Methylomic analysis identifies the involvement of migration and adhesion genes in the ageing of primary haematopoietic stem cells

Oluwatosin Taiwo[1, 2], Gareth A. Wilson[1], Warren Emmett[3], Tiffany Morris[1], Dominique Bonnet[4], Tomas Adejumo[5], Daniel J. Pearce[2#*] and Stephan Beck[1#*]

[1]Medical Genomics, UCL Cancer Institute, University College London, London WC1E 6BT, UK

[2]UCL Institute of Healthy Ageing, University College London, London WC1E 6BT, UK

[3]UCL Genomics, University College London, London WC1E 6BT, UK

[4] Cancer Research UK, London Research Institute, Lincoln's Inn Fields, London WC2A 3LY, UK

[5]Scientific Support Services, Flow Facility, University College London, WC1E 6BT, UK

[#]Equal contributors and senior authors

[*]Correspondence to DJP (daniel.pearce@ucl.ac.uk) or SB (s.beck@ucl.ac.uk)

Abstract

Background: Ageing is a major factor contributing to human morbidity and disease, including cancer. To study the possible involvement of epigenetic changes in ageing, we used murine haematopoiesis as a model system. The key cells determining ageing in this system are thought to be found within the lower side population (LSP) cells of the bone marrow, which are enriched for long-term reconstituting haematopoietic stem cells) (SP-HSCs). We isolated and phenotyped these rare primary SP-HSCs from young, medium-aged and old mice and subjected them to comprehensive methylome (MeDIP-seq) and transcriptome (RNA-seq) analysis.

Results: We found directional (young to old) global loss of DNA methylation of approximately 5% and over 100 significant (FDR <0.2) ageing-specific differentially methylated regions (aDMRs). Enrichment analysis of our transcriptome data revealed significant ($p << 0.0001$) age-dependent changes in pathways and functions associated with cell movement. A number of genes were identified to be both age differentially methylated and differentially expressed, and many of these were associated with migration/adhesion and not previously implicated in ageing. We performed a proof of principle experiment on an aDMR associated with the *Sdpr* gene by ex-vivo culture of isolated SP-HSC in the presence of 5-aza-deoxycytdine and demonstrated dose-dependent negative correlation between promoter methylation and gene expression.

Conclusions: Our data support a model involving epigenetic dysregulation of genes controlling the interaction between SP-HSC and their regulatory niche during ageing.

Keywords

Haematopoietic stem cells, Ageing, Epigenetics, Methylomics, Methylome, Transcriptome, Nano-MeDIP-seq, RNA-seq, DNA methylation and Cell Movement.

Introduction

Diseases of the immune system especially myeloid cancers, inflammatory diseases and immune deficiencies, are major burdens of mammalian ageing. As a result, there is remarkable interest in delineating age-based alterations in primitive cells of the immune system. Haematopoietic stem cells (HSCs) are primitive cells, resident in the bone marrow, which are solely responsible for the maintenance of immune and haematopoietic cells throughout life [1] . The ability of HSCs to self-renew and differentiate into other blood cell types is tightly regulated by both intrinsic mechanisms and the micro-environment or niche in which they reside [2-6]. Several studies have reported a decline in HSC functionality and a differentiation bias towards myelopoiesis with age [7-12]. This occurs concurrently with in an increase in the absolute number and relative frequency of HSCs. Primitive HSCs possess high ABC/G2 pump activity and can be identified via Hoechst dye exclusion assays [13]. HSCs that are identified by their Hoechst exclusion activity are known as side population (SP) cells [13]. In 2007, we reported that the SP subset of murine HSCs accumulates with age [11]. In particular, cells with the highest Hoechst exclusion activity (termed "lower SP cells") increase with age when compared to other subsets of murine stem and progenitor cells [11,14]. As such, analysis of lower SP cells from aged mice and comparison to lower SP from their non-aged counterparts provides an opportunity to examine the nature of age-accumulated murine HSCs that have been identified by a physiological property (ABCG2 activity), rather than surrogate markers. For this reason, we chose to study the ageing of HSCs, which were isolated using the SP method [13]. We have not further purified lower SP cells (e.g. with CD150 and CD48 [15] and will therefore refer to these cells as SP haematopoietic stem cells (SP-HSC). Functional and gene expression changes in ageing HSCs have been well documented [10,12,16]. However, the exact mechanism governing these changes remains to be elucidated. An underlying problem of old age is the decline in the body's ability to respond adequately to damage, a problem attributed to the loss of phenotypic plasticity in somatic stem cells (SSCs). In the event of damage, normal SSCs respond to signals, which

instruct them to differentiate and replace lost cells. This signalling and SSC activation process is likely due to epigenetic regulation of gene expression. DNA methylation (DNAm) is a type of epigenetic modification and a source of genome plasticity, which involves the addition of a methyl group to the carbon-5 position of cytosine bases. This occurs predominantly in a CG context which is enriched at CpG islands (CGIs), but also at other features in the genome [17]. There is a strong possibility that ageing is in part driven by an epigenetic loss of phenotypic plasticity. Indeed, age remains one of the greatest risk factors in carcinogenesis and a key finding in cancer pathogenesis is deleterious changes in genome-wide methylation [18-20]. Furthermore, DNAm has also been shown to be important in HSC regulation and mice lacking enzymes which catalyse DNAm at CG dinucleotides, DNA methyltransferases Dnmt1 and Dnmt3a, have defects in HSC self-renewal and differentiation [21-23].

We previously described an optimised version of methylated DNA immuno-precipitation-based sequencing (MeDIP-seq, [24-30]) with low starting concentrations, termed Nano-MeDIP-seq [31]. Nano-MeDIP-seq is a method which allows the unbiased analysis of genome wide DNA methylation from as little as 50ng DNA. Here, we utilize Nano-MeDIP-seq for the analysis of the SP-HSC methylome and, for the first time, to our knowledge, simultaneously interrogate the methylome and transcriptome of a cell population that is highly enriched for HSCs, in order to explore the underlying causes of changes in HSC functionality during normal ageing.

Results

Isolation of Lower Side Population cells

We sorted lower side population cells (LSPs) from adult female C57b/6 mice at different time points: Young (8 – 12 weeks, referred to as Young), middle-aged (12 months, referred to as Mid) and old (22 - 24 months, referred to as Old). Figure 1A shows the sample collection logic for this study. LSPs have previously been shown to be enriched for HSCs, throughout the murine lifespan [14,32,33]. We enriched for HSCs by selecting for LSPs, which are bone marrow cells defined by the lowest Hoechst staining as determined by their position on the nucleated bone marrow cells (NBMCs) Hoechst profile plot (Fig. 1B). We confirmed, by antibody staining for canonical stem cell markers (Lin^-, $c\text{-}Kit^+$, $Sca\text{-}1^+$ (KLS); $CD150^+$ and $CD48^-$), that LSP cells isolated in this study, (referred to from this point on as SP-HSCs) were enriched for LT-HSCs. (Fig. 1C)

Generation of the SP-HSC methylomes and data quality assessment

We obtained DNA for three separate pools at the aforementioned three time points and generated the corresponding SP-HSC methylomes in triplicates for each age group, using Nano-MeDIP-seq (Fig. 1A). This resulted in an average of 5.25 Gb raw paired-end reads per samples (S.D ±1.15 Gb) of which approximately 90% were successfully paired and aligned to the mouse genome (NCBIM37). After filtering (materials and methods), we obtained an average of 1.92 Gb (S.D ± 0.3 Gb) high quality ($q \geq 10$), uniquely mapping paired-end reads. Approximately 60% of all CpGs in the mouse genome were covered at least 1 fold (Fig. S1), and all MeDIP samples showed a clear CpG enrichment when compared with Input control samples (Average MeDIP enrichment score = 2.73; S.D ± 0.12; n = 9, Average input enrichment score of 1.11; S.D ± 0.03; n = 2). We found a good correlation for all SP-HSC MeDIP samples (average Pearson's correlation of 0.84, S.D ± 0.09, Table S8). One of our

replicates for the 'Old' time point had a lower correlation of approximately 0.7 and was excluded from global analyses to maintain stringency.

Analysis of the ageing SP-HSC methylomes

To investigate DNAm changes in SP-HSC with age, we carried out comparative analyses (materials and methods) on normalised methylome data from the Young, Mid and Old SP-HSC. We detected a global loss of DNAm with age (Fig. 2A) which we quantified to an approximately 5% decrease ($p<0.001$, ,Kolmogorov-Smirnov test) in the average methylation level aggregated over all covered CpG sites in the Old compared to Young samples (Fig. 2B). Age-related gain of DNAm was only observed at CGIs in Old samples (Fig. 2C).

The MeDUSA pipeline [30] was used to determine age-specific differentially methylated regions (aDMRs) between the different time points. 53.8% of the mappable mouse genome had a minimum of 10 reads, which allowed us to confidently detect the presence of aDMRs between our Young, Mid and Old methylomes. A total of 111 significant (FDR < 0.2) aDMRs were found between Old and Young samples, of which, 71 were hypermethylated and 40 hypomethylated in the Old samples (Table. S1). These aDMRs appeared to be randomly distributed across all mouse chromosomes (Fig. 3A), with the exception of chromosome 14 and 16 (where no DMRs were detected) and chromosome 5, where we found a hotspot significantly (p-value < 0.001) enriched for significantly (FDR < 0.01) hypermethylated aDMRs. This hotspot corresponds to the location of at least 4 known members of the *Speer* gene family, which includes 14 genes that encode putative glutamate-rich proteins of unknown function. Speer genes are thought to arise from partial duplications of the Dlg5 gene [34]. Dlg5 is a member of the MAGUK super-family, which are scaffolding proteins that are important in cell adhesion and cell polarity [35].

Using a previously determined cut-off of 100 Kb [20], we were able to associate 97% of all identified aDMRs with genes, henceforth referred to as differentially methylated genes (DMGs) (Table. S1), and observed a progressive change in methylation with age for many aDMGs. Figure 3B shows a DNAm heatmap for all DMRs, displaying progressive and directional DNAm change at several regions in the Young, Mid and Old samples. We used GREAT and Panther analysis programs, for DMR enrichment and DMG enrichment analysis respectively, and did not identify any significant enrichment, suggesting that individual genes or regions, but not particular pathways may be involved in this context.

Using annotation from Ensembl (Ensembl 62), aDMRs were categorised according to their genomic location. Hypermethylated aDMRs were significantly enriched in CGIs (Fishers Exact p-value = 2.33E-09), CGI shores (p-value = 4.30E-09) and exons (p-value = 6.18E-11) and hypo aDMRs were significantly enriched in exons (p-value = 2.14E-06), CGI shores (p-value = 8.39E-05), introns (p-value = 3.6E-03) and intergenic regions (p-value = 0.028). Further analysis using the Ensembl regulatory build (Ensembl 67) revealed that CTCF biding sites were most enriched in the hyper aDMRs and c-MYB sites were most enriched in hypo aDMRs.

Integrated analysis of the SP-HSC methylome and transcriptome

A major advantage of our study is that DNAm and gene expression analysis was conducted on DNA and RNA from the same primary and functionally relevant cells. To assess possible functional effects of the identified aDMRs on gene expression, we performed RNA-seq on mRNAs obtained from the same Young, Mid and Old samples from which the methylomes were generated (Fig. 1A). We analysed 50,601 transcripts for expression and differential expression in three biological replicates each of the Young, Mid and Old samples. For the global analysis, the resulting SP-HSC transcripts were grouped according to four expression levels (0-25, 25-50, 50-75 and 75-100%) and their corresponding mean DNAm levels were determined with MEDIPs (methods) by aggregating DNAm

over three putative promoter regions (-5kb upstream, transcription start site (TSS) and 1st exon), revealing a significant (Kruskal-Wallis, p-value < 0.001) step-wise decrease in promoter methylation with increasing gene expression across all three age groups (Fig. 4). These findings show a strong negative correlation between promoter methylation and gene expression in SP-HSC but no age-specific effect.

Functional analysis of DNA methylation in age-dependent differential gene expression

Of the 50,601 Ensembl transcripts covered by RNA sequencing of ageing SP-HSC, approximately 1290 were found to be significantly (FDR < 0.05) differentially expressed between Young and Old samples, with 430 genes down-regulated and 860 genes up-regulated in Old samples relative to Young samples (Table. S2, Fig. S2A).

Ingenuity pathway analysis (IPA) of the 1290 transcripts revealed significant down-regulation of genes involved in the B cell development pathway (p-value = 2.24E-07) and significant up-regulation of genes involved in pathways such as caveolae mediated endocytosis (p-value = 0.021) and functions such as cell movement (p-value = 5.35E-12) and inflammatory response (p-value < 1E-05). Additionally, transforming growth factor beta 1 (TGFB1), tumour necrosis factor (TNF) and Aryl hydrocarbon receptor (AHR) were identified as the most significantly (p < 1E-08) overrepresented upstream regulators of age-dependent differentially expressed genes. Furthermore, signalling pathway impact analysis (SPIA, [36]) of enriched (KEGG) pathways associated significantly down-regulated SP-HSC transcripts with inhibition of the B-cell receptor signalling pathway (FDR= 0.01) and activation of cytokine-cytokine receptor interaction (FDR = 0.001). Conversely, activation of extracellular matrix receptor interactions (FDR = 0.026) and mammalian circadian rhythm (FDR = 0.026) were associated with significantly up-regulated SP-HSC transcripts with age. Appropriate interaction with the bone marrow niche is critical for HSC regulation, and has been shown to be important in the choice between self-renewal and differentiation [37], as well as HSC mobilization

and homing [6,38,39]. This regulation appears to be altered with age as we found significant (FDR < 0.05) changes in the expression of several adhesion and cell signalling genes, which are involved in HSC-niche interactions, such as Selectins, Integrins, (Proto)cadherins, Adams, Adamts, Matrix metalloproteinases (Mmps) and Cell adhesion molecules (Cams). Additionally, we observed a significant (FDR < 0.05) age-dependent change in the expression of the gene encoding secreted phosphoprotein 1 (SPP1; also known as osteopontin (OPN)) and that of its receptor, CD44. IPA gene ontology analysis of all significantly age-dependent differentially expressed genes, revealed a strong enrichment of genes involved in cell movement (p-value < 1E-10), of which 134 were up-regulated and 85 down-regulated (Table. S4). These genes could be further subdivided into functional categories such as homing, migration, chemotaxis, and infiltration, both up- and down-regulated, while other sub categories like transmigration and extravasation were exclusively down-regulated in HSCs from older animals (Table. S4). Interestingly, analysis of the genes, that were found to be upregulated in this study and in two other studies of HSC ageing (Table S5, [10,12]), revealed cell adhesion to be the most significantly upregulated biological process during HSC ageing (Fig. S3). These findings expose a complex involvement of cell movement, cell adhesion and HSC-niche interactions in HSC ageing.

We also observed a highly significant up-regulation of key genes such as *Wnt4, Jag1, Fgfr2* and *Bmp4* (FDR < 0.001), involved in HSC regulation and self-renewal and validated the differential expression of selected genes involved in the top affected pathways by qRT-PCR (Fig. S2B). These included P-selectin (*Selp*) which has a key role in inflammation and cell movement [40], *Sox4*, a transcription factor crucial for B lymphopoiesis [41], Integrin β3 (*Itgb3*) which is involved in cell adhesion and cell signalling [42,43] and Serum deprivation protein response (*Sdpr*) which is involved in various signalling processes as well as caveolae formation [44,45]. Down-regulation of DNA methyltransferase 1 gene (*Dnmt1*) was also validated by qRT-PCR (Fig. S2B).

14 aDMGs were found to be significantly (FDR < 0.05) differentially expressed (Table. S6). Majority of these aDMGs showed an age progressive increase or decrease in DNAm and gene expression (Fig. S4), suggesting that these genes are differentially regulated with age. Although, the negative correlation between promoter methylation and gene expression is well established, majority of these aDMRs do not occur at promoter regions. To understand the possible association between gene expression and aDMRs, we performed correlation analyses and found significant association with both hypermethylated genes (p-value = 0.0469), and hypomethylated genes (p-value < 0.0001). However, these values suggest that hypomethylation is more significantly associated with changes in gene expression (regardless of whether up- or downregulated). We did not find a significant directional correlation between gene-specific age-differential methylation and gene expression, suggesting that the mechanisms by which DNAm regulates gene expression, during ageing, is complex and remains to be elucidated. For example Trim30a, which was the most significantly (FDR < 0.01) age hypomethylated gene, was found to be significantly downregulated with age (FDR = 3.67E-14). This positive DNAm/gene expression correlation appears counter intuitive; however, the DMR overlaps a U6 non-coding small nuclear RNA (snRNA) (Fig. S5), and it is possible that the loss of DNAm in this region results in the activation of the snRNA, which perhaps causes the down regulation of *Trim30a* through aberrant splicing. We are currently designing a separate study to test this hypothesis. Other notable age-regulated genes include *Itgb3, Nav1, Nav2, Amotl2, Kiss1r* and *Pde8a* (Table. S6), which are all involved in cellular migration [46-51].

Although, the exact mechanism, by which the age differential methylation of specific genes modulates their expression, remains unclear; the integrated methylome/transcriptome analysis described above and shown in Figure 4, suggested that DNAm at gene promoter regions is involved in the regulation of SP-HSC gene expression. We thus sought to identify genes with age-dependent changes in DNAm at the promoter region and concomitant change in gene expression. *Sdpr* is an example of such a gene, showing progressive loss of DNAm with age, as identified by MeDIP-seq and a significant increase in gene expression (FDR < 5E-09) as determined by RNA-seq (Fig. 5A). The *Sdpr*

promoter hypomethylation was validated by bisulfite pyro-sequencing of independent Young, Mid and Old samples (Fig. 5B, Fig. 5C). To demonstrate a possible functional linkage, between the observed methylation and expression changes, we demethylated purified primary SP-HSC by 5-Aza'deoxyCytidine (5-Aza'dC) treatment during a 4 day *ex vivo* culture process. This resulted in a dose-dependent up-regulation of *Sdpr* gene expression (Fig. 5D, p-value = 0.028 for 2 µM treatment). Hypomethylation of the *Sdpr* promoter following 5-Aza'dC treatment was confirmed by bisulfite pyro-sequencing of 2 biological replicates (Fig. 5E, p-value < 0.005). These results, as well as the findings that *Sdpr* was upregulated in Dnmt1-Knockout (KO) mice HSCs [21] and in Dnmt3-KO mice HSCs [23], is consistent with an inverse relationship between DNAm and *Sdpr* gene expression.

Discussion

We conducted the most comprehensive analysis to date of the methylomes and transcriptomes of primitive, primary mouse SP-HSC. This population is highly enriched for HSCs [13] and contains a subset of primitive cells that accumulates with age [52].

We found significant changes in DNAm with age (consistent with studies into human ageing [53,54]), including a 5% reduction in global methylation and more than 100 aDMRs, the majority of which were defined by hypermethylation. A similar age-dependent reduction in global methylation has also been observed in human $CD4^+$ T-cells [55]. The transcriptome analysis revealed well over 1000 differentially expressed genes of which the majority (66%) were up-regulated, which is consistent with the observed global loss of methylation.

Among our novel findings are several cell movement, migration and adhesion genes that were significantly differentially expressed and/or differentially methylated during SP-HSC ageing. Notable examples of age-dependent HSC gene expression alterations include up-regulated targets of the Aryl hydrocarbon receptor (AHR) ligand-dependent nuclear receptor (Table. S3). AHR is known to play a role in HSC regulation and migration through binding of its agonist, 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) [56,57]. Exposure of HSC to TCDD has also been shown to result in increased HSC proliferation and myeloid skewing [58]. Another interesting finding is the age-dependent differential expression of *Spp1* and *CD44*. CD44 is an adhesion molecule that is involved in cell migration and cancer metastasis [59]. Intriguingly, AHR ligand binding has been shown to alter the expression of CD44 [57]. The final example highlighted here is the *Sdpr* gene, which we found to be up-regulated during HSC ageing and demonstrated that this up-regulation may be a result of progressive hypomethylation of the *Sdpr* promoter region with age. One caveat is that since 5-Aza'dC is a global demethylation agent, the effect on SDPR expression may be indirect. However, this proof of principle experiment provides a tool for further analysis of the relationship between DNAm and gene

expression in rare subsets of cells enriched for HSCs. While age-related up-regulation of *Sdpr* in HSCs has previously been observed [10,12], we showed that it is linked to loss of promoter methylation, using an ex-vivo *Dnmt-1* inhibition assay in purified SP-HSC. *Sdpr*, which is also known as *Cavin-2*, was originally identified as a gene whose expression was increased in serum-starved cells [44]; however, an independent role in caveolae formation and cell signalling has since emerged [45]. Caveolae are invaginated lipid rafts that provide a common platform for various signalling molecules, leading to enhanced cell signalling and potent signal transduction. Over-expression of *Sdpr* in cultured cells leads to caveolae elongation [45] and perhaps results in an increased surface area for interacting cell signalling molecules. Interestingly, two out of the top five up-regulated gene networks in our study involved cell-to-cell signalling and interaction. It is therefore possible that age-dependent up-regulation of *Sdpr* is one of the effectors driving this process.

Aberrant up-regulation of genes like *Sdpr* and *Spp1* in ageing HSCs and their subsequent role in abnormal cell signalling, migration and adhesion most likely disrupts the ability of HSCs to interact appropriately with and/or be regulated by the bone marrow micro-environment. The role of CD44 (SPP1 receptor) in carcinogenesis is thought to be mediated by modulating cell migration and adhesion through interaction with integrins [60,61]. Integrins are major cell surface receptors involved in cell-cell and cell-ECM (extra cellular matrix) interactions. We found several members of the Integrin family to be significantly up-regulated during HSC ageing. This includes *Itgb3*, which we found to be significantly hypomethylated with age as well as significantly up-regulated in ageing HSCs (Table S6). The ability of HSCs to respond appropriately to niche-associated signalling is crucial for processes like homing, migration and lodgement. Aged HSCs show reduced ability to home to the bone marrow [11,62,63], which could be due to reduced ability to respond appropriately to regulatory cues within the micro-environment. Additionally, aged early haematopoietic progenitor cells (eHPCs) from old mice showed reduced adhesion to stromal cells and localised more distantly from the BM endosteal niche than eHPCs from young mice [64]. We speculate that age-related defects in HSCs such as reduced lymphopoiesis and expansion of myeloid biased HSCs could be a

result of compromised interactions with the BM, and a sub-optimal response to signals required for normal lymphopoiesis to occur. *Dnmt1* mutant HSCs show global DNA hypomethylation and are unable to sustain lymphopoiesis [22], we suggest that age-related dysregulation HSCs is a consequence of global loss in DNAm.

Conclusions

We have identified several age-dependent alterations in DNAm and corresponding changes in the expression of genes, involved in the regulation of HSC-Niche interactions, such as Itgb3 and Sdpr. Our findings support a model where a shift from homeostatic levels of DNAm in ageing SP-HSC has a wide-spread effect on transcription. This could be a consequence of altered transcription factor binding to epigenetically modified target sites, culminating in abnormal interactions of HSCs with their niche as a result of the dysregulation of key genes involved in HSC migration, homing, adhesion and signalling. Although it is likely that other factors are involved in HSC ageing, our observations led us to conclude that global DNA hypomethylation combined with gene-specific hyper- and/or hypomethylation constitutes an epigenetic mechanism contributing to the decline of HSC functionality with age. While functional validation of these findings remains technically challenging, we were able to demonstrate proof of principle for the *Sdpr*-associated aDMR. Similar assays will be required for other implicated genes to further improve our knowledge of the dynamics of HSC regulation with age.

Materials and Methods

Mice and haematopoietic stem cell preparation

Wild-type Young (8-12 weeks), Mid (~12 month) Old (22-24 months) C57bl/6 mice were sacrificed using an approved method. Mice were originally from the Jackson immune-research laboratories (Bar Harbour, ME) and some of our Old and Mid mice were sourced from Harlan Laboratories (Holland). All of our animals were maintained in a specific pathogen-free animal facility. Whole bone marrow cells were isolated from mouse femur, tibia and in some cases, hip bones, by gentle flushing into 1x HBSS, 10mM Hepes solution (Sigma Aldrich) using a 27G syringe. Erythrocytes were lysed with 3 volumes of ammonium chloride (Stem Cell Technologies,Vancouver, BC, Canada) and debris removed by filtering cell suspension through a 40 μM nylon cell strainer. SP staining was performed as previously described [13]. Briefly, total nucleated bone marrow cells (NBMCs) were resuspended in DMEM staining media at 1 x 10^6 cells per ml and incubated with 5 μg per ml Hoechst 33342 DNA binding dye (Sigma Aldrich) for 90 min at 37^0C with mixing at 30 min intervals. Hoechst stained cells were resuspended in DMEM, 2% FCS and 10 mM HEPES solution at approximately 1 x 10^7 cells/ml. Cells were maintained at 4°C until analysis/sorting and during antibody labelling. Lower SP cells were sorted directly into 1.5 ml eppendorf tubes containing either 1x PBS, or RLT cell lysis buffer (Qiagen) for single RNA extraction or Stemspan Serum-Free Expansion Media (SFEM, Stem Cell Technologies) for *ex-vivo* culture. Approximately 5 x 10^6 Hoechst stained NBMCs were analysed for LT-HSC immunophenotype by SP;KLS;CD150,CD48. Antibody labelling was performed as previously described [11] using the following antibodies: CD150 – phycoerythrin (PE, Biolegend), CD48 – PE/Cy5 (Biolegend), Sca-1 – PE/Cyanine7 (PE/Cy7, Biolegend), CD 117 (c-Kit) – allophycocyanine/Cy7 (APC/Cy7, eBioscience) and Biotinylated Lineage antibody cocktail (anti-CD5, anti-CD11b, anti-B220, anti–Gr-1, and anti–Ter-119), subsequently labelled with Percp-conjugated streptavidin (SA-Percp, BD Biosciences). 2 μg/ml Propidium iodide (PI, Sigma Aldrich) was added to resuspension media to

facilitate the exclusion of dead cells during cell sorting on the Moflo cell sorter (Beckman Coulter) and NBMCs analysis on the LSRII (BD Biosciences). PI was substituted for 1:1000 Dapi (Sigma Aldrich) for post sort analysis on the Cyan ADP (Beckman Coulter).

Ex vivo culture and aza-cytidine treatment of primary SP-HSCs

Lower SP cells were purity sorted into SFEM and cultured as previously described [65] but with minor adjustments to cytokine concentrations. Cells were incubated at 37^0C and 5% CO_2, in a round bottom 96 well plate (corning) at 1000 – 2000 cells per 200 µl SFEM plus 2 µg Heparin and cytokines (3 ng recombinant mouse SCF, 6 ng recombinant mouse TPO, 6 ng recombinant mouse IGF-II, 3 ng recombinant human FGF-I and 30 ng recombinant mouse ANGPTL3). For 5-Aza'dC induced hypomethylation, *ex vivo* cultured cells were treated with two doses of 0 µM – 2 µM 5-Aza'dC. The first dose was applied approximately 18 h after culture was initiated. Cells were maintained for a further 30 h after which 50% of media was replaced with fresh 100 µl SFEM plus 1 µg Heparin and 33.3% of the original cytokine concentrations listed above. 50% of initial 5-Aza'dC dose was added to cells 18 h after media replacement and *ex vivo* culture was continued for another 30 h. LSP cells were cultured *ex vivo* for a total of 4 days.

DNA isolation and methylation analysis

150 – 300 ng of DNA was isolated from 7-8 x 10^4 LSP cells in RLT lysis buffer. 30 - 40 Young mice and 3 - 10 Mid or Old mice were pooled per sample. DNA and RNA were simultaneously extracted from pooled LSPs using the AllPrep DNA/RNA Kit from Qiagen according to manufacturer's instructions. Methylation analysis of extracted DNA was performed by Nano-MeDIP-seq as previously described [31]. MeDIP libraries of 190 – 200 bp was prepared for three biological replicates each for 2, 12 and 24 month old mice and subjected to 36 bp paired end (PE) sequencing on either the Illumina GA IIx ("Old 1", "Old 2" and "Young 1") or the Illumina Hi-seq platform for all other samples. Pyro-

sequencing was used for validation. Briefly, 1 x 10^4 cells per sample were bisulfite converted using EpiTect Plus LyseAll Bisulfite Kit and bisulfite conversion efficiency was found to be at least 99% for all samples as determined by qPCR. Control reactions on unmethylated and *in vitro* methylated samples were also conducted to confirm a linear detection of methylation levels by pyro-sequencing. Pyro-sequencing and PCR amplification of candidate regions were conducted using the PyroMark assay design software and PyroMark PCR kit respectively. Pyro-sequencing was performed using the PyroMark Gold Q96 reagents and the PyroMark Q96 MD Pyro-sequencer. All materials used for Pyro-sequencing were from Qiagen, except for primers which were obtained from Sigma. All methods were performed according to manufacturer's instructions.

RNA purification and transcript analysis

50 – 100 ng high quality RNA was obtained from DNA-depleted lysates of pooled LSP cells as described above. Quality of RNA used in this study was determined by Agilent Bioanalyser RNA Pico assay (Agilent). All RNA had a minimum RIN of 7.5, with the majority above 8. mRNA was isolated from total RNA by oligo dT magnetic bead enrichment and 170 bp (insert size) cDNA libraries were prepared using the Illumina Tru-seq mRNA-seq kit (Illumina). 36 bp PE sequence reads were generated for 3 biological replicates of Young, Mid and Old samples on either the Illumina GA IIx for one Old and Young replicate or the Illumina Hi-seq platform for all other samples. 36 bp Hi-seq mRNA-sequence data was technically validated by independent generation of additional 50 bp PE reads also on the Hi-seq sequencer. Both platforms performed equally well. mRNA isolation, cDNA library preparation and mRNA-sequencing were all performed according to manufacturer's instructions. Differential expression of candidate genes was analysed by quantitative reverse transcriptase- PCR (qRT-PCR). Here, 0.5 – 1 x 10^{-4} cells per sample were converted to cDNA using the SuperScript® III CellsDirect cDNA Synthesis Kit (Invitrogen). qRT-PCR of individual transcripts was then performed using Taq-Man gene expression assays (Applied Biosystems). All steps were

completed according to manufacturer's instructions. Gene expression was normalised to β-Actin housekeeping gene and fold change (FC) was calculated by the ΔΔCt method.

MeDIP-seq data analysis

The generated MeDIP-seq data were analysed using our computational pipeline MeDUSA [30] v1.0.0 and the MEDIPS [30,66] v1.0.0 R bioconductor package. MeDUSA constitutes several discrete stages of analysis. Briefly, BWA v0.5.8 [67] was used to align the paired end sequence data to the reference mouse genome build (NCBIM37) using default settings. Filtering was performed to remove reads failing to map as a proper pair and those pairs in which neither read scored an alignment score of >=10. In cases of non-unique reads, all but one pair was removed. Quality control was performed using the tool FastQC v0.9.4 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) and MEDIPS. Additionally between replicate genome-wide correlations were calculated using QCSeqs from the Useq package [68]. Correlations were calculated using a window size of 500bp, increasing in 250bp increments. A minimum number of 5 reads in a window was required prior to inclusion in the correlation. DMRs were called using the Bioconductor package DESeq [69]. DESeq was run with the estimateDispersions sharing mode set as 'fit-only'. Only regions containing a minimum of 10 reads summed from the cohorts being compared were included for DMR analysis (termed our bump-list).
In order to identify regions of the genome enriched for DMRs, permutation analysis was performed. Our DMRs were placed in 100kb windows across the genome and the most enriched location was identified and the relevant DMR count stored. 111 regions were randomly selected from the bump-list and placed in 100kb windows. The windows were assessed to determine how often a region was found greater than our observed data. This was performed 1000 times in order to calculate an empirical p-value for our region of interest.

Genome-wide methylation scores were generated using the MEDIPS package. The paired reads were not extended and were placed in bins of 50bp across the genome. The CpG coupling vector was generated with fragment lengths of 700bp and using the 'count' function. Circos v0.55 [70] was

used to generate a circular genome plot displaying these methylation scores. Feature annotation was obtained from Ensembl 62 and from the Ensembl Regulatory Build 67. CpG island shores were defined as extending 2000bp upstream and downstream of an annotated CpG island. DMRs and bumps were placed in their relevant features using custom perl scripts and bedtools [71] v2.10.2. We used the Fishers Exact test to identify feature types enriched for DMRs.

RNA-seq data analysis

Transcript counts were generated using RSEM [72] v1.1.21. RSEM provides an opportunity to accurately characterise transcription on the isoform level by taking read mapping uncertainty into account using an appropriate statistical model. RSEM aligns reads against a transcriptome library using BOWTIE [73] and the resulting alignments are used to estimate transcript abundances. The RSEM reference was assembled using the mm9 genome build and the respective UCSC known gene reference annotation. Differential expression of the transcript counts was undertaken using the Bioconductor package, DESeq [69] v2.10 using the recommended parameters. Count data was normalised by library size using the function *estimateSizeFactors*. Variance between the observed data and calculated mean was then calculated with *estimateDispersions*. The expression comparisons between samples were then carried out using a model based on the negative binomial distribution. All transcripts below an adjusted p-value (FDR) of 0.05 were selected for further analysis. This threshold was selected to compensate for the loss of power due to division of read counts amongst transcripts within a gene.

Integrated methylome and transcriptome analysis

Expression and methylation levels were intersected for each transcript. In order to determine if there was a relationship between expression and methylation across the gene, the transcripts were ranked by expression level in the different samples, and methylation levels were calculated for each transcript using MEDIPS. Promoter regions were defined as being 1000bp upstream of the TSS. The

Wilcoxon rank sum test was used to determine the significance of differences in methylation between high and low expressed transcripts.

Gene enrichment analysis

GREAT [74] v2.0.2 was used to investigate the biological significance of DMRs. The genomic co-ordinates of the DMRs were passed to GREAT via the web interface (http://great.stanford.edu/public/html/index.php). The analysis was run using the 'Basal plus extension' parameter for genomic region-gene association. Proximal distance was set to default (5000 bp upstream and 1000bp downstream) and the maximum extension was set at 100 kb. Only annotations that were deemed to be significant by both binomial and hypergeometric tests were considered. Ingenuity pathway analysis (IPA) software was used to determine functional enrichments of age differentially expressed transcripts. Ensembl transcript ID of significantly differentially expressed genes (FDR < 0.05) was used as IPA input. Annotated experimental and prediction based evidence from mammalian species, within the Ingenuity knowledge base, was used to determine functional associations within data sets. IPA analysis was used to identify significantly enriched functions, canonical pathways, gene networks and upstream regulators of differentially expressed data sets. Right-tailed Fisher's exact test was used to determine the statistical significance of over-represented pathways and functions. In addition, hypergeometric distribution and right-tailed Fisher's exact test was used to determine the significance of over-represented gene networks. IPA awards scores to gene networks based on statistical significance, we converted this score to fisher's exact p-values by the formula, $10^{(-score)}$. A threshold of p-value < 0.05 was used to determine statistical significance in this study; this translates to a probability of less than 0.05 that each association is due to chance. Finally, Signalling Pathway Impact Analysis [36] (SPIA v2.7.1) was also used. SPIA uses the information from a list of differentially expressed genes and their log-fold changes together with signalling pathways topology to identify the pathways most associated to the

condition under study. Entrez IDs were obtained for the differentially expressed transcripts identified from the RNA-seq data and SPIA was run on this input with 2000 bootstraps.

Panther v7.2 ‘gene expression tools’ [75] was used to determine enriched biological processes of DMGs and of upregulated genes that were common to this study and previous studies of HSC ageing. A list of age-dependent upregulated HSC genes for the previous studies were obtained from published data by Rossi et.al [10] and Chambers et.al [12]. Ensembl gene IDs were used as input for Panther.

Abbreviations

HSC, haematopoietic stem cell; LT-HSC, Long term reconstituting haematopoietic stem cell; SP, side population; LSP, lower side population; My-bi, myeloid biased; SSCs, somatic stem cells; NBMCs, nucleated bone marrow cells; DNAm, DNA methylation; CGI, CpG island; MeDIP, methylated DNA immuno-precipitation; MeDIP-seq, methylated DNA immuno-precipitation based sequencing; DMR, differentially methylated region; DMG, differentially methylated gene; aDMR, age differentially methylated region; qRT-PCR, quantitative reverse transcriptase polymerase chain reaction; SFEM, Serum-Free Expansion Media; Aza or 5-Aza'dC, 5-Aza'deoxyCytidine; FDR, false discovery rate; S.D, standard deviation; FC, fold change; RIN, RNA integrity number.

Author Contributions

O.T., D.J.P. and S.B. conceived the study. O.T. performed the experiments. G.A.W and W.E. did the bioinformatics and statistical analysis. O.T., G.A.W., W.E., and T.M. analysed data. D.B., D.J.P. and S.B. contributed materials. O.T., T.A. and D.J.P. performed FACS analysis. O.T., S.B and D.J.P. wrote the paper with contributions from all co-authors.

Competing Interest

The authors declare that they have no competing interest

Acknowledgements

OT was supported by a PhD studentship from the UK Medical Research Council. The experimental part was supported by EU-FP7 project IDEAL-HEALTH-2010-259679. DP was supported by The Royal Society (RG2009/R1), EU-FP7 project IDEAL (259679) and the BBSRC (56102). Research in the Beck laboratory was supported by: Wellcome Trust (084071), Royal Society Wolfson Research Merit Award (WM100023) and EU-FP7 projects IDEAL (259679), EPIGENESYS (257082) and BLUEPRINT (282510). We thank all the staff of UCL Biological Service Unit, UCL Genomics and Dr Ayad Eddaoudi of the ICH FACS Laboratory for technical assistance.

Figure legends

Figure 1

(A) Schematic of study design. SP-HSC were isolated from C57bl/6 mice at three time points from pools of 30 – 45 mice and 3 - 10 mice for each replicate of Young and Mid/Old samples respectively. DNA and RNA was concurrently isolated from FACs purified SP-HSC. 150 - 300 ng of DNA was used for methylome analysis by MeDIP-seq and 50 – 100 ng RNA was used for transcriptome analysis by RNA-seq. Three biological replicates were carried out for each step as indicated by (x3). **(B)** Side population (SP) analysis of Hoechst stained bone marrow cells by FACs. Cells were first gated to select for live cells and Hoechst staining (left), to select for lower SP (LSP) cells that show uniformity in their scatter profile (middle and right). **(C)** Confirmation that LSP cells from young, mid and old mice are enriched for LT-HSCs, by their positivity for CD117 (c-Kit), Sca-1 and CD150, as well as negativity for lineage antigens and CD48.

Figure 2

(A) Circos plot of SP-HSC methylomes. Outermost circle shows an ideogram of the mouse genome subdivided by chromosomes. The three blue circles show the Old, Mid and Young SP-HSC methylomes, respectively. The colour code indicates the level of global methylation (dark blue, high and light blue low). The innermost circle depicts CpG density (dark orange, high and light orange low). **(B)** Box plots of global and **(C)** CGI, DNA methylation levels for Young, Mid and Old SP-HSC. DNA methylation levels are expressed as MEDIPS scores.

Figure 3

(A) Location of significant (FDR < 0.2) age-specific differentially methylated regions (aDMRs). aDMRs were mapped to corresponding genomic coordinates on mouse chromosomes and are represented on an Ensemble-generated ideogram (Ensemble, mm9 genome build NCBIM37). Hyper aDMRs are shown in blue and hypo aDMRs in red. **(B)** Methylation heatmap of all aDMRs. Yellow indicates hypomethylation and blue indicates hypermethylation.

Figure 4

Box plots showing negative correlation between promoter methylation (aggregated over -5 kb upstream, TSS and 1st Exon) and gene expression (subdivided into 4 percentiles) in Young, Mid and Old SP-HSC.

Figure 5

(A) Screenshot of MeDIP-seq profiles for Sdpr gene in UCSC genome browser (UCSC mm9). Tracks are colour-coded blue, green and orange for Young, Mid and Old samples, respectively. Three biological replicates are shown for each age group. MeDIP- and RNA-seq data are shown for all samples as indicated. Peak height represents DNA methylation or gene expression levels. The two aDMRs identified in the Sdpr gene are boxed in red and black. The red box depicts the aDMR selected for validation (B-E). This region aligns with other members of the human CAVIN gene family. 'Reg. Feature' denotes the location of the Ensembl promoter associated regulatory feature ENSMUSR00000405109. The promoter associated H3K9ac, H3K36me3, H3K4me3 and PolII are localized to this region. **(B)** Validation of *Sdpr* promoter hypomethylation (aDMR boxed in red (A)) using bisulfite pyro-sequencing. Data was obtained from SP-HSC pools for three biological replicates for Young samples and four biological replicates each for Mid and Old samples. A methylation heatmap for the individaul CpGs assayed, and for each replicate is shown. This data is also displayed as a chart **(C). (D)** Functional analysis of same *Sdpr* aDMR by *ex-vivo* culturing of SP-HSC in the

presence of Aza-deoxycytidine and expression analysis using qRT-PCR. Data was obtained from independent experiments for a total of four biological replicates. Expression values shown are relative to untreated control. **(E)** Validation of same *Sdpr* aDMR for promoter hypomethylation following *ex-vivo* culturing in the presence of Aza-deoxy-cytidine Data is shown for two biological replicates. 2-3 mice were pooled per replicate and average value for two treatment replicates is shown for each biological replicate (replicate 1 and 2 – blue and red bars respectively). FC is fold change. Error bars depict standard error of the mean (S.E.M) in all cases ($^{*}P < 0.05$; $^{**}P < 0.005$; $^{***}P < 0.0005$).

Additional data files

The following additional data are available with the online version of this paper;

Additional data file 1 - Figure S1. CpG coverage plots for Young (Y 1-3), Mid (M1-3) and Old (O 1-3) SP-HSC methylomes: Approximately 60% of all CpGs in the genome were covered at least once and 24-33% were covered at least five times. This depth is typical of MeDIP-seq data and sufficient for differential methylation analysis.

Additional data file 2 - Figure S2. (A) Heatmap of significantly expressed transcripts (FDR < 0.05). **(B)** Validation of differential expression of *Sdpr, Selp, Itgb3, Sox4* and *Dnmt1* by qRT-PCR. Data were normalized to expression of housekeeping gene Beta-Actin. RNA was obtained separately to that used for RNA-Seq.

Additional data file 3 - Figure S3. Analysis of HSC gene expression during ageing. Overrepresented biological processes of genes that were upregulated during HSC ageing and that were common to this study, Chambers et.al [12] and Rossi et.al [10].

Additional data file 4 - Table S1. List of genes associated with aDMRs are shown with FDR scores for: **(A)** hyper aDMRs and **(B)** hypo aDMRs. FDR scores ≥ 20, ≥ 13, ≥ 10 and ≥ 7 correspond to false discovery rates of ≤ 0.01, ≤ 0.05, ≤ 0.1 and ≤ 0.2 respectively.

Additional data file 5 - Table S2. Genes that were significantly differentially expressed between Young and Old samples.

Additional data file 6 - Table S3. AHR target genes that were up-regulated with age.

Additional data file 7 - Table S4. Significantly age-related differentially expressed genes associated with cell movement.

Additional data file 8 - Table S5. A list of genes upregulated during HSC ageing common to this study, Chambers et.al [12] and Rossi et.al [10]

Additional data file 9 - Table S6. List of genes found to be differentially methylated and differentially expressed in ageing SP-HSC. Blue and yellow shadings represent hyper- and hypo-methylation respectively. DMG stands for differentially methylated genes. Log2 FC stands for the log2 fold change in expression values from Young to Old.

Additional data file 10 - Table S7. Oligonucleotide sequences

Additional data file 11 - Figure S4. **(A)** Methylation heatmap of age-differentially expressed aDMGs **(B)** RNA-seq expression levels of age differentially expressed aDMGs. The differentially expressed transcripts are shown.

Additional data file 12 - Figure S5. UCSC track showing the Trim30a aDMR for Young, Mid and Old mice. This DMR overlaps a non-coding RNA (arrowed).

Additional data file 13 - Table S8. Correlation values for MeDIP-seq and RNA-seq samples.

Accession Numbers

The generated MeDIP-seq and RNA-seq data were deposited into the GEO database under accession number GSE41658 .

Reference List

1. Spangrude GJ, Heimfeld S, Weissman IL: **Purification and characterization of mouse hematopoietic stem cells.** *Science* 1988, **241:** 58-62.

2. Park IK, Qian D, Kiel M, Becker MW, Pihalja M, Weissman IL *et al*.: **Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells.** *Nature* 2003, **423:** 302-305.

3. Calvi LM, Adams GB, Weibrecht KW, Weber JM, Olson DP, Knight MC *et al*.: **Osteoblastic cells regulate the haematopoietic stem cell niche.** *Nature* 2003, **425:** 841-846.

4. Zhang J, Niu C, Ye L, Huang H, He X, Tong WG *et al*.: **Identification of the haematopoietic stem cell niche and control of the niche size.** *Nature* 2003, **425:** 836-841.

5. Arai F, Hirao A, Suda T: **Regulation of hematopoiesis and its interaction with stem cell niches.** *Int J Hematol* 2005, **82:** 371-376.

6. Nilsson SK, Johnston HM, Whitty GA, Williams B, Webb RJ, Denhardt DT *et al*.: **Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells.** *Blood* 2005, **106:** 1232-1239.

7. Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL: **The aging of hematopoietic stem cells.** *Nat Med* 1996, **2:** 1011-1016.

8. Sudo K, Ema H, Morita Y, Nakauchi H: **Age-associated characteristics of murine hematopoietic stem cells.** *J Exp Med* 2000, **192:** 1273-1280.

9. Kim M, Moon HB, Spangrude GJ: **Major age-related changes of mouse hematopoietic stem/progenitor cells.** *Ann N Y Acad Sci* 2003, **996:** 195-208.

10. 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:** 9194-9199.

11. Pearce DJ, Anjos-Afonso F, Ridler CM, Eddaoudi A, Bonnet D: **Age-dependent increase in side population distribution within hematopoiesis: implications for our understanding of the mechanism of aging.** *Stem Cells* 2007, **25:** 828-835.

12. 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:** e201.

13. Goodell MA, Brose K, Paradis G, Conner AS, Mulligan RC: **Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo.** *J Exp Med* 1996, **183:** 1797-1806.

14. Challen GA, Boles NC, Chambers SM, Goodell MA: **Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1.** *Cell Stem Cell* 2010, **6:** 265-278.

15. Yilmaz OH, Kiel MJ, Morrison SJ: **SLAM family markers are conserved among hematopoietic stem cells from old and reconstituted mice and markedly increase their purity.** *Blood* 2006, **107:** 924-930.

16. Pearce D, Bonnet D: **Ageing within the hematopoietic stem cell compartment.** *Mech Ageing Dev* 2009, **130:** 54-57.

17. Suzuki MM, Bird A: **DNA methylation landscapes: provocative insights from epigenomics.** *Nat Rev Genet* 2008, **9:** 465-476.

18. Feinberg AP, Vogelstein B: **Hypomethylation distinguishes genes of some human cancers from their normal counterparts.** *Nature* 1983, **301:** 89-92.

19. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P *et al*.: **The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.** *Nat Genet* 2009, **41:** 178-186.

20. Feber A, Wilson GA, Zhang L, Presneau N, Idowu B, Down TA *et al*.: **Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors.** *Genome Res* 2011, **21:** 515-524.

21. Trowbridge JJ, Snow JW, Kim J, Orkin SH: **DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells.** *Cell Stem Cell* 2009, **5:** 442-449.

22. Broske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M *et al*.: **DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction.** *Nat Genet* 2009, **41:** 1207-1215.

23. Challen GA, Sun D, Jeong M, Luo M, Jelinek J, Berg JS *et al*.: **Dnmt3a is essential for hematopoietic stem cell differentiation.** *Nat Genet* 2012, **44:** 23-31.

24. Beck S: **Taking the measure of the methylome.** *Nat Biotechnol* 2010, **28:** 1026-1028.

25. Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H *et al*.: **Quantitative comparison of genome-wide DNA methylation mapping technologies.** *Nat Biotechnol* 2010, **28:** 1106-1114.

26. Butcher LM, Beck S: **AutoMeDIP-seq: A high-throughput, whole genome, DNA methylation assay.** *Methods* 2010.

27. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E *et al*.: **A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis.** *Nat Biotechnol* 2008, **26:** 779-785.

28. Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL *et al*.: **Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications.** *Nat Biotechnol* 2010, **28:** 1097-1105.

29. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL *et al*.: **Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells.** *Nat Genet* 2005, **37:** 853-862.

30. Wilson GA, Dhami P, Feber A, CortÃ¡zar D, Suzuki Y, Schulz R *et al*.: **Resources for methylome analysis suitable for gene knockout studies of potential epigenome modifiers.** *GigaScience* 2012, **1**.

31. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D *et al*.: **Methylome analysis using MeDIP-seq with low DNA concentrations.** *Nat Protoc* 2012, **7:** 617-636.

32. Goodell MA, Rosenzweig M, Kim H, Marks DF, DeMaria M, Paradis G *et al*.: **Dye efflux studies suggest that hematopoietic stem cells expressing low or undetectable levels of CD34 antigen exist in multiple species.** *Nat Med* 1997, **3:** 1337-1345.

33. Pearce DJ, Ridler CM, Simpson C, Bonnet D: **Multiparameter analysis of murine bone marrow side population cells.** *Blood* 2004, **103:** 2541-2546.

34. Spiess AN, Walther N, Muller N, Balvers M, Hansis C, Ivell R: **SPEER--a new family of testis-specific genes from the mouse.** *Biol Reprod* 2003, **68:** 2044-2054.

35. te Velthuis AJ, Admiraal JF, Bagowski CP: **Molecular evolution of the MAGUK family in metazoan genomes.** *BMC Evol Biol* 2007, **7:** 129.

36. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS *et al*.: **A novel signaling pathway impact analysis.** *Bioinformatics* 2009, **25:** 75-82.

37. Arai F, Suda T: **Maintenance of quiescent hematopoietic stem cells in the osteoblastic niche.** *Ann N Y Acad Sci* 2007, **1106:** 41-53.

38. Hattori K, Heissig B, Rafii S: **The regulation of hematopoietic stem cell and progenitor mobilization by chemokine SDF-1.** *Leuk Lymphoma* 2003, **44:** 575-582.

39. Haylock DN, Nilsson SK: **Stem cell regulation by the hematopoietic stem cell niche.** *Cell Cycle* 2005, **4:** 1353-1355.

40. Tedder TF, Steeber DA, Chen A, Engel P: **The selectins: vascular adhesion molecules.** *FASEB J* 1995, **9:** 866-873.

41. Smith E, Sigvardsson M: **The roles of transcription factors in B lymphocyte commitment, development, and transformation.** *J Leukoc Biol* 2004, **75:** 973-981.

42. Jeanes AI, Wang P, Moreno-Layseca P, Paul N, Cheung J, Tsang R *et al*.: **Specific beta-containing integrins exert differential control on proliferation and two-dimensional collective cell migration in mammary epithelial cells.** *J Biol Chem* 2012, **287:** 24103-24112.

43. Lo PK, Kanojia D, Liu X, Singh UP, Berger FG, Wang Q *et al*.: **CD49f and CD61 identify Her2/neu-induced mammary tumor-initiating cells that are potentially derived from luminal progenitors and maintained by the integrin-TGFbeta signaling.** *Oncogene* 2012, **31:** 2614-2626.

44. Gustincich S, Schneider C: **Serum deprivation response gene is induced by serum starvation but not by contact inhibition.** *Cell Growth Differ* 1993, **4:** 753-760.

45. Hansen CG, Bright NA, Howard G, Nichols BJ: **SDPR induces membrane curvature and functions in the formation of caveolae.** *Nat Cell Biol* 2009, **11:** 807-814.

46. Dong H, Osmanova V, Epstein PM, Brocke S: **Phosphodiesterase 8 (PDE8) regulates chemotaxis of activated lymphocytes.** *Biochem Biophys Res Commun* 2006, **345:** 713-719.

47. Hayashi H, Sano H, Seo S, Kume T: **The Foxc2 transcription factor regulates angiogenesis via induction of integrin beta3 expression.** *J Biol Chem* 2008, **283:** 23791-23800.

48. Martinez-Lopez MJ, Alcantara S, Mascaro C, Perez-Branguli F, Ruiz-Lozano P, Maes T *et al*.: **Mouse neuron navigator 1, a novel microtubule-associated protein involved in neuronal migration.** *Mol Cell Neurosci* 2005, **28:** 599-612.

49. Muley PD, McNeill EM, Marzinke MA, Knobel KM, Barr MM, Clagett-Dame M: **The atRA-responsive gene neuron navigator 2 functions in neurite outgrowth and axonal elongation.** *Dev Neurobiol* 2008, **68:** 1441-1453.

50. Stafford LJ, Xia C, Ma W, Cai Y, Liu M: **Identification and characterization of mouse metastasis-suppressor KiSS1 and its G-protein-coupled receptor.** *Cancer Res* 2002, **62:** 5399-5404.

51. Wang Y, Li Z, Xu P, Huang L, Tong J, Huang H *et al*.: **Angiomotin-like2 gene (amotl2) is required for migration and proliferation of endothelial cells during angiogenesis.** *J Biol Chem* 2011, **286:** 41095-41104.

52. Pearce DJ, Anjos-Afonso F, Ridler CM, Eddaoudi A, Bonnet D: **Age-dependent increase in side population distribution within hematopoiesis: implications for our understanding of the mechanism of aging.** *Stem Cells* 2007, **25:** 828-835.

53. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML *et al*.: **Epigenetic differences arise during the lifetime of monozygotic twins.** *Proc Natl Acad Sci U S A* 2005, **102:** 10604-10609.

54. Gentilini D, Mari D, Castaldi D, Remondini D, Ogliari G, Ostan R *et al*.: **Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring.** *Age (Dordr)* 2012.

55. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A *et al*.: **Distinct DNA methylomes of newborns and centenarians.** *Proc Natl Acad Sci U S A* 2012, **109:** 10522-10527.

56. Garrett RW, Gasiewicz TA: **The aryl hydrocarbon receptor agonist 2,3,7,8-tetrachlorodibenzo-p-dioxin alters the circadian rhythms, quiescence, and expression of clock genes in murine hematopoietic stem and progenitor cells.** *Mol Pharmacol* 2006, **69:** 2076-2083.

57. Casado FL, Singh KP, Gasiewicz TA: **Aryl hydrocarbon receptor activation in hematopoietic stem/progenitor cells alters cell function and pathway-specific gene modulation reflecting changes in cellular trafficking and migration.** *Mol Pharmacol* 2011, **80:** 673-682.

58. Singh KP, Wyman A, Casado FL, Garrett RW, Gasiewicz TA: **Treatment of mice with the Ah receptor agonist and human carcinogen dioxin results in altered numbers and function of hematopoietic stem cells.** *Carcinogenesis* 2009, **30:** 11-19.

59. Weber GF, Ashkar S, Cantor H: **Interaction between CD44 and osteopontin as a potential basis for metastasis formation.** *Proc Assoc Am Physicians* 1997, **109:** 1-9.

60. Fujisaki T, Tanaka Y, Fujii K, Mine S, Saito K, Yamada S *et al*.: **CD44 stimulation induces integrin-mediated adhesion of colon cancer cell lines to endothelial cells by up-regulation of integrins and c-Met and activation of integrins.** *Cancer Res* 1999, **59:** 4427-4434.

61. Wang HS, Hung Y, Su CH, Peng ST, Guo YJ, Lai MC *et al*.: **CD44 cross-linking induces integrin-mediated adhesion and transendothelial migration in breast cancer cell line by up-regulation of LFA-1 (alpha L beta2) and VLA-4 (alpha4beta1).** *Exp Cell Res* 2005, **304:** 116-126.

62. Liang Y, Van ZG, Szilvassy SJ: **Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells.** *Blood* 2005, **106:** 1479-1487.

63. Dykstra B, Olthof S, Schreuder J, Ritsema M, de HG: **Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells.** *J Exp Med* 2011, **208:** 2691-2703.

64. Kohler A, Schmithorst V, Filippi MD, Ryan MA, Daria D, Gunzer M *et al*.: **Altered cellular dynamics and endosteal location of aged early hematopoietic progenitor cells revealed by time-lapse intravital imaging in long bones.** *Blood* 2009, **114:** 290-298.

65. Zhang CC, Kaba M, Ge G, Xie K, Tong W, Hug C *et al*.: **Angiopoietin-like proteins stimulate ex vivo expansion of hematopoietic stem cells.** *Nat Med* 2006, **12:** 240-245.

66. Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, Lehrach H *et al*.: **Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage.** *Genome Res* 2010, **20:** 1441-1450.

67. Li H, Durbin R: **Fast and accurate short read alignment with BurrowsΓÇôWheeler transform.** *Bioinformatics* 2009, **25:** 1754-1760.

68. Nix DA, Courdy SJ, Boucher KM: **Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks.** *BMC Bioinformatics* 2008, **9:** 523.

69. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11:** R106.

70. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al*.: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19:** 1639-1645.

71. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26:** 841-842.

72. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics* 2011, **12:** 323.

73. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10:** R25.

74. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB *et al*.: **GREAT improves functional interpretation of cis-regulatory regions.** *Nat Biotechnol* 2010, **28:** 495-501.

75. Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N *et al*.: **PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification.** *Nucleic Acids Res* 2003, **31:** 334-341.