nature research

Corresponding Author: Pirro Hysi, Tim Spector, Manfred Kayser

Corresponding Author: Pirro Hysi, Tim Spector, Manfred Kayser

Date: Feb 4, 2018

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

▶ Experimental design

1. **Sample size**

   Describe how sample size was determined.

   > This work used existing samples, the most important of which were the UK Biobank and 23andMe customers. The UK Biobank was designed specifically for research in epidemiology, is the most powered cohort to date (up to 500,000 currently, of which ~130,000 were used for the purpose of the manuscript) and together with the 23andMe cohort they have exceptional power to detect genetic variants at even effect sizes or minor allele frequencies far beyond common publication benchmarks.

2. **Data exclusions**

   Describe any data exclusions.

   > To avoid issues related to population structure, only individuals of European origin were included in this study.

3. **Replication**

   Describe whether the experimental findings were reliably reproduced.

   > This study is primarily focused on two major cohorts, each exceeding 100,000 subjects. These cohorts reliably replicate each-other's discoveries, but as an additional precaution, these results were compared to results pooled from a meta-analysis of several smaller cohorts, members of the Visigen academic consortium.

4. **Randomization**

   Describe how samples/organisms/participants were allocated into experimental groups.

   > This study did not involve any intervention.

5. **Blinding**

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   > The investigators did not participate in data collection and only analyzed data made available to them.
   > The investigators were blind to any individual genotypic or phenotypic status.

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The <u>exact</u> sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Different software was used at different stages of the analyses. For example, regression models were built and assessed using PLINK, meta-analysis was run using METAL and GWAMA (as results were identical only the former were reported). Conditional analyses and estimates of the proportions of heritability explained were run using the GCTA software. Natural selection signals were assessed using data generated by others using the iHS and XP-EHH methods, downloadable from the 1000 Genomes Browser. R base packages as well as the glmnet, lars, nnet and pROC were used for the prediction models. The gene set enrichment analysis was run on the software Magenta.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

This manuscript reports no experimental results, only statistical analyses

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used in this study.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used in this study

b. Describe the method of cell line authentication used.

No cell lines were used in this study

c. Report whether the cell lines were tested for mycoplasma contamination.

No cell lines and no mycoplasma was used in this study

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No cell lines were used in this study

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

> This manuscript describes only an observational genetic epidemiological work.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

> This was an observational study, involving no experiment.
> The observations were from questionnaires answered by volunteers of European origin, who reported their natural hair color, mostly in adult, or very late childhood age.