BMC Biology

METHODOLOGY ARTICLE   **Open Access**

New targeted approaches for epigenetic age predictions

Yang Han[1,2], Julia Franzen[1,2], Thomas Stiehl[3], Michael Gobs[1,2], Chao-Chung Kuo[1,2], Miloš Nikolić[1,2], Jan Hapala[1,2], Barbara Elisabeth Koop[4], Klaus Strathmann[5], Stefanie Ritz-Timme[4] and Wolfgang Wagner[1,2*]

Abstract

Background: Age-associated DNA methylation changes provide a promising biomarker for the aging process. While genome-wide DNA methylation profiles enable robust age-predictors by integration of many age-associated CG dinucleotides (CpGs), there are various alternative approaches for targeted measurements at specific CpGs that better support standardized and cost-effective high-throughput analysis.

Results: In this study, we utilized 4647 Illumina BeadChip profiles of blood to select CpG sites that facilitate reliable age-predictions based on pyrosequencing. We demonstrate that the precision of DNA methylation measurements can be further increased with droplet digital PCR (ddPCR). In comparison, bisulfite barcoded amplicon sequencing (BBA-seq) gave slightly lower correlation between chronological age and DNA methylation at individual CpGs, while the age-predictions were overall relatively accurate. Furthermore, BBA-seq data revealed that the correlation of methylation levels with age at neighboring CpG sites follows a bell-shaped curve, often associated with a CTCF binding site. We demonstrate that within individual BBA-seq reads the DNA methylation at neighboring CpGs is not coherently modified, but reveals a stochastic pattern. Based on this, we have developed a new approach for epigenetic age predictions based on the binary sequel of methylated and non-methylated sites in individual reads, which reflects heterogeneity in epigenetic aging within a sample.

Conclusion: Targeted DNA methylation analysis at few age-associated CpGs by pyrosequencing, BBA-seq, and particularly ddPCR enables high precision of epigenetic age-predictions. Furthermore, we demonstrate that the stochastic evolution of age-associated DNA methylation patterns in BBA-seq data enables epigenetic clocks for individual DNA strands.

Keywords: Aging, Blood, Buccal swabs, Droplet digital PCR, DNA methylation, Epigenetic, Human, Amplicon sequencing, CTCF

* Correspondence: wwagner@ukaachen.de
[1]Helmholtz-Institute for Biomedical Engineering, Stem Cell Biology and Cellular Engineering, RWTH Aachen University Medical School, Pauwelsstraße 20, 52074 Aachen, Germany
[2]Institute for Biomedical Engineering – Cell Biology, University Hospital of RWTH Aachen, Aachen, Germany
Full list of author information is available at the end of the article

Introduction

During aging, DNA methylation (DNAm) is continuously lost or gained at specific CG nucleotides (CpG sites) of our genome. Conversely, the DNAm levels at multiple CpG sites can be combined to estimate age, and these models are often referred to as "epigenetic clocks" [1]. Epigenetic clocks raise hopes as a biomarker in forensic medicine, to determine the donor age of an unknown specimen or of a person with an allegedly unknown age [2]. On the other hand, accelerated epigenetic aging has been shown to be associated with shorter life expectancy [3–7], and it is liable to be affected by environmental exposure, gender, specific mutations, and diseases [5, 8–10]. Therefore, epigenetic clocks seem to reflect aspects of biological age, which opens perspectives as a surrogate for intervention studies. It is even conceivable that therapeutic regimen in future medicine will rather be stratified by epigenetic age than chronological age. To translate epigenetic biomarkers into an approved medical test, it is advantageous to select a manageable set of informative genomic regions, which can be targeted by DNA methylation assays that are sufficiently fast, cheap, robust, and widely available for clinical diagnostics [11, 12].

Initially, models for epigenetic age predictions were based on Illumina BeadChip data [13, 14]. This microarray platform enables cost-effective and relatively precise DNAm measurements at single-base resolution. In contrast, whole genome bisulfite sequencing (WGBS) or reduced representation bisulfite sequencing (RRBS) does not always cover the same CpG sites and a limited number of reads may entail lower precision of DNAm levels [15, 16]. Another advantage of the Illumina BeadChip technology is that a multitude of publicly available datasets can easily be integrated into the analysis. Various different epigenetic age-predictors have been described that consider up to several hundreds of CpGs [17–19]. The most commonly used age predictor for multiple different tissues is based on 353 CpGs to facilitate epigenetic age predictions with a median error of 3.6 years [19].

For targeted analysis of specific age-associated CpGs, various studies described epigenetic age-predictors based on bisulfite pyrosequencing [18, 20] or by the Sequenom's EpiTYPER assay [21]. Single Base Primer Extension Assay (SNaPshot) was also used by many laboratories for epigenetic age prediction, but the accuracy is apparently lower [22, 23]. Recently, droplet digital PCR (ddPCR) was reported to enable precise DNAm measurements [24, 25], and hence it might facilitate epigenetic age predictions without PCR bias. Barcoded bisulfite amplicon sequencing (BBA-seq), which is based on next generation sequencing, enables multiplexed analysis of PCR amplicons [26, 27]. The strength of BBA-seq lies within the parallelization of multiple DNA sequences on one lane with relatively long amplicons (up to 500 bases), and with very high coverage (usually > 1000 fold) [28, 29]. So far, only few groups described ddPCR [30] and BBA-seq [31, 32] for epigenetic age predictions and a direct comparison of these methods is still elusive.

It is largely unclear, how age-associated DNAm is regulated and if it is functionally relevant, per se. Transcription factors (TFs) or long non-coding RNAs (lncRNAs) might target epigenetic writers, such as DNA methyltransferases (DNMTs) or ten-eleven translocation family enzymes (TETs), to specific sites in the genome [33]. This process may also involve alternative splicing of DNMTs [34]. CCCTC-binding factor (CTCF), as an insulator TFs that is involved in chromatin architecture, has also been shown to be methylation sensitive [35]. If age-associated DNAm was directly regulated by epigenetic writers, it would be anticipated that the DNAm pattern of neighboring CpGs on the same DNA strand is coherently modified. Alternatively, it has been proposed that age-associated DNAm is evoked by "epigenetic drift," which may occur due to stochastic accumulation of errors, e.g., in copying DNAm patterns during cell replication [36–38]. In this case, DNAm on individual DNA strands might rather follow stochastic patterns.

In this study, we have further optimized and directly compared epigenetic age predictors based on pyrosequencing, ddPCR, and BBA-seq of specific age-associated regions. Furthermore, our data indicate that the correlation of age-associated DNAm with chronological age peaks close to CTCF binding sites. Age-associated DNAm is not coherently modified on individual DNA strands and this enabled alternative single-read age-predictors that reflect heterogeneity in epigenetic aging within a specimen.

Results

Selection of age-associated CpGs for blood

To select age-associated CpGs for further comparison, we used a training set of 973 DNAm profiles of healthy human blood samples (1 to 101 years old), which are derived from seven different studies and based on the 450 k Illumina BeadChip platform (Additional file 1: Figure S1A). For further analysis, we excluded CpGs that were associated with a single-nucleotide polymorphism (SNP) or reported to be cross-reactive [39, 40]. To reduce the impact of leukocyte composition [41], we excluded CpGs with high variation in DNAm across hematopoietic subsets ($R > 0.02$ in six different cellular subsets; Additional file 1: Figure S1B) [42]. Furthermore, we excluded CpGs on sex chromosomes, CpGs that are significantly affected by smoking, and those which were no more comprised on the new EPIC Illumina BeadChip microarray. Thus, 313,728 CpGs were used for further selection of age-associated CpGs (Additional file 1: Figure S1C) [39,

40, 43, 44]. To identify candidates that are best suited for targeted DNAm analysis, we selected CpGs by linear correlation with chronological age using Pearson correlation > 0.5 or < − 0.5. Age-associated DNAm might also follow a non-linear monotonic function and therefore we alternatively used Spearman's rank correlation > 0.5 or < − 0.5. Since age-associated DNAm was shown to rather follow a logarithmic pattern, particularly in children [19, 45], we also selected CpGs that correlated with logarithmic age across all samples (Pearson's correlation > 0.5 or < − 0.5). It needs to be taken into account that there is notoriously variation between different Illumina BeadChip datasets due to modifications in sample preparation, handling, and measurement in different studies (Additional file 1: Figure S1D-G). Therefore, we have only considered CpGs that met the corresponding criterion ($R > 0.5$ or $< − 0.5$) in each of the seven individual studies of the training set. Overall, 65 CpGs passed at

least one of these three filter criteria—and despite the different filter criteria, the overlap was remarkably high—with 18 CpGs passing all three thresholds (Fig. 1a). Subsequently, we trained a multivariable model based on the 65 CpGs using functions of transformed chronological age with logarithmic dependence for pediatric and linear dependence for adult donors, as previously described by Horvath et al. [19]. This model provided a high correlation with chronological age in the training set ($R^2 = 0.95$; median error = 3.0 years; Fig. 1b) and in an independent validation set of 3674 blood samples of five different studies ($R^2 = 0.82$; median error = 3.3 years; Fig. 1c). Moreover, the accuracy of our 65 CpG model even outperformed the accuracy of age-predictions using the epigenetic aging signatures with 353 CpGs [19] or 71 CpGs [17], often referred to as Horvath and Hannum clocks, respectively. It even outperformed the more recent skin and blood clock of Horvath et al. [46];

Fig. 1 Selection of age-associated CpGs and targeted analysis with pyrosequencing. **a** Illumina BeadChip profiles of 973 blood samples of 7 studies (all 450 k) were used to select age-associated CpGs by Pearson's correlation (blue), Spearman's rank correlation (red), or Pearson's correlation after logarithmic transformation of age (green). Sixty-five CpGs passed at least one of these thresholds (correlation coefficient: $R > 0.5$ or $R < − 0.5$) and the Venn diagram depicts a very high overlap. **b, c** A multivariable linear model for these 65 age-related CpGs revealed high correlation with chronological age in the training set (**b**; $n = 973$), and in an independent validation set (**c**; $n = 3674$; color code corresponds to different studies as indicated in Supplementary Fig. S3A). **d, e** Epigenetic age prediction based on pyrosequencing of 6 CpGs (**d**) and 9 CpGs (**e**) in blood samples of the training set ($n = 40$; blue) and an independent validation set ($n = 40$; red)
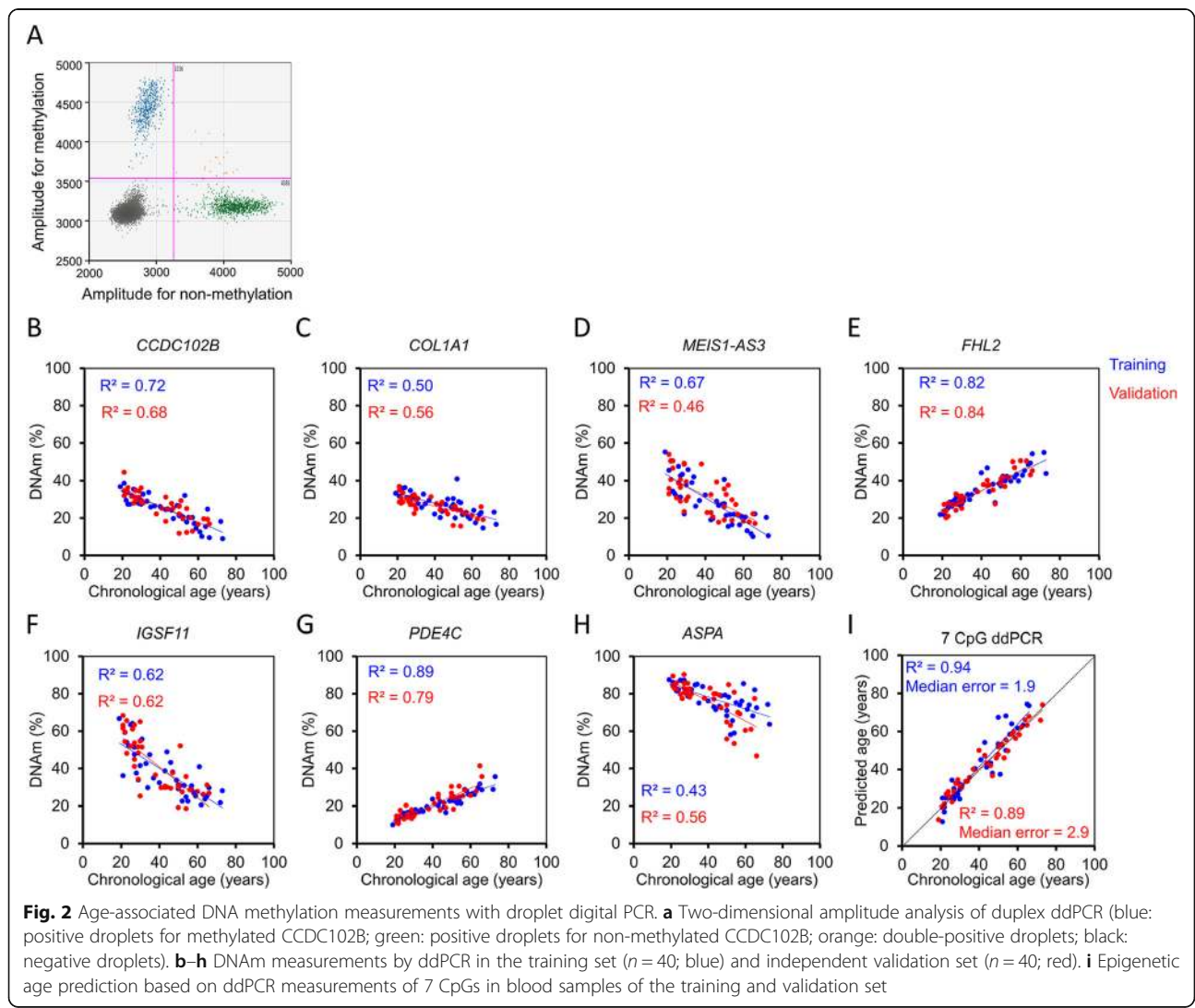
however, the later performed slightly better if we only considered datasets that provided DNAm values for all 391 relevant CpGs (Additional file 1: Figure S2 and S3) [17, 19, 46].

Epigenetic age predictor based on pyrosequencing of specific CpGs

To further narrow down the best-suited CpGs for targeted DNAm measurement, we selected among the 65 age-related CpGs those with the best linear correlation and highest slope in Illumina BeadChip training sets (Additional file 1: Figure S4A). The selected CpGs were associated with the genes of Elongation Of Very Long Chain Fatty Acids Protein 2 (*ELOVL2*), which is well known to reveal high age-association [21], Coiled-Coil Domain-Containing Protein 102B (*CCDC102B*), Four And A Half LIM Domains Protein 2 (*FHL2*), Immunoglobulin Superfamily Member 11 (*IGSF11*), Collagen

Type I Alpha 1 Chain (*COL1A1*), and MEIS1 Antisense RNA 3 (*MEIS1-AS3*). When we analyzed DNAm at these CpGs in 40 blood samples by pyrosequencing, we observed high correlation with chronological age, which was overall higher than for the independent validation set of 450 k Illumina BeadChip profiles (Additional file 1: Figure S4B). A multivariable linear regression model based on pyrosequencing of 40 samples revealed a good correlation with chronological age in an independent validation set of 40 blood samples ($R^2 = 0.86$; Fig. 1d). The correlation was even increased, when we integrated three additional CpGs, which were selected in our previous work (associated with the genes *ITGA2B*, *ASPA*, and *PDE4C* [18]), into the model ($R^2 = 0.89$; Fig. 1e). However, measurement of the independent validation set, which was measured several months later, revealed a small systematic offset at several CpGs (*COL1A1*, *FHL2*, and *IGSF11*; Additional file 1: Figure S4B), and therefore

Fig. 2 Age-associated DNA methylation measurements with droplet digital PCR. **a** Two-dimensional amplitude analysis of duplex ddPCR (blue: positive droplets for methylated CCDC102B; green: positive droplets for non-methylated CCDC102B; orange: double-positive droplets; black: negative droplets). **b–h** DNAm measurements by ddPCR in the training set (*n* = 40; blue) and independent validation set (*n* = 40; red). **i** Epigenetic age prediction based on ddPCR measurements of 7 CpGs in blood samples of the training and validation set

the median error was considerably larger in the validation set (median error = 6.8 years with 6 CpG model and 5.2 years with 9 CpG model). We could not identify the influencing parameters that evoked this small offset, but pyrosequencing appears to be sensitive to small technical variations due to possible PCR bias of methylated versus non-methylated sequences.

Analysis of age-associated DNAm with droplet digital PCR

Droplet digital PCR (ddPCR) is based on randomly distributing a single sample into small droplets, which are then processed for PCR amplification separately. We anticipated that this technology reduces PCR bias, because sequences are amplified with different efficiency and therefore affect the ratio in conventional PCR. For each amplicon two probes were designed that either detect the methylated or the non-methylated sequence and DNAm level was subsequently determined by Poisson distribution (Fig. 2a). Reliable ddPCR assays could be established for the sequences of CCDC102B, COL1A1, MEIS1-AS3, FHL2, PDE4C, ASPA, and IGSF11, while this was hampered for ITGA2B and ELOVL2 due to neighboring CpG sites. The ddPCR results revealed a clear correlation with chronological age at all tested CpGs, which was often even slightly higher than for pyrosequencing (Table 1; Fig. 2b–h). We then generated a multivariable model based on ddPCR measurement of these seven CpGs. This provided relatively reliable age-predictions in an independent validation set of 40 blood samples ($R^2 = 0.89$; Fig. 2i). The median error of these ddPCR age-predictions was only 2.9 years, which was significantly lower than for pyrosequencing ($P = 0.00005$), indicating that age-predictions in independent sets of samples might be more accurate with ddPCR.

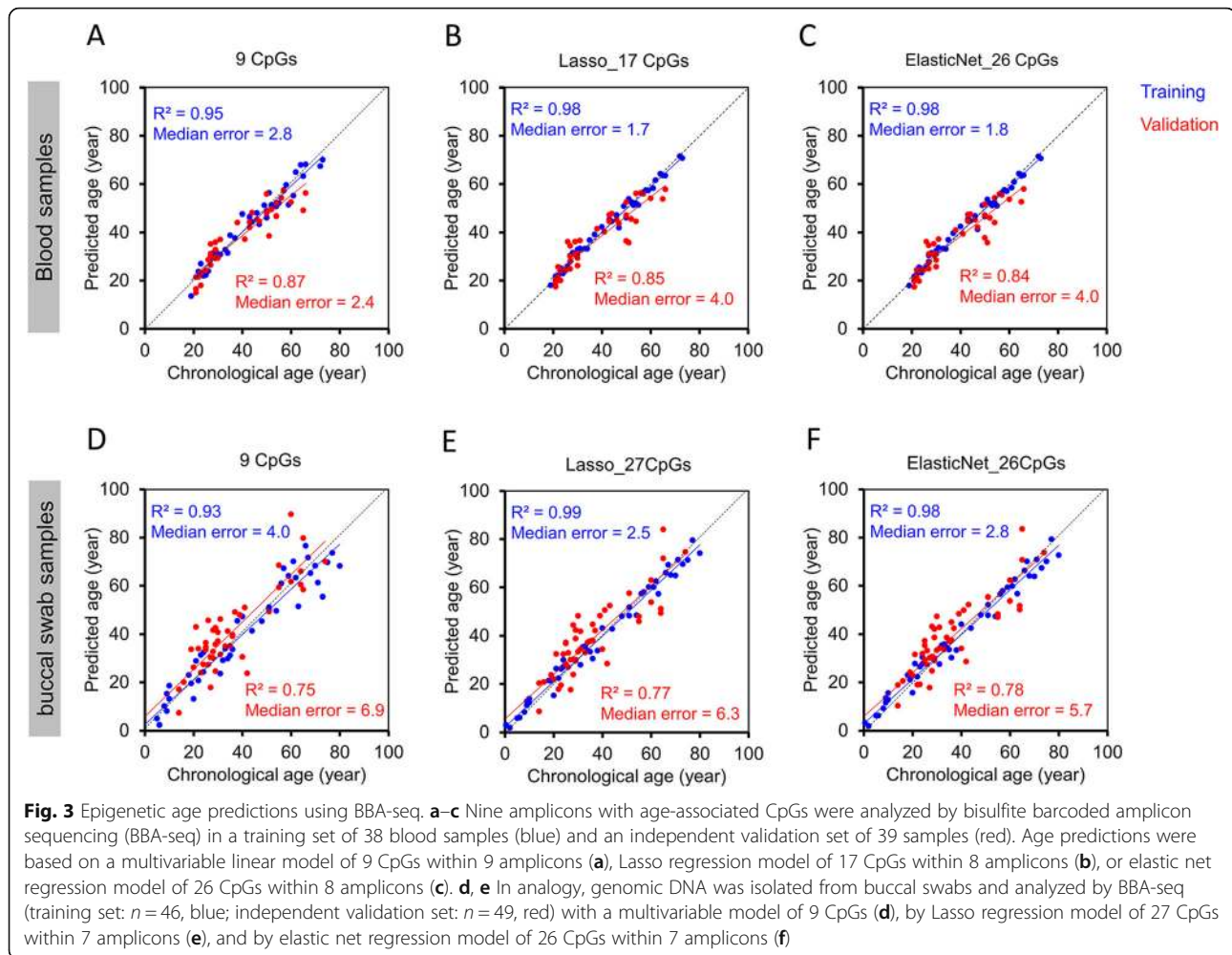### Epigenetic age-predictions with bisulfite barcoded amplicon sequencing

Subsequently, we analyzed DNAm of the nine age-associated regions with deep sequencing technology (Illumina MiSeq). Bisulfite barcoded amplicon sequencing (BBA-seq) was initially performed on a training set of 38 blood samples. Overall, DNAm levels of BBA-seq and pyrosequencing revealed good correlation ($R^2 = 0.92$; Additional file 1: Figure S5), indicating that DNAm measurements were feasible for all amplicons. Furthermore, the correlation of DNAm levels in BBA-seq with chronological age was in a similar range as observed for pyrosequencing or ddPCR (Table 1; Additional file 1: Figure S6). A multivariable linear regression model was then established for the nine CpGs with the highest correlation with chronological age per amplicon. This approach provided high accuracy for age-predictions in the training set ($R^2 = 0.95$; median error = 2.8 years). Unexpectedly, the accuracy of epigenetic age-predictions was even higher in the independent validation set of 39 blood samples ($R^2 = 0.87$; median error = 2.4 years; Fig. 3a). Alternatively, we used all CpGs within the nine amplicons for Lasso (Fig. 3b) and elastic net regression models (Fig. 3c) with 10-fold cross-validation. Both machine learning approaches performed better than the linear regression model on the training set, but the precision was lower on the validation set, which might be attributed to small technical offsets between different BBA-seq runs or overfitting due to the relatively small sample size.

We have then analyzed if the method is also applicable for buccal swab samples, which are a widely used specimen in legal medicine. Due to the different cellular composition, the multivariable model for the 9 CpGs was retrained to provide a good correlation in a training set of 46 buccal swab samples ($R^2 = 0.93$; median error = 4.0

Table 1 Correlation of DNA methylation and chronological age with different targeted approaches

	Pyrosequencing		ddPCR		BBA-seq	
	Training ($n = 40$)	Validation ($n = 40$)	Training ($n = 40$)	Validation ($n = 40$)	Training ($n = 38$)	Validation ($n = 39$)
FHL2[1]	0.76	**0.77**	0.82	**0.84**	0.82	**0.28**
IGSF11	0.53	**0.71**	0.62	**0.62**	0.58	**0.55**
CCDC102B	0.56	**0.7**	0.72	**0.68**	0.60	**0.58**
MEIS1-AS3[1]	0.56	**0.55**	0.67	**0.46**	0.68	**0.55**
ELOVL2	0.45	**0.46**	–	**–**	0.87	**0.63**
COL1A1	0.32	**0.4**	0.5	**0.56**	0.42	**0.51**
ASPA	0.29	**0.57**	0.43	**0.56**	0.37	**0.57**
ITGAB[1]	0.03	**0.15**	–	**–**	0.14	**0.11**
PDE4C[1]	0.8	**0.74**	0.89	**0.79**	0.82	**0.71**
Mean R^2	0.48	**0.56**	0.66	**0.64**	0.59	**0.50**

[1]We have always depicted result for the CpG with the highest Pearson correlation within the amplicon. Thus, the CpGs are not always identical (Additional file 1: Table S5, S7 and S9)

Fig. 3 Epigenetic age predictions using BBA-seq. **a**–**c** Nine amplicons with age-associated CpGs were analyzed by bisulfite barcoded amplicon sequencing (BBA-seq) in a training set of 38 blood samples (blue) and an independent validation set of 39 samples (red). Age predictions were based on a multivariable linear model of 9 CpGs within 9 amplicons (**a**), Lasso regression model of 17 CpGs within 8 amplicons (**b**), or elastic net regression model of 26 CpGs within 8 amplicons (**c**). **d**, **e** In analogy, genomic DNA was isolated from buccal swabs and analyzed by BBA-seq (training set: $n = 46$, blue; independent validation set: $n = 49$, red) with a multivariable model of 9 CpGs (**d**), by Lasso regression model of 27 CpGs within 7 amplicons (**e**), and by elastic net regression model of 26 CpGs within 7 amplicons (**f**)

years), albeit the correlation remained lower in the independent validation set of 49 buccal swab samples ($R^2 = 0.75$; median error = 6.9 years; Fig. 3d). The accuracy could be slightly increased by Lasso (Fig. 3e) and elastic net algorithms (Fig. 3f) that were generated based on all CpGs of the amplicons, but it remained lower than for blood samples. This might be due to the heterogeneous composition of buccal epithelial cells and leukocytes in buccal swab samples [47].
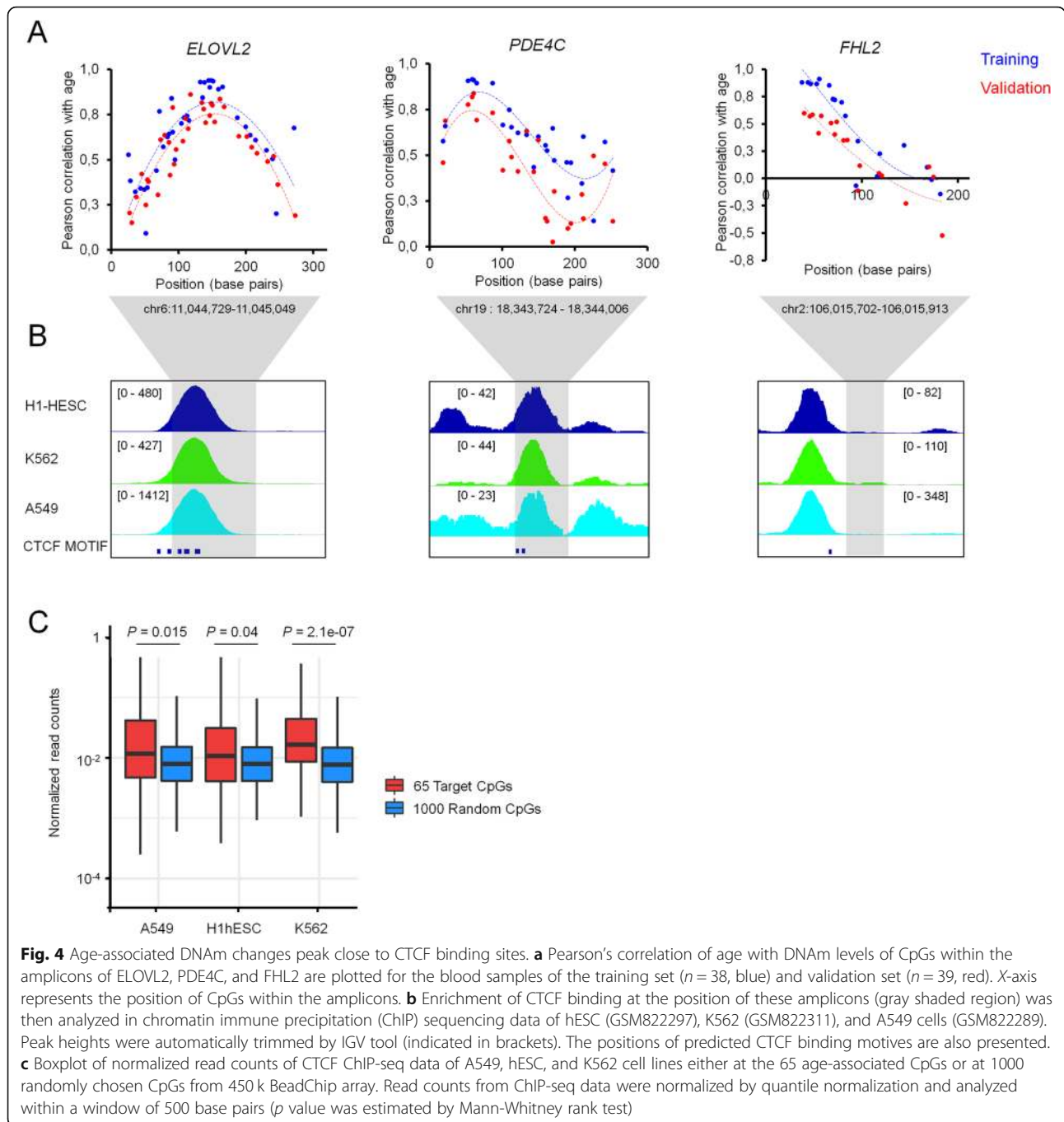
Age-associated DNA methylation changes peak close to CTCF binding sites

One of the advantages of BBA-seq is that amplicons are longer than in pyrosequencing and include measurement of more neighboring CpGs. We have compared the correlation of neighboring CpGs with chronological age, particularly in the amplicons of *ELOVL2*, *PDE4C*, and *FHL2*, which harbor 36, 26, and 18 CpGs, respectively. Plotting of correlation coefficients against the genomic locations revealed curvy distributions (Fig. 4a). Similar distributions, but less distinct, were also observed for the BBA-seq data of buccal swabs

(Additional file 1: Figure S7). Notably, age-associated DNAm changes at these three age-associated regions peaked in vicinity of binding sites of CCCTC-binding factor (CTCF), which is involved in organization of chromatin structure. Furthermore, chromatin immune precipitation (ChIP)-seq data of human embryonic stem cells (hESC; GSM822297), K562 (GSM822311), and A549 cell lines (GSM822289) indicated that CTCF binds close to the peak of age-associated DNAm changes (Fig. 4b). When we analyzed enrichment of our previously identified 65 age-associated CpGs within ChIP-Seq read peaks of CTCF, they were significantly enriched in each of the three ChIP-seq experiments (Fig. 4c). These data may suggest that regulation of age-associated DNAm changes is related to CTCF binding and/or the three-dimensional chromatin conformation [48].

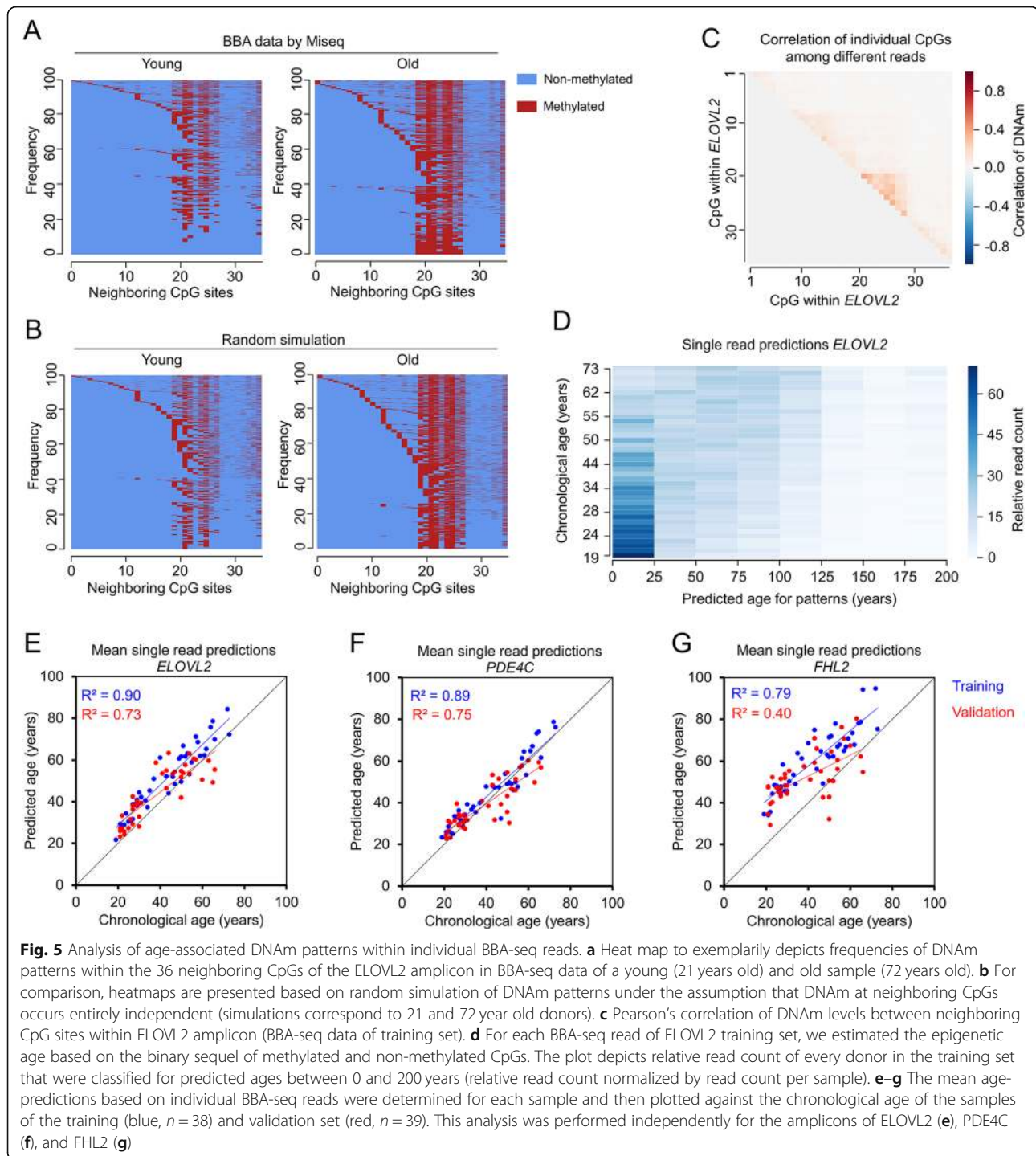## Analysis of age-associated DNAm patterns within individual BBA-seq reads

Age-associated DNAm peaks at specific sites in the genome, but it remained unclear whether neighboring CpGs are coherently modified or not. If DNAm was regulated

Fig. 4 Age-associated DNAm changes peak close to CTCF binding sites. **a** Pearson's correlation of age with DNAm levels of CpGs within the amplicons of ELOVL2, PDE4C, and FHL2 are plotted for the blood samples of the training set ($n = 38$, blue) and validation set ($n = 39$, red). X-axis represents the position of CpGs within the amplicons. **b** Enrichment of CTCF binding at the position of these amplicons (gray shaded region) was then analyzed in chromatin immune precipitation (ChIP) sequencing data of hESC (GSM822297), K562 (GSM822311), and A549 cells (GSM822289). Peak heights were automatically trimmed by IGV tool (indicated in brackets). The positions of predicted CTCF binding motives are also presented. **c** Boxplot of normalized read counts of CTCF ChIP-seq data of A549, hESC, and K562 cell lines either at the 65 age-associated CpGs or at 1000 randomly chosen CpGs from 450 k BeadChip array. Read counts from ChIP-seq data were normalized by quantile normalization and analyzed within a window of 500 base pairs (p value was estimated by Mann-Whitney rank test)

by a targeted protein complex, it would be expected that neighboring CpGs are conjointly modified on individual strands. To address this question, we analyzed the DNAm pattern of individual reads in BBA-seq, which reflect the binary code of methylated and non-methylated cytosines in individual DNA strands. In fact, the DNAm patterns at the age-associated regions were very heterogeneous (Fig. 5a). When we simulated random DNAm patterns based on the likeliness of DNAm at specific

CpGs at a given age, the patterns were similar to the experimental results (Fig. 5b). The correlation in DNAm at individual CpGs was overall very low (Fig. 5c). Thus, age-associated DNAm at neighboring CpGs is apparently not coherently modified and rather seems to be acquired in a stochastic manner.

We then reasoned that DNAm patterns within individual reads might also be used for age-predictions. To this end, we developed a mathematical model based on the

Fig. 5 Analysis of age-associated DNAm patterns within individual BBA-seq reads. **a** Heat map to exemplarily depicts frequencies of DNAm patterns within the 36 neighboring CpGs of the ELOVL2 amplicon in BBA-seq data of a young (21 years old) and old sample (72 years old). **b** For comparison, heatmaps are presented based on random simulation of DNAm patterns under the assumption that DNAm at neighboring CpGs occurs entirely independent (simulations correspond to 21 and 72 year old donors). **c** Pearson's correlation of DNAm levels between neighboring CpG sites within ELOVL2 amplicon (BBA-seq data of training set). **d** For each BBA-seq read of ELOVL2 training set, we estimated the epigenetic age based on the binary sequel of methylated and non-methylated CpGs. The plot depicts relative read count of every donor in the training set that were classified for predicted ages between 0 and 200 years (relative read count normalized by read count per sample). **e–g** The mean age-predictions based on individual BBA-seq reads were determined for each sample and then plotted against the chronological age of the samples of the training (blue, $n = 38$) and validation set (red, $n = 39$). This analysis was performed independently for the amplicons of ELOVL2 (**e**), PDE4C (**f**), and FHL2 (**g**)

BBA-seq data by assigning each DNAm pattern the most likely corresponding age (between 0 and 200 years; Fig. 5d). As anticipated for younger donors, the model revealed a higher number of young read predictions, whereas older donors had more reads that were predicted older. Notably, the mean of strand-specific age-predictions correlated well with the chronological age of the donors in the training and validation sets (Fig. 5e–g).

In fact, in the validation set, the correlation of the mean of single read predictions with chronological age for *ELOVL2*, *PDE4C*, and *FHL2* was even better than the correlation of the mean DNA methylation levels at individual CpGs in amplicons of the BBA-seq data (Table 1). This supports the notion that single-read age-predictors are applicable and that epigenetic clocks tick independently within cells of the same sample.

Discussion

In the advent of new technologies for DNAm measurements, there is a continuous need to revisit, optimize, and validate epigenetic clocks. In this study, we used the linear correlation of age-associated DNAm with chronological age as a proxy for the precision of DNAm measurements. While all methods tested revealed good or even very good correlations with chronological age, they all have their advantages and limitations for the application in epigenetic clocks.

So far, most epigenetic clocks for human samples were derived from Illumina BeadChip datasets [15]. It provides unparalleled opportunity to measure DNAm at single-base resolution, albeit not all CpGs of the genome are represented on these platforms (for instance about 1.7% of the human CpGs are covered by the 450 k Bead-Chip). Our analysis clearly demonstrated that DNAm levels in 450 k Illumina BeadChip data are highly correlated with chronological age at various specific CpGs. For development of our targeted assays, we focused particularly on blood samples and filtered for CpGs with low variation between leukocyte subsets to reduce the impact of the cellular composition. Our predictor with 65 age-associated CpGs provides similar or even better accuracy than other commonly used predictors [17–19]. Furthermore, our signature is also applicable for the current EPIC BeadChip version. While normalization regimen, integration of more CpGs, and machine learning algorithms can further enhance age-predictions, the goal of our selection was to identify suitable CpGs for the subsequent targeted approaches. Focusing on a smaller number of CpGs in targeted assays is a tradeoff between the applicability and accuracy [15]. Targeted analysis is usually faster, more cost-effective, and better applicable for laboratories that do not have immediate access to Illumina BeadChip technology. Furthermore, targeted assays are less dependent on availability of specific BeadChip versions, which is important given the long timeframes needed to develop a standardized procedure towards certification according to in vitro diagnostics regulations (IVDR).

Despite the remarkable linear correlation of age-associated DNAm with age, there is evidence for rather logarithmic association, particularly in childhood [19, 45, 49]. We selected candidate CpGs either for linear correlation (Pearson's correlation), continuous non-linear association (Spearman's rank-order), or by linear correlation with the logarithm of age. To our surprise, there was a very high overlap of selected CpGs with the three filter criteria, indicating that age-associated DNAm changes are generally accelerated in early life and are then rather linearly acquired in adulthood and the elderly. In fact, age-predictions for pediatric cohorts were clearly improved by age-transformation with a logarithmic adjustment, as elegantly described by Horvath [19], and this approach should therefore also be considered for targeted epigenetic age predictors if applied to pediatric cohorts.

Bisulfite pyrosequencing is currently the most popular targeted method for epigenetic age predictions in forensics [50]. This method is relatively simple and it has been shown to have a very high precision in DNAm measurement [11]. The accuracy of our epigenetic age-predictions with pyrosequencing was in a similar range as for Illumina BeadChip models. Nevertheless, the conventional PCR reaction before pyrosequencing can evoke amplification bias for methylated or non-methylated strands [50].

Droplet digital PCR might reduce this technical PCR amplification bias, because the individual droplets are either scored as positive or negative independent of the PCR efficacy. So far, only few studies used ddPCR for DNAm analysis [51, 52], and only one recent study used it for measurement of age-associated DNAm changes in a pediatric cohort [30]. In this study, we performed a direct comparison of epigenetic age-predictions of ddPCR and pyrosequencing, and ddPCR indeed provided significantly lower median error of age-predictions in the validation set. However, fluorescent probes could not be designed for all sequences, particularly if many neighboring CpGs were located in the target region. Overall, DNA methylation measurements with ddPCR appear to be more precise than pyrosequencing and this should be further validated by interlaboratory comparison.

Bisulfite barcoded amplicon sequencing makes use of the technical advances in massive parallel sequencing. Recently, Naue and coworkers described a similar age predictor using massive parallel sequencing [32]. Our comparative approach substantiates the notion that BBA-seq is a powerful method for epigenetic clocks, particularly if multiple samples need to be analyzed in parallel. The correlation of DNAm measurements with chronological age at individual CpGs was similar in BBA-seq data as compared to pyrosequencing or ddPCR. In analogy to pyrosequencing, we observed a systematic off-set between the BBA-seq results of different sequencing runs, which might be attributed to PCR bias. On the other hand, the long BBA-seq amplicons provided insight into DNAm patterns on the same DNA strand. We demonstrate that the correlation of DNAm at neighboring CpGs with chronological age follows a bell-shaped curve. Notably, this correlation often peaked close to CTCF binding sites, which resembles one of the best characterized architectural proteins for the 3D chromatin conformation [48]. This is in line with previous studies indicating that age-associated DNAm changes are enriched at CTCF binding sites [53, 54]. Furthermore, it has been suggested that age-associated

Han *et al. BMC Biology*   (2020) 18:71

Page 10 of 15

hypomethylation, but not hypermethylation, is related with CTCF binding sites across various human tissues [55, 56]—albeit we observed this particularly at the hypermethylated regions in *ELOVL2*, *PDE4C*, and *FHL2*. Another genome scale study supported the notion that CTCF enrichment occurred both in age-associated hyper- and hypomethylated regions [54]. Occupancy with CTCF is tissue-specific and it is influenced by DNAm [57, 58]. In turn, CTCF was also capable to change DNAm status, for instance, by reducing the DNAm level at the vicinity of its binding positions [59]. Thus, the relevance of CTCF-binding for age-associated DNAm changes needs to be further explored.

Furthermore, our analysis of DNAm patterns in individual reads of BBA-seq demonstrated that neighboring CpGs are modified rather independently. We have recently described similar findings for DNAm changes during long-term culture of cells in vitro [29, 60]. While the stochastic changes at neighboring CpGs challenge the view of directed regulation of age-associated DNAm, they may support the notion that this process is evoked by "epigenetic drift," possibly caused by changes in chromatin conformation. The finding of stochastic DNAm changes provided also the basis for our mathematical approach of epigenetic age predictions for individual BBA-seq reads. It is yet unclear if epigenetic aging is accelerated synchronously at different genomic regions within the same cell. Single Cell DNAm analysis has been described using WGBS [61], but the coverage at individual CpGs (including age-associated CpGs) is notoriously very low, and thus analysis of epigenetic clocks is hampered in such datasets. On the other hand, there is evidence that age-associated DNAm is coherently modified in cancer, which possibly reflects the epigenetic state of the tumor-initiating cell [62]. Thus, the epigenetic age predictions based on individual BBA-seq reads seems to reflect heterogeneity of epigenetic aging within a sample.

Conclusions
Our comparative approach demonstrates that targeted analysis of age-associated DNAm via pyrosequencing, ddPCR, and BBA-seq enables similar correlation with chronological age as described for larger signatures on Illumina BeadChip profiles. In the independent validation set, the highest accuracy of epigenetic age-predictions was reached with ddPCR, which might be due to lower PCR bias. Thus, this relatively new method should be considered for further round robin tests to validate the precision and accuracy for epigenetic age-predictions in clinical and forensic application. Furthermore, our analysis of BBA-seq data demonstrated that age-associated DNAm peaks close to specific sites in the genome, particularly at CTCF binding sites. The finding that DNA methylation at neighboring CpGs is not

coherently modified opened the perspective for our alternative single read predictors. The mean of strand-specific age-predictions correlated very well with chronological age. This method will gain additional power with longer reads, as provided by nanopore sequencing, and it will shed additional light into the heterogeneity of cellular aging within a sample.

Methods
Sample collection
Peripheral blood samples of healthy donors ($n = 80$) were obtained from the Department of Transfusion Medicine at the University Hospital of RWTH Aachen. Buccal swab samples were collected with Mastaswab MD555 (Mast Group ltd., Reinfeld, Germany) at the Institute for Legal Medicine of the Heinrich Heine University in Düsseldorf, Germany ($n = 95$). This study has been performed according to the guidelines approved by the local ethics committees of RWTH Aachen University (EK 041/15) and Heinrich Heine University of Düsseldorf (Permit number 4939).

Selection of age-associated CpG sites
In total, 4647 DNAm profiles of human blood samples of 12 different studies (all analyzed on the Human-Methylation450 [450 k] BeadChip platform; no samples with known malignancies) were retrieved from Gene Expression Omnibus Database (GEO; Additional file 1: Table S1) [17, 45, 63–71]. To further select age-associated CpGs for targeted analysis, we excluded (i) 105,454 SNP-associated (provided by the updated 'MASK_general' subset) [40] and 29,233 cross-reactive CpGs [39]; (ii) 26,426 CpGs with high variation across the different hematopoietic subsets in (GSE35069; variances > 0.02) [42]; (iii) 2901 CpGs that are significantly affected by smoking [43, 44]; (iv) 11,648 CpGs of sex chromosomes; and (v) 31,396 CpGs that were not included in the new Illumina BeadChip EPIC platform to facilitate better comparison with future data. The remaining 313,728 CpGs were then filtered for their correlation with chronological age. It has been demonstrated that most age-associated DNA methylation changes can be accurately modeled as a function of logarithmic age, particularly in children [45]. Therefore, we selected CpGs by their correlation with the logarithm of chronological age (cutoffs for all comparisons: $R < -0.5$ or > 0.5). Alternatively, we used Pearson's correlation or Spearman's correlation with chronological age (also $R < -0.5$ or > 0.5), since those approaches were used in other studies before and we anticipated that selection for linear and non-linear DNA methylation changes would provide complementary subsets of CpGs. Due to variation in beta-values between different datasets the corresponding criterion (Pearson's log age, Pearson's age, or

Spearman's age; $R < -0.5$ or > 0.5) needed to be fulfilled in each of the seven different datasets of the training set, individually.

Before establishing the multivariable model for epigenetic age predictions based on the 65 age-associated CpGs, we used a transformed age instead of chronological age, as described before [19]. In brief, adult.age was assigned as 20 years, F function was used for age transformation as follows:

$$F\,(\text{age}) = \log\,(\text{age} + 1) - \log\,(\text{adult.age} + 1),\,\text{if age} \leq \text{adult.age}.$$

$$F\,(\text{age}) = (\text{age} - \text{adult.age})/(\text{adult.age} + 1),\,\text{if age} > \text{adult.age}.$$

Subsequently, we established multivariable regression model (F (transformed age)) based on transformed age on the training sets from Illumina BeadChip (Additional file 1: Table S2) [19]. After estimation of transformed age, the results need to be inversely transformed for epigenetic age-predictions:

$$\begin{aligned} \text{DNAm age} &= \text{inverse}.F\,(\text{transformed age}) \\ &= (1 + \text{adult.age}) \times \exp\,(\text{transformed age}) \\ &\quad -1,\,\text{if transformed age} < 0 \end{aligned}$$

$$\begin{aligned} \text{DNAm age} &= \text{inverse}.F\,(\text{transformed age}) \\ &= (1 + \text{adult.age}) \times \text{transformed age} \\ &\quad + \text{adult.age},\,\text{if transformed age} \geq 0 \end{aligned}$$

To investigate batch effects of the different datasets of the training and validation microarrays, we performed simple mean imputation and principal component analysis with the python package scikit-learn. We used all available CpGs of the datasets within the training (385, 587 CpGs) and validation (462,889 CpGs) groups (only 1000 samples of GSE55763 were included for the validation datasets) or used the 65 CpGs of our age predictor.

Isolation of genomic DNA and bisulfite conversion

Genomic DNA was isolated from 50 μl blood with the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), or from buccal swab with NucleoSpin Tissue Kit (Macherey-Nagel, Düren, Germany). DNA was quantified with a Nanodrop 2000 Spectrophotometer (Thermo Scientific, Wilmington, USA) and 200 ng genomic DNA was bisulfite converted with the EZ DNA Methylation Kit (Zymo Research, Irvine, USA).

Pyrosequencing

Bisulfite converted DNA was used for PCR amplification by PyroMark PCR Kit (Qiagen, Hilden, Germany). Twenty micrograms of PCR products was immobilized to 5 μl Streptavidin Sepharose High Performance Bead (GE Healthcare, Piscataway, NJ, USA), and subsequently annealed to 1 μl sequencing primer (5 μM) for 2 min at 80 °C. PCR and pyrosequencing primers (Metabion, Planegg-Martinsried, Germany) are provided in Additional file 1: Figure S8 and Additional file 1: Table S3. Pyrosequencing was performed on PyroMark Q96 ID System and analyzed with PyroMark Q CpG software (Qiagen). To estimate epigenetic age, we either used a multivariable model based on six CpGs (Additional file 1: Table S4), or a nine CpG model that also considered the three CpGs of our previous work [18] (Additional file 1: Table S5).

Droplet digital PCR (ddPCR)

Droplet digital PCR was performed with a QX200™ Droplet Digital™ PCR System (Bio-Rad, CA, USA). Primers and dual-labeled probes were designed by Primer3Plus software (Additional file 1: Table S6). The reaction mixture consisted of 10 μl of 2X ddPCR Supermix (no dUTP; Bio-Rad), 1 μM of the forward and reverse primers, 250 nM of the probes targeting the methylated and unmethylated DNA sequences, and 25 μg of bisulfite converted DNA in a final volume of 20 μl. Together with 70 μl of droplet generation oil, it was then subjected into a DG8 disposable droplet generation cartridge (Bio-Rad). The water-in-oil droplets were produced by QX200 Droplet Generator (Bio-Rad). Forty microliters of the generated droplets were transferred to the ddPCR 96-well plate (Bio-Rad). After heat sealing with the PX1 PCR Plate Sealer (Bio-Rad), the plate was placed in the C1000 Touch Thermal Cycler (Bio-Rad) for PCR run. The thermal cycling conditions were 95 °C for 10 min, followed by 40 cycles of 94 °C for 30 s and 1 min (2.5 °C/s ramp rate) at 58 °C (*FHL2* and *PDE4C*), 54 °C (*COL1A1*) or 53 °C (*CCDC102B*, *MEIS1-AS3*, *ASPA*, and *IGSF11*) with a 10 min step at 98 °C for enzyme deactivation and a final hold at 4 °C. Subsequently, the plates were read on the QX200 droplet reader (Bio-Rad) and data were analyzed by QuantaSoft 1.7.4 software (Bio-Rad). The percentage methylation of each reaction was calculated by Poisson statistics according to the fraction of positive droplets for methylated and non-methylated probes. The multivariable model for ddPCR is provided in Additional file 1: Table S7.

Bisulfite barcoded amplicon sequencing (BBA-seq)

Target sequences with candidate CpG sites were amplified by PyroMark PCR kit (Qiagen) (Additional file 1: Figure S9). The forward and reverse primers contain handle sequences for the subsequent barcoding step (Additional file 1: Tables S8). PCR conditions were 95 °C for 15 min; 45 cycles of 94 °C for 30 s, 58 °C for 30s, 72 °C for 30s; and then final elongation 72 °C for 10 min. The amplicons of each donor (e.g., of 9 different CpGs) were pooled at equal

concentrations, quantified with Qubit (Invitrogen), and cleaned up with paramagnetic beads from Agencourt AMPure PCR Purification system (Beckman Coulter). Four microliters of PCR products were subsequently added to 21 μl PyroMark Master Mix (Qiagen) containing 10 pmol of barcoded primers (adapted from NEXTflexTM 16S V1-V3 Amplicon Seq Kit, Bioo Scientific, Austin, USA) for a second PCR (95 °C for 15 min; 16 cycles of 95 °C for 30 s, 60 °C for 30s, 72 °C for 30s; final elongation 72 °C for 10 min). PCR products were again quantified by Qubit Kit (Invitrogen), combined in equimolar ratios, and cleaned by Select-a-Size DNA Clean & Concentrator Kit (Zymo Research, USA). Ten picometers DNA library was diluted with 15% PhiX spike-in control and eventually subjected to 250 bp pair-end sequencing on a MiSeq lane (Illunima, CA, USA) using Miseq reagent V2 Nano kit (Illumina). Two samples of the training set and one sample of the validation set were not successfully amplified and therefore not considered for further analysis.

To estimate DNAm levels for each CpG based on BBA-seq data, we used the Bismark tool [72]. The average number of reads per sample and genomic region was approximately 25,000 and only sequences that occurred at least 10 times were further considered. Multivariable models for epigenetic age predictions based on those CpGs that revealed highest correlation with chronological age per amplicon are provided in Additional file 1: Table S9 and S10 (for blood and buccal swabs, respectively). Alternative age prediction models were generated by machine learning as described before [73]. In brief, we applied a penalized lasso and elastic net regression model from glmnet R package on the training set from BBA-seq data. The best-fitted candidate CpGs and model was chosen by 10-fold cross-validation on the training set (Additional file 1: Tables S11 - S14).

Association with CTCF binding sites
Chromatin immune precipitation sequencing data (ChIP-seq) for CTCF in A549 cell lines (GSM822289), H1 human embryonic stem cells (GSM822297), and K562 cell lines (GSM822311) were analyzed. Each CpG site was extended for 250 bp in both directions and quantile normalized read counts of the ChIP-Seq experiments were compared for the 65 age-associated CpG sites in comparison to 1000 randomly chosen CpGs from the 450 k array. Enrichment was estimated by Mann-Whitney rank test. The CTCF binding motifs around the target 65 age-associated CpG sites were predicted by RGT-motif analysis (www.regulatory-genomics.org/motif-analysis) with default parameters.

Simulation of stochastic DNAm patterns
We simulated randomly generated DNAm patterns under the assumption that methylation at neighboring CpGs occurred independently. The probability that the CpG site i of gene X is methylated in the generated patterns is given by a linear function $f_{X,S}(i)$, which was based on correlation of chronological age versus DNAm at i in the training set. For the simulations we used the function *random* from the *python 3* library *random*.

Epigenetic age predictions for individual BBA-seq reads
The following algorithm was developed to estimate epigenetic age based on the binary sequel of methylated and non-methylated CpGs within individual reads of BBA-seq data: Let X be a gene with n_X CpG sites. Using the training set, we estimated the probability for each CpG to be methylated at a given age. Using linear regression, we approximated the methylation frequency of a site i of gene X as a function of age a. This yields a set of n_X linear functions $F_{X,i}$ ($1 \leq i \leq n_X$), where $F_{X,i}(a)$ approximates the methylation frequency of site i of gene X at age a. Since the functions $F_{X,i}$ are linear, they approach infinity or minus infinity if a approaches infinity or minus infinity. Since methylation frequencies assume only values between zero and one, we defined a set of n_X functions $p_{X,i}$ ($1 \leq i \leq n_X$) as follows: $p_{X,i}(a) = F_{X,i}(a)$ if $0 \leq F_{X,i}(a) \leq 1$; if $F_{X,i}(a) < 0$, then $p_{X,i}(a) = 0$; if $F_{X,i}(a) > 1$, then $p_{X,i}(a) = 1$. We interpreted $p_{X,i}(a)$ as the probability that site i of gene X is methylated in a donor of age a. For a given methylation pattern P, we could then calculate the probability $Pr(P,a)$ that pattern P comes from a donor of age a. For this we assumed that methylation of different sites occurs independent from each other. Formally, we obtain $Pr(P,a) = q_1 \cdot \ldots \cdot q_{nx}$, where $q_i = p_{X,i}(a)$ if site i is methylated in pattern P and $q_i = 1 - p_{X,i}(a)$ otherwise. To each detected pattern, we assigned the age a_P that maximizes $Pr(P,a)$ (if the maximum is not unique we use the average of the ages that maximize $Pr(P,a)$). The estimated age of the donor corresponded to the average of the a_P, where we included a given value a_P multiple times if the pattern P is detected multiple times. To practically determine the a_P, we calculated $Pr(P,a)$ for a between 0 and 200 years in steps of 1 year.

Supplementary information

Additional file 1: Figure S1. Demarcation of age-associated CpGs in Illumina BeadChip datasets. **Figure S2.** Comparison with epigenetic ageing clocks of Horvath and Hannum. **Figure S3.** Intersection of probes among different epigenetic ageing clocks. **Figure S4.** Selection of 6 CpG markers for pyrosequencing. **Figure S5.** Comparison of DNAm levels in pyrosequencing and BBA-seq. **Figure S6.** Comparison of age-associated DNAm in ddPCR versus BBA-seq. **Figure S7.** Correlation of DNAm with age in buccal swabs. **Figure S8**. Targeted sequences of pyrosequencing assays. **Figure S9.** Targeted sequences for BBA-seq. **Table S1.** DNAm profiles for candidate CpG selection. **Table S2.** Multivariable 65 CpG model for Illumina BeadChip data. **Table S3.** Primer list for

---

Pyrosequencing assays. **Table S4.** Epigenetic aging signature based on pyrosequencing (6 CpG model). **Table S5.** Epigenetic aging signature based on pyrosequencing (9 CpG model). **Table S6.** Primer list for ddPCR assay. **Table S7.** Multivariable model for ddPCR (7 CpG). **Table S8.** Primer list for BBA-seq assay. **Table S9.** Multivariable model for BBA-seq of blood (9 CpG model). **Table S10.** Multivariable model for BBA-seq of buccal swabs (9 CpG model). **Table S11.** Machine learning model by Lasso (blood). **Table S12.** Machine learning model by ElasticNet (blood). **Table S13.** Machine learning model by Lasso (swab). **Table S14.** Machine learning model by ElasticNet (swab).

Authors' contributions
Y.H. performed pyrosequencing, BBA-seq, and data analysis; J.F., J.H., and M.N. supported BBA-seq data analysis; T.S. for the modeling of single-read prediction; M.G. performed droplet digital PCR and data analysis; and C-C.K. analyzed the CTCF ChIP-seq data. B.K. and S.R. contributed forensic human blood and buccal swab specimens; K.S. provided human blood samples; W.W. conceived the study and analyzed the data; Y.H. and W.W. wrote the manuscript; and all authors read and approved the final manuscript.

Availability of data and materials
In this study, we used publicly available 450 k Illumina BeadChip datasets that have been deposited at Gene Expression Omnibus (GEO) under the accession numbers GSE40279, GSE67705, GSE52588, GSE77445, GSE41169, GSE32148, GSE36064, GSE64495, GSE61496, GSE55763, GSE4286, and GSE125105 (also listed in Additional file 1: Table S1). Furthermore, BBA-seq data of this study is accessible at GEO under the accession number GSE151641.

Ethics approval and consent to participate
This study has been performed according to the guidelines approved by the local ethics committees of RWTH Aachen University (EK 041/15) and Heinrich Heine University of Düsseldorf (Permit number 4939).

Consent for publication
Not applicable

Competing interests
W.W. is cofounder of Cygenia GmbH that can provide service for Epigenetic-Aging-Signatures (www.cygenia.com). Y.H. and J.F. also contribute to this company. Apart from that, the authors declare that they have no competing interests.

Author details
[1]Helmholtz-Institute for Biomedical Engineering, Stem Cell Biology and Cellular Engineering, RWTH Aachen University Medical School, Pauwelsstraße 20, 52074 Aachen, Germany. [2]Institute for Biomedical Engineering – Cell Biology, University Hospital of RWTH Aachen, Aachen, Germany. [3]Interdisciplinary Center for Scientific Computing (IWR), Institute of Applied Mathematics, University of Heidelberg, Heidelberg, Germany. [4]Institute for Legal Medicine, Heinrich Heine University, Düsseldorf, Germany. [5]Institute for Transfusion Medicine, RWTH Aachen University Medical School, Aachen, Germany.

References
1. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD. DNA methylation clocks in aging: categories, causes, and consequences. Mol Cell. 2018;71:882–95.
2. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018;19:371.
3. Lin Q, Weidner CI, Costa IG, Marioni RE, Ferreira MR, Deary IJ, Wagner W. DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. Aging (Albany NY). 2016;8:394.
4. Zhang Y, Hapala J, Brenner H, Wagner W. Individual CpG sites that are associated with age and life expectancy become hypomethylated upon aging. Clin Epigenetics. 2017;9:9.
5. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol. 2015;16:25.
6. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY). 2019;11:303–27.
7. Lund JB, Li S, Baumbach J, Svane AM, Hjelmborg J, Christiansen L, Christensen K, Redmond P, Marioni RE, Deary IJ, Tan Q. DNA methylome profiling of all-cause mortality in comparison with age-associated methylation patterns. Clin Epigenetics. 2019;11:23.
8. Martin-Herranz DE, Aref-Eshghi E, Bonder MJ, Stubbs TM, Choufani S, Weksberg R, Stegle O, Sadikovic B, Reik W, Thornton JM. Screening for genes that accelerate the epigenetic aging clock in humans reveals a role for the H3K36 methyltransferase NSD1. Genome Biol. 2019;20:146.
9. Jeffries AR, Maroofian R, Salter CG, Chioza BA, Cross HE, Patton MA, Dempster E, Temple IK, Mackay DJG, Rezwan FI, et al. Growth disrupting mutations in epigenetic regulatory molecules are associated with abnormalities of epigenetic aging. Genome Res. 2019;29:1057–66.
10. Fiorito G, McCrory C, Robinson O, Carmeli C, Rosales CO, Zhang Y, Colicino E, Dugué P-A, Artaud F, GJ MK. Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis. Aging (Albany NY). 2019;11:2045.
11. Blueprint-consortium. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. Nat Biotechnol. 2016; 34:726–37.
12. Kristensen LS, Hansen LL. PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. Clin Chem. 2009;55:1471–83.
13. Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, Vilain E. Epigenetic predictor of age. PLoS One. 2011;6:e14821.
14. Koch CM, Wagner W. Epigenetic-aging-signature to determine age in different tissues. Aging (Albany NY). 2011;3:1018–27.
15. Wagner W. Epigenetic aging clocks in mice and men. Genome Biol. 2017;18:107.
16. Walker DL, Bhagwate AV, Baheti S, Smalley RL, Hilker CA, Sun Z, Cunningham JM. DNA methylation profiling: comparison of genome-wide sequencing methods and the Infinium Human Methylation 450 Bead Chip. Epigenomics. 2015;7:1287–302.
17. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49:359–67.
18. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, Bauerschlag DO, Jöckel K-H, Erbel R, Mühleisen TW. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol. 2014;15:R24.
19. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14:R115.
20. Zbieć-Piekarska R, Spólnicka M, Kupiec T, Parys-Proszek A, Makowska Ż, Pałeczka A, Kucharczyk K, Płoski R, Branicki W. Development of a forensically useful age prediction method based on DNA methylation analysis. Forensic Sci Int. 2015;17:173–9.
21. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S. Methylation of ELOVL 2 gene as a new epigenetic marker of age. Aging Cell. 2012;11:1132–4.
22. Hong SR, Jung S-E, Lee EH, Shin K-J, Yang WI, Lee HY. DNA methylation-based age prediction from saliva: high age predictability by combination of 7 CpG markers. Forensic Sci Int. 2017;29:118–25.

23. Lee HY, Jung S-E, Oh YN, Choi A, Yang WI, Shin K-J. Epigenetic age signatures in the forensically relevant body fluid of semen: a preliminary study. Forensic Sci Int. 2015;19:28–34.

24. Yu M, Carter KT, Makar KW, Vickers K, Ulrich CM, Schoen RE, Brenner D, Markowitz SD, Grady WM. MethyLight droplet digital PCR for detection and absolute quantification of infrequently methylated alleles. Epigenetics. 2015; 10:803–9.

25. Zemmour H, Planer D, Magenheim J, Moss J, Neiman D, Gilon D, Korach A, Glaser B, Shemer R, Landesberg G. Non-invasive detection of human cardiomyocyte death using methylation patterns of circulating DNA. Nat Commun. 2018;9:1443.

26. Bernstein DL, Kameswaran V, Le Lay JE, Sheaffer KL, Kaestner KH. The BisPCR 2 method for targeted bisulfite sequencing. Epigenetics Chromatin. 2015;8:27.

27. Theophilou G, Paraskevaidi M, Lima KM, Kyrgiou M, Martin-Hirsch PL, Martin FL. Extracting biomarkers of commitment to cancer development: potential role of vibrational spectroscopy in systems biology. Expert Rev Mol Diagn. 2015;15:693–713.

28. Smith AM, Heisler LE, St. Onge RP, Farias-Hesson E, Wallace IM, Bodeau J, Harris AN, Perry KM, Giaever G, Pourmand N. Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. Nucleic Acids Res. 2010;38:e142.

29. Franzen J, Zirkel A, Blake J, Rath B, Benes V, Papantonis A, Wagner W. Senescence-associated DNA methylation is stochastically acquired in subpopulations of mesenchymal stem cells. Aging Cell. 2017;16:183–91.

30. Shi L, Jiang F, Ouyang F, Zhang J, Wang Z, Shen X. DNA methylation markers in combination with skeletal and dental ages to improve age estimation in children. Forensic Sci Int. 2018;33:1–9.

31. Aliferi A, Ballard D, Gallidabino MD, Thurtle H, Barron L, Court DS. DNA methylation-based age prediction using massively parallel sequencing data and multiple machine learning models. Forensic Sci Int. 2018;37:215–26.

32. Naue J, Hoefsloot HC, Mook OR, Rijlaarsdam-Hoekstra L, van der Zwalm MC, Henneman P, Kloosterman AD, Verschure PJ. Chronological age prediction based on DNA methylation: massive parallel sequencing and random forest regression. Forensic Sci Int. 2017;31:19–28.

33. Kalwa M, Hanzelmann S, Otto S, Kuo CC, Franzen J, Joussen S, Fernandez-Rebollo E, Rath B, Koch C, Hofmann A, et al. The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. Nucleic Acids Res. 2016; 44:10631–43.

34. Bozic T, Frobel J, Raic A, Ticconi F, Kuo CC, Heilmann-Heimbach S, Goecke TW, Zenke M, Jost E, Costa IG, Wagner W. Variants of DNMT3A cause transcript-specific DNA methylation patterns and affect hematopoiesis. Life Sci Alliance. 2018;1:e201800153.

35. Wiehle L, Thorn GJ, Raddatz G, Clarkson CT, Rippe K, Lyko F, Breiling A, Teif VB. DNA (de) methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. Genome Res. 2019;29:750–61.

36. Issa J-P. Aging and epigenetic drift: a vicious cycle. J Clin Invest. 2014;124: 24–9.

37. Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Bürkle A, Caiafa P. Reconfiguration of DNA methylation in aging. Mech Ageing Dev. 2015;151:60–70.

38. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A. 2005; 102:10604–9.

39. Y-A C, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8:203–9.

40. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res. 2017;45:e22.

41. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15:R31.

42. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, Soderhall C, Scheynius A, Kere J. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One. 2012;7:e41361.

43. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenetics. 2015;7:113.

44. Teschendorff AE, Yang Z, Wong A, Pipinikas CP, Jiao Y, Jones A, Anjum S, Hardy R, Salvesen HB, Thirlwell C. Correlation of smoking-associated DNA

45. methylation changes in buccal cells with DNA methylation changes in epithelial cancer. JAMA Oncol. 2015;1:476–85.

45. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, Warren ST. Age-associated DNA methylation in pediatric populations. Genome Res. 2012;22:623–32.

46. Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, Felton S, Matsuyama M, Lowe D, Kabacik S. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. Aging (Albany NY). 2018;10:1758.

47. Eipel M, Mayer F, Arent T, Ferreira MR, Birkhofer C, Gerstenmaier U, Costa IG, Ritz-Timme S, Wagner W. Epigenetic age predictions based on buccal swabs are more precise in combination with cell type-specific DNA methylation signatures. Aging (Albany NY). 2016;8:1034–48.

48. Ong C-T, Corces VG. CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet. 2014;15:234.

49. Snir S, Farrell C, Pellegrini M. Human epigenetic aging is logarithmic with time across the entire lifespan. Epigenetics. 2019;14:912–26.

50. Vidaki A, Kayser M. Recent progress, methods and perspectives in forensic epigenetics. Forensic Sci Int. 2018;37:180–95.

51. Pharo HD, Andresen K, Berg KC, Lothe RA, Jeanmougin M, Lind GE. A robust internal control for high-precision DNA methylation analyses by droplet digital PCR. Clin Epigenetics. 2018;10:24.

52. Van Wesenbeeck L, Janssens L, Meeuws H, Lagatie O, Stuyver L. Droplet digital PCR is an accurate method to assess methylation status on FFPE samples. Epigenetics. 2018;13:207–13.

53. Wang Y, Karlsson R, Lampa E, Zhang Q, Hedman ÅK, Almgren M, Almqvist C, McRae AF, Marioni RE, Ingelsson E. Epigenetic influences on aging: a longitudinal genome-wide methylation study in old Swedish twins. Epigenetics. 2018;13:975–87.

54. Yuan T, Jiao Y, de JS ORA, Beck S, Teschendorff AE. An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging. PLoS Genet. 2015;11:e1004996.

55. Day K, Waite LL, Thalacker-Mercer A, West A, Bamman MM, Brooks JD, Myers RM, Absher D. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Genome Biol. 2013;14:R102.

56. Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, Burke G, Post W, Shea S, Jacobs DR Jr, et al. Age-related variations in the methylome associated with gene expression in human monocytes and T cells. Nat Commun. 2014;5:5366.

57. Lai AY, Fatemi M, Dhasarathy A, Malone C, Sobol SE, Geigerman C, Jaye DL, Mav D, Shah R, Li L. DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. J Exp Med. 2010;207:1939–50.

58. Chang J, Zhang B, Heath H, Galjart N, Wang X, Milbrandt J. Nicotinamide adenine dinucleotide (NAD)–regulated DNA methylation alters CCCTC-binding factor (CTCF)/cohesin binding and transcription at the BDNF locus. Proc Natl Acad Sci. 2010;107:21836–41.

59. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011;480:490.

60. Franzen J, Georgomanolis T, Selich A, Stoeger R, Brant L, Fernandez-Rebollo E, Grezella C, Ostrowska A, Begemann M, Rath B. DNA methylation patterns of replicative senescence are strand-specific and reflect changes in chromatin conformation. Preprint at bioRxiv. 2018:445114. https://doi.org/10.1101/445114.

61. Farlik M, Sheffield NC, Nuzzo A, Datlinger P, Schönegger A, Klughammer J, Bock C. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Rep. 2015;10:1386–97.

62. Lin Q, Wagner W. Epigenetic aging signatures are coherently modified in cancer. PLoS Genet. 2015;11:e1005334.

63. Gross AM, Jaeger PA, Kreisberg JF, Licon K, Jepsen KL, Khosroheidari M, Morsey BM, Swindells S, Shen H, Ng CT. Methylome-wide analysis of chronic HIV infection reveals five-year increase in biological age and epigenetic targeting of HLA. Mol Cell. 2016;62:157–68.

64. Bacalini MG, Gentilini D, Boattini A, Giampieri E, Pirazzini C, Giuliani C, Fontanesi E, Scurti M, Remondini D, Capri M. Identification of a DNA methylation signature in blood cells from persons with Down Syndrome. Aging (Albany NY). 2015;7:82.

65. Houtepen LC, Vinkers CH, Carrillo-Roa T, Hiemstra M, Van Lier PA, Meeus W, Branje S, Heim CM, Nemeroff CB, Mill J. Genome-wide DNA methylation levels and altered cortisol stress reactivity following childhood trauma in humans. Nat Commun. 2016;7:10967.

66. Horvath S, Levine AJ. HIV-1 infection accelerates age according to the epigenetic clock. J Infect Dis. 2015;212:1563–73.
67. Harris AR, Nagy-Szakal D, Pedersen N, Opekun A, Bronsky J, Munkholm P, Jespersgaard C, Andersen P, Melegh B, Ferry G. Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. Inflamm Bowel Dis. 2012;18:2334–41.
68. Walker RF, Liu JS, Peters BA, Ritz BR, Wu T, Ophoff RA, Horvath S. Epigenetic age analysis of children who seem to evade aging. Aging (Albany NY). 2015;7:334.
69. Tan Q, Frost M, Heijmans BT, von Bornemann HJ, Tobi EW, Christensen K, Christiansen L. Epigenetic signature of birth weight discordance in adult twins. BMC Genomics. 2014;15:1062.
70. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan S-T, Afzal U, Scott J, Jarvelin M-R, Elliott P. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol. 2015;16:37.
71. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol. 2013;31:142.
72. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. bioinformatics. 2011;27:1571–2.
73. Han Y, Eipel M, Franzen J, Sakk V, Dethmers-Ausema B, Yndriago L, Izeta A, de Haan G, Geiger H, Wagner W. Epigenetic age-predictor for mice based on three CpG sites. eLife. 2018;7:e37462.

Publisher's Note