1   **Title**

2   Novel epigenetic clock for fetal brain development predicts prenatal age for cellular stem cell models

3   and derived neurons.

4   **Authors**

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

6   Flynn[1], Seema C. Namboori[1], Akshay Bhinge[1], Aaron R. Jeffries[1], Joe Burrage[1], Grant W. A. Neilson[1],

7   Emma M. Walker[1], Leo W. Perfect[2], Jack Price[2], Grainne McAlonan[3,4], Deepak P. Srivastava[2,4],

8   Nicholas J. Bray[5], Emma L. Cope[6], Kimberly M. Jones[6], Nicholas D. Allen[6], Ehsan Pishva[1,7], Emma L.

9   Dempster[1], Katie Lunnon[1], Jonathan Mill[1], & Eilis Hannon[1,*]

10   [1] University of Exeter Medical School, College of Medicine and Health, University of Exeter, Exeter,

11   UK.

12   [2] Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and

13   Neuroscience, King's College London, UK.

14   [3] Department of Forensic and Neurodevelopmental Sciences, King's College London, London, UK.

15   [4] MRC Centre for Neurodevelopmental Disorders, King's College London, London, UK.

16   [5] Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff

17   University, Cardiff, UK.

18   [6] School of Biosciences, Cardiff University, Cardiff, UK.

19   [7] Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience

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

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

22   Royal Devon and Exeter Hospital, Barrack Rd, Exeter. EX2 5DW. UK. E-mail: e.j.hannon@exeter.ac.uk

1   **Keywords**

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

3   derived neurons, neuronal precursor cells, DNAm clock

4   **Abstract**

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

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

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

8   during the earliest stages of mammalian development. Epigenetic clocks utilize coordinated age-

9   associated changes in DNA methylation to make predictions that correlate strongly with

10   chronological age. It has been shown that the induction of pluripotency rejuvenates predicted

11   epigenetic age. As existing clocks are not optimized for the study of brain development, we

12   developed the fetal brain clock (FBC), a bespoke epigenetic clock trained in human prenatal brain

13   samples in order to investigate more precisely the epigenetic age of iPSCs and iPSC-neurons. The FBC

14   was tested in two independent validation cohorts across a total of 194 samples, confirming that the

15   FBC outperforms other established epigenetic clocks in fetal brain cohorts. We applied the FBC to

16   DNA methylation data from iPSCs and iPSC-derived neuronal precursor cells and neurons, finding

17   that these cell types are epigenetically characterized as having an early fetal age. Furthermore, while

18   differentiation from iPSCs to neurons significantly increases epigenetic age, iPSC-neurons are still

19   predicted as being fetal. Together our findings reiterate the need to better understand the

20   limitations of existing epigenetic clocks for answering biological research questions and highlight a

21   limitation of iPSC-neurons as a cellular model of age-related diseases.

22   **MAIN TEXT**

23   **Introduction**

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

25   derived cells. iPSCs are obtained by treating skin or blood cells with a set of core pluripotency

1   transcription factors that reprogram the cells to a pluripotent state [1]. Established iPSC lines have

2   the capacity to be further differentiated into specific cell types, including neurons, when treated

3   with the appropriate factors [2–4]. This is of particular interest for neuroscience, as the only

4   alternative cellular model for human neurons are immortalized cell lines. Because immortalized cell

5   lines retain some physiological properties of the cancerous cells they were derived from [5] they do

6   not fully recapitulate the neuronal phenotype. iPSC-derived neurons (iPSC-neurons), on the other

7   hand, express appropriate morphological and neurophysiological properties of neurons and subject

8   to different protocols can be differentiated into a wide range of specific neuronal subtypes [6]. iPSCs

9   and their neuronal derivatives have been widely used to research disorders of the central nervous

10   system, including developmental disorders such as autism and schizophrenia and age-related

11   diseases such as Alzheimer's disease (AD) and Parkinson's disease. However, the extent to which

12   iPSCs and especially iPSC-neurons capture age-associated processes is not known, which is

13   fundamental to the study of age-related diseases. Of specific relevance is the fact that pluripotent

14   cells only occur during the early stages of mammalian development and the effect of differentiation

15   from iPSCs towards neurons on the developmental or aging trajectory of the cellular model [7] has

16   yet to be adequately profiled.

17

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

19   regulate gene expression, and while they are sensitive to environmental stimuli they also underpin

20   key developmental processes [8–10]. There has been much interest and success in capitalizing on

21   these patterns of epigenetic variation to derive individual age predictions from a biological sample.

22   Age predictors based on DNAm, known as "epigenetic clocks" or "DNAm clocks", are widely used to

23   predict the "epigenetic age" of a sample. Epigenetic age, defined here as age predicted by an

24   epigenetic clock, correlates strongly with chronological age, albeit not perfectly, and it has been

25   hypothesized that the deviations from this prediction, referred to as age acceleration, are

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

2   multi-tissue clock (MTC) which was developed using a large number of samples (n > 8000) from 51

3   different tissues and cell types [13]. Overall, the MTC generates reliable predictions of chronological

4   age for most sample-types, although there are potential biases when using Horvath's clock in

5   samples derived from certain tissues, especially the brain [14, 15]. To this end, a number of new

6   DNAm clocks have been developed for specific tissue types, including whole blood [16] and cortex

7   [14], which demonstrate more accurate predictions within the specified tissue. A less established

8   refinement of epigenetics clocks is the application to specific developmental stages, with prenatal

9   samples excluded or underrepresented in most training datasets. Recently, clocks were developed to

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

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

12   accurately predict age in either postnatal brain samples (predominantly middle and older age) or

13   non-brain prenatal samples, these tools have not been thoroughly tested on prenatal brain samples,

14   and it is unknown whether they are able to delineate the earliest stages of brain development.

15

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

17   primary cells from adult donors found that the induction of pluripotency reversed the aging process,

18   with iPSCs predicted as postnatal, with an epigenetic age close to zero (i.e. birth) [13]. As human

19   pluripotent cells only occur during prenatal development, we hypothesize that existing clocks are

20   not sensitive enough to accurately predict iPSCs at prenatal developmental stages. The inability to

21   accurately estimate age during this crucial stage of neurodevelopment limits our ability to profile

22   changes in epigenetic age induced by the differentiation of iPSCs into specific cell-types using

23   already established DNAm clocks. Here we present a novel DNAm clock developed using prenatal

24   brain samples that accurately predicts fetal age, outperforming other DNAm clocks in

25   neurodevelopmental samples. We then apply our clock to iPSCs, iPSC-derived neuronal progenitor

4

1   cells (NPCs) and iPSC-derived neurons, as well as in other cellular stem cell models and derived

2   neuronal cells, to characterize the epigenetic age of these cellular models before and during the

3   differentiation process.

4

5   **Methods**

6   All statistical analyses were performed using R version 3.5.2 (https://www.r-project.org/) [19].

7   ***Development of Fetal Brain Clock (FBC)***

8   *Description of fetal brain samples*

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

10   fetal brain samples (see **Table S1**) of which 194 were processed by our group at the University of

11   Exeter as described previously [20] and 64 were a subset (age < 0 years) of a publicly available

12   dataset downloaded from the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/;

13   GSE74193) [21]. Of the samples processed in Exeter, 154 overlap those included in [20] following

14   additional outlier filtering by principal component analysis, where DNAm was quantified using the

15   Illumina 450K DNA methylation array. The other 40 samples represent additional samples where

16   DNAm was quantified using the Illumina EPIC DNA methylation array using a standard protocol as

17   previously described [14].

18

19   *Data pre-processing and quality control.*

20   All datasets for which raw data was available were pre-processed following a standard quality

21   control (QC) and normalization pipeline as described before [14] using either the R package

22   *wateRmelon* [22] or *bigmelon* [23]. Briefly, samples with low signal intensities or incomplete bisulfite

23   conversion were excluded prior to applying the *pfilter()* function from the *wateRmelon* package,

5

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

2   samples with detection P value >0.05. This was followed by the exclusion of probes known to be

3   affected by SNPs or known to cross-hybridize [24]. QC was finished by quantile normalization using

4   the *dasen()* function of the packages *wateRmelon* or *bigmelon* [22, 23]. To harmonize the age

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

6   most precise unit of age available across the datasets. Where age was provided as weeks post-

7   conception it was transformed to days post-conception by dividing by 7, and where age was

8   reported in (negative) years it was transformed to days post-conception by multiplying by 365 and

9   adding 280. Of note, a few samples (15 out of 258) are actually defined as embryonic (GA < 63 dpc)

10   and not fetal.

11

12   *Fetal brain clock development.*

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

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

15   age median = 99 dpc), while the remaining 25% included in the testing dataset (n = 65, age range =

16   23-153 dpc, age median = 99 dpc) (**Figure S1, Table S1**). There was no overlap in samples between

17   the training and testing dataset. To simplify the FBC development, only probes available in all

18   samples after QC were taken forward (n = 385,069 probes). To develop the fetal brain clock we

19   applied an elastic net (EN) regression model, using the approach described by Horvath [13],

20   regressing chronological age against DNAm level of all available probes. The EN algorithm selects a

21   subset of DNA methylation probes that together produce the optimal prediction of the outcome, in

22   this case chronological age, by combining ridge and LASSO (Least Absolute Shrinkage and Selection

23   Operator) regression. Briefly, ridge regression penalizes the sum of squared coefficients while LASSO

24   penalizes the sum of the absolute values of coefficients. EN is a combination of both methods,

25   where the user specifies the extent of the mixing of the two methods as a number between 0 and 1,

1   in our application this was set to 0.5 [25]. EN was implemented with the R package *GLMnet* [26].

2   The shrinkage parameter lambda was calculated using 10-fold cross-validation on the data, which

3   resulted in a lambda of 3.27.

4

5   *Statistical evaluation of FBC performance*

6   To profile the performance of the FBC, we additionally tested three established DNAm clocks:

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

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

9   *wateRmelon* package [22], where the default estimates the MTC and other clocks (here the GAC and

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

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

12 coefficient with reported chronological age and root mean squared error (RMSE). To investigate

13 potential effects of sex on the predicted epigenetic age, linear models were fitted in the testing and

14 validation datasets with FBC predicted epigenetic age as dependent variable, chronological age and

15 sex as main effects and an interaction of chronological age and sex.

16

17 **Validation of the Fetal Brain Clock (FBC) in additional fetal datasets and adult cortex**

18 To further test the FBC, we used data from 96 additional fetal brain samples currently being

19 assessed by our group (unpublished data), none of which overlapped with either the training data or

20 test data described in the previous section, with DNAm quantified using the Illumina EPIC DNA

21 methylation array. QC and normalization were performed as described above. We also included

22 data from 33 fetal samples from two publicly available datasets on GEO (GSE116754 and GSE90871)

23 [27, 28], where DNAm was quantified using the Illumina 450K DNA methylation array. Pre-processing

24 and QC for the publicly available datasets was not performed in our lab as no raw data was available.

7

1   Age of all samples was converted to dpc as described above. The combined validation dataset has an

2   age range of 42 – 280 dpc with the median at 112 dpc (**Figure S1, Table S1**). To evaluate the

3   performance of the DNAm clocks in adult brain samples we utilized data from the Brains for

4   Dementia Research (BDR) cohort previously generated by our group [14]. Briefly, these data consist

5   of 1,221 samples from 632 donors (age range 41-104 years, median = 84 years), with DNA extracted

6   from the prefrontal cortex (n = 610) and occipital cortex (n = 611). DNAm was quantified using the

7   Illumina EPIC DNAm array, and were pre-processed using a standard QC pipeline as described in [14].

8

9   *Statistical evaluation of FBC performance*

10   The predictive accuracy of the FBC was profiled in each dataset by two measures: Pearson's

11   correlation coefficient with reported chronological age and root mean squared error (RMSE).

12

13   **Testing of Fetal Brain Clock (FBC) in cellular samples**

14

15   *iPSC - neuron samples.*

16   Five different DNAm datasets generated using iPSCs, iPSC-derived NPCs and iPSC-derived neurons

17   were used to characterize epigenetic age of the neuronal cell model, details of which can be found in

18   **Table S1**. For two of these datasets (*Imm, Price*) DNAm data was generated by our lab in Exeter,

19   where DNAm was quantified using the Illumina EPIC DNA methylation array. These were

20   supplemented by three publicly available datasets, downloaded from GEO (*Nazor*, GSE31848,

21   *Sultanov*, GSE105093, and *Fernández-Santiago*, GSE51921) consisting of Illumina 450K DNAm array

22   data [3, 4, 29]. References describing the origin of cell lines and the different methods used for cell

23   culture and differentiation are listed in **Table S1**. Pre-processing and QC for the *Nazor* and

24   *Fernández-Santiago* datasets was not performed in our lab as no raw data was available.

8

1

2 *iPSC – motor neuron samples*

3 A dataset comprised of 23 cellular samples with two iPSC samples and 21 derived motor neurons

4 was generated by our lab, with DNAm quantified using the Illumina EPIC DNAm array (**Table S1**).

5 These data were QC'd following the pipeline described above.

6

7 *ESC – neuron samples*

8 Two publicly available cellular datasets with NPCs and neurons derived from embryonic stem cells

9 (ESCs) were downloaded from GEO (*Nazor*, GSE31848, *Kim* GSE38214) [29, 30]. Both datasets consist

10 of data quantified using the Illumina 450K DNAm array. Pre-processing and QC for both ESC – neuron

11 datasets were not performed in our lab as no raw data was available.

12

13 *Statistical comparison of cellular states.*

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

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

16 ANOVA followed by Tukey HSD multiple comparison (when three cell stages were available), were

17 used. To combine results across all iPSC - neuron datasets, a mixed effects linear model was fitted

18 with predicted epigenetic age as the dependent variable, a fixed effect for cell stage represented as

19 two dummy variables contrasting NPCs vs iPSCs and iPSC-neurons vs iPSCs as and a random effect

20 (i.e. random intercept) for dataset.

21

22 **Results**

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

1   We applied EN regression to genome wide DNAm data from a subset of available prenatal brain

2   samples (n = 193; **Table S1** and **Figure S1**) to develop the fetal brain clock (FBC). 107 DNAm probes

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

4   We found no overlap in the DNAm sites selected for the FBC and DNAm sites used in the other

5   established clocks tested in our analysis. Testing the FBC clock in an independent test dataset of fetal

6   brain samples (**Table S1** and **Figure S1**) to evaluate its performance we found a strong linear

7   relationship between chronological and predicted prenatal age (r = 0.80; **Figure 1A**) with the

8   majority of samples predicted within 15 days of their actual chronological age (RMSE = 14.84 dpc).

9   To benchmark the performance of our clock, we compared it to three existing DNAm clocks:

10   Horvath's MTC [13], Knight's GAC [17] and Lee's CPC [18]. These clocks were selected as they

11   represent either the most well-established algorithm with the broadest applicability (MTC) or were

12   specifically developed to predict pre- and perinatal gestational ages, albeit in non-brain tissue (GAC

13   and CPC). Of note, the MTC only predicted 27 fetal brain samples (41.2%) as prenatal (dpc < 280)

14   with a low correlation between chronological and predicted age (r_{MTC} = 0.06). This correlation is

15   much lower than those reported in the original manuscript when Horvath tested the clock in adult

16   samples [13], highlighting the challenges with extrapolating clocks to samples which were not well

17   represented in model development. By comparison, the GAC and CPC perform better than the MTC,

18   although they have smaller correlation coefficients (r_{GAC} = 0.52 and r_{CPC} = 0.76) and are associated

19   with a larger error ($RMSE_{GAC}$ = 21.32 and $RMSE_{CPC}$ = 60.08) than the FBC. Interestingly, while the

20   predictions from the GAC are more precise, it is not as effective at ranking the samples by age as the

21   CPC. Taken together, these results demonstrate that our novel FBC outperforms existing clocks at

22   predicting age in fetal brain samples, and therefore is the optimal tool available to profile the

23   epigenetic age in models of neuronal development. When applying clocks to the training data, the

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

25   the predictions of the FBC in the training data are more accurate than the predictions in the testing

26   data, reflecting overfitting of the model (**Figure S2**).

1

2   We further tested the FBC in an independent prenatal brain dataset (n = 129, **Table S1**), finding a

3   stronger linear relationship between chronological age and predicted epigenetic age (r = 0.87, **Figure**

4   **2A**), than in the test dataset (**Figure 1A**) but a larger error rate (RMSE = 26.36 dpc). On closer

5   inspection, we observed that this error is mainly driven by a subset of older samples (> 185dpc,

6   **Figure 2A**) in the validation dataset (**Table S1**), that are older than any of the samples in the training

7   data (**Figure 2**). If we limit our analysis to the samples whose chronological age overlaps the range of

8   ages used in the training data (37 - 185 dpc; n = 125) then the error is decreased to 18.94 dpc. The

9   performance of the established clocks in the validation dataset is also comparable to their

10   performance in the testing dataset, with smaller correlations and higher error compared to the FBC

11   and no sample with a predicted prenatal age by the MTC (**Figure 2B-D**).

12

13   Given our previous finding of divergent, sex-specific age trajectories at multiple DNAm sites during

14   prenatal development [20], we tested whether the FBC performed differently between males and

15   females in our testing dataset. Although this analysis initially indicated a significant difference in the

16   correlation with age between males and females ($P_{Sex*Age}$ = 0.0007), this relationship is likely driven

17   by outliers. Indeed, a sensitivity analysis excluding the two samples with youngest and oldest

18   predicted ages produced a non-significant result ($P_{Sex*Age}$ = 0.081). Repeating the analysis in the

19   validation dataset we find a small significant effect of sex on age in the full dataset ($P_{Sex*Age}$ =

20   0.00179), which was driven by the samples older than 185, all of them being female. As samples

21   older than 185 dpc produce inaccurate predictions, they could unfairly bias the analysis of potential

22   sex effects and after removing the them from the analysis, there is no longer a significant effect of

23   sex on the accuracy of the clock ($P_{Sex*Age}$ = 0.95).

24

11

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

2   All four clocks were additionally tested in an adult brain DNAm dataset (**Figure S3**). As expected, the

3   FBC performs poorly in this sample set, with all samples predicted as prenatal albeit at the older end

4   of the spectrum of ages in the training data (range of predicted ages 115-170 dpc). In contrast the

5   MTC performs the best (r_{MTC} = 0.65, $RMSE_{MTC}$ = 20.11 years) as it is the only clock we considered that

6   was developed using adult samples. As with the FBC, the GAC and CPC fail to produce predictions of

7   postnatal age, again reflecting the fact that they were also constructed using data from pre- or

8   perinatal samples.

9

10   *Fetal brain clock captures differences in differentiation of cellular stem cell models 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 multiple cellular studies to characterize epigenetic age in iPSCs and

13   ESCs differentiating towards cortical neurons. All samples were estimated to have a fetal epigenetic

14   age, regardless of cell stage, cell line origin or differentiation protocol. Furthermore, they were

15   predicted to have a "young" fetal age (**Figure 3**) with the iPSCs having a mean predicted age of 75.6

16   dpc (SD = 6.9 dpc; n = 59), the NPCs having a mean age of 79.1 dpc (SD = 11.0 dpc; n = 8) and iPSC -

17   neurons having a mean predicted age of 83.2 dpc (SD = 8.87 dpc; n = 31). To test whether the

18   differentiation process had an effect on the epigenetic age predictions from the FBC we compared

19   the estimated ages between iPSCs and neurons, observing significant differences in all datasets

20   (**Figure 3A**), with neurons being older than iPSCs. For the *Imm* dataset, which included proliferative

21   NPCs as well as postmitotic neurons, we additionally found a significant difference between NPCs

22   and iPSC-neurons (Δ_{mean} = 20.0 dpc, P = 0.00039), but not between iPSCs and NPCs (Δ_{mean} = 10.0 dpc,

23   P = 0.24). In contrast, in the *Nazor* dataset, which only included iPSCs and NPCs, we did find a

24   significant difference in predicted epigenetic age between iPSCs and NPCs (Δ_{mean} = 10.5 dpc, P =

25   0.00734). Meta-analyzing the data across the five studies including iPSCs and iPSC-derived NPCs and

12

1 neurons, we found that iPSC-neurons were predicted to have a significantly advanced epigenetic age

2 compared to iPSCs of about 2 weeks (Δ_{mean} = 13.35 dpc, P = 2.13e-11) but no significant difference

3 was observed between iPSC and NPCs (Δ_{mean} = 4.33 dpc, P = 0.11).

4

5 As well as cortical neurons, we applied the FBC to DNAm data from a study on iPSCs and

6 differentiated motor neurons. The motor neurons had a mean predicted epigenetic age of 79.82 dpc

7 (SD = 6.82 dpc) which is slightly younger than the cortical neurons (mean$_{iPSC-neurons}$ = 83.2 dpc). When

8 comparing the predicted ages between the motor neurons and the iPSCs they were originally

9 derived from, we did not observe a significant difference (Δ_{mean} = 2.94 dpc, P = 0.825, **Figure 3C**).

10

11 Finally, we tested for effects on epigenetic age through the process of differentiation from

12 embryonic stem cells (ESCs) to NPCs and neurons, using two publicly available DNAm datasets (**Table**

13 **S1**). In both datasets we observed a significant increase of predicted epigenetic age from ESC to NPC

14 (**Figure 3D**). Additionally, in the *Kim* dataset, we were able to compare ESC derived neurons, and

15 found a significant difference between ESCs and neurons (Δ_{mean} = 55.42 dpc, P = 0.00015, **Figure 3D**)

16 but no change in epigenetic age from NPC to neuron (Δ_{mean} = -0.58 dpc, P = 0.95).

17 **Discussion**

18 In this study we established a novel epigenetic clock, the fetal brain clock (FBC), to characterize the

19 earliest stages of human neurodevelopment, and applied it to determine the epigenetic age of iPSCs

20 and ESCs and their derived NPCs and neurons. Epigenetic clocks have been widely applied, including

21 for the analysis of both in-vivo and in-vitro models of aging, where they have been shown to

22 correlate with hallmarks of the aging process [31–34]. However, their application has predominantly

23 been focused on studies involving adult samples. Given the lack of fetal brain samples in the

24 development of existing DNAm clocks, prior to this study, there was no optimal method for

1   estimating the age of fetal brain samples from DNAm data, limiting the ability to characterize iPSC-

2   derived neuronal models or other models of neuronal development. We show that, in two non-

3   overlapping independent validation datasets, the FBC generates predictions that correlate strongly

4   with chronological age in prenatal brain samples. Furthermore, it outperforms both a pan-tissue

5   epigenetic clock (Horvath's MTC), and epigenetic clocks focused on the same developmental stage,

6   but based on DNAm profiled in different tissues (Knight's GAC and Lee's CPC) [13, 17, 18]. The FBC

7   outperforms these clocks using both correlation and error statistics (RMSE), indicating that it is not

8   only better at ranking the samples, but it generates more precise estimates.

9

10   As the FBC was predominantly trained on second trimester brain samples, with some first trimester

11   samples, it made less accurate predictions when applied to third trimester samples and performed

12   extremely poorly in adult brain samples. Altogether, this reinforces the findings of previous studies

13   that have also demonstrated that the applicability of DNAm clocks is dependent on the

14   characteristics of the data are were trained on, with the tissue and age range of the training samples

15   of particular relevance [14, 15]. More specifically, we note that while the accuracy of a DNAm clock

16   is typically decreased in tissues not represented in its training data, clocks are completely limited to

17   predicting ages represented in the training data. If the true age of a tested sample lies outside of the

18   age range of the training data, the clock is unable to provide an appropriate prediction with the

19   magnitude of inaccuracy increasing as the true age of the sample becomes more extreme suggesting

20   that, in general, age range is more critical than tissue when training a clock.

21

22   Previous epigenetic clocks have shown that the predicted epigenetic age of iPSCs is significantly

23   lower than the cells from which they are reprogrammed and the chronological age of the donor at

24   sample donation [13]. The induction of pluripotency reprograms the epigenome, including at the loci

25   used in the clock algorithm, ultimately leading to a younger predicted epigenetic age. However, in

1   these analyses the predicted ages remain postnatal, which is unexpected as human pluripotent cells

2   only occur during the early stages of human development and hence we hypothesized that, with an

3   adequately calibrated clock, iPSCs would be expected to be estimated as being early fetal. Applying

4   the FBC to five datasets of iPSCs and iPSC-derived NPCs and neurons, we found this to be the case.

5   iPSCs were estimated as having a mean age of 75.6 dpc, fitting our hypothesis that they represent

6   first trimester developmental stages. These results align with studies that have reported

7   rejuvenation effects on the transcriptome, telomeres and mitochondria of iPSCs following

8   reprogramming [35–37]. In addition, we profiled the effect on predicted epigenetic age following the

9   differentiation of iPSCs towards neurons reporting a small but significant aging effect of 13 dpc. This

10   developmental stage coincides with fetal neurogenesis [38] and suggests that while differentiation

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

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

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

14   [36]. Therefore, it would be interesting to apply our FBC to iNs, iPSCs, iPSC-neurons and their

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

16   preserved in iNs. In addition, we tested the FBC on ESC and ESC-derived neurons. Although our

17   sample size was small, the results also suggested that the differentiation process induced a small

18   aging effect as the ESCs were differentiated into neurons. Altogether, our results indicate that iPSC-

19   neurons may have limited utility for the study of age-related brain diseases, like Alzheimer's disease

20   or other dementias, as many molecular processes related to an aging phenotype may not be

21   recapitulated.

22

23   Epigenetic clocks have been utilized for a wide range of applications [34], with a predominant focus

24   on exploring the biological meaning of deviations between chronological age and epigenetic age. As

25   we have shown the cause of this deviation may result from the use of an inappropriate clock,

1   therefore the FBC is a critical tool for assessing whether epigenetic age acceleration during

2   neurodevelopment is associated with later life outcomes such as disease or in utero exposures (such

3   as maternal smoking). The FBC also has utility for determining the specific developmental stage a

4   model of neurodevelopment recapitulates (e.g. brain organoids or cellular neuronal models) and

5   how different exposures or genetic backgrounds may influence neurodevelopmental processes and

6   aging.

7

8   While a strength of our study is the development of a bespoke clock to optimally profile the

9   epigenetic age of human fetal brain samples, due to the training data predominantly containing

10   second trimester samples the FBC is most accurate for this period of neurodevelopment. We are

11   confident that the FBC has correctly predicted fetal epigenetic ages for the cell lines included in our

12   analysis as the vast majority of the stem cell models and their derived neurons were less than the

13   median age in the training data. This indicates that the predictions are not confounded by saturation

14   of the coefficients. Although we took advantage of previously published data to include all available

15   samples appropriate for addressing our research questions, some of the group sizes, in particular the

16   NPCs, were small and we were not powered to detect significant aging effects as a result of

17   differentiation from iPSCs to NPCs in all datasets. Furthermore, across the different studies there

18   was variation in the estimated prenatal age of each cell state; we hypothesize that these differences

19   results from subtle variation in differentiation protocols, timepoints of cell collection or the

20   definition of NPCs within the respective studies [39]. Despite this, we are confident in the

21   conclusions we report as these study specific effects were controlled for in our analysis.

22

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

24   epigenetic age during neurodevelopment and for precise predictions a bespoke clock trained on

25   fetal brain data is required. Using the FBC to assess the epigenetic age of iPSCs and differentiated

1   neurons, we found that iPSCs and derived NPCs and neurons reflect early prenatal developmental

2   stages. Our findings question the suitability of the iPSC-neurons for the study of aging associated

3   processes.

4

5   **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
ESC	Embryonic Stem Cell
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
NPC	Neuronal precursor cell
QC	Quality control
RMSE	Root mean squared error

6

7   **Declarations**

8   *Ethical approval*

17

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

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

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

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

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

6   human brain tissue from the BDR cohort was granted by the University of Exeter, College of

7   Medicine and Health Research Ethics Committee under reference Mar20/D/009Δ5. Ethical approval

8   for the Price cellular samples is as follows: participants were recruited and methods carried out in

9   accordance to the 'Patient iPSCs for Neurodevelopmental Disorders (PiNDs) study' (REC No

10   13/LO/1218). Informed consent was obtained from all subjects for participation in the PiNDs study.

11   Ethical approval for the PiNDs study was provided by the NHS Research Ethics Committee at the

12   South London and Maudsley (SLaM) NHS R&D Office.

13

14   *Consent for publication*

15   Not applicable

16

17   *Availability of data and materials*

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

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

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

21   GSE116754 and GSE90871. DNAm data for the unpublished fetal brain are available from the

22   corresponding author on request. DNAm data for the cellular datasets, are also available under

23   accession numbers GSE158089, GSE105093, GSE51921, GSE31848 and GSE38214. DNAm data for

24   the cellular dataset *Bhinge* is available from the corresponding author on request. The coefficients of

1   the FBC are included within the article in **Table S2**. The custom R scripts used for this study and to

2   apply the FBC are available at https://github.com/LSteg/EpigeneticFetalClock.

3

4   *Competing interests*

5   The authors declare that they have no competing interests.

6

7   *Funding*

3

4   *Authors' contributions*

9

10   *Acknowledgements*

14

15   **Figure Legends**

16   **Figure 1 The Fetal Brain Clock (FBC) outperforms other DNAm clocks when applied to**

17   **neurodevelopmental samples.** Shown are scatterplots comparing chronological age (x-axis; days

18   post-conception) against (y-axis; days post-conception) predicted epigenetic age calculated using **A**

19   Fetal Brain Clock (FBC), **B** Horvath's multi tissue clock (MTC), **C** Knight's Gestational Age Clock (GAC),

20   and **D** Lee's Control Placental Clock (CPC) in our fetal brain testing dataset (n = 65, age range = 23 −

21   153 dpc). The black line indicates the identity line of chronological and predicted epigenetic age and

22   represents a perfect prediction. Two statistics were calculated to evaluate the precision of each

23   DNAm clock: Pearson's correlation coefficient (r) and the root mean squared error (RMSE).

1

2   **Figure 2 Validation of the Fetal Brain Clock in an independent fetal brain dataset.** Shown are

3   scatterplots comparing chronological age (x-axis; days post-conception (dpc)) against (y-axis; dpc)

4   predicted epigenetic age calculated using **A** Fetal Brain Clock (FBC), **B** Horvath's Multi Tissue Clock

5   (MTC), **C** Knight's Gestational Age Clock (GAC), and **D** Lee's Control Placental Clock (CPC) on data

6   from an independent validation dataset (n = 129, age range = 42 - 280). Two statistics were

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

8   root mean squared error (RMSE). The dashed line in **A** indicates 185 dpc, which is the highest

9   available age in the training dataset of the FBC.

10

11

12   **Figure 3 Comparisons of predicted epigenetic age using the fetal brain clock between cellular**

13   **models throughout differentiation states. A** Boxplots comparing the distribution of predicted

14   epigenetic age (days post-conception) in iPSCs and their derived NPCs and neurons, where each

15   panel represents a different dataset. P values of Tukey HSD corrected ANOVA for the *Imm* dataset

16   and two-sample t-tests for *Price*, *Fernández-Santiago* and *Sultanov* datasets are given. F. -S. =

17   *Fernández-Santiago*. **B** Boxplots of predicted epigenetic age calculated using the FBC where samples

18   are grouped by cell stage (n = 82, 30 iPSCs, 4 NPCs, 48 iPSC-neurons) and colored by dataset. P

19   values from mixed effects model are given for differences between iPSCs and NPCs (non-significant)

20   and iPSC and neurons. **C** Boxplots comparing the predicted epigenetic age in a cohort with iPSCs and

21   derived motor neurons. P values of two-sample t-test are given. **D** Boxplots of predicted epigenetic

22   age by FBC applied on two datasets including ESCs and their derived NPCs and neurons. P values of

23   Tukey HSD corrected ANOVA for the *Kim* dataset and two-sample t-tests for the *Nazor* dataset are

24   given

25

1   **Additional Files**

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

3

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

5

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

7   measured in days post-conception in A) training data, B) testing data and C) validation data.

8

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

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

11   predicted epigenetic age for **A** Fetal Brain Clock (FBC); **B** Horvath's Multi Tissue Clock (MTC); **C**

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

13   training of the FBC. Where necessary, predicted age was converted to days post-conception. The

14   black line indicates the identity line of chronological and predicted epigenetic age and represents a

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

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

17

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

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

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

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

22   independent adult brain dataset. Where necessary, predicted age was converted to years, where 0

23   indicates birth. The black line indicates the identity line of chronological and predicted epigenetic

24   age and represents a perfect prediction. Two statistics were calculated to evaluate the precision of

25   each DNAm clock: Pearson's correlation coefficient (r) and the root mean squared error (RMSE).

26

1 **References**

2 1. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007)
3   Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell
4   131:861–872

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

8 3. Fernández-Santiago R, Carballo-Carbajal I, Castellano G, et al (2015) Aberrant epigenome in
9   iPSC -derived dopaminergic neurons from Parkinson's disease patients. EMBO Mol Med
10   7:1529–1546

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

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

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

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

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

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

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

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

28 12. Hillary RF, Stevenson AJ, McCartney DL, et al (2020) Epigenetic measures of ageing predict
29   the prevalence and incidence of leading causes of death and disease burden. Clin Epigenetics
30   12:115

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

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

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

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

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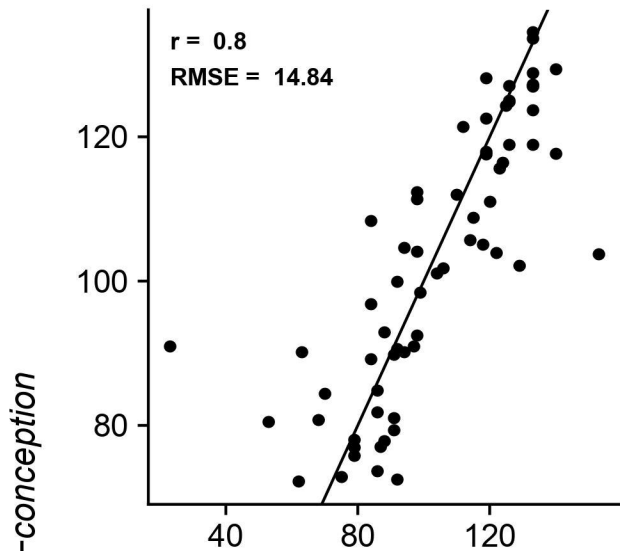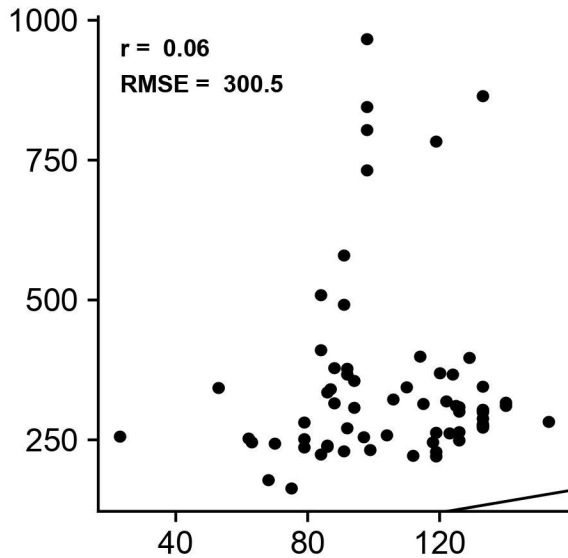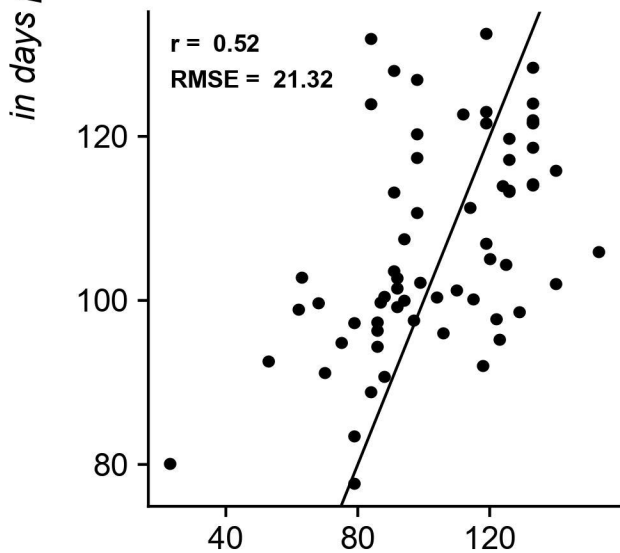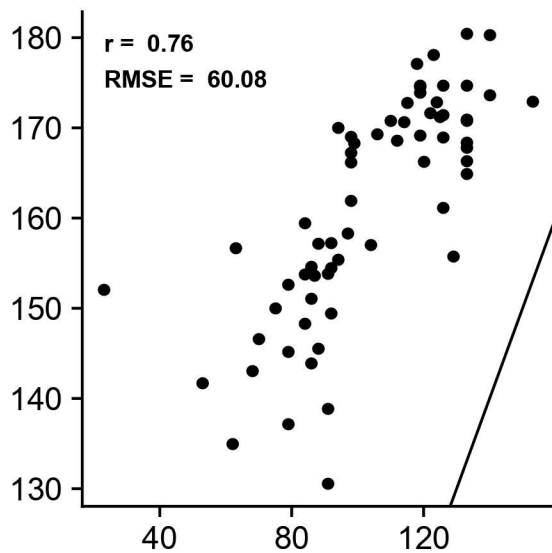40   based on blood methylation data. Genome Biol 17:1–11

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. R Development Core Team (2012) R: A Language and Environment for Statistical Computing.

20. 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

21. 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

22. 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

23. 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

24. 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

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

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

27. Chatterton Z, Hartley BJ, Seok MH, et al (2017) In utero exposure to maternal smoking is associated with DNA methylation alterations and reduced neuronal content in the developing fetal brain. Epigenetics and Chromatin. https://doi.org/10.1186/s13072-017-0111-y

28. Colunga T, Hayworth M, Kreß S, et al (2019) Human Pluripotent Stem Cell-Derived Multipotent Vascular Progenitors of the Mesothelium Lineage Have Utility in Tissue Engineering and Repair. Cell Rep 26:2566-2579.e10

29. Nazor KL, Altun G, Lynch C, et al (2012) Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell Stem Cell 10:620–634

30. Kim M, Park YK, Kang TW, et al (2014) Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. Hum Mol Genet 23:657–667

31. Liu Z, Leung D, Thrush K, Zhao W, Ratliff S, Tanaka T, Schmitz LL, Smith JA, Ferrucci L, Levine ME (2020) Underlying features of epigenetic aging clocks in vivo and in vitro. Aging Cell 19:e13229

32. Horvath S, Raj K (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet 19:371–384

33. Wagner W (2019) The link between epigenetic clocks for aging and senescence. Front Genet. https://doi.org/10.3389/fgene.2019.00303

34. Ryan CP (2020) "Epigenetic clocks": Theory and applications in human biology. Am J Hum Biol e23488

35. 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

36. 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

37. 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

38. Lui JH, Hansen D V., Kriegstein AR (2011) Development and evolution of the human neocortex. Cell 146:18–36

39. 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

40. 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

41. 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

42. 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
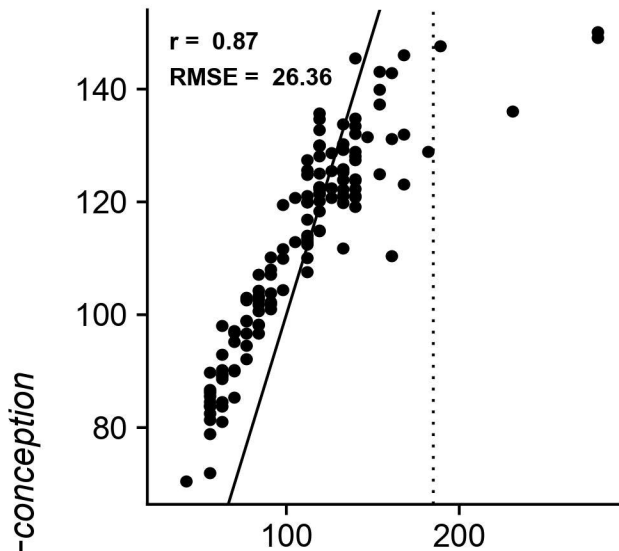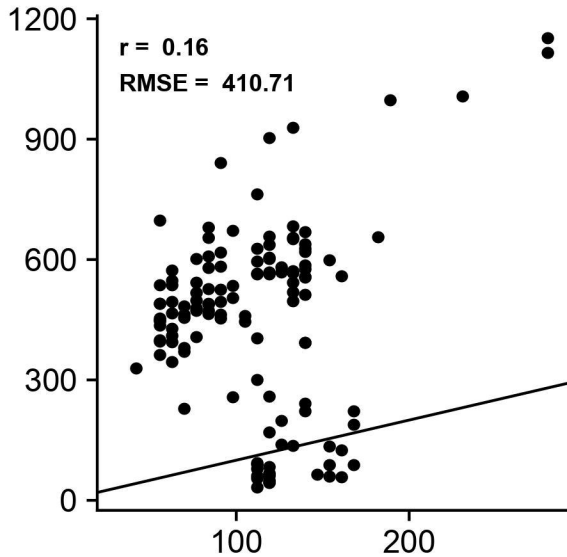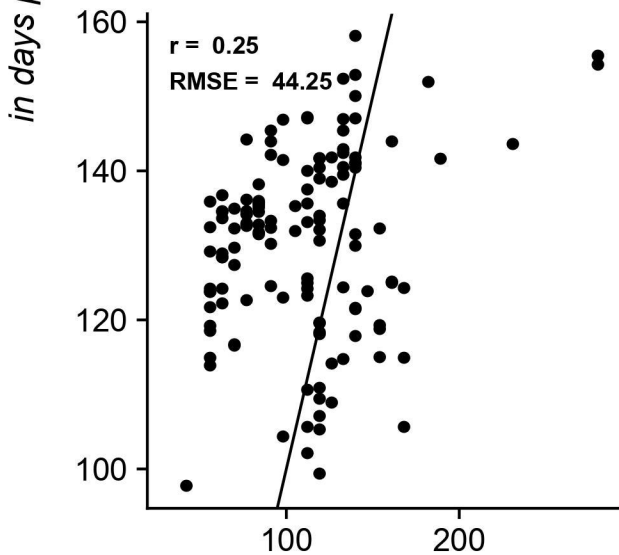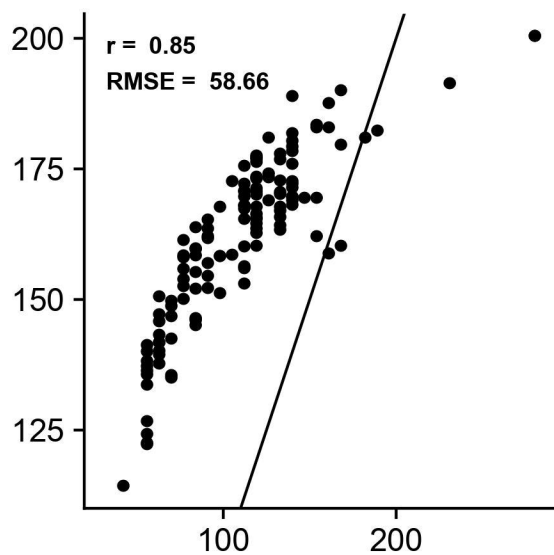
A FBC

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

A FBC

r = 0.87
RMSE = 26.36

B MTC

r = 0.16
RMSE = 410.71

C GAC

r = 0.25
RMSE = 44.25

D CPC

r = 0.85
RMSE = 58.66

Predicted Age
in days post–conception

Chronological Age
in days post–conception

A iPSC - Neuron Datasets

i. Imm   ii. Price   iii. Nazor

iv. F.–S.   v. Sultanov

B iPSC - Neuron Merged

C iPSC - Motor neuron

i. Bhinge

D ESC - Neuron

i. Nazor   ii. Kim