nature portfolio

Corresponding author(s): Marta Melé
Roderic Guigó

Last updated by author(s): Oct 21, 2025

Reporting Summary

Nature Portfolio 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 Portfolio 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 *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 platforms used to generate the long-read sequencing data: Oxford Nanopore Technology (ONT) GridION platform. Basecalling performed in Nvidia Hopper GPUs with CUDA (v12.2) using ONT's dorado tool (v0.5.3). MinKnow (software version 23.11.7)
Data analysis	duplex-tools (v0.3.3) blast (v2.11.0) minimap2 (v2.28) UMI-tools (v1.0.0) porechop (v0.2.4) FLAIR (v2.0.0) ESPRESSO (v1.4.0) IsoQuant (v3.4.1) LyRic (v1.0.4) SQANTI3 (v5.2.1) lr-kallisto (v0.51.1) PyRanges (v0.0.129) ORFanage (v1.2.0) CPAT (v3.0.5) SQANTI Protein (v5.2.1) BLAST protein (2.11.0) VEP (v103)

kb-python (0.28.2)
VCFtools (v0.1.16)
BEDtools (v2.30.0)
gffread (v0.12.8)
R (v4.4.0)
python (v3.7.12)

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All supplementary tables have been deposited in Zenodo at DOI [10.5281/zenodo.14216603](https://doi.org/10.5281/zenodo.14216603).
The lrRNA-seq basecalled data and preprocessed FASTQ (>Q10) data have been deposited in the Array Express database at accession code [E-MTAB-14935](https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-14935).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Dataset is sex balanced. We report the lack of a sex-biased reference gene annotation.
Reporting on race, ethnicity, or other socially relevant groupings	This study refers to groups of individuals with similar genetics and shared recent evolutionary history, as "populations" or "ancestries" interchangeably. Labels for these groups were adapted from the 1000 Genomes or the Coriell Institute for Medical Research cell lines descriptions, with Fig. 1 showing one of the geographical locations where members of each group are predominantly found today rather than original sampling locations74. These labels are not social or ethnic identifiers and have limitations, including the oversimplification of continuous genetic diversity and the admixture present in modern human populations74. Accordingly, labels like European, non-European, African and out-of-African refer to genetic descent of individuals and not to the continent where they were born.
Population characteristics	See previous box.
Recruitment	Our study relies on commercially available biological materials.
Ethics oversight	This study utilized commercially sourced human cell samples, for which institutional review board approval was not required. All experimental procedures were conducted in accordance with ethical guidelines and principles governing research involving commercially obtained human products.

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

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	Sample size of 45 chosen to have at least 4 or 5 samples per population, which is just over the minimum to conduct different kinds of statistical tests.
Data exclusions	2 samples were excluded because an error in their processing was detected. These two samples were resequenced later and are publicly available in ArrayExpress.
Replication	Population specificity was replicated in MAGE dataset. Many discovered transcripts replicated in other studies like CHESS3, ENCODE4 and long-reads GTEx

Randomization	Samples were randomized to prevent group-batch confounding.
Blinding	Because of randomization during batch processing of the samples, blinding was not required.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	All cell lines are human lymphoblastoid cell lines and were obtained from Coriell Institute for Medials Research biobank. GM18906 Female GM12878 Female GM10492 Male GM10494 Male GM22234 Female GM10493 Female GM22299 Male GM18774 Female GM18772 Male GM22300 Male GM10495 Male GM22305 Female GM22304 Female GM19352 Male HG03732 Male GM18486 Male GM12812 Male GM12829 Male GM18561 Male GM12778 Female GM19307 Male GM19328 Female GM19390 Female HG03729 Male HG03719 Female HG01567 Female HG01975 Female GM19446 Female GM18631 Female HG01952 Male GM18489 Female GM18542 Female HG00621 Male HG04216 Female GM19117 Male HG01928 Male GM24385 Male GM12273 Female GM19240 Female

HG04217 Female
HG02261 Male
HG02293 Female
GM19129 Female
GM10496 Female
GM19093 Female
GM18489 Female
HG00621 Male
HG00621 Male
GM24385 Male
GM24385 Male
GM24385 Male
GM12273 Female
GM19240 Female
HG04217 Female
HG02261 Male
HG02261 Male
HG02261 Male
HG02293 Female
GM19129 Female
GM10496 Female
GM19093 Female

Authentication: Cell lines were not formally authenticated after purchase. We did check based on RNA-seq data their sex.

Mycoplasma contamination: All cell lines tested negative for Mycoplasma contamination

Commonly misidentified lines (See ICLAC register): *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Plants

Seed stocks: *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes: *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication: *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*