nature research

Corresponding author(s): Steven A. McCarroll

Last updated by author(s): Mar 12, 2020

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	All data was collected via an Illumina NovaSeq, which generated sequencing BCL files. All subsequent data processing is outlined below.
Data analysis	Picard Tools v2.2 was used for sequence data processing (http://broadinstitute.github.io/picard/). BWA-MEM v0.7.7-r441 was used for alignment (http://bio-bwa.sourceforge.net/). GATK v3.7 was used to call variants (https://gatk.broadinstitute.org/). Custom code in Drop-seq Tools v2.2 was used to format sequencing data to include single-cell barcode information and to call heterozygous data in sperm genomes (https://github.com/broadinstitute/Drop-seq/releases). Genome STRiP v2.0 was used to ascertain read depth for sperm cells (http://software.broadinstitute.org/software/genomestrip/). HapCUT v1 was used for phasing (https://github.com/vibansal/hapcut). Custom code was written in R (www.r-project.org) for this study for data processing and to call and analyze recombination and aneuploidy events; it is available via Zenodo at http://dx.doi.org/10.5281/zenodo.2581595

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

Crossover and aneuploidy data (individual events and counts per donor and/or cell), including source data underlying Figs. 2, 3b-e and Extended Data Figs. 5-9, are available via Zenodo, http://dx.doi.org/10.5281/zenodo.2581570. Raw sequence data are available in the SRA via dbGaP for general research use upon application and approval (study accession number phs001887.v1.p1).

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	We aimed to sequence ~1,000 sperm cells per individual donor to be able to detect rare events (aneuploidy occurs at a frequency of less than a few percent), and the actual number of cells sequenced per donor was a result of experimental fluctuations during library preparation. With this amount of sequencing per donor as a baseline, we then chose to sequence cells from 20 individuals to be able to observe inter-individual differences and to be able to detect trends with moderate power (for example, an analysis of 20 individuals is 80% powered to detect a Pearson's r of 0.58 at p = 0.05).
Data exclusions	No data were excluded from the analyses. For certain analyses, subsets of the data were used to examine questions about those subsets.
Replication	As this was a completely new data set, we did not seek to replicate our findings with a data set of the same type. We did compare our findings of crossover location and frequency to published estimates generated with other methods; our observations were consistent with earlier estimates (detailed in manuscript).
Randomization	N/A - we did not have experimental groups.
Blinding	N/A - we did not have experimental groups.

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	The 20 sperm sample biospecimens analyzed in this study came from from 20 anonymous, karyotypically normal sperm donors at the New England Cryogenic Center (NECC). Per sperm bank policy, donors are 18-38 years old at the time of donation; precise age of donors is not released to researchers. The donors are known only to NECC and not to the researchers. Donor identifiers used in the paper were created specifically for this study and are not linked to any external identifiers.
Recruitment	The biospecimens used were discarded samples from New England Cryogenic Center; the samples had been collected during sperm donation. Donors consented, at the time of donation, for biospecimens to also be used for research purposes. NECC routinely provides such samples to researchers for research that has been IRB-approved at the researchers' home institution. The sperm donors might be subject to self-selection biases such as need for extra income and free time availability for donation, though any specific biases are unknown. If recombination or aneuploidy were correlated with any sperm donor recruitment or self-selection biases, the results of this study would reflect these underlying correlates. For example, the results are only reflective of donors aged 18-38 who had enough free time to act as sperm donors; if individuals of older ages or less free time had different meiotic phenotypes, they would not be captured in this study.
Ethics oversight	The Harvard Faculty of Medicine Office of Human Research Administration reviewed the research protocols (protocols M23743-101 and IRB16-0834) and determined that this research was "Not human subject research", based on the use of discarded biospecimens and the fact that researchers did not have interactions with the biospecimen donors. Harvard has also reviewed and approved our deposition of the data into an NIH respository; dbGaP will assure that access is provided only to researchers with legitimate research uses who agree not to try to re-identify donors based on data in the study.

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