1   **Title**

2   Novel epigenetic clock for fetal brain development predicts fetal epigenetic age for iPSCs and iPSC-

3   derived neurons.

4   **Authors**

5   Leonard C. Steg[1], Gemma L. Shireby[1], Jennifer Imm[1], Jonathan P. Davies[1], Robert Flynn[1], Seema C.

6   Namboori[1], Akshay Bhinge[1], Aaron R.Jeffries[1], Joe Burrage[1], Grant W. A. Neilson[1], Emma M. Walker[1],

7   Leo W. Perfect[2], Jack Price[2], Nicholas J. Bray[3], Emma L. Cope[4], Kimberly M. Jones[4], Nicholas D. Allen[4],

8   Ehsan Pishva[1,5], Jonathan Mill[1], Katie Lunnon[1] & Eilis Hannon[1,*]

9   [1] University of Exeter Medical School, College of Medicine and Health, University of Exeter, Exeter, UK.

10 [2] Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience,

11 King's College London, UK.

12 [3] Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University,

13 Cardiff, UK.

14 [4] School of Biosciences, Cardiff University, Cardiff, UK.

15 [5] Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience

16 (MHeNS), Maastricht University, Maastricht, The Netherlands.

17 * Corresponding author: Eilis Hannon, University of Exeter Medical School, RILD Building Level 3, Royal

18 Devon and Exeter Hospital, Barrack Rd, Exeter. EX2 5DW. UK. E-mail: ej.hannon@exeter.ac.uk

19 **Keywords**

20 Epigenetic clock, DNA methylation, fetal neurodevelopment, induced pluripotent stem cells, iPSC-

21 derived neurons, neuronal precursor cells, DNAm clock

1   **Abstract**

2   Induced pluripotent stem cells (iPSCs) and their differentiated neurons (iPSC-neurons) are a widely

3   used cellular model in the research of the central nervous system. However, it is unknown how well

4   they capture age-associated processes, particularly given that pluripotent cells are only present during

5   the early stages of mammalian development. Epigenetic clocks utilize coordinated age-associated

6   changes in DNA methylation to make predictions that correlate strongly with chronological age, and

7   is has been shown that the induction of pluripotency rejuvenates predicted epigenetic age. As existing

8   clocks are not optimized for the study of brain development, to investigate more precisely the

9   epigenetic age of iPSCs and iPSC-neurons, here, we establish the fetal brain clock (FBC), a bespoke

10 epigenetic clock trained in prenatal neurodevelopmental samples. Our data show that the FBC

11 outperforms other established epigenetic clocks in predicting the age of fetal brain samples. We then

12 applied the FBC to DNA methylation data of cellular datasets that have profiled iPSCs and iPSC-derived

13 neuronal precursor cells and neurons and find that these cell types are characterized by a fetal

14 epigenetic age. Furthermore, while differentiation from iPSCs to neurons significantly increases the

15 epigenetic age, iPSC-neurons are still predicted as having fetal epigenetic age. Together our findings

16 reiterate the need for better understanding of the limitations of existing epigenetic clocks for

17 answering biological research questions and highlight a potential limitation of iPSC-neurons as a

18 cellular model for the research of age-related diseases as they might not fully recapitulate an aged

19 phenotype.

20 **MAIN TEXT**

21 **Introduction**

22 Induced pluripotent stem cells (iPSCs) offer a unique cellular system to investigate disease in human-

23 derived cells. iPSCs are obtained by taking tissue samples (e.g. skin or blood) from patients and treating

24 the cells with a set of core pluripotency transcription factors that reprogram the cells to a pluripotent

25 state [1]. Established iPSC lines have the capacity to be further differentiated into any cell type,

1   including neurons, when treated with the appropriate factors [2–4]. This is of particular interest for

2   neuroscience, as the only alternative cellular model for human neurons are immortalized cell lines.

3   However, as immortalized cell lines retain some physiological properties of the cancerous cells they

4   were derived from, (for example undergoing an infinite number of cell divisions) [5] they do not purely

5   represent the neuronal phenotype. iPSC-derived neurons (iPSC-neurons), on the other hand, express

6   appropriate morphological and neurophysiological properties of neurons and subject to different

7   protocols can be differentiated into a wide range of specific neuronal subtypes [6]. iPSCs and their

8   neuronal derivatives have been widely used to research disorders of the central nervous system,

9   including developmental disorders such as autism and schizophrenia and age-related diseases such as

10   Alzheimer's disease (AD) or Parkinson's disease. However, it is unknown how well iPSCs and especially

11   iPSC-neurons capture age-associated processes, which are fundamental to the study of age-related

12   diseases. Of specific relevance is the fact that pluripotent cells only occur during the early stages of

13   mammalian development and the effect of differentiation from iPSCs towards neurons on the

14   developmental or aging trajectory of the cellular model[7] has yet to be adequate profiled.

15

16   Epigenetic mechanisms, such as DNA methylation (DNAm), are chemical processes that stably regulate

17   gene expression, and while they are sensitive to environmental stimuli, they also underpin key

18   developmental processes [8, 9]. As a consequence, they are not only dynamic over the life course, but

19   are dynamic in a consistent manner across individuals [10]. There has been much interest and success

20   in capitalizing on these patterns of epigenetic variation to predict the age of an individual from a

21   biological sample. Chronological age predictors based on DNAm profiles, known as "epigenetic clocks"

22   or "DNAm clocks", have become commonplace in the epigenetic literature, as they can predict the

23   "epigenetic age" of a sample, which correlates strongly with chronological age. Epigenetic age, defined

24   here as age predicted by an epigenetic clock, correlates strongly with chronological age, albeit not

25   perfectly, and there is interest in whether the deviations from this prediction, referred to as age

1   acceleration, are meaningful in the context of disease [11, 12]. The most well-known epigenetic clock

2   is the Horvath multi-tissue clock (MTC) which was developed using a large number of samples (n >

3   8000) from 51 different tissues and cell types [13]. While in general the MTC generates reliable

4   predictions of chronological age for most samples, there is increasing recognition that its performance

5   is dependent on the characteristics of the training dataset (e.g. tissue/cell type and age range) and

6   greater accuracy can be achieved with clocks trained on more refined sample sets, such as those from

7   a single tissue [14, 15]. To this end, a number of new DNAm clocks have been established based on

8   specific tissue types, like whole blood [16] or brain tissue [14], which demonstrate more accurate

9   predictions within the specified tissue. A less established refinement of epigenetics clocks is the

10   application to specific developmental stages, with embryonic or fetal samples, in particular, either

11   omitted from or underrepresented in most existing clocks. An exception here are algorithms that

12   predict gestational age (GA) of newborns, developed using pre- and perinatal DNAm data from blood

13   samples [17] or placental samples [18]. While existing epigenetic clocks have been shown to accurately

14   predict age in either postnatal brain samples (predominantly adult and older age) or non-brain

15   prenatal samples, these tools have not been through tested on prenatal brain samples and it is

16   unknown whether they are able to delineate the earliest stages of brain development.

17

18   Previous analysis applying the MTC to DNAm data generated from iPSCs and their corresponding

19   primary cells from adult donors found that despite the primary cells predicting the donor's

20   chronological age accurately, the induction of pluripotency reversed the aging process, with the

21   matched iPSCs predicted as postnatal but close to zero [13]. As human pluripotent cells only occur

22   during prenatal development, we hypothesize that existing clocks are not sensitive enough to

23   accurately predict iPSCs at prenatal developmental stages. The inability to accurately estimate age

24   during this crucial stage of neurodevelopment limits our ability to profile changes in epigenetic age

25   induced by the differentiation of iPSCs towards neurons using already established DNAm clocks. Here

1   we present a novel DNAm clock developed using prenatal brain samples that accurately predicts fetal

2   age, outperforming other DNAm clocks in neurodevelopmental samples. We then apply our clock to

3   iPSCs, iPSC-derived neuronal progenitor cells (NPCs) and neurons to characterize the epigenetic age

4   of these cellular models before and throughout differentiation.

5

6   **Methods**

7   *Data pre-processing and quality control.*

8   All statistical analyses were performed using R version 3.5.2 (https://www.r-project.org/). All datasets

9   of which raw data was available were pre-processed by our group following a standard quality control

10   (QC) and normalization pipeline as described before [14] using either the R package *wateRmelon* [19]

11   or *bigmelon* [20]. Briefly, samples with low signal intensities or incomplete bisulphite conversion were

12   excluded prior to applying the *pfilter*() function from the *wateRmelon* package excluding samples with

13   >1 % of probes with a detection P value >0.05 and probes with >1 % of samples with detection P value

14   >0.05. This was followed by the exclusion of probes known to be affected by SNPs or known to cross-

15   hybridize [21] and QC was finished by quantile normalization using the *dasen*() function of the

16   packages *wateRmelon* or *bigmelon* [19, 20].

17

18   *Fetal brain samples.*

19   To develop and profile the performance of our fetal brain clock (FBC), we collated a dataset of 258

20   fetal brain samples (see **Table S1**) of which 194 were processed by our lab and 64 are a subset (age <

21   0 years) of a publicly available dataset downloaded from the Gene Expression Omnibus (GEO;

22   https://www.ncbi.nlm.nih.gov/geo/; GSE74193) [22]. Of the samples processed in our lab, 154 overlap

23   those included in [23] following additional outlier filtering by principal component analysis, where

24   DNAm was quantified using the Illumina 450K DNA methylation array. The other 40 samples are a new

1 dataset where DNAm was quantified using the Illumina EPIC DNA methylation array. To harmonize

2 the age variable across datasets, age was converted into days post-conception (dpc), as it represents

3 the most precise unit of age available in the datasets. Age provided as weeks post-conception was

4 transformed to days post-conception by dividing by 7, and age reported in (negative) years was

5 transformed to days post-conception by multiplying by 365 and adding 280.

6

7 *Training and testing datasets.*

8 To create two separate datasets for the purpose of training and testing the FBC, 75% of the samples

9 from each dataset were randomly assigned into a training dataset (n = 193, age range = 37-184 dpc,

10 age median = 99 dpc), while the remaining 25% were collated into the testing dataset (n = 65, age

11 range = 23-153 dpc, age median = 99 dpc) (**Figure S1**), such that there was no overlap between the

12 training and testing dataset. Of note, a few samples (15 out of 258) would be actually defined as

13 embryonic (GA < 63 dpc) and not fetal. To simplify the FBC development, only probes available in all

14 cohorts and with complete data after QC were taken forward (n = 385,069 probes).

15

16 *Adult samples.*

17 To evaluate the performance of the DNAm clocks in adult brain samples we utilized the Brains for

18 Dementia Research (BDR) cohort DNAm data [14]. Briefly, these data consist of 1,221 samples from

19 632 donors (age range 41-104 years, median = 84 years), with DNA extracted from the prefrontal

20 cortex (n = 610) and occipital cortex (n = 611). DNA methylation was quantified using the Illumina EPIC

21 DNA methylation array, and were pre-processed using our group's standard QC pipeline as described

22 in [14].

23

24 *Cellular samples.*

6

1 Five different DNAm datasets profiling iPSCs, iPSC-derived NPCs and iPSC-derived neurons were used

2 to characterize epigenetic age of the neuronal cell model, details of which can be found in **Table S1**.

3 Three of these datasets (Imm, Price, Bhinge) were generated by our lab, where DNA methylation was

4 quantified using the Illumina EPIC DNA methylation array. These were supplemented by two publicly

5 available datasets, downloaded from GEO (*Sultanov*, GSE105093, and *Fernández-Santiago*, GSE51921)

6 consisting of Illumina 450K DNA methylation array data [3, 4]. References describing the origin of cell

7 lines and the different methods used for cell culture and differentiation are listed in **Table S1**. Pre-

8 processing and QC for the *Fernández-Santiago* dataset was not performed in our lab as no raw data

9 was available.

10

11 *Fetal brain clock development.*

12 To develop the fetal brain clock we applied an elastic net (EN) regression model, using the approach

13 described by Horvath [13], regressing chronological age against DNAm level of all available probes.

14 The EN algorithm selects a subset of DNA methylation probes that together produce the optimal

15 prediction of the outcome, in this case chronological age, by combining ridge and LASSO (Least

16 Absolute Shrinkage and Selection Operator) regression. Briefly, ridge regression penalizes the sum of

17 squared coefficients while LASSO penalizes the sum of the absolute values of coefficients. EN is a

18 combination of both methods, where the user specifies the extent of the mixing of the two methods

19 as a number between 0 and 1, in our application this was set to 0.5 [24]. EN is implemented with the

20 R package *GLMnet* [25]. The shrinkage parameter lambda was calculated using 10-fold cross-validation

21 on the data, which resulted in a lambda of 3.27.

22 *Fetal brain clock evaluation.*

23 To profile the performance of the FBC, we additionally applied three established DNAm clocks:

24 Horvath's multi-tissue clock (MTC) [13], Knight's Gestational Age clock (GAC) [17] and Lee's Control

1   Placental epigenetic clock (CPC) [18]. The clocks were applied using the *agep()* function of the

2   *wateRmelon* package [19], where the default estimates the MTC and other clocks (here the GAC and

3   CPC) can be estimated by providing the necessary coefficient and intercept values. The predictive

4   accuracy of each clock was profiled in each dataset by two measures: Pearson's correlation coefficient

5   with reported chronological age and root mean squared error (RMSE). To investigate potential effects

6   of sex on the predicted epigenetic age a linear model was fitted in the testing dataset with FBC

7   predicted epigenetic age as dependent variable, chronological age and sex as main effects and an

8   interaction of chronological age and sex.

9

10 *Comparison of cellular states.*

11 The FBC was applied to DNAm data for all cellular samples available. To test for differences in

12 predicted epigenetic age between cell stages within each dataset, either two sample t-tests or ANOVA

13 followed by Tukey HSD multiple comparison (when three cell stages were available), were used. To

14 combine results across all datasets, a mixed effects linear model was fitted with predicted epigenetic

15 age as the dependent variable, a fixed effect for cell stage represented as two dummy variables

16 contrasting NPCs vs iPSCs and iPSC-neurons vs iPSCs as and a random effect for dataset.

17

18 **Results**

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

20 We applied EN regression to genome wide DNAm data from a set of prenatal brain samples (n = 193;

21 **Table S1** and **Figure S1**), to develop the fetal brain clock (FBC). A subset of 107 DNAm probes were

22 assigned non-zero coefficients and therefore were selected as the basis of the FBC (**Table S2**). We

23 found no overlap in the DNAm sites selected for the FBC and DNAm sites used in the other clocks we

24 considered. Testing the FBC clock in an independent test dataset of fetal brain samples (**Table S1** and

8

1   **Figure S1**) to evaluate its performance we found a strong linear relationship between chronological

2   and predicted prenatal age (r = 0.80; **Figure 1A**) with the majority of samples predicted within 15 days

3   of their actual chronological age (RMSE = 14.84 dpc). To benchmark the performance of our clock, we

4   compared it to three existing DNAm clocks: Horvath's MTC [13], Knight's GAC [17] and Lee's CPC [18].

5   These clocks were selected as they represent either the most well-established algorithm with the

6   broadest applicability (MTC) or were specifically developed to predict pre- and perinatal gestational

7   ages, albeit in non-brain tissue (GAC and CPC). Of note, the MTC only predicted 27 fetal brain samples

8   (41.2%) as prenatal (dpc < 280) with a very weak correlation between chronological and predicted age

9   (r_{MTC} = 0.06). This correlation is much weaker than those reported in the original manuscript when

10   Horvath tested the clock in adult samples [13], highlighting the challenges with extrapolating clocks

11   to samples which were not well represented in the model development. By comparison, the GAC and

12   CPC perform better than the MTC, although they have smaller correlation coefficients (r_{GAC} = 0.52 and

13   r_{CPC} = 0.76) and are associated with a larger error ($RMSE_{GAC}$ = 21.32 and $RMSE_{CPC}$ = 60.08) than the FBC.

14   Interestingly, while the predictions from the GAC are more precise, it's not as effective at ranking the

15   samples by age as the CPC. Taken together, these results demonstrate that our novel FBC outperforms

16   existing clocks at predicting age in fetal brain samples, and therefore is the optimal tool available to

17   profile the epigenetic age in models of neuronal development. When applying clocks to the training

18   data, the three established clocks produce similar correlations and RMSEs as in the testing data. As

19   expected, the predictions of the FBC in the training data are more accurate than the predictions in the

20   testing data reflecting overfitting of the model (**Figure S2**). Given our previous finding of divergent,

21   sex-specific age trajectories at multiple DNAm sites during prenatal development [23], we tested

22   whether the FBC performed differently between males and females in our testing dataset. Although

23   this analysis initially indicated a significant difference in the correlation with age between males and

24   females (P_{Sex} = 0.0007), on closer inspection we suspected that this was likely driven by outliers.

25   Indeed, a sensitivity analysis excluding the two samples with youngest and oldest predicted ages

26   produced a non-significant result (P_{Sex} = 0.081).

1

2   *Fetal and gestational age clocks are not able to predict adult ages in adult brain tissue*

3   All four clocks were additionally compared using a dataset where DNAm was profiled from adult brain

4   samples (**Figure S3**). As expected, the FBC performs poorly in this sample set, with all samples

5   predicted as prenatal, while the MTC performs the best (r_{MTC} = 0.65, $RMSE_{MTC}$ = 20.11 years) as it is

6   the only clock we considered that was developed using adult samples. As with the FBC, the GAC and

7   CPC fail to produce predictions of post-natal age again reflecting the fact that they were also

8   constructed using data from pre- or perinatal samples.

9

10   *Fetal brain clock captures differences in differentiation of iPSCs towards neurons*

11   Having demonstrated that our novel FBC is the optimal clock to profile age in prenatal brain samples,

12   we applied it to DNAm data from five cellular studies to determine the neuronal development

13   trajectory of iPSCs differentiating towards neurons. All samples were estimated to have a fetal

14   epigenetic age, regardless of cell stage, cell line origin or differentiation protocol. Performing

15   statistical comparisons of the estimated ages between iPSCs and neurons, we observed significant

16   differences for all datasets (**Figure 2**). For the Imm dataset, which also included proliferative NPCs in

17   comparison to postmitotic neurons, we additionally found a significant difference between NPCs and

18   iPSC-neurons (Δ_{mean} = 20.0 dpc, P = 0.00039), but not between iPSCs and NPCs (Δ_{mean} = 10 dpc, P =

19   0.24). Combining the data across the five studies, the observed means of the three cell stages were

20     $mean_{iPSC}$ = 72.14 dpc, $mean_{NPC}$ = 68.5 dpc and $mean_{iPSC-neurons}$ = 84.28 dpc (**Figure 3**). Using a mixed

21   effects model to meta-analyze across studies, we found that iPSC-neurons were predicted to have a

22   significantly advanced epigenetic age compared to iPSCs (Δ_{mean} = 12.14 dpc, P = 6.94E-9) but no

23   significant difference was observed between iPSC and NPCs (Δ_{mean} = -3.64 dpc, P = 0.39).

24

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

11

1   age. However, in these analyses the predicted ages remained postnatal, which is unexpected as

2   human pluripotent cells only occur during the early stages of human development and hence we

3   hypothesized that, with an adequately calibrated clock, iPSCs would be expected to be estimated as

4   early prenatal. Applying the FBC to five datasets of iPSCs and iPSC-derived NPCs and neurons, we

5   found this to be the case, where iPSCs were estimated as having a mean age of 72.1 dpc, fitting our

6   hypothesis that they reflect at least first trimester developmental stages. These results align with

7   studies that have reported rejuvenation effects on the transcriptome, telomeres and mitochondria of

8   iPSCs following reprogramming [26–28]. In addition, we profiled the effect on predicted epigenetic

9   age following the differentiation of iPSCs towards neurons reporting that this increases the mean

10 predicted age to 84.3 dpc. This age coincides with fetal neurogenesis and suggests that while

11 differentiation does induce an aging process, it does not accelerate iPSC-neurons to a postnatal state.

12 Of note, Mertens and colleagues found that, while iPSCs lose age related transcriptomic signatures,

13 induced neurons (iNs; neurons directly reprogrammed from fibroblasts) keep their specific aging

14 signatures [27]. Therefore, it would be interesting to apply our FBC to iNs, iPSCs, iPSC-neurons and

15 their corresponding somatic tissues to verify whether age associated methylation differences are also

16 preserved in iNs. Altogether, our results indicate that iPSC-neurons may be limited when researching

17 age-related diseases, like Alzheimer's disease or other dementias, as many molecular processes

18 related to an aging phenotype, may not be recapitulated.

19

20 While a strength of our study is the development of a bespoke clock to optimally profile prenatal age

21 in human samples obtained during fetal development, the FBC has in all datasets tested, only

22 generated prenatal estimates. Despite this, we are confident that the FBC correctly predicts fetal

23 epigenetic ages for the cellular data as the predictions for both iPSC and iPSC-neurons are comfortably

24 contained within the training data age range. This indicates that the predictions are not induced purely

25 by saturation of the coefficients. In contrast, estimates for the adult brain dataset did lie at the

1 extreme ends, bringing doubt on the meaningfulness of the predictions, an observation that is

2 confirmed by the lack of correlation with chronological age.

3

4 In summary, we demonstrate that established DNAm clocks struggle to capture changes in epigenetic

5 age during neurodevelopment and for precise predictions a bespoke clock is required. To this end, we

6 developed the FBC, a robust predictor of prenatal age in human fetal brain samples. Using this clock

7 to assess the epigenetic age of iPSCs and differentiated neurons, we found that iPSCs and derived

8 NPCs and neurons reflect prenatal developmental stages. Our findings question the suitability of the

9 iPSC-neurons for the study of aging associated processes.

10

11 **List of abbreviations**

Abbreviation	Definition
AD	Alzheimer's Disease
BDR	Brains for Dementia Research
CPC	Control Placental Clock
DNAm	DNA methylation
dpc	Days post-conception
EN	Elastic net
FBC	Fetal brain clock
GA	Gestational age
GAC	Gestation age clock
GEO	Gene Expression Omnibus
iN	Induced neuron
iPSC	Induced pluripotent stem cell
iPSC-neuron	iPSC-derived neuron
LASSO	Least Absolute Shrinkage and Selection Operator
MTC	Multi-tissue clock

13

NPC	Neuronal precursor cell
QC	Quality control
RMSE	Root mean squared error

1

2 **Declarations**

3 *Ethical approval*

4 Ethical approval for the collection of human fetal brain tissue acquired from the Human

5 Developmental Biology Resource was granted by the Royal Free Hospital research ethics committee

6 under reference 08/H0712/34 and Human Tissue Authority material storage license 12220. Ethical

7 approval for collection of human fetal brain tissue for the Medical Research Council Brain Bank was

8 granted under reference 08/MRE09/38. Ethical approval for the work with methylomic data of human

9 brain tissue from the BDR cohort was granted by the University of Exeter, College of Medicine and

10 Health Research Ethics Committee under reference Mar20/D/009Δ5.

11

12 *Consent for publication*

13 Not applicable

14

15 *Availability of data and materials*

16 Raw and normalized DNAm data for the human fetal brain samples used to test and train the fetal

17 brain clock have been submitted to the NCBI Gene Expression Omnibus (GEO;

18 http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE58885, GSE74193 and GSE157908.

19 DNAm data for the cellular dataset are also available under accession numbers GSE158089,

20 GSE105093 and GSE51921. DNAm data for the cellular dataset *Bhinge* is available from the

21 corresponding author on request. The coefficients of the FBC are included within the article in **Table**

14

1 **S2**. The custom R script used for this study is available at

2 https://github.com/LSteg/EpigeneticFetalClock.

3

4 *Competing interests*

5 The authors declare that they have no competing interests.

6

7 *Funding*

19

20 *Authors' contributions*

21 L.C.S, G.L.S. and E.H. designed the work; J.M., E.H., N.J.B, A.B., J.P., N.D.A. and K.L. acquired funding;

22 J.I., J.P.D., R.F., S.C.N., A.R.J., J.B., G.W.A.N., E.M.W., L.W.P., E.L.C., K.M.J., E.P., A.B., N.J.B. and N.D.A

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

2   and E.H. wrote the paper.

3

7

8   **Figure Legends**

9   **Figure 1 Fetal brain clock outperforms other DNAm clocks when applied to neurodevelopmental**

10   **samples.** Shown are scatterplots comparing chronological age (x-axis; days post-conception) against

11   (y-axis; days post-conception) predicted epigenetic age calculated using **A** Fetal Brain Clock (FBC) **B**

12   Horvath's multi tissue clock (MTC); **C** Knight's Gestational Age Clock (GAC); **D** Lee's Control Placental

13   Clock (CPC); in an independent fetal brain sample (n = 65). The black line indicates the identity line of

14   chronological and predicted epigenetic age and represents a perfect prediction. Two statistics were

15   calculated to evaluate the precision of each DNAm clock: Pearson's correlation coefficient (r) and the

16   root mean squared error (RMSE).

17

18   **Figure 2 Comparisons of predicted epigenetic age from fetal brain clock between iPSC**

19   **differentiation states.** Boxplots comparing the distribution of predicted epigenetic age (days post-

20   conception) separated by cellular stage, where each panel represents a different dataset. P values of

21   Tukey HSD corrected ANOVA for the Imm dataset and two-sample t-tests for Price, Fernández-

22   Santiago and Sultanov datasets are given. F. -S. = Fernández-Santiago.

1

2 **Figure 3 iPSCs are significantly younger than iPSC induced neurons using age estimated by the FBC**.

3 Boxplots of predicted epigenetic age calculated using the FBC where samples are grouped by cell stage

4 (n = 82, 30 iPSCs, 4 NPCs, 48 iPSC-neurons) and colored by dataset. P values from mixed effects model

5 are given for differences between iPSCs and NPCs (non-significant) and iPSC and neurons. F.-S. =

6 Fernández-Santiago.

7

8 **Additional Files**

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

10

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

12

13 **Additional File 3: Figure S1 Histogram of age distribution.** Chronological age of the fetal samples

14 measured in days post-conception in A) training data and B) testing data.

15

16 **Additional File 4: Figure S2 Comparison of predictions from the four DNAm clocks in the training**

17 **data (n = 193).** Shown are scatterplots comparing chronological age (x-axis) against age (y-axis)

18 predicted epigenetic age for **A** Fetal Brain Clock (FBC); **B** Horvath's multi tissue clock (MTC); **C** Knight's

19 Gestational Age Clock (GAC); **D** Lee's Control Placental Clock (CPC) in the data used for training of the

20 FBC. Where necessary, predicted age was converted to days post-conception. The black line indicates

21 the identity line of chronological and predicted epigenetic age and represents a perfect prediction.

22 Two statistics were calculated to evaluate the precision of each DNAm clock: Pearson's correlation

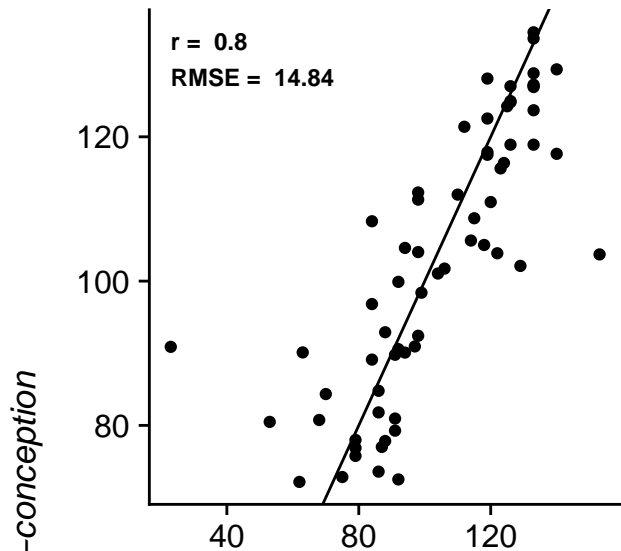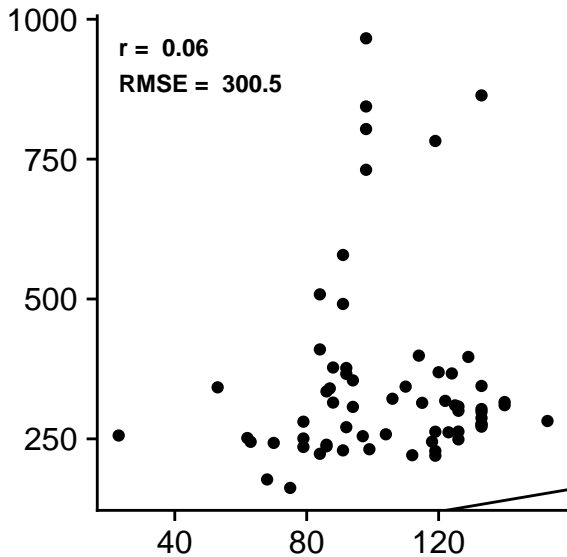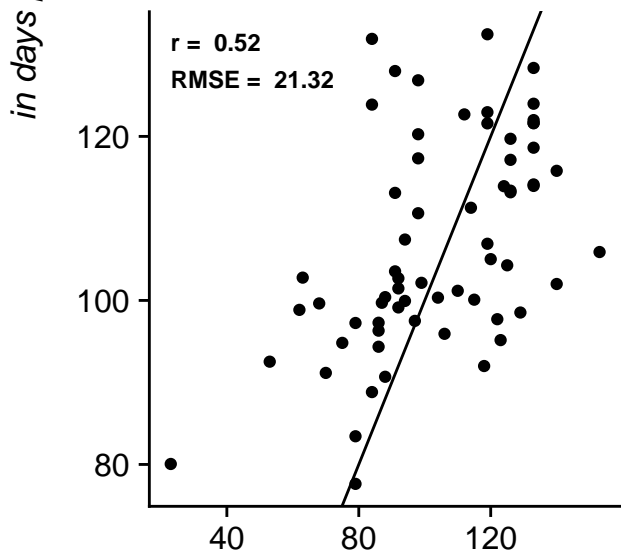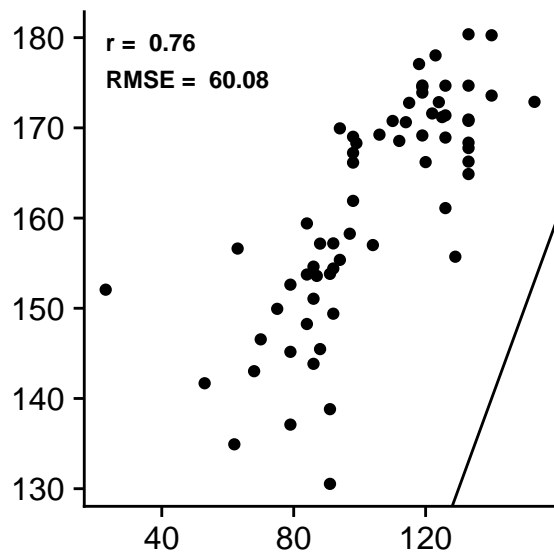23 coefficient (r) and the root mean squared error (RMSE).

24

25 **Additional File 5: Figure S3 Comparison of predictions from the four DNAm clocks in adult brain**

26 **samples (n = 1221).** Shown are scatterplots comparing chronological age (x-axis) against (y-axis)

1 predicted epigenetic age calculated using **A** Fetal Brain Clock (FBC); **B** Horvath's multi tissue clock

2 (MTC); **C** Knight's Gestational Age Clock (GAC); **D** Lee's Control Placental Clock (CPC) in an independent

3 adult brain dataset. Where necessary, predicted age was converted to years, where 0 indicates birth.

4 The black line indicates the identity line of chronological and predicted epigenetic age and represents

5 a perfect prediction. Two statistics were calculated to evaluate the precision of each DNAm clock:

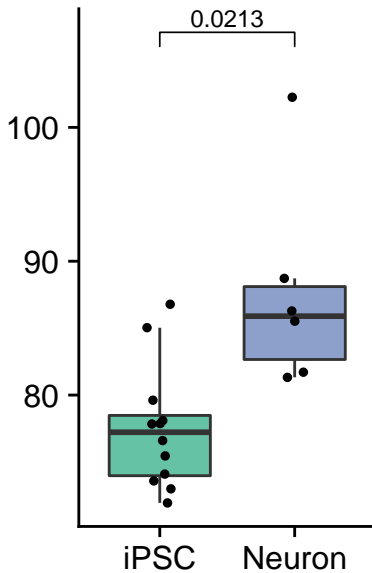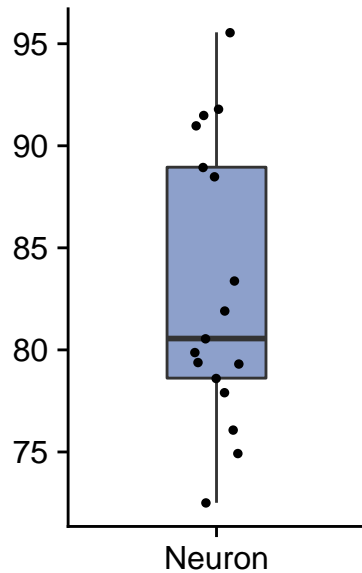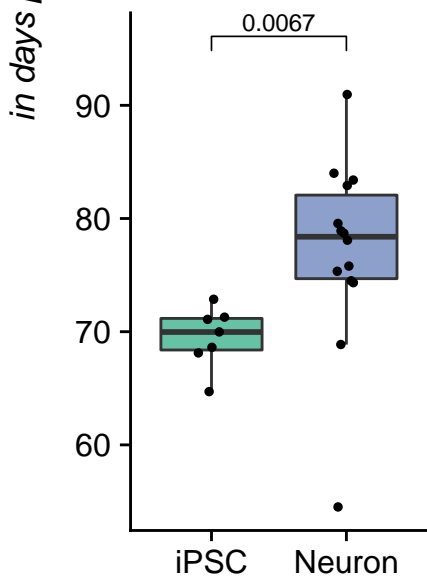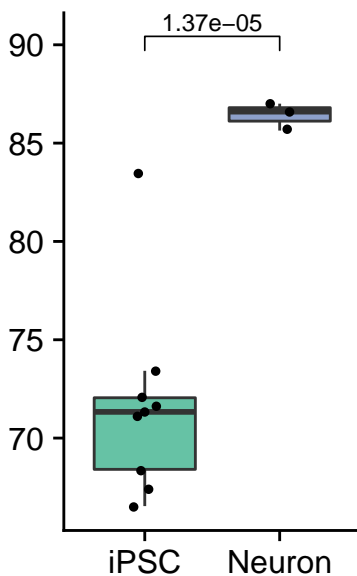6 Pearson's correlation coefficient (r) and the root mean squared error (RMSE).

7

References

9   1.   Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction
10   of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131:861–872

11 2.   Telezhkin V, Schnell C, Yarova P, et al (2016) Forced cell cycle exit and modulation of GABA A ,
12   CREB, and GSK3β signaling promote functional maturation of induced pluripotent stem cell-
13   derived neurons. Am J Physiol Physiol 310:C520–C541

14 3.   Fernández-Santiago R, Carballo-Carbajal I, Castellano G, et al (2015) Aberrant epigenome in
15   iPSC -derived dopaminergic neurons from Parkinson's disease patients. EMBO Mol Med
16   7:1529–1546

17 4.   Sultanov R, Lebedeva O, Arapidi G, Lagarkova M, Kiselev S (2018) Methylation profile of
18   induced pluripotent stem cells generated by integration and integration-free approaches. Data
19   Br 17:662–666

20 5.   Gordon J, Amini S, White MK (2013) General overview of neuronal cell culture. Methods Mol
21   Biol 1078:1–8

22 6.   Arber C, Lovejoy C, Wray S (2017) Stem cell models of Alzheimer's disease: Progress and
23   challenges. Alzheimer's Res Ther. https://doi.org/10.1186/s13195-017-0268-4

24 7.   Morgani S, Nichols J, Hadjantonakis AK (2017) The many faces of Pluripotency: In vitro
25   adaptations of a continuum of in vivo states. BMC Dev Biol 17:1–20

26 8.   Monk D, Mackay DJG, Eggermann T, Maher ER, Riccio A (2019) Genomic imprinting disorders:
27   lessons on how genome, epigenome and environment interact. Nat Rev Genet 20:235–248

28 9.   Zeng Y, Chen T (2019) DNA methylation reprogramming during mammalian development.
29   Genes (Basel). https://doi.org/10.3390/genes10040257

30 10. Smith ZD, Meissner A (2013) DNA methylation: Roles in mammalian development. Nat Rev
31   Genet 14:204–220

32 11. Levine ME, Lu AT, Quach A, et al (2018) An epigenetic biomarker of aging for lifespan and
33   healthspan. Aging (Albany NY) 10:573–591

34 12. Hillary RF, Stevenson AJ, McCartney DL, et al (2020) Epigenetic measures of ageing predict the
35   prevalence and incidence of leading causes of death and disease burden. Clin Epigenetics
36   12:115

13. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol. https://doi.org/10.1186/gb-2013-14-10-r115

14. Shireby G, Davies J, Francis P, et al (2020) Recalibrating the Epigenetic Clock: Implications for Assessing Biological Age in the Human Cortex. bioRxiv 2020.04.27.063719

15. El Khoury LY, Gorrie-Stone T, Smart M, et al (2019) Systematic underestimation of the epigenetic clock and age acceleration in older subjects. Genome Biol 20:283

16. Zhang Q, Vallerga CL, Walker RM, et al (2019) Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med 11:54

17. Knight AK, Craig JM, Theda C, et al (2016) An epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol 17:1–11

18. Lee Y, Choufani S, Weksberg R, et al (2019) Placental epigenetic clocks: Estimating gestational age using placental DNA methylation levels. Aging (Albany NY) 11:4238–4253

19. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. https://doi.org/10.1186/1471-2164-14-293

20. Gorrie-Stone TJ, Smart MC, Saffari A, Malki K, Hannon E, Burrage J, Mill J, Kumari M, Schalkwyk LC (2019) Bigmelon: Tools for analysing large DNA methylation datasets. Bioinformatics 35:981–986

21. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8:203–209

22. Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, Kleinman JE (2015) Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. Nat Neurosci 19:40–47

23. Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CCY, O'Donovan MC, Bray NJ, Mill J (2015) Methylomic trajectories across human fetal brain development. Genome Res 25:338–352

24. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol 67:301–320

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

26. Suhr ST, Chang EA, Tjong J, Alcasid N, Perkins GA, Goissis MD, Ellisman MH, Perez GI, Cibelli JB (2010) Mitochondrial Rejuvenation After Induced Pluripotency. PLoS One. https://doi.org/10.1371/journal.pone.0014095

27. Mertens J, Paquola ACM, Ku M, et al (2015) Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. Cell Stem Cell 17:705–718

28. Lapasset L, Milhavet O, Prieur A, et al (2011) Rejuvenating senescent and centenarian human cells by reprogramming through the pluripotent state. Genes Dev 25:2248–2253

29. Mattis VB, Tom C, Akimov S, et al (2014) HD iPSC-derived neural progenitors accumulate in culture and are susceptible to BDNF withdrawal due to glutamate toxicity. Hum Mol Genet 24:3257–3271

30. Cocks G, Curran S, Gami P, et al (2014) The utility of patient specific induced pluripotent stem cells for the modelling of Autistic Spectrum Disorders. Psychopharmacology (Berl) 231:1079–1088

31. Warre-Cornish K, Perfect L, Nagy R, et al (2020) Interferon-γ signaling in human iPSC–derived neurons recapitulates neurodevelopmental disorder phenotypes. Sci Adv 6:eaay9506

32. Bhinge A, Namboori SC, Zhang X, VanDongen AMJ, Stanton LW (2017) Genetic Correction of SOD1 Mutant iPSCs Reveals ERK and JNK Activated AP1 as a Driver of Neurodegeneration in Amyotrophic Lateral Sclerosis. Stem Cell Reports 8:856–869

A FBC

r = 0.8
RMSE = 14.84

B MTC

r = 0.06
RMSE = 300.5

C GAC

r = 0.52
RMSE = 21.32

D CPC

r = 0.76
RMSE = 60.08

Predicted Age
in days post−conception

Chronological Age
in days post−conception