

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
<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 for the collection of data, as this was an opportunistic study.

Data analysis All code to perform quality control and data analysis is provided in the following Github repos:
https://github.com/macarthur-lab/tx_annotation
https://github.com/macarthur-lab/gnomad_qc
https://github.com/macarthur-lab/gnomad_lof
<https://github.com/konradjk/loftree>
<https://github.com/broadinstitute/gtex-pipeline/>
Hail 0.2 is available at: <https://hail.is/>
RSEM v1.2.22 : <https://deweylab.github.io/RSEM/>
RNA-SeqQC v1.1.8 : <https://github.com/broadinstitute/rnaseqc>
STAR v2.4.2a : <https://github.com/alexdobin/STAR>
Alamut v.2.11 <https://www.interactive-biosoftware.com/alamut-visual/>
Variant Effect Predictor (VEP) v.85 : <https://uswest.ensembl.org/info/docs/tools/vep/index.html>
FUMA GENE2FUNC v1.3.5e : <https://fuma.ctglab.nl/>
R version 3.4.0 : <https://cran.r-project.org/bin/macosx/>

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

All datasets are described in the manuscript or Supplementary Information, including deposition of the full dataset at <https://gnomad.broadinstitute.org>. Data for specific analyses are available publicly at [gs://gnomad-public/papers/2019-tx-annotation/](https://gnomad-public/papers/2019-tx-annotation/) and the specific folders therein for analyses are referenced in the manuscript for ease of recreating analyses with the data provided. There are no restrictions on the aggregate data released.

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](https://www.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 study provides a framework and tool to improve variant interpretation in datasets of any size, no matter how small or large. As a proof of principle, we use one of the largest datasets of human genetic variation, gnomAD and the largest functional genomics dataset GTEx. In other words, this study is opportunistic, and involves secondary use of available genome, exome and transcriptome data. No sample size was predetermined.
Data exclusions	Sample QC and variant QC for gnomAD are described extensively in the supplementary methods of the main manuscript. Notably, individuals with severe pediatric disease, and known first disease relatives of those with severe pediatric disease were excluded. For the analyses in the manuscript, we removed GTEx tissues with low sample numbers, reproductive tissue and non-tissues (ie. cell lines). For the purpose of our manuscript, we did not define pre-exclusion criteria for calculation of pext. However for the analyses in the manuscript, we defined pre-exclusion tissues : we removed reproduction-associated GTEx tissues (endocervix, ectocervix, fallopian tube, prostate, uterus, ovary, testes, vagina), cell lines (transformed fibroblasts, transformed lymphocytes) and any tissue with less than one hundred samples (bladder, brain Cervicalc-1 spinal cord, brain substantia nigra, kidney cortex, minor salivary gland) . This is explained in the Methods section of the manuscript.
Replication	We did not attempt to reproduce any findings in a separate but identical dataset, as no other data set of comparable size exists. However we replicate key findings with a separate isoform quantification tool. We successfully replicate the MAPS results, shown in Supplementary Figure 11. We also use an external fetal dataset to provide additional data.
Randomization	As this was a population-based study, and not a case-control study, no randomization was performed.
Blinding	As this was a population-based study, and not a case-control study, blinding was not relevant.

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

As an opportunistic collection of data, the participants in this study were not selected based on age, gender, or genotypic information. As described above, individuals with severe pediatric disease, and known first disease relatives of those with severe pediatric disease were excluded from gnomAD. As an opportunistic collection of data, the participants in gnomAD were not selected based on age, gender, or genotypic information. The populations are provided in Supplementary Table 7 of the accompanying Karczewski et al., and there are 64,754 females and 76,702 males. These data were obtained primarily from case-control studies of adult-onset common diseases, including cardiovascular disease, type 2 diabetes, and psychiatric disorders.

GTEx v7 collection was similarly opportunistic and has been previously extensively published and reported on. Population characteristics of the data can be found in Reference 2 : GTEx Consortium et al., Genetic effects on gene expression across human tissues. Nature 550, 204 (2017).

Recruitment

As this was an opportunistic secondary use study, we did not recruit any participants.

Ethics oversight

This study was overseen by the Broad Institute's Office of Research Subject Protection and the Partners Human Research Committee, and was given a determination of Not Human Subjects Research.

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