Corresponding author(s): Oscar Esteban <phd@oscaresteban.es>; Krzysztof J. Gorgolewski <krzysztof.gorgolewski@gmail.com>

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	This manuscript uses data publicly available from the OpenfMRI.org/OpenNeuro.org resource (i.e. availability of software used in data collection is responsibility of the original submitters to the repository). Data was collected using datalad version 0.9.1.
Data analysis	All the code, tests, and results of the analyses are available under open-source licenses (BSD-3-clause and MIT for software and CC0 for data derivatives). Preprocessing was conducted using fMRIPrep, versions 1.0.7 and 1.0.8. FMRIPrep uses tools such as AFNI, ANTs, FSL, FreeSurfer, ICA-AROMA, Nilearn, and Nipype. All versions of these software tools are specified in the Online Methods document and in Supplementary Note 3. Data analysis was carried out using FSL 5.0.8, Nilearn 0.4, and Nipype 1.0.

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/reviewers upon request. 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 original data used in this work are publicly available through the OpenNeuro.org platform (formerly, OpenfMRI). Derivatives generated with fMRIPrep in this work are available at https://s3.amazonaws.com/fmriprep/index.html. The expert ratings collected after visual assessment of all reports are available through FigShare (doi:10.6084/m9.figshare.6196994.v3).

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	Experiment 1 utilizes 325 participants, collected from 54 qualifying MRI studies in OpenfMRI. Four participants per qualifying study were selected at random, except for ds0000031 which only has one participant densely sampled. Analysis is fundamentally visual by experts and therefore no power calculation was necessary. Experiment 2 utilizes all 257 participants of ds000030 (OpenfMRI's accession number), establishing a comparative design between the proposed workflow (fMRIPrep) and a widely-adopted alternative (FSL FEAT) in preprocessing the data. Analysis is fundamentally visual, but also included an exploratory test-retest analysis of several measurements. Since the nature of the analysis was exploratory, N=257 was deemed sufficient.
Data exclusions	In Experiment 1 some participants that contained: a) data preprocessed in any way -e.g. skull-stripped T1-weighted MRI-, except defacing which is necessary to share datasets; or b) errors on the BIDS organization of the data and associated metadata. In Experiment 2, 15 participants, for which some image modality was missing (T1w or BOLD) or the task information was lost and inaccessible were removed. The rationale behind these exclusion criteria is that fMRIPrep requires a) unprocessed data; b) a BIDS valid structure for the input datasets; and c) the input dataset must have, at least, one T1w image and one BOLD run per participant. Thus, exclusion criteria were implicitly imposed by the presented tool prior to the start of the study.
Replication	Replication is tracked through continuous integration testing. Three datasets are preprocessed with fMRIPrep after every change done to the codebase, and the MD5 sums of final and intermediate results are checked for identity w.r.t. previous versions of the tool. When some changes break replication of results (i.e. with version changes), then the database of MD5 sums is updated manually by a developer to ensure that no accidental changes are done. Within version reproducibility (or run-to-run reproducibility) is ensured using container technology. Details are provided in the Methods section of the paper and Supplementary Note 4.
Randomization	In Experiment 1, four participants per qualifying study were selected at random, except for ds0000031 which only has one participant densely sampled. For the group-level statistical analysis of Experiment 2, we run an analysis of overlap of statistical maps extracted from disjoint groups, randomly sampled with increasing sample sizes starting from 5 through 120. The resampling was repeated 200 times per sample size.
Blinding	Blinding was not possible, as this study reuses publicly available data.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

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	not measured
Recruitment	opportunistic (publicly shared data)

Magnetic resonance imaging

Experimental design

Design type	The method presented in this manuscript is agnostic to the design type (task/rest, block/random).
Design specifications	The method presented in this manuscript requires, at least: a) unprocessed data; b) a BIDS valid structure for the input datasets; and c) the input dataset must have, at least, one T1w image and one BOLD run per participant.
Behavioral performance measures	not applicable

Acquisition

Imaging type(s)	functional MRI, structural MRI (T1-weighted and T2-weighted), field maps
Field strength	1.5T, 3T
Sequence & imaging parameters	As presented in the paper, the methods proposed identify the sequences and imaging parameters automatically to build up the processing work flow.
Area of acquisition	Whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	Preprocessing software is the core of this contribution. Thus, it is thoroughly described in the main text and methods section.
Normalization	Nonlinear spatial normalization is proposed and described in depth in the main text and methods section.
Normalization template	The methods presented allow to use any available template given certain specifications described in the documentation. By default, results are normalized to the ICBM152 Nonlinear Asymmetric 2009 version c. The software also uses the OASIS and the ICBM152 Linear Symmetric templates. Optionally, NKI template is also available.
Noise and artifact removal	Described in depth within the main text and methods section
Volume censoring	Estimated confounds allow the application of volume censoring, but we did not use volume censoring in our evaluation.

Statistical modeling & inference

Model type and settings	Activity maps per subject were estimated on the task data using a general linear model (GLM). For the one condition under comparison (go - successful stop) one task regressor was included with a fixed duration of 1.5s and an extra regressor was added with equal amplitude, but the duration equal to the reaction time. Again, these regressors were orthogonalized with respect to the fixed duration regressor of the same condition. Predictors were convolved with a double-gamma canonical hemodynamic response function. Temporal derivatives were added to all task regressors to compensate for variability on the hemodynamic response function. Furthermore, the six rigid-motion parameters (translation in 3 directions, rotation in 3 directions) were added as regressors to avoid confounding effects of head motion. We included a high-pass filter (100Hz). Subsequent to the single subject analyses, two random (non-overlapping) subsamples of n subjects were taken and entered into a second level analysis. We vary the sample size n between 10 and 120 (total was 257 subjects). This process is repeated 200 times. We analyzed the group data using

ordinary least squares (OLS) mixed modeling. Subsequently, we threshold the statistical maps, ensuring control of the False Discovery Rate (FDR).

Effect(s) tested

Test-retest reliability of probabilistic and binary indices of activation overlap between groups and Pearson correlation. The only task under analysis was the Stop Signal Task, and the only contrast analyzed was "go - successful stop".

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

Statistic type for inference
(See Eklund et al. 2016)

voxel-wise

Correction

FDR 5%

Models & analysis

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