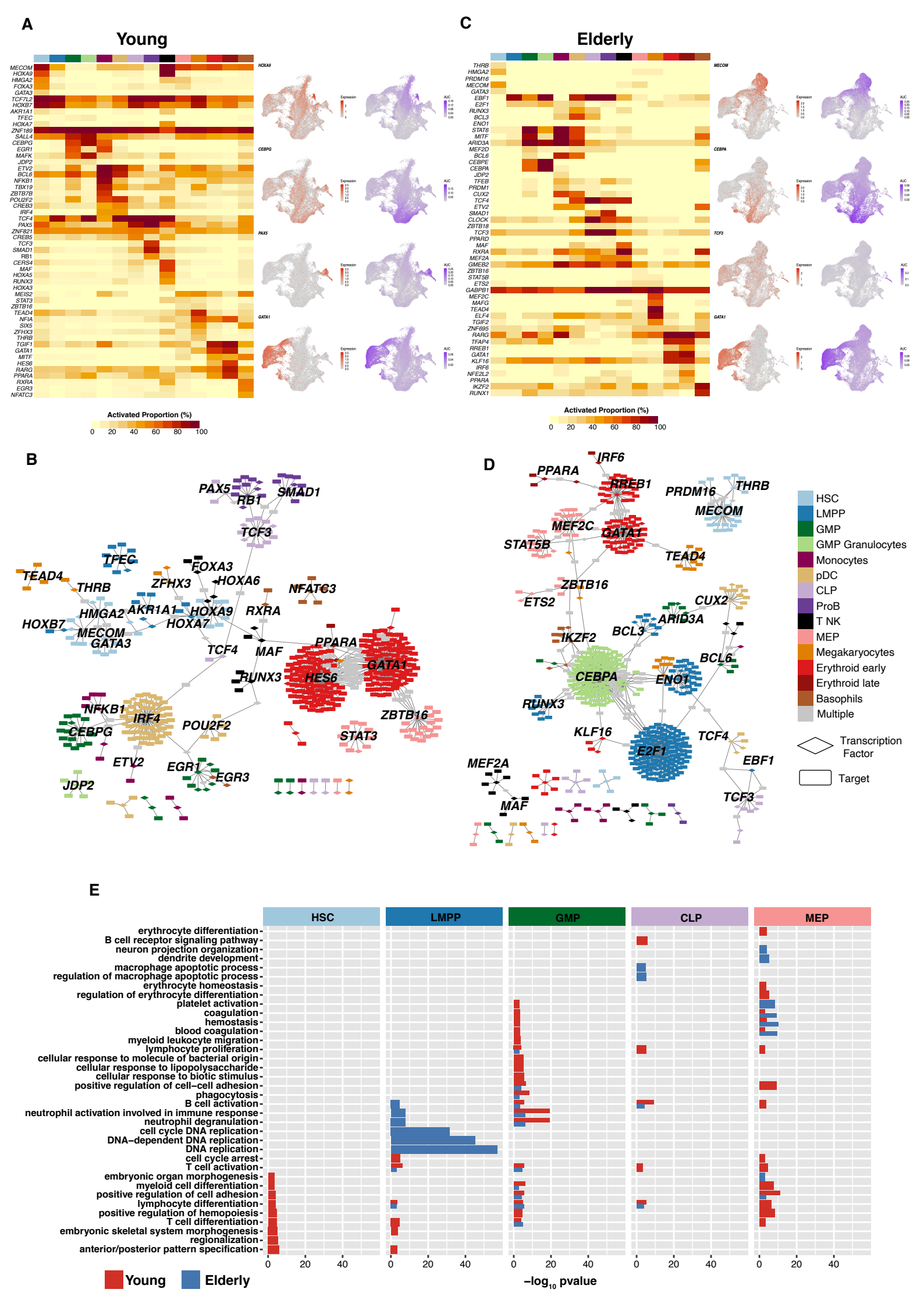
Supplemental Table S3
Cell type proportion test between young donors, elderly donors and conditions

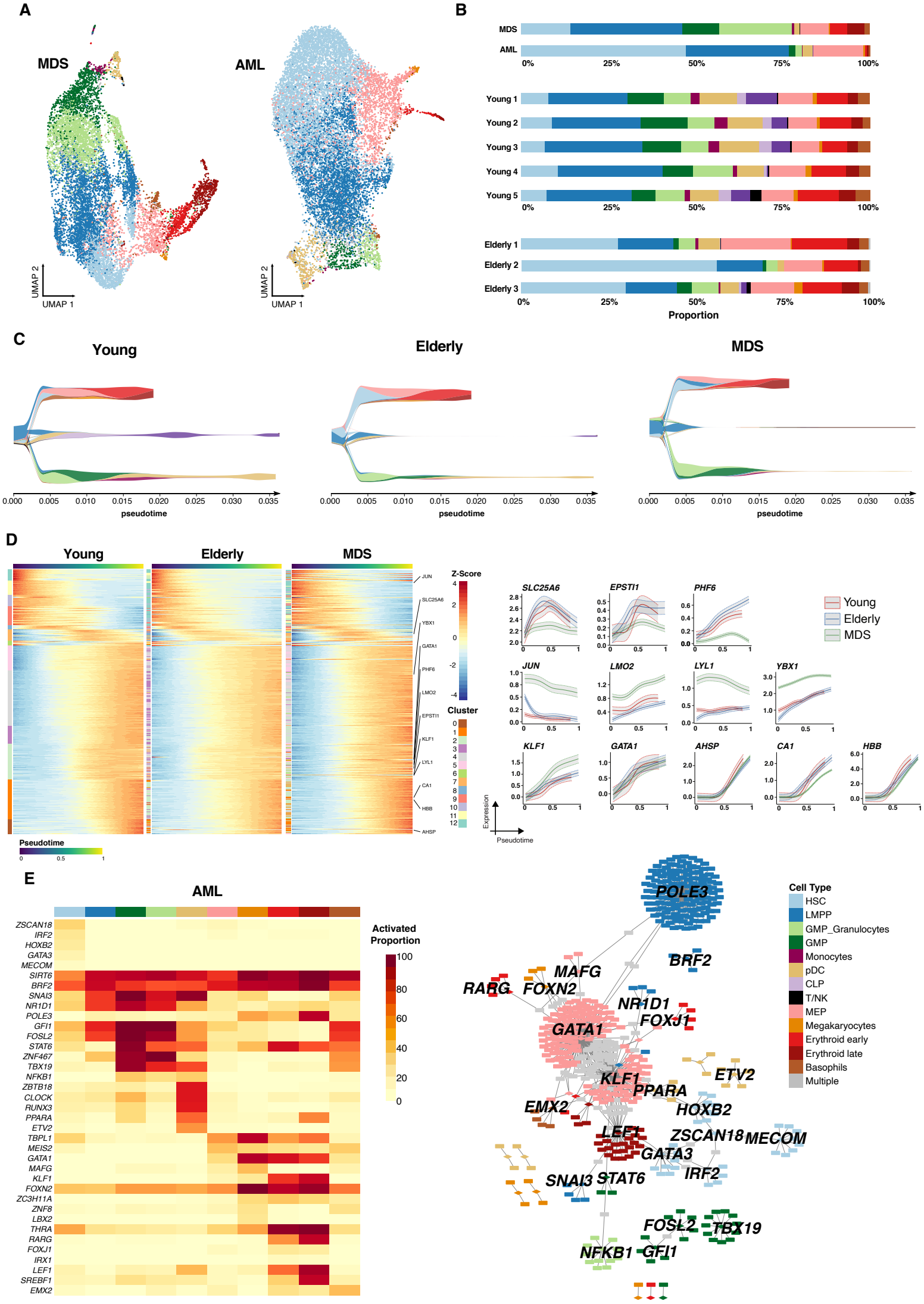
Supplemental Table S4
Differentially expressed genes between young and elderly per cell subpopulation (Adj p.value < 0.01 and |logFC| > 0.1)

Supplemental Table S5
Clinical information from patients used in this study.


A
Young
Elderly
Bone Marrow
FACS
Single cell RNA-seq
Sequencing
Computational Analysis
B
Young
Elderly
UMAP 2
UMAP 1
C
HSC
LMPP
GMP
GMP-Granulocytes
Monocytes
pDC
CLP
Pro-B
T / NK
MEP
Megakaryocytes
Erythroid early
Erythroid late
Basophils
Not assigned
Young
Elderly
Scaled Expression
Percentage
D
Proportion
Young 1
Young 2
Young 3
Young 4
Young 5
Elderly 1
Elderly 2
Elderly 3
0%
25%
50%
75%
100%
E
Enriched group
Young
Elderly
NES abs. value
1.60
2.00
2.40
2.80
P.value
0.01
0.02
0.03
0.04
HSC
LMPP
GMP
GMP Granulocytes
Monocytes
pDC
CLP
ProB
T NK
MEP
Megakaryocytes
Erythroid early
Erythroid late
Basophils
OXIDATIVE PHOSPHORYLATION
ADIPOGENESIS
MYC TARGETS V1
MYC TARGETS V2
FATTY ACID METABOLISM
E2F TARGETS
DNA REPAIR
MTORC1 SIGNALING
GLYCOLYSIS
ALLOGRAFT REJECTION
REACTIVE OXYGEN SPECIES PATHWAY
XENOBIOTIC METABOLISM
UNFOLDED PROTEIN RESPONSE
IL2 STAT5 SIGNALING
COAGULATION
ANGIOGENESIS
MYOGENESIS
KRAS SIGNALING
TGF BETA SIGNALING
COMPLEMENT
IL6 JAK STAT3 SIGNALING
HEDGEHOG SIGNALING
HEME METABOLISM
KRAS SIGNALING DN
APICAL JUNCTION
ESTROGEN RESPONSE UP
HYPOXIA
INFLAMMATORY RESPONSE
P53 PATHWAY
UV RESPONSE EARLY
ESTROGEN RESPONSE DN
G2M CHECKPOINT
EPITHELIAL MESENCHYMAL TRANSITION
UV RESPONSE LATE
APOPTOSIS
MITOTIC SPINDLE
TNFA SIGNALING VIA NFKB


A
Young
Erythroid
Progenitors
Lymphoid
Myeloid
HSC
LMPP
GMP
GMP - Granulocytes
Monocytes
pDC
CLP
Pro-B
T / NK
MEP
Megakaryocytes
Erythroid early
Erythroid late
Basophils
0.000 0.005 0.010 0.015 0.020 0.025 0.030 0.035
pseudotime
B
IRF8
GATA1
EBF1
CSTA
HBB
CD79A
C
Young
Elderly
0.000 0.005 0.010 0.015 0.020 0.025 0.030 0.035
pseudotime
D
Elderly Cell Types
HSC
LMPP
GMP
GMP - Granulocytes
Monocytes
pDC
CLP
Pro-B
T / NK
MEP
Megakaryocytes
Erythroid early
Erythroid late
Basophils
Young
0.000 0.005 0.010 0.015 0.020 0.025 0.030 0.035
pseudotime
E
Young
Elderly
Branch Probability
Pseudotime
F
Reticulum stress response
Antimicrobial response
Electron transport chain
Generation of metabolites
Regulation of actin cytoskeleton
Phagocytosis
Ephrin signaling
Cell cycle progression
Interferon ganma signalling
Antigen processing and presentation
T cell activation
MYCT1
MLLT3
ALDH1A1
TSC22D3
MPO
CST7
PRTN3
CALR
DUSP1
GNAS
CD74
ANXA1
JAML
DDIT4
LGALS3
FOS
JUNB
Z Score
Pseudotime
cluster
G
CST7
PRTN3
MPO
CALR
CD74
GNAS
MYCT1
MLLT3
ALDH1A1
TSC22D3
JUNB
FOS
DUSP1
DDIT4
LGALS3
ANXA1
JAML
Young
Elderly
Expression
Pseudotime


A
Young
C
Elderly
Activated Proportion (%)
0 20 40 60 80 100
B
PAX5
SMAD1
RB1
TCF3
TFEC
FOXA3
TEAD4
THRB
ZFHX3
HOXA6
NFATC3
HMGA2
AKR1A1
HOXA9
RXRA
HOXB7
HOXA7
MECOM
GATA3
TCF4
MAF
PPARA
GATA1
RUNX3
HES6
NFKB1
IRF4
ZBTB16
CEBPG
POU2F2
STAT3
ETV2
EGR1
EGR3
JDP2
D
IRF6
PPARA
RREB1
PRDM16
THRB
MECOM
MEF2C
GATA1
STAT5B
TEAD4
ZBTB16
ETS2
CUX2
ARID3A
BCL3
IKZF2
BCL6
CEBPA
ENO1
RUNX3
KLF16
E2F1
TCF4
EBF1
MEF2A
TCF3
MAF
HSC
LMPP
GMP
GMP Granulocytes
Monocytes
pDC
CLP
ProB
T NK
MEP
Megakaryocytes
Erythroid early
Erythroid late
Basophils
Multiple
Transcription Factor
Target
E
HSC
LMPP
GMP
CLP
MEP
erythrocyte differentiation
B cell receptor signaling pathway
neuron projection organization
dendrite development
macrophage apoptotic process
regulation of macrophage apoptotic process
erythrocyte homeostasis
regulation of erythrocyte differentiation
platelet activation
coagulation
hemostasis
blood coagulation
myeloid leukocyte migration
lymphocyte proliferation
cellular response to molecule of bacterial origin
cellular response to lipopolysaccharide
cellular response to biotic stimulus
positive regulation of cell–cell adhesion
phagocytosis
B cell activation
neutrophil activation involved in immune response
neutrophil degranulation
cell cycle DNA replication
DNA–dependent DNA replication
DNA replication
cell cycle arrest
T cell activation
embryonic organ morphogenesis
myeloid cell differentiation
positive regulation of cell adhesion
lymphocyte differentiation
positive regulation of hemopoiesis
T cell differentiation
embryonic skeletal system morphogenesis
regionalization
anterior/posterior pattern specification
0 20 40
Young
Elderly
−log10 pvalue


A
MDS
AML
UMAP 2
UMAP 1
B
MDS
AML
Young 1
Young 2
Young 3
Young 4
Young 5
Elderly 1
Elderly 2
Elderly 3
0%
25%
50%
75%
100%
Proportion
C
Young
Elderly
MDS
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
pseudotime
D
Young
Elderly
MDS
Z-Score
Cluster
Pseudotime
JUN
SLC25A6
YBX1
GATA1
PHF6
LMO2
EPSTI1
KLF1
LYL1
CA1
HBB
AHSP
Young
Elderly
MDS
Expression
Pseudotime
E
AML
Activated Proportion
ZSCAN18
IRF2
HOXB2
GATA3
MECOM
SIRT6
BRF2
SNAI3
NR1D1
POLE3
GFI1
FOSL2
STAT6
ZNF467
TBX19
NFKB1
ZBTB18
CLOCK
RUNX3
PPARA
ETV2
TBPL1
MEIS2
GATA1
MAFG
KLF1
FOXN2
ZC3H11A
ZNF8
LBX2
THRA
RARG
FOXJ1
IRX1
LEF1
SREBF1
EMX2
Cell Type
HSC
LMPP
GMP_Granulocytes
GMP
Monocytes
pDC
CLP
T/NK
MEP
Megakaryocytes
Erythroid early
Erythroid late
Basophils
Multiple
POLE3
BRF2
MAFG
RARG
FOXN2
NR1D1
FOXJ1
GATA1
ETV2
KLF1
PPARA
EMX2
HOXB2
LEF1
ZSCAN18
MECOM
GATA3
SNAI3
STAT6
IRF2
FOSL2
TBX19
NFKB1
GFI1