

## 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](#).

### 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   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> 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) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> 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

Only sequencing data was analyzed (Illumina/Pacific Biosciences), so no specific software for data collection was used.

Data analysis

Software and workflows developed/adapted for this study (SaaRclust, breakpointR, StrandPhaseR, Snakemake Workflow) are all available under MIT license from the respective public repositories (<https://github.com/daewoooo/SaaRclust>, <https://bioconductor.org/packages/breakpointR>, <https://github.com/daewoooo/StrandPhaseR> devel branch, <https://github.com/ptrebert/project-diploid-assembly>). For the presented analyses, the v0.1dev has been used (<https://github.com/ptrebert/project-diploid-assembly/tree/v0.1dev>).

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. 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

HiFi PacBio reads for HG00731, HG00732, and HG00733 were produced as part of this study and are available from the IGSR FTP ([ftp://1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/HGSVC2/working/20190925\\_PUR\\_PacBio\\_HiFi/](ftp://1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HGSVC2/working/20190925_PUR_PacBio_HiFi/)). Strand-seq data for HG00733 were downloaded from NCBI SRA (BioProject PRJEB12849). Illumina short reads for HG00733 were downloaded from NCBI SRA (BioProject PRJEB9396). The genome assemblies produced in this study are available from [ftp://1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/HGSVC2/working/20200417\\_Marschall-Eichler\\_NBT\\_hap-asm/](ftp://1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HGSVC2/working/20200417_Marschall-Eichler_NBT_hap-asm/).

## 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	<input type="text" value="One genome was selected for assembly. The parent genomes were assembled for quality control."/>
Data exclusions	<input type="text" value="No data was excluded"/>
Replication	<input type="text" value="Not applicable."/>
Randomization	<input type="text" value="Not applicable."/>
Blinding	<input type="text" value="Not applicable."/>

## 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

### Methods

n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	<input type="text" value="All cell lines used in this study (HG00732, HG00731, HG00733) can be obtained from the Coriell Biorepository."/>
Authentication	<input type="text" value="SNV data was compared to previous sequencing data from Chaisson et al. (Nature Communications, 2019), which confirmed the identity of the sample."/>
Mycoplasma contamination	<input type="text" value="Cell lines have been tested before shipping by Coriell. We did not perform further in house tests."/>
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	<input type="text" value="n/a"/>