Same Data - Different Software - Different Results?

Analytic Variability of Group fMRI Results

Alex Bowring, Camille Maumet, Thomas Nichols

Which software?

Which software?

FreeSurfer

BrainVoyager™ QX
The brain analysis and visualisation tool for Linux
Version 2.8

Mango

NITRC

AFNI!

FSL

SPM

Camba
UNIVERSITY OF CAMBRIDGE

ni learn

CARET

Analytic Flexibility

The range of analysis outcomes across different acceptable analysis methods.

- **Choice of Analysis Pipeline (functional MRI)**

Carp J. (2012), Frontiers in Neuroscience

- **Choice of Analysis Software Version**

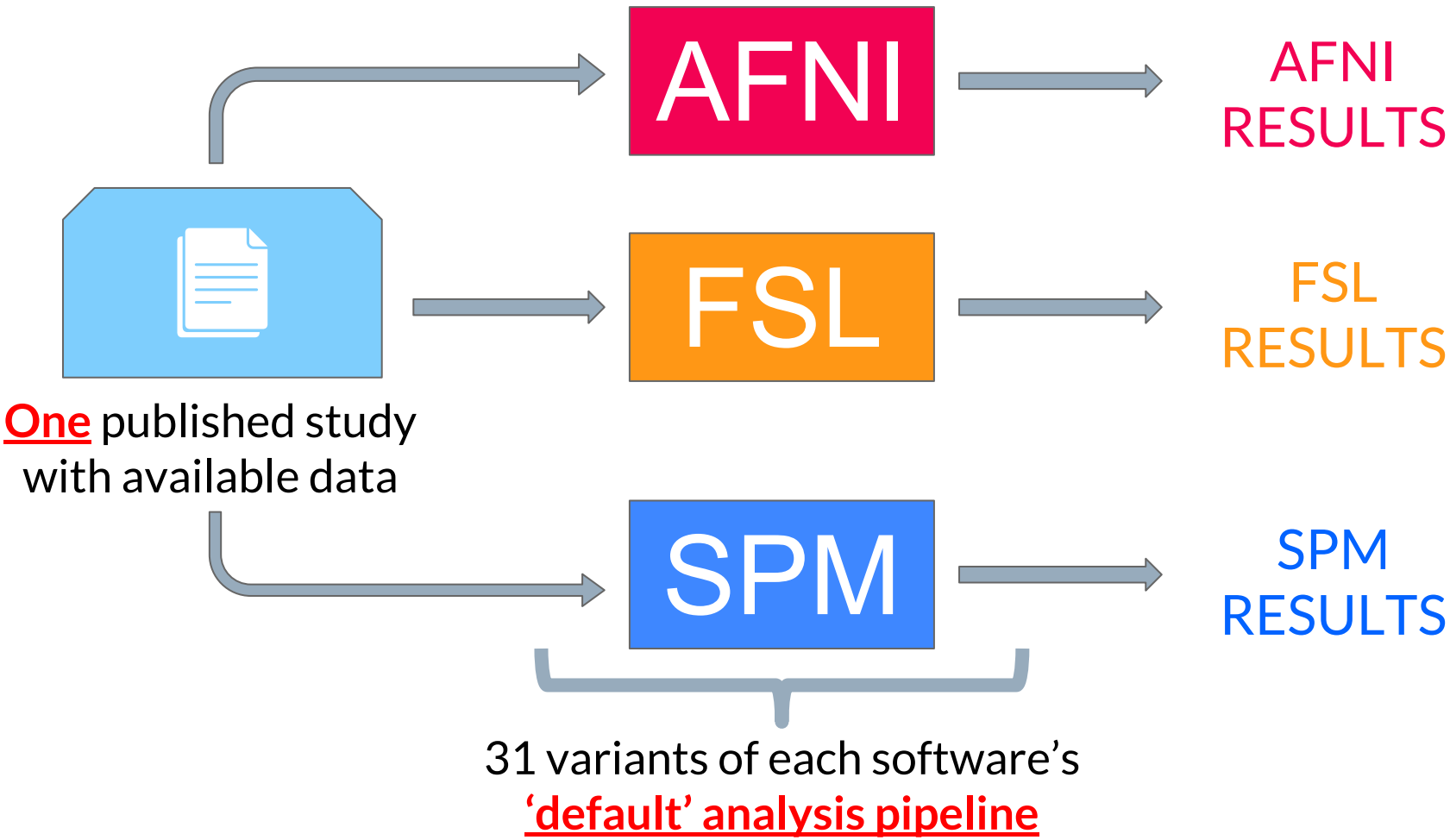
Groenenschild E. et al (2012), PLoS ONE

- **Choice of operating system**

Glatard T. et al (2015), Frontiers in Neuroinformatics

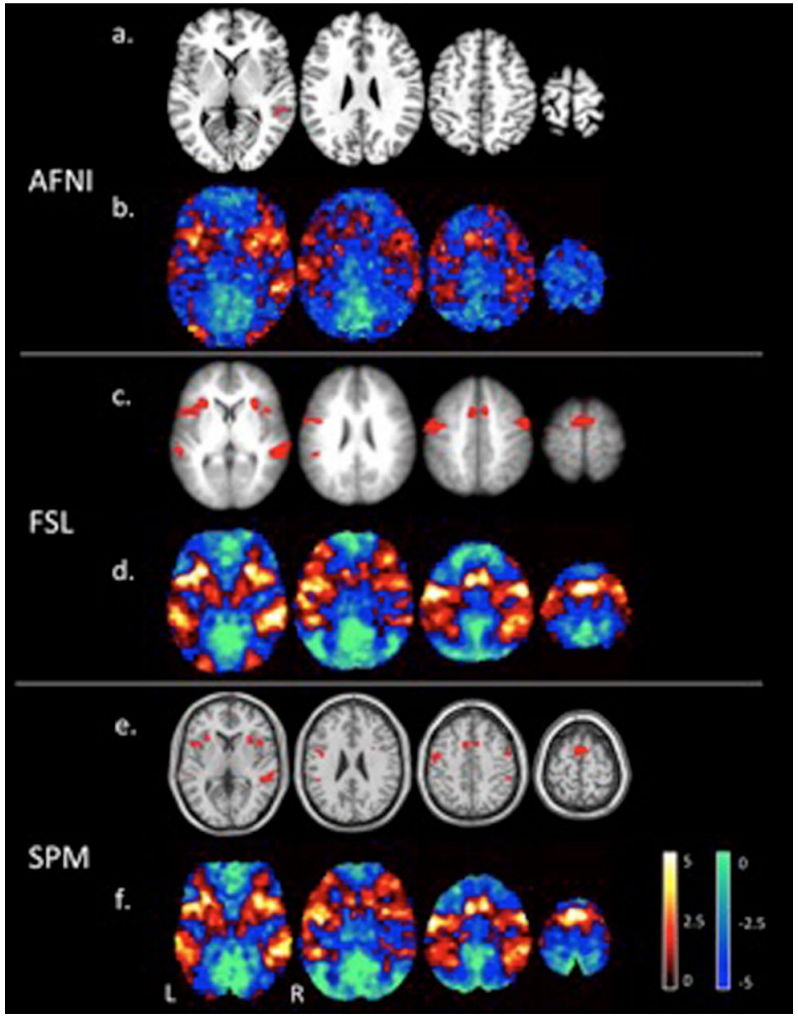
How does the choice of analysis software package impact the analysis results?

Background



Pauli R. et al (2016), Frontiers in Neuroinformatics

Results



- Unthresholded/thresholded statistic maps:

Differences in patterns of spatial activation

RESULTS ARE NOT LIKE-FOR-LIKE!

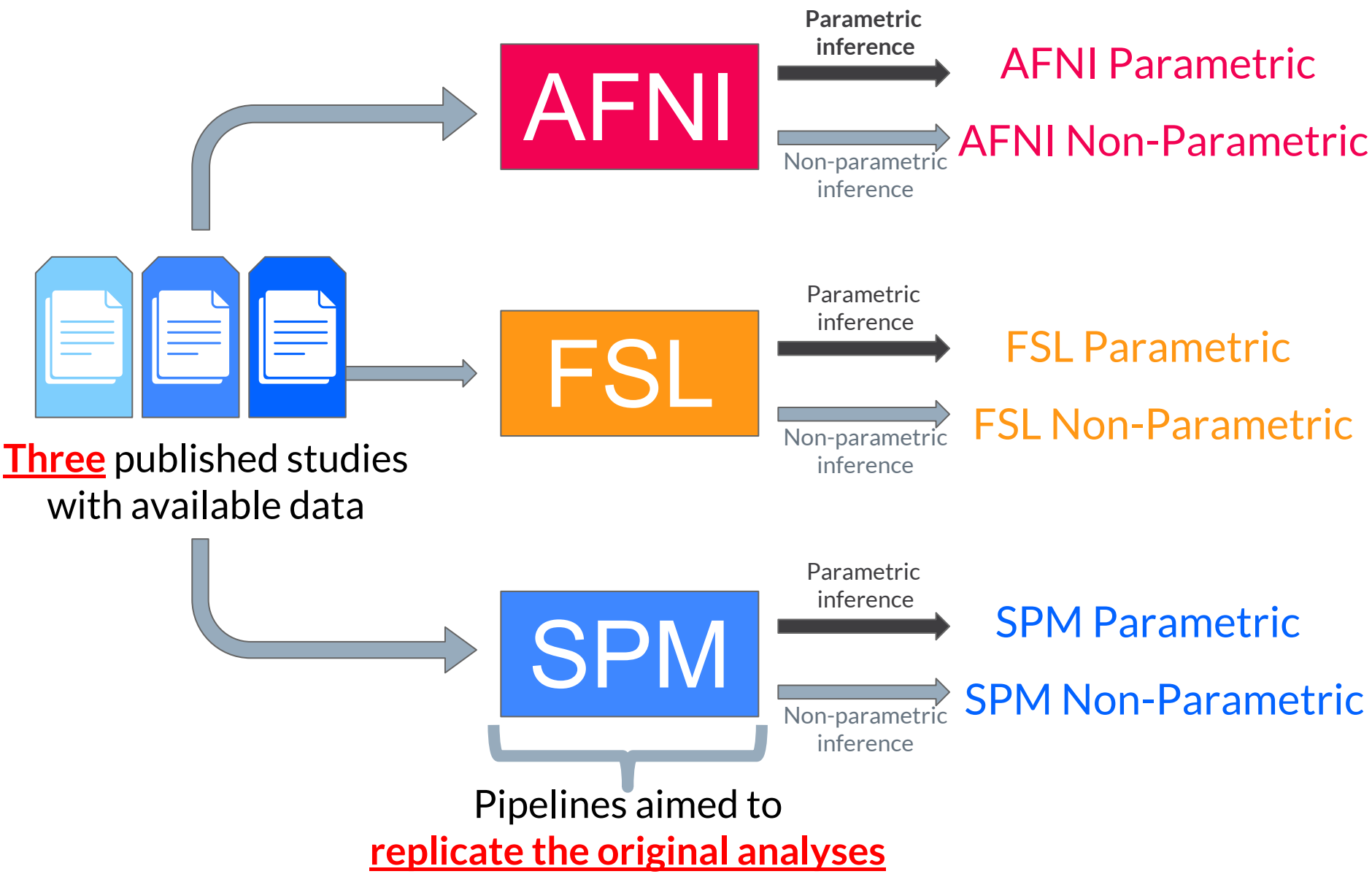
- Differences in default pipelines:

AFNI 4mm FWHM smoothing

FSL/SPM 6mm FWHM smoothing

It would be more interesting if analysis pipelines were identical!

Pauli R. et al (2016), Frontiers in Neuroinformatics



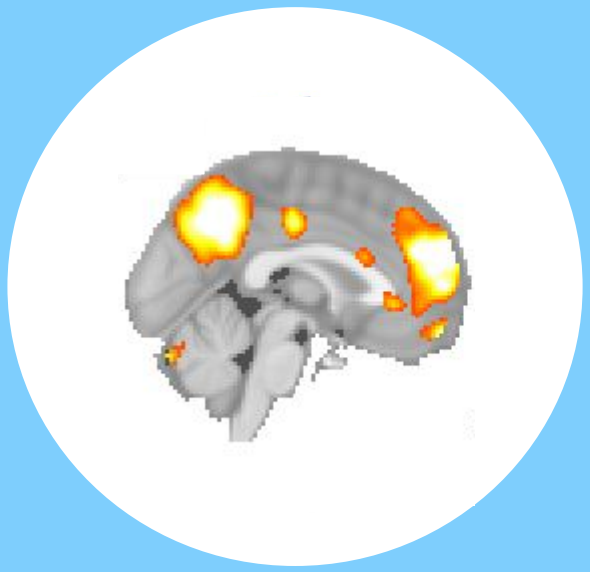
Bowring A. et al (2018), BioRxiv (preprint)

Replicating Publication Analyses

- Three OpenfMRI datasets w/ publications:
  - ds000001
    - Balloon Analog Risk Task
    - N = 16
  - ds000109
    - False-belief Task
    - N = 21
  - ds000120
    - Reward-processing Task
    - N = 17
- Pipelines implemented in software based on info in publications

Maximizing Results Comparability

- Brain extraction
- Non-linear registration to MNI space
- 6 motion regressors added to GLM
- Functional data resampled to 2mm
- Two group-level inference models:
  - Parametric inference
  - Non-parametric inference



Thresholded Statistic Map Comparisons

ds000001 Thresholded Maps

- Balloon Analog Risk Task (N=16)

- Contrast:
  - Pumps demean vs control demean

Cluster-forming threshold:

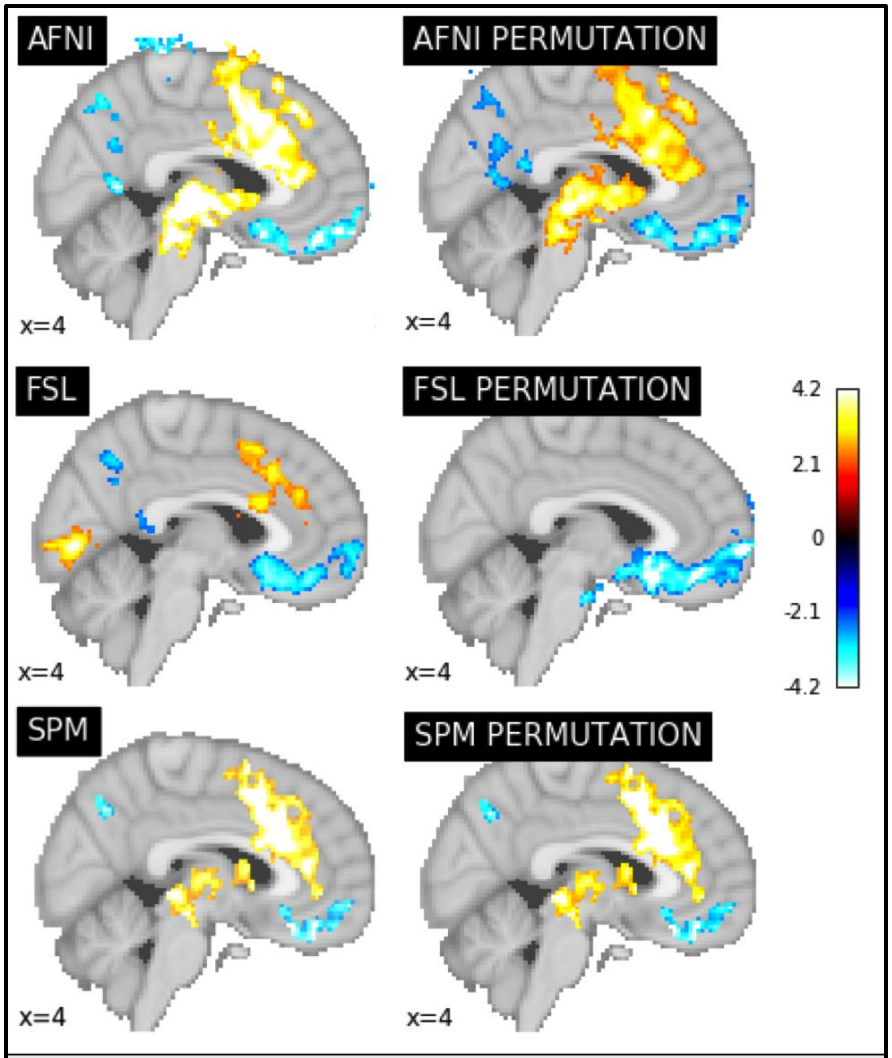
p < 0.01

Clusterwise threshold:

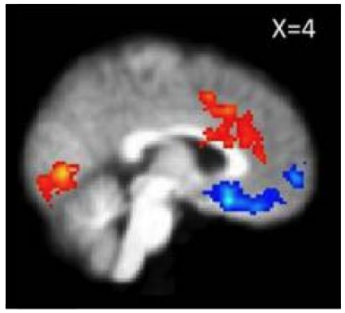
p < 0.05 FWE

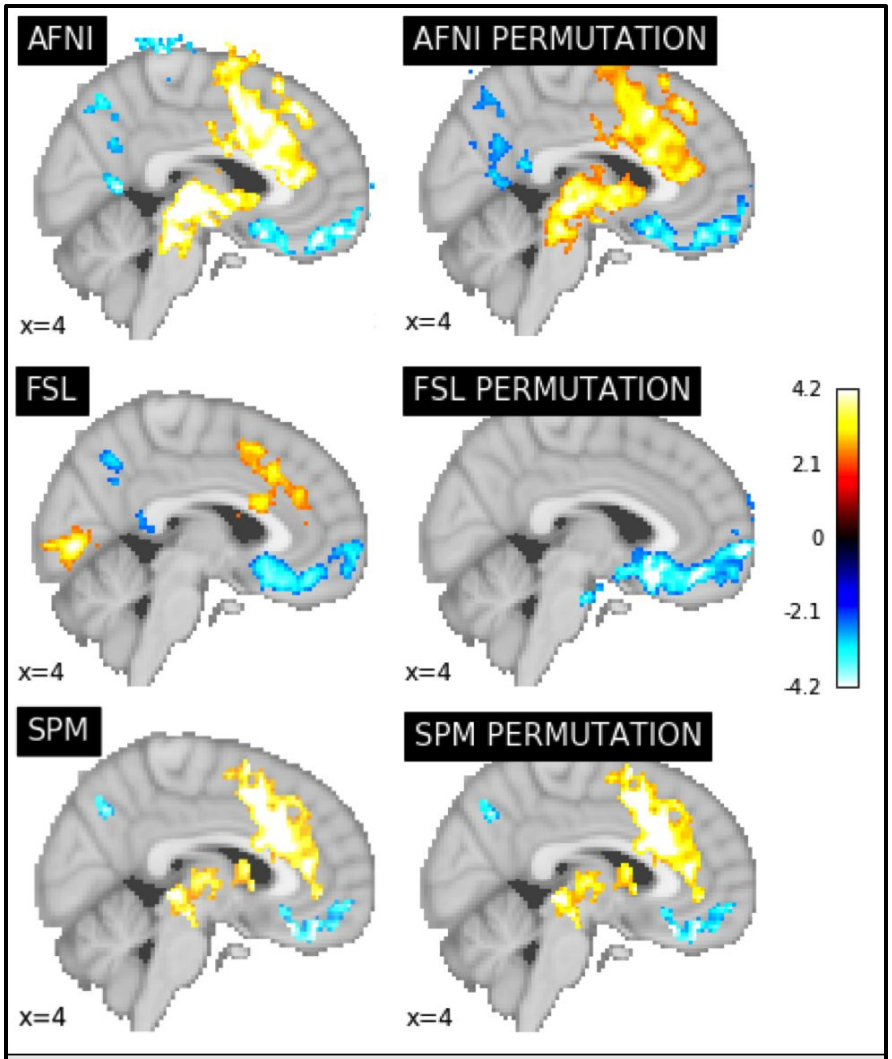
- Publication:

Schonberg T.S. et al (2012), Frontiers in Decision Neuroscience

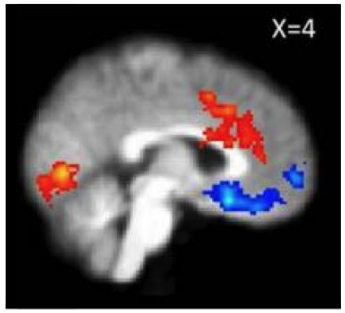


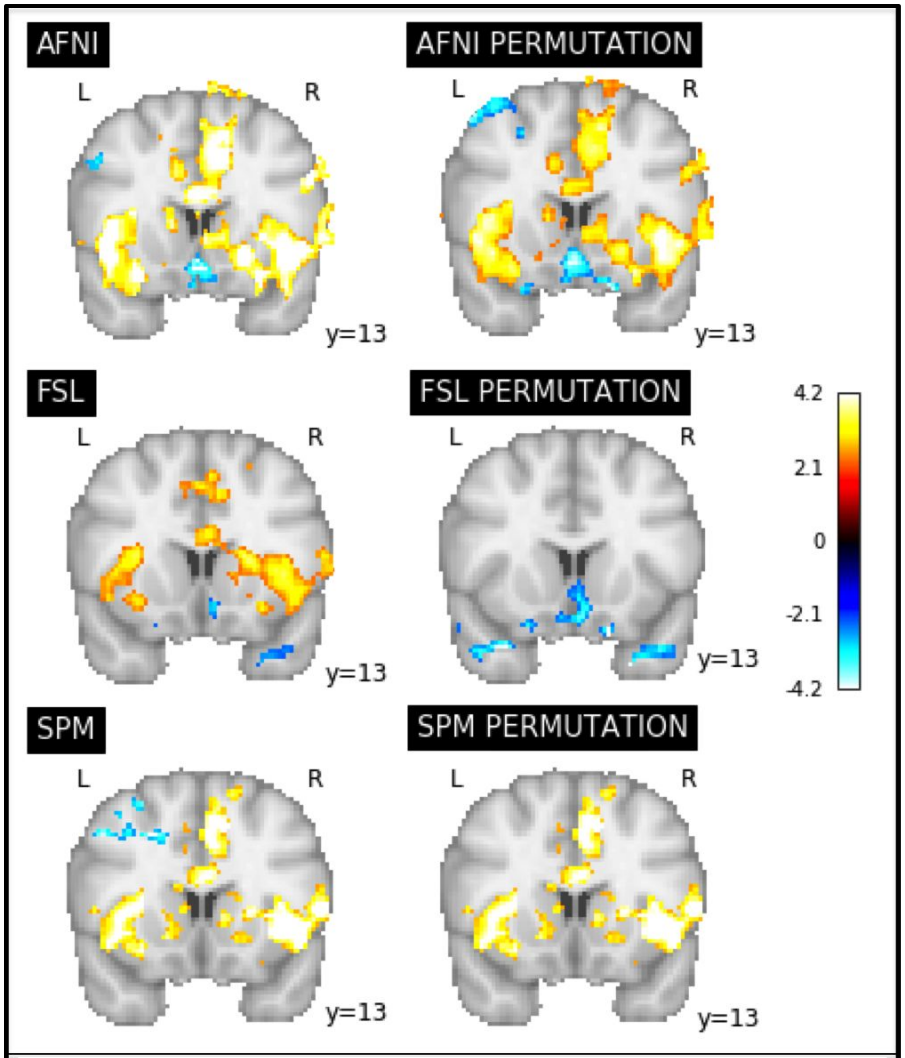
PUBLICATION
(FSL)



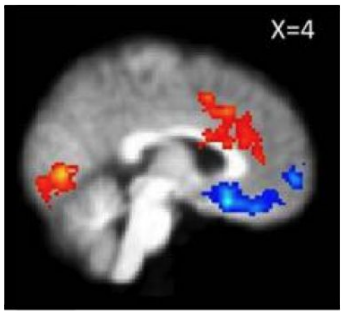


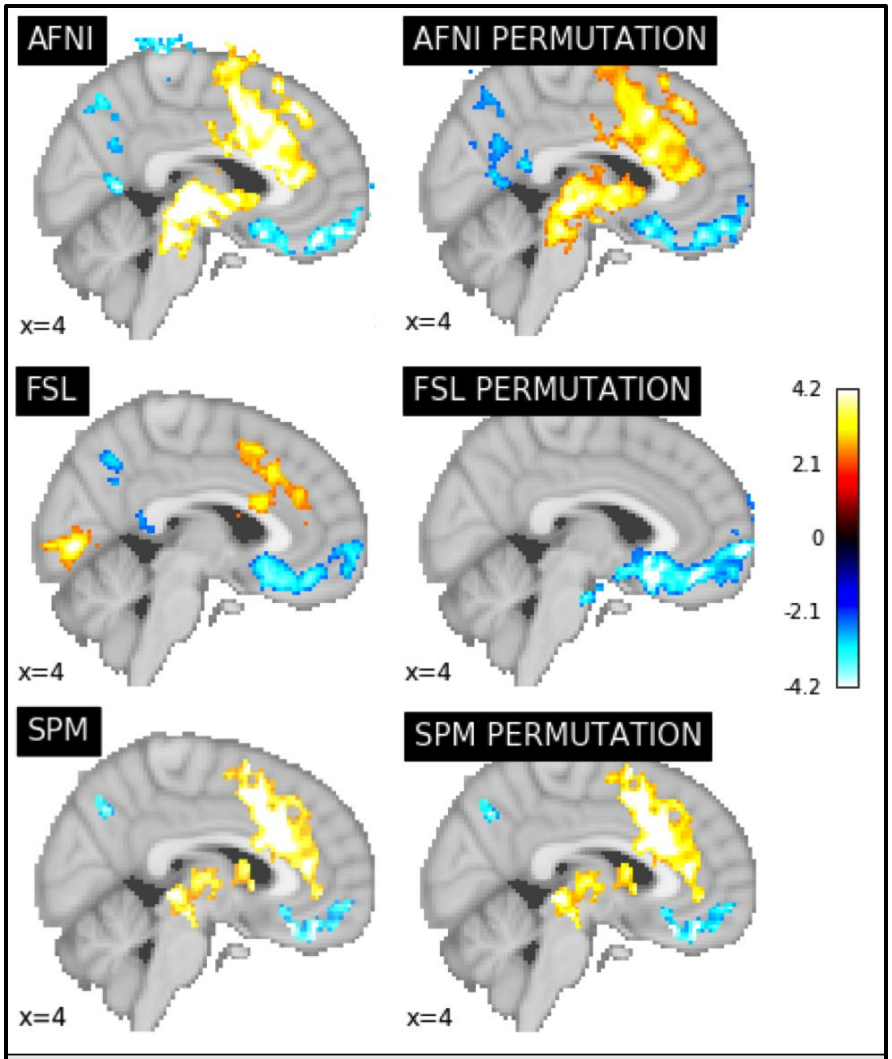
PUBLICATION
(FSL)



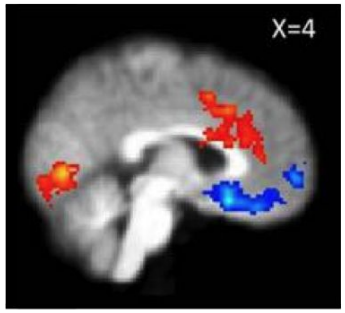


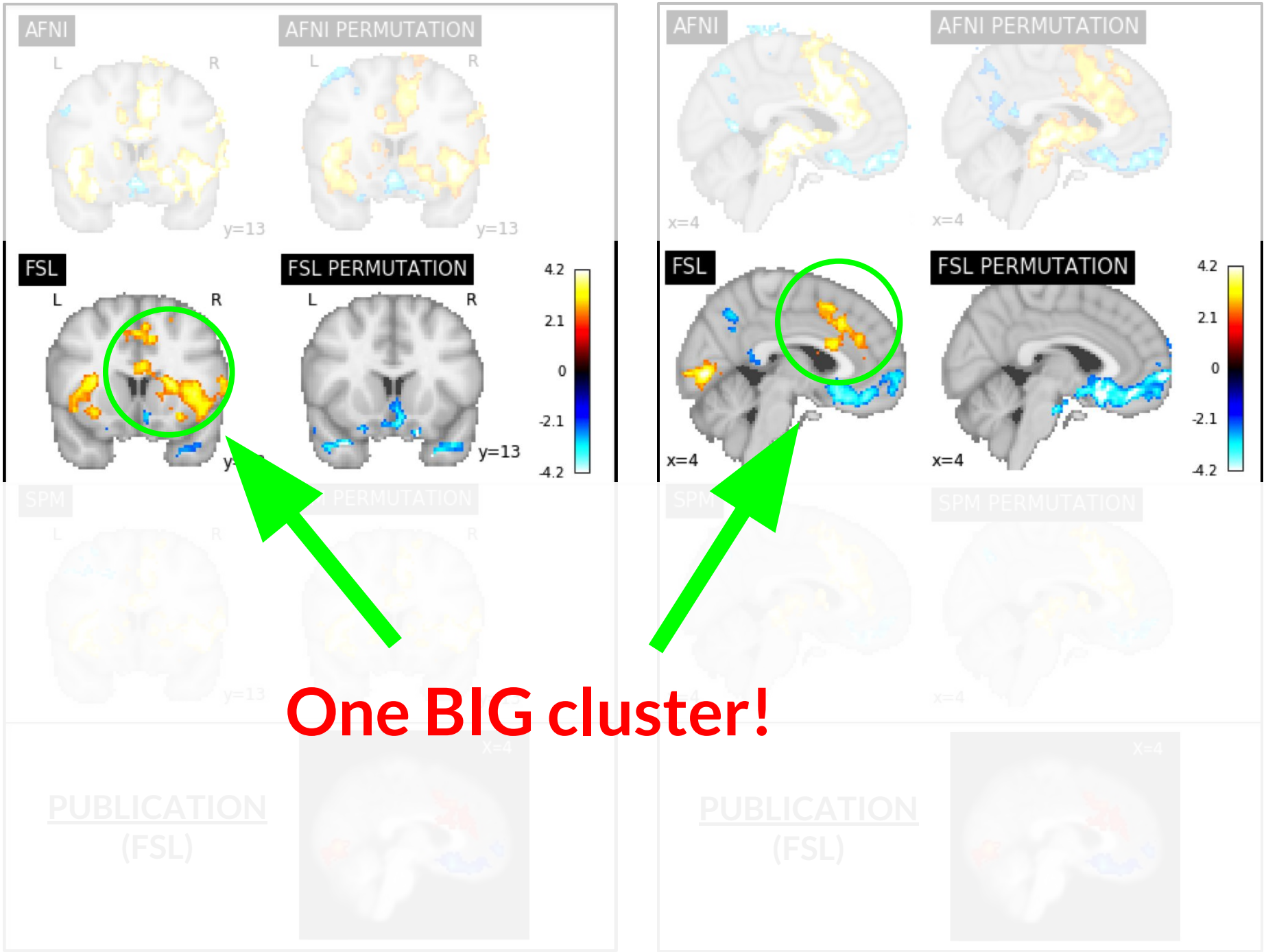
PUBLICATION
(FSL)





PUBLICATION
(FSL)





One BIG cluster!

ds000109 Thresholded Maps

- False-belief Task (N=21)

- Contrast:
  - False photo vs false belief

Cluster-forming threshold:

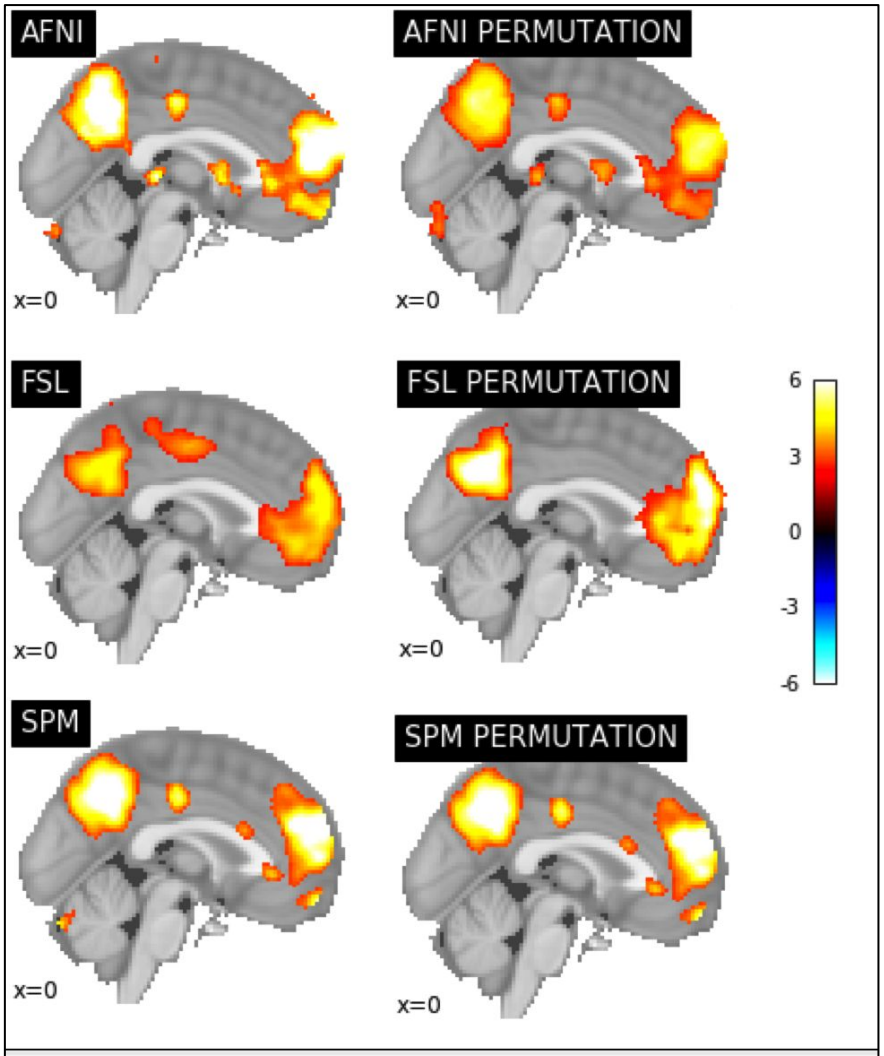
p < 0.005

Clusterwise threshold:

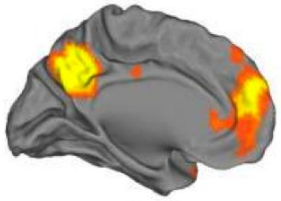
p < 0.05 FWE

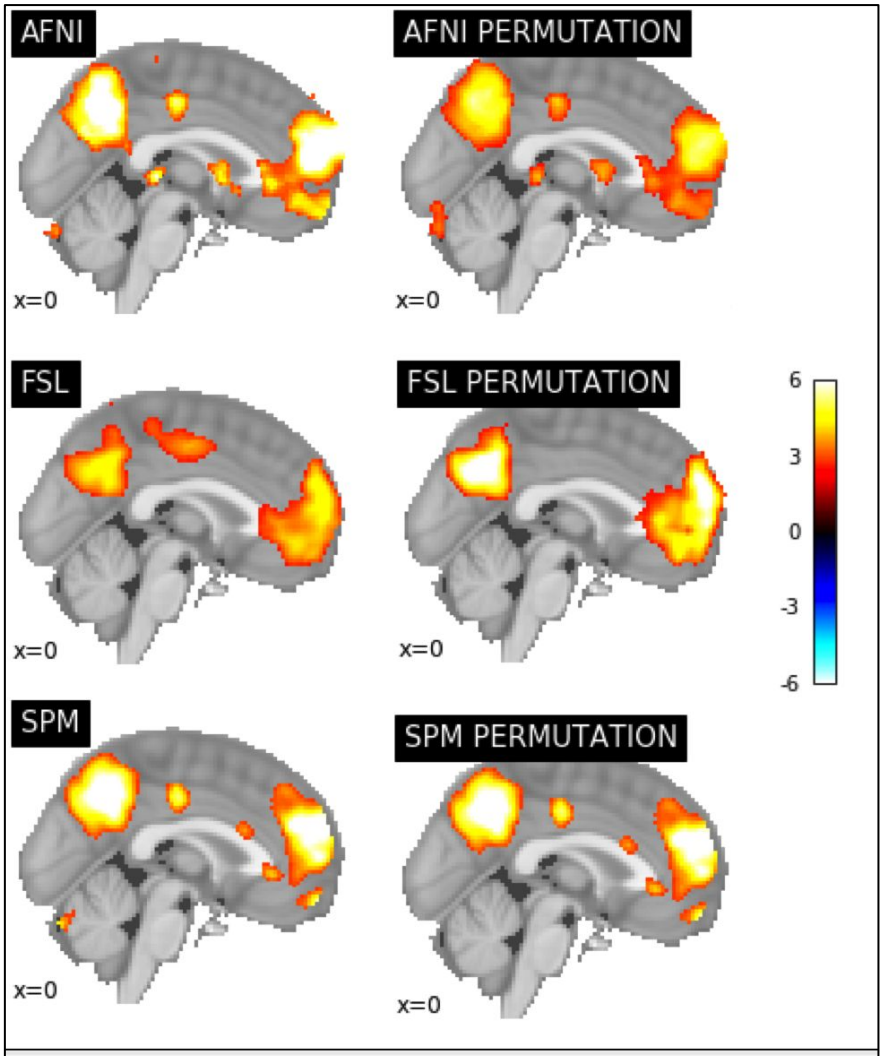
- Publication:

Moran J.M. et al (2012), Journal of Neuroscience

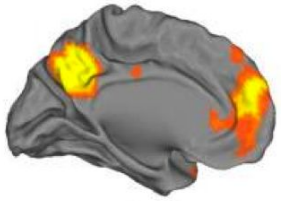


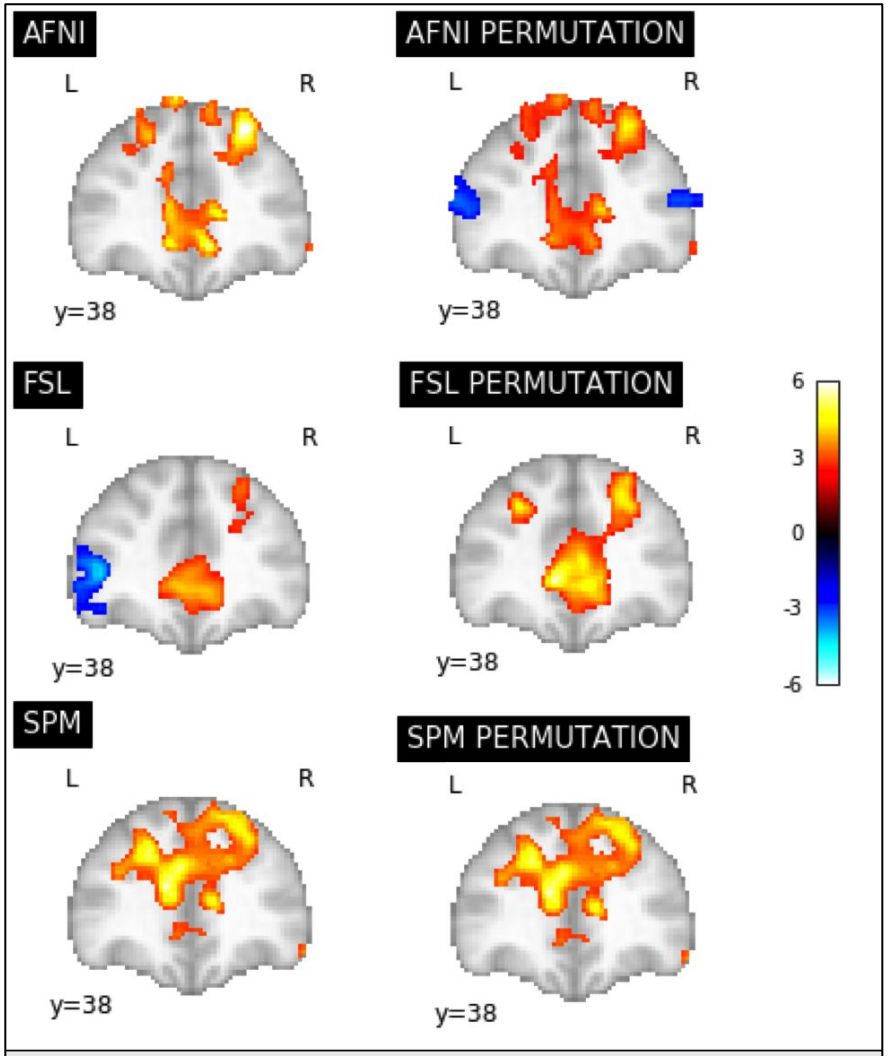
PUBLICATION
(SPM)

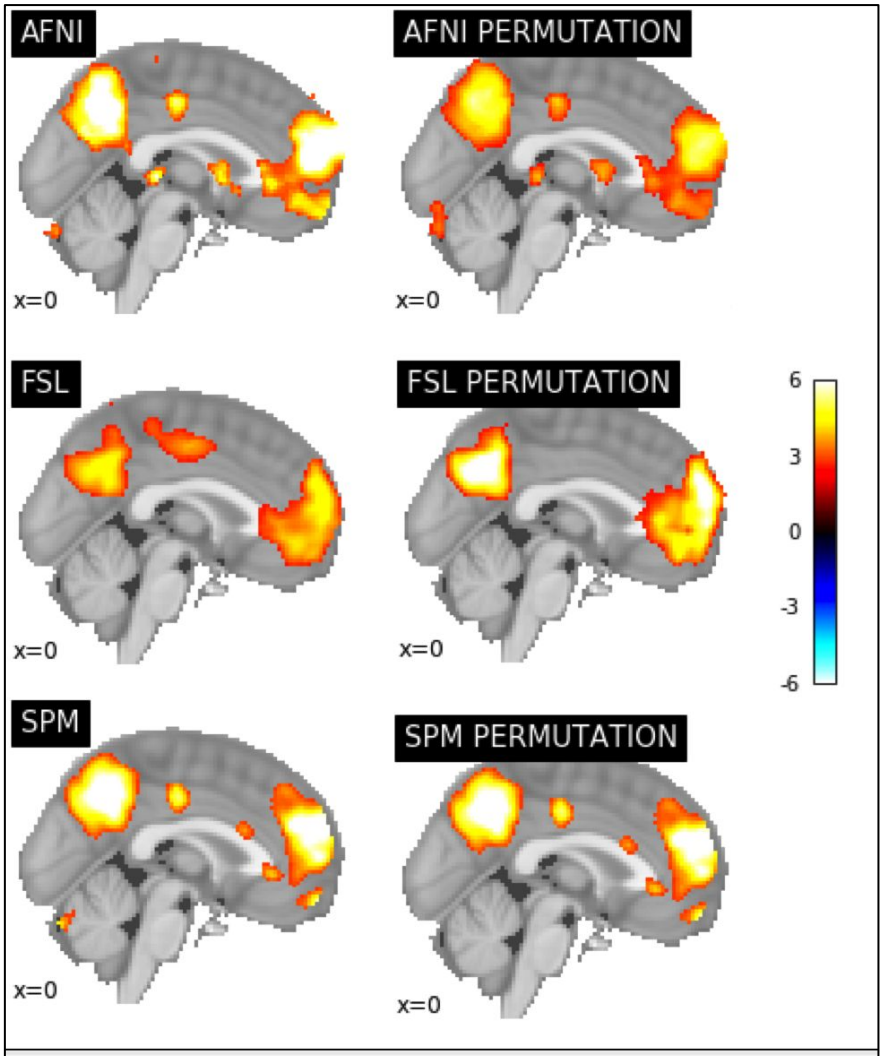




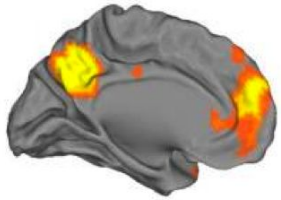
PUBLICATION
(SPM)



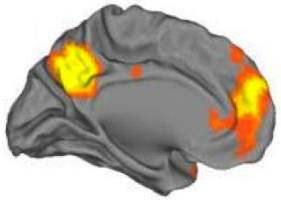


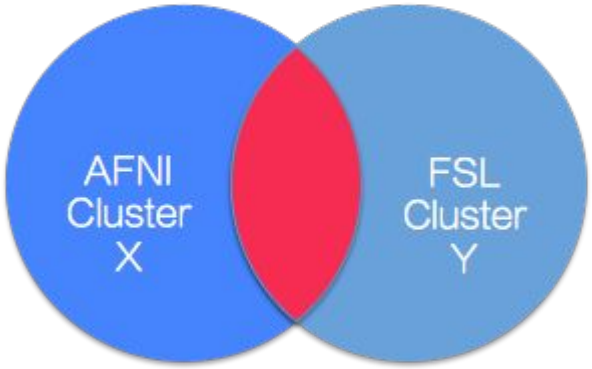


PUBLICATION
(SPM)



PUBLICATION
(SPM)





Dice = $\frac{2|X \cap Y|}{|X| + |Y|}$

- Thresholded Statistic maps
- Assess differences in **locations of activation**
- Measures percentage 'spill-over' of activation between software

Dice Coefficient Comparisons

ds000001 dice (Positive Clusters)

	AFNI	FSL	SPM	AFNI PERM	FSL PERM	SPM PERM
AFNI	-					
FSL	0.258	-				
SPM	0.465	0.381	-			
AFNI PERM	0.850	0.246	0.472	-		
FSL PERM	0.000	0.000	0.000	0.000	-	
SPM PERM	0.477	0.369	0.920	0.519	0.000	-

ds000001 dice (Positive Clusters)

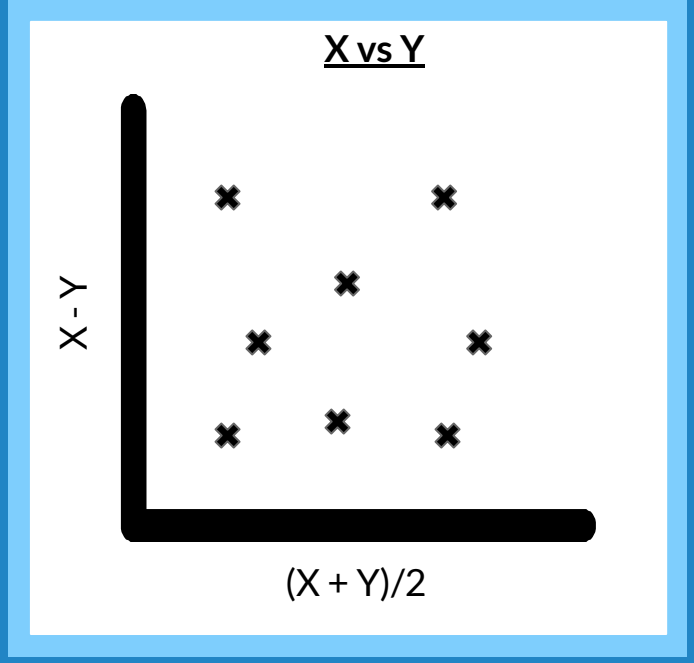
	AFNI	FSL	SPM	AFNI PERM	FSL PERM	SPM PERM
AFNI	-					
FSL	0.258	-				
SPM	0.465	0.381	-			
AFNI PERM	0.850	0.246	0.472	-		
FSL PERM	0.000	0.000	0.000	0.000	-	
SPM PERM	0.477	0.369	0.920	0.519	0.000	-

ds000001 dice (Positive Clusters)

	AFNI	FSL	SPM	AFNI PERM	FSL PERM	SPM PERM
AFNI	-					
FSL	0.258	-				
SPM	0.465	0.381	-			
AFNI PERM	0.850	0.246	0.472	-		
FSL PERM	0.000	0.000	0.000	0.000	-	
SPM PERM	0.477	0.369	0.920	0.519	0.000	-

ds000001 dice (Positive Clusters)

	AFNI	FSL	SPM	AFNI PERM	FSL PERM	SPM PERM
AFNI	-					
FSL	0.258	-				
SPM	0.465	0.381	-			
AFNI PERM	0.850	0.246	0.472	-		
FSL PERM	0.000	0.000	0.000	0.000	-	
SPM PERM	0.477	0.369	0.920	0.519	0.000	-

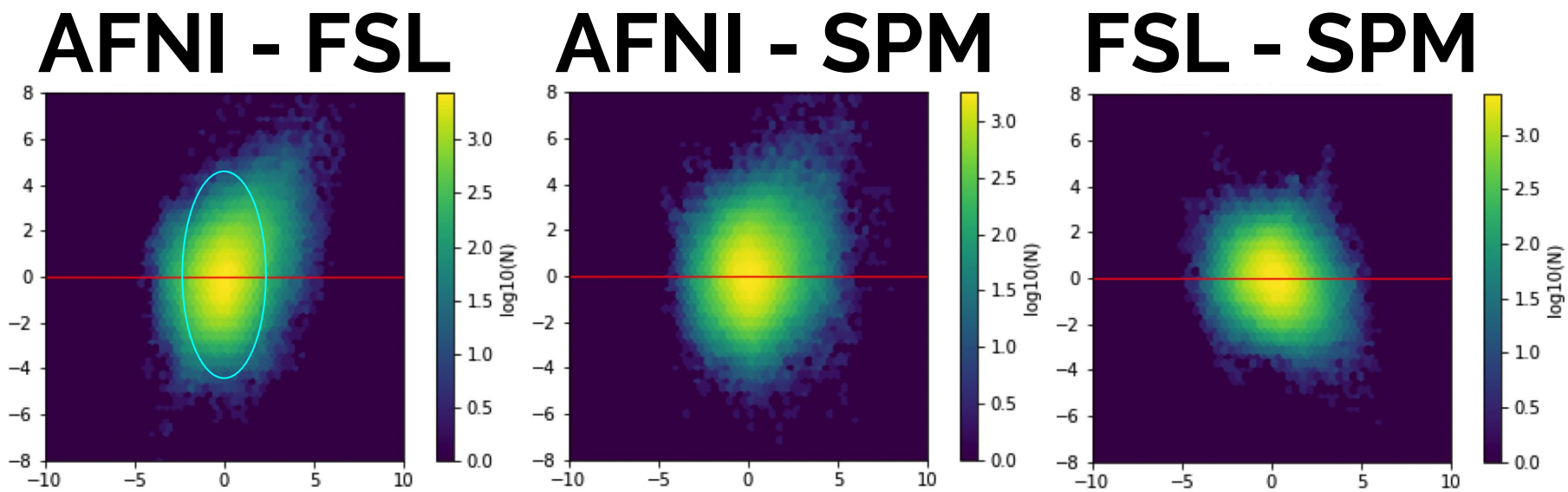


- Unthresholded Statistic maps
- Assess differences in **magnitudes of activation**
- At each voxel, between software:
  Average T-stat vs Diff of T-stats

Bland-Altman Plot Comparisons

ds000001 Parametric BA Plots

Y = Difference of T-statistics



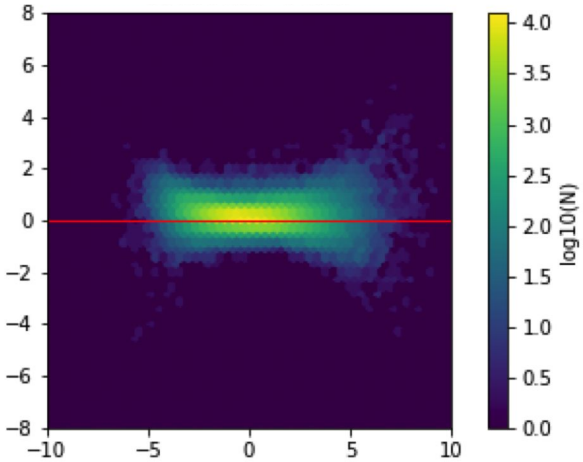
X = Average of T-statistics

- Plots similar to expected variation if **independant data** was fed into each package

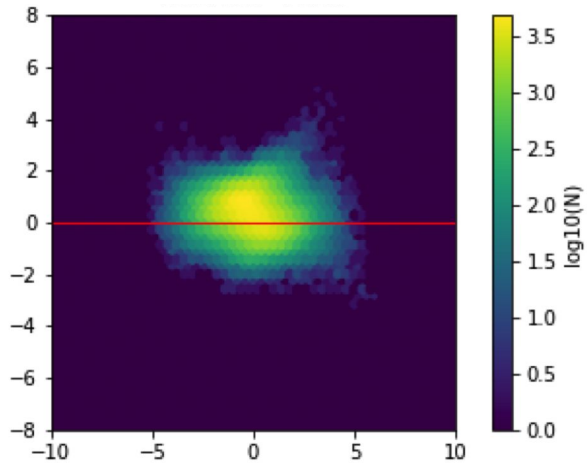
ds000001 Intra-Software BA Plots

Y = Difference of T-statistics

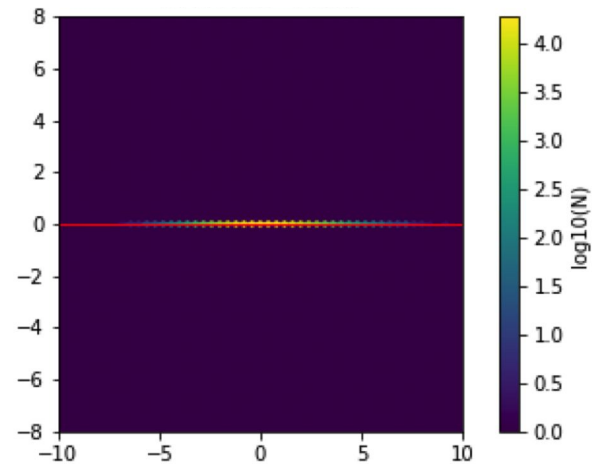
AFNI
PARA - PERM



FSL
PARA - PERM



SPM
PARA - PERM



X = Average of T-statistics

- SPM uses same OLS method in Para/Perm inference models to obtain unthresholded T-stat

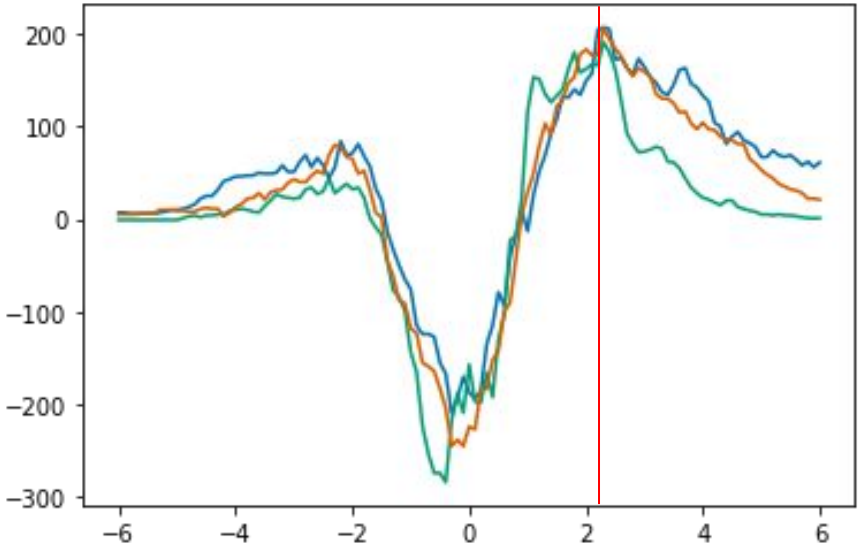
- Thresholded Statistic maps (range of T-values)
- Assess differences in **topological properties of statistic maps**
- For thresholded images: EC = Clusters - Handles + Holes

Euler Characteristic (EC) Comparisons

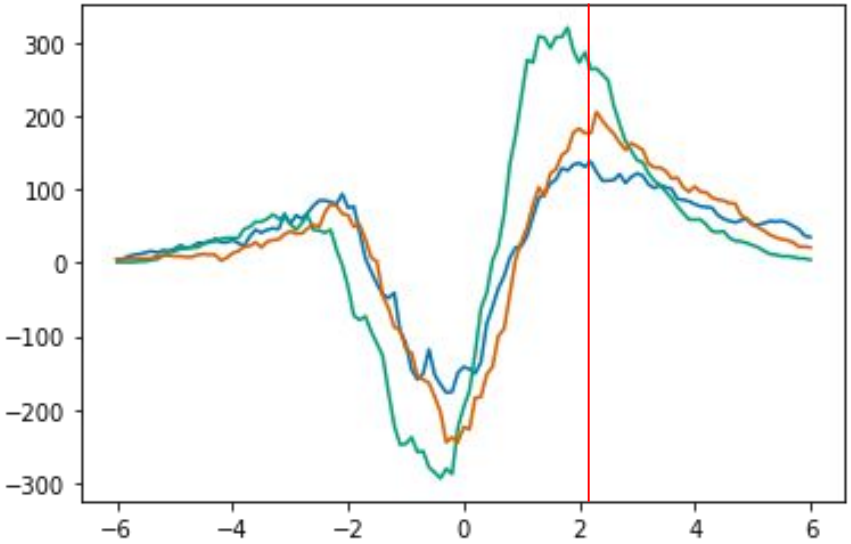
ds000001 Euler Characteristics (ECs)

Y = EC

Parametric ECs



Non-Parametric ECs

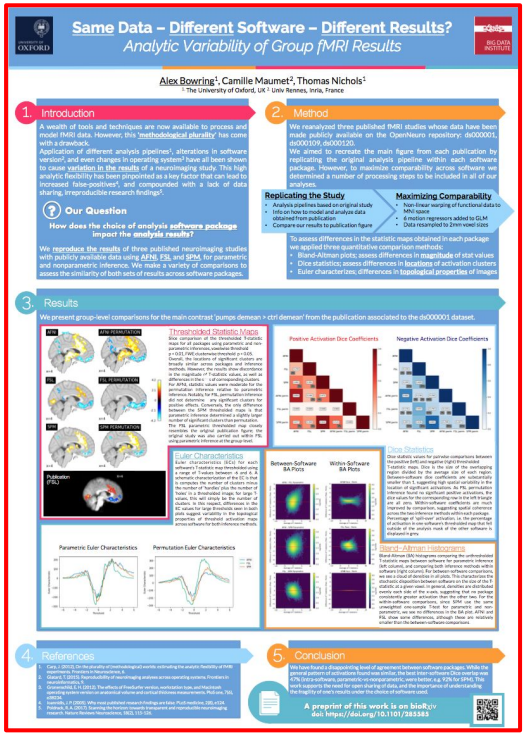


X = Threshold (T value)

Conclusion

- **Large analytic variability of group fMRI results across software packages**
  - Disagreement on significant (de)activation
  - Low dice coefficients
  - ‘Cloud of densities’ in Bland-Altman plots
  - Differences in Euler Characteristic values
- **The measures and metrics introduced here encourage further exploration into software validation**

Thank you!



Poster #2019
Tuesday, Wednesday
Thursday
Standby: 12.45 - 13.45

Email:
alex.bowring@bdi.ox.ac.uk
Twitter: @ABNeuroscience
http://www.nisox.org
#NISOX