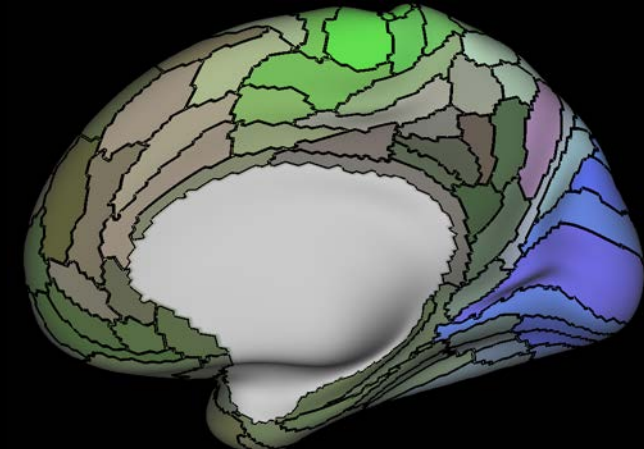
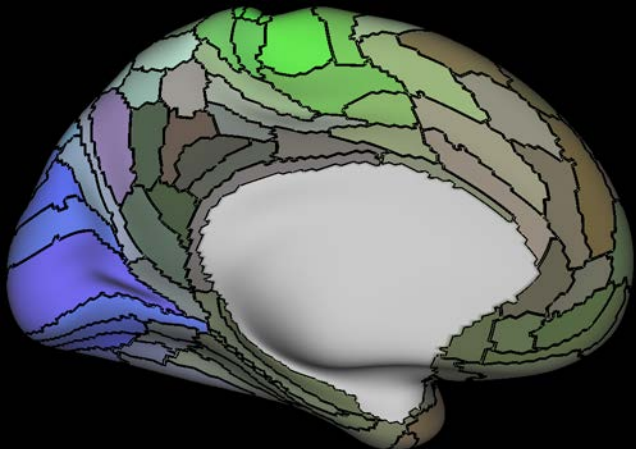


# Multi-modal Classification of Cortical Areas in Individuals and Parcellation Validation

Matt Glasser: Lecture 3 of 3



# Motivation

- In the preceding two lectures, we've learned
  - About the importance of careful preprocessing to preserve the high spatial resolution of the HCP data, even in group averages
  - The multiple non-invasive imaging modalities available in the HCP data and how they can be used for parcellation with a gradient-based approach
- We've discussed the advantages of parcellated analyses
  - And seen a multi-modal parcellation produced using group average HCP data and example parcellated analyses
- Now we'll focus on several outstanding issues
  - What do we do if, despite our best efforts, some cortical areas in some subjects aren't aligned with the group average
  - How do we generate individual subject parcellations including in new subjects not a part of the HCP
  - How do we check that the group parcellation is valid (i.e. that it's reproducible and its areas are indeed different across multiple modalities)

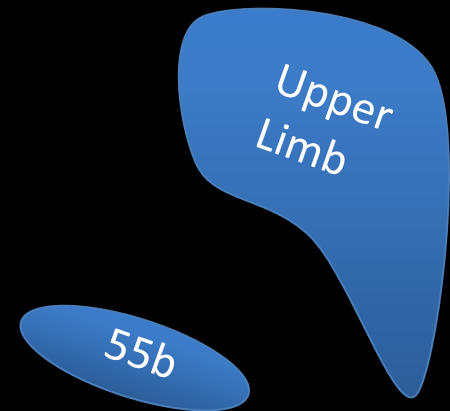
# Lecture Topics

- The problem: residual misalignment of cortical areas even after areal-feature-based registration of some areas in some individuals
- The solution: use a machine learning approach to identify cortical areas in each individual based on their multi-modal areal fingerprints
  - Introduction to Areal Classifier
  - Results of Areal Classifier
- Validation of multi-modal parcellation
- What have we learned from these three lectures?

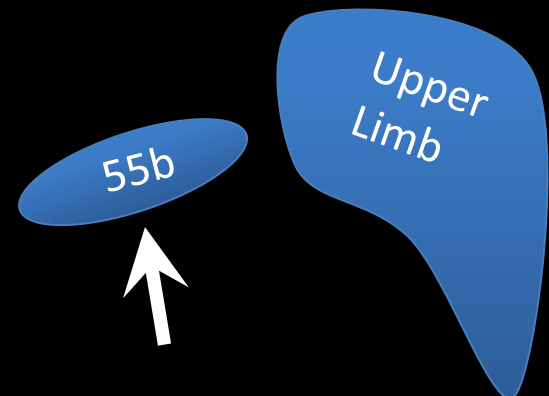
# Example Different Cortical Area Topologies: Shift of 55b (Schematic)

- Most subjects have a particular relationship between area 55b and the upper limb subregion of motor and somatosensory cortex
- A small number of subjects have a very different relationship between these brain regions
- Areal-feature-based surface registration will align the upper limb region preferentially (mainly because it is larger)
  - This leads to an upward shift in 55b
  - If large amounts of distortion are tolerated, registration will attempt to tear the cortical sheet
- Lets look at some real examples...

Typical Pattern

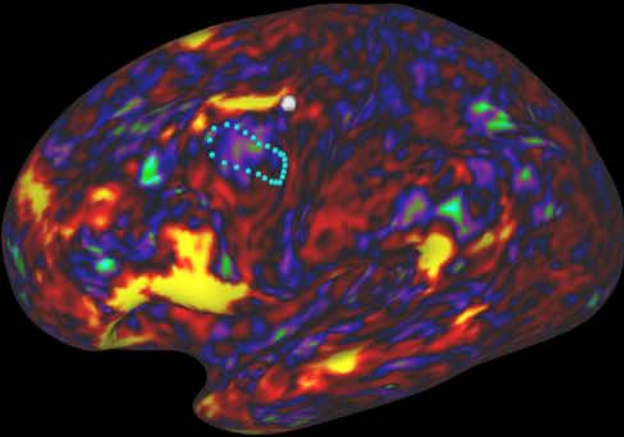


Shifted Pattern

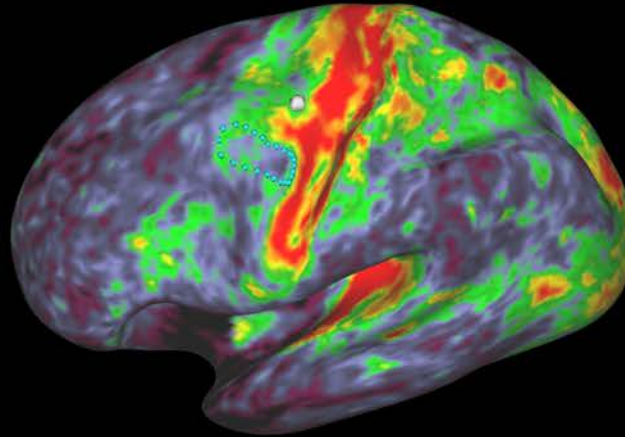


# Example Different Cortical Area Topologies: Shift of 55b (Real Data)

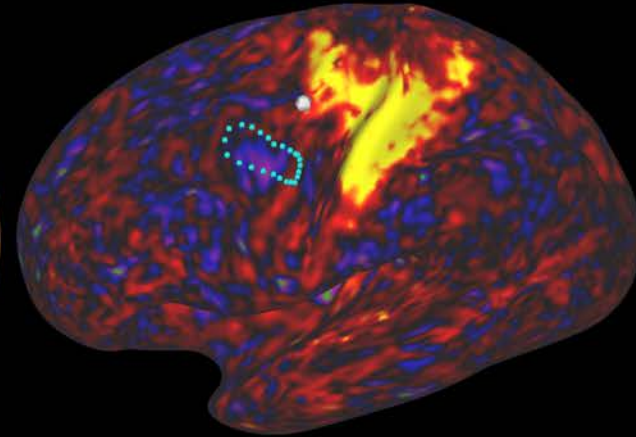
Language Network (ICA d=40)



Myelin Map



Upper Limb (ICA d=40)

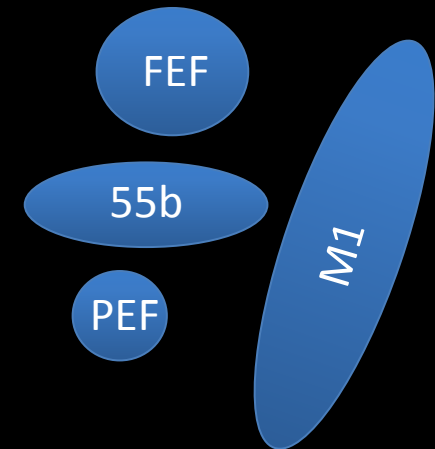


- The typical (average) pattern of 55b (most subjects)
- A typical individual subject
- A shifted individual subject
- Identical twin of shifted individual subject (though twins are not always concordant)

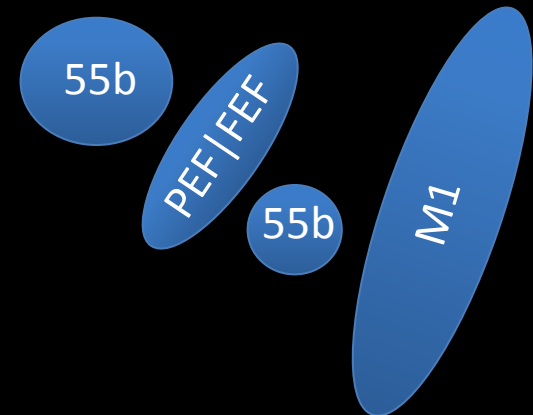
# Example Different Cortical Area Topologies: Split of 55b, Join of FEF and PEF (Schematic)

- Most subjects have a particular relationship between lightly area 55b, heavily myelinated M1, moderately myelinated FEF and PEF
- In some subjects, 55b is split into 2 pieces and FEF and PEF are joined together
- Again, lets consider some real examples...

Typical Pattern



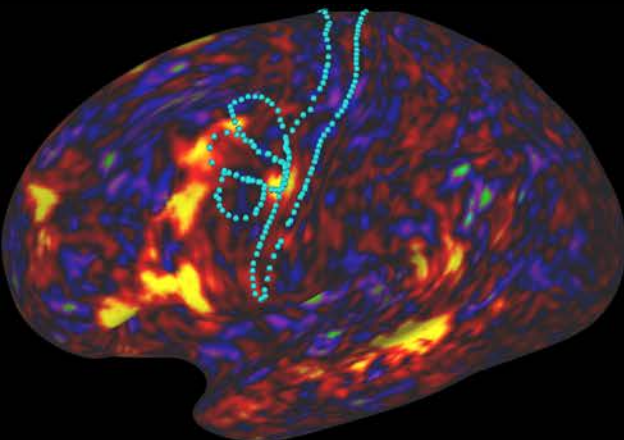
Split Pattern



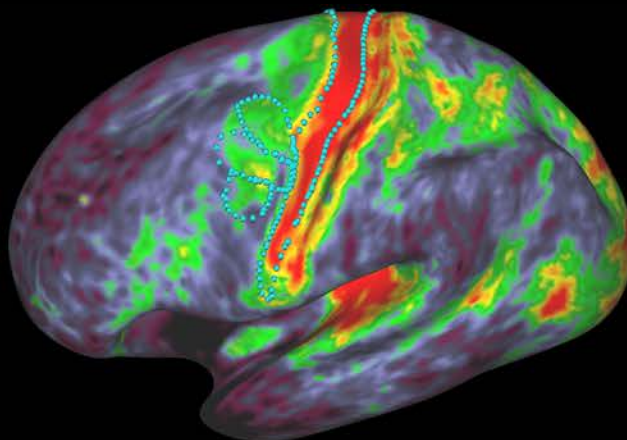


# Example Different Cortical Area Topologies: Split of 55b, Join of FEF and PEF (Real Data)

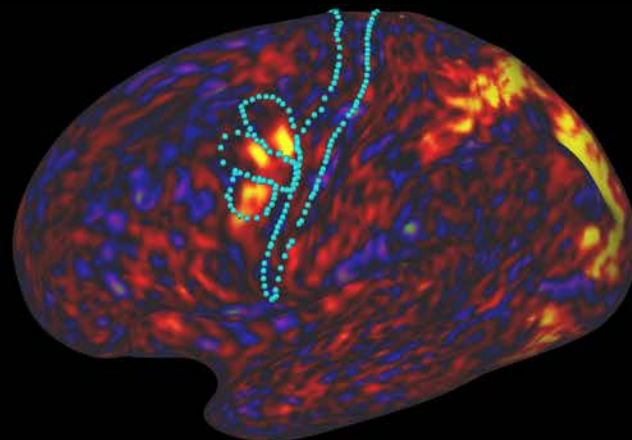
Language Network (ICA d=40)



Myelin Map



Dorsal Stream (ICA d=40)



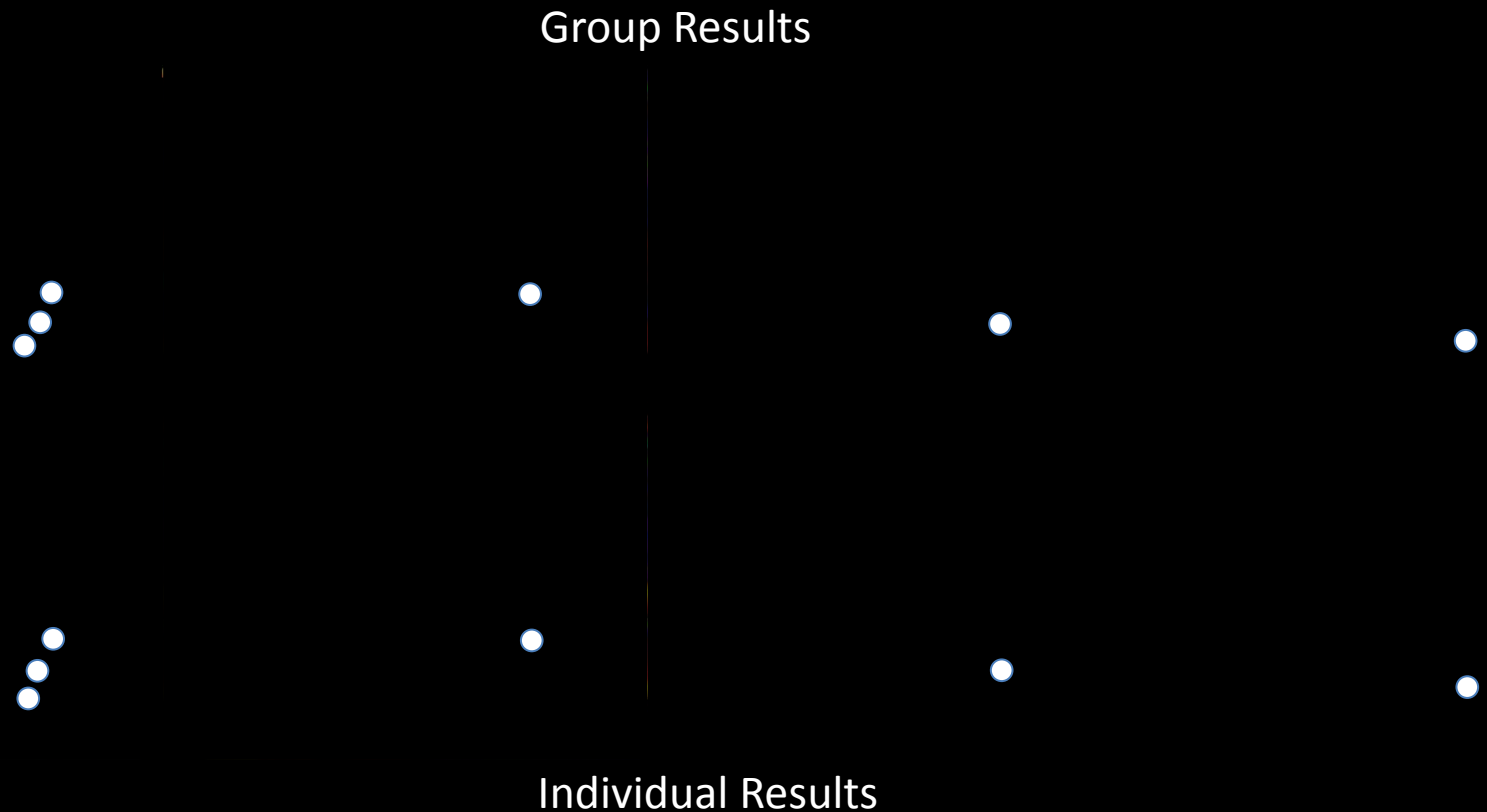
- The typical (average) pattern of 55b (most subjects)
- A typical individual subject
- A split individual subject
  - Note how the myelin and the two RSNs change together
- Another split individual subject

# Different Cortical Area Topologies: Flavors of 55b

- In light of this, is the parcellation of 55b and FEF/PEF correct?
- ~89% of the subjects have the typical topology
  - ~4% have the shifted topology
  - ~6% have the split topology
  - ~1% have something else
- The canonical parcellation should be defined based on the typical topology, but we want to be able to still match alternate topologies as well as possible
- The alternate arrangements are topologically not compatible with the typical topology, so no topology preserving spatial registration can align them
- These topological differences occur more frequently in some parts of cortex than others
  - 55b is a hotspot for this sort of thing
  - As one looks at finer levels of the neuroanatomical hierarchy, such as cortical columns, topological differences likely will become more common



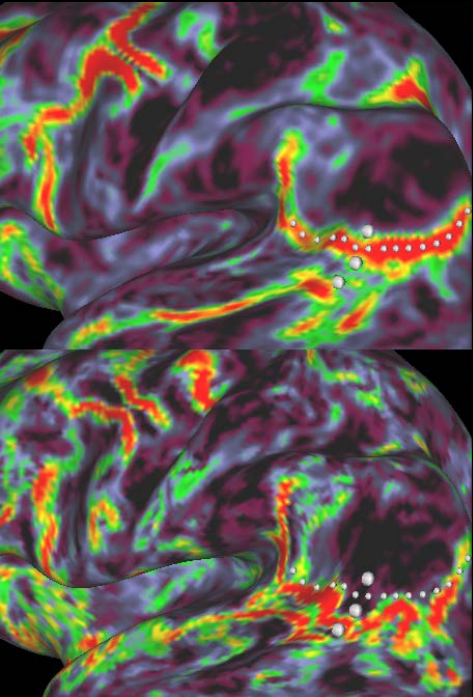
# Another Example Difference in Individual Cortical Area Topology



- Seed 1 looks similar between group and individual
- Seed 2 is very different (in individual it looks like seed 1)
- Seed 3 is similar between individual and group

# Is This Difference Limited To Resting State Functional Connectivity, or Is It Found in an Independent Modality?

Group Results



Individual Results

Gambling Reward

Motor Average

Emotion Faces

- Found in resting state and 3 different tasks scanned on two different days
- Could be related to particular subject's neuro-vascular interaction, but otherwise suggests a topological difference in cortical organization
- As we saw in the previous example, multiple modalities show the same kinds of differences, suggesting these effects are neurobiological

# These Differences Present a Problem For Atlas-based Parcellation

- Generally, atlas based parcellations assume that the surface registration has aligned individuals to the group precisely
  - With a parcellation like Yeo et al (2011) or Gordon et al (2014), one assumes that there are no residual individual differences in brain area locations
  - Individual differences in connectivity, activity, etc could still be due to the brain areas being in different locations
- ICA-based soft parcellations together with dual regression do a better job of accounting for individual variability
  - They fit the parcellation to each subject with the dual multiple regression or more advanced methods like weighted regression
  - However, they are not multi-modal and are not designed to identify brain areas, but rather simply to decompose the data into a set of components that describe the strongest structured signals

# Summary of the Problem

- Not all subjects have the same topological layout of cortical areas, meaning that spatial registration cannot align them
- Area 55b can be shifted or split in a small, but meaningful percent of subjects
- Some areas may even be missing in some subjects
- This presents a problem for atlas-based parcellation approaches
- ICA with dual regression partly addresses this issue, but we would like to be able to create individual multi-modal cortical area parcellations
- Questions about the problem?

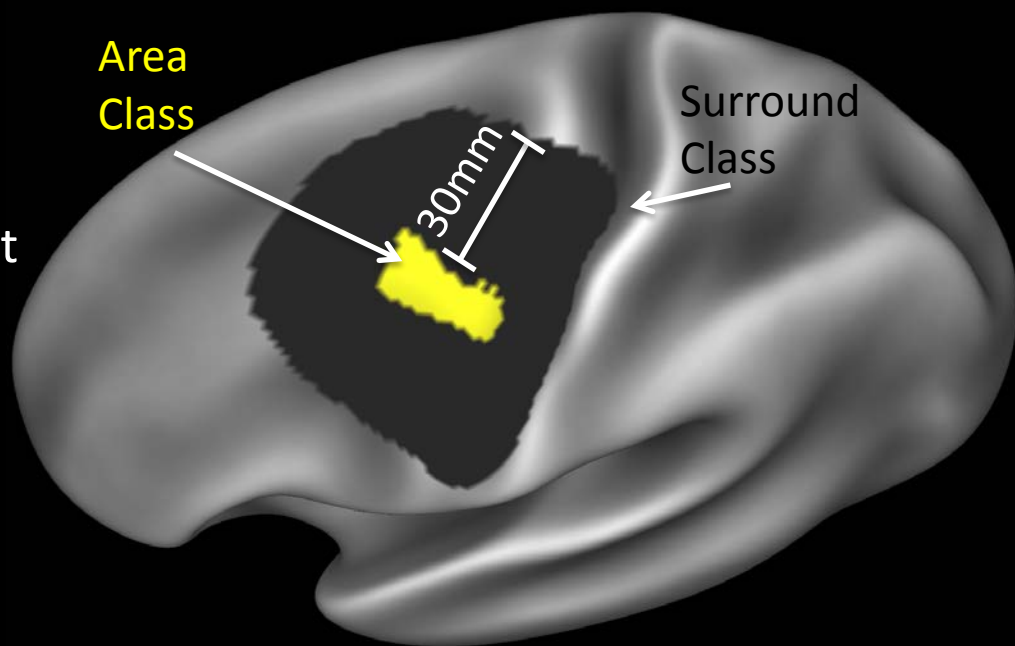
# Lecture Topics

# What Makes a Cortical Area Distinct?

- As we've discussed in previous lectures, a cortical area will have a distinct pattern of one or more of these properties from its neighbors:
  - Architecture
  - Function
  - Connectivity
  - Topography
- These can be combined into a multi-modal areal fingerprint for each cortical area
- Also, we can reasonably make some assumptions about where a cortical area is located (area V1 isn't ever located outside the occipital lobe for example)
  - This helps distinguish some cortical areas that are spatially separated but whose areal fingerprints aren't very distinct (e.g. they are functionally connected, coactivate in task, and have similar myelin content)

# Learning Areal Fingerprints to Delineate and Identify Cortical Areas in Individuals

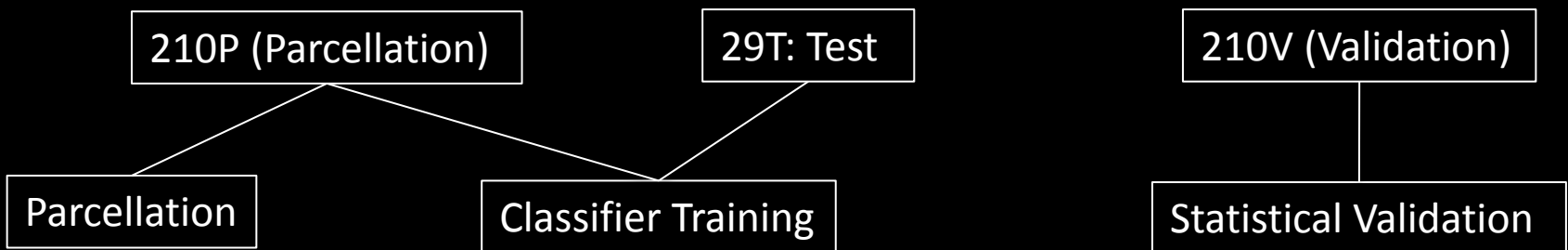
- Recall the discussion of Hacker et al (2013) by Greg this morning to identify 7 resting state networks in individuals using a perceptron
- We can use the same perceptron algorithm they used with a few modifications:
  - Instead of trying to classify 178 areas at once (very hard), classify each area from its surrounding 30mm of cortex (an easier binary problem)
  - 30mm is a reasonable assumption in spite of the residual misalignments
  - We'll combine across classifiers a little later
  - Feed the classifier multi-modal features instead of just functional connectivity maps
  - Use 9 hidden nodes instead of 22 (a binary classifier doesn't need as many)





# Reminder of Study Design

- 210 (P) subjects were used to define the parcellation based on group average gradients across multiple modalities
  - These subjects were also used to train the classifier
- 29 test subjects (can be related to those in the 210P) were used in classifier training to detect overfitting
  - If classifier performance goes down on the 29T subjects test set, during training this suggests the classifier is overfitting to the training set and should be stopped
- 210 (V) subjects were held out for use in all statistical validations (cannot be related to those in the 210P)
  - These subjects played no role making the parcellation or in training the classifier



# What Features Should We Use?

- The classifier will operate on a multi-modal grayordinates X features matrix
- Architecture (Thickness and Myelin)
- Function (Task fMRI):
  - The 86 contrasts have duplicate information and relatively low CNR in individuals
  - Reduce tfMRI dimensionality to 20 while preserving 99%+ of the variance using ICA (on the group tfMRI contrast maps)
  - Also the mean activation map (which ICA doesn't include)
- Connectivity (Resting State fMRI):
  - A dense connectome would be enormous, have a lot of redundant information, and have relatively low CNR
  - Reduce the rfMRI dimensionality to 137 (a local optimum before the ICA appears to start including unstructured noise)
- Topography (Resting State fMRI):
  - Visuotopic maps (processed as described in lecture 2 so that the classifier can use them)
- Artifacts
  - E.g. curvature, maps of veins and signal dropout to help the classifier deal with these issues
- All feature categories are normalized to have the same spatial standard deviation (i.e. a similar starting point)

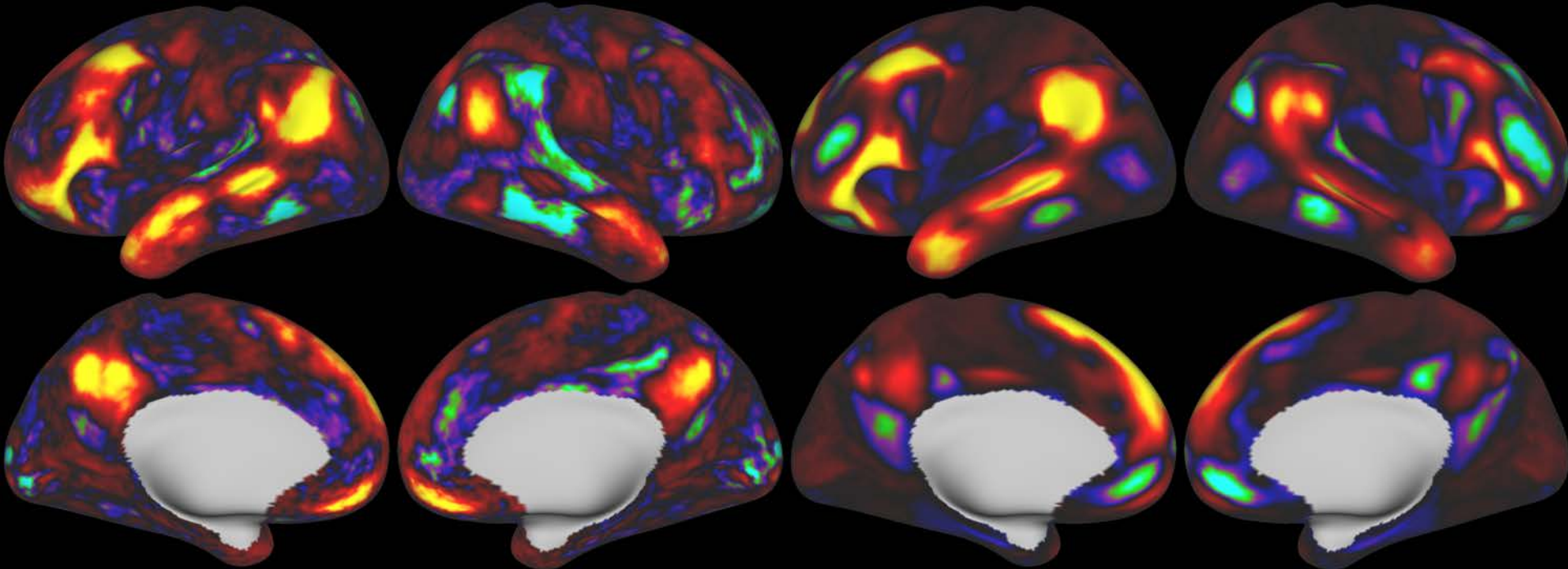
Multi-modal Dense Feature Matrix (one per subject)

Features		Grayordinates
Architecture	Function	
Connectivity	Topography	
Artifacts		

# tfMRI and rfMRI Often Contain Similar Information

tfMRI Contrast ICA (d=20)

rfMRI ICA (d=20)

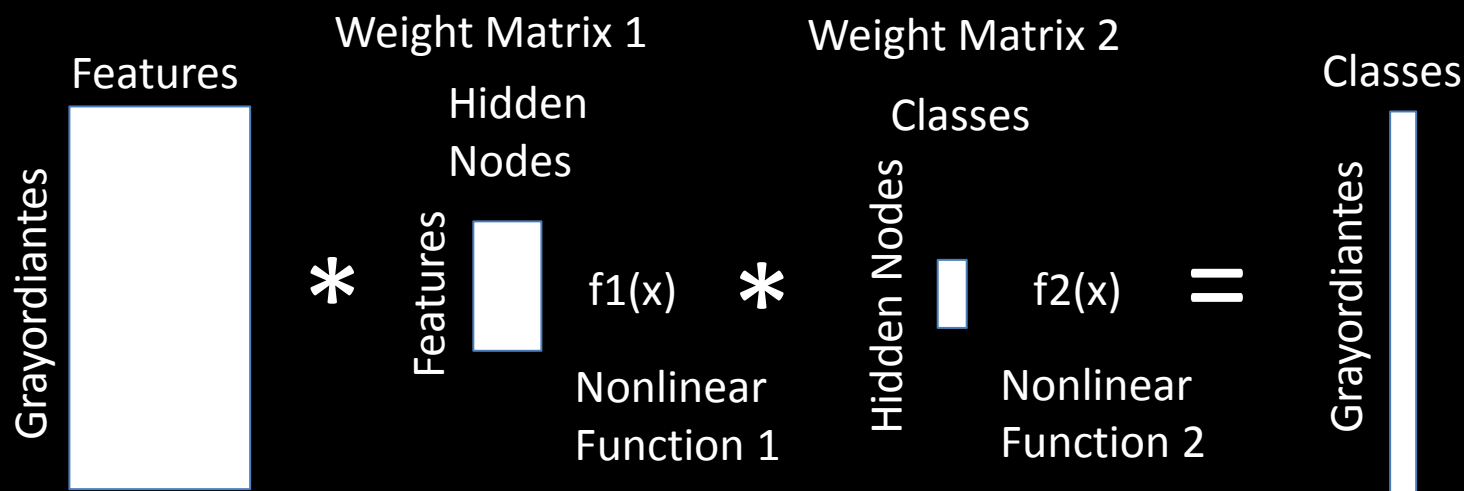


- d=20 ICA was run on the group tfMRI contrast beta maps and compared with d=20 rfMRI ICA
- The tfMRI ICA components often look very similar to the rfMRI ICA components
- Reminiscent of Smith et al 2009, but on much higher quality data
- This may mean that the multi-modal areal classifier will work without the tfMRI data as the information tfMRI provides about areal boundaries appears to be largely duplicated in the rfMRI
  - Not yet tested explicitly, so stay tuned...

# What Training Labels Do We Use?

- We haven't yet defined the individual subject areas, so we need to use the group areal definitions
- We use the group areal definition and the 30mm surrounding the group areal definition as the training labels
- Some areas in some subjects are not aligned with the group areal definitions
  - A short initial run of the classifier detects subjects where the key features for an area differ substantially from the group and then excludes them for that area's classifier training
- Importantly, we don't exclude these subjects when we apply the classifier and we can usually still detect the areas in these often misaligned, excluded subjects
  - The goal is just to make the group training labels more valid for the training and test datasets during classifier training
- In each training iteration, the classifier output is compared with these training labels and the error is back propagated into the classifier according to the learning rate

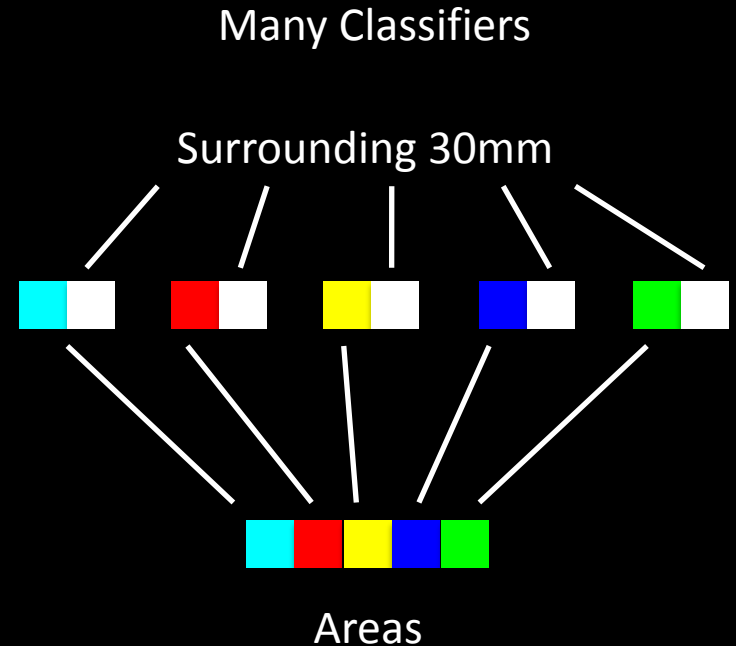
# The Underlying Equation of the Classifier After Training



- In this perceptron, training involves figuring out the values for the two weight matrices so we can distinguish between the area and the 30mm surrounding it
  - So that a particular fingerprint will give high probability for the area and anything else will give high probability for the surround
- Note that the weighting matrices have no idea where the cortical area is (they have no spatial dimension), they just know the pattern in feature space
- The nonlinear activating functions and two weighting matrices allow the classifier to learn non-linear patterns (e.g. XOR)
- Once trained, applying the classification is as simple as running the data through the above equation (a grayordiante's multi-modal feature profile is turned into an areal probability)

# Combining Across Classifiers

- Because we trained a separate classifier for each cortical area, we must combine across classifiers to make a parcellation
  - Each classifier contains an area (colored) and a 30mm surround (white)
- We find the highest area probability for each vertex for an initial hard segmentation
- Then we regularize this segmentation, filling in holes, joining discontinuous pieces of areas, and removing very small islands (less than  $25\text{mm}^2$ )
- The result is the individual subject's parcellation



# What Does Successful Classification of An Area Tell Us?

- If we can successfully classify a cortical area across subjects, this means the area has a distinct pattern of multi-modal features
  - The perceptron could distinguish it from its neighbors across most individuals
- If we are unable to classify the area, this suggests it does not have a unique pattern and may not actually be correctly parcellated
  - There may be a subdivision in the parcellation that is not well enough supported by the data across individuals
- Additionally, the classifier may identify grayordinates that are not of high probability for any cortical area
  - If these form a substantial “hole” in the parcellation (rather than just following the boundaries of areas), a subparcellation may be required
  - Low probability can also occur in regions of high susceptibility artifact
- Thus initial training runs of the classifier on the 210P data were helpful in refining the multi-modal parcellation to address both these issues



# Summary of Areal Classifier Methods

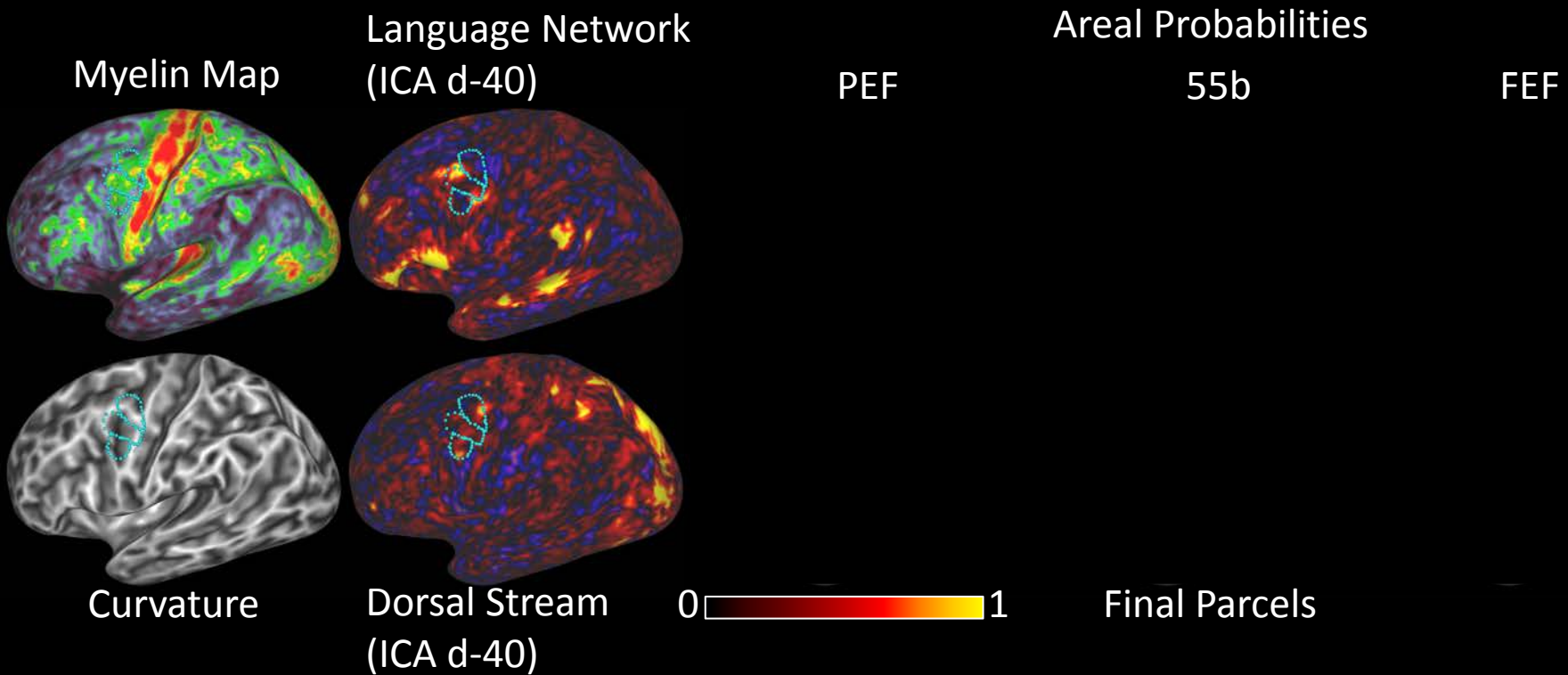
- Areal features like architecture, function, connectivity, and topography make areas distinct
- We want to train a perceptron to recognize the pattern associated with each cortical area as distinct from its surrounding 30mm
- We use dimensionality reduction for tfMRI and rfMRI features to focus on only the unique information and improve to CNR
- We use group training labels and train the classifier to find the values of the two weighting matrices that produce a good separation of the areal and surround classes
- We combine the area classes across classifiers to make the individual subject parcellation, assigning each vertex to the highest probability areal class
- Successful classification indicates that the area is indeed distinct from its neighbors.
- Questions about classifier methods?

# Lecture Topics

# Classifier Results: Outputs

- A probability map for each area (from each classifier) in each subject
- A combined (across classifiers) and regularized individual parcellation for each subject
- Cross-subject probabilistic maps of the location of each area
- Group maximum probability maps from probabilistic areas
- Detection rate for each area
  - Was an area found in the individual within 3x or 0.33x the size of the group area?
- Which features were used by the classifier

# Does the Classifier Identify Misaligned Areas?

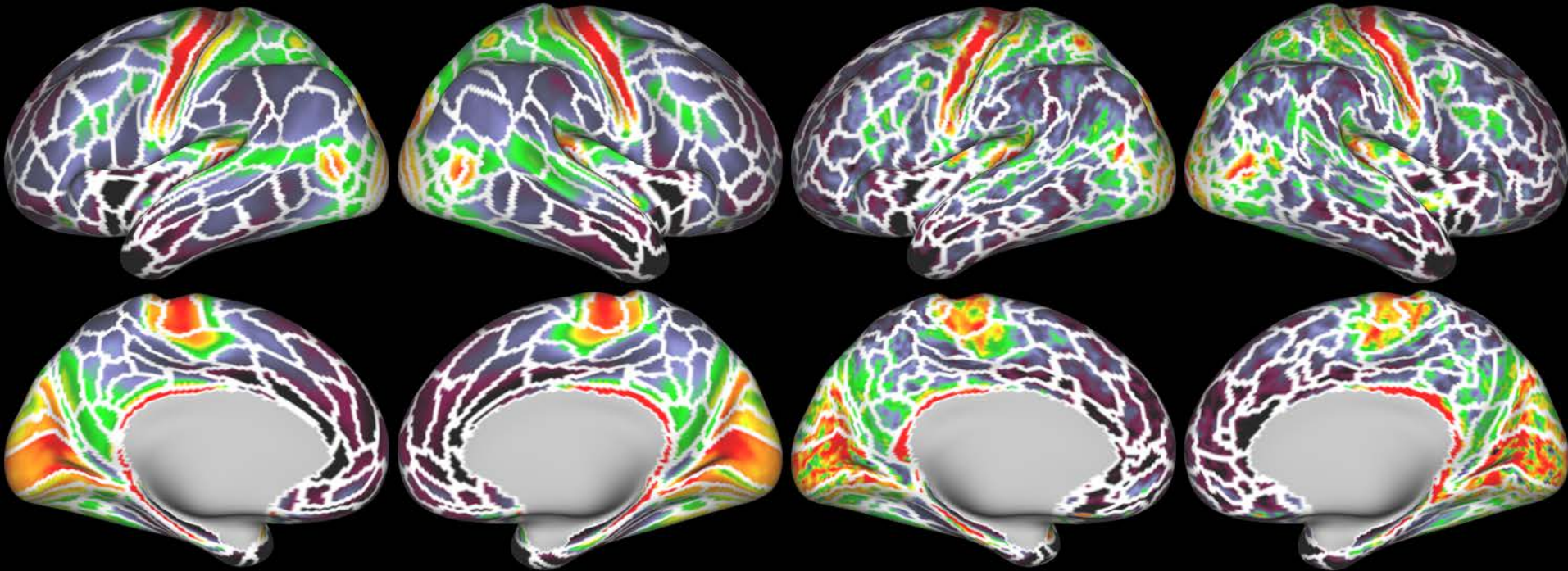


- Individual Areal Features + Folding
- Initial Classifier Probabilities
- Final Parcels (after regularization)
- Typical subject's 55b, PEF, and FEF
- Shifted subject's 55b, PEF, and FEF (55b and FEF swap)
- Split subject's 55b, PEF, and FEF (PEF and FEF are adjacent)

# Original Group Parcellation and Individual Regularized Area MPMs

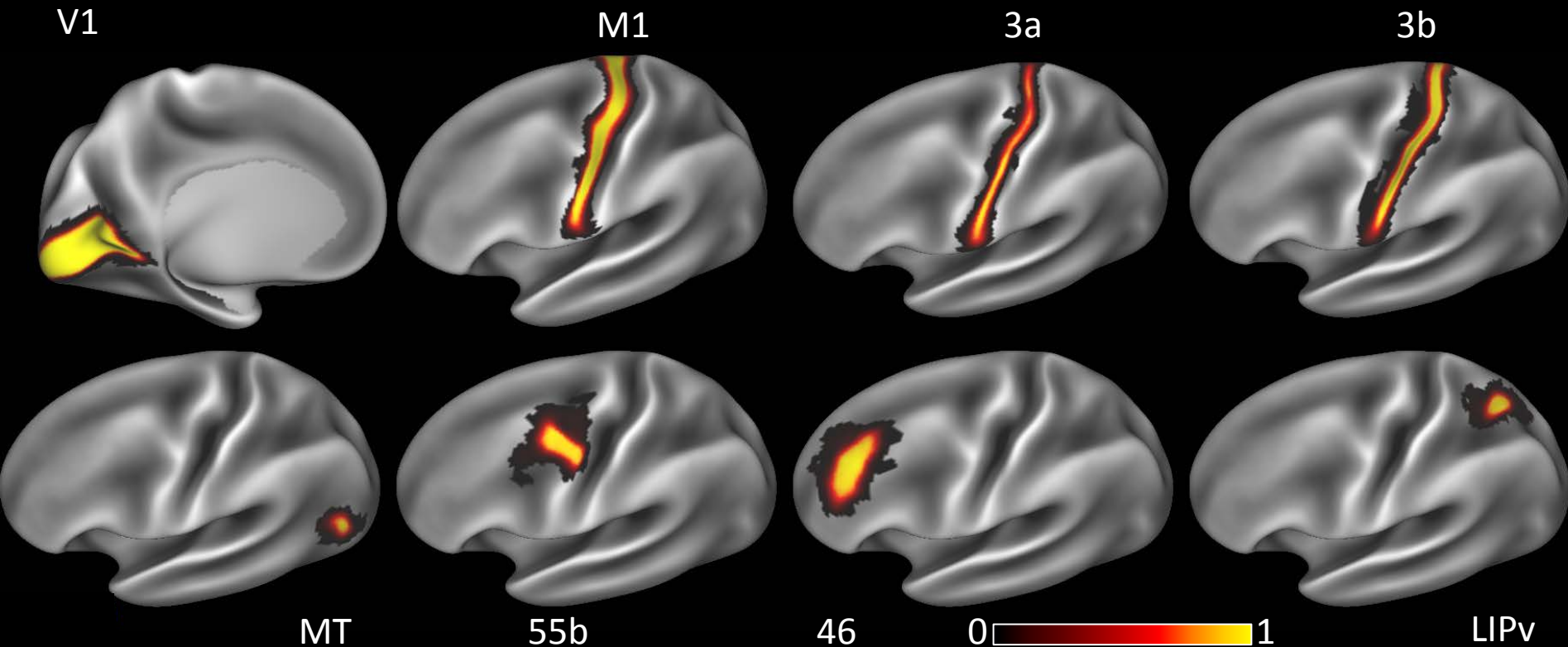
Group Parcellation

Individual Parcellation



- Subject 1
- Subject 2
- Subject 3

# Areal Probabilistic Maps

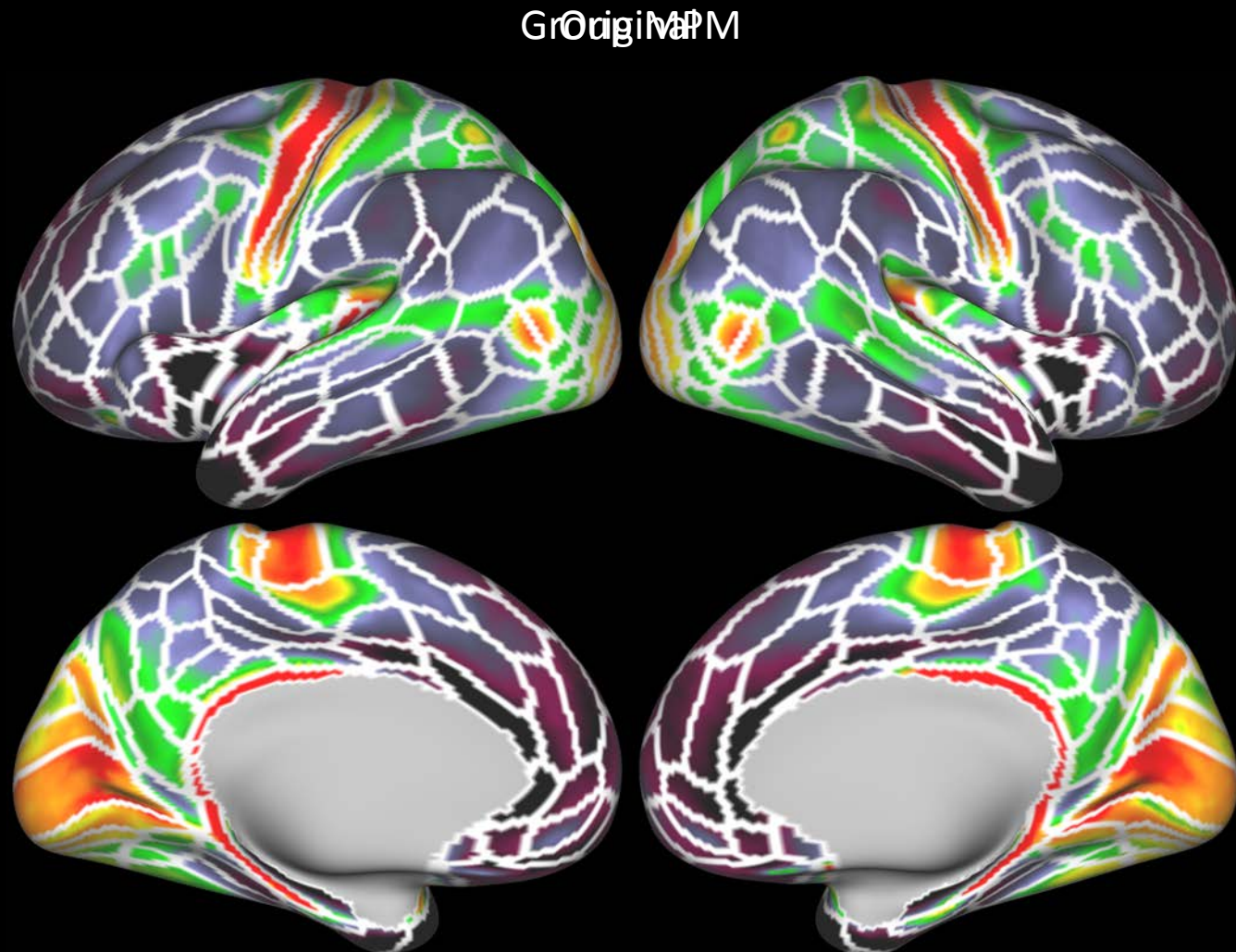


- Some areas have little variability in spatial location
- Others have more
- Recall from lecture 1 though that MT had only 50% overlap from the folding-based registration, whereas it reaches 100% here



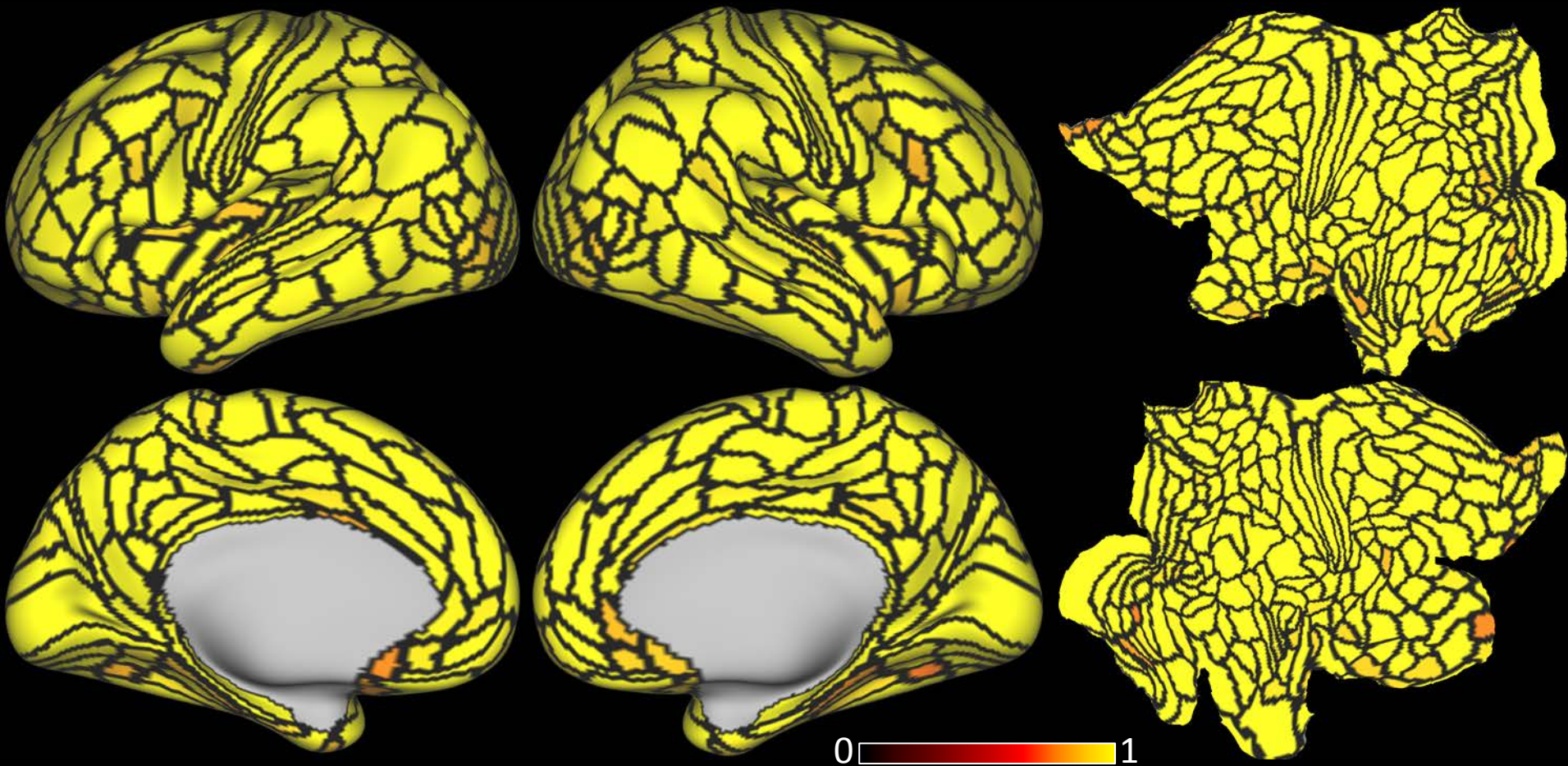
# Original Group Parcellation and 210P Group Average MPM Parcellation

- Original parcellation and group MPM parcellation are not identical, but very similar
- Group MPM parcellation may be better “fit” to the data than the currently hand drawn original parcellation
- We’ll use the group MPM parcellations for any statistics





# Assessing Classifier Performance: Areal Detection Rate in 210P



- Most areas are found in most subjects (between 0.33 and 3x the surface area of the group parcel)

# Determining Which Features Were Used By the Classifier

- One cannot simply invert the classifier equation to find out which features are being used because of the dimensionality reduction (??? Features to 2 classes)
- Instead, one can compute the partial derivatives with respect each feature of each area during the forward propagation (application) of the classifier
- Higher (in absolute value) partial derivatives indicate that the classifier uses the feature more (after appropriate normalizations)

# Determining Which Features Were Used By the Classifier

# Summary of Areal Classifier Results

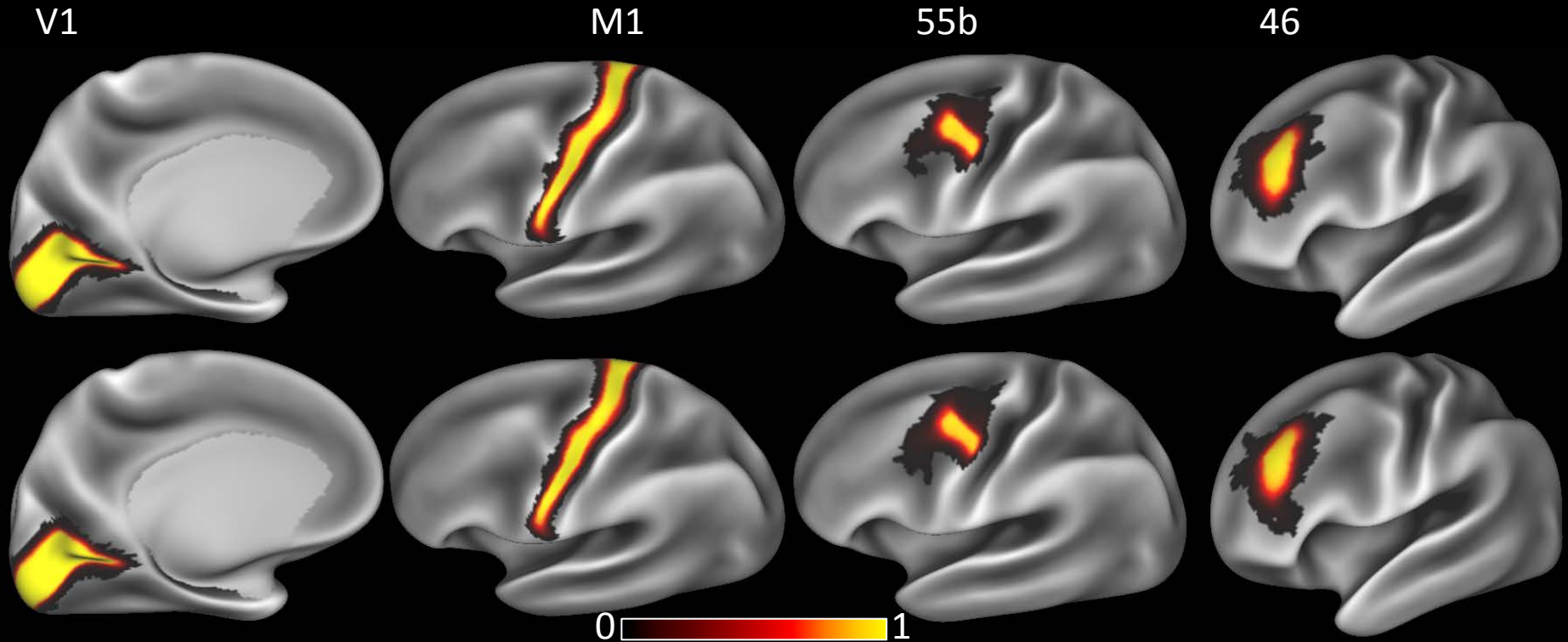
- The areal classifier is able to detect areas based on their multi-modal areal fingerprints, even if they are misaligned with the typical subject pattern
- Individual subject MPM parcellations are different from each other and the group parcellation
- Some areas have very tight probability maps, others are more diffuse
- The group MPM parcellation is very similar to the original hand drawn parcellation
- Areas are detected by the classifier at a high rate in individual subjects
- The classifier can tell us which features it used
- Questions about areal classifier results?

# Lecture Topics

# Statistical Validation of Parcellation

- Already saw that a machine learning classifier could be trained to identify cortical areas in individuals, suggesting that the areas really are different
- How reproducible is the parcellation in an independent validation group of 210 subjects that were not used to generate the parcellation or train the classifier?
- Can one reliably detect the cortical areas in the validation dataset using the areal classifier?
- Do the cross-parcel boundaries in the validation set have large and statistically significant differences in multiple modalities?

# Reproducing the Multi-modal Parcellation Using Only Areal Fingerprints: Probabilistic Maps

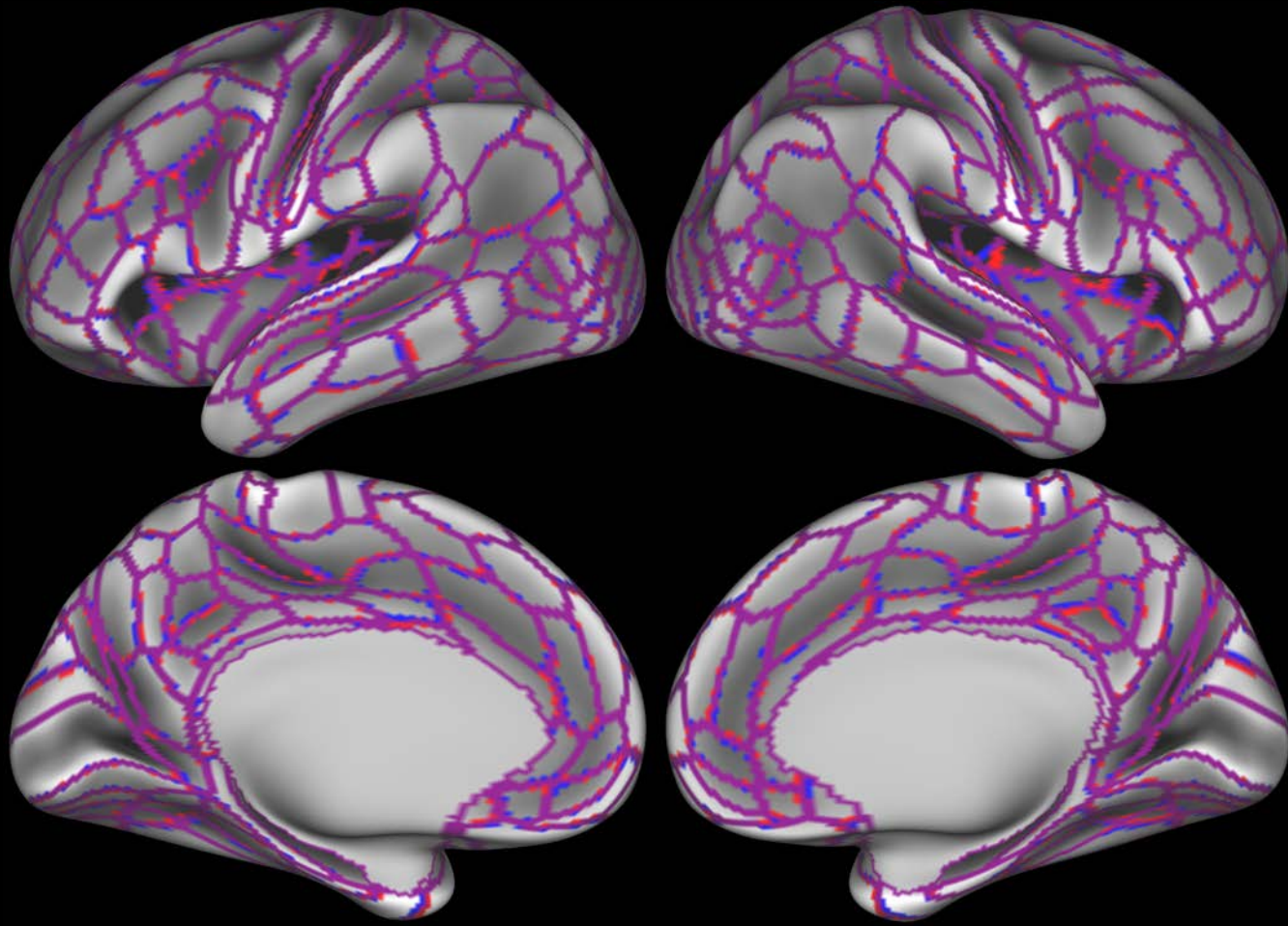


- The trained classifier was applied to the 210P and 210V datasets to generate individual subject parcellations
- These parcellations were averaged across subjects to produce probabilistic areas
- The probabilistic maps are very similar across the two groups

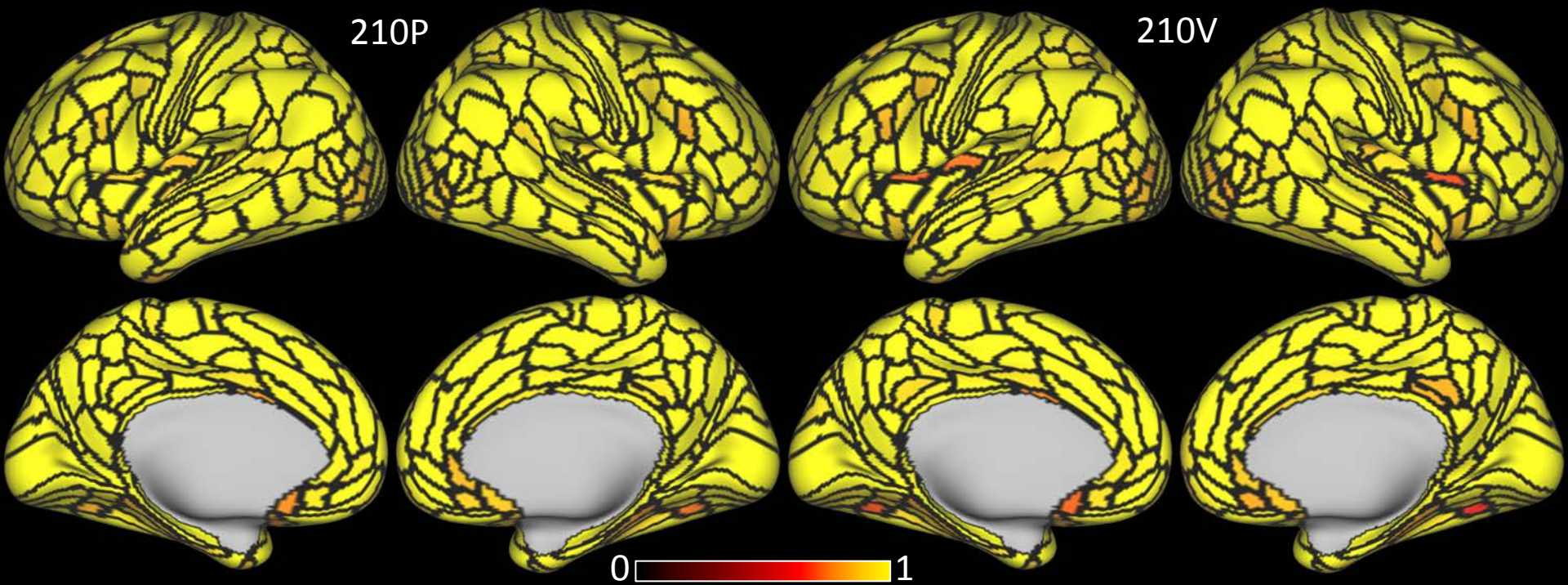


# Reproducing the Multi-modal Parcellation Using Only Areal Fingerprints: MPMs

- The group maximum probability map was computed for both groups
- The areal boundary grayordinates displayed
  - Blue for 210P
  - Red for 210V
  - Purple for both
- The boundaries are in very high agreement
- Correlation of these parcellations is 0.97
  - This is in line with the dense map reproducibilities we saw in lecture 1



# Comparison of Areal Detection Rates in 210P and 210V

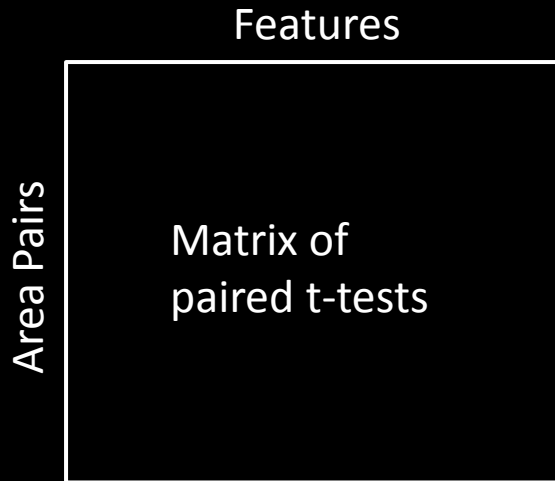


- 210V has very similar areal detection rates as 210P
  - Still most areas in most subjects despite not having been used in the parcellation or classifier training

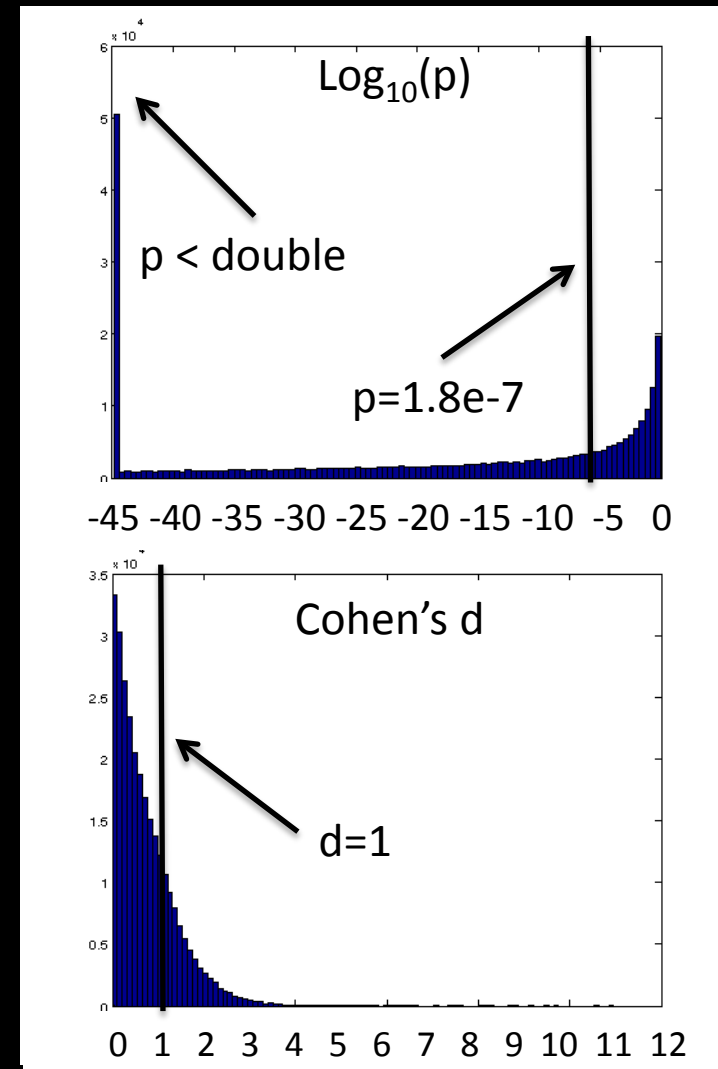
# Are the areal features different across parcel boundaries?

- Find all pairs of areas that are spatially adjacent (i.e. share links on the surface)
  - 178 areas means 523 area pairs (~6 spatial neighbors/area)
- Find some useful features to compare across areas
  - Mean Myelin (1), Thickness with folding removed (1), tfMRI (86), rfMRI Full Connectome (178) = 266 Features
- For each area pair and each feature, compare the means in the two areas across 210V subjects using a paired t-test
- $523 * 266 = 139,118$  comparisons
  - Might seem like a lot of comparisons, but there are 228,483 voxels in the 2mm MNI space brain mask...
  - $0.05 / 139,118 * 2 = 1.8 * 10^{-7}$  Bonferroni corrected p for a two tailed test

# Spatially Adjacent Parcellated Areas Are Very Different in These Features

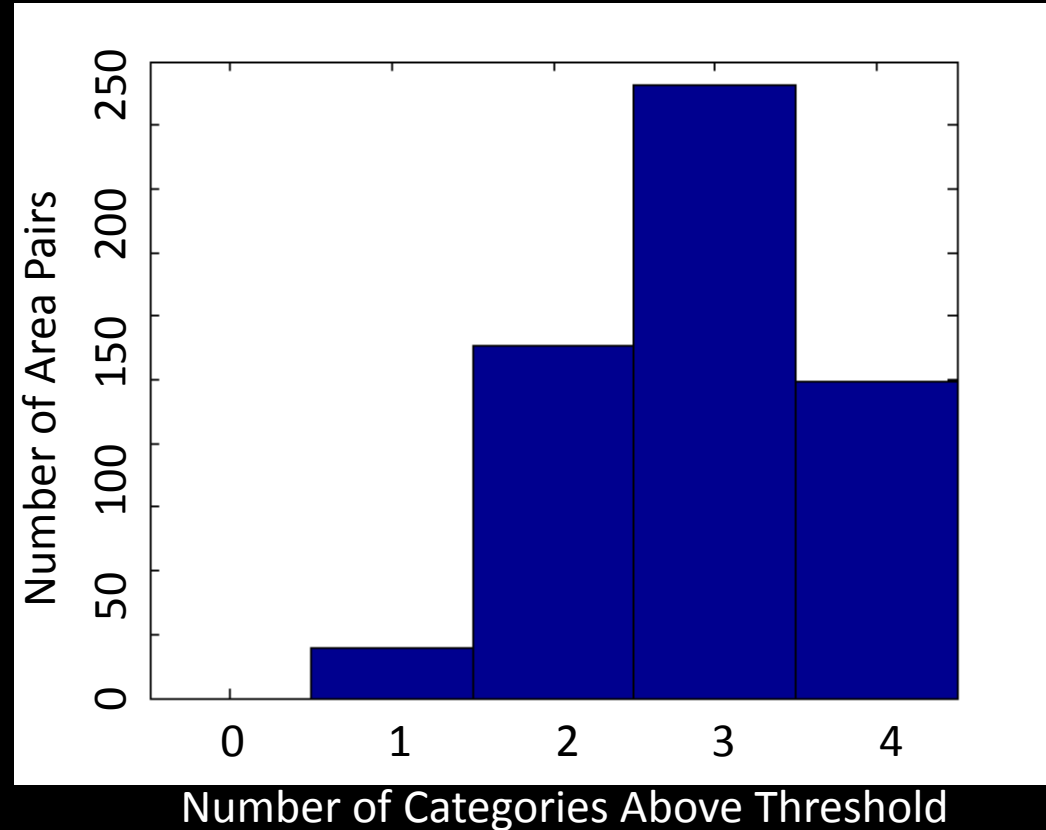


- 64% of the area pairs X features matrix is significant ( $p < 1.8 \times 10^{-7}$ )
  - 18%  $p < \text{double precision float}$
- 28% of the matrix has a very large effect size ( $d > 1$ )



# Spatially Adjacent Area Differences According to Four Independent Feature Categories

- Threshold of  $d > 1$  and  $p < 1.8 * 10^{-7}$  for
- Myelin
- Thickness
- rfMRI
- tfMRI
  - Though technically each task could be called independent



- Most neighboring area pairs have very large & significant differences across more than one category/modality



# Example Area Table: Right Area 46 in “DLPFC”

Border Area	# of Edges	Thickness	Myelin	tfMRI	rfMRI
Area	Edges	Thick	Myelin	tfMRI	rfMRI
9-46d	55	+	-	+	* (53)
p9-46v	42	+	*	* (35)	** (67)
8Av	15	+	*	* (15)	*** (109)
a9-46v	15	+	+	* (2)	** (95)
ifs1	12	-	*	* (26)	*** (32)
8Ad	1	+	*	** (28)	**** (87)

- # of Edges: Number of surface links along areal border

(-) Not Significant

(+) Significant  $d < 1$

(\*) Significant  $d > 1$

(\*\*) Significant  $d > 2$

(\*\*\*) Significant  $d > 3$

(\*\*\*\*) Significant  $d > 4$

# Summary of Parcellation Validation Results

- The areal probabilistic maps are very similar across the 210P and 210V groups
- As a result, the group MPMs are highly overlapping and the parcellations are highly correlated
- The areal detection rate is very similar between the two groups
- The areas have large and statistically significant differences across their boundaries in multiple modalities
- Questions about parcellation validation?

# Lecture Topics



# From Raw Multi-modal Data to Aligned Dense Multi-modal Data

- Raw data has a variety of issues that the HCP's minimal preprocessing pipelines are designed to fix
- The HCP's preprocessing methods (e.g. CIFTI grayordinates) are designed to maintain the high spatial and temporal resolution of the original data
- And at the same time aligning brain areas across subjects and studies (MSMAll, DEDRIFTING)
- Very high reproducibility of even fine details in two independent HCP datasets
- Why not take advantage of our approach in your studies?

# From Dense Multi-modal Data to a Group Average Cortical Parcellation

- The HCP's rich multi-modal dataset has offered an excellent opportunity to parcellate living subjects with a multi-modal approach
- Brain areas can be defined non-invasively based on transitions in architecture, function, connectivity, and topography
- Cortical areas have a wide variety of shapes and sizes
- The last step of parcellation is to try to identify and name the areas according to the extant literature
- Parcellated analyses have many advantages (e.g. simplicity, sensitivity, and power) that you might like to capitalize on

# From a Group Average Parcellation to Individual Subject Parcellations

- Despite our best efforts at registration, some cortical areas in some subjects have different topological relationships from the typical pattern
- A machine learning classifier approach can successfully classify cortical areas based on their multi-modal fingerprints
- It can be used to make individual subject versions of the HCP's multi-modal parcellation
- The parcellation is reproducible across independent groups of subjects
- The cortical areas have large differences in areal features across their borders

# What Kind of Data Do I Need to Acquire to Be Able To Make Use of These Methods?

- Short answer: The same data as you need for the HCP's minimal preprocessing pipelines
- Longer answer:
  - High resolution T1w and T2w images
  - Plenty of fMRI data (fast TR and long duration)
    - tfMRI data likely not essential for classifier, containing largely a subset of the rfMRI information for areal borders
    - If a study is not rfMRI focused but still collects plenty of fMRI, tfMRI, movie fMRI can likely be analyzed as rfMRI for the purpose of MSMAll and areal classification
    - Not yet known the minimum amount, but some analyses have shown that at least 30 mins is required to get somewhat stable connectivity measures and up to 80 mins is beneficial
  - Field Map (spin echo preferred)

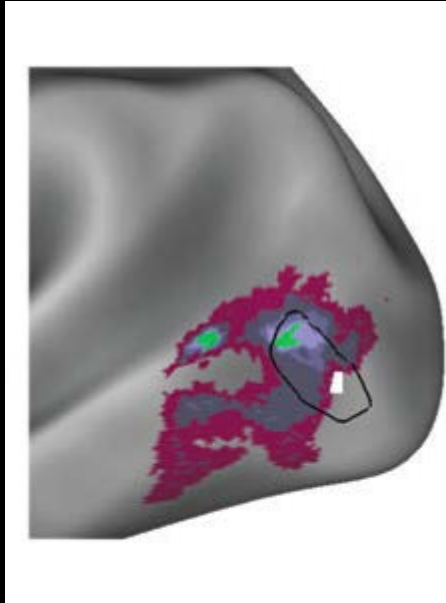
# Volume-based vs Grayordinates-based Analysis Paradigms—an Astronomical Analogy

8 Meter Ground-based Telescope

Dalcanton 2009, Nature

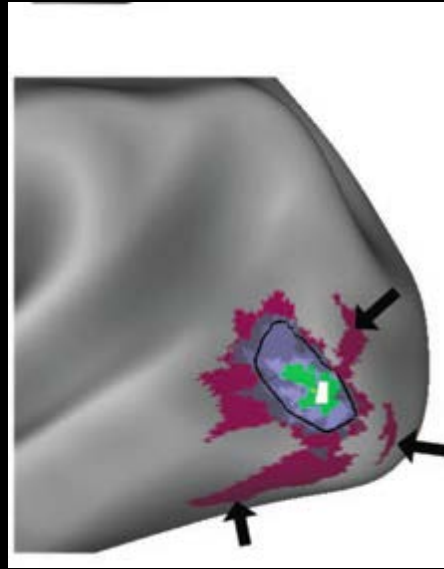
- Ground-based telescopes (even with large mirrors), produce blurry images because of the effects of the earth's atmosphere
- Similarly, when averaged across-subjects, volume-based cortical analyses are blurry because of misalignment of cortical areas
- Space telescopes produce sharp images, because the lack of atmosphere gives them an inherent advantage
- Grayordinates-based analyses also have an inherent advantage as they can better align cortical areas and produce sharper results

# A Final Example of What Can Be Gained: Areas MT and MST

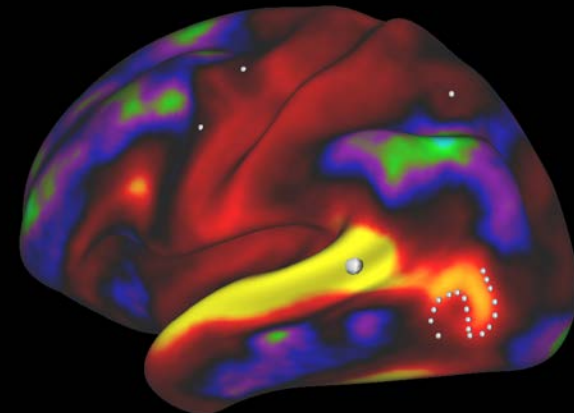
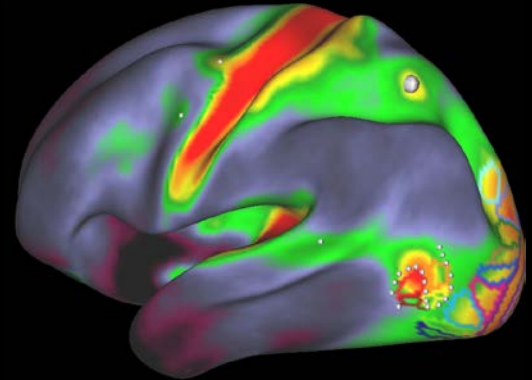
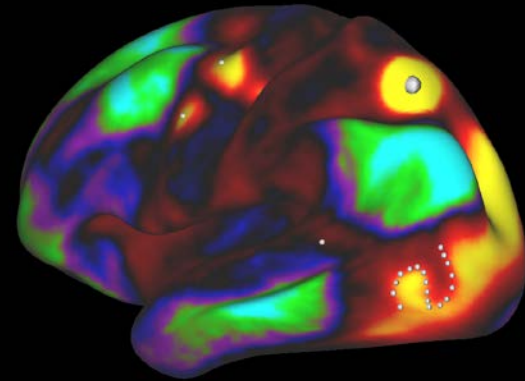


Volume-Based Analysis

Van Essen et al 2012



Surface-based Analysis  
with Folding-based  
Alignment



Surface-based Analysis with Areal-Feature-based  
alignment and dedrifted across studies and  
modalities

# Acknowledgements

## Washington University:

### *Van Essen Lab:*

David Van Essen

Tim Coalson (algorithms)

John Harwell (visualization)

The HCP data collection and informatics team

## University of Minnesota

Gordon Xu

Steen Moeller

Eddie Auerbach

Essa Yacoub

Kamil Ugurbil

## Oxford University

Emma Robinson

Mark Jenkinson

Steve Smith

## University of Parma:

Guy Orban

Rouhi Abdollahi

## Emory University:

Jim Rilling

Todd Preuss (\*)

Longchuan Li



Any Last Questions?