nature portfolio

Corresponding author(s): NCOMMS-24-20423

Last updated by author(s): Apr 12, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was assembled in silico from ten independent cohorts which is described in Table 1. The majority of cohorts (6/10) included in this study were described previously in the initial publication of the PedBE clock, and data were deposited in the publicly available Gene Expression Omnibus (GEO) SuperSeries data repository (accession: GSE137503). Data from another pediatric cohort were also downloaded from GEO (accession: GSE147058). Additional details on data collection from the cohorts are described in Supplementary methods.
Data analysis	Data analysis script will be made available through Kobor Lab Github. All the statistical analysis was performed in R Version 4.0.3. Pediatric epigenetic age was calculated using the publicly available PedBE tool (available from https://github.com/kobor-lab/Public-Scripts/blob/master/PedBE.Md). Horvath pan-tissue and Horvath skin-blood epigenetic clock ages were calculated using the online DNA Methylation Age Calculator developed by the clocks' creator (https://dnamage.genetics.ucla.edu/new)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Provide your data availability statement here.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex, as inferred from DNA methylation intensities on XX XY chromosomes was used in the current study. Inferred sex was concordant with self report and/or parent reported sex. No chromosomally intersex individuals were identified within the study population.
Reporting on race, ethnicity, or other socially relevant groupings	N/a
Population characteristics	Our study population of 3799 typically developing individuals , collected from ten cohorts, ranged in age from 2 months to 20 years old (46.22% females). Range of estimated buccal cell proportions, inferred from DNA methylation algorithm, ranged from 0.5-1 (average 0.85). Five cohorts were ran on the Illumina's 450K platform and five were on the Illumina's EPIC/850K platform. Details on each of the cohorts are presented in Table 1
Recruitment	This was secondary data analysis and no participants were recruited for this analysis
Ethics oversight	The majority of data sets were obtained from the publicly available Gene Expression Omnibus (GEO) repository and details regarding ethic approval of these cohorts were presented in a previous publication 4. In addition, research related to the four unpublished cohorts (SEED, BEPAC, NeuroTox, and OCD cohorts) presented in this manuscript was performed in compliance with local, state, and national regulations for the ethical treatment of human subjects. Ethics approval information is provided in the Supplementary Methods

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was chosen based on available pediatric cohorts in buccal DNA methylation from both publicly available GEO platform and collaborator's willingness to share the data
Data exclusions	With the exception of the OCD cohorts, all buccal samples were collected from typically developing children with estimated buccal cell proportions > 0.50 ,which is expected from cheek swab samples
Replication	Both buccal proportion estimation and epigenetic age estimation were replicated using a multi-method approach (different algorithms/ clocks). But all available data was used for all the analyses
Randomization	N/a
Blinding	N/a

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N/a
Novel plant genotypes	N/a
Authentication	N/a