CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN
International Graduate Program Medical Neurosciences

Spatially Constrained Clustering of Voxels Based on Cross-Individual Consistency

stephan.mauermann@charite.de, 01.11.2016
Supervisor: Dr. Matthias Guggenmos

The Rationale I

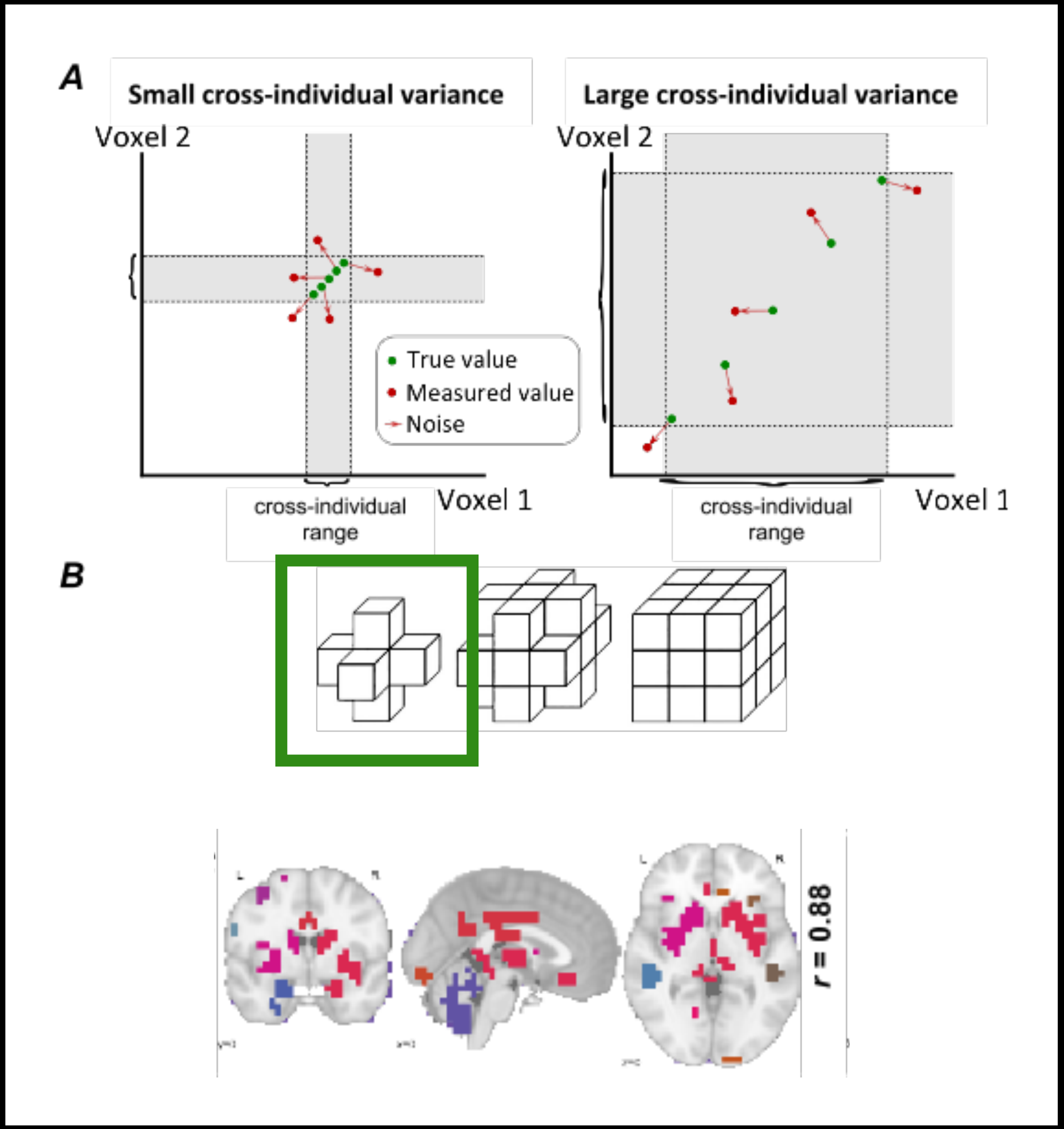
- MRI data is highly complex
- curse of dimensionality => data compression
- conventional:
    - manual definition of regions of interest (ROI)
    - brain atlas-based parcels
    - occluding informative signal, poor fit to data

The Rationale II

- unbiased data-driven reduction of complexity
- good fit to data, capture informative signal
- implementations:
  - feature selection via PearsonMerger (PM)
  - brain parcellation via Spectral Clustering (SC)

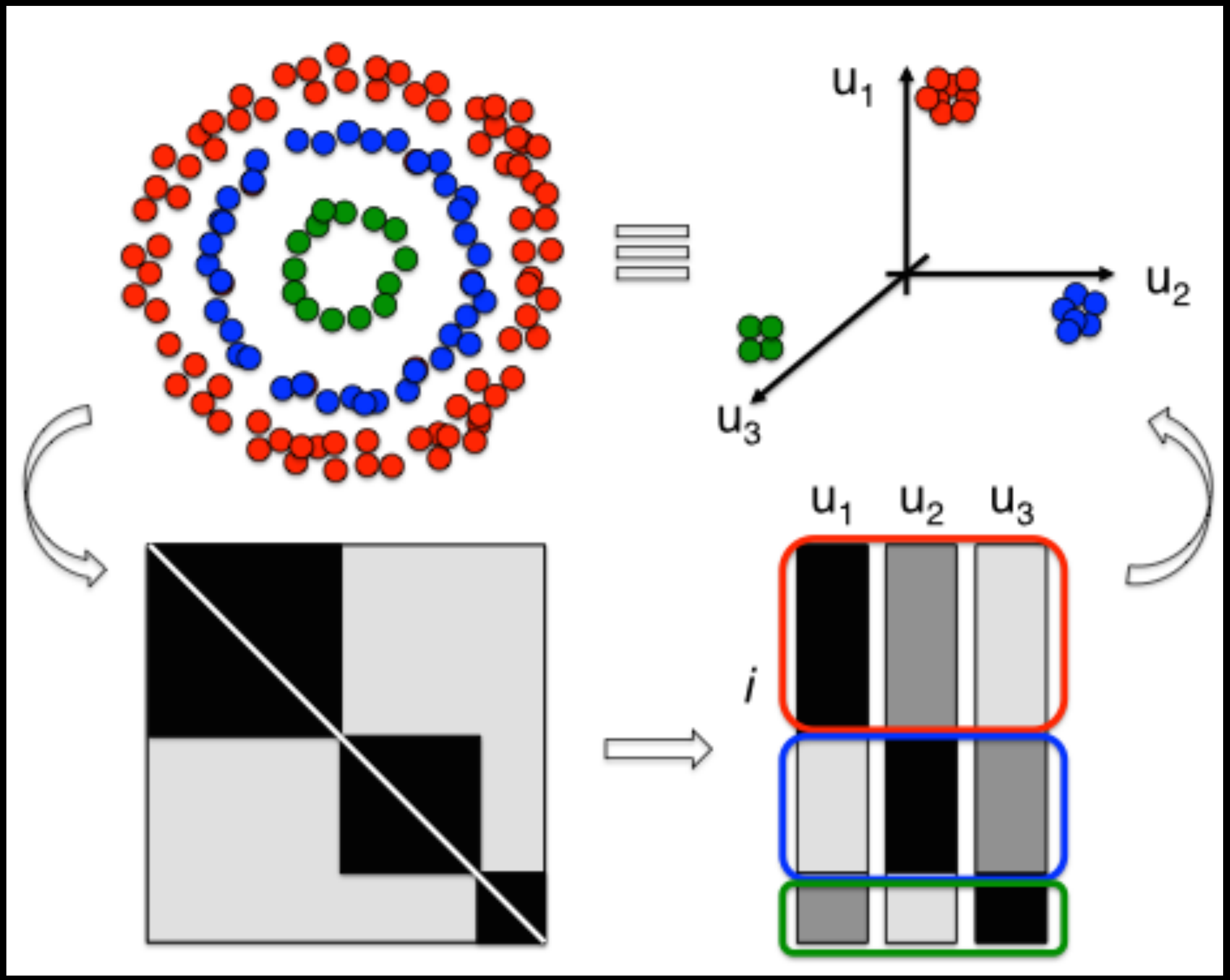
PearsonMerger

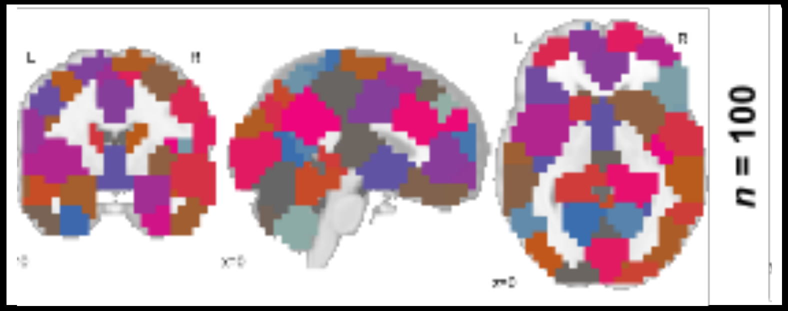
- Pearson correlations of:
  - adjacent voxels
  - across subjects
- merges voxels with:
  - <u>supra threshold correlation</u>
  - <u>large cross-individual variance</u>
- favorable for classification



Spectral Clustering

- uses spectral graph theory:
  - transforms image into graph representation
  - solves the eigenvector problem on Laplacian
  - clusters voxels in eigenspace via k-means
  - clusters all elements!





Testing Ground

- predicting diagnosis of alcohol use disorder (AUD)
- LeAD study: grey matter volume estimates
- classification:
    - averaged voxel values within ROIs/clusters
    - Matthias’ WeiRD algorithm
    - features vote, weighted sum of all votes

PearsonMerger

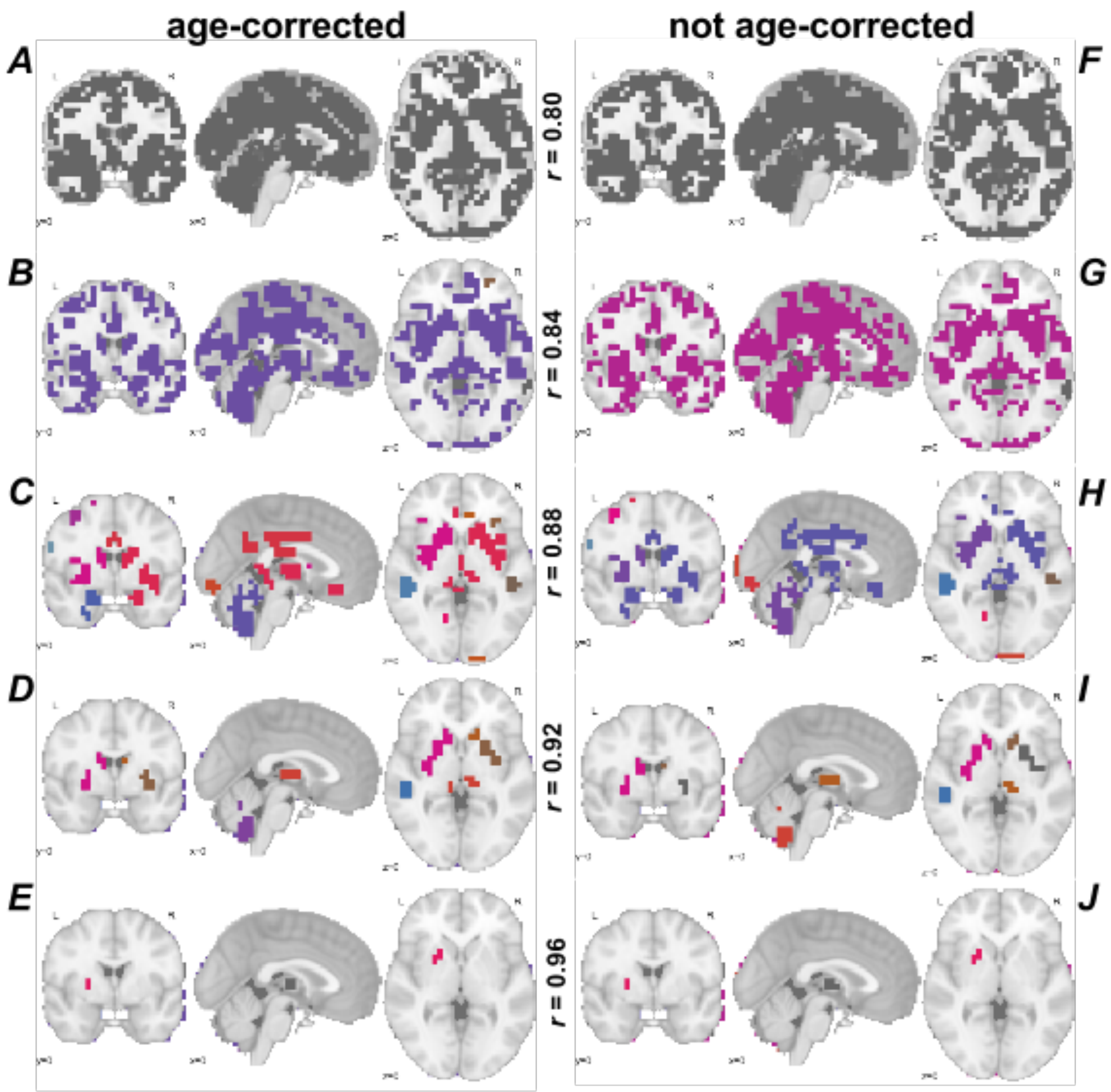


Figure 4| PearsonMerger clusterings at varying Pearson correlation thresholds based on all participants. Images were normalized to zero mean and age-corrected if indicated. **A)** Threshold of *r* = 0.80: one cluster was detected with size of 8429. **B)** Threshold of *r* = 0.84: nine clusters were detected, comprising 603 voxels on average. **C)** Threshold of *r* = 0.88: 29 clusters were created with average size of 83 voxels. **D)** Threshold of *r* = 0.92: eight clusters were detected with average size of 94 voxels. **E)** Threshold of *r* = 0.9062, 23 clusters with average size of 95 detected. **F)** Threshold of *r* = 0.80: one cluster was detected with size of 8397. **G)** Threshold of *r* = 0.84: six clusters were detected, comprising 911 voxels on average. **H)** Threshold of *r* = 0.88: 25 clusters were created with average size of 98 voxels. **I)** Threshold of *r* = 0.92: eight clusters were detected with average size of 92 voxels. **J)** Threshold of *r* = 0.96, two clusters with average size of 22 detected. Clusters are plotted in unique, arbitrary colors. Slice coordinates: x, y, z = (0, 0, 0).

Classification using PM

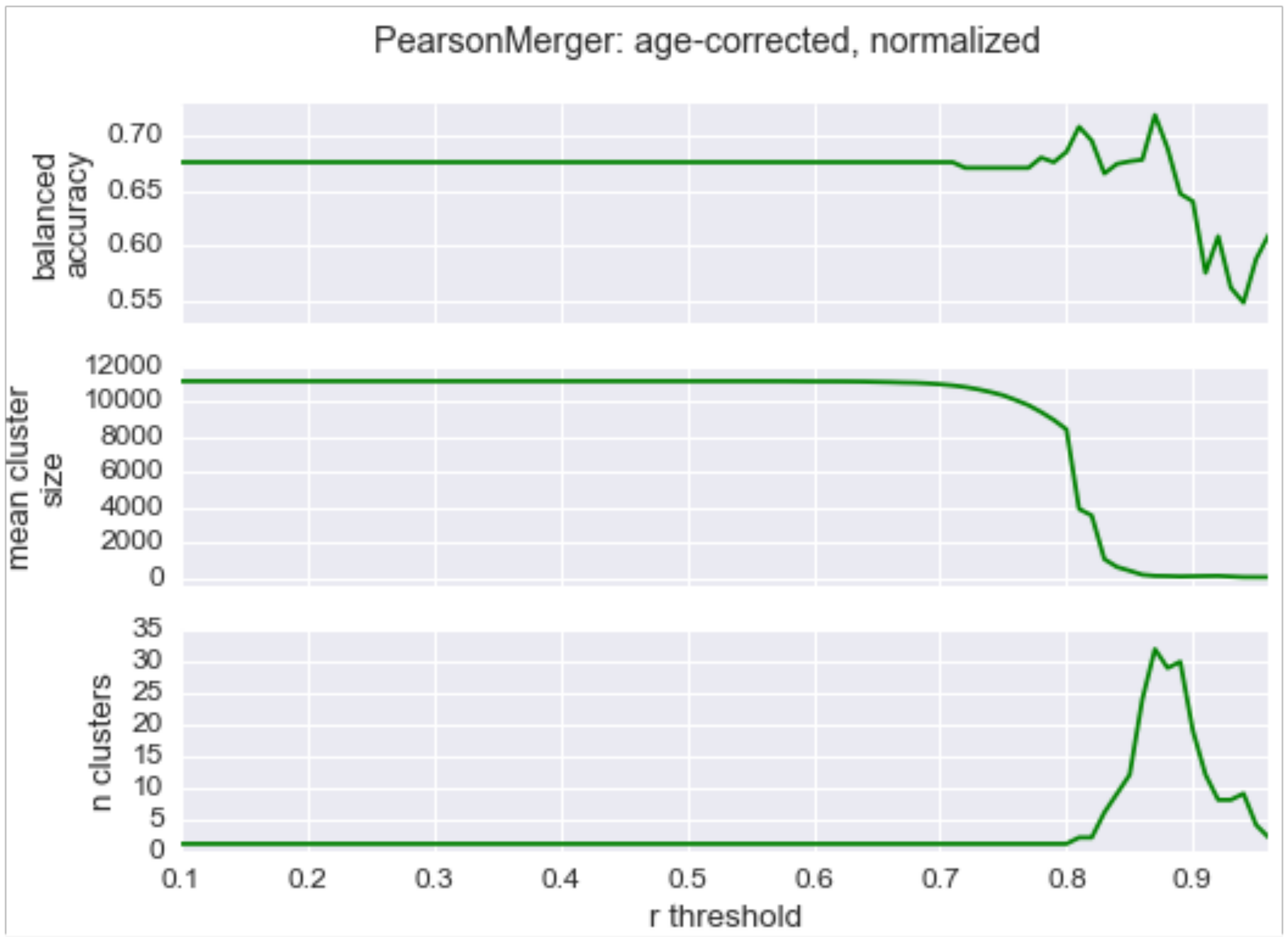
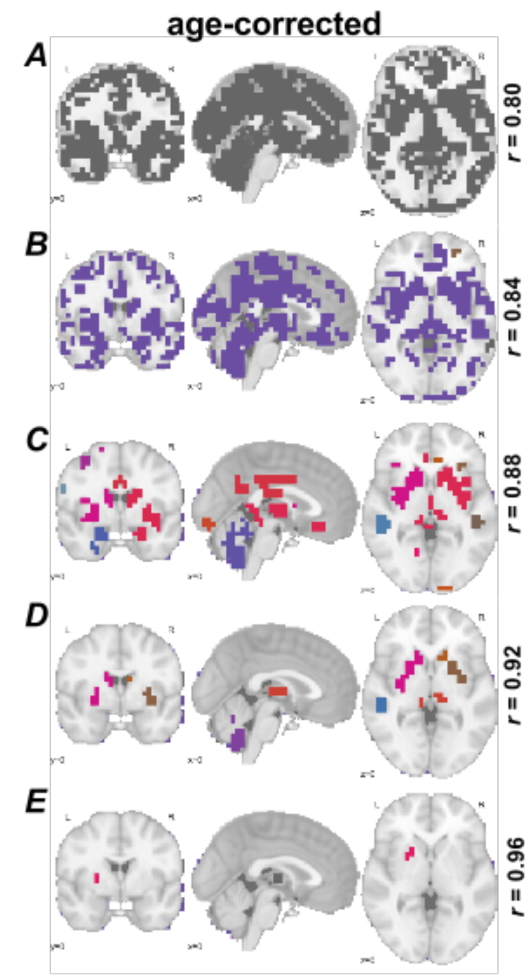
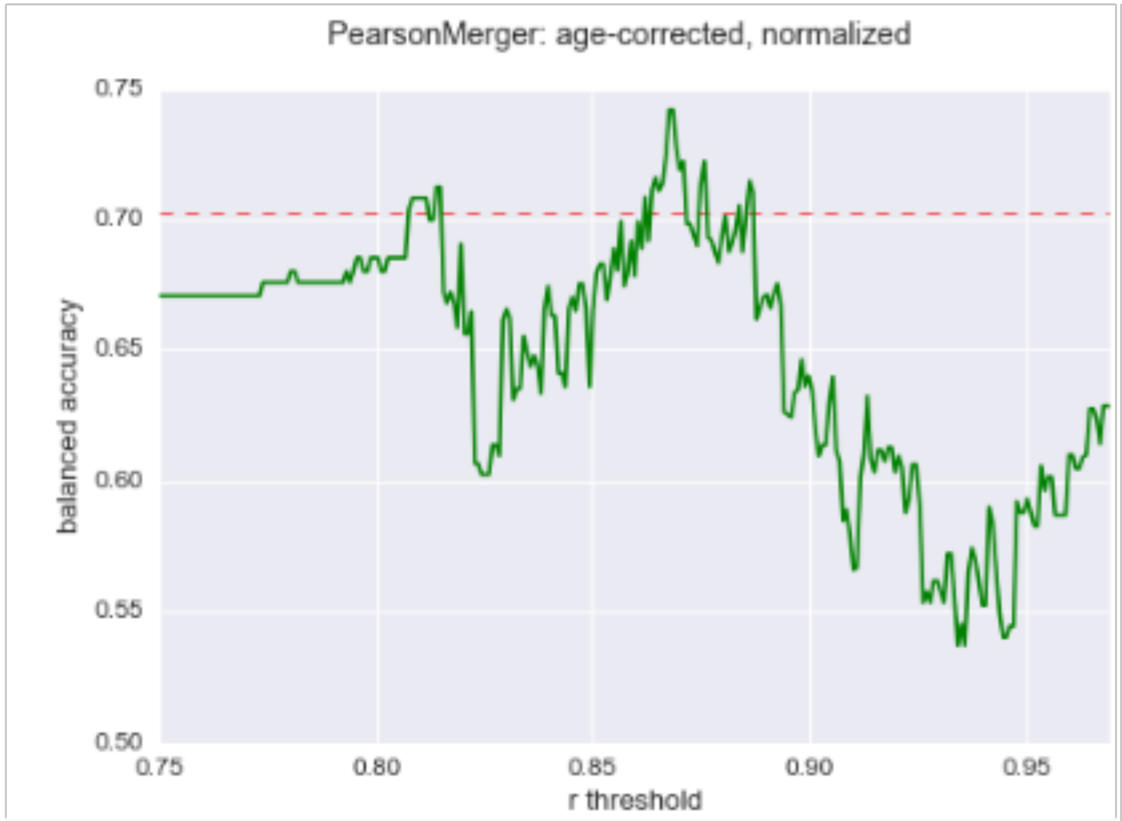


Figure 5| Overview plot for clustering obtained by PearsonMerger. A range of thresholds for Pearson correlation coefficients *r* was tested. As a function of *r*: top graph depicts classification performance (assessed by balanced accuracy scores); middle graph shows the mean size of generated clusters; and bottom graph indicates the number of clusters the algorithm was able to detect. MRI data was normalized to zero mean for all samples and corrected for age.



Classification using PM



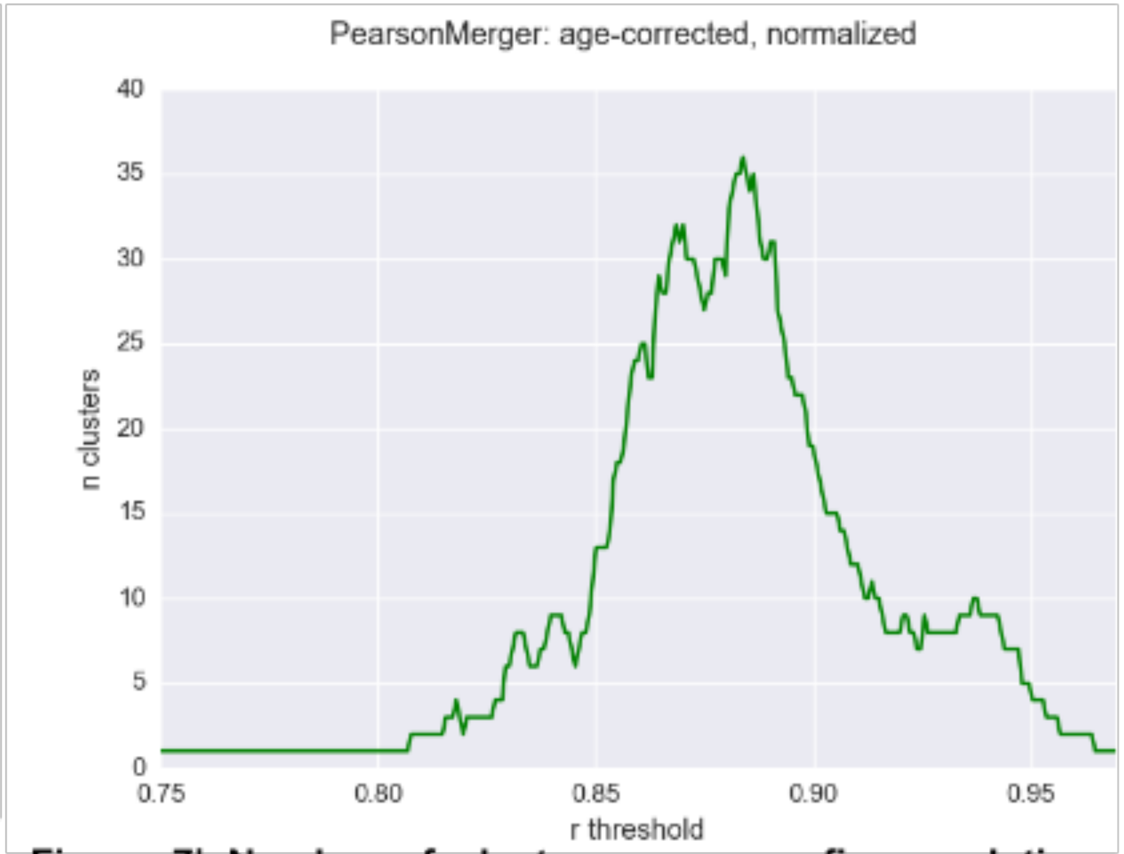


Figure 6| Classification Performance across a fine-resolution Range of Correlation Thresholds. The red dashed line represents classification accuracy on JHU atlas-derived clusters (0.7021). PearsonMerger-based clusters and respective classification scores are drawn in green, optimal performance of 0.7418 is reached at a threshold of r = 0.8677. Data was age-corrected and normalized to zero mean.

Figure 7| Number of clusters across a fine-resolution range of r thresholds. A maximum of 36 clusters was detected at the correlation threshold of r = 0.8838. Data was age-corrected and normalized to zero mean.

Spectral Clustering

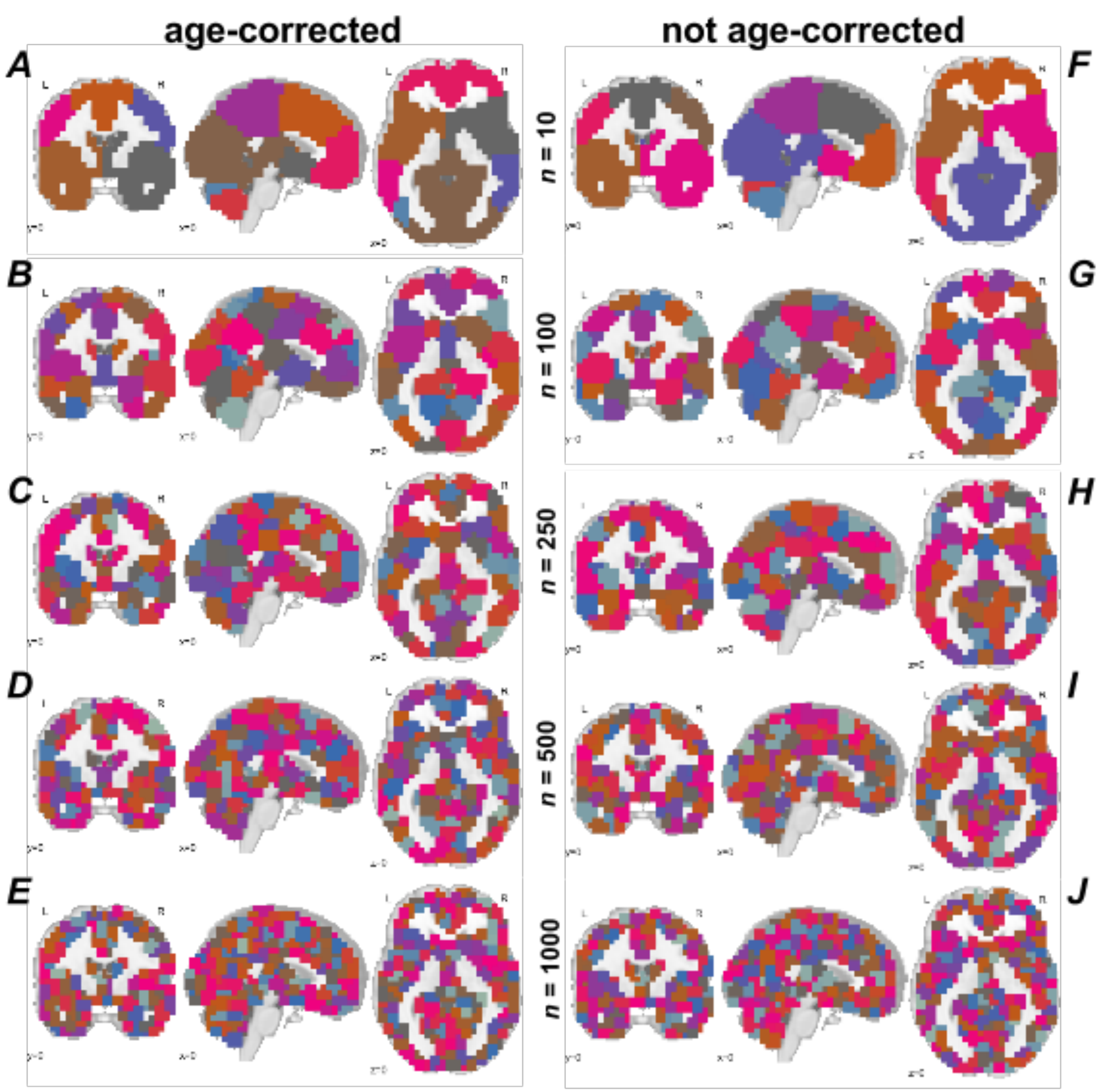


Figure 8| Results of Spectral Clustering at varying cluster numbers based on all participants. Images were normalized to zero mean and age-corrected if indicated. **A, F)** n clusters = 10, mean size of 1117 voxels. **B, G)** n clusters = 100, mean size of 112 voxels. **C, H)** n clusters = 250, mean size of 45 voxels. **D, I)** n clusters = 500, mean size of 22 voxels. **E, J)** n clusters = 1000, mean size of 11 voxels. Clusters are plotted in unique, arbitrary colors. Slice coordinates: x, y, z = (0, 0, 0).

Classification using SC

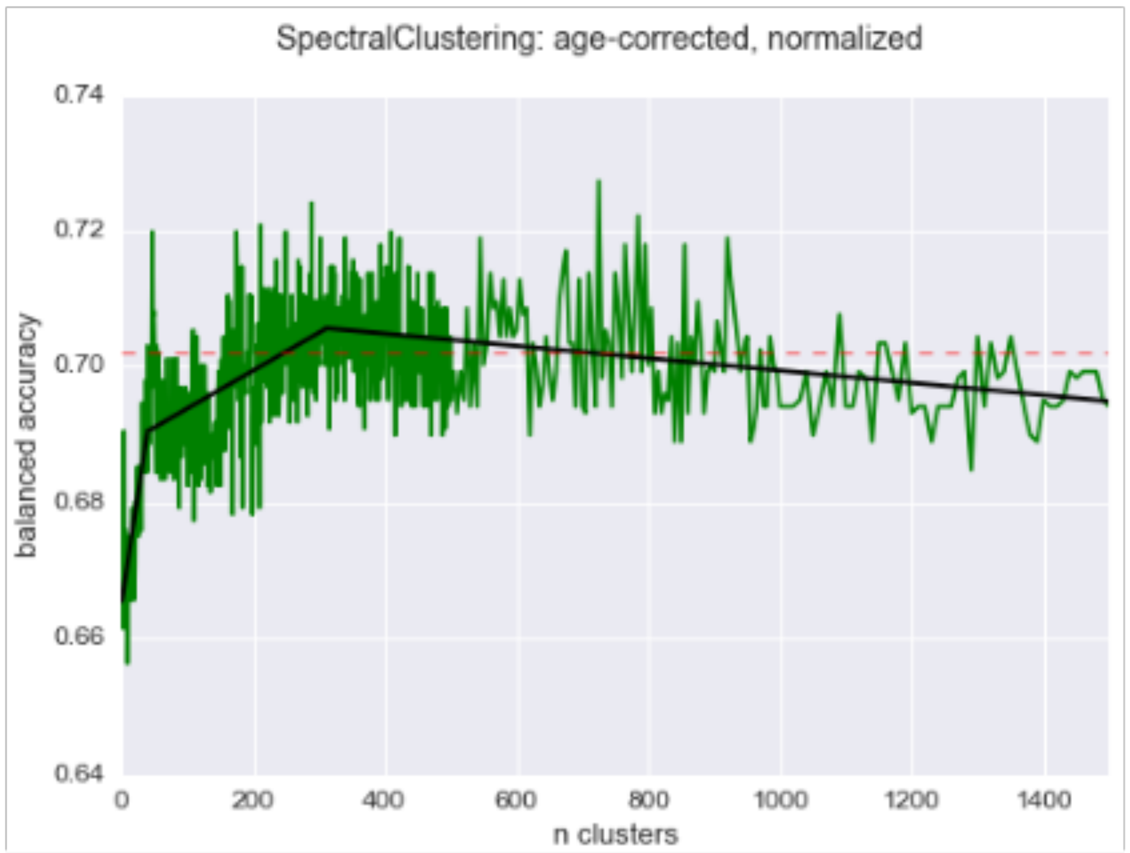
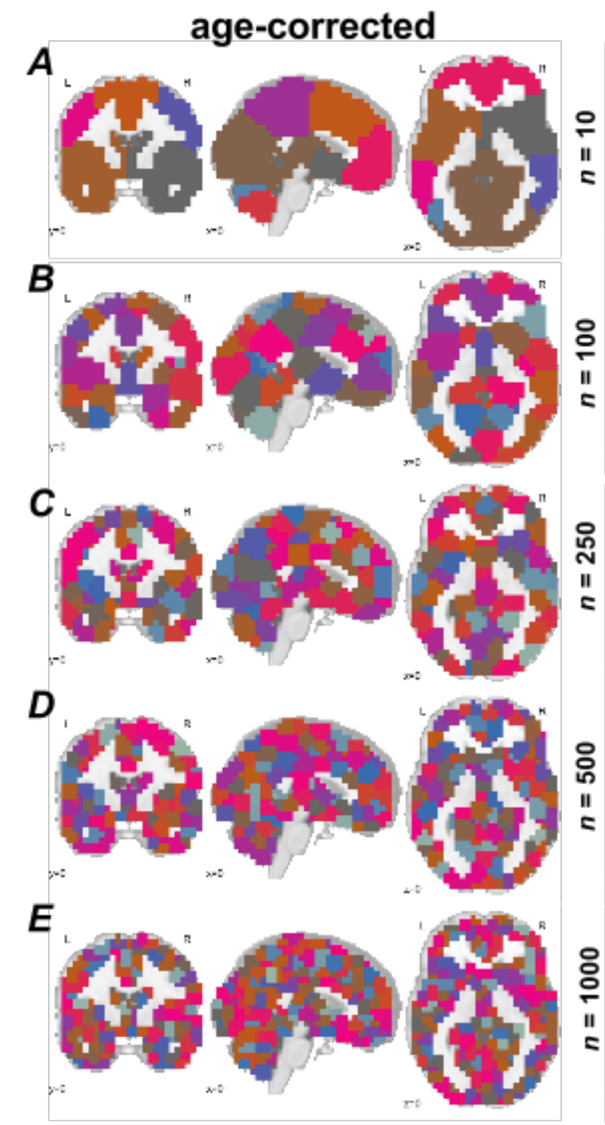


Figure 9| Balanced Accuracy Scores over a range of clusters obtained via Spectral Clustering. Green line represents balanced accuracy scores, red dashed line marks classification performance using the JHU atlas (0.7021), black lines are the result of a change point analysis. Slopes for first line m = 0.00064071 (p = 0), y-intersect n = 0.6648; slope of the second fitted line m = 0.00005587 (p = 0), y-intersect n = 0.6882; and slope of the third line m = -0.00000905 (p = 0), y-intersect n = 0.7085. The first changepoint lies at 39 clusters and the second change point was found at 311 clusters, classification score at second change point is 0.7056. Data was corrected for age and normalized to zero mean prior to clustering.



Results I

- normalisation, age-correction aid classification
- outperformed JHU-atlas based ROIs:
    - *JHU:* 0.7021, *PM:* 0.7418, *SC:* 0.7056
- PM showed best scores:
    - but sensitive to input parameter (r threshold)
- SC with lower classification performance:
    - more stable across parameter space (n clusters)

Results II

- AUD specific regions
  - Guggenmos et al., Accelerated Brain Aging in Alcohol Use Disorder, in preparation:
    - regression models for group contrast
  - partially picked up by PearsonMerger:
    - dorsal anterior cingulate gyrus, frontal gyrus, temporal gyrus, thalamus, insular, fusiform gyrus, cerebellum
- different approaches identified overlapping regions!

Summary

- unbiased and data-driven method to:
  - reduce complexity of sMRI data, create ROIs and brain parcellations
  - enhance classification
  - PM additionally identified AUD-related regions
- future work:
  - more datasets (functional?)
  - more classifiers like SVM or RF

Many thanks to everyone!