nature portfolio

Corresponding author(s): Yoshihiko Suzuki, Shinichi Morishita

Last updated by author(s): May 7, 2026

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	SMRT Link (v11.0) MinKNOW (v21.06.8)
Data analysis	Guppy (v4.2.0) Dorado (v0.5.3) hifiasm (v0.19.8) MUMmer (v4.0.0) purge_dups (v1.2.5) RagTag (v2.1.0) hpp-production_workflows (https://github.com/human-pangenomics/hpp_production_workflows; GitHub commit ID: a4e5a15) seqkit (v2.0.0) Merqury (v1.3) minigraph-cactus (v2.9.0) pggb (nf-core/pangenome v1.1.2) odgi (v0.8.6) vg (v1.57.0) vcfbub (v0.1.1) vcflib (v1.0.9) Truvari (v5.3.0) bcftools (v1.18) Liftoff (v1.6.3)

Immuannot (GitHub commit ID: bca45ef)
winnowmap2 (v2.03)
mosdepth (v0.3.1)
DeepVariant (v1.5.0)
SyRI (v1.7.0)
bedtools (v2.30.0)
RepeatMasker (v4.1.5)
Assembly_HSat2and3_v3.pl (https://github.com/altemose/chm13_hsat, v3)
Biser (v1.4)
asm-to-reference-alignment (https://github.com/mrvollger/asm-to-reference-alignment, v0.1)

Custom scripts and Jupyter Notebooks for data analysis used in this study are available at https://doi.org/10.5281/zenodo.14160240.

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 PacBio HiFi, ONT ultra-long, and Omni-C reads generated in this study at the University of Tokyo have been deposited in the NCBI SRA database under accession code PRJNA1459655 [https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1459655]. ONT ultra-long reads generated at UCSC have also been deposited in the NCBI SRA database under accession code PRJNA731524 [https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA731524]. The 20 assembled haplotype sequences have been deposited in the NCBI GenBank database under accession codes PRJNA1184421 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1184421] – PRJNA1184440 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1184440]. Sequence names in the fasta files of the assembled haplotype sequences follow the format NA<sample_ID>#<hap_ID>#<scaffold_name>; e.g. NA18940#1#chr1_hap2, where 18940 indicates the sample ID and chr1_hap2 denotes the chromosomal scaffold sequence name used in this paper. Reads and assemblies are also available at https://humanpangenome.org/hprc-data-release-2/. The pangenome graphs, segmental duplications, and Immuannot results generated in this study are available at https://doi.org/10.5281/zenodo.14160240. The underlying data for all plots generated in this study are provided in the Source Data file.

Previously published data used in this study are available as follows: HPRC pangenome graphs under accession code PRJNA850430 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA850430]; CPC pangenome graphs from the CPC website [https://pog.fudan.edu.cn/cpc/#/data]; Illumina reads for the ten Japanese samples from the IGSR website [https://www.internationalgenome.org]; small variants called from Illumina reads for the ten Japanese samples from the 1000 Genomes Project website [https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20220422_3202_phased_SNV_INDEL_SV]; and T2T-CHM13 genome sequence and annotations from the CHM13 GitHub repository [https://github.com/marbl/CHM13].

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	All 10 samples were from male individuals, selected to enable chromosome Y haplotype reconstruction. No sex- or gender-based comparative analyses were performed.
Reporting on race, ethnicity, or other socially relevant groupings	All samples were from the 1000 Genomes Project Japanese in Tokyo (JPT) population. No additional race, ethnicity or other socially relevant groupings were assigned or analyzed in this study.
Population characteristics	All 10 samples were Japanese in Tokyo (JPT) samples from the 1000 Genomes Project.
Recruitment	No new participants were recruited for this study.
Ethics oversight	This study was conducted in accordance with the Ethical Guidelines for Medical and Biological Research Involving Human Subjects and was approved by the Research Ethics Committee of the Graduate School of Frontier Sciences, The University of Tokyo (review no. 22-454). The original 1000 Genomes Project samples were collected with informed consent and ethical review.

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	No statistical method was used to predetermine sample size. We selected 10 male Japanese in Tokyo (JPT) samples from the 1000 Genomes Project collection to reconstruct autosomal and sex chromosome haplotypes. This was a resource-generation study, not a hypothesis-testing study requiring a formal power calculation.
Data exclusions	No data were excluded.
Replication	The computational and statistical analyses can be reproducible by using publicly available codes.
Randomization	No randomization or experimental group allocation was performed because this study did not involve interventions or treatment groups. The 10 samples were male Japanese in Tokyo (JPT) samples selected from the 1000 Genomes Project collection to enable chromosome Y haplotype reconstruction.
Blinding	Blinding was not relevant because the study involved no interventions, treatment assignments or subjective outcome assessments.

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	GM18940, GM18943, GM18944, GM18945, GM18948, GM18959, GM18960, GM18967, GM18970 and GM18982, obtained from the 1000 Genomes Project collection at the Coriell Institute for Medical Research.
Authentication	Cell lines were obtained directly from the Coriell Institute for Medical Research. Sample identity was confirmed by comparing the generated sequencing data with publicly available 1000 Genomes Project genotype data.
Mycoplasma contamination	Cell lines were obtained from the Coriell Institute for Medical Research, whose quality-control procedures include mycoplasma detection for cell lines.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used in this study is known to be a commonly misidentified cell line.

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.*