nature portfolio

Corresponding author(s): Manfred Kayser, Fan Liu

Last updated by author(s): Jun 26, 2025

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	No software was used.
Data analysis	MeshMonk is implemented as an open-source toolbox available at https://gitlab.kuleuven.be/mirc/meshmonk. C-GWAS is implemented as an open-source R package available at https://github.com/Fun-Gene/CGWAS. LDSC is implemented as an open-source command line tool available at https://github.com/bulik/ldsc. SpiralNet++ is implemented as an open-source 3D graph auto-encoder available at https://github.com/sw-gong/spiralnet_plus. The latest version of MeshMonk is available at https://doi.org/10.6084/m9.figshare.c.6858271.v1. The latest version of C-GWAS is available at https://doi.org/10.6084/m9.figshare.26360206. Other softwares used in data analysis includes: Plink (https://www.cog-genomics.org/plink2/), FastGWA (https://yanglab.westlake.edu.cn/software/gcta/#fastGWA), FUMA (https://fuma.ctglab.nl/), GCTA-Fst (https://yanglab.westlake.edu.cn/software/gcta/#Fst), R package 'shapes' (version 1.2.7), R package 'irr' (version 0.84.1), R package 'mamba' (version 1.12), R package 'clusterProfiler' (Version 4.6.2), R package 'org.Hs.eg.db' (version 3.16.0), R package 'biomaRt' (version 2.54.1), R package 'MASS' (version 7.3.57), R package 'pROC' (version 1.18.4).

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

Summary statistics data of C-GWAS and MinGWAS, full FUMA results in 253 loci, visualized regional-PVE of 253 lead SNPs, and reconstructed average facial images using 3D graph auto-encoder for 23 traits in face PRS profiles are available at https://doi.org/ 10.6084/m9.figshare.26360206. Single-trait summary statistics of meta-analyses GWAS for the European discovery phase data generated in this study are available at https://doi.org/10.5281/zenodo.13730680. GWAS summary statistics for the East Asian replication phase data are available at https://www.biosino.org/node/project/detail/OEP00002283. The ChIP-seq data of CNCC are available in the Gene Expression Omnibus with accession number GSE70751 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70751). The regulation region annotation data in embryonic craniofacial tissues are available in the Gene Expression Omnibus with accession number GSE97752 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97752). The regulation region annotation data in other general cell types are available in the Roadmap Epigenomics Project (https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state). The regulatory network of CNCC is available at https://github.com/AMSSwanglab/hReg-CNCC. The 1000 Genomes Project phase 3 data are available at https://www.internationalgenome.org/category/phase-3. Other statistical data for generating figures in this study are provided in the Supplementary Information and Supplementary Data. Individual-level data used in this study have different data accessibility conditions based on the cohorts' informed consent. For ALSPAC, data access can be granted via an approved application (http://www.bristol.ac.uk/alspac/researchers/access/) according to defined regulations (https://www.bristol.ac.uk/media-library/sites/alspac/documents/researchers/data-access/ALSPAC_Access_Policy.pdf) including on the time frame of response and access, and queries can be sent to the ALSPAC data team (alspac-data@bristol.ac.uk) and the ALSPAC Executive team (alspac-exec@bristol.ac.uk). For TwinsUK, data accessed can be requested by submitting a formal Data Access Application (http://twinsuk.ac.uk/resources-for-researchers/access-our-data/) according to defined regulations (https://twinsuk.ac.uk/resources-for-researchers/our-data/) including on the time frame of response and access, and queries can be sent to the data access and collaboration manager Victoria Vazquez, at King's College London (Victoria.vazquez@kcl.ac.uk). For RS and the two US datasets (US-P and US-I), informed consent does not allow sharing individual-level data, unless as part of direct collaborations with the cohort owners. For RS, collaboration requests can be directed to Frank J.A. van Rooij at Erasmus MC (f.vanrooij@erasmusmc.nl), for US-P to Mark Shriver at Pennsylvania State University (mds17@psu.edu), and for US-I to Susan Walsh at Indiana University Indianapolis (walshsus@iu.edu); responses to collaboration requests will be answered in a timely fashion, typically one calendar month.

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	No sex- or gender-based analyses were conducted.
Reporting on race, ethnicity, or other socially relevant groupings	The discovery phase of meta-analysis GWAS included 11,662 individuals of European descent. The replication analysis included GWAS summary statistics of 9,674 East Asian individuals.
Population characteristics	The discovery phase of meta-analysis GWAS included five cohorts from the Netherlands (RS), the United Kingdom (TwinsUK and ALSPAC) and the United States (US-P and US-I). Their characteristics were detailed in Supplementary Table 1. The replication analysis included meta-analysis GWAS summary statistics of three cohorts from China. Their characteristics were detailed in Zhang et al. (DOI: 10.1038/s41588-022-01038-7)
Recruitment	Details of the recruitment of the Rotterdam Study, the TwinsUK study and AlSPAC were provided in Methods section. The recruitment of the US-P study and the US-I study are detailed in White et al. (DOI: 10.1038/s41588-020-00741-7). The recruitment of the three cohorts from China (NSPT, NHC and TZL) are detailed in Zhang et al. (DOI: 10.1038/s41588-022-01038-7).
Ethics oversight	The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Dutch Ministry of Health, Welfare and Sport. The Rotterdam Study has been entered into the Netherlands National Trial Register and into the WHO International Clinical Trials Registry Platform and under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians. All participants of the the TwinsUK study provided fully informed consent under a protocol reviewed by the St. Thomas' Hospital Local Research Ethics Committee. Ethical approval for the ALSPAC was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). For the two US cohorts from Pennsylvania (US-P) and Indiana (US-I), institutional review board approval was obtained at each recruitment site of US-P and US-I, and all participants gave their written informed consent before participation. For the US-P sample, the following local ethics approvals were obtained: Urbana-Champaign, IL (PSU IRB 13103); New York, NY (PSU IRB 45727); Cincinnati, OH (UC IRB 2015-3073); Twinsburg, OH (PSU IRB 2503); State College, PA (PSU IRB 44929 and 4320); Austin, TX (PSU IRB 44929); and San Antonio, TX (PSU IRB 1278). For the US-I sample, the following local ethics approvals were obtained: Indianapolis, IN and Twinsburg, OH (IUPUI IRB 1409306349). For the three cohorts from China, all participants provided written informed consent, and all study protocols were approved

by the institutional review boards of the pertinent research institutions. NSPT is a subproject of The National Science & Technology Basic Research Project approved by the Ethics Committee of Human Genetic Resources of School of Life Sciences, Fudan University, Shanghai (14117). NHC was approved by the Ethics Committee of Human Genetic Resources at the Shanghai Institute of Life Sciences, Chinese Academy of Sciences (ER-SIBS-261410-A1801). TZL was approved by the Ethics Committee of Human Genetic Resources at the Shanghai institute of life Sciences, Chinese Academy of Sciences (ER-SIBS-261410).

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 determined by the amount of data available in the cohort study and e GWAS summary data that had predetermined sample sizes. Details are provided in Methods section.
Data exclusions	Data exclusions were based on image quality and missing data, etc. Samples identified as outliers were also excluded as the study focused on normal-range facial shape variation. Details are provided in Methods section.
Replication	Among the 253 lead SNPs (including 97 novel ones) identified by the discovery C-GWAS in Europeans, 176 (including 38 novel ones) were replicated using an LD-based analysis after FDR correction in a recent facial GWAS study involving East Asian individuals from China, as showed in Supplementary Data 5 and detailed in Methods section.
Randomization	Randomization was not applicable to the study as this study was not designed to study treatment effects.
Blinding	Blinding was not applicable to the study as this study was not designed to study treatment effects

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	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*