nature research

Corresponding author(s): Heng Li, Benjamin Neale and Daniel MacArthur

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	No software was used for data collection.
Data analysis	BWA-0.7.15, Bowtie2-2.2.2, minimap2-2.5, FreeBayes-1.0.2, Platypus-0.8.1, Samtools-1.3, GATK-3.5, RTG-2.8.4, htsbox-r346 and Bedtools-2.20.1 were used for data analysis.

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

Illumina reads from this study were deposited to ENA under accession PRJEB13208. Syndip variant calls and confident regions can be acquired from https://github.com/lh3/CHM-eval. Numerical data and gnuplot script to generate Figure 2 are available from the same GitHub repository.

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

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	Only two cell lines meet our requirements: 1) almost homozygous across the whole genome and 2) assembled with PacBio reads
Data exclusions	No data were excluded from the analysis.
Replication	We sequenced four replicates and confirmed that variance in library construction and sequence quality did not affect our general conclusion.
Randomization	No randomization was applied because we did not partition the data.
Blinding	Not applicable because we did not partitioned the data.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Research animals
☒	☐ Human research participants

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	CHM1 and CHM13 DNA was acquired from corresponding cell lines cultivated at the Eichler lab in the University of Washington.
Authentication	Data sequenced by us match public sequence data from NCBI.
Mycoplasma contamination	No Mycoplasma contamination observed in sequence data.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging