

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☐	☒ 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 <i>d</i> , Pearson's <i>r</i>), 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 for data collection beyond standard accession from the public repositories listed in the Data Availability Statement.
Data analysis	The online SV imputation tool developed in this study is available at https://yanglab.westlake.edu.cn/impute_sv. The scripts utilized for reference panel construction and data analysis can be accessed at the Zenodo repository: https://zenodo.org/records/15825718. Other software or packages used in this study are summarized as follows. HiFiAdapterFilt (v2.0.0): https://github.com/sheinasim-USDA/HiFiAdapterFilt Hifiasm (v0.19.8): https://github.com/chhy123/hifiasm QUAST (v5.0.2): https://github.com/ablab/quast Merqury (v1.3): https://github.com/marbl/merqury BUSCO (v5.2.2): https://busco.ezlab.org dipcall (v0.3): https://github.com/lh3/dipcall SVIM-asm (v1.0.2): https://github.com/eldariont/svim-asm PAV (v2.2.4.1): https://github.com/EichlerLab/pav minimap2 (v2.22): https://github.com/lh3/minimap2 unimap: https://github.com/lh3/unimap SAMtools (v1.14): https://github.com/samtools/samtools BCFtools (v1.14): https://github.com/samtools/bcftools SVanalyzer: https://github.com/nhansen/SVanalyzer truvari (v3.4.0): https://github.com/ACEnglish/truvari AnnotSV (v3.4): https://github.com/lgmgeo/AnnotSV

UCSC liftOver tool: <https://genome.ucsc.edu/cgi-bin/hgLiftOver>
 Tandem Repeat Finder (v4.10.0): <https://github.com/Benson-Genomics-Lab/TRF>
 WhatsHap (v1.4): <https://github.com/whatschap/whatschap>
 BEDtools (v2.29.1): <https://github.com/bedtools/bedtools>
 GCTA (v1.93.3): <https://yanglab.westlake.edu.cn/software/gcta/#Overview>
 MINIMAC3 (v2.0.1): <https://github.com/Santy-8128/Minimac3>
 SHAPEIT4 (v4.2): <https://odelaneau.github.io/shapeit4/>
 MINIMAC4 (v1.0.2): <https://github.com/statgen/Minimac4>
 GCTA-fastGWA: <https://yanglab.westlake.edu.cn/software/gcta/index.html#fastGWA-GLMM>
 GCTA-fastGWA-GLMM: <https://yanglab.westlake.edu.cn/software/gcta/index.html#fastGWA-GLMM>
 GCTA-COJO: <https://yanglab.westlake.edu.cn/software/gcta/index.html#COJO>
 PLINK (v2.0): <https://www.cog-genomics.org/plink>
 GCTA-GREML: <https://yanglab.westlake.edu.cn/software/gcta/index.html#GREML>
 FINEMAP (v1.4.1): <http://www.christianbenner.com/>
 WashU Epigenome Browser: <https://epigenomegateway.wustl.edu/>
 region-plot: <https://github.com/pgxcentre/region-plot>
 Cytoscape: <https://cytoscape.org/>
 OSCA: <https://yanglab.westlake.edu.cn/software/osca/#Overview>
 TOPMed Imputation Server: <https://imputation.biodatacatalyst.nih.gov/#!pages/home>
 R (v3.6.3): <https://www.r-project.org>
 ggplot2 (v3.5.1): <https://github.com/tidyverse/ggplot2>

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](#)

The data used and generated in this study are summarized as follows:

The assembly-level SVs and reference panels constructed in this study can be accessed at https://yanglab.westlake.edu.cn/impute_sv/resource.
 The individual-level genotype and phenotype data are available upon formal application to the UKB (<http://www.ukbiobank.ac.uk>).
 The SV- and VNTR-GWAS summary statistics generated in this study are available at https://yanglab.westlake.edu.cn/data/ukb_sv_gwas.
 The HPRCY1 assemblies are available at https://github.com/human-pangenomics/HPP_Year1_Data_Freeze_v1.0.
 The HGSVC3 assemblies are available at https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HGSVC3/working/.
 The CPC assemblies are available at <https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA011422>.
 The APR assemblies are available at <https://www.mbr.u.ac.ae/the-arab-pangenome-reference/>.
 The CN1 assemblies are available at <https://genome.zju.edu.cn/Downloads>.
 The YAO assemblies are available at <https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA017932>.
 The GIAB Tier 1 SV benchmark set for HG002 (v0.6) is available at dbVar (accession: nstd175) and at ftp://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/analysis/NIST_SVs_Integration_v0.6/.
 The GIAB genome stratification (v3.5) is available at <https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/genome-stratifications/>.
 The population-level HGSVC3 Freeze4 SV sets are available at https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HGSVC3/working/20240415_Freeze4.
 The assembly-level HGSVC3 Freeze4 SV sets are available at https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HGSVC3/working/20240307_PAV_VCF/.
 The HPRCY1 pangenome-derived SV sets are available at https://github.com/human-pangenomics/hpp_pangenome_resources.
 SVs called from the moderate-coverage ONT data of 888 1KGP3 individuals are available at <http://1kgp-sv-imputation.s3-website-eu-west-1.amazonaws.com/?prefix=panel/>.
 SVs called from high-coverage ONT data of 100 1KGP3 individuals are available at https://s3.amazonaws.com/1000g-ont/index.html?prefix=Gustafson_etal_2024_preprint_SUPPLEMENTAL/.
 The gnomAD SV set (v4.1) is available at <https://gnomad.broadinstitute.org/downloads>.
 The 1000 Genomes Project Phase 3 genotypes are available at https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20201028_3202_phased/.
 The GTEx (v8) data are available at <https://www.gtexportal.org/home/datasets>.
 The histone ChIP-seq peak and RBP eCLIP peak data are available through the ENCODE portal (<https://www.encodeproject.org>).
 The 18-state Roadmap model annotation data are available at EpiMap (<https://compbio.mit.edu/epimap/>).
 The TAD coordinate data are available at the 3D Genome Browser (<http://3dgenome.fsm.northwestern.edu/publications.html>).
 The deCODE SV-GWAS summary data are available at <https://www.decode.com/summarydata/>.
 The genome partition of LD-independent blocks is available at <https://bitbucket.org/nygcresearch/ldetect-data/>.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Sex was included as a biological variable and used as a covariate in our analyses. The data was obtained from the UK Biobank

Reporting on sex and gender	(UKB), which records sex at recruitment.
Population characteristics	The association analyses were performed on 456,643 UKB individuals of European ancestry.
Recruitment	We analyzed existing UKB data sets. Thus, no recruitment was performed.
Ethics oversight	This study was approved by the Ethics Committee of Westlake University (No. 20200722YJ001).

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](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	For the SV imputation reference panels, 241 individuals were included, and for the SV association analyses, 456,643 UKB individuals of European ancestry were included.
Data exclusions	Participants who withdraw from UK Biobank were excluded.
Replication	The robustness of our results was confirmed through a series of sensitivity analyses with various statistical settings and filtering criteria.
Randomization	The SV reference panel was constructed using all available samples; therefore, randomization was not applicable. Covariates such as sex, age, and genetic principal components were controlled for in the SV association analyses.
Blinding	The SV reference panel was constructed using all available samples, and analyses were performed on existing datasets; therefore, no blinding measures were implemented in this study.

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