



**HAL**  
open science

# The HCP multi-pipeline dataset: an opportunity to investigate analytical variability in fMRI data analysis

Elodie Germani, Elisa Fromont, Pierre Maurel, Camille Maumet

## ► To cite this version:

Elodie Germani, Elisa Fromont, Pierre Maurel, Camille Maumet. The HCP multi-pipeline dataset: an opportunity to investigate analytical variability in fMRI data analysis. 2024. inserm-04356768v2

**HAL Id: inserm-04356768**

**<https://inserm.hal.science/inserm-04356768v2>**

Preprint submitted on 3 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# HCP Multi-Pipeline: a derived dataset to investigate analytical variability in fMRI

Elodie Germani<sup>1,\*</sup>, Elisa Fromont<sup>2</sup>, Pierre Maurel<sup>1,†</sup>, and Camille Maumet<sup>1,†,\*</sup>

<sup>1</sup>Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, Empenn ERL U 1228, F-35000 Rennes, France

<sup>2</sup>Univ Rennes, IUF, Inria, CNRS, IRISA UMR 6074, F-35000 Rennes, France

\*corresponding author(s): Camille Maumet (camille.maumet@inria.fr), Elodie Germani (elodie.germani@irisa.fr)

†Joint senior authorship

## ABSTRACT

Results of functional Magnetic Resonance Imaging (fMRI) studies can be impacted by many sources of variability including: different sampling strategies for the participants, different acquisition protocols and material but also different analytical choices in the processing of the fMRI data. While variability across participants or across acquisition instruments have been extensively studied in the neuroimaging literature, the root causes of analytical variability remain an open question. Here, we share the *HCP Multi-Pipeline dataset* which provides the resulting statistic maps for 24 typical fMRI pipelines on 1,080 participants of the HCP Young Adult dataset. We share both individual and group results - for 1,000 groups of 50 participants - over 5 motor contrasts. We hope that this large dataset covering a wide range of analysis conditions will provide new opportunities to study analytical variability in fMRI.

## Background & Summary

Neuroimaging data, such as functional Magnetic Resonance Imaging (fMRI), can be used for a wide range of applications, including disease diagnosis<sup>1</sup> or brain decoding (*i.e.* identifying stimuli and cognitive states from brain activities)<sup>2</sup>. But the workflows used to analyze these data are highly complex and flexible. Different tools and algorithms were developed over the years, leaving researchers with many possible choices at each step of an analysis<sup>3</sup>. A given series of operations performed on raw fMRI data is referred to as a ‘pipeline’. Task-fMRI analysis pipelines are composed of three high-level stages (preprocessing, subject-level analysis and group-level analysis). Each stage follows a base architecture with multiple processing steps: this sequence can be customized by the addition or removal of specific steps or modified by using different algorithms or set of parameters. Several software packages are also available to run task-fMRI analysis pipelines, for instance SPM<sup>4</sup>, FSL<sup>5</sup> and AFNI<sup>6</sup>, which are the most commonly used.

In the Neuroimaging Analysis Replication and Prediction Study (NARPS)<sup>7</sup>, 70 research teams were asked to analyze the same task-fMRI dataset with their favorite pipeline to answer the same 9 binary research questions investigating the activation of a particular brain region during a specific cognitive task. Each team used a different pipeline, illustrating perfectly how researchers can have different practices to analyze task-fMRI data. In the end, results varied in terms of final activation maps and conclusion to hypotheses. This phenomenon calls for a better understanding of the pipeline-space to try to identify the cause of the observed differences amongst the final results, and their impact on research findings. For instance, in a recent paper<sup>8</sup>, authors explored whether some specific choices in the pipeline (in particular for the first-level statistical modelling) could improve the reliability of fMRI results, showing that specific choices for contrasts and model parameterization meaningfully improved test–retest reliability, both at the individual and group-level.

The pipeline-space is especially large<sup>9</sup> and challenging to explore due to its interaction with other properties of a dataset: for instance, with sample size and sampling uncertainty<sup>10</sup> or with the research question<sup>7</sup>. However, due to the high computational cost of storing and analyzing task-fMRI data, recent studies investigating analytical variability in neuroimaging focused on a restricted number of participants (N=108, N=30, N=15, and N=10 respectively for<sup>3,7,11,12</sup>) and cognitive tasks (one paradigm for<sup>3,7</sup> with respectively k=9 and k=1 contrasts and use of resting-state fMRI for<sup>11,12</sup>).

Multiple efforts for collecting datasets with larger number of participants and tasks have arisen in the field of neuroimaging in the past 10 years with for instance the Human Connectome Project (HCP)<sup>13</sup> or the UK Biobank<sup>14,15</sup>. In particular, the HCP Young Adult most recent releases provide task-fMRI data for more than 1,000 participants and for different tasks and cognitive processes. These data are also available as minimally processed versions, *i.e.* preprocessed using a common pipeline chosen by the HCP collaborators<sup>16</sup>. In brief, this pipeline consists in the following steps: removal of spatial distortions, volumes realignment to correct for subject motion, registration of the functional volumes to the structural one, bias field reduction,

normalization to a global mean and masking using a structural brain mask computed in parallel.

A set of group-level statistic maps of the HCP Young Adult have also been made publicly available (see NeuroVault Collection 457<sup>17</sup> and corresponding publication<sup>18</sup>). These were obtained using data from a subset of the participants (68 subjects scanned during the first quarter (Q1) of Phase II data collection. Z-scored statistic maps are available for all base contrasts (23 different contrasts) and were computed using a single analysis pipeline. This is beneficial for studying individual differences and contrasts, but it does not allow for analytical variability studies, for which multiple pipelines are needed, or to perform other analyses such as group-level analyses that could be used to explore the effect of sampling uncertainty or sample size.

Statistic maps published during the NARPS study<sup>7</sup> are also publicly available on NeuroVault, with one collection per team. For each of the 70 teams, 9 group-level statistic maps are shared (one per research hypothesis) based on two groups of N=54 participants. Additionally, for a limited number of teams (K=4), subject-level contrast maps are also available. The pipeline space studied in this dataset is unconstrained since teams were instructed to use their usual pipelines to analyze the data.

Here, we share the *HCP Multi-Pipeline dataset*, composed of a large number of subject and group-level statistic maps and representing a non-exhaustive but controlled part of the pipeline space. Contrast and statistic maps are made available for the 5 contrasts of the motor task of the Human Connectome Project (HCP Young Adult) for the 1,080 participants of the S1200 release, obtained with 24 analysis pipelines that differ on a predefined set of parameters as typically used in the literature. We also provide group-level contrast and statistic maps for 1,000 randomly sampled groups of 50 participants for each pipeline and contrast.

While solutions have been proposed to standardized fMRI preprocessing (*e.g.* fMRIprep<sup>19</sup>), practitioners still face multiple choices regarding first-level statistical analyses. Here, we focus on a set of parameters that often vary across pipelines and this even when standardized preprocessing are used: smoothing kernels, HRF modelling and the inclusion/exclusion of motion regressors as nuisance covariates. Group-level statistical analyses were performed uniformly for all pipelines.

## Methods

### Raw Data: the Human Connectome Project

This work was performed using data from the Human Connectome Project Young Adult<sup>13</sup>. Written informed consent was obtained from participants and the original study was approved by the Washington University Institutional Review Board. We agreed to the Open Access Data Use Terms<sup>20</sup>.

The HCP Young Adult aimed to study and share data from young adults (ages 22-35) from families with twins and non-twin siblings, using a protocol that included structural and functional magnetic resonance imaging (MRI, fMRI), diffusion tensor imaging (dMRI) at 3 Tesla (3T) and behavioral and genetic testing. The S1200 release includes behavioral and 3T MR imaging data from 1206 healthy young adult participants (1113 with structural MR scans) collected between 2012 and 2015.

Unprocessed anatomical T1-weighted (T1w) and task-fMRI data<sup>21-24</sup> were used in this work. The task-fMRI data includes seven tasks, each performed in two separate runs. Among these tasks, we selected data from the motor task in which participants were presented with visual cues asking them to tap their fingers (left or right), squeeze their toes (left or right) or move their tongue. This task is the simplest one of the tasks performed in the study, and the protocol associated with this task is very standard and robust. We used unprocessed data for the  $N = 1080$  participants who completed this task.

### Analyses pipelines

Multiple preprocessing and first-level analyses were performed on the task-fMRI data, giving rise to 24 different analysis pipelines. These pipelines differed in 4 parameters:

- Software package: SPM (Statistical Parametric Mapping, RRID: SCR\_007037)<sup>4</sup> or FSL (FMRIB Software Library, RRID: SCR\_002823)<sup>5</sup>.
- Smoothing kernel: Full-Width at Half-Maximum (FWHM) was equal to either 5mm or 8mm.
- Number of motion regressors included in the General Linear Model (GLM) for the first-level analysis: 0, 6 (3 rotations, 3 translations) or 24 (the 6 previous regressors + 6 derivatives and the 12 corresponding squares of regressors).
- Presence (1) or absence (0) of the derivatives of the Hemodynamic Response Function (HRF) in the GLM for the first-level analysis. Only the temporal derivatives were added in FSL pipelines and both the temporal and dispersion derivatives were for SPM pipelines.

In the following, we will denote the pipelines by “software-FWHM-number of motion regressors-presence of hrf derivatives”. For instance, pipeline with software FSL, smoothing with a kernel FWHM of 8mm, no motion regressors and no hrf derivatives will be denoted by “fsl-8-0-0”.

All pipelines were implemented using Nipype version 1.6.0 (RRID: SCR\_002502)<sup>25</sup>, a Python project that provides a uniform interface to existing neuroimaging software packages and facilitates interaction between these packages within a single workflow. All pipelines scripts are available online at [https://gitlab.inria.fr/egermani/hcp\\_pipelines](https://gitlab.inria.fr/egermani/hcp_pipelines) and archived on Software Heritage: [swh:1:snp:17870c3d782aa25a7ffdd6165fe27ce6eac6c90b](https://swh:1:snp:17870c3d782aa25a7ffdd6165fe27ce6eac6c90b)

### **Computing environment**

To limit the variability induced by different computer environments and versions of the software packages, we used NeuroDocker (RRID: SCR\_017426)<sup>26</sup> to generate a custom Docker image. To build this image, we chose NeuroDebian<sup>27</sup> and installed the following software packages: FSL version 6.0.3 and SPM12 release r7771. To install Python and Nipype, we created a Miniconda3 environment with Python version 3.8 and multiple packages, such as Nilearn<sup>28</sup> (RRID: SCR\_001362), Nipype and NiBabel (RRID: SCR\_002498)<sup>29</sup>. This docker image is available at [https://hub.docker.com/repository/docker/elodiegermani/open\\_pipeline/general](https://hub.docker.com/repository/docker/elodiegermani/open_pipeline/general) and the command to generate the DockerFile can be found in the README included in the codebase (see links above).

### **Preprocessing**

Preprocessing consisted of the following steps for all pipelines: spatial realignment of the functional data to correct for motion, coregistration of realigned data towards the structural data, segmentation of the structural data, non-linear registration of the structural and functional data towards a common space and smoothing of the functional data. Depending on the software package used, these steps were performed in a different order, following the default behavior of the software package.

In SPM, for each participant, functional data were first spatially realigned to the mean volume using the “Realign: Estimate and Reslice” function with default parameters (quality of 0.9, sampling distance of 4 and a smoothing kernel, 2nd degree B-spline interpolation and no wrapping). Realigned functional data were then coregistered, with the “Coregister: Estimate” function, to the anatomical T1w volume acquired for the participant using Normalized Mutual Information. In parallel, we segmented the different tissue classes of the same anatomical T1w volume using the “Segment” function. The forward deformation field provided by the segmentation step was used to normalize the functional data to a standard space (MNI) (“Normalize: Write” function) with a voxel size of 2mm and a 4th degree B-spline interpolation. Normalized functional data were then smoothed with a Gaussian kernel with different FWHM values depending on the pipeline (5 or 8mm).

In FSL, we reproduced the preprocessing steps used in FEAT<sup>30</sup>. Functional data were realigned to the middle functional volume using MCFLIRT. Brain extraction was applied with BET and we masked the functional data using the extracted mask. We smoothed each run using SUSAN with the brightness threshold set to 75% of the median value for each run and a mask constituting the mean functional. Different values were used for the FWHM of the smoothing kernel depending on the pipeline. We also performed temporal highpass filtering on the functional data with a thresholding value of 100 seconds. In parallel, we computed the transformation matrix to register functional data to anatomical and standard space (MNI) using linear (FLIRT function) and non-linear registration (FNIRT function). Contrary to SPM, the first-level statistical analysis is performed on the smoothed data in subject-space. Only the transformation matrix was computed at this stage, using boundary-based registration and applied on the contrast maps output after the statistical analysis.

### **First level statistical analyses**

To obtain the contrast maps of the different participants and contrasts, we modeled the data using a GLM. Each event was modelled using the onsets and durations provided in the event files of the HCP Young Adult dataset. Six regressors were modeled: cue (which represent the visual cues), right hand, right foot, left hand, left foot and tongue. Each condition was convolved with the canonical HRF. For both SPM and FSL pipelines, we used the Double Gamma HRF (default in SPM).

Different numbers of motion regressors (0, 6 or 24) were included in the design matrix to regress out motion-related fluctuations in the BOLD signal. The modelling of the HRF also varied: Double Gamma HRF with or without derivatives (time+dispersion for SPM and time for FSL).

In SPM, temporal autocorrelations in the BOLD signal timeseries were accounted for by highpass filtering with a 128 seconds filter cutoff and modelling of serial correlation using an autoregressive model of the first order (AR(1)). In FSL, highpass filtering was already performed during preprocessing with a 100 seconds filter cutoff, modelling of serial correlation was also performed using an AR(1) model. Model parameters were estimated using a Restricted Maximum Likelihood approach for both SPM and FSL software packages. Subject-level contrast maps were computed and saved for 5 contrasts (right hand, right foot, left hand, left foot and tongue) for each participant. In the end, for each of the 24 pipelines, we had 5,400 contrast and statistic maps (5 contrasts for each of the 1,080 participants). These maps constituted the subject-level dataset. Figure 1(A) presents the statistic maps for the contrast *right-hand* obtained with the different pipelines for a representative subject.

### **Second level statistical analyses**

Group-level statistical analyses were performed using the contrast maps obtained with the different first-level analyses pipelines. 1,000 groups of 50 participants were randomly sampled among the 1,080 participants. For each analysis pipeline, we performed

one sample t-tests for each group and each contrast in SPM. We purposely used the same second-level analysis method and software for all pipelines in order to focus on first-level analysis differences. For each of the 24 pipelines, the group-level dataset was thus composed of 5,000 contrast maps and statistic maps (5 contrasts for each on the 1,000 groups). Figure 1(B) presents the statistic maps obtained with the different pipelines for one group for the contrast *right-hand*.

## Data Records

The data are accessible on Public nEUro<sup>31</sup> at <https://publicneuro-catalogue.netlify.app/dataset/PN000003:%20HCP%20multipipelines/V1>. The dataset is organized according to the BIDS specification<sup>32</sup>, but to date, there is no established specification for organizing datasets with statistic maps for different pipelines and different processing levels (here, subject and group-levels derivatives). We used a modlisation that is currently under development in the BIDS Extension Proposal: BEP041 - Statistical Model Derivatives.

Briefly, the dataset contains one folder per pipeline (*i.e.* 24 folders in total). Each pipeline folder is composed of two subfolders respectively named “node-L1” and “node-L2” for subject-level and group-level derivatives. The “node-L1” folder includes subject-level folders, named after the BIDS specifications, and the “node-L2” folder is composed of group-level folders, named ‘group-X’ depending on the group number. Group and subject-level folders contain the contrast and statistic maps with the following specification for contrast maps: ‘sub/group-<ID>\_task-<task>\_space-<space>\_contrast-<contrast>\_stat-effect\_statmap.nii.gz’, and T-statistic maps: ‘sub/group-<ID>\_task-<task>\_space-<space>\_contrast-<contrast>\_stat-t\_statmap.nii.gz’.

## Technical Validation

To assess the quality of the statistic maps, we checked that all contrasts led to an activation of the primary motor area (Primary Motor Cortex M1).

As described in Fig. 2, we looked at the significant activations inside the Primary Motor Cortex (M1) of the brain for the statistic maps of each group, each contrast and each pipeline. Our group-level statistic maps were thresholded using a False Discovery Rate (FDR) corrected voxelwise p-value of  $p < 0.05$  and masked using the probabilistic Juelich Atlas<sup>33</sup> available in Nilearn. We selected the region of interest (ROI) corresponding to the Primary Motor Cortex M1, Brodmann Area 4. Depending on the contrast, both left and right hemisfer’s ROI (“tongue”), only the left hemisfer (“right hand” or “right foot”) or only the right hemisfer (“left hand” or “left foot”) ROI were selected, to focus on controlateral activations in the motor cortex.

For each map, we computed the percentage of activation inside the Primary Motor Cortex, which is the percentage of voxels of the ROI that are activated, *i.e.*:

$$\text{Percentage of Activation} = \frac{N_{\text{activated voxels}}}{N_{\text{total voxels}}} \times 100 \quad (1)$$

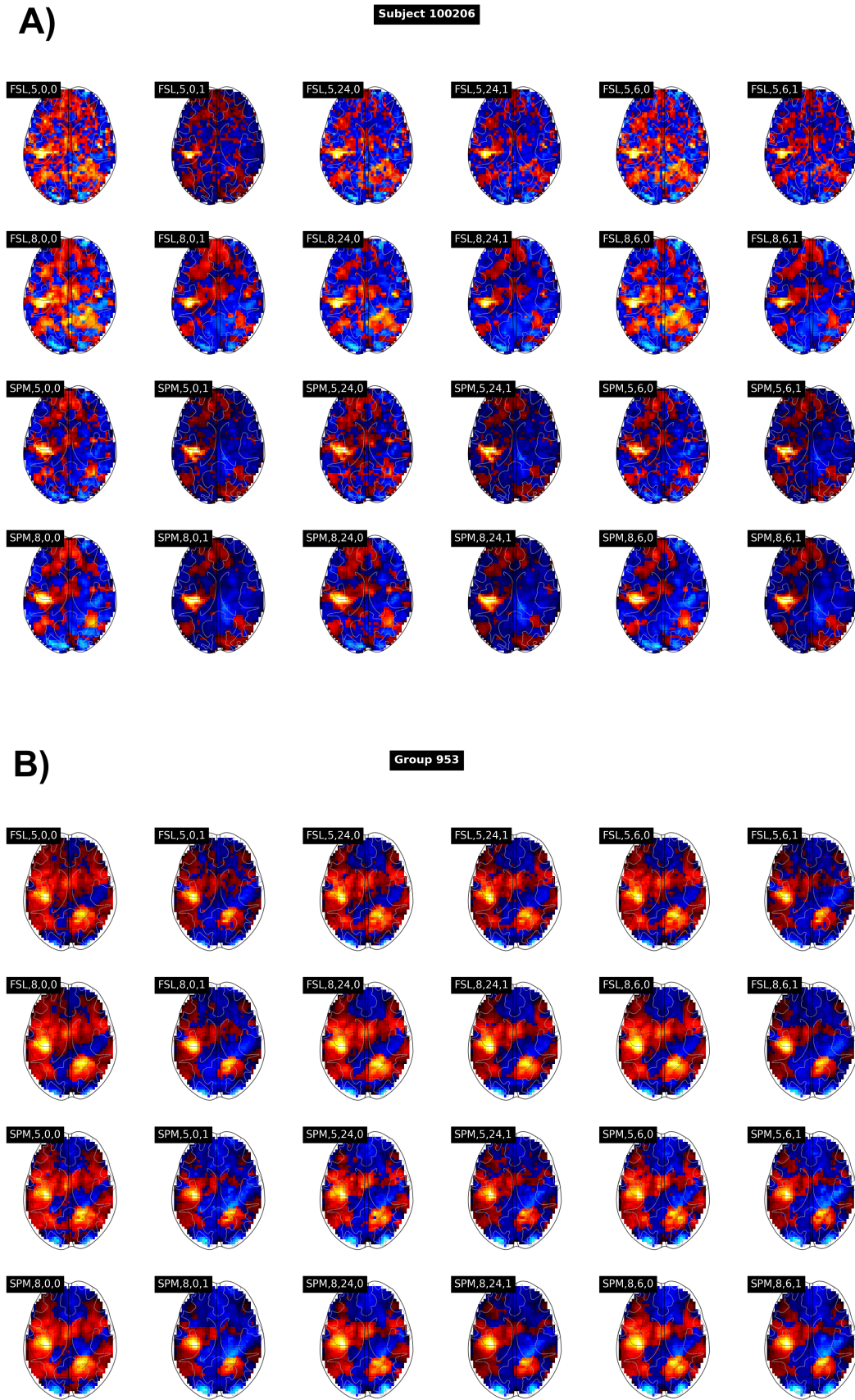
where  $N_{\text{activated voxels}}$  is the number of activated voxels in the ROI and  $N_{\text{total voxels}}$  is the total number of voxels in the ROI.

Fig. 3 represents the distribution of percentages of activation inside the Primary Motor Cortex per contrast for all groups and all pipelines. Results were slightly different across contrasts: mean percentages of activation were all between 20 and 40% but those of contrasts *left-foot* and *right-foot* were below those of *right-hand*, *left-hand* and *tongue*. When looking at the activations of different contrasts in the ROI for one of our group-level statistic maps (see Fig. 4), we could see that the activations detected for the *right-foot* contrast seemed widespread with a small area of high activation. For the *right-hand* contrast, the high activation area was larger and covered nearly the entire ROI. This observation was consistent with the literature<sup>34,35</sup> and with statistic maps obtained from NeuroSynth<sup>35</sup> (RRID:SCR\_006798) in which the identified area of activation of the foot was smaller than the hand one (see Fig. 4). In the Primary Motor Area, the statistic maps of the foot contrasts thus have less activated voxels. Overall, the technical validation was successful. The goal of this quality check was to have a quick estimation of the accuracy of the statistic maps to represent the task performed, thus we chose to define a single ROI covering the entire motor area.

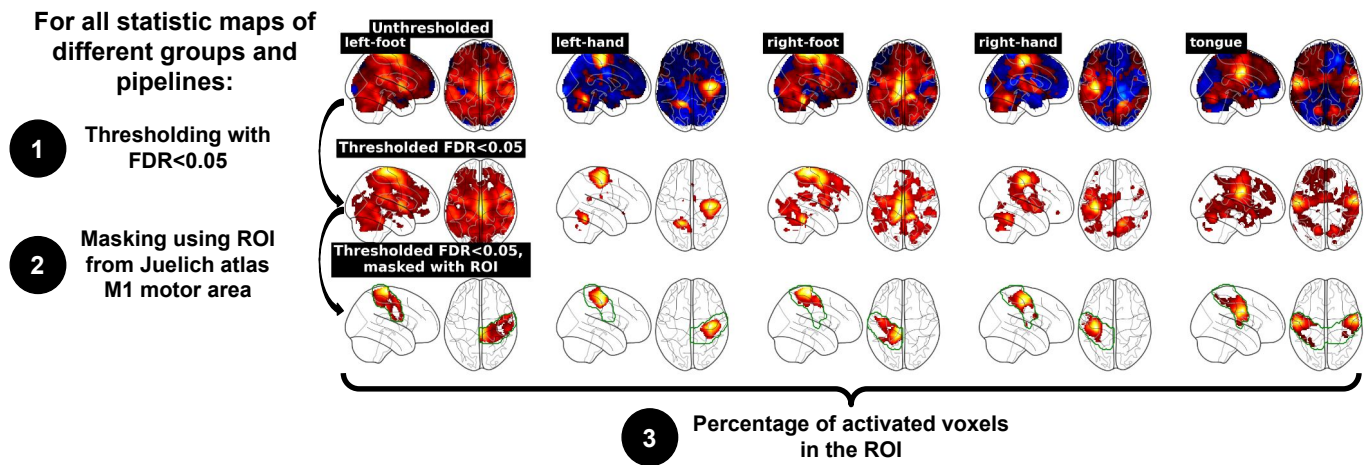
We observed consistent metrics across pipelines, with high percentages of activation for hand and tongue contrasts and lower ones for foot contrasts. An example of the distribution of percentage of activations for all group maps of each contrast is shown in Fig. 5 for the pipeline “spm-5-0-0”. The percentage of activation computed for all groups, pipelines and contrasts are available in Supplementary Files.

## Usage Notes

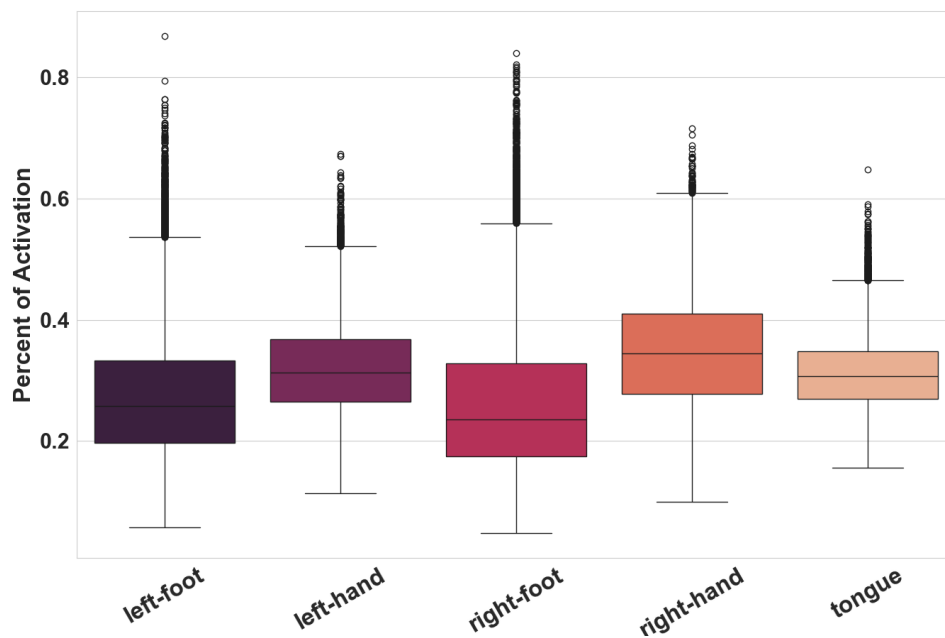
The *HCP Multi-Pipeline dataset* provides researchers with a re-usable dataset of fMRI contrast and statistic maps. The data is accessible on Public nEUro<sup>31</sup> at <https://publicneuro-catalogue.netlify.app/dataset/PN000003:%20HCP%20multipipelines/V1..>



**Figure 1.** Example of subject (A) and group-level (B) statistic maps obtained for subject 100206 and group 953 for each pipeline for the contrast *right-hand*. Pipelines are denoted by  $\langle \text{software} \rangle, \langle \text{FWHM} \rangle, \langle \text{motion regressors} \rangle, \langle \text{hrf derivatives} \rangle$ .



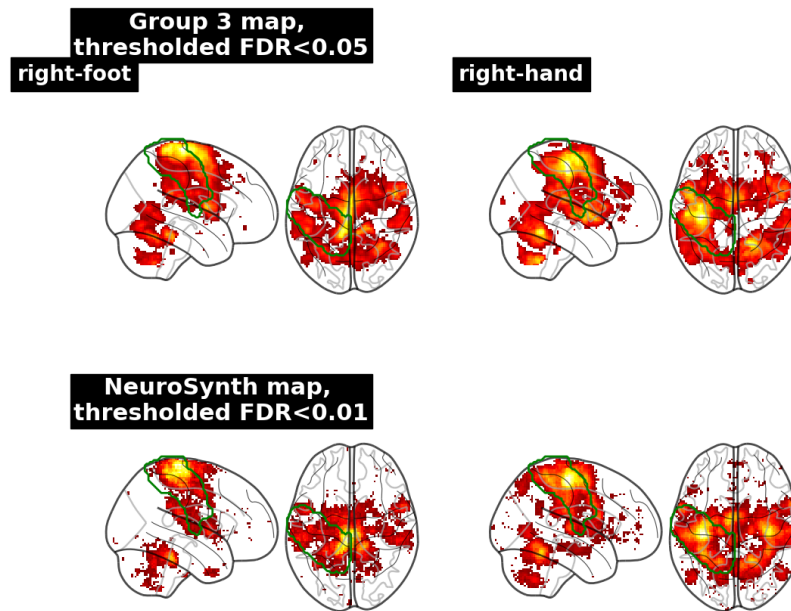
**Figure 2.** Workflow of technical validation of statistic maps. We thresholded the statistic maps of each group, each pipeline and each contrast using a FDR-corrected voxelwise  $p < 0.05$  and masked the thresholded maps using the ROI of the Primary Motor Cortex from Juelich atlas. Then, we computed the percentage of activated voxels inside this ROI.



**Figure 3.** Distribution of mean Percentage of Activation inside the Primary Motor Cortex across all groups and pipelines for the five contrasts under study.

This dataset brings together a wide range of analysis conditions, covering many aspects of inter-subject, inter-groups, inter-contrasts and inter-pipelines variability. Data from 1,080 participants were used to form 1,000 different groups of 50 participants, and 5 contrasts were analyzed with 24 different pipelines, which offers an unprecedented opportunity to explore analytical variability, understand its causes and how it may interact with other sources of variability.

Analytical variability is not limited to neuroimaging and has been studied in many other disciplines<sup>36</sup>, such as psychology<sup>37</sup> or software engineering<sup>38</sup>. These different fields have brought solutions to explore and handle analytical variability. These techniques have begun to be used in neuroimaging, with, for instance, the implementation of continuous integration, a software engineering technique, to facilitate the reproducibility of neuroimaging computational experiments<sup>39</sup> or multiverse analyses



**Figure 4.** Thresholded statistic maps for contrasts *right foot* (right) and *right hand* (left) for group-level analysis of group 3 with pipeline spm-5-0-0 (upper). Percentage of Activation inside the Primary Motor Cortex were respectively 0.34 and 0.41 for the contrasts *right foot* and *right hand*. NeuroSynth activation maps corresponding to the forward inference of the “hand” and “foot” paradigms (lower). Green borders correspond to the motor area ROI.

that help to find the most efficient pipelines depending on the data and the goal of the study<sup>40</sup>.

By sharing directly the results obtained from different analysis strategies, we hope to facilitate the use of these data by researchers from other fields, that could apply their own methods to help explore the neuroimaging analytical space. For instance, this dataset could be used to extend the study from<sup>8</sup> to other tasks. Using the code provided to create the pipelines, other researchers could be able to enhance this dataset with other combinations of parameters, giving rise to other pipelines, or could apply the pipelines to other participants, groups or contrasts.

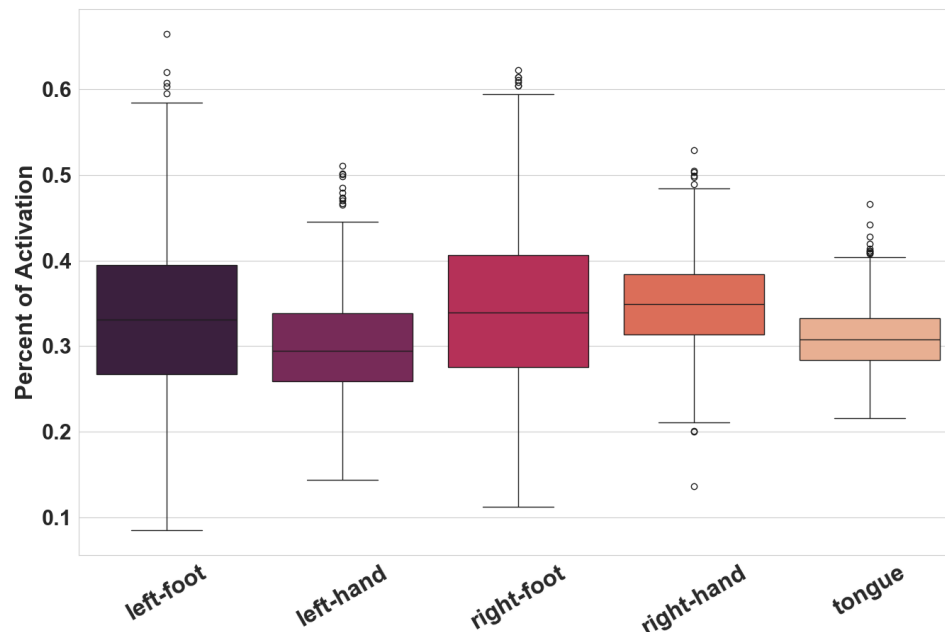
## Code availability

All codes (analysis pipelines and technical validation) were executed in Python v3.8. The executions require the installation of SPM and FSL software packages. To facilitate reproducibility, we provide a NeuroDocker image that can be pulled from Dockerhub and that contains all necessary software packages. The Docker image is available at: [https://hub.docker.com/r/elodiegermani/open\\_pipeline](https://hub.docker.com/r/elodiegermani/open_pipeline). Python scripts to create and run the pipelines as well as to perform technical validation are available publicly at [https://gitlab.inria.fr/egermani/hcp\\_pipelines](https://gitlab.inria.fr/egermani/hcp_pipelines) and archived on Software Heritage at [swh:1:snp:17870c3d782aa25a7ffdd6165fe27ce6eac6c90b](https://swh.inria.fr/obj/fr:17870c3d782aa25a7ffdd6165fe27ce6eac6c90b).

## References

1. Yin, W., Li, L. & Wu, F.-X. Deep learning for brain disorder diagnosis based on fMRI images. *Neurocomputing* **469**, 332–345, <https://www.doi.org/10.1016/j.neucom.2020.05.113> (2022).
2. Firat, O., Oztekin, L. & Vural, F. T. Y. Deep learning for brain decoding. In *2014 IEEE International Conference on Image Processing (ICIP)*, 2784–2788, <https://www.doi.org/10.1109/ICIP.2014.7025563> (IEEE, 2014).
3. Carp, J. On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. *Front. Neurosci.* <https://www.doi.org/10.3389/fnins.2012.00149> (2012).





**Figure 5.** Distribution of Percentage of Activation inside the Primary Motor Cortex for all group-level statistic maps for pipeline “spm-5-0-0” in the different contrast maps.

4. Penny, W., Friston, K., Ashburner, J., Kiebel, S. & Nichols, T. E. *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (2011), elsevier edn.
5. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790, <https://www.doi.org/10.1016/j.neuroimage.2011.09.015> (2012).
6. Cox, R. W. AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Comput. Biomed. Res.* **29**, 162–173, [10.1006/cbmr.1996.0014](https://doi.org/10.1006/cbmr.1996.0014) (1996).
7. Botvinik-Nezer, R. *et al.* Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88, <https://www.doi.org/10.1038/s41586-020-2314-9> (2020).
8. Demidenko, M. I., Mumford, J. A. & Poldrack, R. A. Impact of analytic decisions on test-retest reliability of individual and group estimates in functional magnetic resonance imaging: a multiverse analysis using the monetary incentive delay task, [10.1101/2024.03.19.585755](https://doi.org/10.1101/2024.03.19.585755) (2024).
9. Carp, J. The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage* **63**, 289–300, <https://www.doi.org/10.1016/j.neuroimage.2012.07.004> (2012).
10. Klau, S. *et al.* Comparing the vibration of effects due to model, data pre-processing and sampling uncertainty on a large data set in personality psychology. *Meta-Psychology* **7** (2023).
11. Li, X. *et al.* Moving beyond processing and analysis-related variation in neuroscience. 2021–12 (2021).
12. Xu, T. *et al.* A Guide for Quantifying and Optimizing Measurement Reliability for the Study of Individual Differences, [10.1101/2022.01.27.478100](https://doi.org/10.1101/2022.01.27.478100) (2022).
13. Van Essen, D. *et al.* The human connectome project: A data acquisition perspective. *NeuroImage* **62**, 2222–2231, <https://doi.org/10.1016/j.neuroimage.2012.02.018> (2012).
14. Sudlow, C. *et al.* UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine* **12**, <https://www.doi.org/10.1371/journal.pmed.1001779> (2015).
15. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* **19**, 1523–1536, [10.1038/nn.4393](https://doi.org/10.1038/nn.4393) (2016).

16. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the human connectome project. *NeuroImage* **80**, 105–124, <https://doi.org/10.1016/j.neuroimage.2013.04.127> (2013).
17. Collection n°457. NeuroVault Collection n°457. <https://identifiers.org/neurovault.collection:457>. Accessed: 2023-05-20.
18. Van Essen, D. C. *et al.* The wu-minn human connectome project: An overview. *NeuroImage* **80**, 62–79, <https://doi.org/10.1016/j.neuroimage.2013.05.041> (2013).
19. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116, [10.1038/s41592-018-0235-4](https://doi.org/10.1038/s41592-018-0235-4) (2019).
20. Human connectome project: Data usage agreement. <https://www.humanconnectome.org/study/hcp-young-adult/document/wu-minn-hcp-consortium-open-access-data-use-terms>.
21. Moeller, S. *et al.* Multiband multislice ge-epi at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fmri. *Magn. Reson. Medicine* **63**, 1144–1153, <https://doi.org/10.1002/mrm.22361> (2010).
22. Feinberg, D. A. *et al.* Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. *PLOS ONE* **5**, 1–11, <https://doi.org/10.1371/journal.pone.0015710> (2010).
23. Setsompop, K. *et al.* Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn. Reson. Medicine* **67**, 1210–1224, <https://doi.org/10.1002/mrm.23097> (2012).
24. Xu, J. *et al.* Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband epi. In *Proceedings of the 20th Annual Meeting of ISMRM*, vol. 2306, 1907–1913 (2012).
25. Gorgolewski, K. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinformatics* **15**, <https://www.doi.org/10.5281/zenodo.581704> (2017).
26. Neurodocker.
27. Halchenko, Y. & Hanke, M. Open is Not Enough. Let's Take the Next Step: An Integrated, Community-Driven Computing Platform for Neuroscience. *Front. Neuroinformatics* **6**, <https://www.doi.org/10.3389/fninf.2012.00022> (2012).
28. Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Front. Neuroinformatics* **10.3389/fninf.2014.00014** (2014).
29. Brett, M. *et al.* nipy/nibabel: 3.2.1, [10.5281/zenodo.4295521](https://www.doi.org/10.5281/zenodo.4295521) (2020).
30. Woolrich, M. W., Ripley, B. D., Brady, M. & Smith, S. M. Temporal autocorrelation in univariate linear modeling of fmri data. *NeuroImage* **14**, 1370–1386, <https://doi.org/10.1006/nimg.2001.0931> (2001).
31. Public nEUro.
32. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044, [10.1038/sdata.2016.44](https://doi.org/10.1038/sdata.2016.44) (2016).
33. Amunts, K., Mohlberg, H., Bludau, S. & Zilles, K. Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science* **369**, 988–992, [10.1126/science.abb4588](https://doi.org/10.1126/science.abb4588) (2020).
34. Schott, G. Penfield's homunculus: a note on cerebral cartography. *J Neurol Neurosurg Psychiatry* **56**, 329–333, [10.1136/jnnp.56.4.329](https://doi.org/10.1136/jnnp.56.4.329). (1993).
35. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670, [10.1038/nmeth.1635](https://doi.org/10.1038/nmeth.1635) (2011).
36. Hoffmann, S. *et al.* The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. *Royal Soc. open science* **8**, 201925 (2021).
37. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366, [10.1177/0956797611417632](https://doi.org/10.1177/0956797611417632) (2011).
38. Alférez, M., Acher, M., Galindo, J. A., Baudry, B. & Benavides, D. Modeling variability in the video domain: language and experience report. *Softw. Qual. J.* **27**, 307–347, [10.1007/s11219-017-9400-8](https://doi.org/10.1007/s11219-017-9400-8) (2019).
39. Sanz-Robinson, J., Jahanpour, A., Phillips, N., Glatard, T. & Poline, J.-B. NeuroCI: Continuous integration of neuroimaging results across software pipelines and datasets. 105–116, [10.1109/eScience55777.2022.00025](https://doi.org/10.1109/eScience55777.2022.00025) (2022).
40. Dafflon, J. *et al.* A guided multiverse study of neuroimaging analyses. *Nat. Commun.* **13**, 3758 (2022).

## Acknowledgements

This work was partially funded by Region Bretagne (ARED MAPIS) and Agence Nationale pour la Recherche for the program of doctoral contracts in artificial intelligence (project ANR-20-THIA-0018). Data used in the HCP dataset were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

We thank the Inria open data team and in particular Jozefina Sadowska and Maud Medves who made it possible to share this dataset. We also thank our data protection officers, security, finance and legal teams and in particular Caroline Surquain for her continued support throughout the process. Finally we would like to thank Remi Gau and Chris Markiewicz for their guidance to format the dataset in compliance with BIDS.

## Author contributions

**Elodie Germani:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft. **Elisa Fromont:** Supervision, Writing - Review & Editing. **Pierre Maurel:** Supervision, Writing - Review & Editing. **Camille Maumet:** Supervision, Writing - Review & Editing.

## Competing interests

Camille Maumet is an Editorial Board Member for Scientific Data, handling manuscripts for the journal and advising on its policy and operations. The other authors declare no competing interests.