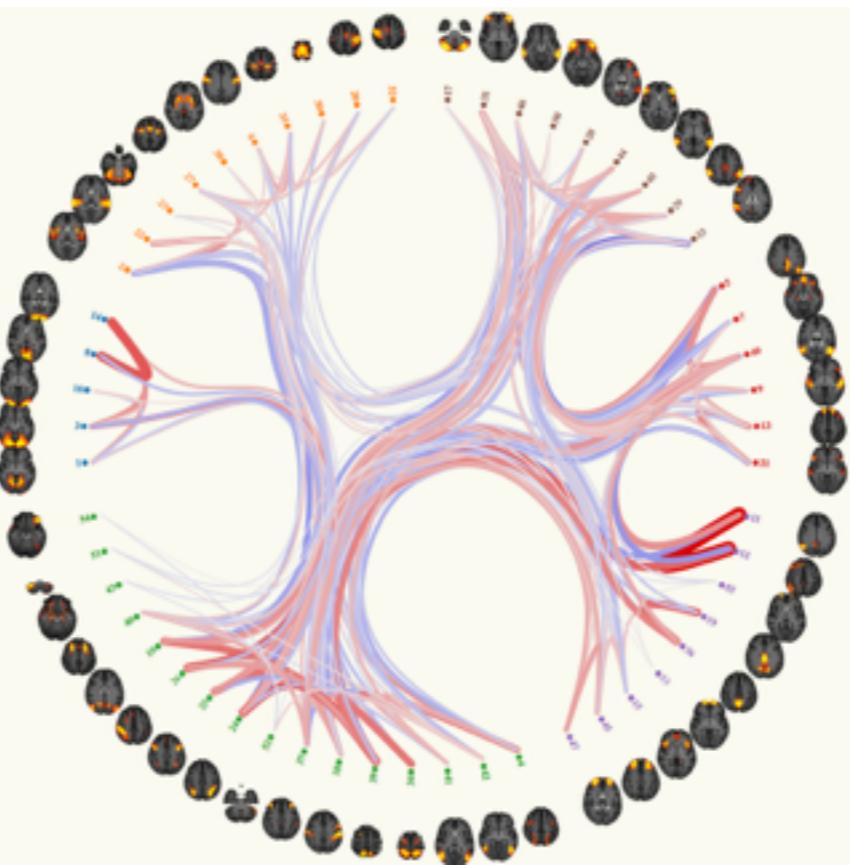
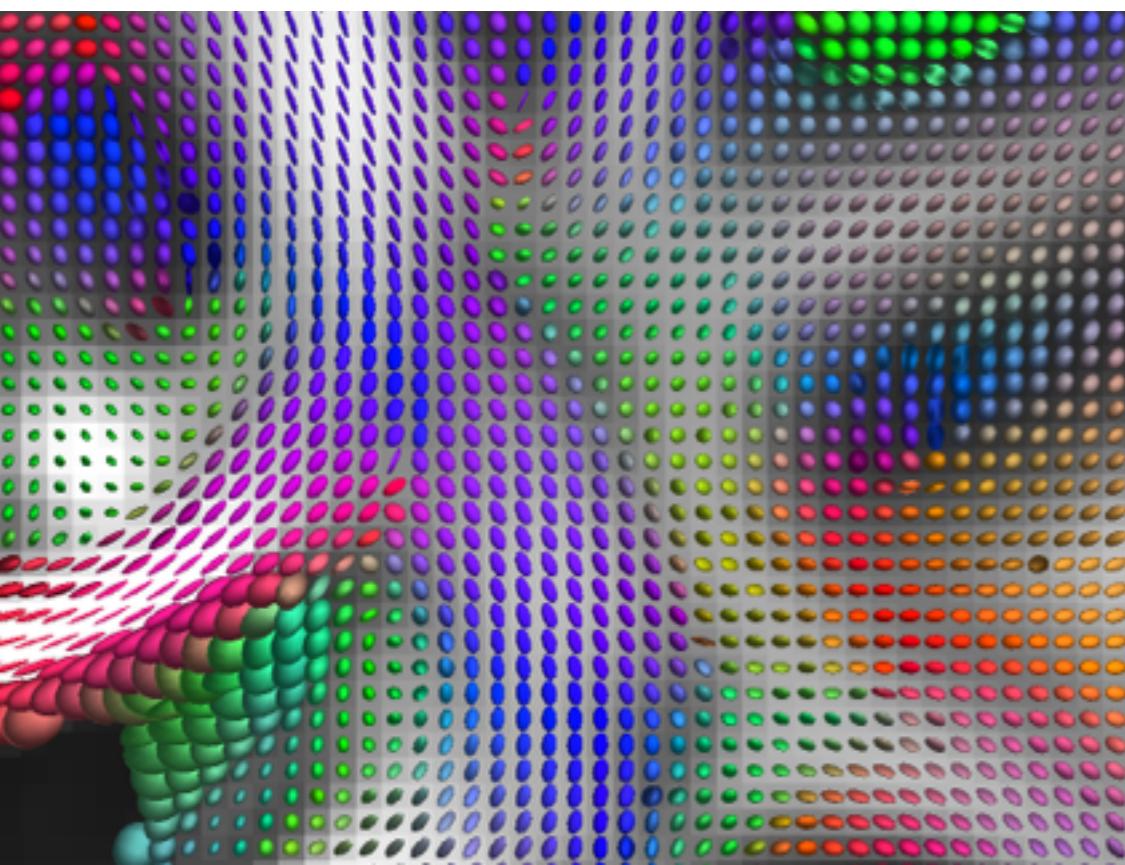
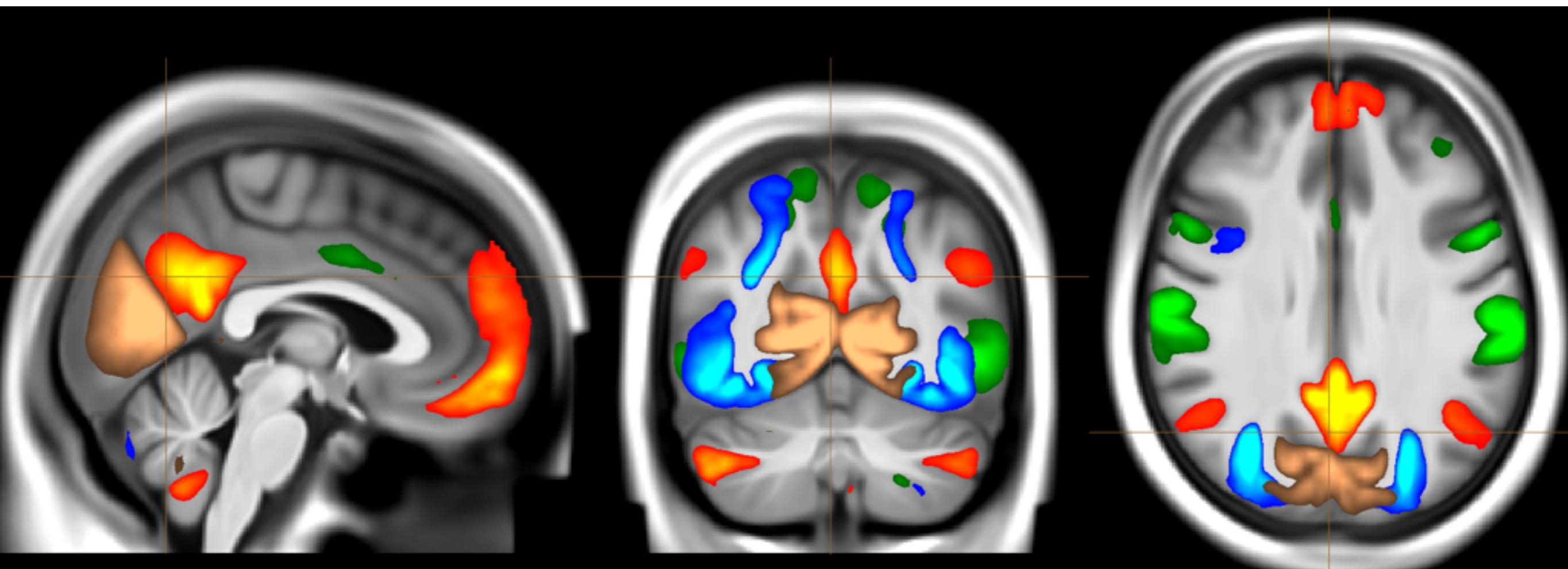


UK Biobank Pipeline: Public release of the first 10,000 datasets

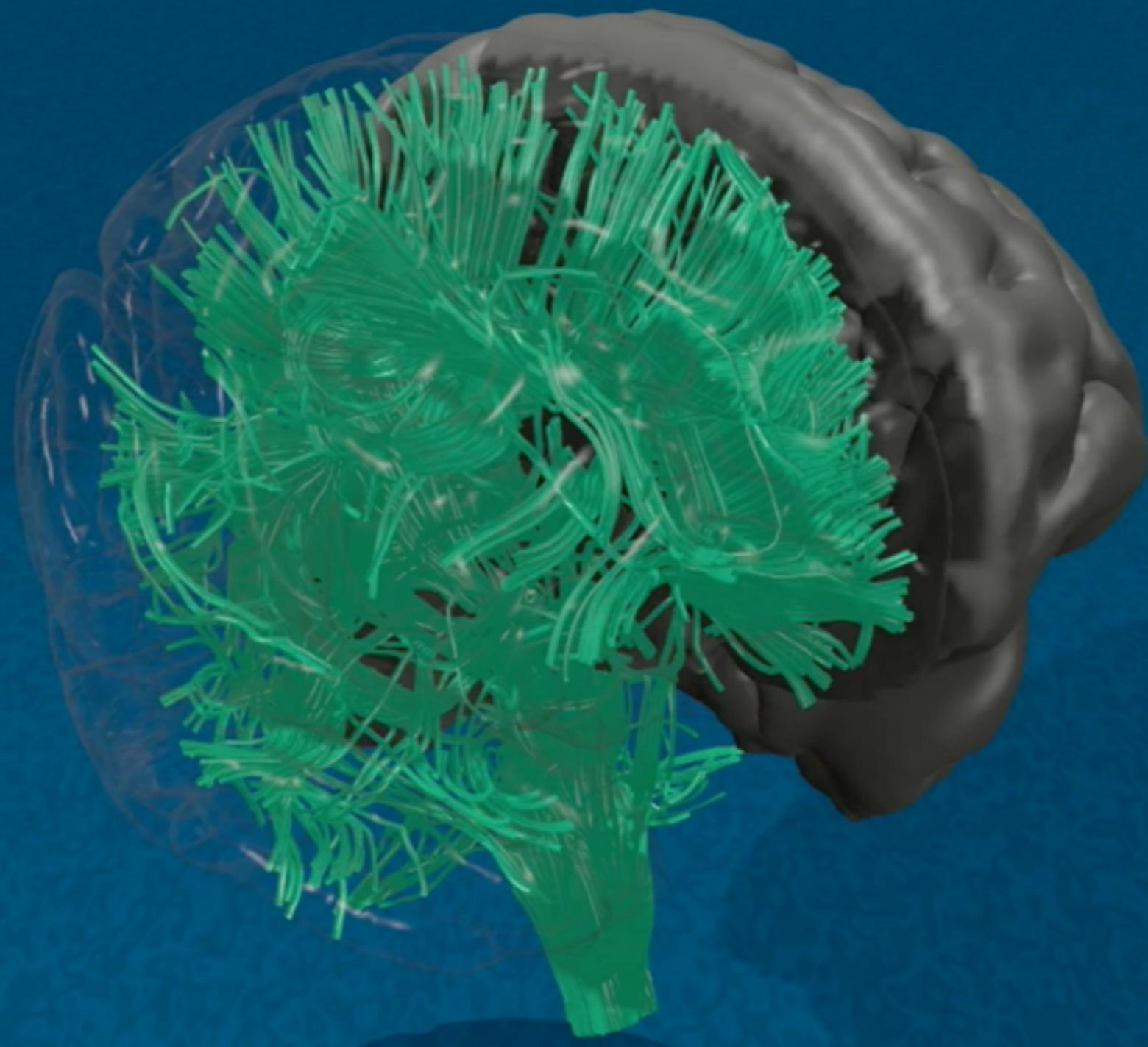
Fidel Alfaro Almagro, FMRIB Oxford



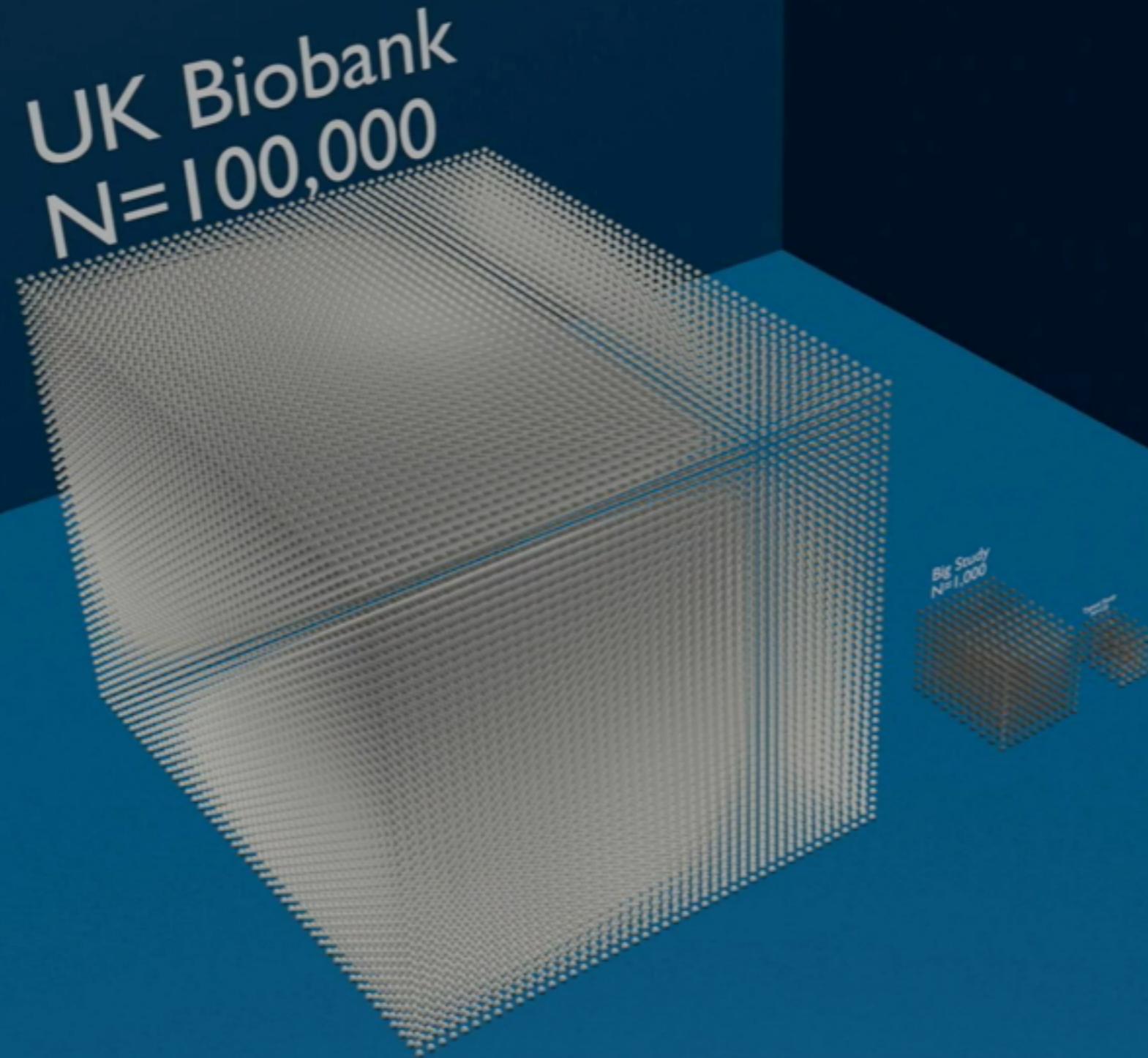
- Prospective epidemiological study: 500,000, 45-75y, UK residents
- Genetic data + biological samples + lifestyle information + health records.
- Discover early markers & risk factors of disease
- A large subset of the subjects are being scanned (13,700 subjects so far).

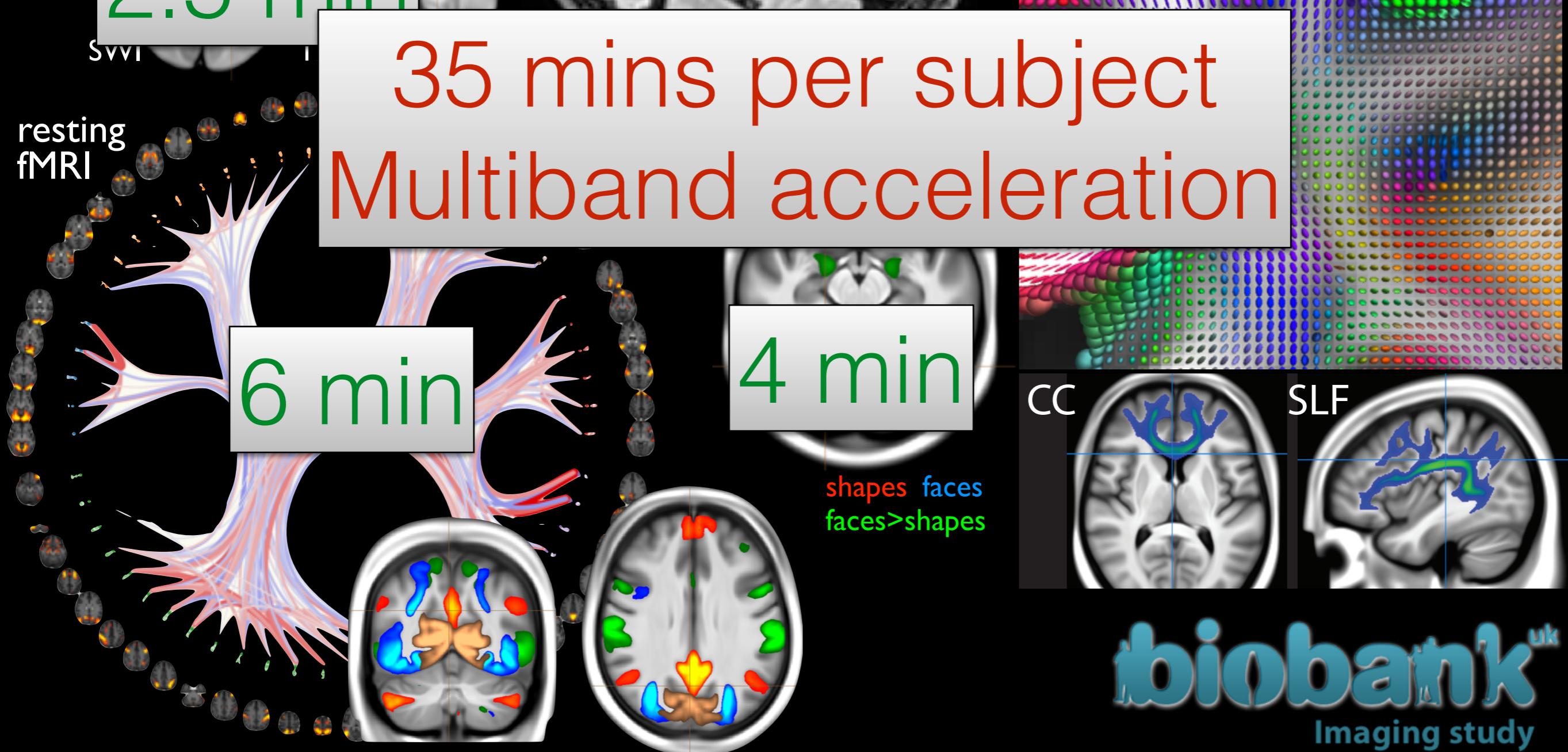
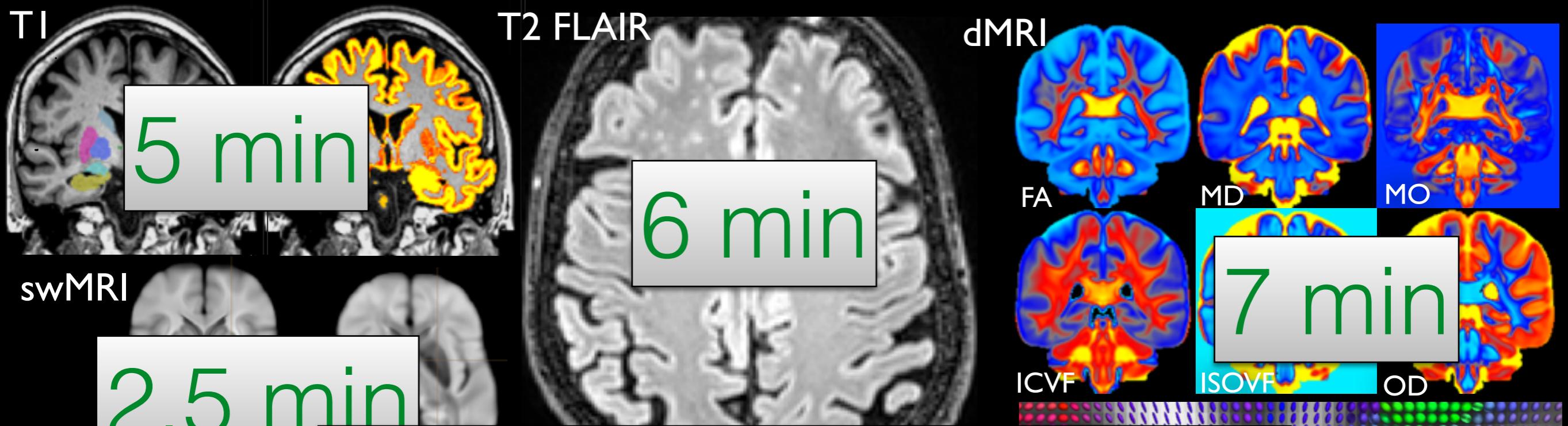


tinyurl.com/UKBiobankMovie

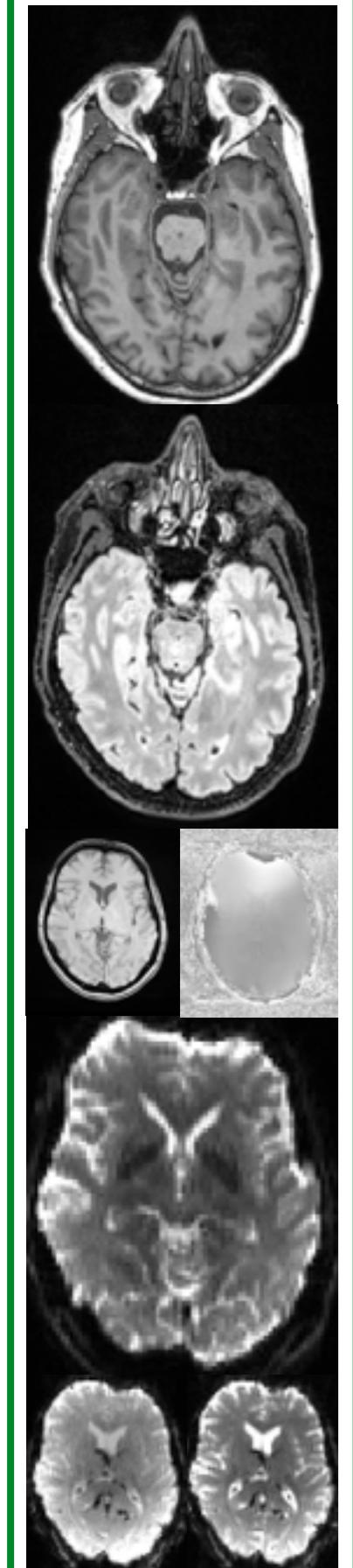


tinyurl.com/UKBiobankMovie

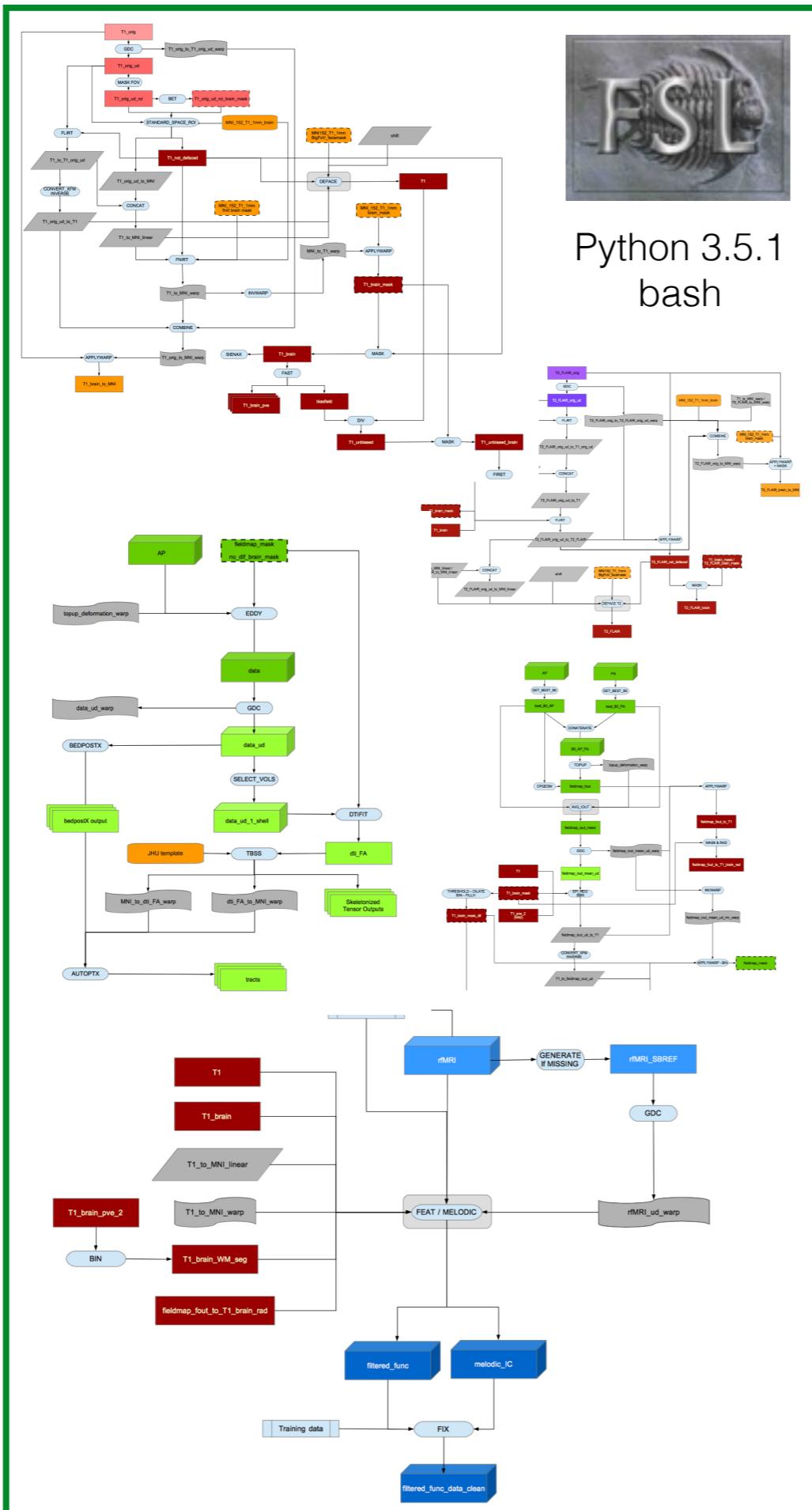




Raw data



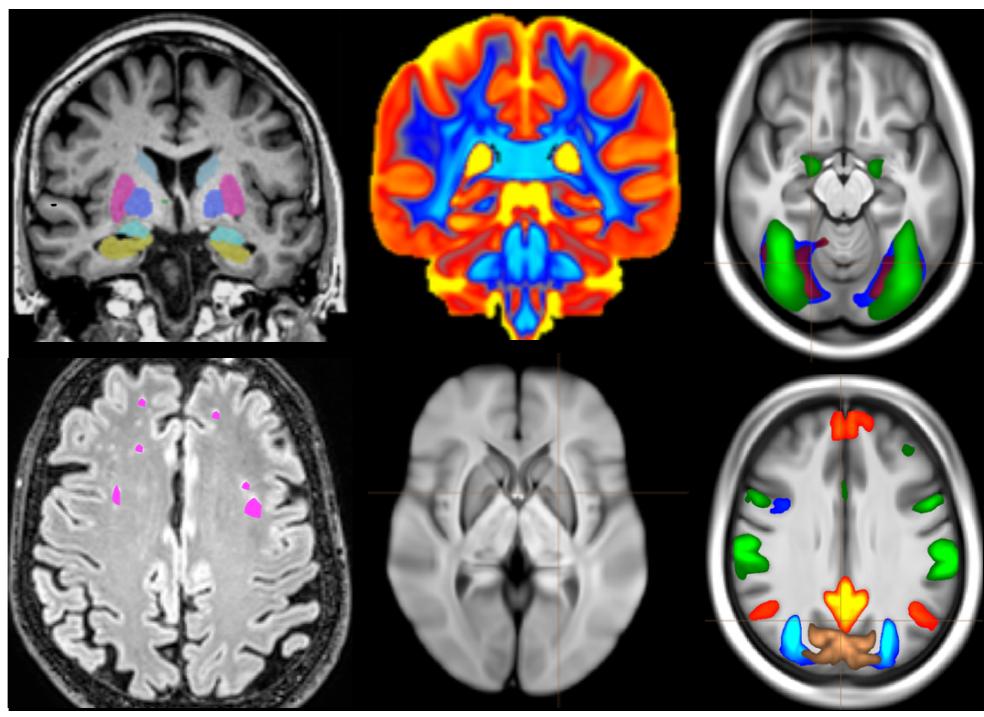
Automated processing



FSL
Python 3.5.1
bash

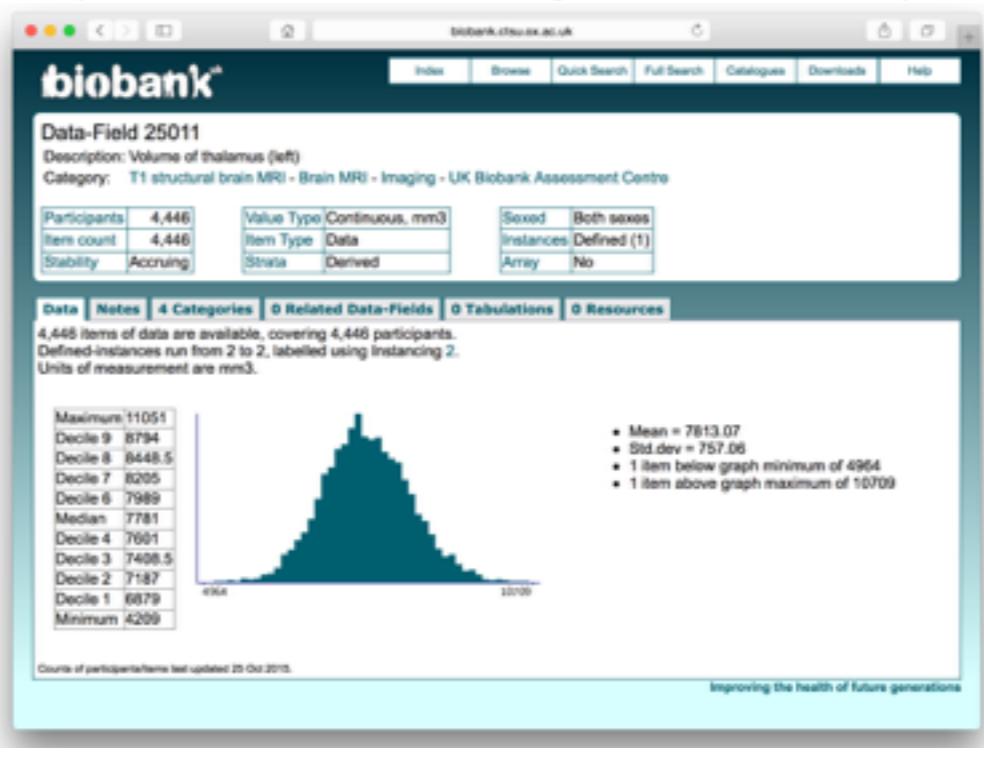


Open-access database



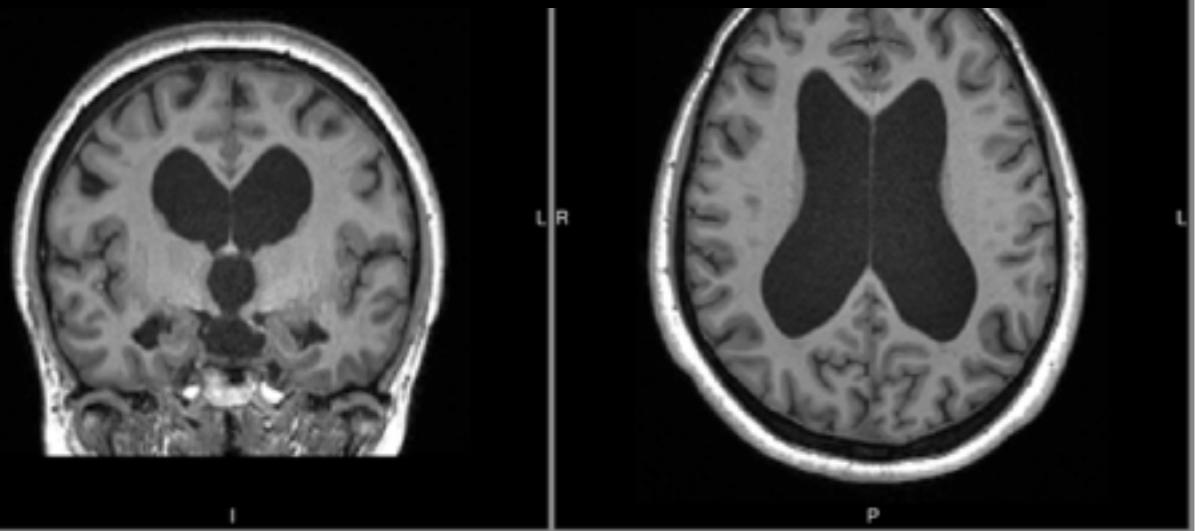
+

Imaging-Derived Phenotypes
(IDPs - summary measures)

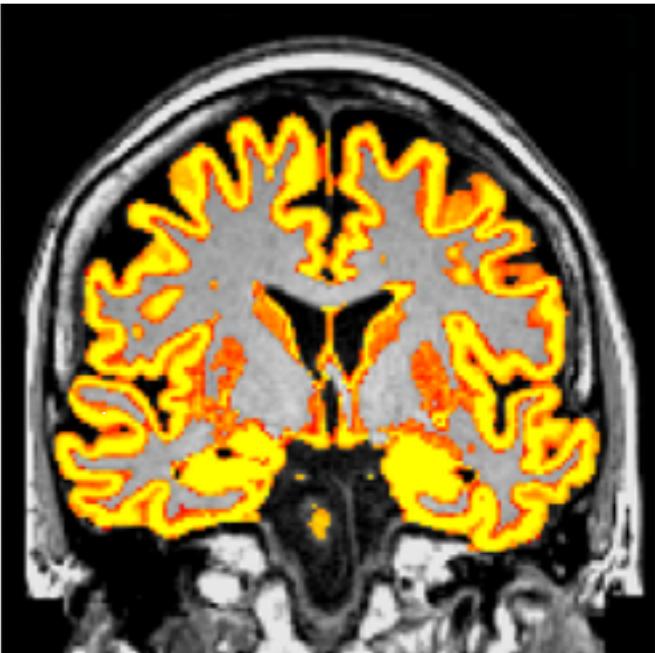


Structural MRI

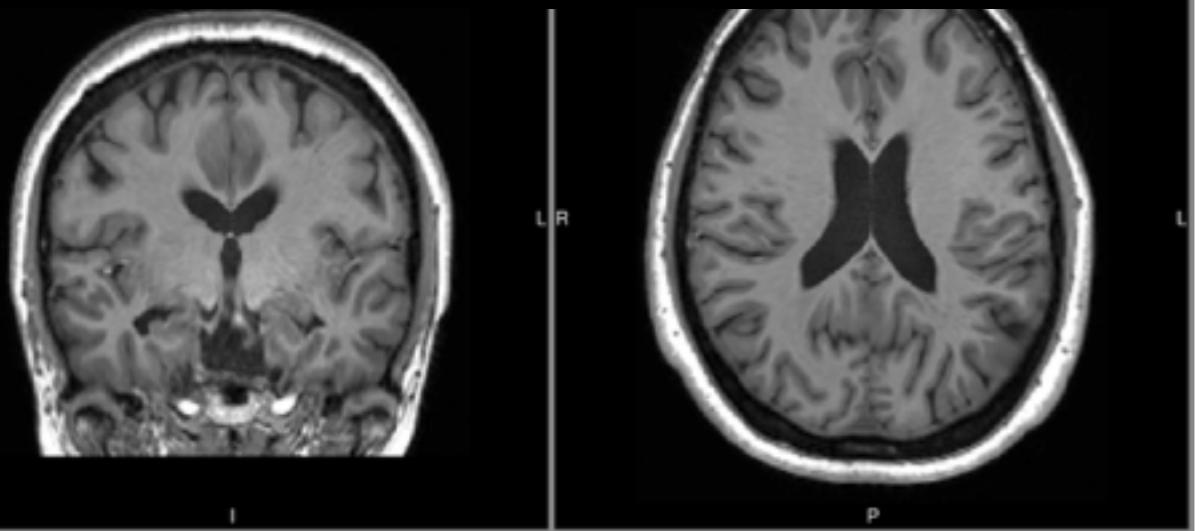
linear registration to standard space



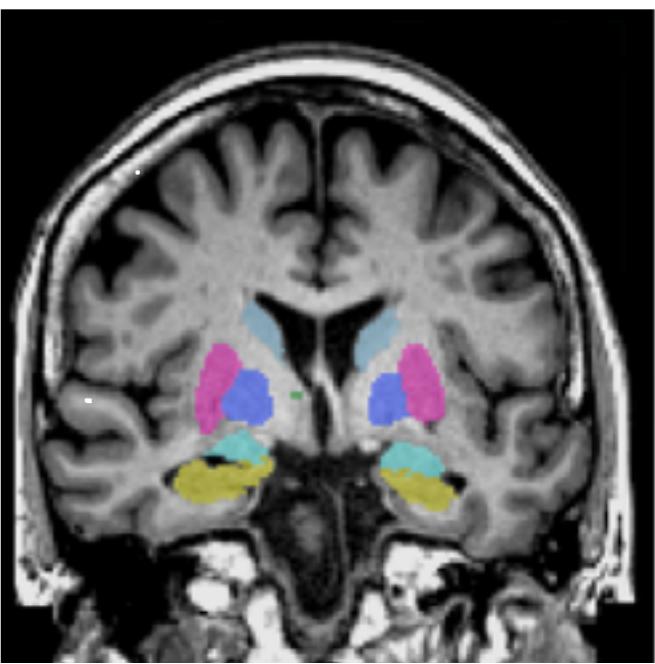
whole-brain
& tissue
volumes



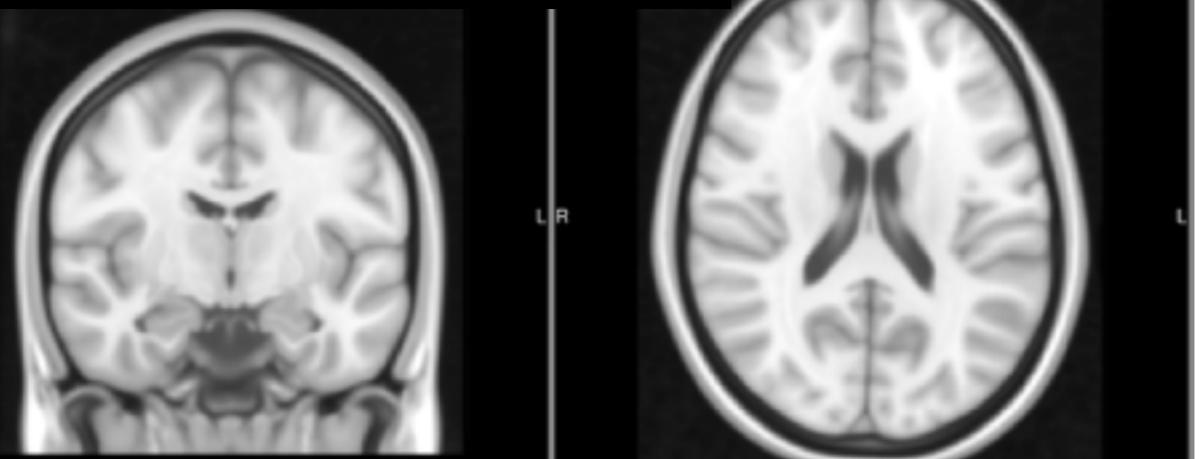
nonlinear registration to standard space



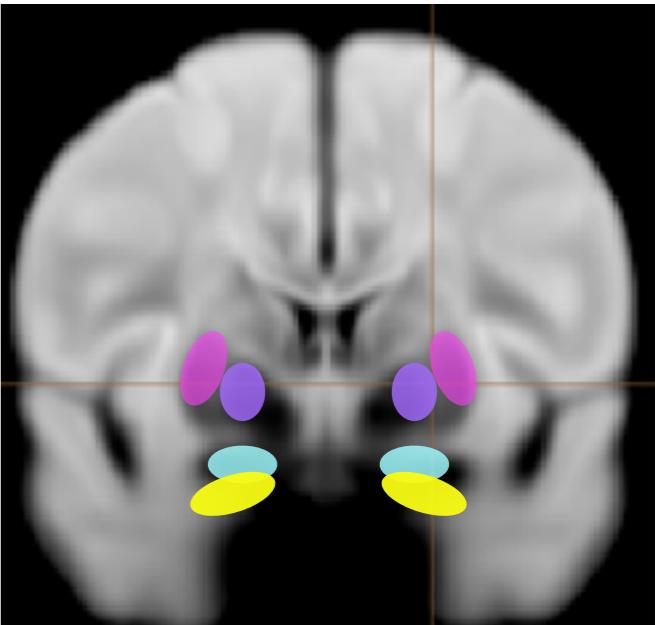
subcortical
volumes



standard space template

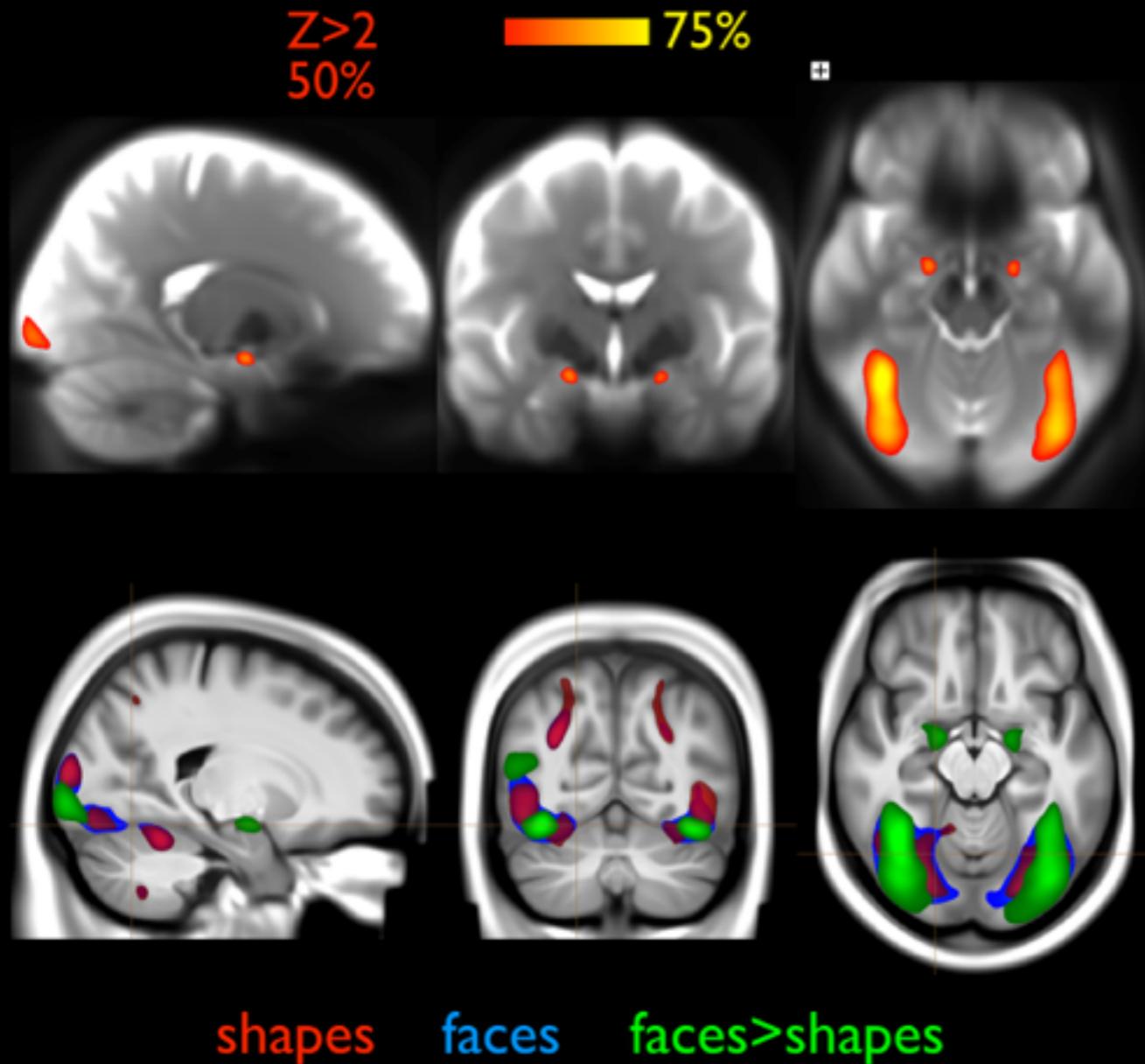


median T2*
within
subcortical
regions



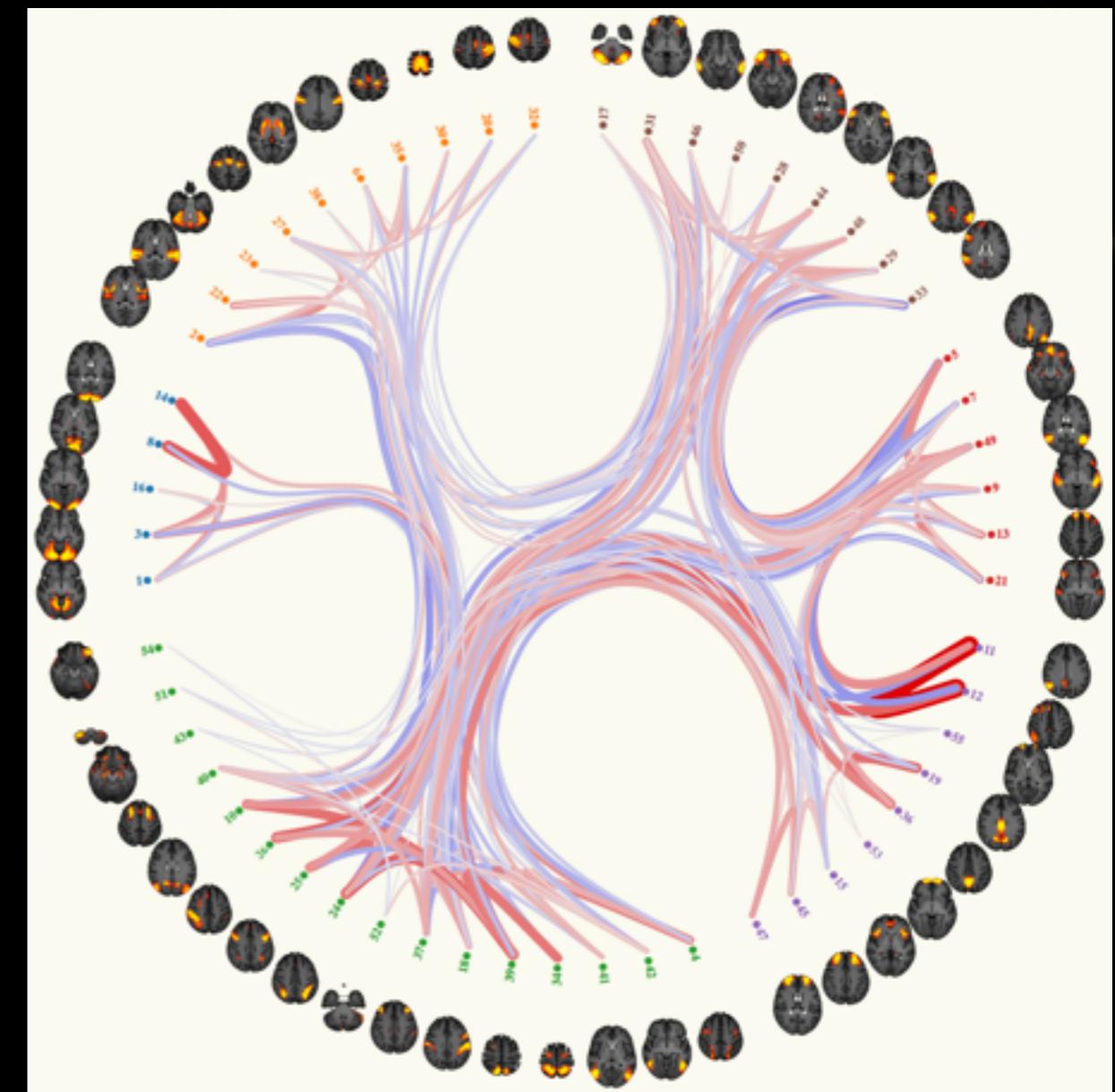
Functional MRI

Task fMRI



HCP's paradigm: Faces/shapes task paradigm. [Hariri et al. 2002. Neuroimage]

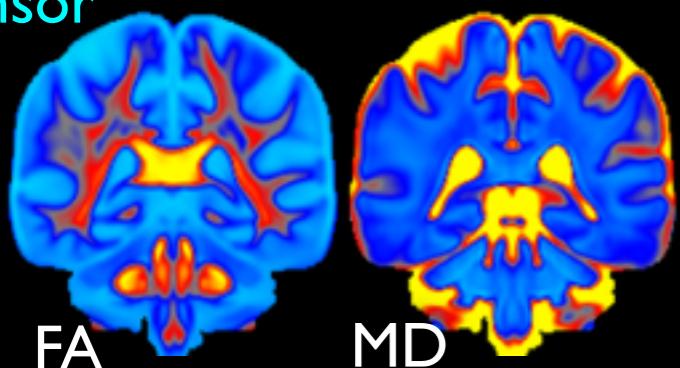
Resting State fMRI



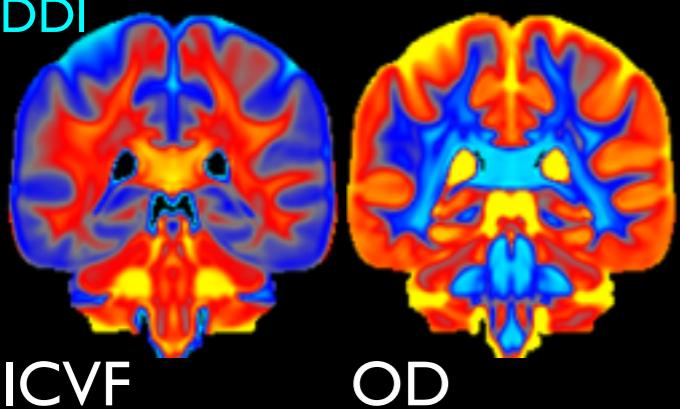
ICA & functional connectivity analysis

Diffusion

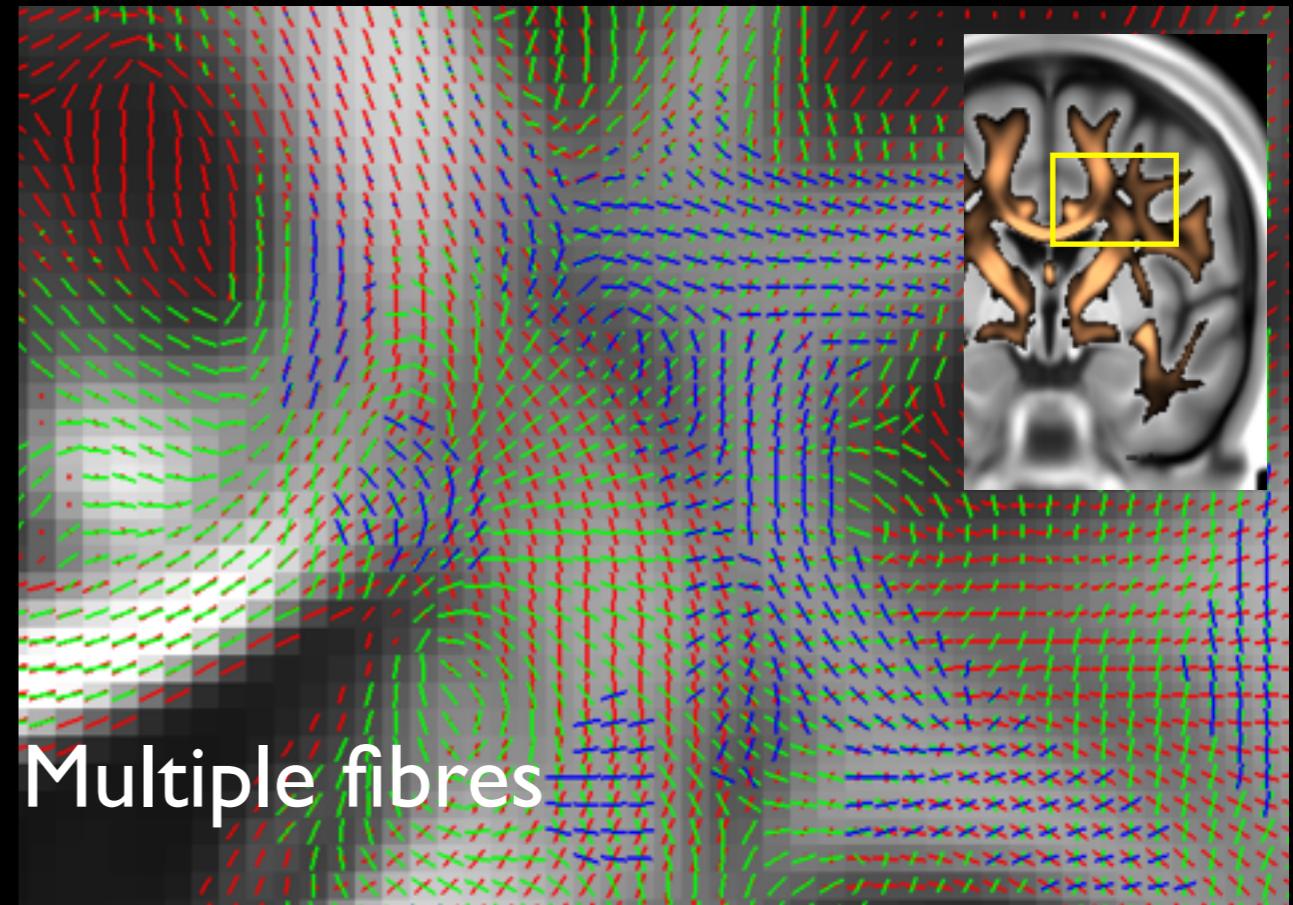
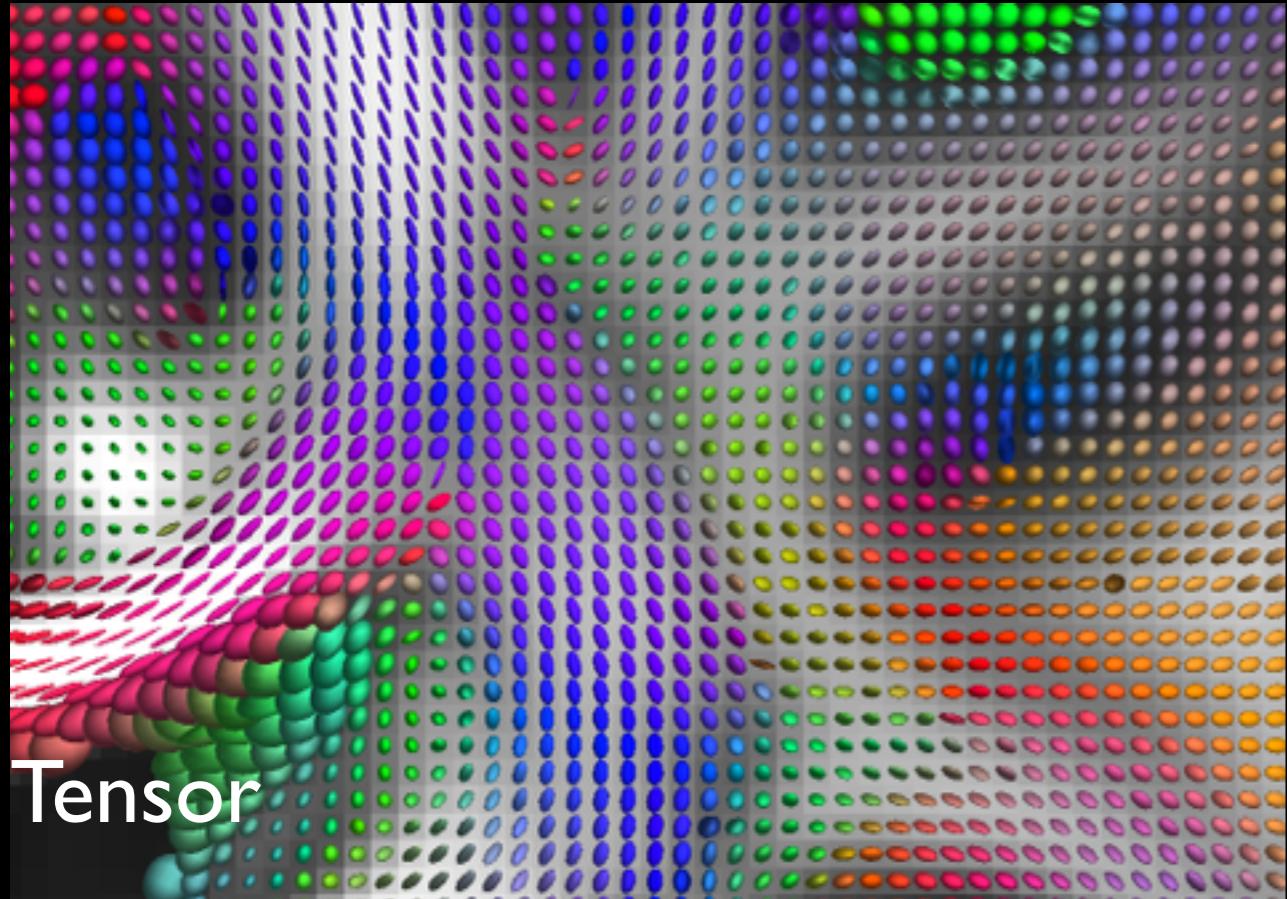
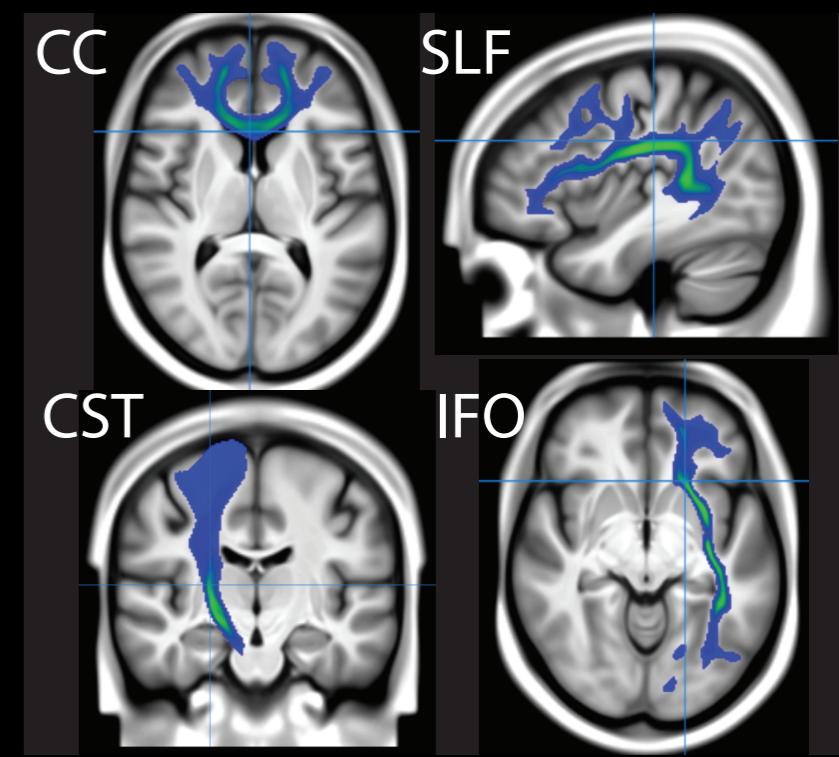
Tensor



NODDI



Tract masks for IDPs

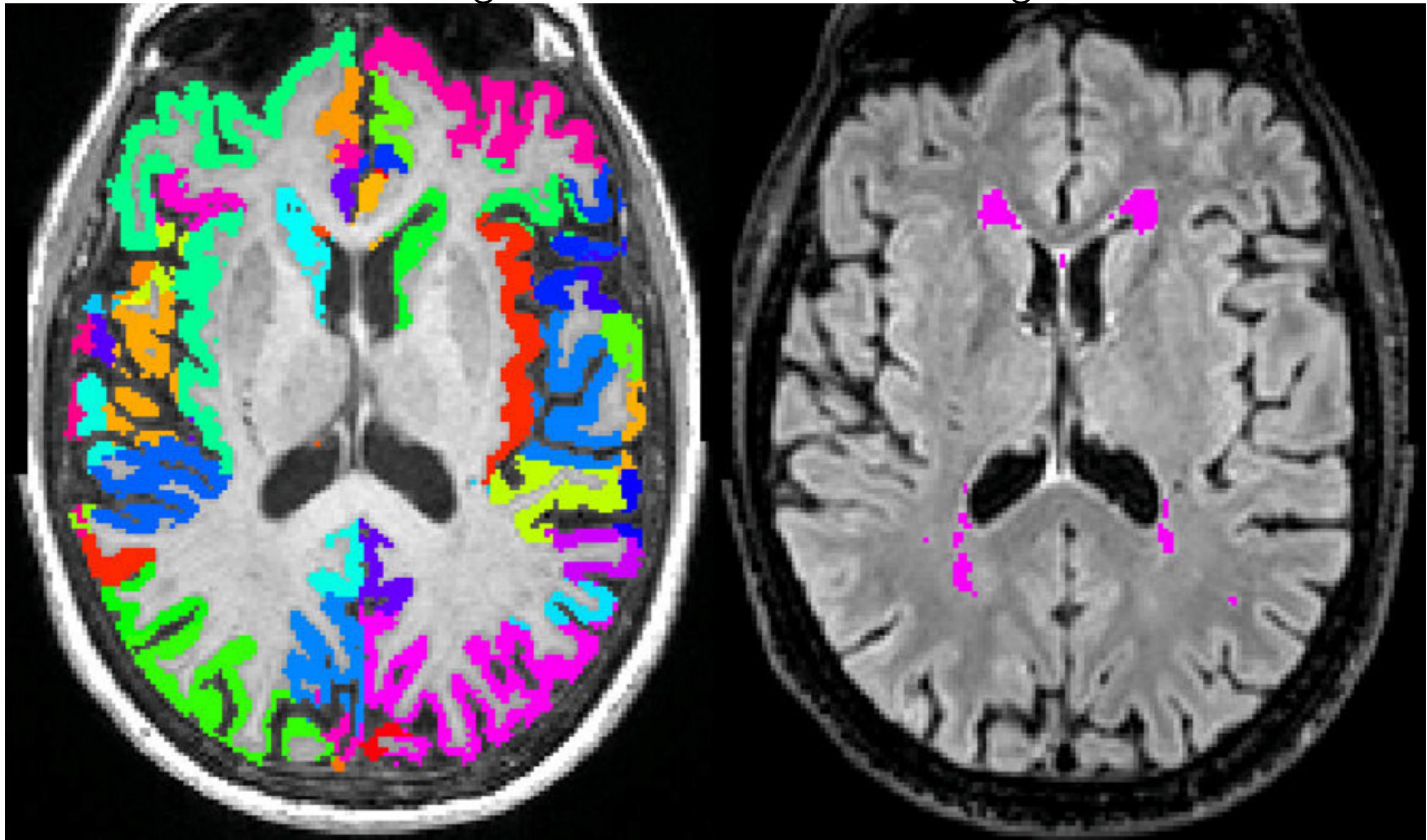


Recent IDPs

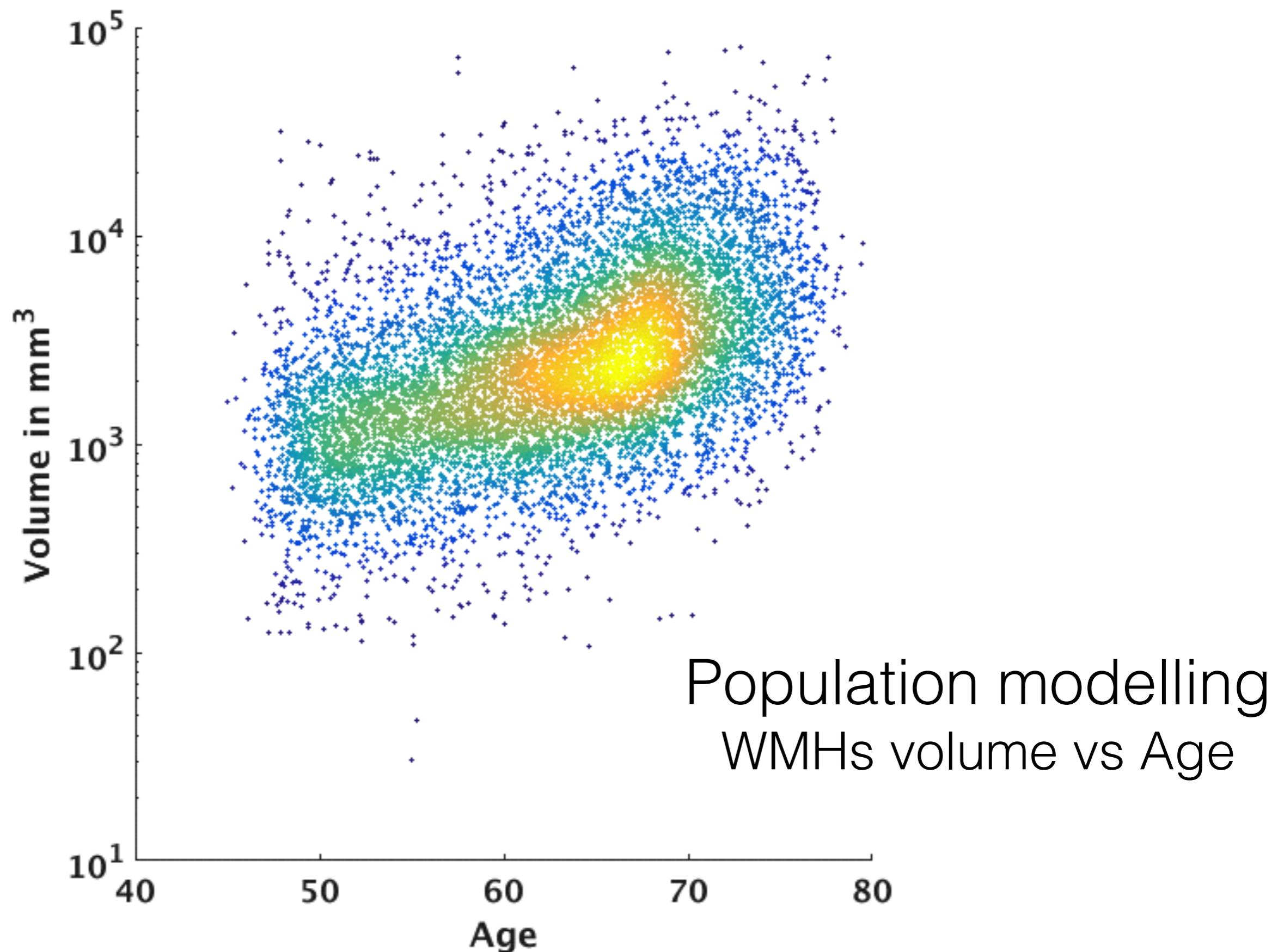
(Imaging Derived Phenotypes)

Volume of grey matter in 139
different brain regions

Volume of WMHs
using BIANCA

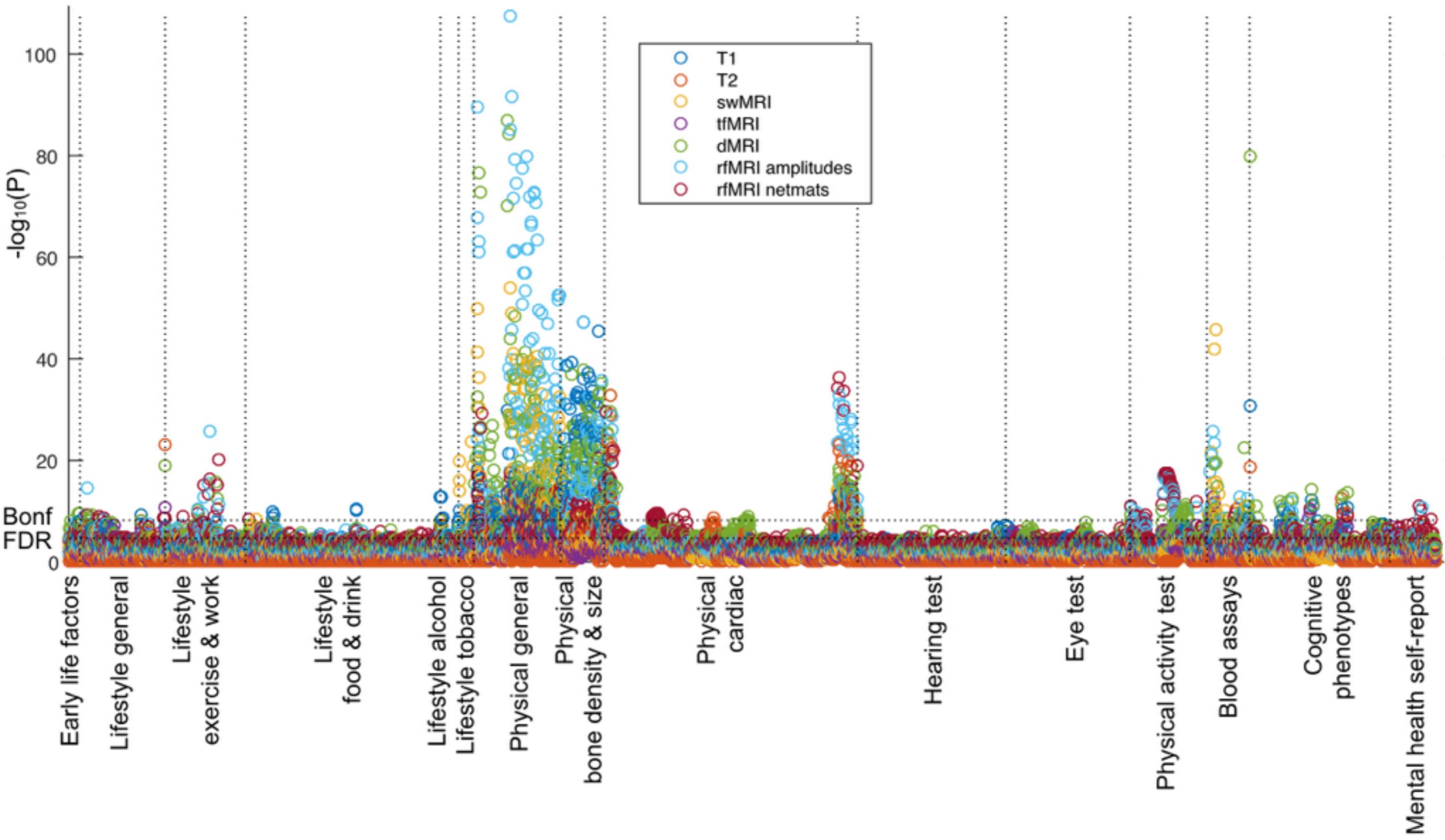


IDP results



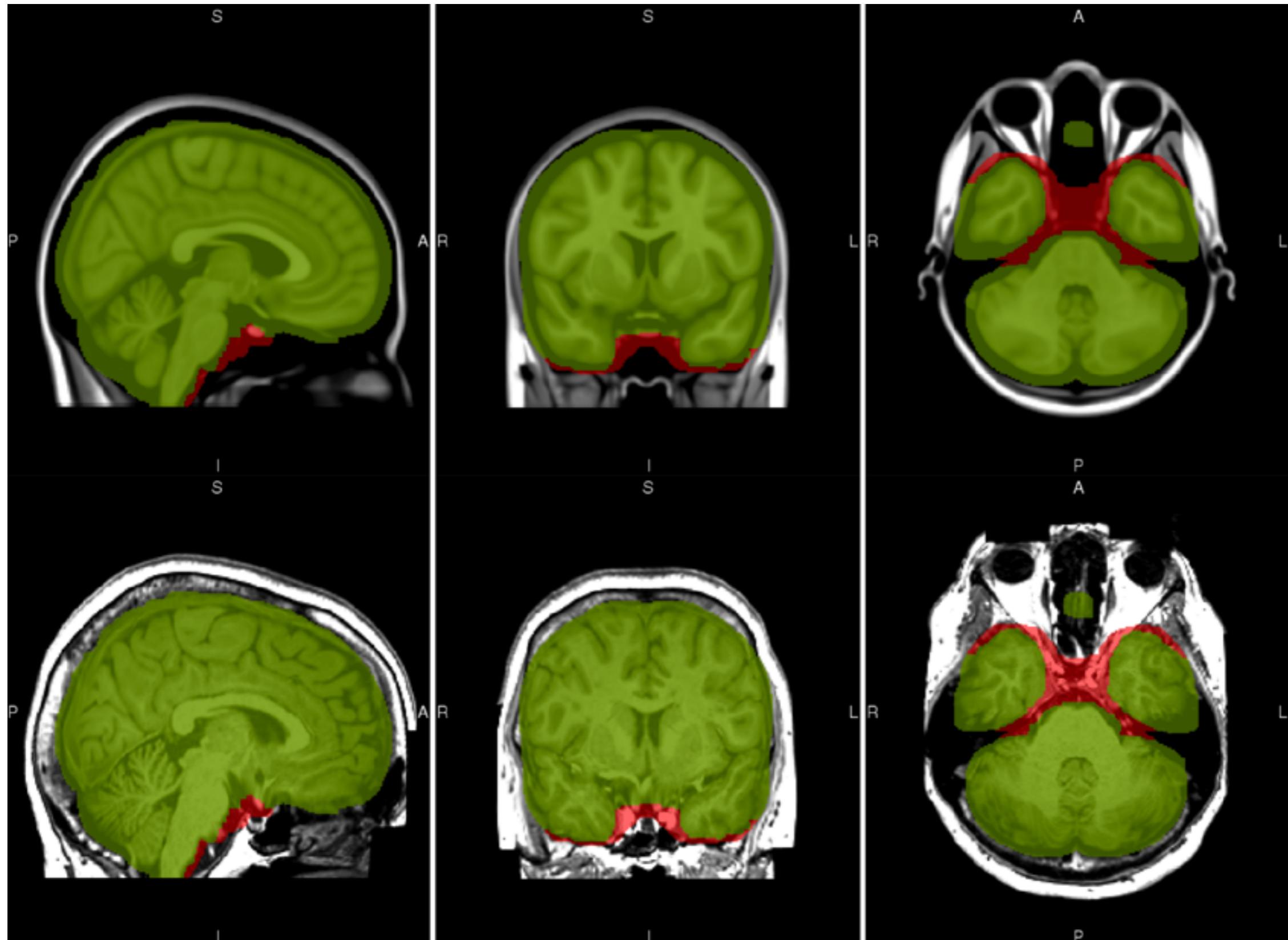
IDP results

8 million univariate associations between IDPs and non-brain-imaging variables (10,000 subjects)

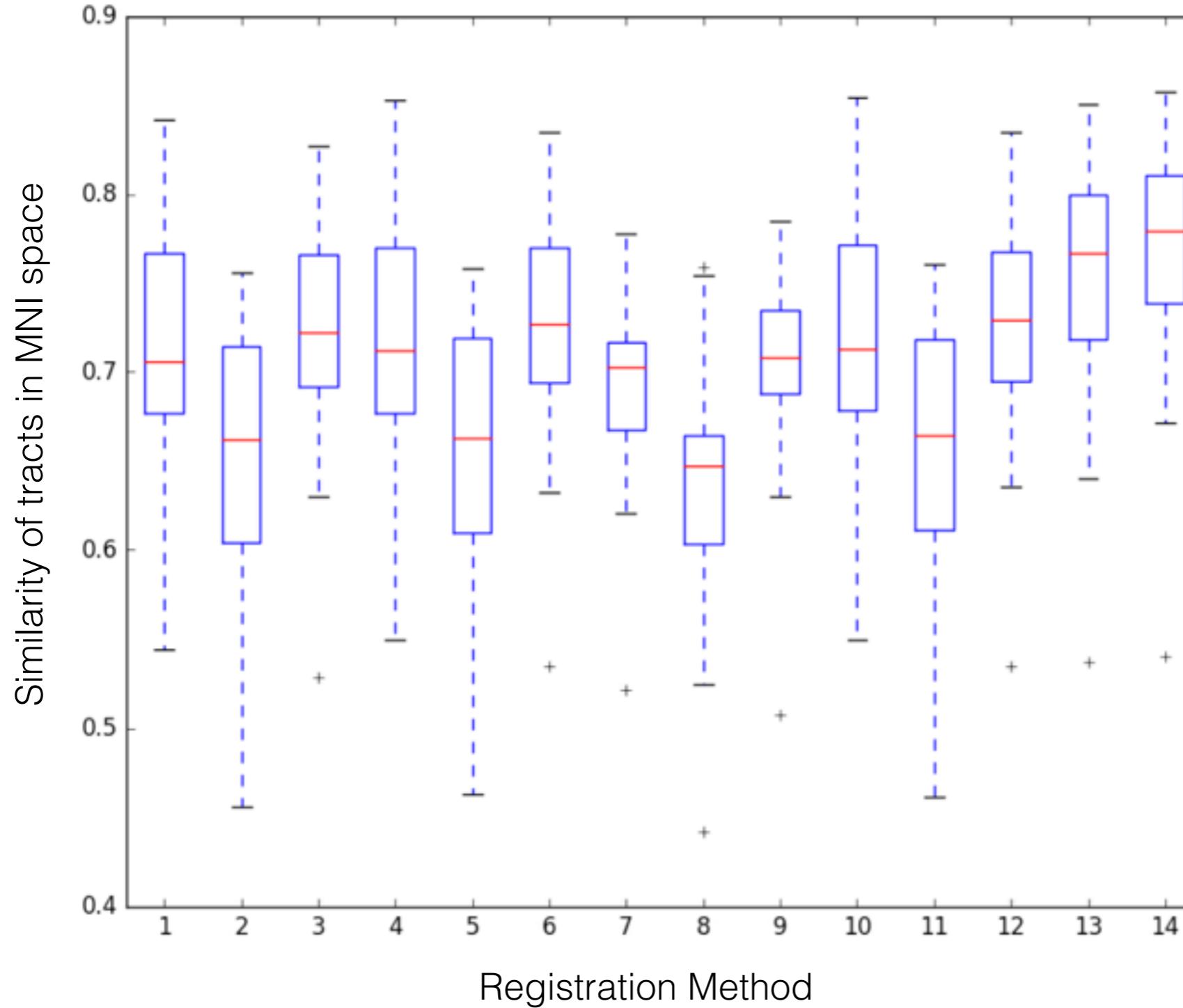


Decisions taken building the pipeline

Non-linear registration to MNI



Decisions taken building the pipeline

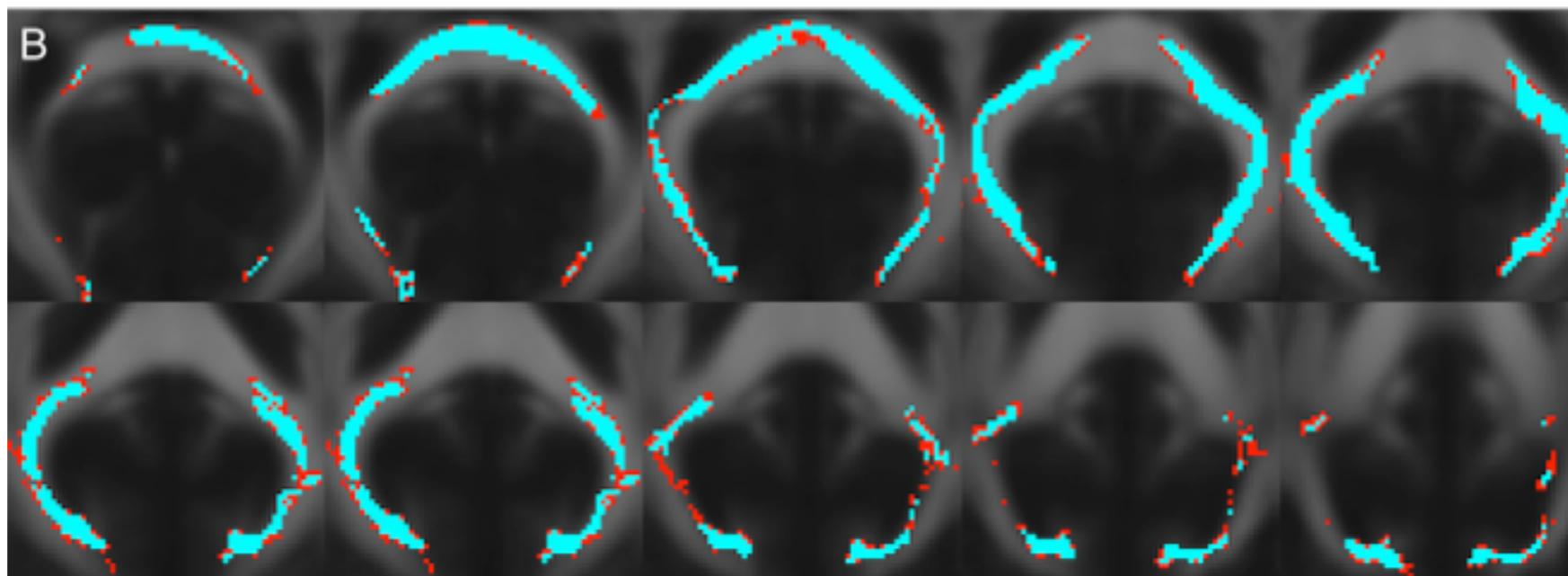
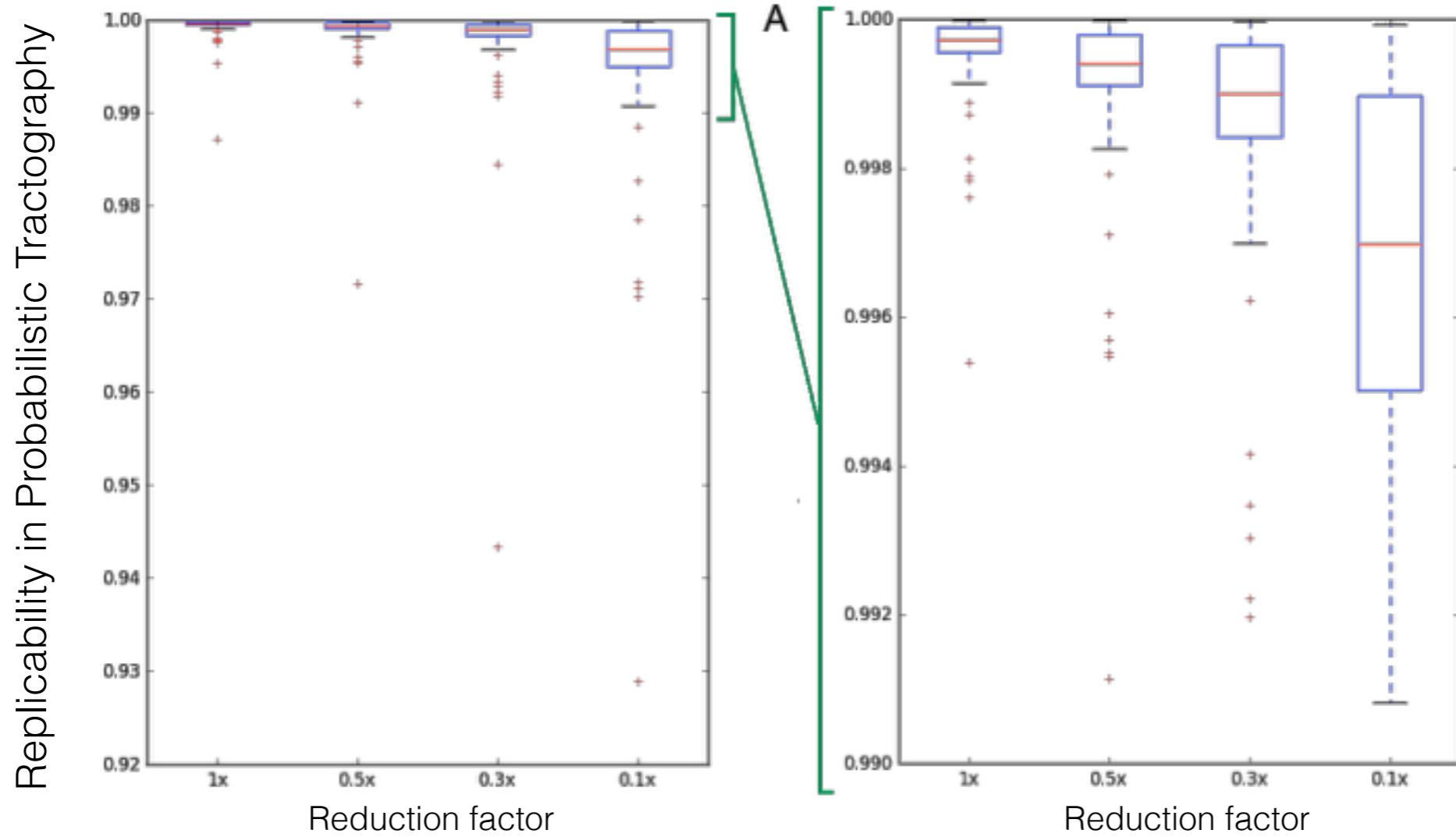


**Registration method
for dMRI to MNI**

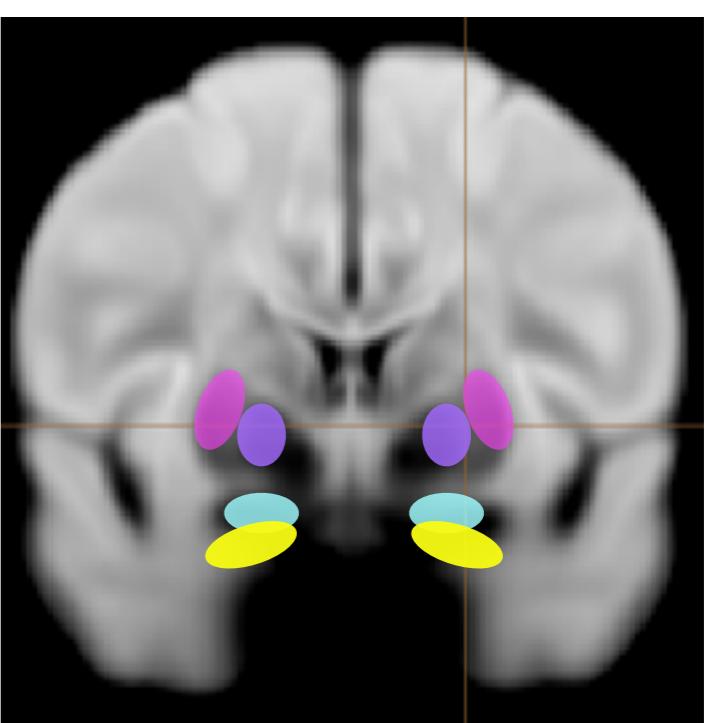
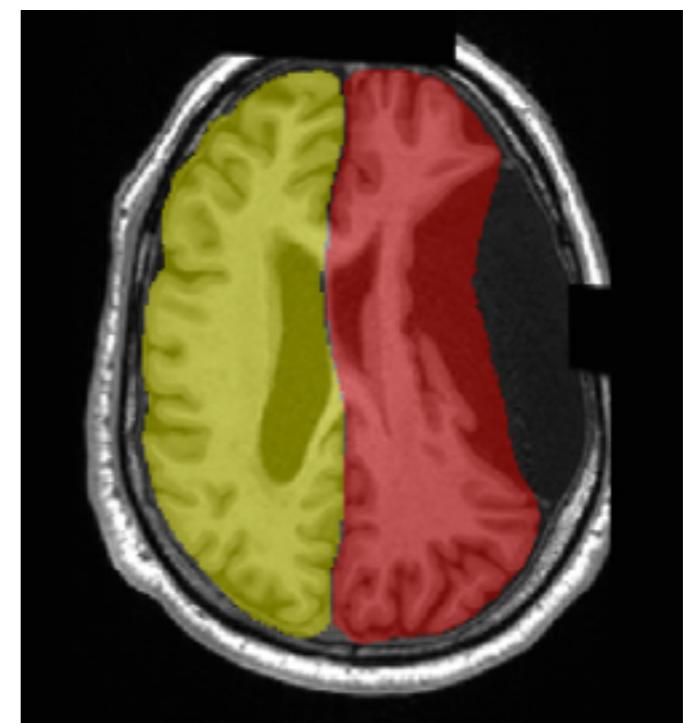
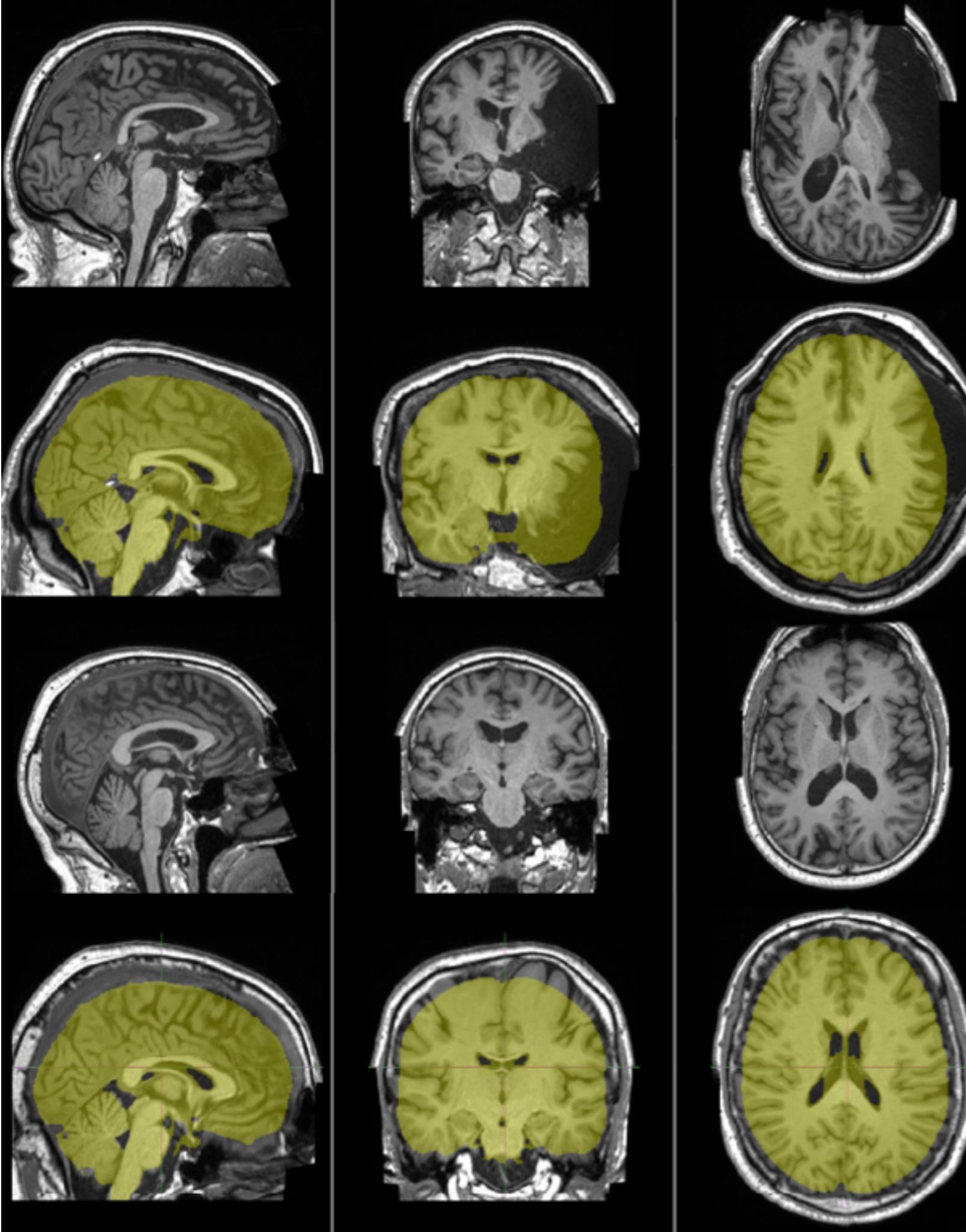
- 1: FA linearly aligned to T1 + T1 non-linearly aligned to MNI.
- 2: FA linearly aligned to T1 + T1's WM non-linearly aligned to MNI's WM.
- 3: FA linearly aligned to T1 + T1's GM non-linearly aligned to MNI's GM.
- 4: Corrected B0 linearly aligned (BBR) to T1 + T1 non-linearly aligned to MNI.
- 5: Corrected B0 linearly aligned (BBR) to T1 + T1's WM non-linearly aligned to MNI's WM.
- 6: Corrected B0 linearly aligned (BBR) to T1 + T1's GM non-linearly aligned to MNI's GM.
- 7: FA non-linearly aligned to T1 + T1 non-linearly aligned to MNI.
- 8: FA non-linearly aligned to T1 + T1's WM non-linearly aligned to MNI's WM.
- 9: FA non-linearly aligned to T1 + T1's GM non-linearly aligned to MNI's GM.
- 10: FA linearly aligned (BBR) to T1 + T1 non-linearly aligned to MNI.
- 11: FA linearly aligned (BBR) to T1 + T1's WM non-linearly aligned to MNI's WM.
- 12: FA linearly aligned (BBR) to T1 + T1's GM non-linearly aligned to MNI's GM.
- 13: FA non-linearly aligned to FA FMRIB58 atlas via a study-specific template.
- 14: FA non-linearly aligned to FA FMRIB58 atlas using high-dimensional FNIRT-based warping.

Decisions taken building the pipeline

Number of seeds per voxel for probabilistic tractography

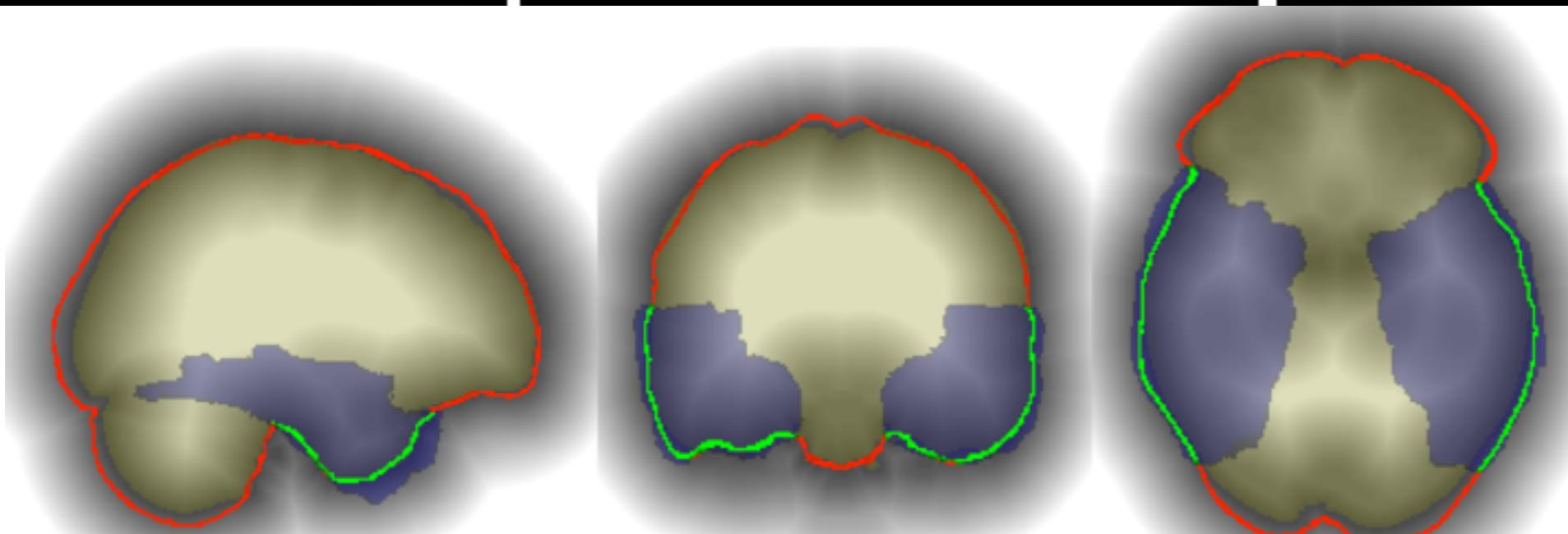
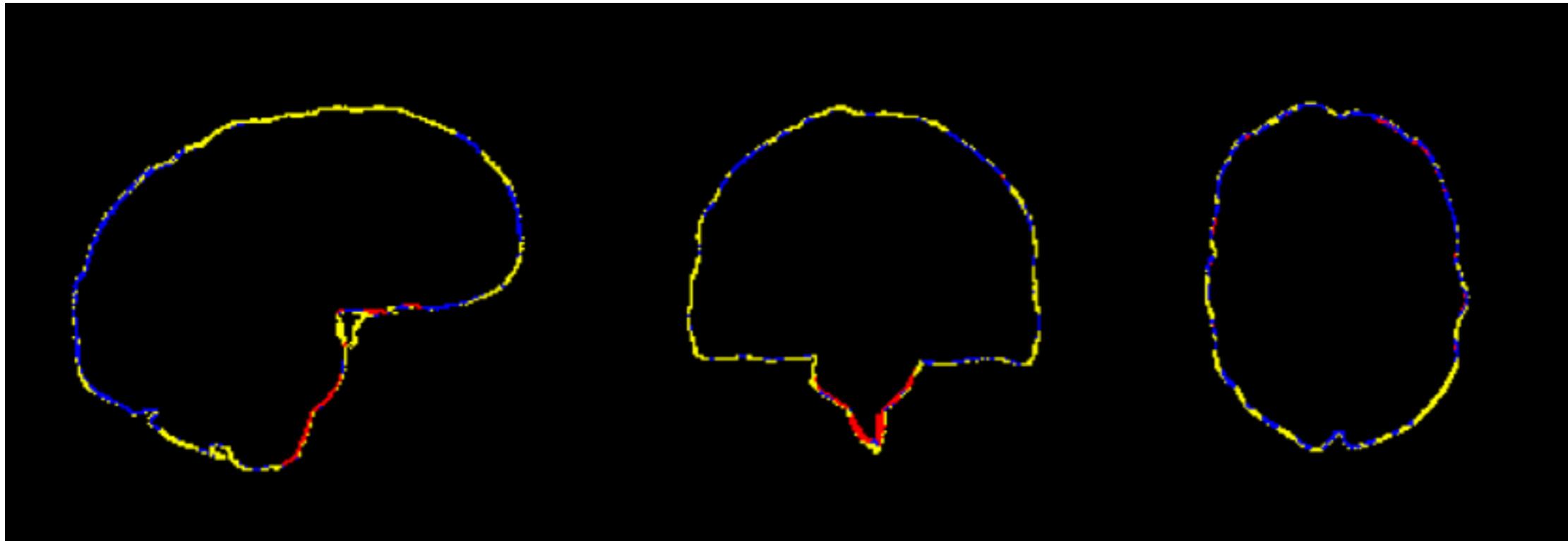


Registration Issues & QC



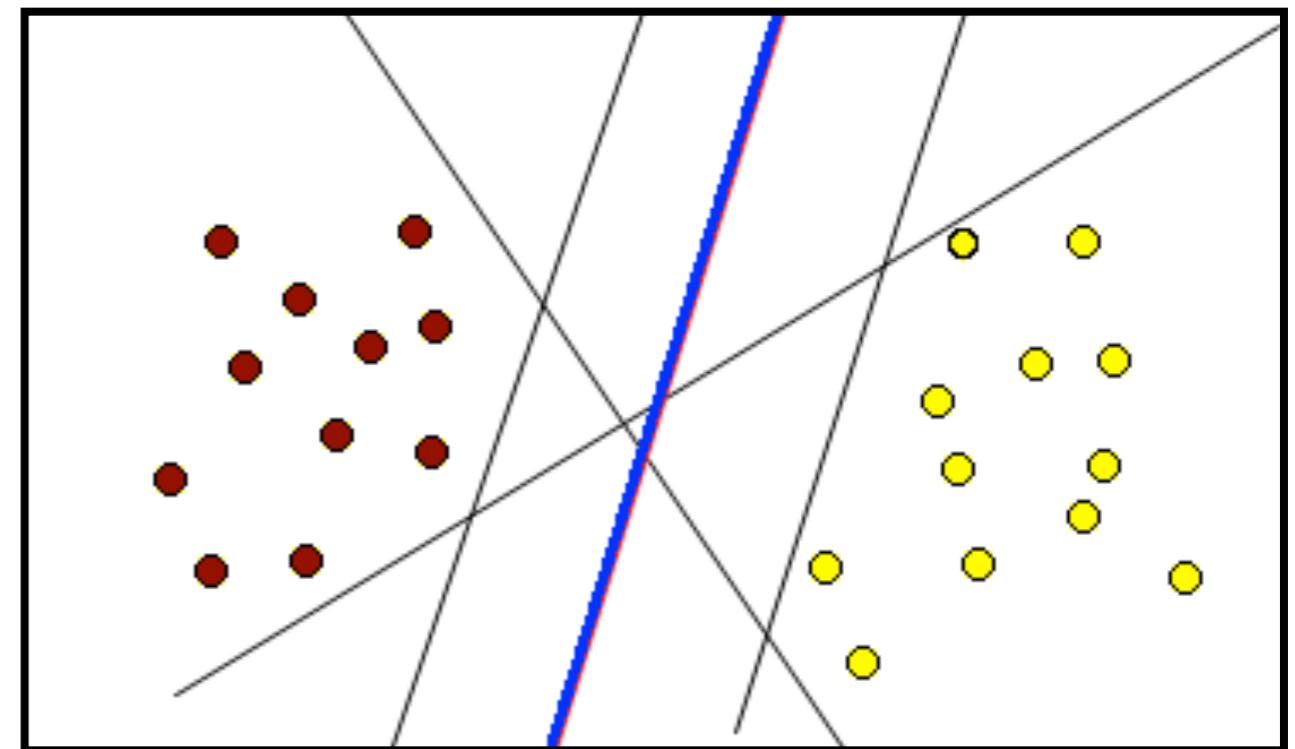
Normalised intensity per
subcortical structure

Some more QC metrics



Quality Control Metrics

- 190 QC features for T1.
- 5815 subjects manually labelled in QC terms.
- 98 (1.68%) bad quality images found.



Ensemble of classifiers

QC Results

10-fold stratified cross validation

		Real Positive	Real Negative	
Classified Positive	TP: 90	FP: 747	14.4%	
	FN: 8	TN: 4970	P. Precision: 0.11	
Classified Negative	Sensitivity: 0.92	Specificity: 0.87	N. Precision: 0.998	
0.13%			Accuracy: 0.87	

Positive = Low quality / artefacts dataset

Negative = Usable dataset



- **October 2015: ~6000 subjects' data were released**
- **February 2017: ~4000 subjects' data were released**
- **Brain imaging**
 - Raw+processed NIFTI images available for all 6 modalities
 - 4350 released IDPs usable by non-imaging-experts
- **4500-subject multimodal brain templates** tinyurl.com/ukbrain
(also: matlab-code and results for IDP processing from Nat Neur paper; files for replicating acquisition protocol)



Papers using Biobank Brain Imaging Data

- Miller et al. (2016). *Multimodal population brain imaging in the UK Biobank prospective epidemiological study*. **Nature Neuroscience**.
- Cox et al. (2016). *Ageing and brain white matter structure in 3,513 UK Biobank participants*. **Nature - Communications**.
- Reus et al. (2017). *Association of polygenic risk for major psychiatric illness with subcortical volumes and white matter integrity in UK Biobank*. **Nature - Scientific Reports**.
- Shen et al. (2017). *Subcortical volume and white matter integrity abnormalities in major depressive disorder: findings from UK Biobank (N=4446)*. **Uploaded to bioRxiv**.
- Wigmore et al. (2017). *Do Regional Brain Volumes and Major Depressive Disorder Share Genetic Architecture: a study in Generation Scotland (n=19,762), UK Biobank (n=24,048) and the English Longitudinal Study of Ageing (n=5,766)*. **Uploaded to bioRxiv**.

OHBM 2017 Abstracts using Biobank Brain Imaging Data (FMRIB).

- Alfaro Almagro et al. *Update on UK Biobank Brain Imaging: First 10,000 subjects and new Imaging Derived Phenotypes*.
- Visser et al. *Subcortical shape analysis using a temporal model reveals nonlinear development of atrophy with age*.
- Heise et al. *APOE genotype affects volume but not iron content of subcortical structures in the UK Biobank population study*.
- Mollink et al. *Fibre dispersion in the corpus callosum relates to interhemispheric functional connectivity*



Data Access <http://www.ukbiobank.ac.uk/register-apply>

- Open for use by researchers worldwide
- Access application needed, primarily to ensure protection of sensitive subject data
- Modest data access fee (~£2.5k including access to imaging data), to ensure that the resource is maintainable indefinitely
- No preferential access to scientists helping run UK Biobank !

Future Big Data Needs



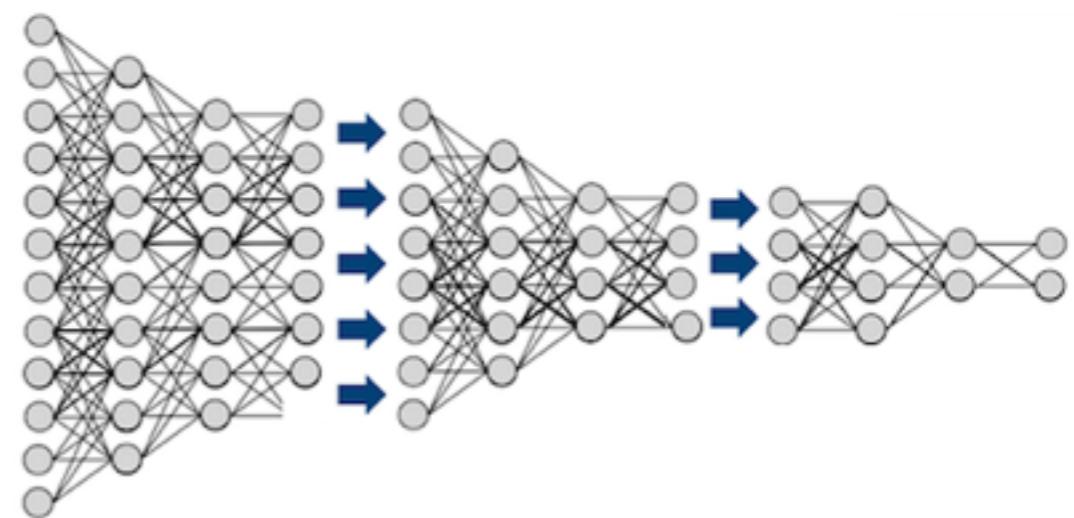
- ~10 GB per subject = ~1 PB total data
- ~27 CPU hours and 0.62 GPU hours per subject.
- Co-modelling IDPs with lifestyle data, genetics & long-term healthcare outcomes (NHS records) will be a huge data/analysis challenge.
- Imaging researchers may run their own from-scratch analyses. Biobank *might* eventually offer “cloud” compute facilities attached to the database

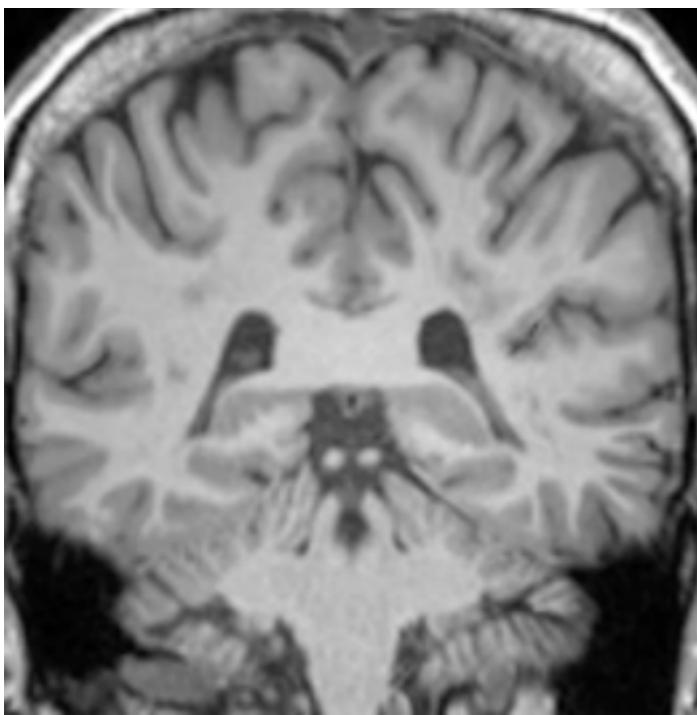
Job	Hours
T1 & T2 Processing	1.9
swMRI Processing	0.13
topup	0.3
dMRI Processing	0.16
TBSS	0.54
NODDI	0.18
Tractography	20
FIX	0.4
Melodic	2.1
FEAT	1
IDP Generation	0.17
Total CPU hours	26.88

Job	Hours
eddy	0.18
bedpostx	0.44
Total GPU hours	0.62

Future Developments

- Improving non-linear registration
- Better autoPtx masks
- Freesurfer ...
- ...and hence HCP pipelines including MSM (Multimodal Surface Matching)
- Cloud Storage / Processing?
- Unsupervised Feature Learning?





THANK YOU

Brain Imaging Contributors

Image processing pipeline: Fidel Alfaro-Almagro, Mark Jenkinson, Jesper Andersson, Stamatis Sotiroopoulos, Saad Jbabdi, Ludovica Griffanti, Gwenaelle Douaud, Eugene Duff, Moises Hernandez Fernandez, Emmanuel Vallee, Gholamreza Salimi-Khorshidi (FMRIB, Oxford)

Scientific direction: Stephen Smith, Karla Miller (FMRIB, Oxford), Paul Matthews (Imperial)

Additional input on acquisitions/protocols/reconstruction/processing: Neal Bangerter (Brigham Young), Kamil Ugurbil, Essa Yacoub, Steen Moeller, Eddie Auerbach (CMRR, U Minnesota), Junqian Gordon Xu (Mount Sinai), David Thomas, Daniel Alexander, Gary Zhang, Enrico Kaden (UCL), Alessandro Daducci (EPFL), Tony Stoecker (Rhineland Study/Bonn), Stuart Clare, Heidi Johansen-Berg (FMRIB, Oxford), Deanna Barch, Greg Burgess, Nick Bloom, Dan Nolan, Michael Harms, Matt Glasser (Washington U), Doug Greve, Bruce Fischl, Jonathan Polimeni (MGH), Andreas Bartsch (Heidelberg), Anna Murphy (Manchester), Fred Barkhof (VU Amsterdam/UCL), Christian Beckmann (Donders Nijmegen), Chris Rorden (U South Carolina), Peter Weale, Iulius Dragonu (Siemens UK), Steve Garratt (Project Manager, UK Biobank Imaging), Sarah Hudson (Lead Radiographer, UK Biobank Imaging)

IT/informatics: Duncan Mortimer, David Flitney, Matthew Webster, Paul McCarthy (FMRIB, Oxford), Alan Young, Jonathan Price, John Miller (CTSU, Oxford)

We are also extremely grateful to all UK Biobank study participants