University of Groningen

Targeted methods for epigenetic age predictions in mice

Han, Yang; Nikolic, Milos; Gobs, Michael; Franzen, Julia; de Haan, Gerald; Geiger, Hartmut; Wagner, Wolfgang

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

scientific reports

Check for updates

OPEN

Targeted methods for epigenetic age predictions in mice

Yang Han[1,2], Miloš Nikolić[1,2], Michael Gobs[1,2], Julia Franzen[1,2], Gerald de Haan[3], Hartmut Geiger[4] & Wolfgang Wagner[1,2 ✉]

Age-associated DNA methylation reflects aspect of biological aging—therefore epigenetic clocks for mice can elucidate how the aging process in this model organism is affected by specific treatments or genetic background. Initially, age-predictors for mice were trained for genome-wide DNA methylation profiles and we have recently described a targeted assay based on pyrosequencing of DNA methylation at only three age-associated genomic regions. Here, we established alternative approaches using droplet digital PCR (ddPCR) and barcoded bisulfite amplicon sequencing (BBA-seq). At individual CG dinucleotides (CpGs) the correlation of DNA methylation with chronological age was slightly higher for pyrosequencing and ddPCR as compared to BBA-seq. On the other hand, BBA-seq revealed that neighboring CpGs tend to be stochastically modified at murine age-associated regions. Furthermore, the binary sequel of methylated and non-methylated CpGs in individual reads can be used for single-read predictions, which may reflect heterogeneity in epigenetic aging. In comparison to C57BL/6 mice the single-read age-predictions using BBA-seq were also accelerated in the shorter-lived DBA/2 mice, and in C57BL/6 mice with a lifespan quantitative trait locus of DBA/2 mice. Taken together, we describe alternative targeted methods for epigenetic age predictions that provide new perspectives for aging-intervention studies in mice.

Aging evokes dynamic changes in DNA methylation (DNAm) at specific CG dinucleotides (CpG)[1]. These epigenetic modifications provide a biomarker for the aging process, which is often referred to as 'epigenetic clock'[2]. They were initially described for humans based on data from Illumina BeadChips[3,4], and in the advent of a fast growing number of such datasets the models were further refined—with signatures of many age-associated CpGs—to provide a very high correlation of predicted and chronological age. Notably, epigenetic clocks for blood seem to reflect aspects of biological age, since the deviation of predicted and chronological age (delta-age) correlates with all-cause mortality[5,6] and it is increased in various diseases, such as obesity[7], Down syndrome[8], Werner Syndrome[9], and HIV infection[10]. Thus, tracking of epigenetic age may also elucidate the impact of drugs or other relevant parameters for the aging process, albeit it is challenging to perform such controlled and long-term aging intervention studies in humans[11].

Mice are one of the most popular mammalian models for aging research. Inbreeding, defined growth conditions, and the shorter life-length of about two years facilitate aging interventions studies with mice that cannot be easily performed in humans. Epigenetic clocks for mice were initially based on whole genome bisulfite sequencing (WGBS) or reduced representation bisulfite sequencing (RRBS)[12]. They were trained for liver, whole blood, or even multi-tissue specimens from mice using hundreds of CpG sites, and they clearly demonstrated that epigenetic clocks in mice are affected by genetic, dietary, or pharmacological interventions[13–15]. However, WGBS and RRBS are relatively labor and cost-intensive and the methods do not always provide enough coverage for all the relevant CpGs, which hampers application of these age-predictors.

To overcome this problems, alternative methods for site-specific analysis of DNAm at few selected age-associated CpGs may be advantageous[12,16]. We have recently described an epigenetic clock that is based on pyrosequencing of DNAm at only three age-associated CpGs to facilitate a high accuracy with chronological age in C57BL/6 mice[17]. Notably, epigenetic aging was significantly accelerated in the shorter-lived DBA/2 mice[17], and in congenic C57BL/6 mice harboring regions of chromosome 11 from DBA/2 mice likely linked to the regulation of lifespan (referred to as Line A mice)[18]. The epigenetic age was also decelerated by systemic administration of a drug that extended murine lifespan[19], implying that the three CpGs might also serve as biomarkers of aging at

[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]Laboratory of Ageing Biology and Stem Cells, European Research Institute for the Biology of Ageing, University Medical Center Groningen, Groningen, the Netherlands. [4]Institute of Molecular Medicine, Ulm University, 89081 Ulm, Germany. ✉email: wwagner@ukaachen.de
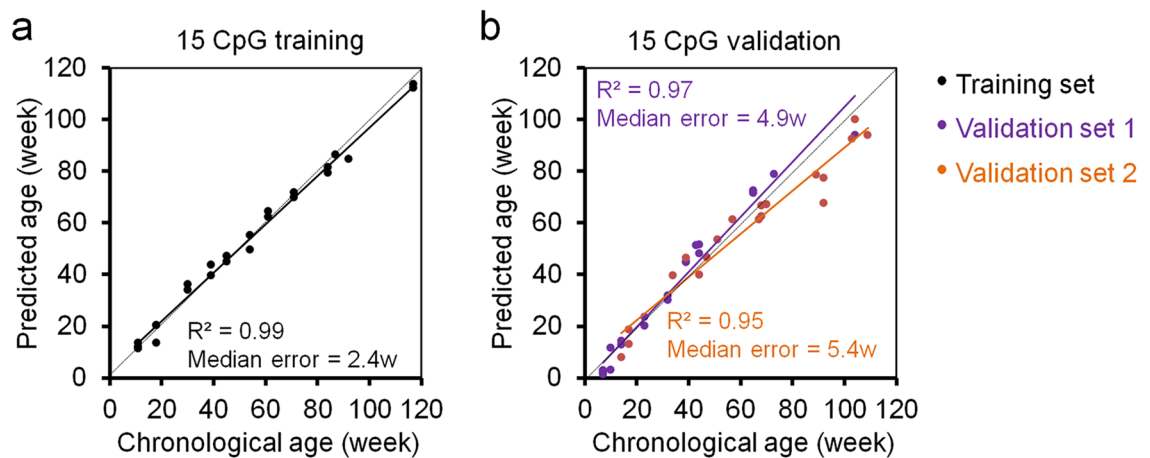
Figure 1. Epigenetic age predictions for pyrosequencing data (15 CpG lasso regression model). (**a**) Multivariable machine learning (Lasso regression) age-predictor based on DNAm levels at 15 CpGs in the genes *Prima1, Hsf4, Kcns1, Gm9312, Gm7325*. Pyrosequencing was performed for 24 C57BL/6 mice (training set) as described before[17]. (**b**) Age predictions with the same model in two independent validation sets: 21 C57BL/6 mice from the University of Ulm and 19 C57BL/6 mice from the University of Groningen (validation sets 1 and 2, respectively). Coefficients of determination (R^2) of DNAm *versus* chronological age and median errors (weeks) are indicated.

least on an C57BL/6 background. While the pyrosequencing based epigenetic clock has proven to be robust and reliable, it is well conceivable that precision, accuracy and applicability can be increased by alternative methods.

Droplet Digital PCR (ddPCR) is a relatively novel targeted approach for DNAm measurement that was reported to provide precise results with less PCR bias[20,21]. Furthermore, barcoded bisulfite amplicon sequencing (BBA-seq), which is based on massive-parallel-sequencing, facilitates DNAm analysis of longer amplicons with more neighboring CpGs and provides insight into the DNAm pattern on individual DNA strands[22]. We have recently demonstrated in BBA-seq data of human blood that the correlation of age with DNAm levels at neighboring CpGs follows a bell-shaped curved[21]. Interestingly, the DNAm pattern of neighboring CpGs was not coherently modified on individual strands, as might be anticipated upon binding of an epigenetic writer, but rather seemed to be evoked by stochastic modifications[21]. Based on this, we developed an epigenetic age-predictor for BBA-seq data of human blood, which was based on the binary sequel of methylated and non-methylated sites in individual reads[21]. This approach might reflect heterogeneity of epigenetic aging within a sample. In this study, we now established and compared such targeted epigenetic clocks also for mice, which are based on pyrosequencing, ddPCR, BBA-seq, or single read predictions.

Results

Alternative epigenetic clocks based on pyrosequencing. In our previous work, we selected nine age-associated genomic regions, which were initially identified for age-predictors based on genome-wide deep-sequencing of DNAm profiles[14,15]. Based on this, we established a 3 CpG model for pyrosequencing measurements in the genes proline rich membrane anchor 1 (*Prima1*), heat shock transcription factor 4 (*Hsf4*) and potassium voltage-gated channel modifier subfamily S member 1 (*Kcns1*)[17]. Age-predictions correlated very well with the chronological age of C57BL/6 mice in a training set (n = 24; R^2 = 0.96; Median error = 3.6 weeks) and in two independent validation sets (n = 21 and 19; R^2 = 0.95 and 0.91; Median error = 5.0 and 5.9 weeks, respectively). We initially also described a 15 CpG model, which considered two additional amplicons of the pseudogene *Gm9312* and myoblast fusion factor (*Gm7325*)[17]. This 15 CpGs model was identified by machine learning and although it provided higher accuracy in the training set (R^2 = 0.99; Median error = 2.4 weeks), this model was not further validated as we anticipated that the very good correlation might rather be due to overfitting[17]. In the present study, we further explored this 15 CpG model by pyrosequencing for the two independent validation sets of C57BL/6 mice (n = 21 and n = 19). In fact, the 15 CpG clock gave slightly better correlation with chronological age and lower prediction error (R^2 = 0.97 and R^2 = 0.95; median error = 4.9 weeks and 5.4 weeks, respectively) than the 3 CpG signature (Fig. 1). Thus, the 15 CpG murine epigenetic aging clock seems to be advantageous, while the need of two additional PCR amplicons and pyrosequencing measurements provides a tradeoff between accuracy and costs.

Age-prediction with droplet digital PCR. Droplet digital PCR (ddPCR) is based on parallel PCR reactions in thousands of micro-droplets and therefore DNAm analysis with this technology may reduce PCR bias for methylated/non-methylated strands that may occur in pyrosequencing (Supplementary Fig. S1a)[21]. Therefore, we have designed ddPCR assay for the same three amplicons for *Prima1, Hsf4,* and *Kcns1*. However, the targeted CpG within the *Hsf4* amplicon was different to the pyrosequencing based 3 CpG predictor, as this was better suitable for the ddPCR probe. DNAm measurements with ddPCR at all three CpGs revealed high correlation with chronological age in 23 C57BL/6 mice of the training set (Fig. 2a–c) and correlated with the
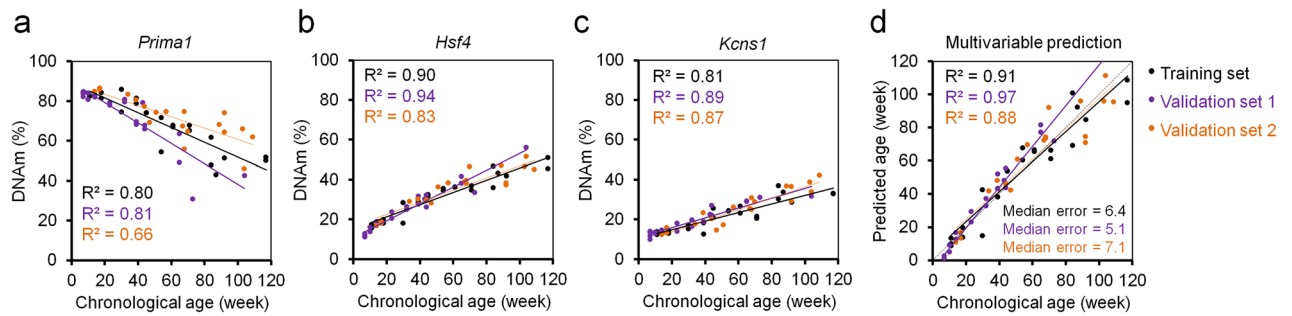
Figure 2. Three CpG epigenetic clock for mice based on droplet digital PCR. Age-associated DNAm was measured with ddPCR at 3 CpGs in the genes *Prima1* (**a**), *Hsf4* (**b**) and *Kcns1* (**c**) in the training set (n = 23) and two independent validation sets (n = 21 and 19) of C57BL/6 mice. (**d**) The measurements of the training set were used for a multivariable model for epigenetic age predictions. Coefficients of determination (R^2) of DNAm *versus* chronological age and median errors (weeks) are demonstrated.

	Prima1	**Hsf4[a]**	**Kcns1[a]**	**Mean R²**
Pyrosequencing				
Training (n = 24)	0.71	0.96	0.84	0.84
Validation1 (n = 19)	0.79	0.96	0.81	0.85
Validation2 (n = 21)	0.69	0.89	0.86	0.81
ddPCR				
Training (n = 23)	0.8	0.9	0.81	0.84
Validation1 (n = 19)	0.81	0.94	0.89	0.88
Validation2 (n = 21)	0.66	0.83	0.87	0.79
BBA-seq				
Training (n = 23)	0.78	0.87	0.85	0.83
Validation1 (n = 19)	0.78	0.91	0.78	0.82
Validation2 (n = 21)	0.64	0.75	0.85	0.75

Table 1. Correlation of DNAm and chronological age in different targeted approaches. [a]CpGs from the amplicons were always selected by the highest Pearson correlation with chronological age, therefore they are not identical in the different sequencing approaches.

DNAm measurements by pyrosequencing (Supplementary Fig. S1b). Based on the ddPCR measurements we determined a multivariable linear regression model that provided reliable age-predictions in the validation sets (R^2 = 0.97 and 0.88; median error 5.1 and 7.1 weeks). These results were slightly less accurate than for the 3 CpG clock by pyrosequencing (Fig. 2d), which might be due to lower age-association in the neighbouring CpGs of *Hsf4*. Either way, the results demonstrate that DNAm measurements with ddPCR are also well suited for epigenetic clocks in mice.

Barcoded bisulfite amplicon sequencing of age-associated regions. Subsequently, we used barcoded bisulfite amplicon sequencing (BBA-seq) to investigate age-associated DNAm in amplicons of *Prima1*, *Hsf4* and *Kcns1*, which covered 4, 12, and 21 neighboring CpGs, respectively. Overall, DNAm measurements correlated in BBA-seq *versus* pyrosequencing (Supplementary Fig. S2), albeit slightly less than ddPCR *versus* pyrosequencing (Supplementary Fig. S1b). Furthermore, the correlation at individual CpGs with chronological age was slightly lower in BBA-seq as compared to pyrosequencing or ddPCR (Table 1). Either way, the three relevant or neighboring CpGs of the pyrosequencing clock also provided a high correlation with chronological age (Fig. 3a-c). The BBA-seq measurements of these three CpGs were then used to train a multivariable linear model and the age-predictions correlated well in the validation sets 1 and 2 (n = 21 and 19; R^2 = 0.95 and 0.91; median error = 6.6 and 10 weeks; Fig. 3d). Alternatively, we considered all CpGs of the three amplicons to generate a Lasso regression model with tenfold cross-validation that considered 7 CpG sites of the three amplicons. The accuracy of age-predictions with this machine learning based model revealed slightly lower median error for the validation sets (n = 21 and 19; R^2 = 0.91 and 0.90; median error = 6.1 and 5.9; Fig. 3e). Taken together, BBA-seq provided similar accuracy in epigenetic age-predictions as pyrosequencing and ddPCR.

Subsequently, we analyzed how DNAm at neighboring CpGs correlates with chronological age. For each CpG within the BBA-seq amplicons of *Prima1*, *Hsf4* and *Kcns1* we determined the correlation with chronological age in the training and validation sets (Fig. 3f–h). This analysis revealed that not only the individual CpGs of our age predictor are age-associated, but also the CpGs in direct vicinity, which is in line with our recent analysis in humans[21].
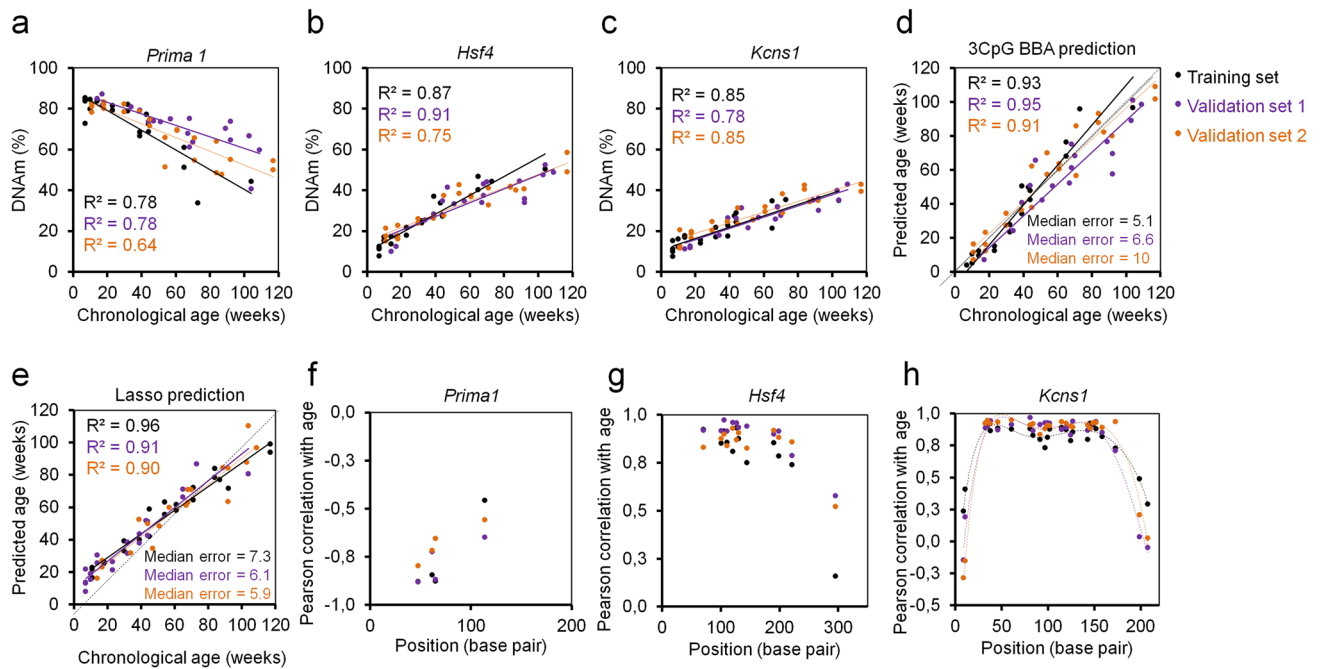
Figure 3. Epigenetic age-prediction by BBA-seq. DNAm levels (%) of three highly age-associated CpGs within three amplicons *Prima1* (**a**), *Hsf4* (**b**) and *Kcns1* (**c**) were determined by barcoded bisulfite amplicon sequencing (BBA-seq). (**d**) Age predictions based on the multivariable linear regression model of three CpGs in the C57BL/6 mice. (**e**) Age predictions with a lasso regression model (7 CpGs in the three age-associated regions), which was trained on the training set of C57BL/6 mice. Coefficients of determination (R^2) of DNAm *versus* chronological age and median errors (weeks) are indicated. (**f–h**) Pearson's correlations of age with DNAm levels of CpGs within the amplicons of *Prima1*, *Hsf4*, and *Kcns1* are plotted for the blood samples of the training set (n = 23) and two independent validation sets (n = 21 and 19). The x-axis represents the position of CpGs within the amplicons.

Epigenetic age predictions for mice based on individual BBA sequencing reads. In contrast to pyrosequencing or ddPCR, BBA-seq provides individual reads with a binary sequel of either methylated or non-methylated CpGs. Heatmaps of DNAm within individual reads indicated that the methylation at neighboring CpGs occurs rather independent of each other (Fig. 4a and Supplementary Fig. S3a). In fact, Pearson´s correlation of DNAm levels between neighboring CpG sites within the three amplicons revealed only moderate correlation in epigenetic modifications (Fig. 4b and Supplementary Fig. S3b), albeit it was slightly higher than previously observed for BBA-seq data in three human age-associated regions[21].

For human BBA-seq data we have recently demonstrated that it is possible to estimate the epigenetic age for individual reads, under the assumption that the age-associated modification of DNAm occurs independently at neighboring CpGs. The mean of all individual read-predictions within a sample correlated with the chronological age (Han et al. 2020). Here, we have analyzed if this was also applicable for murine BBA-seq data. For each BBA-seq read of the three amplicons (*Prima1*, *Hsf4* and *Kcns1*) we estimated the epigenetic age based on the binary sequel of methylated and non-methylated CpGs, using the age-associated correlations at individual CpGs of the training set. Individual reads were predicted between 0 and 200 weeks (Fig. 4c and Supplementary Fig. S3c), which might resemble heterogeneity in epigenetic aging within a given sample. Overall, the 'young' reads were more frequent in young donors, whereas 'old' reads were more frequent in old mice. Notably, the mean of single-read predictions within a sample correlated with the chronological age of the mice (Fig. 4d). Particularly for the amplicons of *Hsf4* and *Kcns1*, which harbor more neighboring CpGs, the mean of single read-predictions correlated good or even better than the DNAm levels at the individual age-associated CpGs (Table 1). Thus, it is possible to estimate the epigenetic age by the binary sequel of methylated and non-methylated CpGs on individual DNA strands, which might also be used as a surrogate for the heterogeneity of epigenetic age within a sample.

Genetic background impacts on epigenetic age-predictions of mice. We have previously demonstrated, that epigenetic age-predictions with our 3 CpG pyrosequencing age-predictor are accelerated in DBA/2 mice, as compared to C57BL/6 mice, which may reflect the different life expectancy of these mouse strains (Han et al., 2018). Furthermore, we demonstrated that age-predictions with this predictor were also accelerated in C57BL/6 mice with quantitative trait locus insertion from DBA/2 into the congenic C57BL/6 chromosome 11, which was expected to be associated with the shorter lifespan of DBA/2 (referred to as Line A mice)[18]. We now determined within the same samples whether the epigenetic age-acceleration was also observed in DBA/2 mice (n = 33) and Line A mice (n = 15) using the BBA-seq approach. In fact, the predictions with either the 3 CpG
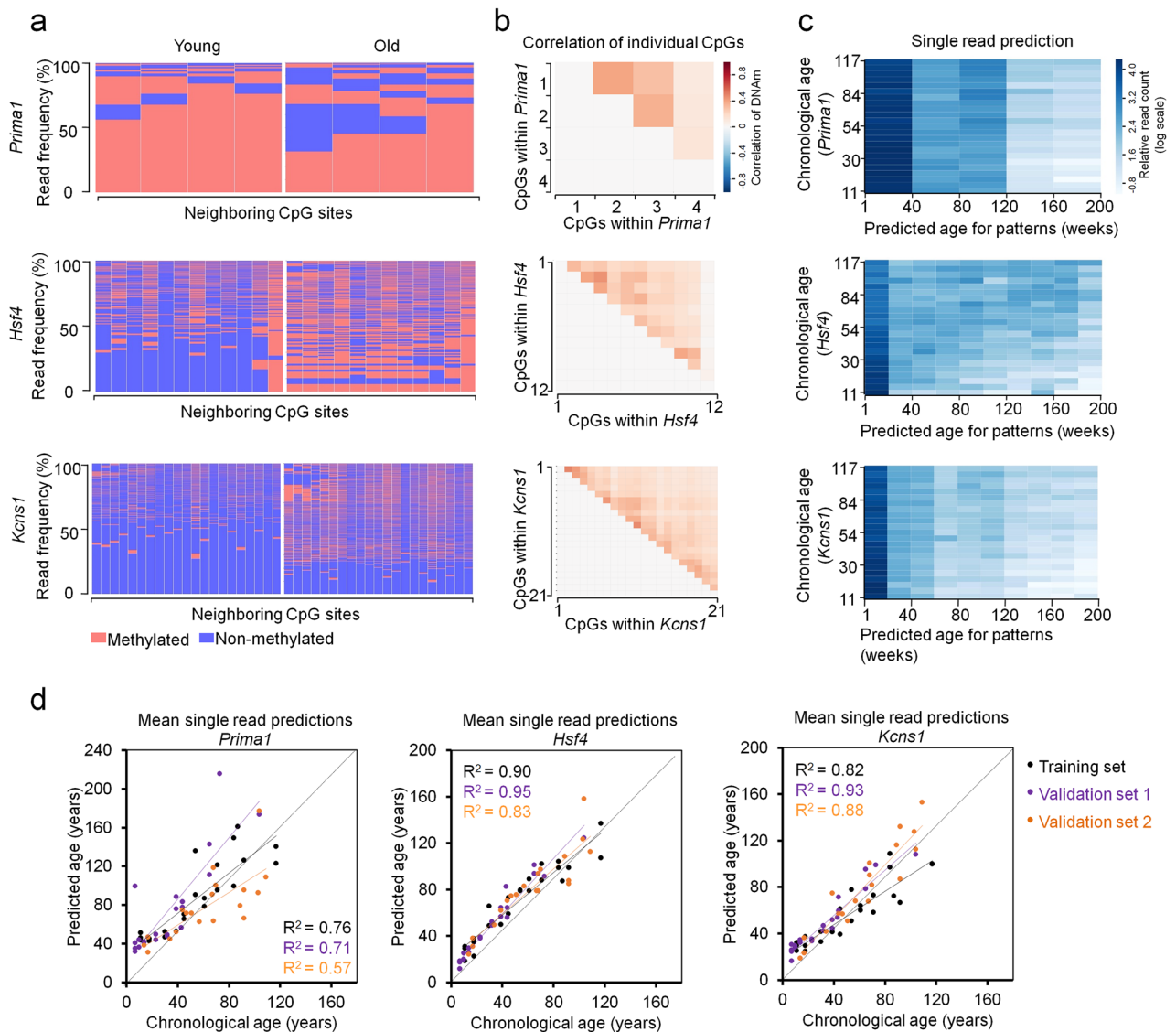
Figure 4. Analysis of age-associated DNAm patterns within individual BBA-seq reads in C57BL/6 mice. (**a**) Frequencies of DNAm patterns in BBA-seq reads (red: methylated; blue: non-methylated) within the amplicons of *Prima1* (4 neighboring CpGs), *Hsf4* (12 neighboring CpGs) and *Kcns1* (21 neighboring CpGs). Samples of one young (11 weeks) and one old C57BL/6 mouse (117 weeks) from the training set are exemplarily depicted. (**b**) Pearson correlation of DNAm among neighboring CpGs within each of the three amplicons in BBA-seq data of the training set. (**c**) Epigenetic ages were estimated for each individual read of the BBA-seq data the training set (n = 23). These single-read predictions were performed for each amplicon based on the binary sequel of methylated and non-methylated CpGs. The heatmaps depict the relative frequency of reads (normalized by the read counts per sample; log scale) that are classified to a specific age category (between 0 and 200 weeks) for each donor in the training set. (**d**) The mean age-predictions based on individual BBA-seq reads of three amplicons were determined for each sample and then plotted against the chronological age of the training (n = 23) and two validation sets (n = 21 and n = 19) of C57BL/6 mice.

BBA-seq, or the 7 CpG BBA-seq Lasso-regression model, provided very similar results as previously observed for the 3 CpG pyrosequencing clock (Fig. 5a,b).

Subsequently, we analyzed the single read patterns of BBA-seq data in DBA/2 and Line A mice. We observed the same random gain or loss of DNAm at neighboring CpGs (Fig. 6a) and a moderate correlation in DNAm at neighboring CpGs (Fig. 6b), as previously observed for C57BL/6 mice. Furthermore, single read predictions within the three amplicons for *Prima1*, *Hsf4* and *Kcns1* (based on the training set of C57BL/6 mice) provided similar heterogeneity and acceleration of age-estimations (Fig. 6c,d). These results indicated that epigenetic aging is generally accelerated within the three age-associated regions in DBA/2 and Line A mice, as compared to C57BL/6 mice.
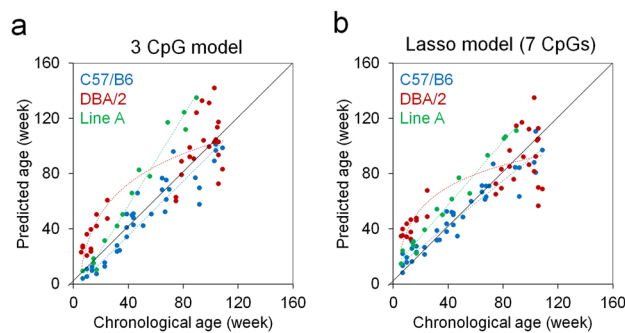
Figure 5. Age-predictions with BBA-seq in mice from different genetic background. DNAm levels were analyzed with BBA-seq in blood samples of 40 C57BL/6 mice of the validation sets, 33 DBA/2 mice, and 15 transgenic C57BL/6 mice with an age-associated region from DBA/2 mice (Line A)[18]. For epigenetic age predictions we either used (**a**) the 3 CpG multivariable model, or (**b**) the lasso regression model based on 7 CpGs of the same three amplicons (*Prima1*, *Hsf4*, *Kcns1)*. As previously described for pyrosequencing, epigenetic age-predictions were logarithmically accelerated in DBA/2 mice[17], and also accelerated in Line A mice[18].

Discussion

Epigenetic clocks are used as a surrogate marker for the process of biological aging. They are therefore valuable tools to gain insight into effects of aging or rejuvenating interventions[23]. To this end, the murine model system enables much better standardization over the life-time than achievable in humans[24]. The targeted assays for epigenetic clocks are easier applicable than epigenetic clocks that are based on genome wide RRBS or WGBS profiles. A bottleneck for the latter is often a low coverage of reads for specific CpG sites. Particularly pyrosequencing and ddPCR seem to provide more precise measures for DNAm levels at individual CpGs[25]. Furthermore, targeted analysis of DNAm only at specific CpGs is faster, facilitates better standardization of procedures, and it is more cost-effective than genome-wide approaches[12,26]. Thus, the targeted assays may be particularly advantageous for larger intervention studies. On the other hand, the number of CpGs to be implemented into epigenetic clocks provides a tradeoff between accuracy, which is generally increased with more age-associated CpGs, *versus* applicability and costs. This is also reflected by the comparison of the 3 CpG *versus* 15 CpG pyrosequencing models. In this regard, larger signatures that are based on genome wide DNAm profiles may be advantageous.

It is not trivial to directly compare the performance of our targeted epigenetic clocks with the other published predictors for WGBS or RRBS data, since the tissues, age-ranges, and methods vary considerably in these studies[13–15,27]. The previously published RRBS and WGBS clocks revealed high precision in the training sets, which markedly decreased when tested on independent samples. For murine blood samples, the blood clock by Petkovich et al. showed the best performance with MAE (mean absolute error) of 8.6 weeks[27]. Our targeted approaches provided similar or sometimes even slightly better accuracy, with an MAE ranging from 4.6 to 12 weeks (or median error 4.9 to 10 weeks).

In our previous work, we demonstrated that robust and reliable epigenetic age-predictions can be achieved by pyrosequencing at three CpGs[17]. We anticipated that the very high correlation of a 15 CpG lasso regression model, which was suggested during the review process, might be due to overfitting with the relatively small training set[17]. In the current study, we revisited this model to demonstrate that it indeed provides higher accuracy and precision than the 3 CpG predictor—however, it also necessitates pyrosequencing of two additional amplicons. It therefore depends on the experimental design and resources which of the pyrosequencing clocks is better suited.

Upon bisulfite conversion, there is a difference in the sequence of methylated and non-methylated DNA and this can entail a PCR bias[28]. Such DNAm sensitive PCR bias might be reduced by ddPCR, since it relies on detection of either methylated or non-methylated DNA in individual droplets, rather than the amplification efficiency[29]. So far, ddPCR is particularly applied for detection and quantification of genetic aberrations. Several studies demonstrated that it also enables precise measurements of DNAm levels[20,29,30], while only few recent studies reported ddPCR assays for epigenetic clocks in humans[21,31]. A major challenge for the establishment of such assays is the design of reliable and specific primers and probes for the bisulfite converted DNA sequences. In this study, we describe a 3 CpG ddPCR assay, that facilitates similar accuracy in age-predictions as the previously described 3 CpG pyrosequencing assay.

Next generation sequencing platforms enable targeted DNAm analysis in a barcoded manner for multiple samples in parallel[21,32]. In this study, we describe that BBA-seq of only three age-associated regions facilitates also reliable epigenetic age-predictions for murine blood samples. Advantages of this approach are the very high coverage and the relatively long target regions (up to 500 base pairs), which may cover more neighboring CpGs than pyrosequencing or ddPCR[22]. Our results confirmed that the correlation of chronological age with DNAm levels follows a bell-shaped curve at neighboring CpGs within about 200 to 400 bases of BBA-seq amplicons[21] – this was particularly observed in amplicons of *Hsf4* and *Kcns1* that comprised more neighboring CpGs. On the other hand, within individual BBA-seq reads there was only a moderate correlation of DNAm at neighboring CpGs. This is further substantiated by the mean single read predictions which clearly correlate with chronological age. Thus, our results support the notion that age-associated genomic regions favor a stochastic accumulation of
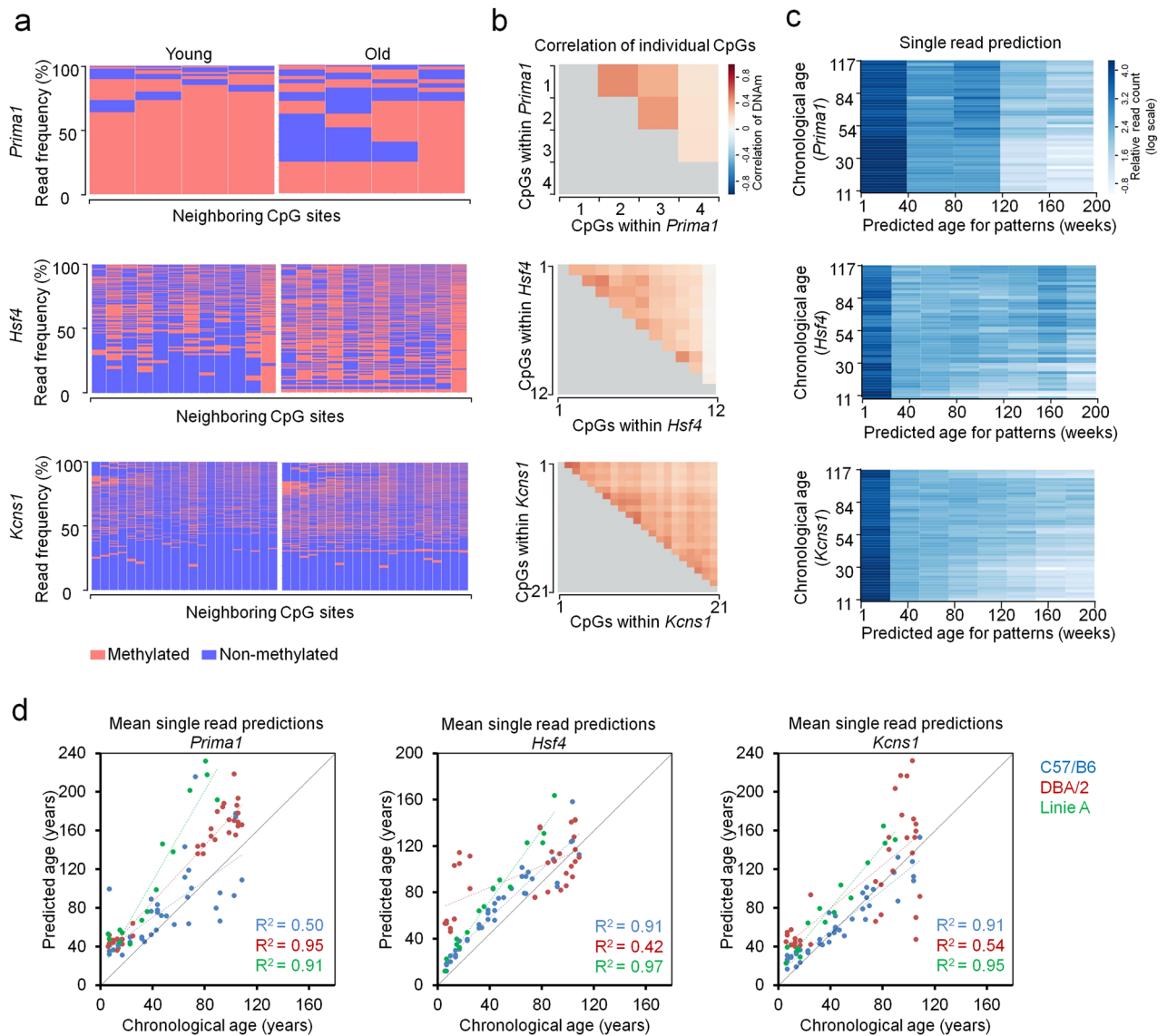
Figure 6. Analysis of age-associated DNAm patterns within individual BBA-seq reads in mice from different genetic background. (**a**) The plots exemplarily display the frequency of DNAm patterns in two DBA/2 mice: one young (7 weeks) and one old DBA/2 mouse(109 weeks). The frequencies of patterns within the amplicons of *Prima1*, *Hsf4* and *Kcns1* were compared, in analogy to Fig. 4a. (**b**) Pearson correlation of DNAm among neighboring CpGs within three amplicons from DBA/2 mice (n = 33). (**c**) Heatmaps of epigenetic age-predictions for individual BBA-seq reads of DBA/2 mice (n = 33). Epigenetic ages were estimated based on the binary sequel of methylated and non-methylated CpGs for three amplicons (read counts were normalized by the read counts per sample and are depicted in log scale). In analogy to Fig. 4c, each read was classified to predicted ages between 0 and 200 weeks. (**d**) The mean of the single read predictions of BBA-seq data was determined for each sample and then plotted against the chronological age of the DBA/2 (n = 33) and line A (n = 15) mice in comparison with validation sets of C57BL/6 mice (n = 40). The linear coefficients of determination (R^2) of DNAm *versus* chronological age are indicated.

DNAm changes, which may be attributed to other epigenetic modifications or higher chromatin order. If age-associated DNAm was directly mediated by epigenetic writers, such as DNMTs or TETs, it might be anticipated that neighboring CpGs are rather coherently modified. The functional relevance of these age-associated DNAm changes remains unclear. Altered promoter methylation with aging was found to be generally un-related to altered gene expression, also in mice[33]. There is evidence, that the epigenetic drift by stochastic DNAm changes in promoters results in degradation of coherent transcriptional networks during aging[34]. In the future, it will be important to better understand and validate how heterogeneity in single BBA-seq read predictions reflects heterogeneity of epigenetic aging within a sample. To this end, it will be interesting to further investigate single-cell DNAm profiles, longer reads that cover multiple age-associated domains (e.g. by nanopore sequencing), or analysis of single-cell derived clones. In the future, nanopore sequencing may provide a more powerful method

for epigenetic age analysis, since it provides very long reads and does not rely on bisulfite conversion. However, so far the precision of DNAm detection in nanopore sequencing is still relatively low[35,36].

Various epigenetic clocks for mice were demonstrated to reflect aspects of biological aging, rather than only chronological aging[27]. It is still not unequivocally proven if specific epigenetic clocks capture such aspects of biological aging better, or if they may rather be influenced by the cellular composition or by direct association of DNAm at individual CpGs with specific diseases. We have recently demonstrated that inhibition of Cdc42 activity extends lifespan in C57BL/6 mice, and this is also reflected by younger age-predictions with our 3 CpG pyrosequencing signature[19]. In this study, we validated that the shorter-lived DBA/2 mice and the Line A mice have also accelerated epigenetic aging in BBA-seq data – in the conventional epigenetic predictors based on DNAm levels as well as in the single-read BBA-seq predictions for all three amplicons. Thus, our 3 CpG signature clearly captures aspects of biological aging in mice. Furthermore, the accelerated epigenetic aging in DBA/2 and Line A mice cannot be attribute to deviations at individual CpGs, but rather affects the entire age-associated region.

Taken together, we further developed and compared targeted epigenetic clocks for mice with pyrosequencing, ddPCR, or BBA-seq. All three methods provided reliable age-predictions with similar accuracy as previously described for RRBS and WGBS clocks. It is difficult to project exact costs and working time of the different methods, because this may vary significantly between local providers and available infrastructure. Furthermore, it depends on the number of samples to be processed in parallel. For orientation, an estimate for costs and time is provided in Supplementary Table S1. For DNAm levels at individual CpGs the measurements with pyrosequencing and ddPCR seemed to correlate slightly better with chronological age than BBA-seq results. On the other hand, the longer reads of BBA-seq gives better insight into neighboring CpGs and facilitates even single-read predictions that may reveal heterogeneity in epigenetic aging within a sample – depending on the availability of instruments and the experimental design all of these methods may now be considered for targeted epigenetic clocks in mice.

Methods

Mouse strains and blood collection.   Blood specimens of C57BL/6J mice of the training set (n = 24) and of the validation set 1 (n = 21), DBA/2J mice (n = 33), and Line A mice (n = 15) were obtained by submandibular bleeding (100–200 µl) of living mice or postmortem from the vena cava at the University of Ulm. One sample from the training set was excluded in the subsequent ddPCR and BBA-seq analysis due to the lack of bisulfite converted DNA. C57BL/6J samples of the validation set 2 (n = 19) were collected at the University of Groningen from the cheek. All mice were fed by ad libitum, and housed under pathogen-free conditions. All experimental protocols were approved by the Institutional Animal Care of the Ulm University as well as by Regierungspräsidium Tübingen and with the Institutional Animal Care and Use Committee of the University of Groningen (IACUC-RUG), respectively. All methods were carried out in accordance with relevant guidelines and regulations.

Genomic DNA isolation and bisulfite conversion.   Genomic DNA from 50 µl murine blood was isolated by the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA was quantified by Nanodrop 2000 Spectrophotometers (Thermo Scientific, Wilmington, USA). 200 ng of extracted genomic DNA was subsequently bisulfite-converted with the EZ DNA Methylation Kit (Zymo Research, Irvine, USA).

Pyrosequencing.   Bisulfite converted DNA was initially subjected to PCR amplification. Primers were purchased at Metabion and the sequences are provided in Supplementary Table S2, as described before[17]. 20 µl PCR products were subsequently immobilized to 5 µl Streptavidin Sepharose High Performance Bead (GE Healthcare, Piscataway, NJ, USA), and were finally annealed to 1 µl sequencing primer (5 µM) for 2 min at 80 °C. Amplicons were sequenced using PyroMark Gold Q96 Reagents (Qiagen) on PyroMark Q96 ID System (Qiagen, Hilden, Germany) and analyzed with PyroMark Q CpG software (Qiagen). The relevant sequences are depicted for the five relevant genomic regions in Supplementary Fig. S4. The 15 CpG model for pyrosequencing data, which was trained by lasso regression with the lambda parameter chosen by cross-fold validation, has been described before[17] and is provided in Supplementary Table S3.

Droplet digital PCR (ddPCR).   DNA methylation analysis by ddPCR was performed with a QX200 Droplet Digital PCR System (Bio-Rad, CA, USA). We used dual-labeled TaqMan hydrolysis probes which recognize either methylated or non-methylated target CpG site. All the primers and probes were designed by Primer3Plus software (Supplementary Table S4). Each 20 µl 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 dual probes, and 25 ng of bisulfite converted DNA. The mixture and 70 µl of droplet generation oil was then subjected into QX200 Droplet Generator (Bio-Rad). 40 µl of the generated droplets were transferred to the ddPCR 96 well plate (Bio-Rad). The plate was heat sealed with the PX1 PCR Plate Sealer (Bio-Rad) and subsequently placed in the C1000 Touch Thermal Cycler (Bio-Rad) for PCR runs as follows: 95 °C for 10 min, 40 cycles of 94 °C for 30 s and 1 min (2.5 °C/s ramp rate) at 55 °C (*Prima1, Kcns1*) or 58 °C (*Hsf4*), followed by 10 min enzyme deactivation step at 98 °C and a final hold at 4 °C. The PCR plate was 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 determined by Poisson statistics according to the fraction of positive droplets for methylated and non-methylated probes. The multivariable regression model for ddPCR is provided in Supplementary Table S5.

Barcoded bisulfite amplicon sequencing (BBA-seq). Target sequences (Supplementary Fig. S5) for *Prima1*, *Hsf4* and *Kcns1* were amplified by PyroMark PCR kit (Qiagen) using forward and reverse primers containing handle sequences for the subsequent barcoding step (Supplementary Table S6). PCR was run under the following conditions: 95 °C for 15 min; 40 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s; and final elongation 72 °C for 10 min. The three amplicons of each donor were pooled at equal concentrations under the quantification of Qubit (Invitrogen), and cleaned up with paramagnetic beads from Agencourt AMPure XP PCR Purification system (Beckman Coulter). 4 µl of pooled products were subsequently added to 21 µl PyroMark Master Mix (Qiagen) containing 10 pmol of barcoded primers (adapted from NEXTflex 16S V1-V3 Amplicon Seq Kit, Bioo Scientific, Austin, USA) for a second amplification (95 °C for 15 min; 16 cycles of 95 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s; final elongation 72 °C for 10 min). PCR products were again quantified by Qubit (Invitrogen), equimolarly pooled, and cleaned up by Select-a-Size DNA Clean & Concentrator Kit (Zymo Research, USA). 10 pM DNA library was prepared following the Denature and Dilute Libraries Guide of Illumina MiSeq System with 15% PhiX spike-in control (Illumina, CA, USA) and eventually subjected to 250 bp pair-end sequencing on a MiSeq lane (Illumina, CA, USA) using Miseq reagent V2 Nano kit (Illumina). We utilized the Bismark tool[37] to determine the DNAm levels for each CpG based on BBA-seq data. The average number of BBA-seq reads per genomic region and sample was approximately 2,000. Multivariable regression models for epigenetic age predictions were generated based on three CpGs that revealed highest correlation with chronological age per amplicon (Supplementary Table S7). Alternatively, we used a penalized regression model from the R package glmnet on the training dataset to establish a predictor with machine learning (Supplementary Table S8). The alpha parameter of glmnet was set to 1 (lasso regression) and the lambda parameter was chosen by cross-fold validation of the training dataset (tenfold cross validation).

Epigenetic age predictions for individual BBA-seq reads. As previously described, we developed an algorithm to estimate epigenetic age based on the binary sequel of methylated and non-methylated CpGs within individual reads of BBA-seq data[21]. In brief, according to the age-associated correlations at individual CpG of the BBA-seq training set, each DNAm pattern with binary sequel of methylation and unmethylation was assigned to their most representative corresponding age (0 to 200 weeks). For each donor, we calculated the mean of strand-specific age-predictions weighted by read counts as final epigenetic age predictions. Further details on the rational and derivation of the mathematical model are provided in our previous work[21].

Statistics and reproducibility. The methylation level were calculated by PyroMark Q CpG software (Qiagen) for pyrosequencing data or by Bismark tool[37] for BBA-seq data. The percentage methylation of each ddPCR reaction was determined by Poisson statistics according to the fraction of positive droplets for methylated and non-methylated probes on QuantaSoft 1.7.4 software (Bio-Rad). Machine learning age prediction models were carried out by the R package glmnet on the training dataset.

Data availability
The BBA-seq data that support the findings of the present study are accessible at NCBI NIH Gene Expression Omnibus (GEO) under the accession number GSE156193.

References

1. Dor, Y. & Cedar, H. Principles of DNA methylation and their implications for biology and medicine. *The Lancet* **392**, 777–786 (2018).
2. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
3. Koch, C. M. & Wagner, W. Epigenetic-aging-signature to determine age in different tissues. *Aging (Albany NY)* **3**, 1018–1027 (2011).
4. Bocklandt, S. *et al.* Epigenetic predictor of age. *PLoS ONE* **6**, e14821 (2011).
5. Marioni, R. E. *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* **16**, 25 (2015).
6. Belsky, D. W. *et al.* Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *Elife* **9**, e54870 (2020).
7. Horvath, S. *et al.* Obesity accelerates epigenetic aging of human liver. *Proc. Natl. Acad. Sci.* **111**, 15538–15543 (2014).
8. Horvath, S. *et al.* Accelerated epigenetic aging in Down syndrome. *Aging Cell* **14**, 491–495 (2015).
9. Maierhofer, A. *et al.* Accelerated epigenetic aging in Werner syndrome. *Aging* **9**, 1143 (2017).
10. Gross, A. M. *et al.* Methylome-wide analysis of chronic HIV infection reveals five-year increase in biological age and epigenetic targeting of HLA. *Mol. Cell* **62**, 157–168 (2016).
11. Fahy, G. M. *et al.* Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell* **18**, e13028 (2019).
12. Wagner, W. Epigenetic aging clocks in mice and men. *Genome Biol.* **18**, 107 (2017).
13. Wang, T. *et al.* Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol.* **18**, 57 (2017).
14. Petkovich, D. A. *et al.* Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell Metab.* **25**, 954–960 (2017).
15. Stubbs, T. M. *et al.* Multi-tissue DNA methylation age predictor in mouse. *Genome Biol.* **18**, 68 (2017).
16. Maegawa, S. *et al.* Caloric restriction delays age-related methylation drift. *Nat. Commun.* **8**, 539 (2017).
17. Han, Y. *et al.* Epigenetic age-predictor for mice based on three CpG sites. *eLife* **7**, e37462 (2018).
18. Brown, A. *et al.* The lifespan quantitative trait locus gene Securin controls hematopoietic progenitor cell function. *Haematologica* **105**, 317–324 (2020).
19. Florian, M. C. *et al.* Inhibition of Cdc42 activity extends lifespan and decreases circulating inflammatory cytokines in aged female C57BL/6 mice. *Aging Cell* **19**, e13208 (2020).

20. Zemmour, H. *et al.* Non-invasive detection of human cardiomyocyte death using methylation patterns of circulating DNA. *Nat. Commun.* **9**, 1443 (2018).
21. Han, Y. *et al.* New targeted approaches for epigenetic age predictions. *BMC Biol.* **18**, 1–15 (2020).
22. Franzen, J. *et al.* Senescence-associated DNA methylation is stochastically acquired in subpopulations of mesenchymal stem cells. *Aging Cell* **16**, 183–191 (2017).
23. Bell, C. G. *et al.* DNA methylation aging clocks: challenges and recommendations. *Genome Biol.* **20**, 249 (2019).
24. Field, A. E. *et al.* DNA methylation clocks in aging: categories, causes, and consequences. *Mol. Cell* **71**, 882–895 (2018).
25. Blueprint-consortium. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat. Biotechnol.* **34**, 726–737 (2016).
26. Schmidt, M., Maié, T., Dahl, E., Costa, I. G. & Wagner, W. Deconvolution of cellular subsets in human tissue based on targeted DNA methylation analysis at individual CpG sites. *BMC Biol.* **18**, 178 (2020).
27. Meer, M. V., Podolskiy, D. I., Tyshkovskiy, A. & Gladyshev, V. N. A whole lifespan mouse multi-tissue DNA methylation clock. *Elife* **7**, e40675 (2018).
28. Warnecke, P. M. *et al.* Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res.* **25**, 4422–4426 (1997).
29. Yu, M., Heinzerling, T. J. & Grady, W. M. DNA methylation analysis using droplet digital PCR. *Methods Mol. Biol.* **1768**, 363–383 (2018).
30. Hindson, C. M. *et al.* Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods* **10**, 1003–1005 (2013).
31. Shi, L. *et al.* DNA methylation markers in combination with skeletal and dental ages to improve age estimation in children. *Forensic Sci. Int.* **33**, 1–9 (2018).
32. Naue, J. *et al.* Proof of concept study of age-dependent DNA methylation markers across different tissues by massive parallel sequencing. *Forensic Sci. Int.* **36**, 152–159 (2018).
33. Hadad, N., Masser, D. R., Blanco-Berdugo, L., Stanford, D. R. & Freeman, W. M. Early-life DNA methylation profiles are indicative of age-related transcriptome changes. *Epigenet. Chromatin* **12**, 58 (2019).
34. Hernando-Herraez, I. *et al.* Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat. Commun.* **10**, 1–11 (2019).
35. Rand, A. C. *et al.* Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Methods* **14**, 411–413 (2017).
36. Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017).
37. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

Author contributions
Y.H.: performed pyrosequencing and BBA-seq, formal analysis, writing of original draft; M.N.: performed BBA-seq data analysis; M.G.: performed ddPCR assays; J.F.: calculated alternative aging models; G.H.: resources and project administration; H.G.: resources, funding acquisition; W.W.: conceptualization, supervision, funding acquisition, writing of original draft, project administration.

Competing interests
W.W. is cofounder of Cygenia GmbH that can provide service for Epigenetic-Aging-Signatures (www.cygenia.com), but the methods are fully described in this manuscript. Y.H. and J.F. also contribute to this company. The other authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-79509-2.

Correspondence and requests for materials should be addressed to W.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.