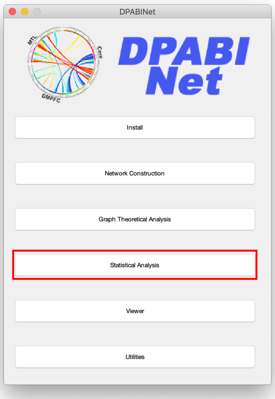
DPABINet: Statistical Analysis and Results Viewing

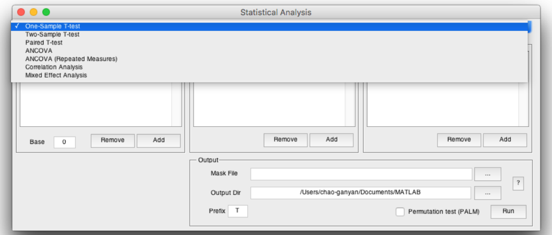
Chao-Gan YAN, Ph.D.
严超赣
ycg.yan@gmail.com
http://rfmri.org
Institute of Psychology, Chinese Academy of Sciences

Outline

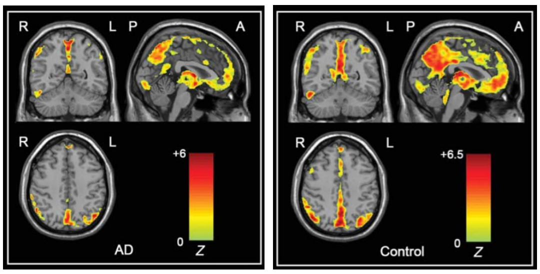
- Statistical Analysis
- Results Viewing



Statistical Analysis

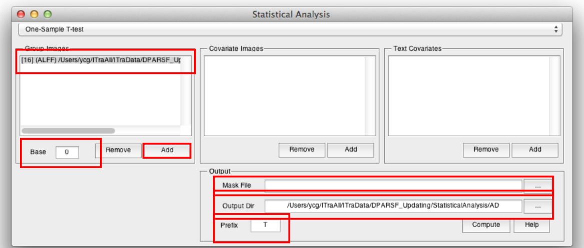


One-Sample T-Test



Wang#, Yan# et al., 2011, Hum Brain Mapp

One-Sample T-Test



T Statistic Matrix

No Mask Needed!

Two-Sample T-Test

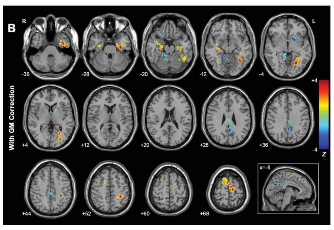


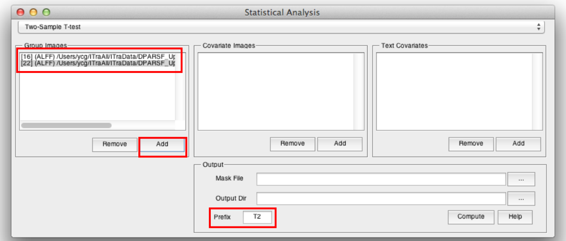
Figure 3.

A: *Z*-statistical difference maps between the AD patients and healthy eldely (without GM correction). The AD patients showed significantly decreased ALFF in the bilateral PCC/PCu, left LN, left OFG, and right ALC, and increased ALFF in the bilateral PHG, bilateral Hip, bilateral SFG, bilateral SMA, left FG, left PoCG, left ITG, and left STG. For the details of the regions, see Table III. **B**: *Z*-statistical difference maps between the AD patients and healthy elderly (with GM correction). The AD patients showed significantly decreased ALFF in the bilateral PCC/PCu, left LN, and right ALC, and increased ALFF in the bilateral PHG, bilateral Hip, bilateral SFG, bilateral SMA, left FG, left PoCG, left ITG, and left STG. For the details of the regions, see Table III. The statistical threshold was set at $|Z| > 1.96$ ($P < 0.05$) and cluster size $>1,404\ mm^3$, which corresponded to a corrected $P < 0.05$. Of note, we showed the two-sample *t*-tests results within a mask showing significant group differences in the ANOVA analysis (Fig. 2A). R, right; L, left. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Wang#, Yan# et al., 2011, Hum Brain Mapp

7

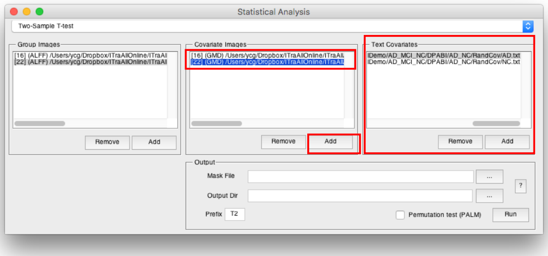
Two-Sample T-Test



T Statistic Matrix: positive corresponds to the mean of Group 1 is greater than the mean of Group 2

8

Two-Sample T-Test



Two-Sample T-Test with covariates: e.g. other matrix

Please make sure the correspondence between the group matrices and the covariate matrices: order

Other covariate can be also specified as text files.

(E.g. head motion (mean FD), age, sex etc.)

9

Paired T-Test

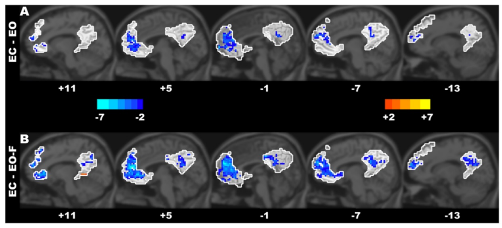
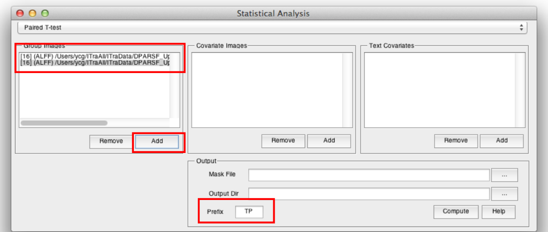


Figure 3. The between-condition differences of the ALFF within the DMN. The ALFF differences were found between the EC and EO conditions (A), and between the EC and EO-F conditions (B). The areas in the white contours denote the ROIs within the DMN. The numbers below the images refer to the x coordinates in the Talairach and Tournoux space. The statistical threshold was set at $|t|>2.093$ ($P<0.05$) and cluster size $>486\ mm^3$, which corresponds to a corrected $P<0.05$.
doi:10.1371/journal.pone.0005743.g003

Yan et al., 2009. PLoS ONE

10

Paired T-Test



T Statistic Matrix

Condition 1 – Condition 2

Please make sure the correspondence

11

ANOVA or ANCOVA

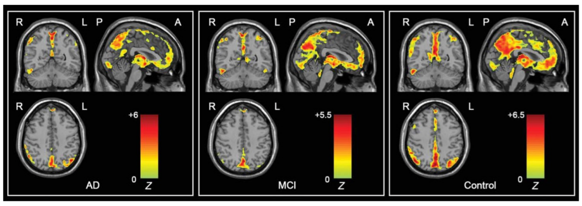


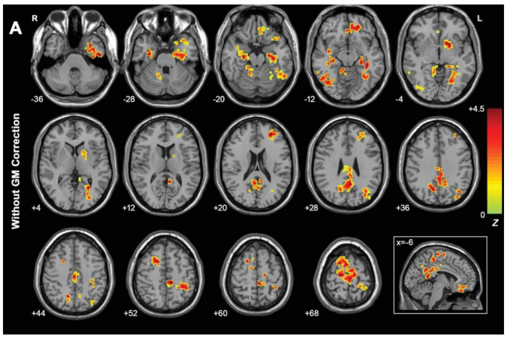
Figure 1.

Within-group ALFF maps within the AD, MCI, and healthy elderly control groups. Visual inspection indicated that the PCC and adjacent PCu had the highest ALFF values within each group and had different strengths among the three groups. The statistical threshold was set at $Z > 3.09$ ($P < 0.001$) and cluster size $>189\ mm^3$, which corresponded to a corrected $P < 0.001$. R, right; L, left; P, posterior; A, anterior. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Wang#, Yan# et al., 2011, Hum Brain Mapp

12

ANOVA or ANCOVA

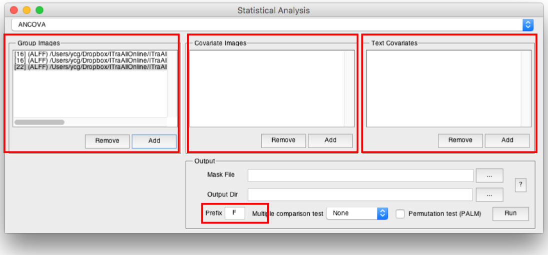


Wang[#], Yan[#] et al., 2011, Hum Brain Mapp

13

13

ANOVA or ANCOVA



ANCOVA: e.g. other matrices

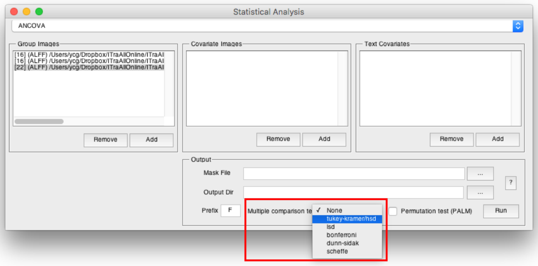
F Statistic Matrix

Other covariate can be also specified as text files.
(E.g. age, brain size, IQ etc.)

14

14

ANOVA or ANCOVA

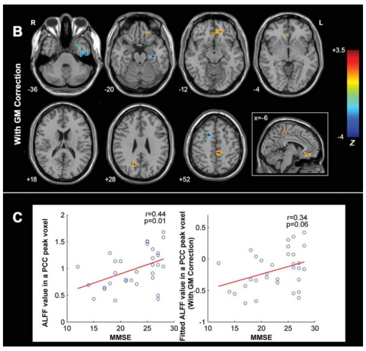


Post-hoc procedures: the corrected p values under a given control procedure for comparing group means of any pairs were calculated (e.g., through Studentized Range statistic for Tukey-Kramer correction) with the same route as MATLAB command multcompare. The p maps were then converted to Z maps according to the Normal inverse cumulative distribution function (norminv), with the sign of group mean differences applied.

15

15

Correlation Analysis

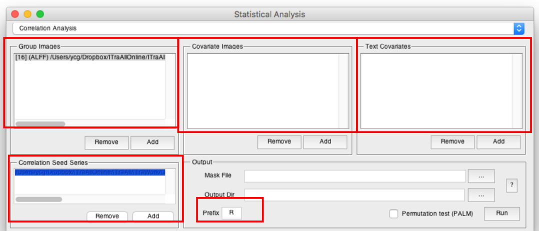


Wang[#], Yan[#] et al., 2011, Hum Brain Mapp

16

16

Correlation Analysis



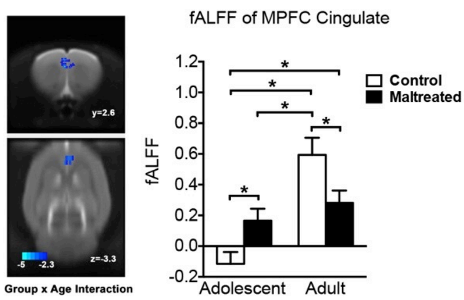
The imaging measure:
FC matrices

Traits: e.g.
MMSE.txt
19
15
...

17

17

Mixed Effect Analysis

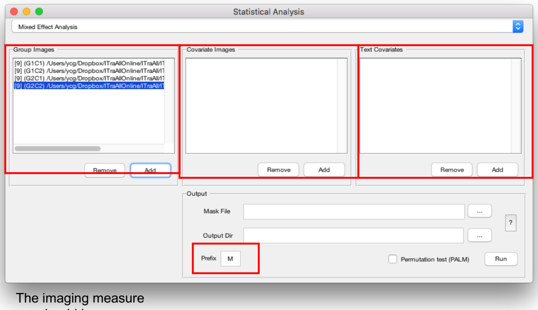


Yan et al., 2016. Translational Psychiatry

18

18

Mixed Effect Analysis



The imaging measure should be:
Group1Condition1
Group1Condition2
Group2Condition1
Group2Condition2

19

19

Mixed Effect Analysis

- *_ConditionEffect_T.mat - the T values of condition differences (corresponding to the first condition minus the second condition) (WithinSubjectFactor)
- *_Interaction_F.mat - the F values of interaction (BetweenSubjectFactor by WithinSubjectFactor)
- *_Group_TwoT.mat - the T values of group differences (corresponding to the first group minus the second group). Of note: the two conditions will be averaged first for each subject. (BetweenSubjectFactor)

20

Statistical Analysis

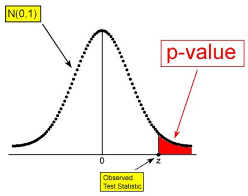
```
function [b_OLS_brain, t_OLS_brain, TF_ForContrast_brain, r_OLS_brain, Header] = y_GroupAnalysis_Image(DependentVolume,Predictor,OutputName,
% function [b_OLS_brain, t_OLS_brain, TF_ForContrast_brain, r_OLS_brain, Header] = y_GroupAnalysis_Image(DependentVolume,Predictor,OutputNam
% Perform regression analysis
% Input:
%   DependentVolume     -   4D data matrix (DimX*DimY*DimZ*DimTimePoints) or the directory of 3D image data file or the filename of one 4D d
%   Predictor - the Predictors M (subjects) by N (traits). SHOULD INCLUDE the CONSTANT column if needed. The program will not add constant c
%   OutputName - the output name. (should not have extention such as .img,.nii)
%   MaskFile - the mask file.
%   CovVolume  [optional] - 4D data matrix (DimX*DimY*DimZ*DimTimePoints) or the directory of image covariates, in which the files should be
%   Contrast [optional] - Contrast for T-test for F-test. 1*ncolX matrix.
%   TF_Flag [optional] - 'T' or 'F'. Specify if T-test or F-test need to be performed for the contrast
%   IsOutputResidual [optional] - 1: output the 4D residuals.
%                    - 0: don't output the 4D residuals
%   Header [optional] - If DependentVolume is given as a 4D Brain matrix, then Header should be designated.

% Output:
%   OutputName_b.nii, OutputName_T.nii     - beta and t value files results
%   OutputName_Residual.nii (optional)     - Residual files
%___________________________________________________________________________
% Written by YAN Chao-Gan 120823.
% The Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA
% Child Mind Institute, 445 Park Avenue, New York, NY 10022, USA
% The Phyllis Green and Randolph Cowen Institute for Pediatric Neuroscience, New York University Child Study Center, New York, NY 10016, USA
% ycg.yan@gmail.com
```

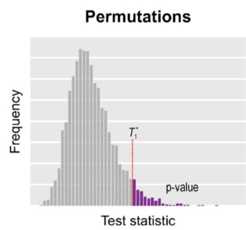
{DPABI_Dir}/StatisticalAnalysis/y_GroupAnalysis_Image.m

21

Get p: Parametric vs. non-parametric



Parametric



Non-Parametric

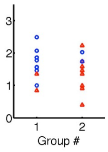
22

22

Non-parametric: permutation

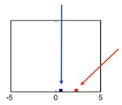
- We can permute the data itself to create a distribution that we can use to test our statistic.
  我们可以对数据本身进行置换，以创建可用于测试统计信息的分布。
  + Makes very few assumptions about the data 对数据做很少的假设
  + Works for any test statistic 适应于任何统计

One re-labelling
第一次重标记



t-value after re-labelling
重标记后t值

$t = 0.67$



Original labelling
原始标记

Let's start collecting them开始收集他们

Winkler et al., 2016. NeuroImage; Converted from FSL course

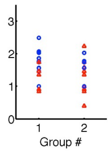
23

23

Non-parametric: permutation

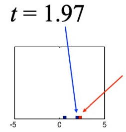
- We can permute the data itself to create a distribution that we can use to test our statistic.
  我们可以对数据本身进行置换，以创建可用于测试统计信息的分布。
  + Makes very few assumptions about the data 对数据做很少的假设
  + Works for any test statistic 适应于任何统计

Second re-labelling
第二次重标记



t-value after re-labelling重标记后的t值

$t = 1.97$



Original labelling
原始标记

And another one另一个

Converted from FSL course

24

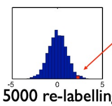
24

Non-parametric: permutation

- We can permute the data itself to create a distribution that we can use to test our statistic.
  我们可以对数据本身进行置换，以创建可用于测试统计信息的分布。
  + Makes very few assumptions about the data 对数据做很少的假设
  + Works for any test statistic 适应于任何统计

Of the 5000 re-labellings, only 90 had a t-value > 2.27 (the original labelling).
在5000个重新标记中，只有90个的t值> 2.27（原始标记）。

I.e. there is only a ~1.8% (90/5000) chance of obtaining a value > 2.27 if there is no difference between the groups
C.f. $p(x \geq 2.27) = 1.79\%$ for t_{18}
即如果各组之间没有差异，则只有1.8%（90/5000）的机会获得> 2.27的值。C.F. t18的p（x 2.27） = 1.79%



Original labelling
原始标记

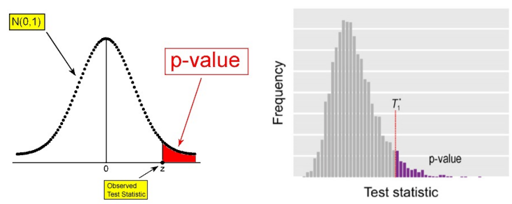
5000 re-labellings. Phew!
5000次重标记，喷！

Converted from FSL course

Get p: Parametric vs. non-parametric

	Parametric	Non-parametric
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Homogenous and Heterogeneous
Typical data	Ratio or Interval	Ordinal or Nominal
Data set relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	Can draw more conclusions	Simplicity; Less affected by outliers

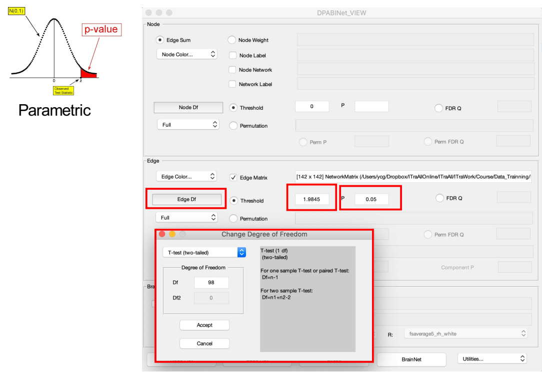
Get p: Parametric vs. non-parametric

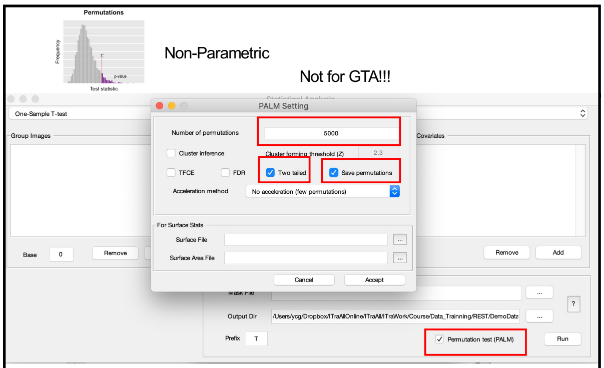


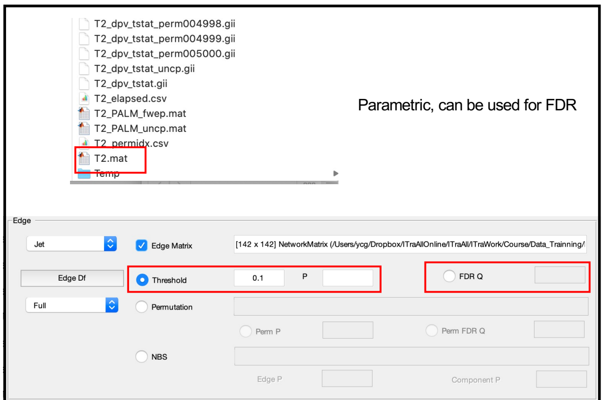
Parametric

Non-Parametric

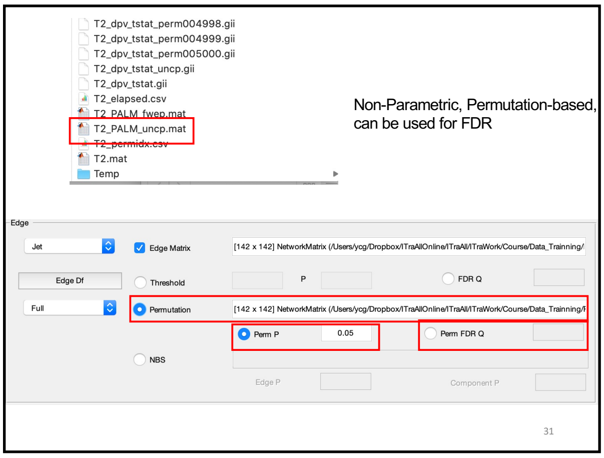
Get p: Parametric vs. non-parametric



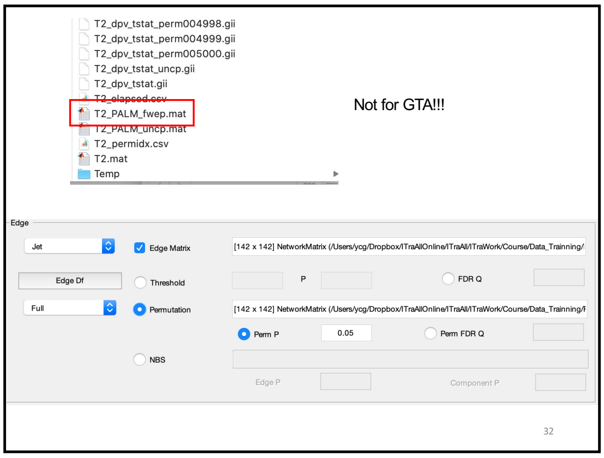




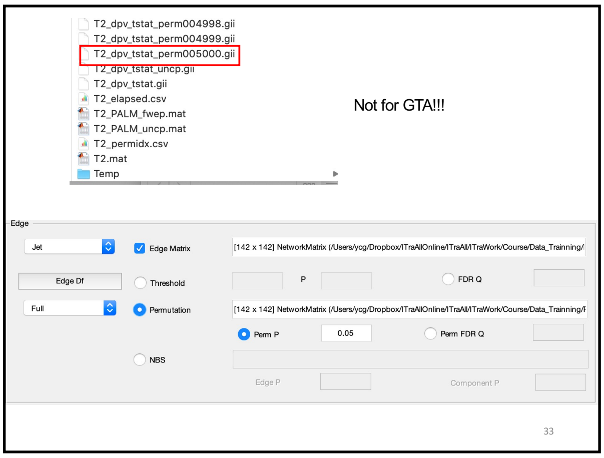
Parametric, can be used for FDR



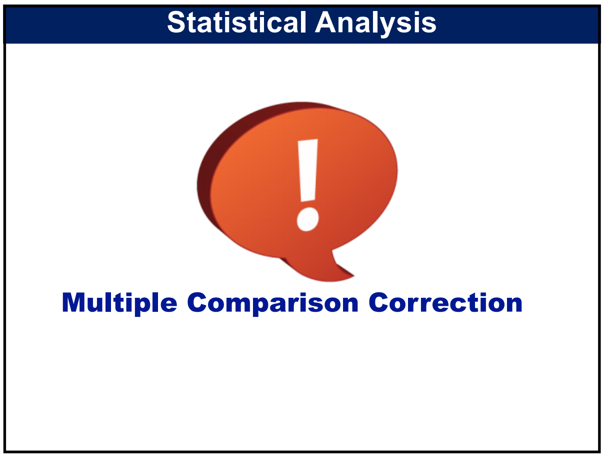
31



32

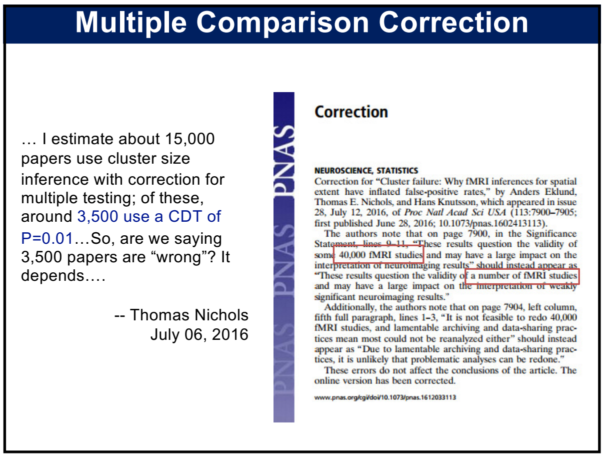


33



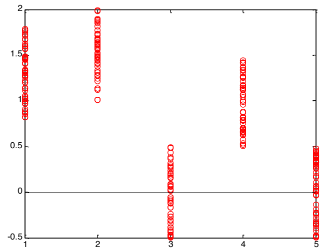
34

35



36

Multiple Comparison Correction



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:

$0.95^5=0.774$

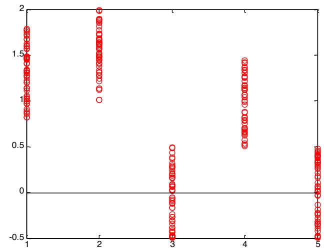
1 - 0.05 = 0.95

37

37

Multiple Comparison Correction

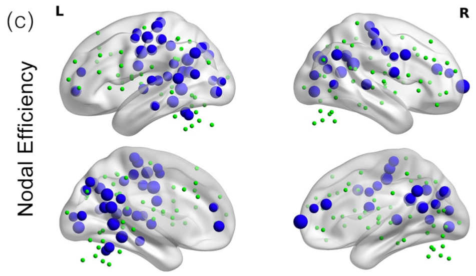
Bonferroni correction: p=0.05/5=0.01



0.01 0.01 0.01 0.01 0.01

38

Multiple Comparison Correction



39

Multiple Comparison Correction

- Bonferroni correction: 0.05/N
- False Discovery Rates (FDR) correction

40

Multiple comparisons

Bonferroni correction

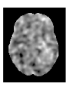
The Bonferroni correction rejects the null hypothesis for each $p_i \leq \alpha/m$, thereby controlling the FWER at $\leq \alpha$

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

Carlo Emilio Bonferroni

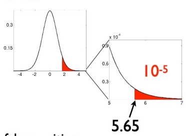
Bonferroni says threshold at α divided by # of tests

阈值应该是α除以测试次数



5255 voxels

$0.05/5255 \approx 10^{-5}$



z-map thresholded at 5.65

阈值在5.65

No false positives. Hurrah for Italy!

无误报。意大利万岁！

Converted from FSL course

41

41

FDR Theory

Number of errors committed when testing m null hypotheses

	Declared non-significant	Declared significant	Total
True null hypotheses	U	V	m_0
Non-true null hypotheses	T	S	$m - m_0$
	$m - \mathbf{R}$	R	m

- False discovery rate $Q_e=E(V/(V+S))=E(V/R)$

Benjamini and Hochberg, 1995, Journal of the Royal Statistical Society

42

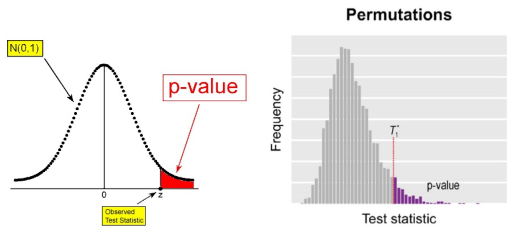
FDR Theory

- Let $H_1, \cdots, H_m$ be the null hypotheses and $P_1, \cdots, P_m$ their corresponding p-values. Order these values in increasing order and denote them by $P_{(1)}, \cdots, P_{(m)}$. For a given q, find the largest k such that $P_{(k)} \leqq kq/m$.

- Then reject (i.e. declare positive) all $H_{(i)}$ for i = 1, ⋯, k.

43

Get p: Parametric vs. non-parametric

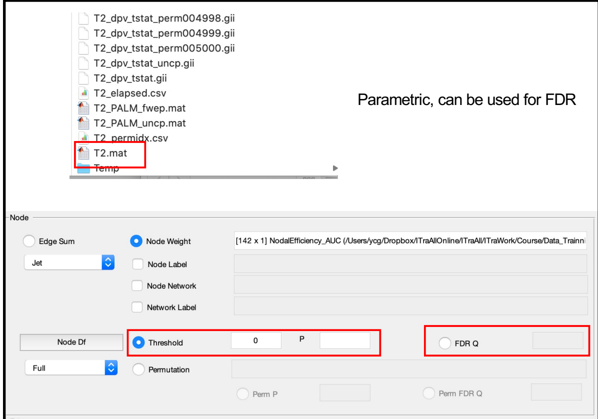


Parametric

Non-Parametric

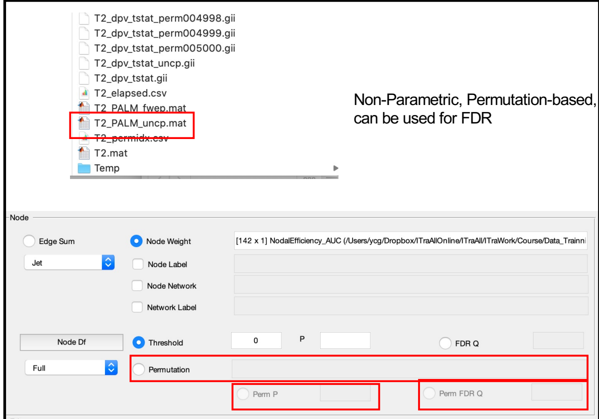
44

44



Parametric, can be used for FDR

45



Non-Parametric, Permutation-based, can be used for FDR

46

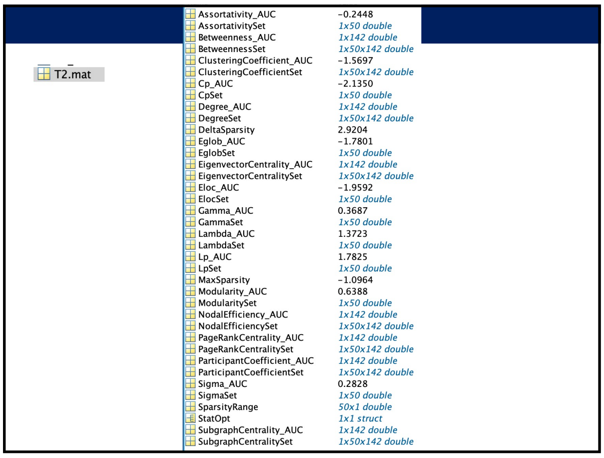
Outline

- Statistical Analysis
- Results Viewing

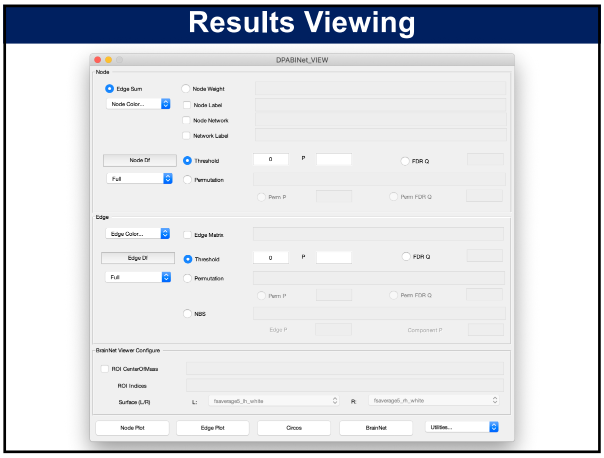
47

47

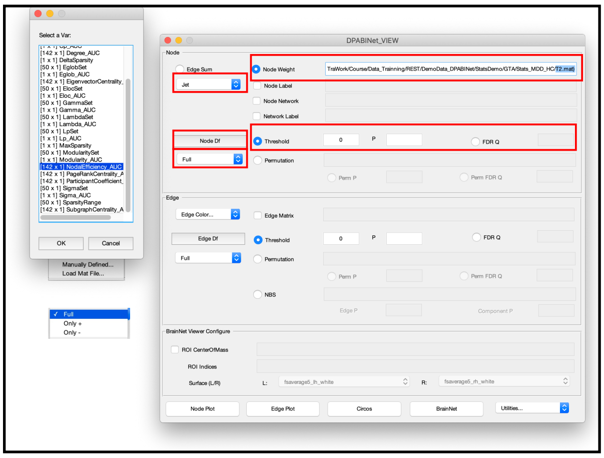
48



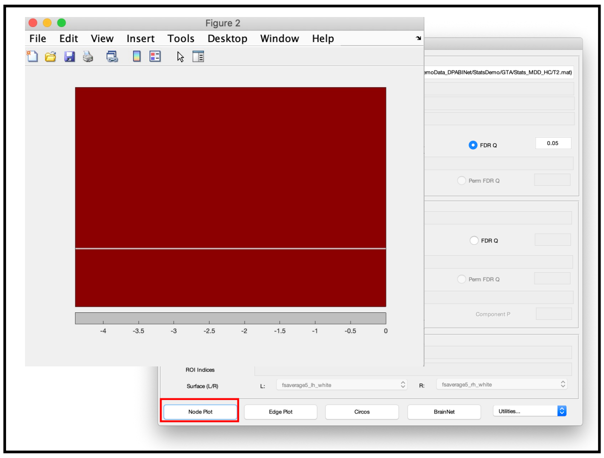
49



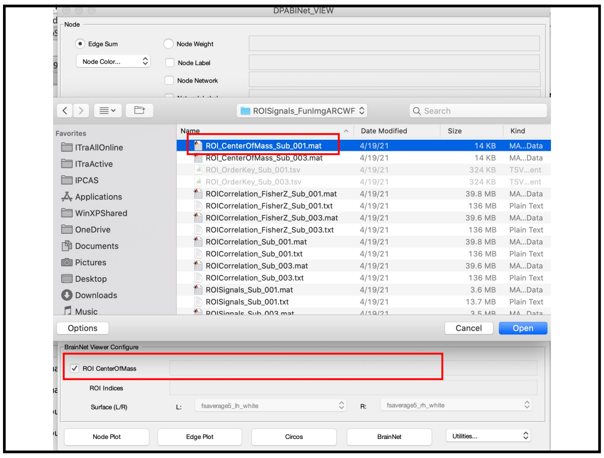
50



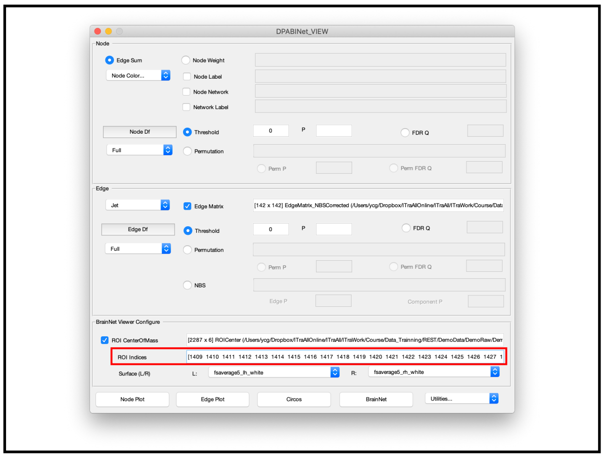
51



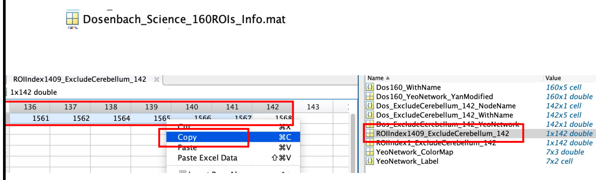
52



53

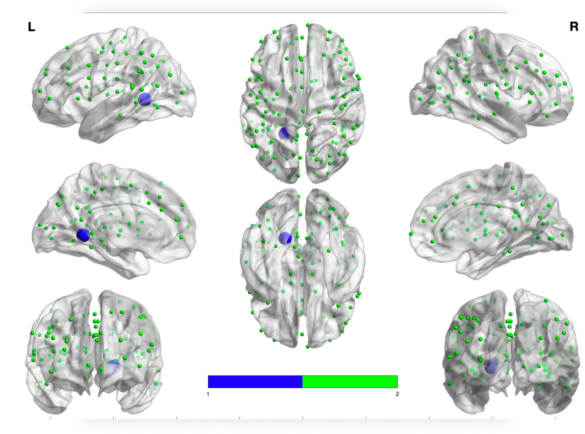


54



55

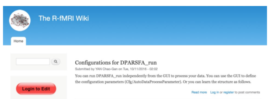
55



56

Further Help

http://rfmri.org/Course



http://rfmri.org/wiki

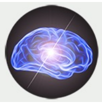
The R-fMRI Journal Club

Official Account: RFMRILab

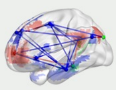
57

57

DPABI/DPABISurf/DPARSF特训营



第九届DPABI/DPABISurf/DPARSF
脑影像基础特训营（云端）通知
2021.3.27～3.29



第一届DPABISurf/DPABINet
脑网络进阶特训营（北京现场）通知
2021.4.24～4.26

定期举办，请关注http://rfmri.org

58

58

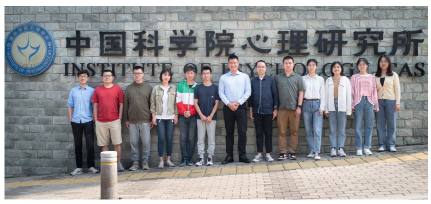
DPABISurf皮层数据预处理业务

特训营学员：99元/例次

59

59

Acknowledgments



Chinese Academy of Sciences
Xi-Nian Zuo
Wei-Wen Wang
Fei Luo
Hangzhou Normal University
Yu-Feng Zang
NYU Child Study Center
F. Xavier Castellanos
Peking University Sixth Hospital
Tian-Mei Si
Jing Liu

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

60

60

Thanks for your attention!

61

61