

RESEARCH ARTICLE

Strong Components of Epigenetic Memory in Cultured Human Fibroblasts Related to Site of Origin and Donor Age

Nikolay A. Ivanov¹, Ran Tao¹, Joshua G. Chenoweth¹, Anna Brandtjen¹, Michelle I. Mighdoll¹, John D. Genova¹, Ronald D. McKay¹, Yankai Jia¹, Daniel R. Weinberger^{1,2,3,4,5}, Joel E. Kleinman¹, Thomas M. Hyde^{1,5,6}, Andrew E. Jaffe^{1,7,8,9*}

1 Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, Maryland, United States of America, **2** McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **3** Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America, **4** Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America, **5** Department of Neurology, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America, **6** Department of Biological Sciences, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America, **7** Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, **8** Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, **9** Center for Computational Biology, Johns Hopkins University, Baltimore, Maryland, United States of America

* andrew.jaffe@libd.org



CrossMark
click for updates

 OPEN ACCESS

Citation: Ivanov NA, Tao R, Chenoweth JG, Brandtjen A, Mighdoll MI, Genova JD, et al. (2016) Strong Components of Epigenetic Memory in Cultured Human Fibroblasts Related to Site of Origin and Donor Age. *PLoS Genet* 12(2): e1005819. doi:10.1371/journal.pgen.1005819

Editor: Bing Ren, Ludwig Institute for Cancer Research, University of California San Diego, UNITED STATES

Received: August 8, 2015

Accepted: December 30, 2015

Published: February 25, 2016

Copyright: © 2016 Ivanov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: DNA methylation data in both raw and processed forms are available on the Gene Expression Omnibus (GEO): GSE77136. RNA sequencing reads (raw data) are available on the Sequencing Read Archive (SRA): SRP068304 (BioProject: PRJNA286856) and the genes and transcript counts (processed data) are available on GEO at the above accession number (GSE77136).

Funding: This work was supported by funding from the Lieber Institute for Brain Development. The funders had no role in study design, data collection

Abstract

Differentiating pluripotent cells from fibroblast progenitors is a potentially transformative tool in personalized medicine. We previously identified relatively greater success culturing dura-derived fibroblasts than scalp-derived fibroblasts from postmortem tissue. We hypothesized that these differences in culture success were related to epigenetic differences between the cultured fibroblasts by sampling location, and therefore generated genome-wide DNA methylation and transcriptome data on 11 intrinsically matched pairs of dural and scalp fibroblasts from donors across the lifespan (infant to 85 years). While these cultured fibroblasts were several generations removed from the primary tissue and morphologically indistinguishable, we found widespread epigenetic differences by sampling location at the single CpG (N = 101,989), region (N = 697), “block” (N = 243), and global spatial scales suggesting a strong epigenetic memory of original fibroblast location. Furthermore, many of these epigenetic differences manifested in the transcriptome, particularly at the region-level. We further identified 7,265 CpGs and 11 regions showing significant epigenetic memory related to the age of the donor, as well as an overall increased epigenetic variability, preferentially in scalp-derived fibroblasts—83% of loci were more variable in scalp, hypothesized to result from cumulative exposure to environmental stimuli in the primary tissue. By integrating publicly available DNA methylation datasets on individual cell populations in blood and brain, we identified significantly increased inter-individual variability in our scalp- and other skin-derived fibroblasts on a similar scale as epigenetic differences between different lineages of blood cells. Lastly, these epigenetic differences did not appear to be driven by somatic mutation—while we identified 64 probable de-novo variants across the 11 subjects, there

and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

was no association between mutation burden and age of the donor ($p = 0.71$). These results depict a strong component of epigenetic memory in cell culture from primary tissue, even after several generations of daughter cells, related to cell state and donor age.

Author Summary

Regenerative medicine specialists have been using a type of cell commonly found in the skin called the fibroblast because it is easily obtained from skin samples, grows well in culture, and can be manipulated in the laboratory to de-differentiate into a primordial state known as the induced pluripotent stem cell. These primitive stem cells can then be transformed into mature tissues, such as liver or pancreas cells. Here we show that fibroblasts, coming from different locations in the same individual, vary significantly in epigenetic marks called DNA methylation, which are involved in the regulation of gene expression. In addition to location-specific patterns of DNA methylation, we also find that fibroblasts from different anatomical locations respond differently in epigenetic patterns related to aging. As the field of regenerative medicine advances, our study demonstrates that deciding upon the source of fibroblasts from an individual to generate new tissues and organs may be an important consideration.

Introduction

DNA methylation (DNAm) at CpG dinucleotides plays an important role in the epigenetic regulation of the human genome, contributing to diverse cellular phenotypes from the same underlying genetic sequence. For example, DNAm levels at particular genomic loci can accurately classify different tissues [1] and even underlying cell types within tissues [2]. These stable cell type- and tissue-discriminating loci appear to represent only a subset of "dynamic" CpGs, approximately 21.8%, actively involved in regulation of gene expression [3]. Changes in these epigenetic patterns across aging have been extensively studied [4], particularly in large studies of whole blood [5–7], but subsets of these age-associated CpGs appear tissue-independent [8].

These epigenetic barcodes also play an important role in cellular reprogramming (the conversion of somatic cells to pluripotent stem cells), a powerful and promising experimental system in biology, genetics and personalized medicine [9]. This epigenetic reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) induces demethylation [10] followed by specific patterns of subsequent DNA methylation that can reflect the original somatic tissue [11]. Fibroblasts are one of the most popular cell types for generating iPSCs [12], particularly from skin, given the relative ease of access to these cells, although other skin-derived cell types such as keratinocytes from the same individual generate similar iPSC lines [13]. Skin, however, is perhaps the most susceptible tissue source in the body to environmentally induced insult, particularly through sunlight and chemical exposures, which can induce changes in epigenetic patterns [14]. The epigenetic "memory" of source tissue for iPSC characterization has been well characterized [11].

In our previous work, we successfully cultured fibroblast lines from the dura mater of post-mortem human donors, a source location largely protected from environmental insult with slowly dividing cells [15]. We compared these cultured fibroblast lines to those derived from scalp samples from the same individuals, and found that the rate of culture success was higher for dura-derived fibroblasts; in some cases only the dura fibroblasts from an individual would

culture. While the resulting cultured cells from these two sampling locations were largely morphologically indistinguishable (see Figure 1 in Bliss et al, 2012 [15]), we hypothesized that increased culture success might have a strong epigenetic component. Previous research has shown that dermal fibroblasts from different locations in the body have distinct gene expression profiles [16], including compared to some non-dermal sources [17], and previous reports have indicated that cultured cells have largely stable epigenomes, with the exception of a small number of loci [18]. We therefore sought to characterize the methylomes and transcriptomes of fibroblasts from these two sampling locations—scalp and dura—from donors across the lifespan.

Here we identify several components of epigenetic “memory” in cultured fibroblasts after multiple passages (i.e. splitting and continuing to grow) where primary tissue originated from two locations in the body. The strongest epigenetic memory was related to sampling location in the body, as we identified widespread DNAm differences at local and regional spatial scales preserved through identical culturing processes. We further find increased stochastic epigenetic variability in cultured fibroblasts from the scalp compared to dura. This increased variability manifested in significant increased quantitative pairwise methylome-wide distances in a combined analysis with publicly available DNAm data on skin fibroblasts [19], pure cell populations from peripheral blood [20], and cells from the dorsolateral prefrontal cortex [21]. Another component of epigenetic memory was related to the age of the donor, including a subset of CpGs that displayed location-dependent changes through aging. The epigenetic differences between these fibroblasts appear to occur largely through epigenetic-dependent mechanisms, as there were few differences in coding sequence across the fibroblasts from the two locations within the same individual. These results demonstrate the effect of epigenetic memory in cultured fibroblasts by sampling location and donor age in morphologically indistinguishable cells.

Results

We measured DNA methylation (DNAm) levels from scalp- and dura-derived cultured fibroblasts in 11 postmortem donors (22 samples) from across the lifespan, ranging from early infancy to 85 years (S1 Fig, S1 Table), using the Illumina HumanMethylation450 microarray (Illumina 450k) [22]. After data processing, normalization, and quality control with the minfi package [23], we obtained normalized data on 21 samples (one dura sample with lower quality was removed prior to across-sample normalization) across 456,513 probes (probes with single nucleotide polymorphisms, SNPs, at the target CpGs or single base extension sites were removed, as were probes on the sex chromosomes, see [Methods](#)).

Strong components of epigenetic memory by primary cell sampling location

We first characterized differences in DNAm levels from cultured fibroblasts derived from different locations (scalp versus dura). Many probes, targeting individual CpGs, were differentially methylated between scalp- and dura-derived fibroblasts—101,989 (22%) at genome-wide significance (false discovery rate, FDR < 5%, see [Methods](#)). These significant DNAm differences between cultured fibroblasts from the scalp and dura were large in magnitude, with 57,704 probes having differences in DNAm levels greater than 10%, and 23,752 with differences greater than 20% (Fig 1A). The directionality of these DNAm differences was balanced, with approximately equal proportions of CpGs showing increased versus decreased methylation in cultured fibroblasts from scalp compared to dura. These differentially methylated probes (DMPs) were widely distributed across the genome, as 18,551 genes (defined by UCSC

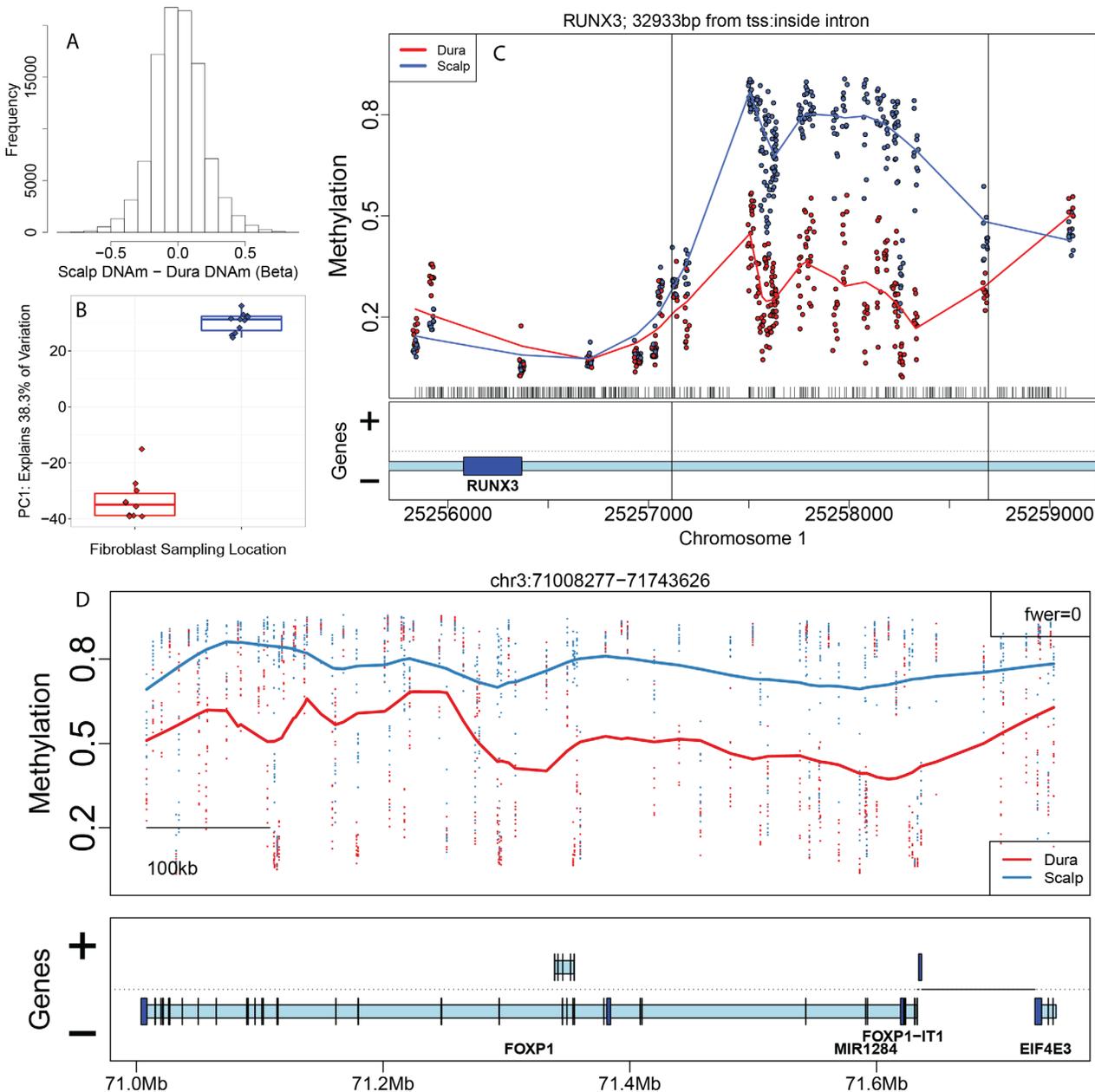


Fig 1. DNA methylation patterns in dura- and scalp-derived fibroblasts. (A) Histogram of difference in DNAm levels at CpGs/probes between scalp and dura derived fibroblasts (on the proportion methylation scale). (B) The first principal component (PC1) of the DNAm data plotted against fibroblast sampling location (scalp versus dura). (C) Example significant differentially methylated region (DMR) that overlaps the gene *RUNX3*, with DNAm levels on the y-axis and genomic coordinates on the x-axis. (D) Example significant DNAm block, with DNAm levels on the y-axis and genomic coordinates on the x-axis. Gene annotation panels in C and D are based on Ensembl annotation—dark blue represents exons and light blue represents introns.

doi:10.1371/journal.pgen.1005819.g001

knownGene database) had at least one DMP within 5 kilobases (kb), as did 33,247 transcripts (see [Methods](#)). These widespread single CpG differences manifest as the largest component of variability in the entire dataset, as the first principal component ([Fig 1B](#), explaining 38% and 62.3% of the variability before and after surrogate variable analysis, SVA [24]) represents the sampling location of these cultured fibroblasts, suggesting a strong epigenetic memory of original cell location.

Since these differentially methylated CpGs tended to cluster in a smaller number of genes, we further identified 697 differentially methylated regions (DMRs) at stringent genome-wide significance (family-wise error rate, FWER < 10%)—these regions were identified based on adjacent probes showing directionally-consistent differences in DNAm > 10% between groups [25] (see [Methods](#)). For example, we identified a region of 24 contiguous probes hypermethylated in scalp-derived fibroblasts within the gene *RUNX3*—a tumor suppressor that plays an integral role in regulating cell proliferation and the rate of apoptosis [26] ([Fig 1C](#), see [S2 Fig](#) and [S2 Table](#) for all significant DMRs). Regional differences, particularly in CpG island shores, previously have been shown to better distinguish tissues and cell types [1] and correlate with neighboring gene expression levels [23] than individual CpGs. Unlike at the single CpG level, which had balanced directionality of differential methylation, the majority of DMRs had higher DNAm levels in fibroblasts derived from scalp compared to those derived from dura (N = 414, 59.4%). Using gene sets defined by biological processes [27], these neighboring genes (within 5 kb) were strongly enriched for morphogenesis (including morphogenesis of the epithelium), developmental processes, cell differentiation, and epithelium and connective tissue development, among other more general gene sets (all $p < 10^{-8}$, [S3 Table](#)).

In addition to the extensive differential methylation at both the CpG and regional level, we identified 243 long-range regions with consistent significant methylation change (FWER < 10%), called “blocks” [28], using an algorithm adapted from whole genome bisulfite sequencing (WGBS) data to Illumina 450k [23]. A representative significant block is shown in [Fig 1D](#) (see [S3 Fig](#) for all significant blocks at FWER < 10%). Blocks have now been identified across many cancer types [29], and tend to associate with higher order chromatin structure including nuclear lamin-associated domains (LADs) [30] and large organized chromatin K9 modification (LOCKs) [28]. The 243 significant blocks in our data represent 41 Mb of sequence and contain 298 annotated genes. These blocks contain 41 of the significant DMRs that differentiate sampling location of the fibroblasts, and more interestingly, every block overlaps at least one “dynamic” cell/tissue DMR identified using WGBS data from Ziller et al (2013) [3].

While these cultured fibroblasts were several generations/passages removed from the primary tissue and morphologically indistinguishable, we nevertheless found widespread epigenetic differences by sampling location of the primary fibroblasts at varying spatial scales, suggesting a strong epigenetic memory of the original cell location.

Epigenetic memory related to original cell location manifests in the transcriptome

We next sought to determine the functional correlates of the widespread epigenetic differences identified between scalp- and dura-derived fibroblasts by performing RNA sequencing (RNA-seq) on polyadenylated (polyA+) mRNA from the same cultured samples (see [Methods](#)). Briefly, we aligned the reads to the transcriptome using TopHat [31] and generated normalized gene counts (as fragments per kb per million mapped reads, FPKM) based on the Illumina iGenome hg19 annotation using the featureCounts software [32]. A median of 88.0% (interquartile range, IQR: 85.5%–88.8%) of reads mapped to the genome, of which a median of 84.7% (IQR: 84.4%–85.5%) mapped to the annotated transcriptome (see [S1 Table](#) for sample-specific percentages). We identified 11,218 expressed genes with average FPKM expression greater than 1.0. Initial clustering of the gene FPKM values separated the fibroblast samples by location in the first principal component (PC), which explained 35.4% of the variance ([S4 Fig](#)), mirroring the first principal component of the DNAm data ([Fig 1B](#)). We could further cluster our samples by sampling location using a set of 337 genes (of which 210 were in our dataset) that were previously identified by Rinn et al [17] to group largely dermal fibroblasts by their anatomical

sites of origin (S5 Fig)—these genes better clustered the samples by sampling location than random sets of 210 genes ($p < 0.001$, see [Methods](#)). Differential expression analysis of the RNA-seq data, independent of the results from the epigenetic analyses above, identified many genes that differed by the source of the primary fibroblast— 5,830 genes at FDR < 5%. Both scalp- and dura-derived fibroblasts expressed high levels of Fibroblast Specific Protein-1 (FSP-1) and this gene was more highly expressed scalp-derived fibroblasts (fold change = 5.5, FDR = 5.6×10^{-6}) in line with increased higher proliferation rates in the scalp-derived versus dura-derived fibroblasts [15]. The differentially expressed genes were strongly enriched for signaling and cell communication, cell proliferation, apoptotic processes, and epithelium development and morphogenesis via gene ontology (GO) analysis (all $p < 10^{-8}$, S4 Table)—these gene sets were similar, and much more significant, to those identified comparing gene expression profiles across positional-identity genes in dermal fibroblasts [17].

We next used the gene expression data as a functional readout of the differentially methylated loci identified between fibroblasts cultured from scalp and dura. The majority of significant DMPs (76,971/101,989, 75.47%) were inside or near (within 5kb of) a UCSC annotated gene, and 28.2% (21,742/76,971) were significantly associated with gene expression levels (at $p < 0.05$). This percentage of DMPs with significant expression readout was elevated (34.9%) among those DMPs with larger DNAm differences by sampling location (greater than 10% difference in DNAm levels). These DMPs were strongly significantly enriched among the CpG sites that associated with expression levels at the $p < 0.05$ (48,062 probes within 5kb of genes, odds ratio, OR = 3.99, $p < 2.2 \times 10^{-16}$) and FDR < 0.05 (6,559 probes within 5kb of genes, OR = 19.54, $p < 2.2 \times 10^{-16}$) significance thresholds. Surprisingly, we found that the DNAm levels at the majority of these expression-associated differentially methylated CpGs tended to be positively associated with gene expression, regardless of overall methylation levels (un-, partially-, or highly-methylated) or their location in the gene (islands, shores and shelves)—these biases towards positive associations were statistically significant for many of these comparisons (see S5 Table, panels A and B). We hypothesize these positive correlations could be due to the probe design of the Illumina 450 (the majority of probes are in lowly methylated regions) combined with the majority of genes having low expression (38.75% had mean FPKMs < 1).

We identified similar associations using transcript-level expression data using the Sailfish program [33] (see [Methods](#)) on the above transcriptome— 76.5% (77,981/101,989) of the DMPs were within 5 kb of a transcript, and 30.4% of them (23,672/77,981) correlated with expression (at $p < 0.05$). 33,247 unique transcripts overlapped or were within 5 kb of DMPs, and of them, 27.0% (8,981/33,247) exhibited significant correlation between DNAm and expression (at $p < 0.05$). The 33,247 transcripts proximal to the DMRs corresponded to 18,699 genes, the majority of which (84.3%, 15,761/18,699) contained more than one transcript. Interestingly, these associations often appear in a transcript-specific manner—6,190 genes (39.3%) had ≥ 1 transcript with significant correlation between DNAm and expression (at $p < 0.05$), with ≥ 1 transcripts that were not associated with nearby CpG levels. These results suggest that genes, and their underlying transcripts, can functionally validate many of the differentially methylated CpGs for sampling location.

Moving beyond individual CpGs, 587/697 (84.2%) DMRs were in or near (<5kb) genes, and many had DNAm levels that were significantly associated with gene expression levels (306/587, 52.1% at $p < 0.05$). For instance, a DMR overlapping an intronic sequence of the *SIM1* gene (Fig 2A) was unmethylated with low corresponding expression of the gene in the cultured fibroblasts from dura, and highly methylated with corresponding high expression levels of the gene in the scalp-derived fibroblasts (Fig 2B and S2 Table). This is in line with previous reports suggesting that gene body methylation levels positively associate with local gene expression [34], unlike CpG island shore methylation that tends to be negatively associated

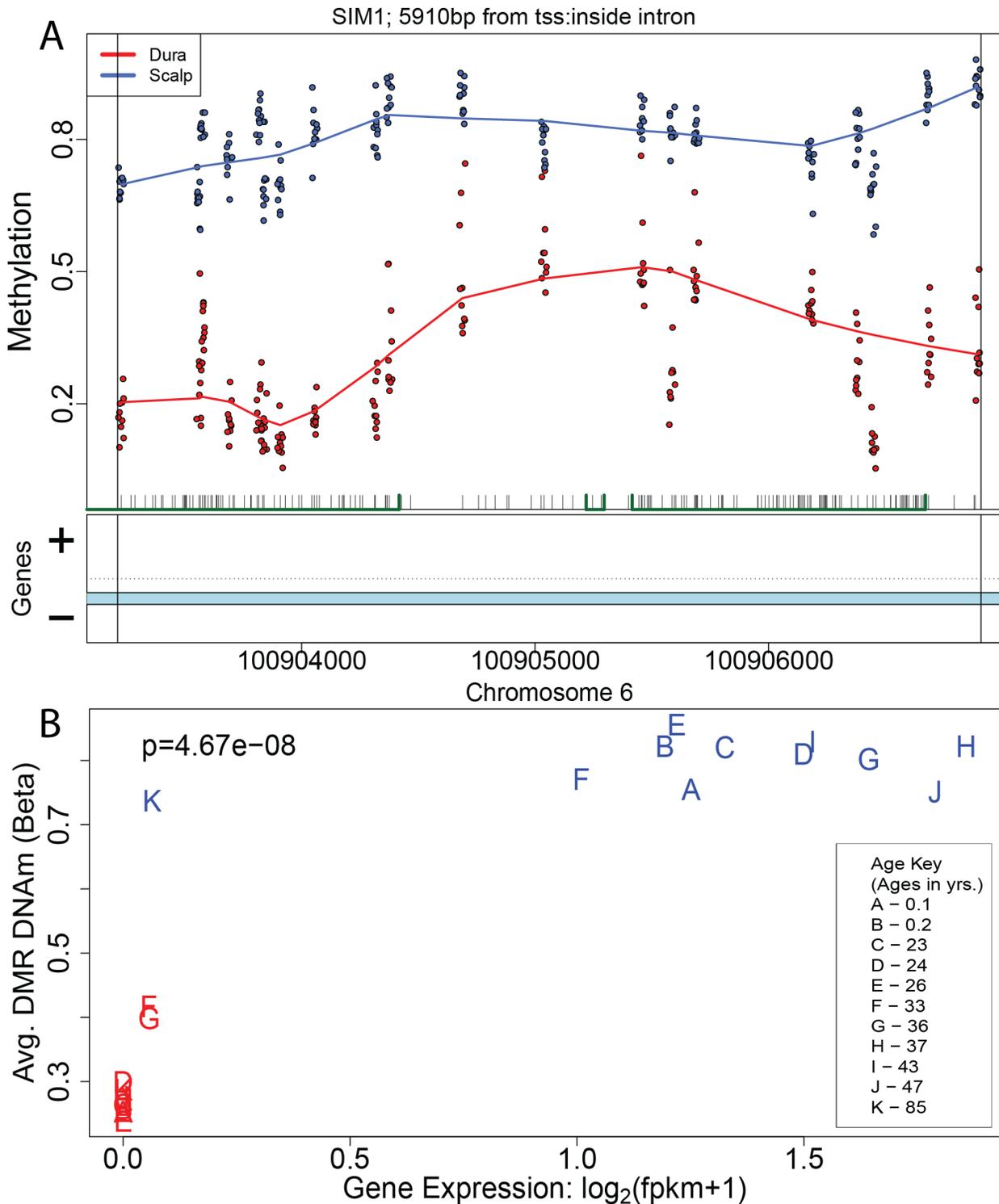


Fig 2. Regional DNA methylation changes manifest in the transcriptome. (A) Plot of the DNAm levels (proportion methylation) of an example significant DMR, which overlaps the gene *SIM1*. (B) Plot of the average DMR DNAm levels versus the expression level of *SIM1*, showing high positive correlation ($p = 4.67 \times 10^{-8}$).

doi:10.1371/journal.pgen.1005819.g002

with gene expression levels [1]. Of the 478 unique genes in or within 5kb of DMRs, the expression of 235 (49.2%) of them was significantly correlated with DNAm ($p < 0.05$). These 235 unique genes tended to exhibit stronger differential expression between the scalp- and dura-derived fibroblasts (median fold change = 1.59, IQR = 1.23–2.68) than individual CpG results, in line with previously published findings [23]. GO analysis on expression-associated genes proximal to DMRs revealed enrichment for multiple important biological processes such as connective tissue development, epithelium morphogenesis and development, cell differentiation (specifically including epithelial cell differentiation), and cell proliferation (including epithelial cell proliferation), among other more general sets (all $p < 10^{-8}$, see S6 Table). Unlike at the single CpG-level, we found that the majority of DMRs in and around the transcriptional start sites of genes (CpGs islands and shores) were negatively correlated with gene expression (S5 Table), in line with previous research [1]. We observed similar methylation-expression associations using transcript-level expression measurements— 312/599 DMRs (52.1%) near ≥ 1 transcripts associated with expression, and like at the single CpG level, found evidence for transcript-specific epigenetic regulation of expression (among 28.9% of genes containing multiple transcripts and associated with DNAm levels within the DMRs).

Lastly, we found that the majority of differentially methylated blocks contained at least one gene and transcript differentially expressed between scalp- and dura-derived fibroblasts. The majority of blocks contained at least one gene ($N = 188/243$, 77.4%); 63.8% ($N = 120/188$) had at least one gene and 66.66% ($N = 124/186$) at least one transcript that was differentially expressed (at $p < 0.05$).

As a representative example, one of the blocks, hypermethylated in scalp-derived fibroblasts, overlaps the *HOXB* gene cluster (Fig 3A), which has previously been shown to play a role in the position identities of fibroblasts [17]. In this block, expression levels of the *HOXB* genes are significantly greater in fibroblasts cultured from scalp than those from dura (Fig 3B), which contrasted previous microarray-based data showing these genes were not expressed in dermal samples taken from the head [17] highlighting the improved precision of RNA-sequencing data to quantify expression levels. Similarly, the 188 significant blocks contained 298 unique genes, and 126 of them (42.3%) were differentially expressed (at $FDR < 0.05$) which is a higher proportion than the rest of the transcriptome (0.42 vs. 0.32, $p = 3.79 \times 10^{-9}$).

Given the strong association between DNAm levels and local expression levels, we sought to more fully examine the epigenetic states of these sampling location-associated DNAm differences. We downloaded chromatin state data (18 states) from the NIH Roadmap Epigenomics Consortium on the four available fibroblast samples (2 primary foreskin, 1 adult dermal, and 1 lung) [35], and mapped our DMPs, DMRs, and blocks for fibroblast sampling location onto these states (S7 Table). The CpGs differentially methylated by sampling location were largely enriched for enhancer chromatin states, including preferential enrichment of genic (EnhG2) and active (EnhA1) enhancer states and depleted for active transcriptional start site (TSS) states (TssA). At the region level, DMRs were largely enriched for bivalent TSS (EnhBiv) and repressive polycomb (ReprPC) states and depleted for transcription (Tx) genic enhancer (EnhG2) states, and blocks were strongly enriched for quiescent (Quies) and heterochromatin (Het) states and depleted for transcriptional states. These enrichments were relatively conserved across the four Roadmap fibroblast samples, further suggesting distinct epigenetic states in scalp- compared to dura-derived fibroblasts. These results suggest that epigenetic memory related to original cell location manifests in genomic state differences and largely reads out in the transcriptome, particularly among regional changes in DNAm related to fibroblast sampling location.

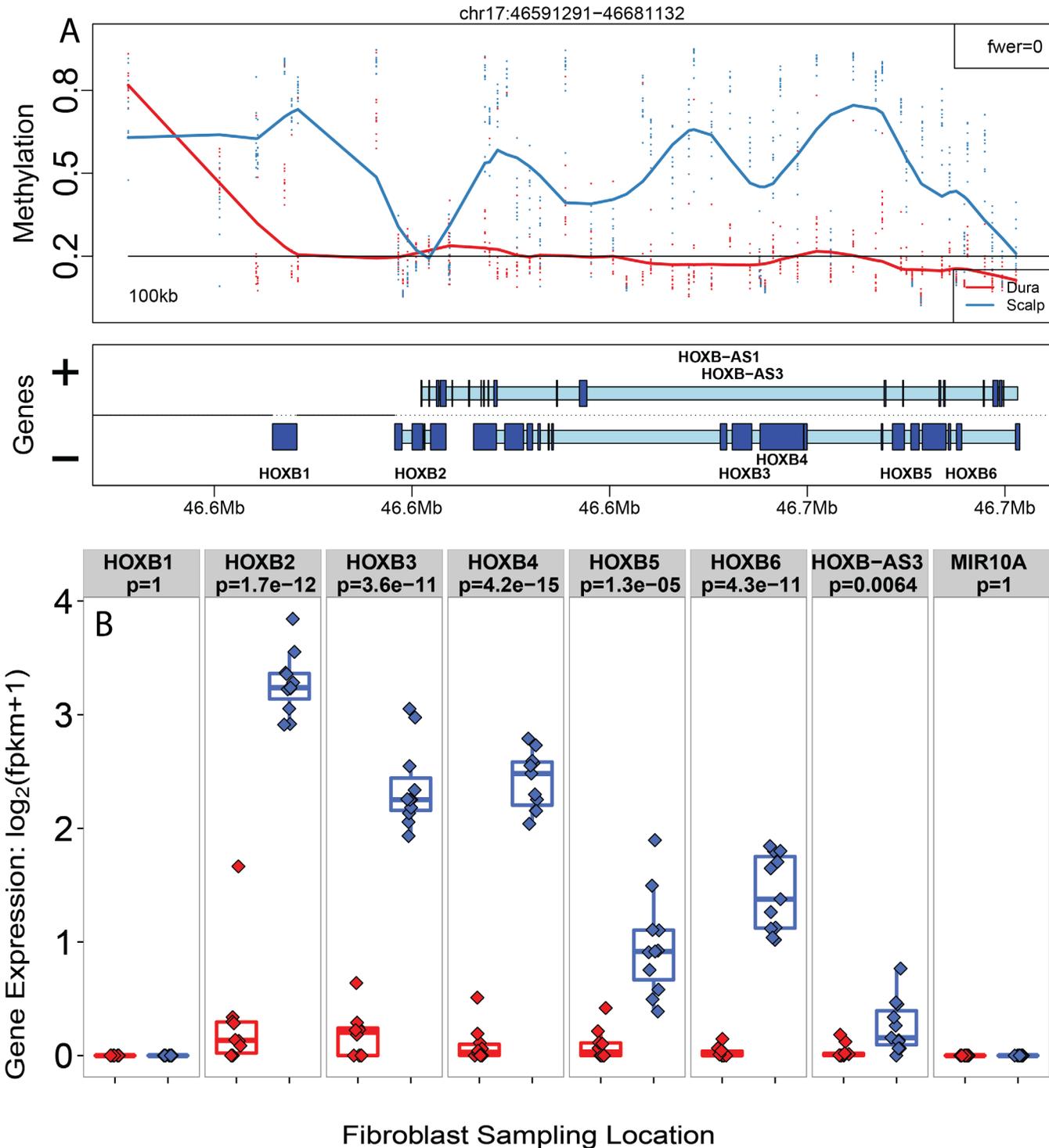


Fig 3. Long-range DNA methylation changes manifest in the transcriptome. (A) Plot of the DNAm levels (proportion methylation) of a significant DNAm block overlapping genes in the *HOX* family. Y-axis: proportion DNAm levels, x-axis: genomic coordinates on chromosome 17. (B) Corresponding expression levels of the *HOX* genes within the DNAm block are more highly expressed in the scalp. Y-axis: \log_2 transformed fragments per kilobase per million mapped (FPKM), x-axis: sampling location.

doi:10.1371/journal.pgen.1005819.g003

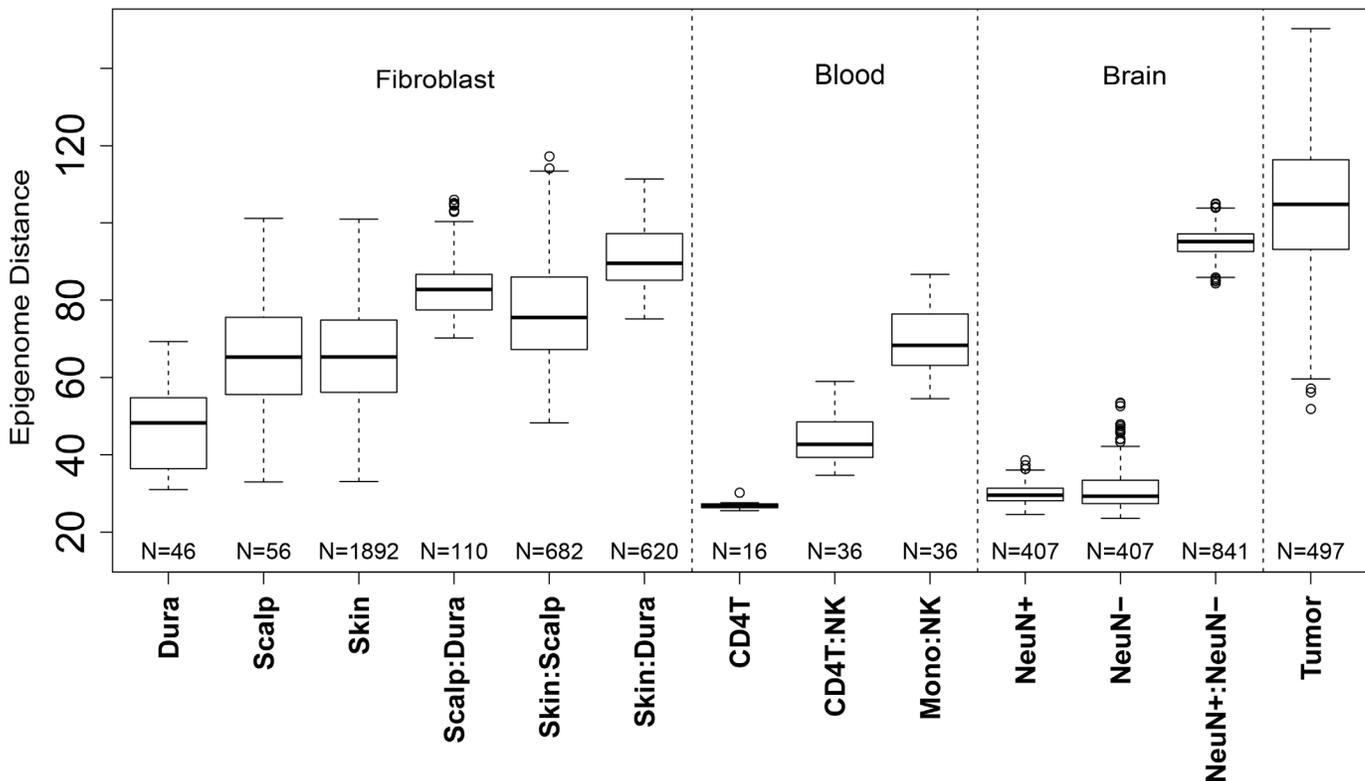


Fig 4. Increased methylome distances within scalp-derived fibroblasts. Y-axis: methylome (Euclidean) distance between pairs of samples stratified by cell and tissue types. CD4T: CD4+ T-cell, NK: natural killer cell, Mono: monocyte, NeuN+: neuronal DLPC cell, NeuN-: non-neuronal DLPC cell.

doi:10.1371/journal.pgen.1005819.g004

Increased stochastic variability in scalp-derived fibroblasts

We hypothesized that scalp-derived fibroblasts might have more variable levels of DNAm than dura-derived fibroblasts, given the chronic exposure to environmental factors (e.g. sunlight, chemicals) in the primary tissue across the lifespan. At the individual CpG level, we tested for differences in variance between the scalp- and dura-derived fibroblasts independent of the underlying mean methylation levels [36] (see [Methods](#) section). While only two probes reached genome-wide significance (at FDR < 0.05) for differences in variance, at marginal levels of significance ($p < 0.05$), fibroblasts cultured from scalp had more variable DNAm levels than fibroblasts cultured from dura ($N = 13,169/16,330, 80.6\%$).

We next sought to characterize methylome-wide patterns of DNAm across these fibroblasts in the context of other diverse cell types. After downloading and normalizing Illumina 450k data from sorted blood [20] and frontal cortex [21], as well as skin-derived fibroblasts [19] and melanoma samples (SKCM) from the Cancer Genome Atlas (TCGA) [37], we computed methylome-wide Euclidean distances between and across each of the 11 cell types (see [Methods](#) section). We noted that these cell types largely cluster by tissue source (brain, blood, and fibroblasts) in the first two principal components and largest dendrogram splits, [S6 Fig](#).

The inter-individual epigenomic distances, and their variability, were much greater in the scalp-derived (as well as skin-derived) fibroblasts than dura-derived fibroblasts ($p = 1.34 \times 10^{-9}$ and $p = 1.77 \times 10^{-14}$ respectively, see [Fig 4](#)). The distances within scalp- and skin-derived fibroblasts were significantly larger than those calculated within pure blood and cortex cell types (p -values range from 1.04×10^{-21} to $< 10^{-100}$). Interestingly, the inter-individual distances between fibroblasts cultured from scalp samples were greater than the distances between different cell

types within a blood cell lineage (e.g. natural killer cells versus CD4+ T-cells) which were previously suggested for different dermal fibroblasts [16] and instead more similar to distances across lineage (e.g. natural kill cells versus monocytes). Note that comparing inter-individual distance between two cell types (e.g. scalp- versus dura-derived fibroblasts) reflects the extensive differential methylation between these two cell types (see Fig 1)—the inter-individual distances are large but the variability in distances was low.

As another example, the distances across scalp-derived fibroblasts were lower than the inter-individual variability between neurons and non-neurons (via NeuN+ sorting), which reflects the extensive methylation differences between these two cell types [21]. As expected, we found the greatest methylome-wide distances and largest inter-individual variability in the melanoma samples [28,36], which highlights the relative scale of these methylome-wide distances (ranging from pure cell types to cancer). These increased epigenomic distances may relate to the rate of cell division, which is non-existent in neuronal cells [38] and infrequent in T-lymphocytes at the population level [39]. The increased epigenetic variability in the scalp samples was further not associated with differences in donor age ($p > 0.05$, S7 Fig), suggesting increased epigenetic stochastic variability in scalp- (and skin-) derived fibroblasts.

Epigenetic memory related to donor age

We hypothesized that a subset of this increased variability might result from age-related divergence in DNAm at individual loci that were differential by sampling location, such that young donors would have lesser difference in DNAm levels, and older donors would have larger differences in DNAm. By fitting linear models on the difference in DNAm levels across sampling location as a function of donor age (see Methods), we identified 7,265 CpGs associated with diverging DNAm levels across aging (at $FDR < 10\%$, S8 Fig). These loci appeared to be clustered into representative patterns of their age-related changes (Fig 5). The majority of these CpGs had significant age-related changes in fibroblasts derived from the scalp (64.0%), but not dura (17.4%), and the magnitude of change across age was larger in scalp-derived fibroblasts—the average change in percent DNAm per decade of life was 3.13% (IQR = 1.81%-4.29%) in fibroblasts derived from scalp compared to 1.13% (IQR = 0.295%-1.61%) in those from the dura mater.

A subset of these CpGs showing sampling location-dependent age-related changes associated with nearby gene expression levels. Most of the probes ($N = 5,185/7,265$, 71.4%) were annotated to 3,553 unique genes (within 5kb) and 21.8% of these ($N = 775/3,555$) showed significant correlation between DNAm and gene expression ($p < 0.05$). These DNAm associated genes were enriched for multiple general developmental processes including cell development, morphogenesis, and differentiation (all $p < 10^{-8}$, S8 Table). Several of the age-related CpGs showing expression association were within genes that are involved in cell proliferation and apoptosis. For instance, DNAm levels at two significant probes inside the gene *TEAD1*, which regulates notochord development and cell proliferation [40], were significantly associated with gene expression levels ($p = 8.60 \times 10^{-4}$ and 0.045, respectively). Another significant DNAm-expression pair ($p = 0.02$) involved *AVEN*, a gene shown to inhibit Caspase activation in apoptosis [41]. Interestingly, while we identified a large number of age-related CpGs, “DNA methylation ages” [8] were very similar to the chronological ages of the samples (see Methods and S9 Fig)—these associations did not differ by sampling location ($p = 0.72$) and there was further no association between “DNA methylation age” and sampling location alone ($p = 0.96$). The age-associated CpGs identified here therefore suggest that altered regulation of DNAm levels across aging occurs primarily in fibroblasts derived from scalp but not from dura, perhaps through altered cell proliferation and apoptosis, and possibly reflecting greater exposure to environmental agents that can affect the methylome.

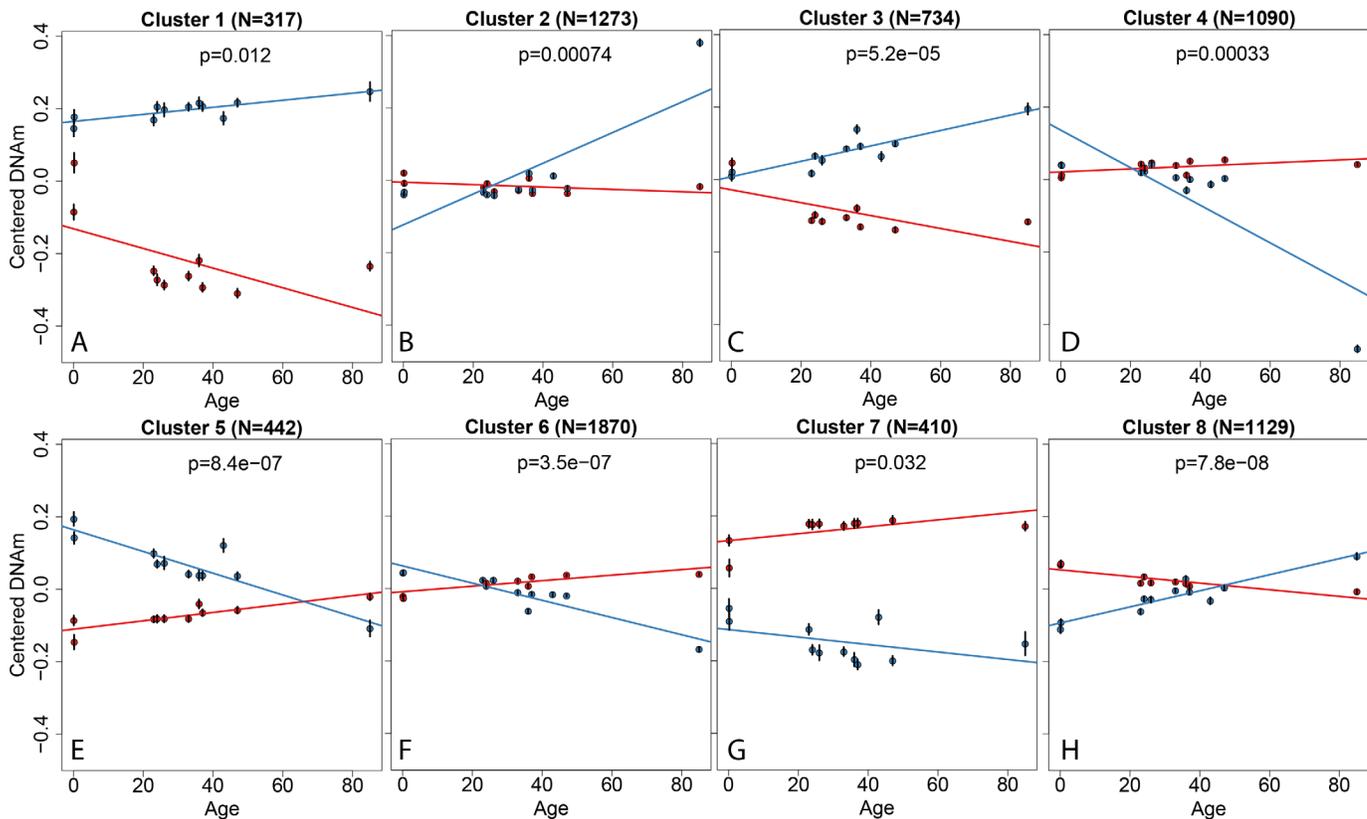


Fig 5. Representative patterns of age-associated changes in DNAm by sampling location. Each panel (A-H) shows mean adjusted expression levels versus age for each of eight clusters of location-specific age-related changes in DNAm levels. Y-axis: mean-centered DNAm levels, x-axis: sample age, p-value represents the statistical interaction between sampling location and age on DNAm levels. N: number of CpGs in the cluster. Vertical lines at each sample indicate ± 3 times the standard error by cluster.

doi:10.1371/journal.pgen.1005819.g005

Epigenetic memory related to sampling location and age do not implicate genetic mosaicism

Lastly, we characterized the expressed sequences of the scalp- and dura-derived fibroblasts within each individual to examine the extent of genetic mosaicism, which may contribute to differences in DNAm through changing the underlying genetic sequence in the fibroblasts taken from scalp. *De novo* variants were called directly from the RNAseq data, and after filtering by many quality metrics (see [Methods](#)) we identified 64 high-confidence candidate variants that were discordant by sampling location in at least a single individual ([S9 Table](#)), including 22 annotated coding variants (13 synonymous and 9 non-synonymous) [42]. We found no association between coding variant burden and subject age ($p = 0.71$, [S10 Fig](#)). These results suggest that many of the location- and age-associated DNAm differences are not due to somatic mosaicism and likely arise through epigenetic mechanisms that are maintained through cell culture and multiple passages.

Discussion

Here we interrogated the methylomes and transcriptomes of pairs of fibroblasts cultured from scalp and dura mater taken from the same individual, in a subject cohort that ranges in age across the human lifespan. These cultured fibroblasts, generations removed from the primary tissue of origin, and with indistinguishable morphology, still maintained strong components of

epigenetic “memory” related to sampling location (scalp versus dura) and differential changes in DNAm levels across aging. The widespread differences in DNAm levels by sampling location were identified at many spatial scales, including single CpGs, differentially methylated regions, blocks, and globally. Furthermore, many of these differences in DNAm levels manifested in the transcriptome, showing significant corresponding differences in expression for genes most proximal to these epigenetic changes. The genes with differences in expression and DNAm levels by sampling location were previously implicated in processes relating to cell proliferation and apoptosis, which likely relate to the function of the fibroblasts in the primary tissue. One might have predicted this outcome, as fibroblasts in the scalp, including those that are cultured, turnover much more rapidly than those in the dura mater [15], which we confirmed here with increased FSP-1 expression in the scalp-derived fibroblasts.

Another component of epigenetic memory in these cultured fibroblasts was related to ages of the donors, where age-related changes occurred differentially by sampling location. These age-associated loci can be clustered into general patterns of epigenetic changes by age and location, all showing significant interaction between donor age and sampling location. While some patterns were expected, such as divergence in DNAm levels from similar levels at birth (clusters 1, 4, 5, and 7), several other clusters showed an unexpected convergence in DNAm across aging (clusters 2 and 3). We do note that the elderly donor (age 85) is influential in both the statistical discovery at individual loci and in some of the subsequent clusters—larger sample sizes can hopefully further define and replicate these observations. Also, while the fibroblasts were analyzed from some subjects with psychiatric disorders, almost all comparisons between scalp and dura sampling locations, and differential changes with age were naturally matched within an individual, reducing the potential impact of diagnostic confounding. Furthermore, a larger sample size would likely identify significant age-related divergence in DNAm at the region level—while we found 7,265 individual CpGs, we found very few DMRs at global significance (6 and 11 DMRs at FWER \leq 10% and 20% respectively). The region-finding approach has been shown to be statistically conservative [25] and the identification of these differential age-related changes by sampling location was based on number of donors ($N = 10$), not the number of observations ($N = 21$). Lastly, while proliferation rates were not measured for these particular fibroblast samples, analyses in a much larger skin biopsy sample ($N = 298$) showed no association between proliferation rates and donor age [43], which was our sampling location with the greater number of age-related changes in DNAm levels.

These age-related changes in cultured fibroblasts are one of the first examples, to our knowledge, of genome-wide significant age-related changes in a pure cell population that is many mitoses and passages from the original donor cells. Many papers have identified widespread age-related changes in heterogeneous cell populations, like blood [5,7], brain [44], and other tissue types [8], which may result in false positives when the underlying cellular composition changes across aging [4]. Other papers have used individual cell populations to validate age-associated loci identified in homogenate tissue at marginal significance [45] or have identified age-related changes in targeted approaches at limited numbers of loci [46].

Similarly, these fibroblasts cultured from the scalp and dura mater were the first example, again to our knowledge, of morphologically indistinguishable cells with vastly different epigenomic profiles. Using epigenomic distances, these two cohorts of fibroblasts were more different in their DNAm patterns than different lineages of blood cells, while less different than neuronal versus non-neuronal cells from the frontal cortex (Fig 4); the cells underlying each comparison have very different morphologies and cellular function. Furthermore, the majority of differences in DNAm levels between scalp- and dura-derived cultured fibroblasts appeared to be determined early in development, prior to early infancy in this sample, and remained stable throughout the lifespan. Of the 101,989 significant DMPs for sampling location, 98,461

(96.5%) were not associated with differential age-related changes. These findings demonstrate strong components of epigenetic memory related to cell location and aging in fibroblasts cultured from the scalp and dura mater from postmortem human donors.

There are important implications from this study for the field of regenerative medicine. If fibroblasts are going to be the source for iPSCs, and ultimately differentiated tissues, the source of these fibroblasts, and their epigenetic characteristics, may be an important consideration. For example, these differences in cellular states in cultured fibroblasts may relate to the number of cell divisions, as skin and scalp fibroblasts have a much quicker turnover than fibroblasts in the dura [15]. The extent of cell division could relate to the epigenomic distances between and across the diverse cell types we have analyzed. Analyses in larger samples of skin biopsy-derived fibroblasts suggest that while donor age does not appear to associate with proliferation rates of fibroblasts, the cultured cells derived from younger donors reprogrammed more readily [43], which presumably has a strong epigenetic component. Further research may better determine the extent of epigenetic memory of cell state of fibroblasts cultured from different locations after the generation of iPSCs and subsequent differentiation into new cell types. As the field of regenerative medicine advances, our study demonstrates that deciding upon the source of fibroblasts from an individual to generate new tissues and organs may be an important consideration. While it was shown that transcriptional variability by tissue of origin was low in iPSCs (13), it was also demonstrated that the DNAm landscape in iPSCs differs greatly by tissue or origin, and this phenomenon may explain the propensity of iPSCs derived from different somatic tissues to differentiate into different lineages (11).

Methods and Materials

Human tissue collection

Human dural and scalp fibroblasts on which the methylation and gene expression studies were performed were obtained from fibroblast lines derived from human post mortem scalp and dura mater tissues. For this study, tissues from 11 individuals were used, with the ages of individuals ranging from 0.1 to 85 years of age (see [S1 Table](#) for additional demographics). The post-mortem tissues from 2 of the subjects were collected by the Lieber Institute for Brain Development (LIBD) and the tissues from the remaining 9 subjects were collected by National Institute for Mental Health (NIMH) (Clinical Brain Disorders Branch (CBDB), Division of Intramural Research Programs (DIRP)). The NIMH tissues were collected from two medical examiners (Washington, DC office and Commonwealth of Virginia, Northern District office); the LIBD tissues were obtained the Office of the Chief Medical Examiner (Baltimore, MD). A preliminary neurological or psychiatric diagnosis was given to each case after demographic, medical, and clinical histories were gathered via a telephone screening on the day of donation. For each case, the postmortem interval (PMI) (the time (in hours) elapsed between death and tissue freezing) was recorded. (See [S1 Table](#) for PMIs and demographics for every subject used in this study). Every case underwent neuropathological examinations to screen for neurological pathology. Additionally, the medical examiner's office performed toxicology analysis of every subject's blood to screen for drugs.

Dura and scalp tissue were collected at the time of autopsy. From the autopsy room, the tissues were transported in separate bags: one containing cerebral dura mater and the other a 1 in X 1 in scalp segment with hair attached. Both bags were transported on wet ice to the lab, where the culture procedure was immediately started.

Scalp and dura tissue cultures

The dura culture medium was prepared out of 1X DMEM (Ref#11960–044, GIBCO) with 10% by volume fetal bovine serum, 2% by volume 100X GlutaMAX (Cat#: 35050, GIBCO), 1% by

volume Penicillin-Streptomycin/Amphotericin solution (Ref# 15140–122, GIBCO), and 1% by volume Gentamicin solution (Cat# 17105–041, Quality Biological). This culture medium was used in all subsequent steps of the dura culturing procedure. The scalp culture medium used for all subsequent steps of the scalp culturing procedure was made the same way except without the 1% Gentamicin. A rinsing solution was prepared out of 1X PBS (pH 7.2) (Ref# 21-040-CV, Corning Life Sciences), 1% by volume Penicillin-Streptomycin/Amphotericin solution (Ref# 15140–122, GIBCO), and 1% by volume Gentamicin (Cat# 17105–041, Quality Biological).

The dissected scalp sample was washed with the rinsing solution three times, the fat tissues were cut away, and all hair was plucked out with forceps. The scalp sample was then placed epidermis side down on a dish and floated with Dispase II enzyme solution (2.4 units of the Dispase II enzyme per mL of PBS, Dispase II enzyme: Cat#17105–041, GIBCO). (Dispase II enzyme is a proteolytic enzyme used to separate the dermis from the epidermis by cleaving the zone of the basement membrane.) The dish was covered with parafilm and foil, and placed in a 37°C incubator for 24 hours. After the 24-hour period, the epidermis was peeled away from the dermis. The dermis was washed with the rinsing solution, dried, and cut into 2–3 mm² pieces. The pieces were placed in a Falcon Easy Grip tissue culture 35×10 mm dish and one drop of scalp culture medium was added to each piece of scalp. The dish was placed in the incubator at 37°C and 5% CO₂ for culturing.

A similar procedure was followed for the dura samples. Dura samples were washed with the rinsing solution three times. Then, a few 2–3 mm² pieces were cut from the dura mater and placed together in an Easy Grip cell culture 35×10 mm dish. One drop of dura culture medium was added to each dura piece. The culture dish was then placed in an incubator (at 37°C and 5% CO₂) for culturing. The medium of each culture was changed to fresh medium 2–3 times per week to promote growth of the fibroblasts. On average, fibroblast cells started to proliferate at 7–14 days, however some samples took longer (up to 3 weeks).

Fibroblast cell cultures

The dura and scalp tissue cultures were monitored under a phase-contrast microscope. When the fibroblast growth reached 90–95% confluence, 1 mL of a 0.25% trypsin solution (Cat#T4049, Sigma) was added to each culture dish, and the cells were incubated for 5 to 8 min. Then, 1 mL of media was added to each dish stop the enzymatic reaction. Next, the contents of each culture dish were transferred into separate 15 mL Falcon conical tubes and 8 mL of media was added to each tube. The conical tubes were centrifuged for 5 min at 1100 rpm. The supernatant was discarded, 5 mL of fresh media was added to each conical tube, and the contents of the tubes were transferred onto separate 25 cm³ cell culture Easy Flasks (Thermo Scientific, Cat# 156367), where they were kept in cultures for 3–5 days in an incubator (at 37°C and 5% CO₂). When the cells reached 90–95% confluence, the cells from each 25 cm³ flask were transferred onto two 75 cm³ cell culture easy flasks (Thermo Scientific, Cat# 156499) and kept in cultures for continued growth. When the cells reached 90–95% confluence, they were incubated with 3 mL of 0.25% trypsin solution for 5 to 8 min, after which 3 mL of fresh culture media was added to stop the enzymatic reaction. Then, the contents of the flasks were transferred into separate 15 mL Falcon conical tube and 4 mL of media was added to each tube. The tubes were centrifuged (5 min, 1100 rpm), the supernatant was discarded and the pellets containing the fibroblasts were removed from the centrifuge tubes and transferred to cryoTube vials (Cat#375418, Thermo Scientific). 0.5 mL of recovery cell culture freezing medium (Cat#12648–010, GIBCO) was added to each vial, after which the vials were insulated with Styrofoam and placed into a -80°C freezer. Later, the tubes were transferred to a -152°C liquid nitrogen freezer.

These frozen dura and scalp fibroblast cells were then used generate DNA methylation and gene expression levels. Genomic DNA was extracted from approximately 3 million cultured human fibroblast cells using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). Bisulfite conversion was performed on 600 ng genomic DNA was done with the EZ DNA methylation kit (Zymo Research).

DNA methylation data generation

DNA methylation landscapes of the dura- and scalp-derived fibroblasts were analyzed using the Illumina HumanMethylation 450 BeadChip array (“450k”). The 450k array interrogates >485,000 DNA methylation sites (probes) and measures the proportion DNA methylation at each target site (the 450k array interrogates both CpG and CH sites). The microarray preparation and scanning were performed in accordance with the manufacturer’s protocols. The resulting data from the 450k consists of R(ed) and G(reen) intensities using two different probe chemistries [22], which we converted to M(ethylated) and U(nmethylated) intensities using the *minfi* Bioconductor package [23], version 1.14.0 using with R version 3.2. One dura sample had lower median probe intensities and was removed prior to normalization and downstream analyses. After quality control (QC), the M and U intensities were normalized separately across samples using stratified quantile normalization [23]. Probes containing common SNPs (based on dbSNP 142) at the target CpG or single base extension site, and probes on the sex chromosomes were removed, leaving 456,513 probes on 21 samples for analysis.

Differential methylation analysis

We determined differential methylation using linear modeling on the normalized DNAm levels, using the model:

$$y_{ij} = \alpha_i + \beta_i Loc_j + \zeta_i SVs_j + \epsilon_{ij} \quad (1)$$

where y_{ij} is the normalized proportion methylation at probe i and sample j , α_i is the proportion methylation in the fibroblasts sampled from the dura mater, β_j is the difference in methylation in the scalp-derived fibroblast, and Loc_j is the sampling location represented by a binary variable (Dura = 0, Scalp = 1). These statistical models were adjusted for surrogate variables (6 SVs) estimated using surrogate variable analysis (SVA) [24].

Differentially methylation probes (DMPs) were identified by fitting Eq 1 to each probe, and obtaining the corresponding moderate t-statistic and p-value using the *limma* package [47]. P-values were adjusted for multiple testing using the false discovery rate (FDR) [48] and significant probes were called were $FDR < 0.05$. Principal component analysis (PCA) was performed after regressing out the surrogate variables from the DNAm levels of each probe, preserving the effect of fibroblast sampling location. Finding differentially methylated regions (DMRs) involves identifying contiguous probes where $\beta \neq 0$ using the *bumphunter* Bioconductor package (version 1.6.0) [25], here requiring $|\beta| > 0.1$ (argument: cutoff = 0.1) and assessing statistical significance using linear modeling bootstrapping with 1000 iterations (argument: null-Method = ‘bootstrap’ and B = 1000). DMRs were called statistically significant when the family wise error rate (FWER) ≤ 0.1 . We identified blocks using the same model as above using the *blockFinder* function in the *minfi* package [23], which collapses nearby CpGs into a single measurement per sample, and then fits Eq.1 above, only here j represents probe group, not probe. Here we again required at least a 10% change in DNAm between groups and assessed statistical significance using the FWER based on 1000 iterations of the linear model bootstrap.

RNA extraction and sequencing

RNA was extracted from the cultured dura and scalp fibroblasts with the RNeasy kit (Qiagen), in accordance with the manufacturer's protocol. RNA molecules were treated with DNase, polyadenylated (polyA+) RNA was isolated, and resulting sequencing libraries were constructed using the Illumina TruSeq RNA Sample Preparation Kit (v2) and sequenced on an Illumina HiSeq 2000. We note that while all samples were run on the same flow cell, the samples were somewhat imbalanced by lane—however, the first PC of the expression data did separate perfectly by sampling location. Sample-specific information on reads and alignments are available in [S1 Table](#).

RNA-seq data generation

Resulting reads were mapped to the genome using TopHat2 [31] using the paired-ends procedure (we used the option—library-type fr-firststrand). Gene counts relative to the UCSC hg19 knownGene annotation were calculated using the featureCounts script of the Subread package (version 1.4.6) [32]. There were 23,710 genes in this annotation, and we dropped 305 genes that were annotated to more than 1 chromosome. Of the remaining 23,405 genes, 18,316 genes had non-zero expression counts in at least one sample. Counts were converted to FPKM (fragments per kilobase per million reads mapped) values to allow comparisons across genes with different lengths and libraries sequenced to different depths. These FPKMs were transformed prior to statistical analysis: $\log_2(\text{FPKM} + 1)$. The log transformed FPKM values were used in all subsequent gene-level analyses.

Next, we used the Sailfish software (33), version 0.7.6, to quantify isoforms from our RNA-seq reads. As a result, we obtained TPM (transcripts per million) values for each isoform, which we log transformed: $\log_2(\text{TPM} + 1)$. The log transformed TPM values were used in all subsequent transcript-level analyses.

RNA-seq data analysis

Differential expression for sampling location was identified using [Eq 1](#) above, where y_{ij} represents transformed expression (rather than DNAm) levels, and different SVs ($N = 4$) were calculated from the expression data.

To test whether we could use a subset of genes to cluster our fibroblasts by sampling location, like reported by Rinn et al [17], we took the 337 genes published by the authors, which they found to group fibroblasts by anatomical location. Of these 337, we used only 210 genes, since a subset of the tabulated genes did not contain gene symbols, another subset was not interrogated by our RNAseq, and yet another subset was not expressed in any of our samples. We then performed Euclidean distance computations and clustering analysis by first using these 210 genes and then repeating the analysis 1000 times using 210 randomly chosen genes.

We carried out gene ontology analysis on the differentially expressed genes with the *GOstats* package [49]. Transformed FPKMs were next used to assess functional significance of differentially methylated features. We mapped the DMPs to genes in the UCSC knownGenes (hg19) and determined which DMPs exhibit correlation between DNAm and gene expression with the *MatrixEQTL* package [50]. We used Pearson's Chi-squared test with Yates' continuity correction to examine whether DMPs are more likely to exhibit correlations between DNAm and gene expression than non-DMPs. We then mapped significant DMRs to genes expressed in the RNA-seq data (e.g. showing non-zero expression levels in ≥ 1 samples), and correlated the average DNAm level within the DMR to the transformed expression level. When multiple genes were within or near a DMR, we retained the gene (and its correlation) with the largest absolute correlation. We carried out gene ontology analysis for the genes proximal to DMRs

with the *GStats* package. For each significant block, we found the UCSC annotated gene(s) containing within the block and their evidence for differential expression as calculated above. We used Pearson's Chi-squared test with Yates' continuity correction to test whether differentially expressed genes were enriched in blocks compared to the rest of the transcriptome. Finally, we analyzed the directionality of DNAm—expression correlations for DMPs and DMRs, as a function of DMR/DMP positions relative to genes. We used the binomial test to access the significance of distributions between positive and negative correlations of DNAm and gene expression.

In addition to gene-level analysis, we studied transcript-level expression and its correlation with DNAm. We carried out the same analysis for isoform expression as for gene-level expression, with the exception that here we used relative isoform abundance values that we obtained with the Sailfish software (see above).

Chromatin state analysis

The 18-chromatin state data, derived using hidden Markov models (HMMs), was obtained for 4 fibroblast samples: samples E055 and E056 (foreskin primary fibroblasts), E126 (adult dermal fibroblast), and E128 (lung fibroblasts) in the Epigenome Roadmap project22 (http://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html). The chromatin states overlapping DMPs, DMRs, and blocks were obtained, and compared to a background of all 450k probes, considered probe groups, and collapsed probe groups respectively. Overlap was assessed based on the total coverage (in base pairs) of the chromatin states. Fold changes for enrichment of > 1.5 fold were highlighted. Prior to carrying out the enrichment analysis, the sex chromosomes and the mitochondrial chromosome were dropped.

Processing of public data and distance calculations

We performed a second larger data processing and normalization procedure on our scalp- and dura-derived fibroblasts after adding data from skin fibroblasts (GSE52025) [19], pure populations of blood [20] and prefrontal cortex cells [21] from the *FlowSorted.Blood.450k* and *FlowSorted.DLPFC.450k* Bioconductor packages respectively, and then melanoma data from TCGA [37]. The M and U channels were combined across all experiments and then normalized with stratified quantile normalization as described above. We then dropped the probes on the sex chromosomes as well as probes that are common SNPs (based on dbSNP 142) as described above. Within the normalized data, we then calculated all pairwise Euclidean distances on the proportion methylation scale, and selected specific comparisons to display in Fig 4.

Differential variability and age related changes by tissue type

We calculated differential variability between scalp and dura CpG DNAm levels using the Levene test [51] and subsequent p-values were adjusted for multiple testing using the FDR. We filtered out the 101,989 genome-wide significant probes showing mean methylation differences by sampling location, as there is a strong mean-variation relationship in DNAm data due to being constrained within 0 and 1 (e.g. gaining methylation from an unmethylated state or losing methylation from a fully methylated state increases variance).

We tested for probes that showed differential age-related divergence in DNAm by fibroblast sampling location. First, we calculated the difference in DNAm between scalp- and dura-derived fibroblasts from the same individual at every probe (creating a 456,513 probe by 10 individual matrix). We then computed 3 surrogate variables (the number estimated by the

SVA algorithm) for a statistical model with donor age, and fit the following linear model:

$$\Delta y_{ij} = \gamma_i + \delta_i \text{Age}_j + \zeta \text{SVs}_j^1 + \varepsilon_{ij} \quad (2)$$

where Δy_{ij} is the difference in DNAm levels between scalp and dura for probe i and individual j , γ_i is the difference in DNAm levels at birth, Age_j is the age of the donor, and δ_i is the change in the difference of DNAm per year of life. We then generated a Wald statistic and corresponding p-value for δ_i and adjusted for multiple testing via the FDR. Post hoc age-related changes, e.g. the change in DNAm levels per year of life, were calculated within the scalp and dura samples. We then associated expression of nearby genes (within 5kb) with the DNAm levels at the probes showing significant age-by-location effects and performed gene ontology on the significant genes with the *GOstats* package [49]. We lastly computed the “DNAm age” of our scalp and dura samples using the R code published by S. Horvath, (available at <https://labs.genetics.ucla.edu/horvath/dnamage/>) and fit a linear model containing main effects of biological age and sampling location, and an interaction term between these two variables on “DNAm age”.

Variant calling

We called expression variants directly from the RNA sequencing alignments using samtools (version 1.1) and mpileup across all samples [52]. We then filtered variants in the resulting variant call format (VCF) file based on coverage (<20), variant distance bias ($p < 0.05$), read position bias ($p < 0.05$), mapping quality bias ($p < 0.05$), base quality bias ($p < 0.05$), inbreeding coefficient binomial test ($p < 0.05$), and homozygote bias ($p > 0.05$). The resulting 64 high quality variants were annotated with SeattleSeq138 [42].

Study approval

For every subject from whom the post-mortem tissues were collected, informed consent was obtained verbally from the legal next-of-kin using a telephone script, and was both witnessed and audiotaped, in accordance with the IRB approved NIMH protocol 90-M-0142 and the Department of Health and Human Services for the State of Maryland (protocol # 12–24).

Data availability

DNA methylation data in both raw and processed forms are available on the Gene Expression Omnibus (GEO): GSE77136. RNA sequencing reads (raw data) are available on the Sequencing Read Archive (SRA): SRP068304 (BioProject: PRJNA286856) and the genes and transcript counts (processed data) are available on GEO at the above accession number (GSE77136).

Supporting Information

S1 Fig. Experimental setup. We took dura and scalp samples from 11 donors ranging from 0.1 to 85 years of age. We then extracted and cultured fibroblasts from these samples, and performed genome-wide DNA methylation and RNA sequencing procedures on these fibroblasts. (PDF)

S2 Fig. DMR plots. DNA methylation levels (proportion methylation) of all 697 significant DMRs (FWER $< 10\%$). (PDF)

S3 Fig. DNA methylation “blocks” plots. DNA methylation levels (proportion methylation) of all 243 significant differentially methylated blocks (FWER $< 10\%$). (PDF)

S4 Fig. Principal component analysis plots. The first principal component (PC1) of the gene expression data plotted against fibroblast sampling location (scalp versus dura). The first PC of the gene expression data mimics the first PC of the DNAm data; both represent sampling location.

(PDF)

S5 Fig. Separation of fibroblasts by anatomical site of origin based on differential expression of a previously reported subset of genes. When we analyzed the expression of 210 genes (which were found to demarcate fibroblasts by anatomical site of origin (17)), our samples separated into categories by their sampling location. When these 210 genes were used, the mean Scalp-Dura Euclidean distance was 27.53; when we performed 1000 iterations taking random subsets of genes, the range of mean Scalp-Dura Euclidean distances was 7.58–14.67.

(PDF)

S6 Fig. Clustering analysis on DNAm data from cells of different tissues. (A) PC1 with respect to PC2 of the DNAm data from the following cells: various cells of the blood; neuronal (NeuN+) and glial (NeuN-) cells from the DLPFC; cultured fibroblasts derived from skin, dura mater, and scalp; cells from a primary solid skin tumor. (B) Cluster dendrogram constructed from the DNAm data from the cells in panel A.

(PDF)

S7 Fig. Epigenomic distance within scalp-derived fibroblasts with respect to age differences between subjects.

(PDF)

S8 Fig. Age related DNAm divergence. DNAm plotted with respect to age for all 7,265 CpGs significantly associated with diverging DNAm levels across aging (at FDR < 10%).

(PDF)

S9 Fig. “DNA methylation age” with respect to chronological age of the scalp- and dura-derived fibroblasts.

(PDF)

S10 Fig. Number of coding variants with respect to subject age.

(PDF)

S1 Table. Tissue donor demographics and RNAseq read alignment data.

(XLSX)

S2 Table. Information on significant DMRs (FWER < 10%).

(XLS)

S3 Table. DMR Gene Ontology. Gene Ontology on genes that overlap or are proximal to (within 5kb) of significant DMRs (FWER < 10%).

(XLS)

S4 Table. Gene Ontology on genes differentially expressed between scalp- and dura-derived fibroblasts (FDR < 5%).

(XLS)

S5 Table. Directionality of correlations between DNA methylation and gene expression.

(XLSX)

S6 Table. DMR Gene Ontology. Gene Ontology on genes that overlap or are proximal to DMRs (within 5 kb) and exhibit significant correlation between gene expression and DNAm

($p < 0.05$).
(XLS)

S7 Table. Fibroblast chromatin state coverage by features.
(XLSX)

S8 Table. Gene Ontology on genes whose expression is correlated with nearby diverging DNAm CpGs. Gene Ontology on genes that overlap or are proximal to (within 5kbs) of CpGs that exhibit location-dependent age-related changes (FDR < 10%) and demonstrate correlation between DNAm and expression ($p < 0.05$).
(XLS)

S9 Table. Candidate exonic variants between scalp- and dura-derived fibroblasts.
(VCF)

Author Contributions

Conceived and designed the experiments: JEK DRW TMH AEJ. Performed the experiments: RT JGC AB JDG. Analyzed the data: NAI AEJ. Contributed reagents/materials/analysis tools: JGC AB MIM JDG RDM YJ. Wrote the paper: AEJ NAI.

References

1. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41: 178–186. doi: [10.1038/ng.298](https://doi.org/10.1038/ng.298) PMID: [19151715](https://pubmed.ncbi.nlm.nih.gov/19151715/)
2. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, et al. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13: 86. doi: [10.1186/1471-2105-13-86](https://doi.org/10.1186/1471-2105-13-86) PMID: [22568884](https://pubmed.ncbi.nlm.nih.gov/22568884/)
3. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, et al. (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500: 477–481. doi: [10.1038/nature12433](https://doi.org/10.1038/nature12433) PMID: [23925113](https://pubmed.ncbi.nlm.nih.gov/23925113/)
4. Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 15: R31. doi: [10.1186/gb-2014-15-2-r31](https://doi.org/10.1186/gb-2014-15-2-r31) PMID: [24495553](https://pubmed.ncbi.nlm.nih.gov/24495553/)
5. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, et al. (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 49: 359–367. doi: [10.1016/j.molcel.2012.10.016](https://doi.org/10.1016/j.molcel.2012.10.016) PMID: [23177740](https://pubmed.ncbi.nlm.nih.gov/23177740/)
6. Johansson A, Enroth S, Gyllenstein U (2013) Continuous Aging of the Human DNA Methylome Throughout the Human Lifespan. *PLoS One* 8: e67378. PMID: [23826282](https://pubmed.ncbi.nlm.nih.gov/23826282/)
7. Yuan T, Jiao Y, de Jong S, Ophoff RA, Beck S, et al. (2015) An Integrative Multi-scale Analysis of the Dynamic DNA Methylation Landscape in Aging. *PLoS Genet* 11: e1004996. doi: [10.1371/journal.pgen.1004996](https://doi.org/10.1371/journal.pgen.1004996) PMID: [25692570](https://pubmed.ncbi.nlm.nih.gov/25692570/)
8. Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biol* 14: R115. PMID: [24138928](https://pubmed.ncbi.nlm.nih.gov/24138928/)
9. Dimos JT, Rodolfa KT, Niakan KK, Weisenthal LM, Mitsumoto H, et al. (2008) Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* 321: 1218–1221. doi: [10.1126/science.1158799](https://doi.org/10.1126/science.1158799) PMID: [18669821](https://pubmed.ncbi.nlm.nih.gov/18669821/)
10. Santos F, Hendrich B, Reik W, Dean W (2002) Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* 241: 172–182. PMID: [11784103](https://pubmed.ncbi.nlm.nih.gov/11784103/)
11. Kim K, Doi A, Wen B, Ng K, Zhao R, et al. (2010) Epigenetic memory in induced pluripotent stem cells. *Nature* 467: 285–290. doi: [10.1038/nature09342](https://doi.org/10.1038/nature09342) PMID: [20644535](https://pubmed.ncbi.nlm.nih.gov/20644535/)
12. Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, et al. (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 1: 55–70. doi: [10.1016/j.stem.2007.05.014](https://doi.org/10.1016/j.stem.2007.05.014) PMID: [18371336](https://pubmed.ncbi.nlm.nih.gov/18371336/)
13. Rouhani F, Kumasaka N, de Brito MC, Bradley A, Vallier L, et al. (2014) Genetic background drives transcriptional variation in human induced pluripotent stem cells. *PLoS Genet* 10: e1004432. doi: [10.1371/journal.pgen.1004432](https://doi.org/10.1371/journal.pgen.1004432) PMID: [24901476](https://pubmed.ncbi.nlm.nih.gov/24901476/)

14. Vandiver AR, Irizarry RA, Hansen KD, Garza LA, Runarsson A, et al. (2015) Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome Biol* 16: 80. doi: [10.1186/s13059-015-0644-y](https://doi.org/10.1186/s13059-015-0644-y) PMID: [25886480](https://pubmed.ncbi.nlm.nih.gov/25886480/)
15. Bliss LA, Sams MR, Deep-Soboslay A, Ren-Patterson R, Jaffe AE, et al. (2012) Use of postmortem human dura mater and scalp for deriving human fibroblast cultures. *PLoS One* 7: e45282. doi: [10.1371/journal.pone.0045282](https://doi.org/10.1371/journal.pone.0045282) PMID: [23028905](https://pubmed.ncbi.nlm.nih.gov/23028905/)
16. Chang HY, Chi JT, Dudoit S, Bondre C, van de Rijn M, et al. (2002) Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc Natl Acad Sci U S A* 99: 12877–12882. PMID: [12297622](https://pubmed.ncbi.nlm.nih.gov/12297622/)
17. Rinn JL, Bondre C, Gladstone HB, Brown PO, Chang HY (2006) Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet* 2: e119. PMID: [16895450](https://pubmed.ncbi.nlm.nih.gov/16895450/)
18. Bork S, Pfister S, Witt H, Horn P, Korn B, et al. (2010) DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging Cell* 9: 54–63. doi: [10.1111/j.1474-9726.2009.00535.x](https://doi.org/10.1111/j.1474-9726.2009.00535.x) PMID: [19895632](https://pubmed.ncbi.nlm.nih.gov/19895632/)
19. Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, et al. (2014) The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol* 15: R37. doi: [10.1186/gb-2014-15-2-r37](https://doi.org/10.1186/gb-2014-15-2-r37) PMID: [24555846](https://pubmed.ncbi.nlm.nih.gov/24555846/)
20. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, et al. (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* 7: e41361. doi: [10.1371/journal.pone.0041361](https://doi.org/10.1371/journal.pone.0041361) PMID: [22848472](https://pubmed.ncbi.nlm.nih.gov/22848472/)
21. Guintivano J, Aryee MJ, Kaminsky ZA (2013) A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* 8: 290–302. doi: [10.4161/epi.23924](https://doi.org/10.4161/epi.23924) PMID: [23426267](https://pubmed.ncbi.nlm.nih.gov/23426267/)
22. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, et al. (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6: 692–702. PMID: [21593595](https://pubmed.ncbi.nlm.nih.gov/21593595/)
23. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, et al. (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30: 1363–1369. doi: [10.1093/bioinformatics/btu049](https://doi.org/10.1093/bioinformatics/btu049) PMID: [24478339](https://pubmed.ncbi.nlm.nih.gov/24478339/)
24. Leek JT, Storey JD (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 3: 1724–1735. PMID: [17907809](https://pubmed.ncbi.nlm.nih.gov/17907809/)
25. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, et al. (2012) Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 41: 200–209. doi: [10.1093/ije/dyr238](https://doi.org/10.1093/ije/dyr238) PMID: [22422453](https://pubmed.ncbi.nlm.nih.gov/22422453/)
26. Bae SC, Choi JK (2004) Tumor suppressor activity of RUNX3. *Oncogene* 23: 4336–4340. PMID: [15156190](https://pubmed.ncbi.nlm.nih.gov/15156190/)
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25–29. PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)
28. Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, et al. (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43: 768–775. doi: [10.1038/ng.865](https://doi.org/10.1038/ng.865) PMID: [21706001](https://pubmed.ncbi.nlm.nih.gov/21706001/)
29. Timp W, Bravo HC, McDonald OG, Goggins M, Umbricht C, et al. (2014) Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med* 6: 61. doi: [10.1186/s13073-014-0061-y](https://doi.org/10.1186/s13073-014-0061-y) PMID: [25191524](https://pubmed.ncbi.nlm.nih.gov/25191524/)
30. Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP (2009) Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet* 41: 246–250. doi: [10.1038/ng.297](https://doi.org/10.1038/ng.297) PMID: [19151716](https://pubmed.ncbi.nlm.nih.gov/19151716/)
31. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14: R36. doi: [10.1186/gb-2013-14-4-r36](https://doi.org/10.1186/gb-2013-14-4-r36) PMID: [23618408](https://pubmed.ncbi.nlm.nih.gov/23618408/)
32. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930. doi: [10.1093/bioinformatics/btt656](https://doi.org/10.1093/bioinformatics/btt656) PMID: [24227677](https://pubmed.ncbi.nlm.nih.gov/24227677/)
33. Patro R, Mount SM, Kingsford C (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol* 32: 462–464. doi: [10.1038/nbt.2862](https://doi.org/10.1038/nbt.2862) PMID: [24752080](https://pubmed.ncbi.nlm.nih.gov/24752080/)
34. Maunakea AK, Chepelev I, Cui K, Zhao K (2013) Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res* 23: 1256–1269. doi: [10.1038/cr.2013.110](https://doi.org/10.1038/cr.2013.110) PMID: [23938295](https://pubmed.ncbi.nlm.nih.gov/23938295/)

35. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317–330. doi: [10.1038/nature14248](https://doi.org/10.1038/nature14248) PMID: [25693563](https://pubmed.ncbi.nlm.nih.gov/25693563/)
36. Jaffe AE, Feinberg AP, Irizarry RA, Leek JT (2012) Significance analysis and statistical dissection of variably methylated regions. *Biostatistics* 13: 166–178. doi: [10.1093/biostatistics/kxr013](https://doi.org/10.1093/biostatistics/kxr013) PMID: [21685414](https://pubmed.ncbi.nlm.nih.gov/21685414/)
37. Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W (2014) Genome-wide analysis of noncoding regulatory mutations in cancer. *Nat Genet* 46: 1160–1165. doi: [10.1038/ng.3101](https://doi.org/10.1038/ng.3101) PMID: [25261935](https://pubmed.ncbi.nlm.nih.gov/25261935/)
38. Yang Y, Herrup K (2007) Cell division in the CNS: protective response or lethal event in post-mitotic neurons? *Biochim Biophys Acta* 1772: 457–466. PMID: [17158035](https://pubmed.ncbi.nlm.nih.gov/17158035/)
39. Tough DF, Sprent J (1994) Turnover of naive- and memory-phenotype T cells. *J Exp Med* 179: 1127–1135. PMID: [8145034](https://pubmed.ncbi.nlm.nih.gov/8145034/)
40. Sawada A, Kiyonari H, Ukita K, Nishioka N, Imuta Y, et al. (2008) Redundant roles of Tead1 and Tead2 in notochord development and the regulation of cell proliferation and survival. *Mol Cell Biol* 28: 3177–3189. doi: [10.1128/MCB.01759-07](https://doi.org/10.1128/MCB.01759-07) PMID: [18332127](https://pubmed.ncbi.nlm.nih.gov/18332127/)
41. Chau BN, Cheng EH, Kerr DA, Hardwick JM (2000) Aven, a novel inhibitor of caspase activation, binds Bcl-xL and Apaf-1. *Mol Cell* 6: 31–40. PMID: [10949025](https://pubmed.ncbi.nlm.nih.gov/10949025/)
42. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461: 272–276. doi: [10.1038/nature08250](https://doi.org/10.1038/nature08250) PMID: [19684571](https://pubmed.ncbi.nlm.nih.gov/19684571/)
43. Paull D, Sevilla A, Zhou H, Hahn AK, Kim H, et al. (2015) Automated, high-throughput derivation, characterization and differentiation of induced pluripotent stem cells. *Nat Methods* 12: 885–892. doi: [10.1038/nmeth.3507](https://doi.org/10.1038/nmeth.3507) PMID: [26237226](https://pubmed.ncbi.nlm.nih.gov/26237226/)
44. Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, et al. (2012) DNA methylation signatures in development and aging of the human prefrontal cortex. *Am J Hum Genet* 90: 260–272. doi: [10.1016/j.ajhg.2011.12.020](https://doi.org/10.1016/j.ajhg.2011.12.020) PMID: [22305529](https://pubmed.ncbi.nlm.nih.gov/22305529/)
45. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, et al. (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20: 434–439. doi: [10.1101/gr.103101.109](https://doi.org/10.1101/gr.103101.109) PMID: [20219945](https://pubmed.ncbi.nlm.nih.gov/20219945/)
46. Chu M, Siegmund KD, Hao QL, Crooks GM, Tavare S, et al. (2008) Inferring relative numbers of human leucocyte genome replications. *Br J Haematol* 141: 862–871. doi: [10.1111/j.1365-2141.2008.07142.x](https://doi.org/10.1111/j.1365-2141.2008.07142.x) PMID: [18410448](https://pubmed.ncbi.nlm.nih.gov/18410448/)
47. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3: Article3.
48. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* 57: 289–300.
49. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* 23: 257–258. PMID: [17098774](https://pubmed.ncbi.nlm.nih.gov/17098774/)
50. Shabalin AA (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28: 1353–1358. doi: [10.1093/bioinformatics/bts163](https://doi.org/10.1093/bioinformatics/bts163) PMID: [22492648](https://pubmed.ncbi.nlm.nih.gov/22492648/)
51. Levene H (1960) Robust tests for equality of variances1. *Contributions to probability and statistics: Essays in honor of Harold Hotelling* 2: 278–292.
52. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)