nature portfolio

Corresponding author(s): Jan Krumsiek

Last updated by author(s): May 7, 2024

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	No software was used
Data analysis	All code used in the analysis of this paper can be found at https://github.com/krumsieklab/autofocus. This code was created with R version 4.2.2. The following R packages are required to run the AutoFocus scripts: cluster (v 2.1.6) dendextend (v 1.17.1) doParallel (v 1.0.17) dplyr (v 1.1.4) DT (v 0.33) foreach (v 1.5.2) ggplot2 (v 3.5.0) glmnet (v 4.1-8) htmlwidgets (v 1.6.4) igraph (v 1.5.1) magrittr (v 2.0.3) mgm (v 1.2-14) networkD3 (v 0.4) parallel (v 4.2.2) plotly (v 4.10.4) RColorBrewer (v 1.1-3) RhpcBLASctl (v 0.23-42) (

reshape2 (v 1.4.4)
shiny(v 1.8.1.1)
shinyalert (v 3.0.0)
shinydashboard (v 0.7.2)
SummarizedExperiment (v 1.28.0)
tidyverse (v 2.0.0)

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

The preprocessed, concatenated QMDiab dataset used in this paper can be found at https://doi.org/10.6084/m9.figshare.23934933.v1.
The ROS/MAP data used in this paper can be obtained from two sources: (1) Metabolomics and proteomics data for the ROS/MAP cohort are available via the AD Knowledge Portal (https://adknowledgeportal.org). The AD Knowledge Portal is a platform for accessing data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. The data, analyses, and tools are shared early in the research cycle without a publication embargo on secondary use. Data is available for general research use according to the following requirements for data access and data attribution (https://adknowledgeportal.org/DataAccess/Instructions). For access to content described in this manuscript see: http://doi.org/10.7303/syn26401311. (2) The full complement of clinical and demographic data for the ROS/MAP cohort are available via the Rush AD Center Resource Sharing Hub and can be requested at https://www.radc.rush.edu.

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	Sex was self-reported from participants in both the QMDiab and ROSMAP datasets. The QMDiab dataset had 192 females had 196 males, and the ROSMAP dataset had 352 females and 148 males. The variable was used as a covariate in the regression analyses of metabolites to both Type 2 Diabetes diagnosis and Alzheimer's Diagnosis.
Reporting on race, ethnicity, or other socially relevant groupings	The QMDiab study population was predominantly of Arab, South Asian, and Filipino descent. The ROSMAP dataset consisted mostly of white, non-Hispanic individuals. These factors (race, ethnicity, country of origin) did not factor into any of our models.
Population characteristics	Our models included by age and BMI as confounding variables as both characteristics have independent relationships with both Type 2 Diabetes and Alzheimer's disease.
Recruitment	ROS: The Religious Orders Study enrolled Catholic nuns, priests and brothers, from more than 40 groups across the United States. As such, they were altruistic and have a history of participating in research projects from which they may derive little to no personal benefit. They lived communally and loss of contact with participants was rare, facilitating the high follow-up and autopsy rates required to ensure internal study validity. Finally, the participants had similar education, socioeconomic and life experiences for most of their adult lives. This allowed for tighter control of these potentially confounding variables in analyses of incident AD and cognitive decline. MAP: Study participants were primarily recruited from retirement communities throughout northeastern Illinois. The study primarily enrolled residents of continuous care retirement communities. Several features of these facilities and the study design enhance dthe validity and generalizability of the study. Because the only exclusion was the inability to sign the Anatomical Gift Act, and because all clinical evaluations are performed as home visits, co-morbidities common in population-based epidemiologic studies were well represented; this reduced the “healthy volunteer effect” seen in many cohort studies.
Ethics oversight	Institutional Review Boards of Hamad Medical Corporation (HMC) in Doha, Qatar and Weill Cornell Medicine-Qatar (WCM-Q). Institutional review board of Rush University Medical center

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

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	n = 388 in the QMDiab cohort n = 500 in the ROSMAP cohort
Data exclusions	Pre-exclusion criteria: in QMDiab, if >20% of metabolite measurements were missing, the sample was excluded. In ROSMAP, if >25% of the metabolite measurements were missing, the sample was excluded
Replication	Datasets containing multiple omic platforms, especially metabolomic and proteomic, are incredibly rare to find. QMDiab and ROSMAP are one of a kind datasets, and replication requires datasets of similar makeup. Due to the lack of these types of datasets, replication is next to impossible.
Randomization	Data comes from observational datasets, not controlled trials. Therefore randomization does not apply here.
Blinding	Blinding was not necessary to the observational datasets.

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

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A