Genetic Architecture of Human Organs through Imaging Genetics: Challenges and Opportunities

University of North Carolina at Chapel Hill

Hongtu Zhu

Departments of Biostatistics, Statistics, Computer Science,

Genetics, and Radiology, and Biomedical Research Imaging Center

Thank Drs. Bingxin Zhao, Tengfei Li, Owen Jiang, Yun Li, Jason Stein, Patrick Sullivan, Stephen Smith, Paul Thompson, Tom Nichols, and Xin Wang.

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

CONTENTS

Part I

Genetic Architecture of Human Organs and Associated Diseases

Part II

Foundation Models for Medical Imaging Analysis

Part III

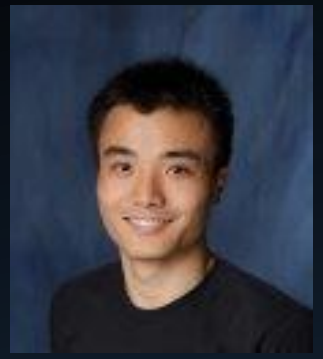
Highly Efficient Imaging Genetics Methods

Part I

Genetic Architecture of Human Organs and Associated Diseases

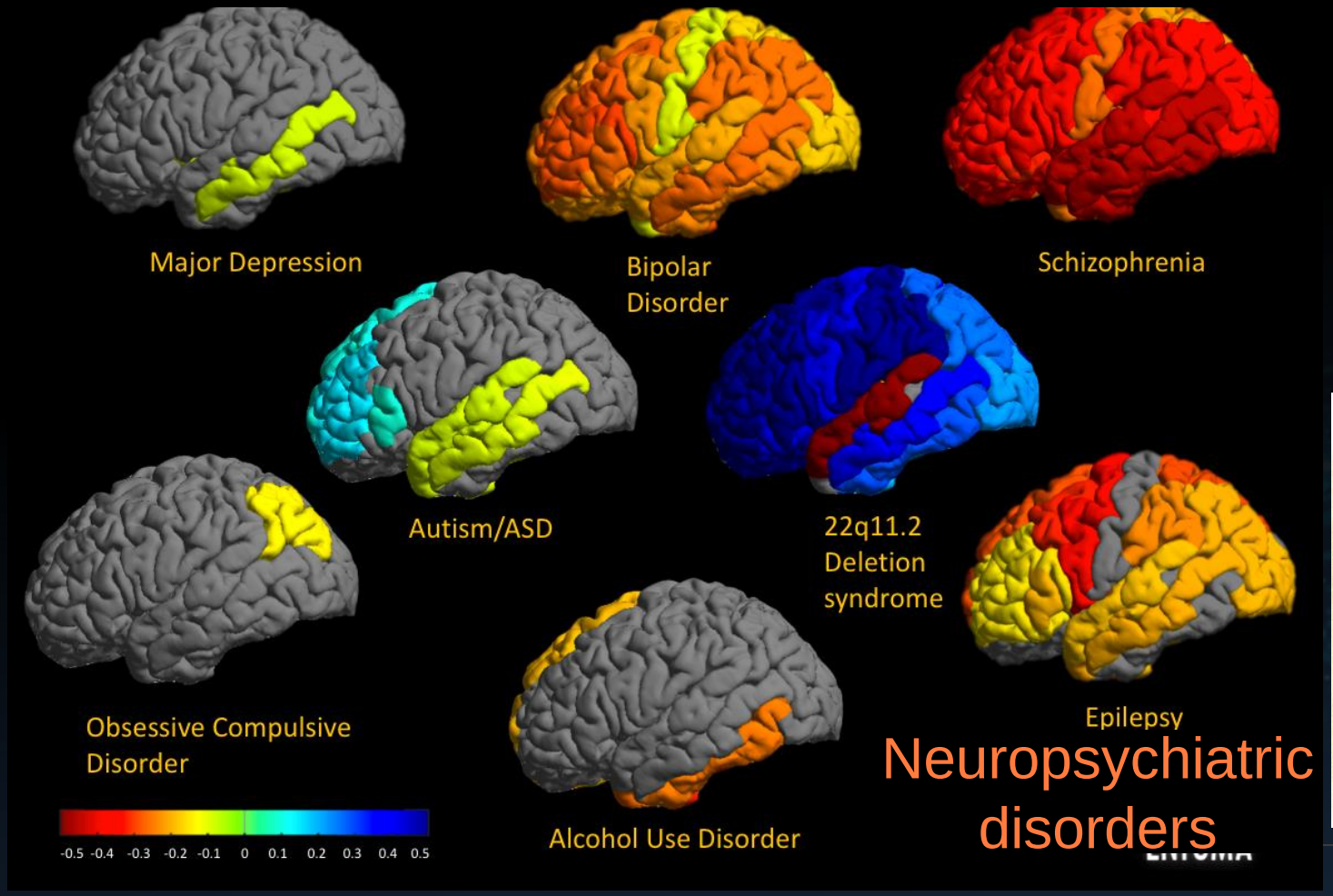
“The best thing about being a statistician is that you get to play in everyone's backyard.”

- John Tukey -

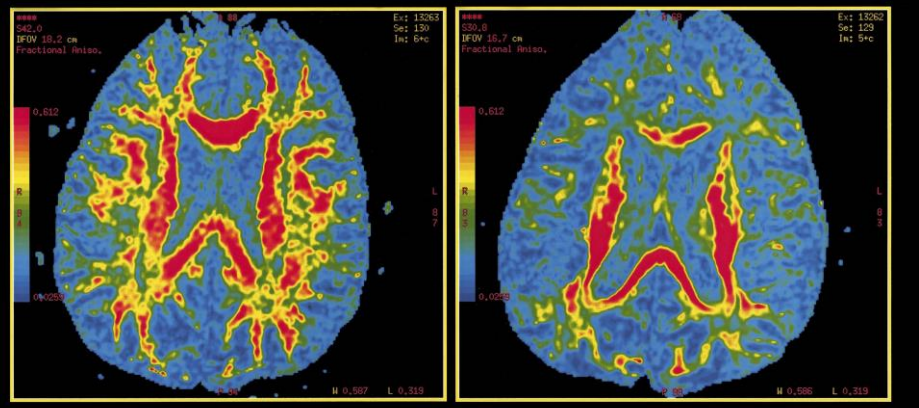


Brain Imaging for Brain Disorders

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

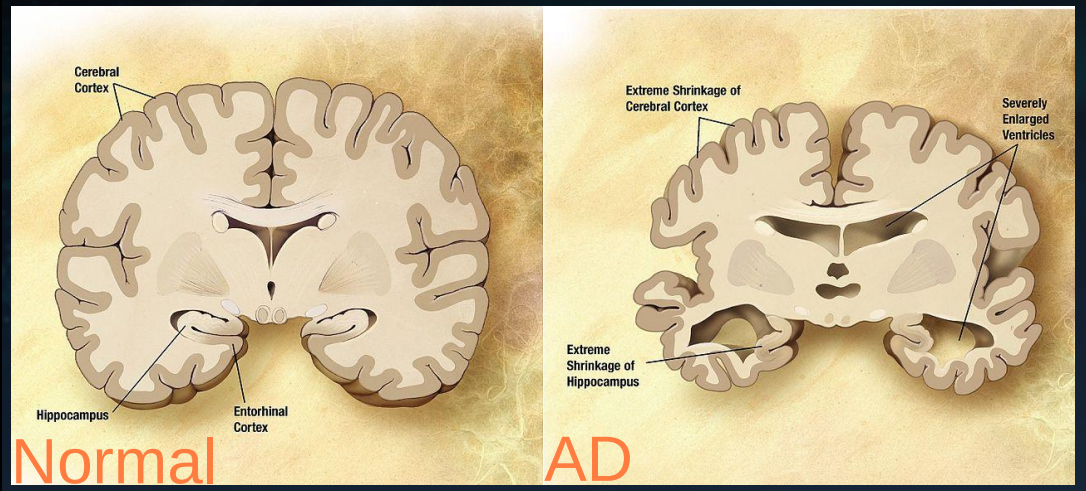


Neuropsychiatric disorders



Normal AD

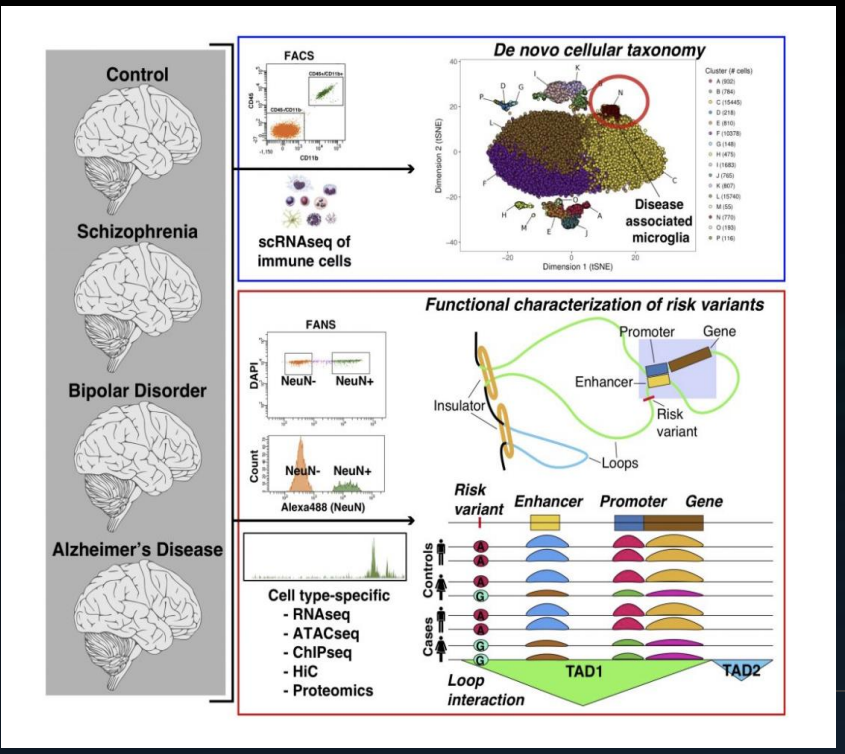
Alzheimer's disease (AD) is associated with brain shrinkage



Genetics of Brain Disorders

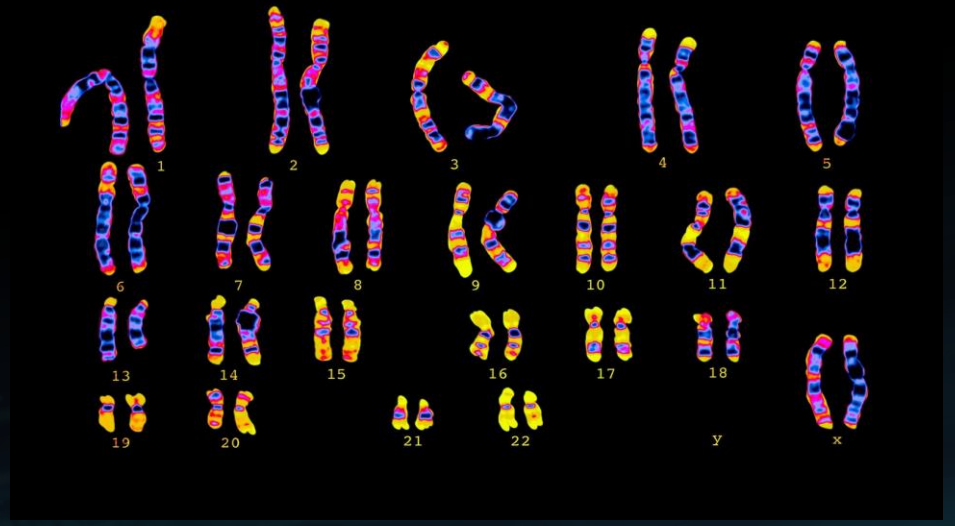
Most major brain disorders (like AD) are **heritable complex traits/diseases**

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



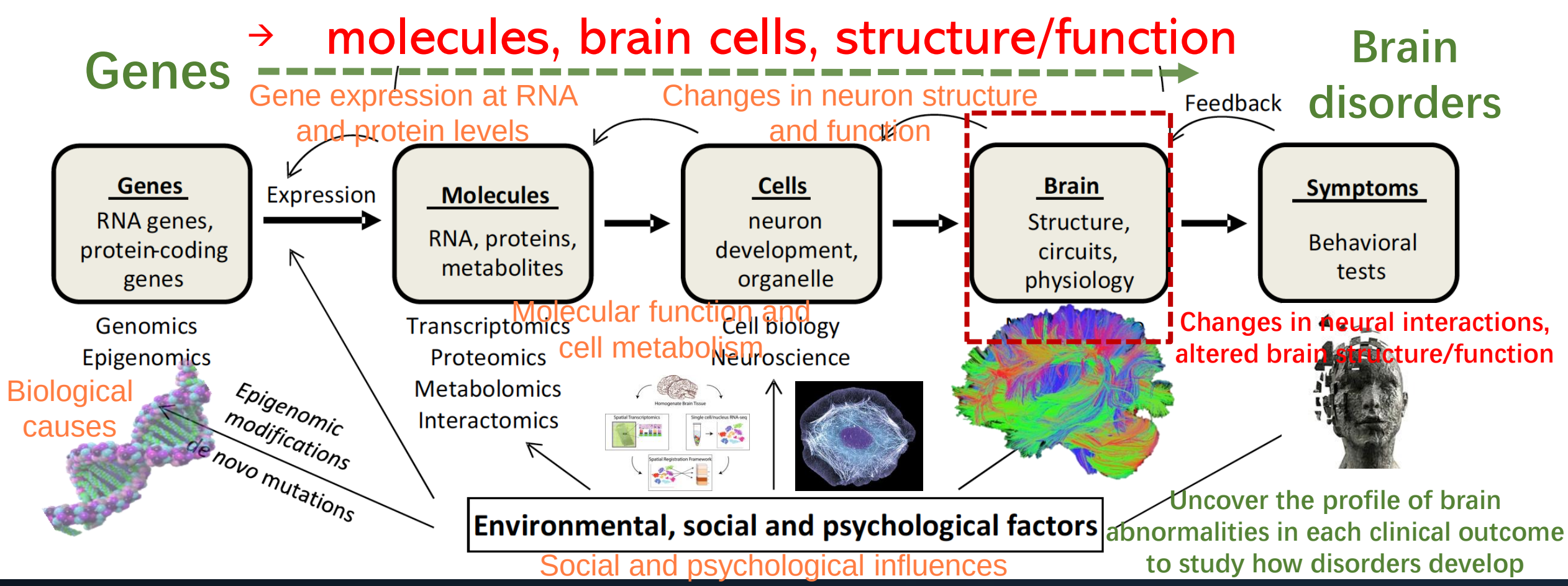
Complex traits/diseases (many genes, environmental factors, complex functional mechanism)

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



Brain Imaging Genetics Paradigm

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



“Big Data” Imaging Cohorts

“Big data” Brain imaging datasets become available in recent few years

Systematically collect publicly available individual-level data for **> 120k** individuals

Build the largest database in this field

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)*

HBCD *(Age [0,10])*

HCP-D *(Age [5,22])*

IMAGEN *(Age [14,22])*

NACC *(Age [36+])*

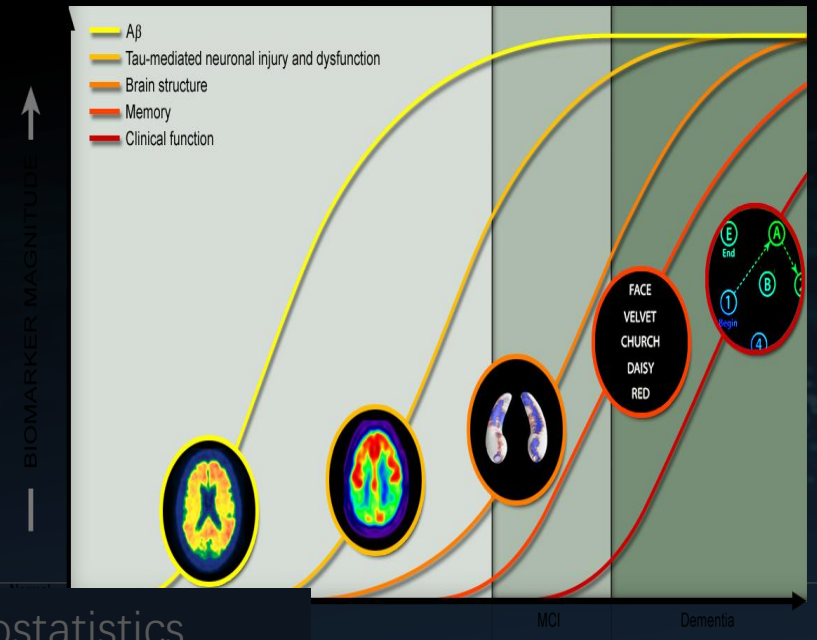
HCP-A *(Age [36+])*

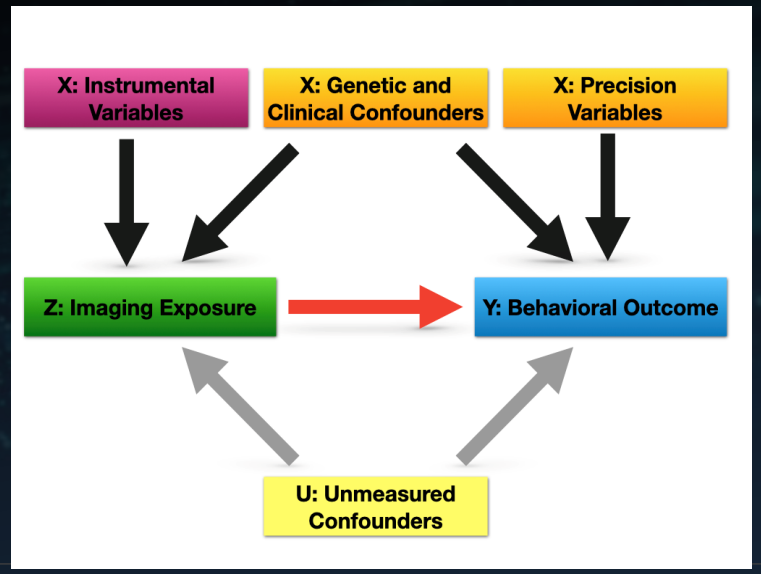
ADNI *(Age [55,92])*

Why Imaging Traits?

For a heterogeneous, clinically defined disorder, the endophenotype (or imaging traits) is 'closer to the underlying biology,'

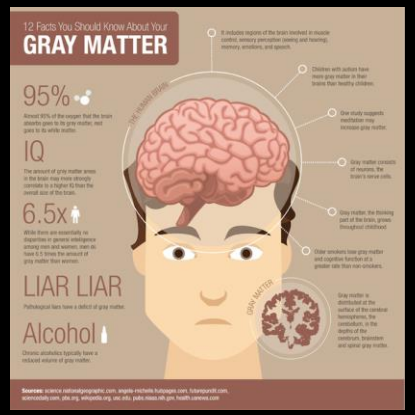
- **Be reproducible and heritable.**
- **Being informative about disorder risk.**
- **Providing mechanistic connections linking genetic variation to clinical measures.**
- **Some imaging traits (or brain circuits) may be treatable (e.g., ECT, TMS).**
- **Increasing the power of genetic search for polygenic genetic architecture.**



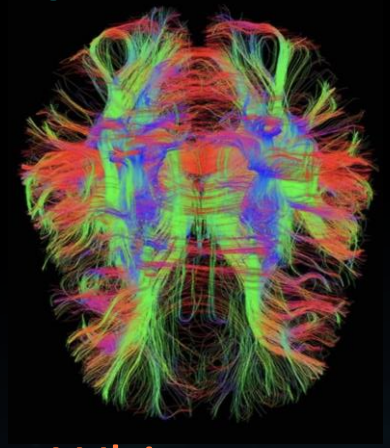


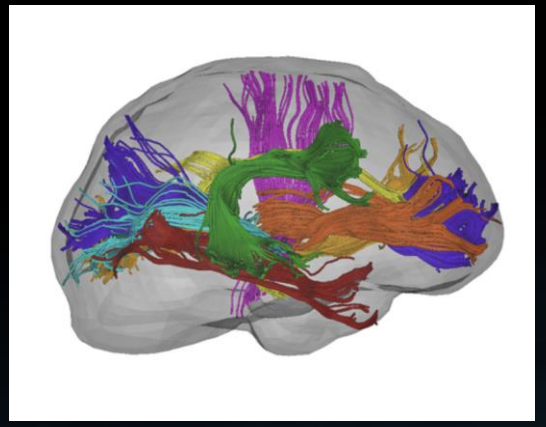
Brain Imaging Modality Examples

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

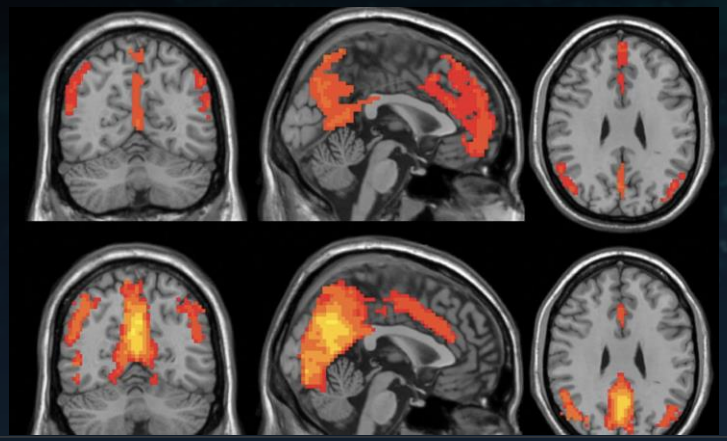


Cortical and subcortical structures





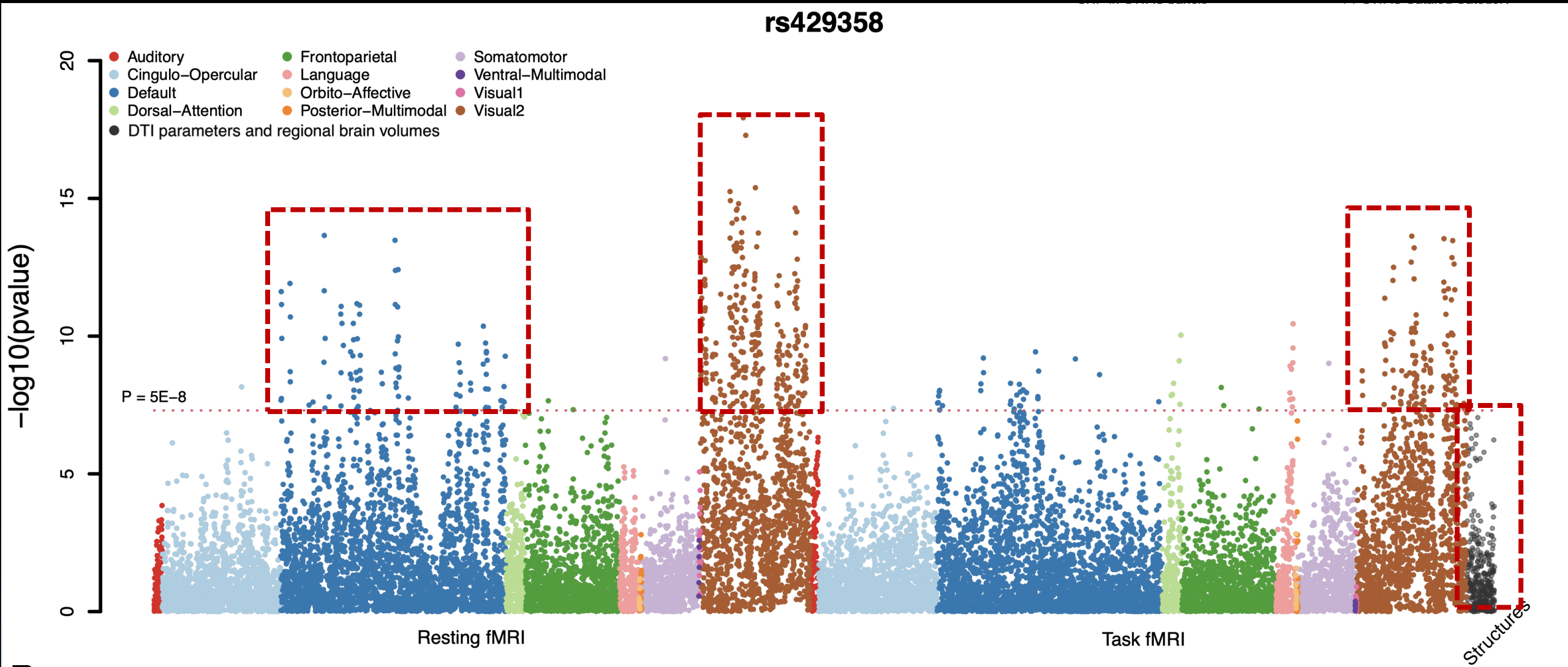
White matter microstructure (Structural connectivity, diffusion MRI)



Functional networks (Functional connectivity, functional MRI)

APOE-associations across functional networks

observations: 1) Enriched in the secondary visual and default mode networks;
2) Stronger connections in fMRI than in structural MRI.



It's just a beginning

Publications (2018+)

Heart–brain connections: Phenotypic and genetic insights from magnetic resonance images. ***Science*** 380, abn6598 (2023). LINK.

Genetic influences on the shape of brain ventricular and subcortical structures (2022). **medRxiv**, :

Common variants contribute to intrinsic human brain function networks (2022). ***Nature Genetics***.

Genetic influences on the intrinsic and extrinsic functional organizations of the cerebral cortex (2021). **medRxiv**, 21261187. LINK

Common genetic variation influencing human white matter microstructure (2021). ***Science***, 372–6548. LINK

Transcriptome–wide association analysis of brain structures yields insights into pleiotropy with complex neuropsychiatric traits (2021). ***Nature Communications***, 842872. LINK

Genome–wide association analysis of 19,629 individuals identifies variants influencing regional brain v –architecture with cognitive and mental health traits (2019). ***Nature Genetics***, *51*(11), 1637–1644. LINK

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

Hundreds of associated genetic variants for 2100+ neuroimaging traits across six modalities: (grey matter volume, white matter microstructure, resting-state functional connectivity+fMRI, task fMRI, shape, heart)

We make our research results publicly available by building the following resources.

If you are interested in other summary–level data from our analyses or have any questions or comments, feel free to contact Bingxin Zhao (bingxin@purdue.edu) or Hongtu Zhu (htzhu@email.unc.edu).

1. Imaging Genetics Online Server

We build a GWAS browser using the PheWeb tool to explore GWAS results for massive functional, structural, and diffusion neuroimaging traits. Currently, we support GWAS results of 2104 traits trained in the UKB British cohort (n~34,000), including

1. **635** ENIGMA–DTI parameters of brain white matter **(diffusion MRI)**
2. **376** ANTS regional brain volumes **(structural MRI)**
3. **191** ICA–based functional MRI traits **(rs–fMRI(ICA))**
4. **300** parcellation–based functional MRI **(task/rs–fMRI(Glasser360))**

Genetics discovery in human brain by big data integration

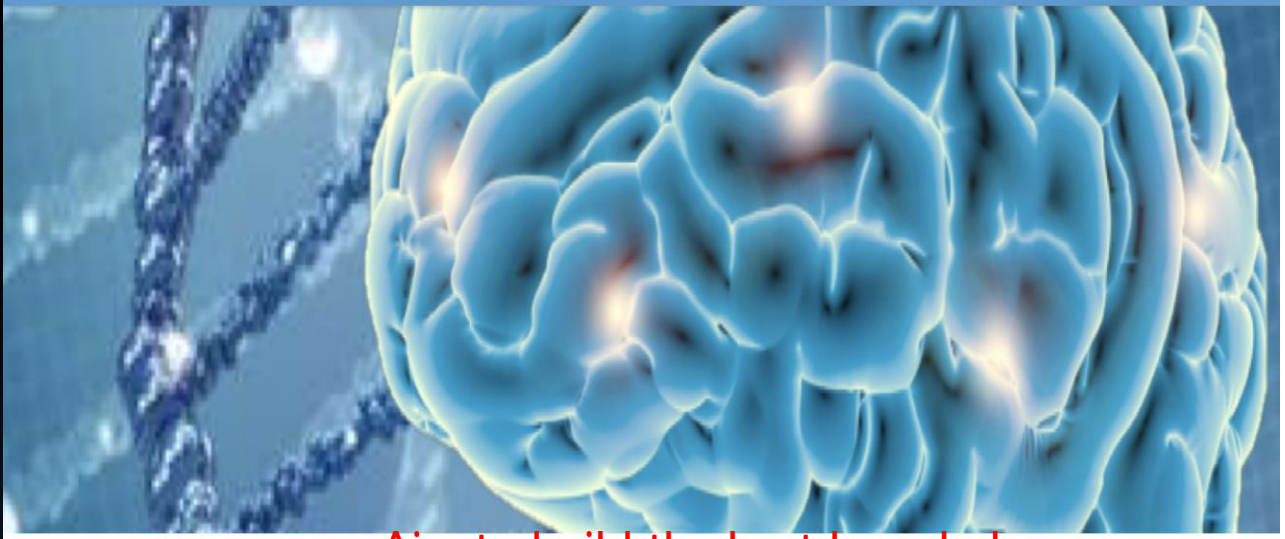
Brain Imaging Genetics Knowledge Portal

Brain Imaging Genetics Knowledge Portal (BIG-KP)

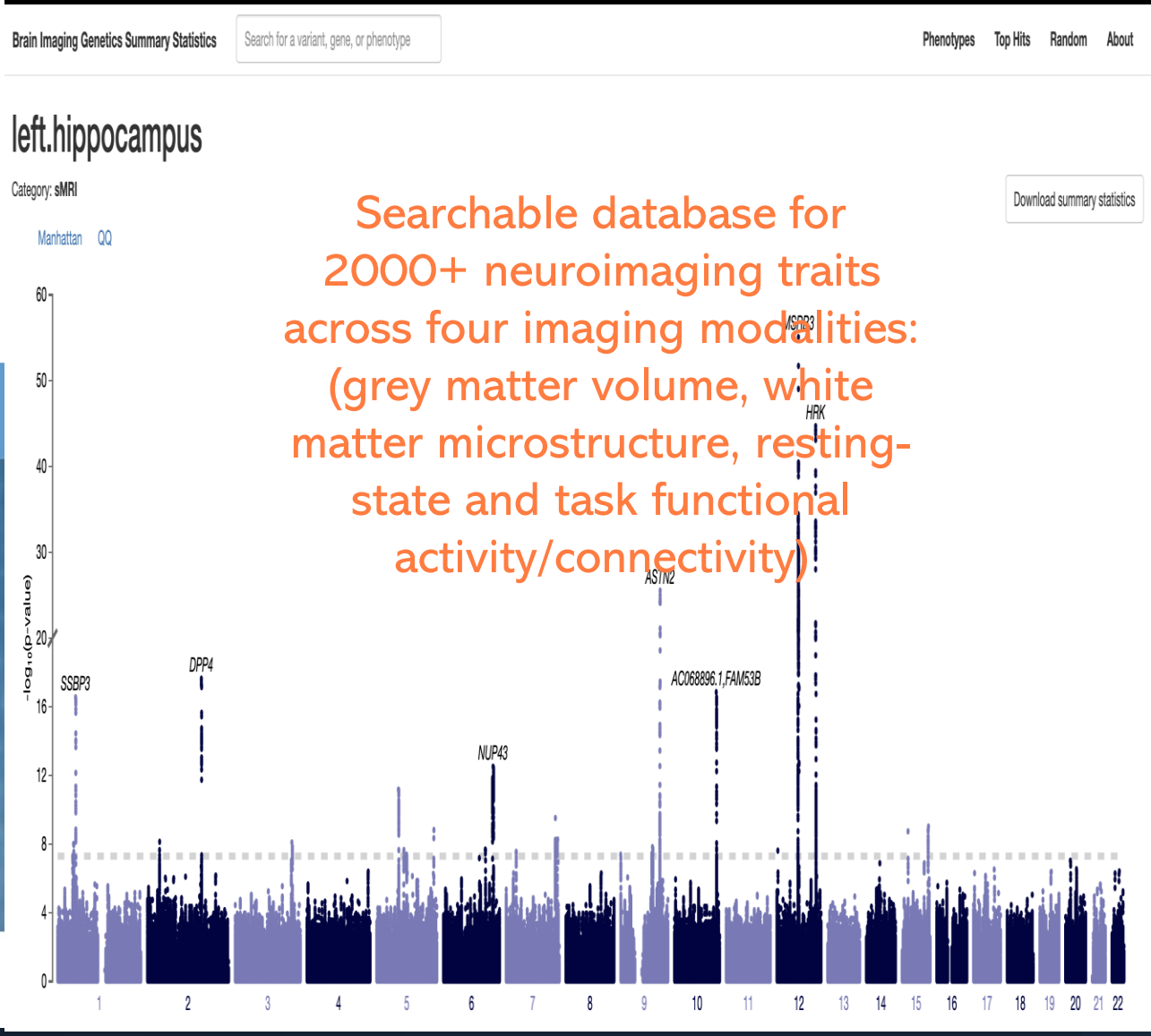
Genetics Discoveries in Human Brain by Big Data Integration

bigkp.org

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

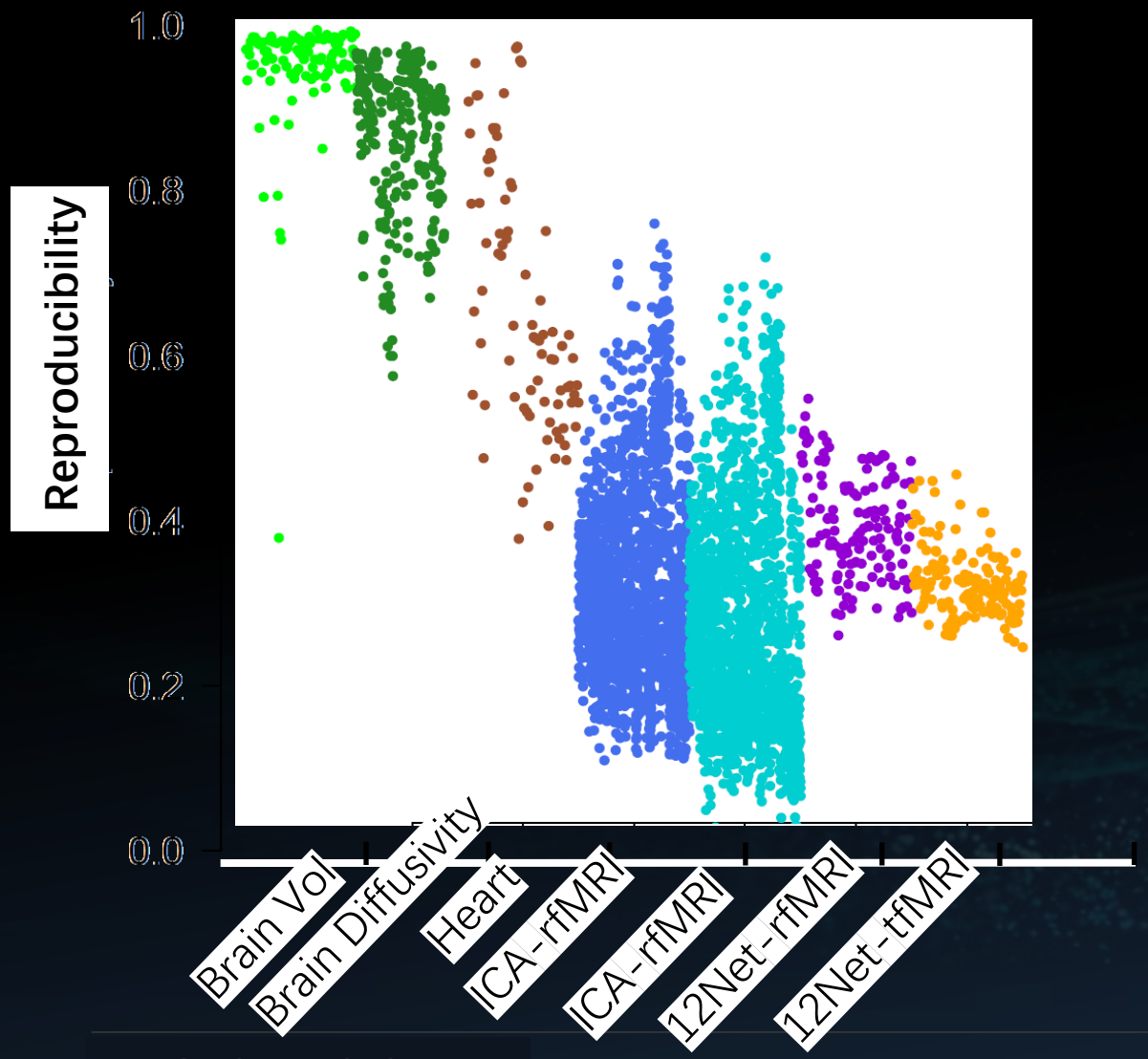


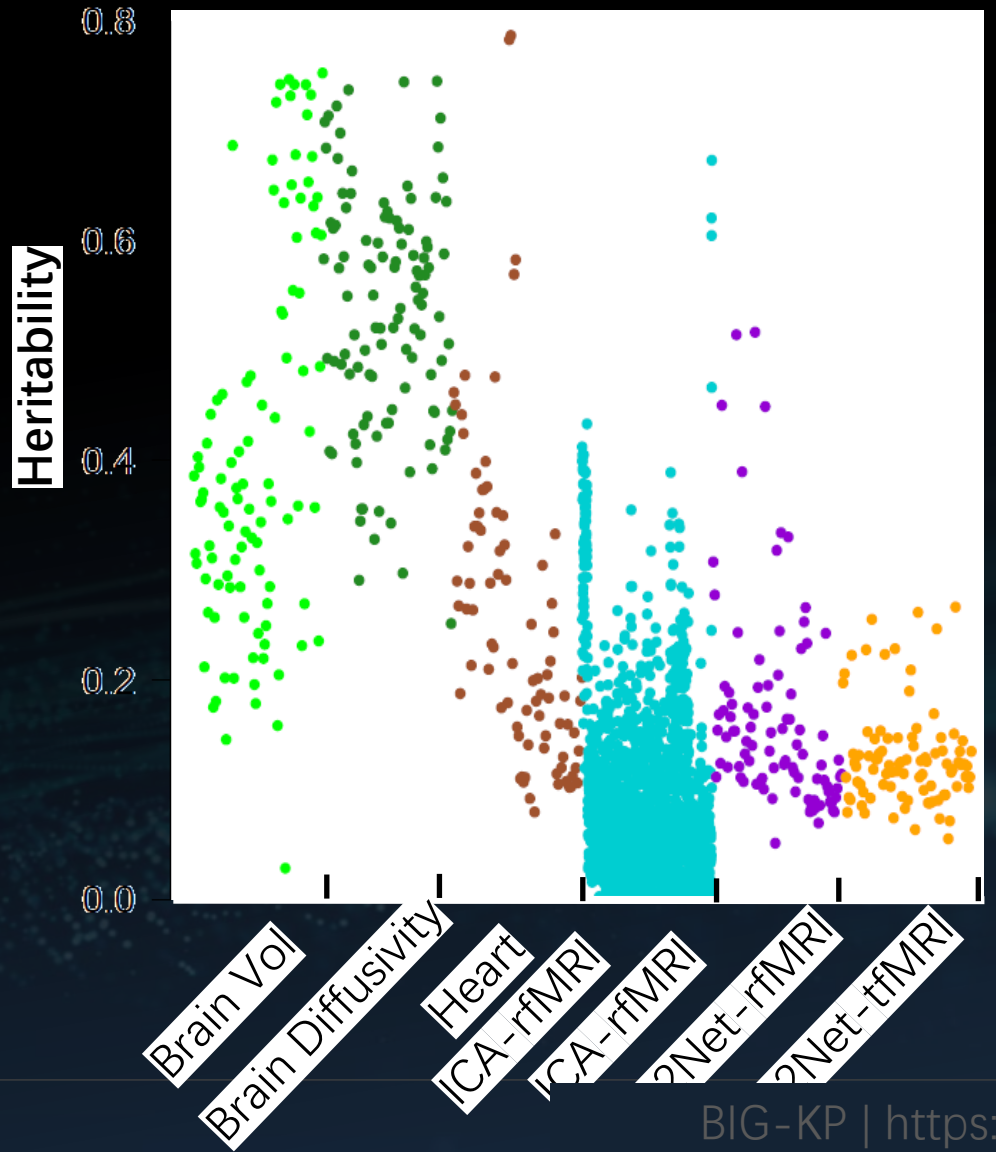
Aim to build the best knowledge database of neuroimaging genetics



Searchable database for 2000+ neuroimaging traits across four imaging modalities: (grey matter volume, white matter microstructure, resting-state and task functional activity/connectivity)

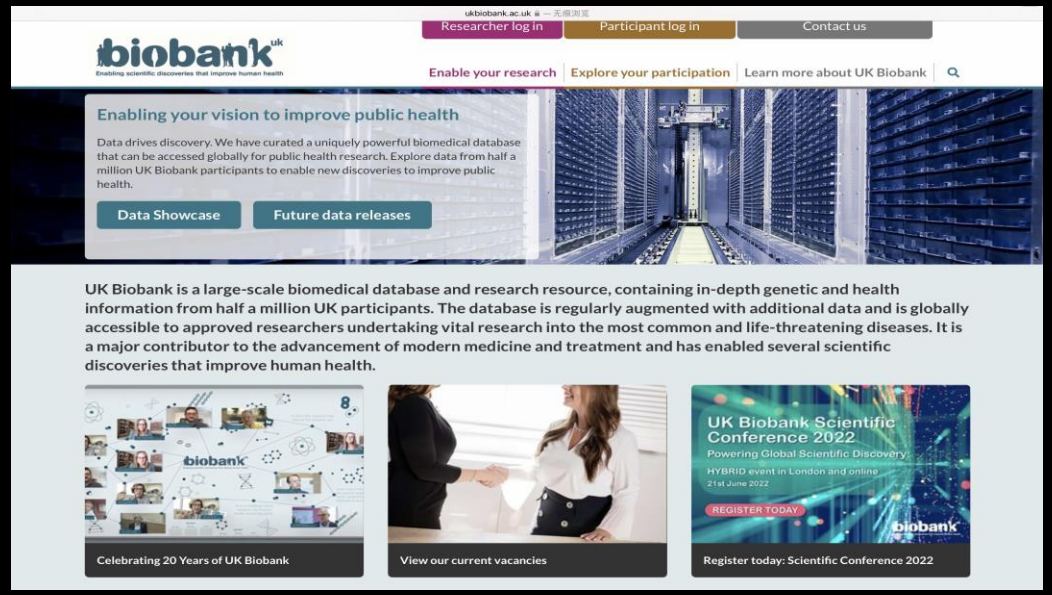
IG: Reproducibility and Heritability



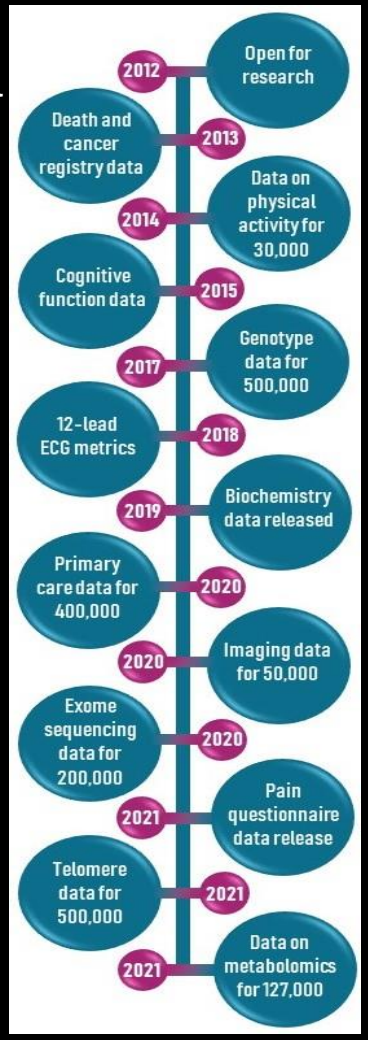


The UK Biobank Study

UK Biobank has collected and continues to collect extensive environmental, lifestyle, and genetic data on half a million participants.



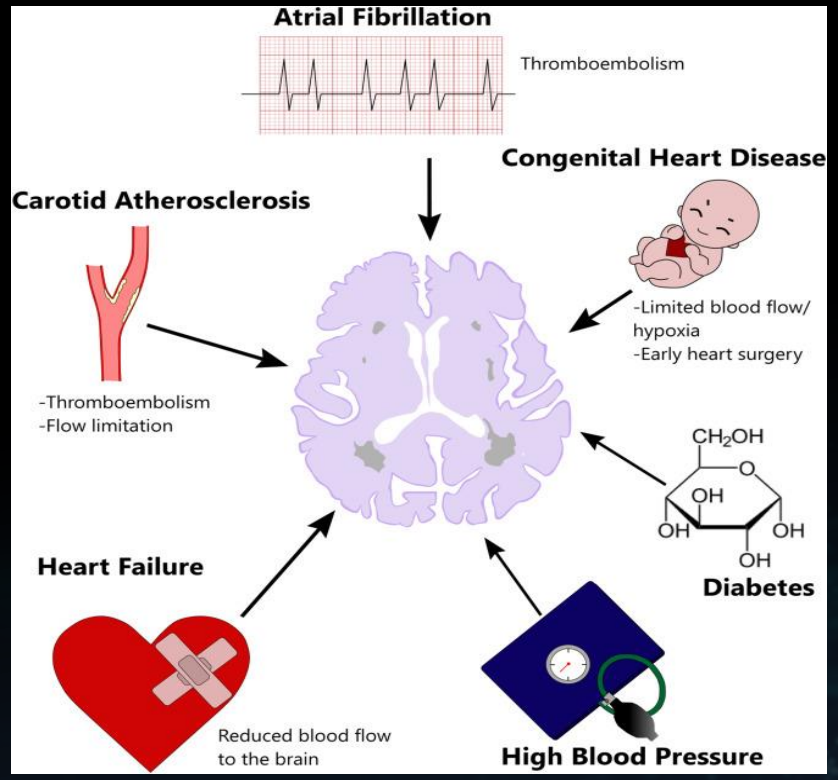
2006-now



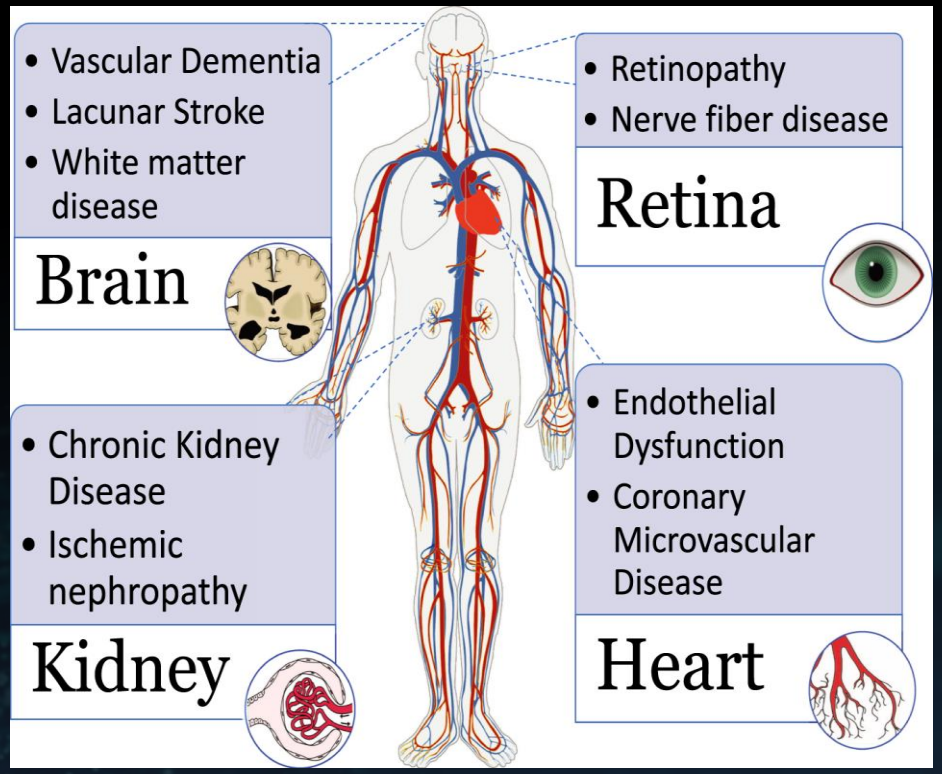
- **Imaging:** Brain, heart and full body MR imaging, plus full body DEXA scan of the bones and joints and an ultrasound of the carotid arteries. The goal is to image 100,000 participants, and to invite participants back for a repeat scan some years later.
- **Genetics:** Genotyping, whole exome sequencing & whole genome sequencing for all participants.
- **Health linkages:** Linkage to a wide range of electronic health-related records, including death, cancer, hospital admissions and primary care records.
- **Biomarkers:** Data on more than 30 key biochemistry markers from all participants, taken from samples collected at recruitment and the first repeat assessment.
- **Activity monitor:** Physical activity data over a 7-day period collected via a wrist-worn activity monitor for 100,000 participants plus a seasonal follow-up on a subset.
- **Online questionnaires:** Data on a range of exposures and health outcomes that are difficult to assess via routine health records, including diet, food preferences, work history, pain, cognitive function, digestive health and mental health.
- **Repeat baseline assessments:** A full baseline assessment is undertaken during the imaging assessment of 100,000 participants.
- **Samples:** Blood & urine was collected from all participants, and saliva for 100,000.

Multiorgan Dysfunction Syndromes

Imaging: help understand the complex interplay between brain and other human organs and their underlying genetic overlaps



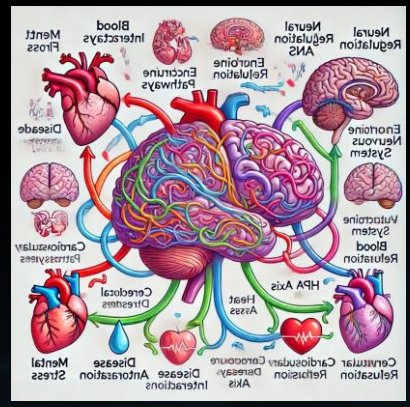
Possible causal factors of brain structure changes, resulting in brain disorders like stroke, dementia and cognitive impairment



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

The Brain-Heart Axis

The brain-heart axis refers to the bidirectional communication between the brain and the heart, playing a crucial role in regulating physiological functions and maintaining overall health.



Neural Regulation:

- **Autonomic Nervous System (ANS):** regulate heart rate, blood pressure, and cardiac output.
- **Vagus Nerve:** reduce heart rate and promoting relaxation.

Endocrine Pathways:

- **Hypothalamic-Pituitary-Adrenal (HPA) Axis:** Influences heart function through the release of hormones, affecting blood pressure and cardiovascular health.
- **Catecholamines:** Adrenaline and noradrenaline released during stress increase heart rate and cardiac output.

Blood Flow and Oxygen Supply:

- **Cerebral Perfusion:** The heart ensures a continuous supply of oxygenated blood to the brain, essential for cognitive functions and neural health.
- **Cerebral Autoregulation:** Mechanisms that maintain stable blood flow to the brain despite changes in systemic blood pressure.

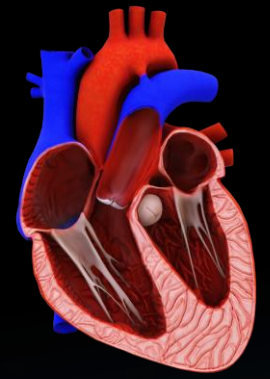
The Brain-Heart Axis

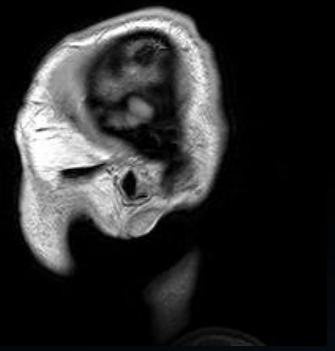
Disease Interactions:

- **Cardiovascular Diseases:** Conditions like atrial fibrillation and heart failure are linked to brain diseases such as stroke, dementia, and cognitive impairment due to reduced cerebral perfusion.
- **Mental Disorders:** Mental illnesses, including schizophrenia, bipolar disorder, epilepsy, and depression, increase the risk of CVD.

Acute Mental Stress:

- **Impact on Cardiovascular Health:** Acute stress can cause vascular inflammation and increase the risk of atherosclerosis due to stress-induced leukocyte migration.

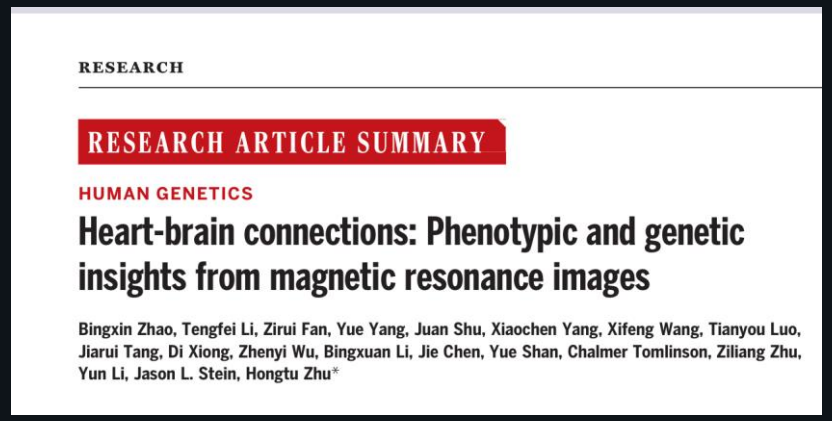




Research Significance:

- **Integrated Treatment Approaches:** Lead to better treatments for neurocardiological disorders.
- **Comprehensive Studies:** A need for larger studies to provide a complete picture of the structural and functional links between heart and brain health.

Overview



RESEARCH

RESEARCH ARTICLE SUMMARY

HUMAN GENETICS

Heart-brain connections: Phenotypic and genetic insights from magnetic resonance images

Bingxin Zhao, Tengfei Li, Zirui Fan, Yue Yang, Juan Shu, Xiaochen Yang, Xifeng Wang, Tianyou Luo, Jiarui Tang, Di Xiong, Zhenyi Wu, Bingxuan Li, Jie Chen, Yue Shan, Chalmer Tomlinson, Ziliang Zhu, Yun Li, Jason L. Stein, Hongtu Zhu*

Genetics

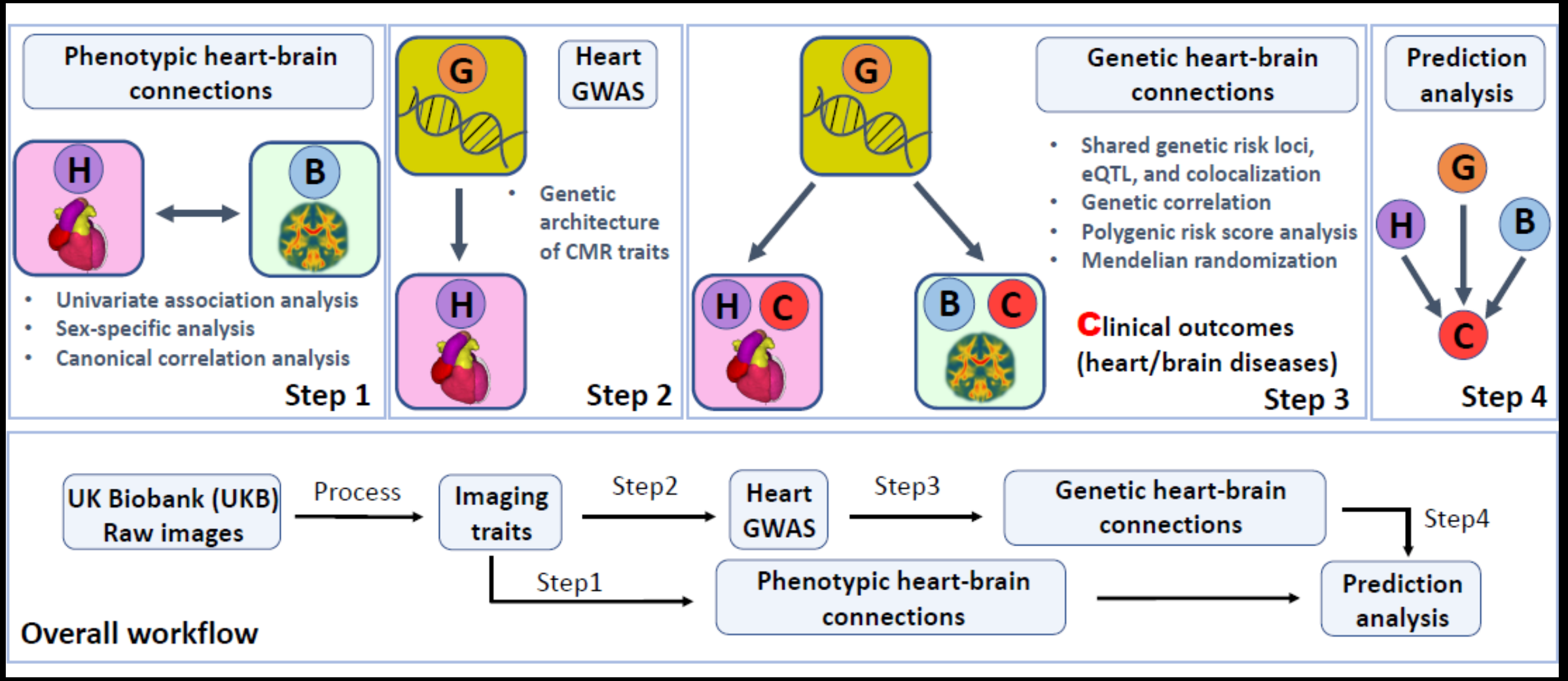
Left ventricle
Right ventricle
Left atrium
Right atrium
Ascending aorta
Descending aorta

Heart health
(measured by CMR)

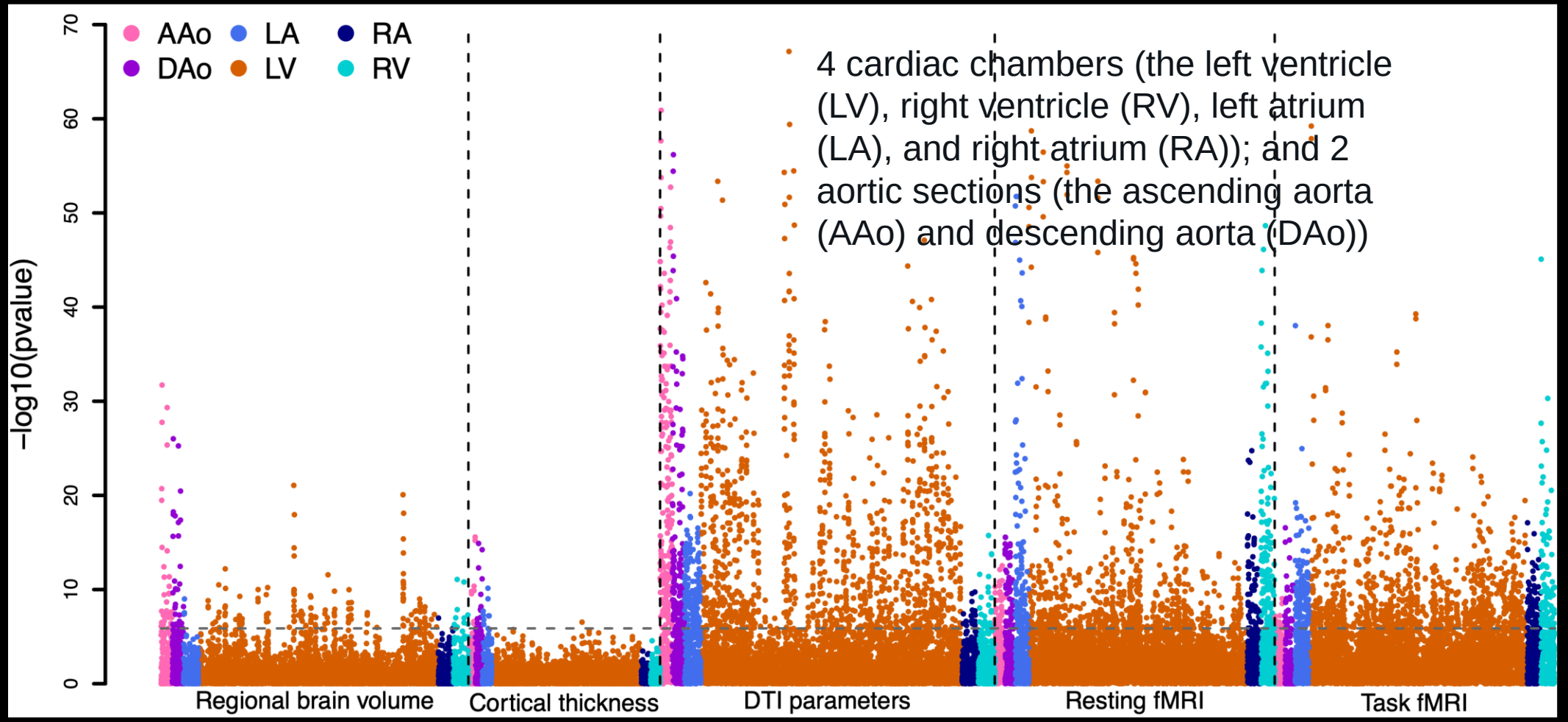
Brain health
(measured by brain MRI)

Structural MRI
Diffusion MRI
Resting functional MRI
Task functional MRI

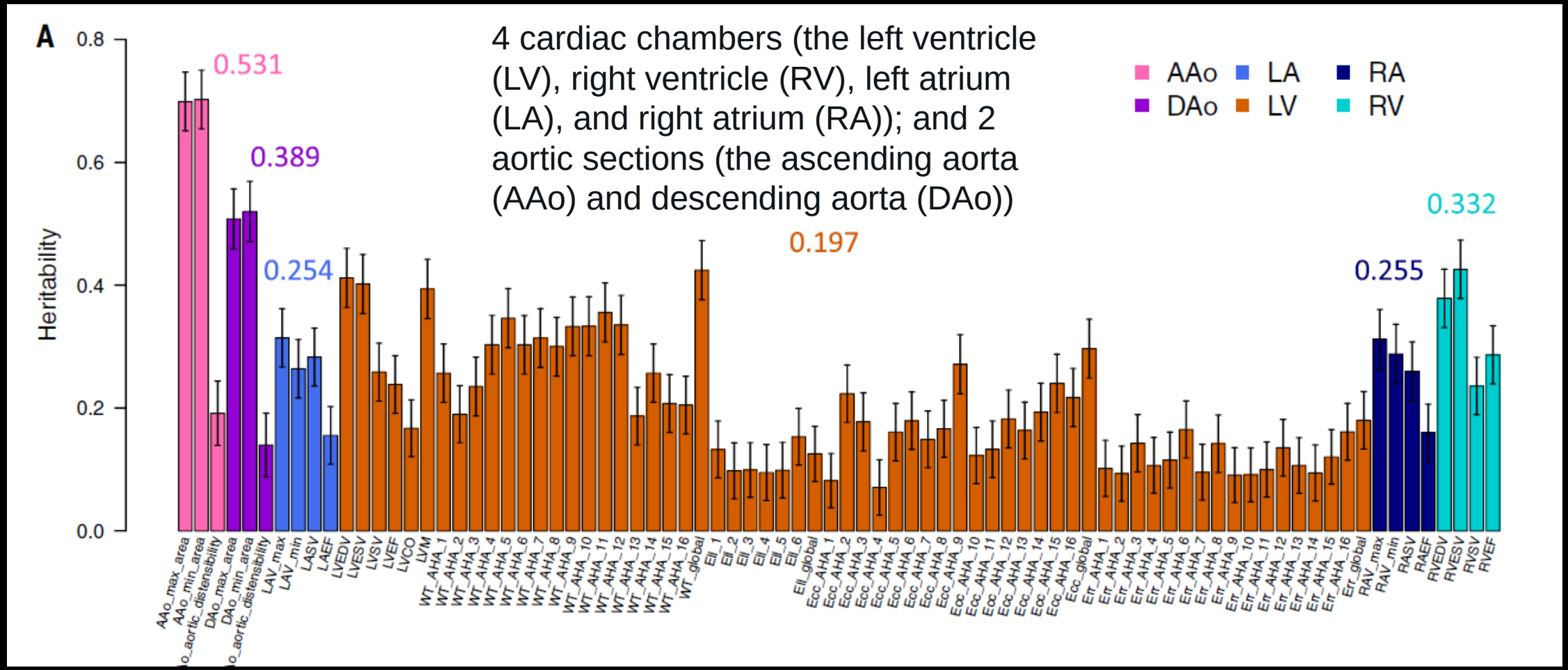
Overview


Phenotypic heart-brain connections
H
B
Univariate association analysis
Sex-specific analysis
Canonical correlation analysis
Step 1
G
Heart GWAS
Genetic architecture of CMR traits
H
Step 2
G
H C
B C
Genetic heart-brain connections
Shared genetic risk loci, eQTL, and colocalization
Genetic correlation
Polygenic risk score analysis
Mendelian randomization
Clinical outcomes (heart/brain diseases)
Step 3
Prediction analysis
G
H
B
C
Step 4
UK Biobank (UKB) Raw images
Process
Imaging traits
Step2
Heart GWAS
Step3
Genetic heart-brain connections
Step4
Step1
Phenotypic heart-brain connections
Prediction analysis
Overall workflow

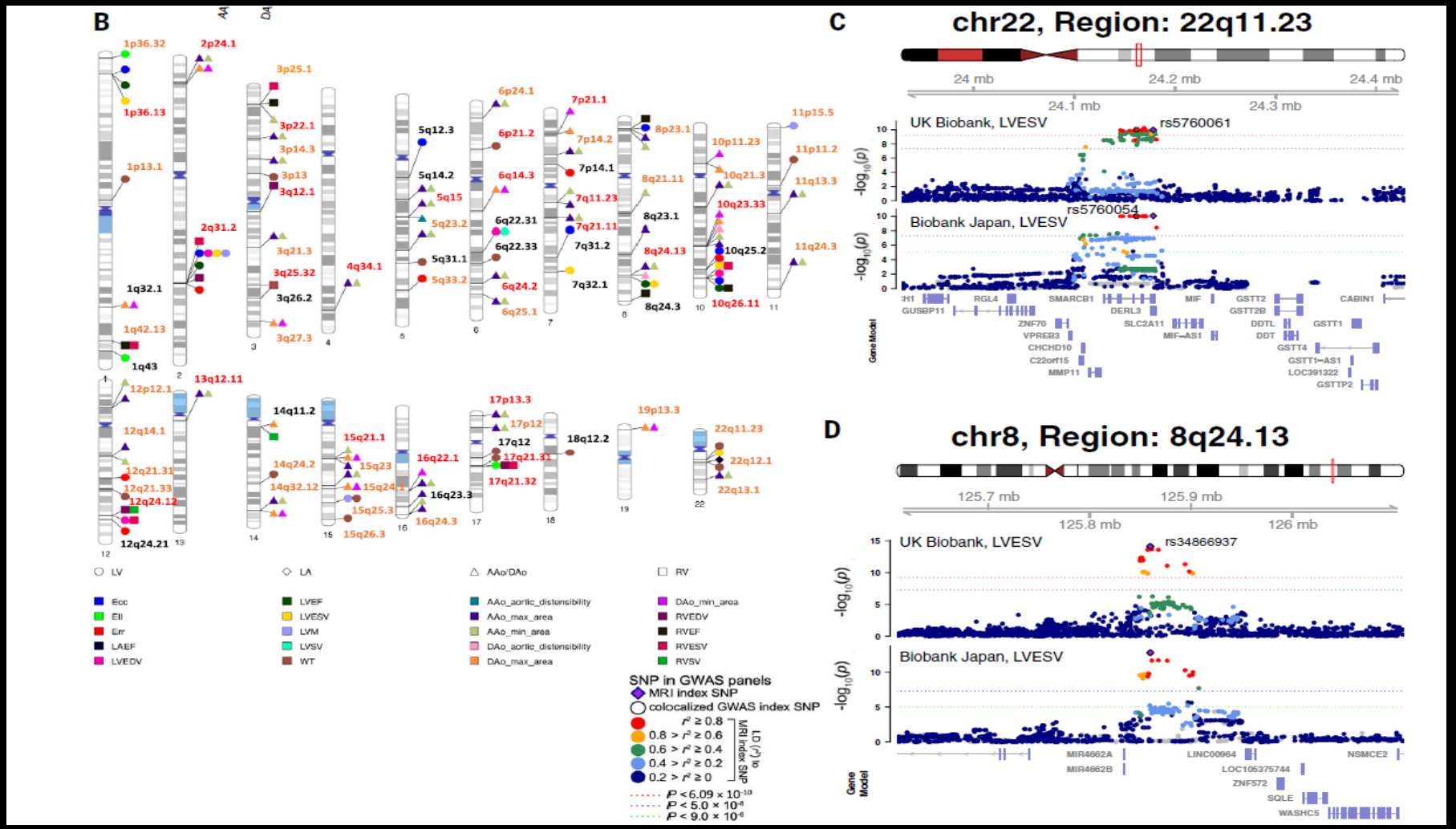
Phenotypic Heart-Brain Connections



SNP heritability of 82 CMR traits



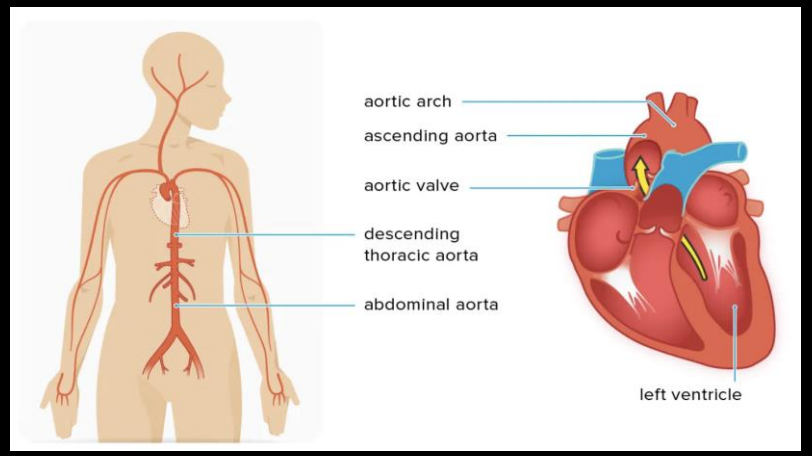
Genetics of CMR traits in the UKB



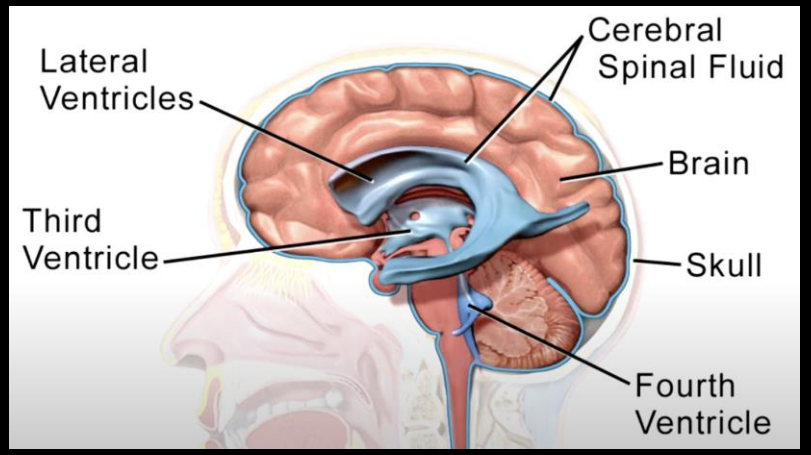
Ideogram of 80 genomic regions associated with CMR traits

LV end-systolic volume (LVESV)

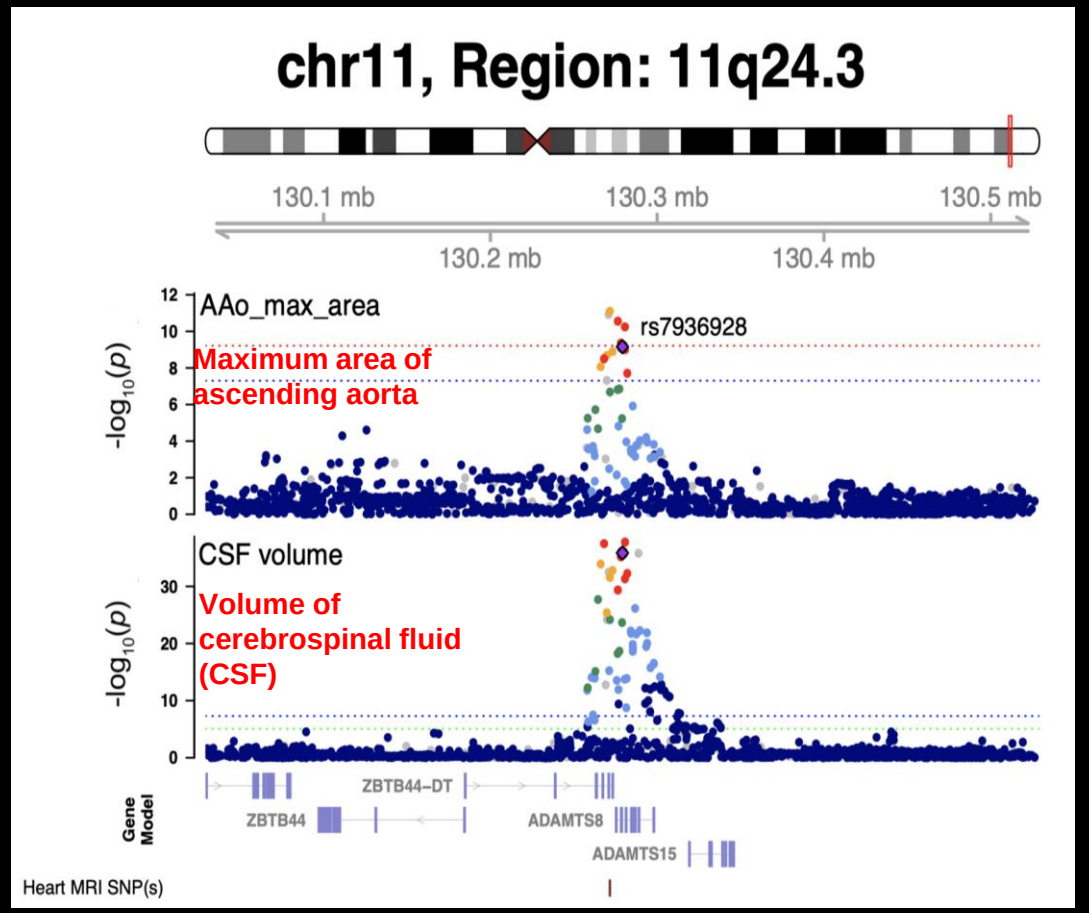
Selected genetic loci



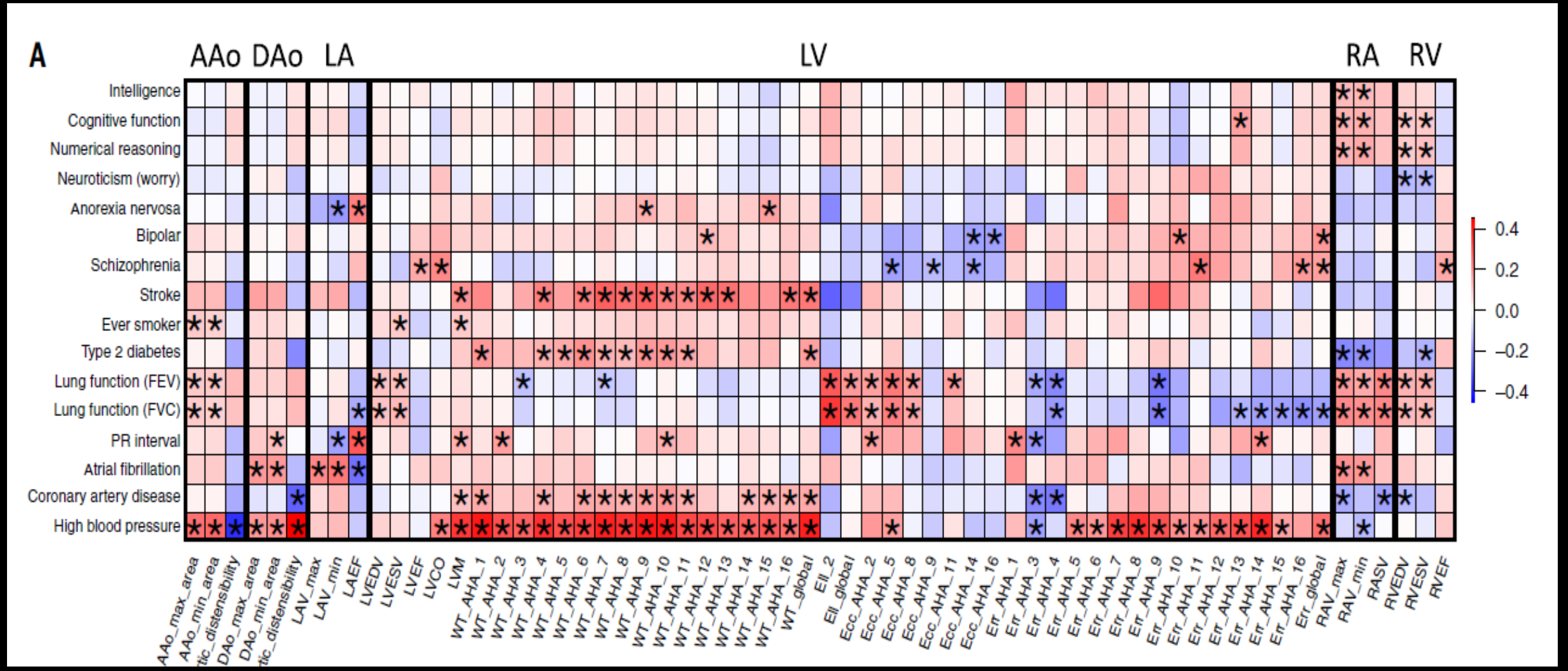
Heart MRI



Brain MRI

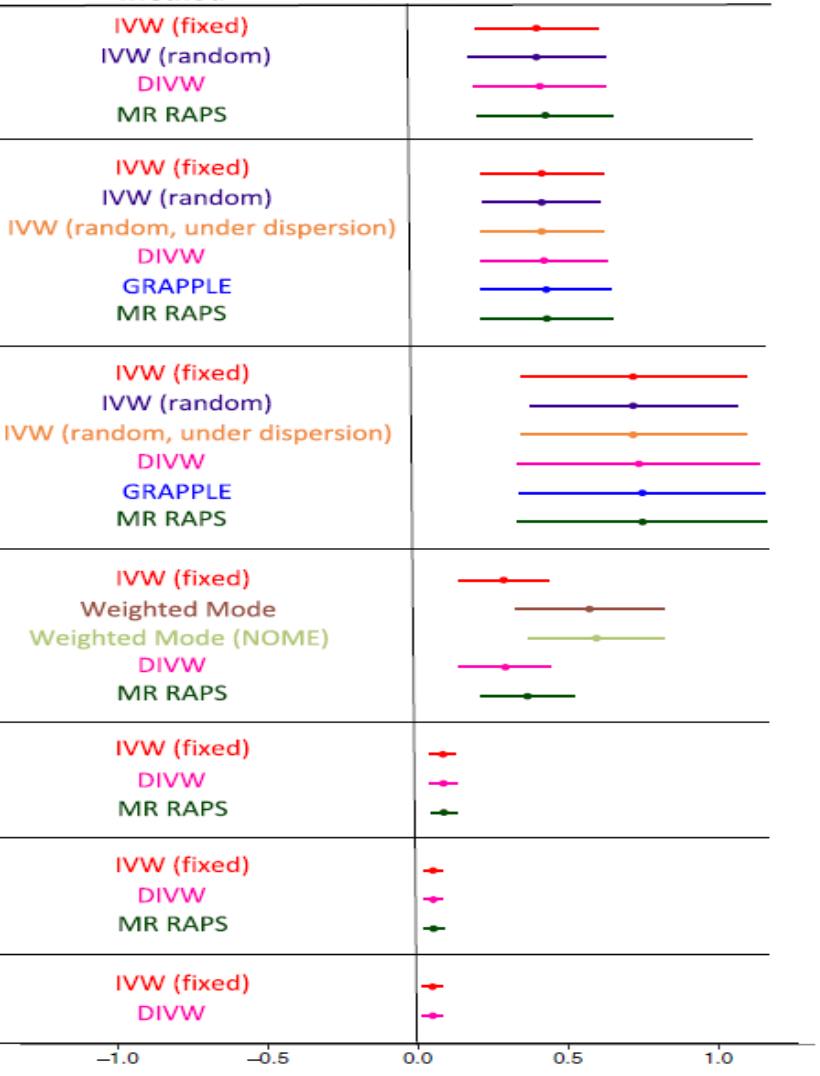


Genetic Correlations



MR: Causal heart-brain relationships

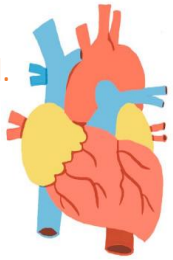
Exposure	Category	Outcome	#IVs	Method
WT AHA 9	LV	Bipolar disorder	7	IVW (fixed) IVW (random) DIVW MR RAPS
LVESV	LV	Cross disorders	12	IVW (fixed) IVW (random) IVW (random, under dispersion) DIVW GRAPPLE MR RAPS
WT AHA 12	LV	Cross disorders	6	IVW (fixed) IVW (random) IVW (random, under dispersion) DIVW GRAPPLE MR RAPS
AAo min area	AAo	Cross disorders	34	IVW (fixed) Weighted Mode Weighted Mode (NOME) DIVW MR RAPS
WT AHA 11	LV	Depression	7	IVW (fixed) DIVW MR RAPS
AAo max area	AAo	Psychiatric diseases (FinnGen)	47	IVW (fixed) DIVW MR RAPS
AAo min area	AAo	Psychiatric diseases (FinnGen)	50	IVW (fixed) DIVW



Potential heart-brain relationships

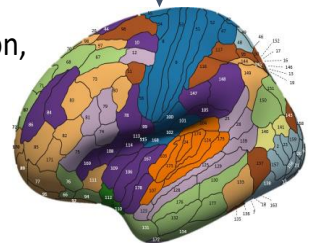
Coronary artery disease (Krantz et al. (*12*), PMID: 24677165)
Atrial fibrillation (Kwok et al. (*6*), PMID: 21383328)
Heart failure (Vogels et al. (*9*), PMID: 17174152)
Myocardial infarction (Levine et al. (*11*), PMID: 33486973)
Ventricular arrhythmias (Abisse et al. (*13*), PMID: 21920534)
Left ventricular hypertrophy (Papadopoulos et al. (*3*), PMID: 32635685)

Shared genetic and genomic factors
(For example, shared risk loci, Broce et al. (*2*), PMID: 30413934)



Cardiovascular disease/dysfunction
(For example, chronic systemic hypoperfusion, Moroni et al. (*5*), PMID: 29946567)

Mental health disorders and negative psychological factors
(For example, mental stress, Hinterdobler et al. (*18*), PMID: 34279021)



Heart health (measured by CMR)

Brain health (measured by brain MRI)

Stroke (Gardener et al. (*1*), PMID: 26481296)
Alzheimer's Disease (Broce et al. (*2*), PMID: 30413934)
Cognitive impairment (Abete et al. (*4*), PMID: 25107566)
Schizophrenia (Jindal et al. (*14*), PMID: 16327872)
Bipolar disorder (Nielsen et al. (*15*), PMID: 33128044)
Depression (Tawakol et al. (*16*), PMID: 28088338)
Epilepsy (Verrier et al. (*17*), PMID: 32109857)

Aging and specific disease processes
(For example, diabetes, Jensen et al. (*20*), PMID: 31522551)

Cardiovascular risk factors
(For example, smoking, high blood pressure, high cholesterol, and unhealthy diet, Cox et al. (*19*), PMID: 30854560)

Mental health issues-induced behavioral and biological processes
(For example, smoking initiation and dysregulation of the autonomic nervous system, Levine et al. (*11*), PMID: 33486973)

MORE THA
滴滴

The Brain-Eye Axis

The **brain-eye axis** refers to the complex and interconnected relationship between the brain and the eyes, essential for visual perception, processing, and overall neurological function.

Anatomical Connections:

- **Optic Nerve:** The **optic nerve** transmits visual information from the retina to the brain's visual centers.
- **Retina as a Brain Extension:** The **retina** contains photoreceptor cells that detect light and initiate vision.

Visual Processing:

- **Primary Visual Cortex (V1)** in the occipital lobe. V1 processes visual information received from the eyes.
- **Higher Visual Areas:** For object recognition, spatial awareness, and motion detection.

Eye Movements and Brain Function:

Oculomotor Control: Eye movements are controlled by cranial nerves (III, IV, VI). **Attention and Cognition:** The brain coordinates eye movements.

Feedback Loops:

- **Top-Down Control:** The brain modulates visual processing based on contextual information, attention, and experience.
- **Autonomic Regulation:** The brain regulates the pupil and lens to control how much light enters the eye and focuses on objects at different distances.

The Brain-Eye Axis

Health and Disease

Neurodegenerative Diseases: Changes in the retina and optic nerve can be early indicators of neurological diseases like Alzheimer's, Parkinson’s, and multiple sclerosis.

Visual Disturbances: Conditions like glaucoma, optic neuritis, and retinal degeneration affect both vision and brain function, showing how interdependent the brain and eyes are.

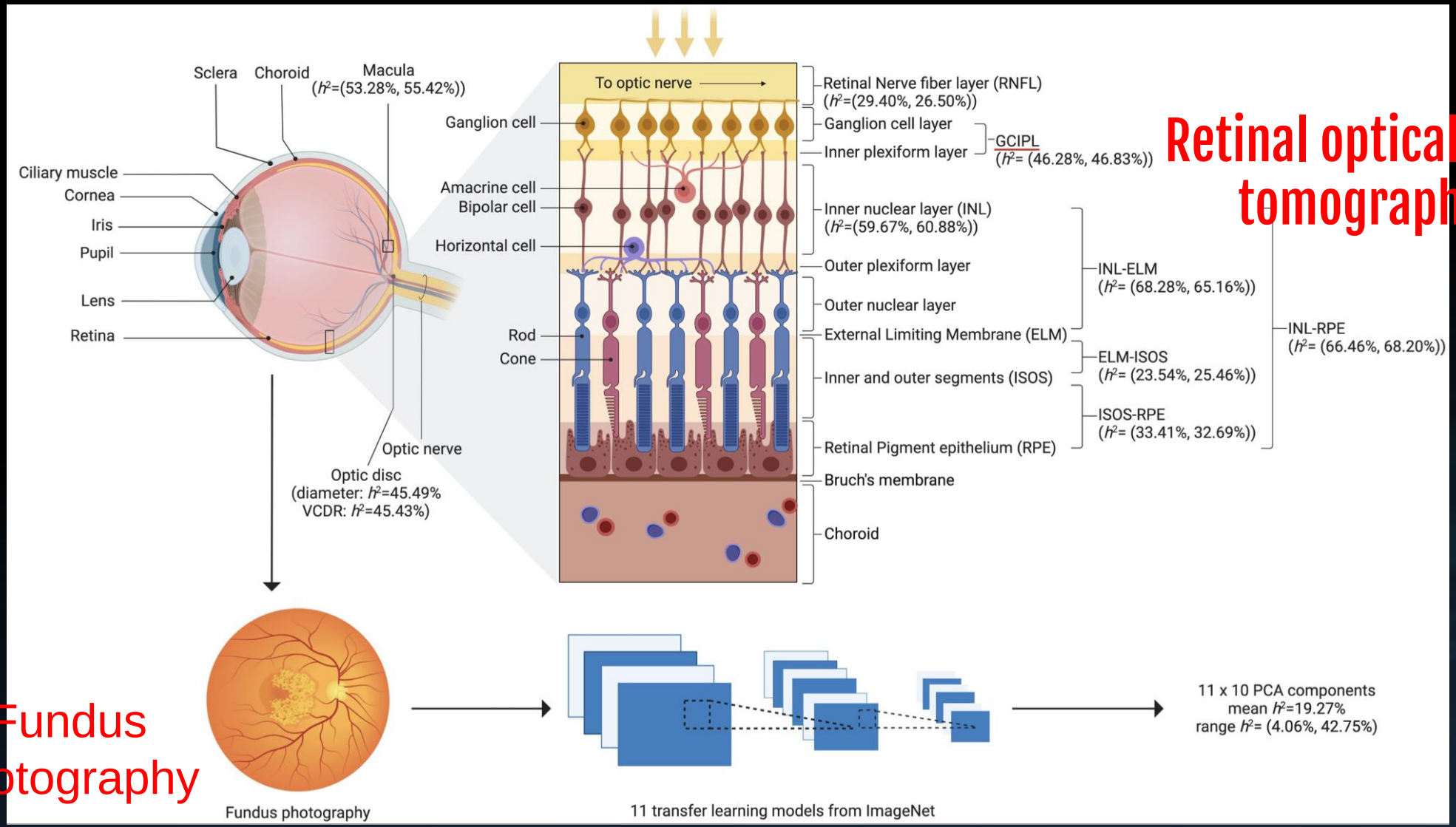
Stroke and Vision: Strokes that affect the brain’s visual pathways can lead to visual field defects and other disturbances.

Clinical Relevance

Diagnostics: Eye health and retinal imaging can serve as non-invasive markers for detecting brain health, especially in diseases like Alzheimer's and multiple sclerosis.

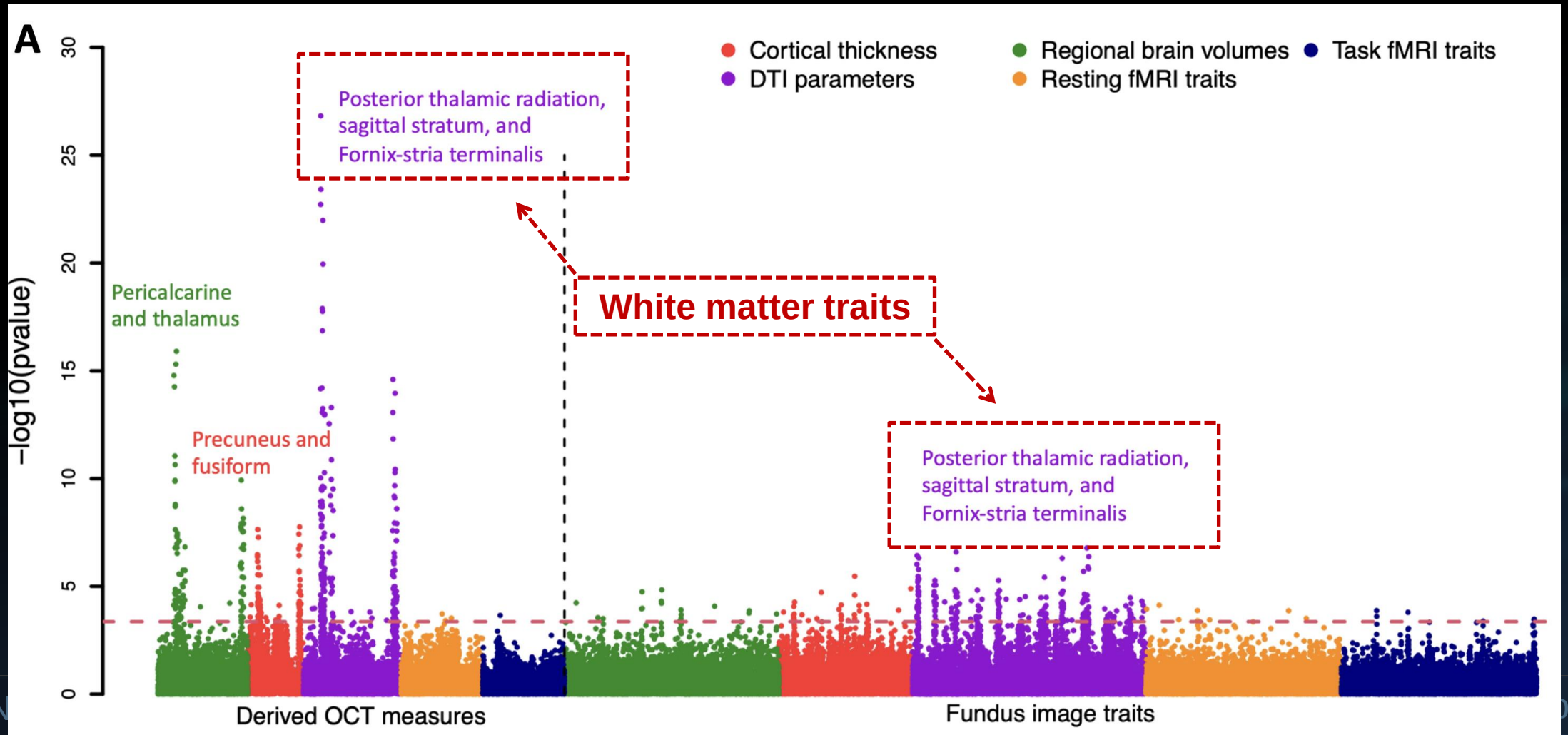
Neuro-Ophthalmology: This field bridges neurology and ophthalmology, focusing on conditions that affect both the brain and eyes, such as optic neuropathies and visual processing disorders.

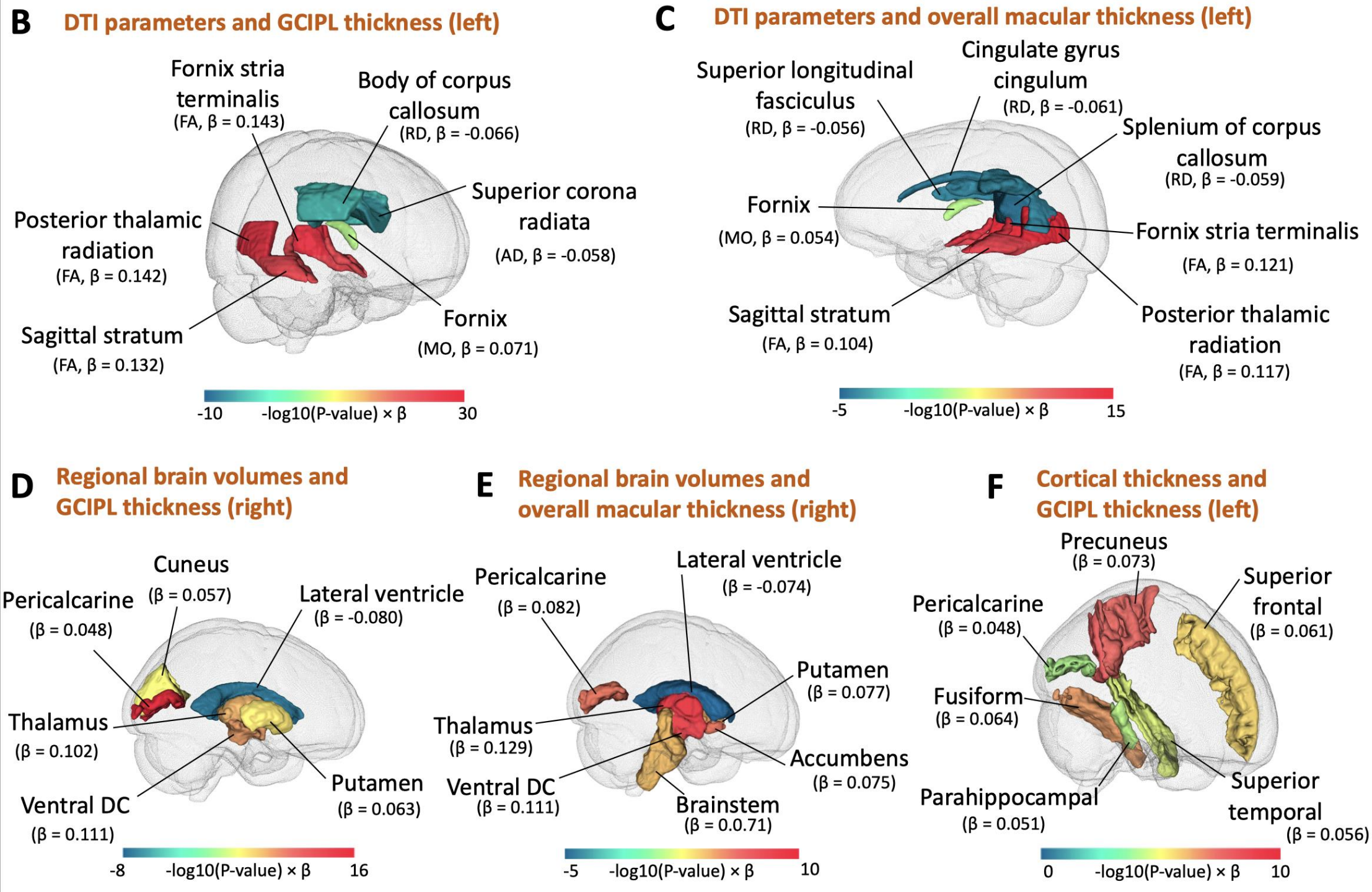
Eye as a window to the brain?



Phenotypic Eye-Brain Connections

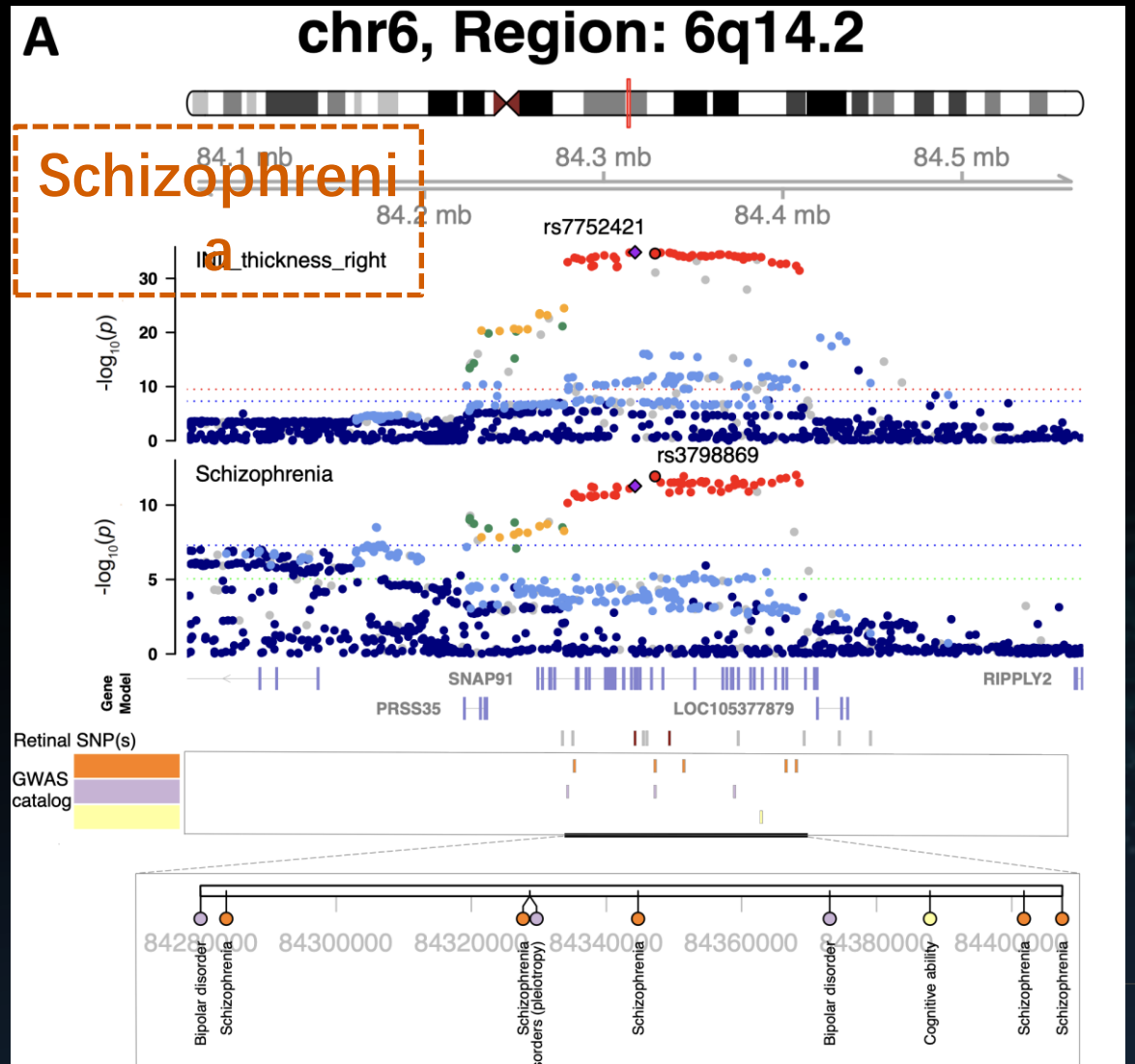
Both OCT and fundus imaging traits are associated with regional brain volumes, cortical thickness, and white matter microstructures.

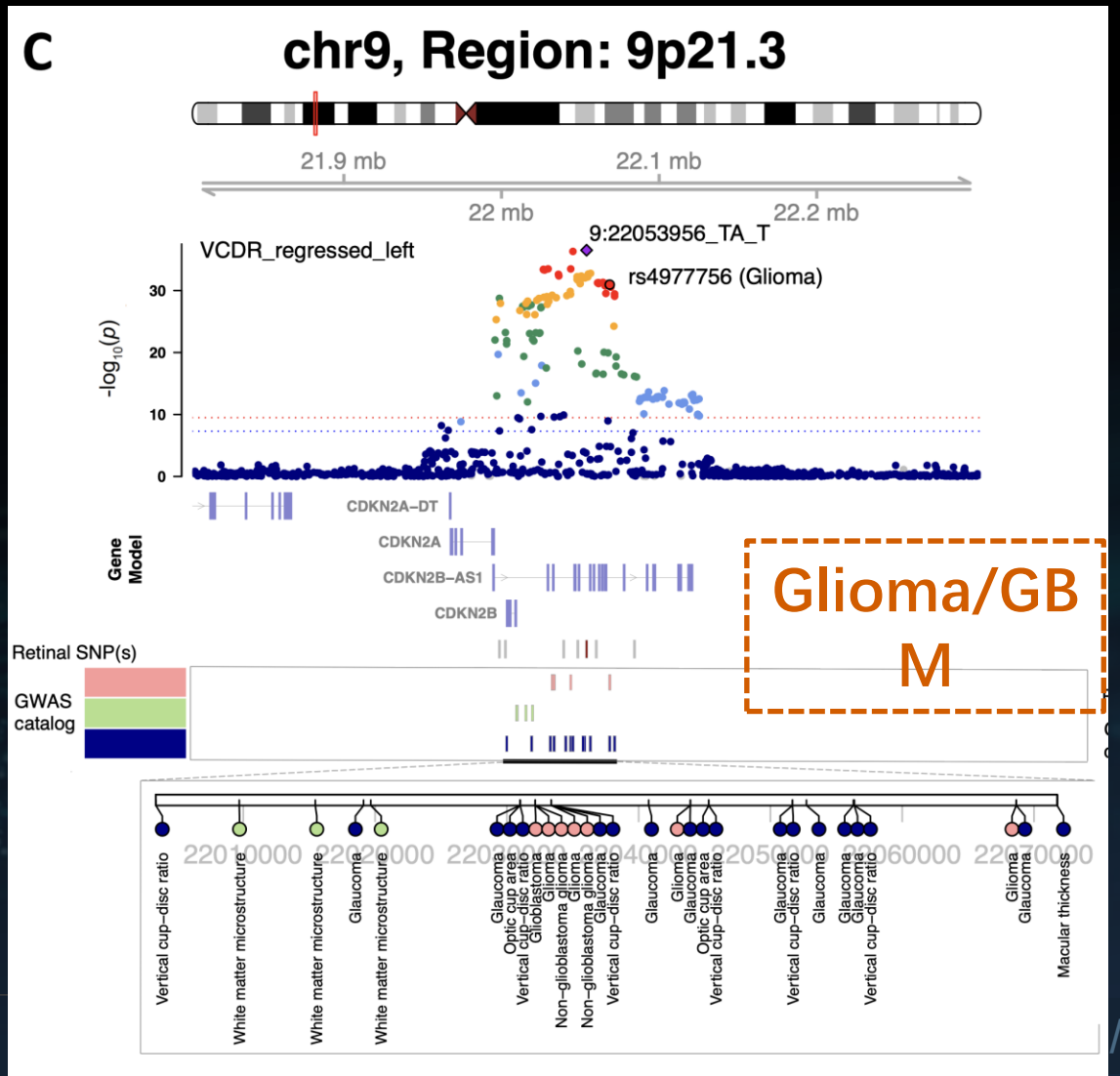



B
DTI parameters and GCIPL thickness (left)
Fornix stria terminalis (FA, β = 0.143)
Body of corpus callosum (RD, β = -0.066)
Superior corona radiata (AD, β = -0.058)
Posterior thalamic radiation (FA, β = 0.142)
Sagittal stratum (FA, β = 0.132)
Fornix (MO, β = 0.071)
-10
-log10(P-value) × β
30
C
DTI parameters and overall macular thickness (left)
Superior longitudinal fasciculus (RD, β = -0.056)
Cingulate gyrus cingulum (RD, β = -0.061)
Splenium of corpus callosum (RD, β = -0.059)
Fornix (MO, β = 0.054)
Fornix stria terminalis (FA, β = 0.121)
Sagittal stratum (FA, β = 0.104)
Posterior thalamic radiation (FA, β = 0.117)
-5
-log10(P-value) × β
15
D
Regional brain volumes and GCIPL thickness (right)
Cuneus (β = 0.057)
Pericalcarine (β = 0.048)
Lateral ventricle (β = -0.080)
Thalamus (β = 0.102)
Putamen (β = 0.063)
Ventral DC (β = 0.111)
-8
-log10(P-value) × β
16
E
Regional brain volumes and overall macular thickness (right)
Pericalcarine (β = 0.082)
Lateral ventricle (β = -0.074)
Putamen (β = 0.077)
Thalamus (β = 0.129)
Accumbens (β = 0.075)
Ventral DC (β = 0.111)
Brainstem (β = 0.0.71)
-5
-log10(P-value) × β
10
F
Cortical thickness and GCIPL thickness (left)
Precuneus (β = 0.073)
Superior frontal (β = 0.061)
Pericalcarine (β = 0.048)
Fusiform (β = 0.064)
Parahippocampal (β = 0.051)
Superior temporal (β = 0.056)
0
-log10(P-value) × β
10

Genetic Eye-Brain Connections

Examples of shared genetic influence between eye and brain disorders





Part II

Foundation Models for Medical Imaging Analysis

”Oddly, we are in a period where there has never been such a wealth of new statistical problems and sources of data. The danger is that if we define the boundaries of our field in terms of familiar tools and familiar problems, we will fail to grasp the new opportunities.”

- Leo Breiman -

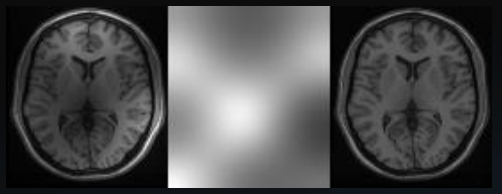
Image Processing Analysis Methods

How to enhance and extract signals of interest in imaging data?

Deconvolution

$$\frac{S(\mathbf{q})}{S_0} = \int P(\mathbf{r}, \Delta) e^{i\mathbf{q}.\mathbf{r}} d\mathbf{r}; \quad \mathbf{q} = \gamma \delta g \mathbf{u}$$

➢ Image Reconstruction



➢ Image Enhancement

❖ Image Segmentation

❖ Image Registration

Structural Learning

Engineering | Statistics | Machine Learning | Mathematics

Structural Learning

Image Segmentation

- Organ parcellation
- Localization of pathology
- Surgical planning
- Image-guided interventions
- Computer-aided diagnosis
- Quantification of organ change

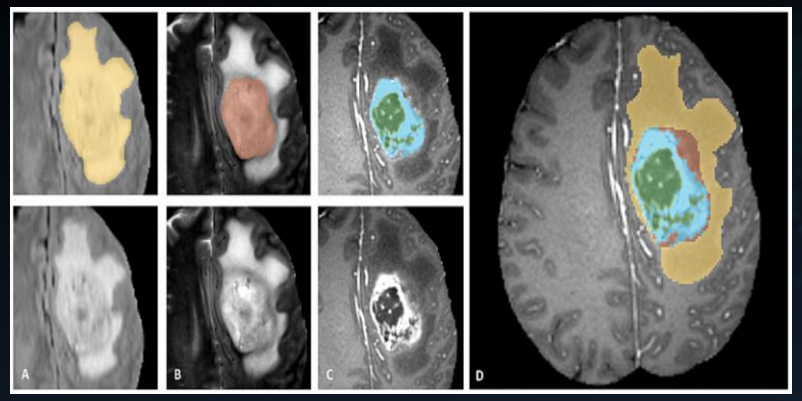
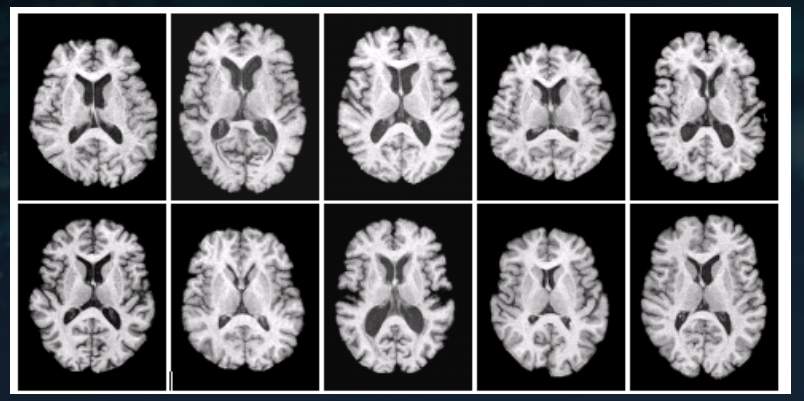


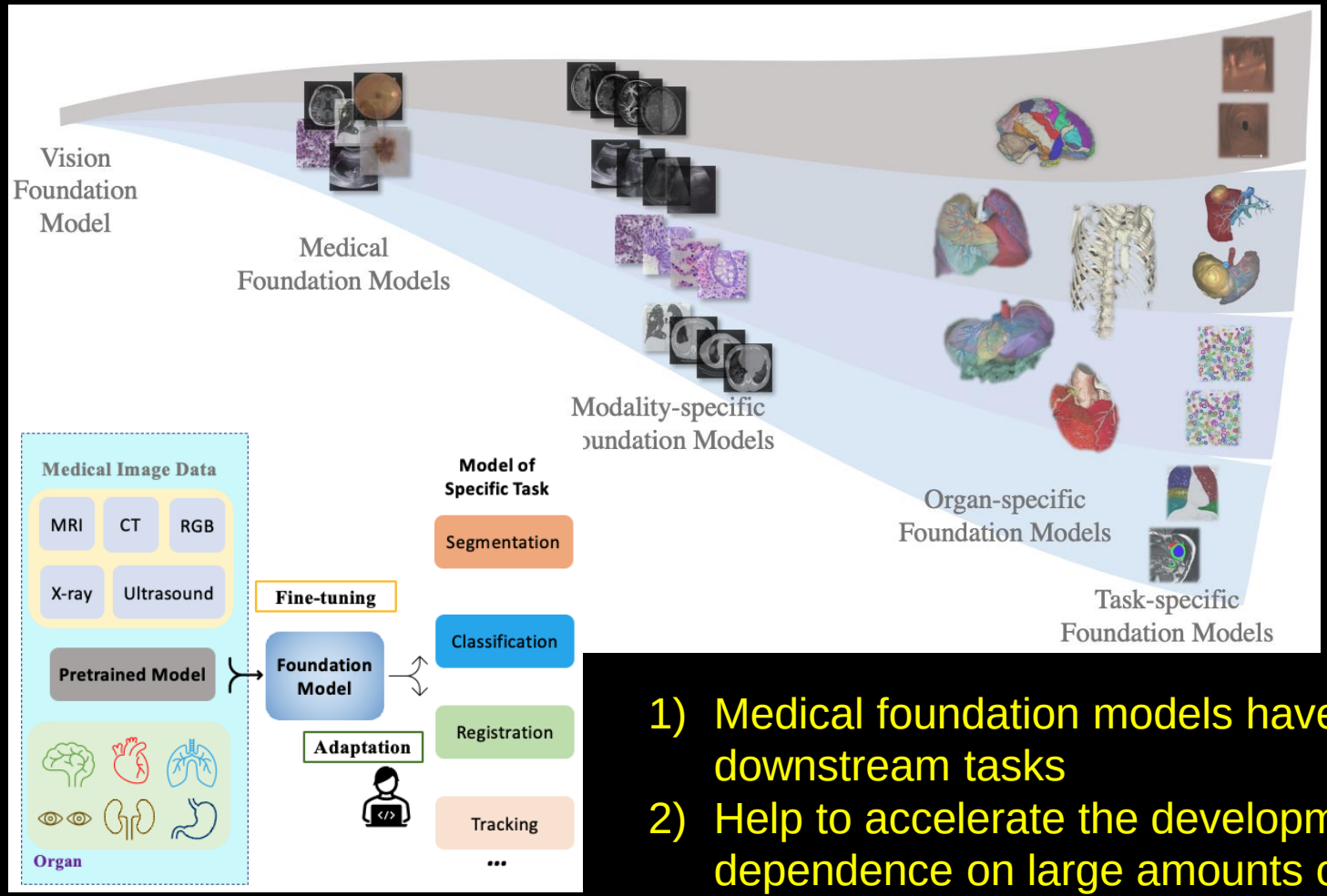
Image Registration

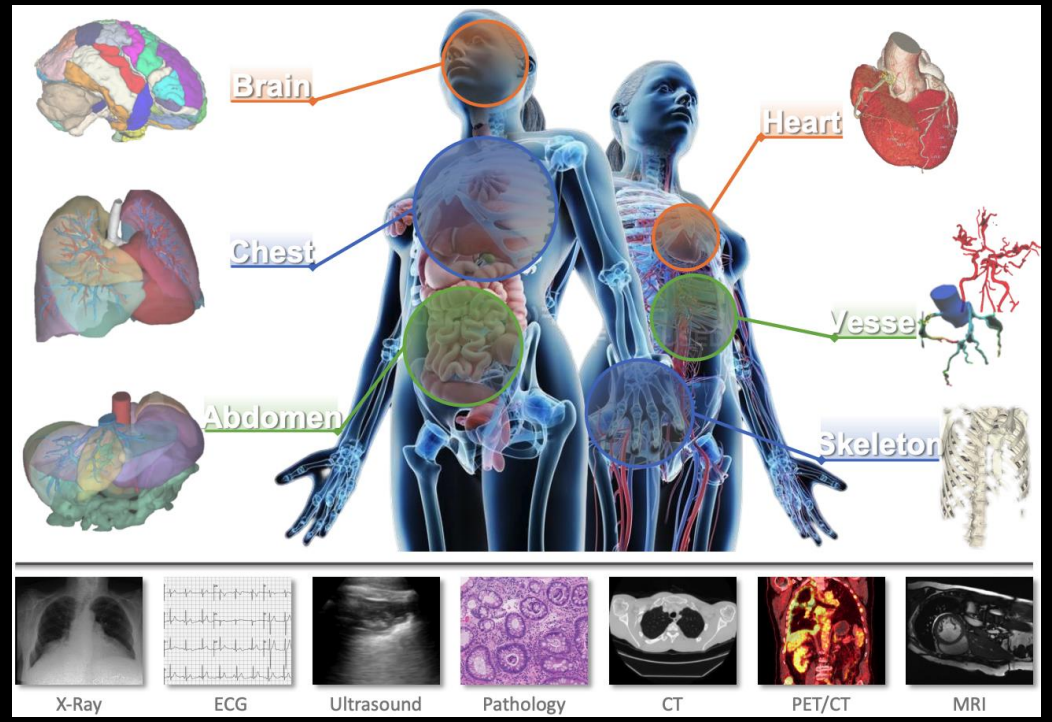
- Organ atlas
- Localization of pathology
- Automated image segmentation
- Multimodal fusion
- Population analysis
- Quantification of organ changes



Medical Image Foundation Model (MIFM)

Spectrum of foundation models in medical image analysis





1) Medical foundation models have immense potential in solving a wide range of downstream tasks
2) Help to accelerate the development of accurate and robust models, reduce the dependence on large amounts of labeled data

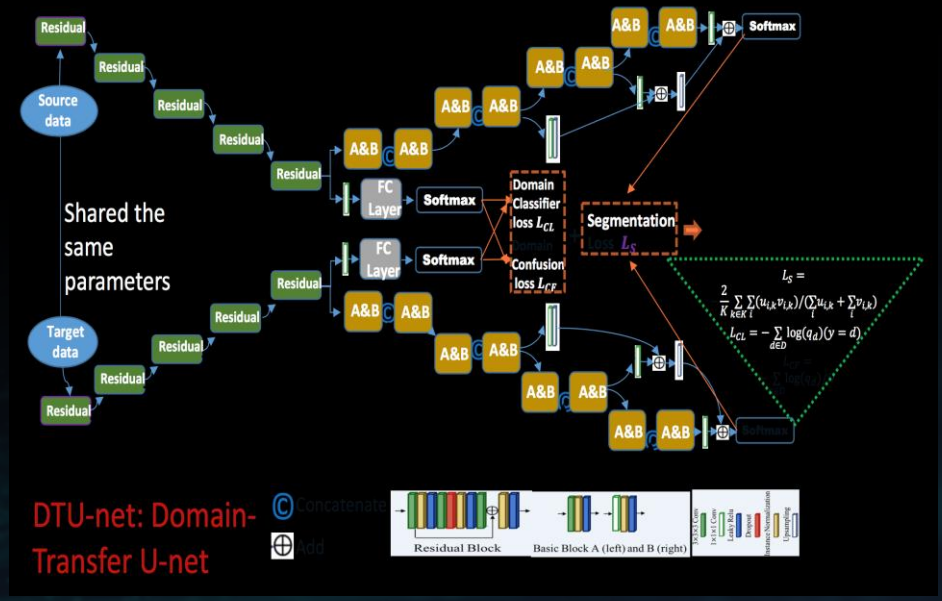
MORE THAN A JOURNEY

Shaoting Zhang and Dimitris Metaxas. On the challenges and perspectives of foundation models for medical image analysis. Medical Image Analysis, 2023.

AI for Image Segmentation

Segmentation Annotation

U-Nets

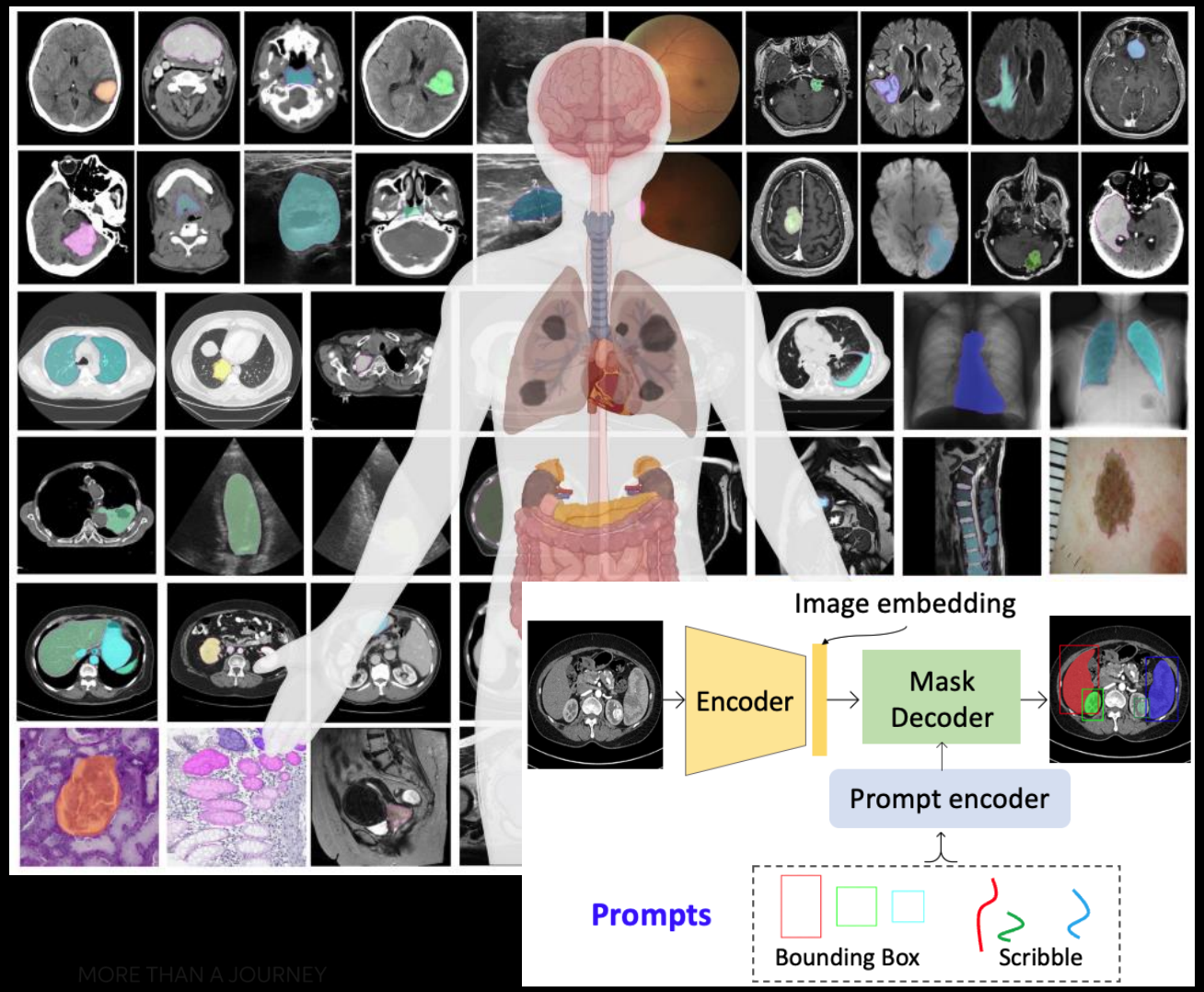


Liu, Q., Xu, Z., Bertasius, G., & Niethammer, M. (2023). SimpleClick: Interactive Image Segmentation with Simple Vision Transformers. ICCV., 22290-22300. 2023.

R. Azad *et al.*, “Medical Image Segmentation Review: The success of U-Net.” arXiv, Nov. 27, 2022.
Minaee, Shervin, et al. "Image segmentation using deep learning: A survey." *IEEE PAMI* 44.7 (2021): 3523-3542.

MIFM for Segmentation

MedSAM: Segment Anything in Medical Images



1) Developed on a large-scale medical image dataset with 1,570,263 image-mask pairs, covering 10 imaging modalities and over 30 cancer types.

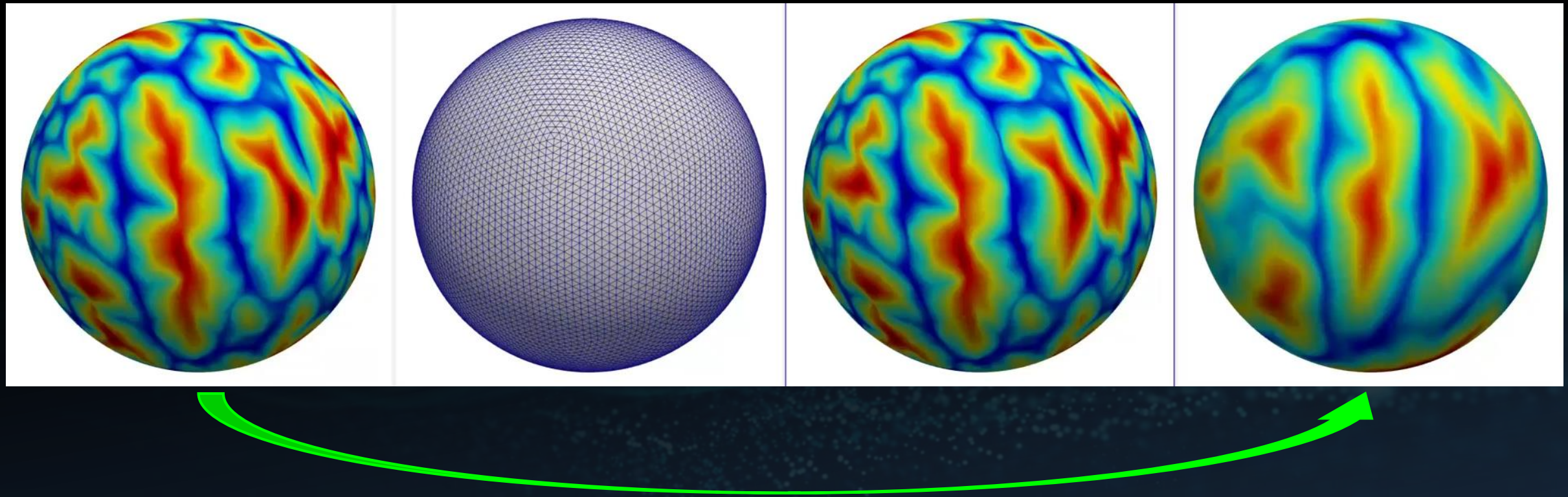
2) Evaluation on 86 internal validation tasks and 60 external validation tasks, demonstrating better accuracy and robustness than modality-wise specialist models.

3) Delivering accurate and efficient segmentation across a wide spectrum of tasks.

MORE THAN A JOURNEY

Jun Ma, …, Bo Wang. Segment Anything in Medical Images. Nature Communications, 2024.

Superfast Spherical Surface Registration

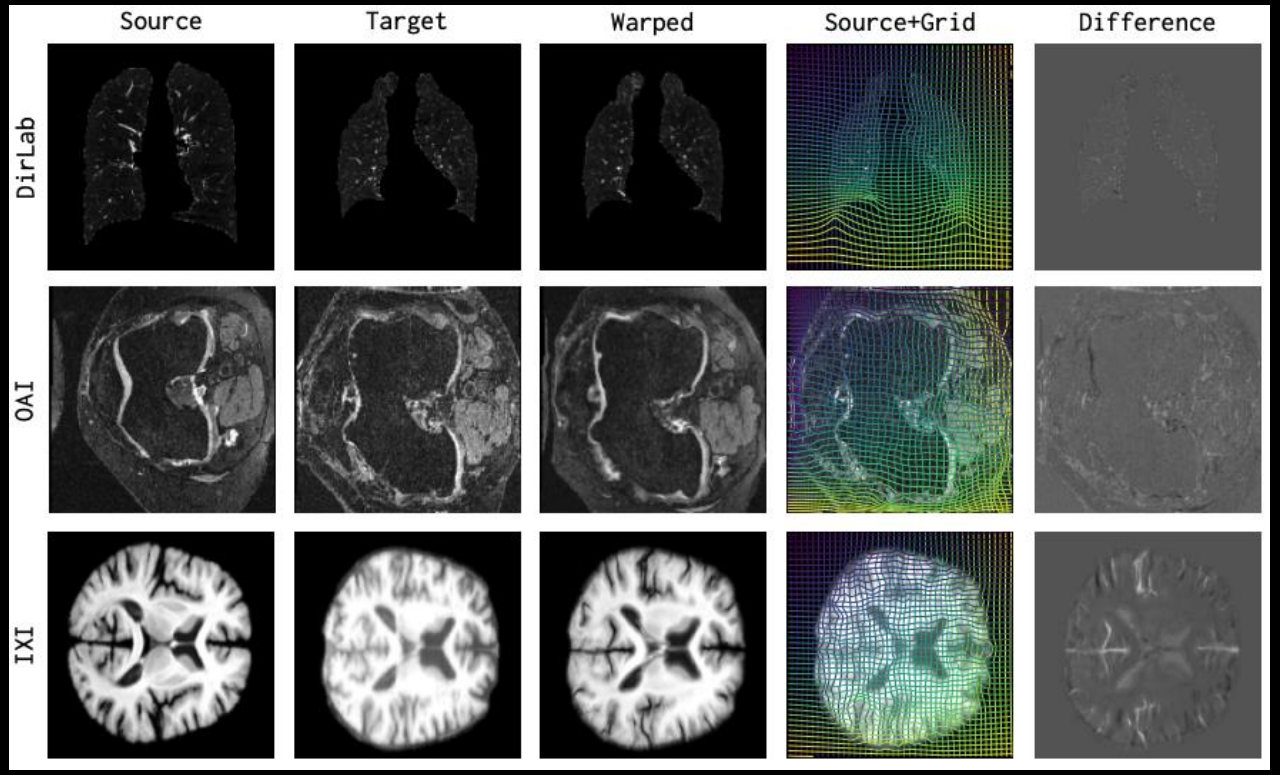


Zhao F, Wu Z, Wang F, Lin W, Xia S, Shen D, Wang L, Li G. S3Reg: Superfast Spherical Surface Registration Based on Deep Learning. IEEE Trans Med Imaging 2021; 40(8): 1964-1976.

MIFM for Registration

uniGradICON：A Foundation Model for Medical Image Registration

Example uniGradICON Registrations



1) Great performance across multiple datasets which is Not feasible for current learning-based registration methods

2) Zero-shot capabilities for new registration tasks suitable for different anatomical regions, and modalities

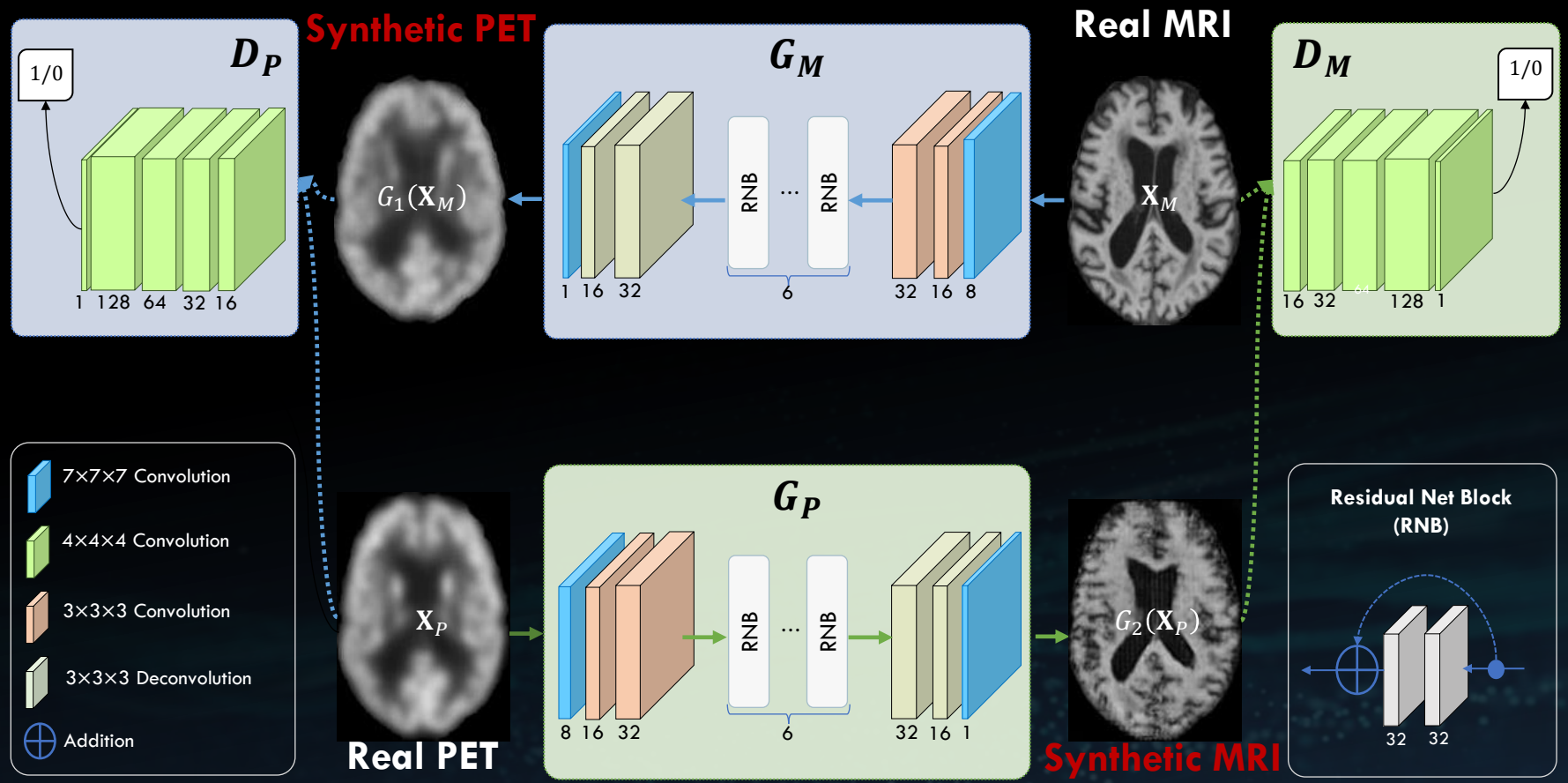
3) A strong initialization for finetuning on out-of-distribution registration tasks

MORE THAN A JOURNEY

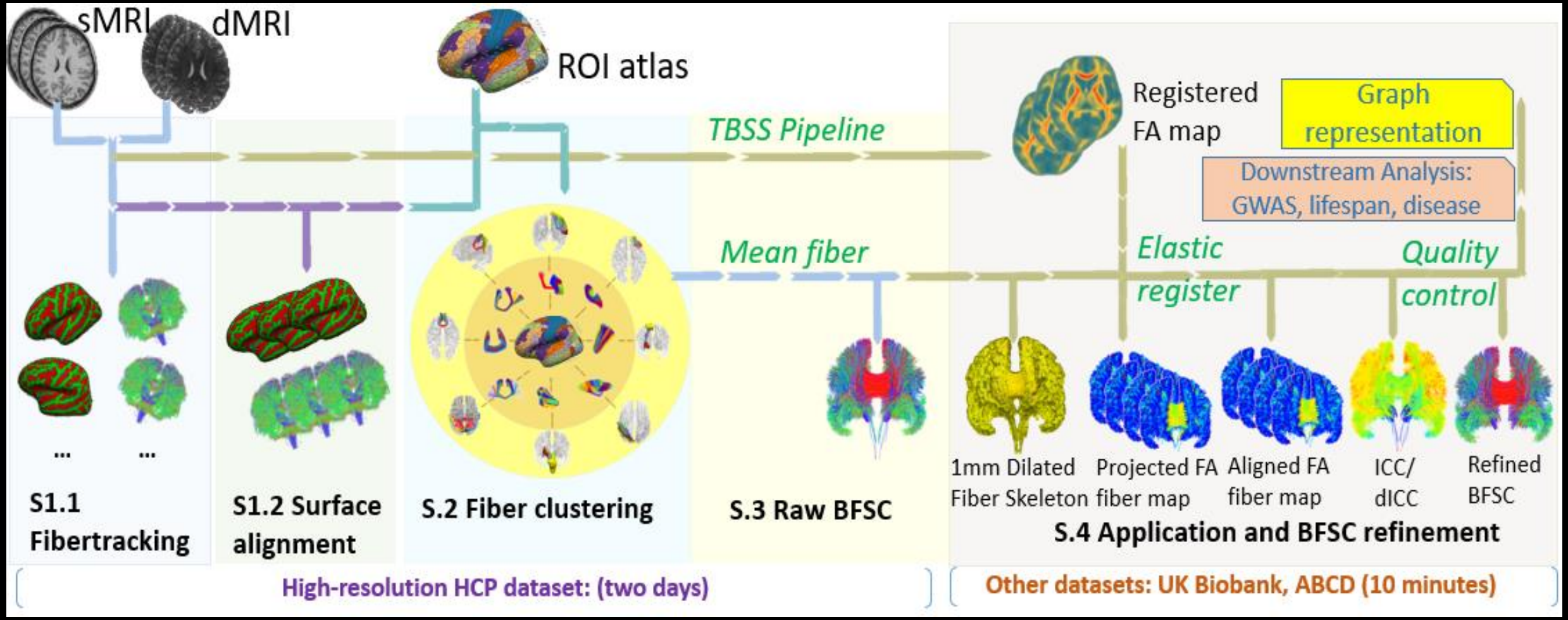
Lin Tian, …, Marc Niethammer. uniGradICON: A Foundation Model for Medical Image Registration. MICCAI, 2024.

滴滴

Cross-Modality Image Synthesis



CBSS: Connectome-based Spatial Statistics

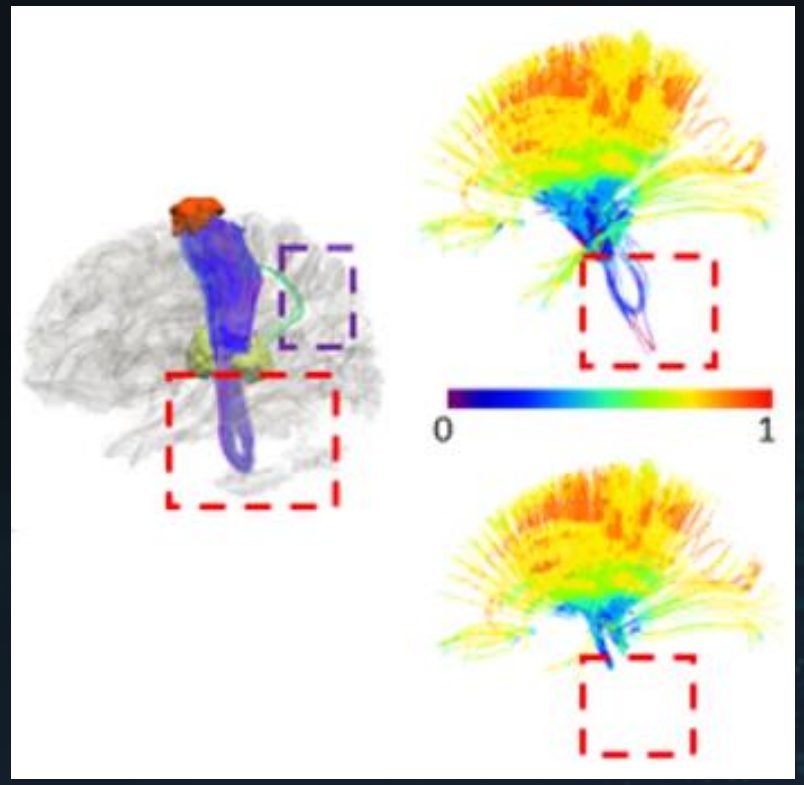


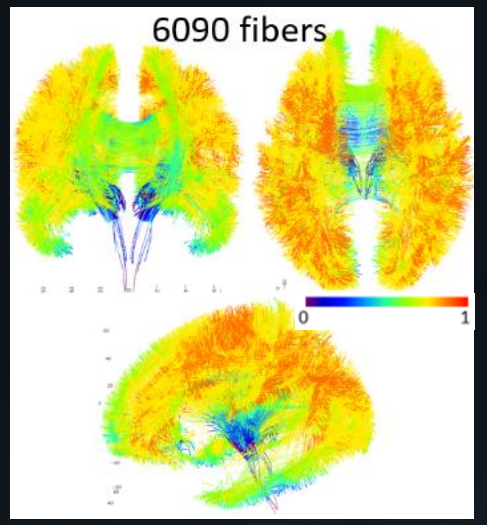
Stage1: BFSC

Stage2: CBSS

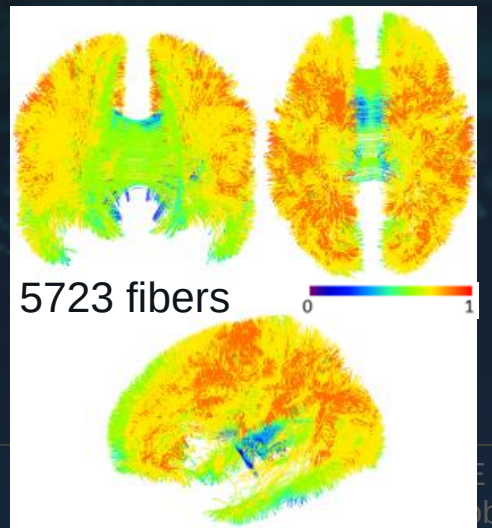
CBSS: Reproducibility and QC

ICCs of FA along fibers





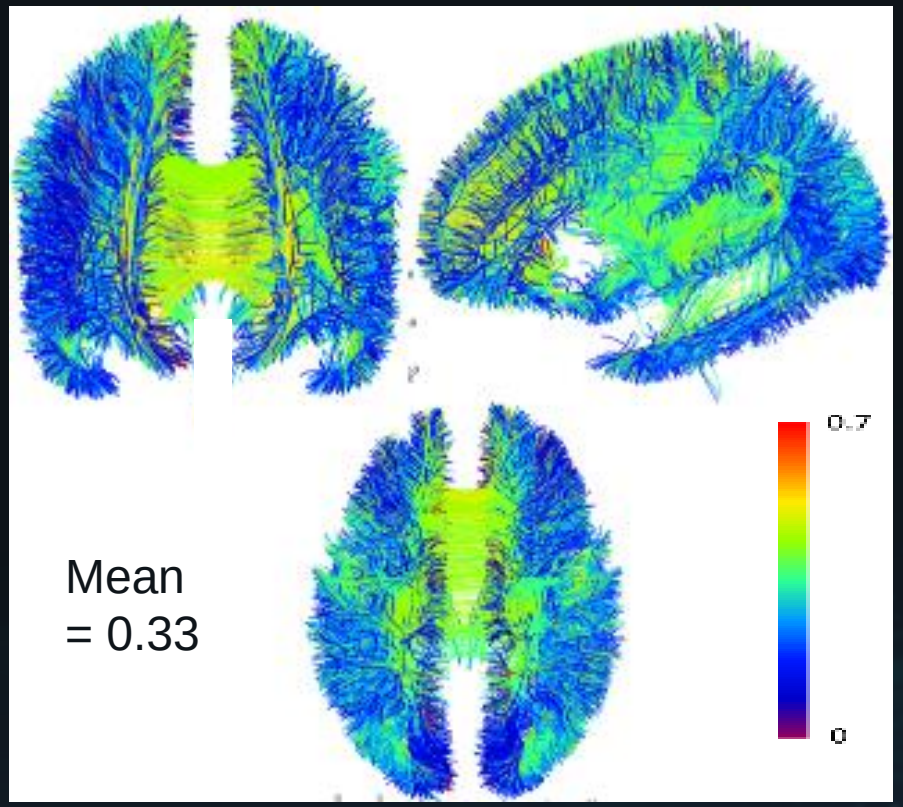
Mean ICC=0.74



Mean ICC=0.76

Downstream analysis: Heritability

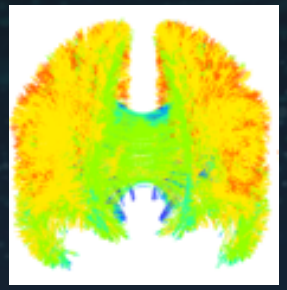
Our results



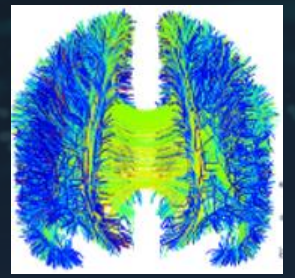
Literature

Phenotype	h^2	h^2 SE	Intercept	Intercept SE	Lambda	Mean chi^2	Ratio	Ratio_SE
FC-Visual	1.30%	2.00%	1.00	0.01	1.01	1.01	0.33	0.75
FC-Somatomotor	9.50%	2.20%	0.99	0.01	1.04	1.04	<0	NA
FC-Dorsal Attention	8.50%	2.20%	1.00	0.01	1.05	1.04	0.04	0.16
FC-Ventral Attention	7.30%	2.20%	1.00	0.01	1.03	1.04	0.04	0.17
FC-Limbic	6.30%	2.10%	0.99	0.01	1.02	1.02	<0	NA
FC-Frontoparietal	3.40%	1.90%	1.00	0.01	1.02	1.02	0.22	0.31
FC-Default	10.70%	2.20%	1.00	0.01	1.04	1.05	<0	NA
SC-Visual	12.90%	2.40%	1.02	0.01	1.07	1.08	0.20	0.09
SC-Somatomotor	14.30%	2.10%	1.01	0.01	1.07	1.07	0.11	0.09
SC-Dorsal Attention	14.90%	2.20%	1.00	0.01	1.06	1.07	<0	NA
SC-Ventral Attention	10.00%	2.20%	1.00	0.01	1.04	1.05	0.02	0.15
SC-Limbic	15.40%	2.10%	1.00	0.01	1.06	1.07	<0	NA
SC-Frontoparietal	13.50%	2.40%	0.99	0.01	1.06	1.06	<0	NA
SC-Default	14.10%	2.20%	1.00	0.01	1.06	1.06	<0	NA

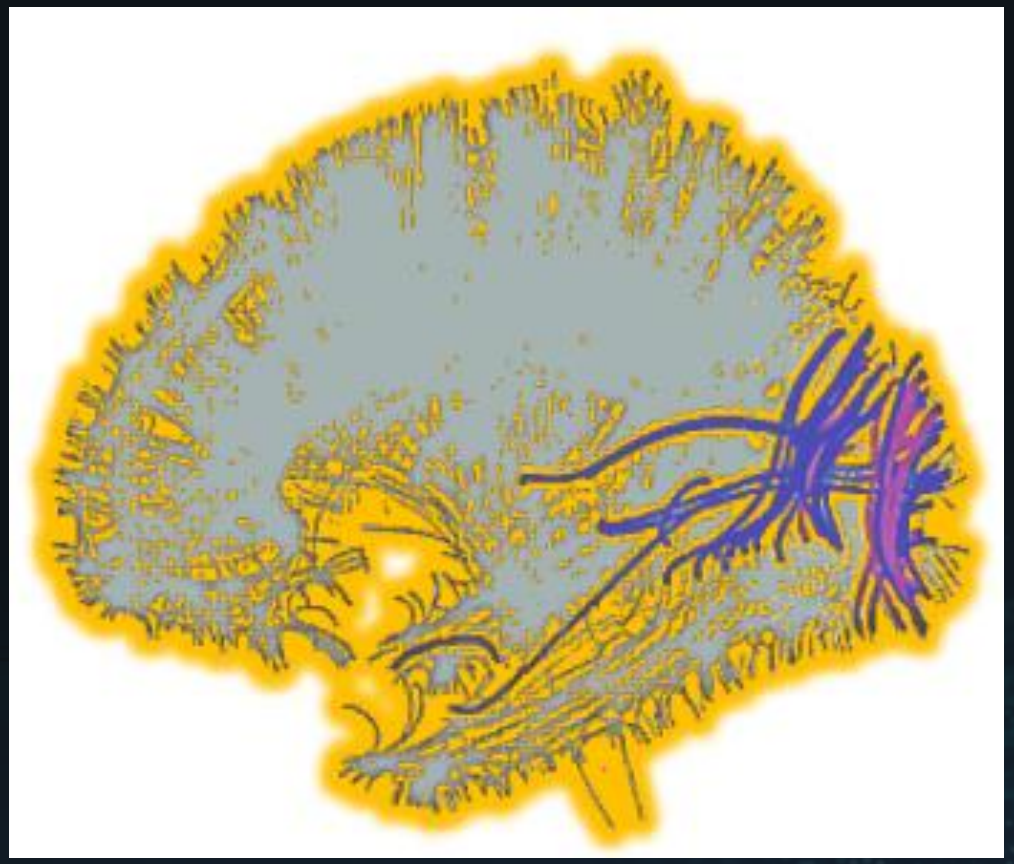
Reproducibility



Heritability



Downstream analysis: APOE Assoc.



- CBA (connectome-based analysis)

For each fiber extract the mean FA

- Structural connectivities (FA) of fibers within visual network and between visual and subcortical networks are negatively affected.

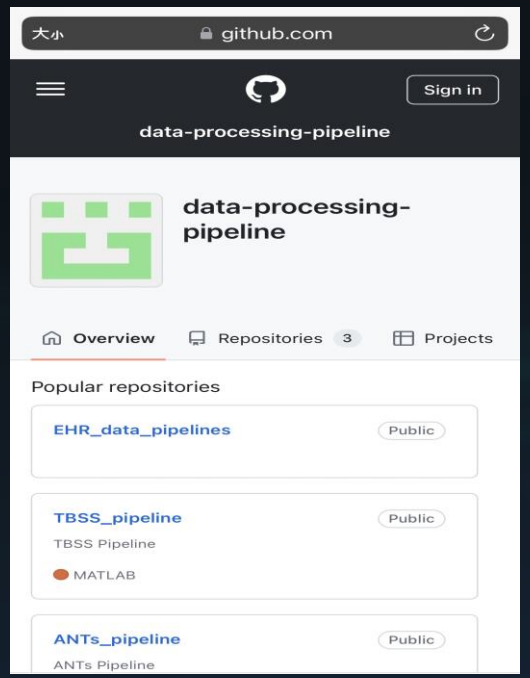
Linear models:

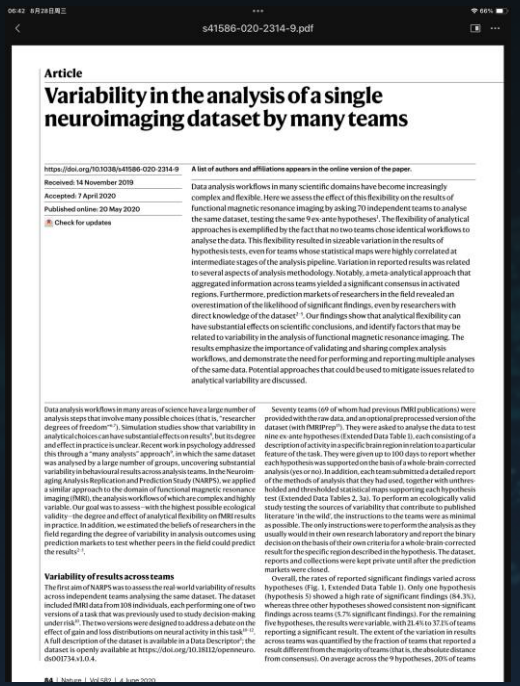
Fiber-Mean FA~ APOE4 + Age + Sex+ Age*Sex+ site + education + top 10 genetic PCs

Evaluating and Sharing Pipelines

More good data processing pipeline papers

https://github.com/data-processing-pipeline





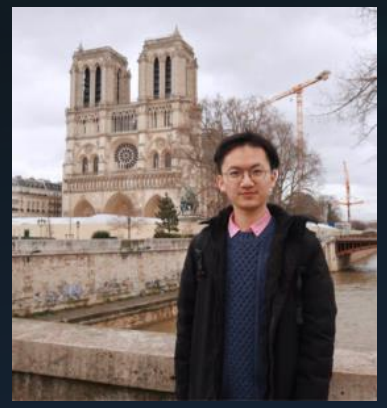
- Good datasets for comprehensive evaluation. There is no publicly available, high-quality imaging datasets with detailed annotation information that cover a large spectrum of segmentation tasks in health care.
- How to quantify the uncertainty and generalizarability of brain atlas as well as segmentation and registration tools?
- How to develop foundational models for various segmentation and registration tasks?

Part III

Highly Efficient Imaging Genetic Methods

"Statistical methods become truly impactful when they are not just about fitting models, but about driving discovery—by making complex imaging data interpretable, reproducible, and applicable in clinical practice.

- chatGPT-



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

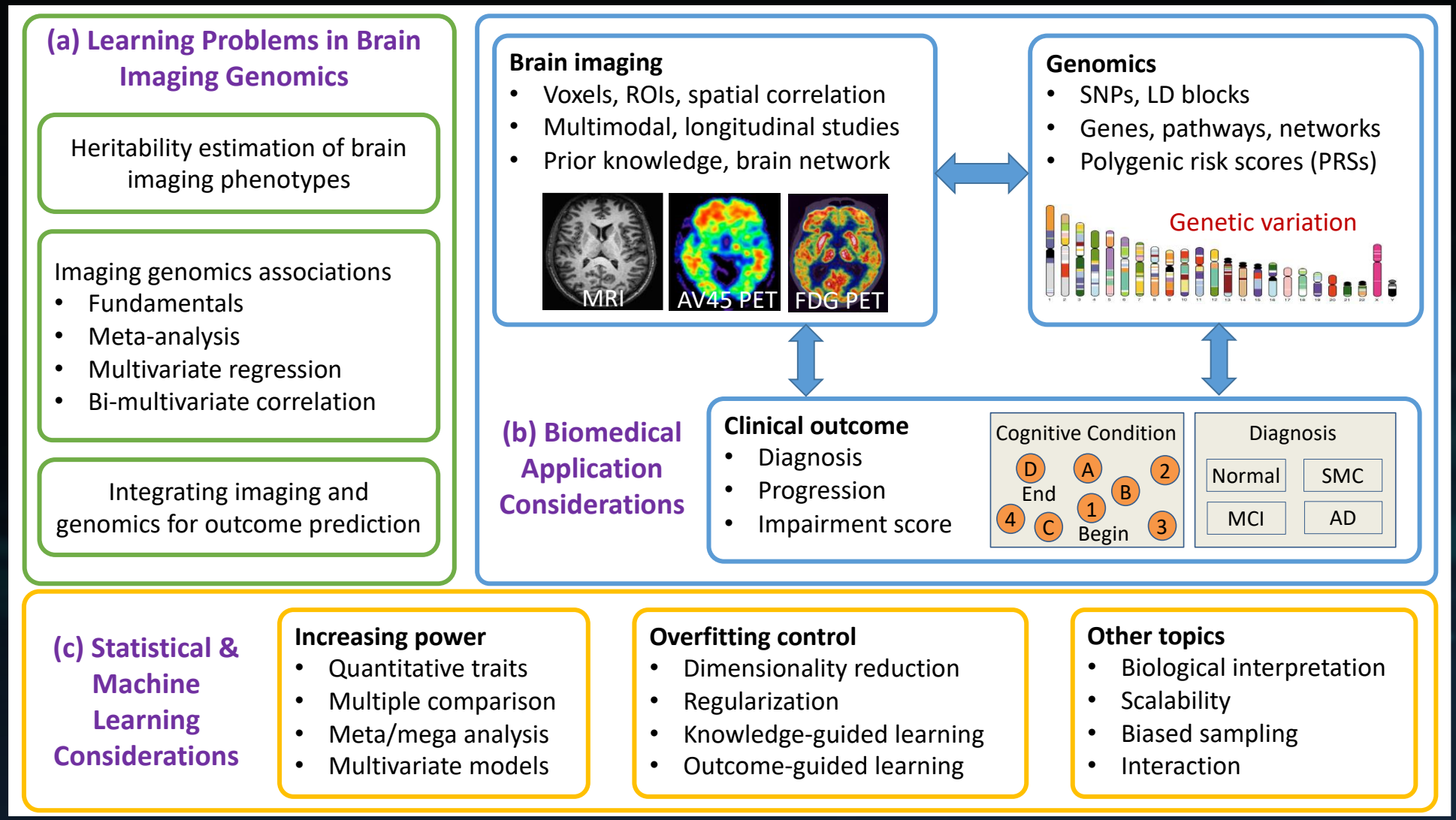
Integrate external genetics/genomics data (e.g., the GTEx, Hi-C chromatin interactions) to uncover new biological insights

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

Brain Imaging Genetics: Learning Problems



(Shen & Thompson Proc. of the IEEE 2020)

Imaging-derived phenotypes (IDPs)

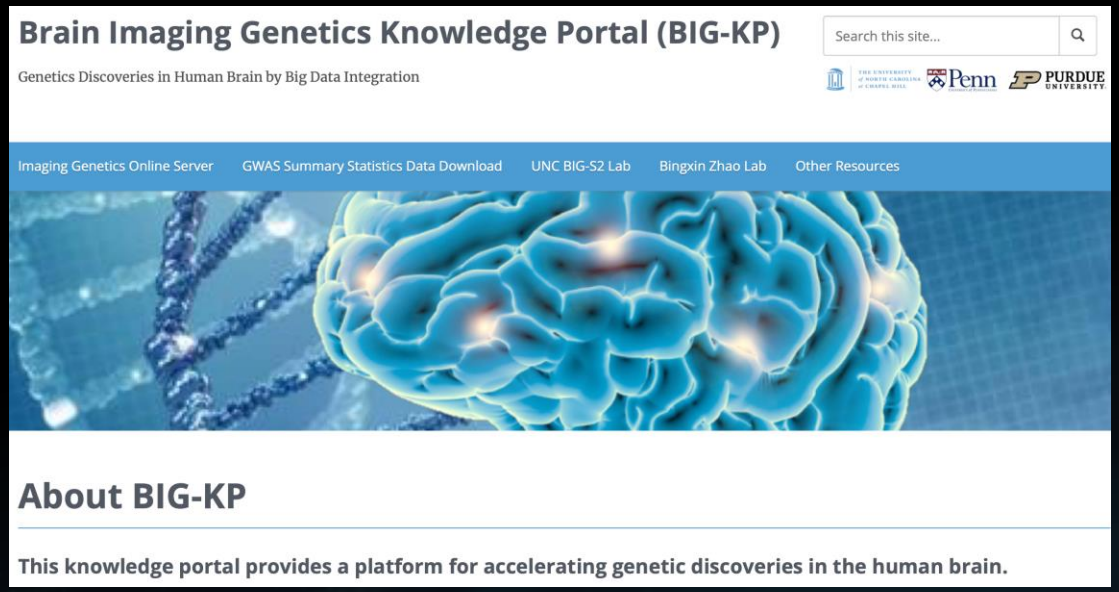
- Parcellating the human brain based on predefined atlases; aggregating voxels within each ROI (e.g., taking average)
  - Restricted by atlas
  - Removed local variations

- Conducting independent component analysis (ICA); extracting components as phenotypes

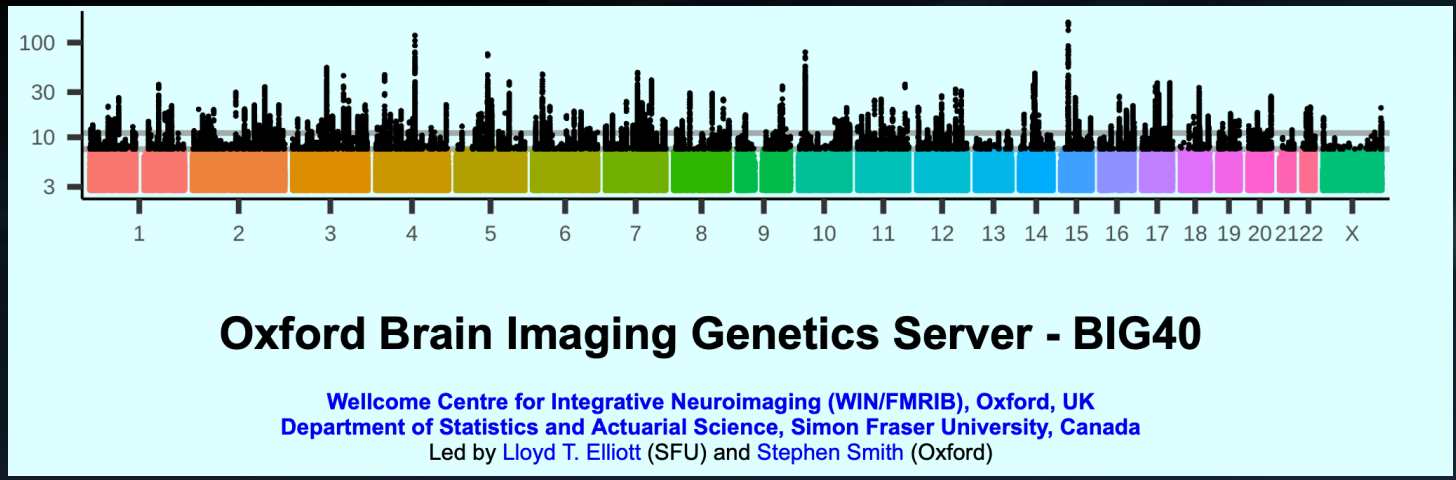
- Extracting features from deep learning models (e.g., Autoencoder)
  - Difficult to interpret the connection between the human brain and genetic markers

Missing localized genetic architecture in brain structure and function

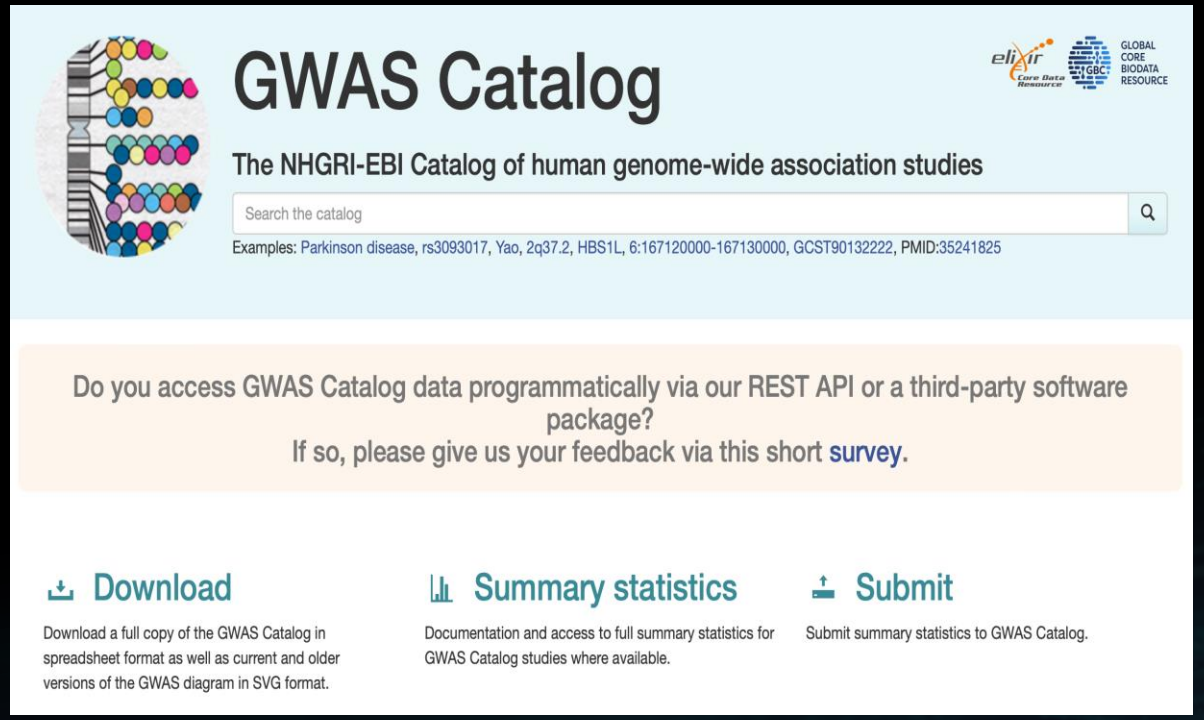
Neuroimaging Genetic Knowledge Portals

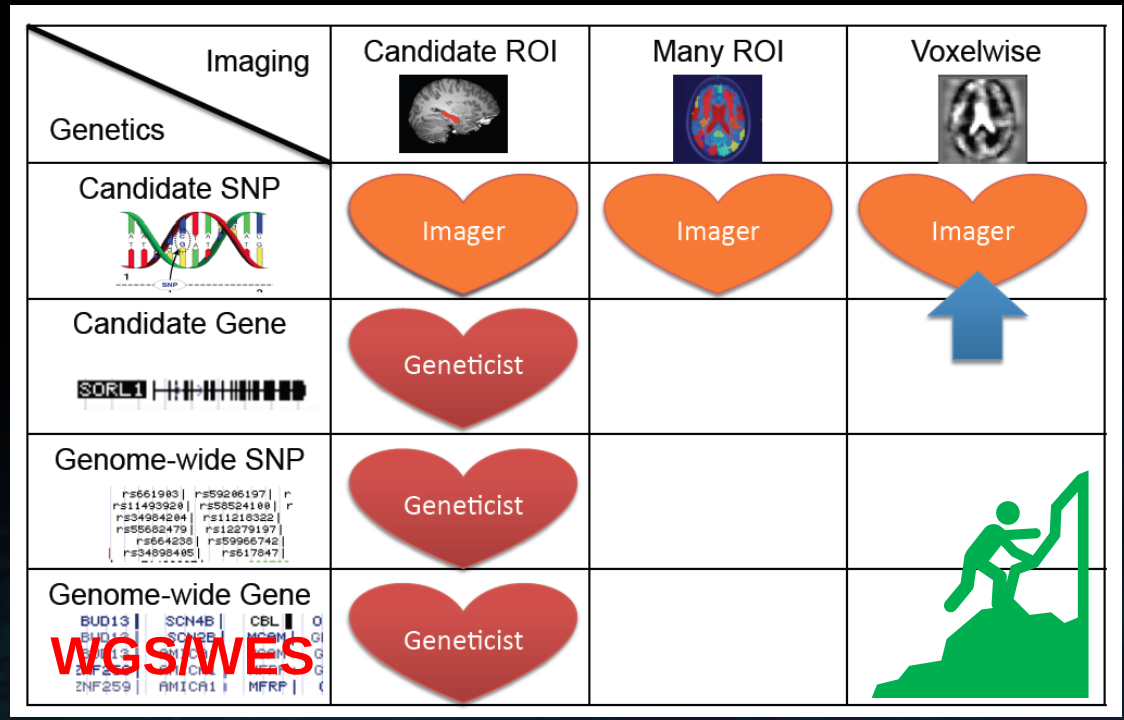


- 5000+ full
- summary statistics



Computational and Statistical Challenges





As of 2024-09-13

- 73,886 published studies
- 97,102 full summary statistics

10,000-1M

Resources Challenges

NEWS / **COMMUNITY**

Neuroimaging researchers pen statement protesting UK Biobank data-access changes

The signatories asked the organization to grant all imaging researchers a data-download exemption until the cloud platform can accommodate their processing needs.

BY **CALLI MCMURRAY**
30 JULY 2024 | 4 MIN READ

10000x50$=0.5M$; 0.1Mx50$=5M$; 1Mx50$=50M$

Challenges in Voxel-level GWAS

- **Computationally intractable**
  - High dimensionality of both imaging ($10^4 \sim 10^6$) and genetic data (10^8)
- **Difficult to share summary statistics**
  - The data volume can be 100 times larger than the current size of GWAS catalog
- **Multiple comparison in hypothesis testing across voxels**
  - Voxels are highly correlated
- **Noises in imaging data**
  - Reducing statistical power and reliability
  - Intensifying challenges in identifying genetic architectures underlying brain structures and functions

Current Solutions & Limitations

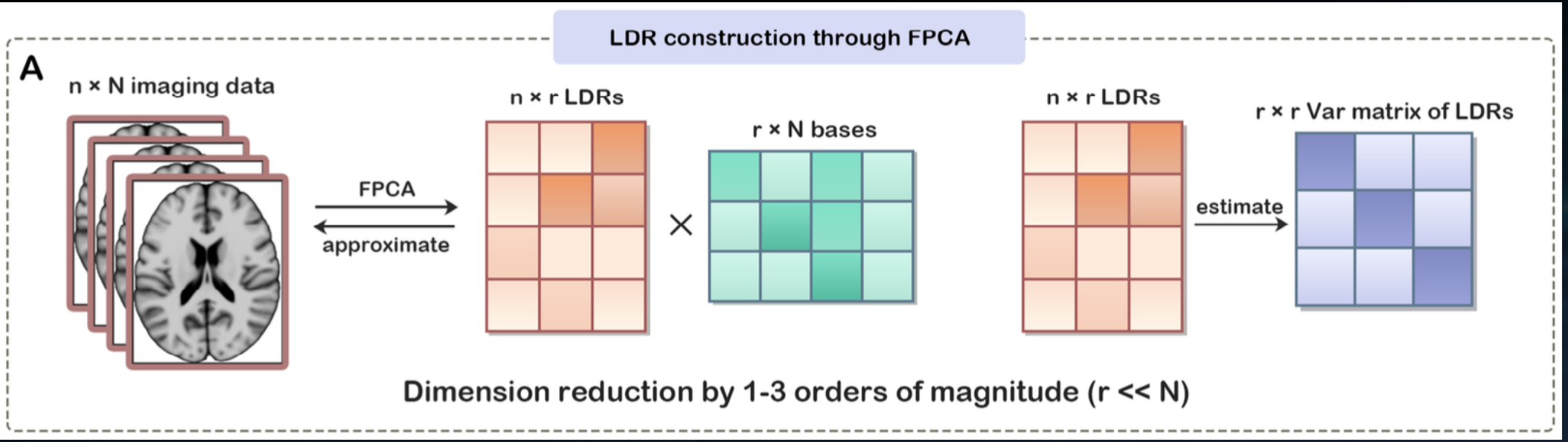
- Considering multiple genetic markers simultaneously
- Screening candidate markers
- Using rank-reduction or Bayesian techniques

Cannot generate whole-genome summary statistics needed for secondary analyses

- Voxel/local heritability
- Genetic correlation between voxels and with other phenotypes
- Colocalization analysis
- Mendelian randomization

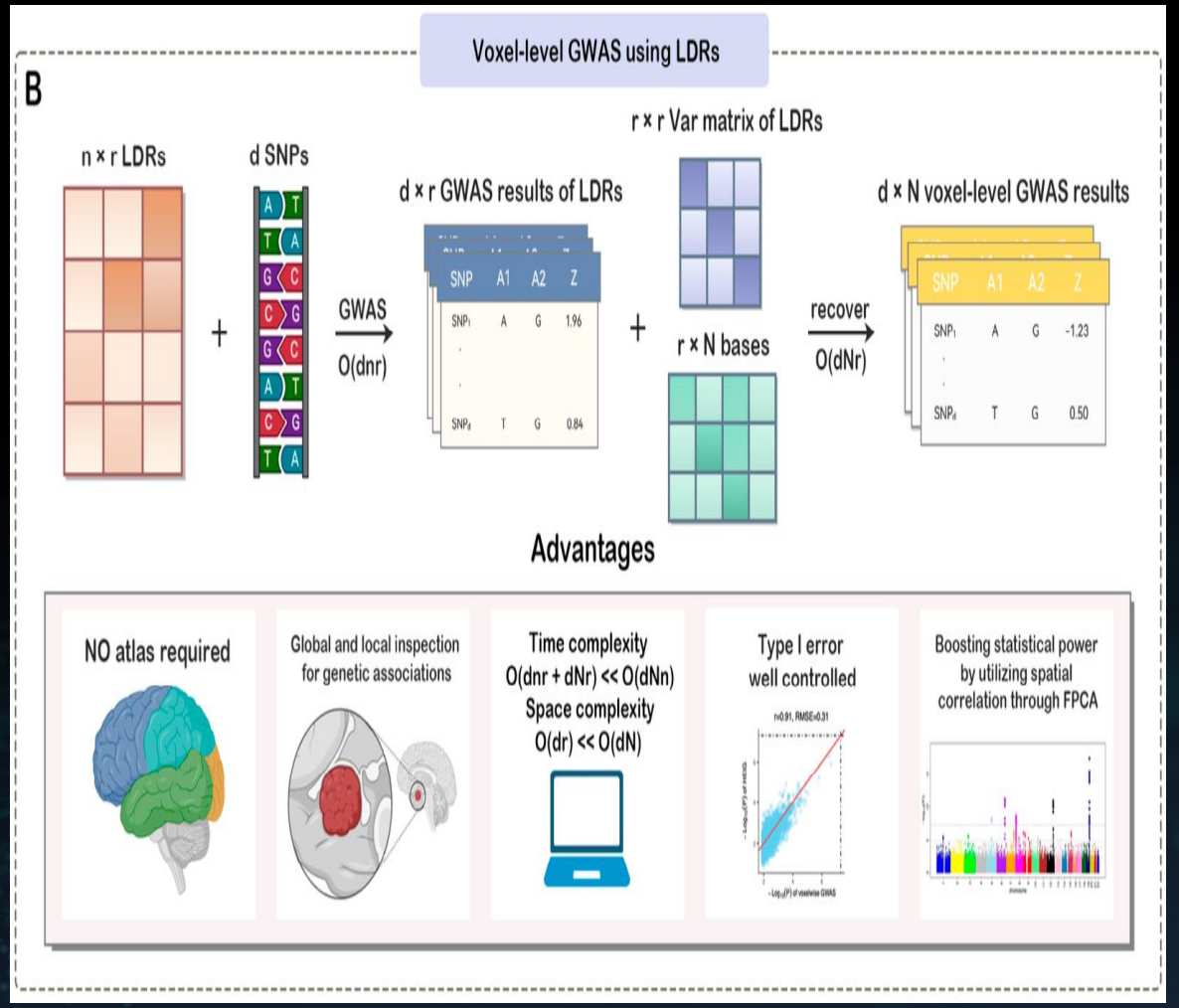
Highly efficient imaging genetics (HEIG)

- Using functional PCA (or other DR methods) to derive low-dimensional representations (LDRs)
- Dimension reduction by 2-3 orders of magnitude
- Removing white noise while preserving majority of imaging signals

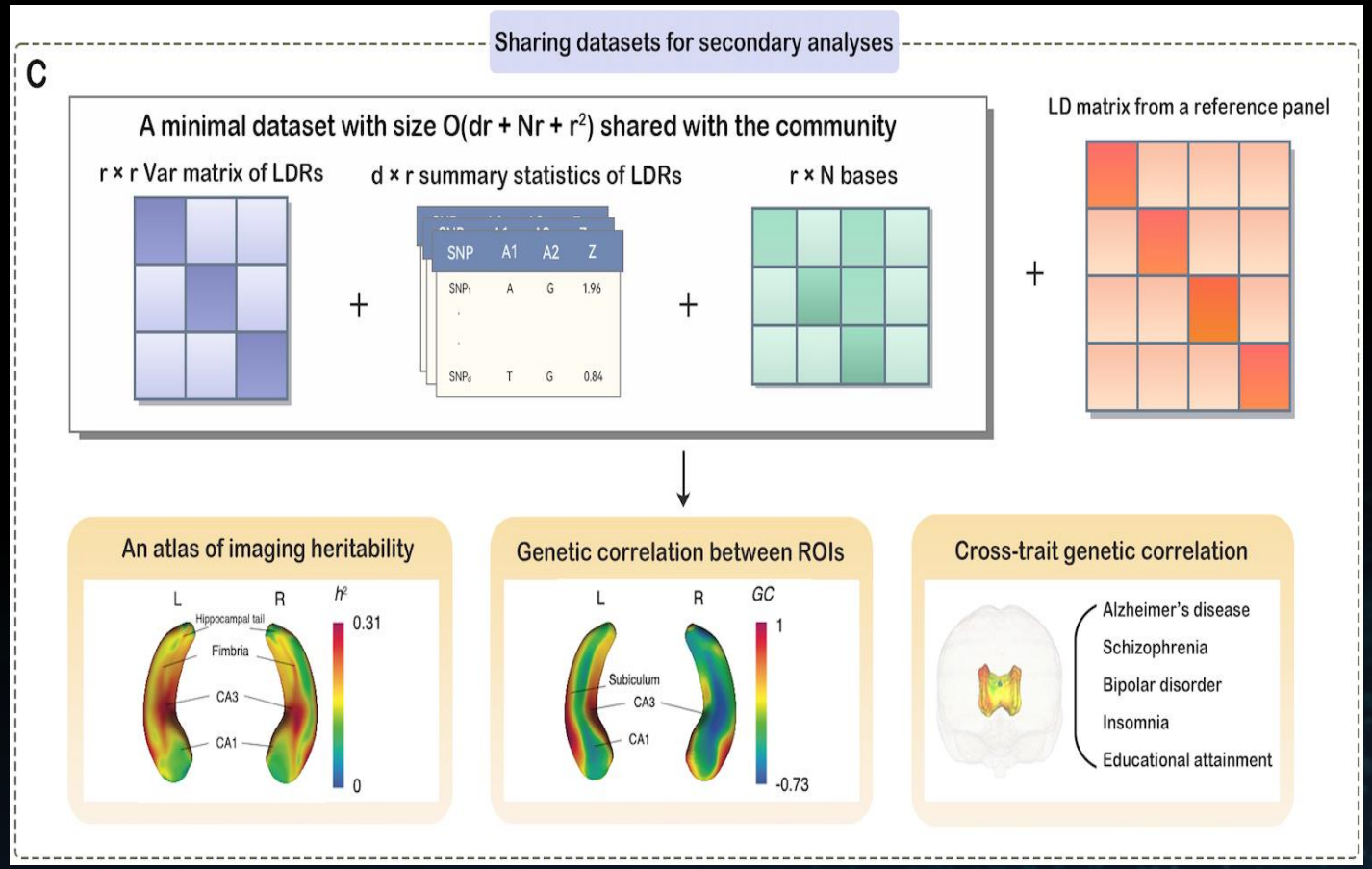


Highly efficient imaging genetics (HEIG)

- Conducting GWAS on LDRs and reconstructing voxel-level summary statistics
- Global and local inspection for genetic associations without atlas
- Reducing time and memory consumption by over 100 times
- Effectively protecting the type I error and boosting statistical power
- Using the effective number of independent voxels to adjust for multiple comparison



Highly efficient imaging genetics (HEIG)



- Sharing minimal datasets with the community for all secondary analyses
- Proposing a unified estimator to investigate voxel heritability, genetic correlation between voxels, and cross-trait genetic correlation between voxels and non-imaging phenotypes

Real data analysis

Input data

- UK Biobank 33,324 unrelated European subjects
- Hippocampus shape (15,000 vertices in each of left and right hemispheres)
- White matter (WM) microstructure (21 tracts, from 88 to 3,503 voxels, 32,217 in total)
- Surface curvature (the whole brain, 59,412 vertices)
- Genetic data: 7.8 million common variants

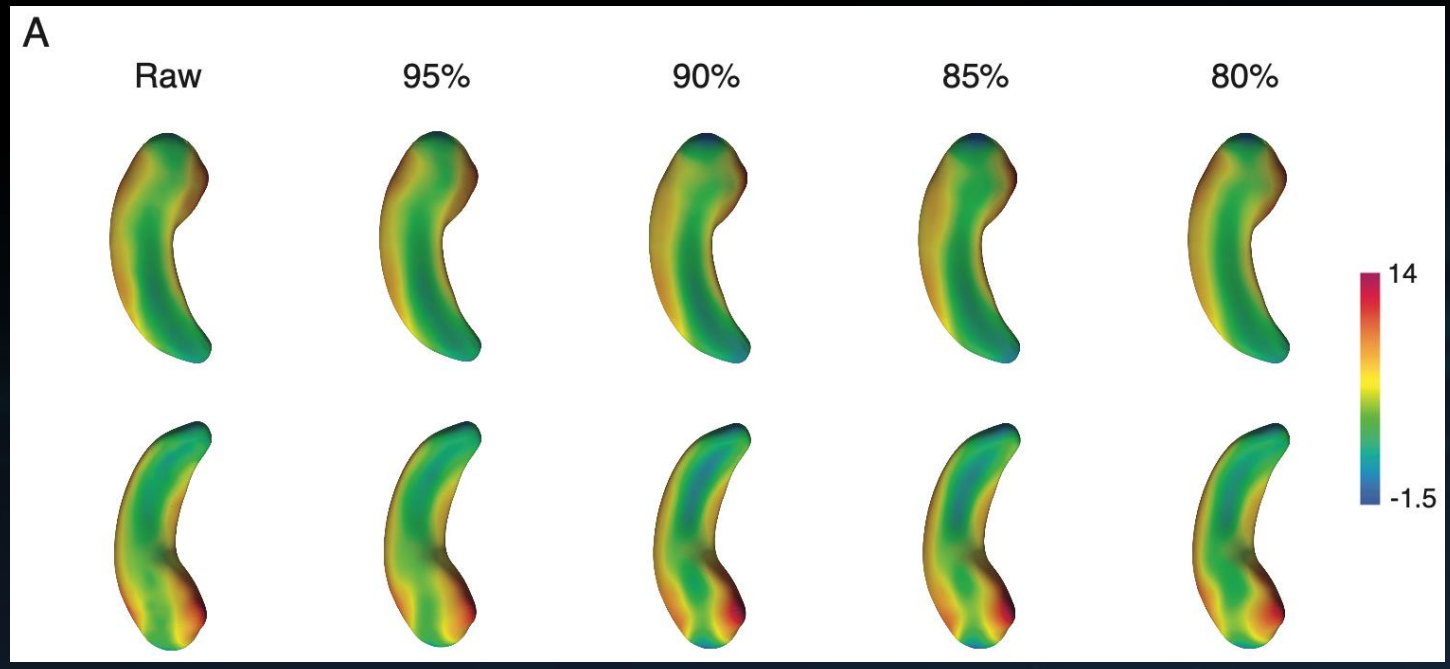
Constructing low-dimensional representations (LDRs)

- Hippocampus shape: 25 (0.13%) LDRs
- WM microstructure: 49 (4.1%) LDRs
- Surface curvature: 1,750 (2.9%) LDRs
- Using only 49, 1,034, and 1,750 LDRs to capture signals in 30,000, 32,217, and 59,412 voxels/vertices, respectively

Real data analysis

High correlation between the raw and reconstructed images using LDRs

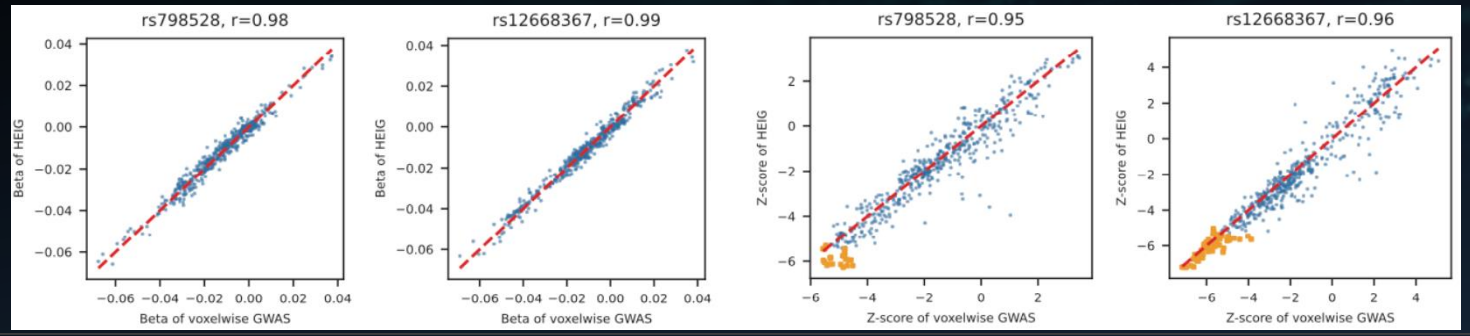
- Hippocampus shape: 0.93 with 25 (0.13%) LDRs
- WM microstructure: 0.86 with 49 (4.1%) LDRs
- Surface curvature: 0.85 with 1,750 (2.9%) LDRs

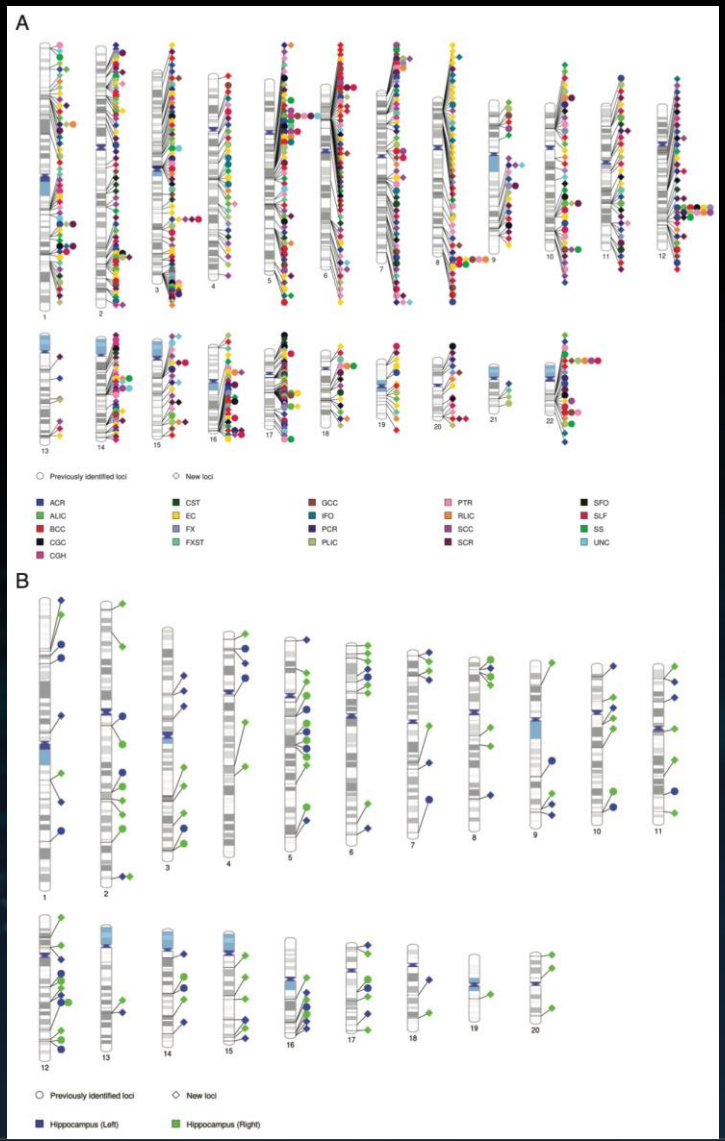


Real data analysis

Voxel-level GWAS – hippocampus and WM microstructure

- GWAS on 1,083 LDRs generated 95 GB of data, 133 times less than storing every voxel-variant pair (12,611 GB)
- Identified 815 significant loci, of which 94 and 540 are novel for hippocampus shape and WM microstructure, respectively, compared to IDP-based studies
- In a replication study (n=6,168), 39% of loci can be replicated
- Consistent with GWAS on voxels (raw data), but boosting statistical power for significant associations

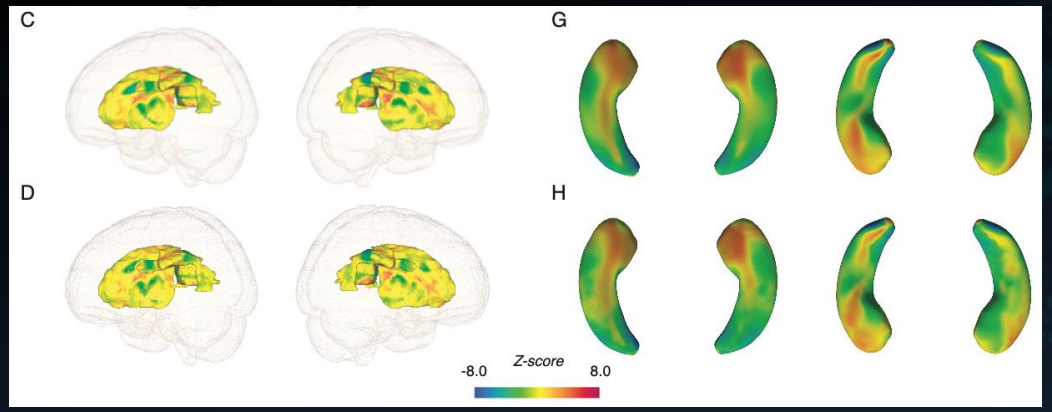


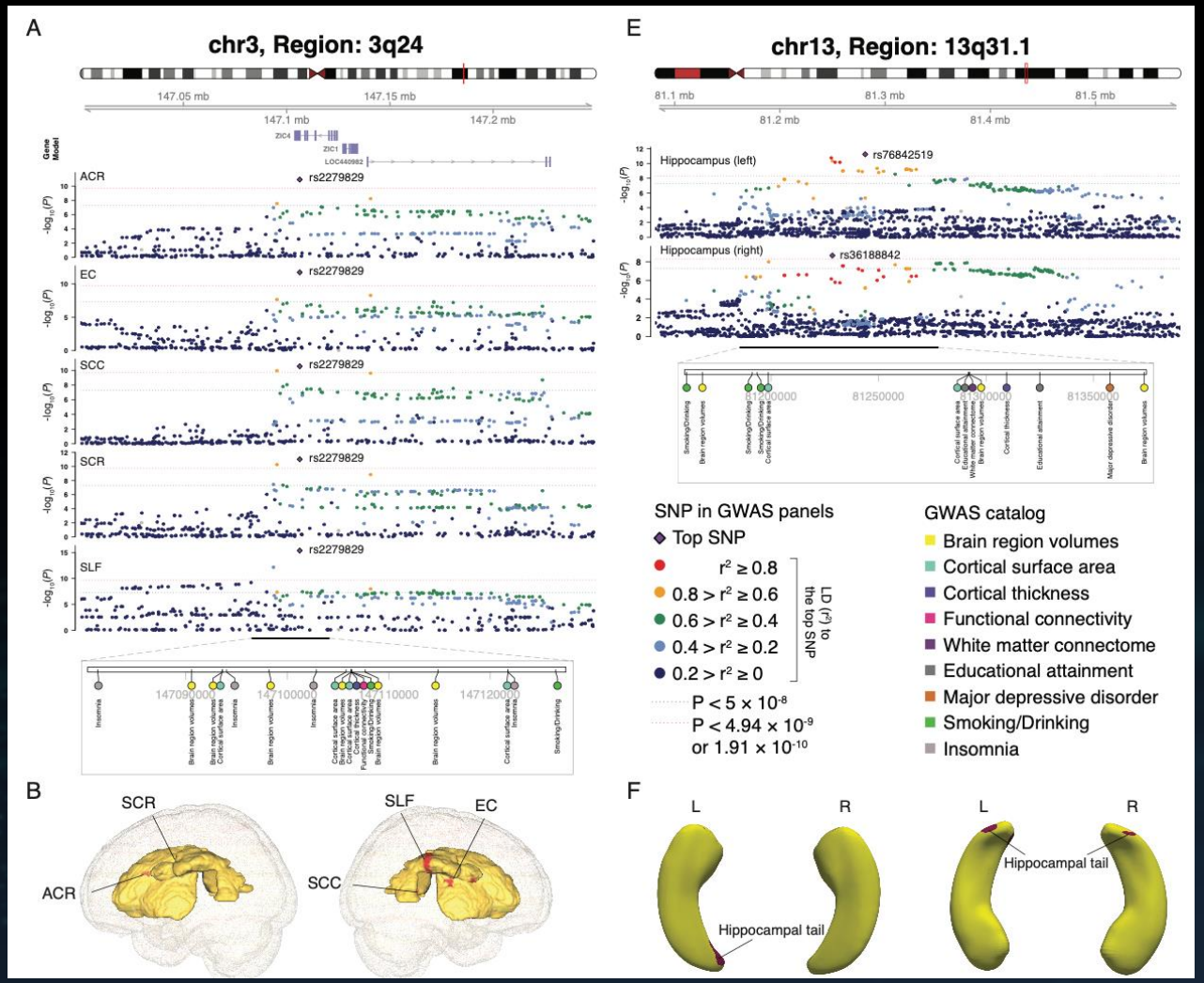


Real data analysis

Voxel-level GWAS

- Discovered novel loci and colocalizations with brain-related phenotypes including insomnia, major depressive disorder, and smoking/drinking
- Z-score correlation = 0.98 with GWAS directly on voxels

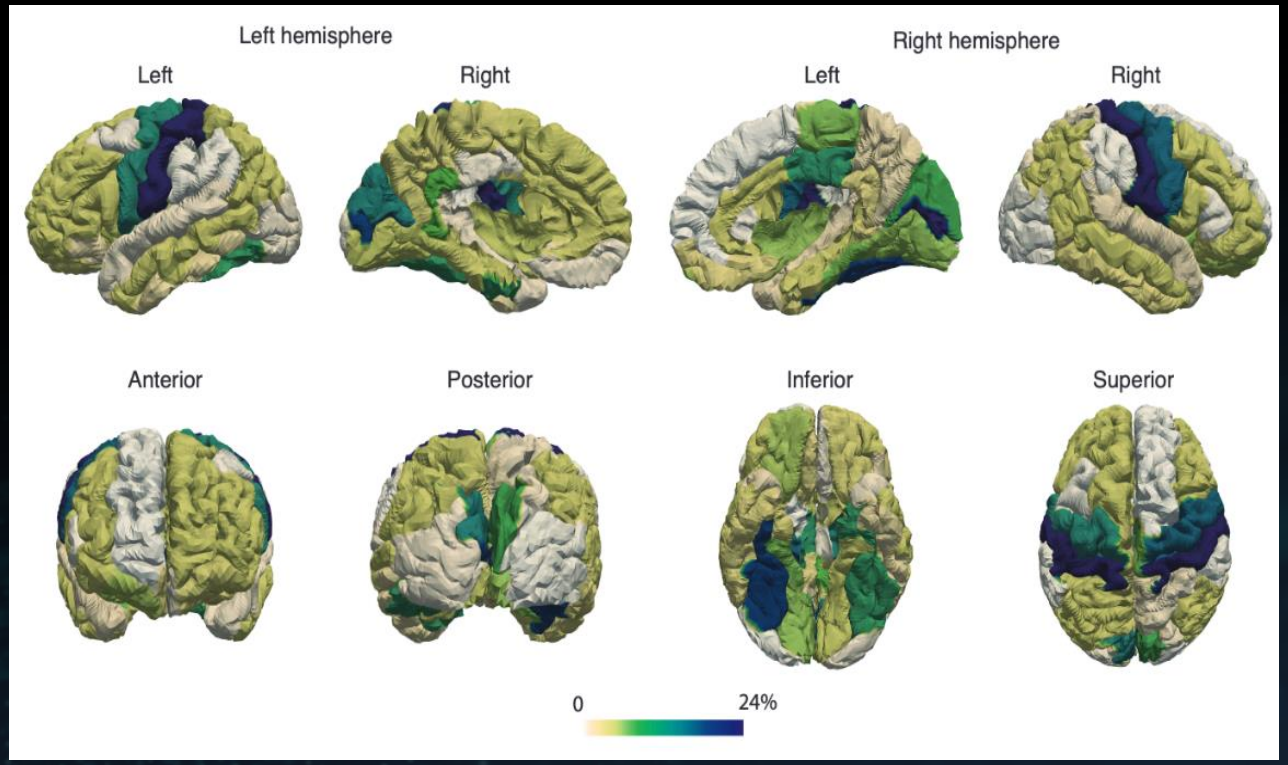




Real data analysis

Voxel-level GWAS – surface curvature

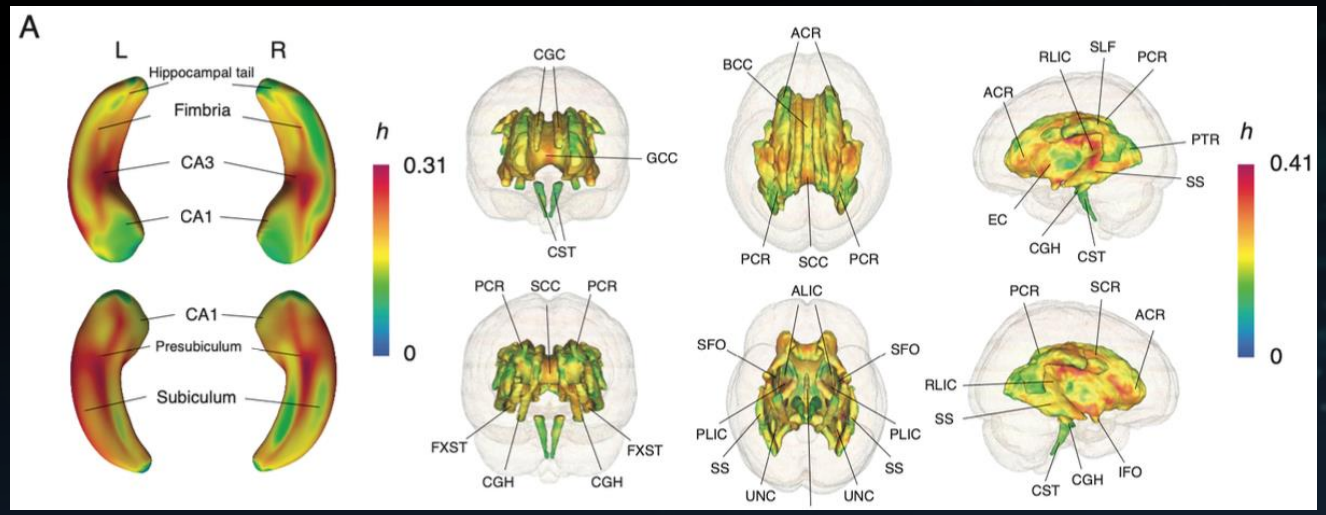
- GWAS on 1,750 LDRs generated 87 GB of data to share, 138 times less than storing every vertex-variant pair (12,042 GB)
- Identified 1,431 significant loci, of which 912 (63.7%) were replicated (n=12,431)
- Postcentral and precentral regions were enriched with significant vertices associated with at least one SNP

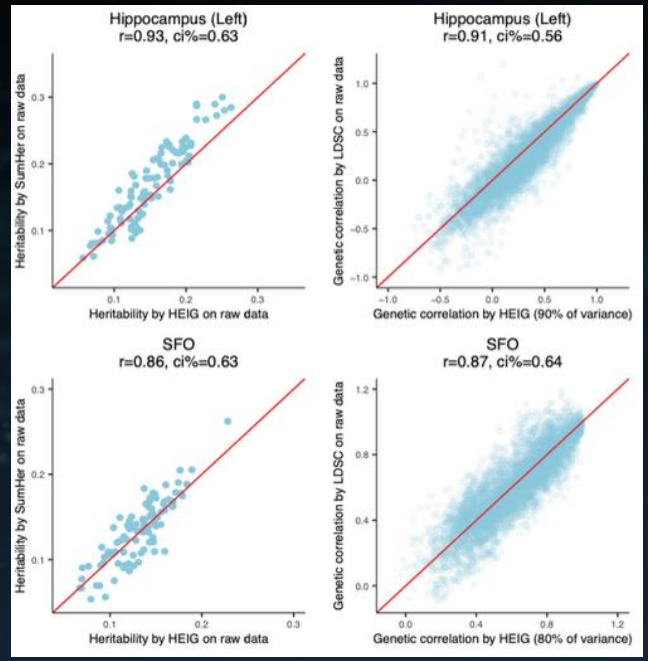


Real data analysis

Voxel heritability and genetic correlations

- Constructed an atlas of heritability using LDR summary statistics and an LD matrix
- Identified shared genetic bases between subregions
- Highly consistent with SumHer in heritability estimates and with LDSC in genetic correlation estimates, despite distinct methodology

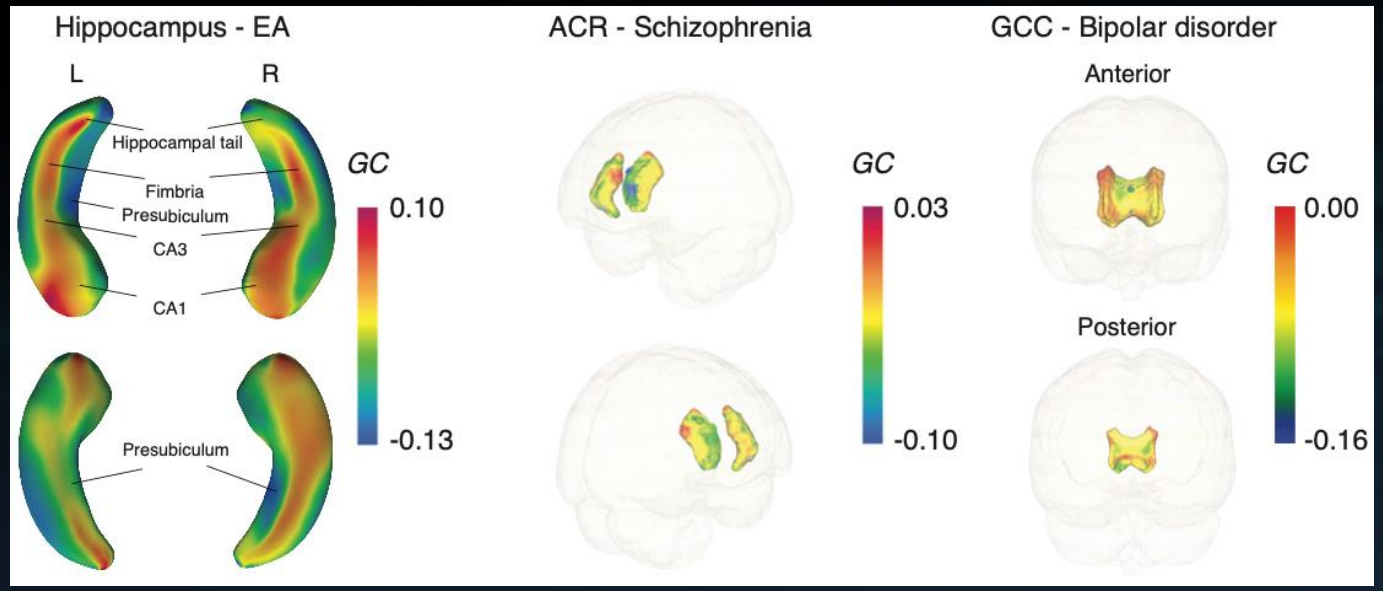




Real data analysis

Cross-trait genetic correlations with 14 brain-related phenotypes

- Hippocampus (presubiculum) with educational attainment (-13%, P = 0.001)
- Right anterior corona radiata with schizophrenia (-9%, P = 4.3×10^{-5})
- Genu of the corpus callosum with bipolar disorder (-16% ~ 0%, se = 2.6%)



Computational efficiency

Comparing with REGENIE using 193 voxels, 33,324 subjects, and 7.8 million SNPs

- HEIG took 13 CPU hours and generated 4.1 GB of data for sharing
- REGENIE took 107 CPU hours and generated 127 GB of data of all variant-voxel pairs

Benchmarking study using the surface area data of the whole brain with 59,412 vertices, 15,752 subjects, and 7.8 million SNPs

- HEIG took 2.8 hours to construct 1750 LDRs (4 CPUs in parallel)
- GWAS for 1750 LDRs using REGENIE took 56 hours (16 CPUs in parallel)
- Voxel-level GWAS using HEIG took 32 hours (4 CPUs in parallel)
- Generated 87 GB of data for sharing
- HEIG took 20 minutes for estimating 59,412 vertex heritability and genetic correlation

Contributions

- Reducing time and storage consumption by 2-3 orders of magnitude
- Efficiently generating and sharing summary statistics for voxel-level analysis
- Enhancing statistical power and genetic influence by removing white noise
- Investigating genetic architectures underlying both global and local patterns in the human brain without predefined atlases
- Straightforward extension to other organs and high-dimensional datasets with strong correlation

Acknowledgement

Funding: U.S. NIH Grants MH116527, R01AG082938-01.U01AG079847, 1R01AG085581 and R01AR082684.

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.