Epigenetic drift during long-term culture of cells *in vitro*

Julia Franzen[1], Theodoros Georgomanolis[2], Anton Selich[3], Chao-Chung Kuo[1], Reinhard Stöger[4], Lilija Brant[2], Melita Sara Mulabdić[1], Eduardo Fernandez-Rebollo[1], Clara Grezella[1], Alina Ostrowska[1], Matthias Begemann[5], Miloš Nikolić[1], Björn Rath[6], Anthony D. Ho[7], Michael Rothe[3], Axel Schambach[3], Argyris Papantonis[2,8], Wolfgang Wagner[1]

[1] Helmholtz-Institute for Biomedical Engineering, RWTH Aachen University Medical School, 52074 Aachen, Germany

[2] Center for Molecular Medicine, University of Cologne, 50931 Cologne, Germany

[3] Institute of Experimental Hematology, Hannover Medical School, 30625 Hannover, Germany

[4] School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5RD, United Kingdom

[5] Institute of Human Genetics, Medical Faculty, RWTH Aachen University, 52074, Aachen, Germany

[6] Department for Orthopedics, RWTH Aachen University Medical School, 52074 Aachen, Germany

[7] Internal Medicine Department of Hematology, Oncology and Rheumatology, Heidelberg University Medical Center, 69120 Heidelberg, Germany

[8] Institute of Pathology, University Medical Center Göttingen, 37075 Göttingen, Germany

Corresponding author: Wolfgang Wagner, E-mail: wwagner@ukaachen.de

Abstract

Culture expansion of primary cells evokes highly reproducible DNA methylation (DNAm) changes at specific sites in the genome. These changes might be due to an directly regulated epigenetic process, or to gradual deregulation of the epigenetic state, which is often referred to as “epigenetic drift”. We have identified CG dinucleotides (CpGs) that become continuously hyper- or hypomethylated in the course of culture expansion of mesenchymal stem cells (MSCs) and other cell types. During reprogramming into induced pluripotent stem cells (iPSCs) particularly the culture-associated hypomethylation is reversed simultaneously with age-associated and pluripotency-associated DNAm changes. Bisulfite barcoded amplicon sequencing (BBA-seq) demonstrated that upon passaging the DNAm patterns of neighboring CpGs become more complex without evidence of continuous pattern development and without association to oligoclonal subpolulations of MSCs at later passages. Circularized chromatin conformation capture (4C) revealed reproducible changes in nuclear organization between early and late passages, while there was no preferential interaction with other genomic regions that also harbor culture-associated DNAm changes. Chromatin immunoprecipitation of CTCF did not show significant differences during long-term culture of MSCs, however culture-associated hypermethylation was enriched at CTCF binding sites and hypomethylated CpGs were devoid of CTCF. Taken together, our results indicate that DNAm changes during culture-expansion resembles epigenetic drift, which seems to occur in relation to chromatin conformation.

Keywords: Cellular Aging, CTCF, DNA methylation, epigenetic, hairpin, hemimethylation, mesenchymal stem cells, long-term culture, 5mC, 4C

Introduction

Cell preparations are often expanded *in vitro* for many passages to achieve enough material for basic research or cellular therapy. However, culture expansion has severe and continuous impact on the growth, morphology, gene expression, metabolomics, and function of primary cells, until they ultimately enter a state of replicative senescence (1, 2). This is of particular relevance for the more than 1,000 registered clinical trials with mesenchymal stem/stromal cells (MSCs; www.clinicaltrials.gov), which usually use high cell doses (often about 10^6 cells/kg body weight) and therefore necessitate excessive culture expansion (3, 4). Furthermore, long-term culture of cells is often required for basic research and this may hamper reproducibility of results. Tracking the impact of long-term culture is therefore an important aspect for quality control of cell preparations.

Long-term culture is reflected by highly reproducible DNA methylation (DNAm) changes at specific sites in the genome (5, 6). We have previously demonstrated that DNAm levels at only six cytosine/guanine dinucleotides (CpG sites) can be used to estimate passage numbers and cumulative population doublings (7, 8). It therefore provides a reliable biomarker to estimate the state of culture-associated modifications. In our previous studies, we referred to these epigenetic modifications as “senescence-associated DNAm changes”, albeit it is unclear if the culture-associated DNAm changes are linked to the state of cellular senescence – we therefore changed the terminology into “culture-associated DNAm changes”. Furthermore, it is so far unclear how DNAm patterns evolve during culture expansion and why they occur at specific genomic regions.

On first sight, the culture-associated DNAm changes seem to be related to the DNAm changes that are acquired during aging of the organism (9). In fact, there is some overlap with epigenetic clocks for aging, but the two processes can be clearly discerned and it is

yet unclear how the DNAm changes are governed. It is generally anticipated that DNAm changes during development are regulated by epigenetic writers - such as the *de novo* methyltransferases DNMT3A and DNMT3B, or TET methylcytosine dioxygenases (10). If an epigenetic writer is targeted to a specific site in the genome the neighboring CpGs will most likely also be modified. Alternatively, the DNAm changes that accumulate during culture expansion might not be directly regulated but rather reflect dysregulation, as suggested for “epigenetic drift” during aging of the organism (11-13). A better understanding of how DNAm changes evolve in culture expansion might shed light into the underlying process.

Furthermore, it is generally thought that DNAm patterns are identical on the forward and the reverse DNA strand (Watson and Crick strands) (14). Using hairpin-bisulfite PCR (16) several studies have demonstrated that, despite the general preference for concordant DNA methylation on both strands, certain sites are specifically methylated on only one strand (17-19). Such hemimethylation can be stably inherited over several passages and has been associated with CCCTC-binding factor (CTCF)/cohesin binding sites (20). CTCF has been linked with changes in chromatin conformation and cellular senescence (22, 23). We have demonstrated that senescence entry upon extensive culture expansion is associated with a reorganization of CTCF into large senescence-induced CTCF clusters (SICCs) (22). However, it remains unclear if and how the binding of CTCF changes during long-term culture.

In this study, we further investigated if culture-associated DNAm changes are caused by a targeted regulatory mechanism or rather by epigenetic drift.

Results

DNA methylation changes to track the process of culture expansion

We have previously identified DNAm changes during culture expansion of mesenchymal stromal cells (MSCs) based on 27k Illumina BeadChip datasets, and thereby established a 6 CpG predictor to estimate passage numbers specifically for MSCs (7). In continuation of this work, we utilized the meanwhile available 450k Illumina BeadChip datasets, which interrogate approximately sixteen times more CpGs than the 27k version of the chip, and we also considered datasets of additional types of primary cells. We compiled 63 DNAm profiles of human primary MSCs (n = 45), fibroblasts (n = 5), and human umbilical vein endothelial cells (n = 13) with precise information on passage numbers (Supplemental Table S1). To identify individual candidate CpGs that become continuously hyper- or hypomethylated during culture expansion of these primary cell types, we filtered CpGs by Pearson correlation with passage number with an initial cutoff of $R > 0.7$ or $R < -0.7$: 646 and 2,442 CpGs passed these criteria, respectively (Supplemental Table S2). To further refine the list of candidates, we used more stringent filter criteria ($R > 0.8$ and $R < -0.8$; and a linear regression slope $m > 0.02$) to select 15 hypermethylated and 15 hypomethylated sites (Supplemental Table S3).

To develop a robust biomarker to estimate passage number based on targeted DNAm analysis with pyrosequencing, we then focused on two hyper- and two hypomethylated CpGs that cooperated best for prediction of passage numbers in the microarray training dataset. The four chosen CpGs were related to the genes *Arachidonate 12-Lipoxygenase* (*ALOX12,* cg03762994), *Docking Protein 6* (*DOK6,* cg25968937), *Leukotriene C4 Synthase* (*LTC4S,* cg26683398) and *TNNI3 Interacting Kinase* (*TNNI3K,* cg05264232; Fig 1A). The long-term culture associated DNAm changes at these CpGs were then tested and validated by pyrosequencing in MSCs, fibroblasts, and HUVECs at various passages

(n = 44). Samples of the training set were all cultured until growth arrest and senescence was checked by staining with senescence-associated β-galactosidase (Supplemental Fig. S1A, B). Based on pyrosequencing results, we generated a multivariable model for epigenetic predictions of passage number (R^2 = 0.81; Fig. 1B and Supplemental Fig. S1C). Ten times 10-fold cross-validation of the pyrosequencing training dataset resulted in a R^2 = 0.84 and a root mean squared error (RMSE) of 3.9 passages. Subsequently, our epigenetic predictor for long-term culture was validated on an additional independent set of samples (n = 83; R^2 = 0.74; Fig. 1C and Supplemental Fig. S1D) by pyrosequencing. Thus, DNAm analysis at these four CpGs facilitates relative precise estimation of passage numbers and can be utilized for different cell types.

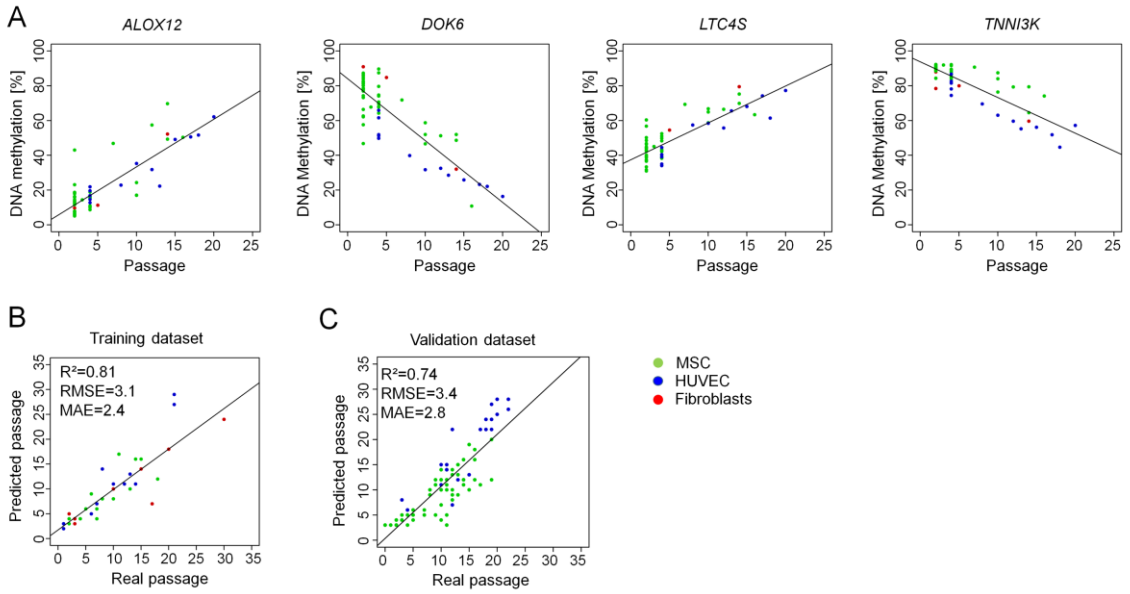

Figure 1. DNA methylation changes during long-term culture.
A) Four CpGs were selected that revealed continuous changes during culture expansion in DNAm profiles of MSCs, fibroblasts, and HUVECs (all 450k BeadChip data). **B)** DNAm at these four CpGs was then analyzed in a training set (n = 44) by pyrosequencing and the results were used to train a multivariable model to estimate passage numbers (RMSE = root mean squared error; MAE = mean average error). **C)** This predictor was validated on pyrosequencing results of an independent set of samples (n = 83).

Hypomethylation during long-term culture is reversed as an early event during cell reprogramming into induced pluripotent stem cells

Culture-associated DNAm changes are reversed in fully reprogrammed induced pluripotent stem cells (iPSCs) (24, 25), but the kinetics of this epigenetic rejuvenation have not yet been addressed and it was unclear if this occurs simultaneously with re-setting of age-associated DNAm (26, 27). We therefore utilized publicly available DNAm profiles of TRA-1-60 positive cells at various time points after retroviral reprogramming of fibroblasts with *OCT3/4*, *SOX2*, *KLF4* and *c-MYC* (28). Particularly CpGs that are hypomethylated at later passages become re-methylated between day 15 and day 20 after reprogramming, whereas hypermethylated sites are not consistently demethylated and sometimes gain methylation during reprogramming (Fig. 2A and B, Supplemental Fig. S2A). Notably, these epigenetic changes occur in parallel to the epigenetic modifications at pluripotency-associated CpGs (Supplemental Fig. S2B) (29). We utilized the four new and five former CpG sites of our epigenetic signatures, which were represented by the 450k Illumina Microarray (Supplemental Fig. S2A) for predictions of passage numbers. While these estimates increased during culture expansion of MSCs, they declined around day 20 after reprogramming (Fig. 2C).

Importantly, culture-associated DNAm changes are distinct from age-associated DNAm changes, which correlate with chronological age rather than with passage numbers (9). Age-related epigenetic signatures (27) are overall reset at the same time course (Supplemental Fig. S2C), a finding which is in line with another recent study (30). In fact, the DNAm changes related to pluripotency, culture expansion and ageing follow the same kinetics (Fig. 2D and E) and the Pearson correlation of these DNA methylation changes is highly significant ($p < 2.2*10^{-16}$).

During re-differentiation of iPSCs towards MSCs (iMSCs) there is an inverse switch in epigenetic patterns of pluripotency and culture expansion around day 7. While the long-term culture associated DNAm changes are then continuously acquired upon differentiation of iPSCs, this is not observed for age-associated signatures (25) (Fig. 2A, B). Conversely, estimation of passage numbers gradually increased upon differentiation of MSCs towards iMSCs (25) (Fig. 2C). Taken together, epigenetic rejuvenation occurs simultaneously at aging and culture-associated CpGs. In contrast, to aging-associated DNAm changes (25, 31) the culture-associated epigenetic modifications are then gradually reacquired over multiple passages of iMSCs.

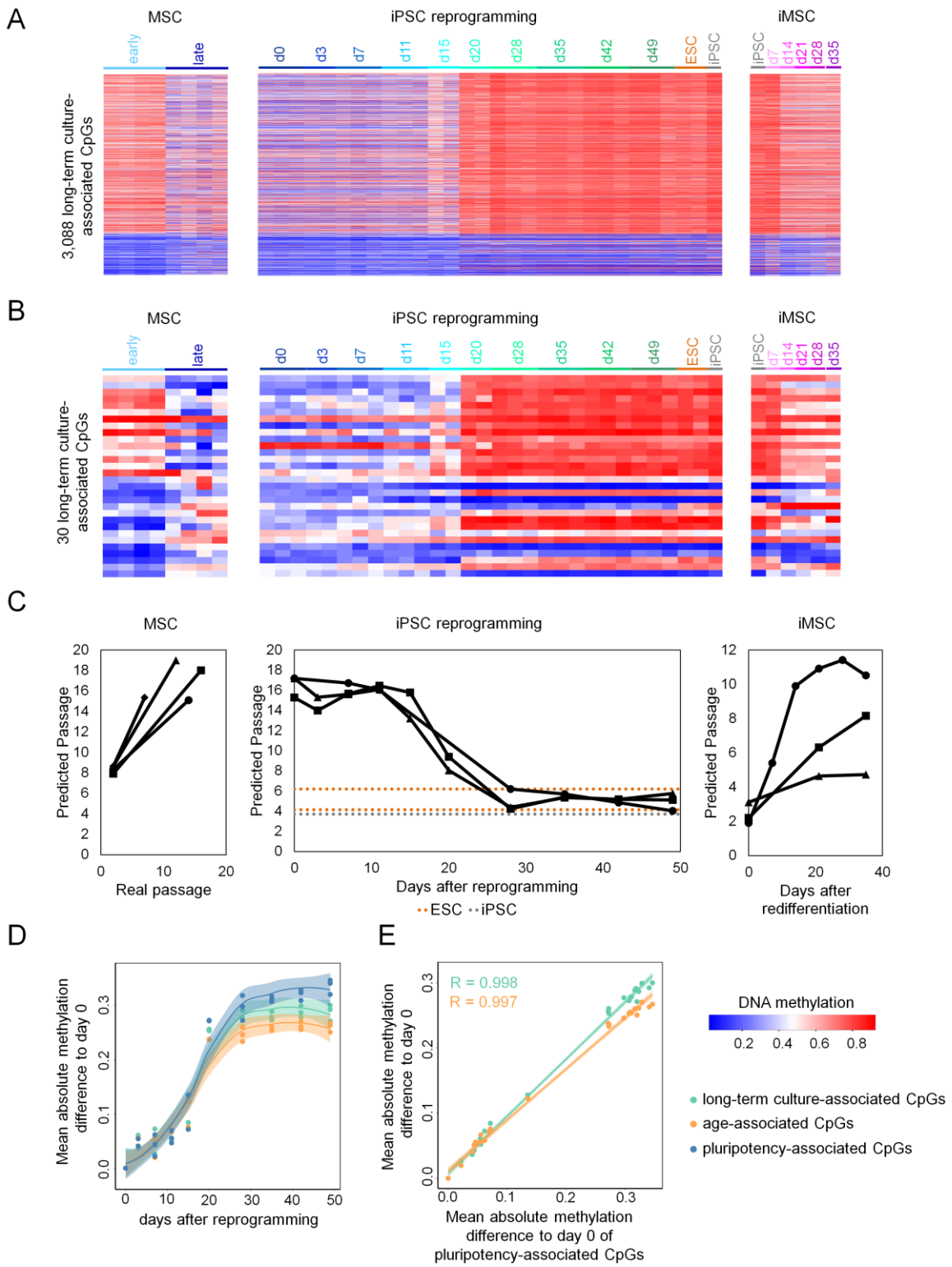

Figure 2. DNA methylation kinetics during reprogramming into iPSCs and re-differentiation to iMSCs. A and B) DNAm changes were analyzed in 3,088 (A) or 30 (B) culture-expansion-associated CpG sites using DNAm profiles of MSCs of early (P2) and late (P7 to P16) passage (GSE37067 (24); left), in a dataset that analyzed DNAm changes at various time points during reprogramming of fibroblasts (GSE54848 (28); central), and

during re-differentiation to iPSC-derived MSC (iMSC; GSE54767 (25); right). **C)** Passage predictions of the three datasets described in A) and B). Passage predictions were calculated by using the mean of the predicted passages for the four new and five former CpG sites of our epigenetic signatures shown in Supplemental Fig. 2A. **D and E)** To further quantify the coincidence of pluripotency- with culture- or age-associated DNAm changes, we calculated the mean absolute methylation difference of all donors at each day of reprogramming to day zero. Culture-associated (3,088 CpGs), age-associated (99 CpGs), and pluripotency-associated CpGs (1,432 CpGs) follow the same trend in methylation changes (D) and the differences of long-term culture- and age-associated sites are highly correlated to those of pluripotency-associated sites (E, Pearson correlation R = 0.999 and 0.997, respectively).

DNA methylation patterns do not reflect MSC clonality

MSC preparations are heterogeneous and there is evidence that individual subclones become dominant at later passages (33), which might contribute to culture-associated DNAm changes. Therefore, we aimed for a better understanding how DNAm patterns at neighboring CpGs evolve during culture-expansion and how this is affected by the clonal composition of MSCs. We anticipated that tracking of DNAm patterns over several passages would provide insights into the changing composition of subclones within MSC preparations. To address this question, we used samples from a previously published study (33): Umbilical cord derived MSCs from two donors were transduced with lentiviral vectors containing random barcodes and three different fluorescent proteins. Flow cytometry and deep sequencing demonstrated that the diversity of cellular subsets declines and that senescent passages became oligoclonal (Fig. 3A and Supplemental Fig. S3A, B). We used barcoded bisulfite amplicon sequencing (BBA-Seq) to investigated DNAm patterns at the four culture-associated CpGs identified above, as well as the six CpGs of our previous predictor for replicative senescence (associated with the genes *CASR*, *CASP14*, *GRM7*, *KRTAP13.3*, *PRAMEF2* and *SELP*) (7). The combined BBA-seq measurements were then used to predict passage numbers. Taking all passages into account, the epigenetic estimations correlated well with the number of passages (R^2 = 0.87, Fig. 3B). In contrast to pyrosequencing, BBA-Seq facilitates analysis of the

succession of methylated and non-methylated CpGs within individual reads. To this end, we focused on those amplicons that comprised several neighboring CpGs on the BBA-Seq reads (*GRM7*, *CASR*, *LTC4S*, *DOK6* and *ALOX12)*. DNAm patterns became more diverse during culture expansion, which is also reflected by the increased Shannon index (Fig. 3C and Supplemental Fig S3B - D). Thus, the development of DNAm patterns during long-term culture is independent of the oligoclonal composition of MSCs at later passages.

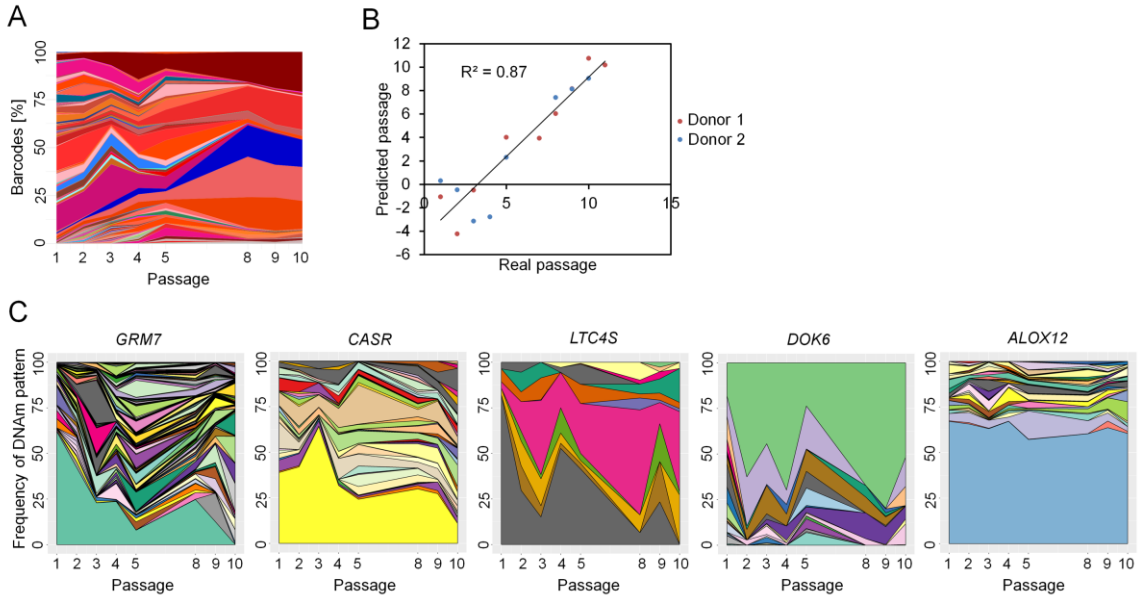

Figure 3. DNA methylation patterns are not affected by MSC clonality.
A) Mesenchymal stem cells from umbilical cord were labelled with barcoded-RGB-vectors to correlate DNAm patterns with the composition of cellular subsets (33). Deep sequencing analysis of the random barcodes demonstrates that the MSCs at passage 10 are oligoclonal. **B)** These samples were subsequently used for DNAm analysis by barcoded bisulfite amplicon sequencing (BBA-Seq) at ten culture-associated CpGs. The predicted passage numbers based on these DNAm levels correlated with real passage numbers. **C)** Changes in the frequency of different DNAm patterns over the passages are depicted for the neighboring CpGs within the amplicons of *GRM7* (22 CpGs), *CASR* (7 CpGs), *LTC4S* (4 CpGs), *DOK6* (7 CpGs), and *ALOX12* (9 CpGs). Each colored area corresponds to the frequency of one specific DNAm pattern, while the sum of all pattern frequencies adds up to 100%.

Culture-associated DNA methylation patterns evolve stochastically at neighboring CpGs

To gain further insight if culture-associated DNAm is regulated by targeted DNAm or rather by indirect epigenetic drift, we analyzed if DNAm is coherently modified at neighboring CpGs. If an epigenetic writer is targeted to a specific site in the genome the neighboring CpGs will most likely be coherently modified. To further investigate the dynamics of DNAm patterns we focused on the amplicon of *GRM7*, which comprised the highest number of CpG sites. The DNAm patterns fluctuated over subsequent passages (Supplemental Fig. S3D) and there was no evidence for continuous development of culture-associated modifications at this genomic region. In fact, the modifications seemed to be acquired randomly and there was hardly any correlation in DNAm between neighboring CpGs at single read level (maximal Pearson correlation R = 0.34, Fig. 4A). On the other hand, several neighboring CpGs within the amplicon show also very high correlation with passage numbers (Fig. 4B).

If the methylation changes at neighboring CpGs are acquired rather independently, then it should be possible to estimate probabilities for passage numbers also for the individual BBA-seq reads, based on the binary sequel of methylated and non-methylated CpGs (34). To this end, we utilized the correlation with passage number at individual CpGs in the training set to establish a predictor based on single sequencing reads, as described in our previous work (34). The algorithm estimates the likelihood for each pattern/read to belong to any passage between 0 and 50. By this approach we clearly detect a high heterogeneity of sequenced DNAm patterns in samples of early and late passages. In tendency, samples at later passages comprised more BBA-seq reads that were predicted to correspond to higher passage numbers (Fig. 4C). To validate our findings we used BBA-Seq data of the *GRM7* amplicon from our previous study (8). Notably, the mean of single read predictions

showed clear correlations between real and predicted passage numbers for training and validation set (R^2 = 0.88 and 0.72, respectively; Fig. 4D).

To address the question if DNAm patterns are identical on both complementary DNA strands we ligated hairpin oligonucleotides to connect the forward and reverse strands of individual DNA molecules (Fig. 4E) (16). These hairpins also comprised a unique molecular identifier (UMI) in the loop region to adjust for potential PCR bias (Supplemental Fig. S4A). Eight out of the ten culture-expansion-associated regions encompassed suitable endonuclease restriction sites for targeted hairpin-ligation and could be further analyzed by BBA-Seq with primers specific for these hairpins (*CASR*, *GRM7*, *KRTAP13.3*, *PRAMEF2, SELP*, *DOK6*, *LTC4S*, *TNNI3K*). As a control we considered an additional genomic region that was generally methylated (associated with the genes *C12orf12*). The accuracy of epigenetic predictions of passage numbers was similar when reads with the same UMI were only considered once (Supplemental Fig. S4B), indicating that potential PCR bias during amplification does not have major impact on the mean DNAm levels. Subsequently we compared the DNAm patterns of the two complementary DNA strands. While we observed similar stochastic DNAm patterns as with conventional BBA-Seq, these patterns were overall faithfully shared between both DNA strands (Fig. 4F). Distinct CpGs of the long-term culture associated sites exhibited slightly higher frequencies of hemimethylation than the frequencies observed at the control site (e.g. *DOK6*, *SELP*, *PRAMEF2*, *KRTAP13.3*, Supplemental Fig. S4C).

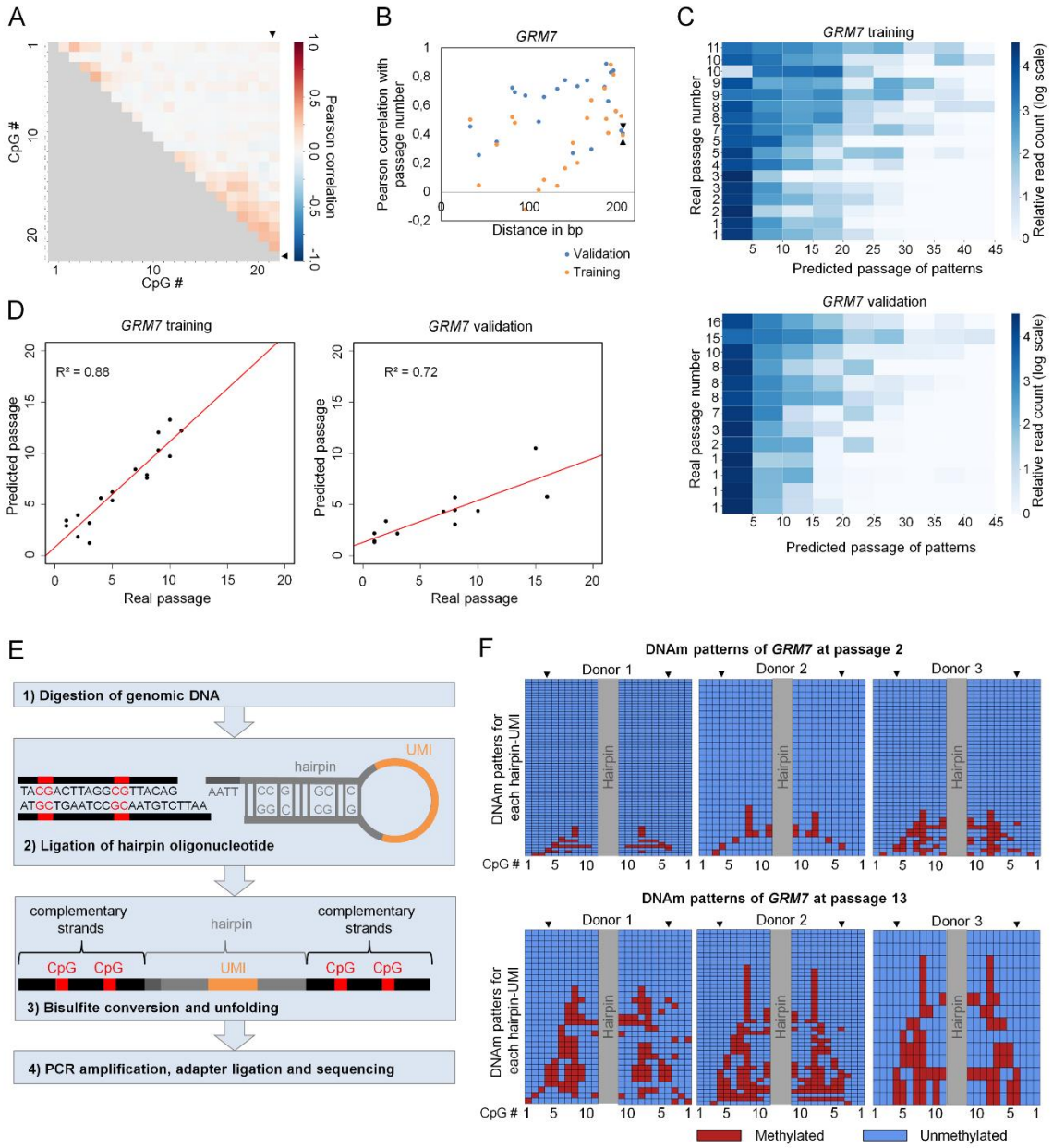

Figure 4. DNA methylation of neighboring CpG sites in the amplicon of *GRM7*.
A) Pearson correlation of DNAm levels at neighbouring CpG sites of the BBA-Seq amplicon of *GRM7*. The CpG site included in the Epigenetic-Culture-Expansion-Signature is CpG #22, depicted by black arrow heads. **B)** Pearson correlation with passage number of neighbouring CpGs in *GRM7*. The CpG site included in the Epigenetic-Culture-Expansion-Signature is the CpG at 207 bp, depicted by black arrow heads. Correlations are depicted for a training set (the two RGB labeled donors depicted in Fig. 3A and Supplemental Fig 3A) and a validation set of our previous study (8). **C)** Heatmaps of single read predictions for the amplicon of *GRM7* show a high heterogeneity of predictions within each of the samples of the training (upper panel) and validation set (lower panel). **D)** The mean of the single read predictions correlated with real passage numbers. **E)** Schematic

presentation of DNAm analysis on complementary DNA strands using hairpin ligation and BBA-Seq (UMI = unique molecular identifier). **F)** DNAm patterns of eleven neighboring CpGs in *GRM7* are depicted on complementary strands (MSC of three different donors in early and late passages). The CpG site included in the Epigenetic-Culture-Expansion-Signature is CpG #4, depicted by black arrow heads on both strands. Each row represents patterns for an UMI in the hairpin.

Chromatin interactions at genomic regions with culture-associated DNAm changes

Subsequently, we analyzed if genomic regions with gains or losses of DNAm might preferentially interact between different chromatin loops, which might be indicative for a co-regulation. To this end, we investigated chromatin interactions of four genomic regions with culture-associated DNAm changes (*ALOX12*, *LTC4S*, *CASR* and *KRTAP13.3*) using circular chromatin conformation capture (4C). Two independent MSC preparations at early (P2 and P3) and late (P7 and P9) passages revealed overall reproducible interaction profiles (Fig. 5A). For downstream analyses we only considered highly interacting regions that were categorized as “nearbait” (10 MB around the “bait” locus of interest) and as “cis” (all cis-contacts on the same chromosome), respectively. Although trans-chromosomal analysis showed a high number of reproducible interactions (Fig. 5B) we excluded these sites from further analysis due to their high background signal, as commonly observed in such studies. The number of interacting sites in nearbait and cis remained similar with only a moderate increase between early and late passages (Fig. 5C and D). However, many of the interacting sites showed reproducible and significant differences between early and late passages (p value < 0.05; Fig. 5E and F).

To gain insight into how chromatin interactions are reflected on DNAm we used datasets of MSCs at early passage (n = 5; P2 - P3) and pre-senescent passage (n = 5; P7 – P13; GSE37067) (24). We investigated the mean DNAm levels within the interacting regions of our four different culture-expansion associated sites. Overall the mean DNAm levels within these interaction sites were similar to the average DNAm of the corresponding

chromosome, indicating that there was no preferential binding to either methylated or non-methylated regions. However, nearbait interactions of *KRTAP13.3*, which becomes hypomethylated upon long-term culture, preferably formed with non-methylated regions both in early and late passages (Fig. 5G). Notably, none of the four culture-associated genomic regions revealed significant enrichment of interaction with other culture-associated CpGs (Fig. 5H). Furthermore, gains and losses in chromatin interactions were not enriched in regions with culture-associated DNAm changes and thus the changes of DNA methylation do not seem to affect changes in chromatin interactions (Fig. 5I and J).

We and others previously demonstrated that hypomethylation during long-term culture occurs preferentially at lamina-associated domains (LADs) (6, 36, 37). This is supported by the observation that only the hypomethylated region of *KRTAP13.3* was enriched in interactions with LADs, whereas such interactions were depleted for hypermethylated regions (*ALOX12*, *CASR* and *LTC4S*; Fig. 4K). Taken together, culture-associated DNAm occurs at genomic regions that reveal significant and reproducible changes in chromatin conformation, but do not show enriched interaction with other genomic regions with culture-associated DNAm changes. It is therefore rather unlikely that the culture-associated CpGs are synchronously regulated at the interaction sites of different chromatin loops.

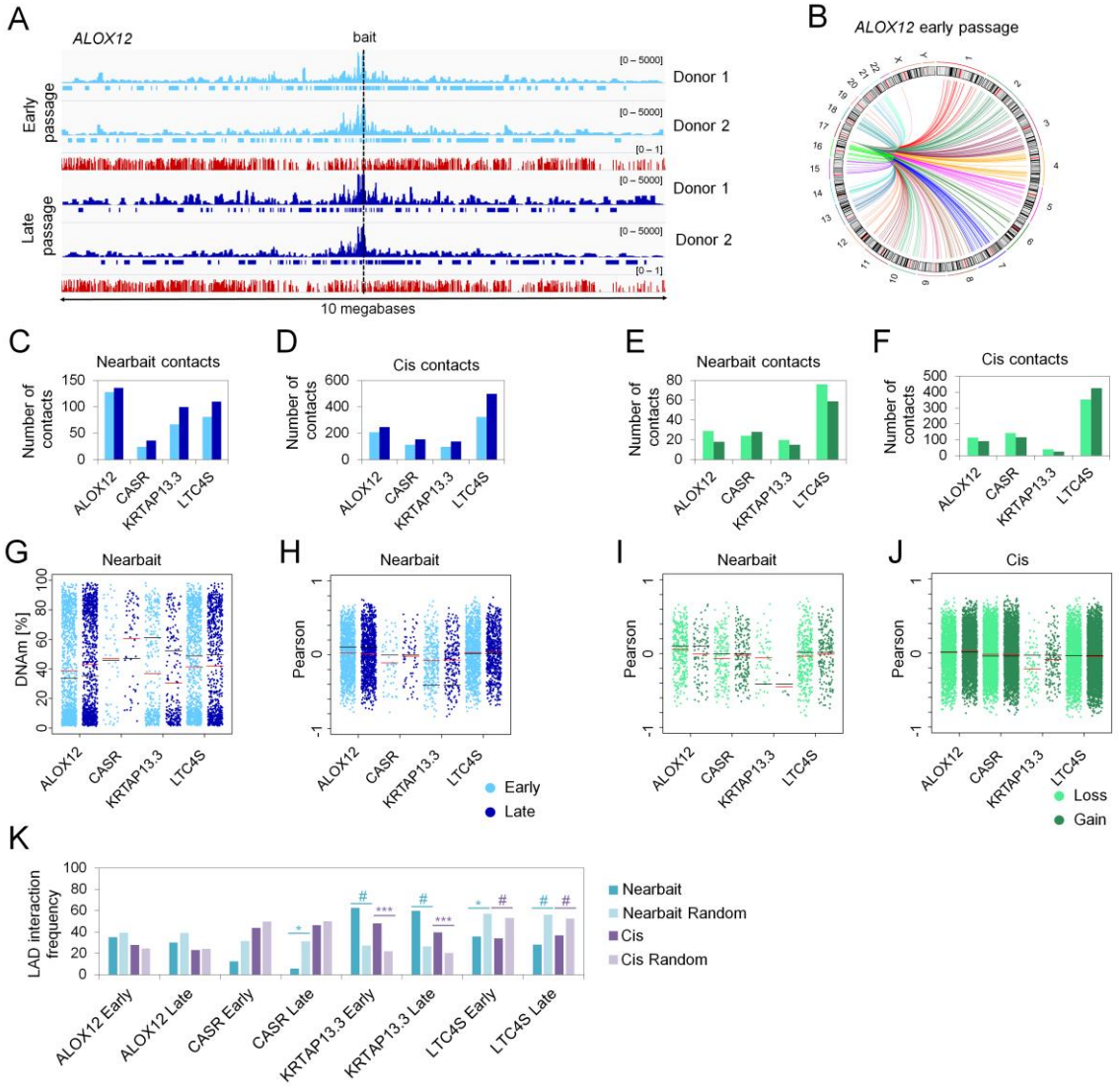

Figure 5. Circular chromatin conformation capture (4C) of culture-associated CpGs. A) Integrative Genomics Viewer (IGV) overview of the bait region of *ALOX12* in MSCs of two donors at early (P2 and P3, light blue) and late (P7 and P9, dark blue) passage. Sequencing peaks are presented as normalized counts. High interacting regions called by 4Cker tool are indicated by horizontal bars beneath the peaks. Mean methylation levels of CpGs on the Illumina 450k BeadChip are depicted for MSCs at early and late passages (GSE37067; red) (24). **B)** Circos-plot of highly interacting regions across different chromosomes (trans) is exemplarily depicted for *ALOX12* (interactions are reproducible in two MSC preparations at early passage). **C)** Numbers of highly interacting regions within the nearbait regions of the four culture-expansion-associated CpGs were similar at early and late passages. **D)** Numbers of cis-interacting regions (within the same chromosome). **E**, **F)** Number of differential contacts (p-value < 0.05) between early and late passages of all four CpG sites in nearbait and cis locations. **G)** DNAm levels of all CpG sites within high interacting regions in early and late passages. Red lines indicate mean DNAm levels of all CpG sites within interacting regions, black lines depict mean DNAm of all CpG sites on the 450k BeadChip within the nearbait region. **H)** The association of DNAm with

passage numbers (Pearson correlation) is presented for all CpG sites within high interacting regions. Mean correlations are depicted for CpG sites at highly interacting regions (red) and all CpGs within the nearbait regions (black). **I, J)** Association of DNAm with passage numbers (Pearson correlation) is presented for all CpG sites at differential contact regions within nearbait and cis (red and black lines in analogy to H). **K)** Enrichment of lamina-associated domains (LADs) within high interacting sites of the four baits compared to random background regions. Significance was tested by Fisher's exact test (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; # $p < 0.0001$).

DNA methylation changes during long-term culture are related to CTCF binding sites

During culture expansion and upon entry into replicative senescence the cell nuclei become much larger and CTCF was shown to reorganize into large senescence-induced CTCF clusters (SICCs) in HUVEC and IMR90 cells (22). When we analyzed the distribution of CTCF in MSCs by fluorescence microscopy we also observed increased co-localization of CTCF in the larger nuclei of senescent MSCs (Fig. 6A). We then asked the question if CTCF binding at specific genomic locations is also changed during culture expansion. To address this question, we performed chromatin immune precipitation (ChIP) with MSCs of early and late passage (n = 3). Overall, the ChIP seq peaks were in line with previous data of embryonic stem cell derived MSCs (38) (Supplemental Fig. S5A) and they centered clearly around predicted CTCF binding motives (Fig. 6B). We were then looking for differential CTCF peaks between early and late passage. Transforming the data onto M (log ratio) and A (mean average) scales (MA-plot) revealed that there are no highly abundant differential peaks between early and late passages (Fig. 6C). Furthermore, Spearman correlation of normalized read counts provided further evidence that ChIP seq profiles of early and late passages are highly correlated (Supplemental Fig. S5B). Thus, CTCF binding appears to be relatively stable during culture expansion, despite dramatic increase in nuclear size, reproducible chromatin conformation changes, and reorganization of SICCs.

We then analyzed if chromatin interactions of the four culture-associated regions in our 4C data were related to CTCF sites. In fact, nearbait and cis-interacting regions of two hypermethylated sites (*ALOX12* and *LTC4S*) exhibited significant enrichment of binding motifs for CTCF and CCCTC-binding factor like (CTCFL; Supplemental Fig. S5C and D). Similar results were observed using our CTCF Chip-seq data of MSCs (Supplemental Fig. S5E).

Subsequently, we analyzed if CpGs that become either hyper- or hypomethylated during culture expansion are related to CTCF binding sites. To provide enough genomic regions for such statistical analysis, we used our set of 646 hyper- and 2,442 hypomethylated CpGs that were selected with the less stringent filter criteria as described above. In fact, hypermethylated CpGs were enriched at CTCF binding sites, whereas genomic regions that become hypomethylated were almost devoid of CTCF (Fig. 6D). This was consistent and highly significant ($p < 10^{-51}$) for all three donors (Fig. 6E). For comparison, we have also analyzed 2,000 randomly chosen CpGs from the Illumina 450k BeadChip. Since CTCF-binding motives often comprise CpGs there was also a CTCF-ChIP-seq peak around these random CpGs, but significantly less than for the culture-associated hypermethylated, and significantly more than for the hypomethylated CpGs (Fig. 6E). These results indicate that the formation of CTCF associated chromatin loops is relevant for the stochastic gains and losses of DNAm during culture expansion.

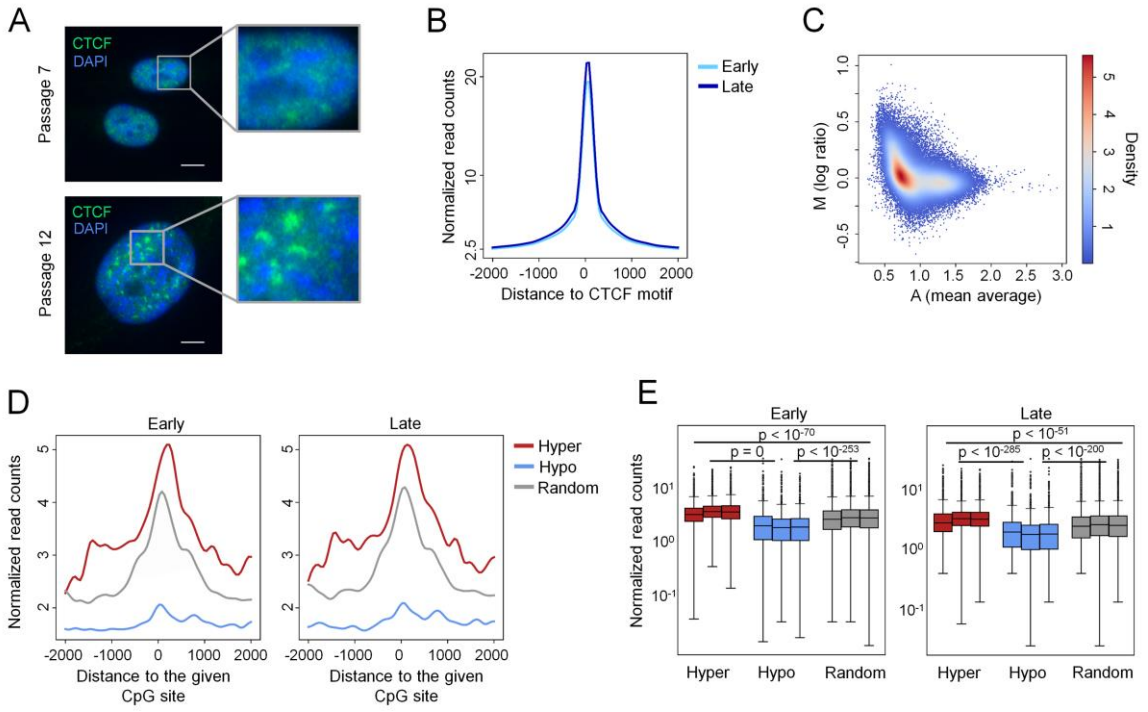

Figure 6. CTCF binding in cells of early and late passages.
A) High resolution microscopy pictures of MSCs at median passage (P7) and late passage (P12). Staining of cell nuclei with DAPI and CTCF reveals co-localization of CTCF particularly in the larger nuclei of late passage MSCs (size bar = 5 µm). **B)** CTCF ChIP-seq signals of cells in early (P2) and late (P8 - P14) passage are centered around predicted CTCF binding motifs. **C)** MA plot of CTCF ChIP-seq signals of cells in early (P2) and late (P8 - P14) passage. Mean average (A) is the mean of the normalized read counts of early and late passages, whereas log ratio (M) is the ratio of the normalized read counts of late passages over early passages. D) Lineplots depict the CTCF ChIP-seq signals of cells in early (P2) and late (P8 - P14) passage centered around hyper- or hypomethylated culture-associated CpGs or 2,000 randomly picked CpGs of the Illumina 450k BeadChip. **E)** Statistical testing of the ChIP-seq signal differences of three donors in early (P2) and late (P8 - P14) passages at the CpG subsets shown in D). CTCF ChIP-seq signals at hypermethylated CpGs are significantly enriched in comparison to randomly chosen CpGs, while CTCF ChIP-seq signals at hypomethylated sites are significantly depleted (Dunn´s test on the quantile normalized reads of the 2,000 bp window).

Discussion

The continuous and highly reproducible nature of culture-associated DNAm changes may suggest that this process is tightly controlled (40). Nevertheless, our results support the

notion that the process is rather associated with stochastic epigenetic drift, which does not involve site-specific targeting of regulatory protein complexes: 1) Resetting of culture-associated DNAm during reprogramming into iPSCs occurs synchronously with the epigenetic changes in pluripotency genes, which is directly linked to the epigenetic transition itself. Furthermore, it occurs simultaneously with resetting of age-associated DNAm, which has also been attributed to epigenetic drift (11-13). 2) The culture-associated DNAm patterns do not reflect MSC clonality and become even more diverse in oligoclonal cell preprations at late passages. If these epigenetic modifications were evoked by targeted regulation in individual subclones specific patterns should become dominant, too. 3) BBA-seq analysis demonstrated that culture-associated DNAm patterns do not develop in an additive manner at neighboring CpGs and this was further supported by hairpin sequencing of the complementary DNA strands. If an epigenetic writer is targeted to a specific site in the genome the neighboring CpGs will most likely also be modified, as observed for CRISPR guided approaches of epigenetic writers that coherently modify neighboring CpGs (41). 4) There was no evidence that culture-associated DNAm is coherently modified at interacting chromatin domains. It is known that DNMTs accumulate in replication foci or punctate heterochromatic foci (43, 44) and hence, it might be speculated that a targeted mechanism of an epigenetic writer would also involve interaction of culture-associated chromatin domains. If culture-associated DNAm changes are not governed by targeting of epigenetic writers, there needs to be another – yet unknown – mechanism that disposes specific genomic regions to epigenetic drift.

Cell culture is evidently associated with major changes in chromatin structure. The nucleus becomes much larger at later passages, while chromatin volume decreases due to extensive reorganization of hetero- and euchromatin conformation (35, 45). Nuclear depletion of HMGB2 and its induction of CTCF clustering are early events on the path to

replicative senescence, which disturb the chromosomal 3D organization (22). Our 4C analysis supports the notion that the 3D chromatin structure undergoes highly reproducible changes during culture expansion. It has been demonstrated that CTCF occupancy, which is to some extent cell-type specific, is also linked to differential DNA methylation (46). Furthermore, some CTCF sites may function as a bifurcation point defining the differential methylation landscape (47). In this regard, it was unexpected that we did not observe clear differences in the CTCF ChIP-seq data of early *versus* late passages. On the other hand, hypermethylation seems to occur preferentially at CTCF binding sites, whereas hypomethylation occurs apart from CTCF binding sites and preferentially at LADs. Our results support the notion that the given chromatin conformation favors site-specific epigenetic drift over subsequent passages. The DNAm changes during long-term culture may therefore be indirectly mediated by the histone code, insulators, chromatin loops and the overarching nuclear structure.

The functional relevance of DNAm changes is still unclear. Culture-associated DNAm is only partly associated with gene expression changes of corresponding genes (24, 49, 50). Yet, it would be an oversimplification to only consider transcriptional regulation as being functionally relevant. Chromatin conformation, loop structures, histone modifications, and lamina-associated domains may favor modulation of DNAm at specific sites in the genome – and on the other hand culture-associated DNAm may stabilize such chromatin features. Different levels of chromatin organization will have to be considered to fully understand the underlying mechanism that drive epigenetic drift during long-term culture.

Materials and Methods

Identification of culture-associated CpGs

We compiled 64 published and newly generated DNAm datasets of untreated primary cells with reliable information on passage numbers (all Illumina 450k Methylation BeadChip; Supplemental Table S1). For further analysis we excluded CpGs from X and Y chromosomes and performed k-nearest neighbor imputation and quantile normalization using the R packages impute and lumi (51), respectively. An outlier test using the R package car was performed resulting in exclusion of one sample (GSM1004625). Initially, we selected culture-associated CpGs based on Pearson correlation of DNAm levels (beta values) and passage numbers at a threshold of $R > 0.7$ or $R < -0.7$ (646 hyper- and 2,442 hypomethylated CpGs, respectively; Supplemental Table S2). Subsequently, we used a more stringent threshold to identify the best candidates for targeted analysis: $R > 0.8$ or $R < -0.8$; and the slope of regression $m > 0.02$ (highest ranked 15 hyper- and 15 hypomethylated CpG sites; Supplemental Table S3). Using the R package leaps we calculated the best multivariable linear regression model including two hypo- and two hypermethylated CpG sites.

Analysis of DNAm changes during reprogramming into iPSCs

The kinetics of culture-associated DNAm changes during reprogramming of fibroblasts into iPSCs were investigated using the dataset of Ohnuki *et al.* (GSE54848; 450k BeadChip) (28). For comparison we used DNAm profiles of MSCs at early and late passage (GSE37067) (24) as well as DNAm profiles during re-differentiation of iPSCs to iMSCs (GSE54767) (25). We focused on three different sets of culture-associated CpGs: 1) 30 CpGs (filtered by $R > 0.8$ or $R < -0.8$, $m > 0.02$, described above); 2) 3,088 CpGs (filtered by $R > 0.7$ or $R < -0.7$, described above); and 3) 9 CpGs of our epigenetic

signatures (*KRTAP13.3*, *SELP*, *CASP14*, *TNNI3K*, *DOK6*, *CASR*, *GRM7*, *LTC4S*, *ALOX12*) (7) . Furthermore, we focused on 1,432 pluripotency-associated CpGs (29) and 99 age-associated CpGs (27), as well as 2,000 random CpG sites of the Illumina 450k BeadChip. Heatmaps were produced with R package gplots, Pearson correlation and corresponding p-values were calculated with R package stats.

Cell culture

Mesenchymal stem cells were isolated from the bone marrow of donors undergoing orthopedic surgery (n = 11; BM-MSC RWTH Aachen) (52), from bone marrow aspirates of allogeneic hematopoietic stem cell donors (n = 8; BM-MSC University of Heidelberg) (53), from subcutaneous adipose tissue lipoaspirates (n = 3, AT-MSC RWTH Aachen) (54) and from umbilical cord pieces (n = 2; UCP-MSC; University of Hannover) (33). Fibroblasts were isolated from dermis (n = 4; RWTH Aachen) (7). Human umbilical vein endothelial cells were isolated from umbilical cords of healthy donors after cesarean sections (HUVECS; n = 4; RWTH Aachen) (55) or obtained from Lonza (n = 3, Basel, Switzerland) (8). All samples were taken after informed and written consent and the study was specifically approved by the ethics committees of the corresponding Universities. All cell preparations were thoroughly characterized (including morphology, immunophenotype, and three lineage *in vitro* differentiation potential) and culture conditions were used as described in detail in our previous work (7, 8, 33, 52-54). In addition, HUVECS were cultured on 0.1% gelatin in M199 medium (Thermo Fisher, Waltham, USA) supplemented with 20% fetal calf serum (FCS) (Gibco Thermo Fisher), 1 % penicillin/streptomycin (PAA)**,** 0.1 % heparin (5000 IU/ml, Ratiopharm) and 50 µg/ml endothelial cell growth supplement (ECGS) (Sigma-Aldrich, St. Louis, USA)**.** For long-term culture all cells were passaged at approximately 90% confluency and reseeded at

10,000 cells/cm^2. Beta-galactosidase (SA-ß-gal) staining was performed using the Senescence Detection Kit ab65351 (Abcam, Cambridge, UK). Cumulative population doublings (cPD) were calculated as described before (54).

Pyrosequencing

Genomic DNA was isolated with the NucleoSpin Tissue kit (Macherey&Nagel, Düren, Germany) and bisulfite converted using the EZ DNA Methylation kit (Zymo Research, Irvine, CA, USA). Pyrosequencing was performed on a PyroMark ID System (Biotage, Uppsala, Sweden). Primers for pyrosequencing were designed with the PSQ assay design software (Biotage; Supplemental Table S4). DNAm levels were determined with the Pyro-Q-CpG Software (Biotage). To train the epigenetic predictor on pyrosequencing data we divided the pyrosequencing samples into a training and validation set (Supplemental Tables S5 and S6). The multivariable linear regression model based on DNAm levels (β-values) at the four CpGs in α = *ALOX12* (cg03762994), β = *DOK6* (cg25968937), γ = *LTC4S* (cg26683398) and δ = *TNNI3K* (cg05264232) was as follows:

$$\text{Predicted passage} = 39.0341 – 10.9266\ \alpha - 0.4219\ \beta + 5.8979\ \gamma - 38.889\ \delta$$

Finally, we used the R package caret (56) to perform a 10-fold cross-validation on the training dataset.

Lentiviral Barcode-RGB marking

Primary MSCs in umbilical cord tissue were transduced with three different lentiviral vectors containing the fluorescent proteins mCherry (red), Venus (green) or Cerulean (blue) and a barcode of 16 random and 15 vector-backbone-specific nucleotides, as previously published (33). Clonal dynamics were assessed at each passage by flow

cytometry using a BD LSR II flow cytometer (BD Biosciences, Heidelberg, Germany) and by deep sequencing of PCR amplified barcodes using ion torrent sequencing. Area plots were produced with the R package ggplot2. The Shannon-Index was calculated with the following formula:

$$H` = -\sum_i p_i * \ln p_i$$

where p_i is the proportion of distinct RGB clones or DNAm patterns, respectively.

Barcoded-bisulfite-amplicon sequencing

Bisulfite converted DNA was used for a nested PCR using the PyroMark PCR kit (Qiagen; primers are provided in Supplemental Table S7). The second PCR added barcoded Illumina adapters that allowed to distinguish donors and passages, as described before (8). Amplicons were pooled and sequenced on an Illumina MiSeq lane with the v2 nano reagents (Illumina) in 250 PE mode. Bisulfite converted sequencing data was analyzed using TrimGalore, Bismark (57) and bowtie2 (58). Mean sequencing coverage of amplicons was approximately 3,900 reads per amplicon. Further pattern analysis and visualization was performed with custom perl and R scripts or with R package ggplot2 for area plots. Pearson correlation of neighboring CpGs and the corresponding heatmap were produced with the python packages scipy and seaborn, respectively. Single read predictions were performed as described before (34). In short, single reads of the BBA-Seq amplicons were assigned to their most likely passage number (from 0 to 50) based on their binary sequel of methylated and unmethylated CpG sites. Probabilities of passage numbers were based on linear regression models at each CpG site, retrieved from the training dataset. Finally, we calculated the mean passage number for each sample based

on all sequencing reads. Further details on the rational and derivation of the mathematical model are provided in our previous work (34).

Analysis of hemimethylation

Hemimethylation analysis was modified from a protocol by Laird et al. (16). Genomic DNA (4 µg, three donors in passages 2 and 13) was digested with restriction enzymes that cut close to our CpG of interest: AccI (*CASR*, *TNNI3K*, *C12orf12*), DdeI (*KRTAP13.3*, *LTC4S*) and CviQI (*DOK6*, *GRM7*, *PRAMEF2*, *SELP*, *GNAS*). Hairpin linkers (Supplemental Table S8) were denatured at 95°C and subsequently folded by slow cooling to room temperature. Ligation was performed with 4,000 U ligase and 3.3 µM hairpin linker DNA over night at 16° C. The ligated DNA was denatured with 0.3 M NaOH at 42°C for 15 min and 99°C for 2 min before a 0.4 g/ml sodium bisulfite and 1 mg/ml hydroquinone solution was added. Bisulfite conversion was performed at 55°C overnight with 10 intervening denaturation steps at 99°C to prevent renaturation of hairpin structures. The regions of interest were subsequently amplified by PCR using the PyroMark PCR kit (Qiagen; primers listed in Supplemental Table S9). Illumina adapters were ligated using the GeneRead DNA Library I Core Kit (Qiagen) and GeneRead DNA I Amp Kit (Qiagen). Final library cleaning was performed with the Select-a-size DNA Clean & Concentrator kit (Zymo Research). MiSeq v2 nano reagents (Illumina) were used for library dilution to 4 nM and 20% PhiX were spiked in to increase sequencing diversity. Sequencing was performed on an Illumina MiSeq in 250 PE mode and analyzed as described above. Directional clustering of UMI tools was used to group the unique molecular identifiers (59). Mean sequencing coverage of hairpin amplicons was approximately 9,600 reads per amplicon.

Circular Chromatin Conformation Capture (4C)

Ten million cells from two MSC preparations at early (P2 and 3) and late (P7 and 9) passages were cross-linked with 4% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA, USA) for 10 min., harvested into ice cold PBS with 0.125 M glycine, and frozen under a protease inhibitor cocktail (Roche, Basel, Suisse). 4C-seq was performed as described before (60) using *ApoI* as the primary and *DpnII* as the secondary restriction enzyme. Bait-specific primers for the circularized inverse PCR are listed in Supplemental Table S10. Amplicons were sequenced on a HiSeq2500 platform (Illumina), mapped to the reference genome (hg19) and analyzed with the Hidden-Markov-Model-based tool *4Cker* (61). Numbers of sequenced reads per sample ranged from approximately 0.7 to 12 million reads with a mean sequenced read number of approximately 7 million reads per sample. *4Cker* corrects for increasing signal noise in *trans* chromosomal interactions and far-cis chromosomal interactions by adaptive window sizes. We used the kth-next-neighbor adaptive window sizes of k = 5 for nearbait (10 MB around the “bait” region of interest) and cis interaction analysis and k = 20 for *trans* chromosomal interactions. We focused particularly on the high-interacting reads that were called in both replicates. Differential interactions were called with *4Cker*, which uses DESeq2 and a FDR corrected p-value < 0.5 to call significant differential interactions. Circos-plots were generated with the R package RCircos (62). For comparison we used DNAm profiles of MSCs at early (n = 5) and late passages (n = 5; GSE37067) (24). CTCF and CTCFL enrichment in interacting regions was tested with the RGT motif enrichment tool (http://www.regulatory-genomics.org/motif-analysis/introduction). Motif enrichment was tested within high-interacting regions called by *4Cker* ranging approximately from 1.1 to 336 kilobases and a mean region length of 23 kilobases within nearbait and from 1.3 to 950 kilobases with a mean region length of 86 kilobases within cis regions. Enrichment of interactions with

lamina-associated domains (LADs) (48) were analyzed in comparison to randomly chosen regions of similar sizes by Fisher's exact test using R stats.

Fluorescence microscopy of CTCF

Staining of CTCF (Rabbit polyclonal anti-CTCF, AB_2614983; Active Motive) and counterstaining with DAPI was performed as described in detail before (22). For image acquisition, a widefield LeicaDMI 6000B with a HCX PL APO 63x/1.40 (Oil) objective was used.

CTCF ChIP-seq

Chromatin Immunoprecipitation of CTCF was performed using the ChIP-IT PBMC kit with the fixation protocol of the ChIP-IT High Sensitivity kit (both Active Motif, Carlsbad, CA, USA). In brief, MSCs of early passage (P 2, n = 3) and late passage (P 8 – 14, n = 3) were cross-linked with 1.2 % formaldehyde (AppliChem, Darmstadt, Germany) for 15 min at room temperature. After cell lysis, chromatin fragmentation was performed for 20 min (20 cycles a´ 30 sec on/30 sec off, total on-time 10 min) with a Biorupter® Pico Sonicator Device (Diagenode). 5 - 33 µg of sonicated chromatin was incubated with 4 µg ChIP-validated CTCF antibody (Active Motif) overnight on an end-to-end rotator at 4°C. For immunoprecipitation, samples were incubated with protein G agarose beads (Active Motife) at 4°C for 3 hours. Finally, the cross-link was reversed, the DNA was proteinase K digested and purified.

Sequencing and library preparation were conducted by the IZKF Sequencing Core Facility of RWTH Aachen University Medical school. The samples were sequenced in paired-end mode on the Illumina NextSeq 500. Quality of the resulting reads was checked by FastQC (Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc) before

reads were aligned to the hg19 genome assembly using bowtie2 (58). Reads approximately exhibited a 97% concordant alignment rate in all samples and sequencing depth ranged from 38 to 170 million reads. We used MACS2 (63) for peak calling on each sample with the default parameters against the input control. All the peaks were filtered by signal > 50 and selected with the overlapping of potential CTCF binding motif, which was obtained by RGT motif analysis (http://www.regulatory-genomics.org/motif-analysis). Enrichment of interactions between 4C high interacting regions and CTCF ChIP-seq peaks of early and late passages were analyzed in comparison to randomly chosen regions of similar sizes by Fisher's exact test using R stats. The normalized coverage of each sample was calculated by Deeptools (64) bamCoverage with normalization of Reads Per Kilobase per Million mapped reads (RPKM), these normalized coverages were used for down-stream analyses. The correlation heatmap across the samples was generated by Deeptools multiBigwigSummary. All the potential CTCF binding sites were evaluated by the normalized coverage profiles and shown in MA plot by comparing early passages with late passages. The lineplots were calculated by a sliding window (size = 200 bp with step = 100 bp) on the normalized read counts on the extended sites (length = 2,000 bp for each direction) of either CTCF motifs or various sets of CpGs. Statistical analysis was performed against these random CpG sites using the Kruskal-Wallis Test with post-hoc Dunn's test and Benjamini & Hochberg adjustment.

Declarations

Funding

This work was particularly supported by the Else Kröner-Fresenius-Stiftung (AP and WW: 2014_A193, www.ekfs.de), by the Deutsche Forschungsgemeinschaft (WW: WA 1706/8-

1; WA 1706/12-1; AP: UoC Advancer Research grant of the DFG Excellence Initiative, MR: RO 5102/1-1, www.dfg.de), by the German Ministry of Education and Research (WW: VIP+, 03VP06120, www.bmbf.de), by the Interdisciplinary Center for Clinical Research within the faculty of Medicine at the RWTH Aachen University (WW: IZKF O3-3, www.medizin.rwth-aachen.de/cms/Medizin/Die-Fakultaet/Einrichtungen/~dgun/IZKF-Aachen/), and by CMMC core funding (AP, www.cmmc-uni-koeln.de).

Conflict of interests

WW is cofounder of Cygenia GmbH (www.cygenia.com), which can provide service for epigenetic analysis to other scientists. JF contributes to this company, too. All other authors do not have a conflict of interest to declare.

Ethics approval and consent to participate

All human cell samples were taken after informed and written consent was obtained from donors and the study was specifically approved by the ethics committees of RWTH Aachen University Medical School (permit numbers: EK300/13, EK163/07, EK 187/08) and Hannover Medical School.

Authors' contributions

JF contributed to experimental design, analysis and writing of the manuscript; TG, LB and AP performed 4C experiments; MM performed CTCF ChIP-seq experiment and C-CK analyzed the ChIP-seq data; AnS, MR and AxS performed barcoded-RGB-vector experiments and MN supported the analysis; RS helped to establish hairpin BBA-Seq; EFR, CG and AO contributed to long term cell culture and cellular characterization; MB

helped to sequence data; BR and ADH contributed important material. WW contributed to experimental design, data analysis and writing of the manuscript.

Acknowledgements

We would like to thank all donors for their valuable support of our research and Minseung Choi (University of Washington, Washington, USA) for his input to the analysis of hairpin BBA-Seq.

References

1. Campisi J, d'Adda di Fagagna F. Cellular senescence: when bad things happen to good cells. Nat Rev Mol Cell Biol. 2007;8(9):729-40.

2. Fernandez-Rebollo E, Franzen J, Goetzke R, Hollmann J, Ostrowska A, Oliverio M, et al. Senescence-Associated Metabolomic Phenotype in Primary and iPSC-Derived Mesenchymal Stromal Cells. Stem cell reports. 2020;14(2):201-9.

3. Galipeau J, Sensébé L. Mesenchymal Stromal Cells: Clinical Challenges and Therapeutic Opportunities. Cell stem cell. 2018;22(6):824-33.

4. Pittenger MF, Discher DE, Péault BM, Phinney DG, Hare JM, Caplan AI. Mesenchymal stem cell perspective: cell biology to clinical progress. npj Regenerative Medicine. 2019;4(1):22.

5. Bork S, Pfister S, Witt H, Horn P, Korn B, Ho AD, et al. DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. Aging cell. 2010;9(1):54-63.

6. Cruickshanks HA, McBryan T, Nelson DM, Vanderkraats ND, Shah PP, van Tuyn J, et al. Senescent cells harbour features of the cancer epigenome. Nature cell biology. 2013;15(12):1495-506.

7. Koch CM, Joussen S, Schellenberg A, Lin Q, Zenke M, Wagner W. Monitoring of cellular senescence by DNA-methylation at specific CpG sites. Aging cell. 2012;11(2):366-9.

8. Franzen J, Zirkel A, Blake J, Rath B, Benes V, Papantonis A, et al. Senescence-associated DNA methylation is stochastically acquired in subpopulations of mesenchymal stem cells. Aging Cell. 2017;16(1):183.

9. Wagner W. The Link Between Epigenetic Clocks for Aging and Senescence. Front Genet. 2019;10:303.

10. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 2019;20(10):590-607.

11. Teschendorff AE, West J, Beck S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? Hum Mol Genet. 2013;22(R1):7-15.

12. Hernando-Herraez I, Evano B, Stubbs T, Commere PH, Jan Bonder M, Clark S, et al. Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. Nat Commun. 2019;10(1):4361.

13. Li Y, Tollefsbol TO. Age-related epigenetic drift and phenotypic plasticity loss: implications in prevention of age-related human diseases. Epigenomics. 2016;8(12):1637-51.

14. Ushijima T, Watanabe N, Okochi E, Kaneda A, Sugimura T, Miyamoto K. Fidelity of the methylation pattern and its variation in the genome. Genome Res. 2003;13(5):868-74.

15. Jones PA, Liang G. Rethinking how DNA Methylation Patterns are Maintained. Nature reviews Genetics. 2009;10(11):805-11.

16. Laird CD, Pleasant ND, Clark AD, Sneeden JL, Hassan KM, Manley NC, et al. Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. Proc Natl Acad Sci U S A. 2004;101(1):204-9.

17. Choi M, Genereux DP, Goodson J, Al-Azzawi H, Allain SQ, Simon N, et al. Epigenetic memory via concordant DNA methylation is inversely correlated to developmental potential of mammalian cells. PLoS genetics. 2017;13(11):e1007060.

18. Patino-Parrado I, Gomez-Jimenez A, Lopez-Sanchez N, Frade JM. Strand-specific CpG hemimethylation, a novel epigenetic modification functional for genomic imprinting. Nucleic Acids Res. 2017;45(15):8822-34.

19. Arand J, Wossidlo M, Lepikhov K, Peat JR, Reik W, Walter J. Selective impairment of methylation maintenance is the major cause of DNA methylation reprogramming in the early embryo. Epigenetics & Chromatin. 2015;8(1):1.

20. Xu C, Corces VG. Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. Science. 2018;359(6380):1166-70.

21. Sharif J, Koseki H. Hemimethylation: DNA's lasting odd couple. Science. 2018;359(6380):1102-3.

22. Zirkel A, Nikolic M, Sofiadis K, Mallm J-P, Brackley CA, Gothe H, et al. HMGB2 Loss upon Senescence Entry Disrupts Genomic Organization and Induces CTCF Clustering across Cell Types. Molecular Cell. 2018;70(4):730-44.e6.

23. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell. 2007;128(6):1231-45.

24. Koch C, Reck K, Shao K, Lin Q, Joussen S, Ziegler P, et al. Pluripotent Stem Cells Escape From Senescence-Associated DNA Methylation Changes. Genome Res. 2013;23(2):248-59.

25. Frobel J, Hemeda H, Lenz M, Abagnale G, Joussen S, Denecke B, et al. Epigenetic Rejuvenation of Mesenchymal Stromal Cells Derived from Induced Pluripotent Stem Cells. Stem Cell Reports. 2014;3(3):414-22.

26. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.

27. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biology. 2014;15(2):R24-R.

28. Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, et al. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(34):12426-31.

29. Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, et al. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell stem cell. 2012;10(5):620-34.

30. Olova N, Simpson DJ, Marioni RE, Chandra T. Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity. Aging cell. 2019;18(1):e12877.

31. Sheng C, Jungverdorben J, Wiethoff H, Lin Q, Flitsch LJ, Eckert D, et al. A stably self-renewing adult blood-derived induced neural stem cell exhibiting patternability and epigenetic rejuvenation. Nat Commun. 2018;9(1):4047.

32. Frobel J, Hemeda H, Lenz M, Abagnale G, Joussen S, Denecke B, et al. Epigenetic rejuvenation of mesenchymal stromal cells derived from induced pluripotent stem cells. Stem Cell Reports. 2014;3(3):414-22.

33. Selich A, Daudert J, Hass R, Philipp F, von Kaisenberg C, Paul G, et al. Massive Clonal Selection and Transiently Contributing Clones During Expansion of Mesenchymal Stem Cell Cultures Revealed by Lentiviral RGB-Barcode Technology. Stem Cells Translational Medicine. 2016;5(5):591-601.

34. Han Y, Franzen J, Stiehl T, Gobs M, Kuo C-C, Nikolić M, et al. New targeted approaches for epigenetic age predictions. BMC Biology. 2020;18(1):71.

35. Chandra T, Ewels Philip A, Schoenfelder S, Furlan-Magaril M, Wingett Steven W, Kirschner K, et al. Global Reorganization of the Nuclear Landscape in Senescent Cells. Cell Reports. 2015;10(4):471-83.

36. Hänzelmann S, Beier F, Gusmao EG, Koch CM, Hummel S, Charapitsa I, et al. Replicative senescence is associated with nuclear reorganization and with DNA methylation at specific transcription factor binding sites. Clinical Epigenetics. 2015;7(1):19.

37. Shah PP, Donahue G, Otte GL, Capell BC, Nelson DM, Cao K, et al. Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. Genes & Development. 2013;27(16):1787-99.

38. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015;518(7539):331-6.

39. Gu Y, Li T, Ding Y, Sun L, Tu T, Zhu W, et al. Changes in mesenchymal stem cells following long-term culture in vitro. Molecular medicine reports. 2016;13(6):5207-15.

40. Xie W, Kagiampakis I, Pan L, Zhang YW, Murphy L, Tao Y, et al. DNA Methylation Patterns Separate Senescence from Transformation Potential and Indicate Cancer Risk. Cancer Cell. 2018;33(2):309-21 e5.

41. Josipovic G, Tadic V, Klasic M, Zanki V, Beceheli I, Chung F, et al. Antagonistic and synergistic epigenetic modulation using orthologous CRISPR/dCas9-based modular system. Nucleic Acids Res. 2019;47(18):9637-57.

42. Lövkvist C, Dodd IB, Sneppen K, Haerter JO. DNA methylation in human epigenomes depends on local topology of CpG sites. Nucleic Acids Research. 2016;44(11):5123-32.

43. Ren W, Fan H, Grimm SA, Guo Y, Kim JJ, Yin J, et al. Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. Proc Natl Acad Sci U S A. 2020;117(31):18439-47.

44. Dukatz M, Holzer K, Choudalakis M, Emperle M, Lungu C, Bashtrykov P, et al. H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to its Chromatin Interaction. J Mol Biol. 2019;431(24):5063-74.

45. Criscione SW, De Cecco M, Siranosian B, Zhang Y, Kreiling JA, Sedivy JM, et al. Reorganization of chromosome architecture in replicative cellular senescence. Science Advances. 2016;2(2):e1500882.

46. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. 2012;22(9):1680-8.

47. Wiehle L, Thorn GJ, Raddatz G, Clarkson CT, Rippe K, Lyko F, et al. DNA (de)methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. Genome Res. 2019;29(5):750-61.

48. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008;453(7197):948-51.

49. Sidler C, Woycicki R, Kovalchuk I, Kovalchuk O. WI-38 senescence is associated with global and site-specific hypomethylation. Aging (Albany NY). 2014;6(7):564-74.

50. Sakaki M, Ebihara Y, Okamura K, Nakabayashi K, Igarashi A, Matsumoto K, et al. Potential roles of DNA methylation in the initiation and establishment of replicative senescence revealed by array-based methylome and transcriptome analyses. PLoS ONE. 2017;12(2):e0171431.

51. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics (Oxford, England). 2008;24(13):1547-8.

52. Fernandez-Rebollo E, Mentrup B, Ebert R, Franzen J, Abagnale G, Sieben T, et al. Human Platelet Lysate versus Fetal Calf Serum: These Supplements Do Not Select for Different Mesenchymal Stromal Cells. Scientific Reports. 2017;7(1):5132.

53. Horn P, Bork S, Diehlmann A, Walenda T, Eckstein V, Ho A, et al. Isolation of human mesenchymal stromal cells is more efficient by red blood cell lysis. Cytotherapy. 2008;10(7):676-85.

54. Cholewa D, Stiehl T, Schellenberg A, Bokermann G, Joussen S, Koch C, et al. Expansion of Adipose Mesenchymal Stromal Cells is Affected by Human Platelet Lysate and Plating Density. Cell transplantation. 2011;20(9):1409-22.

55. Baudin B, Bruneel A, Bosselut N, Vaubourdolle M. A protocol for isolation and culture of human umbilical vein endothelial cells. Nature protocols. 2007;2(3):481-5.

56. Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software. 2008;28(5):26.

57. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics (Oxford, England). 2011;27(11):1571-2.

58. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

59. Smith T, Heger A. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome research. 2017;27(3):491-9.

60. Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, et al. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. Nature protocols. 2013;8(3):509-24.

61. Raviram R, Rocha PP, Muller CL, Miraldi ER, Badri S, Fu Y, et al. 4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments. PLoS computational biology. 2016;12(3):e1004780.

62. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC Bioinformatics. 2013;14:244-.

63. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biology. 2008;9(9):R137.

64. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic acids research. 2016;44(W1):W160-W5.