OXFORD

*Cerebral Cortex*, 2023, **33**, 4013–4025
https://doi.org/10.1093/cercor/bhac323
Advance access publication date 14 September 2022
**Original Article**

# Accurate machine learning prediction of sexual orientation based on brain morphology and intrinsic functional connectivity

Benjamin Clemens[1,2,†,*], Jeremy Lefort-Besnard[3,†], Christoph Ritter[4], Elke Smith[5], Mikhail Votinov[1,2], Birgit Derntl[6,7], Ute Habel[1,2], Danilo Bzdok[8,9,10,11]

[1]Department of Psychiatry, Psychotherapy and Psychosomatics, Faculty of Medicine, RWTH Aachen, Pauwelsstr. 30, 52074 Aachen, Germany,
[2]Research Center Jülich, Institute of Neuroscience and Medicine: JARA-Institute Brain Structure Function Relationship (INM 10), Wilhelm-Johnen-Strase, 52428 Jülich, Germany,
[3]Unicaen, Inserm, Comete, Gip Cyceron, 1400 Caen, Normandie, France,
[4]Interdisciplinary Center for Clinical Research (IZKF), RWTH Aachen University, Pauwelsstr. 30, 52074 Aachen, Germany,
[5]Biological Psychology, Department of Psychology, University of Cologne, Bernhard-Feilchenfeld-Str. 11, 50969 Cologne, Germany,
[6]Department of Psychiatry and Psychotherapy, University of Tübingen, Calwerst. 14, 72076 Tübingen, Germany,
[7]Werner Reichardt Center for Integrative Neuroscience (CIN), University of Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany,
[8]McConnell Brain Imaging Centre, McGill University, 3801 University Rue, Montreal Quebec H3A 2B4, Canada,
[9]Department of Biomedical Engineering, McGill University, 3775 University Rue, Montreal Quebec H3A 2B4, Canada,
[10]Faculty of Medicine, Montreal Neurological Institute (MNI) and Hospital, McGill University, 3801 University Rue, Montreal Quebec H3A 2B4, Canada,
[11]Mila–Quebec Artificial Intelligence Institute, 6666 Rue St-Urbain #200, Montreal Quebec H2S 3H1, Canada

*Corresponding author: Pauwelsstrasse 30, 52074 Aachen, Germany. Email: bclemens@ukaachen.de

[†]Benjamin Clemens and Jeremy Lefort-Besnard contributed equally to the manuscript and share first-authorship.

**Background**: Sexual orientation in humans represents a multilevel construct that is grounded in both neurobiological and environmental factors. **Objective**: Here, we bring to bear a machine learning approach to predict sexual orientation from gray matter volumes (GMVs) or resting-state functional connectivity (RSFC) in a cohort of 45 heterosexual and 41 homosexual participants. **Methods**: In both brain assessments, we used penalized logistic regression models and nonparametric permutation. **Results**: We found an average accuracy of 62% (±6.72) for predicting sexual orientation based on GMV and an average predictive accuracy of 92% (±9.89) using RSFC. Regions in the precentral gyrus, precuneus and the prefrontal cortex were significantly informative for distinguishing heterosexual from homosexual participants in both the GMV and RSFC settings. **Conclusions**: These results indicate that, aside from self-reports, RSFC offers neurobiological information valuable for highly accurate prediction of sexual orientation. We demonstrate for the first time that sexual orientation is reflected in specific patterns of RSFC, which enable personalized, brain-based predictions of this highly complex human trait. While these results are preliminary, our neurobiologically based prediction framework illustrates the great value and potential of RSFC for revealing biologically meaningful and generalizable predictive patterns in the human brain.

*Key words*: fMRI; machine learning; predictive modeling; sexual orientation; resting-state functional connectivity (RSFC).

## Introduction

In the quest for the origins of sexual orientation, scientists have long grappled with the question about the degree to which sexual orientation is biologically determined or socially constructed, i.e. shaped by cultural expectations and personal experiences. While there are arguments and vocal proponents for both (i) biological (Swaab 2007, 2008; Savic et al. 2010; Roselli 2018) and (ii) social or environmental (Butler 1990; Knauer 2000; Eskridge Jr. 2005, 2008) influences that shape sexual orientation, a combined influence of both factors may provide a more realistic answer. However, there is not a single specific theory of sexual orientation that has received overwhelming agreement among scientists.

Today, most experts are sympathetic to the view that the development of sexual orientation seems to result from complex interaction effects between genes, hormones, peers, and social norms, thus perhaps a prime example of the interplay between nature and nurture (Hines et al. 2004; Jorge 2010; Bailey et al. 2016; Balthazart 2016). For clarity and in accordance with previous research (Clemens et al. 2020, 2021; Shah et al. 2012), we will refer to "biological sex" as the sex assigned to an individual at birth and to "sexual orientation" as the sexual attraction or sexual preference. The term heterosexual refers to men or women who are sexually or emotionally attracted to individuals of the other biological sex. Instead, the term homosexual refers to individuals who are sexually or emotionally attracted to individuals of the same biological sex (American Psychological Association 2008).

With respect to putative biological influences, sexual orientation has been linked to neuroendocrine expression levels (Gladue et al. 1984), neurotransmitter systems (Liu et al. 2011), and genetic liability (Hamer et al. 1993). These factors contribute to the shaping

**Received:** April 13, 2022. **Revised:** July 20, 2022. **Accepted:** July 21, 2022

of brain structure and function, especially during early development. As a motivation for the present investigation, it is therefore possible and plausible that brain measurements of hetero- and homosexual individuals also feature discernible anatomical and functional characteristics. Indeed, neuroimaging studies investigating brain morphology and intrinsic functional coupling repeatedly reported differences between heterosexual and homosexual individuals (Frigerio et al. 2021). Earlier brain-imaging research has documented group differences based on both structural (Ponseti et al. 2007; Savic and Lindström, 2008; Witelson et al. 2008; Abé et al. 2014; Votinov et al. 2021; Manzouri and Savic 2018) and functional (Hu et al. 2008; Paul et al. 2008; Ponseti et al. 2009; Zeki et al. 2010; Kagerer et al. 2011; Perry et al. 2013; Sylva et al. 2013; Safron et al. 2017, 2018; Manzouri and Savic 2018; Folkierska-Żukowska et al. 2020) brain measurements. All aforementioned functional neuroimaging studies were task-based functional magnetic resonance imaging (fMRI) studies, i.e. presenting visual or auditory stimuli to evoke sex-specific responses and to analyze differences in brain activity.

Overall, these neuroimaging results paint a colorful picture of research on sexual orientation. Drawing generalizable and robust conclusions from these previous studies has previously been hampered by small sample sizes and the heterogeneity of experimental task designs. For instance, Folkierska-Zukowska et al. (2020) used a mental rotation task in groups of 23 participants. These authors found significant differences in a subset of brain regions, including the right superior frontal gyrus, right angular gyrus, right amygdala/parahippocampal gyrus, and bilaterally in the middle temporal gyrus and precuneus. Safron et al. (2017) examined the activation patterns in response to both erotic pictures and videos in groups of 11 participants. The study did not find distinct brain activity patterns between hetero- and homosexual individuals. Kagerer et al. (2011) focused on the group differences involved in the processing of sexual stimuli in groups of 11 participants. They did not find any group differences either. Hu et al. (2008) as well as Paul et al. (2008) used real-time visual stimulation using groups of 10 participants. The first research team found differences in the left angular gyrus, left caudate nucleus, right pallidum, bilateral lingual gyrus, right hippocampus, and right parahippocampal gyrus. The second research team, however, found similar patterns of activation in both groups. More importantly, most of these task-based fMRI studies did perform a whole-brain GLM. Only a few picked regions of interest (ROIs) based on previous research findings to perform their analysis. This heterogeneity among task-design and analytical choices renders comparison among findings difficult. As a consequence, the possibility of examining sexual orientation in a task-free context, charting resting-state functional connectivity (RSFC), has started to raise attention, thus opening a new window of opportunity to revisit the brain correlates of sexual orientation. As compared to task-based fMRI investigations, RSFC fMRI studies are advantageous for several reasons: They are more easy and cheaper to conduct; they provide much better signal to noise ration; and they allow for a substantially larger suitable population of participants, as no complex set of instructions is required.

In our particular approach, we confronted the question of whether sexual orientation can be reliably predicted from structural and functional brain measurements. If clearly distinguishable brain phenotypes for heterosexual and homosexual individuals are detectable, it should be practically feasible for predictive pattern classification to tell apart an individuals' sexual orientation based solely on a brain scanning session. The present machine learning study put this possibility to a direct test in a rarely available sample of heterosexual and homosexual men and women. In contrast to previous efforts on this topic, our goal is not to design experimental stimuli to induce specific neural responses, localize the associated changes in brain function, and merely compare these changes between homo- and heterosexual groups. Instead, our present investigation provides the first spatially unbiased, whole-brain machine learning approach of male and female homo- and heterosexual participants.

## Materials and methods

### Participants

The magnetic resonance imaging (MRI) data that presented the basis for the present investigation were recorded as part of a larger research project designed to elucidate the neurobiological substrates of gender identity and sexual orientation. Within this project, several fMRI tasks and resting-state fMRI were acquired, which have been described elsewhere (Junger et al. 2013; Clemens et al. 2017, 2020; Smith et al. 2018). For the resting-state condition, participants were asked to relax in the scanner, keep their eyes open, and avoid falling asleep, which we confirmed in postscan interviews. In total, 86 participants took part in the present study: 45 heterosexual participants (HE) and 41 homosexual participants (HO). Basic demographic information on the sample can be found in Table 1. Biological sex was evenly distributed in both homo- and heterosexual participants: 22 heterosexual men, 23 heterosexual women, 22 homosexual men, and 19 homosexual women. All participants were recruited via public announcement in Aachen (Germany) and the surrounding area. Sexual orientation was systematically assessed by a single self-report question ("Do you identify as primarily hetero- or homosexual?"). To rule out the possibility of including individuals with bisexual orientation, only participants who unambiguously indicated either homo- or heterosexual orientation in the screening were eligible for our study. This was verified first by the answer given to the aforementioned screening question (Do you identify as primarily hetero- or homosexual?). Second, the interviewer double-checked the correctness of the

**Table 1.** Demographic information in the participant sample.

| | Heterosexual (HE) | Homosexual (HO) | *P*-value |
|---|---|---|---|
| Participants (men/women) | 45 (22/23) | 41 (22/19) | — |
| Age (SD) | 32 (10) | 28 (6) | <0.05 |
| Years of education (SD) | 15 (3) | 16 (2) | >0.05 |

The table lists demographic information about our participants, divided by groups. For age and education, we present means, with standard deviations (SDs) in brackets. Both age and years of education were compared to check for significant group differences using standard *t*-tests, with significant differences only present for age. As part of the confound-removal procedure, variance that could be explained by the factors "age," "biological sex," and "years of education" was regressed out from both the MRI and the fMRI signals.

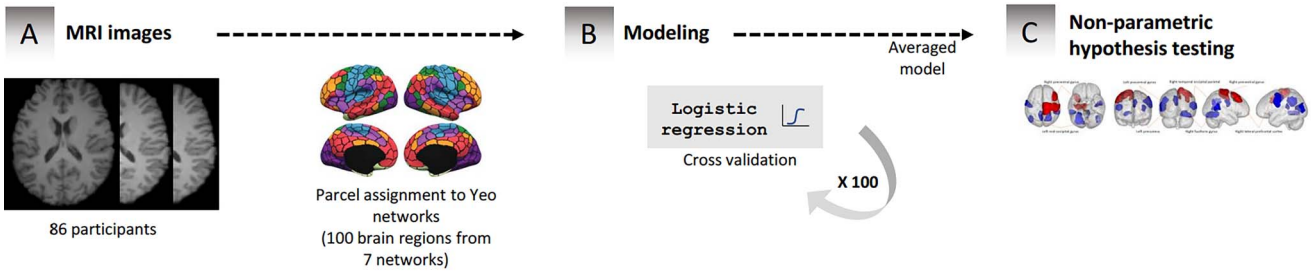


**Fig. 1.** Analysis workflow for prediction from brain morphology. A) Using the Schaefer Yeo reference atlas (Schaefer et al. 2018), quantitative measures of gray-matter volume differences were extracted within the 100 ROIs of the atlas in the 86 participants. B) Linear predictive models were implemented using 5-fold crossvalidation to predict sexual orientation from the GMVs. The crossvalidation procedure was repeated 100 times to ensure stability. The final coefficients and accuracies were then averaged. C) Statistical significance for weights in the final model was assessed and results were overlaid onto the MNI-152 brain for visualization.

answer by verbally asking participants whether they are sure and unambiguous of their indicated response.

The German version of the Structured Clinical Interview of the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (Wittchen et al. 1997) was applied by a board-certified psychologist to ensure exclusion of participants with any mental health diagnoses. Further exclusion criteria were the presence of neurological disorders, other medical conditions affecting the brain metabolism, and first-degree relatives with a history of mental health diagnosis. The local Ethics Committee of the Medical Faculty of RWTH Aachen University approved the study (EK 088/09). Participants were financially compensated and each participant gave written informed consent for participation.

## Brain imaging data acquisition

Using a 3 Tesla Siemens Trio Scanner (Siemens Medical Systems, Erlangen, Germany) at the Department of Psychiatry, Psychotherapy and Psychosomatics of the RWTH Aachen University Hospital, the following sequences covering the entire brain were obtained for each participant: (i) 4-min T1-weighted magnetization prepared rapid gradient echo 3D measurement (time repetition [TR] = 1,900; time echo [TE] = 2.52; TI = 900; $\alpha$ = 9°; FoV=250 mm$^2$; voxel size: 1 × 1 × 1 mm; slices = 176) and (ii) a 6.2-min T2*-weighted echo-planar imaging resting-state condition (TR = 3,000; TE = 35, $\alpha$=84°; FoV = 192 mm; voxel size: 3 × 3 × 3 mm; 44 slices; gap 15%; 64 × 64 matrix; repetitions = 124).

## Image processing

T1 anatomical images were cropped with FMRIB's Software Library's (FSL's) "robustfov" tool to remove neck and lower head, reoriented to match the orientation of the MNI152 standard template with the "fslreorient2std" tool, and nonbrain tissue was removed using Brain Extraction Tool (BET; Smith 2002). Using FMRI Expert Analysis Tool Version 6.00, part of FSL, all preprocessing steps for the fMRI data were carried out using a standard pipeline, which is comparable to previous studies (Satterthwaite et al. 2013; Clemens et al. 2017). This pipeline included spatial and temporal data normalization and special consideration of in-scanner head motion. We discarded the first 3 images of each functional series to avoid T1 saturation effects; the remaining 121 volumes were submitted to downstream analyses. We applied the following signal processing steps: motion correction using Motion Correction FMRIB's linear image registration tool (Jenkinson et al. 2002), nonbrain removal using BET (Smith 20202), spatial smoothing using a Gaussian kernel of full-width at half-maximum of 6 mm, grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, and high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 180 s). Low-pass filtering was explicitly kept to a minimum to preserve high-frequency content and to keep the largest frequency range possible. Registration to high-resolution structural and/or standard space images was carried out using FLIRT (FMRIB's linear image registration tool; Jenkinson and Smith, 2001).

Registration from subject to standard space was then refined using FMRIB's nonlinear image registration tool. Additional preprocessing steps included masking of nonbrain voxels, voxel-wise demeaning of the data, and normalization of data to Montreal Neurological Institute (MNI) space. To further reduce spurious correlations associated with head motion in the 86 participants, variation that could be explained by head motion was

removed from each voxel's time series at the single-subject level. Adhering to previously published studies (Chai et al. 2012; Satterthwaite et al. 2013; Kernbach et al. 2018), we helped remove nuisance-related variation in the brain signals based on 24 regressors of no interest: (i) 6 motion parameters derived from the image realignment, (ii) their 6 first derivatives, and (iii) their respective 12 squared terms. This regression approach has been shown to increase the specificity and sensitivity of functional connectivity analyses and to detect valid signal correlation at rest (Satterwaite et al. 2013).

## Signal extraction for feature space construction

For extracting relevant signals from a functional or structural brain scan, 100 ROIs from the Schaefer Yeo reference atlas (Schaefer et al. 2018) served as the topographic atlas definition. These anatomical regions were used to extract average MRI signals from the voxels belonging to a given ROI. Note that for the sake of reproducibility and comparability, we extracted MRI signals in the structural and functional settings from the same reference atlas (i.e. the Schaefer Yeo atlas). This atlas is readily available online for replication and reuse. In RSFC, each ROI was represented by the average blood oxygen level–dependent (BOLD) signal (121 time-series) across all voxels of that ROI. In structural brain data, each ROI was represented by the average gray matter volume (GMV) across all voxels belonging to a particular ROI. In sum, the resulting feature spaces for further analysis were composed of as many brain features as ROIs in the Schaefer atlas (100 ROIs in total) for the participants. All ROI-wise functional or structural time series were transformed into z-scores by mean centering and unit-variance scaling. As part of the confound-removal procedure, variance that could be explained by the factors age, biological sex, years of education, and handedness was regressed out from the corresponding brain features at the subject-by-subject level.

## Machine learning prediction of sexual orientation from structural brain measures

We analyzed the relative importance of the GMV signal per ROI for predicting sexual orientation using a L2-penalized sigmoid-loss linear predictive model. The L2 shrinkage regularization (default hyperparameter lambda = 1.0) was used to reduce the risk of overfitting, which can render the models' prediction of future observation unreliable (Okser et al. 2014). The L2-penalized logistic regression model estimated the separating hyperplane (i.e. a linear function) in input space that best allowed distinguishing between homosexual and heterosexual participants. The outcome to be predicted was defined by being homosexual (encoded as 0) or being heterosexual (encoded as 1). The model parameters were then fit to optimally predict sexual orientation based on all the ROI GMV using 5-fold crossvalidation. That is, the participants were divided into 5 balanced data splits (fold), each preserving the percentage of participants of both classes (homosexual and heterosexual). In each crossvalidation fold, the predictive model was fitted on 80% of the participants (the training set) and was assessed on the left-out 20% (the test set) of the participants (Hastie et al. 2001). To ensure the stability of the model coefficients, this 5-fold crossvalidation was repeated 100 times (model fitting and testing). The obtained percentages of correctly classified test participants and coefficients were then averaged and these are described in our Results section. A visualization of all analysis steps for brain strcutural data can be found in Fig. 1.

## Machine learning prediction of sexual orientation from functional brain measures

As in the GMV setting, a L2-penalized sigmoid-loss linear predictive model (lambda = 1.0) was used to assess the importance of all the ROI RSFC patterns for predicting sexual orientation. Extending upon the GMV setting, this algorithm was used as stacking classifier. Stacking is a way to build an ensemble algorithm from multiple classifications models. That is, multiple learners (L2-penalized logistic regressions) were used to build an intermediate prediction. A new instance of that algorithm (another L2-penalized logistic regression) was estimated on the same training participants using these intermediate predictions (probabilistic class assignments that are naturally afforded by our pattern-learning model) to predict the same target (sexual orientation). The final model (composite model) is placed on the top of the outcome predictions of the other lower-level models (base models). Put more simply, stacking enables to learn how to best combine the predictions from a collection of learning algorithms. This analytical framework makes it possible to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone (Wolpert, 1992; Breiman, 1996). In this way, we capitalized on the stacking ensemble model to get the most out of our available participant sample. The following steps were performed for the machine learning prediction of sexual orientation from functional brain measures:

### *Pearson correlation coefficients to estimate the functional coupling strengths*

For each participant, the 121 averaged BOLD signals had been extracted in each of the 100 ROIs (see above). A Pearson correlation matrix with all pairwise region-region association strengths was computed for each subject. A Pearson correlation is a measure of linear correlation between the time series of 2 given ROIs (Benesty et al. 2009). Here, we computed the Pearson correlation between the time series (121 scans per participant) of each ROI included in our analysis (100 ROIs). Therefore, we obtained a (symmetric) matrix of size 100 * 100 for each participant. We then extracted every correlation of each ROI. That is, we extracted a total of 99 correlation values per ROI for each participant. Note

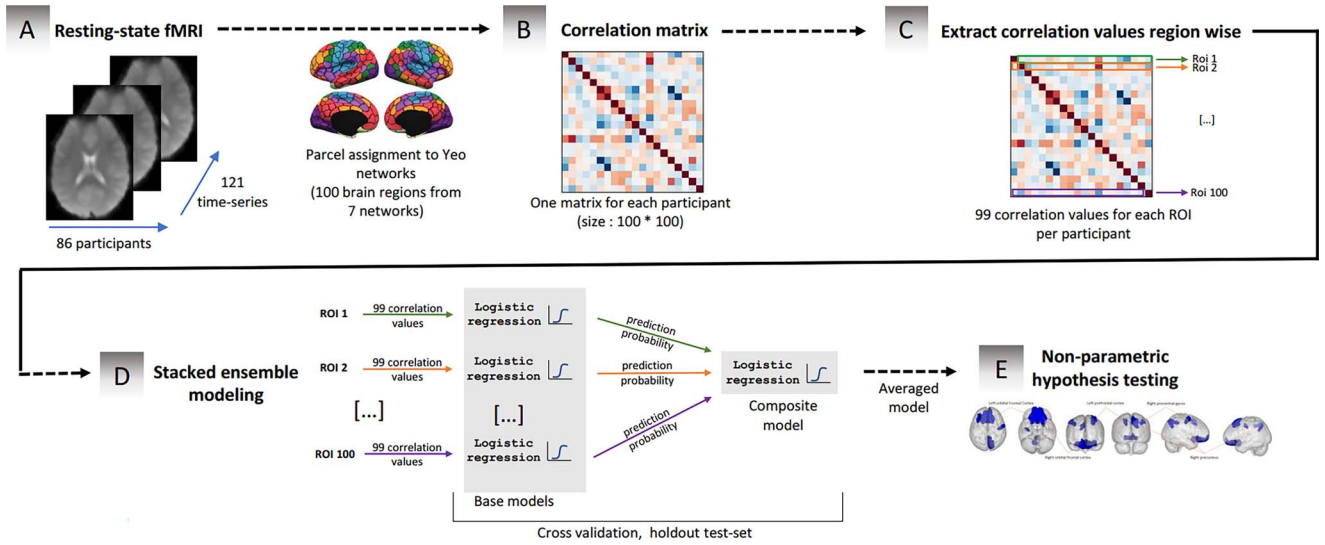


**Fig. 2.** Analysis workflow for prediction from intrinsic function coupling using stacking ensemble modeling. A) Using the Schaefer Yeo reference atlas (Schaefer et al., 2018), the 121 brain scan time points of BOLD signals were extracted in each of the 100 ROIs for each of the 86 participants. B) The Pearson correlation between the averaged bold signal values (121 scans) of each ROI (100 ROIs) were computed. Then, C) for each ROI and participant, the 99 estimates of functional connectivity strength with the remaining ROIs were extracted. D) Separately for each ROI, the 99 correlation values per subject were used to fit a L2-penalized pattern-learning algorithm on 80% of the data. The built logistic regression (i.e. the 100 base models) was then used to predict sexual orientation (homosexual vs. heterosexual) in the left-out participants (20%). The obtained predictions of the base models were combined for training a top-level stacking model (i.e. the composite model). E) statistical significance for weights in the final averaged model (across the previous 5-fold crossvalidated composite models) was assessed and results were overlaid onto the MNI-152 brain for visualization.

that we did not extract 100 coupling estimates as we did not use the correlation with the ROI itself, which would always be equal to 1. The 99 ensuing connectivity values per ROI per subject were then used as input in the base models to classify sexual orientation, with 0 encoding homosexual and 1 encoding heterosexual individuals (see Fig. 2).

### *Base models for ROI-wise prediction of the target phenotype*

Separately for each ROI, we used the 99 correlation values per subject to fit an L2-penalized logistic regression algorithm on 80% of the participants (the training sample) in a 5-fold crossvalidation framework. Then, we used the formed logistic regression model to predict sexual orientation (homosexual vs. heterosexual) in the left-out participants (the test sample). The evaluation of sexual orientation status in new participants yielded practically relevant predictions because the algorithm did not visit these participants during model building process (Gabrieli et al. 2015; Bzdok et al. 2019). Thus, each base model predicted the probability for a given participant to be homo- or heterosexual from the functional brain data correlation of one particular ROI. The independent (probabilistic) sexual orientation status predictions of each ROI correlation with the remaining ROIs served as input for the integrated model (Karrer et al. 2019). In other words, the probabilistic prediction outputs of the base models were values ranging between 0 and 1, indicating a probability of predicting a homosexual or a heterosexual participant. The continuous prediction probabilities were stored and served as intermediate input features for a new model fit.

### *Stacking models for ensemble prediction of the target phenotype*

The previous specific predictions of the base models were combined for training a more powerful predictive model in an integrative modeling approach. The built ensemble model considered the separate probabilities of all ROI correlations with the remaining set of ROIs at the same time for sexual orientation prediction. In other words, the continuous prediction probabilities were used to fit a L2-penalized logistic regression algorithm on 80% of the participants (the same identical training sample as in the previous step) again in a 5-fold crossvalidation fashion. It is important to appreciate that the model estimations of the base models as well as the composite model were carried out on only 80% of the participants (the training set). Each fold of the crossvalidation procedure contained 20% of independent participants, the same 20% in the first and the second steps of the stacking analysis. This modeling strategy ensured that no information leaked into the fitting process and guarded the analysis workflow against circular analysis. Finally, we averaged weights and accuracies across the 5 CV folds for the final model fit. A visualization of all analysis steps for brain functional data can be found in Fig. 2.

## Testing for significance

A L2-penalized predictive model based on GMV and a stacked L2-penalized logistic regression based on RSFC were conducted separately. Statistical significance for weights in each of the 2 final models was assessed based on *P* values derived through a rigorous nonparametric permutation approach using the model weights as the test statistic (Nichols and Holmes 2002; Efron 2012).

Relying on minimal modeling assumptions, a valid null distribution was derived for the achieved weights resulting from the logistic regression fit. In 100 permutation iterations, the brain volume matrix was held constant, while the sexual orientation underwent participant-wise random shuffling. While the constructed surrogate brain signals preserved the statistical structure idiosyncratic to the MRI-derived brain signals, they were permitted to selectively destroy the signal property related to the logistic regression weight to be tested. The empirical distribution generated in this manner reflected the null hypothesis of random association between the volume and sexual orientation across participants. The beta coefficients were recorded in each iteration. The *P* values were obtained given the distance between the original beta values and the mean beta values obtained during the permutation iterations. Note that significance in this setting does not imply single contribution of the ROI with a significant weight in the prediction. Indeed, in linear models, the target value is modeled as a linear combination of all the features. In other words, the models exploited information carried by all included ROIs to predict sexual orientation. Significance here demonstrates the systematic contribution of a ROI at hand for contributions to distinguishing between the 2 groups but does not mean that other ROIs did not contribute at all. This analysis setting is an instance of the prediction-inference dilemma (Bzdok and Yeo 2017; Bzdok et al. 2019, 2020): The modeling goals of identifying single-variable effects and maximizing prediction performance in unseen subjects are incongruent.

### Code availability

All analysis scripts of the present study are readily accessible to the reader online (https://github.com/JLefortBesnard/SexualOrientation_2020).

## Results

### Brain structure

Our findings indicated a prediction of sexual orientation slightly above chance level, revealing 8 significant weights in the GMV setting ($P < 0.05$). The mean accuracy of the averaged GMV models, incorporating only structural MRI data, was 62.31%, with a standard error of 6.72%. This indicates an overall prediction of sexual orientation slightly above chance level. Figure 3 shows those brain regions for which GMV were highly weighted for predicting heterosexual or homosexual individuals. There were 5 significant ROIs ($P < 0.05$) that consistently contributed to predicting homosexual orientation. That is, model parameter values corresponding to these ROIs were repeatedly exceeding chance level as part of the overall model for predicting homosexual orientation of single individuals. These most discriminatory ROIs included the left middle occipital gyrus (coefficient = −0.55), left precentral gyrus (coefficient = −0.94), left precuneus/posterior cingulate cortex (coefficient = −0.59), right fusiform gyrus (coefficient = −0.75), and right temporal occipital-parietal cortex (coefficient = −0.62). On the other hand, there were 3 significant ROIs ($P < 0.05$) that consistently contributed to predicting heterosexual orientation. That is, model parameter values corresponding to these ROIs were repeatedly exceeding chance level as a part of the overall model for predicting the heterosexual orientation in single individuals. These most discriminatory ROIs included 2 clusters in the right precentral gyrus (coefficient = 0.72 and 0.62) and the right lateral prefrontal cortex (coefficient = 0.50). Thus, GMV in visual cortices, somatosensory cortices, and parts of the DMN were informative about predicting sexual orientation, with lateralized results in the somatosensory cortices. Detailed information on the significant GMV regions is provided in Table 2. Figure 4A depicts the receiver operating characteristic curve (ROC), which measures the classification performance at various threshold settings, while Fig. 4B depicts the confusion matrix, with the percentage of correct classification and misclassification, based on structural MRI data, for both homo- and heterosexual participants. Overall, 64.6% of the homosexual and 60.5% of the heterosexual individuals were correctly classified by our L2-penalized logistic regression model for GMV. On the other hand, 35.4% of the homosexual and 39.5% of the heterosexual participants were misclassified. Thus, classification accuracy was similar for both hetero- and homosexual participants and in both cases only slightly above the chance level of 50%.

### Brain function

The mean accuracy of the averaged models incorporating only fMRI data was 91.76%, with a standard error of 9.89. Figure 5 shows those brain regions that were highly weighted for predicting heterosexual or homosexual individuals. There were 5 significant weights ($P < 0.05$) indicating that the respective ROIs were informative for predicting homosexual orientation: left orbito-frontal cortex (coefficient = −0.51), left prefrontal cortex (coefficient = −0.58), right precentral gyrus (coefficient = −0.69), right orbito-frontal cortex (coefficient = −0.44), and right precuneus (coefficient = −0.49). None of the ROIs indicating high relevance for the prediction of heterosexual orientation were significant. Detailed information on the significant RSFC regions is provided in Table 3. Figure 6A depicts the ROC curve, which measures the classification performance at various threshold settings, while Fig. 6B depicts the confusion matrix, with the percentage of correct classification and misclassification, based on RSFC, for both homo- and heterosexual participants. Overall, 82.9% of homosexual and 100% of heterosexual individuals were correctly classified by our logistic regression model. Thus, all heterosexual participants were correctly classified, whereas 17.1% of homosexual participants were misclassified as heterosexual.

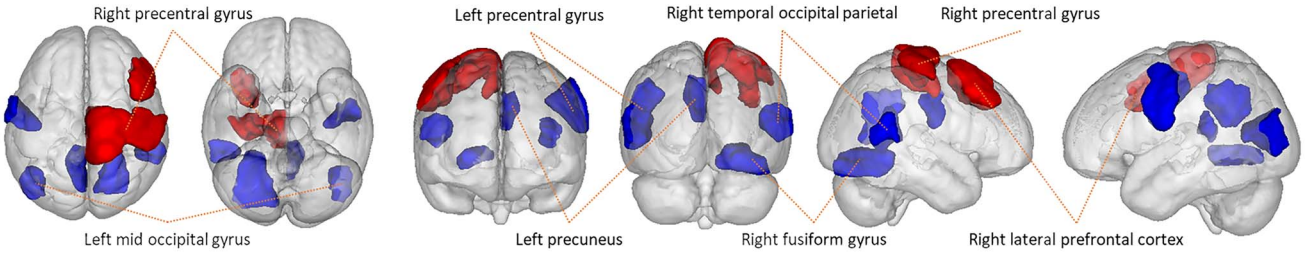


**Fig. 3.** GMV results for predicting sexual orientation. GMV of these regions overlaid on the MNI-152 template brain showed statistically significant weights in the prediction of sexual orientation. The blue regions contributed to detecting homosexual individuals, while the red regions contributed to detecting heterosexual individuals.

**Table 2.** MNI coordinates of the significant brain regions in the structural MRI analysis and their associated coefficients from the modeling.

| Significant ROIs | Area | Network | x | y | z | Coefficient |
|---|---|---|---|---|---|---|
| Left midoccipital gyrus | 7 | Visual | −28 | −76 | −14 | −0.55 |
| Left precentral gyrus | 4 | Somatomotor | −56 | −16 | 6 | −0.94 |
| Left precuneus/PCC | 2 | Default mode | −6 | −46 | 12 | −0.59 |
| Right right fusiform gyrus | 2 | Visual | 32 | −26 | −22 | −0.75 |
| Right precentral gyrus | 6 | Somatomotor | 44 | −18 | 0 | −0.62 |
| Right precentral gyrus | 8 | Somatomotor | 64 | −22 | 0 | 0.72 |
| Right temporal occipital parietal | 1 | Salience/ventral attention | 60 | −46 | 6 | 0.62 |
| Right lateral prefrontal cortex | 4 | Control | 44 | 46 | −14 | 0.5 |

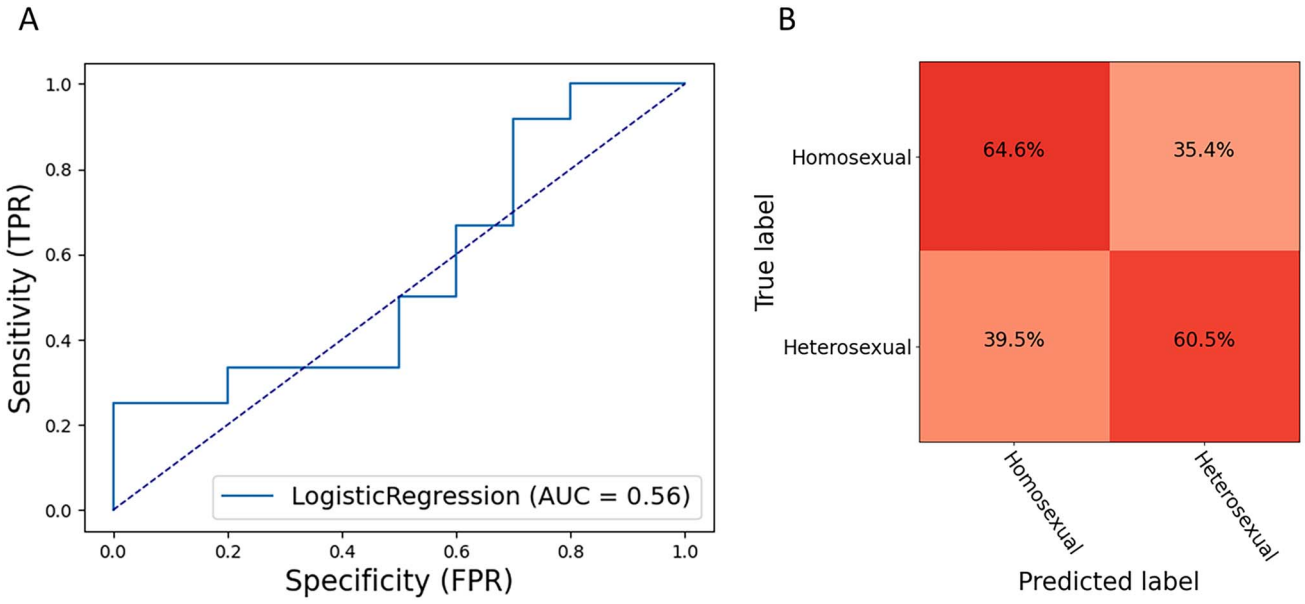


**Fig. 4.** ROC curve and confusion matrix of the sexual orientation prediction based on GMV. A) ROC curve plot of true positive rate versus false positive rate. The area under the ROC curve (AUC) measures the classification performance at various threshold settings. The higher the AUC, the better the model is at distinguishing between homosexual and heterosexual participants. Our model gave an AUC of 0.56 which can be considered as acceptable. B) The y-axis shows each of the 2 targets (homosexual or heterosexual), while the x-axis shows each of the 2 predicted labels. The upper left and lower right squares display the percentage of correct classification, while the upper right and lower left squares display the percentage of misclassification. Here, 64.6% of homosexual participants and 60.5% of heterosexual participants were correctly classified by the linear model. In each case, the percentage of correctly classified participants was higher than the percentage of misclassified participants, suggesting a fair performance of the model.

## Discussion

Here, we have developed a machine learning pipeline to provide a proof of principle that an individual's sexual orientation could be predictable based on noninvasive brain measurements alone. We report a predictive accuracy for structural MRI data which was slightly above the chance level. This brain-imaging modality may thus be less suitable to achieve highly accurate predictions of sexual orientation. Previous studies investigating sexual orientation using structural MRI focused primarily on the cortical and subcortical thicknesses as well as diffusion tensor imaging modalities. They also did not use a consistent approach to answer their question. Most studies

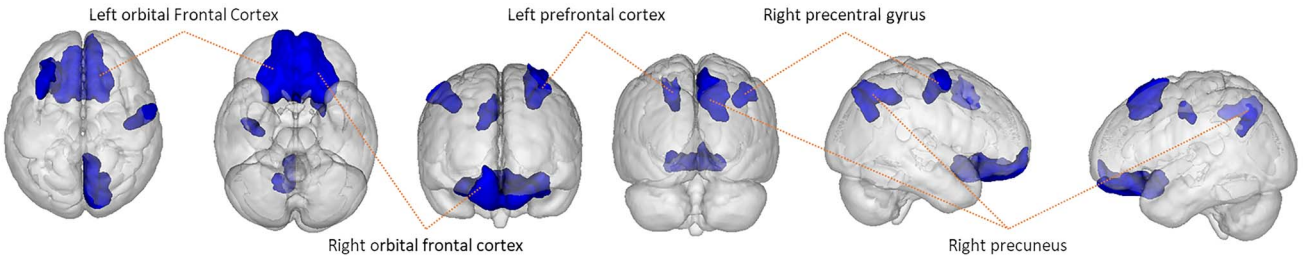


**Fig. 5.** RSFC results for predicting sexual orientation**.** RSFC of these regions overlaid on the MNI-152 template brain showed statistically significant weights in the prediction of sexual orientation. The blue regions contributed to detecting homosexual individuals.

**Table 3.** MNI coordinates of the significant brain regions in the RSFC analysis and their associated coefficients from the modeling.

| Significant ROIs | Area | Network | x | y | z | Weight |
|---|---|---|---|---|---|---|
| Left orbito-frontal cortex | 1 | Limbic | −14 | 18 | −22 | −0.51 |
| Left prefrontal cortex | 7 | Default mode | −20 | 62 | −8 | −0.58 |
| Right precentral gyrus | 5 | Somatomotor | 56 | −10 | 4 | −0.69 |
| Right orbito-frontal cortex | 1 | Limbic | 22 | 14 | −22 | −0.44 |
| Right precuneus | 1 | Control | 16 | −66 | 32 | −0.49 |

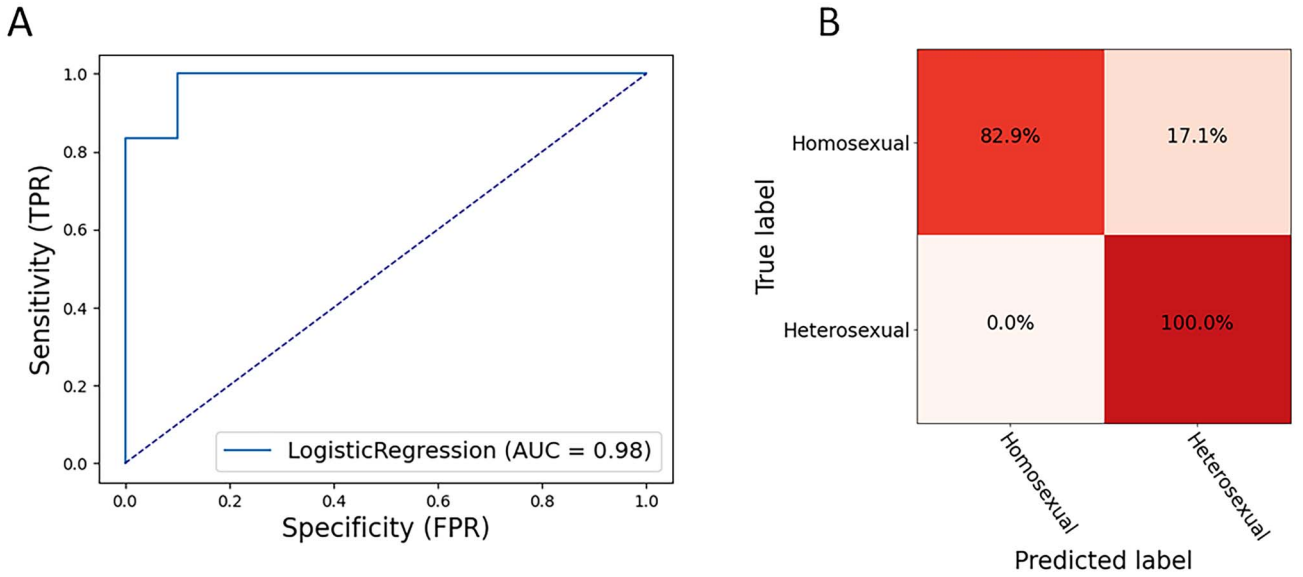


**Fig. 6.** ROC curve and confusion matrix of the sexual orientation prediction based on RSFC. A) ROC curve plot of true positive rate versus false positive rate. The AUC measures the classification performance at various threshold settings. The higher the AUC, the better the model is at distinguishing between homosexual and heterosexual participants. Our model gave an AUC of 0.98 which can be considered as very good. B) The y-axis shows each of the 2 targets (homosexual or heterosexual), while the x-axis shows each of the 2 predicted labels. The upper left and lower right squares display the percentage of correct classification, while the upper right and lower left squares display the percentage of misclassification. Here, 82.9% of homosexual participants and 100% of heterosexual participants were correctly classified by the linear model. In each case, the percentage of correctly classified participants was considerably higher than the percentage of misclassified participants, suggesting a very good performance of the model.

reported cortical thickness differences between groups (Abé et al. 2014; Manzouri and Savic 2019), while others have questioned differences in subcortical volumes, hemispheric asymmetries or using diffusion modalities (Frigerio et al. 2021). As a result, when comparing previous results with the predictive ROIs we found in our study, we could not find matching brain regions for distinguishing sexual orientation. In contrast with the structural analysis, using only RSFC to predict sexual orientation, resulted in a prediction accuracy of 92% (±9.89). Part of the explanation for this discrepancy in prediction accuracy might reside in the fundamentally different nature of the 2 windows into the human central nervous system as well as associated methodological constraints. By means of a stacking modeling strategy (Karrer et al. 2019), the pattern-learning problem (i.e. predicting sexual orientation) in the RSFC analyses was tackled with different models which learned some part of the information contained in the functional connectivity but not all the information. The new model, in turn, learned from the intermediate predictions and thus improved the overall performance. Due to inherent differences in the data

structure, this complex and recursive learning strategy could only be applied in the RSFC setting. Therefore, a better accuracy in the RSFC setting might be expected for 2 reasons: (i) the stacking approach drastically improved prediction performance and (ii) as there are much more dimensions in the RSFC settings, there also are more variations.

Focusing on RSFC, we demonstrated for the first time, that a simple 5-min resting-state scan of the human brain may provide a sufficiently strong basis to predict sexual orientation in healthy participants with high accuracy. Adding to the corpus of previous neuroimaging studies, our results are compatible with the interpretation that sexual orientation is linked to certain patterns of RSFC which can be used for the prediction of this highly complex human trait. This observation potentially supports a neurobiological component of sexual orientation: Such an accurate prediction of homo- versus heterosexual orientation is only plausible if RSFC allowed extracting predictive rules to tell apart the 2 sexual orientations. Many previous investigations in the field were task-related fMRI studies whose results may have been sensitized to the fact that they employ experimental tasks that are specifically designed to elicit brain responses related to sexual orientation (e.g. presenting erotic stimuli of male and female persons). In task-based fMRI studies, deviations in the neural activity between hetero- and homosexual participants are supposed to be present predominately because of the nature of the task itself. Examining RSFC offers a more native und unbiased approach toward the connectomic basis of sexual orientation, as it eliminates the potential confounding influence of specific tasks.

While the differences in RSFC between homo- and heterosexual participants allow for a rather accurate prediction of sexual orientation, our study cannot speak to the question of what caused these group differences. Future longitudinal extensions of our present investigation might be able to trace out the trajectory of sexual orientation and its brain-behavior correspondence by comparing the predictive accuracy of our pattern-learning algorithms from the beginning of puberty to later adulthood. Indeed, individual life experience can shape the brain in multiple ways, thus brain plasticity throughout lifetime should be studied as potential confounding variable as much as a variable of interest in future research. The lived experiences of homosexual individuals can be quite different, and differences in the experience-dependent plasticity could also shape the brain in multiple ways. In order to speak to nature/nurture issues, we would need to observe trait-effects for neurobiological properties for which experience-dependent plasticity is an unlikely explanation. Nevertheless, we still want to discuss the evidence of biological factors that are differentially reflected in the brains of individuals who identify as hetero- and homosexual, bringing to the surface the possible brain substrates that provide the basis for the predictive rules extractable in the present study.

In agreement with previous studies, one might speculate that the high predictive accuracy of brain functional connectivity observed here might be the result of key biological influencing factors, including specific genetic factors and prenatal hormones (Ganna et al. 2019). With respect to genetic influences, both twin studies and molecular genetic studies have suggested that the genetic influences on sexual orientation might be rather subtle (Eskridge Jr. et al. 2005; Vasey et al. 2013). Several deoxyribonucleic acid linkage studies, in sibling-pairs with 1 homosexual brother, showed an increased sharing of the Xq28 chromosomal region (Hu et al. 1995; Sanders et al. 2015). Other studies (Rice et al. 1999; Mustanski et al. 2005), however, failed to replicate this result, and no significant genetic linkage has been reported for female sexual orientation. A genome-wide association study (GWAS), including data from 24,000 individuals with self-reports of sexual orientation, revealed that the genetic marker closest to statistical significance was located on the pericentromeric chromosome 8 (Drabant et al. 2012). The seminal study by Ganna et al. (2019), comprising genetic data from 477,522 individuals, concluded that all included genetic variants accounted for maximally 25% of variation. Only 2 of the 5 significant loci were associated with homosexual behavior and were significant for the combined sample of male and female individuals: rs11114975-12q21.31 and rs10261857-7q31.2. Thus, no GWAS study so far allows for a meaningful prediction of an individual's sexual orientation. As for the genetic underpinnings of other complex behavioral phenotypes, it is expected that the singular influence of any sexual orientation related genetic locus will be minimal at the single-subject level. In contrast to potential genetic contributions, theories implicating hormonal influences on sexual orientation appear to be more broadly validated (Zucker et al. 1996; Bailey et al. 2016). In the prenatal period of development, during which fetal brains are particularly sensitive to hormonal influences, there seems to be an underexposure to prenatal androgens in homosexual men and an overexposure to prenatal androgens in homosexual women (Ellis and Ames 1987; Mustanski et al. 2002; Bailey et al. 2016). This rationale is supported by early hormonal manipulation in rats and mice and by studies on the sexual orientation of individuals with atypical hormonal development. Thus, a corpus of existing research suggests that genes and hormones probably have some neurobiological influence on the shaping of sexual orientation and the accompanying brain functional changes. But previous studies suggest that social and environmental factors also play a role (Bailey et al. 2016), which is supported by longitudinal studies reporting fluidity in sexual orientation (Savin-Williams and Ream 2007).

In comparison to previous RSFC studies on sexual orientation, the present study provides the first spatially unbiased, whole-brain machine learning approach of male and female homo- and heterosexual participants.

By contrast, most previous resting-state investigations employed a comparative approach and focused solely on seed-based analyses or a priori-defined resting-state networks, trying to demonstrate increased or decreased functional connectivity in relation to sexual orientation. Several of these previous studies have provided valuable insights, enabling basic inference based on group comparisons and testing of specific hypotheses of localized brain functional differences. But only a data-driven, whole-brain, machine learning approach can transcend this comparative approach and enable us to detect generalizable predictive patterns related to sexual orientation in RSFC.

Thus, although the nature of our results is fundamentally different from previous studies, we still want to examine any potential overlap or discrepancy with previous RSFC studies. Previous studies examining RSFC repeatedly reported an involvement of the ACC, medial prefrontal cortex, caudate, putamen, and the precuneus (Savic and Lindström 2008; Manzouri and Savic 2018; Manzouri and Savic 2019). For these midline regions, reduced connectivity was typically found in homosexual participants. A potential interpretation of these aforementioned findings might be that self-referential thoughts and self-related perception are different in homosexual individuals. The aspect of self-related perception makes sense, given that the aforementioned areas form part of the DMN, which is specifically related to self-referential processing. Our RSFC analyses seems in agreement with these previous findings since we also found key nodes of the DMN to be predictive for sexual orientation in the present study. This specifically includes the left precuneus and precentral gyrus as well as several bilateral prefrontal areas. Specifically, the precuneus is directly connected to functional brain networks processing visual and pheromonal stimuli as well as sexual arousal (Berglund et al. 2006; Witelson et al. 2008; Zhang et al. 2012). Thus, it seems plausible that the precuneus in combination with the prefrontal cortex is involved in processing aspects of sexual orientation. RSFC differences in primary and secondary visual cortices were also found in a previous study (Hu et al. 2013). A potential interpretation for these previous findings in visual processing areas might be the differences in visuospatial orientation between heterosexual and homosexual individuals (Wegesin 1998; Rahman and Wilson 2003; Rahman et al. 2011). Here, we did not find significant effects in visual processing areas in our RSFC data. We suggest that for the specific prediction of sexual orientation based on RSFC, higher-order association areas as part of the DMN are more important and informative than visual processing areas.

Using a well-controlled, state-of-the-art machine learning framework, the present study provides factual information on how well sexual orientation can be predicted based on real neurobiological evidence. We provide initial support for the idea that sexual orientation represents a complex construct which is more strongly related to RSFC. Our quantitative modeling approach based on individual brain anatomical/connectivity patterns potentially enables personalized predictions, thus going beyond strict group comparisons between homo- and heterosexual groups. The accuracies of these predictions do vary substantially in our study, depending on the specific brain parameter being used (GMV vs. RSFC). As outlined earlier, there might be specific methodological reasons for this discrepancy in predictive accuracy. In addition, the lower prediction accuracy for GMV, compared to RSFC, might also indicate that sexual orientation is less strongly implemented and measurable in the anatomical features of the brain. Instead, sexual orientation can be more accurately described based on dynamic features of RSFC, indicating that it represents a complex construct which is jointly shaped by an interplay of personal experiences, social factors, and neurobiological variables. In general, brain anatomy is considered to be more rigid and less flexible than brain functional connectivity, which therefore might be more important for shaping the way in which we perceive, process, and interpret information.

We are fully aware, that the topic being studied here intersects with other sensible areas of personal life and may have political or civil implications for sexual minority groups. Some homosexual individuals might be skeptical regarding the predictive nature of its findings. Any kind of research on sexual orientation has become increasingly controversial in modern times due to political, cultural, and social implications. From a scientific point of view, sexual orientation has been understudied because it is politically controversial. Members of the homosexual community might fear that academic research on sexual orientation will be misconstrued to substantiate agendas of oppression or exclusion: Outside the context of balanced and nuanced scientific discussions, results might be used to justify discrimination against homosexual individuals. In light of these aforementioned issues, we want to make it abundantly clear that this is not our intention and that we clearly distance ourselves from such discriminatory perspectives. We believe that LGBTQIA+ rights are absolutely fundamental and that every human being should be allowed to be attracted to whomever he or she wants. We explicitly do not pursue any political agenda with our research and are interested in the topic purely from a scientific perspective. We believe that specifically such highly controversial areas could greatly benefit from a scientific and evidence-based approach and discussion. We try to be as transparent about the limitations and shortcomings of the present study as possible (see below). We want to clearly point out that our results should only be interpreted as pointing toward the rich diversity of human sexual orientation. Most importantly, to prevent any misinterpretation, we want to explicitly state that our results do not point toward a role for

oppression or discrimination on the basis of sexual orientation.

Limitations to keep in mind when interpreting results from the present study include a limited sample size compared to other neuroimaging investigations with wider subject populations. It has indeed been noted before that smaller participant samples can give better out-of-sample model performances than what can be obtained from larger, more diverse populations (Benkarim et al. 2022; Woo et al. 2017). Thus, replication in separate and larger datasets in future research will be important to consolidate our results. However, we opted for an analysis strategy which learns from data how to best combine individual ROI-wise predictions to obtain more accurate and robust models. In other words, though a larger sample size would benefit this approach, the stacking framework allowed to exploit the current sample to its fullest (Karrer et al. 2019). Nevertheless, in the future, we hope to join forces with other groups investigating neuroimaging efforts toward lifestyle traits of individuals with the prospect of fusing multiple datasets to achieve larger sample sizes. Large, general-purpose population datasets, such as ENIGMA or the UK Biobank, are not ideally suited for studying sexual orientation because it is almost never assessed in a systematic manner. Furthermore, our study was focused on 1 self-report question for determining sexual orientation. As a next step in future investigations in this direction, using Kinsey scores would provide a number of advantages. These indicators have the potential to provide a more precise picture and allow for direct integration with common and advanced brain-imaging derived assessments. Furthermore, this research predicted sexual orientation in a purely task-free context. Such a stimulus-independent setting allows studying more directly intrinsic functional connectivity—one of the most high-yield brain measurements at our disposal to capture a person's specific brain "fingerprint" (Finn et al. 2015)—and may permit to readily increase sample size in future studies given the reduced logistic costs. However, it is important to note that task-based settings may evoke some discriminant brain correlates that cannot be uncovered using RSFC.

## Conclusion

As the key result of our present machine learning investigation, a snapshot of functional coupling links alone may be sufficient to accurately predict sexual orientation. Such neurobiologically based information, which is (completely) free of task influences or other social desirability aspects, supports the notion of clearly distinguishable brain phenotypes for heterosexual and homosexual individuals. We thereby shed new light on the brain correlates of sexual orientation in a task-free context, focusing more on the patterns and differences of intrinsic functional connectivity. These results indicate that, aside from self-reports, RSFC offers neurobiological information valuable for the highly accurate prediction of sexual orientation. Of course, the complex results presented here warrant additional studies and it would be premature to draw firm conclusions. Additional studies should be conducted to test the replicability and generalizability of the observed findings. Nevertheless, our study illustrates the value and potential of RSFC for biologically meaningful and generalizable predictive patterns that provide clues to the functional organization of the human brain.

## Acknowledgements

The authors thank Sabine Bröhr, Cordula Kemper, and Maria Peters for their assistance and support. Specifically, we are indebted to all participants whose support and participation made this study possible.

## Funding

This study was supported by the German Research Foundation (DFG: HA 3202/7-1) and the Brain Imaging Facility of the Interdisciplinary Centre for Clinical Research of the Faculty of Medicine at RWTH Aachen University, Germany. D.B. was supported by the Brain Canada Foundation, through the Canada Brain Research Fund, with the financial support of Health Canada, National Institutes of Health (NIH R01 AG068563A) and the Canadian Institute of Health Research (CIHR 438531). D.B. was also supported by the Healthy Brains Healthy Lives initiative (Canada First Research Excellence Fund), Google (Teaching and Research Award), and by the CIFAR Artificial Intelligence Chairs program (Canada Institute for Advanced Research).

*Conflict of interest statement*: Benjamin Clemens serves as scientific advisor for Dionysus Digital Health, Inc. and holds shares of this company.

## Data availability

Restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. The data are, however, available from the authors upon reasonable request and with the permission of the funding agency (DFG) and the Ethics Committee of the Medical Faculty of RWTH Aachen University.

## References

Abé C, Johansson E, Allzén E, Savic I. Sexual orientation related differences in cortical thickness in male individuals. *PLoS One*. 2014:9(12):e114721.

American Psychological Association. *Answers to your questions: for a better understanding of sexual orientation & homosexuality*. Washington (DC): (CO):: American Psychological Association; 2008. pp. 163–208. http://www.apa.org/topics/lgbt/orientation.aspx

Bailey JM, Vasey PL, Diamond LM, Breedlove SM, Vilain E, Epprecht M. Sexual orientation, controversy, and science. *Psychol Sci Public Interest*. 2016:17(2):45–101.

Balthazart J. Sex differences in partner preferences in humans and animals. *Philos Trans R Soc Lond B Biol Sci*. 2016:371(1688): 20150118.

Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. In: *Noise reduction in speech processing*. Berlin, Heidelberg (DE): Springer; 2009. pp. 1–4

Benkarim O, Paquola C, Park B-y, Kebets V, Hong S-J, Vos R, de Wael S, Zhang BTT, Yeo ME, Ge T. Population heterogeneity in clinical cohorts affects the predictive accuracy of brain imaging. *PLoS Biol*. 2022:20(4):e3001627.

Berglund H, Lindstrom P, Savic I. Brain response to putative pheromones in lesbian women. *Proc Natl Acad Sci*. 2006:103: 8269–8274.

Butler J. Gender trouble, feminist theory, and psychoanalytic discourse. *Feminism/postmodernism*. 1990:327:x.

Breiman L. Stacked regressions. *Mach Learn*. 1996:24(1):49–64.

Bzdok D, Engemann D, Thirion B. Inference and prediction diverge in biomedicine. *Patterns*. 2020:1(8):100119. https://doi.org/10.1016/j.patter.2020.100119.

Bzdok D, Nichols TE, Smith SM. Towards algorithmic analytics for large-scale datasets. *Nat Mach Intell*. 2019:1(7):296–306.

Bzdok D, Thomas Yeo BT. Inference in the age of big data: future perspectives on neuroscience. *NeuroImage*. 2017:155(July):549–564. https://doi.org/10.1016/j.neuroimage.2017.04.061.

Chai XJ, Castañón AN, Öngür D, Whitfield-Gabrieli S. Anticorrelations in resting state networks without global signal regression. *NeuroImage*. 2012:59(2):1420–1428.

Clemens B, Junger J, Pauly K, Neulen J, Neuschaefer-Rube C, Frölich D, Mingoia G, Derntl B, Habel U. Male-to-female gender dysphoria: gender-specific differences in resting-state networks. *Brain Behavior*. 2017:7(5):e00691.

Clemens B, Derntl B, Smith E, Junger J, Neulen J, Mingoia G, Schneider F, Abel T, Bzdok D, Habel U. Predictive pattern classification can distinguish gender identity subtypes from behavior and brain imaging. *Cereb Cortex*. 2020:30(5):2755–2765.

Clemens B, Votinov M, Puiu AA, Schüppen A, Hüpen P, Neulen J, Derntl B, Habel U. Replication of previous findings? Comparing gray matter volumes in transgender individuals with gender incongruence and cisgender individuals. *J Clin Med*. 2021:10(7):1454.

Drabant EM, Kiefer AK, Eriksson N, Mountain JL, Francke U, Tung JY, Hinds DA, Do CB. Genome-wide association study of sexual orientation in a large, web-based cohort. *Poster presented at the 2012 Annual Meeting of the American Society for Human Genetics*, San Francisco, CA. 2012.

Efron B. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Cambridge, UK: Cambridge University Press; 2012. p. 1

Ellis L, Ames MA. Neurohormonal functioning and sexual orientation: a theory of homosexuality heterosexuality. *Psychol Bull*. 1987:101(2):233–258.

Eskridge WN Jr. Body politics: Lawrence v. Texas and the constitution of disgust and contagion. *Fla L Rev*. 2005:57:1011.

Eskridge WN Jr. A pluralist theory of the equal protection clause. *U Pa J Const L*. 2008:11:1239.

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Todd Constable R. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci*. 2015:18(11):1664–1671.

Folkierska-Żukowska M, Rahman Q, Marchewka A, Wypych M, Drozdziel D, Skoloswki DW. Male sexual orientation, gender nonconformity, and neural activity during mental rotations: an fMRI study. *Sci Rep*. 2020:10:18709.

Frigerio A, Ballerini L, Hernández MV. Structural, functional, and metabolic brain differences as a function of gender identity or sexual orientation: a systematic review of the human neuroimaging literature. *Arch Sex Behav*. 2021:50:3329–3352.

Gabrieli JD, Ghosh SS, Whitfield-Gabrieli S. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*. 2015:85(1):11–26.

Ganna A, Verweij KJ, Nivard MG, Maier R, Wedow R, Busch AS, Abdellaoui A, Guo S, Sathirapongsasuti JF, Lichtenstein P, et al. Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior. *Science*. 2019:365(6456):eaat7693.

Gladue BA, Green R, Hellman RE. Neuroendocrine response to estrogen and sexual orientation. *Science*. 1984:225(4669):1496–1499.

Hamer DH, Hu S, Magnuson VL, Hu N, Pattatucci AM. A linkage between DNA markers on the X chromosome and male sexual orientation. *Science*. 1993:261(5119):321–327.

Hastie T, Tibshirani R, Friedman J. *Linear discriminant analysis. The elements of statistical learning*. New York City, USA: Springer; 2001. p. 84

Hines M, Brook C, Conway GS. Androgen and psychosexual development: core gender identity, sexual orientation, and recalled childhood gender role behavior in women and men with congenital adrenal hyperplasia (CAH). *J Sex Res*. 2004:41(1):75–81.

Hu S, Pattatucci AML, Patterson C, Li L, Fulker DW, Cherny SS, Kruglyak L, Hamer DH. Linkage between sexual orientation and chromosome Xq28 in males but not in females. *Nat Genet*. 1995:11(3):248–256.

Hu SH, Wie N, Wang QD, Yan LQ, Wei EQ, Zhang MM, Hu JB, Huang ML, Xu Y. Patterns of brain activation during visually evoked sexual arousal differ between homosexual and heterosexual men. *Am J Neuroradiol*. 2008:29(10):1890–1896.

Hu S, Xu D, Peterson B, Wang Q, He X, Hu J, Xu X, Wei N, Long D, Huang M, et al. Association of cerebral networks in resting state with sexual preference of homosexual men: a study of regional homogeneity and functional connectivity. *PLoS One*. 2013:8(3):e59426.

Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal*. 2001:5(2):143–156.

Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*. 2002:17(2):825–841.

Jorge JC. The embryology of gender. *J LGBT Youth*. 2010:7(4):310–319.

Junger J, Pauly K, Bröhr S, Birkholz P, Neuschaefer-Rube C, Kohler C, Schneider F, Derntl B, Habel U. Sex matters: neural correlates of voice gender perception. *NeuroImage*. 2013:79:275–287.

Kagerer S, Klucken T, Wehrum S, Zimmermann M, Schienle A, Walter B, Vaitl D, Stark R. Neural activation toward erotic stimuli in homosexual and heterosexual males. *J Sex Med*. 2011:8(11): 3132–3143.

Karrer TM, Bassett DS, Derntl B, Gruber O, Aleman A, Jardri R, Laird AR, Fox PT, Eickhoff SB, Simon B, et al. Brain-based ranking of cognitive domains to predict schizophrenia. *Hum Brain Mapp*. 2019:40(15):4487–4507.

Kernbach JM, Yeo BT, Smallwood J, Margulies DS, De Schotten MT, Walter H, Sabuncu M, Holmes A, Gramfort VG, et al. Subspecialization within default mode nodes characterized in 10,000 UK Biobank participants. *Proc Natl Acad Sci*. 2018:115(48): 12295–12300.

Knauer NJ. Homosexuality as contagion: from the well of loneliness to the boy scouts. *Hofstra L Rev*. 2000:29:401.

Liu Y, Si Y, Kim JY, Chen ZF, Rao Y. Molecular regulation of sexual preference revealed by genetic studies of 5-HT in the brains of male mice. *Nature*. 2011:472(7341):95–99.

Manzouri A, Savic I. Cerebral sex dimorphism and sexual orientation. *Hum Brain Mapp*. 2018:39(3):1175–1186.

Manzouri A, Savic I. Possible neurobiological underpinnings of homosexuality and gender dysphoria. *Cereb Cortex*. 2019:25(5): 2084–2101.

Mustanski BS, Chivers ML, Bailey JMA. Critical review of recent biological research on human sexual orientation. *Annu Rev Sex Res*. 2002:13(1):89–140.

Mustanski BS, DuPree MG, Nievergelt CM, Bocklandt S, Schork NJ, Hamer DHA. Genomewide scan of male sexual orientation. *Human Genetic*. 2005:116(4):272–278.

Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*. 2002:15(1):1–25.

Okser S, Pahikkala T, Airola A, Salakoski T, Ripatti S, Aittokallio T. Regularized machine learning in the genetic prediction of complex traits. *PLoS Genet*. 2014:10(11):e1004754.

Paul T, Schiffer B, Zwarg T, Krüger TH, Karama S, Schedlowski M, Forsting M, Gizewski ER. Brain response to visual sexual stimuli in heterosexual and homosexual males. *Hum Brain Mapp*. 2008:29(6): 726–735.

Perry D, Walder K, Hendler T, Shamay T, Soory SG. The gender you are and the gender you like: sexual preference and empathic neural responses. *Brain Res*. 2013:1534:66–75.

Ponseti J, Siebner HR, Klöppel S, Wolff S, Granert O, Jansen O, Hubertus MM, Bosinski HA. Homosexual women have less grey matter in perirhinal cortex than heterosexual women. *PLoS One*. 2007:2(8):e762.

Ponseti J, Granert O, Jansen O, Wolff S, Mehdorn H, Bosinski H, Siebner H. Anatomy/physiology: assessment of sexual orientation using the hemodynamic brain response to visual sexual stimuli. *J Sex Med*. 2009:6(6):1628–1634.

Rahman Q, Wilson GD. Large sexual-orientation-related differences in performance on mental rotation and judgement of line orientation tasks. *Neuropsychology*. 2003:17(1):25.

Rahman Q, Newland C, Smyth BM. Sexual orientation and spatial position effects on selective forms of object location memory. *Brain Cogn*. 2011:75(3):217–224.

Rice G, Anderson C, Risch N, Ebers G. Male homosexuality: absence of linkage to microsatellite markers at Xq28. *Science*. 1999:284(5414): 665–667.

Roselli CE. Neurobiology of gender identity and sexual orientation. *J Neuroendocrinol*. 2018:30(7):e12562.

Safron A, Sylva D, Klimaj V, Rosenthal AM, Li M, Walter M, Bailey JM. Neural correlates of sexual orientation in heterosexual, bisexual, and homosexual men. *Sci Rep*. 2017:7(1):1–15.

Safron A, Klimaj V, Sylva D, Rosenthal AM, Li M, Walter M, Bailey JM. Neural correlates of sexual orientation in heterosexual, bisexual, and homosexual women. *Sci Rep*. 2018:8(1): 1–14.

Sanders A, Martin E, Beecham G, Guo S, Dawood K, Rieger G, Badner JA, Gershon ES, Krishnappa RS, Kolundzija B, et al. Genome-wide scan demonstrates significant linkage for male sexual orientation. *Psychol Med*. 2015:45(7):1379–1388.

Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, Eickhoff SB, Hakonarson H, Guir RC, Gur RE et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*. 2013:64:240–256.

Savic I, Lindstrom P. PET and MRI show differences in cerebral asymmetry and functional connectivity between homo- and heterosexual subjects. *Proc Natl Acad Sci*. 2008:105(27):9403–9408.

Savic I, Garcia-Falgueras A, Swaab DF. Sexual differentiation of the human brain in relation to gender identity and sexual orientation. *Prog Brain Res*. 2010:186:41–62.

Savin-Williams RC, Ream GL. Prevalence and stability of sexual orientation components during adolescence and young adulthood. *Arch Sex Behav*. 2007:36:385–394.

Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo XN, Holmes AJ, Yeo BT. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb Cortex*. 2018:28(9):3095–3114.

Shah NM, Jessel TM, Sanes JR. Sexual differentiation of the nervous system. *Principles of neural science*. 2012:5:1306–1327.

Smith E, Junger J, Pauly K, Kellermann T, Neulen J, Neuschaefer-Rube C, Derntl B, Habel U. Gender incongruence and the brain–Behavioral and neural correlates of voice gender perception in transgender people. *Horm Behav*. 2018:105:11–21.

Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp*. 2002:17(3):143–155.

Swaab DF. Sexual differentiation of the brain and behavior. *Best Pract Res Clin Endocrinol Metab*. 2007:21(3):431–444.

Swaab DF. Sexual orientation and its basis in brain structure and function. *Proc Natl Acad Sci*. 2008:105(30):10273–10274.

Sylva D, Safron A, Rosenthal AM, Reber PJ, Parrish TB, Bailey JM. Neural correlates of sexual arousal in heterosexual and homosexual women and men. *Horm Behav*. 2013:64(4):673–684.

Vasey PL, Parker JL, VanderLaan DP. Comparative reproductive output of androphilic and gynephilic males in Samoa. *Arch Sex Behav*. 2013:43(2):363–367.

Votinov M, Goerlich KS, Puiu AA, Smith E, Nickl-Jockschat T, Derntl B, Habel U. Brain structure changes associated with sexual orientation. *Sci Rep*. 2021:11(1):1–10.

Wegesin DJ. A neuropsychologic profile of homosexual and heterosexual men and women. *Arch Sex Behav*. 1998:27(1):91–108.

Witelson SF, Kigar DL, Scamvougeras A, Kideckel DM, Buck B, Stanchev P, Bronskill M, Black S. Corpus callosum anatomy in right-handed homosexual and heterosexual men. *Arch Sex Behav*. 2008:37(6):857–863.

Wittchen HU, Zaudig M, Fydrich T. *Skid, Strukturiertes klinisches interview für DSM-IV. Achse I und II*. Gottingen, Germany: Handanweisung; 1997.

Woo C-W, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci*. 2017:20(3):365–377.

Wolpert DH. Stacked generalization. *Neural Netw*. 1992:5(2):241–259.

Zeki S, Romaya JP. The brain reaction to viewing faces of opposite- and same-sex romantic partners. *PLoS One*. 2010:5(12):e15802.

Zhang S, Li CS. Functional connectivity mapping of the human precuneus by resting state fMRI. *NeuroImage*. 2012:59:3548–3562.

Zucker KJ, Bradley SJ, Oliver G, Blake J, Fleming S, Hood J. Psychosexual development of women with congenital adrenal hyperplasia. *Horm Behav*. 1996:30(4):300–318.