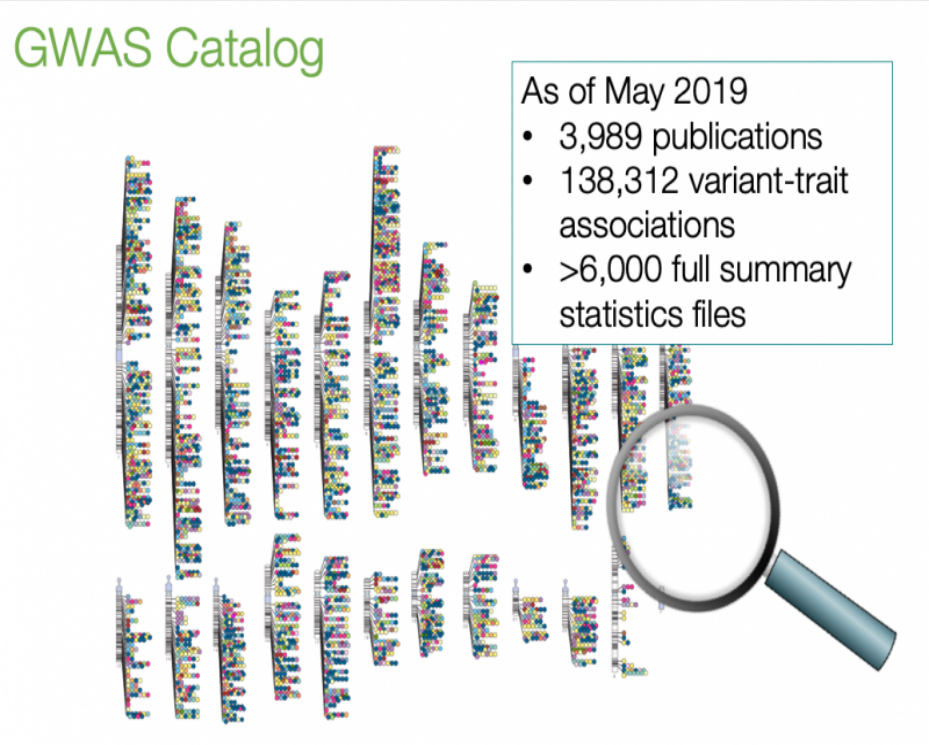
Multi-trait genome-wide analyses of the brain imaging phenotypes in UK Biobank

Chong Wu
Department of Statistics
Florida State University

ASHG 2019

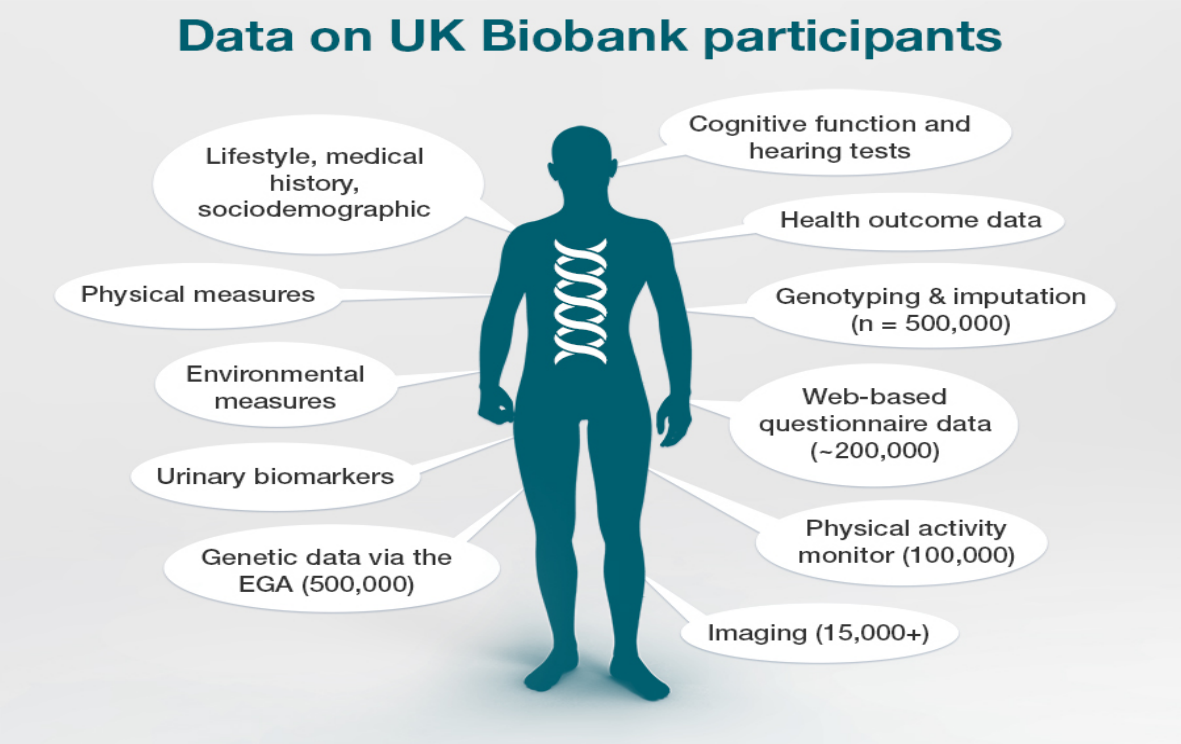
Oct. 16, 2019

Introduction



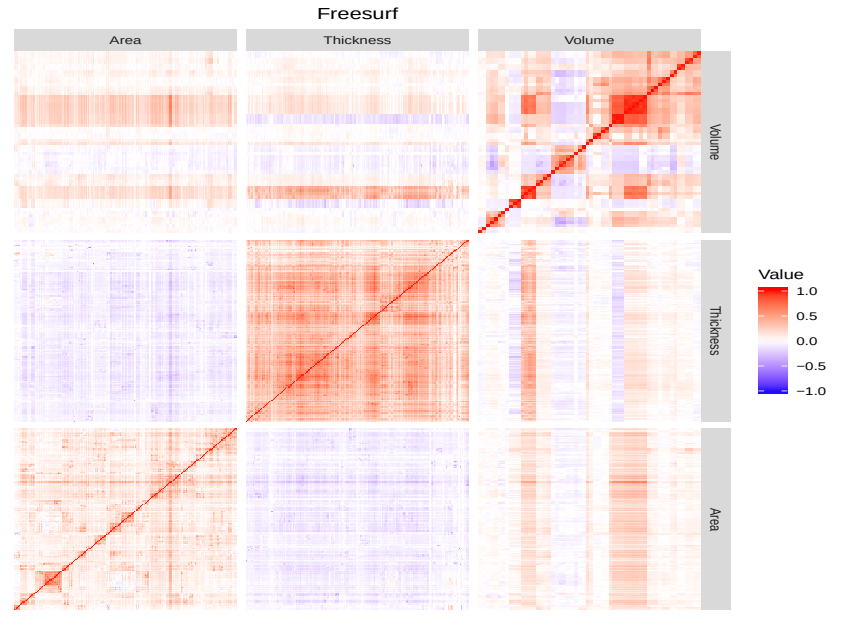
- "missing heritability" problem
- Many genetic variants are associated with multiple traits
- Multi-trait association tests

UK Biobank data



- Deep phenotyping data
- 3,144 brain image-derived phenotypes (IDPs) (Elliott et al. Nature, 2018)

Challenges



- Most existing studies analyze less than ten traits jointly
- For deep phenotyping data, we have many traits
- Some traits are highly correlated
- Individual-level data may not available

Goals

Develop a new multi-trait association test that

- enables a joint analysis of an arbitrary number (e.g. hundreds) of traits
- yields well-controlled Type 1 error rates
- achieves robust high power across different scenarios
- can apply to GWAS summary statistics
- computationally efficient

Outline

- Background
- Methods
- Results
- Discussion

Model set-up

- Suppose we have Z scores across p traits of interest for SNP j, $\mathbf{Z}_j = (Z_{j1}, Z_{j2}, \ldots, Z_{jp})$
- $\beta = (\beta_1, \ldots, \beta_p)'$ be the marginal effect sizes of the SNP j for p traits
- $H_0 : \beta = 0$ vs. $H_1 : \beta_j \neq 0$ for at least one $j \in \{1, 2, \ldots, p\}$
- Under the null, $\mathbf{Z}_j \sim N(0, R)$, where R is the trait correlation matrix

adaptive multi-trait association test (aMAT)

- Estimating trait correlation matrix R by LD score regression (LDSC)
- Constructing a class of multi-trait association tests (MAT)
- Constructing an adaptive test called aMAT to maintain robust power across different scenarios

MAT

- Chi-squared test: $T_{\chi^2} = \mathbf{Z}'\hat{R}^{-1}\mathbf{Z}$
- Challenge: when analyzing hundreds of traits or highly correlated traits jointly, $\hat{R}$ is often near singular
- $\hat{R} = U\Sigma U'$ (SVD)
- $\hat{R}^{+}_{\gamma} = U\Sigma^{+}_{\gamma}U'$
- Only keep the largest k singular values such that $\sigma_1/\sigma_k < \gamma$
- $T_{\mathrm{MAT}(\gamma)} = \mathbf{Z}'\hat{R}^{+}_{\gamma}\mathbf{Z}$

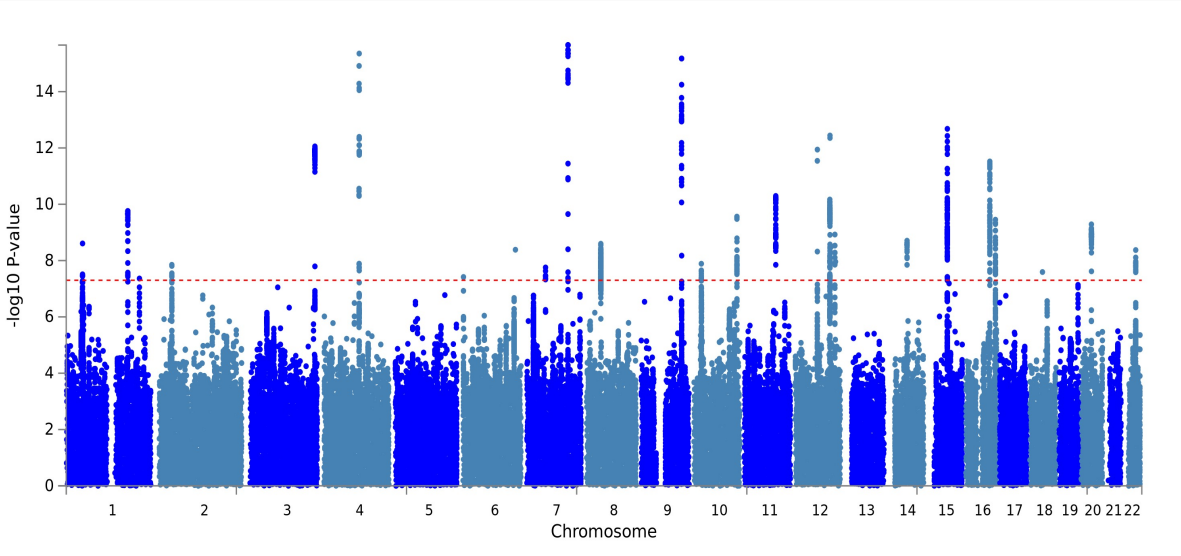
aMAT

- There is no uniformly most powerful test
- MAT(1) achieves high power when the first PC captures the majority association signals across p traits
- When most PCs have weak signals, MAT with larger γ will be more powerful
- $T_{\text{aMAT}} = \min_{\gamma \in \Gamma} p_{\text{MAT}(\gamma)}$, where $\Gamma = \{1, 10, 30, 50\}$

Outline

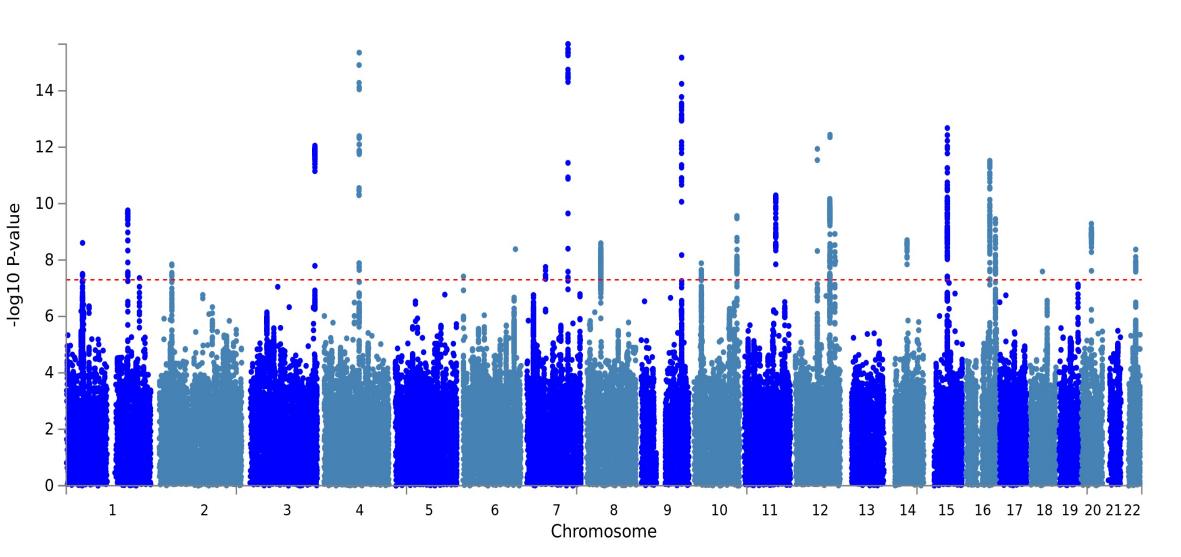
- Background
- Methods
- **Results**
- Discussion

Analysis of UK Biobank brain imaging GWAS summary data



- For illustration, we focus on the results of analyzing the group of 58 Freesurfer volume IDPs
- Among about 10 million SNPs, aMAT identified 801 significant SNPs, 453 of which were ignored by any individual IDP tests at the 5×10^{-8} significance level

Analysis of UK Biobank brain imaging GWAS summary data



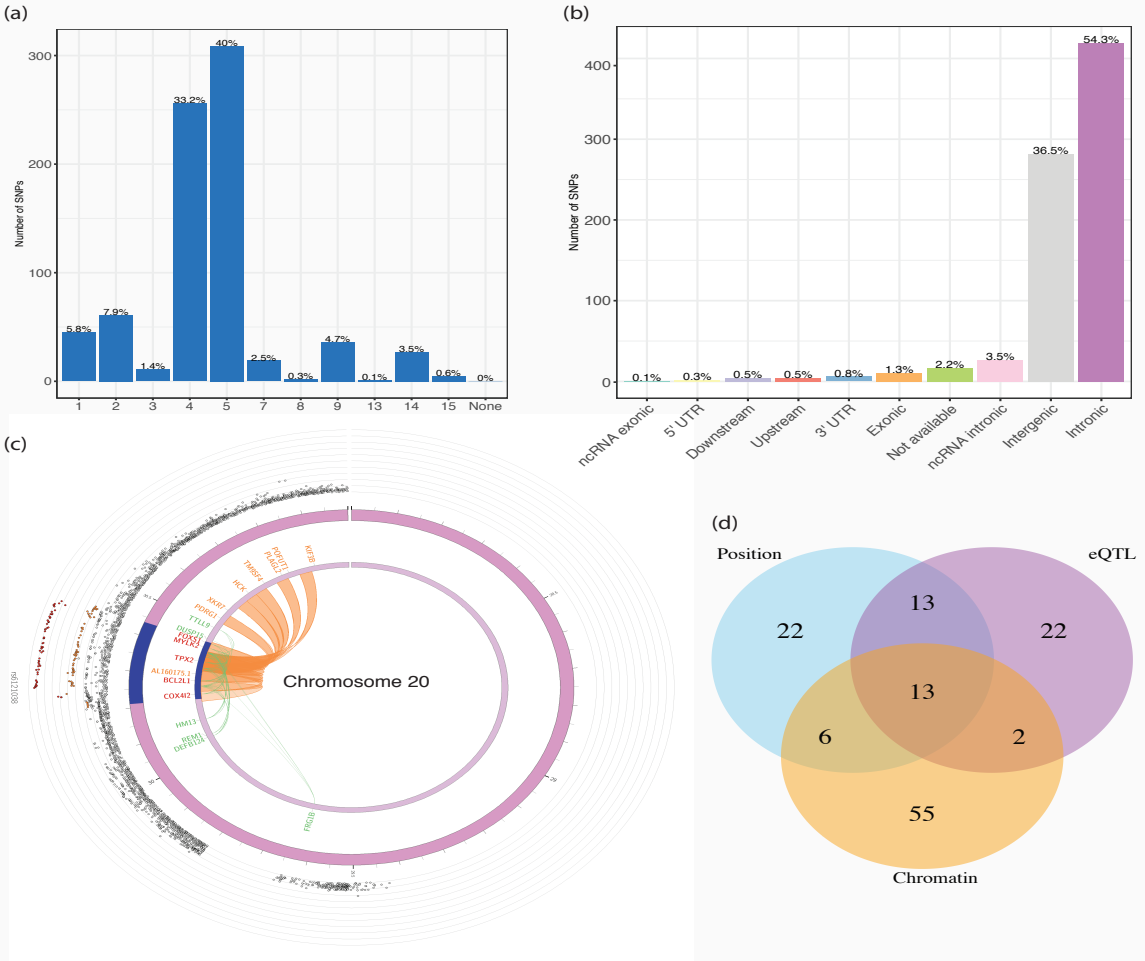
- 28 lead SNPs, located in 24 distinct risk loci
- Among these 28 lead SNPs, 13 SNPs (46.4%) were missed by any individual IDP tests

Replication of aMAT-identified loci

Replicate by the ENIGMA consortium (Hibar et. al, Nature, 2015)

- GWAS summary statistics of seven subcortical volumes in up to 13,171 subjects
- Among 28 lead SNPs, 13 SNPs showed nominally significant association results (two-tailed binomial test $P = 2.2 \times 10^{-10}$); four loci showed genome-wide significant association results ($P = 6.3 \times 10^{-30}$)

Functional annotation of genetic variants



Functional annotation of genetic variants

- Relevant SNPs were chromatin states 4 (33.2%) and 5 (40.0%), indicating effects on active transcription
- Five genome-wide significant SNPs (rs10507144, rs3789362, rs4646626, rs6680541, and rs2845871) had a high observed probability of a deleterious variant effect (CADD score > 20)
- The identified genes were enriched in many GWAS catalog reported volume related gene sets, including dentate gyrus granule cell layer volume $P = 1.5 \times 10^{-13}$ and hippocampal subfield CA4 volume $P = 1.5 \times 10^{-13}$

Outline

- Background
- Methods
- Results
- Discussion

Discussion

- Multi-trait analysis is different from cross phenotype or pleiotropy effect analysis, where the null hypothesis is at most one trait is associated with the SNP
- aMAT is a general framework and can be easily extended to incorporate other multi-trait methods such as MTAG, N-GWAMA, and HIPO
- Codes: https://github.com/ChongWu-Biostat/aMAT
- Manuscript: https://www.biorxiv.org/content/10.1101/758326v1.abstract

Acknowledgment

- RCC@FSU
- ENIGMA and Elliott et al. that made their GWAS summary data available
- Looking for collaborators who are interested in imaging genetics, Alzheimer's disease, and analyzing UK Biobank individual data

Thank you!