A powerful framework for integrating eQTL/imaging and GWAS summary data

Wei Pan, Zhiyuan Xu, Chong Wu

Division of Biostatistics, University of Minnesota

September 15, 2017

Background

- Uncovered risk loci only account for a small proportion of the heritability for each complex trait;
- One obvious but costly approach is to have a larger sample size, motivating meta- or mega-analyses by large consortia.
- Complementary strategy is to use multiple endophenotypes, intermediate between genetics and the disease.

Background

- Since the identified SNPs or genes may or may not be associated with the disease, e.g. AD, further studies are still needed to confirm or refute a suggestive link based on imaging endophenotypes.
- The sample size of a typical GWAS with imaging traits is still much smaller than those of other GWAS with clinical trait.

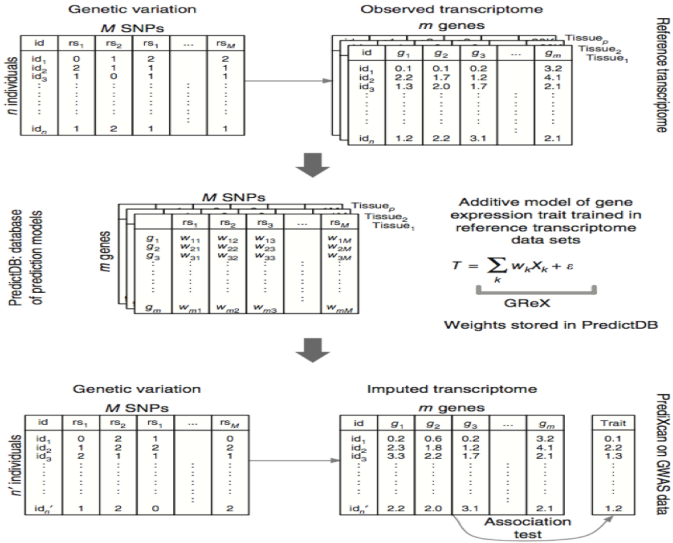
Our idea:

Inspired by TWAS, we propose IWAS, which uses an imaging endophenotype to construct weights for a weighted gene-based GWAS test.

Transcriptome-wide association study

- Two methods, PrediXcan/TWAS (Nature genetics 2015, 2016) similar idea;
- predict/impute gene-expression with SNPs as predictors;
- test association b/w a trait and imputed gene-expression;
- we propose a novel re-formulation of PrediXcan/TWAS.

PrediXcan idea



TWAS/PrediXcan idea cont.

- Build a prediction model for genetically regulated expression (GRex): $Y^* = \sum_{j=1}^{p} w_j X_j^* + \epsilon$, where Y^* is gene-expression.
- for a given gene for subject i, predict the GReX of the gene using the SNPs around that gene: $\widehat{\text{GReX}}_i = \sum_{j=1}^{p} \hat{w}_j X_{i,j}$;
- test association between a trait and predicted gene-expression: $g(E(Y_i)) = \beta_0 + \widehat{\text{GReX}}_i \beta_c = \beta_0 + \sum_{j=1}^{p} \hat{w}_j X_{i,j} \beta_c$ with null hypothesis H_0: $\beta_c = 0$.

Novel reformulation

- Consider a GLM: $g(E(Y_i)) = \beta_0 + \beta' X_i = \beta_0 + \sum_{j=1}^{p} X_{i,j}\beta_j$ with $H_0 : \beta = (\beta_1, ..., \beta_p)' = 0$;
- replace $X_{i,j}$ by the weighted genotype scores $\hat{w}_j X_{i,j}$;
- PrediXcan = TWAS = Sum test (Pan 2009).
- $U^* = (U_1^*, ..., U_p^*)' = \sum_{i=1}^{n} X_i'(Y_i - \hat{\mu}_i^0)$;
  $U = (U_1, ..., U_p)' = WU^* = \sum_{i=1}^{n} WX_i'(Y_i - \hat{\mu}_i^0)$,
  where $W = \text{Diag}(\hat{w}_1, ..., \hat{w}_p)$

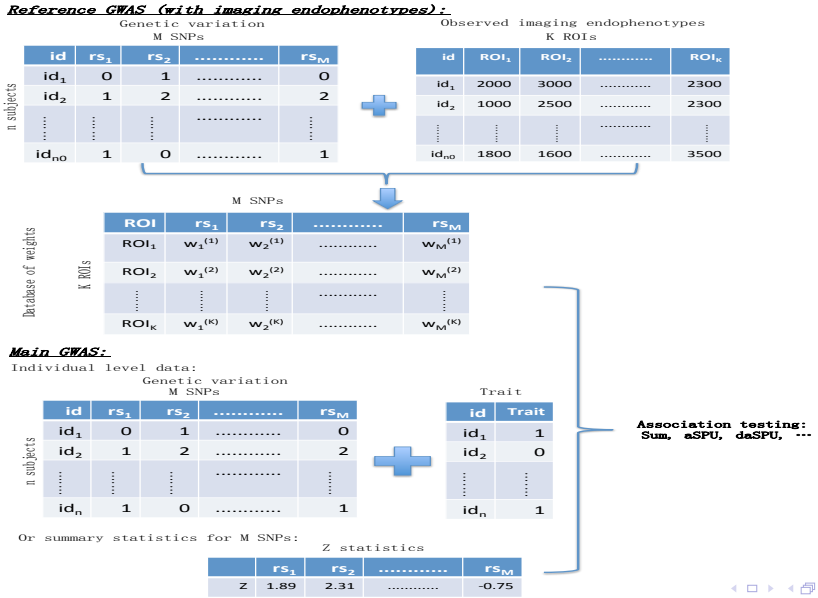
aSPU test with a single set of weight

- Sum test: $T_{\text{Sum}} = \sum_{j=1}^{p} U_j$;
  SSU test: $T_{\text{SSU}} = U^T U = \sum_{j=1}^{p} U_j^2$
- More generally, for an integer $\gamma \geq 1$, an SPU(γ) test is defined as: $T_{\text{SPU}(\gamma)} = \sum_{j=1}^{p} U_j^{\gamma}$;
- for an even integer $\gamma \to \infty$,
  $T_{\text{SPU}(\gamma)} \propto \left(\sum_{j=1}^{p} |U_j|^{\gamma} \right)^{1/\gamma} \to \max_j |U_j| = T_{\text{SPU}(\infty)}$.
- $T_{\text{aSPU}} = \min_{\gamma \in \Gamma} P_{\text{SPU}(\gamma)}$, where $P_{\text{SPU}(\gamma)}$ is the p-value of the SPU(γ) test.

Apply methods to summary statistics

- Suppose that $Z_j = \hat{\beta}_j / \mathrm{SE}_j$ is the Z-statistic for association between the GWAS trait and SNP j, where $\hat{\beta}_j$ is the estimated (marginal and signed) association effect and SE_j is its standard error.
- simply redefine $U = WZ$ with $Z = (Z_1, Z_2, ..., Z_p)'$, then proceed as before.
- We use a reference sample (e.g. the 1000 Genome Project samples) to estimate linkage disequilibrium (LD) among the SNPs and thus the correlation matrix for Z and U (Kwak and Pan 2016; Gusev et al 2016).

IWAS: Integrate imaging endophenotypes into GWAS



Constructing the weights

- We used the ADNI-1 data set.
- We Ran quality control first: MAF > 0.05. All the ambiguous SNPs with alleles A/T or C/G were removed.
- Five covariates were also included: baseline age, gender, baseline education (in years), handedness (left or right), and baseline intracranial volume.

Constructing the weights

- We Built a prediction model using elastic net for each of the 14 endophenotypes, the gray matter volumes of the 12 regions of interest (ROIs) related to the default mode networks (DMN) and those of hippocampus, due to the possible relatedness of the above ROIs to AD.
- We calculated the squared Pearson correlation, r^2, between the predicted and observed endophenotype values in the dataset, and selected only those genes with $r^2 > 0.01$.

Apply methods to IGAP summary data

- The International Genomics of Alzheimer's Project (IGAP), stage 1 data;
- $\sim$ 7 million SNPs; $\sim$ 54,000 subjects.
- Derive imaging weights from ADNI data;
- gene expression weights were downloaded from PrediXcan database.

IWAS results

Table: The numbers of the significant AD-associated genes identified by IWAS and TWAS with the IGAP data. The numbers a/b in each cell indicate the total number of significant genes/number of significant genes that overlap at least a genome-wide significant SNP.

	Imaging weights				Gene-exp weights	
	1	2	3	4	Blood	Hippo
SPU(1)	8/8	65/64	41/40	73/64	13/5	11/4
SPU(2)	10/10	66/66	40/40	67/67	15/11	7/5
aSPU	10/10	70/69	40/40	77/69	16/9	13/5

Apply methods to lipids 2013 summary data

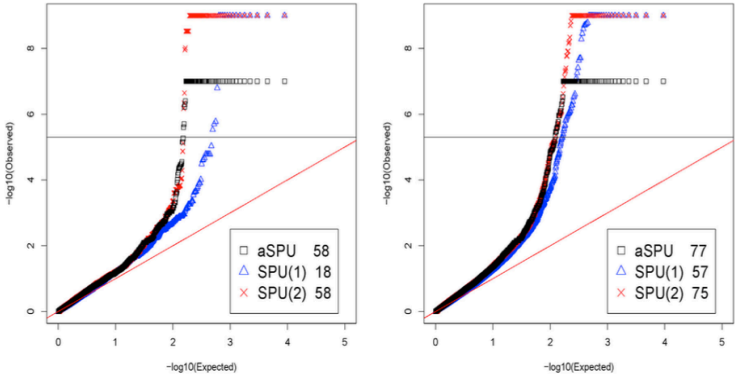


Fig. 5. Q-Q plots for (a) IWAS and (b) TWAS testing for gene-HDL association applied to the 2013 lipid GWAS summary statistics. The **numbers** in the second column in each legend box of each panel indicate the *numbers of the genome-wide significant genes* identified by each method.

Discussion

- We have developed a powerful adaptive test (aSPU) to integrate GWAS and eQTL data;
- PrediXcan and TWAS, can be regarded as a special case of our proposed test;
- incorporating weights derived from other sources of endophenotypes (e.g. imaging phenotypes).

Reference

- Gamazon, E.R et al. (2015) A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics* **47**, 1091-1098.
- Gusev, A et al. (2016) Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**, 245-252.
- Xu, Zhiyuan, et al. (2017) Imaging-wide association study: Integrating imaging endophenotypes in GWAS. *NeuroImage*.

Thank you!

Questions?