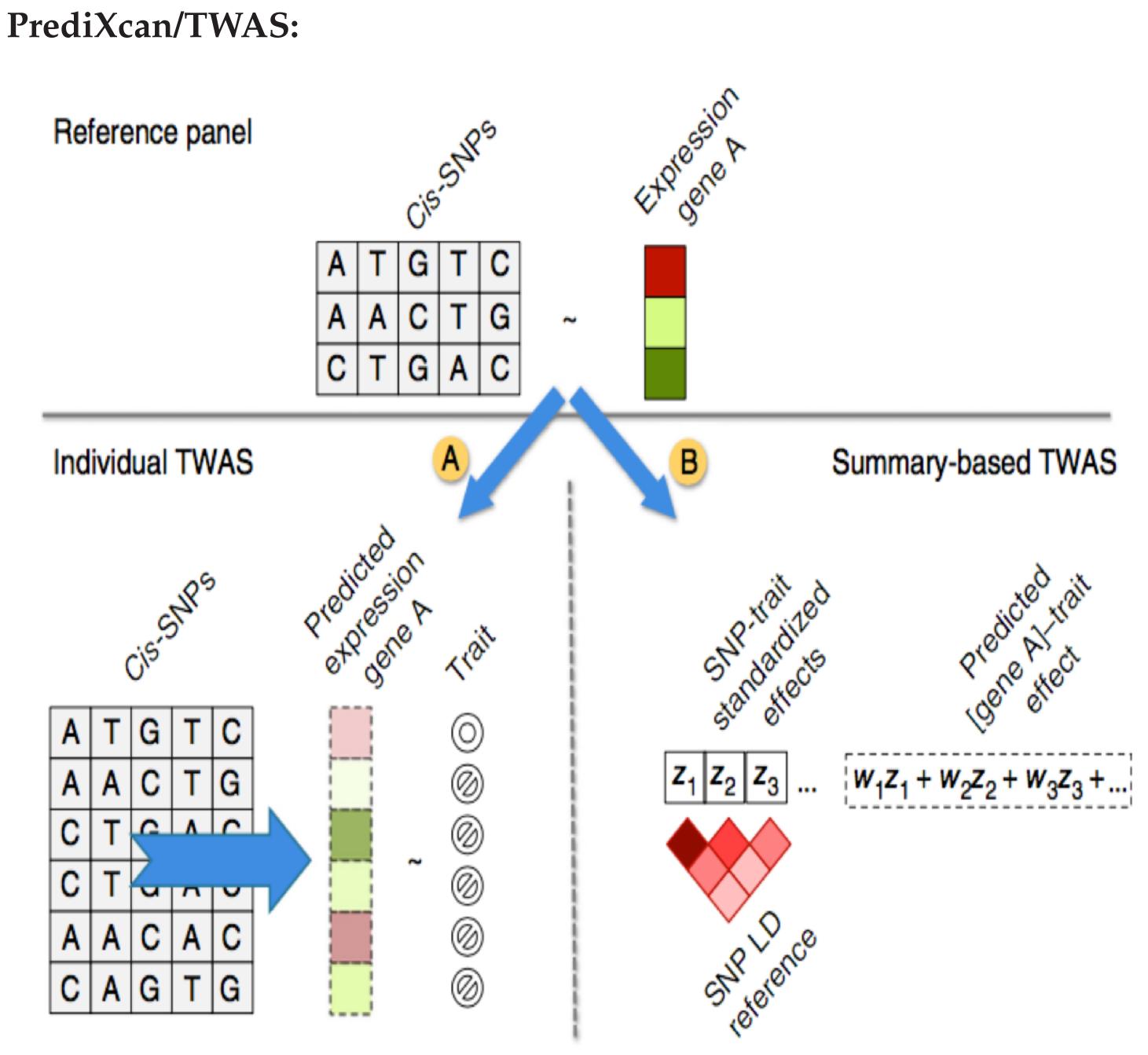# Integrating eQTL data with GWAS summary statistics

Chong Wu and Wei Pan   wuxx0845@umn.edu
Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455

Background

PrediXcan/TWAS:
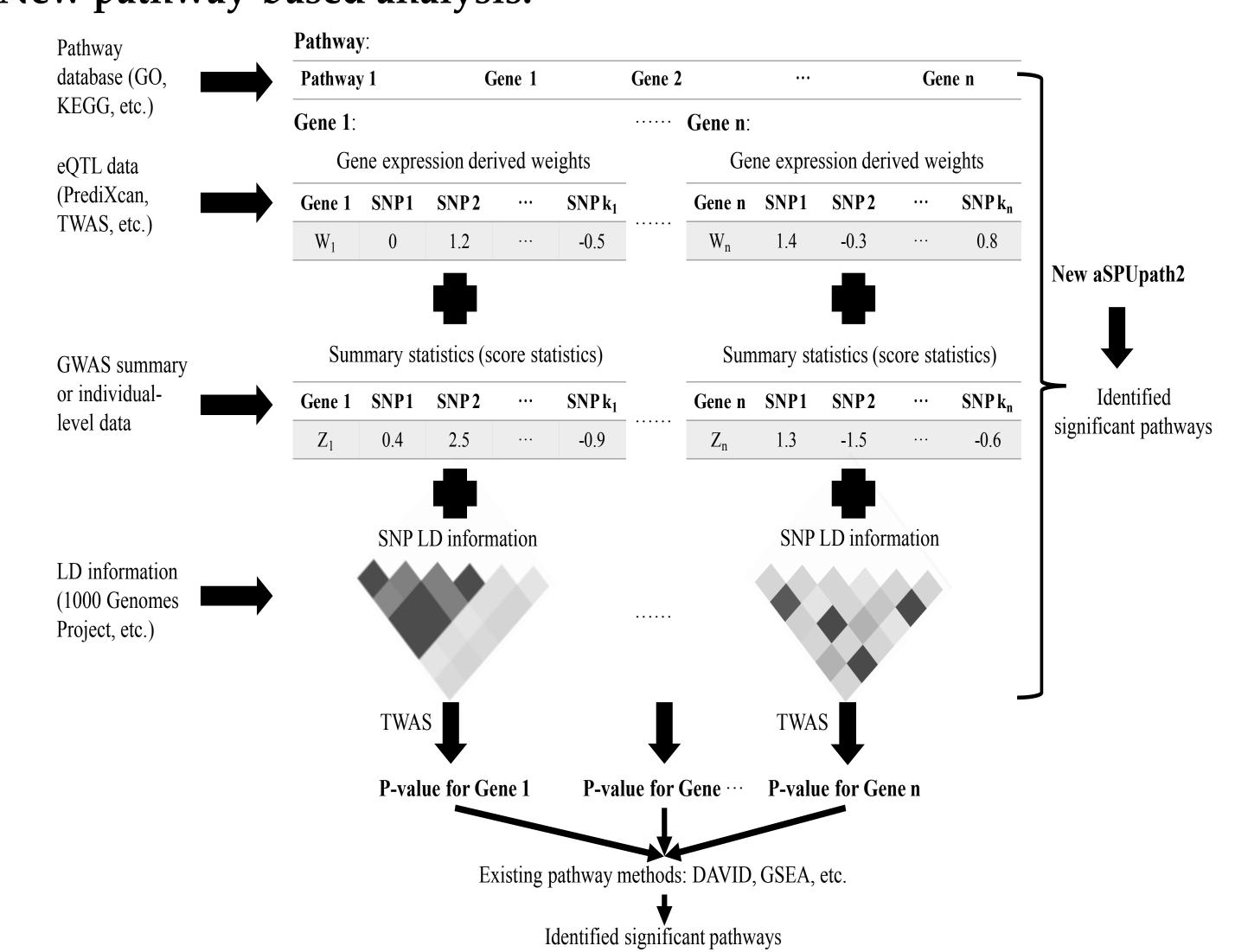

copyright @ Gusev et. al. (2016)

New methods

New gene-based analysis:

- PrediXcan = TWAS = (weighted) Sum test;

- Sum test will lose statistical power under many situations;

- More powerful tests, such as SSU and aSPU, can be applied.

New pathway-based analysis:

Remark:

- aSPUpath2 formula:

$$\text{PathSPU}(\gamma) = \sum_{g=1}^{n} \sum_{k=1}^{k_g} (W_{gk}^{s} Z_{gk})^{\gamma},$$

$$\text{aSPUpath2} = \min_{\gamma \in \{1,2\}} P_{\text{PathSPU}(\gamma)}.$$

- In pathway analysis, one would aim to identify those containing many associated genes or SNPs with only weak effects that cannot be detected by single SNP- or gene-based analysis.

- p-values can be calculated by an asymptotics-based method.

Applications: SCZ associated genes

Two GWAS summary datasets:

- The SCZ1 data: contains the meta-analyzed summary statistics based on 20,899 individuals;

- The SCZ2 data: based on 150,064 individuals.

Results:

Table: The numbers of the significant genes identified by analyzing the SCZ1 data.

	YFS	NTR	CMC	Combined
TWAS	14/11/14	13/8/13	16/10/13	51/31/43
SSU	31/25/26	27/19/26	39/25/34	108/67/95
aSPU	29/26/26	23/16/22	21/16/21	87/64/79

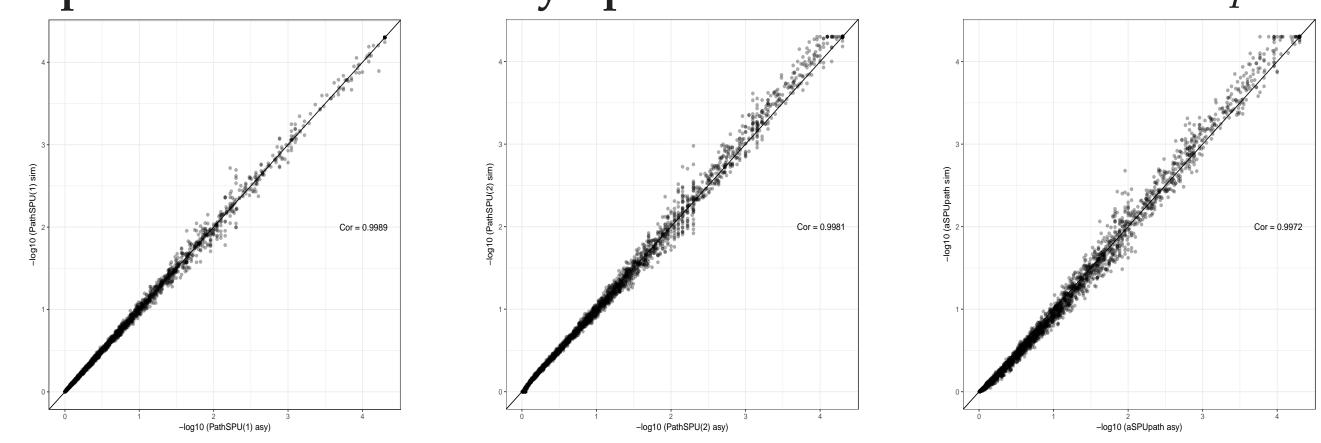Table: The numbers of the significant genes identified by analyzing the SCZ2 data.

	YFS	NTR	CMC	Combined
TWAS	63/19/46	56/17/37	69/21/50	202/63/142
SSU	127/40/94	100/22/61	124/32/85	381/108/255
aSPU	105/40/83	85/24/55	110/34/82	314/110/234

Note: The numbers a/b/c in each cell indicate the numbers of (a) the significant genes; (b) the significant genes covering at least one genome-wide significant SNP within 500 kb in the SCZ1 data; (c) the significant genes covering at least one genome-wide significant SNP within 500 kb in the SCZ2 data.

- Some newly identified genes: *FOXN2*, *MSRA*, *PAX5*, *MAP7D1*, *CNN3*, *GABPB2*, *TBC1D5*, *IK*, *CXXC5*, *MCO6*, *DND1*, and *TMCO6*.

Applications: SCZ associated pathways

Comparison between the asymptotics- and simulation-based p-values:

The significant and novel gene sets containing no significant genes as identified by aSPUpath2 (based on the SCZ2 data):

ID	Description	# genes
GO:1902710	GABA receptor complex	18
GO:1901661	quinone metabolic process	29
GO:0043162	ubiquitin-dependent protein catabolic process	18
GO:0016339	calcium-dependent cell-cell adhesion	27
GO:0030315	T-tubule	45
GO:0007528	neuromuscular junction development	36
GO:0003143	embryonic heart tube morphogenesis	62
GO:0007569	cell aging	67
GO:0035050	embryonic heart tube development	73
GO:0004181	metallo-carboxypeptidase activity	27
hsa00590	Arachidonic acid metabolism	56
GO:0051279	regulation of release of sequestered calcium ion into cytosol	75
GO:0072665	protein localisation to vacuole	46
GO:0010880	regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum	25

Software & Advertisement

Related software:

- http://wuchong.org/aspupath2.html

- http://wuchong.org/TWAS.html

Advertisement: Chong Wu (a biostatistics Ph.D. candidate at the University of Minnesota) is looking for an Assistant Professor or Research Scientist position. Please contact me at wuxx0845@umn.edu if you have any suitable positions.