

## Reporting Summary

Nature Research 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 Research 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



Phenotype data were collected by online questions and surveys on the 23andMe Research platform (<https://customercare.23andme.com/hc/en-us/articles/212881977-23andMe-Research-Surveys-and-Questions>) from consented 23andMe Research Participants (<https://customercare.23andme.com/hc/en-us/articles/212195708-Research-Participation-and-Consent>). Genetic data were the de-identified individual-level genetic information of the consented 23andMe Research Participants (<https://www.23andme.com/howitworks/>).

#### Data analysis



Genetic and GWAS analyses were performed on 23andMe Research platform. It has been described in numerous peer-reviewed publications (<https://research.23andme.com/publications/>). Additional statistical analyses, figures, and tables were generated on R v3.3.

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 Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

23andMe provides access to full summary statistics from published GWAS analyses through a Data Transfer Agreement that protects the privacy of our participants'

## 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](https://www.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	The sample size was determined based on GWAS and PRS statistical powers.
Data exclusions	<p>Genetic data was processed through 23andMe GWAS pipeline; Genotyped and imputed variants were excluded based on Quality Control statistics.</p> <p>Only the phenotype data of 23andMe participants who completed the full surveys and questions were included in the analyses. No other exclusion criteria were used.</p>
Replication	We included in the main manuscript and supplementary material comparisons to published and independent datasets.
Randomization	The manuscript included Demographic statistic comparison between the training and validation sets.
Blinding	Group allocation for the training and validation sets was determined by the date of online survey completion. After consenting to research, 23andMe Research Participants can complete the surveys at anytime.

## 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

### Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involved in the study   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms            |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern           |

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involved in the study                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	<p>The cohort used in this study is restricted to 23andMe Research Participants from European ancestry. Ancestry for 23andMe Research Participants was determined via genetics.</p> <p>The demographic characteristics of this cohort was described in the main manuscript and the supplementary material.</p>
Recruitment	The 23andMe Research cohort is derived from the 23andMe Customers. Every 23andMe Research Participants had the opportunity to complete the online surveys (non-targeted surveys).
Ethics oversight	23andMe Research Participants provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent Review Services ( <a href="http://www.eandireview.com">www.eandireview.com</a> ).

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