Rodent Automated Bold Improvement of EPI Sequences (RABIES): A standardized image processing and data quality platform for rodent fMRI.

Gabriel Desrosiers-Gregoire[1,2], Gabriel A. Devenyi[2,3], Joanes Grandjean[4,5], M. Mallar Chakravarty[1,2,3,6]

1. Integrated Program in Neuroscience, McGill University, Montreal, Canada
2. Cerebral Imaging Center, Douglas Mental Health University Institute, Montreal, Canada
3. Department of Psychiatry, McGill University, Canada.
4. Donders Institute for Brain, Cognition, and Behaviour, Donders Institute, The Netherlands
5. Department of Medical Imaging, Radboud University Medical Center, The Netherlands
6. Biological & Biomedical Engineering, McGill University, Canada

ABSTRACT

Functional magnetic resonance imaging (fMRI) in rodents holds great potential for advancing our understanding of brain networks. However, rodent fMRI lags the human community with regards to the availability of image processing softwares to support reproducibility, a significant challenge facing neuroimaging. To address these shortcomings, we introduce RABIES, a novel pipeline designed along best practices for reproducibility and transparency handling processing stages from image registration to data quality assessment during analysis. First, we present a robust registration workflow validated across both acquisition sites and species. Leveraging insights from the human literature, we introduce a novel framework for data quality assessment of single scans and during group-level network analysis. This framework will support the development of standardized high-quality acquisition protocols and harmonization of findings between studies. Taken together, the RABIES software will allow the emerging community to adopt reproducible practices and foster progress in translational neuroscience.

INTRODUCTION

Functional magnetic resonance imaging (fMRI) experiments in rodents have enabled important investigations regarding brain-wide activity (Chuang & Nasrallah, 2017). Unlike many neuroscientific tools, fMRI allows for whole brain coverage using longitudinal experimental designs. This presents a significant advantage for investigations in model organisms as fMRI analyses can be coupled with other techniques like transcriptomics (Mills et al., 2018), calcium imaging (using wide-field (Lake et al., 2020) or fibre-photometry methods (Schlegel et al., 2018)), optogenetics (Grandjean et al., 2019), and more. This represents a significant opportunity for neuroscientific investigation that allows for the bridging of measures of brain function across different spatial and temporal resolution.

Identifying brain networks in the rodent may be critical for establishing functional correspondence to canonical brain networks in humans (Grandjean et al., 2020; Sforazzini et al., 2014; Sierakowiak et al., 2015). Reproducibility is a major concern for the neuroimaging community, leading to elegant solutions for human research applications (Botvinik-Nezer et al., 2020; Eklund et al., 2016; Poldrack et al., 2017; Turner et al., 2018), but complete tailored solutions have not yet emerged for rodents. Reproducibility concerns are now emerging within the rodent community (Grandjean et al., 2020, 2022; Ioanas et al., 2021; Mandino et al., 2020), echoing the human litterature on aspects of image processing, analysis and data quality assessment, but also raising additional challenges pertaining to rodent acquisition, including whether to scan animals awake or sedated and the selection of anesthetic type (Ioanas et al., 2021; Mandino et al., 2020). Without careful consideration, each stage of the study design can introduce confounds into downstream analysis, and left uncorrected, these factors can bias downstream statistical inference (Grandjean et al., 2020; Power et al., 2012).

The distribution of open-source software platforms has become a cornerstone of reproducibility for the human community, for instance with the introduction of pipelines such as fMRIPrep (Esteban et al., 2019) and C-PAC (Craddock et al., 2013). However, these tools optimized for human applications present several limitations when adapted for rodent images. Developing a robust image processing workflow for rodent acquisitions presents new challenges because of the smaller brain size and the site variability in field strengths (4.7T, 7T, 9.4T, 11.7T), coil type (room-temperature vs cryogenic) and shape (i.e. surface vs saddle vs quadrature coil), acquisition parameters (Grandjean et al., 2020; Ioanas et al., 2021). This leads to downstream variability in acquisition artefacts, anatomical contrast, and the extent of brain coverage. Recently, different rodent-adapted software pipelines were introduced for either mice and rats (Celestine et al., 2020; Diao et al., 2021; Ioanas et al., 2021; Lee et al., 2021), but a thorough validation of preprocessing quality across rodent acquisition sites and species is still missing. Additionally, following image processing, the stage of image analysis requires considerations regarding confound correction and data quality. In the human litterature, careful investigations into confound characterization have led to the identification and correction of data quality inconsistencies (Power et al., 2012, 2017; Satterthwaite et al., 2013), and increasingly, recommendations are becoming available in this regard for data cleaning and quality control (Ciric et al., 2018; Parkes et al., 2018). Comparatively in rodents, previous work described confound sources using independent component analysis (ICA) and demonstrated the efficacy

of ICA denoising at reducing inter- and within-subject variability (Diao et al., 2021; Zerbi et al., 2015), but systemic guidelines for determining the influence of confounds on analysis results are missing. A recent meta-analysis demonstrated substantial variability across mouse fMRI sites in their ability to detect canonical brain networks reliably (Grandjean et al., 2020). Using seed-based connectivity to establish reliability (e.g.: a seed in the anterior cingulate cortex should show connectivity with the retrosplenial cortex), only 38% of all scans displayed the expected connectivity features. Despite partial associations with acquisition parameters, the sources leading to analysis failures remained largely unclear. These findings were also recently reproduced in rats, incorporating datasets from 46 different sites and identifying expected connectivity in 40.8% of all scans (Grandjean et al., 2022). The authors found, however, that by standardizing key aspects of scan acquisition across 20 sites, this could be improved to 61.8%, highlighting the importance of improving standards for acquisition and data quality evaluation.

In light of these current issues, we present here Rodent Automated Bold Improvement of EPI Sequences (RABIES; publicly available to the community https://github.com/CoBrALab/RABIES), a novel software platform that integrates the latest computational framework to deliver a solution for preprocessing, confound correction, and analysis, tailored for rodent neuroimaging. RABIES was recently applied to successfully process and analyze datasets from 46 rat fMRI acquisition sites (Grandjean et al., 2022). Built around principles of reproducibility and transparency, RABIES aims to provide rodent fMRI research with an accessible and reliable image processing platform complemented with data diagnostic tools encouraging best practices for quality control. This is achieved by innovating on two fronts which are covered in two distinct results sections: 1) implementing an adaptive image registration workflow offering consistent high-quality outputs across sites and species, with minimal user intervention, and 2) leveraging the human literature to develop data diagnostic features allowing the characterization of confounds in individual scans, and then building on these observations to provide formal recommendations for the quality control of network analysis.

RESULTS

PART 1: A ROBUST AND REPRODUCIBLE IMAGE PROCESSING WORKFLOW FOR RODENT FMRI

An fMRIPrep-inspired software design following best practices for reproducible science and transparency

RABIES is an open-source software largely inspired by fMRIPrep's design. We made three important software design choices that follow best practices for reproducibility in

neuroimaging research: 1) the input data format requires that users follow the BIDS standards (K. J. Gorgolewski et al., 2016); 2) the preprocessing workflow automatically generates visual outputs for QC; and 3) the distribution of the software via containerized versions on Docker and Singularity platforms (https://hub.docker.com/r/gabdesgreg/rabies).

Also inspired from fMRIPrep, the pipeline's core architecture is built using nipype (K. Gorgolewski et al., 2011). Nipype orders image processing steps into an execution graph, termed workflow, which allows both for optimal parallelization, maximizing efficiency, and for a flexible organization of the pipeline according to user inputs. Using information from the BIDS input format, RABIES' preprocessing workflow re-organizes around dataset-specific features, including the number of subjects, the availability of structural images, and the presence of multiple fMRI sessions. RABIES encourages transparency by providing extensive log reports and version control accompanied with automated visual outputs from each failure-prone preprocessing steps (e.g. inhomogeneity correction or commonspace registration) within a PNG format. These figures are ideal for use in presentation and publication.

The major pipeline steps include a robust registration workflow conducting preprocessing prior to group-level analysis (**figure 1A**), QC outputs for verification of every registration operation (**figure 1B**), confound correction (**figure 1C**), and the availability of analysis methods (**figure 1D**). These steps are further described in the following sections.

A generalizable registration workflow across rodent species and acquisition sites

To evaluate the robustness of our registration workflow, we used multi-site fMRI that spanned several field strengths (4.7T, 7T, 9.4T, and 11.7T), different coil types (room temperature and cryogenically cooled coils), differing anesthesia protocols (awake, isoflurane [iso], medetomidine [med], med and iso, or halothane), ventilation protocols (free breathing and ventilated), and rodent species (mice and rats) (23 publicly available datasets total, 441 mouse and 232 rat echo-planar imaging (EPI) scans, 29 EPI scans per dataset on average; **sup. table 1**). The goal of the registration framework is to align all EPI data to a common space that also contains a whole brain parcellation. The parcellation can be defined as the default parcellations available in RABIES for mouse (Dorr et al., 2008; Richards et al., 2011; Steadman et al., 2014; Ullmann et al., 2013), or can be input by the user. The registration workflow accommodates EPI datasets with or without the presence of a neuroanatomical scan, but uses the neuroanatomical scan where possible. We have identified critical processing steps in the pipeline that should be verified to ensure proper completion of the registration workflow (**method section 4; sup. figure 1)**); these include 1) masking of both EPI and anatomical scans for inhomogeneity correction; 2) a dataset-specific template generated through group-wise linear and nonlinear alignment of anatomical scans or EPI scans when anatomical scans are not available, which we will refer here as the unbiased template (https://github.com/CoBrALab/optimized_antsMultivariateTemplateConstruction; (Avants et al., 2011)); 3) registration of the unbiased template to the defined reference atlas for anatomical parcellation (Avants et al., 2008); and 4) nonlinear alignment between the EPI and anatomical scans (Avants et al., 2008; Wang et al., 2017). The

workflow thus regroups four key registration operations that must be visually inspected in the automated QC report offered by RABIES (**figure 1B**).

We sought to ensure high-quality commonspace alignment to the reference atlas in all 23 datasets. This necessitated several innovations (**sup. table 2**) to account for the extensive variability in inhomogeneity artefacts, brain coverage and anatomical contrast of the EPI images (**sup. figure 2**). We developed a scale-independent registration algorithm, wrapped around antsRegistration (Avants et al., 2008), which automatically adapts parameters to species-specific physical dimensions[1]. Brain masking and altas registration in datasets lacking structural images was improved by creating an EPI mouse template and computing masks using Otsu thresholding (Otsu, 1979) rather than registration for certain datasets. Group-wise registration was also improved by integrating regions of interest defined by the improved brain mask. Despite a common workflow, mild differences in preprocessing parameters per dataset were necessary to successfully process every scan (listed in **sup. table 3)**.

In the improved workflow, failures were marginal (**sup. table 1; sup. figure 3**). For the 17 mouse datasets lacking structural images, we achieved a 99.6% (254/255) success rate for the inhomogeneity correction of EPIs, 99.6% (253/254) success during cross-subject alignment to the unbiased template. For the 3 mouse datasets including supporting structural scans, we obtained 100% success rate for both structural (110/110) and EPI (185/185) inhomogeneity correction. We achieved a success rate of 99.1% (109/110) for the alignment to the unbiased template and 100% (185/185) for the cross-modal alignment of the EPI to the structural scan. Finally, in the 3 rats datasets (1 dataset only included EPI scans), 98.9% (88/89) and 100% (230/230) of scans passed inhomogeneity correction for anatomical and EPI scans respectively, and we obtained 99.6% (230/231) success rate for cross-subject alignment and 98.9% success rate during cross-modal alignment. Detailed reports for each dataset can be found in **sup. table 1**, and the visual QC report from each preprocessing run is provided in the supplementary material. We consider that this high success rate can address the significant preprocessing needs within the rodent fMRI community, despite the current cross-site variability in data quality.

RABIES integrates several standardized post-processing functionalities from confound correction to analysis

To provide a fully standardized workflow, RABIES integrates post-processing options for confound correction (**figure 1C; method section 5)** and analysis (**figure 1D; methods section 7**). The confound correction is designed to respect guidelines from the workflow in (Power et al., 2014), which includes scrubbing (or spike censoring) (Power et al., 2012), confound regression (Friston et al., 1996), frequency filtering (Satterthwaite et al., 2013) and spatial smoothing. Additionally, we adapted the ICA AROMA algorithm (Pruim et al., 2015), a well-used motion-detection algorithm in the human fMRI literature. Each confound correction step is optional, and together they cover most standard strategies in the literature. Following confound correction, analysis options can be conducted across individual scans, including seed-based

[1]antsRegistration_affine_SyN.sh;https://github.com/CoBrALab/minc-toolkit-extras/blob/master/antsRegistration_affine_SyN.sh

correlation (B. Biswal et al., 1995), whole-brain functional connectivity matrix generation (B. B. Biswal et al., 2010), and dual regression (Beckmann et al., 2009; Nickerson et al., 2017). To increase flexibility, outputs from each pipeline stage (including preprocessing) can be exported in standardized NifTi or CSV formats for use in software packages external to RABIES. This design choice is part of an effort to facilitate the reproducibility of post-processing steps, acknowledged as a significant challenge in fMRI studies (Botvinik-Nezer et al., 2020), and supports the development of analysis QC standards (see below).

PART 2: ASSESSMENT OF DATA QUALITY THROUGH SCAN DIAGNOSIS AND QUALITY CONTROL OF NETWORK ANALYSIS

A high-dimensional spatiotemporal diagnosis for data quality characterization in single scans

The reliable assessment of data quality, after preprocessing, is critical for mitigating the influence from confounds known to corrupt or bias statistical inference (Ciric et al., 2018; Power et al., 2012; Satterthwaite et al., 2013). We identified metrics that provide information on the relationship between the underlying BOLD signal and confounds that may bias this signal. These include temporal features and spatial features that RABIES combines into an automatically generated report for each scan derived after the confound correction step, which we term the *spatiotemporal diagnosis* (**table 1; figure 2**). We selected a dataset that groups mouse fMRI scans with known contribution of motion artefacts (9 awake) and other scans with very minimal contributions from motion (10 anesthetized) acquired on the same 11.7T scanner (termed REST-AWK dataset) (**methods section 9**). We leveraged these known differences in acquisition to support characterizing data quality features observed in the diagnosis across different scans. The spatiotemporal diagnosis was inspected for every scan and compared between confound correction strategies (**methods section 12**). Below we detail the interpretation of each feature included in the diagnosis, and contrast examples of corrupted and uncorrupted scans to highlight the successful distinction between desirable BOLD sources and confounds using the spatiotemporal diagnosis:

A) **Carpet plot** (Power et al., 2017)**:** the entire fMRI timeseries are displayed in a time by voxel 2D matrix. This allows to visualize global fluctuations in signal intensity, which can be a proxy for various global artefacts.
B) **The translation and rotation head motion parameters:** those are the 6 rigid body parameters estimated during preprocessing, and allow tracking of head position across scan duration.
C) **The framewise displacement and the temporal shifts in global signal from the root-mean-square of the timeseries' temporal derivative (DVARS)** (Power et al.,

2012)**:** Framewise displacement quantifies movement between consecutive frames, which reveals the timing and amplitude of spontaneous motion, whereas DVARS reveals shifts in global fMRI signal intensities. In the corrupted scan from **figure 2**, we can clearly observe the important impact of motion spikes, which corrupt every temporal features including brain networks timecourses modelled in **E)**. In contrast, the uncorrupted scan has a comparatively homogeneous carpet plot, framewise displacement and DVARS spikes are absent, and network timecourses are minimally aligned with motion indices. In line with previous literature (Power et al., 2012), we observed that confound regression alone does not correct the spiking effect in the corrupted scan (**sup. figure 5**), motivating the use of spike censoring. In **figure 2**, we do so based on both framewise displacement and DVARS (**see method section 10**) which appropriately removes spiking effects when combined with confound regression.

D) **Markers of confounded fMRI signal fluctuation:** A representative timecourse is derived within a set of anatomical masks (edge, white matter and CSF masks) using the L2-norm across voxels. Each of these anatomical regions are susceptible to motion (Pruim et al., 2015), and the white matter and CSF signal can reveal physiological confounds (Ciric et al., 2018), thus offering broader insights into potential confound sources. Furthermore, the diagnosis leverages the confound regression step, where nuisance timecourses (e.g. the 6 realignment parameters) are fitted against the fMRI signal at each voxel (**method section 7)** to obtain a modelled timecourse representing the confounded portion of the fMRI signal. To obtain an average confound timecourse across the brain, we compute the L2-norm across voxels. The proportion of variance explained by confound regression is also provided. These features allow both to visualize confound effects, and evaluate whether confound regression appropriately modelled confounds detected from other temporal features.

E) **Dual regression timecourses** (Beckmann et al., 2009; Nickerson et al., 2017)**:** An important addition to the diagnosis involves detecting the presence of known BOLD sources associated with resting-state networks using dual regression (Beckmann et al., 2009). Group-ICA provides spatial priors of the fMRI signal, and a subject-specific version of each prior is derived with two linear regression steps. The first regression provides a temporal timecourse for each prior, and then these timecourses are regressed to the data to obtain a subject-specific spatial component for each prior. In the RABIES diagnosis, we used group-ICA priors representing known rodent resting-state networks (**methods section 8**) to obtain both a temporal **E**) and spatial **J**) estimate of expected BOLD sources. In **E)**, the diagnosis reports the averaged timecourses from a selected set of BOLD network ICA priors as well as for a set of confound ICA priors.

F) Spatial distribution in signal variability ($BOLD_{SD}$): The first spatial feature of the diagnosis is the signal variability (standard deviation) at each voxel. In **figure 2**, we observe that signal variability is largely homogeneous without the major influence of confounds, but displays instead the anatomical signature of motion (i.e. anatomical edges) in the corrupted scan. This map thus offers an index of whether significant confounds are contributing to the signal.

G) Confound regression variance explained (CR_{SD}): Here, the variance explained from confound regression is quantified at each voxel by taking the standard deviation from the

modelled confound timecourse (see **D)**). This allows to contrast spatially the amplitude of confound effects. CR_{SD} is the only spatial feature revealing the expected signature of motion along anatomical edges for both uncorrupted and corrupted examples (although more prominent in the corrupted scan). This suggests that this feature can both reliably and specifically reflect the presence of confounds, better so than alternative features.

H) Confound regression variance explained proportion (Ciric et al., 2018)**:** Similar to CR_{SD}, but showing instead the proportion of variance explained (R^2).

I) Global signal covariance (Power et al., 2017)**:** The correlation with the global signal was used previously to identify globals sources of confounds (Power et al., 2017). Here we select instead the covariance to preserve the inhomogeneous distribution of voxel variance, which is relevant for the detection of confounds in **F)**. By visualizing the spatial contrast in the covariance of each voxel with the global signal timecourse, we can clearly observe the mixed nature of the global signal, which is known to capture the effect from both confounds (Power et al., 2017) and neural activity (Schölvinck et al., 2010). Indeed, in the uncorrupted scan, the global signal has a predominant contrast along the cortical gray matter, with anatomical features reminiscent of brain networks. On the other hand, in the corrupted scan, the global signal clearly reflects the spatial contrast of the motion confound, whereas applying confound correction effectively recovers a predominant cortical contrast. This feature can thus offer an index of whether confounding sources or network activity are dominant in the data.

J) Dual regression spatial maps (Beckmann et al., 2009; Nickerson et al., 2017)**:** Finally, the diagnosis shows the spatial network maps fitted using dual regression from a selected set of brain networks. These fits provide insights into the quality of network analysis, and how they may affect downstream statistical analyses. We observe in **figure 2** that only visualizing such network estimates can be misleading. Despite a clear effect from the motion spike in the temporal features (**figure 2E**), the spatial outputs preserve spatial features from the prior networks (**figure 2J**). Given the risk of overfitting with dual regression (**sup. Figure 7**), this raises concerns over the reliability of such network estimates, and suggests that we cannot reliably estimate the integrity of network analysis by visually inspecting spatial outputs in a single scan.

Outside of motion spikes, we could derive similar conclusions for two alternative sources of confound related to slow global fluctuations (**sup. figure 8**). By relying on the two spatial features of signal variability, i.e. $BOLD_{SD}$ and CR_{SD}, we find evidence that one of the confounds is associated to a ventricular origin and the other to a vascular origin, but the temporal features reveal that only the latter significantly influences network estimates from dual regression. In all of these confound cases, the temporal features reveal key mechanistic interactions between the presence of each confound and other aspects of the data, including network analysis. Subsequently, complementing the temporal information, CR_{SD} can identify the specific source of confounds by inspecting its anatomical features within the brain. Finally, the $BOLD_{SD}$ and the global signal covariance can reveal whether signal variability is dominated by confound sources or network activity, and whether confounds persist after applying a correction. These observations demonstrate the complementarity of temporal and spatial features in the

interpretation of signal sources, and together, these features allow the diagnosis of most major sources of confounds and can support their correction in rodent fMRI data.

Quality control of mouse fMRI datasets during network analysis.

To complement the subjective nature of scan-level observations, they should be followed by a quantitative assessment of confound effects on connectivity analysis at the group level to summarize data quality in a dataset (Power et al., 2012; Satterthwaite et al., 2013). However, defining interpretable and reliable quality control measures during analysis is challenging, and there is no widely adopted standard. To exemplify group-level diagnosis, we leveraged two mouse resting state dataset (REST-AWK, N = 113 scans; Multisite, N = 253 scans (Grandjean et al., 2020)). We introduce 3 main criteria which we argue are essential to establish that network analysis relates to the intended network of study, and not to spurious sources (**figure 3**). Here, we consider network analysis using dual regression and seed-based connectivity (**sup. figure 13**), which are both among the most standard and interpretable resting-state analysis techniques with a clear relationship to canonical networks (Grandjean et al., 2020).

- **Criterion 1:** *Canonical resting-state networks can be identified at the group level*.

  Brain networks previously defined across numerous independent datasets and experimental conditions represent the most robust basis we currently possess for the comparison of studies. In criterion 1, we thus ensure that the modelled network in the dataset corresponds to the expected canonical network of study. For this purpose, a representative network average is derived from the dataset and compared to a brain map of the canonical network to ensure correspondence (**figure 3A**) (**methods section 15** for the selection of the canonical network). Absence of such a network is a clear indication of dataset-level problems.

- **Criterion 2:** *Network variability across individuals is specific to the canonical network of study*.

  Reproducing the canonical network from the group average (criterion 1) is insufficient to ensure that statistical analyses involving individual variability, such as group comparisons, will be primarily driven by the network of interest. Indeed, we observed in several confounded datasets spurious signatures of individual variability despite reproducing the network with the average (**sup. figure 10**). This issue may arise in particular for model-driven analyses, such as dual regression, where overfitting can reproduce the modelled network in the data average, but not the variability maps (**sup. figure 8**). Furthermore, in other datasets, we observed a lack of individual variability altogether, suggesting a lack of underlying network activity (**figure 3B**). Such issues could arise from acquisition challenges such as low scanner signal-to-noise ratio or deep anesthesia (Nasrallah et al., 2014). In contrast, a high-quality dataset expresses clear overlap between individual variability and the canonical network, which justifies the expectation to find such anatomical alignment between variability and the modelled network.

- **Criterion 3:** *The relationship between connectivity measures and confounds should be minimal*.

  In the confound correction literature, it is recommended to measure the association between confound and connectivity measures to ensure minimal corruption of statistical inference. A large set of such metrics have been previously suggested (Ciric et al., 2018), but for this criterion, we selected a subset of three which aim to cover together most failure cases: mean framewise displacement, CR_{SD} and temporal degrees of freedom (tDOF). Mean framewise displacement is possibly the most commonly used metric. However, it is limited to motion effects and does not localize confounding effects anatomically. Conversely, we previously demonstrated that the CR_{SD} can reliably capture a wider variety of confound sources with anatomical specificity (**figure 2G, sup. figure 9**). Lastly, unequal tDOF between different scans, resulting for instance from frame censoring during confound correction, can introduce further biases into the evaluation of connectivity strength (**sup. figure 10**). Thus, to evaluate the effects of confounds, each confound metric is correlated (Pearson's r) with connectivity measures across subjects (**figure 3C**).

In summary, these 3 criteria aim to ensure that network analysis does relate to individual measures of a canonical network, and not to spurious sources. If all criteria are met, we argue that false positives or negatives during downstream analysis should be reasonably mitigated. The brain maps necessary for the evaluation of the criteria (i.e. the dataset average, individual variability and confound correlation maps) are organized to form an analysis quality report and can be automatically generated within RABIES. Each brain map was additionally spatially smoothed at 0.3 mm Full-Width at Half Maximum Gaussian to measure spatially homogeneous trends. The evaluation of those criteria is designed to allow the quality control of analysis prior to the interpretation of statistical results, and we explore its application across datasets in the next section.

The inspection of analysis quality reveals divergences across imaging sites

After defining a procedure to evaluate the quality control of network analysis, we next sought to estimate the prevalence of analysis quality issues with current acquisition paradigms for mouse fMRI. For this purpose, an analysis quality report was generated for each dataset from 17 different sites (Grandjean et al., 2020) in addition to the REST-AWK dataset, where each dataset was processed with a minimal confound correction using framewise displacement censoring and 6 motion regressors. We contrast analysis quality between two of the most discussed canonical networks in the litterature: the somatomotor network and default mode network. Analyses were repeated with seed-based connectivity in supplementary material (**sup. figure 13**).

Each analysis quality report was visually inspected, and across datasets, we observed failures related to each of the 3 criteria. To quantify performance on criterion 1 and 2, the dataset average and individual variability maps were thresholded to include the top 4% of voxels

with highest values, and their specificity was evaluated relative to the canonical network map (also thresholded) using Dice Coefficient between the thresholded areas. Criterion 3 was quantified by taking the absolute mean correlation within the thresholded area of the canonical network, for each confound metric of mean framewise displacement, CR_{SD} or temporal degrees of freedom, providing an effect size for confounds. We then selected fixed thresholds for this set of network specificity and effect size measures to classify criterion successes and failures (table in **figure 4**). The threshold values were selected to discriminate outcomes that were clearly failing from successful ones based on visual inspection. Doing so, we categorized each report into a successful, uncertain or failing outcome. A successful outcome clearly respects all 3 criteria, a failing one clearly fails on at least one criterion, and an uncertain one presents concerns, but should not be dismissed as a clear failure. In **figure 4A**, we report analysis quality outcomes, where the rate of failures (11/19 and 13/19 for the somatomotor and default mode networks respectively) dominates over the successful and uncertain outcomes. To better understand this high failure rate, we carefully inspected the spatiotemporal diagnoses of single scans in all 19 datasets, looking more specifically for evidence of network activity or residual confounds based on the $BOLD_{SD}$, global signal covariance and dual regression network maps (**figure 4B; sup. table 5**). We could find consistent high quality scans in 3 datasets, all of which were acquired under the medetomidine-isoflurane anesthetic mix and using a Cryoprobe for imaging. Conversely, in 5 datasets there was no sign of network activity in the global signal or network maps, and in 2 other datasets there were heavy confounds affecting the specificity of network fits. For the remaining 9 datasets, the outcome was unclear: either there was minimal evidence of network activity or the impact of confounds on connectivity was difficult to estimate subjectively.

We next sought to determine whether outcomes can be improved with confound correction. We repeated the analysis quality report classification across 8 variations of confound correction, each corresponding to an additional correction on top of framewise displacement censoring and regression of 6 motion parameters (**method section 13**), and counted the number of improvements (Failed -> Uncertain -> Success) or losses (Success -> Uncertain -> Failed) (**sup. table 6**). Overall, we find that adding further corrections leads to important variations in outcomes, and yet, we cannot identify any alternative strategy which leads to a clear and uniform improvement in data quality. Instead, improvements tend to be specific to the dataset and analysis, and are almost always accompanied with a similar number of losses among other datasets (except for the WM/CSF regression which led to 6 improvements for 3 losses). However, the impact of the corrections was not the same between criteria: whereas corrections tend to improve criterion 3, they also tend to worsen criteria 1 and 2 (**figure 4C**). This effect is most pronounced with global signal regression, which also yielded the worst outcomes (12 losses for one improvement). Confound regression can remove signatures of brain networks (Bright & Murphy, 2015), thus the global signal may remove the network activity previously observed in the global signal covariance maps from the spatiotemporal diagnosis (**figure 2I**). By grouping datasets with at least one success across CR strategies (**sup. table 7**), we can obtain at best success rates of 6/19 for each network analysis (**sup. figure 11**). Among those optimal outcomes, the high quality datasets have high success rates of 5/6, whereas the heavily confounded datasets had only 1/4 success and there wasn't a single success among the datasets lacking network activity.

In a final analysis, we investigate whether failures along criterion 2 could be explained by lack of statistical power, given the relatively low sample sizes of the multi-site datasets (15 scans per dataset). Using the high quality anesthetized group from REST-AWK dataset (totalizing 59 scans), we randomly subsampled the dataset at varying sample sizes and we found an important impact of sample size on network specificity measures of criterion 2 (**figure 4D**). Hence, increasing sample size would likely improve success rates on criterion 2 across datasets. We also find, however, that the correlation with framewise displacement tends to be inflated at low sample sizes, suggesting that this metric has limited reliability at low sample sizes. This effect is observed because positive and negative correlations are inflated at low sample sizes, and since the absolute value is taken for criterion 3, this leads to larger effects (the trend is not observed if the signed correlation is used **sup. figure 12**). The remaining quality metrics are minimally impacted by sample size. Exactly how to design an experiment to improve network detectability is a matter of great interest and requires properly designed power analyses which is outside of the scope of this manuscript.

In summary, we report high failure rates with regards to our suggested analysis quality criteria when conducting standard network analyses across mouse fMRI datasets. The spatiotemporal diagnosis allows us to identify striking differences in the presentation of network activity or confounds across datasets, and supports corresponding criteria failures at the dataset level. Rather than providing general improvements in outcomes, varying confound correction strategies tends to improve criterion 3, but can also include a trade-off of network detectability for criteria 1 and 2. We show that criterion 2 is dependent on statistical power, and thus, while some datasets are unlikely to be recovered at post-processing (e.g. datasets lacking network activity), others may pass analysis quality criteria by combining larger sample sizes with more stringent confound correction.

DISCUSSION

We introduce a robust image processing and analysis pipeline adapted for specific challenges to rodent fMRI. Previous preprocessing pipelines were introduced for rodent fMRI (Celestine et al., 2020; Ioanas et al., 2021; Lee et al., 2021). The strengths of the proposed approach include a thorough validation across acquisition sites and species, spanning large differences in image quality, together with the integration of state-of-the-art tools for confound correction, analysis and data quality assessment. The software was designed in line with best practices for reproducibility and transparency: open source code on Github (https://github.com/CoBrALab/RABIES) accompanied with thorough documentation (https://rabies.readthedocs.io/en/stable/), data inputs in BIDS format, containerized distribution under Docker (https://hub.docker.com/r/gabdesgreg/rabies), together with standardized and easily shared PNG outputs for quality control.

Motivated by the current variability in results from network analysis between mouse fMRI sites (Grandjean et al., 2020, 2022), we further develop new tools for data quality assessment and provide novel insights into the underlying sources of data quality failures leading to downstream issues with network analyses. These tools are building upon an extensive body of

human fMRI literature on confounds and data quality metrics. Scan-level data quality reports of similar nature as the spatiotemporal diagnosis can be found in previous work in humans (e.g. (Power et al., 2017)), but to our knowledge, such a report has not been formally integrated as part of an existing image processing software. Furthermore, there does not currently exist in the greater fMRI literature a standardized statistical framework for controlling the influence of data quality on downstream group analysis. Several metrics have been put forward for evaluating the severity of confound relationships with analysis (Ciric et al., 2018), but there is no consensus regarding which set of assumptions should constitute a standard for analysis quality. As a final effort towards standardization, we formulate a set of statistical criteria for group-level network analysis which should be met to ensure comparable analysis streams between studies. Informed by detailed observations from the spatiotemporal diagnosis, we introduce comprehensive metrics to evaluate each criterion. This analysis quality evaluation revealed further details as to the source of the discrepancies between mouse fMRI datasets, i.e., whether connectivity analysis is corrupted by confound sources or lacking statistical power for network detectability. As with the quality control of preprocessing steps, the spatiotemporal diagnosis and the analysis quality reports can be automatically generated within RABIES and easily provided along with publications. The integration of these data quality reports aims to share best-practices with the wider community of fMRI researchers, and foster future discussion regarding the necessity of data quality assessment prior to analysis in the broader literature.

The reproducibility of the proposed workflow for image processing and quality control remains limited by its reliance on subjective visual inspection and associated parameter selection for registration and confound correction. To address the substantial variability in image contrast and artefacts across rodent acquisition protocols, it remains necessary to provide flexible registration parameters which must be handled by the user. A set of recommendations are provided online to navigate the range of registration challenges (https://rabies.readthedocs.io/en/stable/registration_troubleshoot.html). Limitations surrounding the visual assessment of registration quality (Craddock et al., 2013; Esteban et al., 2019) and the selection of confound correction strategies are not unique to RABIES, but are shared with human work. Otherwise, for the analysis quality control stage, we aimed to classify the quality outcomes using a set of fixed thresholds, but these thresholds were not able to perfectly distinguish failing and successful outcomes, and a third category of uncertain outcome was needed. These thresholds won't apply to any future study, and instead, we recommend relying on visual inspection of the analysis reports, complemented with the spatiotemporal diagnosis, to accurately estimate data quality.

We provide the community with a reliable tool for conducting image processing and analysis, thereby allowing harmonization of computational practices across laboratories and fostering reproducible research. This advancement is timely in light of current inconsistencies in methodological and reporting practices for rodent fMRI. Previous initiatives, such as fMRIPrep, have already been transformative in those regards for human work. The RABIES analysis quality control framework will be crucial for the development of high quality data acquisition protocols (Grandjean et al., 2022). Providing these reports, automatically generated from RABIES, along with publications will improve transparency of results and prevent false positives/negatives. Equipped with RABIES, rodent fMRI is poised to make the greatest advances towards addressing pressing issues in translational neuroscience.

ACKNOWLEDGMENTS

We thank the fMRIPrep team for their leadership in providing standards and guidelines in the development of open source neuroimaging softwares. We also thank the ICA-AROMA developper for making their code open access, which allowed us to integrate the algorithm with minor adaptations. We are grateful to all members of the mouse multi-center initiative from (Grandjean et al., 2020) as well as other members of the rodent fMRI community who publicly shared data. Without these recent initiatives, this work would not have been possible. We also thank Daniel Gallino for suggesting the software acronym, and the RABIES users who provided feedback to correct issues and enhance the distribution of the software. Finally, we acknowledge our funding sources. The Fonds de recherche du Québec – Nature et technologies (FRQNT) scholarship for providing salary support to G.D.-G., the Dutch Research Council grant OCENW.KLEIN.334 supports the work of JG and MMC receives salary support from the les Fonds de recherches du Québec – Santé (FRQS). Furthermore, MMCs research is supported by the McGill University's Healthy Brains for Healthy Lives (a Canada First Research Excellence Fund program), Canadian Institutes of Health Research, and the National Sciences and Engineering Research Council of Canada.

FIGURES

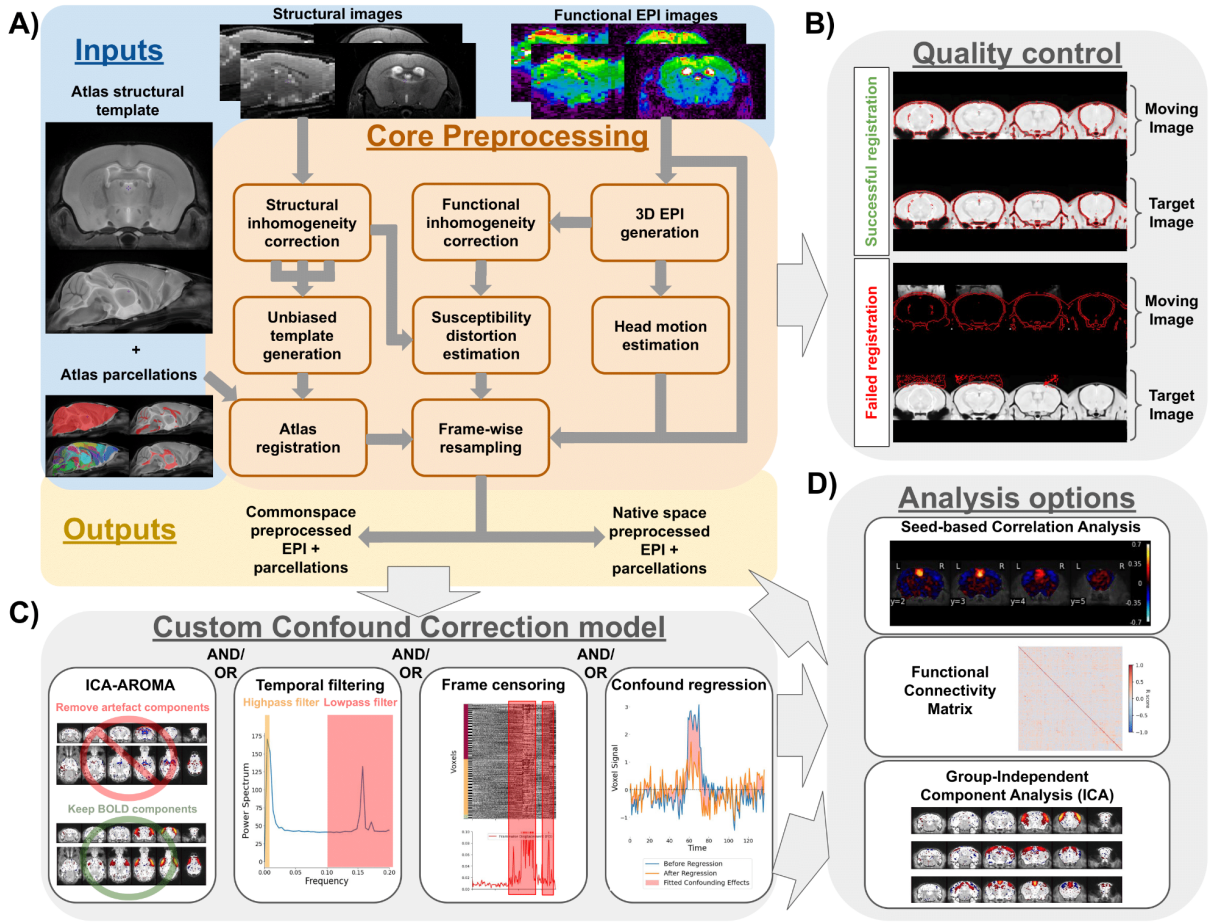


Figure 1: RABIES allows the preprocessing, quality control, confound correction and analysis of rodent fMRI datasets. **A)** Schematic organization of the preprocessing pipeline. RABIES takes as inputs a dataset of EPI functional scans along with their structural scan (optional) in BIDS format, together with an external reference structural atlas containing anatomical masks and labels. The core preprocessing architecture conducts commonspace alignment, head motion realignment and susceptibility distortion correction (detailed in **method section 1**). **B)** Along the execution of the preprocessing pipeline, RABIES generates automatically PNG images allowing the visual quality control of each failure-prone preprocessing step (**method section 4**). **C)** Following preprocessing, an array of confound correction strategies are made optional and can be customized according to user needs (**method section 5**). **D)** After applying confound correction, a final workflow is made available to conduct basic resting state analysis (**method section 7**).

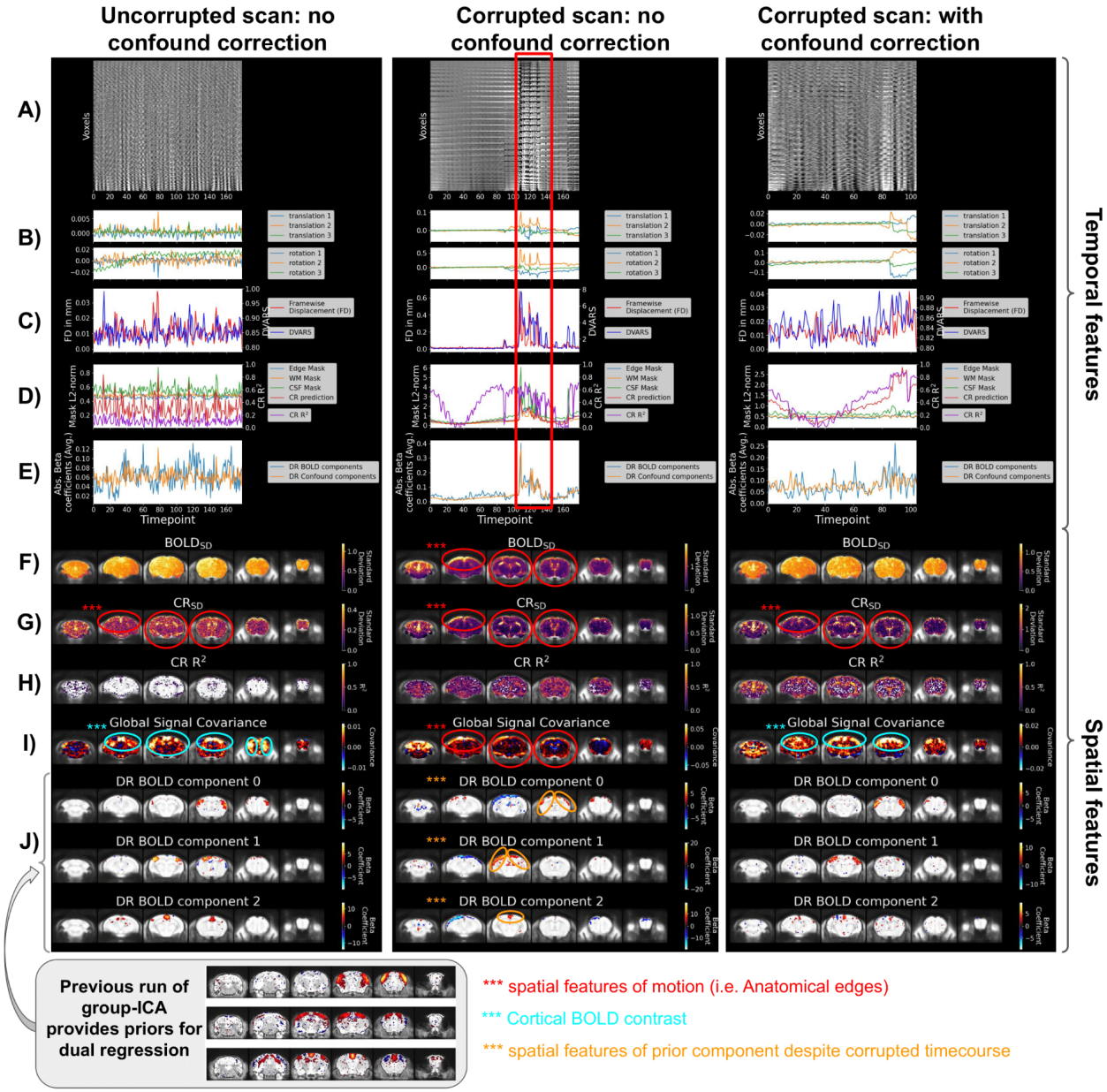


Figure 2: The spatiotemporal diagnosis generated by RABIES for 3 exemplified cases: an uncorrupted scan with no confound correction applied (first column), a motion-corrupted scan without confound correction (second column) and the same corrupted scan after applying confound correction (third column). The confound correction for the third example consisted of spike censoring using FD and DVARS combined, followed by confound regression (CR) of the 6 motion realignment parameters. For each diagnosis, the top half represents the set of temporal features, and the bottom half the set of spatial features. Each spatial map is represented along 6 cortical slices, overlapped onto the anatomical template in common space. **A)** The carpet plot, representing all voxel timeseries in a 2D voxel by time matrix. **B)** The 6 motion realignment parameters estimated during preprocessing. **C)** The framewise displacement (FD) and DVARS timecourses. **D)** The timecourses of signal fluctuations within anatomical masks (edge, WM or CSF) or fluctuations modelled across all voxels using CR. If CR was not applied during confound correction, the linear regression is still computed using the 6 realignment parameters

to always provide every feature of the RABIES diagnosis. **E)** The absolute values of the temporal linear coefficient from dual regression (DR) are averaged from either across BOLD components or across confound components. **F)** The standard deviation at each voxel ($BOLD_{SD}$). **G)** The standard deviation estimated on the predicted confound timecourse from CR (CR_{SD}). **H)** The proportion of variance explained from confound regression. **I)** The covariance of each voxel with the global signal timecourse. **J)** The spatial linear coefficients measured from DR for a given BOLD component. Each DR map is thresholded to include the top 4% of the voxels with the highest values.

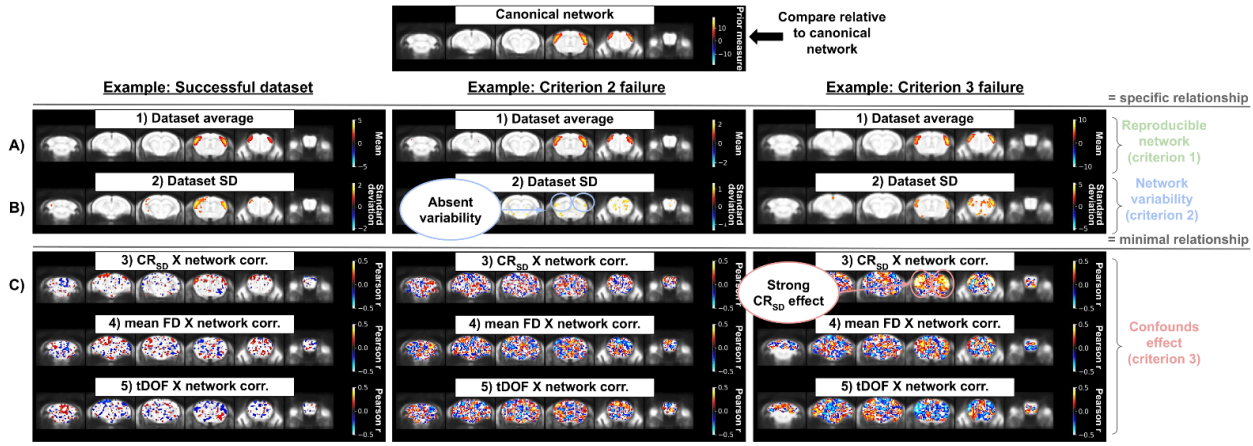
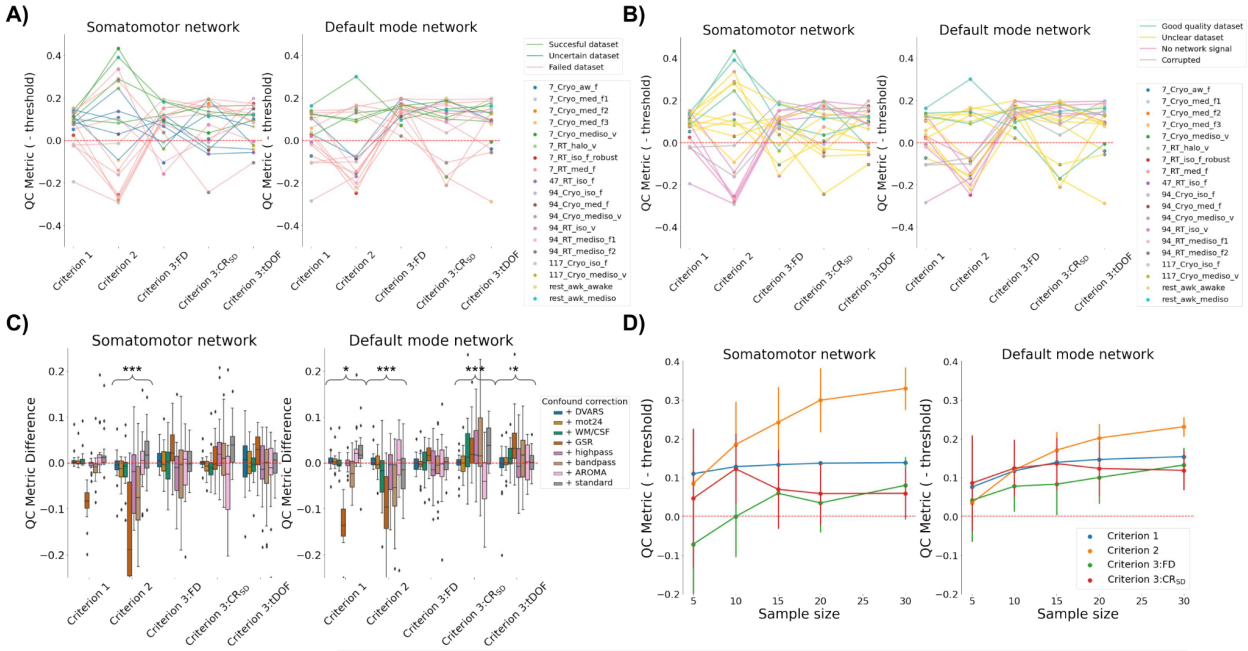


Figure 3: The 5 brain maps forming the analysis quality report, and their affiliated 3 criteria. Examples are shown for three different quality outcomes from the analysis of the somatomotor canonical network using dual regression. A region of interest for the network is defined by thresholding the map to include 4% of the voxels with the highest values. Criteria 1 and 2 are evaluated respectively from **A)** the dataset mean and **B)** the dataset standard deviation (SD) in connectivity values across subjects for each voxel (here connectivity values are linear coefficients from dual regression). Both maps are thresholded similarly to the canonical network to evaluate specificity using Dice coefficient. **C)** The relationship to confounds, i.e. criterion 3, is evaluated using three different metrics of confounds: mean framewise displacement (FD), CR_{SD}, and temporal degrees of freedom (tDOF). Each metric is correlated (Pearson's r) with connectivity values across subjects, for each voxel, forming a correlation brain map (thresholded at r>0.1). Each brain map is spatially smoothed at 0.3mm Full-Width at Half Maximum.



QC metric	Thresholds			
	DR:S1N	SBC:S1N	DR:DMN	SBC:DMN
Criterion 1 - Dice overlap for the dataset mean	**Failure:** Dice<0.75	**Failure:** Dice<0.65	**Failure:** Dice<0.6	**Failure:** Dice<0.55
Criterion 2 - Dice overlap for dataset STD	**Uncertain:** Dice<0.3 **Failure:** Dice<0.2			
Criterion 3:FD - correlation for mean FD	**Uncertain:** Absolute mean correlation>0.20			
Criterion 3:CR$_{SD}$ - correlation for CR$_{SD}$	**Failure:** Absolute mean correlation>0.30			
Criterion 3:tDOF - correlation for tDOF				

Figure 4: Evaluation of analysis quality across 19 mouse fMRI datasets using dual regression on the somatomotor and default mode networks. 5 quality control (QC) metrics are derived from the analysis quality report (one metric per brain map) to quantify performance. The metrics and their associated thresholds for determining the category of quality outcome are listed in the table. **A)** Analysis quality outcomes across datasets/analyses. Each QC metric is divided by its associated threshold to align their respective threshold for successful outcome at 0 (dotted red line). The analysis quality report for a given dataset is linked by a line, and the line is color-coded according to the classified outcome (successful in green, uncertain in blue and failed in red). **B)** Same plot as in A), but datasets are instead categorized based on observation of the spatiotemporal diagnosis across scans (described in **sup. table 5**). **C)** The influence of additional confound correction on QC metrics. The barplots show the difference between the initial result in A) and the results after additional confound corrections (listed in the legend), where each bar represents the trend across all 19 datasets. A positive difference (values above 0) represents an improvement in quality, and vice-versa for negative values. The trends towards improved or worsened quality are evaluated (separately for each network) across QC metrics using a linear mixed effect model, modeling the QC metric difference as a function of criterion (*p<0.05,**p<0.01 ,***p<0.001). For those statistics, all confound correction strategies are pooled together to reflect general trends. All significant trends are negative for criteria 1 and 2, and positive for criterion 3. **D)** The effect of sample size on QC metrics was evaluated in the REST-AWK anesthetized group (tDOF was not considered, since there was negligible variation in this dataset). The dataset was randomly subsampled at varying sample sizes (5, 10, 15, 20 and 30) for 50 iterations at each sample size, and the analysis quality report was computed on

each subsample. Each point corresponds to the average across the 50 iterations and its associated error bar (standard deviation).

METHODS

1) RABIES core preprocessing

The preprocessing of functional magnetic resonance imaging (fMRI) scans prior to analysis consists of, at minimum, the anatomical alignment of scans to a common space, head realignment to correct for motion, and the correction of susceptibility distortions arising from the echo-planar imaging (EPI) acquisition of functional scans. The core preprocessing pipeline in Rodent Automated Bold Improvement of EPI Sequences (RABIES) carries each of these steps with state-of-the-art processing tools and techniques.

To conduct common space alignment, structural images, which were acquired along the EPI scans, are initially corrected for inhomogeneities and then registered together to allow the alignment of different MRI acquisitions. This registration is conducted by generating an unbiased data-driven template through the iterative linear and non-linear registration of each image to the dataset consensus average (using ANTs' registration algorithm (Avants et al., 2008) with Mattes Mutual Information and cross-correlation objective functions for linear and non-linear stages respectively), where the average gets updated at each iteration to provide an increasingly representative dataset template (https://github.com/CoBrALab/optimized_antsMultivariateTemplateConstruction; (Avants et al., 2011)). The finalized template after the last iteration provides a representative alignment of each MRI session to a template that shares the acquisition properties of the dataset (e.g. brain shape, field of view, anatomical contrast, ...), making it a stable registration target for cross-subject alignment. This newly-generated unbiased template is then itself registered to an external reference atlas to provide both an anatomical segmentation and a common space comparable across studies defined from the provided reference atlas.

The remaining preprocessing involves the EPI image. A volumetric EPI image is first derived using a trimmed mean across the EPI frames, after an initial motion realignment step. Using this volumetric EPI as a target, the head motion parameters are estimated by realigning each EPI frame to the target using a rigid registration. To correct for EPI susceptibility distortions, the volumetric EPI is first subjected to an inhomogeneity correction step, and then registered non-linearly to the anatomical scan from the same MRI session, which allows to calculate the required geometrical transforms for recovering brain anatomy (Wang et al., 2017). Finally, after calculating the transformations required to correct for head motion and susceptibility distortions, both transforms are concatenated into a single resampling operation (avoiding multiple resampling) which is applied at each EPI frame, generating the preprocessed EPI timeseries in native space (Esteban et al., 2019). Preprocessed timeseries in common space are also generated by further concatenating the transforms allowing resampling to the reference atlas.

The preprocessing workflow of RABIES is illustrated in **figure 1A**, and each associated module is described in detail in the RABIES documentation (https://rabies.readthedocs.io/en/stable/). When structural images are not provided by the dataset, the volumetric EPI replaces the structural image during the generation of the unbiased template.

2) Datasets for preprocessing benchmarking

The robustness of RABIES preprocessing was evaluated by sampling a range of datasets from different sites to attempt capturing representative data samples across the rodent fMRI community. We accessed 20 open mouse fMRI datasets, 17 of which did not include structural scans, allowing for testing the workflow without structural scans. The effectiveness of the pipeline was also tested on 3 independent rat fMRI datasets. All datasets are openly accessible (see **sup. table 1** for links), and acquisition details for each dataset can be found in their respective original publication. For mouse preprocessing using structural scans, we used the DSURQE ex-vivo MRI atlas (Dorr et al., 2008; Richards et al., 2011; Steadman et al., 2014; Ullmann et al., 2013), and the Fischer 344 in vivo atlas (Goerzen et al., 2020) was used for rat preprocessing. We generated a mouse EPI atlas for the preprocessing of mouse datasets lacking structural images (see **method section 3** below).

3) Generation of a EPI atlas

We selected 5 mouse datasets (i.e. 117_Cryo_mediso_v, 7_Cryo_med_f1, 7_Cryo_med_f2, 7_Cryo_aw_f, 7_RT_med_f in **sup. table 1**) from the multi-site study by Grandjean and colleagues (Grandjean et al., 2020). We selected those datasets for their complete brain coverage, minimal susceptibility distortions and variable anatomical contrast. Using the RABIES workflow, each EPI scan was first corrected for inhomogeneity and an unbiased template was generated combining all 5 datasets. The unbiased template was then registered to the DSURQE mouse atlas using a non-linear registration (Avants et al., 2008) to propagate brain masks and anatomical labels onto the newly-generated EPI template. This EPI atlas is openly available online (https://doi.org/10.5281/zenodo.5118030), and used by default by RABIES as reference atlas when executing the EPI-only preprocessing workflow.

4) Preprocessing quality control

During preprocessing, a set of PNG images are generated to conduct visual QC of each failure-prone preprocessing step (**sup. figure 1**). To visualize the inhomogeneity correction of either the structural or the EPI image, the input MRI image is shown before and after the correction together with the overlap of the brain mask used for the final correction, thus allowing to visualize the quality of inhomogeneity correction. Failures can be noticed either if important inhomogeneities remain or if brain masking fails. Then for the QC of each registration step, the report shows the overlap between the moving image and the target image. These registration steps include alignment to the generated unbiased template, the alignment between the unbiased template and the reference common space atlas, and the registration between anatomical and functional images for susceptibility distortion correction.

The entire QC report from each dataset was visually inspected to record the number of failures at each registration step. Registration failures were considered when there was a clear misalignment of at least one brain structure. Minimal misalignments that subsisted because of EPI distortions and poor anatomical contrast were tolerated, given that very few EPIs have sufficient contrast to achieve specific anatomical alignment and complete correction for distortions. In **sup. table 1**, additional notes on preprocessing quality are provided to detail

dataset-specific limitations (e.g. partial misalignment to the reference atlas in highly distorted regions). All QC reports inspected are provided along this paper in the supplementary material.

5) RABIES confound correction workflow

The workflow for confound correction is structured following best practices found in human literature, largely based on recommendations from Power and colleagues (Power et al., 2014). Here, we detail the execution of the confound correction workflow, which requires specific considerations regarding the order and combination of multiple correction steps. These considerations aim to avoid inefficient corrections, or the re-introduction of confounds into the data at later steps (Lindquist et al., 2019). Several options are provided, and if selected, are applied in the following sequence:

(1) First, spike censoring temporal masks are derived from framewise displacement and/or DVARS thresholds (**method section 11** for details). The temporal masking is applied on both blood oxygen level-dependent (BOLD) timeseries and nuisance regressors before any other correction step to exclude motion spikes which may bias downstream corrections, in particular, detrending, frequency filtering and confound regression (Power et al., 2014).

(2) If spike censoring was applied in 1), all scans in the dataset can be set to retain the same final number of frames to account for downstream analysis biases due to unequal tDOF. To do so, a pre-set final number of frames is defined, and the number of extra frames is calculated after the application of spike censoring and the removal of edge artefacts in 5). A corresponding set of extra frames is then randomly selected and removed.

(3) Linear detrending is applied to both BOLD timeseries and nuisance regressors.

(4) If desired, motion sources can be cleaned using a rodent-adapted ICA-AROMA classifier (Pruim et al., 2015) (**method section 6**). ICA-AROMA is applied prior to frequency filtering to remove further effects of motion than can result in ringing after filtering (Carp, 2013; Pruim et al., 2015).

(5) Next, frequency filtering requires particular considerations when applied after spike censoring, which results in missing timepoints. Conventional filters cannot account for missing data, and to address this issue, Power and colleagues introduced a method for simulating missing timepoints (Power et al., 2014). This method relies on an adaptation of the Lomb-Scargle periodogram, which allows estimating the frequency composition of timeseries with missing data points, and from that estimation, missing timepoints can be simulated while preserving the frequency profile (Mathias et al., 2004). Missing data is thereby simulated with preserved frequency properties for both BOLD timeseries and nuisance regressors prior to frequency filtering.

(6) Following the simulation, frequency filtering (highpass, lowpass or bandpass) is applied using a 3rd-order Butterworth filter, and timepoints can be removed at each edge of the timeseries to account for edge artefacts following filtering (Power et al., 2014). Edge artefacts were visualized in simulated data (**sup. figure 4**), and from those observations we conclude that removing 30 seconds at each edge and using a 3rd order filter are optimal. The nuisance regressors are also filtered to ensure orthogonality between the frequency filters and subsequent confound regression, which can otherwise re-introduce removed confounds

(Lindquist et al., 2019). After frequency filtering, the temporal masks from spike censoring are re-applied to remove the simulated timepoints.

(7) After frequency filtering, remaining confound effects are modelled and removed using ordinary least squares (OLS) linear regression of the nuisance regressors onto the BOLD timeseries (i.e. confound regression) (more details in **methods section 8**).

(8) After previous confound correction steps, timeseries can be scaled according to one of three options: 1) based on background noise variance, 2) the global standard deviation calculated by including all voxels, or 3) by applying voxelwise standardization (each voxel is separately scaled to unit variance). We selected the global standard deviation over the voxelwise standardization to avoid 'over-scaling' voxels expressing confounds, as observed in the $BOLD_{SD}$ (**figure 2F**). We explore the alternative scaling with voxelwise standardization when varying confound correction (**method section 13**).

(9) Timeseries are spatially smoothed using a Gaussian smoothing filter.

Importantly, each confound correction step (with the exception of linear detrending) is optional when using RABIES. Currently, there are no generalizable confound correction workflow, and rather, the optimal confound correction strategy may well be largely study-specific. The issue of confound correction for resting-state fMRI remains largely unresolved among human literature, and is only beginning to be studied in rodents. Thus, a flexible workflow remains crucial, despite concerns regarding reproducibility.

6) Rodent-adapted ICA-AROMA

The original code for the algorithm (https://github.com/maartenmennes/ICA-AROMA) was adapted to function without the hard-coded human priors for anatomical masking and parameter thresholds for component classification. Following an initial independent component analysis (ICA) decomposition of the data using FSL's MELODIC algorithm (Beckmann & Smith, 2004), four features are extracted from each ICA component spatial map for classification: 1) the temporal correlation between the component and motion parameters, 2) the fraction of the frequency spectrum within the high frequency range, and the fraction of the component within 3) the cerebrospinal fluid (CSF) mask and 4) the brain edge mask. The component is classified as motion if the CSF or high frequency content fractions are above a given threshold, or if classified by a pre-trained linear classifier combining the brain edge fraction and motion correlation. To adapt the original algorithm with RABIES, the CSF mask is inherited from the rodent reference atlas and the edge mask is automatically generated from the brain mask, the threshold for high frequency content was increased as rodent can express higher BOLD frequencies, particularly under medetomidine (Grandjean et al., 2014), and the linear classifier was retrained. To select the new parameters, we manually classified motion and network components, derived from a set of scans from the REST-AWK group anesthetized under a medetomidine-isoflurane mixture, and selected parameters to successfully classify clear motion components while avoiding false classification of brain networks.

7) RABIES analysis functionalities

Here we detail how each connectivity analysis available within RABIES can be conducted following the completion of the confound correction workflow.

- **Seed-based connectivity** (B. Biswal et al., 1995): An anatomical seed of interest is selected, and the mean timecourse within the seed is correlated (Pearson's r) with each voxel timeseries to measure the brain connectivity map for that seed. In RABIES, this is repeated for each scan independently, to obtain a subject-level connectivity map that can then be used for hypothesis testing.
- **whole-brain connectivity matrix** (B. B. Biswal et al., 2010): the whole brain parcellation during RABIES preprocessing is used to extract a mean timecourse at each parcel. Then, all parcel timecourses are cross-correlated (Pearson's r) to obtain a whole-brain matrix describing the connectivity strength between every region pair. As with SBC, in RABIES, a connectivity matrix is computed for each scan.
- **Group independent component analysis (ICA)** (Beckmann & Smith, 2004): the timeseries from all scans are concatenated in common space, and ICA decomposition is conducted on those concatenated timeseries to identify spatial sources of covariance (components) prevalent in the entire group. Among those components, resting-state networks as well as sources of confounds can be identified (example in **sup. figure 5**). RABIES uses FSL's MELODIC ICA algorithm (Beckmann & Smith, 2004).
- **Dual regression (DR)** (Beckmann et al., 2009; Nickerson et al., 2017): DR is a technique building on the group ICA algorithm which allows the modelling of previously detected group-level components back onto single scans. The DR algorithm consists of two consecutive linear regression steps. First, components from group ICA are regressed against an individual scan timeseries to obtain an associated timecourse for each component. Second, a second regression using these component timecourses are regressed again against the scan's timeseries, allowing to finally derive scan-specific spatial maps for each corresponding ICA component. Those linear coefficients can allow for describing individual-specific features of resting-state networks, such as amplitude or spatial location.

8) Confound timecourses predicted from confound regression

In RABIES, the linear regression for confound regression is solved through OLS regression:

$$OLS(X,Y) = (X^T X)^{-1} X^T Y = B,$$

where X is the nuisance regressors matrix (time by nuisance regressor), Y is the scan timeseries matrix (time by voxel), and B is the estimated linear coefficient matrix (nuisance regressor by voxel).

The regression of confounds is then conducted by subtracting the fitted confound effect from the original scan timeseries as follows:

$$Y' = Y - XB,$$

where Y' is the cleaned timecourse matrix (time by voxel). Here, the fitted confound effect XB corresponds to the predicted confound timeseries matrix (time by voxel) used in the RABIES spatiotemporal diagnosis.

9) Group-ICA priors for dual regression

The group-ICA components used as priors for dual regression (available online https://zenodo.org/record/5118030/files/melodic_IC.nii.gz) were obtained using FSL's MELODIC algorithm (Beckmann & Smith, 2004). MELODIC was run with 30 components on the combined data from the anesthetized and awake mouse groups, where confound correction consisted of highpass filtering at 0.01 Hz, FD and DVARS censoring (**methods section 11**), and confound regression was applied with the 6 rigid motion parameters together with white matter (WM)/CSF mask signals. The group-ICA maps were visually inspected, and inspired by suggestions from Zerbi and colleagues (Zerbi et al., 2015), each component was classified into a rodent brain network, confound source or other (**sup. figure 5**). More specifically, confound components display signatures of motion (i.e. edge effects), bilateral asymmetry, high loadings into WM and CSF tissues, wide-spread effects across the brain or affect only single slices. Components are classified as brain networks only if they are anatomically-localized with bilateral symmetry, with no evidence of confound signature, and are reproducing features of previously reported ICA brain networks in (Zerbi et al., 2015).

10) Dataset with anesthetized and awake mice (REST-AWK)

To explore signatures of confounds in mouse fMRI, we relied on a dataset with a group of anesthetized mice under a medetomidine-isoflurane mixture equipped with mechanical ventilation (i.e. controlling for motion) (N=10) and a group of physically-restrained, freely-breathing, awake mice (N=9) (https://openneuro.org/datasets/ds001653/versions/1.0.2). The group of awake mice was conditioned to the restraint system for 1 week prior to scanning. fMRI (gradient echo EPI, 0.18x0.15 mm and 90x60 matrix in-plane, 28 0.4 mm slices, TR=1.2 s, 180 volumes) were acquired on a 11.7T Bruker animal scanner. For each subject, data was acquired over 2 sessions where 3 consecutive fMRI scans were acquired per session, giving a total of 60 and 53 functional scans for the anesthetized and awake group respectively (one awake scan was lost during preprocessing QC). Despite the availability of structural scans (T2 RARE, 0.085x0.09 mm and 200x100 matrix in-plane, 28 0.4 mm slices), for all analysis results shown, the dataset was processed only using EPI scans for comparability with the multicenter dataset in later analysis (see below).

11) Method for spike censoring using framewise displacement and/or DVARS

Introduced in Power and colleagues (Power et al., 2012), spike censoring (or scrubbing) consists of removing frames that exceed a certain threshold in framewise displacement. As in their original publication, spike censoring using framewise displacement with RABIES will remove all frames exceeding a specified threshold in mm, together with one backward and two forward frames (Power et al., 2012).

Further inspection of the spatiotemporal diagnosis in the awake condition revealed that spike censoring based on a framewise displacement alone was insufficient to remove certain forms of global signal spikes, as measured using DVARS (Power et al., 2012), affecting the dual regression estimates (**sup. figure 7**). This motivated the implementation of a secondary spike

censoring strategy based on DVARS instead, where frames expressing outlier values in DVARS are removed. This is conducted by iteratively removing frames which present outlier DVARS values above or below 2.5 standard deviations until no more outliers are detected.

12) Temporal degrees of freedom (tDOF) estimation

Following confound correction, the remaining number of tDOF is calculated for each scan to account for potential effects of tDOF during analysis QC (see below). To calculate tDOF, the number of timepoints remaining after spike censoring is subtracted with the number of components removed running ICA-AROMA, and by the number of regressors applied during confound correction. When the number of frames was matched across scans (**methods section 5**, step 2), the tDOF estimation is conducted as if the frames were not matched, to still allow for evaluation of a residual bias from tDOF removal in the analysis quality control.

13) Confound correction strategies investigated

The REST-AWK dataset and the 17 datasets from the multi-site study (Grandjean et al., 2020) were processed over a range of confound correction strategies to derive scan-level spatiotemporal diagnosis and to conduct network analysis quality control. In all conditions, we applied linear detrending, scaling based on global standard deviation and spatial smoothing using a 0.3 mm Full-Width at Half Maximum Gaussian filter. After censoring timepoints, scans which had less than ⅔ of the original number of frames were removed from further analysis.

To evaluate the features derived using the spatiotemporal diagnosis and to investigate the correction of confounds, the REST-AWK dataset was processed with 1) no confound correction, 2) spike censoring using framewise displacement, 3) regression of 6 motion parameters, 4) spike censoring + the motion regression, and 5) spike censoring using both framewise displacement and DVARS + the motion regression.

For the subsequent evaluation of network analysis quality control across all 17 multi-site datasets and the two groups from REST-AWK, datasets were all corrected at minimum using spike censoring with framewise displacement together with regression of 6 motion parameters. Every dataset which presented a significant bias from tDOF in the analysis quality report (mean correlation>0.15), either with the somatomotor or default mode network, was also processed to include matching numbers of frames across scans (**methods section 5**, step 2; example in **sup. figure 10**). Additional corrections were tested to attempt improving the quality of network analysis, and consisted of 1) spike censoring using DVARS (in addition to framewise displacement), 2) regression of 24 motion parameters (6 parameters + their derivative + all 12 terms squared (Friston et al., 1996)), 3) regression of WM and CSF mean signals, 4) regression of the global signal, 5) highpass filtering at 0.01Hz, 6) bandpass filtering at 0.01-0.2Hz, 7) applying ICA-AROMA and 8) voxelwise standardization of voxel timeseries. The thresholds for framewise displacement censoring were specific to each dataset, as detailed in **method section 14**. When frequency filtering was applied, 30 seconds were removed at each end of the timeseries to avoid edge artefacts.

Both the scan spatiotemporal diagnoses and the analysis quality report for all confound correction strategies and datasets are provided in supplementary material for consultation.

14) Selection of dataset-specific thresholds for framewise displacement censoring

We observed that framewise displacement values did not provide a consistent baseline between different datasets, potentially as a consequence of varying image contrast which could affect the head motion realignment algorithm. Thus, to ensure that only motion spikes were removed, the framewise displacement timecourse was visually inspected across scans for each dataset, and a threshold was estimated based on values associated with isolated spikes in the timecourse. The thresholds are listed in **sup. table 3**.

15) Definition of canonical network brain maps for analysis quality control

To conduct the proposed analysis quality control framework, we needed to define the canonical networks. Two rodent canonical networks were investigated: the somatomotor network and the default mode network (Grandjean et al., 2020; Liska et al., 2015; Zerbi et al., 2015). For seed-based connectivity analysis, the somatomotor and default mode networks were mapped using a seed in the right primary somatosensory cortex and anterior cingulate cortex respectively. For dual regression, we selected the group-ICA components corresponding to each network.

To conduct analysis quality control on seed-based connectivity analysis (**sup. figure 13**), we needed to compute a representative average brain map for each network from the set of scans across datasets. To do so, we selected each dataset reaching a successful analysis quality outcome in at least one of the alternative confound correction strategies during dual regression analysis. The selection of datasets was repeated for each network.

REFERENCES

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41.

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044.

Beckmann, C. F., Mackay, C. E., Filippini, N., & Smith, S. M. (2009). Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *NeuroImage*, *47*(Suppl 1), S148.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, *23*(2), 137–152.

Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., Dogonowski, A.-M., Ernst, M., Fair, D., Hampson, M., Hoptman, M. J., Hyde, J. S., Kiviniemi, V. J., Kötter, R., Li, S.-J., … Milham, M. P. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(10), 4734–4739.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *34*(4), 537–541.

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88.

Bright, M. G., & Murphy, K. (2015). Is fMRI “noise” really noise? Resting state nuisance regressors remove variance with network structure. *NeuroImage*, *114*, 158–169.

Carp, J. (2013). Optimizing the order of operations for movement scrubbing: Comment on Power et al

[Review of *Optimizing the order of operations for movement scrubbing: Comment on Power et al*]. *NeuroImage*, *76*, 436–438.

Celestine, M., Nadkarni, N. A., Garin, C. M., Bougacha, S., & Dhenain, M. (2020). Sammba-MRI: A Library for Processing SmAll-MaMmal BrAin MRI Data in Python. *Frontiers in Neuroinformatics*, *14*, 24.

Chuang, K.-H., & Nasrallah, F. A. (2017). Functional networks and network perturbations in rodents. *NeuroImage*, *163*, 419–436.

Ciric, R., Rosen, A. F. G., Erus, G., Cieslak, M., Adebimpe, A., Cook, P. A., Bassett, D. S., Davatzikos, C., Wolf, D. H., & Satterthwaite, T. D. (2018). Mitigating head motion artifact in functional connectivity MRI. *Nature Protocols*, *13*(12), 2801–2826.

Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S. S., Yan, C., Li, Q., Lurie, D., Vogelstein, J., Burns, R., & Others. (2013). Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). *Frontiers in Neuroinformatics*, *42*, 10–3389.

Diao, Y., Yin, T., Gruetter, R., & Jelescu, I. O. (2021). PIRACY: An Optimized Pipeline for Functional Connectivity Analysis in the Rat Brain. *Frontiers in Neuroscience*, *15*, 602170.

Dorr, A. E., Lerch, J. P., Spring, S., Kabani, N., & Henkelman, R. M. (2008). High resolution three-dimensional brain atlas using an average magnetic resonance image of 40 adult C57Bl/6J mice. *NeuroImage*, *42*(1), 60–69.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(28), 7900–7905.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-Related effects in fMRI time-series. In *Magnetic Resonance in Medicine* (Vol. 35, Issue 3, pp. 346–355). https://doi.org/10.1002/mrm.1910350312

Goerzen, D., Fowler, C., Devenyi, G. A., Germann, J., Madularu, D., Chakravarty, M. M., & Near, J.

(2020). An MRI-Derived Neuroanatomical Atlas of the Fischer 344 Rat Brain. *Scientific Reports*, *10*(1), 6952.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*, 13.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*, 160044.

Grandjean, J., Canella, C., Anckaerts, C., Ayrancı, G., Bougacha, S., Bienert, T., Buehlmann, D., Coletta, L., Gallino, D., Gass, N., Garin, C. M., Nadkarni, N. A., Hübner, N. S., Karatas, M., Komaki, Y., Kreitz, S., Mandino, F., Mechling, A. E., Sato, C., … Gozzi, A. (2020). Common functional networks in the mouse brain revealed by multi-centre resting-state fMRI analysis. *NeuroImage*, *205*, 116278.

Grandjean, J., Corcoba, A., Kahn, M. C., Upton, A. L., Deneris, E. S., Seifritz, E., Helmchen, F., Mann, E. O., Rudin, M., & Saab, B. J. (2019). A brain-wide functional map of the serotonergic responses to acute stress and fluoxetine. *Nature Communications*, *10*(1), 350.

Grandjean, J., Desrosiers-Gregoire, G., Anckaerts, C., Angeles-Valdez, D., Ayad, F., Barrière, D. A., Blockx, I., Bortel, A. B., Broadwater, M., Cardoso, B. M., Célestine, M., Chavez-Negrete, J. E., Choi, S., Christiaen, E., Clavijo, P., Colon-Perez, L., Cramer, S., Daniele, T., Dempsey, E., … Hess, A. (2022). StandardRat: A multi-center consensus protocol to enhance functional connectivity specificity in the rat brain. In *bioRxiv* (p. 2022.04.27.489658). https://doi.org/10.1101/2022.04.27.489658

Grandjean, J., Schroeter, A., Batata, I., & Rudin, M. (2014). Optimization of anesthesia protocol for resting-state fMRI in mice based on differential effects of anesthetics on functional connectivity patterns. *NeuroImage*, *102 Pt 2*, 838–847.

Ioanas, H.-I., Marks, M., Zerbi, V., Yanik, M. F., & Rudin, M. (2021). An optimized registration workflow and standard geometric space for small animal brain imaging. *NeuroImage*, *241*, 118386.

Lake, E. M. R., Ge, X., Shen, X., Herman, P., Hyder, F., Cardin, J. A., Higley, M. J., Scheinost, D.,

Papademetris, X., Crair, M. C., & Constable, R. T. (2020). Simultaneous cortex-wide fluorescence Ca2+ imaging and whole-brain fMRI. *Nature Methods*, *17*(12), 1262–1271.

Lee, S.-H., Broadwater, M. A., Ban, W., Wang, T.-W. W., Kim, H.-J., Dumas, J. S., Vetreno, R. P., Herman, M. A., Morrow, A. L., Besheer, J., Kash, T. L., Boettiger, C. A., Robinson, D. L., Crews, F. T., & Shih, Y.-Y. I. (2021). An isotropic EPI database and analytical pipelines for rat brain resting-state fMRI. *NeuroImage*, *243*, 118541.

Lindquist, M. A., Geuter, S., Wager, T. D., & Caffo, B. S. (2019). Modular preprocessing pipelines can reintroduce artifacts into fMRI data. *Human Brain Mapping*, *40*(8), 2358–2376.

Liska, A., Galbusera, A., Schwarz, A. J., & Gozzi, A. (2015). Functional connectivity hubs of the mouse brain. *NeuroImage*, *115*, 281–291.

Mandino, F., Cerri, D. H., Garin, C. M., Straathof, M., van Tilborg, G. A. F., Chakravarty, M. M., Dhenain, M., Dijkhuizen, R. M., Gozzi, A., Hess, A., Keilholz, S. D., Lerch, J. P., Shih, Y.-Y. I., & Grandjean, J. (2020). Animal Functional Magnetic Resonance Imaging: Trends and Path Toward Standardization. *Frontiers in Neuroinformatics*, *13*, 78.

Mathias, A., Grond, F., Guardans, R., Seese, D., Canela, M., & Diebner, H. H. (2004). Algorithms for spectral analysis of irregularly sampled time series. *Journal of Statistical Software*, *11*(1), 1–27.

Mills, B. D., Grayson, D. S., Shunmugavel, A., Miranda-Dominguez, O., Feczko, E., Earl, E., Neve, K. A., & Fair, D. A. (2018). Correlated Gene Expression and Anatomical Communication Support Synchronized Brain Activity in the Mouse Functional Connectome. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(25), 5774–5787.

Nasrallah, F. A., Lew, S. K., Low, A. S.-M., & Chuang, K.-H. (2014). Neural correlate of resting-state functional connectivity under α2 adrenergic receptor agonist, medetomidine. *NeuroImage*, *84*, 27–34.

Nickerson, L. D., Smith, S. M., Öngür, D., & Beckmann, C. F. (2017). Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Frontiers in Neuroscience*, *11*, 115.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66.

Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *NeuroImage*, *171*,

415–436.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews. Neuroscience*, *18*(2), 115–126.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341.

Power, J. D., Plitt, M., Laumann, T. O., & Martin, A. (2017). Sources and implications of whole-brain fMRI signals in humans. *NeuroImage*, *146*, 609–625.

Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, *112*, 267–277.

Richards, K., Watson, C., Buckley, R. F., Kurniawan, N. D., Yang, Z., Keller, M. D., Beare, R., Bartlett, P. F., Egan, G. F., Galloway, G. J., Paxinos, G., Petrou, S., & Reutens, D. C. (2011). Segmentation of the mouse hippocampal formation in magnetic resonance images. *NeuroImage*, *58*(3), 732–740.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256.

Schlegel, F., Sych, Y., Schroeter, A., Stobart, J., Weber, B., Helmchen, F., & Rudin, M. (2018). Fiber-optic implant for simultaneous fluorescence-based calcium recordings and BOLD fMRI in mice. *Nature Protocols*, *13*(5), 840–855.

Schölvinck, M. L., Maier, A., Ye, F. Q., Duyn, J. H., & Leopold, D. A. (2010). Neural basis of global resting-state fMRI activity. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(22), 10238–10243.

Sforazzini, F., Schwarz, A. J., Galbusera, A., Bifone, A., & Gozzi, A. (2014). Distributed BOLD and CBV-weighted resting-state networks in the mouse brain. *NeuroImage*, *87*, 403–415.

Sierakowiak, A., Monnot, C., Aski, S. N., Uppman, M., Li, T.-Q., Damberg, P., & Brené, S. (2015). Default mode network, motor network, dorsal and ventral basal ganglia networks in the rat brain: comparison to human networks using resting state-fMRI. *PloS One*, *10*(3), e0120345.

Steadman, P. E., Ellegood, J., Szulc, K. U., Turnbull, D. H., Joyner, A. L., Henkelman, R. M., & Lerch, J. P. (2014). Genetic effects on cerebellar structure across mouse models of autism using a magnetic resonance imaging atlas. *Autism Research: Official Journal of the International Society for Autism Research*, *7*(1), 124–137.

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, *1*, 62.

Ullmann, J. F. P., Watson, C., Janke, A. L., Kurniawan, N. D., & Reutens, D. C. (2013). A segmentation protocol and MRI atlas of the C57BL/6J mouse neocortex. *NeuroImage*, *78*, 196–203.

Wang, S., Peterson, D. J., Gatenby, J. C., Li, W., Grabowski, T. J., & Madhyastha, T. M. (2017). Evaluation of Field Map and Nonlinear Registration Methods for Correction of Susceptibility Artifacts in Diffusion MRI. *Frontiers in Neuroinformatics*, *11*, 17.

Zerbi, V., Grandjean, J., Rudin, M., & Wenderoth, N. (2015). Mapping the mouse brain with rs-fMRI: An optimized pipeline for functional network identification. *NeuroImage*, *123*, 11–21.