

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	Oxford Nanopore Technologies MinKNOW (v25.03.7) and iVision (v4.5).
Data analysis	All custom code used in this project are publicly available at the following GitHub repositories: AssemblyRepairer (v1.0; https://github.com/logsdon-lab/AssemblyRepairer); CenMAP (v0.4.3.1; https://github.com/logsdon-lab/CenMAP); CenStats (v0.0.10; https://github.com/logsdon-lab/CenStats); NucFlag (v0.3.3; https://github.com/logsdon-lab/NucFlag); CenPlot (v0.1.4; https://github.com/logsdon-lab/CenPlot); Snakemake-HumAS-SD (v0.1.0; https://github.com/logsdon-lab/Snakemake-HumAS-SD); DiMeLo-seq-Viz (v0.1.0; https://github.com/logsdon-lab/Snakemake-DiMeLo-seq-Viz); CUT&RUN-Viz (v0.0.1; https://github.com/logsdon-lab/Snakemake-CutNRUN); Pan-centromeric mutation rate estimation (v0.1.0; https://github.com/logsdon-lab/Snakemake-MutationRate); and Cencyclopedia (v0.0.1; https://github.com/logsdon-lab/Cencyclopedia). Other publicly available software used in this study include: minimap2 (v2.28), pbmm2 (v1.13.1), BEDTools (v2.31.1), SAMtools (v1.22 and v1.9), RepeatMasker (v4.1.0 and v4.1.6), Dfam (v3.8), Pysam (v0.22.0), MUSCLE (v3.8.31), L1ME-AID (v1.2.0-beta), dna-brnn (v0.1-r65), trf (v4.09), StringDecomposer (no version specified), CDR-Finder (v1.0.4), IGV (v2.19.5), ModDotPlot (v0.8.4), HMM-Flagger (v1.1.0), GAVISUNK (v1.0), Meryl (v1.4.1), Merqury (v1.3), SVbyEye (v0.99.0), R (v4.4.2), Gepard (v2.1.0), modkit (v0.3.2), cutadapt (v5.1), seqkit (v2.10.0), bwa mem (v0.7.19), deepTools (v3.5.6), MAFFT (v7.526), IQ-TREE (v2.3.6), iTOL (v7.2.2), subseq (no version specified), ARCKmerFinder (v1.1) and Adobe Photoshop (v25.11.0).

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 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](#)

All datasets generated and/or used in this study are listed in Supplementary Table 24 with their BioProject, accession # (if available), and/or URL. For convenience, we also list their BioProjects and/or URLs here: 65 original human whole-genome assemblies generated by the HGSVC (PRJEB76276 and PRJEB83624); 65 human whole-genome assemblies with repaired centromeric contigs (PRJNA1309774); PacBio HiFi and ONT data from 65 humans generated by the HGSVC (PRJEB58376, PRJEB75216, PRJEB77558, PRJEB75190, PRJNA698480, PRJEB75739, PRJEB36100, PRJNA988114, PRJNA339722, PRJEB41778, ERP159775); Illumina data from 65 humans generated by the HGSVC (PRJNA200694, ERP114329, ERP120144); CENP-A and IgG DiMeLo-seq data from 6 humans (SRR35424754, SRR35424755, SRR35269316, SRR35269317, SRR35269314, SRR35269315, SRR35144393, SRR35144394, SRR35424752, SRR35424753, SRR35048265, SRR35048266); CENP-A and IgG CUT&RUN ONT data from GM20355 (SRR35048263, SRR35048264); 227 original human whole-genome assemblies generated by the HPRC (PRJNA730822); PacBio HiFi data from 227 humans generated by the HPRC (PRJNA701308, PRJNA731524); ONT data from 227 humans generated by the HPRC (PRJNA701308, PRJNA731524); 28 human whole-genome assemblies, PacBio HiFi data, and ONT data from the CEPH 1463 pedigree (PRJEB86317 or dbGaP accession: phs003793.v1.p1); Element data from the CEPH 1463 pedigree (PRJEB3246); CHM13 whole-genome assembly (PRJNA559484); CHM1 whole-genome assembly (PRJNA975207); Altai Neanderthal (PRJEB1265); Vindija Neanderthal (PRJEB21157); Chagyrskaya Neanderthal (<http://ftp.eva.mpg.de/neandertal/Chagyrskaya/>); Denisovan (PRJEB3092); and chimpanzee (AG18354) whole-genome assembly (PRJNA916736). In addition, centromere sequences and annotations for the 65 whole-genome assemblies containing repaired centromeres (HGSVC Phase 3) as well as 227 whole-genome assemblies (HPRC Release 2) and 17 CEPH 1463 whole-genome assemblies containing repaired centromeres consented for public release are available from figshare (<https://doi.org/10.6084/m9.figshare.29921990>). Finally, the centromeres described in this study can be visualized on www.cencyclopedia.com, an interactive and dynamic centromere browser.

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

We analyzed 320 human samples, including 159 males (46,XY) and 161 females (46,XX).

Reporting on race, ethnicity, or other socially relevant groupings

We report analyses of publicly available human genome assemblies generated by the Human Genome Structural Variation Consortium (HGSVC), Human Pangenome Reference Consortium (HPRC), and the Platinum Pedigree Consortium and their associated population groups, as established and described by the 1000 Genomes Project (<https://www.internationalgenome.org/category/population/>).

Population characteristics

The individuals included in the study represent 28 human populations, included 26 defined by the 1000 Genomes Project (ACB, ASW, ESN, GWD, LWK, MSL, YRI, CLM, MXL, PEL, PUR, CDX, CHB, CHS, KHV, JPT, CEU, FIN, GRB, IBS, TSI, BEB, GIH, ITU, PJL, and STU) and 2 defined by the International HapMap Consortium and Human Genome Structural Variation Consortium (MKK and ASK).

Recruitment

Individuals included in this study belong to the following population groups: African Caribbean in Barbados (ACB); African Ancestry in Southwest USA (ASW); Esan in Nigeria (ESN); Gambian in Western Division – Mandinka (GWD); Luhya in Webuye, Kenya (LWK); Maasai in Kinyawa, Kenya (MKK); Mende in Sierra Leone (MSL); Yoruba in Ibadan, Nigeria (YRI); Colombian in Medellin, Colombia (CLM); Mexican Ancestry in Los Angeles, California, USA (MXL); Peruvian in Lima, Peru (PEL); Puerto Rican in Puerto Rico (PUR); Chinese Dai in Xishuangbanna (CDX); Han Chinese in Beijing, China (CHB); Han Chinese South, China (CHS); Kinh in Ho Chi Minh City, Vietnam (KHV); Japanese in Tokyo, Japan (JPT); Ashkenazi (ASK); Utah residents (CEPH) with Northern and Western European ancestry (CEU); Finnish in Finland (FIN); British from England and Scotland (GBR); Iberian Populations in Spain (IBS); Toscani in Italia (TSI); Bengali in Bangladesh (BEB); Gujarati Indians in Houston, Texas, USA (GIH); Indian Telugu in the UK (ITU); Punjabi in Lahore, Pakistan (PJL); and Sri Lankan Tamil in the UK (STU).

Ethics oversight

The HPRC and HGSVC lymphoblastoid cell lines are available from the Coriell Institute for Medical Research for research purposes and are covered by the appropriate ethics approvals by the Coriell Institute. The Platinum Pedigree cells lines are approved and overseen by the Institutional Review Board of the University of Utah under IRB_00065564.

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	We analyzed 65 human whole-genome assemblies generated by the HGSVC, 227 human whole-genome assemblies generated by the HPRC, 28 human whole-genome assemblies belonging to the Platinum Pedigree, and 1 chimpanzee whole-genome assembly generated by the Telomere-to-Telomere (T2T) Consortium. Sample size was defined based on the availability of consented participants in the respective research studies..
Data exclusions	No data were excluded.
Replication	Computational experiments are deterministic and are, therefore, reproducible. Each wet-lab experiment was performed at least two independent times successfully.
Randomization	N/A: this was not an interventional trial.
Blinding	N/A: this was not an interventional trial.

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

Antibodies

Antibodies used	Mouse monoclonal anti-CENP-A (3-19) antibody (ADI-KAM-CC006, Enzo; 1:50 dilution) Rabbit polyclonal anti-IgG antibody (ab171870, Abcam; 1:50 dilution) Rabbit polyclonal anti-CENP-C antibody (ABE1957, Millipore; 1:100 dilution) Goat anti-rabbit Alexa Fluor Plus 647 antibody (A32733TR, ThermoFisher Scientific; 1:500 dilution)
Validation	The anti-CENP-A antibody was generated against a synthetic peptide consisting of amino acids 3-19 of CENP-A, and mutation of this epitope in human cells prevents antibody binding (Logsdon et al., JCB, 2015). The anti-IgG antibody lacks specificity for known cellular antigens and has been validated by the manufacturer for ChIP applications via Western blot. The anti-CENP-C antibody was tested in both wild-type and CENP-C knockdown HeLa cells and showed time-dependent loss following CENP-C knockdown (Falk et al., Science, 2015).

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Human cell lines HG00512, HG00513, HG00514, HG02769, HG03065, and GM20355 were purchased from the Coriell Institute for Medical Research (Camden, NJ).
Authentication	Human cell lines HG00512, HG00513, HG00514, HG02769, HG03065, and GM20355 are part of the NHGRI Sample Repository for Human Genetic Research at the Coriell Institute for Medical Research (Camden, NJ) and was authenticated using a multiplex PCR assay with six autosomal microsatellite markers.
Mycoplasma contamination	Human cell lines HG00512, HG00513, HG00514, HG02769, HG03065, and GM20355 are negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

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, mosaicism, off-target gene editing) were examined.