Biobank-scale Brain Imaging Genetics:

Clinical and Methodological Advances

University of North Carolina at Chapel Hill

Hongtu Zhu

Joint works with all members of the UNC BIG-S2 Lab,

Bingxin Zhao, Yun Li, Stephen Smith, and Jason Stein

https://www.med.unc.edu/big-s2

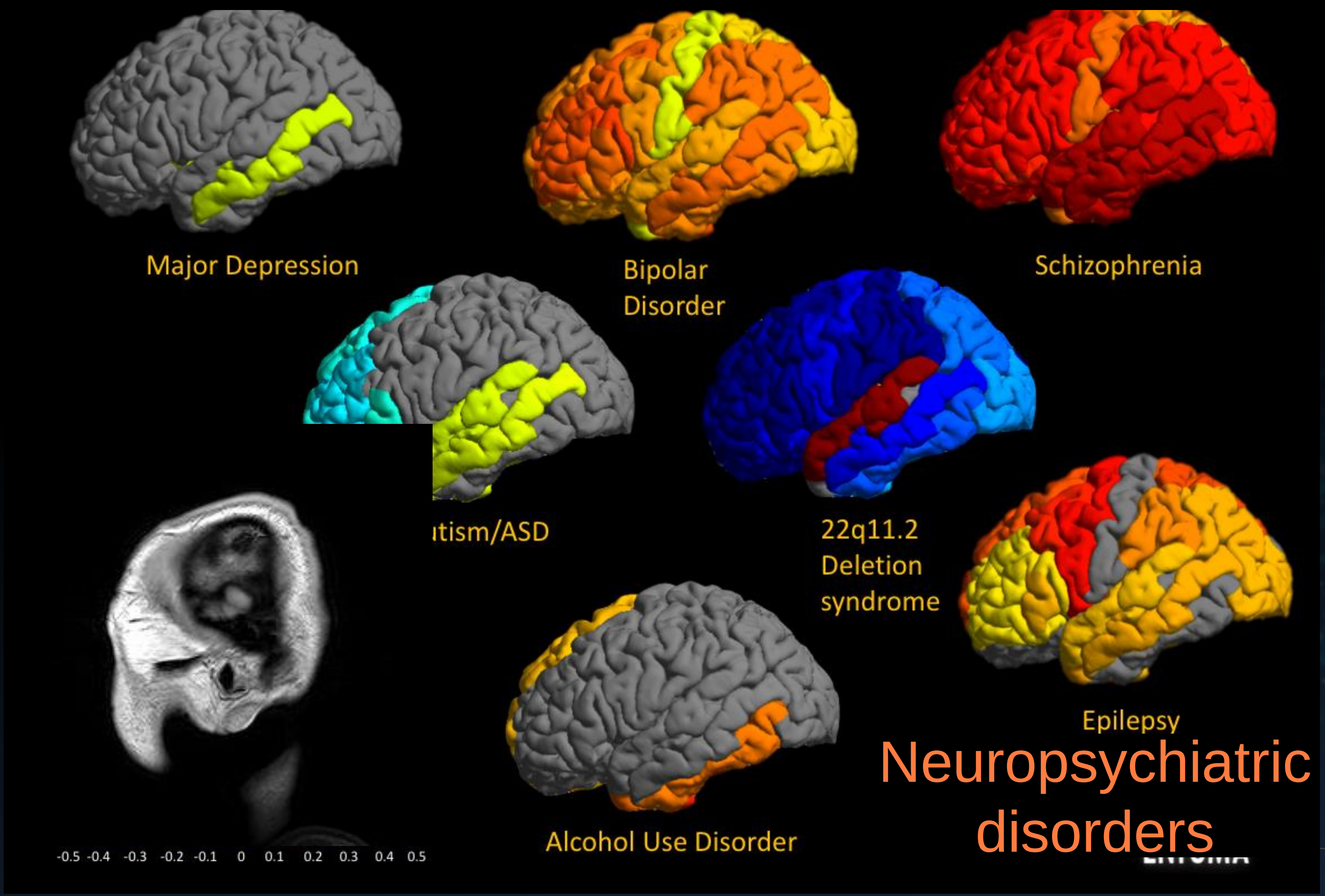
CONTENTS

Part I

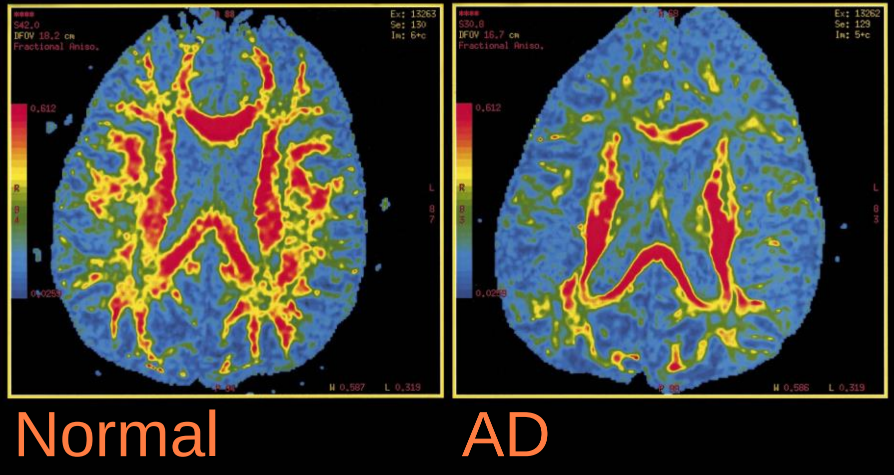
Big Data in Imaging Genetics

Brain Imaging for Brain Disorders

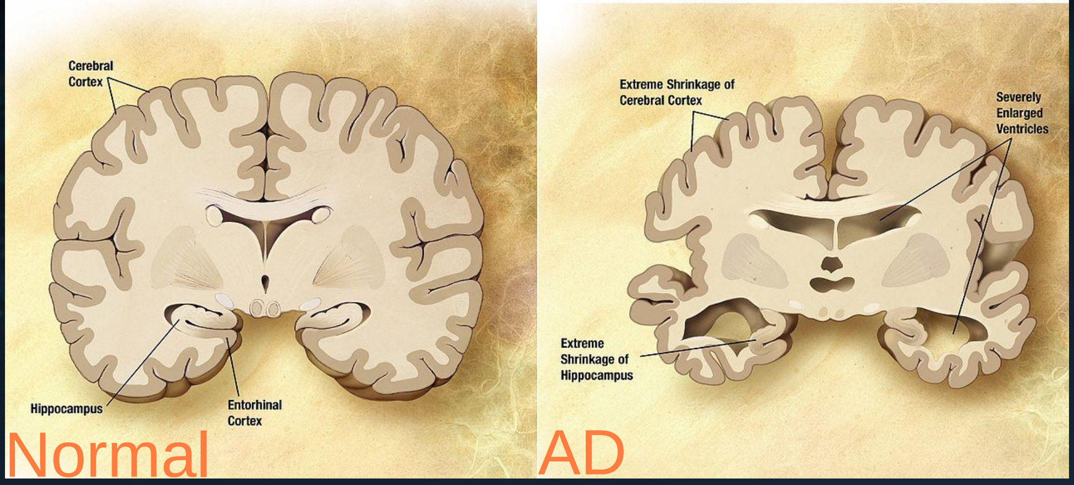
Capture the brain structure and function changes associated with major brain-related disorders and normal development



Neuropsychiatric disorders



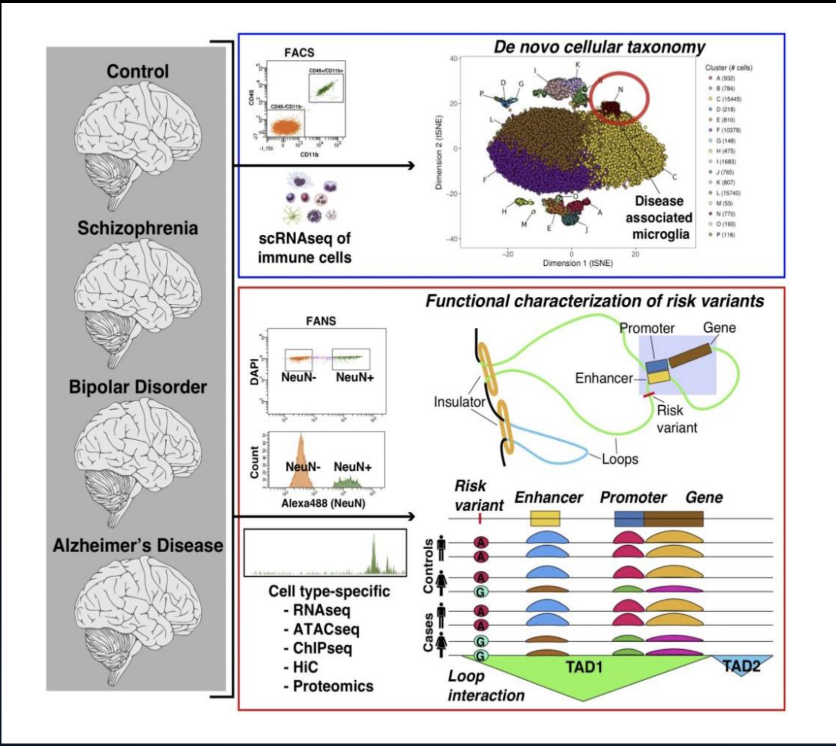
Alzheimer's disease (AD) is associated with brain shrinkage

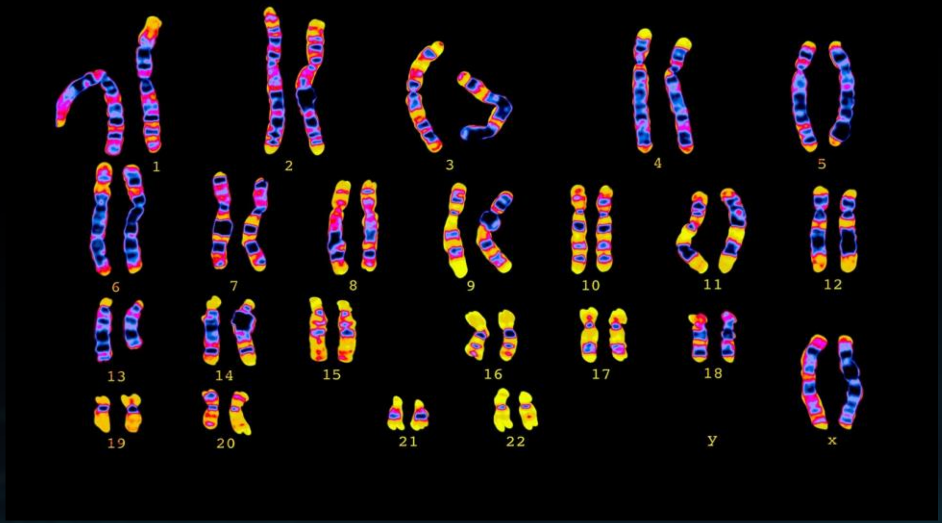


Genetics of Brain Disorders

Most major brain disorders (like AD) are **heritable complex traits/diseases**

Together 50%-70% of AD risk
75%-90% of ADHD risk
60%-85% of Schizophrenia risk
~80% of Autism Spectrum Disorder (ASD) risk





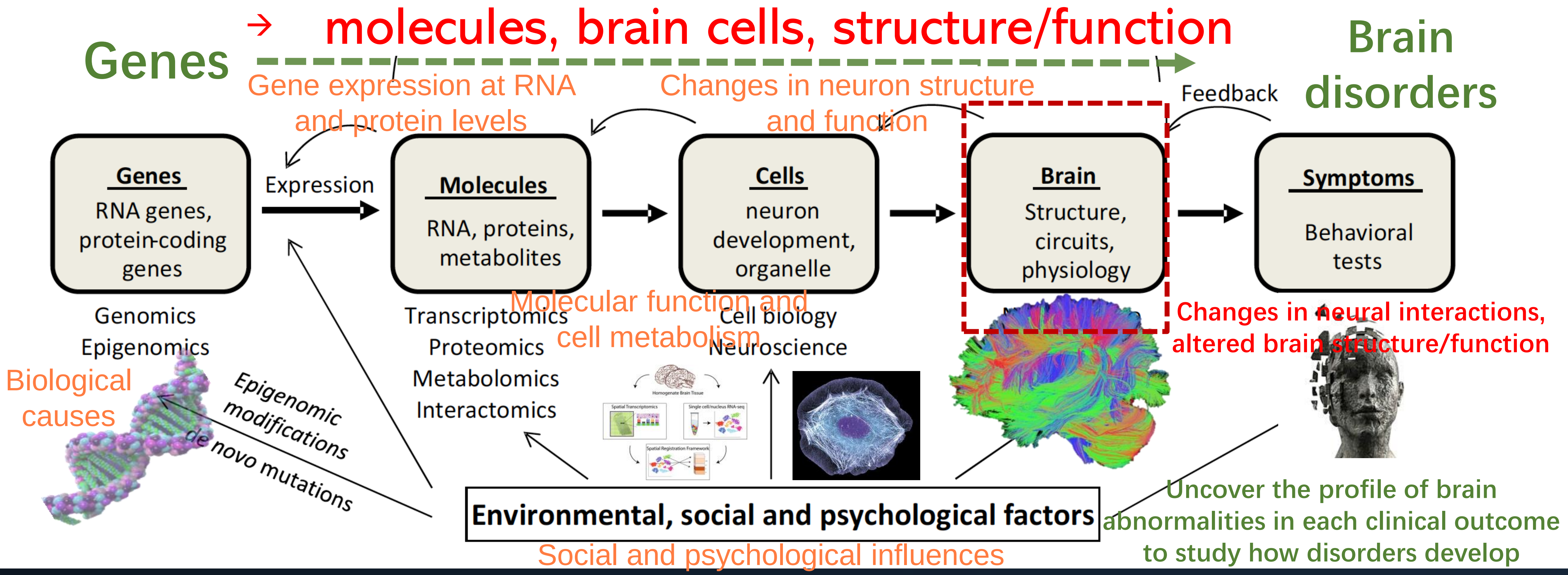
Complex traits/diseases (many genes, environmental factors, complex functional mechanism)

Genetic signals are non-spare and weak:
Need large sample size to detect weak signals

Many genes contribute to the risk of AD
(polygenic genetic architecture)
(small but nonzero contribution)

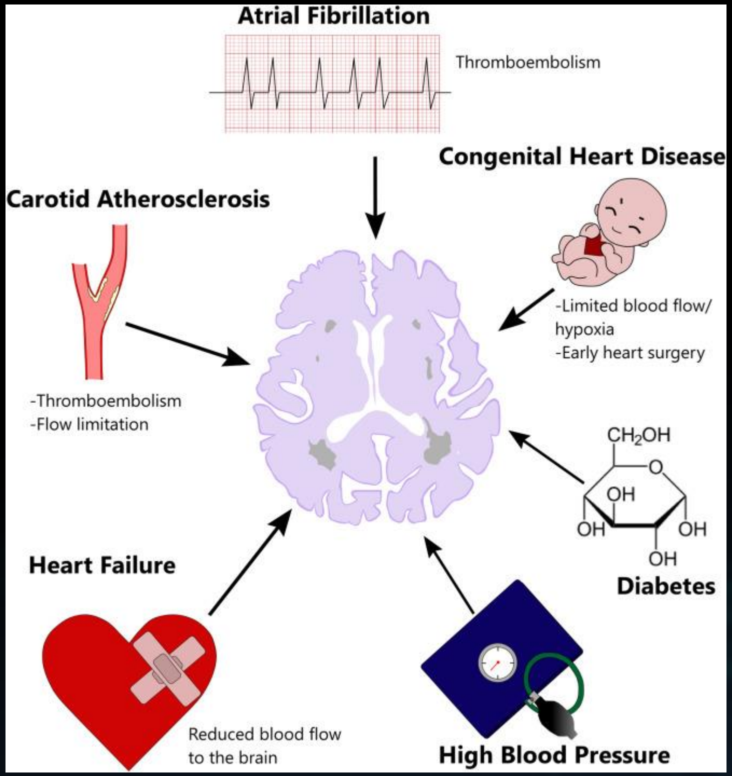
Brain Imaging Genetics Paradigm

Neuroimaging: an important component to help understand the complex biological pathways of brain disorders

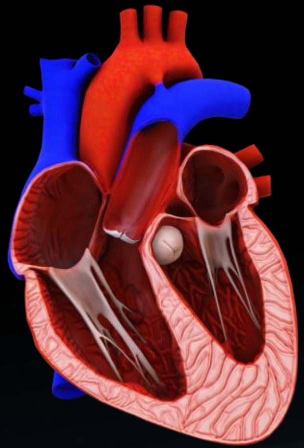


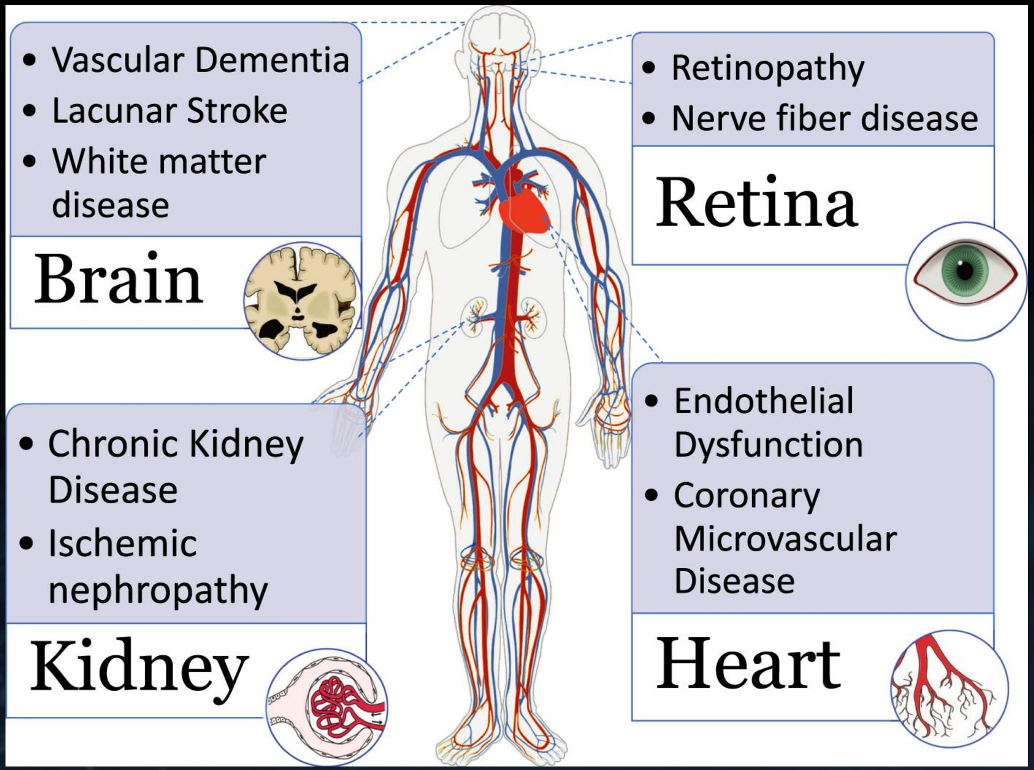
Cardiovascular Disease & Brain Health

(Neuro)imaging: help understand the complex interplay between brain and other human organs and their underlying genetic overlaps



Possible causal factors of brain structure changes, resulting in brain disorders like stroke, dementia and cognitive impairment





Many diseases (e.g., microvascular disease, high blood pressure) are multisystem disorders

Long-term Challenges in Brain Imaging Genetics

- Traditionally, neuroimaging data are expensive and have very limited sample size (n ~ 100)
- On the other hand, genetic risk factors are typically dense and have small effect size, and thus need large sample size to detect
- Imaging batch effects/confounders (e.g., image acquisition, processing procedures, and software)

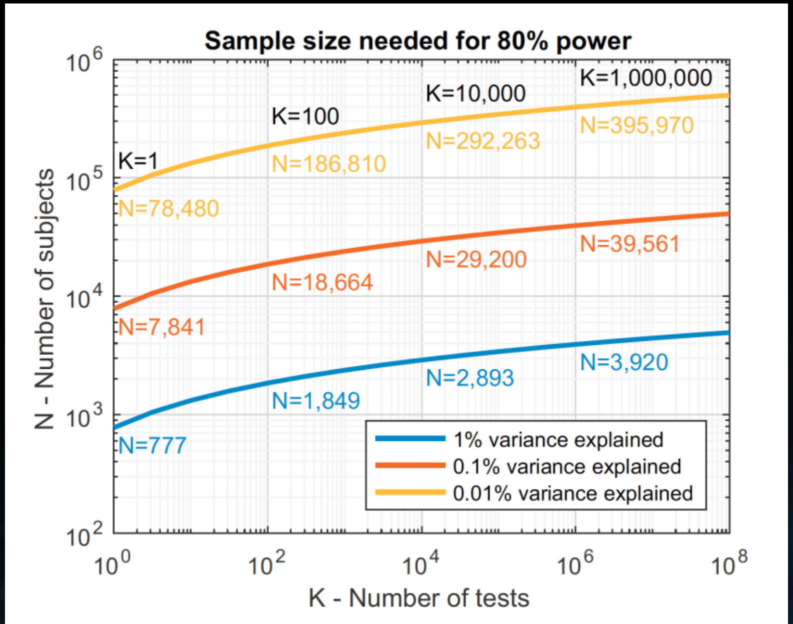


Table 1. Selected Examples of Imaging Confounds with a Subset of Image Artifacts and Potential Correlates

Confound	Example Effects on MRI Data	Potential Artifactual Correlates	Comments
Head motion	Striping, ringing, blurring, dMRI dropout, low SNR, biased connectivity	Diseases (PD, ADHD) and aging correlate with increased head motion	Relates to head size; may be estimated from and partially corrected in fMRI and dMRI
Breathing rate/depth	Changes in fMRI contrast, SNR, distortion and dropout (due to B0)	COPD, heart conditions, BMI, exercise levels, some fMRI tasks	Can cause changes in real and apparent head motion and blood oxygenation/flow
Blood pressure	BOLD contrast (fMRI) and vascular compartment size (dMRI)	Functional connectivity (fMRI), and white matter microstructure (dMRI) in disease	–
Age	Structural atrophy (cortical thinning, ventricle enlargement) influences voxel partial volume effects	Non-volumetric imaging measures; interaction with disease progression	If age is not of explicit interest, it should generally be included as a confound
Scanner hardware	Differences in SNR, contrast or artifact as a function of site or date (all MRI modalities)	Other measures varying with site or date	Can occur even in studies run with "identical" hardware
Operator inconsistency	Differences in SNR, artifacts, distortion, coverage	Other measures varying with site or date	Even with automated protocol, subject placement or instructions can vary

SNR, signal to noise ratio; PD, Parkinson's disease; ADHD, attention deficit hyperactivity disorder; COPD, chronic obstructive pulmonary disease; BMI, body mass index.

"Big Data" Brain Imaging Genetics Cohorts

"Big data" Brain imaging genetics datasets become available in recent few years

Systematically collect publicly available individual-level data for **> 50k** individuals

Build the largest database in this field

Aging Brain

BCP
(Age [0,5])

PING
(Age [3,21])

ABCD
(n ~ 10k, Age [9,11])

PNC
(Age [14,29])

HCP
(Age [22,35])

UK Biobank
(n ~ 100k [Ongoing], Age [40,69]))

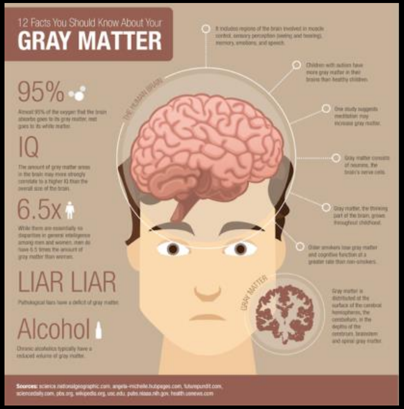
RADC
(Age > 65)

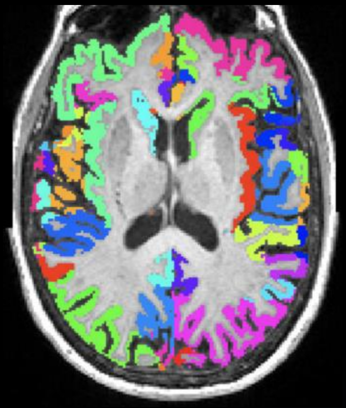
ADNI
(Age [55,92])

Brain Development

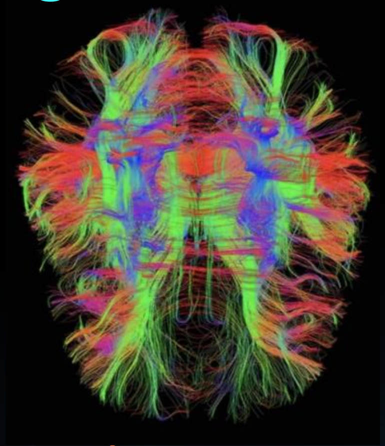
Brain Imaging Modality Examples

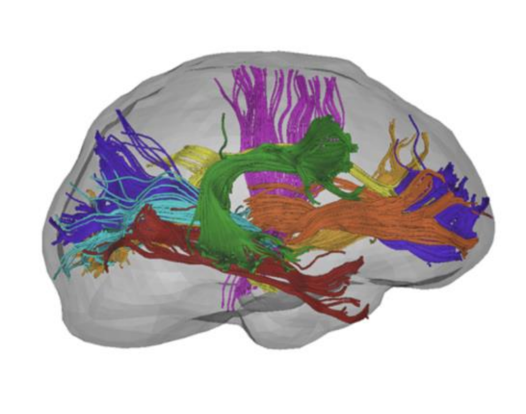
Harmonize tools/pipelines to consistently generate the full spectrum of neuroimaging features



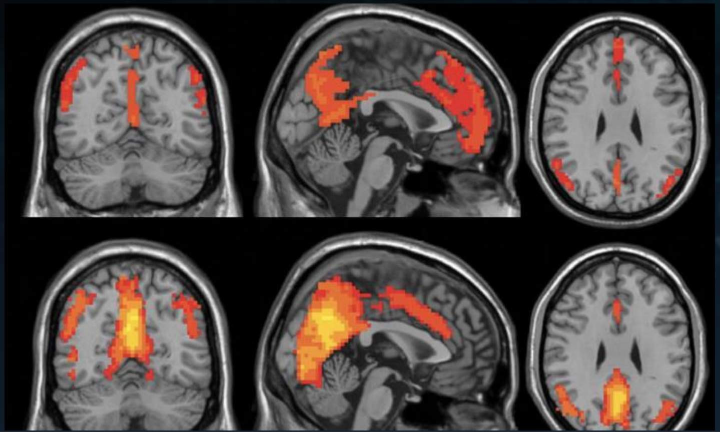


Cortical and subcortical structures





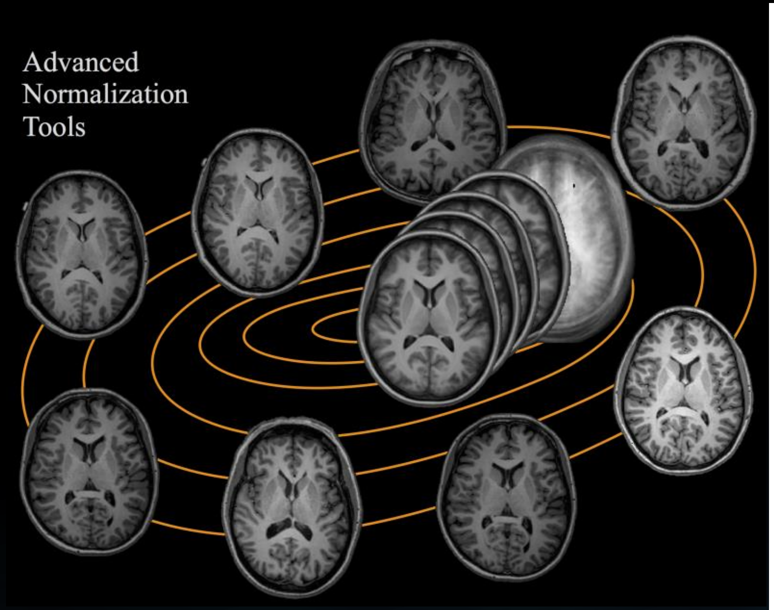
White matter microstructure (Structural connectivity, diffusion MRI)

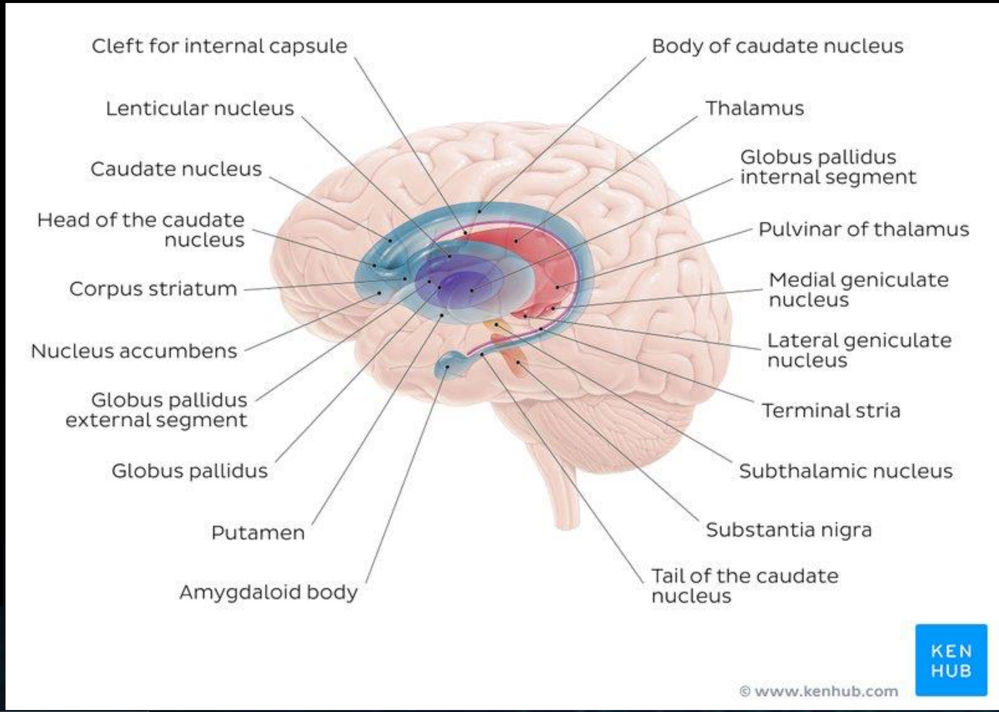


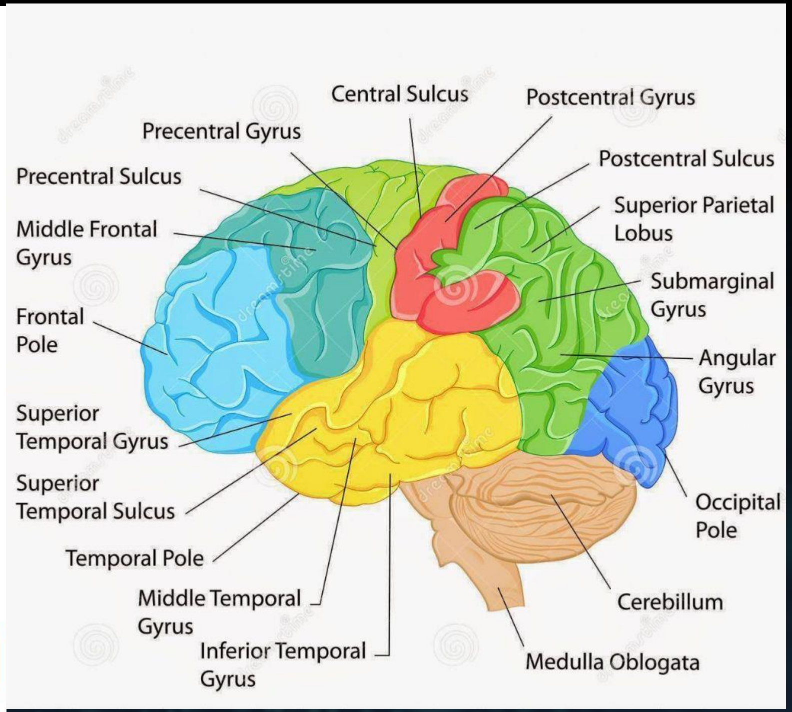
Functional networks (Functional connectivity, functional MRI)

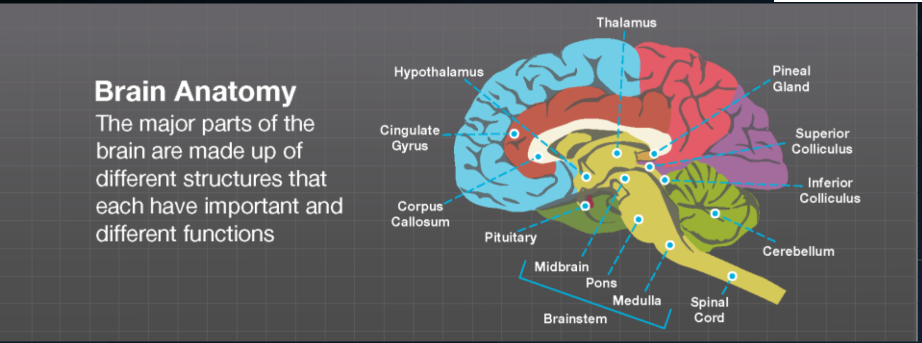
Regional Brain Volumes and Shape

Generate regional brain volumes and shape representations for 98 pre-specified brain regions and total grey matter, white matter, and brain volumes









Subcortical structures
(deep within the brain)

Cortical structures
(outer layer of the cerebrum)

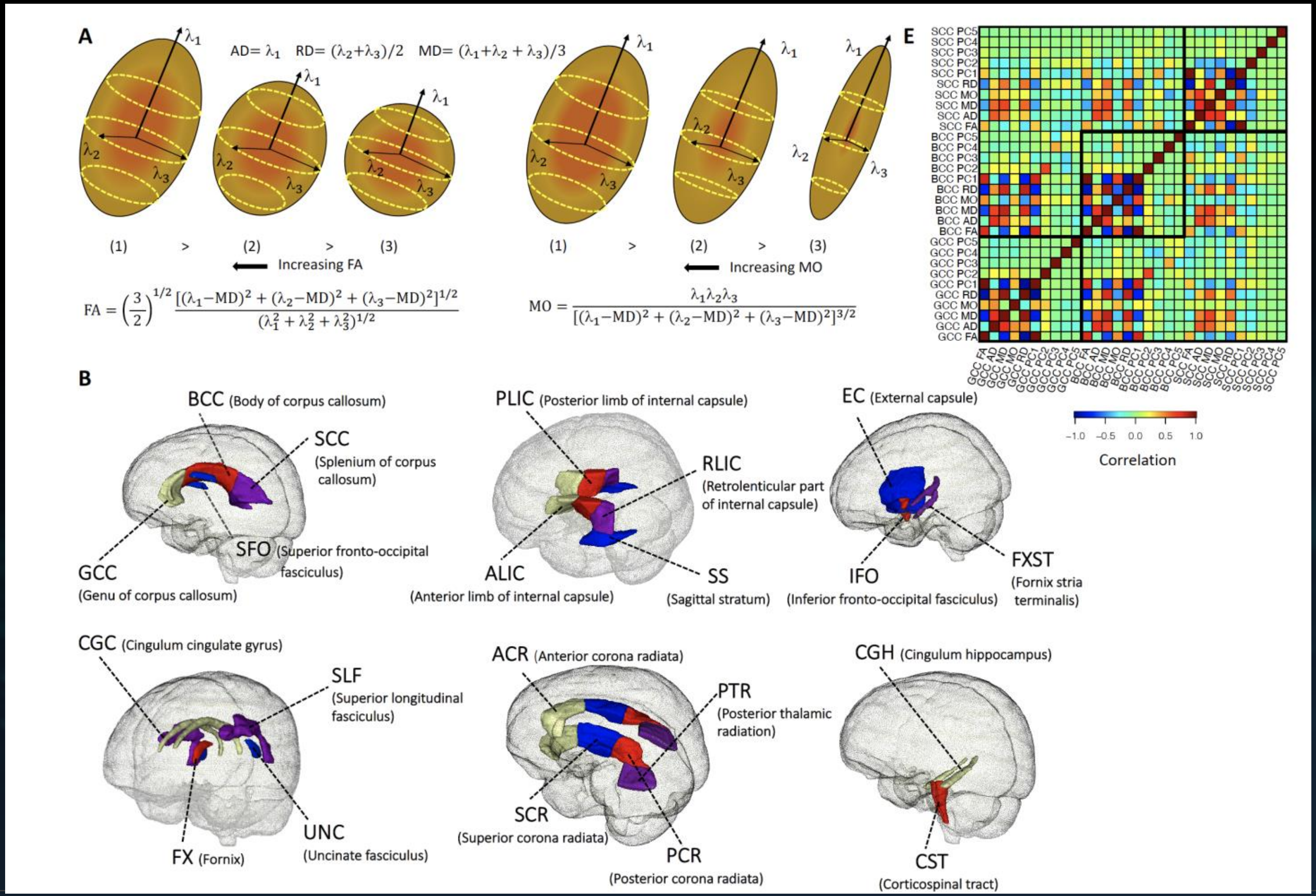
White Matter Microstructure

5 white matter microstructure measures (DTI parameters) for 21 white matter tracts

21 white matter tracts from ENIGAMA-DTI pipeline

fractional anisotropy (FA)
mean diffusivity (MD),
axial diffusivity (AD),
radial diffusivity (RD), and
mode of anisotropy (MO)

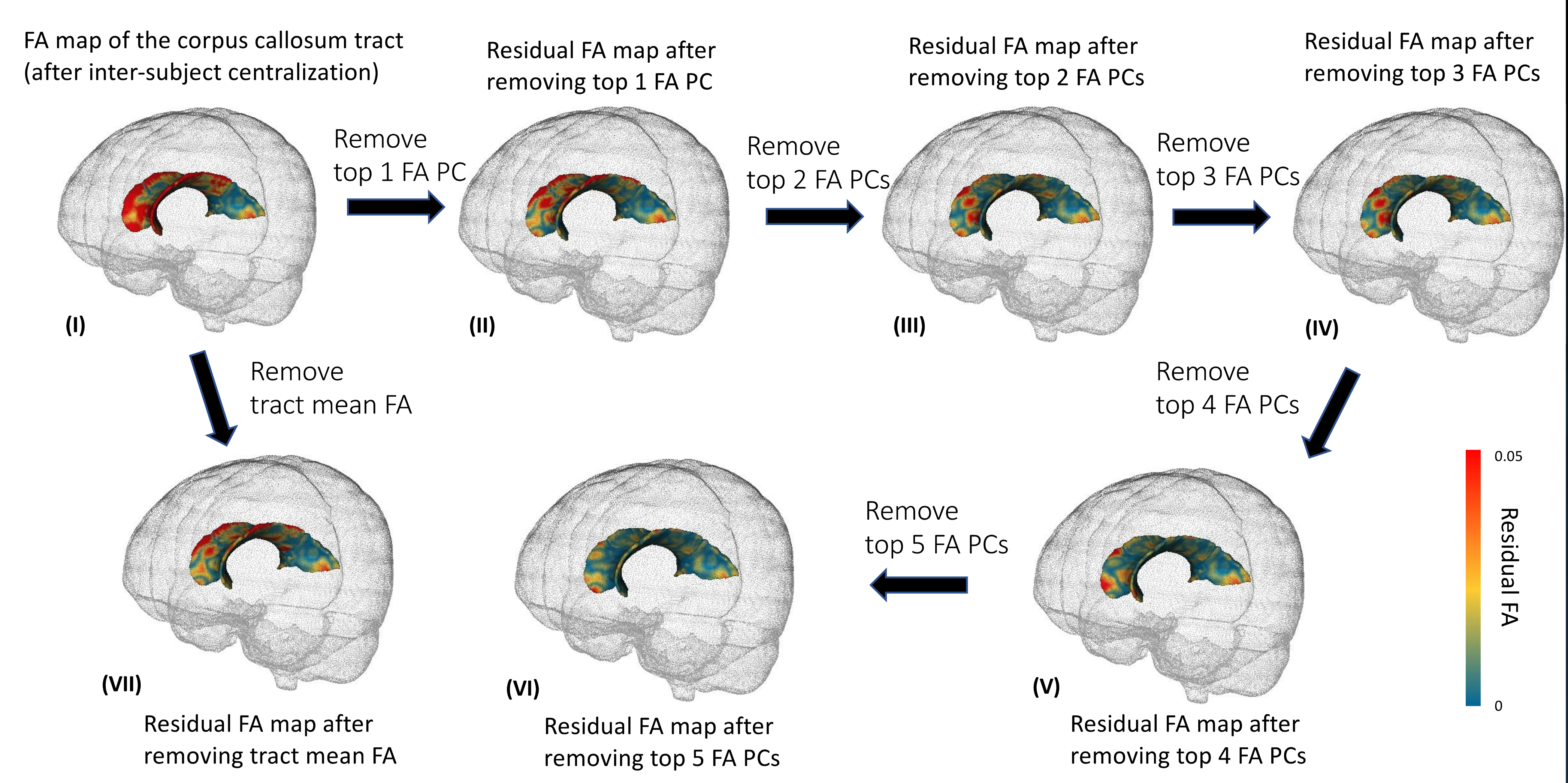
sensitive to specific types of microstructural changes and have also been widely used in clinical research



White Matter Microstructure

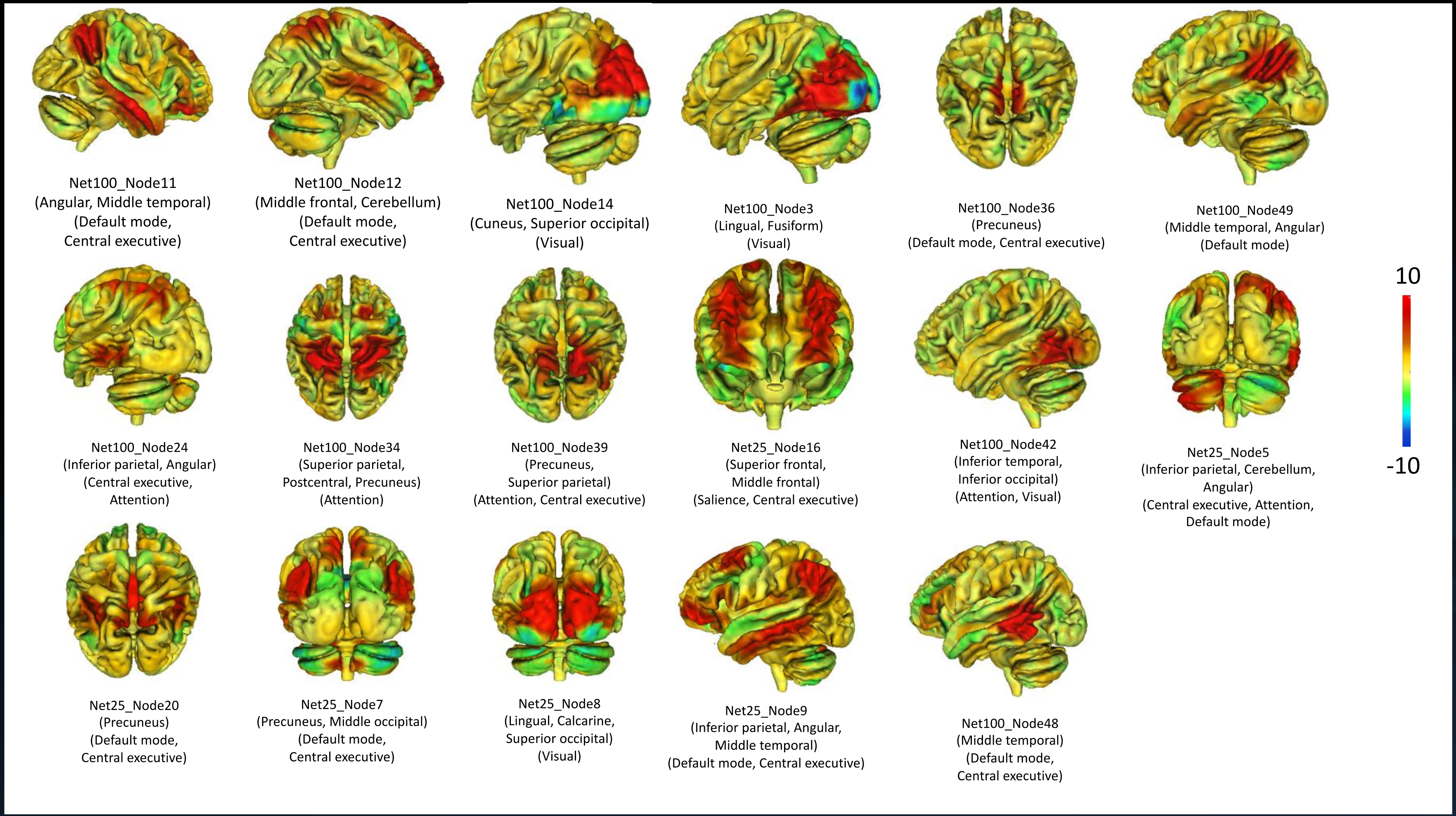
Tract-specific functional principal component analysis (FPCA) to capture major variations within each white matter tract

Beyond conventional tract-averaged mean



Resting/task functional MRI (fMRI)

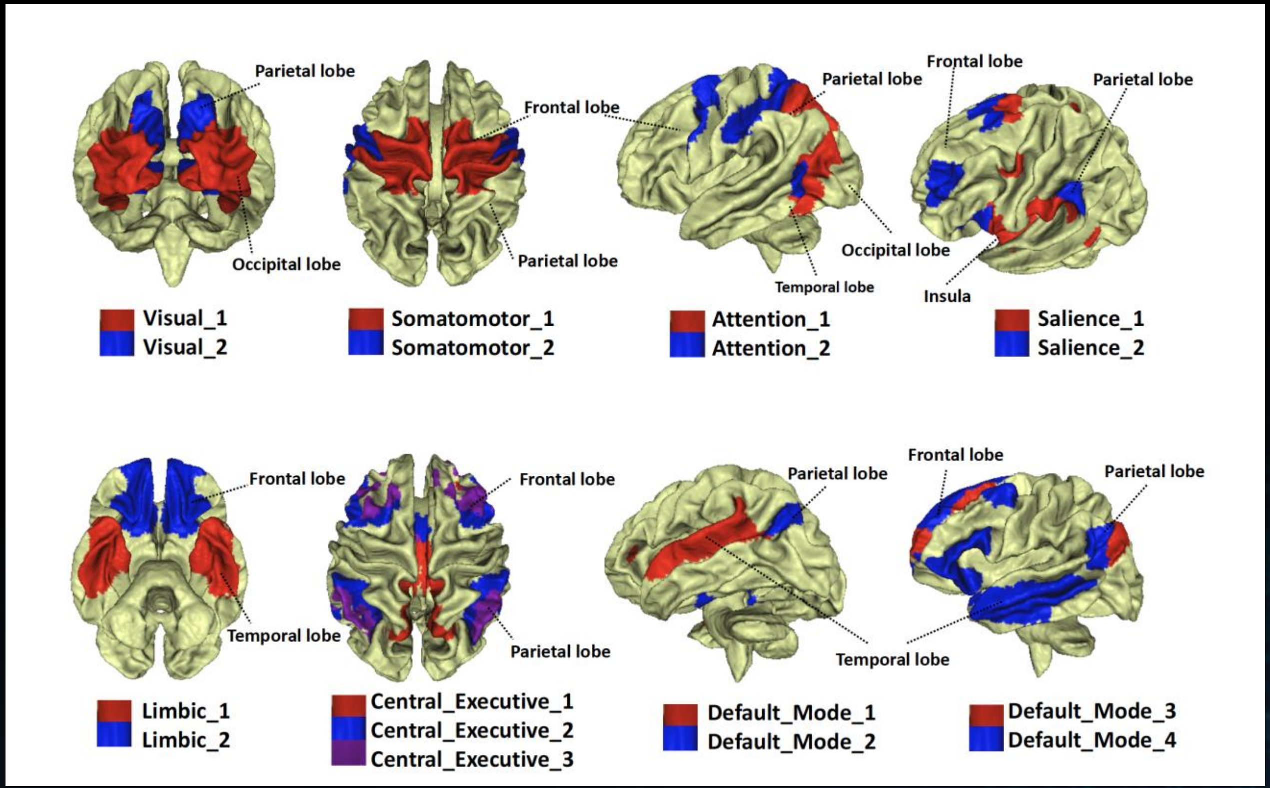
Independent component analysis (ICA)-based methods to form 76 functional regions and generate 1,701 functional connectivity traits



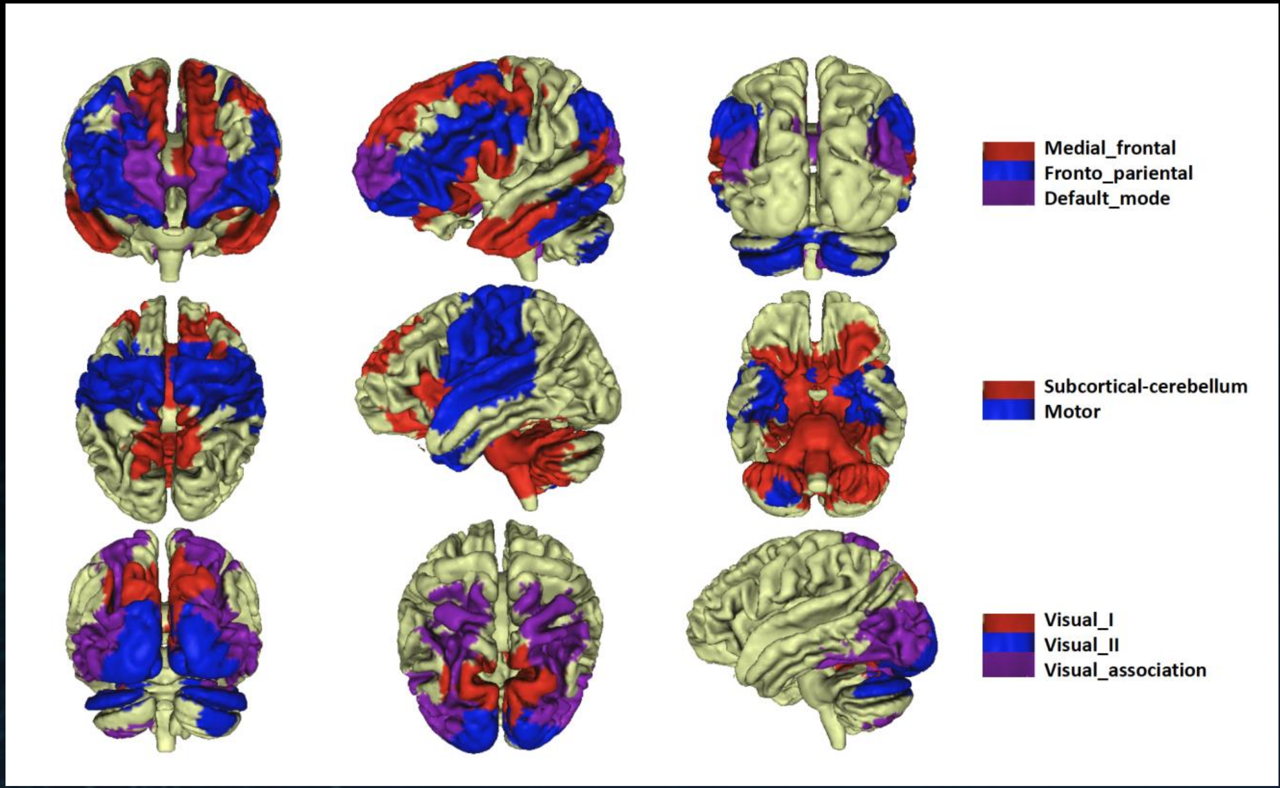
Characterize major functional brain regions and their connectivity

Resting/task functional MRI (fMRI)

Map 76 ICA-nodes onto automated anatomical labeling (AAL) parcellation and pre-defined functional networks



Seven networks in Yeo et al., 2011



Eight networks in Finn et al., 2015

Brain Imaging Genetics Data Analysis

Association tests	Identify and replicate novel genetic factors associated with brain structure and function
Causal inference/ Mediation analysis	Analyze the genetic links among brain structure, brain function, cognition, and major brain disorders.
Data integration	Integrate external genetics/genomics data (e.g., the GTEx, Hi-C chromatin interactions) to uncover new biological insights
Predictive model	Perform out-of-sample risk prediction for brain disorders using genetics, genomics, and imaging data

1) Output high-quality novel clinical findings
2) Identify, model, and address important statistical problems
3) Share our summary-level data/results to the research community

Part II

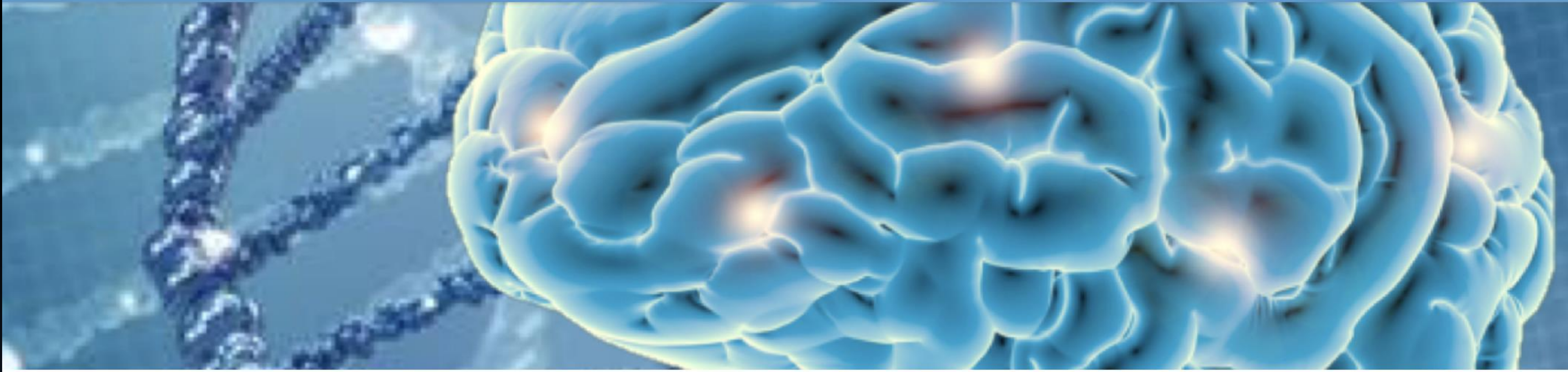
Novel Clinical findings

Brain Imaging Genetics Knowledge Portal (BIG-KP)

Genetics Discoveries in Human Brain by Big Data Integration

bigkp.org

Imaging Genetics Online Server | GWAS Summary Statistics Data Download | UNC BIG-S2 Lab | BIG-S2 Github | Other Resources



Aim to build the best knowledge database of neuroimaging genetics

GWAS Locus Browser

Brain Imaging Genetics Summary Statistics | Search for a variant, gene, or phenotype | Phenotypes | Top Hits | Random | About

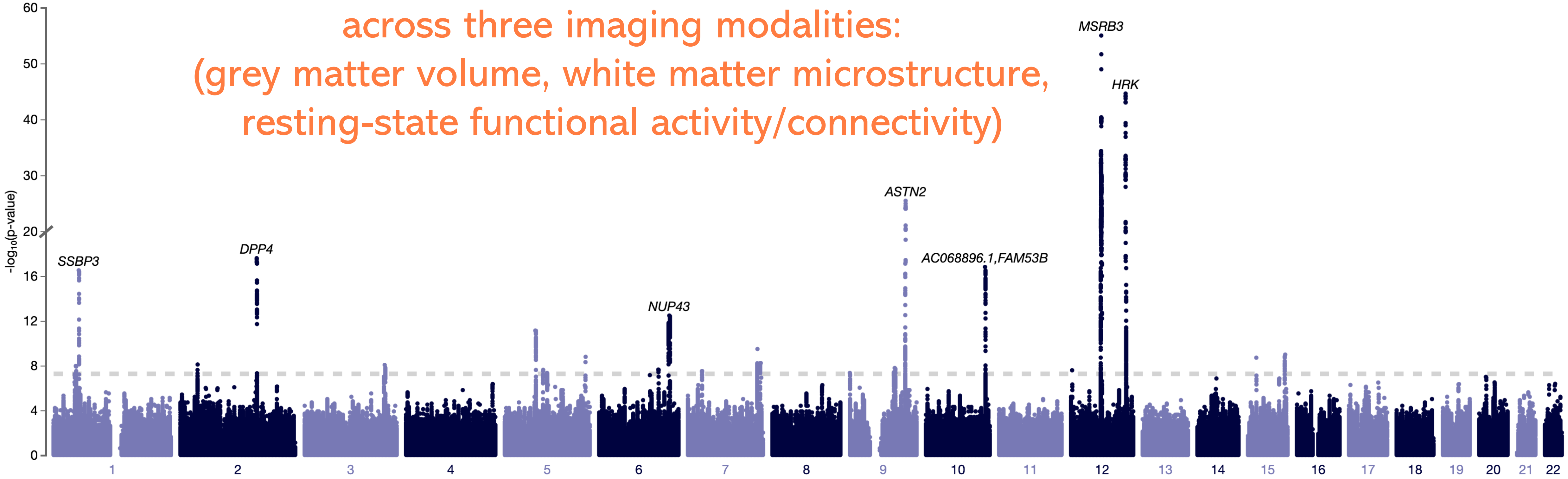
left.hippocampus

Category: **sMRI**

Download summary statistics

Manhattan QQ

Searchable database for 516 neuroimaging traits across three imaging modalities: (grey matter volume, white matter microstructure, resting-state functional activity/connectivity)



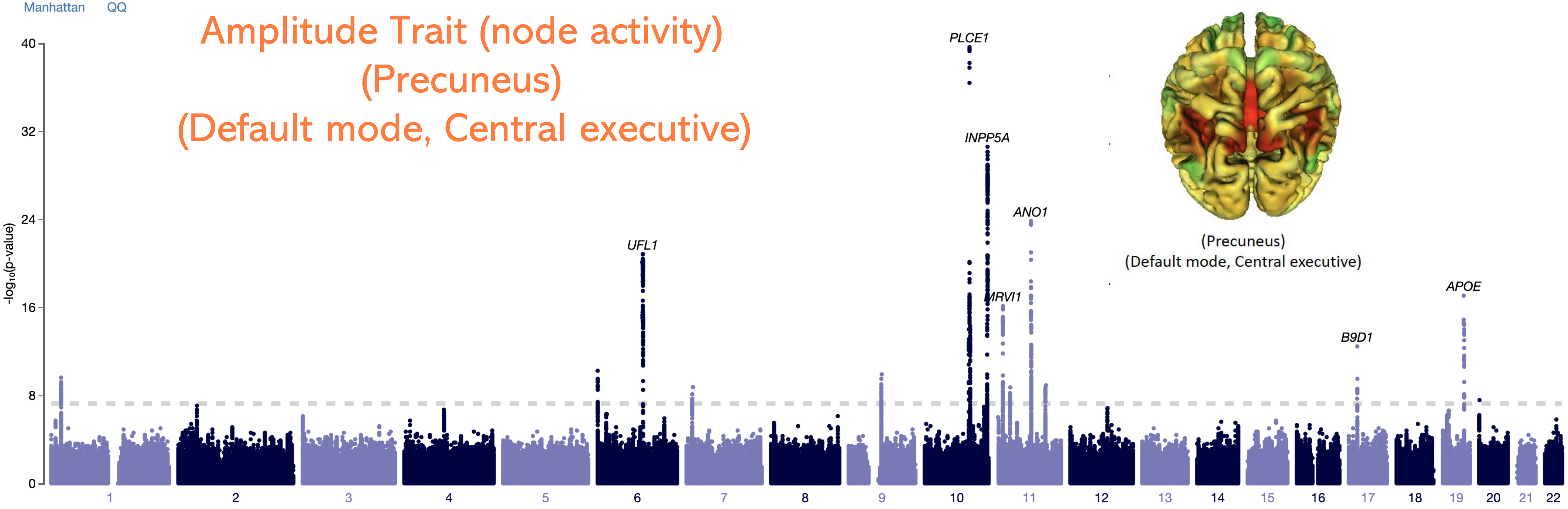
GWAS Locus Browser

Net25_Node20

Category: **rs-fMRI**

Manhattan QQ

Download summary statistics



GWAS Summary Statistics

The full set of GWAS summary statistics have been made freely available to the research community

GWAS summary statistics for 215 tract-specific diffusion tensor imaging (DTI) parameters

- **Sample size:** n=33,292
- **Version**: July 15, 2020
- Download Summary Statistics:

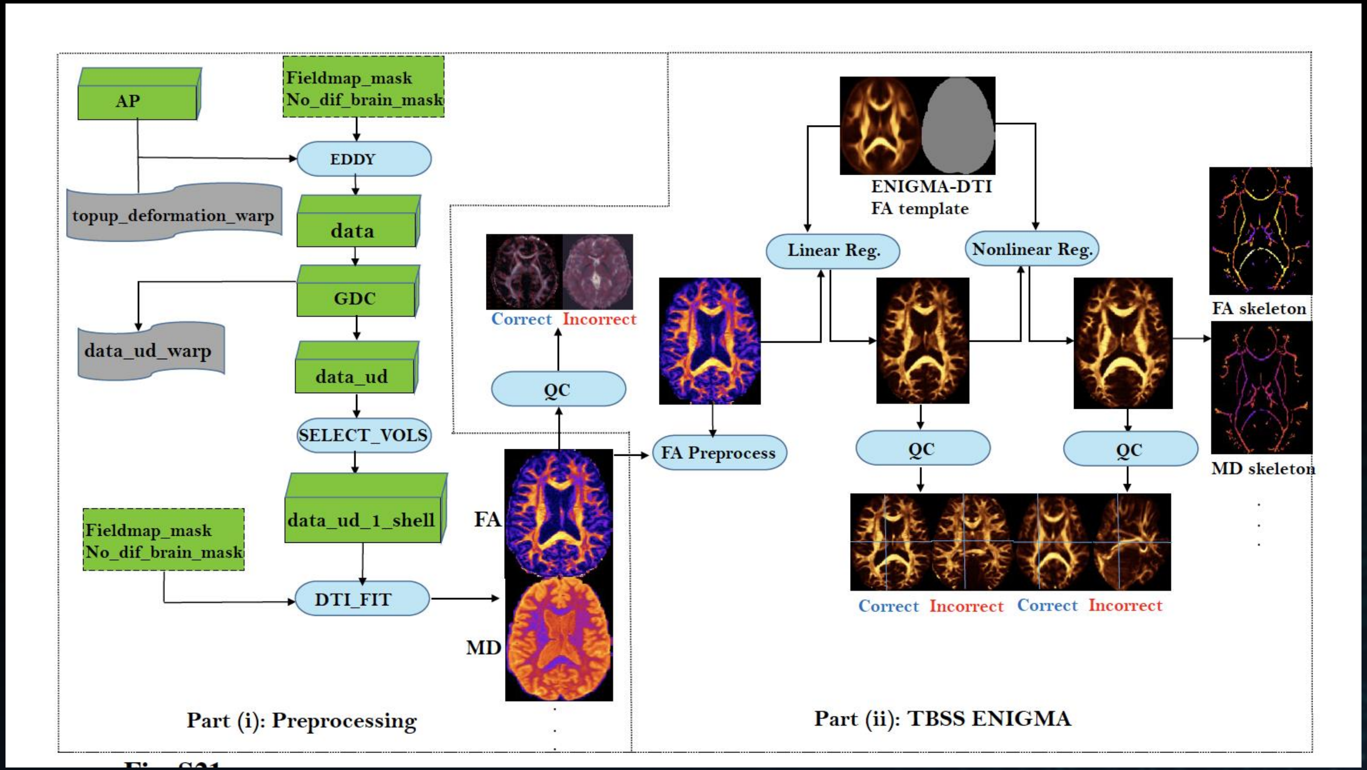
```
wget --no-check-certificate --content-disposition https://raw.githubusercontent.com/stat-yyang/sumsta
wget -i DTI.list
```

- **Description**: readme
- **Citation**: Zhao et al. (2020) Common genetic variation influencing human white matter microstructure. Preprint available at https://doi.org/10.1101/2020.05.23.112409.

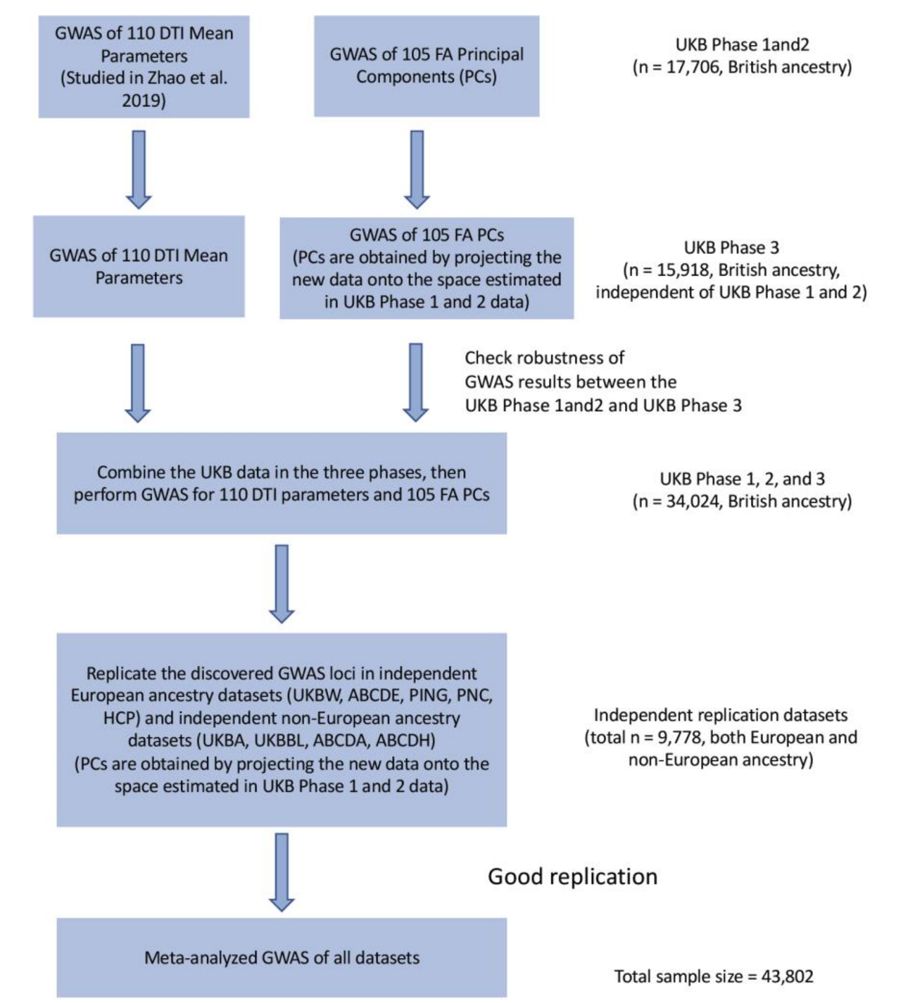
Resources with the largest sample size
(> 3,400 page views since Sep 2019)

GWAS of White Matter Tracts

Overview of the ENIGMA-DTI pipeline and the multiple-stage design in GWAS



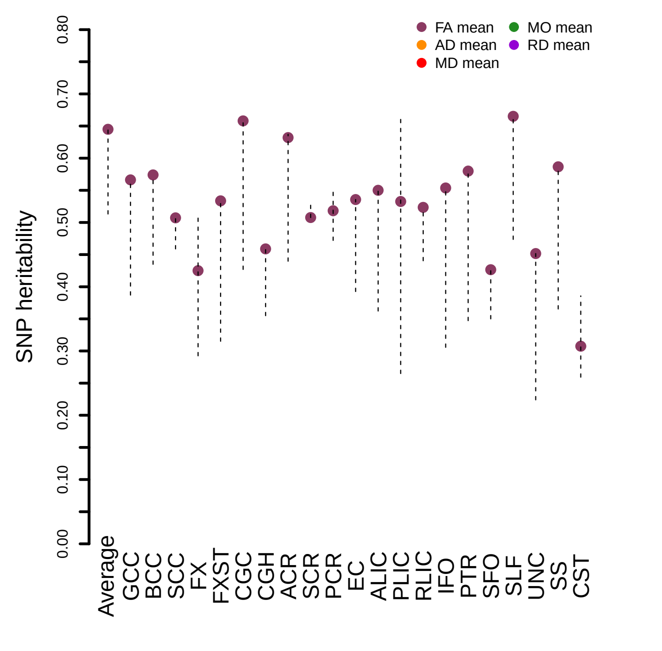
Apply the same pipeline in different datasets (UKB, ABCD, PING, PNC, HCP)

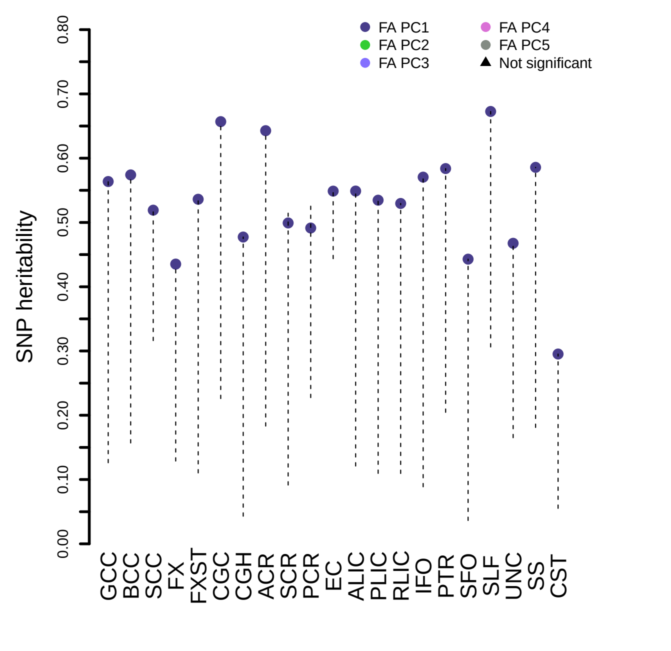


Genetic Architecture of White Matter

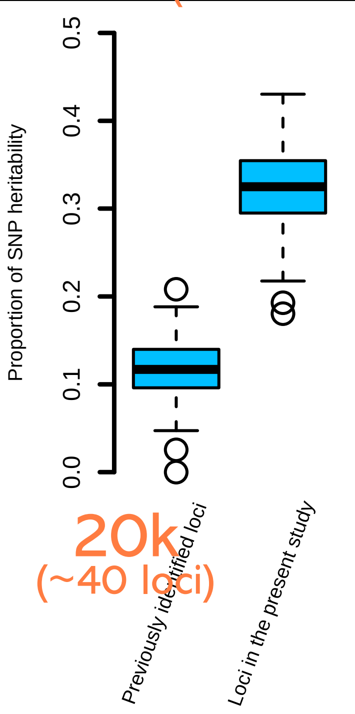
We observed 109 novel genomic regions (151 in total, $P < 2.3e\text{-}10$, 5e-8/215) associated with white matter microstructure

Heritability $h^2 \sim 45\%$



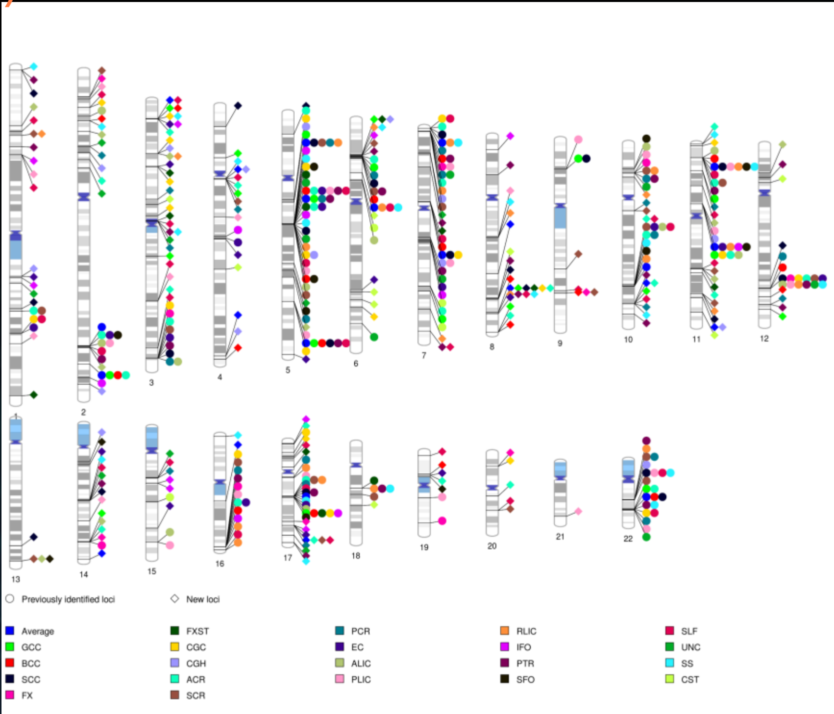


40k
(~150 loci)



20k
(~40 loci)

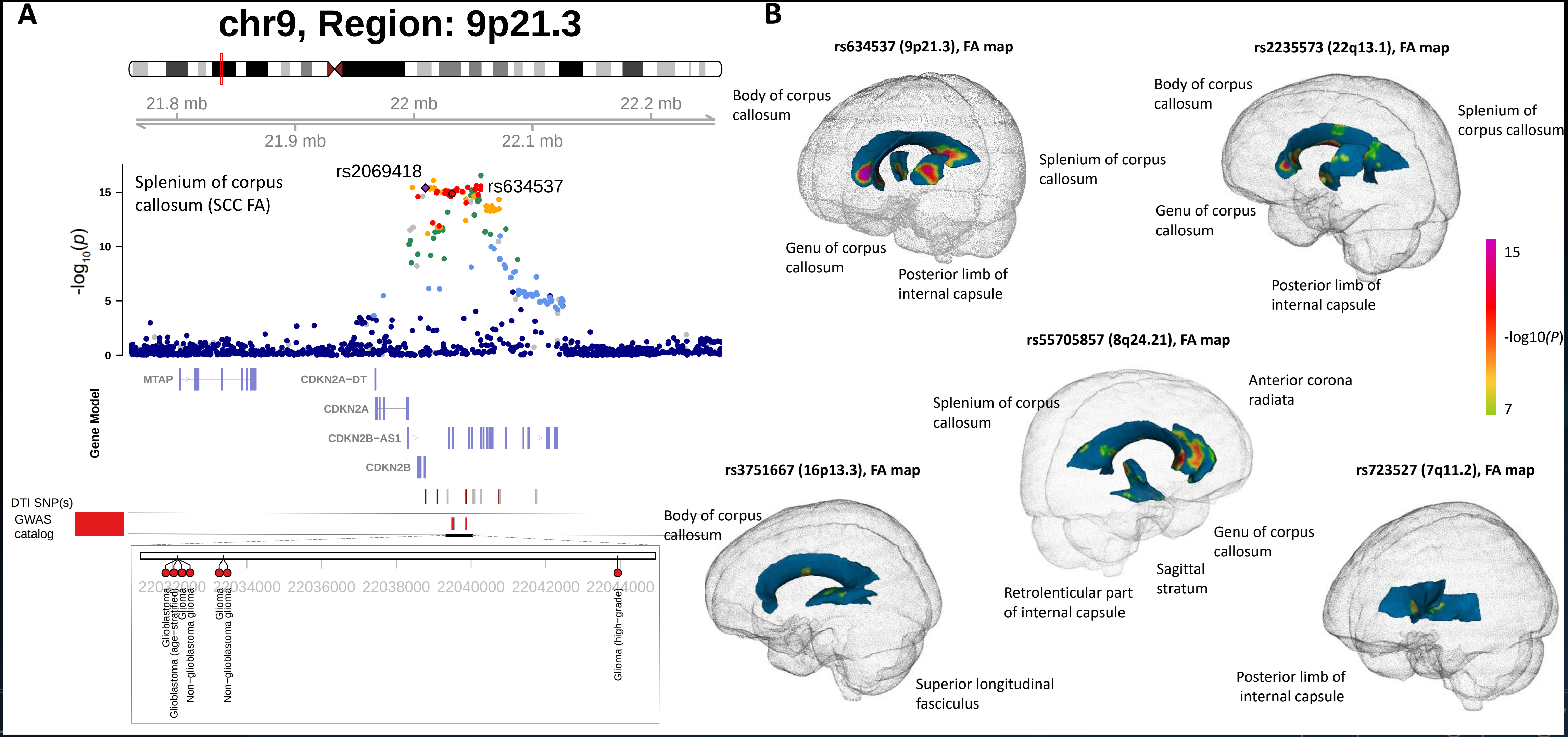
Ideogram of genomic regions



Sample size is essential for gene discovery of traits with highly polygenic genetic architecture

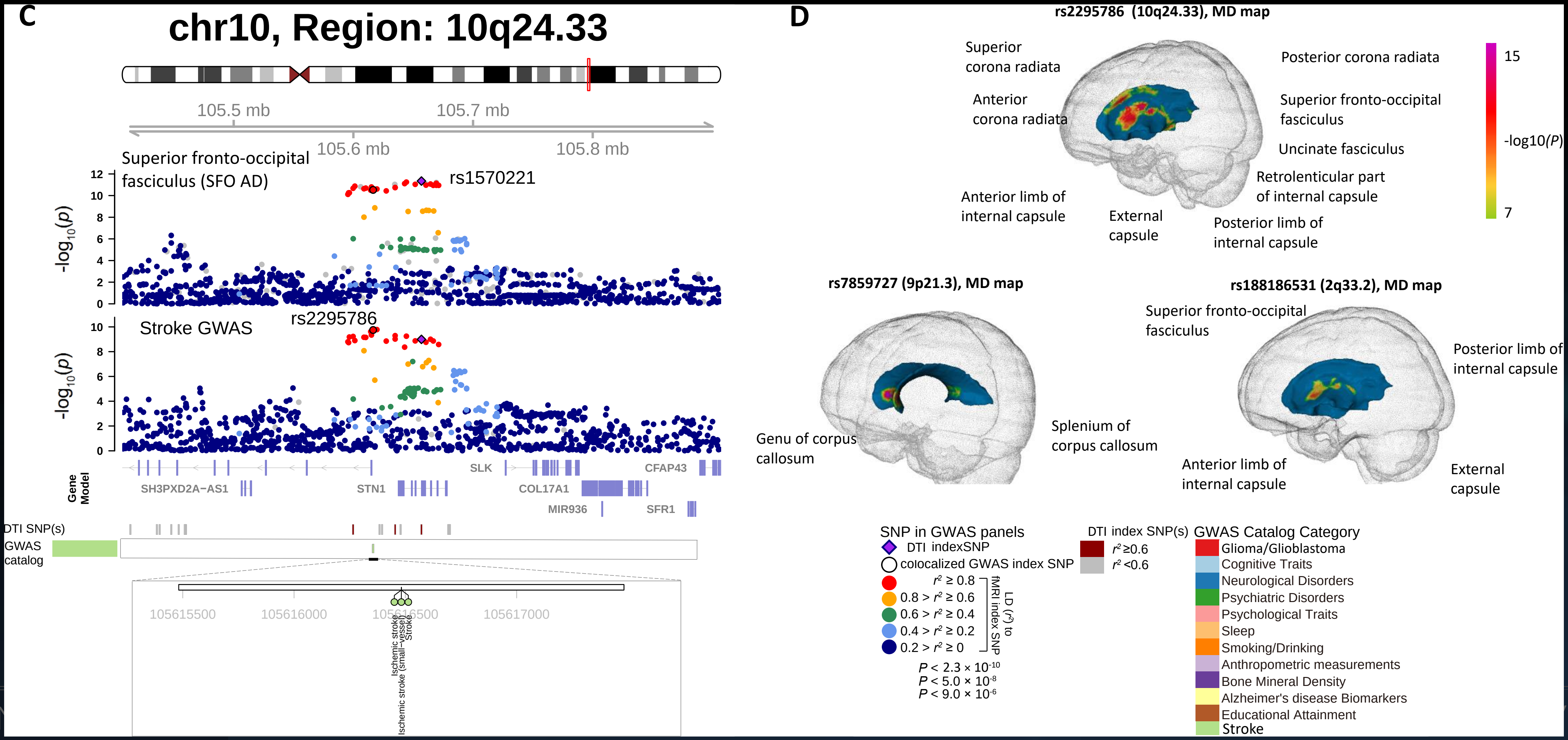
Colocalization with Glioma/GBM

For the 25 known genomic risk regions of Glioma/GBM, 11 are associated with white matter microstructure



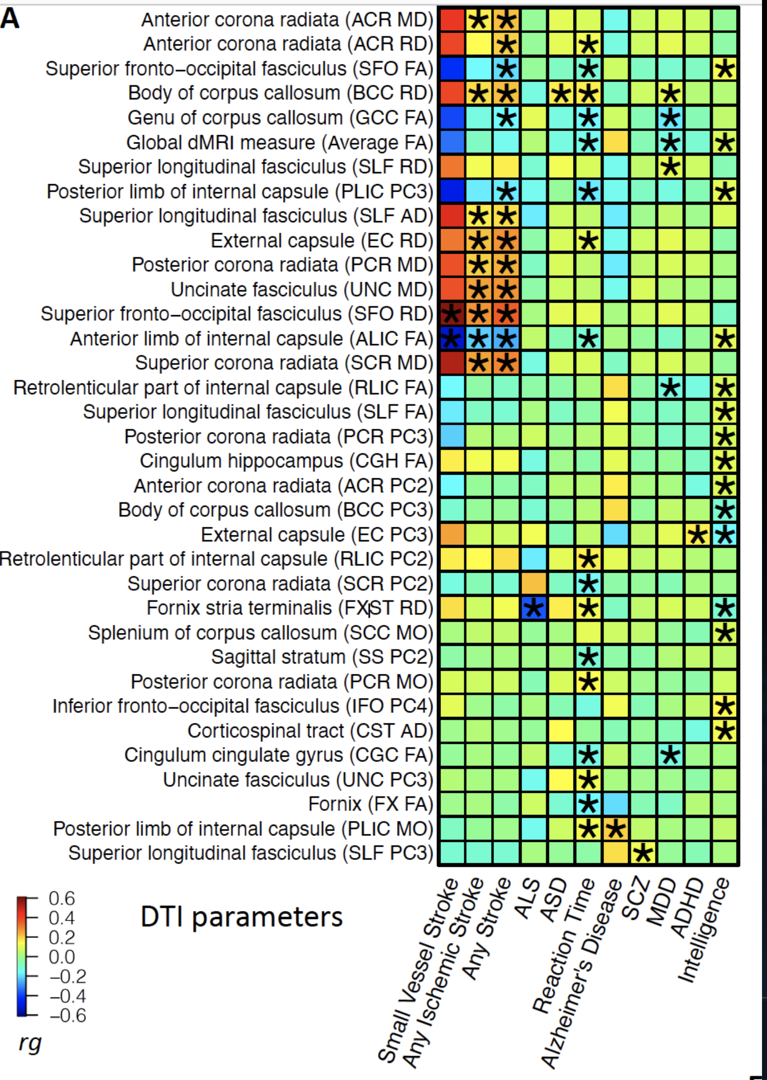
Colocalization with Stroke

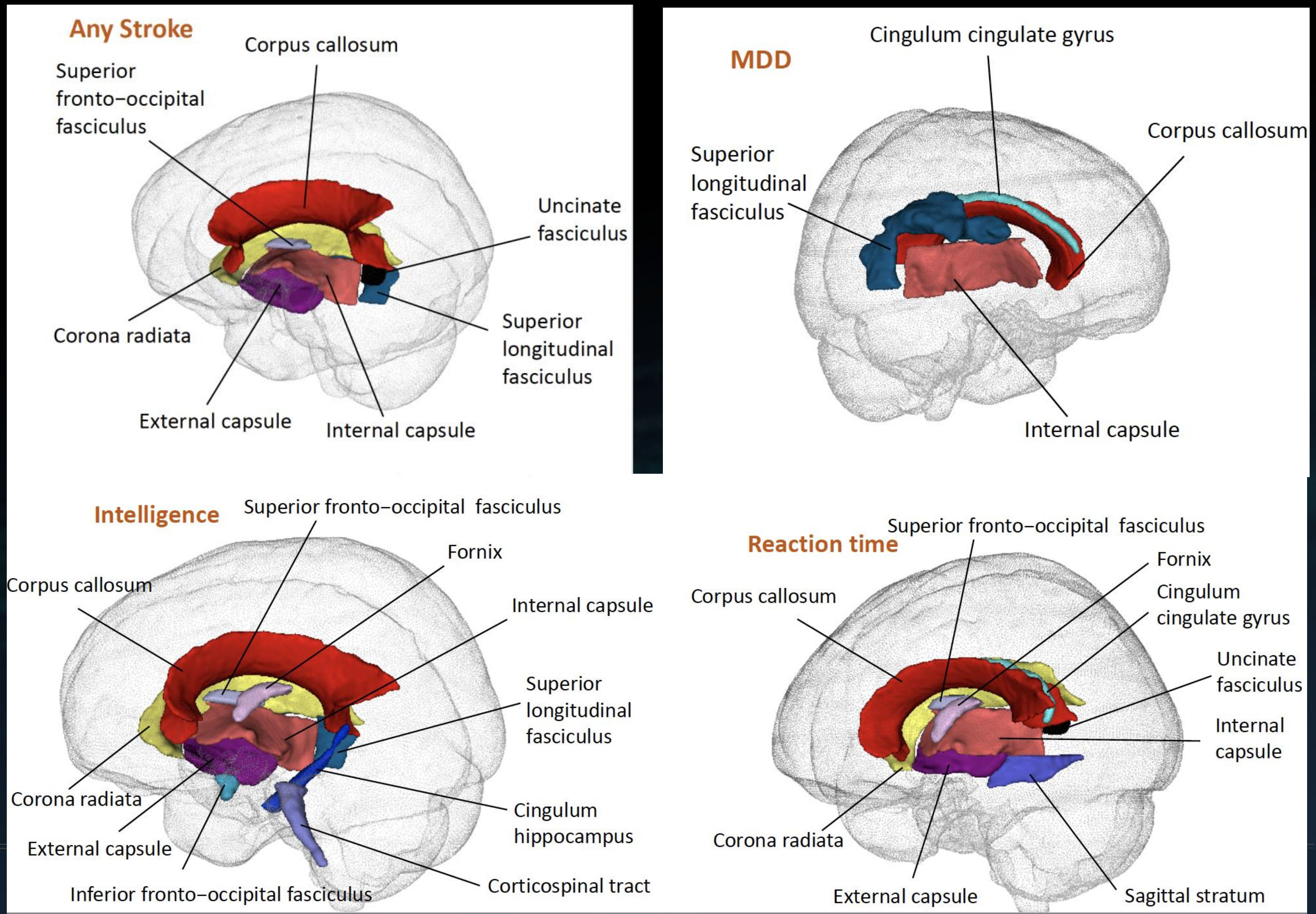
Genetic colocalizations among vascular risk factors (e.g., obesity, diabetes, high blood pressure), white matter microstructure, and stroke



Genetic Correlations with Brain Disorders

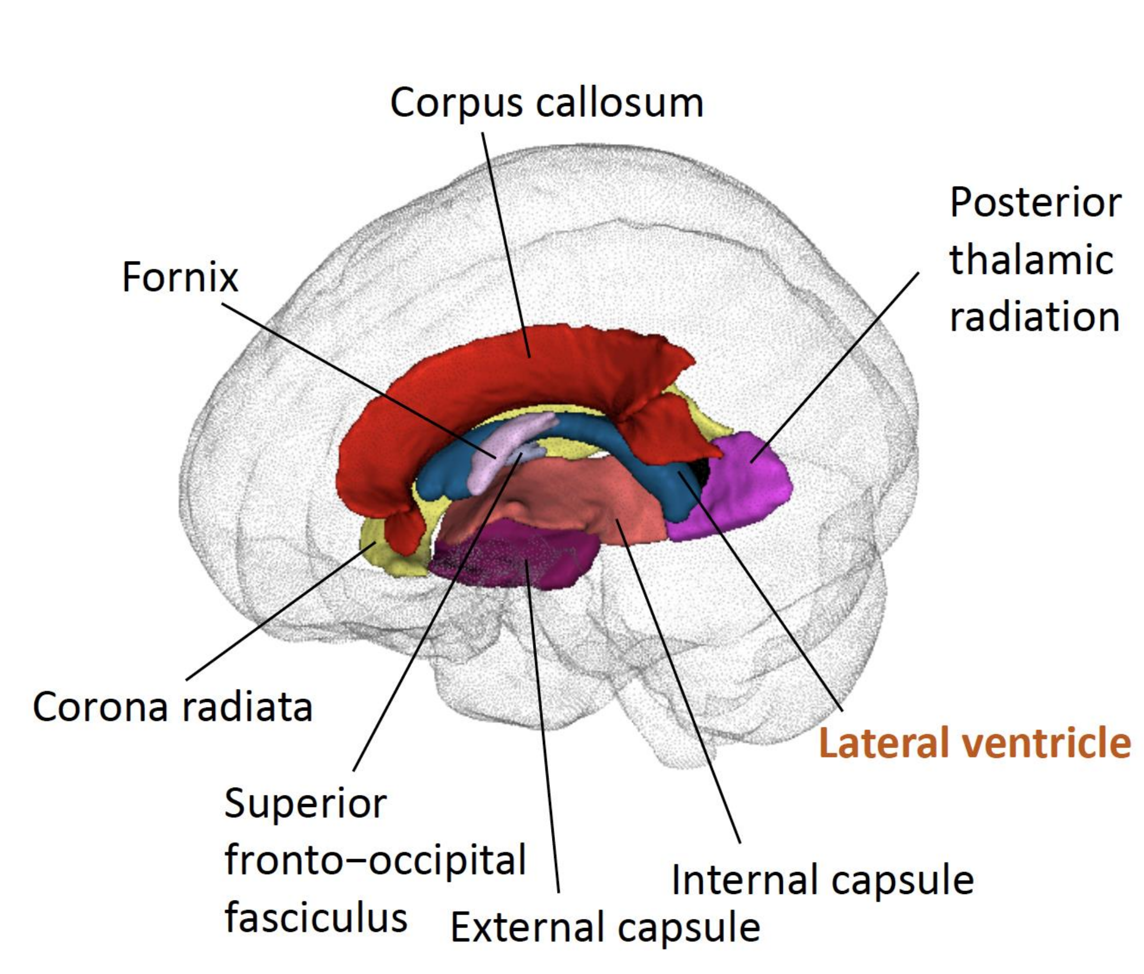
Strong genetic correlation between white matter microstructure and small vessel stroke subtype

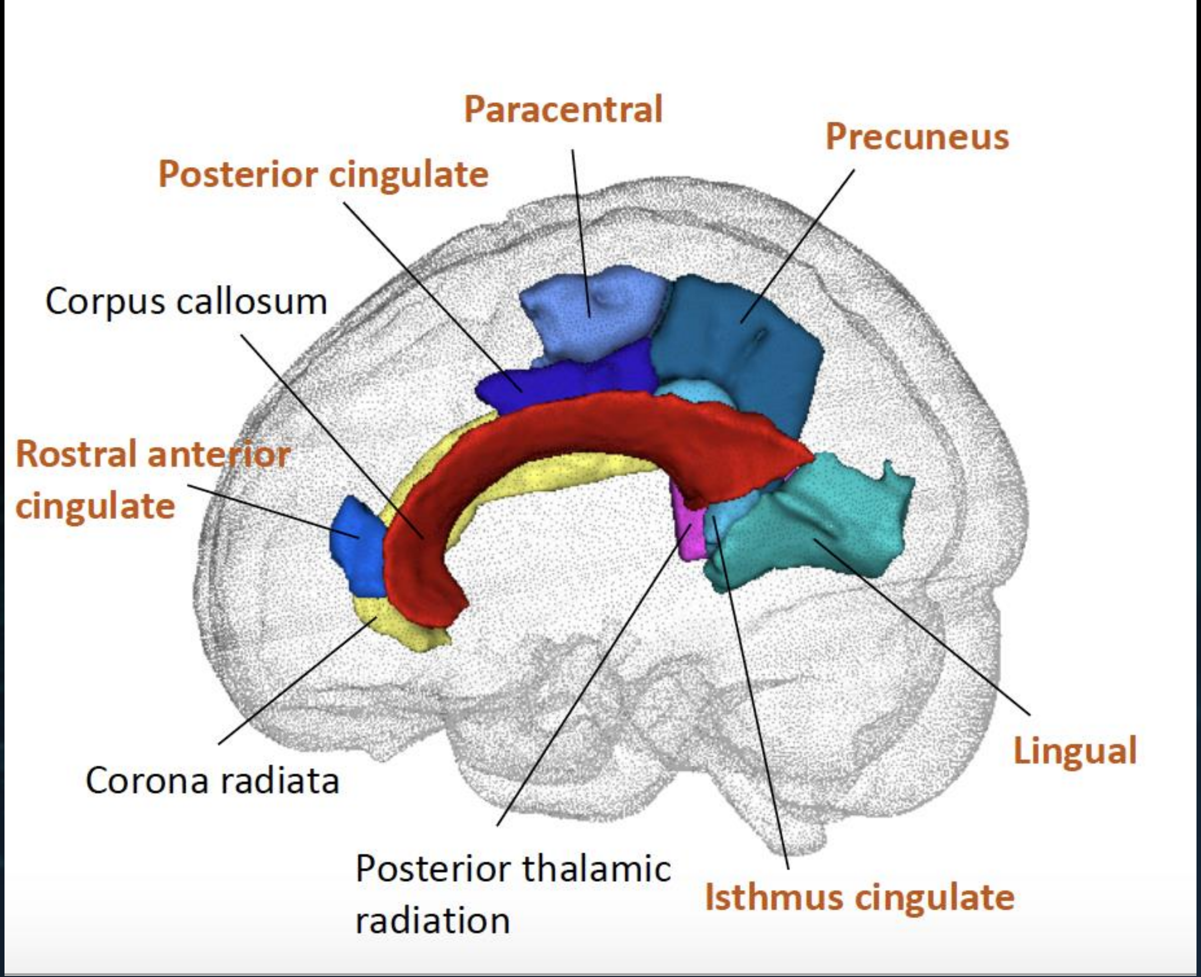




Spatial Overlap with Brain Structures

Strong genetic correlation between white matter microstructure and the grey matter volume of neighboring regions

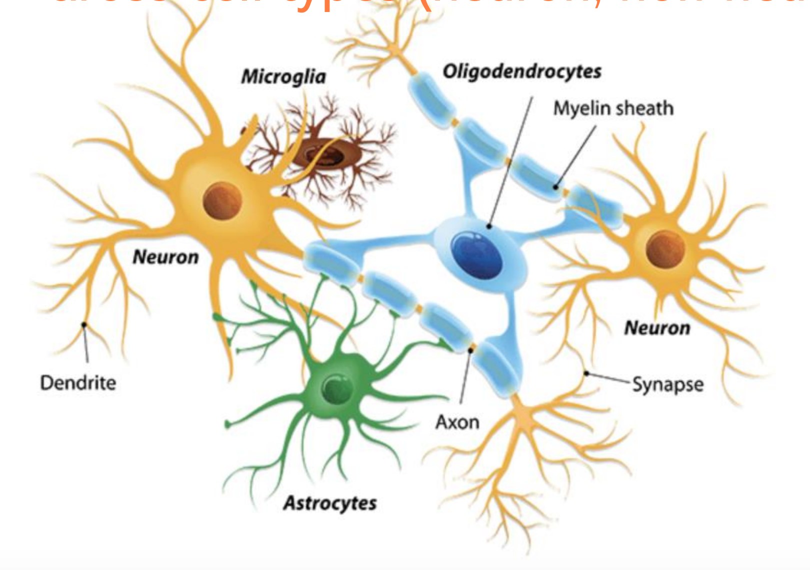


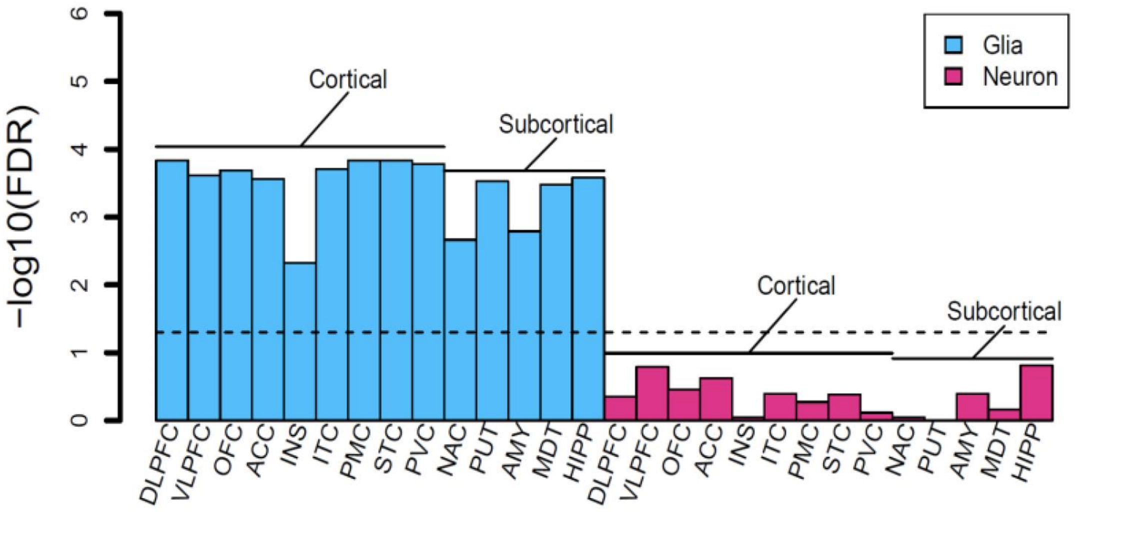


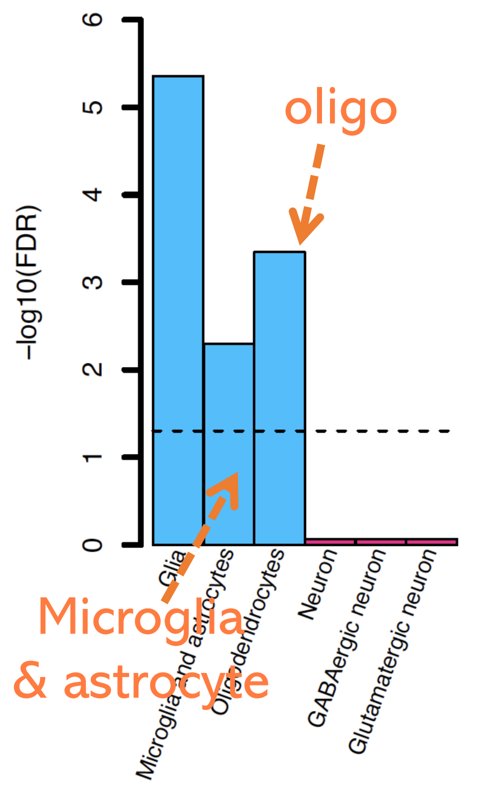
Heritability Enrichment in Brain Cells

Identify brain cell types where genetic variation leads to changes in white matter connectivity

Gross cell types (neuron, non-neuron [glia, including oligo, microglia, astrocyte])





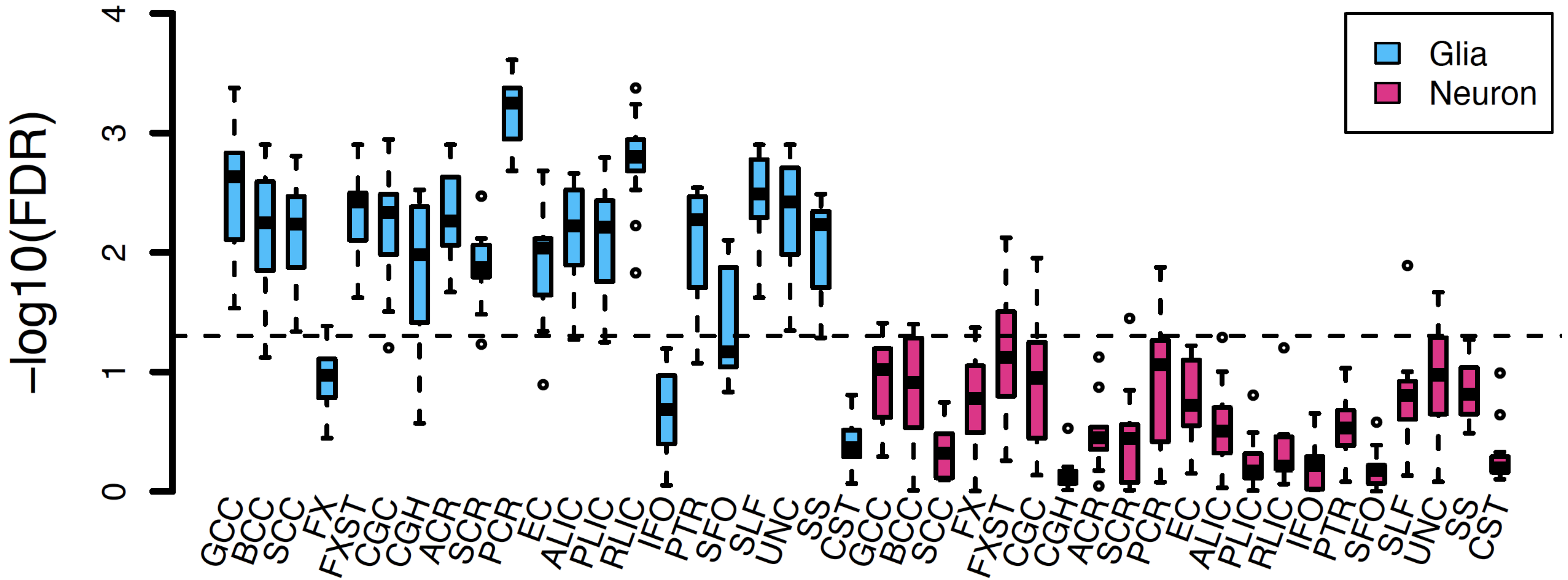


White matter is largely composed of glial cell types (oligo, microglia, astrocyte)

Oligo annotation accounted for 10.4% heritability while only composed 0.3% of the genetic variants

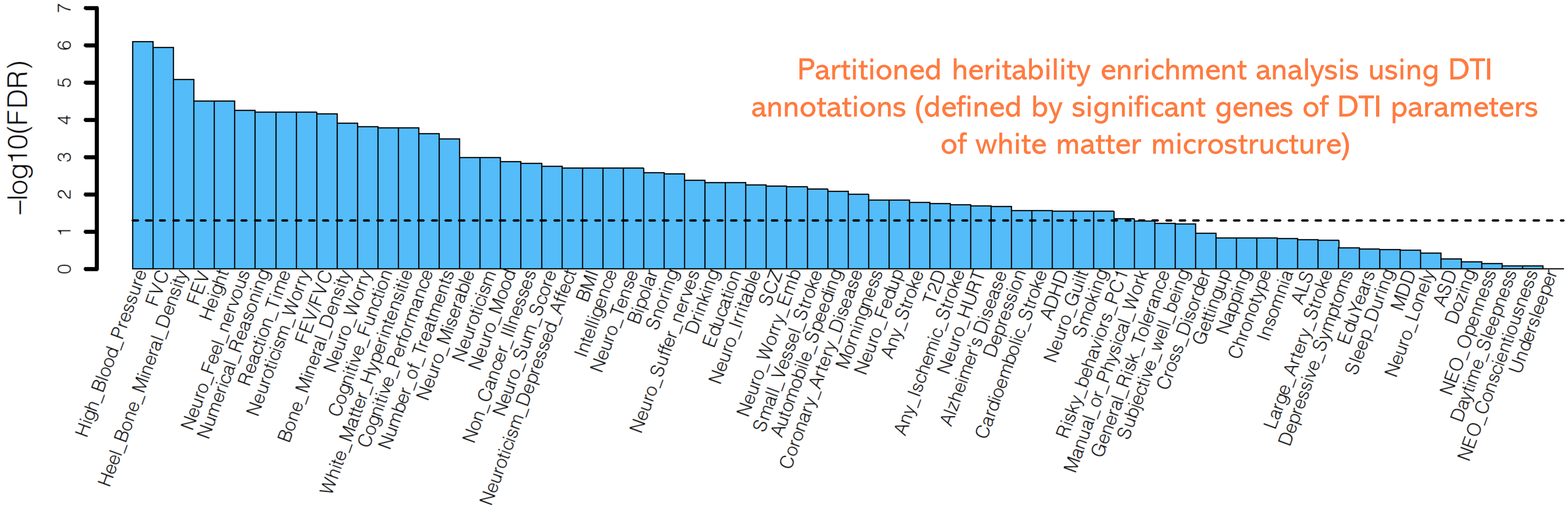
Heritability Enrichment in Brain Cells

Glial cell enrichment was widely observed in white matter tracts and was most significant in posterior corona radiata (PCR), posterior limb of internal capsule (PLIC), and genu of corpus callosum (GCC)



DTI annotation enrichment

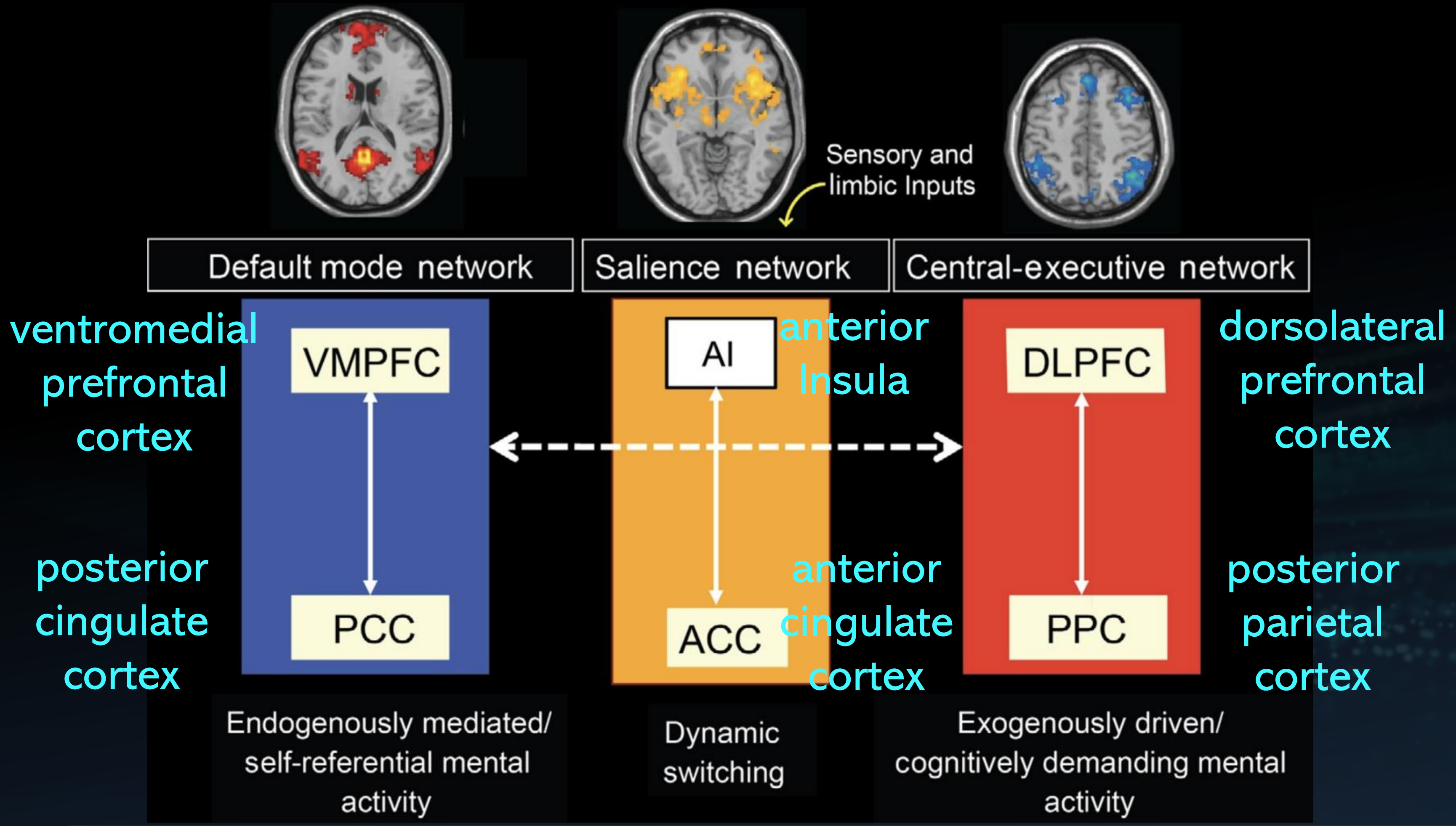
Heritability of 49 complex traits was significantly enriched in genetic regions influencing white matter microstructure, such as stroke, schizophrenia, ADHD, bipolar Alzheimer's Disease, T2D, high blood pressure, and coronary artery disease



Partitioned heritability enrichment analysis using DTI annotations (defined by significant genes of DTI parameters of white matter microstructure)

Triple Network Model of Psychopathology

The salience network (SN) plays a crucial role in dynamic switching between the central executive (CE) and default mode (DM) networks

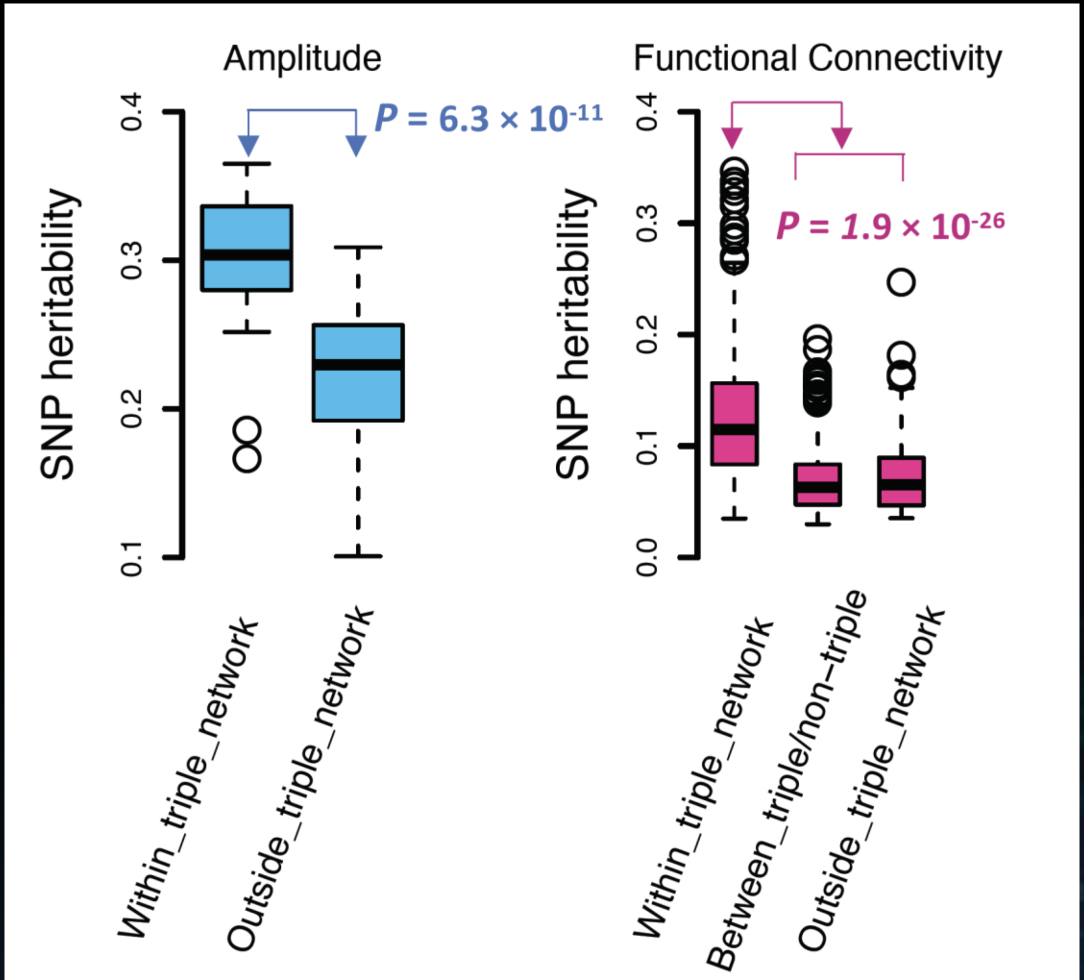


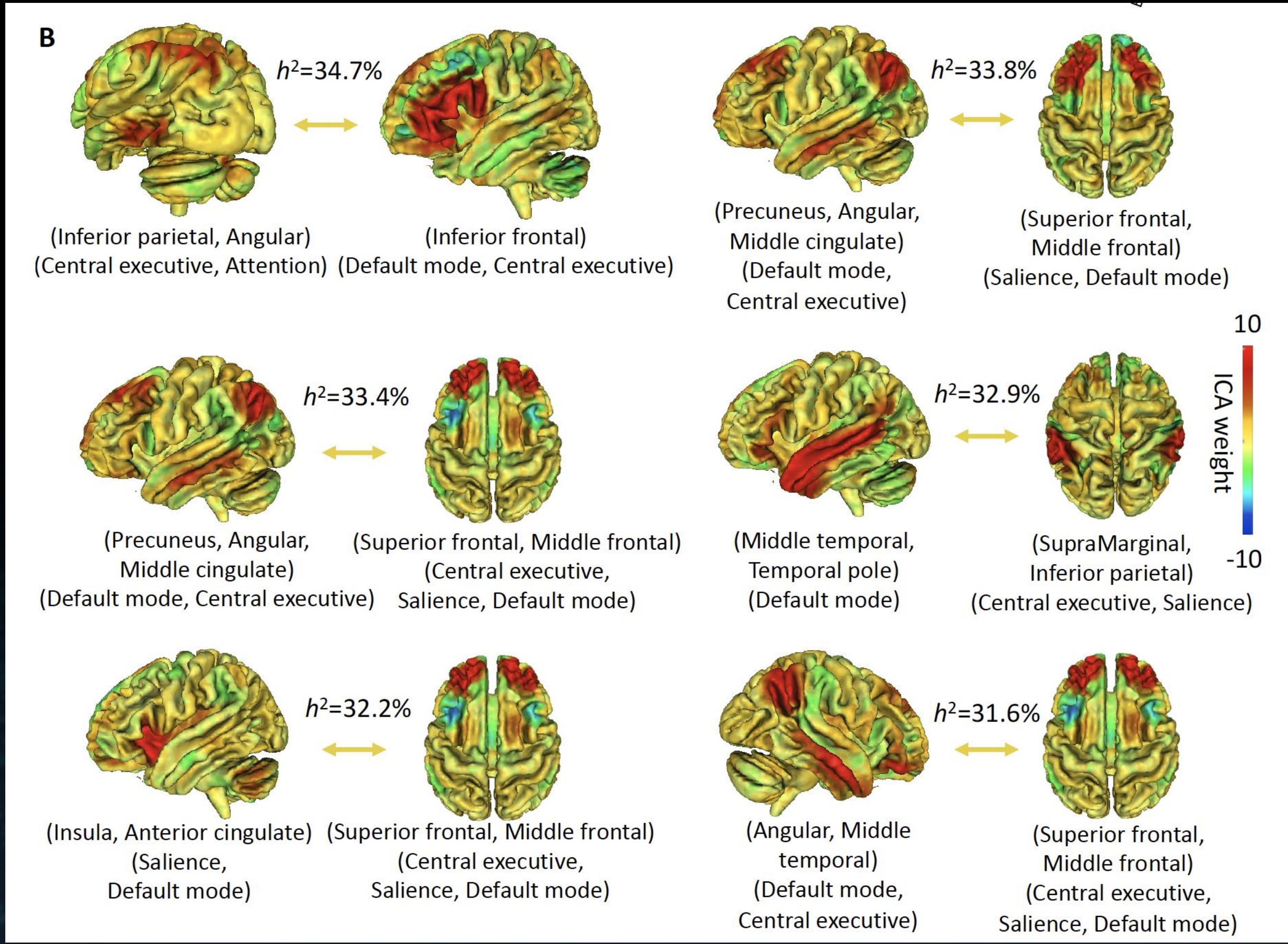
Three core functional networks that support efficient cognition

Related to major brain disorders, such as Alzheimer's disease (AD), Parkinson's disease (PD), and major depressive disorder (MDD)

Genetics of the Triple Networks

Higher heritability than other functional networks (e.g., motor, vision)

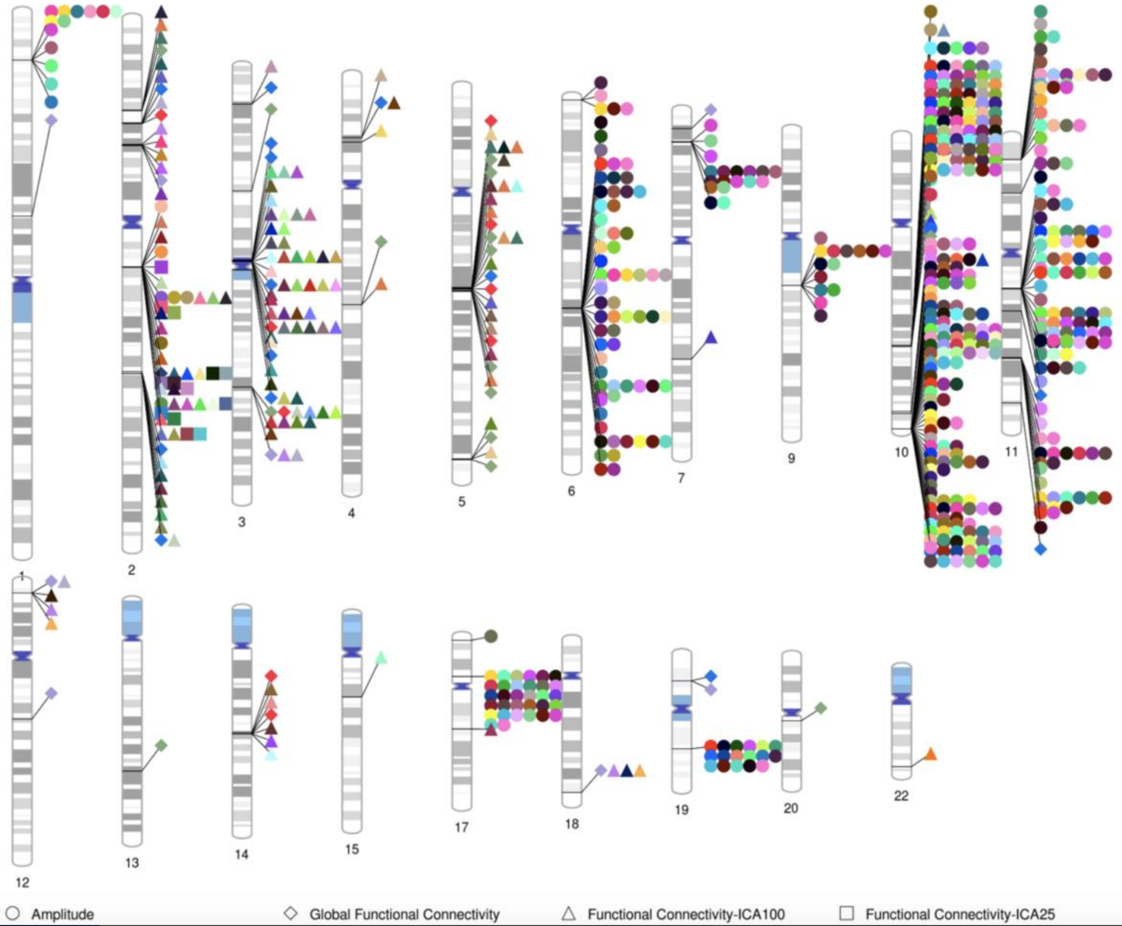


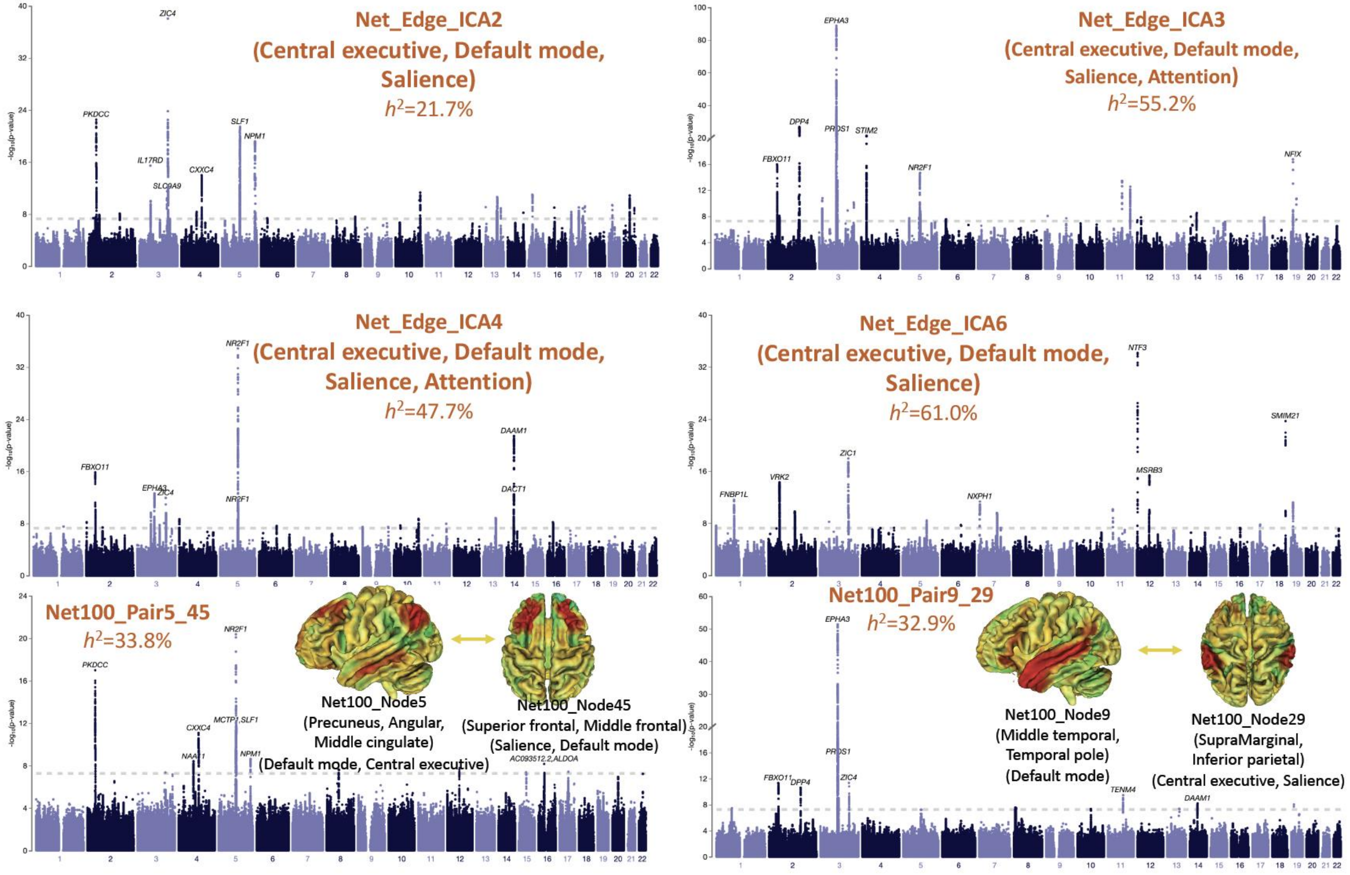


The level of genetic control is higher in the triple networks, which closely control multiple cognitive functions and affect major brain disorders

Genetics of Functional Brain

Ideogram of the loci influencing rsfMRI traits of intrinsic brain activity at the significance level 2.8e-11 (5e-8/1777)

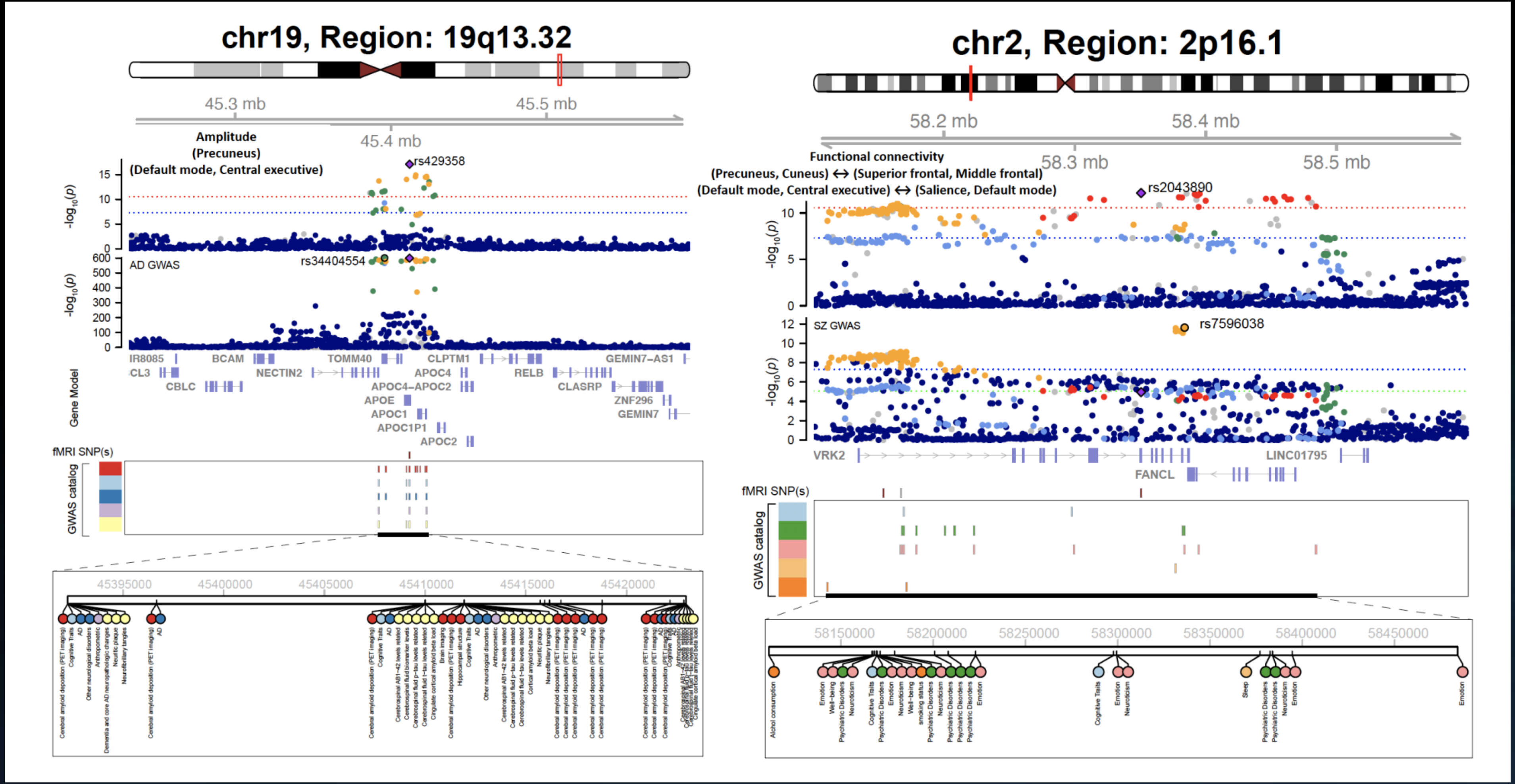




Colocalization with AD and SCZ

Colocalization between brain function in the default mode (DM) and central executive (CE) networks with Alzheimer's disease (AD) and Schizophrenia (SCZ)

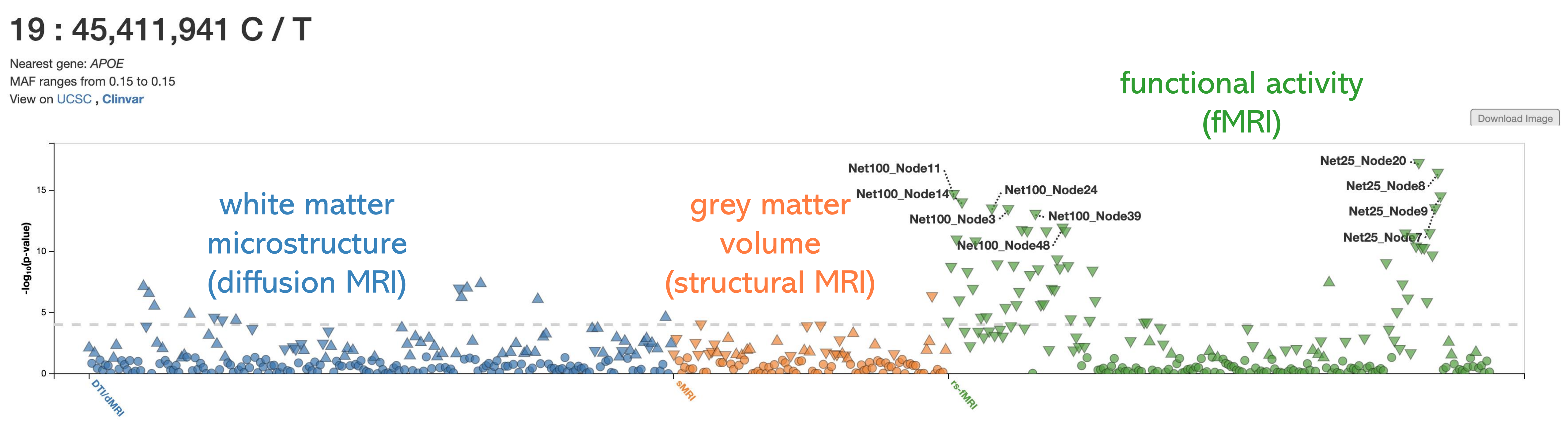
Alzheimer's disease *(APOE)*



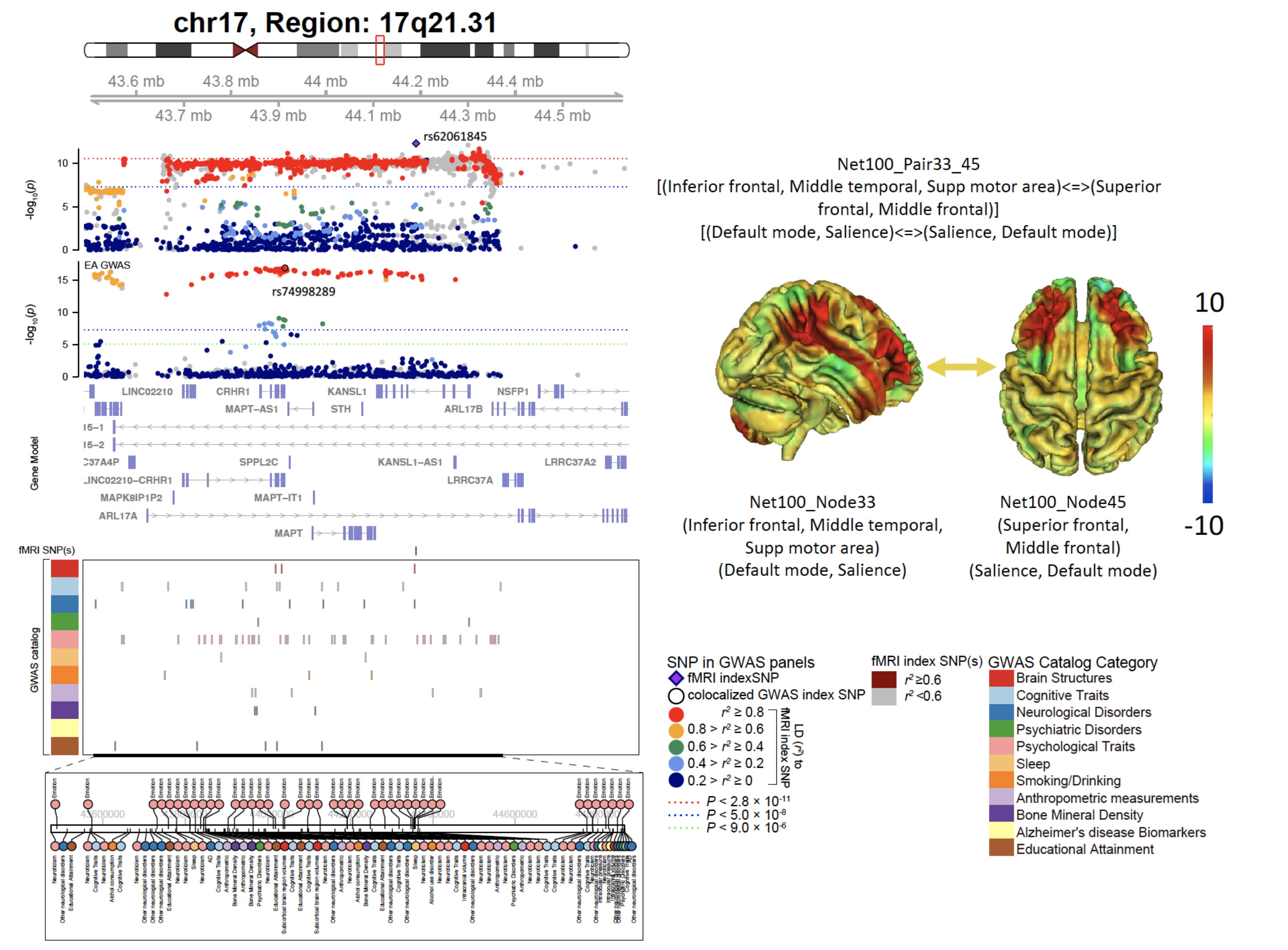
Schizophrenia

Colocalization at *APOE*

APOE gene has stronger genetic relationships with brain function than brain structures



Colocalization at 17q21.31 regions



Neurological disorders
(e.g., Parkinson's disease, Alzheimer's disease, corticobasal degeneration)

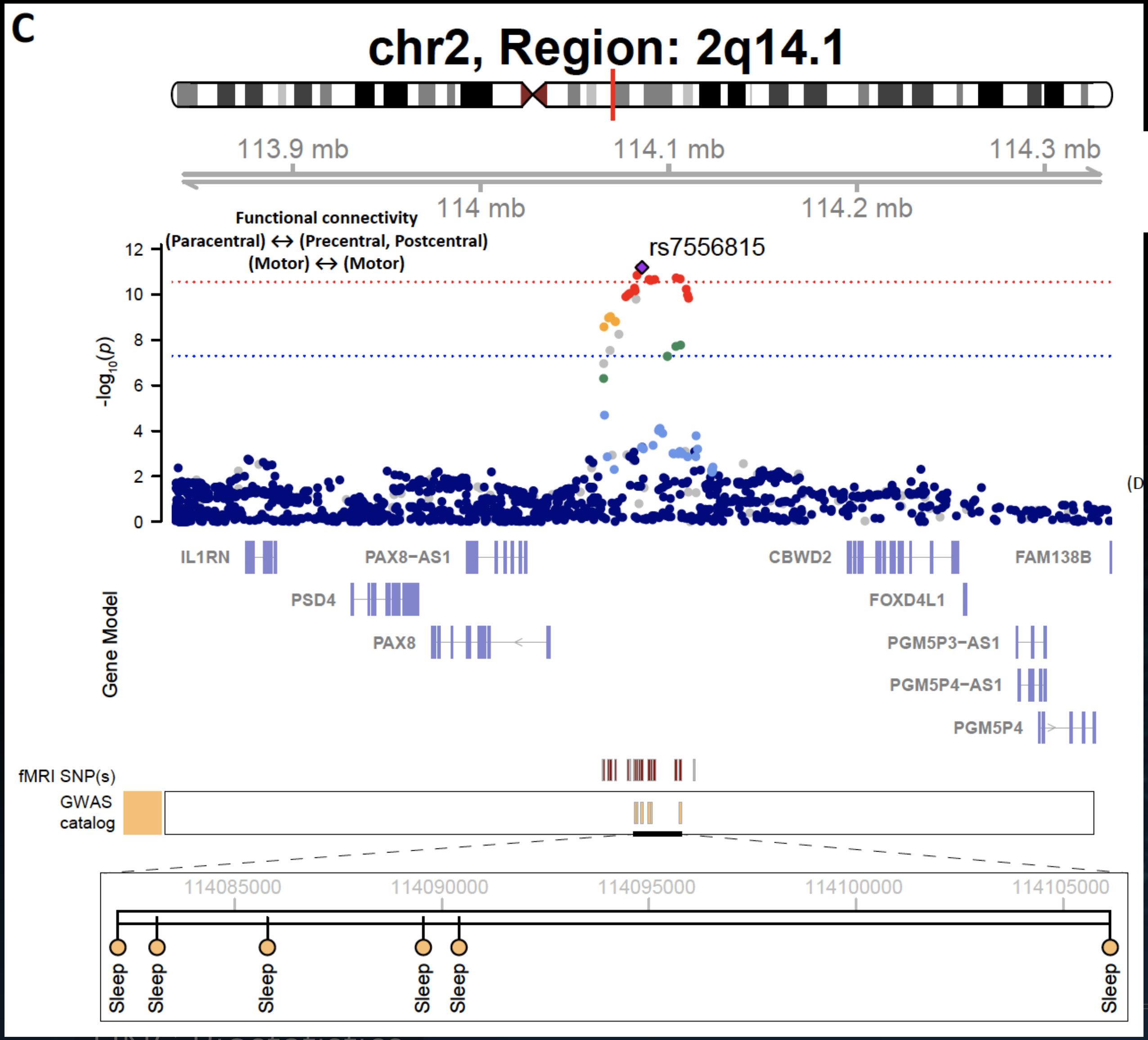
Psychiatric disorders
(e.g., autism spectrum disorder, depression)

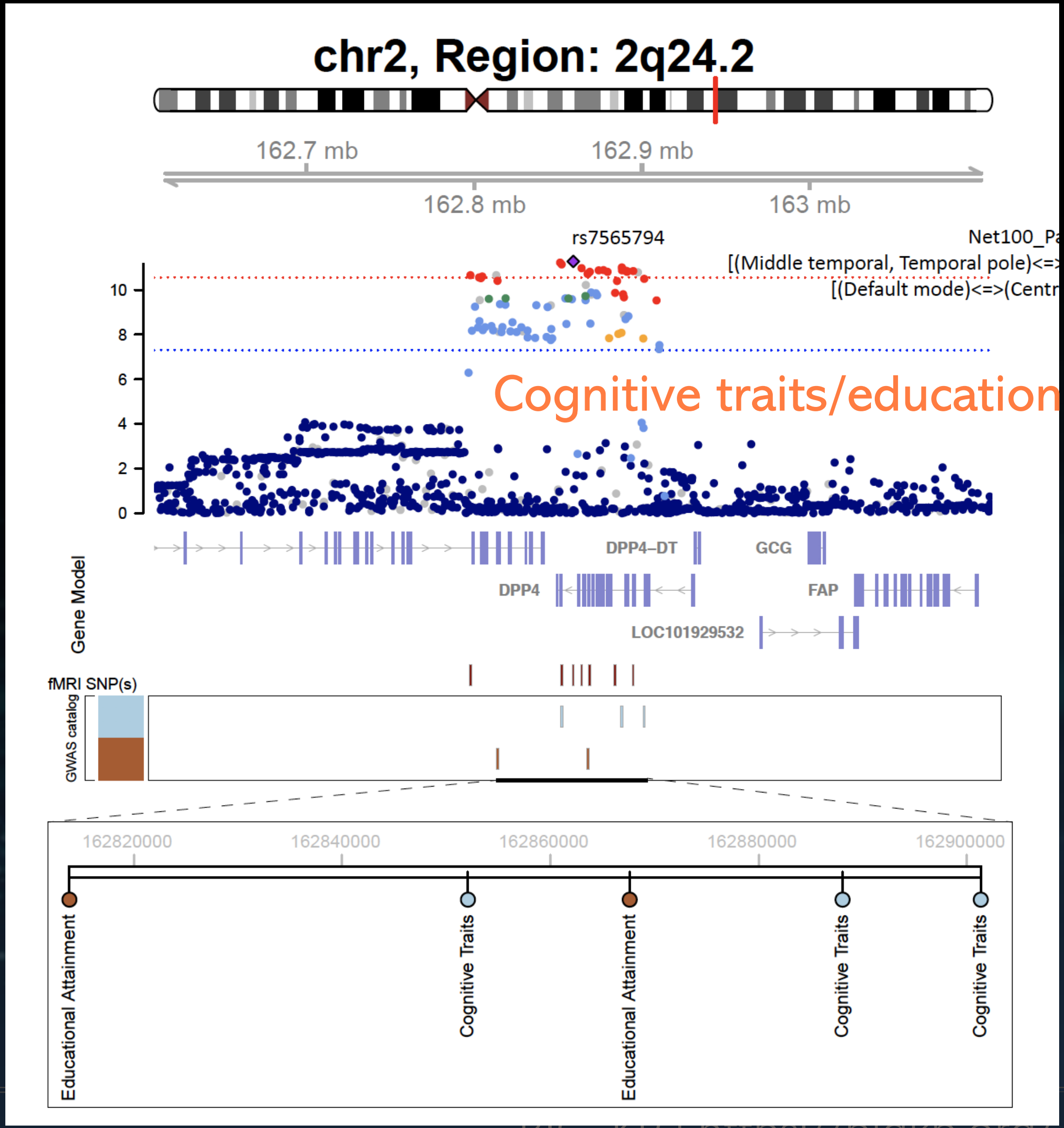
Education, cognitive ability

Psychological traits (e.g., neuroticism)

Alcohol use disorder

Colocalization with Sleep and Cognition

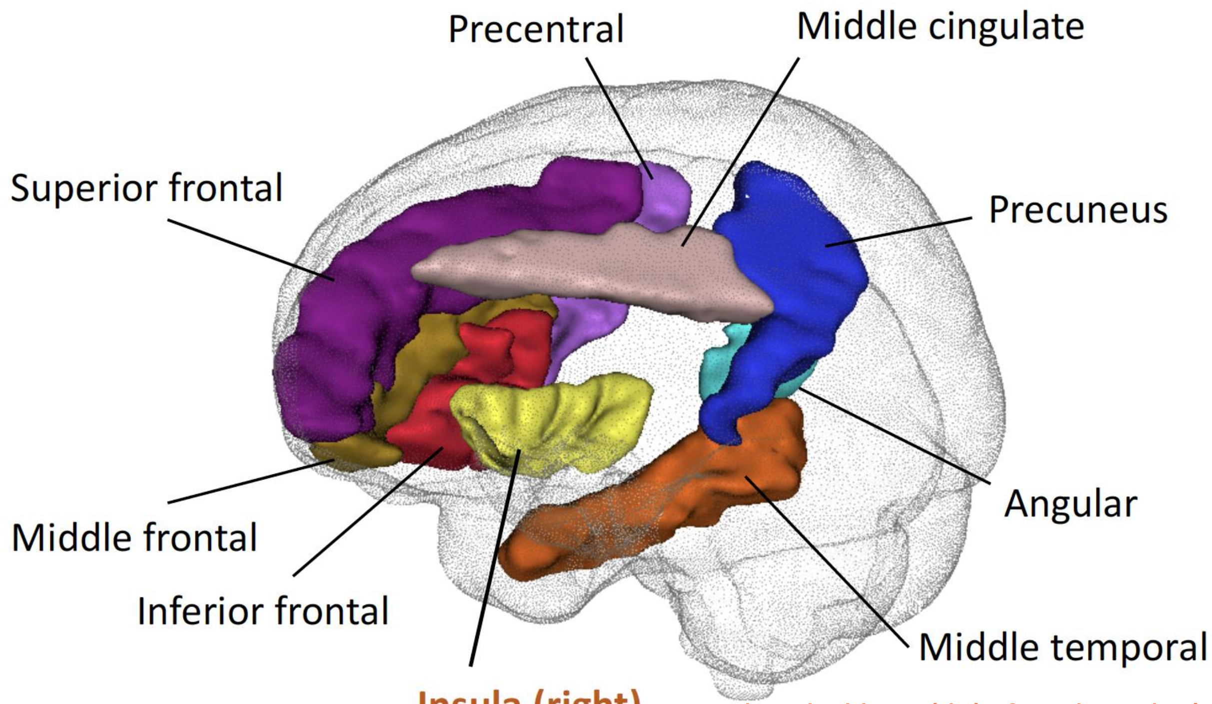




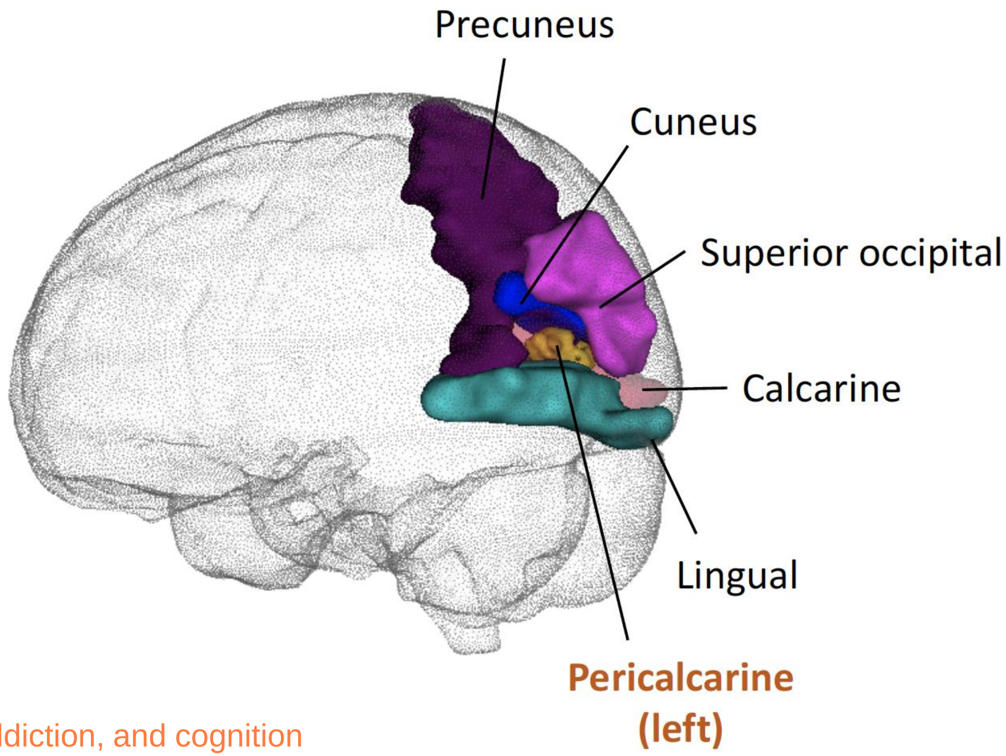
Genetic Overlap with Brain Structures

Shared genetic influences between functional connectivity of default mode and central executive networks and insula volume

Location of the right insula and its neighboring brain regions whose functional connectivity strengths were genetically correlated with the right insula volume



left pericalcarine volume was genetically correlated with the connectivity strengths among its neighboring regions



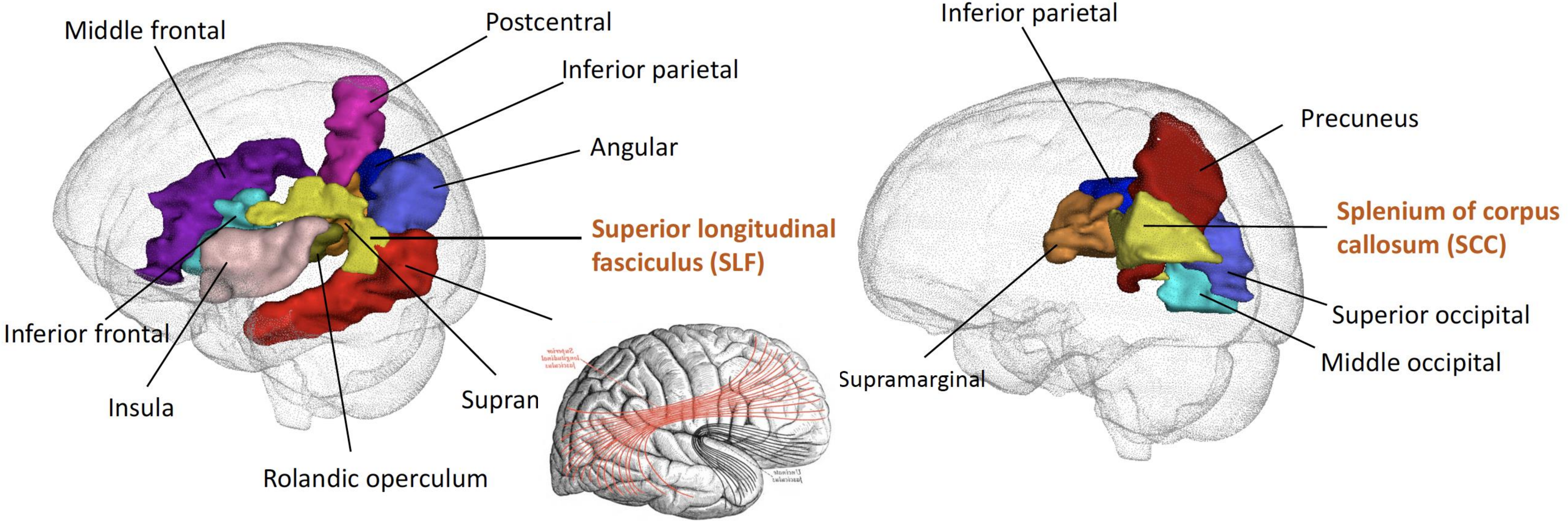
associated with multiple functions, including emotion, addiction, and cognition through extensive connections to neocortex, the limbic system, and amygdala

Spatial colocalizations between regional brain volumes and their genetically correlated functional connectivity traits

Spatial Overlap of Genetics Effects

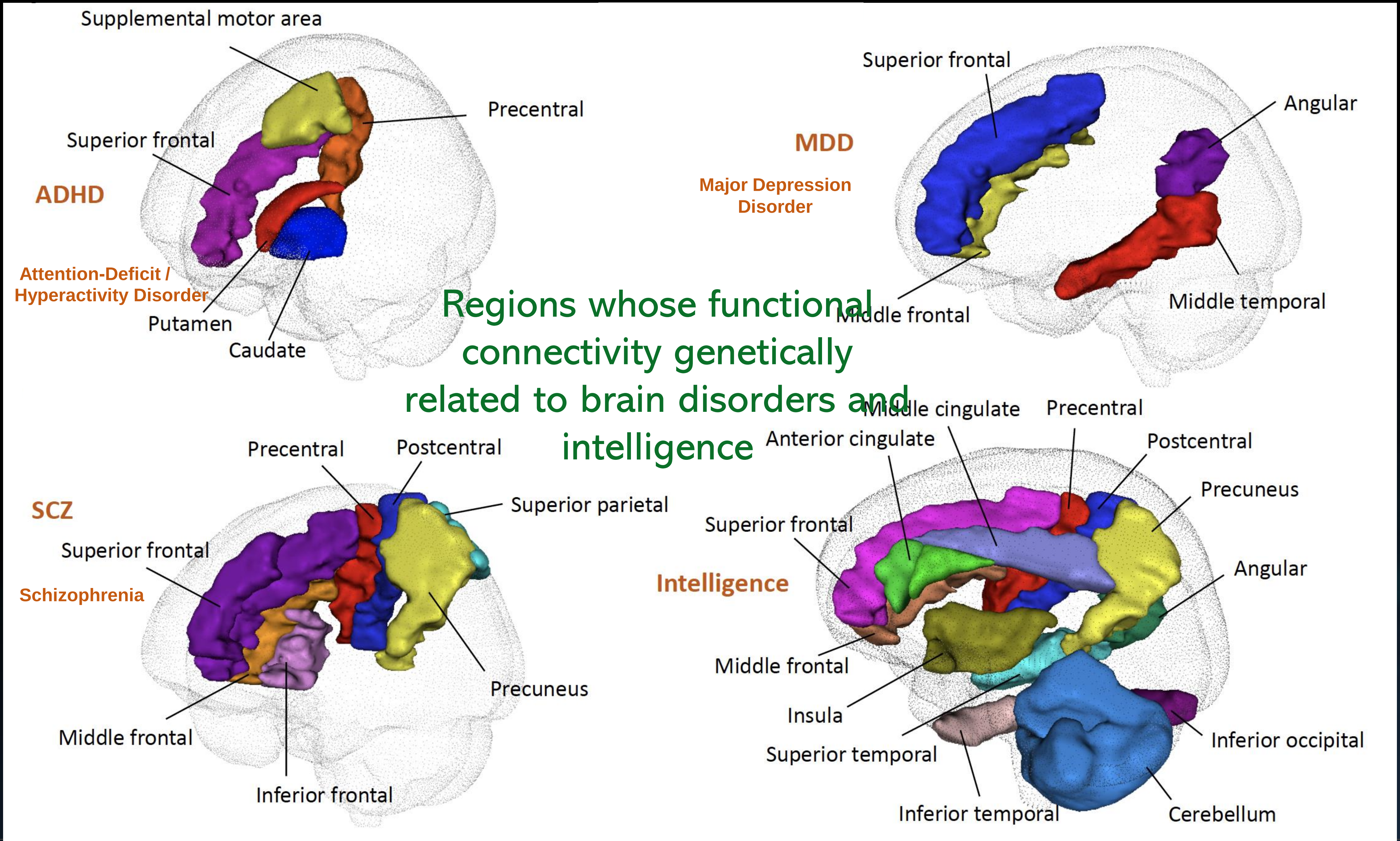
Shared genetic influences between brain functional connectivity and structural connectivity

Location of the SLF and its neighboring brain regions whose functional connectivity strengths were genetically correlated with the structural connectivity of SLF

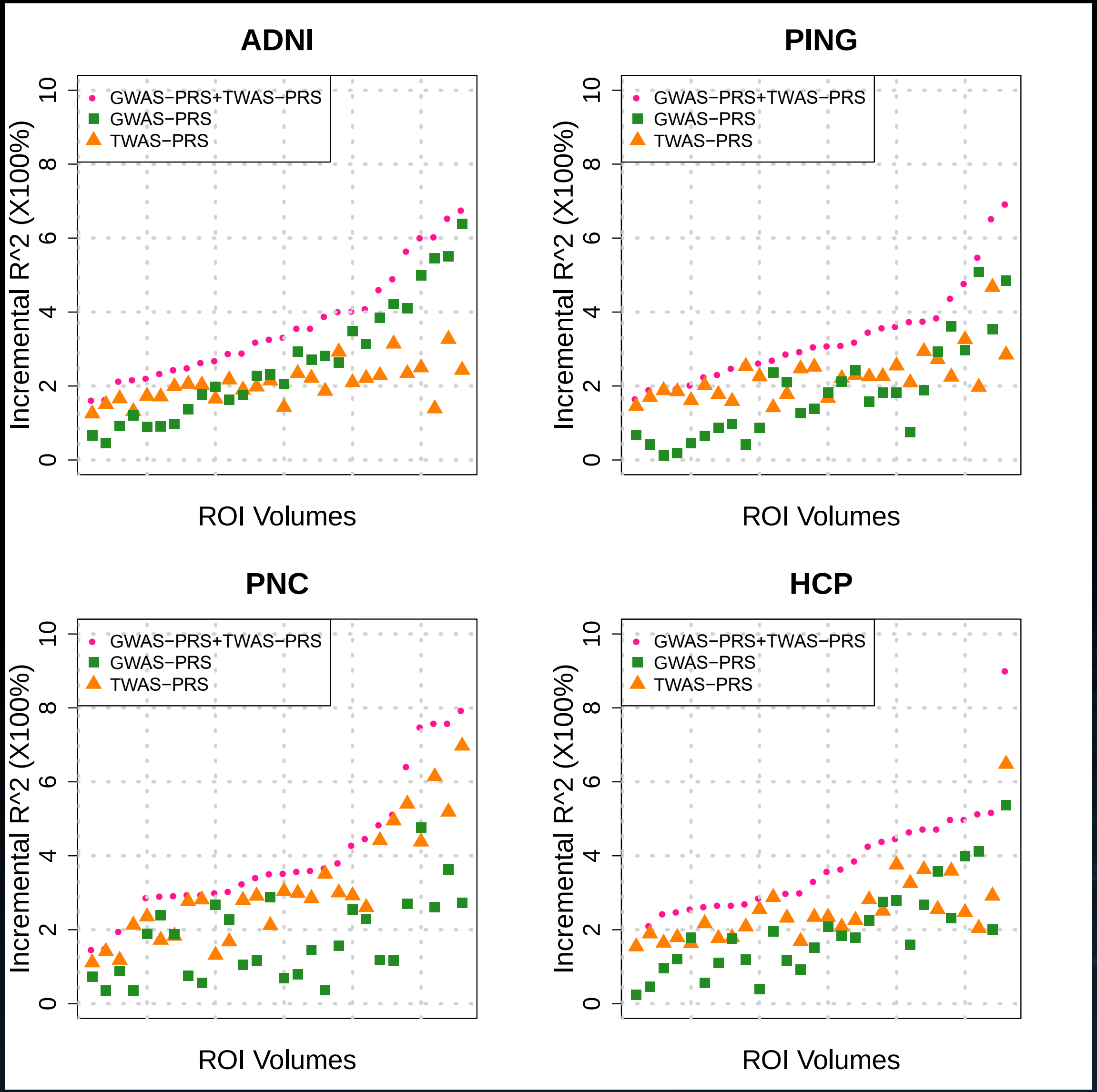


Genetic evidence on how distributed functional networks communicate across large distances

Regions whose functional connectivity genetically related to brain disorders and intelligence



Integrating Gene Expression to PRS



Gene expression-informed gene-level PRS + GWAS PRS has higher prediction accuracy

Construct gene-level PRS (polygenic risk scores) by leveraging gene expression reference panels (e.g., GTEx) in TWAS

It's just a beginning

Publications (2018+)

Common variants contribute to intrinsic functional architecture of human brain (2020). ***bioRxiv***, 229914. LINK

Common genetic variation influencing human white matter microstructure (2020). ***bioRxiv***, 112409. LINK

Transcriptome-wide association analysis of 211 neuroimaging traits identifies new genes for brain structures and yields insights into the gene-level pleiotropy with other complex traits (2019). ***bioRxiv***, 842872. LINK

Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits (2019). ***Nature Genetics***, *51*(11), 1637-1644. LINK [Cover Feature]

Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n= 17,706) (2019). ***Molecular Psychiatry***, in press. LINK

Heritability of regional brain volumes in large-scale neuroimaging and genetic studies (2018). ***Cerebral Cortex***, 29(7), 2904-2914. LINK

Genetics discovery in human brain by big data integration

Ongoing/Future Directions

- Causal relationships among disease, brain structures, and brain functionalities (e.g., the genetic pathway among vascular risk factors, white matter, and stroke)
- Build optimal models for complex traits and diseases prediction using imaging and genetics data (e.g., deep learning)
- Compare and identify the best practical strategy and pipelines to process different neuroimaging modalities (e.g., ICA for fMRI)
- Model brain changes and genetics effects across the life span
- Align and integrate different neuroimaging modalities

Part III

Methodological Challenges

Brain Imaging Genetics: Learning Problems

(a) Learning Problems in Brain Imaging Genomics

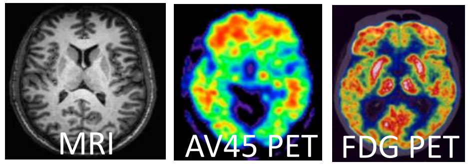
Heritability estimation of brain imaging phenotypes

Imaging genomics associations
- Fundamentals
- Meta-analysis
- Multivariate regression
- Bi-multivariate correlation

Integrating imaging and genomics for outcome prediction

(b) Biomedical Application Considerations

Brain imaging
- Voxels, ROIs, spatial correlation
- Multimodal, longitudinal studies
- Prior knowledge, brain network



Genomics
- SNPs, LD blocks
- Genes, pathways, networks
- Polygenic risk scores (PRSs)

Clinical outcome
- Diagnosis
- Progression
- Impairment score

Cognitive Condition: D, A, 2, End, B, 1, 4, C, Begin, 3

Diagnosis	
Normal	SMC
MCI	AD

(c) Statistical & Machine Learning Considerations

Increasing power
- Quantitative traits
- Multiple comparison
- Meta/mega analysis
- Multivariate models

Overfitting control
- Dimensionality reduction
- Regularization
- Knowledge-guided learning
- Outcome-guided learning

Other topics
- Biological interpretation
- Scalability
- Biased sampling
- Interaction

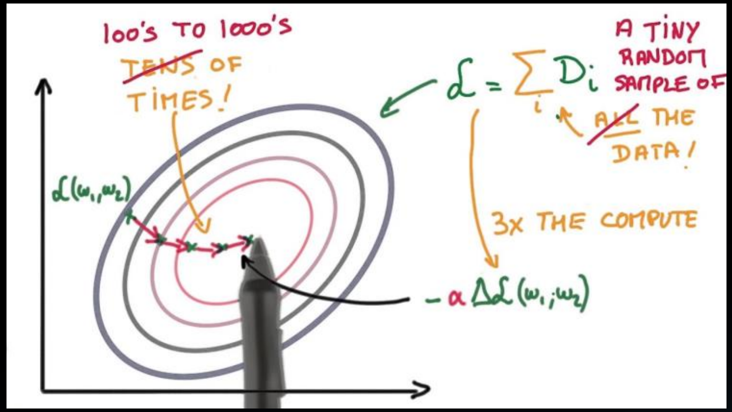
(Shen & Thompson Proc. of the IEEE 2020)

Methodological Challenges

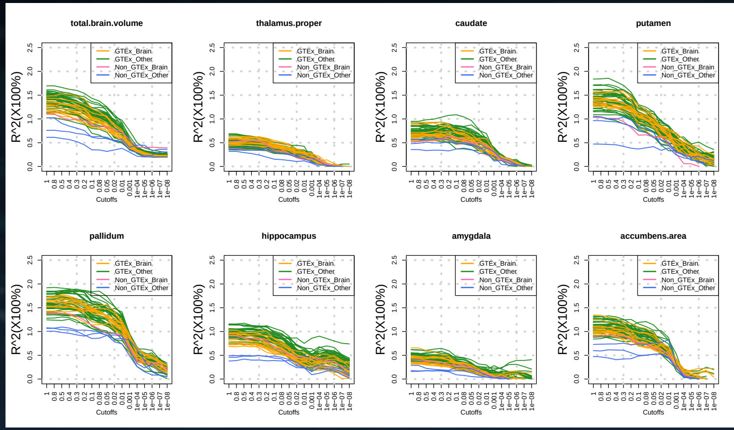
Multiple Biobanks Integration
(e.g., Heterogeneity in global populations)

Omics Data Integration
(e.g., new tech, biological pathway)

Database/
Tool/Theory for
Brain Imaging
Genetics

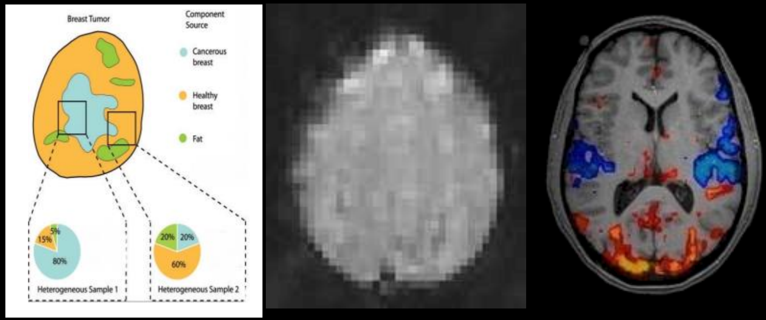


New Computational Tools
(e.g., challenge of dense signal in biobank-scale database)

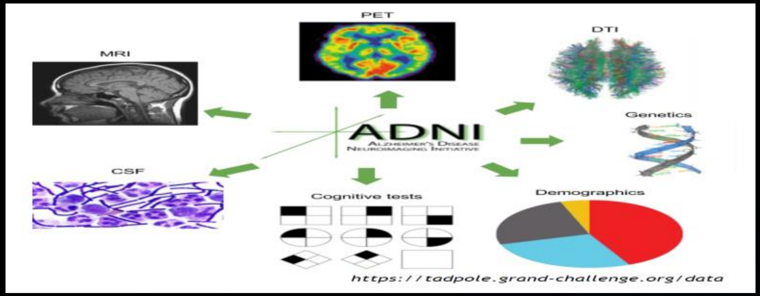


Advanced Methods for Dense Signals
(e.g., deep learning)

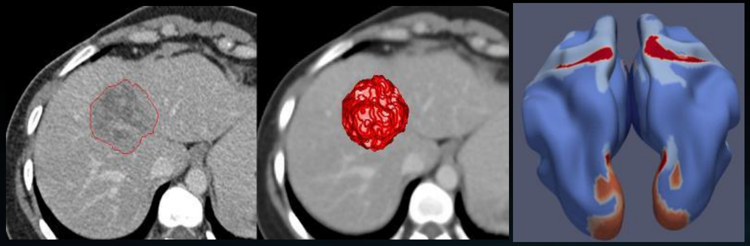
Ecological Layout



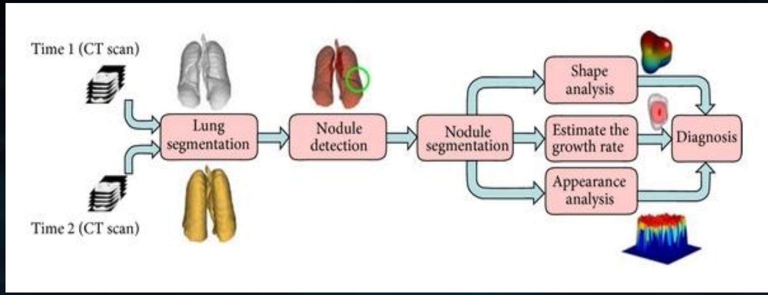
Deconvolution



Integration

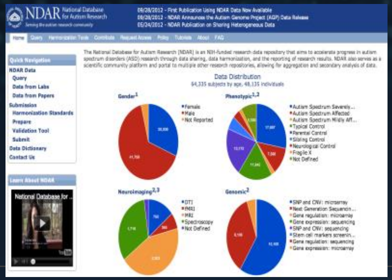


Learning



Prediction

Large-scale Database



Acknowledgement

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

PURDUE UNIVERSITY | Department of Statistics

Brain Imaging Genetics Knowledge Portal (BIG-KP)

Genetics Discoveries in Human Brain by Big Data Integration

bigkp.org

Funding: U.S. NIH Grants MH086633 and MH116527

Pictures: Copyrights belong to their own authors and/or holders.

Data: We thank Bingxin Zhao, Tengfei Li and other members of the UNC BIG-S2 lab (https://med.unc.edu/bigs2/) for processing the neuroimaging data.

UK Biobank resource application number: 22783.