Genome Biology

RESEARCH   **Open Access**

CrossMark

Multi-tissue DNA methylation age predictor in mouse

Thomas M. Stubbs[1], Marc Jan Bonder[2], Anne-Katrien Stark[3], Felix Krueger[4], BI Ageing Clock Team, Ferdinand von Meyenn[1*], Oliver Stegle[2*] and Wolf Reik[1,5,6*]

Abstract

Background: DNA methylation changes at a discrete set of sites in the human genome are predictive of chronological and biological age. However, it is not known whether these changes are causative or a consequence of an underlying ageing process. It has also not been shown whether this epigenetic clock is unique to humans or conserved in the more experimentally tractable mouse.

Results: We have generated a comprehensive set of genome-scale base-resolution methylation maps from multiple mouse tissues spanning a wide range of ages. Many CpG sites show significant tissue-independent correlations with age which allowed us to develop a multi-tissue predictor of age in the mouse. Our model, which estimates age based on DNA methylation at 329 unique CpG sites, has a median absolute error of 3.33 weeks and has similar properties to the recently described human epigenetic clock. Using publicly available datasets, we find that the mouse clock is accurate enough to measure effects on biological age, including in the context of interventions. While females and males show no significant differences in predicted DNA methylation age, ovariectomy results in significant age acceleration in females. Furthermore, we identify significant differences in age-acceleration dependent on the lipid content of the diet.

Conclusions: Here we identify and characterise an epigenetic predictor of age in mice, the mouse epigenetic clock. This clock will be instrumental for understanding the biology of ageing and will allow modulation of its ticking rate and resetting the clock in vivo to study the impact on biological age.

Keywords: DNA methylation, Epigenetics, Ageing/aging, Ovariectomy, Epigenetic clock, High fat diet, Chronological age, Biological age, Prediction, Model

Background

Ageing describes the progressive decline in cellular, tissue and organismal function during life, which ultimately drives age-related diseases and limits lifespan [1]. From a biological perspective, ageing is associated with numerous changes at the cellular and molecular level [2], including epigenetic changes, that is modifications of DNA or chromatin that do not change the primary nucleotide sequence. At present, it is not clear which epigenetic changes are causative or correlative, but these mechanisms are of particular interest due to their reversibility, suggesting that rejuvenation might be possible, at least in principle [3, 4].

Recently, age-correlated DNA methylation changes at discrete sets of CpGs in the human genome have been identified and used to predict age [5–7]. These 'epigenetic clocks' can estimate the DNA methylation age in specific tissues [5] or tissue-independently [6] and can predict mortality [8] and time to death [9]. These findings have sparked intense interest regarding the role of DNA methylation in the ageing process and also opened up a number of key questions. Interestingly, while initially designed to predict chronological age, there is evidence that the epigenetic age also reflects biological age and is predictive of functional decline [10–22]. This suggests that the observed methylation signatures might be caused by an intrinsic biological ageing process. One suggestion has been that the methylation clock 'measures the cumulative

* Correspondence: vonmeyenn@babraham.ac.uk; oliver.stegle@ebi.ac.uk; wolf.reik@babraham.ac.uk
[1]Epigenetics Programme, The Babraham Institute, Cambridge CB22 3AT, UK
[2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton CB10 1SD, UK
Full list of author information is available at the end of the article

Stubbs *et al. Genome Biology* (2017) 18:68

Page 2 of 14

effect of an epigenetic maintenance system' [6], a system that is of critical importance and regulated at multiple levels [23, 24]. As such, further insights into the mechanistic properties of this underlying process are of key relevance to understand ageing in more detail and will also be instrumental for the design of future interventions. Consequently, a methylation clock that is applicable to animal models more amenable to experimental interventions would be of considerable importance.

Importantly, while a very small number of age-correlated methylation changes at selected sites in the mouse genome have been reported [25], it is not known whether such an epigenetic ageing clock is conserved between species or a unique property of humans and some closely related primates [6]. Given the general occurrence of the ageing process across the animal kingdom [1], differences in the mechanistic properties of such a clock could explain differences in median lifespan between closely related species. Here we have generated high-resolution methylomes from the experimentally tractable mouse across a wide range of tissues and ages. We find that discrete DNA methylation changes correlate with chronological age and are associated with biological functions. Based on these findings, we generated a multi-tissue age predictor for the mouse, characterised its properties, and demonstrated that it can be applied to inform other studies by applying it to publicly available datasets, including key biological interventions.

Results and discussion
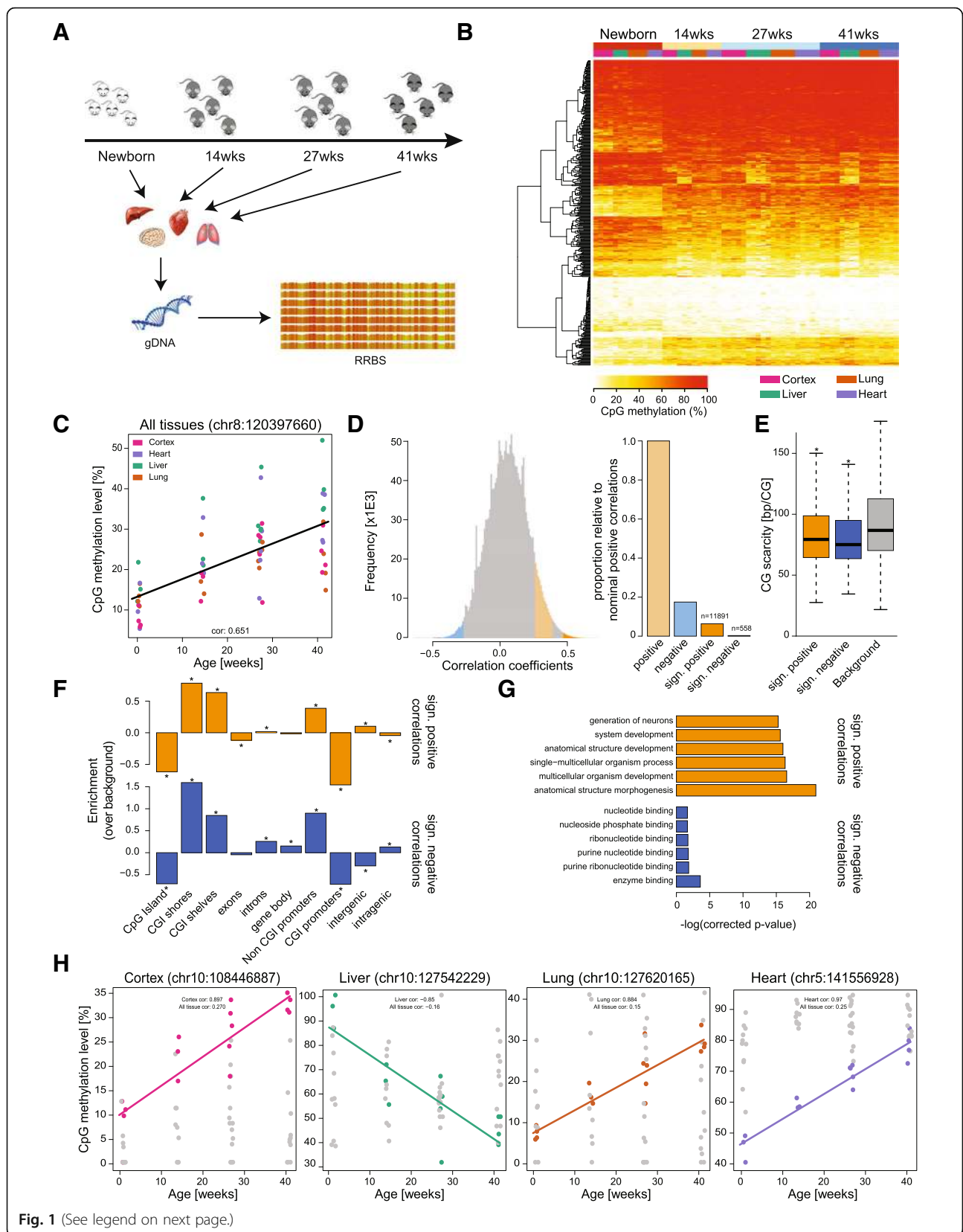### DNA methylation changes in mice correlating with age
In order to study age-associated DNA-methylation changes in the mouse over a wide range of ages and tissues, we collected liver, lung, heart and brain (cortex) samples from newborn to 41-week-old mice (Fig. 1a). To reduce genetic variability and hormonal variations, we restricted our cohort to male C57BL/6-BABR mice and sampled three to five animals per time point. In total, we collected 62 samples (Additional file 1) and extracted genomic DNA for methylation analysis from them.

We generated reduced-representation bisulfite sequencing (RRBS) libraries of all samples to be able to assess DNA methylation changes at a wide range of CG sites and sequenced these to 15× genomic coverage on average. RRBS represents a good compromise between sequencing costs, CG sites measured and fold genomic coverage obtained. To improve the quantification results, we optimised the standard RRBS library preparation protocol [26] and were able to achieve very low duplication rates and high genomic coverage. On average, more than 1.23 million CG sites in each sample were covered at least five-fold and of these 0.73 million CG sites had more than five-fold coverage in all samples analysed. Global CpG methylation levels in newborns

were around 43% and did not differ significantly between tissues. In the older samples, average methylation levels were slightly higher (~45%) but did not show major differences between ages or tissues (Additional file 2A). Global methylation levels measured by RRBS are generally lower than whole-genome bisulfite sequencing (WGBS) estimates, as the method enriches for hypomethylated CpG islands (CGIs) [26]. We also observed low levels of non-CpG methylation in all non-brain tissues (Additional file 2B). In agreement with the notion that de novo methylation activity in non-dividing cells results in accumulation of CHH methylation, we found that adult cortex samples had higher CHH methylation levels than newborn cortex samples [27]. Together our samples represent the most comprehensive dataset thus far of matched single base resolution methylomes in mice across multiple tissues and ages. Importantly, a hierarchical clustering analysis using Manhattan distances (Additional file 2C) clearly separated the samples by tissue (with the exception of newborn lung samples which clustered together with adult heart samples), highlighting key tissue-specific methylation signatures [28].

A correlation analysis showed that DNA methylation at a substantial number of CG sites across all tissues correlated with age (Spearman's correlation, with a multiple testing corrected p value < 0.05) (Fig. 1b and c). As expected, the majority of these sites showed age-correlated methylation changes in all or at least three tissues (Additional file 2D), suggesting that age-dependent DNA methylation changes at specific sites occur in a coordinated manner across tissues. The correlation values between age and DNA methylation at discrete CG sites were both positive and negative and normally distributed (Fig. 1d). Overall, we identified more positive correlations, but this skewing is likely to be the result of the slight global hypomethylation in the newborn samples and represents the general tendency for gain of DNA methylation during development.

To understand whether the underlying sequence composition or genomic context was relevant to the changes observed, we analysed CpG density (Bonferroni corrected two-tailed t-test, p value < 0.05) and genomic context (Binomial test, p value < 0.05) at the significantly correlating sites. While CpGs changing DNA methylation with age were on average in regions with higher CpG density (lower CpG scarcity), the CpG density was not predictive of the methylation changes (AUC = 0.58 or 0.61; see 'Methods' for details) (Fig. 1e). We also found a strong depletion of sites with significant correlations with age at CGIs and CGI rich promoters and conversely a strong increase at CGI-shores and CGI-shelves (4 kb around CGIs) (Fig. 1f). This indicates that tightly controlled regulatory regions, such as CGIs, are relatively protected from age-associated DNA methylation

Stubbs *et al. Genome Biology* (2017) 18:68

Page 3 of 14

Fig. 1 (See legend on next page.)

Stubbs *et al. Genome Biology* (2017) 18:68

Page 4 of 14

(See figure on previous page.)

Fig. 1 DNA methylation changes correlate with age. **a** Overview of the Babraham dataset. Ages of the mice are emphasised by colour. Tissues (liver, lung, heart and cortex) were isolated from mice at four distinct time points (newborn, 14 weeks, 27 weeks and 41 weeks). DNA was isolated from these tissues and reduced-representation bisulfite (RRBS) libraries were made. **b** *Heatmap* of the top 500 tissue independent age-associated corre-lations. Highlighted are ages and tissues, CG sites were clustered by Euclidean distance. **c** Single CG sites within the genome are correlated with age. Shown is an example site (chr8:120397660), with a Spearman correlation with age of 0.65. Tissues are highlighted by *colour*. Jitter is for aesthetic pur-poses only. **d** Overview of Spearman correlations calculated over all tissues. The distributions of correlation estimates are shown in the histogram and proportionate numbers of correlations are shown in the *bar plot*. Nominal correlations are highlighted (*p* value < 0.05) in *light orange* (positive) and *light blue* (negative). Significant correlations are highlighted (q-value < 0.05) in *orange* (positive) and *blue* (negative). Numbers provided above the bar plot represent the number of single CG sites within the given category. **e** Surrounding CpG content (± 500 bp) of the significantly correlated sites. Back-ground was calculated from all CG sites with five-fold coverage in 90% of samples; CpG scarcity is defined as the average CpG distance within this 1 kb region. *p Bonferroni corrected *p* value < 0.05. **f** Enrichment of a correlated CG site falling within a given genomic element. Background was all CG sites with five-fold coverage in 90% of samples. Tested using binomial test; *p value < 0.05. **g** Gene Ontology (GO) analysis of the significantly correlated CG sites. GO terms are plotted against −log(corrected *p* value). Positive correlations are shown in *orange* and negative in *blue*. The six most significant GO terms are shown. **h** Tissue-specific Spearman correlations with age. An example is provided of a tissue-specific correlation with age in cortex, liver, lung and heart. Correlations for these CG sites are provided for all tissues combined and for the tissue in question. Jitter is for aesthetic purposes only

changes, while regions with intermediate CpG density are more prone to changes. A Gene Ontology (GO) ana-lysis of the genes closest (max. 4 kb distance) to CpGs with highly significant positive age-dependent changes in DNA methylation revealed a significant enrichment of genes associated with the terms 'anatomical structure morphogenesis', 'anatomical structure development' and 'developmental process' (Fig. 1g). Genes close to sites with significant negative age-related correlations were enriched in terms containing nucleotide and enzyme binding. These results suggest that the age-related changes in DNA methylation could alter various import-ant biological processes and future experiments may re-veal the regulatory relevance and association with gene expression. In particular, increased DNA methylation at developmentally relevant genes suggests that the ageing process may restrict expression of developmental genes.

We next analysed age correlations in each tissue inde-pendently and identified a large number of sites, which showed exclusively tissue specific changes in DNA methylation with age (Fig. 1h and Additional file 2E). Interestingly, only a small fraction of these were shared between three or more tissues, indicating that additional age-related DNA methylation changes are characteristic for each tissue and might relate to its intrinsic biological function (Additional file 2F). In particular, GO analysis of genes close to the sites with tissue-specific changes revealed highly different GO term enrichments, indicat-ing their unique tissue-specific regulation.
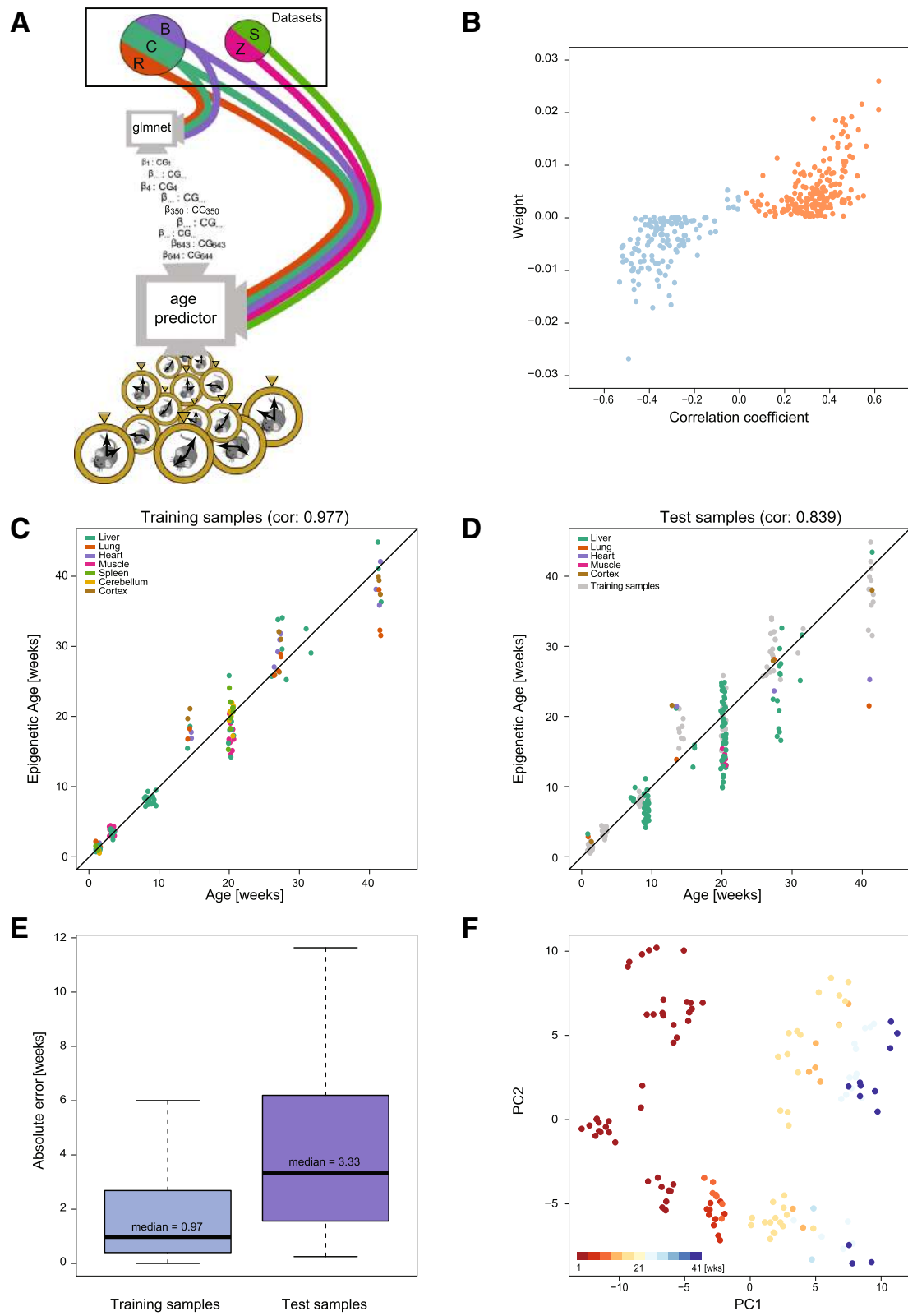
The marked differences between the global methylation levels of newborn and adult samples (Additional file 2A) prompted us to also analyse the datasets excluding new-born samples (Additional file 3). A number of sites showed a reversed directionality of age-dependent methylation changes if we excluded the newborn samples from the ana-lysis (highlighted in pink and green; Additional file 3A and B); however, the majority of sites (highlighted in blue and orange/red) showed the same correlation characteristics independent of the inclusion or exclusion of newborn

samples (Additional file 3C). The correlations between age and DNA methylation in the adult datasets were both posi-tive and negative and the correlation values were normally distributed (Additional file 3D), but none of the correla-tions were significant (Additional file 3E; Spearman's correlation, with a multiple testing corrected *p* value cutoff of < 0.05). Although the analysis was based on all tissues, only a small number of the age-correlated methylation changes in the adult samples were common to all tissues (*p* < 0.005; Additional file 3F). A subsequent analysis of age correlations in each tissue independently identified a large number of tissue specific methylation changes (*p* < 0.005; Additional file 3G and H), suggesting that changes in DNA methylation in adults are primarily driven by tissue-specific processes, while tissue-independent methylation changes are associated with global developmental pro-cesses (Fig. 1g).

DNA methylation levels at a discrete set of CpGs are predictive of age

Having found that methylation at many individual CG sites did change in an age-dependent manner, we decided to generate an epigenetic age predictor in mice (Fig. 2a). In addition to our own datasets, we also included previ-ously published datasets comprising RRBS libraries from liver, lung, muscle, spleen and cerebellum samples from male and female C57BL/6 mice aged newborn to 31 weeks [29–32]. An overview of these datasets can be found in Additional file 4. In short, 129 healthy samples were used to define the training set, with the remaining 189 making up an independent test set, including two datasets that were generated by different labs and hence experimentally independent from the datasets used to train the model. All samples were processed as described in the 'Methods'. We excluded CpGs located on either of the sex chromo-somes or in the mitochondrial genome prior to further processing and analysis.

We used an elastic-net regression model to predict log-transformed chronological age, measured in days.

Stubbs *et al. Genome Biology* (2017) 18:68

Page 5 of 14

Fig. 2 (See legend on next page.)

Stubbs *et al. Genome Biology* (2017) 18:68

Page 6 of 14

(See figure on previous page.)

Fig. 2 Prediction of chronological age from a mouse epigenetic clock. **a** *Flow chart* to illustrate the steps taken in defining the model and testing it. Datasets are displayed as segments of a *circle*. They are coloured to correspond with later figures, namely: Reizel in *brown* (R), Cannon in *green* (C), Babraham in *purple* (B), Zhang in *pink* (Z) and Schillebeeckx in *light green* (S). The two independent datasets are displayed as segments in a separate circle to those datasets utilised for the training phase. The flow of methylation data is shown as colour-coded *lines*. Training occurs at the node (screen) with the caption: 'glmnet'. The chosen CG sites and their corresponding weighting are passed on to the prediction tool itself (node with the caption: 'epigenetic age predictor'. Test data enter this prediction tool and age predictions are outputted, as displayed by the pocket watches exiting this node. **b** *Scatter plot* depicting weight of chosen clock sites against their age-associated correlation, *blue* are negatively correlated sites and *orange* are positively correlated sites. **c** Training set ages as predicted by the model, *x-axis* shows the actual age and *y-axis* shows the predicted age, coloured by tissue. Jitter is used to represent the experimental error in age estimates. **d** Test set ages as predicted by the model, *x-axis* shows the actual age and *y-axis* shows the predicted age, coloured by tissue. **e** *Box plot* of the absolute error in the training (*left*) and test (*right*) samples. The median absolute error is indicated. **f** Principal component representation of the sites used in the clock, coloured by age in weeks

Only CG sites covered by at least 5 reads in all samples were used (~18 k sites; see 'Methods' for details). Following cross-validation to optimise the model parameters, the final predictor was based on 329 CG sites (Additional files 5A and 6); the sites in this predictor will be referred to as mouse (multi-tissue) clock sites. The model selects the most informative sites but allows for some redundancy to increase robustness and it infers weights for each individual site [33]. The model weights across sites are depicted in Additional file 5B. One implication of this approach is that the clock sites do not necessarily represent the strongest age-correlating sites characterised above (Fig. 2b). Similar to the human age predictor described by Hannum et al. [5], the initial starting methylation levels of the mouse clock sites are somewhat predictive of the directionality of their methylation changes with age (Additional file 5C).

As expected, the exponent of the weighted average of the DNA methylation levels of the 329 selected CG sites was highly correlated with (chronological) age of the individual samples within the training dataset (Fig. 2c). Notably, using unobserved test samples, our mouse epigenetic clock was able to accurately predict chronological age in various tissues (Fig. 2d) and across multiple independent datasets (Additional file 7A). The accuracy of the model predictions was also independent of sequencing depth of the test samples (Additional file 7B), provided there was a mean coverage per CG site of 5 reads or more. This indicates that minor technical variations and coverage differences are well tolerated and consequently our mouse epigenetic clock model can be applied to a wide range of different settings. This was evident by the fact that the predictor was able to accurately estimate age in two completely independent test datasets [31, 32] (Additional file 7C).

The mouse clock performed well across all tissues and ages tested, with an age correlation of 0.839 and median absolute error (MAE) of 3.33 weeks in the test data (Fig. 2d and e), corresponding to less than 8.5% error relative to the oldest ages (41 weeks) analysed. In order to compare the accuracy with the human epigenetic

clock [6], we calculated the MAE as a proportion of the expected lifespan of a mouse (>100 weeks) and found it to be similar to that reported for the human clock (assuming an average human lifespan of 85 years). It is worthy of note that, similar to the human clock, the performance of our mouse age-predictor varied between young and old mice. In young animals (< 20 weeks), the model predictions were much more accurate, with a MAE of 2.14 weeks in the test samples. In mice aged 20 weeks or older, the MAE was 4.66 weeks (Additional file 7D). We also attempted to include publicly available WGBS datasets, including samples from 24-month-old animals [34], to test our age predictor at older ages. Unfortunately, the sequencing depth in most WGBS datasets is significantly lower than in RRBS datasets (mean coverage < 5-fold), thus precluding their analysis. It is expected that future high coverage datasets will help to further improve the accuracy of the mouse clock and test its performance at older ages.

To get further insights into the architecture of our multi-tissue age predictor, we performed a principal component analysis (PCA) of the variation within the 329 selected sites in the training datasets. Ninety percent of the observed variability was explained by 69 principal components (PCs) (Additional file 7E), of which two PCs (PC1 and PC13) displayed a clear age relation ($p < 0.05$). PC1 captured age-dependent changes and showed a good separation of samples by age; PC2 separated liver samples from the other tissue samples (Fig. 2f and Additional file 7F). This analysis highlights that the major variation within the selected CG sites in the training set is governed by numerous factors, including tissue type and age. However, dataset effects (i.e. technical variations) are not among the major drivers of variation.

Next, we characterised the clock sites in more detail. The CG sites were distributed across all autosomes with no specific enrichment in any chromosome (Additional file 8A). Similar to the age-correlated CG sites, we found a strong depletion over CGIs and CGI promoters but also over non-CGI promoters (Additional file 8B), suggesting that the clock sites were specifically depleted in

Stubbs *et al. Genome Biology* (2017) 18:68

Page 7 of 14

regulatory regions. CGI shores and intergenic regions showed increased enrichment in clock sites, whereas the CpG density around the clock sites did not show any differences compared to other random CG sites (Additional file 8C). The 329 clock sites did also not show any specific GO enrichment (not shown), suggesting that the sites selected by the model might not represent a unique biological function and are instead associated with various biological functions.

Age predictions using the human clock sites in mouse

Given the similarities between our mouse epigenetic age predictor and the previously described human epigenetic clock [6], we asked whether the specific genomic loci described in human could be used to predict age in mouse samples too. First, we attempted to lift over the genomic locations of the 353 human clock sites to the mouse, by defining regions ± 500 bp around the sites. We were able to lift over 328 regions, of which 175 regions (in the following referred to as 'Horvath clock regions in mouse') were covered by at least one CG site in our dataset. Methylation levels at these 175 Horvath clock regions in mouse were only weakly correlated with age (Additional file 9A), which is also true for the Horvath clock sites in human [6].

Next, we assessed whether the 175 Horvath clock regions in mouse could be used for age prediction. Using a ridge model, we were able to generate an age prediction model with a MAE of 11.2 weeks (Fig. 3), indicating that the methylation levels at the 175 Horvath clock regions in mouse contain age-related information and are predictive of age in the mouse. The directionality of the mouse-specific and human-specific weighting for any individual CG site within the Horvath clock regions in mouse were completely unrelated (Additional file 9B), suggesting that the human clock [6] is not fully conserved in mouse. Remarkably, we found that we could also generate age prediction models from matched random regions (see 'Methods' for details) with an average MAE of 10.6 weeks (Fig. 3), highlighting the fact that methylation changes in many genomic regions are age-related and potentially predictive.

In summary, the predictive accuracy of the sites corresponding to the human clock in mouse samples was significantly lower than our mouse epigenetic clock. These differences are in part due to technical differences in defining the clock sites, but do also highlight species-specific differences, which result in different age-dependent methylation changes at discrete loci and might point to different 'ticking' rates of human and mouse ageing.

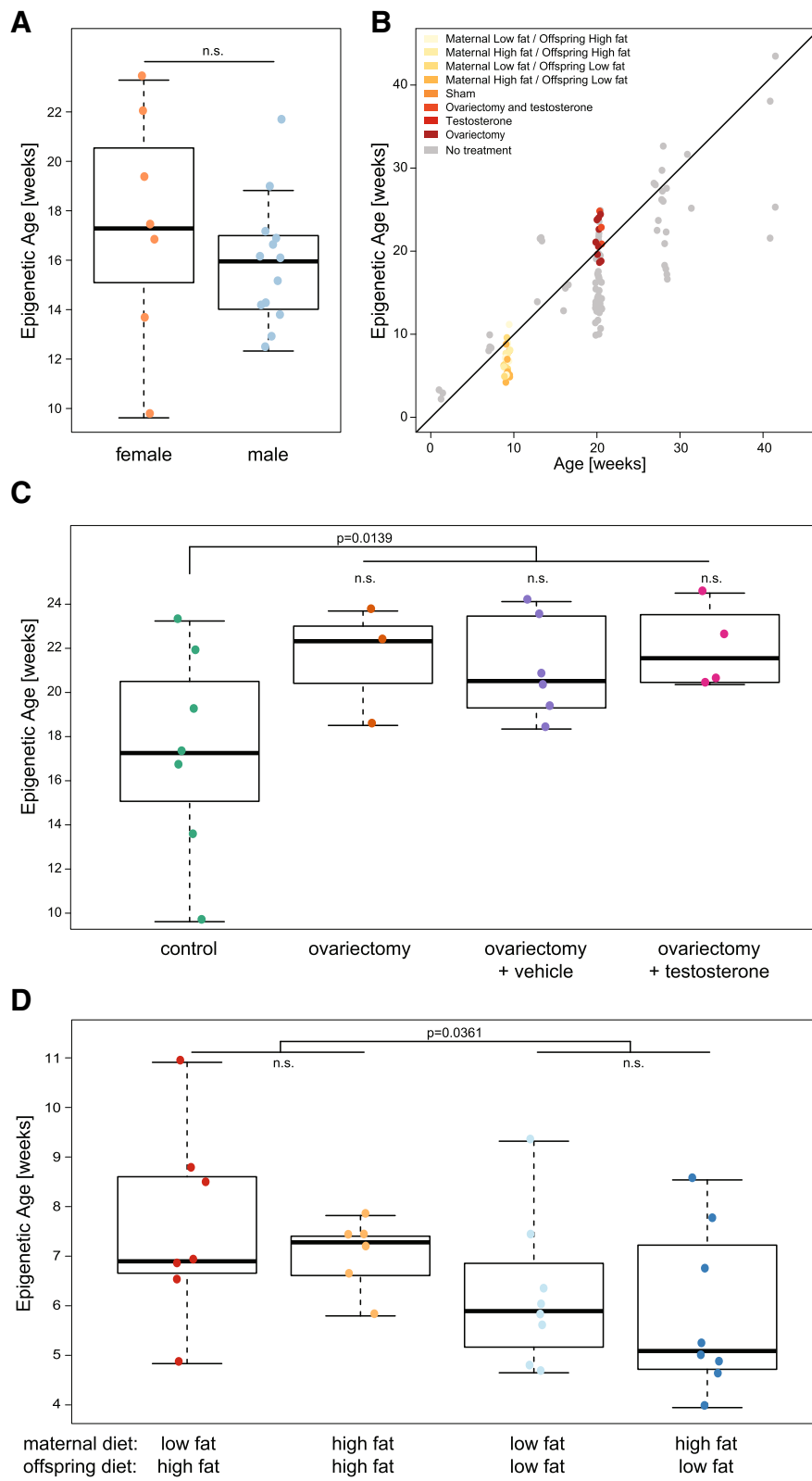### DNA methylation age is altered in ovariectomised females and by diet

Given the accuracy of our mouse epigenetic clock to predict chronological age in healthy individuals, we

Fig. 3 Prediction evaluation of the human Horvath clock sites in mouse. Shown is the MAE of the age prediction model generated using the lifted-over human clock sites [6] (*red line*). The distribution (*blue*) shows the MAEs of 1000 age-prediction models generated using random selections of 329 regions (see 'Methods' for further details)
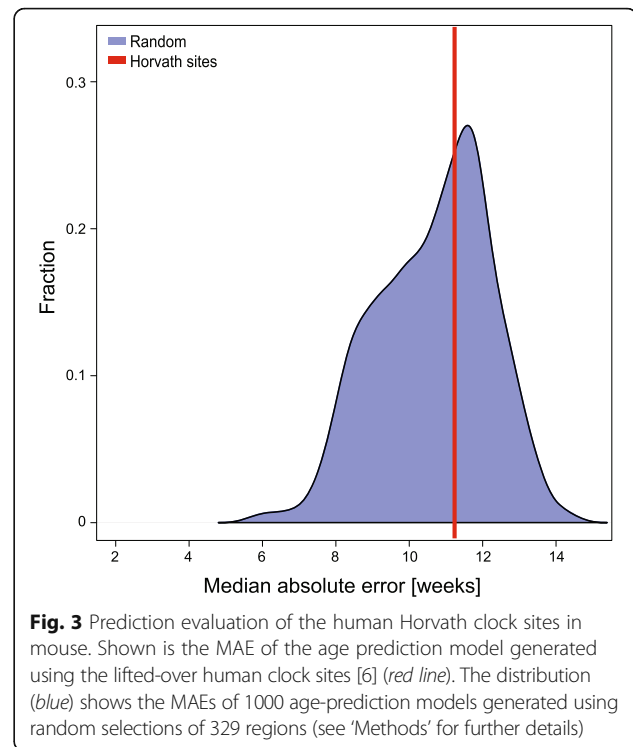
asked whether epigenetic age in the mouse would be affected by gender, diet or other biological interventions. Since it has been reported that DNA methylation shows gender-specific patterns [30], we compared the predicted DNA methylation age of female and male mice. Although the training data were predominantly composed of male samples (> 72%), we did not find significant differences in the estimated ages between sexes (Fig. 4a and Additional file 10), highlighting the robustness of the model and showing that gender-specific differences are not skewing the predictions of the mouse epigenetic clock.

We next analysed publicly available RRBS datasets of samples with various biological interventions [29, 30]. Overall the epigenetic age predictions were well related to the true chronological age of the individual mice (Fig. 4b), but at closer inspection we identified important differences, which depended on treatment.

Reizel et al. performed ovariectomy in females, treated the mice with testosterone (5 mg/mL) or vehicle [30] and analysed their liver DNA methylome by RRBS. Ovariectomy decreases average lifespan in female rats [35] and indeed we found that ovariectomy increased epigenetic age significantly in mice (Fig. 4c). Additional testosterone or vehicle treatment had no significant effect on the epigenetic age and was also shown to not alter lifespan in rats [35]. This suggests that hormonal differences in mice affect 'biological age', as measured by

Stubbs *et al. Genome Biology* (2017) 18:68

Page 8 of 14

Fig. 4 (See legend on next page.)

Stubbs *et al. Genome Biology* (2017) 18:68

Page 9 of 14

(See figure on previous page.)
Fig. 4 Methylation age is affected by biological interventions. **a** Predicted age of 20-week-old liver samples [30], shown separately for males and females. Statistical test performed: t-test, *p* value of 0.58. **b** Age prediction in test samples from various biological intervention studies. *X-axis* shows the actual age and *y-axis* shows the predicted age, coloured by study. **c** Age prediction in normal female liver samples and samples which underwent an ovariectomy and were administered with either vehicle alone or testosterone [30]. Statistical test performed: Unpaired two-tailed t-test performed to assess the impact of ovariectomy, *p* value of 0.014. **d** Age prediction in diet perturbation study [29]. Liver samples from animals with following treatments were analysed: maternal high fat diet followed by adult high fat diet, maternal high fat diet followed by adult low fat diet, maternal low fat diet followed by adult high fat diet, and maternal low fat diet followed by adult low fat diet. Statistical test performed: two-way ANOVA performed, *p* values displayed where significant

our mouse epigenetic clock; similarly in humans, breast tissue is found to have accelerated epigenetic ageing [6].

In a second study, Cannon et al. [29] characterised the effects of lipid content in maternal and offspring diet on body weight, physiology and DNA methylation status in the liver and reported a significant effect of maternal diet on the susceptibility of the offspring to become obese and develop signs of metabolic disease. The greatest adverse effects were observed when the mothers were fed a low-fat diet and the offspring a high-fat diet. By applying our mouse epigenetic clock to these liver samples, we identified a strong dependency of the predicted epigenetic age on diet (Fig. 4d). Offspring on a high-fat diet showed accelerated epigenetic ageing, which had a tendency to be further exacerbated if the mothers were fed a low-fat diet. This result suggests that biological ageing is modulated by diet and possibly intergenerational or transgenerational effects and that epigenetic age is a potential powerful measure of biological function.

Conclusions

In this study, we have generated the thus far most comprehensive set of matched single base resolution methylomes in mice across multiple tissues and ages. This resource allowed us to study the correlation between DNA methylation changes and ageing in detail. Importantly, we were able to establish, for the first time, a mouse epigenetic clock, which estimates age based on the methylation state at 329 discrete CG sites throughout the mouse genome. Our novel epigenetic clock is performing similar to the human epigenetic clock and allowed us to assess (epigenetic) age in unrelated methylation datasets. Worthy of note, the 329 sites of our mouse epigenetic clock perform significantly better in predicting age in mouse samples than the sites in the mouse genome corresponding to the human Horvath clock sites. So far, our epigenetic clock has been developed and tested from tissue samples up to 41 weeks of age and future experiments and datasets will be required to assess its accuracy in older mice. The epigenetic clock and the comprehensive set of methylomes are available to the ageing research community and will enable mechanistic and intervention studies in the experimentally tractable mouse model system.

Importantly, we found that the mouse methylation clock is affected by biological interventions and as such we suggest that the prediction of the clock reflects not only chronological age but also biological age. It will be exciting to test the consequences of manipulations of the ticking rate of the epigenetic clock on biological function, in particular the possible reversibility of ageing associated functional decline. We believe that this work will help scientists gain insights into the biology of ageing and potentially help to discover novel strategies to extend lifespan and health span in the future.

Methods

Babraham Institute C57BL/6 J/Babr Ageing Mouse Colony

Mice were bred and maintained in the Babraham Institute Biological Services Unit (BI BSU) under Specific Opportunistic Pathogen Free (SOPF) conditions. C57BL/6 J mice (supplied by Charles River Laboratories) were imported into the BI BSU by embryo transfer and bred there (/Babr) as a SOPF colony in plastic film isolators. After weaning, male mice were transferred to individually-ventilated cages in groups of between two and five mice. All mice were fed CRM (P) VP diet (Special Diet Services) ad libitum and received sunflower seeds, poppy seeds or millet at cage-cleaning as part of their environmental enrichment. The health of mice was monitored closely and any mouse exhibiting clinical signs of ill-health or distress that persisted for more than three days was culled. In this way, all the mice maintained remained 'sub-threshold' with respect to the UK Home Office severity categorisation. Maintained, assessed and defined in this way, mice in this ageing colony remained healthy for a median period of approximately 99 weeks. Clinical signs necessitating culling included dermatitis (excessive scratching), fight wounds and muscle weakness, among several others. Post-mortem analysis sometimes revealed neoplasias (especially lymphomas and haemangiosarcomas) typical of this strain. Any mice exhibiting any gross pathology upon post-mortem examination was excluded from this study.

Sample collection: Babraham dataset

Cortex, heart, liver and lung samples from the Babraham Institute C57BL/6 J/Babr ageing mouse colony were collected at four different ages: newborn (<1 week), 14

Stubbs *et al. Genome Biology* (2017) 18:68

Page 10 of 14

weeks, 27 weeks and 41 weeks. All tissues were snap-frozen directly after isolation. Genomic DNA was isolated from ~10 mg frozen tissue using the DNeasy Blood & Tissue Kit (Qiagen). A total of 62 samples were collected, processed and further analysed. The resulting dataset are referred to here as Babraham dataset.

Library preparation

RRBS libraries were prepared from isolated DNA following published protocols [36]. Briefly, RRBS libraries were prepared by MspI digestion of 100–500 ng genomic DNA, followed by end-repair and T-tailing using Klenow Exo- (Fermentas). Adapter ligation was performed overnight (in-house adapters) using T4 DNA Ligase (NEB), followed by a clean-up step using AMPure XP beads (Agencourt, 0.9×). Subsequently, libraries were bisulfite-treated according to the manufacturer's instructions (Sigma Imprint Kit; 2 step protocol) and purified using an automated liquid handling robotic system (Agilent Bravo). The libraries were amplified using KAPA HiFi Uracil HotStart DNA Polymerase (KAPA Biosystems), indexing the samples with individual primers. All amplified libraries were purified (AMPure XP beads, 0.8×) and assessed for quality and quantity using High-Sensitivity DNA chips on the Agilent Bioanalyzer. High-throughput sequencing of all libraries was carried out with a 75 bp paired-end protocol on a HiSeq 2000 instrument (Illumina).

Methylation data processing

For datasets generated in this study, raw paired-end FastQ files were pre-processed to remove the first 13 bp from their 5' ends, containing unique molecular identifiers (UMI) sequence tags. Both Read 1 and Read 2 UMIs and fixed sequences were written into the read IDs. These trimmed reads were then subjected to adapter and quality trimming with Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; v0.4.2; options: –paired –three_prime_clip_R1 15 –three_prime_clip_R2 15, to also remove potential UMI and fixed tag sequences from the 3' ends). The trimmed files were then aligned to the mouse genome (GRCm38) using Bismark [37]; v0.16.3, default parameters). The mapped sequences were deduplicated by chromosomal position as well as the UMI sequences of both Read 1 and Read 2 (no mismatches tolerated) using the tool UmiBam (https://github.com/FelixKrueger/Umi-Grinder; v0.0.1; options: –bam –dual_umi). These UMI-deduplicated BAM files were then further processed with the Bismark Methylation Extractor (default parameters) to yield Bismark coverage files.

Calling of methylation at single CG sites

Mean methylation levels of each CG site covered in each sample were calculated from the Bismark coverage files. In addition, a read count was conducted for each CG site in each sample, so that filtering could be done based on this information in downstream analysis.

Statistical analysis of age association at single CG sites

For the statistical analysis presented in Fig. 1, we first filtered site that had a mean coverage of < 2 reads or > 100 reads. This was done to remove spurious reads from library preparation and potential mapping artefacts. For the remaining sites, any site that was covered with < 5 reads in a sample was replaced with NA. Before calculating age association, we further filtered for sites, such that the sites were at least covered in 90% of samples (n = 1,921,569). Ages in days were used when computing the Spearman correlation for each site using the R implementation of the Spearman correlation and multiple testing correction was performed using the Q-value package [38]. For the tissue-specific analysis, a further filtering step was conducted to ensure that there were at least four samples being considered for each correlation test. For data exploration, we used PCA analysis, on sites (n = 729,785), which were covered in all samples at 5×.

Predicting significance of age relation based on CpG density

To assess if CpG density is predictive of the methylation age relations, we tried to predict if a relation between age and methylation level would be significant based on CpG density. To do so, we used CpG density as a linear predictor using multiple thresholds to predict if a methylation age relation would be significant or not using the AUC function for the pROC library [39].

Genomic enrichment analysis of significant age-associated sites

Normalised likelihood was calculated as:

$$Normalised\,likelihood,\,at\,x = \left(\frac{s}{b} \times \frac{B}{S}\right) - 1$$

where:
s = number of significant sites at a given x
S = total number of significant sites
b = number of background sites at a given x
B = total number of background sites
CpG scarcity
CpG scarcity was calculated as:

$$CpG\,scarcity = \frac{200}{no.\,of\,CG\,sites\,witin\,a\,200bp,\,centered\,on\,a\,CpG\,of\,interest}$$

GO analysis of neighbouring genes

Neighbouring genes were defined for single CG sites that were within 4 kb of a gene. GO terms were defined

Stubbs *et al. Genome Biology* (2017) 18:68

Page 11 of 14

using the gprofiler online software [40]. For the GO enrichment analysis, a background gene list was made consisting of the neighbouring genes (max distance of 4 kb) for all sites considered in that analysis. Significant GO terms were ordered by *p* value and the top six GO terms are shown.

Human-comparative analysis

We defined 1 kb windows around the 21 k CpG sites that are interrogated in both the 27 k and 450 k MethylArray [6], to ensure that the sites could be faithfully lifted over. These sites were then lifted over from the human genome to the corresponding regions in the mouse genome (GRCm38). In particular, 91% of the sites selected by Horvath [6] for the human clock were successfully lifted, i.e. 329 of 353. To be able to contrast the human Horvath clock sites [6] to other sites in the mouse genome, we chose to use all ~21 k sites, of which we were able to lift over 19 k regions. 175 regions corresponding to the 353 clock sites and 10 k regions corresponding to all ~21 k sites were covered in the Babraham dataset. Adding additional datasets (e.g. Reizel et al. [30]) reduced the number of regions covered dramatically.

Using these sites, we first assessed the age association by comparing the correlation to age in the Horvath clock regions verses random selections of all lifted-over regions. After this we built an age prediction model based on the 175 covered regions corresponding to the Horvath clock sites. For this, we built a ridge model as implemented in glmnet, by fixing the alpha parameter to 0. The predictor reaches a MAE of 11.2 weeks. To compare this to background, we built a thousand random models, picking a random set of 329 regions, regardless of coverage, from the 19 k regions we could lift over. The average MAE was 10.65 weeks in these random models.

Predicting age in mice

Dataset overview (see Additional file 1)

For defining a more generalizable age predictor, we included four additional external RRBS datasets, which were downloaded from the GeneExpressionOmnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/): Reizel et al. [30] (GSE60012; n = 173); Cannon et al. [29] (GSE52266; n = 40); Zhang et al. [31] (GSE80761; n = 4); and Schillebeeckx et al. [32] (GSE45361; n = 23). The datasets were processed in the following manner: Raw FastQ files were trimmed with Trim Galore (v0.4.2; parameters: –rrbs) and then aligned to the mouse genome (GRCm38) with Bismark (v0.16.3; default parameters). The aligned BAM files did not undergo deduplication but were processed directly with the Bismark Methylation Extractor (default parameters) to yield Bismark coverage files.

In the following sections a short description of the dataset is given.

Reizel et al. [30]

The Reizel study consists of 173 samples originating from four different tissues: liver, muscle, cerebellum and spleen; data were collected at six time points ranging from 1 to 31 weeks. In the original study, gender and tissue specificity of demethylation during ageing has been studied. In the study, a perturbation based on castration and restoring testosterone levels after castration has been performed. For the development of the methylation clock, the perturbations were not taken into the training; these were left for the test-set. Further information can be found in Additional file 1 and the original publication. After QC, there were 143 samples left.

Cannon et al. [29]

The Cannon study consists of 40 samples all originating from liver at the age of nine weeks. In the original study, the effect of maternal diet on the metabolism of adult offspring was studied. In our study, we selected part of the data to be in our training set to reflect the nine-week time point (n = 5). The other part of the data is used to assess the effect of diet on ageing. Further information can be found in Additional file 1 and the original publication. After QC, there were 36 samples left.

Zhang et al. [31]

The Zhang study consists of four samples all originating from liver at the age between six and eight weeks. In the original study, methylation differences between different strains of mice, and the difference between mouse and zebrafish DNA-methylation levels are assessed. In our study, these samples were used as a validation to see how the predictor works for an unobserved time point. The age of these mice has been set to seven weeks in our study. Further information can be found in Additional file 1 and the original publication. After QC, there were four samples left.

Schillebeeckx [32]

The Schillebeeckx study consists of 23 samples all originating from the liver, the adrenal gland and from endometrial cancer. The mice were ovariectomised at the age of three to four weeks; samples were collected after an additional three months. In the original study, a laser capture microdissection RRBS method was introduced. For our study, we selected the liver samples, which were generated using normal RRBS; after QC, three samples were left. In our study, these samples were used as a validation to see how the predictor works for an unobserved time point. The age of these mice has been set to 16 weeks in our study. Further information can be found in Additional file 1 and the original publication.

Age prediction

To predict the mouse age, we adopted a similar approach to those utilised in human studies [5, 6], namely

Stubbs *et al. Genome Biology* (2017) 18:68

Page 12 of 14

an elastic-net regression model. First, we intersected the sites, which were available with more than five-fold coverage in all training and test samples used, totalling to 17,992 sites. By selection of the sites available in all datasets, we hope to have selected a set of methylation sites, which will be present in most RRBS studies, irrespective of size selection and data handling. In addition, we filtered out both sex chromosomes (X and Y) and the mitochondrial genome to ensure that the model would not be sex-specific nor hampered by the unreliability of mitochondrial genome bisulfite conversion. After selection of the sites and samples, we have used a quantile normalization to normalize the methylation values, followed by a standardization, putting the mean methylation per site to 0 and the standard deviation to 1.

For the predictor, we used the elastic-net generalised linear model as implemented in the GLMNET package [33]. To optimise the alpha, defining the elastic net mixing parameter (1 for lasso to 0 for ridge) and to optimise the lambda, the regularisation parameter, we used a double-loop cross-validation setup. This setup is described in Ronde et al. [41]. We have trained the model to predict the log transformed mouse age (in weeks); three weeks were added before log transforming the ages to be able to predict samples pre-birth.

For the training set we selected 129 healthy samples from the Babraham, Reizel and Cannon studies; sample details are in Additional file 4. By using an internal ten-fold cross-validation, in the inner loop the optimal alpha (0.05) and optimal lambda (0.93) were identified. In the outer loop the actual performance of the predictor was scored (as assessed by the mean squared error). After the cross-validation in the training, we built the final model on all 129 samples to get our final model. To get to our final model, we have taken the betas as derived from GLMNET for the selected sites (329) and trained a quadratic function using the nls function in R to transform the raw prediction scores (sum of the product of the beta weights multiplied by their respective methylation level) to the log age in weeks. The final function used is:

$$log(age) = 0.1207x^2 + 1.2424x + 2.5440$$

where x is the summed beta score per sample.

A set of healthy and treated samples, originating from the same three studies, and the Schillebeeckx and Zhang studies were used to assess the usability of the final model. The MAE of the prediction was found to be 3.33 weeks. Furthermore, the model has been used to assess the influence of diet on the methylation age, using the Cannon training samples and the influence of male and female castration on the methylation age.

MouseEpigeneticClock script

To predict methylation age from new samples we have generated an easy to use R project and deposited it under GNU General Public License at Zenodo [42] and as a GitHub project: https://github.com/EpigenomeClock/MouseEpigeneticClock.

Additional files

Additional file 1: List of all samples collected in this study, as well as all additional publicly available datasets used. (XLS 27 kb)

Additional file 2: Characterisation of ageing associated methylation changes at discrete CpGs, related to Fig. 1. (PDF 974 kb)

Additional file 3: Characterisation of ageing associated methylation changes at discrete CpGs excluding newborn samples, related to Fig. 1. (PDF 1117 kb)

Additional file 4: Table of Training and Test samples. (XLS 26 kb)

Additional file 5: Properties of the mouse epigenetic clock sites, related to Fig. 2. (PDF 1310 kb)

Additional file 6: Table of mouse epigenetic clock sites. (XLS 94 kb)

Additional file 7: Additional detail on the mouse epigenetic clock, related to Fig. 2. (PDF 152 kb)

Additional file 8: Information on the datasets properties of the clock sites, related to Fig. 2. (PDF 153 kb)

Additional file 9: Testing age relation of the human clock sites in the mouse, related to Fig. 3. (PDF 538 kb)

Additional file 10: Predicted age of test samples as coloured by sex, related to Fig. 4. (PDF 95 kb)

Abbreviations
CGI: CpG island; GO: Gene Ontology; MAE: Median absolute error; PCA: Principal component analysis; RRBS: Reduced-representation bisulfite sequencing; WGBS: Whole-genome bisulfite sequencing

Stubbs *et al. Genome Biology* (2017) 18:68

Page 13 of 14

[b]Immunology Programme, The Babraham Institute, Cambridge CB22 3AT, UK
[c]Epigenetics Programme, The Babraham Institute, Cambridge CB22 3AT, UK
[d]BSU, The Babraham Institute, Cambridge CB22 3AT, UK
[e]Current address: MRC Institute of Genetics & Molecular Medicine, The University of Edinburgh, Edinburgh EH4 2XU, UK
[f]Current address: Instituto de Medicina Molecular, 1649-028 Lisboa, Portugal

Funding

Authors' contributions

TMS, FvM and WR designed the study; TMS, FvM and AK-S collected samples and performed experiments; TMS, MJB and OS analysed data and developed the model; FK performed processing of data; TMS, MJB, FvM, OS and WR interpreted data and wrote the manuscript. All authors approved the final manuscript.

Competing interests

WR is a consultant and shareholder of Cambridge Epigenetix.

Ethics approval

All animal work carried out as part of this study is covered by a UK Home Office Project Licence held by W.R. under the 1986 Animal (Scientific Procedures) Act and approved by the Babraham Research Campus Animal Welfare and Ethical Review Body.

Accession numbers

The accession numbers for the next-generation-sequencing data reported in this study are GEO: GSE93957.

Publisher's Note

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

Author details

[1]Epigenetics Programme, The Babraham Institute, Cambridge CB22 3AT, UK. [2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton CB10 1SD, UK. [3]Immunology Programme, The Babraham Institute, Cambridge CB22 3AT, UK. [4]Bioinformatics Group, The Babraham Institute, Cambridge CB22 3AT, UK. [5]Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK. [6]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK.

References

1. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153:1194–217.
2. Booth LN, Brunet A. The aging epigenome. Mol Cell. 2016;62:728–44.
3. Benayoun BA, Pollina EA, Brunet A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. Nat Rev Mol Cell Biol. 2015;16:593–610.
4. Ocampo A, Reddy P, Martinez-Redondo P, Platero-Luengo A, Hatanaka F, Hishida T, et al. In vivo amelioration of age-associated hallmarks by partial reprogramming. Cell. 2016;167:1719–33. e12.
5. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49:359–67.
6. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14:R115.
7. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol. 2014;15:R24. BioMed Central Ltd.
8. Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, et al. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol. 2015;16:25.
9. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai P-C, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging. 2016;8:1844–65.
10. Horvath S, Levine AJ. HIV-1 infection accelerates age according to the epigenetic clock. J Infect Dis. 2015;212:1563–73.
11. Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D, et al. Accelerated epigenetic aging in Down syndrome. Aging Cell. 2015;14:491–5.
12. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schönfels W, Ahrens M, et al. Obesity accelerates epigenetic aging of human liver. Proc Natl Acad Sci U S A. 2014;111:15538–43.
13. Marioni RE, Shah S, McRae AF, Ritchie SJ, Muniz-Terrera G, Harris SE, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. Int J Epidemiol. 2015;44:1388–96.
14. Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. Aging. 2015;7:1130–42.
15. Horvath S, Pirazzini C, Bacalini MG, Gentilini D, Di Blasio AM, Delledonne M, et al. Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. Aging. 2015;7:1159–70.
16. Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. Aging. 2015;7:690–700.
17. Levine ME, Lu AT, Bennett DA, Horvath S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. Aging. 2015;7:1198–211.
18. Lu AT, Hannon E, Levine ME, Hao K, Crimmins EM, Lunnon K, et al. Genetic variants near MLST8 and DHX57 affect the epigenetic age of the cerebellum. Nat Commun. 2016;7:10561.
19. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. Genome Biol. 2016;17:171. BioMed Central.
20. Horvath S, Langfelder P, Kwak S, Aaronson J, Rosinski J, Vogt TF, et al. Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. Aging. 2016;8:1485–512.
21. Vidal-Bralo L, Lopez-Golan Y, Mera-Varela A, Rego-Perez I, Horvath S, Zhang Y, et al. Specific premature epigenetic aging of cartilage in osteoarthritis. Aging. 2016;8:2222–31.
22. Carroll JE, Irwin MR, Levine M, Seeman TE, Absher D, Assimes T, et al. Epigenetic aging and immune senescence in women with insomnia symptoms: findings from the Women's Health Initiative Study. Biol Psychiatry. 2017;81:136–44.
23. von Meyenn F, Iurlaro M, Habibi E, Liu NQ, Salehzadeh-Yazdi A, Santos F, et al. Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. Mol Cell. 2016;62:848–61.
24. Iurlaro M, von Meyenn F, Reik W. DNA methylation homeostasis in human and mouse development. Curr Opin Genet Dev. 2017;43:101–9.
25. Spiers H, Hannon E, Wells S, Williams B, Fernandes C, Mill J. Age-associated changes in DNA methylation across multiple tissues in an inbred mouse model. Mech Ageing Dev. 2016;154:20–3.
26. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 2005;33:5868–77.
27. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic reconfiguration during mammalian brain development. Science. 2013;341:1237905.
28. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, et al. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet. 2009;5:e1000602.
29. Cannon MV, Buchner DA, Hester J, Miller H, Sehayek E, Nadeau JH, et al. Maternal nutrition induces pervasive gene expression changes but no detectable DNA methylation differences in the liver of adult offspring. PLoS ONE. 2014;9:e90335.
30. Reizel Y, Spiro A, Sabag O, Skversky Y, Hecht M, Keshet I, et al. Gender-specific postnatal demethylation and establishment of epigenetic memory. Genes Dev. 2015;29:923–33.
31. Zhang C, Hoshida Y, Sadler KC. Comparative epigenomic profiling of the DNA methylome in mouse and zebrafish uncovers high interspecies divergence. Front Genet. 2016;7:110.
32. Schillebeeckx M, Schrade A, Löbs A-K, Pihlajoki M, Wilson DB, Mitra RD. Laser capture microdissection-reduced representation bisulfite sequencing (LCM-RRBS) maps changes in DNA methylation associated with gonadectomy-induced adrenocortical neoplasia in the mouse. Nucleic Acids Res. 2013;41:e116.

Stubbs *et al. Genome Biology* (2017) 18:68

Page 14 of 14

33. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1–22.

34. Gravina S, Dong X, Yu B, Vijg J. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. Genome Biol. 2016;17:150.

35. Asdell SA, Doornenbal H, Joshi SR, Sperling GA. The effects of sex steroid hormones upon longevity in rats. J Reprod Fertil. 1967;14:113–20.

36. Smallwood SA, Tomizawa S-I, Krueger F, Ruf N, Carli N, Segonds-Pichon A, et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet. 2011;43:811–4.

37. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–2.

38. Storey JD. A direct approach to false discovery rates. J R Stat Soc Ser B (Stat Methodol). 2002;64:479–98.

39. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf. 2011;12:77. BioMed Central.

40. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler–a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007;35:W193–200.

41. de Ronde JJ, Bonder MJ, Lips EH, Rodenhuis S, Wessels LFA. Breast cancer subtype specific classifiers of response to neoadjuvant chemotherapy do not outperform classifiers trained on all subtypes. PLoS ONE. 2014;9:e88551.

42. Stubbs TM, Bonder MJ. EpigenomeClock/MouseEpigeneticClock. Zenodo. 2017. http://doi.org/10.5281/zenodo.438976.