

## 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

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 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) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 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><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | 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 in the collection of data for this study.
Data analysis	The following software was used in the analysis of data for this study: Python 3.8.17 (a full list of dependencies and Python libraries used in this study are shared in our publicly available Github repository), bedtools v2.31.1, BCFtools v1.20, Variant Effect Predictor (VEP) v105, Biorender.

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](#)

Regional nonsense constraint annotations and summary statistics are available in Supplementary Dataset 3, and for download at [https://github.com/alexblakes/regional\\_nonsense\\_constraint/blob/main/data/final/regional\\_nonsense\\_constraint.tsv](https://github.com/alexblakes/regional_nonsense_constraint/blob/main/data/final/regional_nonsense_constraint.tsv). Regional constraint annotations are also publicly available through the DECIPHER portal [URL to be added]. Source data are provided with this paper. For Figure 6, export of Source Data from the Genomics England Research

Environment is not permitted. These data are available to registered users of the National Genomic Research Library at [/re\\_gecip/shared\\_allGeCIPs/AlexBlakes/regional\\_nonsense\\_constraint/figure\\_6.tsv](#).

Data from the National Genomic Research Library (NGRL) used in this research are available within the secure Genomics England Research Environment. Access to NGRL data is restricted to adhere to consent requirements and protect participant privacy.

Access to NGRL data is provided to approved researchers who are members of the Genomics England Research Network, subject to institutional access agreements and research project approval under participant-led governance. For more information on data access, visit: <https://www.genomicsengland.co.uk/research>

## 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	No sex- or gender-based analyses were performed.
Reporting on race, ethnicity, or other socially relevant groupings	No analyses based on race, ethnicity, or other groupings were performed.
Population characteristics	<p>This study used sequencing data from participants in:</p> <ul style="list-style-type: none"> <li>- The gnomAD study v4.1 (<a href="https://gnomad.broadinstitute.org/about">https://gnomad.broadinstitute.org/about</a>)</li> <li>- Rare disease cohorts, including: <ul style="list-style-type: none"> <li>- Individuals recruited to the 100,000 Genomes Project</li> <li>- Individuals recruited to the NHS Genomic Medicine Service</li> <li>- Individuals recruited to sequencing studies through the Deciphering Developmental Disorders study, GeneDx, and Radboud University Medical Center.</li> </ul> </li> </ul>
Recruitment	Participants in the rare disease cohorts were recruited on the basis that they or a family member were affected by a rare, presumably genetic, condition or developmental disorder.
Ethics oversight	<p>The aggregation and release of summary data from the exomes and genomes collected by the Genome Aggregation Database has been approved by the Mass General Brigham IRB (protocol 2013P001339, "Large-scale aggregation of human genomic data").</p> <p>Genomics England has approval from the HRA Committee East of England – Cambridge South (REC Ref 14/EE/1112).</p>

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](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	Sample size was determined by the number of participants belonging to each cohort.
Data exclusions	De novo variants with any possibility of duplication between cohorts were conservatively excluded from the study, as described in the methods. No other data were excluded.
Replication	Data were produced using the largest / most highly powered datasets available at present, and can be expanded as cohort sizes increase.
Randomization	Randomisation was not applicable to this study because participants were not subject to any treatment or exposure.
Blinding	Blinding was not possible or applicable for this retrospective analysis.

## 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 &amp; experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A