nature research

Corresponding author(s): David R. Rank, Michael W. Hunkapiller

Last updated by author(s): Jun 18, 2019

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
☐	☒ 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	Sequencing data was collected using Pacific Biosciences Sequel Instrument Control Software (ICS) (version 6.0.0) , on instrument Primary analysis (version 6.0.0), and CCS (version 3.0.0).
Data analysis	pbmm2 0.10.0; minimap2 2.14-r883; bedtools v2.27.1; samtools 1.3.1; daligner commit 381fa920; GATK 4.0.6.0; DeepVariant 0.7.1; WhatsHap v0.17; hap.py 0.3.10; RTG Tools 3.8.2; pbsv 2.1.0; Sniffles 1.0.10; SURVIVOR 1.0.3; Manta 1.4.0; Delly 0.7.6; Truvari commit 600b4ed7; Canu 1.7.1; FALCON kit 1.2.0; FALCON 0.7; wtdbg2 2.2; segDupPlots commit d34df78e; Arrow 2.2.2; BUSCO 3.0.2; custom scripts available at https://github.com/PacificBiosciences/hg002-ccs/

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

Data is available in NCBI BioProject PRJNA529679. CCS reads are available on NCBI SRA with accession SRX5327410. Small variant calls are available on NCBI dbSNP with accessions ss3783301452-ss3798736595. Structural variant calls are available on NCBI dbVar with accession nstd167. The trio binned Canu assemblies are available on NCBI Assembly with accessions GCA_004796485.1 (maternal) and GCA_004796285.1 (paternal). Alignments to GRCh37 are available at ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/AshkenazimTrio/HG002_NA24385_son/PacBio_CCS_15kb/ or https://bit.ly/2RW1b3I. Additional data, including all assemblies and a track hub for the UCSC Genome Browser, is available at https://downloads.pacbcloud.com/public/publications/2019-HG002-CCS.

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	One human genome was sequenced and analyzed. This sample has been extensively sequenced on several platforms and is a NIST Genome in a Bottle benchmark sample.
Data exclusions	All data that passed quality filters in the manufacturer's data collection pipeline was used in the study. No data was excluded.
Replication	To assay reproducibility and quality, data analysis was performed on subsampled data as described in the manuscript and is consistent across groupings.
Randomization	There is no allocations into groups as there was only one sample sequenced.
Blinding	Blinding was not required for this study as there is only one sample. For DeepVariant (a deep learning algorithm), training and model selection were performed without considering chromosome 20, and then models were evaluated for chromosome 20.

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	One 'personally identifying genetic information (PIGI)' consented human sample obtained from NIST (NA24385/HG002) was sequenced. The cell line for this sample was created from a 45 yr (at sampling) male. Remarks on the sample are: Participant (huAA53E0) in the Personal Genome Project: http://www.personalgenomes.org. History of blue rubber bleb nevus syndrome; central serous chorioretinopathy; cystoid macular degeneration; hemangioma; migraine with aura; narcolepsy; sleep paralysis; same subject as GM26105 (stem cell); mother is GM24143 (Lymph) / GM26077 (stem cell); father is GM24149 (Lymph).
Recruitment	The sample was recruited by the Personal Genome Project: http://www.personalgenomes.org and is distributed by the National Institute of Standards and Technology. The sample has been consented as mentioned above.
Ethics oversight	Ethics oversight is managed by the Personal Genome Project.

Note that full information on the approval of the study protocol must also be provided in the manuscript.