Reproducibility of R-fMRI Metrics on the Impact of Different Strategies for Multiple Comparison Correction and Small Sample Size

Xiao Chen
陈晓
chenxiao@paych.ac.cn
The R-fMRI Lab, Institute of Psychology, Chinese Academy of Sciences

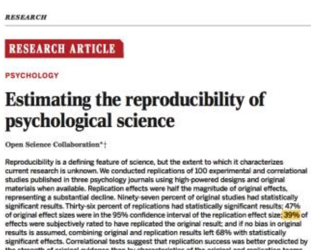
1

Outline

- **Introduction**
- Materials and Methods
- Results
- Discussion

Introduction

"Reproducibility Crisis"


RESEARCH

RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

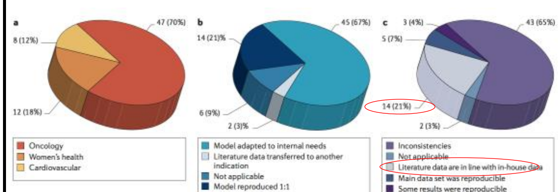
Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

Open Science Collaboration, 2015. Science

3

Introduction

False findings may be the majority majority of published research claims



Analysis of the reproducibility of published data in 67 in - house projects

Prinz et al., 2011. Nat Rev Drug Discov

4

Introduction

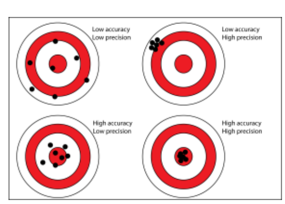
Button et al., 2013. Nat Rev Neurosci

Poldrack, et al., 2017. Nat Rev Neurosci

5

Introduction

Defining reproducibility

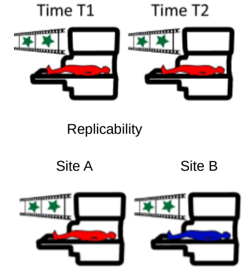


Test-retest reliability

Time T1 Time T2

Replicability

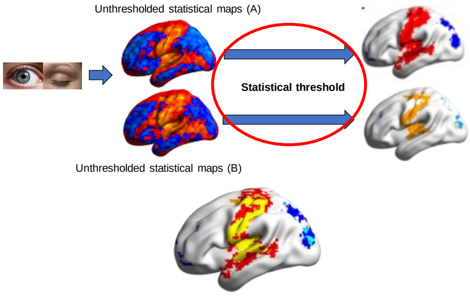
Site A Site B



6

Introduction

Defining reproducibility



7

Introduction

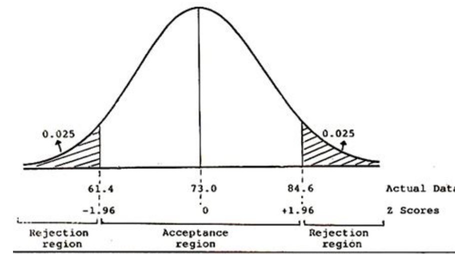
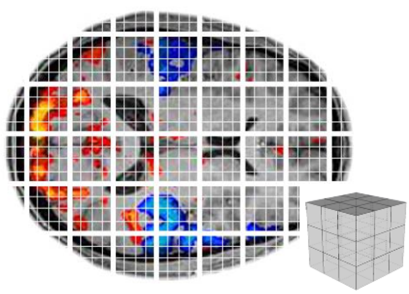


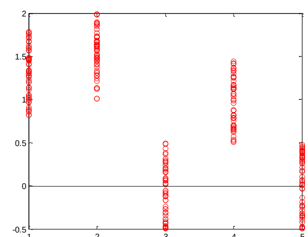
Figure. Critical region for the test of significance.

8

Reproducibility and Multiple Comparison Correction



9



P=0.05 P=0.05 P=0.05 P=0.05 P=0.05

0.95 0.95 0.95 0.95 0.95

Probability of not getting a false positive result:

0.95^5=0.774

1 - 0.05 = 0.95 1 - 0.05 = 0.95 1 - 0.05 = 0.95 1 - 0.05 = 0.95 1 - 0.05 = 0.95

10

Reproducibility and Multiple Comparison Correction

Multiple Comparisons

Bonferroni correction

The Bonferroni correction rejects the null hypothesis for each $p_i \leq \alpha/m$, thereby controlling the FWER at $\leq\alpha$

$$\mathrm{FWER} = P\left\{\bigcup_{i=1}^{m_0}\left(p_i \leq \frac{\alpha}{m}\right)\right\} \leq \sum_{i=1}^{m_0}\left\{P\left(p_i \leq \frac{\alpha}{m}\right)\right\} = m_0\frac{\alpha}{m} \leq m\frac{\alpha}{m} = \alpha.$$

Carlo Emilio Bonferroni

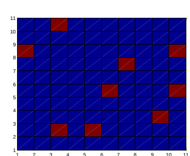
11

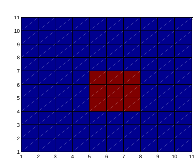
Reproducibility and Multiple Comparison Correction

Multiple Comparisons

Gaussian Random Field Theory Correction

Monte Carlo simulations (AlphaSim)

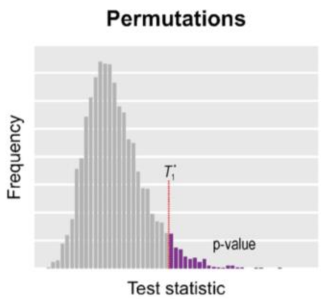




12

Reproducibility and Multiple Comparison Correction

Permutation Test



Ronald Aylmer Fisher

Winkler et al., 2016. Neuroimage

13

Reproducibility and Multiple Comparison Correction

Threshold-Free Cluster Enhancement (TFCE)

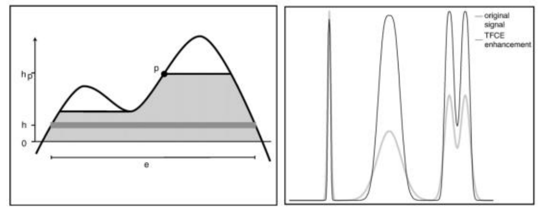


Fig. 1. Illustration of the TFCE approach. Left: the TFCE score at voxel *p* is given by the sum of the scores of all incremental supporting sections (one such is shown as the dark-grey band) within the area of "support" of *p* (light grey). The score for each section is a simple function of its height *h* and extent *e*. Right: example input image and TFCE-enhanced output. The input contains a focal, high signal, a much more spatially extended, lower, signal and a pair of overlapping signals of intermediate extent and height. The TFCE output has the same maximal values for all three cases, and preserves the distinct local maxima in the third case.

Smith et al., 2009. Neuroimage

14

Multiple Comparison Correction

PNAS

CrossMark

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund[a,b,c,1], Thomas E. Nichols[d,e], and Hans Knutsson[a,c]

[a]Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, S-581 85 Linköping, Sweden; [b]Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, S-581 83 Linköping, Sweden; [c]Center for Medical Image Science and Visualization, Linköping University, S-581 83 Linköping, Sweden; [d]Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom; and [e]WMG, University of Warwick, Coventry CV4 7AL, United Kingdom

Edited by Emery N. Brown, Massachusetts General Hospital, Boston, MA, and approved May 17, 2016 (received for review February 12, 2016)

Technology

15 years of brain research has been invalidated by a software bug, say Swedish scientists

- **Up to 70% of fMRI analyses produce at least one false positive, challenging the validity of over 40,000 studies.**

Eklund et al., 2016. PNAS

15

Reproducibility and Multiple Comparison Correction

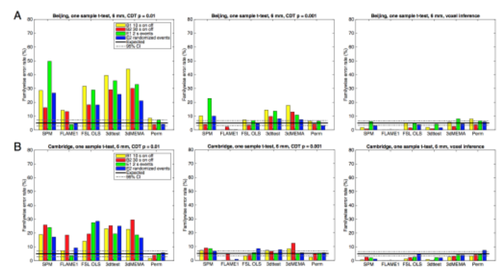


Fig. 1. Results for one-sample *t* test, showing estimated FWE rates for (*A*) Beijing and (*B*) Cambridge data analyzed with 6 mm of smoothing and four different activity paradigms (B1, B2, E1, and E2), for SPM, FSL, AFNI, and a permutation test. These results are for a group size of 20. The estimated FWE rates are simply the number of analyses with any significant group activation divided by the number of analyses (1,000). From *Left* to *Right*: Cluster inference using a cluster-defining threshold (CDT) of $P = 0.01$ and a FWE-corrected threshold of $P = 0.05$, cluster inference using a CDT of $P = 0.001$ and a FWE-corrected threshold of $P = 0.05$, and voxel inference using a FWE-corrected threshold of $P = 0.05$. Note that the default CDT is $P = 0.001$ in SPM and $P = 0.01$ in FSL (AFNI does not have a default setting).

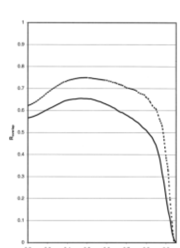
Eklund et al., 2016. PNAS

16

Introduction

Statistical thresholds

Reproducibility is highly sensitive to the statistical threshold used to define significance

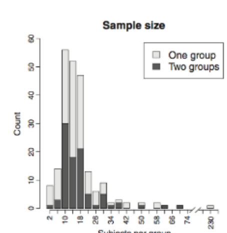


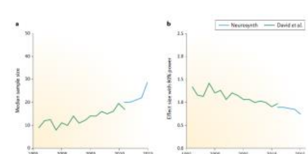
Rombouts et al., 1998

17

Introduction

Small samples in neuroscience





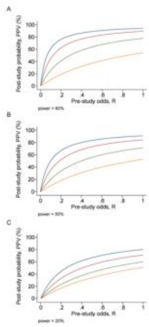
Median sample size: 15 for one group studies and 14.75 per group for two group studies (Carp, 2012)

Poldrack et al., 2017

18

Introduction

Low power studies are unlikely reflecting a true effect



John P. A. Ioannidis

19

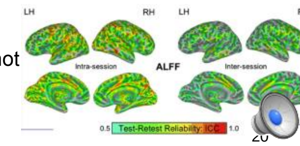
Introduction

Small samples in neuroscience

- The financial issue

- Good test-retest reliability

…However, reliability cannot guarantee replicability



20

Introduction

Summary

- The impact of multiple comparison correction strategy (considering FWER) on reproducibility (test-retest reliability and replicability)
- The impact of sample size on reproducibility (test-retest reliability)

21

Outline

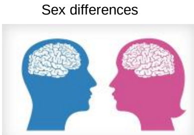
- Introduction
- **Materials and Methods**
- Results
- Discussion

Introduction

Defining reproducibility

We sought to propose a quantitative method to calculate reproducibility of R-fMRI metrics

Sex differences



Eyes open eyes closed (EOEC) differences

23

Materials and Methods

Participants and Imaging Protocols

CORR
CONSORTIUM FOR RELIABILITY AND REPRODUCIBILITY

Consortium for Reliability and Reproducibility (CORR)

1000 Functional Connectomes Project (FCP)

24

Materials and Methods

CORR dataset

Sample size: 420 (212 M vs. 208 F)
Scanned 2 times
Inclusion criteria (from 549):
Age between 18 and 32
No extreme head motion
No poor T1 or functional images, low quality normalization or inadequate brain coverage

1000 Functional Connectomes Project (FCP) dataset

Sample size: 716 (296 M vs. 420 F)
Same inclusion criteria

Beijing EOEC1 dataset

Sample size: 48
Eyes-open vs. eyes-closed
Same Inclusion criteria

Beijing EOEC2 dataset

Sample size: 20
Eyes-open vs. eyes-closed
Same inclusion criteria

Chen, Lu, Yan*, 2017. Human Brain Mapping

Materials and Methods

Preprocessing

1. The first 10 volumes were discarded
2. Slice-timing correction
   shifted to the slice at the mid-point of each TR
3. Realignment
   six-parameter (rigid body) linear transformation
   two-pass procedure
4. Co-registration and segment
   six degree-of-freedom linear transformation without re-sampling
5. Transformation from native space to MNI space
   Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra tool (DARTEL)

26

Materials and Methods

Nuisance Regression

A General Linear Regression Model including: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$

1. Head motion
   Friston 24-parameter model and mean FD
2. Global Signal Regression (GSR)
   Results both with and without GSR were evaluated
3. Other sources of spurious variance
   WM and CSF signals
4. Linear trends

Temporal bandpass filtering (0.01–0.1 Hz)
   All time series except for ALFF and fALFF analyses

27

Materials and Methods

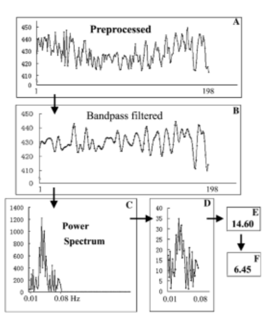
A Broad Array of R-fMRI Metrics

ALFF:
The mean of amplitudes within a specific frequency domain (here, 0.01–0.1Hz) from a fast Fourier transform of a voxel's time course
fALFF:
A normalized version of ALFF and represents the relative contribution of specific oscillations to the whole detectable frequency range
ReHo:
A rank-based Kendall's coefficient of concordance that assesses the synchronization among a given voxel and its nearest neighbors' (here, 26 voxels) time courses
Degree Centrality:
The number or sum of weights of significant connections for a voxel. The weighted sum of positive correlations with a threshold of r>0.25
VMHC:
The functional connectivity between any pair of symmetric inter-hemispheric voxels

28

Computational Methodology

Amplitude of low frequency fluctuations

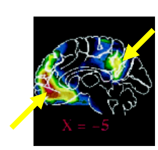


Zang et al., 2007

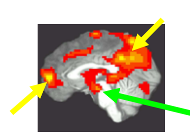
29

Computational Methodology

ALFF



PET
(Raichle et al., 2001)



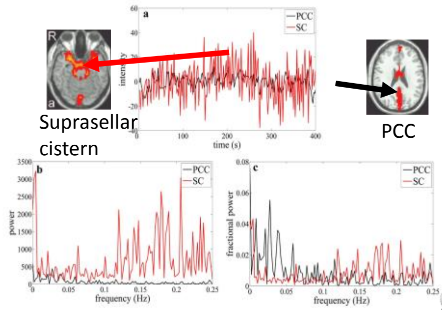
ALFF
(Zang et al., 2007)

noise

30

Computational Methodology

Improvement: fractional ALFF



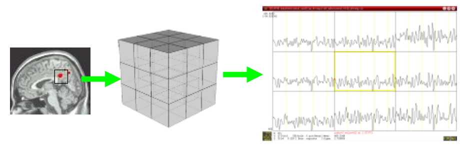
Zou et al., 2008. J Neurosci Methods

31

Computational Methodology

Regional Homogeneity (ReHo)

Similarity or coherence of the time courses within a functional cluster



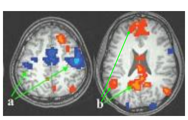
$$W = \frac{\Sigma (R_i)^2 - n(\bar{R})^2}{\frac{1}{12}K^2(n^3 - n)}$$

(Zang et al., 2004)

32

Computational Methodology

ReHo: motor task state vs. pure resting state



Rest > Motor

Motor > Rest

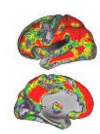
a) Higher ReHo in bilateral primary motor cortices during motor task
b) Higher ReHo in default mode network (PCC, MPFC, IPL) during rest *(Raichle et al., 2001; Greicius et al., 2003)*

(Zang et al., 2004)

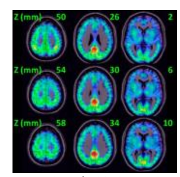
33

Computational Methodology

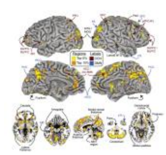
Degree centrality



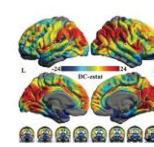
Buckner et al., 2009. J Neurosci



Tomasi et al., 2010. PNAS



Cole et al., 2010. Neuroimage

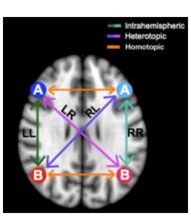


Zuo et al., 2011. Cereb Cortex

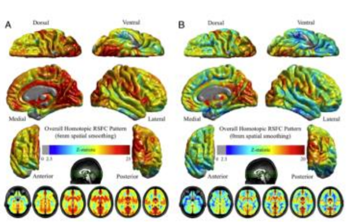
34

Computational Methodology

- **Voxel-mirrored homotopic connectivity (VMHC)**



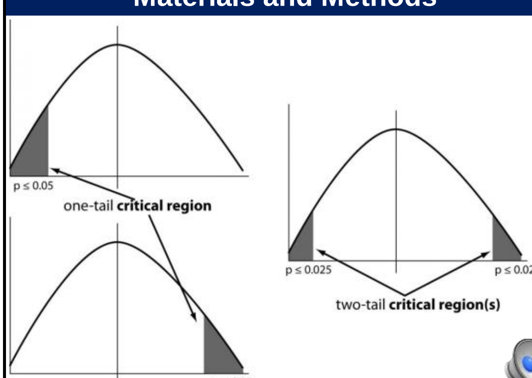
Gee et al., 2011



Zuo et al., 2010

35

Materials and Methods



Materials and Methods

Evaluating FWER of Different Strategies to Correct for Multiple Comparisons

FCP dataset: Beijing site

106 females

40 females

20 females ⟷ 20 females

Multiple Comparison Correction

Permutation, repeated for 1000 times

Significant differences? False Positive!

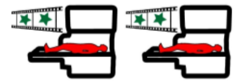
37

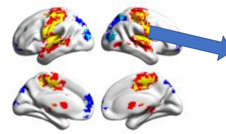
Materials and Methods

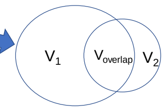
Test-retest reliability

Sex differences in test and retest

Time T1 Time T2







Statistical significant voxels

V_1: significant voxels in test

V_2: significant voxels in retest

$V_{overlap}$: voxels significant in both test and retest

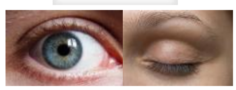
$$Dice = \frac{2 \times V_{overlap}}{V_1 + V_2}$$

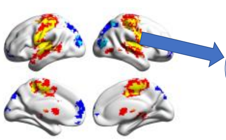
38

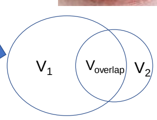
Materials and Methods

Replicability

Site A Site B







Statistical significant voxels

V_1: significant voxels in site A

V_2: significant voxels in site B

$V_{overlap}$: voxels significant in both site A and site B

$$Dice = \frac{2 \times V_{overlap}}{V_1 + V_2}$$

39

Materials and Methods

Influences of Sample Size on Test–Retest Reliability, Sensitivity and PPV

"SWU 4" site in the CORR dataset

k/2 116 males

k/2 105 females

randomized 100 times

randomized 100 times

Session 1 ⟷ Session 1

Test statistical map

Test-retest reliability (Dice coefficient)

Retest Statistical map

Session 2 ⟷ Session 2

$k \in \{30,40,50,60,70,80,90,100,120,140,160,180,200\}$

40

Materials and Methods

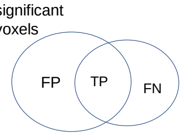
Influences of Sample Size on Test–Retest Reliability, Sensitivity and PPV

CORR session 1 ⟷ CORR session 2 ⟷ FCP dataset

PT with TFCE

"Gold standard"

Statistical significant voxels



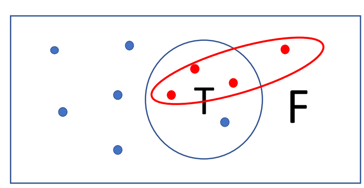
Gold standard

$$Sensitivity = \frac{TP}{TP + FN}$$

$$PPV = \frac{TP}{TP + FP}$$

41

Introduction



Positive Predictive Value, PPV

After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true

PPV = (1 − β)R/(R − βR + α)

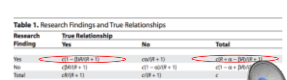


Table 1. Research Findings and True Relationships

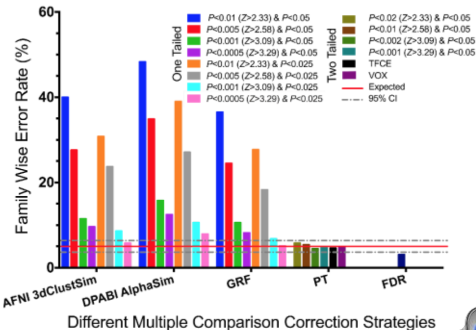
Research Finding	True Relationship Yes	No	Total
Yes	c(1 − β)R/(R + 1)	cα/(R + 1)	c(R + α − βR)/(R + 1)
No	cβR/(R + 1)	c(1 − α)/(R + 1)	c(1 − α + βR)/
Total	cR/(R + 1)	c/(R + 1)	c

DOI: 10.1371/journal.pmed.0020124.t001

Outline

- Introduction
- Materials and Methods
- **Results**
- Discussion

Results



FWER calculated with ALFF without GSR (4mm FWHM smoothing kernel)

44

Results

(One-tailed twice)		AFNI 3dClustSim		DPABI AlphaSim		GRF	
Voxel threshold	Cluster threshold	FWER	Cluster size	FWER	Cluster size	FWER	Cluster size
$P < 0.01$ ($Z > 2.33$)	$P < 0.05$	40.0%	66.05 ± 0.73	48.3%	60.24 ± 1.68	36.5%	69.35 ± 1.09
$P < 0.005$ ($Z > 2.58$)	$P < 0.05$	27.6%	43.59 ± 0.42	34.9%	39.45 ± 1.13	24.5%	46.70 ± 0.75
$P < 0.001$ ($Z > 3.09$)	$P < 0.05$	11.5%	19.98 ± 0.34	15.8%	18.40 ± 0.61	10.6%	21.29 ± 0.46
$P < 0.0005$ ($Z > 3.29$)	$P < 0.05$	9.6%	14.53 ± 0.25	12.5%	13.93 ± 0.54	8.2%	15.82 ± 0.39
$P < 0.01$ ($Z > 2.33$)	$P < 0.025$	30.8%	74.50 ± 1.14	39.0%	67.72 ± 2.36	27.7%	78.96 ± 1.24
$P < 0.005$ ($Z > 2.58$)	$P < 0.025$	23.7%	47.01 ± 0.59	27.1%	44.48 ± 1.60	18.3%	53.48 ± 0.85
$P < 0.001$ ($Z > 3.09$)	$P < 0.025$	8.6%	22.63 ± 0.25	10.6%	21.00 ± 0.87	6.8%	24.94 ± 0.41
$P < 0.0005$ ($Z > 3.29$)	$P < 0.025$	5.8%	17.33 ± 0.22	7.9%	16.03 ± 0.71	5.1%	18.51 ± 0.50

	Voxel Threshold	Cluster Threshold	Family Wise Error Rate										
			ALFF	fALFF	ReHo	DC	VMHC	ALFF with GSR	fALFF with GSR	ReHo with GSR	DC with GSR	VMHC with GSR	ALFF (8mm smoothed)
Smoothness (mm, FWHM)			7.94×7.31×6.86	7.34×7.42×7.20	9.36×8.72×8.39	7.86×7.97×7.81	6.31×6.87×6.61	7.99×7.31×6.84	7.32×7.41×7.19	9.24×8.56×8.18	8.06×8.16×8.09	6.11×6.61×6.37	11.88×11.53×11.68
AFNI 3dClustSim (One Tailed)	$P < 0.0005$ ($Z > 3.29$)	$P < 0.025$	5.8%	6.1%	7.3%	8.5%	6.0%	5.3%	6.6%	6.9%	6.8%	6.4%	5.5%
DPABI AlphaSim (One Tailed)			7.9%	8.3%	8.5%	10.2%	9.0%	7.8%	7.7%	7.8%	8.3%	9.6%	6.9%
Gaussian Random Field (One Tailed)			5.1%	5.5%	4.9%	7.4%	5.2%	4.8%	5.9%	5.3%	5.1%	6.4%	4.4%
PT Cluster Extent Correction (Two Tailed)	$P < 0.02$ ($Z > 2.33$)	$P < 0.05$	5.8%	3.6%	5.8%	4.6%	5.2%	4.8%	3.9%	5.5%	5.2%	4.3%	5.3%
	$P < 0.01$ ($Z > 2.58$)	$P < 0.05$	5.4%	4.0%	5.7%	4.6%	5.5%	5.3%	3.8%	5.3%	5.0%	4.5%	5.4%
	$P < 0.002$ ($Z > 3.09$)	$P < 0.05$	4.5%	4.1%	5.3%	4.8%	4.2%	4.5%	5.0%	5.1%	4.7%	4.3%	4.4%
	$P < 0.001$ ($Z > 3.29$)	$P < 0.05$	4.8%	4.5%	4.5%	4.9%	3.4%	4.3%	4.8%	5.4%	4.2%	3.9%	4.1%
PT Threshold-Free Cluster Enhancement (TFCE)			4.6%	3.9%	5.7%	5.0%	4.3%	5.3%	4.2%	5.5%	4.7%	4.8%	4.6%
PT Voxel-Wise Correction (VOX)			4.9%	4.9%	5.7%	3.9%	4.7%	6.0%	4.5%	5.6%	4.0%	4.6%	3.9%
FDR Correction			3.1%	3.4%	4.4%	2.4%	3.9%	4.1%	2.8%	3.6%	2.4%	3.5%	1.6%

45

Results

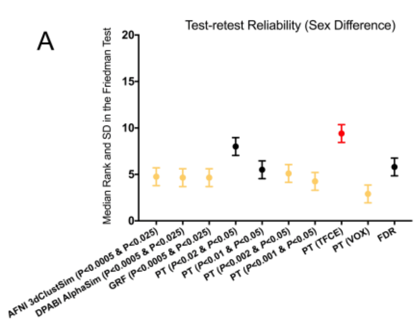
Test-retest reliability of between-subject sex difference

	Voxel threshold	Cluster threshold	Test-retest reliability (dice coefficient)									
			ALFF	fALFF	ReHo	DC	VMHC	ALFF with GSR	fALFF with GSR	ReHo with GSR	DC with GSR	VMHC with GSR
AFNI 3dClustSim (one-tailed)	$P < 0.0005$ ($Z > 3.29$)	$P < 0.025$	0.65	0.51	0.50	0.34	0.39	0.64	0.48	0.44	0.28	0.24
DPABI AlphaSim (one-tailed)			0.65	0.51	0.49	0.34	0.39	0.64	0.48	0.45	0.27	0.27
GRF (one-tailed)			0.64	0.51	0.50	0.35	0.39	0.65	0.48	0.43	0.28	0.24
PT cluster extent correction (two-tailed)	$P < 0.02$ ($Z > 2.33$)	$P < 0.05$	0.65	0.70	0.56	0.45	0.40	0.62	0.68	0.45	0.30	0.40
	$P < 0.01$ ($Z > 2.58$)	$P < 0.05$	0.67	0.66	0.52	0.32	0.33	0.60	0.63	0.46	0.27	0.32
	$P < 0.002$ ($Z > 3.09$)	$P < 0.05$	0.63	0.55	0.51	0.36	0.38	0.63	0.52	0.47	0.23	0.32
	$P < 0.001$ ($Z > 3.29$)	$P < 0.05$	0.64	0.51	0.48	0.37	0.38	0.64	0.48	0.44	0.28	0.26
PT TFCE			0.68	0.75	0.54	0.48	0.44	0.66	0.74	0.44	0.31	0.42
PT VOX			0.66	0.34	0.48	0.37	0.22	0.65	0.31	0.38	0.11	0.14
FDR correction			0.64	0.67	0.54	0.39	0.37	0.63	0.64	0.47	0.23	0.29

- ◆ Moderate test-retest reliability
- ◆ ALFF, fALFF, ReHo are better than DC and VMHC

46

Test-retest Reliability



Chen, Lu, Yan[*], 2017. Human Brain Mapping

212 M vs. 208 F × 2 times

47

Results

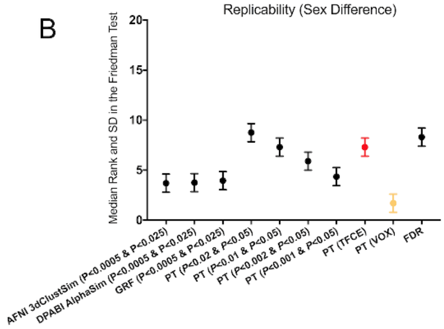
Replicability of between-subject sex difference

	Voxel threshold	Cluster threshold	Replicability (dice coefficient)									
			ALFF	fALFF	ReHo	DC	VMHC	ALFF with GSR	fALFF with GSR	ReHo with GSR	DC with GSR	VMHC with GSR
AFNI 3dClustSim (one-tailed)	$P < 0.0005$ ($Z > 3.29$)	$P < 0.025$	0.12	0.10	0.07	0.07	0.01	0.10	0.11	0.02	0.08	0.02
DPABI AlphaSim (one-tailed)			0.13	0.09	0.07	0.07	0.02	0.10	0.11	0.02	0.08	0.02
GRF (one-tailed)			0.13	0.10	0.07	0.07	0.01	0.10	0.11	0.02	0.08	0.02
PT cluster extent correction (two-tailed)	$P < 0.02$ ($Z > 2.33$)	$P < 0.05$	0.21	0.13	0.14	0.17	0.05	0.21	0.06	0.12	0.22	0.10
	$P < 0.01$ ($Z > 2.58$)	$P < 0.05$	0.19	0.11	0.11	0.16	0.02	0.17	0.09	0.08	0.24	0.08
	$P < 0.002$ ($Z > 3.09$)	$P < 0.05$	0.14	0.10	0.08	0.11	0.02	0.12	0.10	0.03	0.05	0.03
	$P < 0.001$ ($Z > 3.29$)	$P < 0.05$	0.12	0.10	0.07	0.07	0.01	0.10	0.11	0.02	0.08	0.02
PT TFCE			0.25	0.06	0.13	0.20	0.01	0.25	0.03	0.09	0.26	0.02
PT VOX			0.02	0.00	0.01	0.00	0.00	0.01	0.05	0.00	0.00	0.00
FDR correction			0.15	0.06	0.11	0.09	0.02	0.13	0.04	0.05	0.08	0.00

- ◆ Poor replicability

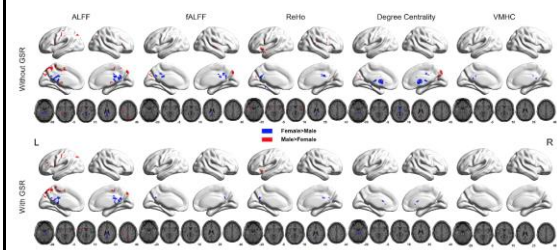
48

Results



49

Results



◆ Female's PCC demonstrate more spontaneous activity than male

50

Results

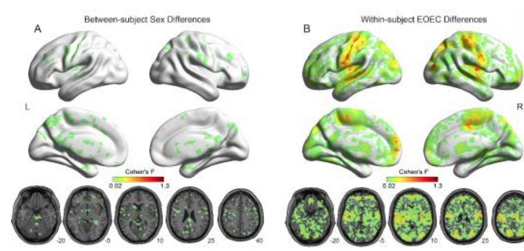
Replicability of within-subject EOEC difference

	Voxel threshold	Cluster threshold	Replicability (dice coefficient) ALFF	fALFF	ReHo	DC	VMHC	ALFF with GSR	fALFF with GSR	ReHo with GSR	DC with GSR	VMHC with GSR
AFNI 3dClustSim (one-tailed)	$P < 0.0005$ ($Z > 3.29$)	$P < 0.025$	0.15	0.11	0.26	0.03	0.10	0.14	0.11	0.31	0.07	0.10
DPABI AlphaSim (one-tailed)			0.15	0.11	0.26	0.03	0.10	0.14	0.11	0.31	0.07	0.09
GRF (one-tailed)			0.15	0.11	0.27	0.04	0.10	0.14	0.11	0.30	0.05	0.10
PT cluster extent correction (two-tailed)	$P < 0.02$ ($Z > 2.33$)	$P < 0.05$	0.46	0.27	0.44	0.24	0.21	0.41	0.30	0.49	0.28	0.17
	$P < 0.01$ ($Z > 2.58$)	$P < 0.05$	0.39	0.24	0.40	0.20	0.16	0.35	0.21	0.48	0.18	0.21
	$P < 0.002$ ($Z > 3.09$)	$P < 0.05$	0.22	0.16	0.32	0.06	0.14	0.19	0.16	0.35	0.09	0.12
	$P < 0.001$ ($Z > 3.29$)	$P < 0.05$	0.15	0.11	0.27	0.04	0.10	0.14	0.11	0.30	0.05	0.09
PT TFCE)			0.49	0.31	0.45	0.29	0.20	0.46	0.32	0.47	0.30	0.20
PT VOX			0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00
FDR Correction			0.09	0.00	0.29	0.03	0.08	0.12	0.00	0.34	0.12	0.10

◆ Higher than between-subject sex difference but still not moderate

51

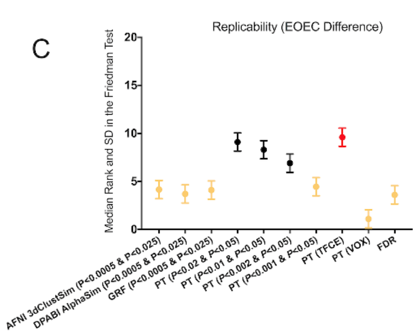
Results



Within-subject design has larger effect size

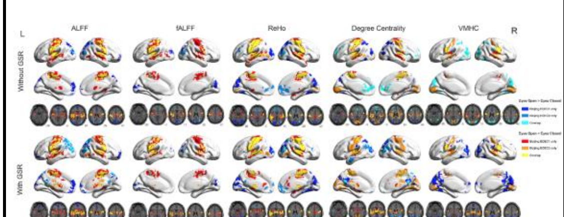
52

Results



53

Results



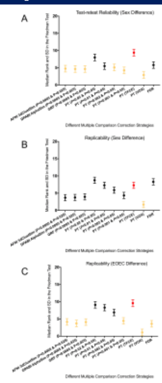
◆ Eyes open > Eyes closed in bilateral occipital cortices

◆ Eyes open < Eyes closed in bilateral pre- and post-central gyrus

54

PT with TFCE outperforms

Permutation test TFCE, a strict multiple comparison correction strategy, reached the best balance between family-wise error rate (under 5%) and test-retest reliability / replicability

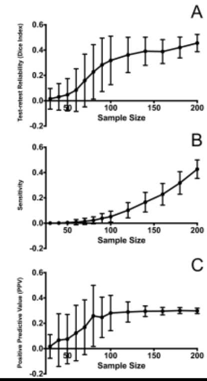


Chen, Lu, Yan*, 2017. Human Brain Mapping

55

Sample Size Matters

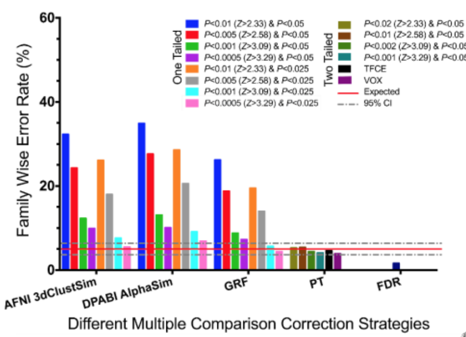
Randomly draw k subjects from the "SWU 4" site in the CORR dataset, which has two sessions of 116 males and 105 females



Chen, Lu, Yan*, 2017. Human Brain Mapping

56

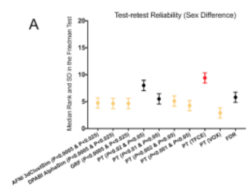
Results



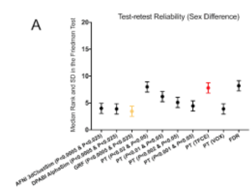
FWER calculated with ALFF without GSR (8mm FWHM smoothing kernel)

57

Results



4mm smoothing

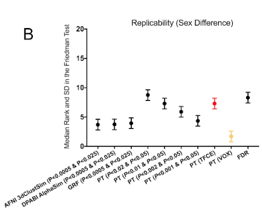


8mm smoothing

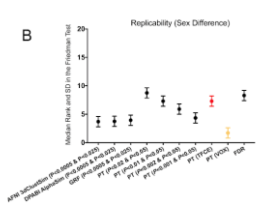
Friedman test of 10 different correction strategies on test-retest reliability regarding sex difference

58

Results



4mm smoothing

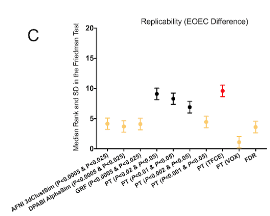


8mm smoothing

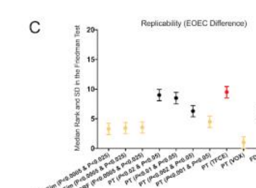
Friedman test of 10 different correction strategies on replicability regarding sex difference

59

Results



4mm smoothing

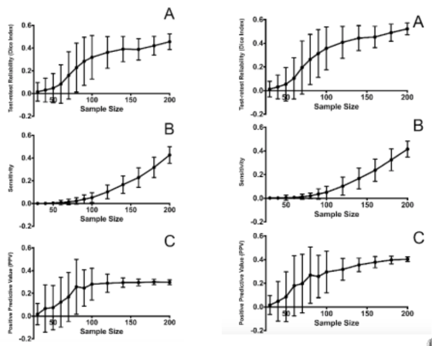


8mm smoothing

Friedman test of 10 different correction strategies on replicability regarding EOEC difference

60

Results



4mm smoothing 8mm smoothing

61

Outline

- Introduction
- Materials and Methods
- Results
- **Discussion**

Discussion

Main findings:

- Liberal correction strategies yield unacceptable high FWERs
- PT with TFCE reach the best balance between FWER and reproducibility
- Between-subject design has moderate test-retest reliability but poor replicability
- Within-subject design has better replicability but still not moderate
- Larger sample size increases reproducibility, sensitivity as well as PPV

63

Discussion

What correction strategy can be used?

According to FWER…

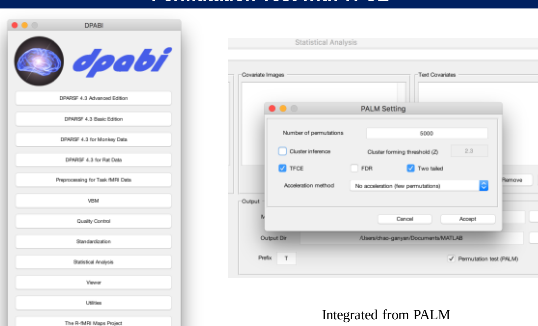
- GRF correction with strict p values (voxel wise P<0.0005 and cluster wise P<0.025 for each tail)
- Four kinds of PT with extent thresholding
- PT with TFCE
- PT with VOX
- FDR correction

According to reproducibility…

Strict strategies cannot achieve moderate reproducibility, except PT with TFCE

64

Permutation Test with TFCE



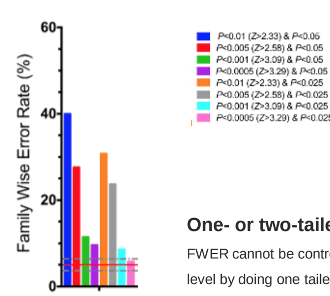
Integrated from PALM
(Winkler et al. 2016. Neuroimage)

Yan* et al., 2016. Neuroinformatics

ESI Top 1% highly cited (>60 times)

65

Discussion



One- or two-tailed?

FWER cannot be controlled to the norminal level by doing one tailed correction twice

66

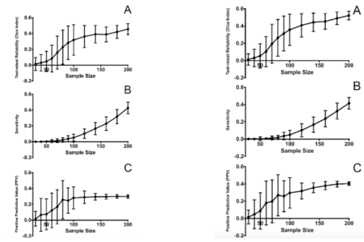
Discussion

Sample size (k)	Test–retest reliability (dice index)	Sensitivity	PPV
30	0.02 ± 0.08	0.001 ± 0.004	0.02 ± 0.09
40	0.03 ± 0.11	0.001 ± 0.01	0.07 ± 0.21
50	0.05 ± 0.13	0.004 ± 0.01	0.07 ± 0.19
60	0.08 ± 0.17	0.01 ± 0.02	0.12 ± 0.22
70	0.16 ± 0.21	0.01 ± 0.02	0.17 ± 0.22
80	0.23 ± 0.22	0.02 ± 0.03	0.26 ± 0.24
90	0.28 ± 0.21	0.04 ± 0.04	0.25 ± 0.16
100	0.32 ± 0.19	0.05 ± 0.04	0.28 ± 0.14
120	0.36 ± 0.14	0.10 ± 0.06	0.29 ± 0.08
140	0.39 ± 0.11	0.17 ± 0.08	0.29 ± 0.04
160	0.39 ± 0.09	0.23 ± 0.09	0.30 ± 0.03
180	0.42 ± 0.08	0.32 ± 0.09	0.30 ± 0.02
200	0.46 ± 0.07	0.43 ± 0.07	0.30 ± 0.02

Results from a sample size <80 (40 per group) should be considered preliminary, given their low reliability (< 0.23), sensitivity (< 0.02) and PPV (< 0.26)

67

Discussion



4mm smoothing   8mm smoothing

Larger smoothing kernel (8 mm) improves reproducibility and is more likely reflecting true effect

68

Discussion

Take-home message:

- PT with TFCE reach the best balance between FWER and reproducibility
- Within-subject design has better replicability and larger effect size
- Larger sample size (>80, 40 in each group) increases reproducibility, sensitivity as well as PPV

69

Discussion

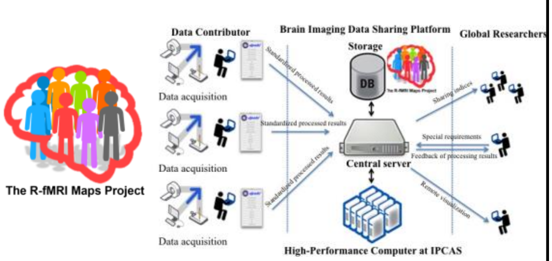
All statistical maps have been shared through the R-fMRI Maps project (http://rfmri.org/maps)

Key source code have been shared through (https://github.com/Chaogan-Yan/PaperScripts/tree/master/Chen_2017_HBM)

Thus our findings could be easily reproduced by any researchers

70

The R-fMRI Maps Project



Part of the Human Brain Data Sharing Initiative (HBDSI), IPCAS

71

The R-fMRI Maps Project

Shared data of 4770 subjects:

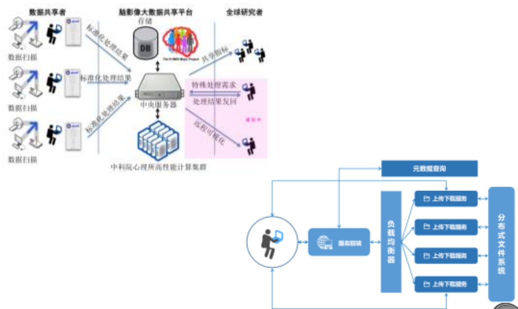
1. Amplitude of low frequency fluctuations (ALFF)
2. Fractional ALFF (fALFF)
3. Regional Homogeneity (ReHo)
4. Voxel-mirrored homotopic connectivity (VMHC)
5. Degree Centrality (DC)
6. Functional Connectivity Matrices
   a. Automated Anatomical Labeling (AAL) atlas
   b. Harvard-Oxford atlas
   c. Craddock's clustering 200 ROIs
   d. Zalesky's random parcellations
   e. Dosenbach's 160 functional ROIs

In addition, gray matter, white matter and CSF density and volume files were shared

Downloaded by 593 researchers

72

The R-fMRI Maps Project

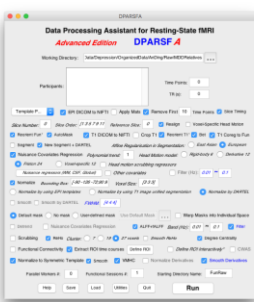


73

REST-meta-MDD

74

DPARSF



Yan and Zang, 2010.
Front Syst Neurosci.
Cited: > 1200 times

Haris I. Sair
Johns Hopkins
University

COMMUNICATION

Reporting of Resting-State Functional Magnetic Resonance Imaging Preprocessing Methodologies

Twelve different software packages were used in the 100 studies. Many articles utilized the use of multiple software for analysis. The most commonly used software was SPM (...), followed by DPARSF (29%) and FSL (25%). Other les... monly used software included AFNI and various M... toolboxes, such as the GIFT toolbox and the Conn to...

DPARSF Citations

Year	First Author	Journal
2016	Wei Cheng	Brain
2016	Clifford I. Workman	Neuropsychopharmacology
2016	Yamamura, T	Translational Psychiatry
2016	Hermesdorf, M.	Human Brain Mapping
2016	Haichen Yang	Journal of Affective Disorders
2016	Shuang Zhang	BMC psychiatry
2016	Lian Du	The Journal of ECT
2016	Chienhan Lai	Brain Imaging and Behavior
2016	Lijuan Wang	Journal of International Medical Research
2016	Zhenghua Hou	Frontiers in aging neuroscience
2016	Dongsheng Lv	Journal of Affective Disorders
2016	Yunhui Chen	Data in Brief
2016	Xue Du	PsychiatryResearch:Neuroimaging
2016	Workman C.	Psychological Medicine
2016	Coutinho, J. F.	Brain Imaging and Behavior
2016	Borchardt,V	Human Brain Mapping
2016	Zhan Feng	Progress in Neuro-Psychopharmacology
2016	Ho N. S. P	Journal of Affective Disorders
2016	Chuanchih Yang	Neural Plasticity
2016	Li Wang	Metabolic Brain Disorder
2016	Li Wang	Neuroscience Letters
2016	Song, Z	Translational Psychiatry
2016	Zhongwei Guo	Experimental Gerontology
2016	Lian Du	Scientific reports
2016	Liliana Ramona Dem	European Archives of Psychiatry and Clin
2016	Zhenghua Hou	PsychiatryResearch:Neuroimaging
2015	Yu Chen	Journal of Affective Disorders
2015	Xiao Yang	Neuropsychiatric Disease and Treatment
2015	Wenbin Guo	Australian & New Zealand Journal of Psy
2015	Chienhan Lai	Journal of Affective Disorders
2015	Mingli Li;ChaohuaHu	Journal of Affective Disorders
2015	Chunhong Liu	Behavioural Brain Research
2015	Ying Wang	Acta Psychiatrica Scandinavica
2015	Li Wang, Mingrui Xia	Human Brain Mapping
2015	Wenbin Guo	Journal of Affective Disorders
2015	Wenbin Guo	Journal of Affective Disorders
2015	Liu Chonghun	Brain research
2015	Hongna Zheng	BioMed Research International
2015	Yinfeng Zhang	Complementary Therapies in Medicine
2015	Yingying Yue	BioMed Research Interna
2015	Yingying Yin	Journal of Affective Diso
2015	Yingying Yin	Journal of Neural Trans
2015	Jinfeng Wu	Frontiers in human ne
2015	Ying Wang	Bipolar Disorders
2015	Jakub Späti	Neuropsychopharmacology
2015	Desmond J. Oathes	Biological psychiatry

> 1200 times

108 depression studies were identified

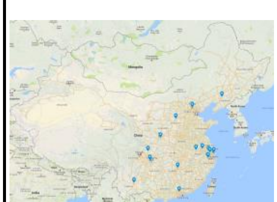
Supported by 33 PIs from 24 Hospitals

REST-meta-MDD Conference

赵靖平 中南大学湘雅二医院
谢光荣 中南大学湘雅二医院
刘哲宁 中南大学湘雅二医院
王传跃 安定医院
陈炜 浙江大学邵逸夫医院
司天梅 北京大学第六医院
蒙华庆 重庆医科大学
臧玉峰 杭州师范大学
傅小兰 中科院心理所
刘勋 中科院心理所
……

77

REST-meta-MDD



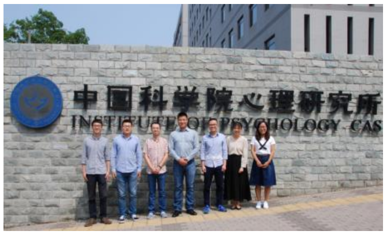
北京大学第六医院、北京安定医院、中南大学湘雅二医院、四川大学华西医院、东南大学附属中大医院、浙江大学医学院附属第一医院与附属邵逸夫医院、南京脑科医院、苏州广济医院、暨南大学附属第一医院、中国医学科学研究院、上海精神卫生中心、无锡市精神卫生中心、中国医学科学院、中国医科大学、重庆医科大学、安徽医科大学、山西医科大学、昆明医学院、西安交通大学、西南大学、复旦大学、杭州师范大学……

Accumulated R-fMRI Maps of 1300 MDD data and 1128 healthy controls

The biggest in the world

78

Acknowledgements



Funding
- National Natural Science Foundation of China
- National Key R&D Program of China
- Chinese Academy of Sciences

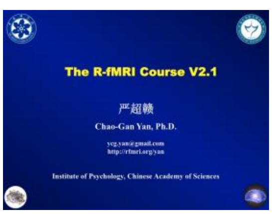
NYU Child Study Center
F. Xavier Castellanos

79

Thank you for your attention!

80

Further Help



http://rfmri.org/Course

http://rfmri.org/wiki

The R-fMRI Journal Club

Official Account: RFMRILab

81

Preprints of the R-fMRI Network

Preprints of the R-fMRI Network (PRN) is a preprint, open-access, free-submission, open-discussion, community funded Preprints of R-fMRI related research. The goal of PRN is to supplement the peer reviewed journal publication system – by more rapidly communicating the latest research achievements across the global.

F1000Research — F1000Research 2015, 3:313 Last updated: 21 AUG 2015

SOFTWARE TOOL ARTICLE

REVISED **PRN: a preprint service for catalyzing R-fMRI and neuroscience related studies [v2; ref status: indexed, http://f1000r.es/5qy]**

Chao-gan Yan[1-4], Qingyang Li[4], Lei Gao[4,5]

[1]The Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA
[2]Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Rd, Chaoyang District, Beijing, 100101, China
[3]Department of Child and Adolescent Psychiatry, New York University Langone Medical Center, New York, NY, USA
[4]Editorial Office of PRN, the R-fMRI Network, Inc., New York, NY, USA
[5]Department of Radiology, the First Affiliated Hospital of Nanchang University, Nanchang, China

ChinaXiv 中国科学院科技论文预发布平台

82

数据分析与深度培训

DEEP BRAIN

静息态功能磁共振成像深度数据分析

功能磁共振成像越来越成为一种主流的科研手段，然而功能磁共振的数据分析却是一项具有高度挑战性的工作。海量的原始数据，繁多的分析步骤，复杂的分析方法都让研究者们无所适从。恰当的分析方法可以从普通的数据中挖掘出富有创新性的结果，而不适当的分析则可能让精心收集的数据黯然失色。深度大脑公司联合中国科学院 The R-fMRI Lab 的专业脑功能成像研究团队推出一站式功能磁共振数据分析解决方案，助您从容应对功能磁共振数据带来的挑战。

http://deepbrain.com

DEEP BRAIN

静息态功能磁共振成像数据处理深度特训

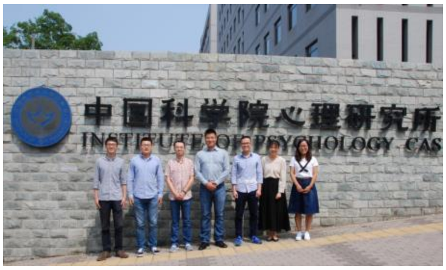
从您见到这条消息开始，您便将有机会与中国科学院 The R-fMRI Lab 的静息态功能磁共振专家团队共同探索大脑的奥秘！深度跟组特训期间，您将会亲身体验：

- 数据处理　专家指导下高效学习静息态功能磁共振成像数据处理
- 思路设计　与国际知名专家讨论形成研究思路
- 论文撰写　系统的 SCI 论文写作训练

The R-fMRI Lab

WeChat Official Account: RFMRILab

Acknowledgments



Chinese Academy of Sciences
Xi-Nian Zuo
Hangzhou Normal University
Yu-Feng Zang
NYU Child Study Cente
F. Xavier Castellanos

Child Mind Institute
Michael P. Milham

Funding
- National Natural Science Foundation of China
- National Key R&D Program of China
- Chinese Academy of Sciences

85

Thanks for your attention!

86