Corresponding author(s): Dr. Gloria M. Sheynkman
Dr. David E. Hill

Last updated by author(s): 4/7/2020

# 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<br>*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<br>*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 | Standard protocols for Illumina and PacBio sequencing were used - references cited in Methods |
| Data analysis | Software or programs used for data analysis: Iso-Seq 3 (bioinformatic pipeline available from PacBio), Cupcake (code available at https://github.com/Magdoll/cDNA_Cupcake, v11.0.0), SQANTI2 (code available at https://github.com/Magdoll/SQANTI2), Kallisto (v0.44.0), Bowtie2 (v2.2.3), SAMtools (v1.2), RSEM (v1.2.29), STAR (v2.4.0), SMRTAnalysis (v6.0), minimap2 (v2.11-r797), bedtools (v2.27.1), liftOver (no version, used in UCSC browser September 2nd, 2018), SDS (v2.4.1).<br>Custom Python and R scripts were used for post-result processing and are available upon reasonable request. |

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

The data supporting the findings from this study are included either in the manuscript or its associated supplementary files. Sequencing data has been deposited to the Sequence Read Archive (SRA) and can be found under BioProject PRJNA615244. The processed PacBio sequencing data (SQANTI2 output) have been deposited to the Zenodo database (https://zenodo.org). All data are also available from the authors upon request.
The CAGE peak data, which is publicly available from the FANTOM consortium, may be found using the following link:
http://fantom.gsc.riken.jp/5/datafiles/latest/extra/CAGE_peaks/hg19.cage_peak_phase1and2combined_coord.bed.gz

# 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 sizes were based on the number of available human ORF clones. Capture experiments were carried out with increasing numbers of probes to determine an optimal number of genes to target in pooled samples.<br>All capture experiments done with the purpose of comparing conditions were done in technical triplicate. Experiments done for the purpose of analytical benchmarking were performed in technical duplicate. Capture experiments performed for the purpose of isoform discovery was performed once. |
| Data exclusions | No sequence data was excluded from the analysis. |
| Replication | All attempts at replication were successful. |
| Randomization | Not applicable to our experimental design. |
| Blinding | Not applicable to our experimental design. |

# 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 |