nature research

Corresponding author(s): Charleston W.K. Chiang, John Novembre

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 Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	The data generation pipeline and associated softwares used were detailed in a previous publication (Sidore et al. Nature Genetics 2015). Any additional data used in this publication were generated in the same pipeline to ensure data uniformity. For variant calling of the high coverage individuals, we used samtools (v0.2) and bcftools (v1.2).
Data analysis	Softwares used for data analysis include PLINK (v1.08), EIGENSTRAT (v.5), ADMIXTURE (v1.22 and v1.3), EEMS (v.0.0.0.9), Admixtools (v3.0), ALDER (v1.03), SMC++ (v1.9.3), and MSMC (v0.1.0).

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/reviewers upon request. 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

Allele frequency summary data analyzed in the study will be deposited to EGA under accession number EGAS00001002212. The disaggregated individual-level sequence data for 1887 samples - from adult volunteers of the SardiNIA cohort longitudinal study - analysed in this study are from Sidore et al (2015) and are available from dbGAP under project identifier phs000313.v3.p2. The remaining individual-level sequence data are from a case-control study of autoimmunity from across Sardinia, and per the obtained consent and local IRB, these data are only available for collaboration by request from the project leader (Francesco Cucca, Consiglio Nazionale delle Ricerche, Italy).

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	We used all available data and the maximally unrelated subsets as detailed in the Methods for population genetic analysis
Data exclusions	Individuals appearing as the outliers in quality of imputed genotypes upon visual inspection were excluded (8 out of 3,514). Genetically identified related individuals were also excluded.
Replication	General findings are consistent with results from limited Sardinian samples from external reference datasets such as HGDP or SGDP.
Randomization	For subset analyses, a randomly selected subsets are used. In EEMS analysis, we specifically selected individuals with all four grandparents born in the same location to reduce impact of recent motility in human populations (see Methods).
Blinding	The investigators were not blinded. There is no testing of an intervention vs. placebo group, and thus no blinding is necessary.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	We included in this study all individuals from the SardiNIA/Progenia longitudinal study of aging based in the Ogliastra region and from the case-control studies of Multiple Sclerosis and Type 1 Diabetes across the general population of Sardinia. For the SardiNIA study, over 6000 individuals older than 13 years of age were recruited from four villages in Lanusei, Sardinia. For the case-control study cohort, we required individuals to have at least three Sardinian grandparents.
Recruitment	Details of the sample recruitments are described in prior publications (Sidore et al. Nature Genetics 2015, Sanna et al. Nat.

Recruitment	Genet. 2010, Zoledziewska et al. Genes Immun. 2013, and cited in the current report). As this is a population genetic analysis, any bias in recruitment, if present, is unlikely to impact the estimates of allele frequencies genome-wide.