nature portfolio

Corresponding author(s): Peter Krawitz

Last updated by author(s): March 11, 2022

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	For individuals additionally participating in the PEDIA study, scores of the analysis of the portrait images by artificial intelligence (PEDIA) were collected via the PEDIA web service (https://www.pedia-study.org/).
Data analysis	All analysis that has been conducted on the data is described in the main manuscript and the supplement. The regarding code can be found in the publicly available github respository cited in the manuscript (https://github.com/Ax-Sch/TNAMSE_exomes). In particular the code for the lasso analysis (scripts/figure2_lasso_analysis_TNAMSE_cohort.R) which is the basis for the provided webservice (http://tnamse.de) as well as the code for the analysis of the pedia cohort (scripts/Pedia_Analysis_manusscript.R) can be found. The analysis were conducted in the statistics software R (version 4.0.5).

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 data shown in the manuscript can be found in the main manuscript, the supplement or in our public github repository (https://github.com/Ax-Sch/TNAMSE_exomes/tree/main/resources).

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	In total, 5652 individuals with a suspected rare disorder were enrolled in TRANSLATE-NAMSE by centers for rare diseases at ten German university hospitals over a period of three years (2018-2020). A cohort of 1577 patients underwent exome sequencing with recommendation of multidisciplinary teams. In this manuscript we report results for those 1577 patients of which 211 patients additionally consented to the analysis of their potraits by an artificial intelligence tool (PEDIA).
Data exclusions	No data were excluded.
Replication	All findings of the analysis can be reproduced by the provided code (https://github.com/Ax-Sch/TNAMSE_exomes). We validated the findings of the LASSO model on an external patient cohort (shown in the supplement).
Randomization	For the LASSO analysis the patients were randomly assigned to training (80%) and test (20%) data sets. All other analyses were conducted on the complete cohort.
Blinding	Splitting data into training and test sets is equivalent to blinding data for the machine.

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
☐	☒ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

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	Male and female as well as pediatric and adult participants were analyzed in the study. All patients underwent exome sequencing while a subset of patients additionally consented to image analysis of their potraits by artificial intelligence (PEDIA). Age (below or above 18 years), sex, the sequencing laboratory as well as the use of the PEDIA workflow could confound the diagnostic yield and the data were therefore included as confounders in the LASSO analysis.
Recruitment	Participants were enrolled in the TRANSLATE-NAMSE study at ten centers for rare diseases at German university hospitals and recommended to undergo exome sequencing by multidisciplinary teams.
Ethics oversight	This study is governed by the approval of the following Institutional Review Boards: Charité–Universitätsmedizin Berlin, Germany (EA2/140/17); UKB Universitätsklinikum Bonn, Germany (Lfd.Nr.386/17); Universitätsklinikum Essen, University Duisburg-Essen, Germany (17-7774-BO); Universitätsklinikum Heidelberg, Germany (S-499/2017); Universitätsklinikum Tübingen, Germany (643/2017BO1); Universität zu Lübeck, Germany (17-272); Ludwig-Maximilians-Universität München, Germany (17-640); Ärztekammer Hamburg, Germany (MC-316/17); Technische Universität Dresden, Germany (AK 464122017). The authors have obtained written informed consent from the patients or their guardians.

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Exome-sequencing in TRANSLATE-NAMSE was not a clinical trial, as no health-related interventions, according to the ICMJE definition, were conducted. Therefore, this study was not registered as a clinical trial.
Study protocol	n/a because the study was not registered as a clinical trial
Data collection	see Recruitment in Human Research Participants and Data collection in Software and Code.
Outcomes	The outcome was defined as wheter a molecular diagnosis could be established or not per case.