A comprehensive transcriptome signature of murine hematopoietic stem cell aging

Arthur Flohr Svendsen[1,2], Daozheng Yang[1,2], Seka Lazare[1], Erik Zwart[1], Albertina Ausema[1], Gerald de Haan[1,3*], Leonid V. Bystrykh[1,3]

[1]European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands.

[2] These authors contributed equally

[3] Lead Author

* Correspondence: g.de.haan@umcg.nl

Highlights

1. A meta-analysis of murine HSC transcriptomes identifies a robust aging signature;
2. The HSC aging signature is highly enriched for cell membrane-related transcripts;
3. HSCs become transcriptionally activated upon aging.
4. P-selectin functionally contributes to HSC aging

Summary

To determine whether a consistent pattern of transcriptional deregulation in aging murine hematopoietic stem cells (HSC) exists, we collected all available transcriptome studies of aged HSCs, adding our own unpublished data. Cross-validation of all datasets identified a core list of consistently differentially expressed genes; the HSC aging signature. Despite heterogeneity between individual studies, the aging signature is robust and has reached saturation. Our analysis also indicates that HSCs become transcriptionally activated upon aging. Unexpectedly, the signature consists largely of membrane-associated transcripts, including many cell surface molecules previously not associated with HSC biology. We validated that *Selp*, the top aging gene, is not only a marker for aged HSCs but is functionally involved in age-associated HSC functional decline. We share the aging signature as an online resource with the community and demonstrate its value by confirming that exposure to sympathomimetics, and deletion of Dnmt3a/b, molecularly resembles HSC rejuvenation or aging, respectively.

Keywords

aging, hematopoiesis, stem cells, transcriptomics, meta-analysis, signature, Selp, p-selectin

Introduction

Aging is a natural and time-dependent process, leading to the functional decline of many cell types and tissues. Such decline is associated with a wide array of age-associated diseases, including those of the hematopoietic system , such as a reduced ability to combat infections, anemias, and myeloproliferative diseases and leukemias.

Previous studies using C57BL/6 mice as an aging model have documented that aged hematopoietic stem cells (HSCs) – the cell population which gives rise to all mature hematopoietic populations – have less regeneration potential when compared to their young counterparts (de Haan et al., 1997; Dykstra et al., 2011; Morrison et al., 1996; Rossi et al., 2008; Sudo et al., 2000), suggesting stem cell exhaustion, which has been classified as one of the hallmarks of aging (López-Otín et al., 2013; Singh et al., 2019). Paradoxically, the absolute number of HSCs increases upon aging, but their overall contribution to formation of mature blood populations, at the clonal level, is restricted and lineage-skewed (Dykstra et al., 2011; Verovskaya et al., 2014, 2013).

The underlying molecular mechanisms causing these age-associated changes have been well-studied. To note, transplantation of aged HSCs into young recipients does not improve HSC function (Morrison et al., 1996; Sudo et al., 2000), suggesting that cell-intrinsic aspects play a pivotal role in HSC aging. Previously, genome reorganization and epigenetic modifications were implicated as causative factors (Chambers et al., 2007; Cheung et al., 2018; Rossi et al., 2005). Such alterations are expected to result in transcriptional changes, and indeed, multiple studies have aimed to compare transcriptome changes of young and aged HSCs (Beerman and Rossi, 2015; Bersenev et al., 2012; Chambers et al., 2007; Flach et al., 2014; Grover et al., 2016; Kirschner et al., 2017; Kowalczyk et al., 2015; Mann et al., 2018; Maryanovich et al., 2018; Noda et al., 2009; Norddahl et al., 2011; Quere et al., 2014; Rossi et al., 2005; Sun et al., 2014; Wahlestedt et al., 2013).

These studies have suggested a plethora of distinct aging-associated mechanisms. These include aberrant regulation of genes involved in lymphoid potential and platelet differentiation (Grover et al., 2016; Rossi et al., 2005); elevated inflammation and stress responses (Chambers et al., 2007; Mann et al., 2018); chromatin remodeling/DNA methylation (Beerman et al., 2013; Chambers et al., 2007; Sun et

al., 2014; Wahlestedt et al., 2013); changes in cell cycle and maintenance of quiescence (Kirschner et al., 2017; Kowalczyk et al., 2015; Noda et al., 2009) and age-associated replicative stress in HSCs (Flach et al., 2014). Taken together, it remains unclear whether all suggested mechanisms contribute equally to HSC aging, and why different studies report different mechanisms.

Although a substantial number of HSC aging transcriptome studies has been published, no cross-validation of all transcriptomic HSC aging data has yet been carried out. In fact, it is unknown whether a common HSC aging transcriptome exists, and if so, which genes are involved.

In order to address question, we identified 15 studies in which transcriptomes of young and aged murine HSCs were compared, and added to these our own unpublished transcriptome analysis. Despite uncovering considerable heterogeneity among different studies, by employing a novel approach we were able to identify a robust and biologically relevant HSC transcriptomic aging signature.

This signature is mainly composed of genes that encode for membrane-associated proteins, many with unknown function(s) in HSC biology. A large fraction of the aging signature genes is upregulated in aged HSCs, indicating transcriptional activation in aged HSCs. Furthermore, we demonstrate that the top age-associated gene – *P-selectin* (*Selp)* – does not only mark aged HSCs, but its upregulation contributes to functional impairment of HSCs. We provide an open web-based resource to the community, and use two independent datasets to demonstrate how our approach can be used to validate novel biological paradigms.

Results

Systematic collection of publications reporting gene expression changes in aged HSC

In order to identify all publications which have reported on HSC aging and to select all appropriate studies, we took advantage of PubMed's API E-utility **(Figure 1A)**. This approach allowed us to systematically search for keywords and terms including "aging" and "hematopoietic stem cells" and any relevant variations **(Supplemental text)**. This query resulted in 1710 different publications. We next selected publications which contained differentially expressed gene analyses comparing young and physiologically aged HSCs, specifically in the C57BL/6 mouse strain. We further refined the search by selecting only those studies which contained raw data

deposited at the Gene Expression Omnibus (GEO) database. Ultimately, this refined search resulted in 15 published studies that met our criteria. Additionally, we also included our own unpublished dataset in this list, culminating in 16 HSC aging studies that were further used in this systematic analysis **(Figure 1B/Table S1)**. In order to identify differentially expressed genes (DEGs) the various studies used different sequencing platforms, ranging from microarray hybridization, to bulk RNA-sequencing (bRNA-seq) and single-cell RNA-sequencing (scRNA-seq)**(Figure 1B/Table S1)**.

We assessed the age groups that were used in all publications and found consistent definitions for young and old age intervals **(Figure 1B)** where young mice had a median age of 3.03 ± 1.04 (SD) months and aged mice a median age of 22 ± 3.43 months **(Figure S1A)**.

We employed two alternative approaches to perform the analysis. First, we compared the self-reported, pre-analyzed lists of DEGs from the selected manuscripts and referred to it as 'meta-analysis'. In parallel, we re-analyzed all available raw data in the most uniform way, these are referred to as 're-analysis' **(Figure 1A/Table S1)**.

Individual studies identify highly variable numbers of DEGs

We first focused on the number of DEGs and observed a striking variation in terms of numbers of reported DEGs across different studies, in both microarray and RNA-seq analyses **(Figure 1C)**. Whereas some articles reported thousands of genes to be differentially expressed, others only reported dozens. When DEGs from the meta-analysis and re-analysis approaches were compared, both demonstrated a significant correlation between fold-change (FC) expression values, indicating that our re-analysis of the raw data did not deviate drastically from the originally reported data sets **(Figure S1D).**Thus, these study-to-study variations might result from differences in mouse gender, group size, exact age, HSC immunophenotyping, intrinsic differences between expression platforms, or unknown environmental effects, including differences in mouse facilities or breeding companies from which the mice originated **(Table S1)**. Principal component analysis (PCA) performed in the re-analyzed sets demonstrated that studies cluster according to their sequencing platform **(Figure S1B)**. Further, each sequencing platform also demonstrated clear differences in terms of fold-changes of DEGs **(Figure S1C)**. Overall, this analysis

demonstrates that the majority of variation is of technical origin rather than resulting from biological factors.

Most reported age-associated HSC genes have low reproducibility

Collectively, all 16 studies reported a total of 6314 and 5443 genes being differentially expressed between young and aged HSCs in the meta-analysis and re-analysis data, respectively. We then performed a cross-comparison across all studies, assessing the reproducibility – *i.e.* consistency – for every gene ever reported to be differentially expressed upon aging. The vast majority of reported genes demonstrated poor reproducibility, as almost 80% of all putative age-associated genes were reported in a single study and not in any other study **(Figure S2A/B)**. Moreover, pair-wise analysis of all different studies showed that some particular studies had little to no overlapping genes with any other study **(Figure S2C/D).**

Identification of an HSC aging signature

Despite poor reproducibility between individual data sets, we wondered whether it would be possible to identify a set of the most consistent DEGs across all studies. To this end we ranked all reported genes by their degree of consistency – the frequency in which genes were reported to be differentially expressed across studies **(Figure 2A)**. This set revealed a limited number of highly consistent genes that were regularly documented to be differentially expressed upon HSC aging. In the re-analyzed set, the single most consistently upregulated gene was *Selp,* which was present in all 12 studies. In addition, *Mt1* and *Nupr1* have been identified in 11 out of 12 studies. We provide a complete gene list with their respective ranking in **Table S2**. For further analysis we restricted the list of most consistent aging genes to the ones which were reported consistently in ≥ 4 studies, resulting in a list of 142 genes for the meta-analysis and 220 genes for the re-analyzed data set **(Figure S2E/ Table S2)**.

The aging signature is robust and confers higher resolution into age-associated genes than individual studies

In order to assess how many publications were required to derive a robust list of consistent aging genes, we studied how gene consistency depended on the number

of publications taken for analysis. Therefore, we modelled all possible combinations of most consistent genes from the re-analyzed data, starting from 1 dataset at a time to all combinations of 11 datasets out of 12 **(Figure 2B)**. This analysis generated multiple consistencies for the same genes across different permutations. The consistency scores where then expressed as coefficient of variation (CV). We observed that when the analysis was performed with a small number of datasets, CV values for genes were high. As the number of datasets increased, variation started to diminish. When 6 datasets or more were used, the number of consistent genes also became stable **(Figure 2B/ supplemental text)**. Complementarily, we also modeled the predicted resolution – how many aging genes can be found – for the same range of studies **(Figure 2C)**. In line with our empirical observations, the model predicted that the number of studies that we used for our analyses – 11 for meta-analysis and 12 for re-analysis – has a high resolution. This indicates that with the current number of included studies, >90% of all the age-associated genes has been found. It also implies that the genes consistently reported in ≥ 4 studies from the HSC aging gene list – now referred to as the "aging signature" (AS) – remains stable. Moreover, it is unlikely that addition of novel HSC transcriptome data sets will significantly alter the aging signature.

We wondered whether any of the reported datasets could serve as a reference for HSC aging. We evaluated the degree of similarity between the AS and the individual studies from which it is derived **(Figure 2D)**. All studies demonstrated approximately the same similarity to the AS, with no clear outliers. Any single study had approximately equal and limited resolution to serve as a transcriptomic aging reference – no single study encompasses all aging genes – but collectively they contribute to the robustness of the AS. We conclude that the AS can be used as a reference for DEGs upon HSC aging.

The aging signature reveals high enrichment for membrane-related processes and HSC transcriptional activation

Since the re-analyzed data demonstrated higher resolution in terms of number of reported genes **(Figure 2A)**, and provided more parameters – such as fold-changes, *p*-values, and number of overlapping studies – we restricted subsequent analyses to this signature.

An in-depth analysis of the genes which comprise the aging signature demonstrated two remarkable features. First, the large majority of the genes encode for membrane-associated proteins. Second, there is a disproportionate number of upregulated genes.

When genes were classified by the cellular localization of the encoded protein, ~50% of the AS genes were found to encode for membrane proteins **(Figure 3A)**. These included many cell adhesion molecules (*Selp*, *Alcam*, *Cd34, Dsg2, Itgb3*) and cytokine receptors (*Csf2rb, Ebi3, Flt3, Ghr, Il1rapl2, Osmr),* many of which had not been previously associated with HSC biology. Complementary gene network analysis demonstrated that the proteins which are coded by AS genes are also localized in multiple cell compartments **(Figure 3B)**.

Beyond their physical location, we set out to biologically annotate the role and molecular nature of the 220 AS genes. Gene Ontology (GO) analysis using all AS DEGs identified a total of 46 enriched GO function terms **(Table S3)**. As expected, the most significant GO terms revolve around membrane associated functions **(Figure S3A)**.

Next, in order to gain further insight into the processes which are aberrant upon aging, we performed additional analyses, this time separating up- and downregulated genes **(Figures 3C and 3D)**. Interestingly, some GO terms were present in both up and down regulated GO analysis. Such terms revolved around response to external and internal stimuli, cell activation, and response to oxygen-containing compounds, among others **(Figure S3B).**

Within upregulated processes, we observed significant enrichment of different GO terms associated with cell adhesion and activation, and the organization of cell junction and cell projection. (**Figure 3C)**. Within the downregulated processes we identified significant enrichment for DNA replication, cytokine signaling, tissue homeostasis, and cell proliferation **(Figure 3D)**. Pathway analysis also demonstrated analogous pathways **(Figure S3C)**.

Since one of the top enriched GO terms found in the upregulated genes was associated with cell activation, we wondered whether this would also be reflected in transcriptional activation. We observed that 181 genes (~80%) from the AS were upregulated **(Figure 3E)**. In further validation in young and aged LT-HSCs (LSK $CD48^{-}$ $CD150^{+}$) **(Figure S3D)**, we observed a ~3-fold higher overall RNA yield per cell in aged HSCs **(Figure S3E).** We confirmed this by intracellular staining of sorted

young and aged LT-HSCs with RNA-specific dye (SYTO) **(Figure S3F)**. In order to investigate if increased RNA abundance in aged LT-HSCs was directly linked to increased transcription, we performed intracellular staining for RNA Polymerase II (Pol-II) in young and aged HSCs and observed increased RNA Pol-II activity in aged cells **(Figure 3F)**, indicating that aged LT-HSCs are in fact more transcriptionally active than their young counterparts.

Taken together, the aging signature demonstrated that aged HSCs undergo adaptation at cellular and functional levels through membrane-related processes and that aged HSCs demonstrate higher transcriptional activation.

The top aging gene – Selp – functionally contributes to HSC aging.

We sought to provide biological validation that genes in the AS actually contribute to HSC aging. To this end, we focused on the *Selp*, only gene which was consistently reported to be upregulated in all re-analyzed studies. *Selp* encodes for the the adhesion molecule P-selecting, which will be referred as SELP.

Firstly, we assessed SELP expression in young and aged hematopoietic progenitor subsets by flow cytometry. Although SELP was expressed by different populations, only primitive LT-HSCs showed a significant increase upon aging **(Figure 4A and S4A)**. Thus, age-associated increase of *Selp* expression is restricted to the most primitive cell compartment.

Following, as we found aged LT-HSCs to become activated **(Figure 3E and F)**, we employed a Tet-On H2B-GFP reporter mouse to determine the proliferative behavior of $SELP^{low}$ and $SELP^{high}$ LT-HSCs. To this end mice were pulsed with doxycycline for 6 weeks to induce the expression of the GFP transgene in LT-HSCs, and chased for 14 weeks to enable dilution of the GFP reporter in actively cycling cells. GFP fluorescence was significantly higher in $SELP^{low}$ LT-HSCs, indicating that $SELP^{high}$ LT-HSCs were more actively cycling **(Figure 4B).**

In order to investigate differences in multi-lineage differentiation and long-term engraftment between differentially expressing SELP cell fractions, 1000 $SELP^{low}$ or 1000 $SELP^{high}$ LT-HSCs **(Figure S4B)** were competitively transplanted in lethally irradiated young recipients. Whereas overall engraftment level was similar for $SELP^{low}$ and $SELP^{high}$ LT-HSCs, we found that $SELP^{high}$ LT-HSCs produced more myeloid cells, suggesting that $SELP^{high}$ HSCs are myeloid biased **(Figure 4C)**.

It has been well described that aged HSCs have an impaired response to inflammatory agents (Haas et al., 2015; Yamashita and Passegué, 2019). To assess this characteristic for LT-HSCs that differentially express *Selp*, we sorted SELP^{low} and SELP^{high} LT-HSCs and induced *in vitro* myeloid differentiation in the presence of inflammatory inducing agents such TNF-α and LPS. The differentiation of SELP^{high} HSCs towards myeloid cells (MAC-1^{+}CD16/32^{+}) was significantly delayed compared to similarly aged SELP^{low} LT-HSCs **(Figure 4D and S4C)**.

To determine whether increased *Selp* expression contributed to impaired functioning of LT-HSCs, we ectopically overexpressed *Selp* in young LT-HSCs, and competitively transplanted these in young recipients **(Figure 5A).** Whereas *Selp* overexpression did not affect engraftment of leukocytes and platelets, erythroid engraftment was almost completely absent **(Figure 5A and S5A).** To further dissect this anemia phenotype, we carefully characterized various distinct erythroid cell populations in the bone marrow of these transplanted mice **(Figure 5B)**. Seventeen weeks post-transplant primary recipients were sacrificed and bone marrow cells were analyzed. In the GFP^{-} bone marrow fraction (i.e. bone marrow cells in which *Selp* was not overexpressed), ~60% of the cells expressed the pan-erythroid marker TER-119. Within the TER-119^{+} fraction the Ery-A (TER-119^{+}FSC^{high}CD71^{+}), Ery-B (TER-119^{+}FSC^{low}CD71^{+}), and Ery-C (TER-119^{+}FSC^{low}CD71^{-}) subpopulations were present in roughly equal frequencies **(Figure 5C)**. In contrast, in the GFP^{+} –*Selp* overexpressing bone marrow cells– erythroid development was severely perturbed. The fraction of TER-119^{+} bone marrow cells was reduced to ~20%, and within this erythroid compartment the final erythroid Ery-B and Ery-C populations were significantly reduced. Indeed, the most committed TER-119^{+}FSC^{low}CD71^{-} Ery-C population was virtually absent **(Figure 5C)**. The same effect was observed when erythroid engraftment was compared between mice transplanted with LT-HSCs transduced with an empty vector (EV) and recipients receiving *Selp* overexpressing (SELP OE) HSCs **(Figure S5B)**. To explore to what extent this erythroid phenotype is reminiscent of that in normally aged mice we carefully characterized these erythroid cell populations in the bone marrow of young and aged mice. Similar as in SELP OE recipients, TER-119^{+} and Ery-C populations frequencies were significantly reduced in aged mice **(Figure 5D)**.

Collectively, high levels of SELP in LT-HSCs functionally contribute to myeloid bias, delayed inflammatory response, and anemia. This demonstrates that SELP is not only a marker of aged HSCs, but its up-regulation contributes to the functional impairment of aged HSCs. This reinforces the idea that AS genes can have a functional and significant contribution to the age-associated functional decline of HSC.

The aging signature includes genes involved in traits associated with HSC aging mechanisms

HSC aging has been previously reported to be associated with cell stress, cell cycling, inflammation and epigenetic drift (Chambers et al., 2007; Flach et al., 2014; Kirschner et al., 2017; Kowalczyk et al., 2015; Sun et al., 2014). It was therefore surprising that neither GO nor pathway analyses of the AS immediately revealed a significant involvement of these previously proposed molecular HSC aging mechanisms. However, as GO analysis only identifies genes that are statistically enriched, we employed an alternative approach to detect any gene which could possibly be related to previously proposed mechanisms.

We started by collecting all associated GO terms associated with every gene present in the aging signature, thus assembling an AS GO term database **(Figure 6A)**. We found a total number of 2336 GO terms which were associated with all 220 AS genes. Upon inspection of the number of AS genes that were present in individual GO terms it became apparent that the majority of GO terms are very specific and contain less than 20 AS genes **(Figure 6A, lower panel)**.

Next, we searched for key terms in the AS GO term database that represent previously described HSC aging mechanisms. For GO terms associated with increased inflammatory state upon aging (Chambers et al., 2007), we searched for “inflammation”. For mechanisms of HSC aging related to replicative stress (Flach et al., 2014), or cell cycle activation (Kirschner et al., 2017; Kowalczyk et al., 2015), we searched for “cell cycle”; For epigenetic drift (Chambers et al., 2007; Hidalgo et al., 2012; Sun et al., 2014) we searched for ‘histone’ **(Figure 6B)**. Finally, we also searched for the key term “membrane” as a validation since the AS is mainly comprised of membrane-associated genes **(Figures 3A/B/C)**. The combined search of these 4 key terms yielded a total of 168 GO terms, which represents ~7% of the AS GO term database. However, these relatively few associated GO terms actually

contain 160 AS genes (~72%), suggesting that most of the AS is associated with one or more of these 4 categories of HSC aging mechanisms **(Figure 6B)**.

We identified 18 genes associated with "inflammation" including *Selp*, which we showed to be associated with a reduced response to LPS and TNF-α **(Figure 4D);** 15 with "cell cycle" and 5 genes associated with "histone" **(Figure 6C)**. All gene symbols found in each category can be found in **Supplemental Figure 6**. Although these 3 terms were broadly represented, the large majority of genes (149) were associated with "membrane", once again denoting that significant changes occur in the HSC membrane upon aging **(Figure 6C)**.

We then correlated the genes found in these targeted searches with their consistency scores from the aging signature **(Figure 6D)**. Collectively, this alternative approach demonstrates that evidence for previously reported HSC aging mechanisms can be retrieved from the AS. However, the most significant process associated with HSC aging, from a transcriptomic perspective, is the aberrant expression of membrane-associated genes.

Independent validation and use of the aging signature as a reference for age-associated studies

Since our aging signature builds on findings of almost 20 years of research in the HSC aging field, we aimed to test whether it can be used as a reference to analyze existing and future gene expression profiles that report on HSC aging. Although there are already several statistical methods available for gene sets comparisons, there is no consensus on which one is most appropriate. Taking this into consideration, we developed two different quantitative analyses where custom transcriptome sets can be compared against the aging signature. The most common approach, the gene set enrichment analysis (GSEA) algorithm (Subramanian et al., 2005) cannot be used since it requires a fully ranked list of genes. In contrast, the aging signature list is semi-ordered **(Figure S7A)**.To circumvent this obstacle, we developed an alternative algorithm, which we refer to as discrete gene set enrichment (dGSE), which is able to cope with semi-ordered (discrete) lists of genes **(Supplemental text)**. We also developed an alternative and more simple method, termed Rank Score (RS), which computes the degree of similarity between any experimental gene expression list and the aging signature **(Supplemental text)**.

In order to benchmark both methods, we individually tested all 12 HSC aging transcriptome studies that we used in this study against the aging signature **(Figure 7A)**. As expected, we observed that the different data sets had different degrees of enrichment scores when compared to the aging signature, in accordance with the degree of similarity previously tested **(Figure 2D)**. Some sets demonstrated a very robust enrichment, whereas a milder enrichment was observed for other studies. The application of both algorithms enables insights into the overall contribution of the independent studies to the collective aging signature.

Notably, our algorithms allow for direct comparison of any other gene list with its degree of similarity to the aging signature. As a proof of concept, we tested both algorithms with transcriptomic data sets from two unrelated publications, in which claims on HSC aging were made. We used transcriptomic data from HSCs that were presumably "rejuvenated" by a sympathomimetic drug acting on adrenoreceptor β3 (Maryanovich et al., 2018) and expression data from HSCs that were derived from *Dmnt3a/b* deficient mice that supposedly show accelerated aging phenotypes (Challen et al., 2014). As a negative control, we included a transcriptomic signature of embryonic stem cells (ES) (Wong et al., 2008). As expected, the ES signature demonstrated no enrichment for aging genes, using either of the assessment methods **(Figure 7B)**. In contrast, when we compared the list of genes that were downregulated in "rejuvenated" old HSCs to the AS, we found strong enrichment. Similarly, the transcriptome of young HSCs isolated from Dmnt3a/b KO mice was enriched for HSC aging genes, suggesting that both β3 sympathomimetic signaling and loss of *Dnmt3a/b* molecularly phenocopy normal HSC aging **(Figure 7B)**.

Discussion

In this paper, we compared 16 independent studies, all of which have reported on transcriptomic changes in highly purified young and physiologically aged murine HSCs. We demonstrated that these studies are highly heterogeneous with little overlap of reported DEGs. Nevertheless, collectively, these independent data sets were found to be highly valuable and allowed us to identify a comprehensive, robust, and stable transcriptomic signature of HSC aging. Additional age-associated HSC transcriptomic studies, although they may slightly improve the resolution of AS gene consistency, are predicted to provide minimal contribution towards the identification

of completely protein-coding genes**.** It should be emphasized that our current analysis disregards non-coding RNAs, such as long non-coding and micro-RNAs.

The HSC aging signature that we uncover is highly enriched for membrane-associated transcripts, many of which have as yet unknown roles in HSC (aging) biology. Many adhesion molecules such as *Selp ,Alcam, Itgb3, Neo1, Slamf1* and *Vwf* are upregulated in aged HSCs. As such, the involvement of cell surface molecules provides insight into the long-standing discussion in the field whether HSC aging results from cell-intrinsic or cell-extrinsic molecular alterations; our data indicate that it is both. The strong enrichment for cell membrane-associated transcripts in aged HSCs suggests that physiologically aged HSCs communicate differently with their immediate environment compared to their young counterparts. Indeed, it is also evident that the bone marrow (BM) niche undergoes remodeling upon aging, suggesting a role for cell-extrinsic factors (Ho et al., 2019; Morrison and Scadden, 2014; Pinho and Frenette, 2019).

The identification of the HSC aging signature does not resolve whether altered expression of (or a collection of) genes encoding for cell membrane associated proteins are the cause or consequence of HSC aging. However, we do demonstrate that the top HSC aging gene, *Selp*, is functionally involved in HSC aging. Whereas transplantation experiments have previously shown that *Selp*-deficient (Selp$^{-/-}$) HSCs display a competitive advantage over wild-type HSCs (Sullivan et al., 2011), here we document that high SELP expression levels impairs HSCs functionality. SELP^{high} HSCs are less responsive to inflammatory signals; they are myeloid-biased, and show strongly reduced erythroid potential. This demonstrates that *Selp* is not merely a marker of aged HSCs, but also is involved in their age-dependent functional decline.

Future functional studies are required to assess whether perturbation of AS genes can improve functioning of aged stem cells. The recent finding of rejuvenating aged HCSs by β3 sympathomimetic activity (Maryanovich et al., 2018) – which we show leads to significant repression of aging genes – suggests that this may be possible.

Evidence supporting previously proposed HSC aging theories, such as changes in cell cycle, epigenetic drift, and replicative stress, were not immediately emergent from the enrichment analysis of the AS. However, direct GO term searches on genes in the AS revealed that these features are present in the AS signature as well. At this

moment it is difficult to distinguish a hierarchy between established categories of aging processes. As an example, we did identify *Dmnt3b* as being consistently downregulated in multiple studies. DNMT3B and DNMT3A are methyltransferases with analogous functions (Challen et al., 2014), and the latter has been strongly implicated in age-associated clonal hematopoiesis (ARCH), a process known to occur in human hematopoietic aging (Jaiswal et al., 2014). It is of interest that our analysis reveals that the transcriptome of HSCs deficient for both *Dntm3a* and *Dnmt3b* is significantly enriched for genes that are also perturbed during physiological aging. Although ARCH and *Dnmt3* loss-of-function mutations have not been observed in aged mice, it has been established that DNA methylation is molecularly associated with HSC aging (Beerman et al., 2013; Sun et al., 2014). It has become clear that clocks based on differentially methylated genomic loci in peripheral blood cells result in surprisingly accurate estimates of chronological age, suggesting that DNA methylation patterns are altered in the hematopoietic system in a coordinated manner (Field et al., 2018; Han et al., 2018). In this context, it is relevant to note that many differentially expressed HSC aging genes were upregulated and that aged HSCs were transcriptionally activated, implying that gene regulation control, likely DNA and/or histone methylation play a role in HSC aging .

Finally, as a resource we provide protocols and a user-friendly and interactive website – https://agingsignature.webhosting.rug.nl/ – which contains all AS data presented in this study, and also allows users to execute custom queries and compare new datasets against the aging signature that we present here **(Figure 7C and S7B)**. This resource will be valuable in assisting the aging research community to verify genuine involvement of genes in physiological HSC aging. Furthermore, it will be important to assess to what extent the aging signature is also representative of HSCs isolated from other mouse strains, and from humans. It should be emphasized that our current analysis is restricted to mRNAs, and disregards non-coding RNAs, such as long non-coding and micro-RNAs which are likely to be also involved in HSC aging. Finally, it also remains to be determined whether transcriptomic signatures in other aging tissues in murine, human, or other aging model organisms show similarities with the signature that we present. Such studies will be crucial to provide more insight into the molecular causes of aging.

Author Contributions

Conceptualization, A.F.S., D.Y., S.S., G.d.H. and L.V.B.; Methodology, A.F.S., D.Y., and L.V.B.; Investigation, A.F.S., , D.Y., S.S., A.A. and L.V.B.; Formal Analysis, A.F.S., D.Y., S.S. and L.V.B.; Data Curation, A.F.S., E.Z., L.V.B.; Resources, E.Z. and A.A.; Writing – Original Draft, A.F.S.; Writing – Review & Editing, A.F.S., D.Y., G.d.H. and L.V.B.; Supervision, G.d.H. and L.V.B.

Acknowledgements

We thank T. Bijma, G.Mesander and H.J. Teunis from UMCG Flowcytometry Unit facilities for their assistance on cell sorting; Klaas Sjollema from UMCG Microscopy and Imaging Center (UMIC) for assistance with confocal microscopy; B Bakker, M Gerritsen, E Verovskaya and all members of the Ageing Biology and Stem Cells laboratory for discussions; L Saba for manuscript revision. This work was supported by the Netherlands Organization for Scientific Research/Mouse Clinic for Cancer and Aging, the Landsteiner Foundation for Blood Transfusion Research (LSBR1703), a China Student Council Fellowship awarded to D.Y. and Marriage, a EU FP7 Marie Curie Initial Training Network (Contract number 316964).

Declaration of Interests

The authors declare no competing interests.

Figure Title and Legends

Figure 1: Systematic analysis reveals heterogeneity in transcriptomic studies of young and aged murine HSCs.

A) Representation of decision tree used to acquire and analyze the different publications used for the systematic analysis.

B) Overview of the age groups used in each of the selected studies. Color codes refer to the transcriptome platform used. Squares next to each study indicate in which kind of analysis each study was included.

C) Number of differentially expressed genes (DEGs) reported by each study. Left panel shows the number of DEGs as reported by the authors in their respective manuscripts. Right panel shows the number of DEGs identified by us using a uniform re-analysis. Red bars indicate the number of genes that are upregulated upon aging, while blue bars indicate downregulated genes.

Figure 2: The identification of a robust hematopoietic stem cell aging signature

A) Overview of number of differentially expressed HSC aging genes. Bars represent the number of DEGs, subdivided by the number of different studies by which they were found (consistency). The actual number of DEG genes are shown in each bar. Purple and green bars indicate meta-analysis and re-analysis, respectively.

B) Variation of consistent HSC aging genes across different studies subsets. Dots represent the different coefficient of variation (CV) values and mean consistency for individual genes **(see supplemental text for further details).**

C) Predicted resolution to discover consistently reported genes. The model demonstrates that >90% of all HSC aging genes have been identified in the two analytic approaches. Purple and green dashed lines indicate meta-analysis and re-analysis, respectively.

D) Extent of overlap between all independent studies and the HSC aging signature (centered red circle, AS). The size of the circles represents the number of DEGs identified in each study, and the distance from the center represents the degree of similarity of each study to the HSC aging signature. Studies that show a higher degree of similarity are closer to the AS.

Figure 3: The HSC aging signature is highly enriched for membrane-associated genes and results largely from transcriptional activation

A) Classification of the proteins of the aging signature by cellular compartment. The localization of each protein was determined using UNIPROT classification. Colors of the bars represent upregulation (red) or repression (blue) of genes in the AS. Numbers next to the bars indicate the frequency of genes found in each category.

B) Network representation of the aging genes based on the cellular compartment where the encoded proteins are predicted to be localized. The size of each circle represents the number of genes found in each category. Connections between each cellular compartment reflect the number of genes which are shared per cellular compartment.

C/D) GO analysis of biological process for AS genes. Numbers in each bar demonstrate the number of genes found per GO term. Red (panel C) and blue (panel D) bars represent GO terms found to be significantly enriched in upregulated and downregulated AS genes, respectively. GO terms are selected based on a FDR < 0.05 cut-off.

E) Volcano plot depicting expression levels of AS genes. 181 upregulated and 39 downregulated genes are represented. Dots represent the average of either adjusted p-values (adjpval) and fold-changes (FC) per gene in different studies.

F) Protein levels of RNA Pol-II in individual young (6 months) and aged (24 months) LT-HSCs. Right panel shows the MFI quantification of young (n = 120 cells) and aged HSCs (n = 139 cells). Scale bars in IF images are 5 µm. ±SD is shown. **** $p < 0.0001$.

Figure 4: *Selp* (P-Selectin) is the top aging signature candidate and contributes to HSC aging

A) SELP protein expression level across different hematopoietic progenitors in young (black) and aged (red) mice. Top panel shows fold-changes of aged (n = 3 mice) samples normalized to their young (n = 3 mice) counterparts. Dots show biological replicates while the bars represent the mean values for the given population. The lower panel displays a representative FACS plot for SELP expression in young and aged LT-HSCs population.

B) SELP expression is associated with increased HSC cycling. The left panel shows the experimental approach. Right panels show H2B-GFP expression in DOX treated mice (n = 4 mice for each group) and control mice that were not treated (n = 2 for each group). Right upper panel shows the H2B-GFP expression in LT-HSCs subdivided in $SELP^{low}$ (blue) and $SELP^{high}$ (orange). The far right panel shows that $Selp^{low}$ LT-HSCs preferentially retained the GFP label, which is diluted in $Selp^{high}$ LT-HSCs.

C) $Selp^{high}$ LT-HSCs display a myeloid bias. The upper panel shows the experimental approach for competitive transplantation. $SELP^{low}$ (blue) and $SELP^{high}$ (orange) LT-HSCs were purified from aged mice and competitively transplanted (n = 3 mice for each group). While overall engraftment levels were similar, at 18 weeks post-transplant aged $Selp^{high}$ LT-HSCs produced more myeloid cells.

D) $SELP^{high}$ LT-HSCs show delayed *in vitro* myeloid differentiation. $SELP^{low}$ (blue) and $SELP^{high}$ (orange) LT-HSCs (n = 2 donor mice/group) were cultured for 10 days in the presence of PBS (control), TNF-α or LPS and the kinetics of myeloid differentiation ($Mac1^{+}/Gr1^{+}$) were quantified (n=3 replicate wells/experiment). $Selp^{high}$ LT-HSCs showed delayed response to both LPS and TNF-α conditions.

For all panels, ±SD is shown. ns = non-significant; * $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$. n indicates biological replicates

Figure 5: Ectopic expression of Selp impairs erythroid development

A) Selp overexpression impairs erythroid development *in vivo*. Upper panel shows the experimental approach; LT-HSCs were isolated from young mice and retrovirally transduced either with an empty vector (EV, grey) or a vector overexpressing *Selp* (SELP OE, green). Total white blood, platelets and erythrocyte counts from donor (CD45.2^{+} GFP^{+}) cells. Whereas SELP-overexpressing LT-HSC provided normal white blood cells numbers, platelets were mildly reduced and erythrocytes were virtually absent.

B) Schematic representation of flow cytometry gating strategy for erythroid populatations.

C) SELP OE recipients show impairment of erythroid differentiation. Left panel displays flow cytometry gating for the different erythroid populations from transduced (GFP^{+}) – SELP overexpressing – and untransduced (GFP^{-}) donor cells (labeled in green and grey, respectively) (n = 3 per group). Right panel shows the quantification of each population in each group. Within the GFP^{+} population, the number of Ter-119^{+} cells, Ery-B, and Ery-C population is significantly reduced.

D) Aged mice have reduced Ter-119^{+} and Ery-C populations. Left panel displays flow cytometry gating for the different erythroid populations from young (black) and aged (red) mice (n = 3 per group). Right panel shows the quantification of each population in each group. Aged mice show significant decrease in Ter-119^{+} cells, Ery-C populations, similarly to SELP OE recipients.

For all panels, ±SD is shown. ns = non-significant; * $p < 0.05$; *** $p < 0.001$; **** $p < 0.0001$. n indicates biological replicates

Figure 6: The Aging Signature uncovers new and previously proposed HSC aging mechanisms

A) Schematic representation of how associated GO terms were selected. Upper panel: All known GO terms for each individual gene were extracted from UNIPROT for each individual AS gene and collected as the Aging Signature GO terms. Lower panel: number of AS genes (x-axis) present in different GO terms (y-axis).

B) Searching for common mechanisms associated with HSC aging. Upper panel: The entire set of all associated GO terms from all genes (AS GO terms dataset) served as a database for a key term search. Each key term which represents a different mechanism of AS aging has its own list of associated GO terms. Lower left panel: 168 GO terms were found to be associated with inflammation, cell cycle, histones, and membranes. Lower right panel: out of these 168 GO terms, 160 AS genes (approximately 70% of AS) were found.

C) Number of AS genes found to be associated with membrane, inflammation, histone and cell cycle.

D) Bubble plot demonstrates the number of genes in each consistency category by their key term. The size and color or each circle demonstrates the number of genes found.

Figure 7: Using the HSC aging signature as a resource

A) Using Rank Score (RS) and discrete gene set enrichment (dGSE) analyses to compare existing datasets to the HSC aging signature. Left panel: graphical representation of the dGSE enrichment analyses from our own generated HSC transcriptome data was compared to the HSC aging signature. Right panel: Plots of RS (upper) and dGSE (lower) enrichment analyses of all individual studies that comprise the AS. Numbers refer to the studies shown on the right.

B) Searching for an aging signature using RS and dGSE analyses in independent datasets. Left panel: Lack of enrichment for aging genes in an embryonic stem cell signature from *Wong et al.* Middle panel: Significant enrichment for a HSC aging signature among genes downregulated in "rejuvenated" HSCs from Maryanovich *et al*. Right panel: Significant enrichment for a HSC signature among differentially expressed genes in HSCs isolated from young DNMT3A/B KO mice from Challen *et al.* The right-most panels show combined RS and dGSE scores for each test set.

C) Schematic representation of online tools which are available for custom testing of gene expression data sets to search for molecular features of aging. These tools are available at https://agingsignature.webhosting.rug.nl/

STAR METHODS

Lead Contact And Materials Availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Prof. Dr. Gerald de Haan (g.de.haan@umcg.nl). This study did not generate new unique reagents and novel insights are provided in the main and supplemental text

Data and Code Availability

Data generated for this study and collected data is available at Gene Expression Omnibus (GEO). Studies accession numbers used for the analysis are provided in Table S1. Original RNA-sequencing data are deposited under the accession number GSE128050.

We provide two resources. Firstly, we provide a friendly use code-free WEB tool site (http://agingsignature.webhosting.rug.nl/) where users can run their own enrichment analysis and explore the Aging List (AL) and/or Aging Signature (AS). The second resource regards scripts and data analysis, which can be found at GitHub (https://github.com/LeonidBystrykh/data-for-manuscript). It contains source files for differentially expressed data (meta and re-analyzed from all used publications). The scripts show major steps in assembling Aging List, Aging Signature. Two applications, find the gene on the list, and check for the enrichment of the set are also provided.

Experimental Model

Mice

All experiments were approved by the Central Commission for animal Testing (CCD AVD105002015286) and by Animal Ethical Committee (CDP 15286-01-01 and and 15286-02-013).Young (4-5 months) and aged (20-24 months) male C57BL/6, C57BL/6.SJL (CD45.1^{+}, 10-14 weeks) and young C57BL/6J-kitW-41J/kitW-41J (W41) (2-3 months) (originally from European division of Harlan) were obtained from the Central Animal Facility (CDP) at the University Medical Center Groningen. TetOP-H2B-GFP mice were a kind gift from Prof. Dr. Rob Coppes (University Medical Center of Groningen, Groningen, Netherlands). H2B-GFP expression was induced by providing mice (6 weeks) with food containing doxycycline (Dox) for 6 weeks. Mice were then fed with normal food for 14 weeks. Mice were kept in group

cages and housed under conventional in a temperature and day cycle controlled conditions.

Method Details

Flow Cytometry

Bone marrow was isolated from the tibia, femur, pelvis, sternum and spine by crushing, and red blood cells were lysed with erylysis buffer. For LT-HSC isolation, lysed bone marrow was stained with lineage (Lin) markers antibodies (B220, CD3, GR-1, MAC-1 and TER-119) conjugated with Alexa 700, C.KIT-PE, SCA-1(CD117)-BV421 or Pacific Blue, CD48-Alexa 647, and CD150-PECy7 (all from Biolegend). For $Selp^{low}$ and $Selp^{high}$ HSC isolation, lysed bone marrow was stained with Lin-Alexa 700, C-KIT-FITC (Biolegend), SCA-1-BV421, CD48-Alexa 647, CD150-APCFire 780 (Biolegend), SELP-PE (eBioscience). All cells were isolated on MoFlo Astrios or XDP cell sorters (Beckman Coulter).

For SELP expression on LT-HSCs, ST-HSCs and MPPs, lysed bone marrow was stained with Lin-Alexa 700, C-KIT-PE, SCA-1-BV421, CD48-Alexa 647, CD150-BV605 (Biolegend), SELP-PE. For SELP expression on CMPs, GMPs and MEPs, lysed bone marrow was stained with Lin-Alexa 700, C.KIT-APC (Biolegend), SCA-1-BV421, CD34-FITC (BD Biosciences), CD16/32-PECy7 (eBioscience) and SELP-PE. For H2B-GFP mice, lysed bone marrow was stained with Lin-Alexa 700, C-KIT-PE, SCA-1-BV421, CD48-Alexa 647, CD150-PECy7, SELP-BV605. For *in vitro* myeloid differentiation, cells were stained with C-KIT-PE, SCA-1-BV421, MAC.1-Alexa 700 and CD16/32-PECy7. For erythroblasts staining, bone marrow was flushed by 0.2% BSA/PBS from femur and was stained with CD71-PE (BioLegend) and TER-119-Alexa 700(Koulnis et al., 2011). For white blood cells chimerism, blood samples were lysed with erylysis buffer and were then stained with CD45.1-PB, CD45.2-PE, CD3-APC, B220-Alexa 700 and MAC-1/GR-1-PECy7 (all from BioLegend). For red blood cells and platelets chimerism, blood samples were stained with Ter119-Alexa 700 and CD41-BV605 (Biolegend). All samples were analyzed on BD FACSCanto™ II (BD Biosciences).

All data were analyzed with Flowjo (V10).

Retroviral constructs and transduction

Selp was amplified from C57BL/6J LT-HSC cDNA (forward primer: 5'-ATAAGAATGCGGCCGCGAGGACATGGCTGGC-3'; reverse primer: 5'-ACGCGTCGACTTAAGGGGTTGGGTCATATG-3') and then cloned into SF91–IRES–GFP vector (gift from C. Baum, Hannover Medical School, Germany). Virus was produced by transfecting Platinum-E retroviral packaging cells (Cell Biolabs, Inc.) according to the manufacture protocol. Viral supernatant was harvested 48 hours after transfection. LT-HSCs were transduced as previously described (Verovskaya et al., 2014). Cells were harvested 20-22 hours after transduction for transplantation. A small aliquot of cells was kept in culture for 5 days to determine transduction efficiency.

Transplantation

For SELP^{low} and SELP^{high} transplantation, 1000 SELP^{low} and SELP^{high} LT-HSCs were isolated from old CD45.2^{+} C57BL/6 mouse and were transplanted together with 2 million CD45.1^{+} W41 bone marrow into lethally irradiated (9 Gy) CD45.1^{+} C57BL/6.SJL recipients, respectively.

For *Selp* overexpression transplantation, EV (SF91–IRES–GFP empty) and SELP OE CD45.2^{+} cells harvested after retroviral transduction were transplanted together with CD45.1^{+} 2 million W41 bone marrow into lethally irradiated (9 Gy) CD45.1^{+} C57BL/6.SJL recipients, respectively. Transplanted mice treated with ciprofloxacin (0.1mg/L) for 10 days. Blood samples were taken for chemerism analysis from retro-orbital plexus by heparinised Hematocrit Capillary Tubes (Praxisdienst).

TNF-α and LPS treatment

SELP^{low} and SELP^{high} LT-HSCs were sorted into collection tubes, respectively, and then aliquoted to 3 technical replicates for each condition. Cells were cultured in 96-well plates in StemSpan (StemCell Technologies) supplemented with 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco), 10% FCS, 300 ng/mL SCF (PeproTech), 20 ng/mL IL-11 (R&D Systems), 1 ng/mL Flt-3 ligand (PeproTech), 10 ng/mL GM-CSF (R&D Systems), 20 ng/mL TNF-α (R&D Systems) or 100 ng/mL LPS (Sigma). Medium was refreshed 30% every other day. Cells were analyzed by at BD FACSCanto™ II at day 6, day 8 and day 10. Cells were cultured at 37℃ / 5% CO_2 conditions.

RNA-sequencing

RNA was isolated from 15000 LT-HSC (Lin$^-$SCA-1$^+$C-KIT$^+$CD48$^-$CD150$^+$) using the Nucleospin XS kit (Macherey Nagel) and quantified on Bionalyzer using RNA Pico 6000 Kit (Agilent). Ribosomal depletion was performed using a modified version of RiboZero Kit (Illumina). 300pg ribosomal-depleted RNA was used as input into TotalScript RNA-Seq Kit (Epicentre). Libraries were pooled and sequenced to 30-50 million reads on HiSeq 2500.

Immunofluorescence Staining

Images were acquired on a Leica Sp8 confocal microscope and quantified on Fiji Image J.

RNA content

4000-6000 LT-HSCs were seeded onto spots on an adhesion immunofluorescent slide (VWR). For RNA staining, cells were fixed with 100% methanol for 10 minutes and washed with PBS. Cells were subsequently stained with 1:10000 SYTO RNA Select Green Fluorescent Stain (ThermoFisher) for 20 minutes at room temperature. After washing, coverslips were mounted with ProLong Diamond Antifade Mountant with DAPI.

RNA pol II

For RNA-Polymerase II staining, cells were fixed and permeabilized with Fixation/Permeabilization Solution Buffer (BD Bioscience) for 20 minutes on ice. After washing, cells were blocked first with Endogenous Biotin-Blocking Kit (Thermo Fisher) followed by 4% BSA for 30 minutes. Cells were stained with 1:100 biotin mouse monoclonal RNA Polymerase II antibody (Novus Bio) at 4°C overnight. Cells were then washed with 0.1% Triton-X-100 and stained with 1:1000 streptavidin Alexa-488 secondary antibody. After washing, coverslips were mounted with ProLong Diamond Antifade Mountant with DAPI.

Data Acquisition

Pubmed was screened for the following combination of search keywords: “hematopoietic stem” AND (“cell” OR “cells”) AND (“ageing” OR “aging”). The retrieved list was checked for the availability of GEO Datasets. Those papers which also contained original GEO submissions were selected, and further screened for presence of gene expression data in young and physiologically aged mice. In addition, we manually added supplementary information from (Noda et al., 2009), for which we do not know any publicly available data submission.

Meta-analysis

All data for differential expression (DE) for meta-analysis studies were retrieved from provided DE tables from manuscripts and reformatted from files provided by authors and used “as is”.

Differential expression re-analysis

Three types of expression data were used: expression micro-arrays, bulk RNA-seq, and single-cell RNA-seq. For all re-analyzed sets, only samples regarding young and aged samples were taken into consideration. No batch effects were assumed or taken for correction.

Micro-arrays analysis

Raw data was downloaded from Gene Expression Omnibus (GEO) using available GEO number and analyzed for DE genes using a similiar script structure from online GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/). Samples were log2 transformed and quantile normalized using limma R-package. Lists of DE genes were cut by adjusted p-value at 0.05.

Bulk RNA-seq analysis

FASTQ files were downloaded from GEO. Following, sequenced reads were aligned to the mouse reference genome (mm10, GENCODE M13) using STAR 2.7.0f (Dobin et al., 2013). Strand specificity was determined using Picard Tools 2.18.0. For differential expression analysis edgeR 3.24.3(Robinson et al., 2009) was used. Genes with average read count <5 were filtered out; "upperquartile" method was used for library normalization; data dispersion were estimated using function

estimateGLMRobustDisp(), significance of differential expression was adjusted by BH method, cut-off values for DE analysis were set by FDR < 0.05.

Single-cell RNA-seq analysis

Data for single cell RNA-seq was downloaded as provided by authors from GEO, preprocessed using Scater package (McCarthy et al., 2017), outliers and cycling cells were removed. The remaining cells were analyzed by Seurat package by comparing two groups of cells, young and aged as provided by authors. No further subgrouping of data was considered. As above, DEG lists were cut by FDR < 0.05.

Coefficient of Variation of consistent subseted datasets

First, we extracted all gene names from the re-analyzed Aging List (AL) for consistency >1, yielding approximately 1000 genes. For a set of 12 re-analyzed DE lists we generated series of all possible combinations starting from 1 list, and ending with 11 lists (out of 12 in total). Through all permuted combinations we counted mean frequencies of citations for every predefined gene name, and also CV (stdev/mean)

for the same genes. The scatter plots represents the 1000 genes by their mean consistency rank (x-axis) and CV (y-axis).

Radial plot of individual studies distances to AS

Degrees of similarity on the Figure 2D (radial plot) were calculated as follows.

we estimated relative frequency (RF) – the number of overlapping genes to a given studie to the AS –The distance is then estimated as:

$$dist = 1 - RF$$

Node sizes represent the total number of a DE gene list in each data set.

Resolution modelling

Resolution – ranging from 0 to 1 - was calculated as the fraction of the consistency genes of subseted studies given a number of studies. Assuming 2 publications, then the gene of interest is either present once, not at all, therefore the resolution of the score is 1/2. The resolution can be written in general terms as:

$$\text{resolution} = 1 - \frac{1}{n}$$

n = number of studies

Rank Score

Rank score is calculated for any particular gene list using simple counting algorithm. To a given study, genes with consistency score > 1 (reproducible gene) are summed, otherwise (consistency score < 1, non-reproducible) it is penalized with a score of -1. Thus, the sum of all gene scores is taken across entire list as follows:

$$RS = \sum_{n=1}^{length(test)} \begin{cases} Sc_n & Sc_n > 1 \\ -1 & Sc_n = 1 \end{cases}$$

Test = test gene list

Sc = score: number of times gene was found in all lists (consistency).

Discrete enrichment tool (dGSE)

Independently from RS, dGSE provides a score, which is based not just on presence of overlapping genes, but also takes into consideration the order of appearance based on a ranked list. Similarly to GSEA enrichment (Subramanian et al., 2005), it also utilizes a concept of counting genes from a reference list – in this case AL is taken as such – on top of the test gene list. The test gene list subjected for analysis is first ranked by p-value, most significant on the top. Next it counts the genes from the AL found in discrete subsets of the analyzed set (therefore named discrete gene set enrichment). The subset sizes of the full list are always equal to the sums of sizes of the subgroups in the Aging Signature (beginning from most frequent genes to the least frequent). In our particular case it will be 1 (for *Selp*),1+2 (for *Selp* and [*Mt1*, *Nupr1*]), 1+2+2, 1+2+2+3 and so on.

This analysis comes in two versions. One is calculating enrichment score for the defined window size.

The enrichment score is calculated per fixed window of AS size, 220 genes for re-analysed data. In order to access the overall Enrichment score (EnSc) we firstly determine the number of “observed” and “expected” genes. Observed genes are the number of overlapping genes between the AS and the test set. Expected genes, is defined as the fraction of genes present on the test set and the AS relative to AL

(which is 220/5443 (genes in AS divided by AL). Once observed and expected genes are computed, they are subtracted and normalized to the total number of reported genes, AL, and the result is given as %. The following formula shows the expression used:

$$\mathrm{EnSc} = \frac{length(AS \cup test) - \frac{length(test) \times length(AS)}{length\ (AL)}}{length(test)}$$

Test = test gene list

AS = aging signature genes

AL = long list genes

Length(test) = full length of gene list in the test also found in AL

For graphical representation of the EnSc, we perform the same calculation but instead of performed the score calculation for fixed window (AS, 220 genes), we calculate the EnSc for all groups. Thus, the line demonstrate the cumulative EnSc and the red bars at the bottom demonstrate the position of the genes present in the test set in reference to the ranked AL. It gives a visual impression how well any particular set is enriched for AS.

Quantification and Statistical Analysis

For figures which statistical analysis was applied, data is presented as biological replicates as mean of *N* events, exact *N* values are provided in each figure legend. For experimental data, data was analyzed using the two-tailed, unpaired *Student's t test* function or RM two-way ANOVA with Sidak's multiple comparison test in Prism (GraphPad) for the comparison of two or more groups. Acceptance for *p*-values was set at 0.05 and the significance of each test is provided in each figure legend.

Additional Resources

Aging signature WEB tool site: http://agingsignature.webhosting.rug.nl/

GitHub with all resources: https://github.com/LeonidBystrykh/data-for-manuscript

Supplemental Figure Legends

Figure S1 - Related to Figure 1

A) Representation of mouse age distribution used in all selected studies. Individual dots represent the age (in months) of each age groups from all studies; box plot and violin plots show the distribution of the age groups.

B) Re-analyzed studies cluster by sequencing platform. PCA of normalized and centralized re-analyzed datasets. PC1 and PC2, which are responsible for the majorty of the variation, are shown. Studies are color labelled according to the sequencing platform used.

C) Differences in fold-change across the re-analyzed studies. Boxplots demonstrate the mean and quartiles for each gene fold-change. Black dots demonstrate individual genes which reside outside of lower and upper quartiles. Only DEGs are shown.

D) Correlation between fold-changes found in studies which are present in both meta and re-analyzed analysis. Dots represent individual DEG genes found in both analysis, and lines represent linear curve fitting and its R^2 is shown.

Figure S2 - Related to Figure 2

A) Combined number of consistently reported DEGs across studies. Dots represent the number of DEGs (y-axis) and in how many studies they were reported to be differentially expressed (x-axis). Lines represent the curve fitting to each analysis. Purple and green dots/line indicate meta-analysis and re-analysis, respectively.

B) Predicted and observed number of log-transformed DEGs per kind of analysis. Dots represents the number of times that each gene was found in respect to studies. Purple and green dots/line indicate meta-analysis and re-analysis, respectively.

C/D) Heatmaps showing the pair-wise comparison of studies in meta and re-analyzed analysis. Numbers indicate the number of overlapping genes each study has with another study. The color of each square represents the number of overlapping studies. Fraction of total overlaps represents the percentage of reported genes which overlapped with any other study.

E) Bar plot depicting the number of all reported DEGs. Purple and green bars demonstrate the number of genes which are reported in ≥ 4 studies in meta and re-analysis, respectively. Gray bars indicate the number of DEGs which were found in ≤ 3 studies.

Figure S3 – Related to Figure 3

A) Gene Ontology enrichment analysis of all genes in the aging signature. All genes, regardless of their directionality were used for GO analysis. Represented are the top 10 terms from unfiltered GO “molecular function” category. Numbers in each bar demonstrate the number of genes found per term.

B) Common GO biological process in up and downregulated AS genes. Dots represent GO terms and their FDR values in each analysis are shown; dowregulated genes only (x-asis) and upregulated genes only (y-axis) .

C) Pathway analysis of AS genes divided by directionality. KEGG and REACTOME significantly enriched terms are displayed. Numbers in each bar demonstrate the number of genes found per term. Red and blue bars represent GO terms found to be significantly enriched in upregulated and downregulated AS genes, respectively. Terms were selected based on a FDR < 0.05 cutoff.

D) Isolation of LT-HSCs. Upper panel shows a schematic representation of the FACS gating strategy for isolation of LT-HSCs. Middle and lower panel show the population frequencies for young (middle panel, black) and aged (lower panel, red) mice.

E) RNA content per LT-HSC cell from young and aged mice (n = 8 per group).

F) Quantification of MFI of SYTO of sorted LT-HSC cells derived from and young (black, n = 120 cells) and aged (red, n= 139 cells).

For panels D, E, ± SD is shown. **** $p < 0.0001$.

Figure S4 – Related to Figure 4

A) Isolation of different hematopoietic populations. Left upper panel shows a schematic representation of the FACS gating strategy for isolation of LT-HSC, ST-HSC and MPP; Right upper panel depicts the gating for GMP, CLP and MEP. Middle and lower panels show the population frequencies for young (middle panel, black) and aged (lower panel, red) mice.

B) Isolation of SELP^{low} and SELP^{high} LT-HSC. Similar FACS gating strategy was used for LT-HSC **(see above)**. SELP^{low} and SELP^{high} fractions of LT-HSC population were defined as the lower 30% and the higher 30% SELP-expressing cells.

C) *In vitro* myeloid differentiation of SELP^{low} and SELP^{high} LT-HSCs at day 6, 8 and 10 of culture. Flow cytometry plots show the myeloid population, marked by My (MAC.1/GR.1) and CD16/32 cells.

Figure S5 – related to Figure 5

A) BM flow cytometry plots of recipient mice transplanted with EV or SELP OE cells. Upper panel depicts the gating strategy for different cell populations: total white blood cells (left), platelets (middle) and erythrocytes (right). Cells were measured using CD45.2 (donor cells); CD41(platelet marker) and TER-119 (erythrocytes). Middle and lower panels indicate EV and SELP OE, respectively.

B) SELP OE recipients show impairment of erythroid differentiation. Left panel displays flow cytometry gating for the different erythroid populations from EV and SELP PE recipients. Right panel shows the quantification of each population in each group.

Figure S6 – Related to Figure 6

Gene symbols of all AS genes. Each section demonstrates the gene symbols of genes found in each key term search and also the gene symbols of AS genes which were not associated with any of the key terms.

Figure S7 – related to Figure 7

A) Differences in gene list ranking. Upper panel shows a gene list with continuous values (such as FDR, logFC and other metrics) and is used to rank genes in a continuous order. Lower panel shows a discrete rank gene list, where multiple genes can have the same rank order.

B) The Aging Signature website (version 0.14). On the center of the panel, the frontpage of the website is displayed. The dashboard contains all the links and descriptions to the different functionalities that the website has to offer. Several illustrative examples are shown. Upper right panel (red) shows the "Aging Signature" module, where all 220 genes which compose the re-analyzed aging signature are displayed, together with gene descriptions, consistency scores, logFC. Lower left panel (green) shows the "Search" function where the result of a single or multiple gene search provides information whether genes are either present in the AS, AL or not reported or found. Lower right (turquoise) displays the "dGSE" function where users can provide their own list of genes which will be compared to the AS. The analysis results in different enrichment plots for both enrichment and rank scores. The Aging signature and the different studies of which it is comprised are plotted together with the user input as a reference. Test sets used in this manuscript (Maryanovich and Challen) are also present as examples.

Supplemental Table Titles

Table S1: Overall information of collected transcriptome studies used in this study

Table S2: Gene symbols for meta and re-analyzed aging signature and aging list

Table S3: Gene Ontology (GO) and pathway analysis of re-analyzed aging signature

References

Beerman, I., Bock, C., Garrison, B.S., Smith, Z.D., Gu, H., Meissner, A., Rossi, D.J., 2013. Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. Cell Stem Cell 12, 413–425. doi:10.1016/j.stem.2013.01.017

Beerman, I., Rossi, D.J., 2015. Epigenetic Control of Stem Cell Potential during Homeostasis, Aging, and Disease. Cell Stem Cell 16, 613–625. doi:10.1016/j.stem.2015.05.009

Bersenev, A., Rozenova, K., Balcerek, J., Jiang, J., Wu, C., Tong, W., 2012. Lnk deficiency partially mitigates hematopoietic stem cell aging. Aging Cell 11, 949–959. doi:10.1111/j.1474-9726.2012.00862.x

Challen, G.A., Sun, D., Mayle, A., Jeong, M., Luo, M., Rodriguez, B., Mallaney, C., Celik, H., Yang, L., Xia, Z., Cullen, S., Berg, J., Zheng, Y., Darlington, G.J., Li, W., Goodell, M.A., 2014. Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. Cell Stem Cell 15, 350–364. doi:10.1016/j.stem.2014.06.018

Chambers, S.M., Shaw, C.A., Gatza, C., Fisk, C.J., Donehower, L.A., Goodell, M.A., 2007. Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. PLoS Biol. 5, 1750–1762. doi:10.1371/journal.pbio.0050201

Cheung, P., Vallania, F., Warsinske, H.C., Donato, M., Schaffert, S., Chang, S.E., Dvorak, M., Dekker, C.L., Davis, M.M., Utz, P.J., Khatri, P., Kuo, A.J., 2018. Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. Cell 173, 1385-1397.e14. doi:10.1016/j.cell.2018.03.079

de Haan, G., Nijhof, W., Van Zant, G., 1997. Mouse strain-dependent changes in frequency and proliferation of hematopoietic stem cells during aging: correlation between lifespan and cycling activity. Blood 89, 1543–50.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. doi:10.1093/bioinformatics/bts635

Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M., de Haan, G., 2011. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. J. Exp. Med. 208, 2691–2703. doi:10.1084/jem.20111490

Field, A.E., Robertson, N.A., Wang, T., Havas, A., Ideker, T., Adams, P.D., 2018.

DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. Mol. Cell. doi:10.1016/j.molcel.2018.08.008

Flach, J., Bakker, S.T., Mohrin, M., Conroy, P.C., Pietras, E.M., Reynaud, D., Alvarez, S., Diolaiti, M.E., Ugarte, F., Forsberg, E.C., Le Beau, M.M., Stohr, B.A., Méndez, J., Morrison, C.G., Passegué, E., 2014. Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. Nature 512, 198–202. doi:10.1038/nature13619

Grover, A., Sanjuan-Pla, A., Thongjuea, S., Carrelha, J., Giustacchini, A., Gambardella, A., Macaulay, I., Mancini, E., Luis, T.C., Mead, A., Jacobsen, S.E.W., Nerlov, C., 2016. Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. Nat. Commun. 7. doi:10.1038/ncomms11075

Haas, S., Hansson, J., Klimmeck, D., Loeffler, D., Velten, L., Uckelmann, H., Wurzer, S., Prendergast, Á.M., Schnell, A., Hexel, K., Santarella-Mellwig, R., Blaszkiewicz, S., Kuck, A., Geiger, H., Milsom, M.D., Steinmetz, L.M., Schroeder, T., Trumpp, A., Krijgsveld, J., Essers, M.A.G., 2015. Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors. Cell Stem Cell 17, 422–434. doi:10.1016/j.stem.2015.07.007

Han, Y., Eipel, M., Franzen, J., Sakk, V., Dethmers-Ausema, B., Yndriago, L., Izeta, A., de Haan, G., Geiger, H., Wagner, W., 2018. Epigenetic age-predictor for mice based on three CpG sites. Elife 7. doi:10.7554/eLife.37462

Hidalgo, I., Herrera-Merchan, A., Ligos, J.M., Carramolino, L., Nuñez, J., Martinez, F., Dominguez, O., Torres, M., Gonzalez, S., 2012. Ezh1 is required for hematopoietic stem cell maintenance and prevents senescence-like cell cycle arrest. Cell Stem Cell 11, 649–662. doi:10.1016/j.stem.2012.08.001

Ho, Y.-H., del Toro, R., Rivera-Torres, J., Rak, J., Korn, C., García-García, A., Macías, D., González-Gómez, C., del Monte, A., Wittner, M., Waller, A.K., Foster, H.R., López-Otín, C., Johnson, R.S., Nerlov, C., Ghevaert, C., Vainchenker, W., Louache, F., Andrés, V., Méndez-Ferrer, S., 2019. Remodeling of Bone Marrow Hematopoietic Stem Cell Niches Promotes Myeloid Cell Expansion during Premature or Physiological Aging. Cell Stem Cell 25, 407-418.e6. doi:10.1016/j.stem.2019.06.007

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B.G.,

Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., Higgins, J.M., Moltchanov, V., Kuo, F.C., Kluk, M.J., Henderson, B., Kinnunen, L., Koistinen, H.A., Ladenvall, C., Getz, G., Correa, A., Banahan, B.F., Gabriel, S., Kathiresan, S., Stringham, H.M., McCarthy, M.I., Boehnke, M., Tuomilehto, J., Haiman, C., Groop, L., Atzmon, G., Wilson, J.G., Neuberg, D., Altshuler, D., Ebert, B.L., 2014. Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. 371, 2488–2498. doi:10.1056/NEJMoa1408617

Kirschner, K., Chandra, T., Kiselev, V., Flores-Santa Cruz, D., Macaulay, I.C., Park, H.J., Li, J., Kent, D.G., Kumar, R., Pask, D.C., Hamilton, T.L., Hemberg, M., Reik, W., Green, A.R., 2017. Proliferation Drives Aging-Related Functional Decline in a Subpopulation of the Hematopoietic Stem Cell Compartment. Cell Rep. 19, 1503–1511. doi:10.1016/j.celrep.2017.04.074

Koulnis, M., Pop, R., Porpiglia, E., Shearstone, J.R., Hidalgo, D., Socolovsky, M., 2011. Identification and analysis of mouse erythroid progenitors using the CD71/TER119 flow-cytometric assay. J. Vis. Exp. i, 6–11. doi:10.3791/2809

Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L., Regev, A., Tirosh, I., Haas, B.J., Regev, A., Wagers, A.J., Kowalczyk, M.S., Rao, T.N., Heckl, D., Ebert, B.L., Dixit, A., 2015. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res. 25, 1860–1872. doi:10.1101/gr.192237.115

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., Kroemer, G., 2013. The hallmarks of aging. Cell 153, 1194. doi:10.1016/j.cell.2013.05.039

Mann, M., Mehta, A., de Boer, C.G., Kowalczyk, M.S., Lee, K., Haldeman, P., Rogel, N., Knecht, A.R., Farouq, D., Regev, A., Baltimore, D., 2018. Heterogeneous Responses of Hematopoietic Stem Cells to Inflammatory Stimuli Are Altered with Age. Cell Rep. 25, 2992-3005.e5. doi:10.1016/j.celrep.2018.11.056

Maryanovich, M., Zahalka, A.H., Pierce, H., Pinho, S., Nakahara, F., Asada, N., Wei, Q., Wang, X., Ciero, P., Xu, J., Leftin, A., Frenette, P.S., 2018. Adrenergic nerve degeneration in bone marrow drives aging of the hematopoietic stem cell niche. Nat. Med. 24, 782–791. doi:10.1038/s41591-018-0030-x

McCarthy, D.J., Campbell, K.R., Lun, A.T.L., Wills, Q.F., 2017. Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 33, 1179–1186. doi:10.1093/bioinformatics/btw777

Morrison, S.J., Scadden, D.T., 2014. The bone marrow niche for haematopoietic stem cells. Nature. doi:10.1038/nature12984

Morrison, S.J., Wandycz, A.M., Akashi, K., Globerson, A., Weissman, I.L., 1996. The aging of hematopoietic stem cells. Nat. Med. 2, 1011–1016. doi:10.1038/nm0996-1011

Noda, S., Ichikawa, H., Miyoshi, H., 2009. Hematopoietic stem cell aging is associated with functional decline and delayed cell cycle progression. Biochem. Biophys. Res. Commun. 383, 210–215. doi:10.1016/j.bbrc.2009.03.153

Norddahl, G.L., Pronk, C.J., Wahlestedt, M., Sten, G., Nygren, J.M., Ugale, A., Sigvardsson, M., Bryder, D., 2011. Accumulating mitochondrial DNA mutations drive premature hematopoietic aging phenotypes distinct from physiological stem cell aging. Cell Stem Cell 8, 499–510. doi:10.1016/j.stem.2011.03.009

Pinho, S., Frenette, P.S., 2019. Haematopoietic stem cell activity and interactions with the niche. Nat. Rev. Mol. Cell Biol. doi:10.1038/s41580-019-0103-9

Quere, R., Saint-Paul, L., Carmignac, V., Martin, R.Z., Chretien, M.-L., Largeot, A., Hammann, A., Pais de Barros, J.-P., Bastie, J.-N., Delva, L., 2014. Tif1 regulates the TGF- 1 receptor and promotes physiological aging of hematopoietic stem cells. Proc. Natl. Acad. Sci. 111, 10592–10597. doi:10.1073/pnas.1405546111

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. doi:10.1093/bioinformatics/btp616

Rossi, D.J., Bryder, D., Zahn, J.M., Ahlenius, H., Sonu, R., Wagers, A.J., Weissman, I.L., 2005. Cell intrinsic alterations underlie hematopoietic stem cell aging.

Rossi, D.J., Jamieson, C.H.M., Weissman, I.L., 2008. Stems Cells and the Pathways to Aging and Cancer. Cell. doi:10.1016/j.cell.2008.01.036

Singh, P.P., Demmitt, B.A., Nath, R.D., Brunet, A., 2019. The Genetics of Aging: A Vertebrate Perspective. Cell 177, 200–220. doi:10.1016/j.cell.2019.02.038

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550. doi:10.1073/pnas.0506580102

Sudo, K., Ema, H., Morita, Y., Nakauchi, H., 2000. Age-associated characteristics of

murine hematopoietic stern cells. J. Exp. Med. 192, 1273–1280. doi:10.1084/jem.192.9.1273

Sullivan, C., Chen, Y., Shan, Y., Hu, Y., Peng, C., Zhang, H., Kong, L., Li, S., 2011. Functional ramifications for the loss of p-selectin expression on hematopoietic and leukemic stem cells. PLoS One 6, 1–10. doi:10.1371/journal.pone.0026246

Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., Wang, H., Le, T., Faull, K.F., Chen, R., Gu, H., Bock, C., Meissner, A., Göttgens, B., Darlington, G.J., Li, W., Goodell, M.A., 2014. Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. Cell Stem Cell 14, 673–688. doi:10.1016/j.stem.2014.03.002

Verovskaya, E., Broekhuis, M.J.C., Zwart, E., Ritsema, M., Van Os, R., De Haan, G., Bystrykh, L. V., 2013. Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. Blood 122, 523–532. doi:10.1182/blood-2013-01-481135

Verovskaya, E., Broekhuis, M.J.C., Zwart, E., Weersing, E., Ritsema, M., Bosman Lisette J., J., Van Poele, T., De Haan, G., Bystrykh, L. V., 2014. Asymmetry in skeletal distribution of mouse hematopoietic stem cell clones and their equilibration by mobilizing cytokines. J. Exp. Med. 211, 487–497. doi:10.1084/jem.20131804

Wahlestedt, M., Norddahl, G.L., Sten, G., Ugale, A., Frisk, M.A.M., Mattsson, R., Deierborg, T., Sigvardsson, M., Bryder, D., 2013. An epigenetic component of hematopoietic stem cell aging amenable to reprogramming into a young state. Blood 121, 4257–4264. doi:10.1182/blood-2012-11-469080

Wong, D.J., Liu, H., Ridky, T.W., Cassarino, D., Segal, E., Chang, H.Y., 2008. Module Map of Stem Cell Genes Guides Creation of Epithelial Cancer Stem Cells. Cell Stem Cell 2, 333–344. doi:10.1016/j.stem.2008.02.009

Yamashita, M., Passegué, E., 2019. TNF-α Coordinates Hematopoietic Stem Cell Survival and Myeloid Regeneration. Cell Stem Cell 25, 357-372.e7. doi:10.1016/j.stem.2019.05.019

Figure 1

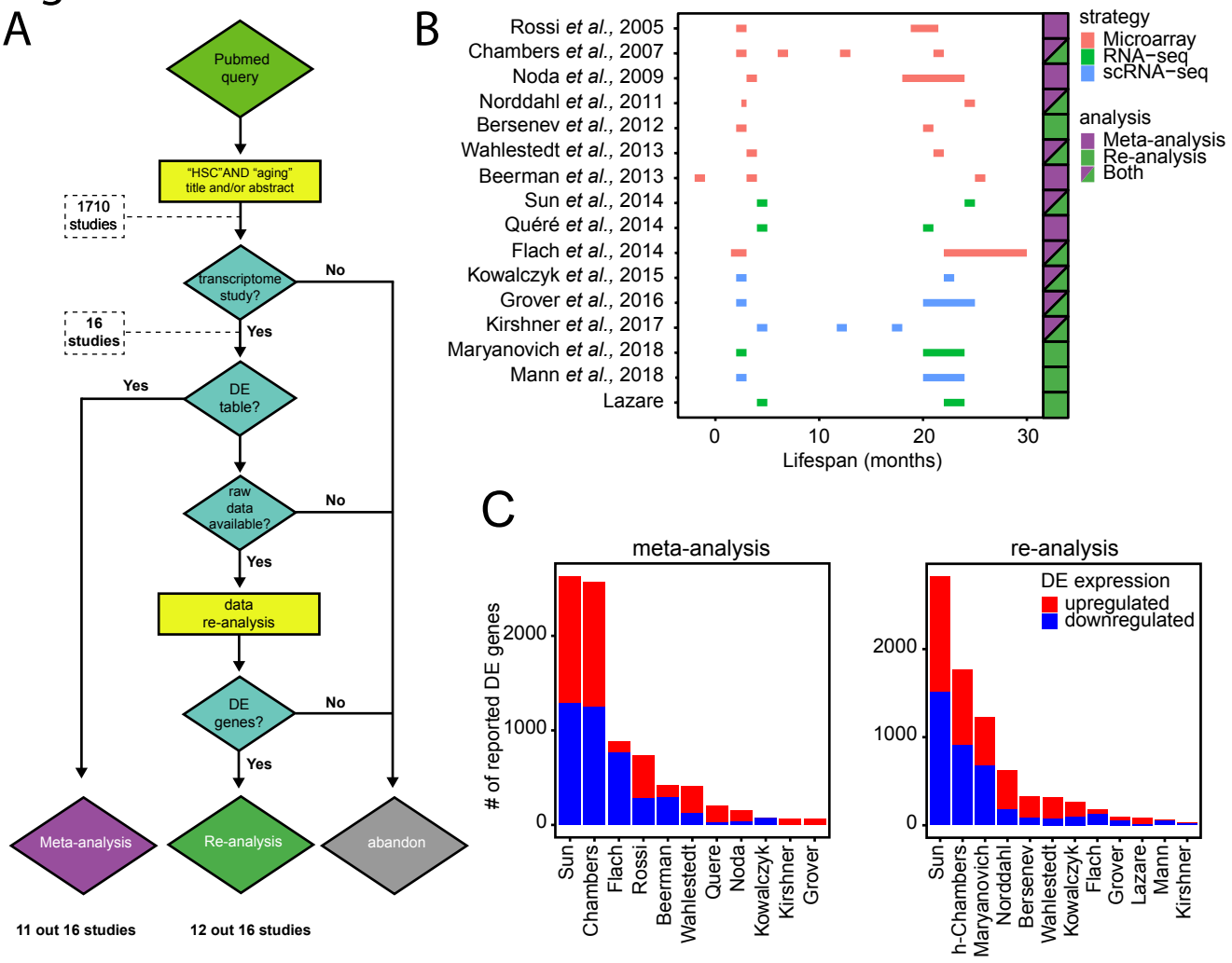

A
Pubmed query
"HSC"AND "aging" title and/or abstract
1710 studies
transcriptome study?
No
16 studies
Yes
DE table?
Yes
raw data available?
No
Yes
data re-analysis
DE genes?
No
Yes
Meta-analysis
Re-analysis
abandon
11 out 16 studies
12 out 16 studies
B
Rossi et al., 2005
Chambers et al., 2007
Noda et al., 2009
Norddahl et al., 2011
Bersenev et al., 2012
Wahlestedt et al., 2013
Beerman et al., 2013
Sun et al., 2014
Quéré et al., 2014
Flach et al., 2014
Kowalczyk et al., 2015
Grover et al., 2016
Kirshner et al., 2017
Maryanovich et al., 2018
Mann et al., 2018
Lazare
0
10
20
30
Lifespan (months)
strategy
Microarray
RNA−seq
scRNA−seq
analysis
Meta-analysis
Re-analysis
Both
C
meta-analysis
re-analysis
# of reported DE genes
0
1000
2000
DE expression
upregulated
downregulated
Sun
Chambers
Flach
Rossi
Beerman
Wahlestedt
Quere
Noda
Kowalczyk
Kirshner
Grover
h-Chambers
Maryanovich
Norddahl
Bersenev
Lazare
Mann

Figure 2

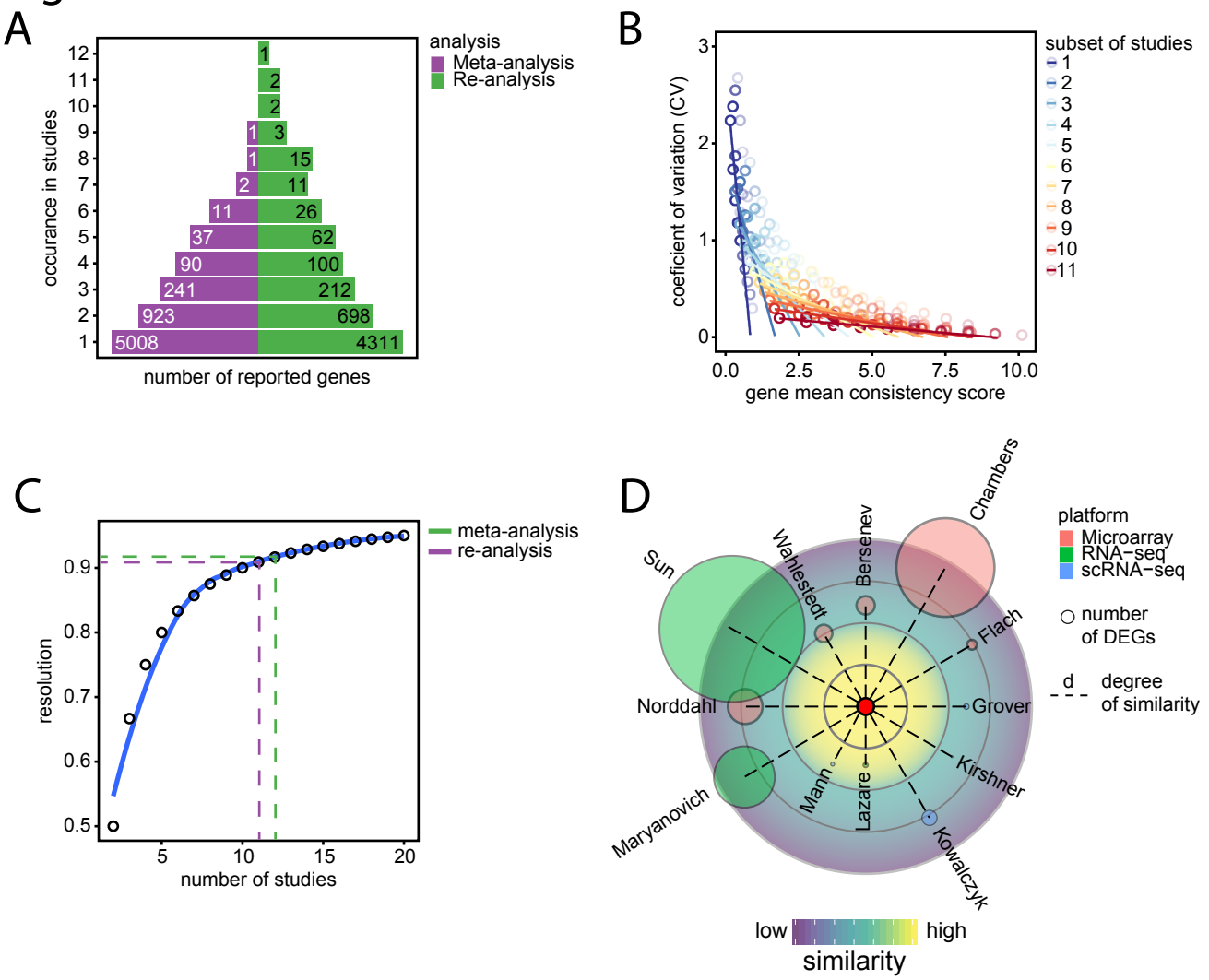

A
analysis
Meta-analysis
Re-analysis
occurance in studies
number of reported genes
B
subset of studies
coeficient of variation (CV)
gene mean consistency score
C
meta-analysis
re-analysis
resolution
number of studies
D
platform
Microarray
RNA−seq
scRNA−seq
number of DEGs
degree of similarity
Sun
Wahlestedt
Bersenev
Chambers
Flach
Norddahl
Grover
Kirshner
Maryanovich
Mann
Lazare
Kowalczyk
low
high
similarity

Figure 3

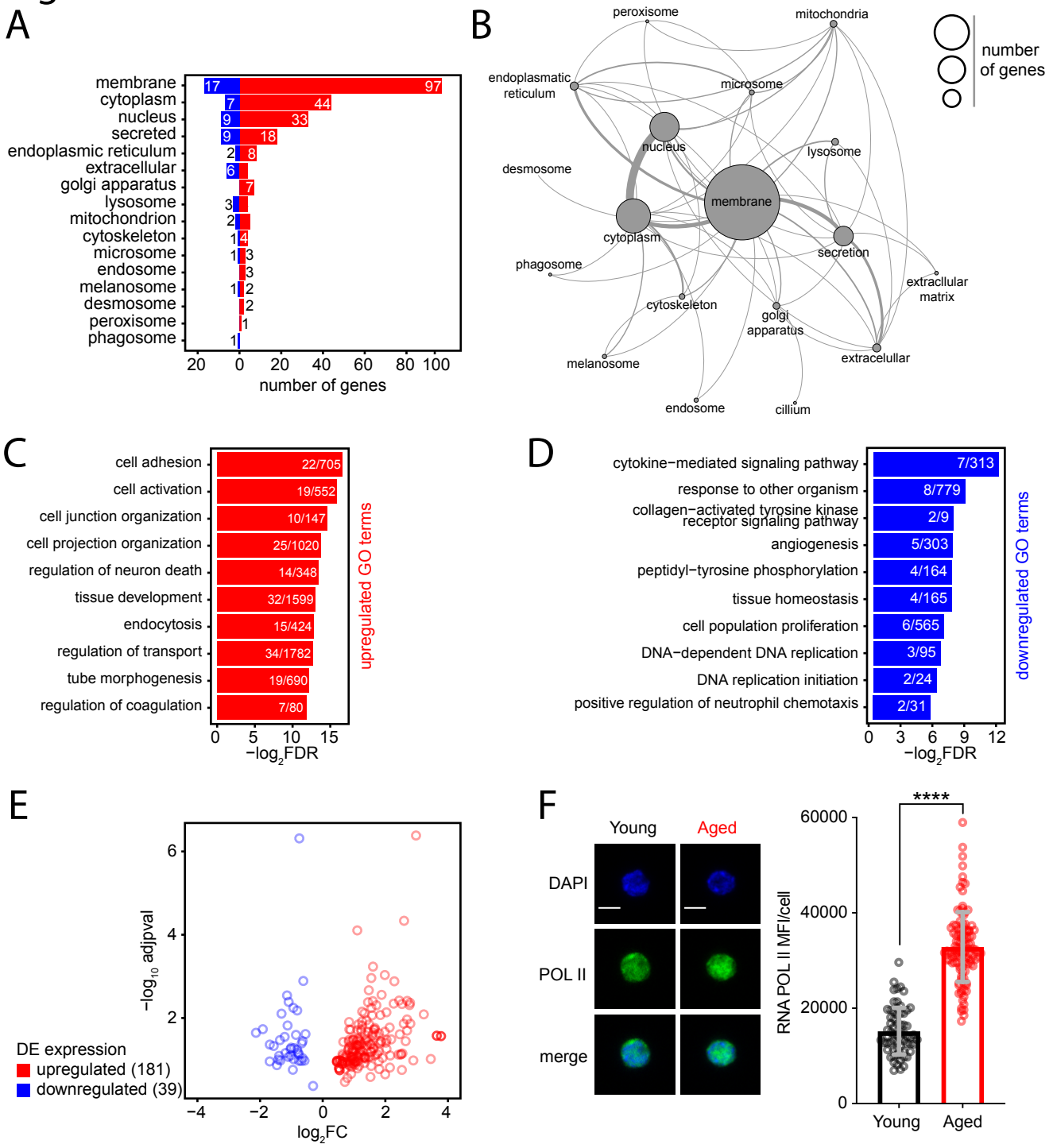

A
membrane
cytoplasm
nucleus
secreted
endoplasmic reticulum
extracellular
golgi apparatus
lysosome
mitochondrion
cytoskeleton
microsome
endosome
melanosome
desmosome
peroxisome
phagosome
number of genes
B
peroxisome
mitochondria
endoplasmatic reticulum
microsome
nucleus
lysosome
desmosome
membrane
cytoplasm
secretion
phagosome
extracllular matrix
cytoskeleton
golgi apparatus
melanosome
extracelullar
endosome
cillium
number of genes
C
cell adhesion
cell activation
cell junction organization
cell projection organization
regulation of neuron death
tissue development
endocytosis
regulation of transport
tube morphogenesis
regulation of coagulation
22/705
19/552
10/147
25/1020
14/348
32/1599
15/424
34/1782
19/690
7/80
upregulated GO terms
−log2FDR
D
cytokine−mediated signaling pathway
response to other organism
collagen−activated tyrosine kinase receptor signaling pathway
angiogenesis
peptidyl−tyrosine phosphorylation
tissue homeostasis
cell population proliferation
DNA−dependent DNA replication
DNA replication initiation
positive regulation of neutrophil chemotaxis
7/313
8/779
2/9
5/303
4/164
4/165
6/565
3/95
2/24
2/31
downregulated GO terms
−log2FDR
E
−log10 adjpval
log2FC
DE expression
upregulated (181)
downregulated (39)
F
Young
Aged
DAPI
POL II
merge
RNA POL II MFI/cell
****
Young
Aged

Figure 4

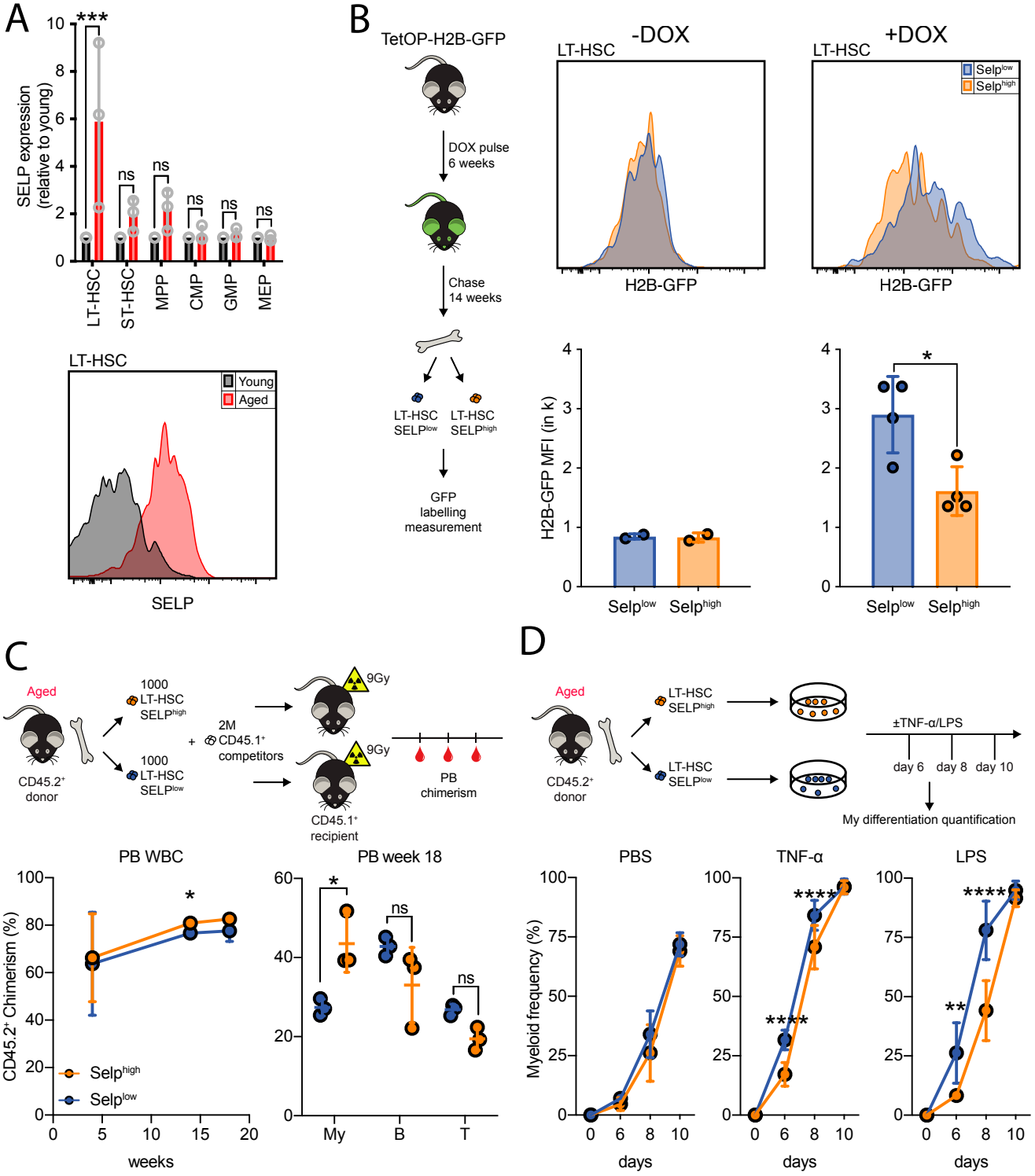

A
SELP expression (relative to young)
***
ns
LT-HSC
ST-HSC
MPP
CMP
GMP
MEP
Young
Aged
SELP
B
TetOP-H2B-GFP
DOX pulse 6 weeks
Chase 14 weeks
LT-HSC SELPlow
LT-HSC SELPhigh
GFP labelling measurement
-DOX
+DOX
H2B-GFP
Selplow
Selphigh
H2B-GFP MFI (in k)
*
C
Aged
CD45.2+ donor
1000 LT-HSC SELPhigh
1000 LT-HSC SELPlow
2M CD45.1+ competitors
9Gy
CD45.1+ recipient
PB chimerism
PB WBC
PB week 18
CD45.2+ Chimerism (%)
weeks
My
B
T
D
LT-HSC SELPhigh
LT-HSC SELPlow
±TNF-α/LPS
day 6
day 8
day 10
My differentiation quantification
PBS
TNF-α
LPS
Myeloid frequency (%)
days
****
**

Figure 5

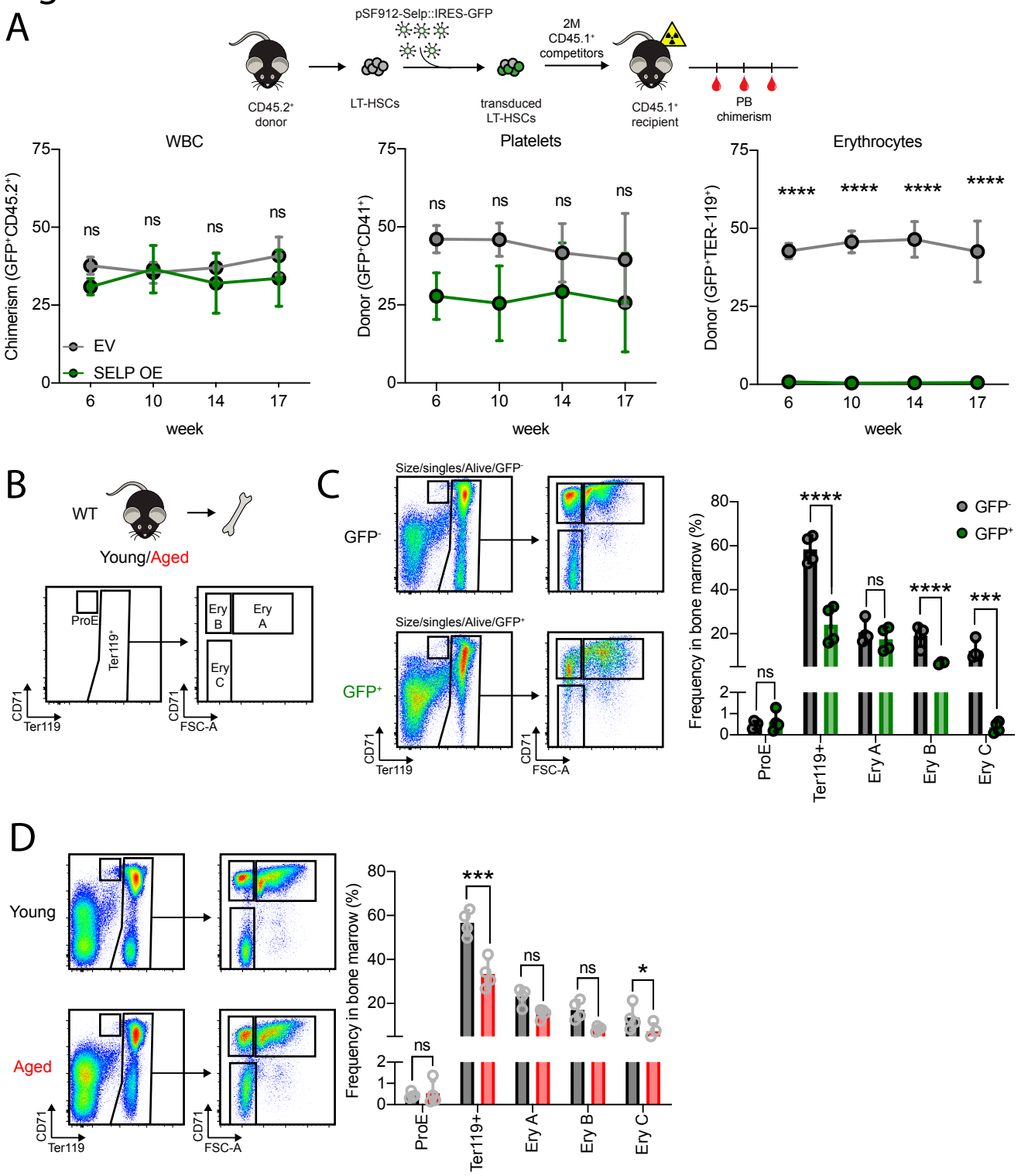

A
pSF912-Selp::IRES-GFP
2M CD45.1+ competitors
CD45.2+ donor
LT-HSCs
transduced LT-HSCs
CD45.1+ recipient
PB chimerism
WBC
Platelets
Erythrocytes
Chimerism (GFP+CD45.2+)
Donor (GFP+CD41+)
Donor (GFP+TER-119+)
ns
****
EV
SELP OE
week
B
WT
Young/Aged
ProE
Ter119+
Ery B
Ery A
Ery C
CD71
Ter119
FSC-A
C
Size/singles/Alive/GFP-
Size/singles/Alive/GFP+
GFP-
GFP+
Frequency in bone marrow (%)
***
ProE
Ter119+
Ery A
Ery B
Ery C
D
Young
Aged
*

Figure 6

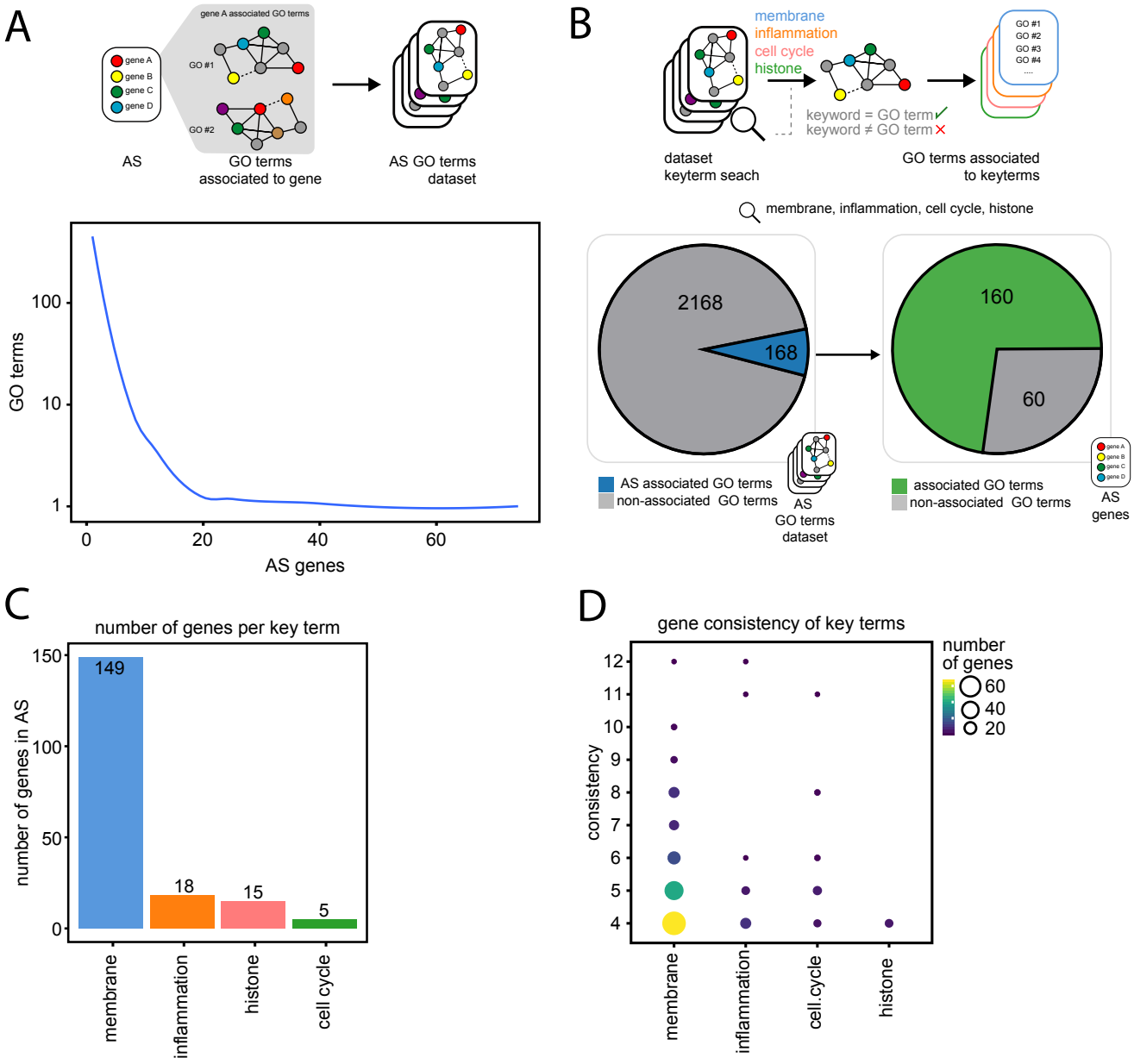
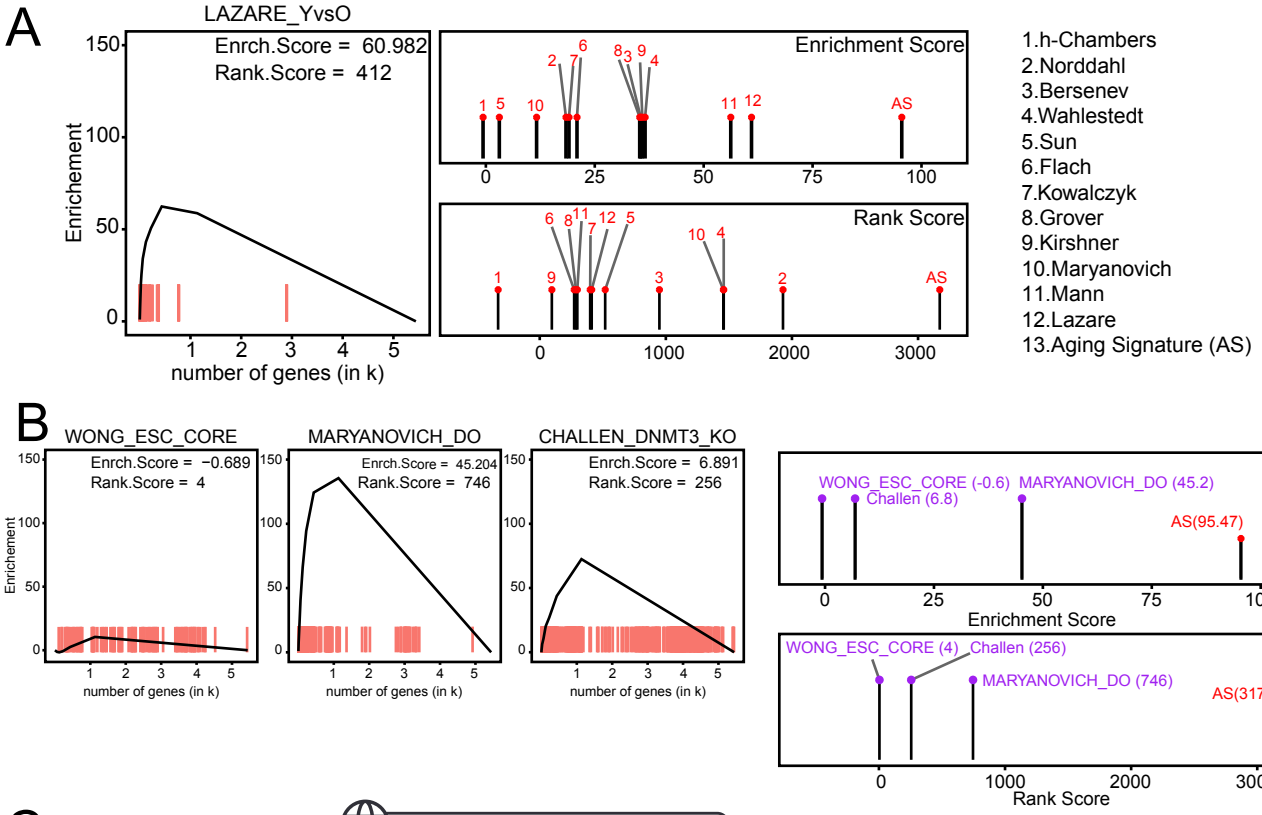
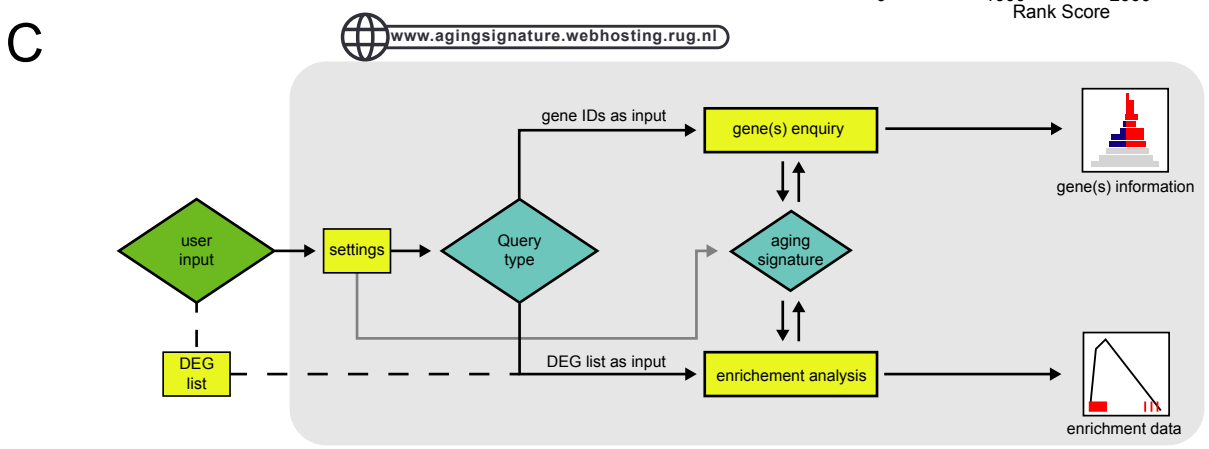

A
gene A
gene B
gene C
gene D
gene A associated GO terms
GO #1
GO #2
AS
GO terms
associated to gene
AS GO terms
dataset
GO terms
AS genes
B
membrane
inflammation
cell cycle
histone
keyword = GO term
keyword ≠ GO term
GO #1
GO #2
GO #3
GO #4
dataset
keyterm seach
GO terms associated
to keyterms
membrane, inflammation, cell cycle, histone
2168
168
160
60
AS associated GO terms
non-associated GO terms
AS
GO terms
dataset
associated GO terms
non-associated GO terms
AS
genes
C
number of genes per key term
number of genes in AS
149
18
15
5
membrane
inflammation
histone
cell cycle
D
gene consistency of key terms
consistency
number
of genes
60
40
20
membrane
inflammation
cell.cycle
histone

Figure 7


A
LAZARE_YvsO
Enrch.Score = 60.982
Rank.Score = 412
Enrichment
number of genes (in k)
Enrichment Score
Rank Score
AS
1.h-Chambers
2.Norddahl
3.Bersenev
4.Wahlestedt
5.Sun
6.Flach
7.Kowalczyk
8.Grover
9.Kirshner
10.Maryanovich
11.Mann
12.Lazare
13.Aging Signature (AS)
B
WONG_ESC_CORE
Enrch.Score = −0.689
Rank.Score = 4
MARYANOVICH_DO
Enrch.Score = 45.204
Rank.Score = 746
CHALLEN_DNMT3_KO
Enrch.Score = 6.891
Rank.Score = 256
WONG_ESC_CORE (-0.6)
MARYANOVICH_DO (45.2)
Challen (6.8)
AS(95.47)
WONG_ESC_CORE (4)
Challen (256)
MARYANOVICH_DO (746)
AS(3172)
Rank Score


C
www.agingsignature.webhosting.rug.nl
gene IDs as input
gene(s) enquiry
gene(s) information
user input
settings
Query type
aging signature
DEG list
DEG list as input
enrichement analysis
enrichment data