nature portfolio

Corresponding author(s): Pieter C. Dorrestein
Wout Bittremieux

Last updated by author(s): 2023/11/14

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 for data collection.
Data analysis	Data analysis was performed using the online GNPS platform (https://gnps.ucsd.edu). Custom analysis code was implemented in Python 3.8, and uses NumPy (version 1.19.2), SciPy (version 1.5.2), Pandas (version 1.1.3), and statsmodels (version 0.13.1) for scientific data processing, Pyteomics (version 4.4.0) to interface the UNIMOD repository, and matplotlib (version 3.5.1), Seaborn (version 0.11.0), spectrum_utils (version 0.3.4), Jupyter notebooks, and Cytoscape (version 3.9.1.) for visualization purposes. The corresponding code is available on GitHub at https://github.com/bittremieux/gnps_suspect_library and on Zenodo at https://doi.org/10.5281/zenodo.6459282 under the open source BSD-3-Clause license.

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

Source data are provided with this paper.

All of the data involved in this work are publicly available through GNPS/MassIVE:

GNPS living data molecular networking

- GNPS living data (version November 17, 2020): https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=25cc4f9135c6428aabe1f41a9e54c369&view=advanced_view
- Living data global molecular network: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=4f69e11bfb544010b2c4225a255f17ba

Spectrum annotation using the nearest neighbor suspect spectral library

- Spectral library searching using the default GNPS libraries only:
  - Part 1: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=308b3393a2b2401e8c9b562152531b4c
  - Part 2: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=18cf4e521f9b4124af54d7e3d837a888
  - Part 3: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=c0249eb6a52e4ea993b03de90a509b35
  - Part 4: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=debd3bbb51f6490394e905e13779f295
  - Part 5: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=8cdb4d7d1a784f5bb4f99e4c31564cd1
  - Part 6: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a9e7e4b1b8104416a39142fd6072e02a
  - Part 7: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=334ed0d944844e90b71d6151d4e74263
  - Part 8: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b55aef34c0bd4d78a1f3952f7c49a52c
- Spectral library searching using the default GNPS spectral libraries and the nearest neighbor suspect spectral library:
  - Part 1: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=064be855f46e407f9f5fcbe652c8b9d5
  - Part 2: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=d243afb8f233490886bb8ab5eedcf8b8
  - Part 3: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=febab54db7a14af6b451ab5e5789785f
  - Part 4: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=eba0dfe63a464b0a924fd5e373917b37
  - Part 5: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=95b541cb3be54d08a0b14367554630ca
  - Part 6: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=1df48f2dc7c443fc9364dfc8b28f6b47
  - Part 7: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b7f8c3d47a464b53ab94f1780f56c893
  - Part 8: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=50e3d8ae4e004f989862fcc9d1353534

Evaluation of suspect use cases

- Molecular networking of apratoxin suspects: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=5c41693f607d4b4cabbcfbbf5b9bcf86
- Molecular networking of azithromycin suspects: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=e91e2e44e3234f08bb3d7f3f16d5f782
- Molecular networking of flavonoid suspects: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=38a1bd60bd094c8a97cf49d822e7f853
- Molecular networking of home environment personal care products: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=890e39f28140470ab0598c77cc5c048e
- Spectral library searching of Alzheimer's disease data
  - Using the default GNPS spectral libraries only: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b55aef34c0bd4d78a1f3952f7c49a52c
  - Using the default GNPS spectral libraries and the nearest neighbor suspect spectral library: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=50e3d8ae4e004f989862fcc9d1353534

Additionally, all relevant data files have been deposited to a permanent Zenodo archive at https://doi.org/10.5281/zenodo.8282733.

Potential explanations for the observed delta masses were partially sourced from the UNIMOD database of protein modifications (https://www.unimod.org/), complemented with a community-curated list of delta masses (Supplementary Table 3).

Metabolomics and clinical data for the ROSMAP clinic cohorts are also available via the AD Knowledge Portal (https://adknowledgeportal.org) and through request to Dr. David Bennett at Rush University who provided brain samples used for analysis. 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: https://doi.org/10.7303/syn30255033.1.

The nearest neighbor suspect spectral library is freely available under the CC0 license at https://gnps.ucsd.edu/ProteoSAFe/gnpslibrary.jsp?library=GNPS-SUSPECTLIST and archived on Zenodo at https://doi.org/10.5281/zenodo.8282733. Additionally, it can be used for any data analysis task on GNPS by selecting it from the CCMS_SpectralLibraries > GNPS_Propogated_Libraries > GNPS-SUSPECTLIST > GNPS-SUSPECTLIST.mgf path in the GNPS file selector dialog. Step-by-step instructions are also provided on GitHub at https://github.com/bittremieux/gnps_suspect_library and on the GNPS Documentation website.

Individual spectra are accessible by their Universal Spectrum Identifiers (USIs). The spectra displayed in Figure 3, Figure 4, Supplementary Figure 5, and Supplementary Figure 6 are:

- Hexanoylcarnitine, C6:0: mzspec:GNPS:GNPS-LIBRARY:accession:CCMSLIB00003135669
- Hexenoylcarnitine, C6:1: mzspec:MSV000085561:011c:scan:2864
- Benzoylcarnitine: mzspec:MSV000085561:010c:scan:2829
- Dodecanedioylcarnitine, C12-DC: mzspec:MSV000082650:M031_48:scan:1501
- 3-Hydroxybutyrylcarnitine reference: mzspec:GNPS:TASK-015e9e338c5649a7af6715af2be98e2f-spectra/specs_ms.mgf:scan:1
- 3-Hydroxybutyrylcarnitine suspect: mzspec:MSV000082049:20_51:scan:106
- 3-Hydroxyhexanoylcarnitine reference: mzspec:GNPS:TASK-015e9e338c5649a7af6715af2be98e2f-spectra/specs_ms.mgf:scan:4
- 3-Hydroxyhexanoylcarnitine suspect: mzspec:MSV000085561:018b:scan:2609
- Apratoxin A: mzspec:GNPS:GNPS-LIBRARY:accession:CCMSLIB00000424840
- Apratoxin D: mzspec:GNPS:GNPS-LIBRARY:accession:CCMSLIB00000424841
- Apratoxin F: mzspec:GNPS:GNPS-LIBRARY:accession:CCMSLIB00000070287

- Apratoxin C: mzspec:MSV000086109:BF9_BF9_02_57124.mzML:scan:722
- Apratoxin A – 30.010 Da: mzspec:MSV000086109:BD5_dil2x_BD5_01_57213:scan:760
- Apratoxin F – 30.010 Da: mzspec:MSV000086109:BC11_dil2x_BC11_02_57176:scan:736
- Apratoxin A – 26.015 Da: mzspec:MSV000086109:BD5_dil2x_BD5_01_57213:scan:614
- Apratoxin A – 14.016 Da: mzspec:MSV000086109:BD11_BD11_02_57022:scan:591
- Azithromycin: mzspec:GNPS:GNPS-LIBRARY:accession:CCMSLIB00005434451
- 3'-O(desmethyl)azithromycin: mzspec:MSV000084132:Pos_C18_Aq7:scan:977
- Apigenin-8-C-hexosylhexoside: mzspec:GNPS:GNPS-LIBRARY:accession:CCMSLIB00004698180
- 7-O-methylapigenin-6-C-hexoside + 132.042 Da: mzspec:GNPS:TASK-38a1bd60bd094c8a97cf49d822e7f853-spectra/specs_ms.mgf:scan:1573560
- Apigenin-8-C-hexosylhexoside – 30.010 Da: mzspec:GNPS:TASK-38a1bd60bd094c8a97cf49d822e7f853-spectra/specs_ms.mgf:scan:1559636
- Apigenin-8-C-hexosylhexoside – 31.991 Da: mzspec:GNPS:TASK-38a1bd60bd094c8a97cf49d822e7f853-spectra/specs_ms.mgf:scan:1559563
- Apigenin-8-C-hexosylhexoside – 46.005 Da: mzspec:GNPS:TASK-38a1bd60bd094c8a97cf49d822e7f853-spectra/specs_ms.mgf:scan:1543689

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	N/A.
Reporting on race, ethnicity, or other socially relevant groupings	N/A.
Population characteristics	For the Alzheimer's disease acylcarnitine analysis, we used public untargeted metabolomics data (MassIVE dataset identifier MSV000086415) from 514 individuals with and without Alzheimer's disease (360 Alzheimer's disease patients, 154 healthy subjects).
Recruitment	Only publicly available data were used.
Ethics oversight	The original brain samples were acquired by Dr. David Bennett at Rush University.

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

Life sciences study design

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

Sample size	All publicly available untargeted metabolomics datasets on GNPS/MassIVE at the time of the research (November 2021) were included, consisting of 1335 public projects.
Data exclusions	No data was excluded.
Replication	Replication of the spectrum annotation results was performed once on a temporally independent collection of untargeted metabolomics datasets on GNPS/MassIVE that was deposited at a later data, consisting of 72 public projects. This independent replication achieved highly similar spectrum annotation rates.
Randomization	There was no allocation to different experimental groups.
Blinding	Not applicable as group allocation was not relevant during creation of the nearest neighbor suspect spectral library.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*

Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority,*

	the date of issue, and any identifying information).
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	*State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.*
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	*For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Reporting on sex	*Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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 ICMJEguidelines for publication of clinical research and a completedCONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session
(e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

(See Eklund et al. 2016)

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.