nature portfolio

Corresponding author(s): Eric Verdin

Last updated by author(s): Mar 14, 2023

# 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<br>*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<br>*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 | As described in Methods, data were primarily collected from the Gene Expression Omnibus (GEO) database. |
| Data analysis | Minfi R package was utilized for generating DNA methylation beta values from .idat files. Glmnet R package was used for training the elastic net model used for chronological age prediction. Impute, watermelon, statix, DEGreport, and ggplot packages were used for data pre-processing and visualization. The methylclock package was utilized for age predictions for the Hannum, Horvath, PhenoAge (methylation-based), and Horvath Skin & Blood clocks. |

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

DNA methylation profiles generated in this study will be submitted to the GEO public repository prior to publication. Code used to generate the results in this study will be made public on GitHub prior to publication.

## Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

| | |
|---|---|
| Reporting on sex and gender | This study profiled cells from sixteen individuals in total, of whom nine are biological men and seven are biological women. |
| Population characteristics | Peripheral blood mononuclear cells obtained for this study were obtained from anonymous individuals donating plasma to a clinic in San Francisco. Ages of the donors varied (listed in deposited data and within the study description) and all donors tested positive for cytomegalovirus. |
| Recruitment | Peripheral blood mononuclear cells obtained for this study were obtained from anonymous individuals donating plasma to a clinic in San Francisco. Due to the screening process required to donate plasma, donors are likely healthier than the average individual. |
| Ethics oversight | No ethics approval was required for this study. |

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 | For our studies, we performed a power calculation estimate whereby we sought to be able to detect a five year difference in epigenetic age prediction using currently available epigenetic clocks between different immune subsets, with an assumption of three years for standard deviation of age prediction (estimated using previous literature). This indicated we would need at least six donors, with an alpha of .05 and a power of 80%. This informed both of the datasets in the study - the first dataset of seven individuals, and the second of nine. |
| Data exclusions | One naive T cell data point was excluded from analysis in the first data set due to human error in pipetting during sample transfer from one 96-well plate to the plate sent for DNA methylation analysis. No data points were excluded from the second dataset. |
| Replication | The second dataset generated in this study replicates and reinforces the first dataset generated in this study. Data from other studies are used to replicate the findings observed in this study. |
| Randomization | Samples were randomly divided into training and test sets for the purposes of developing and testing our machine learning model. |
| Blinding | Blinding was not relevant to this study, as no intervention was performed. |

# 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 |

## Methods

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

# Antibodies

Antibodies used: Information is provided in Supplementary Tables.

Validation: All antibodies have validation performed on the manufacturer's website. The T cell staining cocktail was also validated in our laboratory through RNA-Seq analysis after sorting (unpublished data) where the sorted fraction accurately recapitulated previously known transcriptomic profiles of the appropriate cell subsets.

# Flow Cytometry

## Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

Sample preparation: Detailed sample preparation steps are listed in Methods under "Immune cell isolation, sorting, and DNA extraction." Briefly, PBMCs were extracted using a Ficoll gradient. They were then washed and then either blocked with IgG and stained immediately with an antibody cocktail or processed via magnetic enrichment and then blocked with IgG and stained with antibody cocktail.

Instrument: BD FACSAriaTM II was used for cell sorting; Cytek AuroraTM was used for spectral flow phenotyping.

Software: FlowJo v10 was utilized for flow cytometry analysis.

Cell population abundance: Post-sort purity was at least 97% for every cell population. At least 1000 cells were analyzed in the post-sort.

Gating strategy: For all gates, cells were initially defined with an inclusive FSC/SSC gate that only excluded <50k FSC (small cell debris). A singlet cell gate was applied using SSC-H and SSC-A, followed by a live gate based on staining negative for a LIVE/DEAD stain.

For sorting T cells, T cells were identified based on expression of CD3, and CD4/CD8 were split based on uniquely expressing one marker. Naive/CM/EM/TEMRA populations were identified using CD28 and CD45RO.

For sorting B cells, NK cells, and monocytes, CD3 was used as a dump gate. B cells were identified via expression of CD19 and NK/monocytes were gated as a part of the "non B-cell" CD19- fraction. B cells were then subdivided as memory or naive on the basis of IgD and CD27. Non-B cells were subdivided as CD56dim CD16+ NK cells based on CD56 and CD16 and classical monocytes were gated on the basis of CD14 and CD16.

For PBMC phenotyping, PBMCs were gated as CD45+. T cells were identified as CD3+ T cells, and then CD4+ and CD8+ T cells were identified by their respective markers. Naive/CM/EM/TEMRA cells were gated using CD45RA and CCR7. For CD3- cells, B cells and non-B cells were gated on the basis of CD19. B cells were then analyzed as naive, class-switched or non-class-switched memory on the basis of CD27 and IgD. Non-B cells were gated as monocytes on the basis of HLA-DR and then into non-classical, classical and intermediate fractions by CD14 and CD16 staining. Non-B cells were gated as NK cells using CD56 and then as CD56dimCD16+ NK cells using CD56 and CD16.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.