

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 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
<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. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P 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 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 No software was used

Data analysis Scripts used to develop the CMRG benchmark and generate figures and tables for the manuscript are being made available at <https://github.com/usnistgov/cmrg-benchmarkset-manuscript>. The assembly previously developed, which was used as the basis of this benchmark, was from hifiasm v0.11.

A variety of open source software was used for variant calling for the evaluations of the benchmark: NextDenovo2.2-beta.0, DRAGEN 3.6.3, NeuSomatic's submission for the PrecisionFDA truth challenge v212, [BWA-MEM48 version 0.7.17-r1188] (<https://github.com/lh3/bwa>), [GATK version gatk-4.1.4.1] (<https://gatk.broadinstitute.org/hc/en-us>); Parabricks_DeepVariant ([Parabricks Pipelines DeepVariant v3.0.0_2] (<https://developer.nvidia.com/clara-parabricks>)); Sentieon (DNAscope) version sentieon_release_201911 (<https://www.sentieon.com/products/#dnaseq>); BWA-MEM+Strelka2 ([BWA-MEM version 0.7.17-r1188] (<https://github.com/lh3/bwa>) + [Strelka2 version 2.9.10] (<https://github.com/Illumina/strelka>), BWA-Mem 48(ver. 0.7.8), Picard tools (<https://broadinstitute.github.io/picard/>) (ver. 1.83), GATK50 (ver. 3.4-0), GATK (ver. 3.5), BWA-MEM v0.7.15-r1140, SAMtools51 v1.3, Picard v2.10.10, GATK v3.8, DELLY52 v0.8.5, GRIDSS53 v2.9.4, LUMPY54 v0.3.1, Manta55 v1.6.0, and Wham56 v1.7.0, NanoPlot58 v1.27.0, Filtlong v0.2.0, minimap238,58 v2.17-r941, cuteSV v1.0.8, Sniffles59 v1.0.12, SURVIVOR 57 v1.0.7, BWA v0.7.15, GATK v3.6, Java v1.8.0_74 (OpenJDK), Picard Tools v2.6.0, Sambamba61 v0.6.7, Samblaster62 v0.1.24, Samtools v1.9, DeepVariant v1.0, LiftOff32 v1.4.0.

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 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

The PacBio HiFi reads used to generate the hifiasm assembly for the benchmark are in the NCBI SRA with accessions SRR10382245, SRR10382244, SRR10382249, SRR10382248, SRR10382247, and SRR10382246. The v1.00 benchmark VCF and BED files, as well as Liftoff gene annotations, assembly-assembly alignments, and variant calls, are available at https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/AshkenazimTrio/HG002_NA24385_son/CMRG_v1.00/, and as a DOI at <https://doi.org/10.18434/mds2-2475>. This is released as a separate benchmark from v4.2.1 because it includes a small fraction of the genome, it has different characteristics from the mapping-based v4.2.1, and v4.2.1 only includes small variants. Using v4.2.1 and the CMRG benchmarks as two separate benchmarks enables users to obtain broader performance metrics for most of the genome and for a small set of particularly challenging genes, respectively. The masked GRCh38 reference, recently updated to version 2 with additional false duplications from the Telomere to Telomere Consortium, is under <https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/references>. We recommend using v3.0 GA4GH/GIAB stratification bed files intended for use with hap.py when benchmarking, available under <https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/genome-stratifications/>. These stratifications include bed files corresponding to false duplications and collapsed duplications in GRCh38. All data have no restrictions, as the HG002 sample has an open consent from the Personal Genome Project.

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	This benchmark was created for a single sample because only one sample with all the data needed was available
Data exclusions	No data were excluded
Replication	Replication was not needed here because the goal of the study was to establish a benchmark, and this benchmark was evaluated by 11 independent methods and curators
Randomization	Randomization was not relevant since this was a single sample benchmark.
Blinding	Blinding was not relevant or possible because this is a widely used benchmark sample, and all Genome in a Bottle data are open and public as soon as possible

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s) For the 10x Genomics and Oxford Nanopore sequencing and Bionano mapping, the GM24385 (RRID:CVCL_1C78) cell line was

Cell line source(s)	obtained from the Coriell Institute for Medical Research National Institute for General Medical Sciences cell line repository. For the Illumina and Pacific Biosciences sequencing, NIST RM 8391 DNA was used, which was prepared from a large batch of GM24385 to control for differences arising during cell growth. For binning reads into paternal and maternal haplotypes, Illumina sequencing of DNA from NIST RM 8392 (HG002-HG004) was used. DNA was extracted from cell lines publicly available as GM24149 (RRID:CVCL_1C54) and GM24143 (RRID:CVCL_1C48) at the Coriell Institute for Medical Research National Institute for General Medical Sciences cell line repository.
Authentication	Variants from whole genome sequencing were compared to previous variant calls from this cell line
Mycoplasma contamination	Cell lines tested negative for mycoplasma by Coriell prior to aliquotting NIST Reference Material 8391
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study