Title

Novel epigenetic clock for fetal brain development predicts fetal epigenetic age for iPSCs and iPSC-derived neurons.

Authors

Leonard C. Steg¹, Gemma L. Shireby¹, Jennifer Imm¹, Jonathan P. Davies¹, Robert Flynn¹, Seema C. Namboori¹, Akshay Bhinge¹, Aaron R. Jeffries¹, Joe Burrage¹, Grant W. A. Neilson¹, Emma M. Walker¹, Leo W. Perfect², Jack Price³, Nicholas J. Bray³, Emma L. Cope⁴, Kimberly M. Jones⁴, Nicholas D. Allen⁴, Ehsan Pishva¹⁵, Jonathan Mill¹, Katie Lunnon¹ & Eilis Hannon¹*¹

¹University of Exeter Medical School, College of Medicine and Health, University of Exeter, Exeter, UK.
²Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, UK.
³Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK.
⁴School of Biosciences, Cardiff University, Cardiff, UK.
⁵Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Maastricht, The Netherlands.

* Corresponding author: Eilis Hannon, University of Exeter Medical School, RILD Building Level 3, Royal Devon and Exeter Hospital, Barrack Rd, Exeter. EX2 5DW. UK. E-mail: ej.hannon@exeter.ac.uk

Keywords

Epigenetic clock, DNA methylation, fetal neurodevelopment, induced pluripotent stem cells, iPSC-derived neurons, neuronal precursor cells, DNA methylation clock.
Abstract

Induced pluripotent stem cells (iPSCs) and their differentiated neurons (iPSC-neurons) are a widely used cellular model in the research of the central nervous system. However, it is unknown how well they capture age-associated processes, particularly given that pluripotent cells are only present during the early stages of mammalian development. Epigenetic clocks utilize coordinated age-associated changes in DNA methylation to make predictions that correlate strongly with chronological age, and is has been shown that the induction of pluripotency rejuvenates predicted epigenetic age. As existing clocks are not optimized for the study of brain development, to investigate more precisely the epigenetic age of iPSCs and iPSC-neurons, here, we establish the fetal brain clock (FBC), a bespoke epigenetic clock trained in prenatal neurodevelopmental samples. Our data show that the FBC outperforms other established epigenetic clocks in predicting the age of fetal brain samples. We then applied the FBC to DNA methylation data of cellular datasets that have profiled iPSCs and iPSC-derived neuronal precursor cells and neurons and find that these cell types are characterized by a fetal epigenetic age. Furthermore, while differentiation from iPSCs to neurons significantly increases the epigenetic age, iPSC-neurons are still predicted as having fetal epigenetic age. Together our findings reiterate the need for better understanding of the limitations of existing epigenetic clocks for answering biological research questions and highlight a potential limitation of iPSC-neurons as a cellular model for the research of age-related diseases as they might not fully recapitulate an aged phenotype.

MAIN TEXT

Introduction

Induced pluripotent stem cells (iPSCs) offer a unique cellular system to investigate disease in human-derived cells. iPSCs are obtained by taking tissue samples (e.g. skin or blood) from patients and treating the cells with a set of core pluripotency transcription factors that reprogram the cells to a pluripotent state [1]. Established iPSC lines have the capacity to be further differentiated into any cell type,
including neurons, when treated with the appropriate factors [2–4]. This is of particular interest for neuroscience, as the only alternative cellular model for human neurons are immortalized cell lines. However, as immortalized cell lines retain some physiological properties of the cancerous cells they were derived from, (for example undergoing an infinite number of cell divisions) [5] they do not purely represent the neuronal phenotype. iPSC-derived neurons (IPSC-neurons), on the other hand, express appropriate morphological and neurophysiological properties of neurons and subject to different protocols can be differentiated into a wide range of specific neuronal subtypes [6]. iPSCs and their neuronal derivatives have been widely used to research disorders of the central nervous system, including developmental disorders such as autism and schizophrenia and age-related diseases such as Alzheimer’s disease (AD) or Parkinson’s disease. However, it is unknown how well iPSCs and especially iPSC-neurons capture age-associated processes, which are fundamental to the study of age-related diseases. Of specific relevance is the fact that pluripotent cells only occur during the early stages of mammalian development and the effect of differentiation from iPSCs towards neurons on the developmental or aging trajectory of the cellular model[7] has yet to be adequate profiled.

Epigenetic mechanisms, such as DNA methylation (DNAm), are chemical processes that stably regulate gene expression, and while they are sensitive to environmental stimuli, they also underpin key developmental processes [8, 9]. As a consequence, they are not only dynamic over the life course, but are dynamic in a consistent manner across individuals [10]. There has been much interest and success in capitalizing on these patterns of epigenetic variation to predict the age of an individual from a biological sample. Chronological age predictors based on DNAm profiles, known as “epigenetic clocks” or “DNAm clocks”, have become commonplace in the epigenetic literature, as they can predict the “epigenetic age” of a sample, which correlates strongly with chronological age. Epigenetic age, defined here as age predicted by an epigenetic clock, correlates strongly with chronological age, albeit not perfectly, and there is interest in whether the deviations from this prediction, referred to as age
acceleration, are meaningful in the context of disease [11, 12]. The most well-known epigenetic clock is the Horvath multi-tissue clock (MTC) which was developed using a large number of samples (n > 8000) from 51 different tissues and cell types [13]. While in general the MTC generates reliable predictions of chronological age for most samples, there is increasing recognition that its performance is dependent on the characteristics of the training dataset (e.g. tissue/cell type and age range) and greater accuracy can be achieved with clocks trained on more refined sample sets, such as those from a single tissue [14, 15]. To this end, a number of new DNAm clocks have been established based on specific tissue types, like whole blood [16] or brain tissue [14], which demonstrate more accurate predictions within the specified tissue. A less established refinement of epigenetics clocks is the application to specific developmental stages, with embryonic or fetal samples, in particular, either omitted from or underrepresented in most existing clocks. An exception here are algorithms that predict gestational age (GA) of newborns, developed using pre- and perinatal DNAm data from blood samples [17] or placental samples [18]. While existing epigenetic clocks have been shown to accurately predict age in either postnatal brain samples (predominantly adult and older age) or non-brain prenatal samples, these tools have not been through tested on prenatal brain samples and it is unknown whether they are able to delineate the earliest stages of brain development.

Previous analysis applying the MTC to DNAm data generated from iPSCs and their corresponding primary cells from adult donors found that despite the primary cells predicting the donor’s chronological age accurately, the induction of pluripotency reversed the aging process, with the matched iPSCs predicted as postnatal but close to zero [13]. As human pluripotent cells only occur during prenatal development, we hypothesize that existing clocks are not sensitive enough to accurately predict iPSCs at prenatal developmental stages. The inability to accurately estimate age during this crucial stage of neurodevelopment limits our ability to profile changes in epigenetic age induced by the differentiation of iPSCs towards neurons using already established DNAm clocks. Here
we present a novel DNAm clock developed using prenatal brain samples that accurately predicts fetal
age, outperforming other DNAm clocks in neurodevelopmental samples. We then apply our clock to
iPSCs, iPSC-derived neuronal progenitor cells (NPCs) and neurons to characterize the epigenetic age
of these cellular models before and throughout differentiation.

**Methods**

*Data pre-processing and quality control.*

All statistical analyses were performed using R version 3.5.2 (https://www.r-project.org/). All datasets
of which raw data was available were pre-processed by our group following a standard quality control
(QC) and normalization pipeline as described before [14] using either the R package *watermelon* [19]
*or bigmelon* [20]. Briefly, samples with low signal intensities or incomplete bisulphite conversion were
excluded prior to applying the *pfilter()* function from the *watermelon* package excluding samples with
>1 % of probes with a detection P value >0.05 and probes with >1 % of samples with detection P value
>0.05. This was followed by the exclusion of probes known to be affected by SNPs or known to cross-
hybridize [21] and QC was finished by quantile normalization using the *dasen()* function of the
packages *watermelon* or *bigmelon* [19, 20].

**Fetal brain samples.**

To develop and profile the performance of our fetal brain clock (FBC), we collated a dataset of 258
fetal brain samples (see Table S1) of which 194 were processed by our lab and 64 are a subset (age <
0 years) of a publicly available dataset downloaded from the Gene Expression Omnibus (GEO;
https://www.ncbi.nlm.nih.gov/geo/; GSE74193) [22]. Of the samples processed in our lab, 154 overlap
those included in [23] following additional outlier filtering by principal component analysis, where
DNAm was quantified using the Illumina 450K DNA methylation array. The other 40 samples are a new
dataset where DNAm was quantified using the Illumina EPIC DNA methylation array. To harmonize the age variable across datasets, age was converted into days post-conception (dpc), as it represents the most precise unit of age available in the datasets. Age provided as weeks post-conception was transformed to days post-conception by dividing by 7, and age reported in (negative) years was transformed to days post-conception by multiplying by 365 and adding 280.

Training and testing datasets.

To create two separate datasets for the purpose of training and testing the FBC, 75% of the samples from each dataset were randomly assigned into a training dataset (n = 193, age range = 37-184 dpc, age median = 99 dpc), while the remaining 25% were collated into the testing dataset (n = 65, age range = 23-153 dpc, age median = 99 dpc) (Figure S1), such that there was no overlap between the training and testing dataset. Of note, a few samples (15 out of 258) would be actually defined as embryonic (GA < 63 dpc) and not fetal. To simplify the FBC development, only probes available in all cohorts and with complete data after QC were taken forward (n = 385,069 probes).

Adult samples.

To evaluate the performance of the DNAm clocks in adult brain samples we utilized the Brains for Dementia Research (BDR) cohort DNA methylation data [14]. Briefly, these data consist of 1,221 samples from 632 donors (age range 41-104 years, median = 84 years), with DNA extracted from the prefrontal cortex (n = 610) and occipital cortex (n = 611). DNA methylation was quantified using the Illumina EPIC DNA methylation array, and were pre-processed using our group’s standard QC pipeline as described in [14].

Cellular samples.
Five different DNAm datasets profiling iPSCs, iPSC-derived NPCs and iPSC-derived neurons were used to characterize epigenetic age of the neuronal cell model, details of which can be found in Table S1. Three of these datasets (Imm, Price, Bhinge) were generated by our lab, where DNA methylation was quantified using the Illumina EPIC DNA methylation array. These were supplemented by two publicly available datasets, downloaded from GEO (Sultanov, GSE105093, and Fernández-Santiago, GSE51921) consisting of Illumina 450K DNA methylation array data [3, 4]. References describing the origin of cell lines and the different methods used for cell culture and differentiation are listed in Table S1. Pre-processing and QC for the Fernández-Santiago dataset was not performed in our lab as no raw data was available.

Fetal brain clock development.

To develop the fetal brain clock we applied an elastic net (EN) regression model, using the approach described by Horvath [13], regressing chronological age against DNAm level of all available probes. The EN algorithm selects a subset of DNA methylation probes that together produce the optimal prediction of the outcome, in this case chronological age, by combining ridge and LASSO (Least Absolute Shrinkage and Selection Operator) regression. Briefly, ridge regression penalizes the sum of squared coefficients while LASSO penalizes the sum of the absolute values of coefficients. EN is a combination of both methods, where the user specifies the extent of the mixing of the two methods as a number between 0 and 1, in our application this was set to 0.5 [24]. EN is implemented with the R package GLMnet [25]. The shrinkage parameter lambda was calculated using 10-fold cross-validation on the data, which resulted in a lambda of 3.27.

Fetal brain clock evaluation.

To profile the performance of the FBC, we additionally applied three established DNAm clocks: Horvath’s multi-tissue clock (MTC) [13], Knight’s Gestational Age clock (GAC) [17] and Lee’s Control
Placental epigenetic clock (CPC) [18]. The clocks were applied using the agep() function of the watermelon package [19], where the default estimates the MTC and other clocks (here the GAC and CPC) can be estimated by providing the necessary coefficient and intercept values. The predictive accuracy of each clock was profiled in each dataset by two measures: Pearson’s correlation coefficient with reported chronological age and root mean squared error (RMSE). To investigate potential effects of sex on the predicted epigenetic age a linear model was fitted in the testing dataset with FBC predicted epigenetic age as dependent variable, chronological age and sex as main effects and an interaction of chronological age and sex.

Comparison of cellular states.

The FBC was applied to DNAm data for all cellular samples available. To test for differences in predicted epigenetic age between cell stages within each dataset, either two sample t-tests or ANOVA followed by Tukey HSD multiple comparison (when three cell stages were available), were used. To combine results across all datasets, a mixed effects linear model was fitted with predicted epigenetic age as the dependent variable, a fixed effect for cell stage represented as two dummy variables contrasting NPCs vs iPSCs and iPSC-neurons vs iPSCs as and a random effect for dataset.

Results

Fetal brain clock outperforms existing DNAm clocks at predicting age of prenatal brain samples.

We applied EN regression to genome wide DNAm data from a set of prenatal brain samples (n = 193; Table S1 and Figure S1), to develop the fetal brain clock (FBC). A subset of 107 DNAm probes were assigned non-zero coefficients and therefore were selected as the basis of the FBC (Table S2). We found no overlap in the DNAm sites selected for the FBC and DNAm sites used in the other clocks we considered. Testing the FBC clock in an independent test dataset of fetal brain samples (Table S1 and
Figure S1) to evaluate its performance we found a strong linear relationship between chronological and predicted prenatal age ($r = 0.80$; Figure 1A) with the majority of samples predicted within 15 days of their actual chronological age (RMSE = 14.84 dpc). To benchmark the performance of our clock, we compared it to three existing DNAm clocks: Horvath’s MTC [13], Knight’s GAC [17] and Lee’s CPC [18]. These clocks were selected as they represent either the most well-established algorithm with the broadest applicability (MTC) or were specifically developed to predict pre- and perinatal gestational ages, albeit in non-brain tissue (GAC and CPC). Of note, the MTC only predicted 27 fetal brain samples (41.2%) as prenatal (dpc < 280) with a very weak correlation between chronological and predicted age ($r_{\text{MTC}} = 0.06$). This correlation is much weaker than those reported in the original manuscript when Horvath tested the clock in adult samples [13], highlighting the challenges with extrapolating clocks to samples which were not well represented in the model development. By comparison, the GAC and CPC perform better than the MTC, although they have smaller correlation coefficients ($r_{\text{GAC}} = 0.52$ and $r_{\text{CPC}} = 0.76$) and are associated with a larger error (RMSE$^\text{GAC} = 21.32$ and RMSE$^\text{CPC} = 60.08$) than the FBC. Interestingly, while the predictions from the GAC are more precise, it’s not as effective at ranking the samples by age as the CPC. Taken together, these results demonstrate that our novel FBC outperforms existing clocks at predicting age in fetal brain samples, and therefore is the optimal tool available to profile the epigenetic age in models of neuronal development. When applying clocks to the training data, the three established clocks produce similar correlations and RMSEs as in the testing data. As expected, the predictions of the FBC in the training data are more accurate than the predictions in the testing data reflecting overfitting of the model (Figure S2). Given our previous finding of divergent, sex-specific age trajectories at multiple DNAm sites during prenatal development [23], we tested whether the FBC performed differently between males and females in our testing dataset. Although this analysis initially indicated a significant difference in the correlation with age between males and females ($P_{\text{Sex}} = 0.0007$), on closer inspection we suspected that this was likely driven by outliers. Indeed, a sensitivity analysis excluding the two samples with youngest and oldest predicted ages produced a non-significant result ($P_{\text{Sex}} = 0.081$).
Fetal and gestational age clocks are not able to predict adult ages in adult brain tissue

All four clocks were additionally compared using a dataset where DNAm was profiled from adult brain samples (Figure S3). As expected, the FBC performs poorly in this sample set, with all samples predicted as prenatal, while the MTC performs the best ($r_{MTC} = 0.65$, $RMSE_{MTC} = 20.11$ years) as it is the only clock we considered that was developed using adult samples. As with the FBC, the GAC and CPC fail to produce predictions of post-natal age again reflecting the fact that they were also constructed using data from pre- or perinatal samples.

Fetal brain clock captures differences in differentiation of iPSCs towards neurons

Having demonstrated that our novel FBC is the optimal clock to profile age in prenatal brain samples, we applied it to DNAm data from five cellular studies to determine the neuronal development trajectory of iPSCs differentiating towards neurons. All samples were estimated to have a fetal epigenetic age, regardless of cell stage, cell line origin or differentiation protocol. Performing statistical comparisons of the estimated ages between iPSCs and neurons, we observed significant differences for all datasets (Figure 2). For the Imm dataset, which also included proliferative NPCs in comparison to postmitotic neurons, we additionally found a significant difference between NPCs and iPSC-neurons ($\Delta_{mean} = 20.0$ dpc, $P = 0.00039$), but not between iPSCs and NPCs ($\Delta_{mean} = 10$ dpc, $P = 0.24$). Combining the data across the five studies, the observed means of the three cell stages were $mean_{iPSC} = 72.14$ dpc, $mean_{NPC} = 68.5$ dpc and $mean_{iPSC-neurons} = 84.28$ dpc (Figure 3). Using a mixed effects model to meta-analyze across studies, we found that iPSC-neurons were predicted to have a significantly advanced epigenetic age compared to iPSCs ($\Delta_{mean} = 12.14$ dpc, $P = 6.94E-9$) but no significant difference was observed between iPSC and NPCs ($\Delta_{mean} = -3.64$ dpc, $P = 0.39$).
Discussion

In this study we established a novel epigenetic clock, the fetal brain clock (FBC), to specifically profile the earliest stages of human neurodevelopment and applied it to determine the epigenetic age of iPSCs and iPSC-derived NPCs and neurons. Given the lack of fetal brain samples in the development of existing DNAm clocks prior to this study, there was no tool optimized for estimating the age of fetal brain samples from DNAm data limiting the ability to characterize neuronal models or indeed any model of neuronal development. We showed that, in an independent test dataset, the FBC generates predictions that correlate strongly with chronological age in prenatal brain samples. Furthermore, it outperforms both a pan-tissue epigenetic clock (Horvath’s MTC), and epigenetic clocks focused on the same developmental stage, but based on DNAm profiled in different tissues (Knight’s GAC and Lee’s CPC) [13, 17, 18]. The FBC outperforms these clocks using either correlation or error statistics (RMSE), indicating that it is not only better at ranking the samples, but it generates more precise estimates. In contrast, the FBC, performed poorly in an adult brain dataset. Altogether, this reinforces the findings of previous studies that have also demonstrated that the applicability of DNAm clocks is dependent on the characteristics of the dataset(s) they were trained on, with the tissue and age range of the training samples of particular relevance [14, 15]. More specifically, we note that while the accuracy of a DNAm clock is typically decreased in tissues not represented in its training data, clocks are limited to predicting ages represented in the training data. If the testing sample lies outside of the age range of the training data, the clock is unable to provide an appropriate prediction, suggesting that, in general, age range is more important than tissue when training a clock.

Previous use of epigenetic clocks has shown that the predicted epigenetic age of iPSCs is significantly younger than the somatic tissue the cells originate from and the chronological age of the donor at sample donation [13]. This highlights that the induction of pluripotency reprograms the epigenome, including at the loci used in the clock algorithm, ultimately leading to a younger predicted epigenetic
However, in these analyses the predicted ages remained postnatal, which is unexpected as human pluripotent cells only occur during the early stages of human development and hence we hypothesized that, with an adequately calibrated clock, iPSCs would be expected to be estimated as early prenatal. Applying the FBC to five datasets of iPSCs and iPSC-derived NPCs and neurons, we found this to be the case, where iPSCs were estimated as having a mean age of 72.1 dpc, fitting our hypothesis that they reflect at least first trimester developmental stages. These results align with studies that have reported rejuvenation effects on the transcriptome, telomeres and mitochondria of iPSCs following reprogramming [26–28]. In addition, we profiled the effect on predicted epigenetic age following the differentiation of iPSCs towards neurons reporting that this increases the mean predicted age to 84.3 dpc. This age coincides with fetal neurogenesis and suggests that while differentiation does induce an aging process, it does not accelerate iPSC-neurons to a postnatal state. Of note, Mertens and colleagues found that, while iPSCs lose age related transcriptomic signatures, induced neurons (iNs; neurons directly reprogrammed from fibroblasts) keep their specific aging signatures [27]. Therefore, it would be interesting to apply our FBC to iNs, iPSCs, iPSC-neurons and their corresponding somatic tissues to verify whether age associated methylation differences are also preserved in iNs. Altogether, our results indicate that iPSC-neurons may be limited when researching age-related diseases, like Alzheimer’s disease or other dementias, as many molecular processes related to an aging phenotype, may not be recapitulated.

While a strength of our study is the development of a bespoke clock to optimally profile prenatal age in human samples obtained during fetal development, the FBC has in all datasets tested, only generated prenatal estimates. Despite this, we are confident that the FBC correctly predicts fetal epigenetic ages for the cellular data as the predictions for both iPSC and iPSC-neurons are comfortably contained within the training data age range. This indicates that the predictions are not induced purely by saturation of the coefficients. In contrast, estimates for the adult brain dataset did lie at the
extreme ends, bringing doubt on the meaningfulness of the predictions, an observation that is confirmed by the lack of correlation with chronological age.

In summary, we demonstrate that established DNAm clocks struggle to capture changes in epigenetic age during neurodevelopment and for precise predictions a bespoke clock is required. To this end, we developed the FBC, a robust predictor of prenatal age in human fetal brain samples. Using this clock to assess the epigenetic age of iPSCs and differentiated neurons, we found that iPSCs and derived NPCs and neurons reflect prenatal developmental stages. Our findings question the suitability of the iPSC-neurons for the study of aging associated processes.

List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Alzheimer’s Disease</td>
</tr>
<tr>
<td>BDR</td>
<td>Brains for Dementia Research</td>
</tr>
<tr>
<td>CPC</td>
<td>Control Placental Clock</td>
</tr>
<tr>
<td>DNAm</td>
<td>DNA methylation</td>
</tr>
<tr>
<td>dpc</td>
<td>Days post-conception</td>
</tr>
<tr>
<td>EN</td>
<td>Elastic net</td>
</tr>
<tr>
<td>FBC</td>
<td>Fetal brain clock</td>
</tr>
<tr>
<td>GA</td>
<td>Gestational age</td>
</tr>
<tr>
<td>GAC</td>
<td>Gestation age clock</td>
</tr>
<tr>
<td>GEO</td>
<td>Gene Expression Omnibus</td>
</tr>
<tr>
<td>iN</td>
<td>Induced neuron</td>
</tr>
<tr>
<td>iPSC</td>
<td>Induced pluripotent stem cell</td>
</tr>
<tr>
<td>iPSC-neuron</td>
<td>iPSC-derived neuron</td>
</tr>
<tr>
<td>LASSO</td>
<td>Least Absolute Shrinkage and Selection Operator</td>
</tr>
<tr>
<td>MTC</td>
<td>Multi-tissue clock</td>
</tr>
</tbody>
</table>
NPC Neuronal precursor cell
QC Quality control
RMSE Root mean squared error

Declarations

Ethical approval

Ethical approval for the collection of human fetal brain tissue acquired from the Human Developmental Biology Resource was granted by the Royal Free Hospital research ethics committee under reference 08/H0712/34 and Human Tissue Authority material storage license 12220. Ethical approval for collection of human fetal brain tissue for the Medical Research Council Brain Bank was granted under reference 08/MRE09/38. Ethical approval for the work with methylomic data of human brain tissue from the BDR cohort was granted by the University of Exeter, College of Medicine and Health Research Ethics Committee under reference Mar20/D/009Δ5.

Consent for publication

Not applicable

Availability of data and materials

Raw and normalized DNAm data for the human fetal brain samples used to test and train the fetal brain clock have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE58885, GSE74193 and GSE157908. DNAm data for the cellular dataset are also available under accession numbers GSE158089, GSE105093 and GSE51921. DNAm data for the cellular dataset Bhinge is available from the corresponding author on request. The coefficients of the FBC are included within the article in Table
The custom R script used for this study is available at https://github.com/LSteg/EpigeneticFetalClock.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was in part funded by the Simons Foundation (SFARI) grant 573312 awarded to JM. The human embryonic and fetal material was provided by the Joint MRC Wellcome Trust Human Developmental Biology Resource (http://www.hdbr.org), which is funded by Medical Research Council grant G0700089 and Wellcome Trust grant GR082557. The BDR is jointly funded by Alzheimer's Research UK and the Alzheimer's Society in association with the Medical Research Council. The South West Dementia Brain Bank is part of the Brains for Dementia Research program, jointly funded by Alzheimer's Research UK and Alzheimer's Society, and is also supported by BRACE (Bristol Research into Alzheimer's and Care of the Elderly) and the Medical Research Council. E.H., and J.M. were supported by Medical Research Council grant K013807. G.S. was supported by a PhD studentship from the Alzheimer's Society. N.J.B. was supported by the Medical Research Council grant MR/L010674/2. N.D.A. was supported by Medical Research Council grant N013255/1.

Authors’ contributions

acquired and processed samples and data; L.C.S, G.L.S. and E.H analyzed and interpreted data; L.C.S.
and E.H. wrote the paper.

Acknowledgements

We would like to gratefully acknowledge all donors and their families for the tissue provided for this
study.

Figure Legends

Figure 1 Fetal brain clock outperforms other DNAm clocks when applied to neurodevelopmental
samples. Shown are scatterplots comparing chronological age (x-axis; days post-conception) against
(y-axis; days post-conception) predicted epigenetic age calculated using A Fetal Brain Clock (FBC) B
Horvath’s multi tissue clock (MTC); C Knight’s Gestational Age Clock (GAC); D Lee’s Control Placental
Clock (CPC); in an independent fetal brain sample (n = 65). The black line indicates the identity line of
chronological and predicted epigenetic age and represents a perfect prediction. Two statistics were
calculated to evaluate the precision of each DNAm clock: Pearson’s correlation coefficient (r) and the
root mean squared error (RMSE).

Figure 2 Comparisons of predicted epigenetic age from fetal brain clock between iPSC
differentiation states. Boxplots comparing the distribution of predicted epigenetic age (days post-
conception) separated by cellular stage, where each panel represents a different dataset. P values of
Tukey HSD corrected ANOVA for the Imm dataset and two-sample t-tests for Price, Fernández-
Santiago and Sultanov datasets are given. F. -S. = Fernández-Santiago.
Figure 3 iPSCs are significantly younger than iPSC induced neurons using age estimated by the FBC. Boxplots of predicted epigenetic age calculated using the FBC where samples are grouped by cell stage (n = 82, 30 iPSCs, 4 NPCs, 48 iPSC-neurons) and colored by dataset. P values from mixed effects model are given for differences between iPSCs and NPCs (non-significant) and iPSC and neurons. F.-S. = Fernández-Santiago.

Additional Files

Additional File 1: Table S1 Summary of fetal, adult and cellular datasets.

Additional File 2: Table S2 Intercept and coefficients of the fetal brain clock.

Additional File 3: Figure S1 Histogram of age distribution. Chronological age of the fetal samples measured in days post-conception in A) training data and B) testing data.

Additional File 4: Figure S2 Comparison of predictions from the four DNA methylation clocks in the training data (n = 193). Shown are scatterplots comparing chronological age (x-axis) against age (y-axis) predicted epigenetic age for A Fetal Brain Clock (FBC); B Horvath’s multi tissue clock (MTC); C Knight’s Gestational Age Clock (GAC); D Lee’s Control Placental Clock (CPC) in the data used for training of the FBC. Where necessary, predicted age was converted to days post-conception. The black line indicates the identity line of chronological and predicted epigenetic age and represents a perfect prediction. Two statistics were calculated to evaluate the precision of each DNA methylation clock: Pearson’s correlation coefficient (r) and the root mean squared error (RMSE).

Additional File 5: Figure S3 Comparison of predictions from the four DNA methylation clocks in adult brain samples (n = 1221). Shown are scatterplots comparing chronological age (x-axis) against (y-axis)
predicted epigenetic age calculated using **A** Fetal Brain Clock (FBC); **B** Horvath’s multi tissue clock (MTC); **C** Knight’s Gestational Age Clock (GAC); **D** Lee’s Control Placental Clock (CPC) in an independent adult brain dataset. Where necessary, predicted age was converted to years, where 0 indicates birth.

The black line indicates the identity line of chronological and predicted epigenetic age and represents a perfect prediction. Two statistics were calculated to evaluate the precision of each DNAm clock: Pearson’s correlation coefficient (r) and the root mean squared error (RMSE).

### References


Predicted Age in days post-conception

A FBC

\[ r = 0.8 \]
\[ \text{RMSE} = 14.84 \]

B MTC

\[ r = 0.06 \]
\[ \text{RMSE} = 300.5 \]

C GAC

\[ r = 0.52 \]
\[ \text{RMSE} = 21.32 \]

D CPC

\[ r = 0.76 \]
\[ \text{RMSE} = 60.08 \]
Predicted Age in days post-conception

A Imm

B Price

C Bhinge

D F.-S.

E Sultanov

Cell Stage