nature methods

ARTICLES

https://doi.org/10.1038/s41592-018-0235-4

fMRIPrep: a robust preprocessing pipeline for functional MRI

Oscar Esteban[1*], Christopher J. Markiewicz[1], Ross W. Blair[1], Craig A. Moodie[1], A. Ilkay Isik[2], Asier Erramuzpe[3], James D. Kent[4], Mathias Goncalves[5], Elizabeth DuPre[6], Madeleine Snyder[7], Hiroyuki Oya[8], Satrajit S. Ghosh[5,9], Jessey Wright[1], Joke Durnez[1], Russell A. Poldrack[1,10] and Krzysztof J. Gorgolewski[1,10*]

Preprocessing of functional magnetic resonance imaging (fMRI) involves numerous steps to clean and standardize the data before statistical analysis. Generally, researchers create ad hoc preprocessing workflows for each dataset, building upon a large inventory of available tools. The complexity of these workflows has snowballed with rapid advances in acquisition and processing. We introduce fMRIPrep, an analysis-agnostic tool that addresses the challenge of robust and reproducible preprocessing for fMRI data. fMRIPrep automatically adapts a best-in-breed workflow to the idiosyncrasies of virtually any dataset, ensuring high-quality preprocessing without manual intervention. By introducing visual assessment checkpoints into an iterative integration framework for software testing, we show that fMRIPrep robustly produces high-quality results on a diverse fMRI data collection. Additionally, fMRIPrep introduces less uncontrolled spatial smoothness than observed with commonly used preprocessing tools. fMRIPrep equips neuroscientists with an easy-to-use and transparent preprocessing workflow, which can help ensure the validity of inference and the interpretability of results.

fMRI is a commonly used technique for mapping human brain activity[1]. However, the blood-oxygen-level-dependent (BOLD) signal measured by fMRI is typically mixed with non-neural sources of variability[2]. Preprocessing identifies the nuisance sources and reduces their effect on the data[3,4], and further addresses particular imaging artifacts and the anatomical localization of signals[5]. For instance, slice-timing[6] correction, head-motion correction, and susceptibility distortion correction (SDC) address particular artifacts, whereas co-registration and spatial normalization are concerned with signal localization (Supplementary Note 1). Extraction of a signal that is faithful to the underlying neural activity is crucial to ensure the validity of inference and the interpretability of results[7]. Thus, a primary goal of preprocessing is to reduce sources of false positive errors without inducing excessive false negative errors. An example of a false positive error familiar to most researchers is activation observed outside the brain due to faulty spatial normalization. As a more practical example, unaccounted-for head motion in resting-state fMRI data can generate systematic correlations that could be misinterpreted as functional connectivity[8]. Conversely, false negatives can result from a number of preprocessing failures, such as anatomical misregistration across individuals, which reduces statistical power.

Workflows for preprocessing of fMRI data produce two broad classes of outputs. First, preprocessed time series are derived from the original data after the application of retrospective signal corrections, spatiotemporal filtering, and resampling onto a target space appropriate for analysis (e.g., a standardized anatomical reference). Second, experimental confounds are additional time series such as physiological recordings and estimated noise sources that are useful for analysis (e.g., to be modeled as nuisance regressors). Some commonly used confounds include motion parameters, framewise displacement[9], spatial s.d. of the data after temporal differencing[8], and global signals, among others. Preprocessing may include further steps for denoising and estimation of confounds, such as dimensionality-reduction methods based on principal component analysis or independent component analysis. Two corresponding instances of these techniques are component-based noise correction (CompCor[10]) and automatic removal of motion artifacts (AROMA[11]).

The neuroimaging community is well equipped with tools that implement the majority of the individual steps of preprocessing described so far (Table 1). These tools are readily available in software packages such as AFNI[12], ANTs[13], FreeSurfer[14], FSL[15], Nilearn[16], and SPM[17]. Despite the wealth of accessible software and multiple attempts to outline best practices for preprocessing[2,5,7,18], the large variety of data-acquisition protocols has led to the use of ad hoc pipelines customized for nearly every study[19]. In practice, the neuroimaging community lacks a preprocessing workflow that reliably provides high-quality and consistent results from diverse datasets.

Results

fMRIPrep is a robust and convenient tool that enables researchers and clinicians to prepare both task-based and resting-state fMRI data for analysis. Its outputs allow for a range of applications, including within-subject analysis using functional localizers, voxel-based analysis, surface-based analysis, task-based group analysis, resting-state connectivity analysis, and others.

[1]Department of Psychology, Stanford University, Stanford, CA, USA. [2]Max Planck Institute for Empirical Aesthetics, Hesse, Germany. [3]Computational Neuroimaging Lab, Biocruces Health Research Institute, Bilbao, Spain. [4]Neuroscience Program, University of Iowa, Iowa City, IA, USA. [5]McGovern Institute for Brain Research, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [6]Montreal Neurological Institute, McGill University, Montreal, QC, Canada. [7]Department of Psychiatry, Stanford Medical School, Stanford University, Stanford, CA, USA. [8]Department of Neurosurgery, University of Iowa Health Care, Iowa City, IA, USA. [9]Department of Otolaryngology, Harvard Medical School, Boston, MA, USA. [10]These authors contributed equally: Russell A. Poldrack, Krzysztof J. Gorgolewski. *e-mail: phd@oscaresteban.es; krzysztof.gorgolewski@gmail.com

Table 1 | State-of-the-art neuroimaging offers a catalog of readily available software tools

Preprocessing task	Included with fMRIPrep	Alternatives (not included in fMRIPrep)
Anatomical T1-weighted brain extraction	antsBrainExtraction.sh (ANTs)	bet (FSL), 3dSkullstrip (AFNI), MRTOOL (SPM plug-in)
Anatomical surface reconstruction	recon-all (FreeSurfer)	CIVET, BrainSuite, Computational Anatomy (SPM plug-in)
Head-motion estimation (and correction)	MCFLIRT (FSL)	3dvolreg (AFNI), spm_realign (SPM), cross_realign_4dfp (4dfp), antsBrainRegistration (ANTs)
Susceptibility-derived distortion estimation (and unwarping)	3dqwarp (AFNI)	FUGUE and topup (FSL); FieldMap and HySCO (SPM plug-ins)
Slice-timing correction	3dTshift (AFNI)	slicetimer (FSL), spm_slice_timing (SPM), interp_4dfp (4dfp)
Intrasubject registration	bbregister (FreeSurfer), FLIRT (FSL)	3dvolreg (AFNI), antsRegistration (ANTs), Coregister (SPM GUI)
Spatial normalization (intersubject co-registration)	antsRegistration (ANTs)	@auto_tlrc (AFNI), FNIRT (FSL), Normalize (SPM GUI)
Surface sampling	mri_vol2surf (FreeSurfer)	SUMA (AFNI), MNE, Nilearn
Subspace projection denoising (e.g., independent or principal component analysis)	MELODIC (FSL), ICA-AROMA	Nilearn, LMGS (SPM plug-in)
Confounds	In-house implementation	fsl_motion_outliers (FSL), TAPAS PhysIO (SPM plug-in)
Detection of non-steady states	In-house implementation	Ad hoc implementations, manual setting

fMRIPrep integrates best-in-breed tools for each of the preprocessing tasks that its workflow covers, except for steps implemented as part of the development of fMRIPrep (in-house implementations). Tasks listed in the first column are described in detail in Supplementary Note 1.

A modular design alongside BIDS allows for a flexible, adaptive workflow. fMRIPrep is composed of sub-workflows that are dynamically assembled into different configurations depending on the input data. These building blocks combine tools from widely used, open-source neuroimaging packages (Table 1). The workflow engine Nipype[20] is used to stage the workflows and deal with execution details (such as resource management). The workflow comprises two major blocks, separated into anatomical MRI and fMRI processing streams (Fig. 1). The Brain Imaging Data Structure[21] (BIDS; Supplementary Note 2) allows fMRIPrep to identify the structure of the input data and gather the available metadata (e.g., imaging parameters) without manual intervention. fMRIPrep self-adapts to dataset irregularities such as missing acquisitions or runs through a set of heuristics.

Visual reports ease quality control and maximize transparency. Users can assess the quality of preprocessing with an individual report generated per participant (Supplementary Fig. 1). Reports contain dynamic and static mosaic views of images at different quality control points along the preprocessing pipeline. Written in HTML, reports can be opened with any web browser, are amenable to integration in online science services (e.g., OpenNeuro (https://openneuro.org/) and CodeOcean[22]), and maximize shareability between peers. These reports effectively minimize the amount of time required to assess the quality of the results. As an additional transparency enhancement, reports include a citation boilerplate that follows established guidelines[23] and gives due credit to all authors of all of the individual pieces of software used in fMRIPrep.

Highlights of fMRIPrep in the neuroimaging context. fMRIPrep is analysis-agnostic with respect to currently available analysis choices, as it supports a range of higher-level analysis and modeling options. Alternative workflows such as afni_proc.py (AFNI[12]), FEAT (FSL[15]), C-PAC[24] (configurable pipeline for the analysis of connectomes), Human Connectome Project (HCP[25]) Pipelines[26], and the Batch Editor extension to SPM are not agnostic because they prescribe particular methodologies to analyze the preprocessed data. Limitations to compatibility with downstream analysis derive from the coordinate space of the outputs and the regular (volume) versus irregular (surface) sampling of the BOLD signal. For example, HCP Pipelines supports surface-based analyses on subject or template space. Conversely, C-PAC and FEAT are volume-based only. Although afni_proc.py is volume-based by default, pre-reconstructed surfaces can be manually set for sampling of the BOLD signal before analysis. fMRIPrep allows a multiplicity of output spaces including subject space and atlases for both volume-based and surface-based analyses. Whereas fMRIPrep avoids processing steps that might limit further analysis (e.g., spatial smoothing), other tools are designed to perform preprocessing that supports specific analysis pipelines. For instance, C-PAC carries out several processing steps toward the connectivity analysis of resting-state fMRI. Further advantages of fMRIPrep include 'fieldmap-less' SDC, community-driven development and high standards of software engineering, and a focus on reproducibility (Methods).

fMRIPrep yields high-quality results from diverse data. We iteratively maximized the robustness and overall quality of the results generated by fMRIPrep by using a two-stage validation framework (Supplementary Fig. 2). In phase I for fault discovery, we tested fMRIPrep on a set of 30 datasets from OpenfMRI (https://openfmri.org/) (Supplementary Table 1). Because data of substandard quality are likely to degrade the outcomes of image processing, we used MRIQC[27] to select the set of test images. Phase I concluded with the release of fMRIPrep version 1.0 on 6 December 2017. We addressed quality assurance and reliability validation in phase II, thereby improving the aggregate quality of the results (Fig. 2). After phase II, 50 datasets out of 54 were rated above the 'acceptable' average quality level. The remaining four datasets were above the 'poor' level and at or near the 'acceptable' rating. We monitored the individual evolution of every dataset at seven quality control points (Supplementary Figs. 2 and 3) and report some examples of issues resolved during validation (Supplementary Results 1). Phase II concluded with the release of fMRIPrep version 1.0.8 on 22 February 2018.

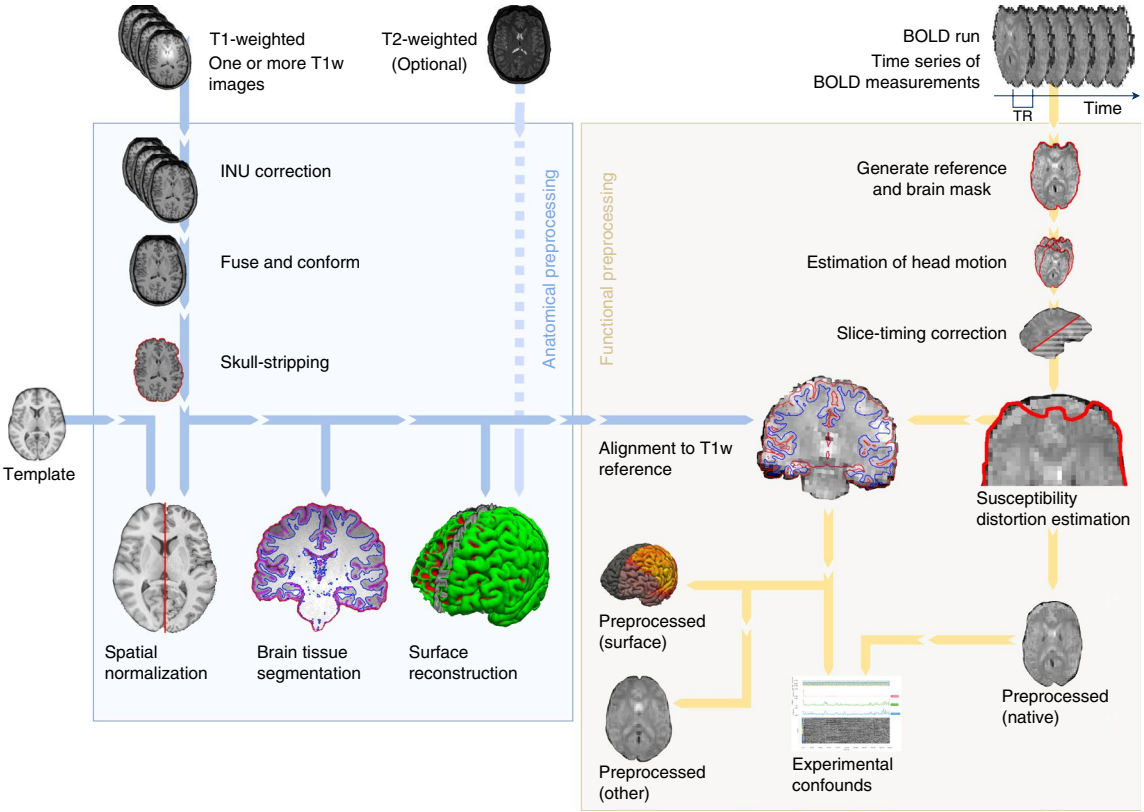

Fig. 1 | fMRIPrep is an fMRI preprocessing tool that adapts to the input dataset. Leveraging BIDS[21], the software self-adjusts automatically, configuring the optimal workflow for the given input dataset. Thus, no manual intervention is required to locate the inputs (one T1-weighted (T1w) image and one BOLD series), read acquisition parameters (such as the repetition time (TR) and the slice acquisition-times), or find additional acquisitions intended for specific preprocessing steps (such as field maps and other alternatives for the estimation of the susceptibility distortion).

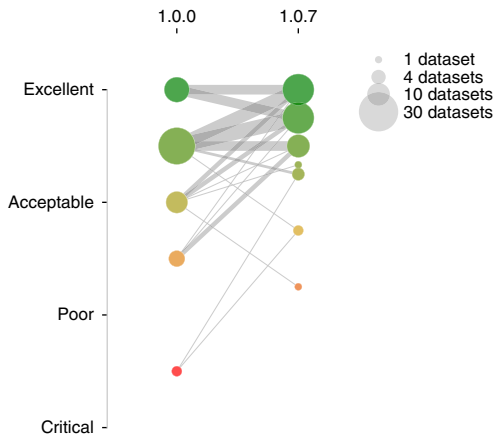

Fig. 2 | Integration of visual assessment into the software testing framework effectively increases the quality of results. In an early assessment of quality using fMRIPrep version 1.0.0, the overall rating of two datasets was below the cutoff for the 'poor' category, and that of four datasets was below the 'acceptable' level (left-hand column of colored circles). After some outstanding issues detected in the early assessment were addressed, the overall quality of processing improved substantially (right-hand column of colored circles), and no datasets were below the 'poor' quality level. Only two datasets were rated below the 'acceptable' level in the second assessment (using fMRIPrep version 1.0.7).

fMRIPrep prevents loss of spatial accuracy via smoothing. We demonstrate that the focus on robustness against data irregularity does not come at the cost of the quality of the preprocessing outputs. Moreover, the preprocessing outcomes of FSL's FEAT are smoother than those of fMRIPrep (Fig. 3a). Although preprocessed data were resampled to an isotropic voxel size of $2.0 \times 2.0 \times 2.0\,mm^3$, the smoothness estimation (before the prescribed smoothing step) for fMRIPrep was less than 4.0 mm, very close to the original resolution of $3.0 \times 3.0 \times 4.0\,mm^3$ of these data. We calculated s.d. maps in MNI (Montreal Neurological Institute) space[28] for the temporal average map derived from preprocessing with both alternatives. Visual inspection of these variability maps (Fig. 3b) revealed higher anatomical accuracy of fMRIPrep compared with that of FEAT, probably reflecting the combined effects of a more precise spatial normalization scheme and the application of fieldmap-less SDC. fMRIPrep outcomes were better aligned with the underlying anatomy in regions typically warped by susceptibility distortions such as the orbitofrontal lobe, as demonstrated in close-ups (Supplementary Fig. 4). We also compared preprocessing done with fMRIPrep and FEAT in two common fMRI analyses. First, we carried out within-subject statistical analysis using FEAT—which provides both preprocessing and first-level analysis, separable via configuration file—on both sets of preprocessed data. Second, we conducted a group statistical analysis using ordinary least-squares mixed modeling (FLAME[29]; FSL). In both experiments, we applied identical analysis workflows and settings to both preprocessing alternatives. The first-level analysis showed that the thresholded

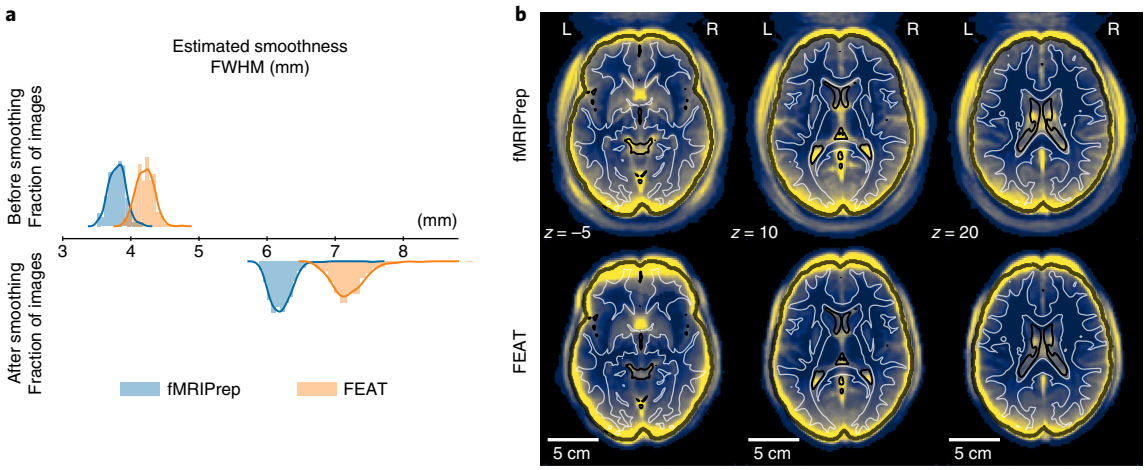

Fig. 3 | fMRIPrep affords researchers finer control over the smoothness of their analysis. a, Estimation of the spatial smoothness of data before and after the initial smoothing step of the analysis workflow confirmed that results of preprocessing with FEAT were intrinsically smoother. **b**, The s.d. maps of averaged BOLD time series showed greater variability around the brain outline (represented by a black contour) for data preprocessed with FEAT compared with fMRIPrep. This effect is generally associated with a lower performance of spatial normalization[28]. Reference contours correspond to the brain tissue segmentation of the MNI atlas. $N = 257$ independent participants.

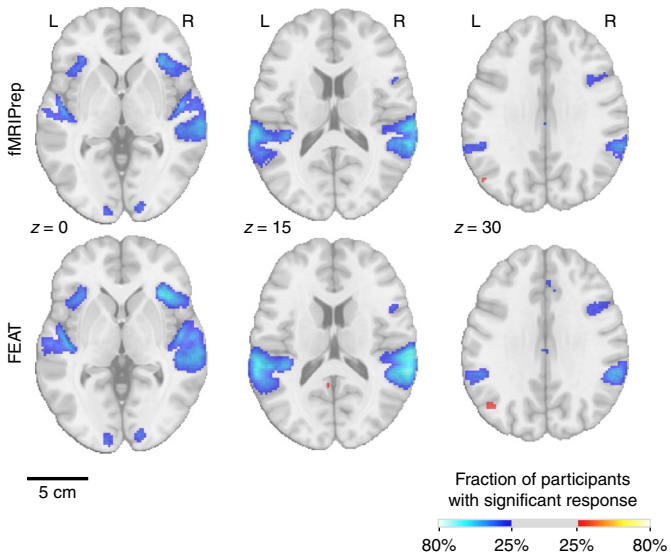

Fig. 4 | Activation count maps from fMRIPrep are better aligned with the underlying anatomy than those from FEAT. The mosaics show thresholded activation count maps for the go versus successful-stop contrast in the 'stopsignal' task after preprocessing with either fMRIPrep (top) or FSL's FEAT (bottom), with identical single-subject statistical modeling. Both tools produced similar activation maps, with the fMRIPrep results being slightly better aligned with the underlying anatomy. Cold hues represent negative activation (response inhibition). Warm hues represent positive activation. $N = 257$ independent participants.

activation count maps for the go versus successful-stop contrast in the 'stopsignal' task were similar (Fig. 4). The results from both pipelines identified activation in the same regions. However, because data preprocessed with FEAT were smoother, the results from fMRIPrep were more local and better aligned with the cortical sheet. The overlap of statistical maps, as well as Pearson's correlation, was tightly related to the smoothing of the input data. In the group analysis, fMRIPrep and FEAT performed equivalently (Supplementary Results 2).

Discussion

fMRIPrep is an fMRI preprocessing workflow developed to excel at four aspects of scientific software: robustness to data idiosyncrasies, high quality and consistency of results, maximal transparency, and ease of use. We stress that fMRIPrep has been developed with the best software engineering principles, which are fundamental to ensure software reliability. The pipeline is easy to use for researchers and clinicians without extensive computer engineering experience, and produces comprehensive visual reports.

In a comparison with FSL's FEAT, fMRIPrep achieved higher spatial accuracy and introduced less uncontrolled smoothness. Group P statistical maps differed only in their smoothness (sharper for fMRIPrep). The fact that first-level and second-level analyses resulted in small differences between fMRIPrep and our ad hoc implementation of a FEAT-based workflow indicates that the individual preprocessing steps perform similarly when they are fine-tuned to the input data. This justifies the need for fMRIPrep, which autonomously adapts the workflow to the data without error-prone manual intervention. To a limited extent, this also mitigates some concerns and theoretical risks that arise from analytical degrees of freedom[19] available to researchers. fMRIPrep stands out among pipelines because it automates the adaptation to the input dataset without compromising the quality of the results.

One limitation of this work is the use of visual and semivisual assessments of the quality of preprocessing outcomes. Although some frameworks have been proposed for the quantitative evaluation of preprocessing on task-based (e.g., NPAIRS[30]) and resting-state[31] fMRI, they impose a set of assumptions on the test data and the workflow being assessed that limit their suitability in general. The modular design of fMRIPrep defines an interface for each processing step, which permits programmatic evaluation of the many possible combinations of software tools and processing steps. This will also enable the use of quantitative testing frameworks to pursue the minimization of type I errors without an associated increase in type II errors. The range of possible applications for fMRIPrep also presents some boundaries. For instance, images with a very narrow field of view often do not contain enough information for standard image-registration methods to work correctly. OpenfMRI datasets representing images with reduced fields of view were excluded from the evaluation because they are not yet fully supported by fMRIPrep. Extension of fMRIPrep's support for these particular images is

already a future line of the development roadmap. fMRIPrep may also underperform for particular populations (e.g., infants) or for brains that show nonstandard structures such as tumors, resected regions, or lesions. Despite these challenges, fMRIPrep performed robustly on data from a study involving simultaneous MRI and electrocorticography, which are challenging to analyze owing to the massive BOLD-signal dropout near the implanted cortical electrodes (Supplementary Fig. 5). In addition, fMRIPrep's modular architecture makes it straightforward to extend the tool to support specific populations or new species by providing appropriate atlases of their brains. This would be particularly interesting as a means to adapt the workflow to data collected from rodents and nonhuman primates in the future.

Approximately 80% of the analysis pipelines investigated by Carp[19] were implemented using AFNI[12], FSL[15], or SPM[17]. Ad hoc pipelines adapt the basic workflows provided by these tools to the particular dataset at hand. Although workflow frameworks like Nipype[20] ease the integration of tools from different packages, these pipelines are typically restricted to just one of these alternatives (AFNI, FSL, or SPM). Otherwise, scientists can adopt the acquisition protocols and associated preprocessing software of large consortia[26,32] like the HCP and the UK Biobank[33]. The off-the-shelf applicability of these workflows is contravened by important limitations on the experimental design. Therefore, researchers typically opt to recode their custom preprocessing workflows with nearly every new study[19]. That practice entails a 'pipeline debt', which requires an investment in proper software engineering to ensure acceptable correctness and stability of the results (e.g., continuous integration testing) and reproducibility (versioning, packaging, containerization, etc.). A trivial example of this risk is the leakage of 'magic numbers' that are hard-coded in the source (e.g., a crucial imaging parameter that has inadvertently changed from one study to the next one). Until fMRIPrep, an analysis-agnostic approach was lacking.

The rapid increase in the volume and diversity of data, as well as the evolution of available techniques for processing and analysis, presents an opportunity for considerable advancement of research in neuroscience. The drawback resides in the need for progressively more complex analysis workflows that rely on decreasingly interpretable models of the data. Such context encourages 'black-box' solutions that efficiently perform a valuable service but do not provide insights into how the tool has transformed the data into the expected outputs. Black boxes obscure important steps in the inductive process mediating between experimental measurements and reported findings. This way of moving forward risks producing a future generation of cognitive neuroscientists who have become experts in sophisticated computational methods but have little to no working knowledge of how their data were transformed through processing. Transparency is often identified as a remedy for these problems. fMRIPrep ascribes to 'glass-box' principles, which are defined in opposition to the many different facets or levels at which black-box solutions are opaque. The visual reports that fMRIPrep generates are a crucial aspect of the glass-box approach. Their quality control checkpoints represent the logical flow of preprocessing, allowing scientists to critically inspect and better understand the underlying mechanisms of the workflow. A second transparency element is the citation boilerplate that formalizes all details of the workflow and provides the versions of all involved tools along with references to the corresponding scientific literature. A third asset for transparency is thorough documentation that delivers additional details on each of the building blocks represented in the visual reports and described in the boilerplate. Further, fMRIPrep has been open-source since its inception: users have access to all of the incremental additions to the tool through the history of the version-control system. The use of GitHub (https://github.com/poldracklab/fmriprep) grants access to the discussions held during development, allowing one to see how and why the main design decisions were made. The modular design of fMRIPrep enhances its flexibility and improves transparency, as the main features of the software are more easily accessible to potential collaborators. In combination with some coding style and contribution guidelines, this modularity has enabled multiple contributions by peers and the creation of a rapidly growing community that would be difficult to nurture behind closed doors. A number of existing tools have implemented elements of glass-box philosophy (e.g., visual reports in FEAT, documentation in C-PAC, the open-source community of Nilearn), but the complete package (visual reports, educational documentation, reporting templates, and a collaborative open-source community) is still rare among scientific software. fMRIPrep's transparent and accessible development and reporting aim to better equip fMRI practitioners to conduct reliable, reproducible statistical analyses with a high-standard, consistent, and adaptive preprocessing instrument.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41592-018-0235-4.

Received: 27 April 2018; Accepted: 29 October 2018;
Published online: 10 December 2018

References

1. Poldrack, R. A. & Farah, M. J. Progress and challenges in probing the human brain. *Nature* **526**, 371–379 (2015).
2. Power, J. D., Plitt, M., Laumann, T. O. & Martin, A. Sources and implications of whole-brain fMRI signals in humans. *Neuroimage* **146**, 609–625 (2017).
3. Lindquist, M. A. The statistical analysis of fMRI data. *Stat. Sci.* **23**, 439–464 (2008).
4. Caballero-Gaudes, C. & Reynolds, R. C. Methods for cleaning the BOLD fMRI signal. *Neuroimage* **154**, 128–149 (2017).
5. Strother, S. C. Evaluating fMRI preprocessing pipelines. *IEEE Eng. Med. Biol. Mag.* **25**, 27–41 (2006).
6. Sladky, R. et al. Slice-timing effects and their correction in functional MRI. *Neuroimage* **58**, 588–594 (2011).
7. Ashburner, J. Preparing fMRI data for statistical analysis. In: M. Filippi, ed. *fMRI Techniques and Protocols* (pp. 151–178. Humana Press, New York, 2009).
8. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).
9. Power, J. D. et al. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).
10. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).
11. Pruim, R. H. R. et al. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* **112**, 267–277 (2015).
12. Cox, R. W. & Hyde, J. S. Software tools for analysis and visualization of fMRI data. *NMR. Biomed.* **10**, 171–178 (1997).
13. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
14. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
15. Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
16. Abraham, A. et al. Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **8**, 14 (2014).
17. Friston, K. J., Ashburner, J., Kiebel, S. J., Nichols, T. E. & Penny, W. D. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. (Academic Press, London, 2006).
18. Power, J. D., Plitt, M., Kundu, P., Bandettini, P. A. & Martin, A. Temporal interpolation alters motion in fMRI scans: magnitudes and consequences for artifact detection. *PLoS ONE* **12**, e0182939 (2017).
19. Carp, J. The secret lives of experiments: methods reporting in the fMRI literature. *Neuroimage* **63**, 289–300 (2012).
20. Gorgolewski, K. et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinform.* **5**, 13 (2011).

21. Gorgolewski, K. J. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044 (2016).
22. Esteban, O. et al. FMRIPrep: a robust preprocessing pipeline for functional MRI. *Code Ocean* https://doi.org/10.24433/CO.ed5ddfef-76a3-4996-b298-e3200f69141b (2018).
23. Poldrack, R. A. et al. Guidelines for reporting an fMRI study. *Neuroimage* **40**, 409–414 (2008).
24. Sikka, S. et al. Towards automated analysis of connectomes: the configurable pipeline for the analysis of connectomes (C-PAC). 5th INCF Congress of Neuroinformatics, Munich, Germany, 10–12 September 2012.
25. Van Essen, D. C. et al. The Human Connectome Project: a data acquisition perspective. *Neuroimage* **62**, 2222–2231 (2012).
26. Glasser, M. F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
27. Esteban, O. et al. MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE* **12**, e0184661 (2017).
28. Calhoun, V. D. et al. The impact of T1 versus EPI spatial normalization templates for fMRI data analyses. *Hum. Brain. Mapp.* **38**, 5331–5342 (2017).
29. Beckmann, C. F., Jenkinson, M. & Smith, S. M. General multilevel linear modeling for group analysis in FMRI. *Neuroimage* **20**, 1052–1063 (2003).
30. Strother, S. C. et al. The quantitative evaluation of functional neuroimaging experiments: the NPAIRS data analysis framework. *Neuroimage* **15**, 747–771 (2002).
31. Karaman, M., Nencka, A. S., Bruce, I. P. & Rowe, D. B. Quantification of the statistical effects of spatiotemporal processing of nontask fMRI data. *Brain Connect* **4**, 649–661 (2014).
32. Alfaro-Almagro, F. et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400–424 (2018).
33. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536 (2016).

Acknowledgements

This work was supported by the Laura and John Arnold Foundation (R.A.P. and K.J.G.), the NIH (grant NBIB R01EB020740, S.S.G.), NIMH (R24MH114705 and R24MH117179, R.A.P.), and NINDS (U01NS103780, R.A.P.). J.D. has received funding from the European Union's Horizon 2020 research and innovation program under Marie Sklodowska-Curie grant agreement 706561. The authors thank S. Nastase and T. van Mourik for their thoughtful open reviews of a preprint version of this paper.

Author contributions

O.E. contributed with conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, and writing (original draft, review, and editing). C.J.M. contributed with conceptualization, data curation, methodology, software, validation, and writing (review and editing). R.W.B. contributed with software, validation, and writing (review and editing). C.A.M. contributed with methodology, software, and writing (review and editing). A.I.I. contributed with software and writing (review and editing). A.E. contributed with software and writing (review and editing). J.D.K. contributed with investigation, methodology, software, visualization, and writing (review and editing). M.G. contributed with software and writing (review and editing). E.D. contributed with software and writing (review and editing). M.S. contributed with software and writing (review and editing). H.O. contributed with data acquisition and writing (review and editing). S.S.G. contributed with conceptualization, software, and writing (review and editing). J.W. contributed with conceptualization and writing (review and editing). J.D. contributed with formal analysis, investigation, methodology, software, and writing (review and editing). R.A.P. contributed with conceptualization, formal analysis, investigation, methodology, validation, supervision, resources, funding acquisition, and writing (original draft, review, and editing). K.J.G. contributed with conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, supervision, resources, funding acquisition, and writing (original draft, review, and editing).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0235-4.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to O.E. or K.J.G.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Data. *Data used in the validation of fMRIPrep.* Participants were drawn from a multiplicity of studies[9,34–90] available in OpenfMRI, accessed on 30 September 2017. Studies were sampled uniformly (four participants each), except for ds000031, which consisted of only one participant. Data selection criteria are described below. Magnetic resonance imaging (MRI) data were acquired at multiple scanning centers, with the following frequencies of vendors: ~70% Siemens, ~14% Philips, ~14% GE. Data were acquired by 1.5T and 3T systems running varying software versions. Acquisition protocols, as well as the particular acquisition parameters (including relevant BOLD settings such as the repetition time (TR), the echo time (TE), the number of TRs, and the resolution), also varied with each study. However, only datasets that included at least one T1-weighted (T1w) and one BOLD acquisition per subject run were included. Datasets containing BIDS errors (ds000210) and degenerate data (many T1w images of ds000223 are skull-stripped) at the time of access were discarded. Similarly, datasets comprising very narrow field of view (FoV) BOLD scans (ds000172, ds000217, and ds000232) were also excluded. In total, 54 datasets (46 single-session and 8 multi-session datasets) were included in this assessment (Supplementary Table 1).

This evaluation covered 54 studies out of a total of 58 studies in OpenfMRI that included the two required imaging modalities (T1w and BOLD). Therefore, by covering 93% of the studies in OpenfMRI, we ensured considerable heterogeneity in terms of acquisition protocols, settings, instruments, and parameters, which was necessary to demonstrate the robustness of fMRIPrep against the variability in input data features.

Data used in the comparison to FSL FEAT. We reused the UCLA Consortium for Neuropsychiatric Phenomics LA5c study[34], a dataset that is publicly available on OpenfMRI under data accession ds000030. During the experiment, subjects performed six tasks, a block of resting-state fMRI, and underwent two anatomical scans. The study includes imaging data of a large group of healthy individuals from the community, as well as samples of individuals diagnosed with schizophrenia, bipolar disorder, and attention-deficit or hyperactivity disorder. As described in the data descriptor[34], MRI data were acquired on one of two 3T Siemens Trio scanners located at the Ahmanson–Lovelace Brain Mapping Center (syngo MR B15) and the Staglin Center for Cognitive Neuroscience (syngo MR B17). fMRI data were collected with an echo-planar imaging (EPI) sequence (slice thickness, 4 mm; 34 slices; TR, 2 s; TE, 30 ms; flip angle, 90°; matrix, 64×64; FoV, 192 mm; oblique slice orientation). Additionally, one T1w image per participant is available (MPRAGE; TR, 1.9 s; TE, 2.26 ms; FoV, 250 mm; matrix, 256×256; sagittal plane; slice thickness, 1 mm; 176 slices). For this experiment, only images including both the T1w and the functional scans corresponding the Stop Signal task (referred to as 'stopsignal') were included (totaling $N=257$ participants).

For the stopsignal task, participants were instructed to respond quickly to a 'go' stimulus. During some of the trials, at unpredictable times, a stop signal would appear after the stimulus was presented. During those trials, the subject had to inhibit any planned response. In this experiment, we specifically looked into the difference between the brain activation during a successful stop trial and a go trial (contrast: go – stopsuccess). Thus, we expected to see brain regions responsible for response inhibition (negative) and motor response (positive). Further details on the task are available with the dataset descriptor[34].

The fMRIPrep workflow. *Preprocessing anatomical images.* The T1w image is corrected for intensity nonuniformity with N4BiasFieldCorrection[91] (ANTs) and skull-stripped with antsBrainExtraction.sh (ANTs). Skull-stripping is done through co-registration to a template, with two options available: the OASIS template[92] (default) and the NKI template[93]. Using visual inspection, we have found that this approach outperforms other common approaches, which is in agreement with previous reports[26]. When several T1w volumes are found, the intensity-nonuniformity-corrected versions are first fused into a reference T1w map of the subject with mri_robust_template[94] (FreeSurfer). Brain surfaces are reconstructed from the subject's T1w reference (and T2w images if available) with recon-all[95] (FreeSurfer). The brain mask estimated previously is refined with a custom variation of a method (originally introduced in Mindboggle[96]) to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray matter (GM). Both surface reconstruction and subsequent mask refinement are optional and can be disabled to save run time when surface-based analysis is not needed. Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template[97] (version 2009c) is done through nonlinear registration with antsRegistration[98] (ANTs), using brain-extracted versions of both the T1w reference and the standard template. ANTs was selected because of its superior performance in terms of volumetric group level overlap[99]. Brain tissues—cerebrospinal fluid (CSF), white matter (WM), and GM—are segmented from the reference, brain-extracted T1w image with FAST[100] (FSL).

Preprocessing functional runs. For every BOLD run found in the dataset, a reference volume and its skull-stripped version are generated via an in-house methodology (Supplementary Note 3). Then, head-motion parameters (volume-to-reference transform matrices and corresponding rotation and translation parameters) are estimated with MCFLIRT[101] (FSL). From among several alternatives (Table 1), MCFLIRT was chosen because its results are comparable to those of other tools[102] and it stores the estimated parameters in a format that facilitates the composition of spatial transforms to achieve one-step interpolation (see below). If slice-timing information is available, BOLD runs are (optionally) slice-time corrected with 3dTshift (AFNI[12]). When fieldmap information is available or the experimental fieldmap-less correction is requested (see the section "Fieldmap-less susceptibility distortion correction" below), SDC is carried out with the appropriate methods (Supplementary Fig. 6). This is followed by co-registration to the corresponding T1w reference using boundary-based registration[103] with nine degrees of freedom (to minimize remaining distortions). If surface reconstruction is selected, fMRIPrep uses bbregister (FreeSurfer). Otherwise, the boundary-based co-registration implemented in FLIRT (FSL) is applied. In our experience, bbregister yields the better results[103] owing to the high resolution and topological correctness of the GM/WM surfaces driving registration. To support a large variety of output spaces for the results (the native space of BOLD runs, the corresponding T1w space, FreeSurfer's fsaverage spaces, the template used as a target in the spatial normalization step, etc.), one can combine the transformations between spaces. For example, for the generation of preprocessed BOLD runs in template space (e.g., MNI), the following transforms are concatenated: head-motion parameters, the warping to reverse susceptibility distortions (if calculated), BOLD-to-T1w, and T1w-to-template mappings. The BOLD signal is also sampled onto the corresponding participant's surfaces with mri_vol2surf (FreeSurfer) when surface reconstruction is being performed. Thus, these sampled surfaces can easily be transformed onto different output spaces available by concatenating transforms calculated throughout fMRIPrep and internal mappings between spaces calculated with recon-all. The composition of transforms allows for a single-interpolation resampling of volumes with antsApplyTransforms (ANTs). Lanczos interpolation is applied to minimize the smoothing effects of linear or Gaussian kernels[104]. Optionally, ICA-AROMA can be used and corresponding 'non-aggressively' denoised runs are then produced. When ICA-AROMA is enabled, the time series are first smoothed and then denoised according to the description of the original method[11].

Extraction of nuisance time series. To avoid restriction of fMRIPrep's outputs to particular analysis types, the tool does not perform any temporal denoising by default. Nonetheless, it provides researchers with a diverse set of confound estimates that could be used for explicit nuisance regression or as part of higher-level models. This lends itself to decoupling of preprocessing and behavioral modeling, as well as evaluation of the robustness of final results across different denoising schemes. A set of physiological noise regressors is extracted for the purpose of component-based noise correction (CompCor[10]). Principal components are estimated after high-pass filtering of the BOLD time series (using a discrete cosine filter with a 128-s cutoff) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). Six tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavy erosion of the brain mask, which ensures that it does not include cortical GM regions. For aCompCor, six components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Framewise displacement and spatial s.d. of the data after temporal differencing (known as 'DVARS') are calculated for each functional run, both with their implementations in Nipype (according to the definitions by Power et al.[8]). Three global signals are extracted within the CSF, the WM, and the whole-brain masks using Nilearn[16]. If ICA-AROMA[11] is requested, the 'aggressive' noise regressors are collected and placed within the corresponding confounds files. Because the non-aggressive cleaning with ICA-AROMA is done after extraction of other nuisance signals, the aggressive regressors can be used to orthogonalize those other nuisance signals to avoid the risk of reintroducing nuisance signal within regression. In addition, a non-aggressive version of preprocessed data is also provided, as this variant of ICA-AROMA denoising cannot be applied with only nuisance regressors.

Fieldmap-less susceptibility distortion correction. Many legacy and current human fMRI protocols lack the magnetic resonance fieldmaps necessary to apply standard SDC methods. As described in Supplementary Fig. 6, the BIDS dataset is queried to discover whether extra acquisitions containing fieldmap information are available. When no fieldmap information is found, fMRIPrep adapts the fieldmap-less correction for diffusion EPI images introduced by Wang et al.[105]. They propose using the same-subject T1w reference as the undistorted target in a nonlinear registration scheme. To maximize the similarity between the T2* contrast of the EPI scan and the reference T1w data, the intensities of the latter are inverted. To regularize the optimization of the deformation field, only displacements along the phase-encoding direction are allowed, and the magnitude of the displacements is modulated using priors. To our knowledge, no other existing pipeline applies fieldmap-less SDC to BOLD images. Further details on the integration of the different SDC techniques, and particularly this fieldmap-less option, are presented in Supplementary Note 3.

fMRIPrep is thoroughly documented, community-driven, and developed with high standards of software engineering. Preprocessing pipelines are generally well

documented; however, the extreme flexibility of fMRIPrep makes its proper documentation substantially more challenging. As is true of other scientific software communities, fMRIPrep contributors pledge to keep the documentation thorough and updated along coding iterations. Packages also differ on the involvement of the community: whereas fMRIPrep includes researchers in the decision-making process and invites their suggestions and contributions, other packages have a more closed model where the feedback from users is more limited (e.g., a mailing list). In contrast to other pipelines, fMRIPrep is community-driven. This paradigm allows the fast adoption of advances on fMRI preprocessing, which tend to render existing workflows (including fMRIPrep) obsolete. For example, fMRIPrep initially performed slice-timing correction before head-motion correction, but we adapted the tool to the recent recommendations of Power et al.[18] upon a user's request. This model has allowed the user base to grow rapidly and has enabled substantial third-party contributions to be included in the software, such as support for processing of multi-echo datasets. The open-source nature of fMRIPrep has permitted frequent code reviews that are effective in enhancing the software's quality and reliability[106]. Supplementary Note 4 describes how the community interacts, discusses the code-review process, and underscores how the modular design of fMRIPrep successfully facilitates contributions from peers. Finally, fMRIPrep undergoes continuous integration testing (Supplementary Note 4), a technique that has recently been proposed as a means to ensure reproducibility of analyses in computational sciences[107,108]. Additional comparison points, such as the graphical user interface of several preprocessing workflows, are given in Supplementary Note 5.

Ensuring reproducibility with strict versioning and containers. For enhanced reproducibility, fMRIPrep fully supports execution via the Docker (https://docker.com) and Singularity[109] container platforms. Container images are generated and uploaded to a public repository for each new version of fMRIPrep. These containers are released with a fixed set of software versions for fMRIPrep and all its dependencies, maximizing run-to-run reproducibility in an easy way. This helps to address the widespread lack of reporting of specific software versions and the large variability of software versions, which threaten the reproducibility of fMRI analyses[19]. Except for C-PAC, alternative pipelines do not provide official support for containers. The adoption of the BIDS-Apps[107] container model makes fMRIPrep amenable to a multiplicity of infrastructures and platforms: PC, high-performance computing, Cloud, and so on.

Validation of fMRIPrep on diverse data. Our general validation framework (Supplementary Fig. 2) implements a testing plan elaborated prior to the release of version 1.0 of the software. The plan is divided into two validation phases in which different data samples and validation procedures are applied. Supplementary Table 1 describes the data samples used in each phase. In phase I, we ran fMRIPrep on a manually selected sample of participants that we viewed as potentially challenging to the tool's robustness, exercising the adaptiveness to the input data. In phase II we focused on visual assessment of the quality of preprocessing results for a large and heterogeneous sample.

Methodology and test plan. To ensure that fMRIPrep fulfills the specifications on reliability and scientific-software standards, the tool undergoes a thorough acceptance testing plan. The plan is structured in three phases: the discovery of faults, evaluation of the robustness, and the final phase at the full coverage of OpenfMRI. We note that an additional early test phase (phase 0) was conducted as a proof of concept for the tool.

During validation phase I, in which fault-discovery testing was carried out, a total of 120 subjects from 30 different datasets (Supplementary Table 1) were manually identified as low quality by MRIQC[27]. Data showing substandard quality are known to be likely to degrade the outcomes of image processing, and therefore they are helpful in tests of software reliability. This subsample of OpenfMRI underwent preprocessing in the Stampede2 supercomputer of the Texas Advanced Computer Center in Austin, TX. Results were visually inspected and failures were reported in the GitHub repository. Once software faults were fixed, fMRIPrep 1.0.0 was released and phase II of validation was launched.

In validation phase II, focused on quality assurance and reliability testing, the coverage of OpenfMRI was extended to 54 available datasets (Supplementary Table 1), with four participants selected at random per dataset (with replacement of participants covered in phase I). A total of 325 participants were preprocessed in the Sherlock cluster of Stanford University (Stanford, CA). Validation phase II integrated a protocol for the screening of results into the software testing (Supplementary Fig. 2). Three raters evaluated each participant's report according to the protocol described below. Their ratings are made available with the corresponding reports for scrutiny.

Each visual report generated in phase II was inspected by one expert (C.J.M., K.J.G., or O.E., selected randomly) at seven quality checkpoints: (i) overall performance, (ii) surface reconstruction from anatomical MRI, (iii) T1w brain mask and tissue segmentation, (iv) spatial normalization, (v) brain mask and regions of interest (ROIs) for CompCor application in native BOLD space ('BOLD ROIs'), (vi) intrasubject BOLD-to-T1w co-registration, and (vii) SDC. Experts were instructed to assign a score on a scale from 1 (poor) to 3 (excellent) at each quality control point. A special rating score of 0 (unusable) was assigned to tasks that failed in a critical way that hampered further preprocessing. Poor (1) was assigned when fMRIPrep did not critically fail at the task but the outcome was likely to negatively affect downstream analysis. For example, when fieldmap-less correction unwarped in the expected direction but some distorted areas remained (or were overcorrected), then a rating of acceptable (2) was assigned. Finally, a rating of excellent (3) was assigned when the expert did not notice any substantial defect that would indicate a lower rating. Supplementary Fig. 3 shows the evolution of the quality ratings at the seven checkpoints at the beginning and completion of phase II (indicated by versions 1.0.0 and 1.0.7, respectively).

Comparison to an alternative preprocessing tool. For comparison, data were preprocessed with two alternative pipelines: fMRIPrep 1.0.8 and FSL's FEAT 5.0.10. We then carried out identical analyses of each dataset preprocessed with either pipeline. In the first-level analysis, we calculated a *t*-statistic map per participant for the task under analysis ($N=257$). Second-level analyses were conducted in a specific resampling scheme to allow for statistical comparison between the pipelines: two random (non-overlapping) subsets of n participants were repeatedly entered into a group-level analysis. The first step was the experimental manipulation resulting in two conditions: (1) the data were preprocessed with fMRIPrep, and (2) the data were preprocessed with FEAT. The next two steps were identical for both conditions.

Preprocessing. Preprocessing with fMRIPrep is described using the corresponding citation boilerplate (Supplementary Note 3). We configured FEAT using its graphical user interface and generated a template.fsf file, which can be found in GitHub (https://github.com/oesteban/misc/tree/16660df9fe80d20107b6abd7fc8ce1f4946791e6/fsl-feat). We manually extended execution to all participants in our sample, creating the script fsl_feat_wrapper.py, which accompanies the template.fsf file in GitHub. As can be seen in the template.fsf file, we disabled band-pass filtering and spatial smoothing to make the results of preprocessing comparable. Both processing steps (temporal filtering and spatial smoothing) were implemented in a common, subsequent analysis workflow described below. Additionally, we manually configured the ICBM 152 Nonlinear Asymmetrical template[97] version 2009c as a target for spatial normalization. Finally, we manually resampled the preprocessed BOLD files into template space using FSL's FLIRT.

To investigate the spatial consistency of the average BOLD signal across participants, we calculated s.d. maps in MNI space for the temporal average map[28] derived from preprocessing with both alternatives.

We used AFNI's 3dFWHMx to estimate the (average) smoothness of the data at two checkpoints: (i) before the first-level analysis workflow, and (ii) after application of a 5.0-mm full-width at half-maximum (FWHM) spatial smoothing, which was the first step of the analysis workflow described below.

First-level statistical analysis. We analyzed the stopsignal task data using FSL and AFNI tools, integrated in a workflow using Nipype. Spatial smoothing was applied with AFNI's 3dBlurInMask with a Gaussian kernel of FWHM = 5 mm. Activity was estimated using a general linear model with FSL's FEAT. For the one condition under comparison (go – successful), one task regressor was included with a fixed duration of 1.5 s. An extra regressor was added with equal amplitude, but with the duration equal to the reaction time. 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 three directions, rotation in three directions) were added as regressors to avoid confounding effects of head motion. We included a high-pass filter (100 Hz) in FSL's FEAT.

The statistical map for each participant was binarized at $z=\pm1.65$ (which corresponds to a two-sided test value of $P<0.1$). Then, the average of these maps was computed across participants. The average negative map (the percentage of subjects showing a negative effect with $z<-1.65$) was subtracted from the average positive map to indicate the direction of effects. High values in certain regions and low values in other regions indicate a good overlap of activation between subjects.

Second-level statistical analysis. Subsequent to the single-subject analyses, two random, non-overlapping subsamples of n subjects were taken and entered into a second-level analysis. We varied the sample size n of groups between 10 and 120. We ran the group-level analyses based on two variants of the first level: with a prescribed smoothing of 5.0-mm FWHM, and without such a smoothing step. The resampling process was repeated 200 times per group sample size and smoothing condition. To investigate the implications of either pipeline on the group analysis use case, we ran the same ordinary least-squares mixed modeling using FSL's FLAME on each of two disjoint subsets of randomly selected subjects and resampling repetition. We calculated several metrics of spatial agreement on the resulting maps of (uncorrected) *P*-statistical values. We also calculated the spatial agreement of the thresholded statistical maps, binarized with a threshold chosen to control for the false discovery rate (FDR) at 5% (using FSL's FDR command).

Ethical compliance. We complied with all relevant ethical regulations. This study reused publicly available data acquired at many different institutions. Protocols for all of the original studies were approved by the corresponding ethical boards.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Software availability. fMRIPrep's source code is available at GitHub (https://github.com/poldracklab/fmriprep). We use Zenodo to generate new digital object identifiers for each new release of fMRIPrep, the latest being version 1.1.4 (https://doi.org/10.5281/zenodo.1340696). fMRIPrep is licensed under the BSD 3-Clause "New" or "Revised" License. Software is distributed as a Python package (https://pypi.org/project/fmriprep/), as a Docker container (https://hub.docker.com/r/poldracklab/fmriprep/), and as a CodeOcean capsule[22].

Data availability

All original data used in this work are publicly available through the OpenNeuro 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 (https://doi.org/10.6084/m9.figshare.6196994.v3). Source data for Fig. 3 are available online.

References

34. Poldrack, R. A. et al. A phenome-wide examination of neural and cognitive function. *Sci. Data* **3**, 160110 (2016).
35. Schonberg, T. et al. Decreasing ventromedial prefrontal cortex activity during sequential risk-taking: an fMRI investigation of the balloon analog risk task. *Front. Neurosci.* **6**, 80 (2012).
36. Aron, A. R., Gluck, M. A. & Poldrack, R. A. Long-term test-retest reliability of functional MRI in a classification learning task. *Neuroimage* **29**, 1000–1006 (2006).
37. Xue, G. & Poldrack, R. A. The neural substrates of visual perceptual learning of words: implications for the visual word form area hypothesis. *J. Cogn. Neurosci.* **19**, 1643–1655 (2007).
38. Tom, S. M., Fox, C. R., Trepel, C. & Poldrack, R. A. The neural basis of loss aversion in decision-making under risk. *Science* **315**, 515–518 (2007).
39. Xue, G., Aron, A. R. & Poldrack, R. A. Common neural substrates for inhibition of spoken and manual responses. *Cereb. Cortex* **18**, 1923–1932 (2008).
40. Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J. & Poldrack, R. A. Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J. Neurosci.* **27**, 3743–3752 (2007).
41. Foerde, K., Knowlton, B. J. & Poldrack, R. A. Modulation of competing memory systems by distraction. *Proc. Natl Acad. Sci. USA* **103**, 11778–11783 (2006).
42. Gorgolewski, K. J., Durnez, J. & Poldrack, R. A. Preprocessed Consortium for Neuropsychiatric Phenomics dataset. *F1000 Res.* **6**, 1262 (2017).
43. Laumann, T. O. et al. Functional system and areal organization of a highly sampled individual human brain. *Neuron* **87**, 657–670 (2015).
44. Alvarez, R., Jasdzewski, G. & Poldrack, R. A. Building memories in two languages: an fMRI study of episodic encoding in bilinguals. In SfN Neuroscience (Orlando, FL, US, 2002).
45. Poldrack, R. A. et al. Interactive memory systems in the human brain. *Nature* **414**, 546–550 (2001).
46. Kelly, A. M. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X. & Milham, M. P. Competition between functional brain networks mediates behavioral variability. *Neuroimage* **39**, 527–537 (2008).
47. Mennes, M. et al. Inter-individual differences in resting-state functional connectivity predict task-induced BOLD activity. *Neuroimage* **50**, 1690–1701 (2010).
48. Mennes, M. et al. Linking inter-individual differences in neural activation and behavior to intrinsic brain dynamics. *Neuroimage* **54**, 2950–2959 (2011).
49. Haxby, J. V. et al. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
50. Hanson, S. J., Matsuka, T. & Haxby, J. V. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? *Neuroimage* **23**, 156–166 (2004).
51. Duncan, K. J., Pattamadilok, C., Knierim, I. & Devlin, J. T. Consistency and variability in functional localisers. *Neuroimage* **46**, 1018–1026 (2009).
52. Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist, M. A. & Ochsner, K. N. Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron* **59**, 1037–1050 (2008).
53. Moran, J. M., Jolly, E. & Mitchell, J. P. Social-cognitive deficits in normal aging. *J. Neurosci.* **32**, 5553–5561 (2012).
54. Uncapher, M. R., Hutchinson, J. B. & Wagner, A. D. Dissociable effects of top-down and bottom-up attention during episodic encoding. *J. Neurosci.* **31**, 12613–12628 (2011).
55. Gorgolewski, K. J. et al. A test-retest fMRI dataset for motor, language and spatial attention functions. *Gigascience* **2**, 6 (2013).
56. Repovš, G. & Barch, D. M. Working memory related brain network connectivity in individuals with schizophrenia and their siblings. *Front. Hum. Neurosci.* **6**, 137 (2012).
57. Repovš, G., Csernansky, J. G. & Barch, D. M. Brain network connectivity in individuals with schizophrenia and their siblings. *Biol. Psychiatry* **69**, 967–973 (2011).
58. Walz, J. M. et al. Simultaneous EEG-fMRI reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem. *J. Neurosci.* **33**, 19212–19222 (2013).
59. Walz, J. M. et al. Simultaneous EEG-fMRI reveals a temporal cascade of task-related and default-mode activations during a simple target detection task. *Neuroimage* **102**, 229–239 (2014).
60. Conroy, B. R., Walz, J. M. & Sajda, P. Fast bootstrapping and permutation testing for assessing reproducibility and interpretability of multivariate fMRI decoding models. *PLoS ONE* **8**, e79271 (2013).
61. Walz, J. M. et al. Prestimulus EEG alpha oscillations modulate task-related fMRI BOLD responses to auditory stimuli. *Neuroimage* **113**, 153–163 (2015).
62. Velanova, K., Wheeler, M. E. & Luna, B. Maturational changes in anterior cingulate and frontoparietal recruitment support the development of error processing and inhibitory control. *Cereb. Cortex* **18**, 2505–2522 (2008).
63. Padmanabhan, A., Geier, C. F., Ordaz, S. J., Teslovich, T. & Luna, B. Developmental changes in brain function underlying the influence of reward processing on inhibitory control. *Dev. Cogn. Neurosci.* **1**, 517–529 (2011).
64. Geier, C. F., Terwilliger, R., Teslovich, T., Velanova, K. & Luna, B. Immaturities in reward processing and its influence on inhibitory control in adolescence. *Cereb. Cortex* **20**, 1613–1629 (2010).
65. Cera, N., Tartaro, A. & Sensi, S. L. Modafinil alters intrinsic functional connectivity of the right posterior insula: a pharmacological resting state fMRI study. *PLoS ONE* **9**, e107145 (2014).
66. Woo, C.-W., Roy, M., Buhle, J. T. & Wager, T. D. Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS Biol.* **13**, e1002036 (2015).
67. Smeets, P. A. M., Kroese, F. M., Evers, C. & de Ridder, D. T. D. Allured or alarmed: counteractive control responses to food temptations in the brain. *Behav. Brain. Res.* **248**, 41–45 (2013).
68. Pernet, C. R. et al. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage* **119**, 164–174 (2015).
69. Verstynen, T. D. The organization and dynamics of corticostriatal pathways link the medial orbitofrontal cortex to future behavioral responses. *J. Neurophysiol.* **112**, 2457–2469 (2014).
70. Bursley, J. K., Nestor, A., Tarr, M. J. & Creswell, J. D. Awake, offline processing during associative learning. *PLoS ONE* **11**, e0127522 (2016).
71. Gabitov, E., Manor, D. & Karni, A. Learning from the other limb's experience: sharing the 'trained' M1 representation of the motor sequence knowledge. *J. Physiol. (Lond.)* **594**, 169–188 (2016).
72. Gabitov, E., Manor, D. & Karni, A. Patterns of modulation in the activity and connectivity of motor cortex during the repeated generation of movement sequences. *J. Cogn. Neurosci.* **27**, 736–751 (2015).
73. Gabitov, E., Manor, D. & Karni, A. Done that: short-term repetition related modulations of motor cortex activity as a stable signature for overnight motor memory consolidation. *J. Cogn. Neurosci.* **26**, 2716–2734 (2014).
74. Lepping, R. J., Atchley, R. A. & Savage, C. R. Development of a validated emotionally provocative musical stimulus set for research. *Psychol. Music* **44**, 1012–1028 (2016).
75. Park, C.-A. & Kang, C.-K. Sensing the effects of mouth breathing by using 3-tesla MRI. *J. Korean Phys. Soc.* **70**, 1070–1076 (2017).
76. Iannilli, E. et al. Effects of manganese exposure on olfactory functions in teenagers: a pilot study. *PLoS ONE* **11**, e0144783 (2016).
77. Kim, J., Wang, J., Wedell, D. H. & Shinkareva, S. V. Identifying core affect in individuals from fMRI responses to dynamic naturalistic audiovisual stimuli. *PLoS ONE* **11**, e0161589 (2016).
78. Tétreault, P. et al. Brain connectivity predicts placebo response across chronic pain clinical trials. *PLoS Biol.* **14**, e1002570 (2016).
79. Chakroff, A. et al. When minds matter for moral judgment: intent information is neurally encoded for harmful but not impure acts. *Soc. Cogn. Affect. Neurosci.* **11**, 476–484 (2016).
80. Koster-Hale, J., Saxe, R., Dungan, J. & Young, L. L. Decoding moral judgments from neuralrepresentations of intentions. *Proc. Natl Acad. Sci. USA* **110**, 5648–5653 (2013).
81. Gao, X. et al. My body looks like that girl's: body mass index modulates brain activity during body image self-reflection among young women. *PLoS ONE* **11**, e0164450 (2016).

82. Romaniuk, L., Pope, M., Nicol, K., Steele, D. & Hall, J. Neural correlates of fears of abandonment and rejection in borderline personality disorder. *Wellcome Open Res.* **1**, 33 (2016).
83. Cohen, A. D., Nencka, A. S., Lebel, R. M. & Wang, Y. Multiband multi-echo imaging of simultaneous oxygenation and flow timeseries for resting state connectivity. *PLoS ONE* **12**, e0169253 (2017).
84. Dalenberg, J. R., Weitkamp, L., Renken, R. J., Nanetti, L. & Ter Horst, G. J. Flavor pleasantness processing in the ventral emotion network. *PLoS ONE* **12**, e0170310 (2017).
85. Roy, A. et al. The evolution of cost-efficiency in neural networks during recovery from traumatic brain injury. *PLoS ONE* **12**, e0170541 (2017).
86. Gordon, E. M. et al. Precision functional mapping of individual human brains. *Neuron* **95**, 791–807.e7 (2017).
87. Veldhuizen, M. G. et al. Integration of Sweet Taste and Metabolism Determines Carbohydrate Reward. *Curr. Biol.* **27**, 2476–2485.e6 (2017).
88. Greene, D. J. et al. Behavioral interventions for reducing head motion during MRI scans in children. *Neuroimage* **171**, 234–245 (2018).
89. Nastase, S. A. et al. Attention selectively reshapes the geometry of distributed semantic representation. *Cereb. Cortex* **27**, 4277–4291 (2017).
90. Kanazawa, Y. et al. Phonological memory in sign language relies on the visuomotor neural system outside the left hemisphere language network. *PLoS ONE* **12**, e0177599 (2017).
91. Tustison, N. J. et al. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
92. Marcus, D. S. et al. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* **19**, 1498–1507 (2007).
93. Nooner, K. B. et al. The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry. *Front. Neurosci.* **6**, 152 (2012).
94. Reuter, M., Rosas, H. D. & Fischl, B. Highly accurate inverse consistent registration: a robust approach. *Neuroimage* **53**, 1181–1196 (2010).
95. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).
96. Klein, A. et al. Mindboggling morphometry of human brains. *PLoS Comput. Biol.* **13**, e1005350 (2017).
97. Fonov, V., Evans, A., McKinstry, R., Almli, C. & Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* **47**, S102 (2009).
98. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image. Anal.* **12**, 26–41 (2008).
99. Klein, A. et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* **46**, 786–802 (2009).
100. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).
101. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
102. Oakes, T. R. et al. Comparison of fMRI motion correction software tools. *Neuroimage* **28**, 529–543 (2005).
103. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).
104. Lanczos, C. Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B: Numerical Analysis* **1**, 76–85 (1964).
105. Wang, S. et al. Evaluation of field map and nonlinear registration methods for correction of susceptibility artifacts in diffusion MRI. *Front. Neuroinform.* **11**, 17 (2017).
106. McIntosh, S., Kamei, Y., Adams, B. & Hassan, A. E. The impact of code review coverage and code review participation on software quality: a case study of the Qt, VTK, and ITK projects. In: P. Devanbu, S. Kim, M. Pinzger eds. *Proc. 11th Working Conference on Mining Software Repositories, MSR 2014* (pp. 192–201. ACM, New York, 2014).
107. Gorgolewski, K. J. et al. BIDS apps: improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Comput. Biol.* **13**, e1005209 (2017).
108. Beaulieu-Jones, B. K. & Greene, C. S. Reproducibility of computational workflows is automated using continuous analysis. *Nat. Biotechnol.* **35**, 342–346 (2017).
109. Kurtzer, G. M., Sochat, V. & Bauer, M. W. Singularity: scientific containers for mobility of compute. *PLoS ONE* **12**, e0177459 (2017).

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