Integrative approaches for large-scale transcriptome-wide association studies

Alexander (Sasha) Gusev

Harvard T.H. Chan School of Public Health

How does the transcriptome inform GWAS?

- Top SNPs and polygenic heritability enriched at eQTLs

- **New locus discovery**: aggregating effects into single expression instrument, fewer tests

- **Fine mapping**: Genes are more interpretable units than tagging SNPs

☹ Ignores effects that are entirely non-genic

Wright et al. 2014 Nat Genet; Torres et al. 2014 AJHG; Lappalainen et al. 2013 Nature

What kind of data is needed?

- Evaluating expression in large GWAS is still prohibitive but summary GWAS data from >100k samples are readily available.

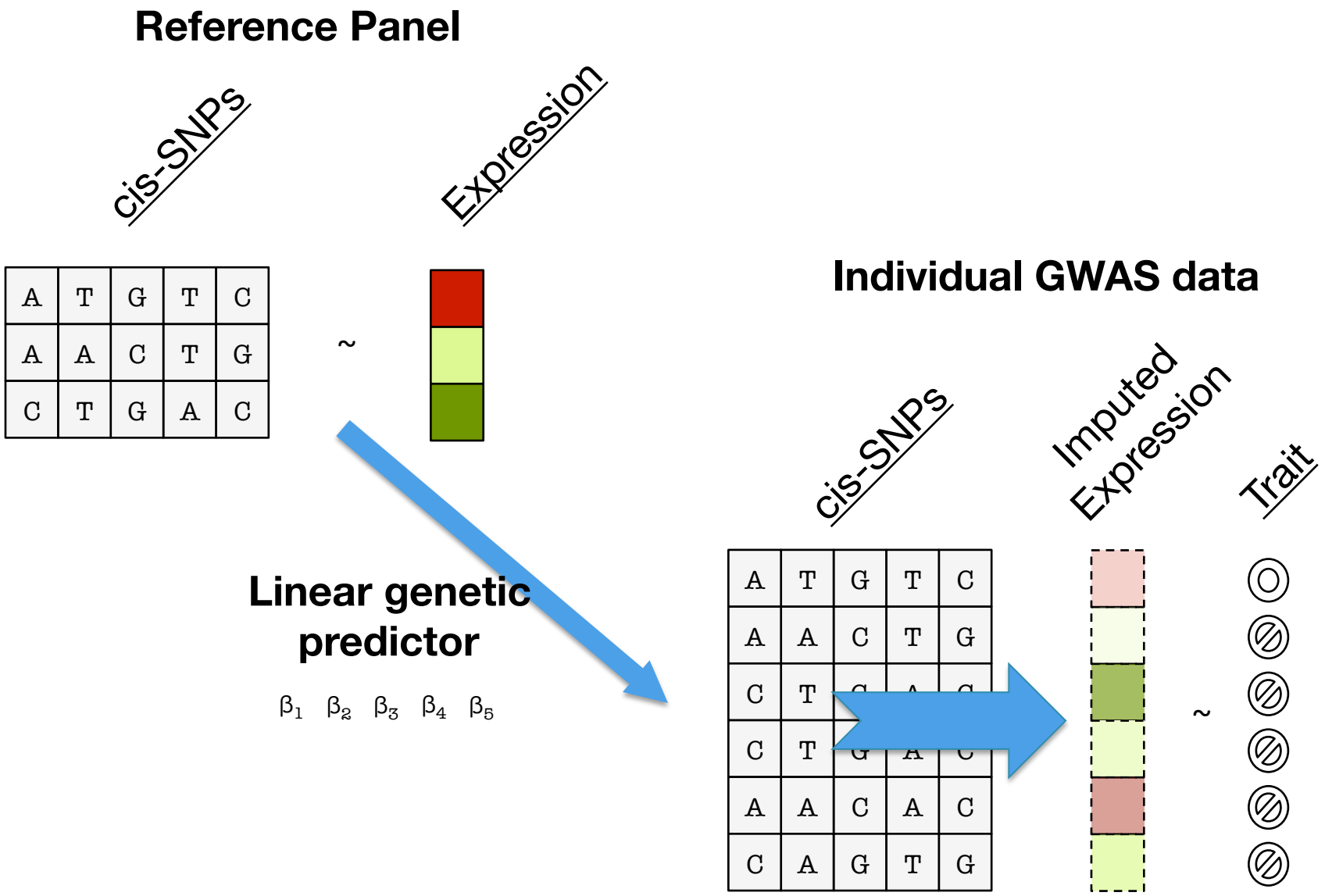
- Expression measured in >1k samples from multiple tissues.

- Goal: **Transcriptome-wide Association Study (TWAS):**
  (1) Impute cis-genetic component of expression
  (2) associate imputed expression to trait

Outline

1. Imputing gene expression & individual TWAS
2. Summary-based TWAS
3. TWAS on real phenotypes

Model: Imputation with individual data



Gusev et al. (in revision) *Nat Gen*; **See also**: Gamazon et al. 2015 *Nat Gen*

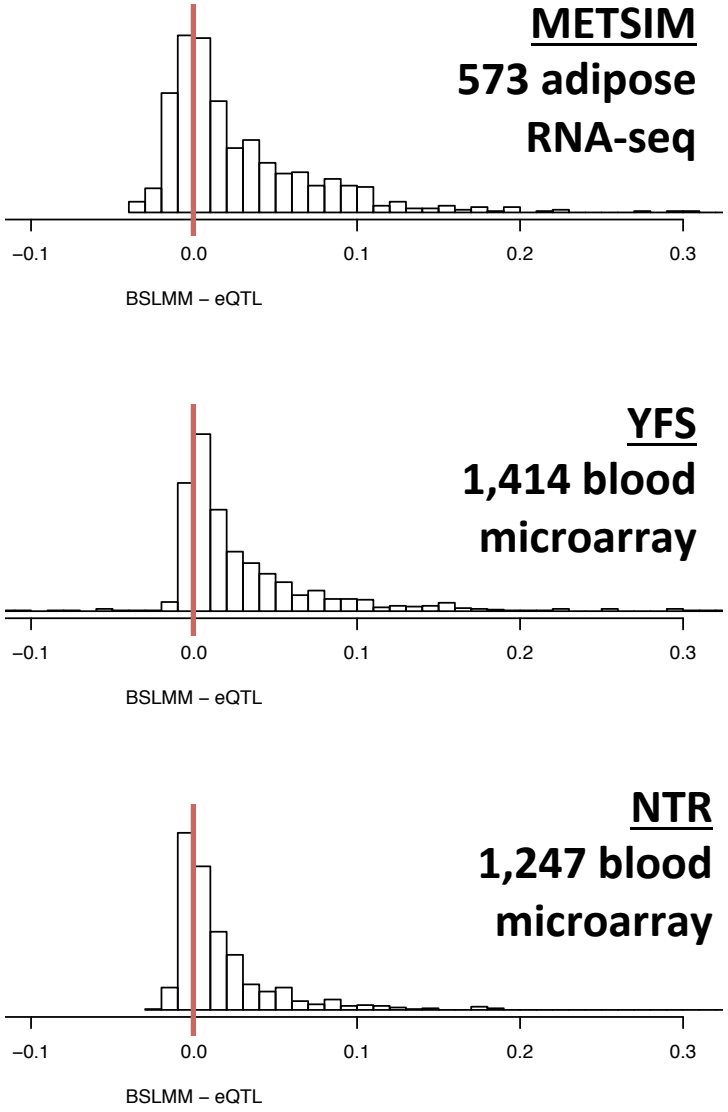
Long tail of improved imputation

Cross-validation in heritable genes from three datasets:

- eQTL explains ~50% of cis effect!
- TWAS consistently better predictor, 1.3x on average
- 25% of genes have **2x** better prediction
- Larger samples further increase relative imputation accuracy

See poster: ***1018F - Mancuso***

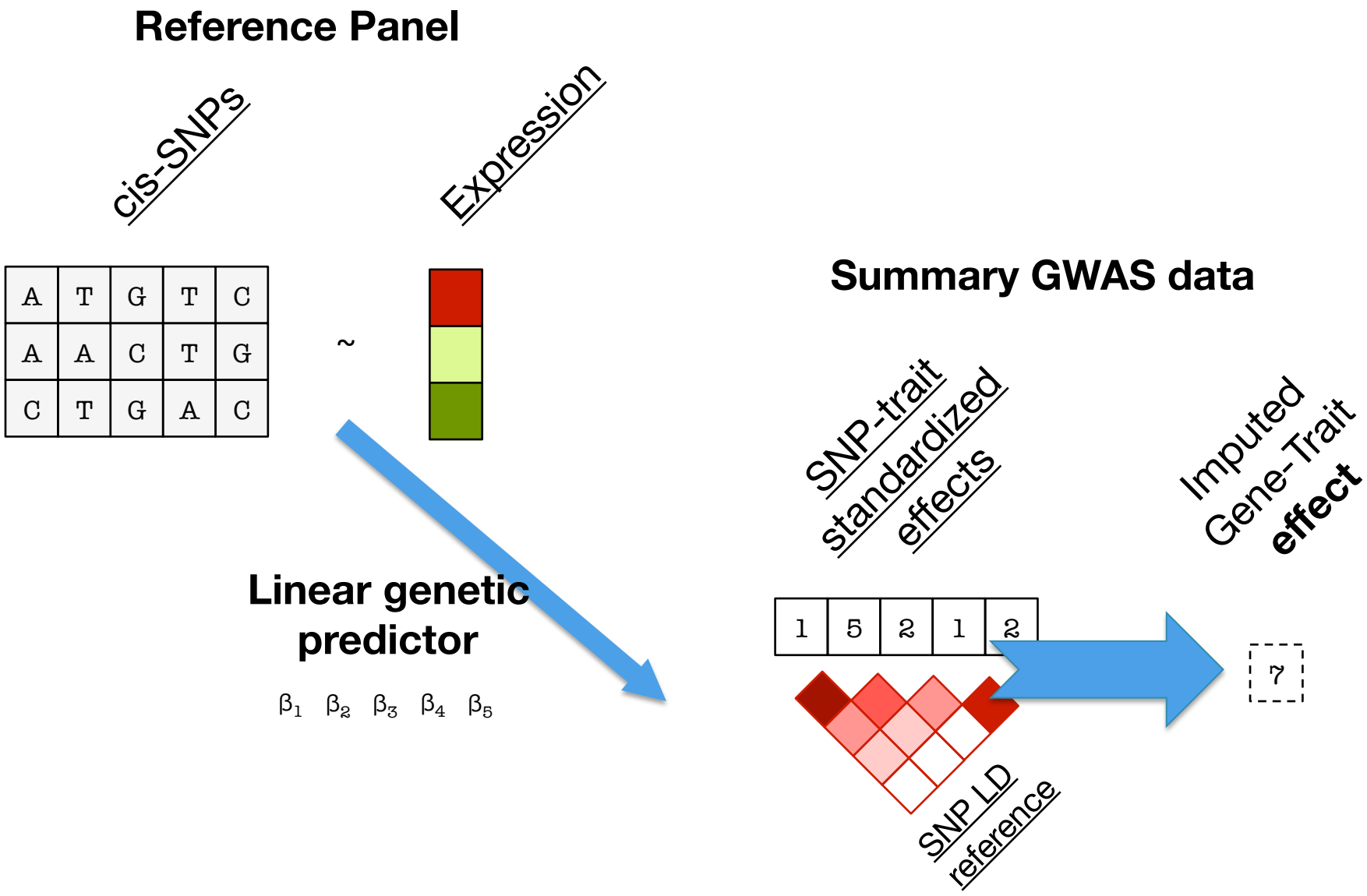
TWAS– eQTL accuracy



Outline

1. Imputing gene expression & individual TWAS
2. Summary-based TWAS
3. TWAS on real phenotypes

Summary-based TWAS

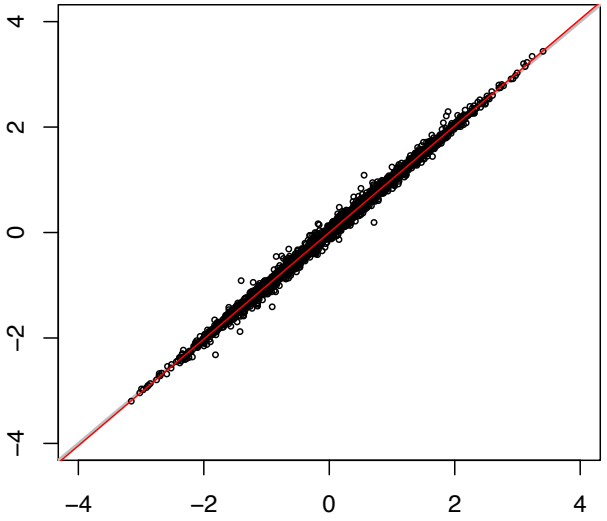


Gusev et al. (in revision) *Nat Gen*

Accurate associations from summary data

Estimate Z-score of [height ~ cis-expression]
with summary and individual data

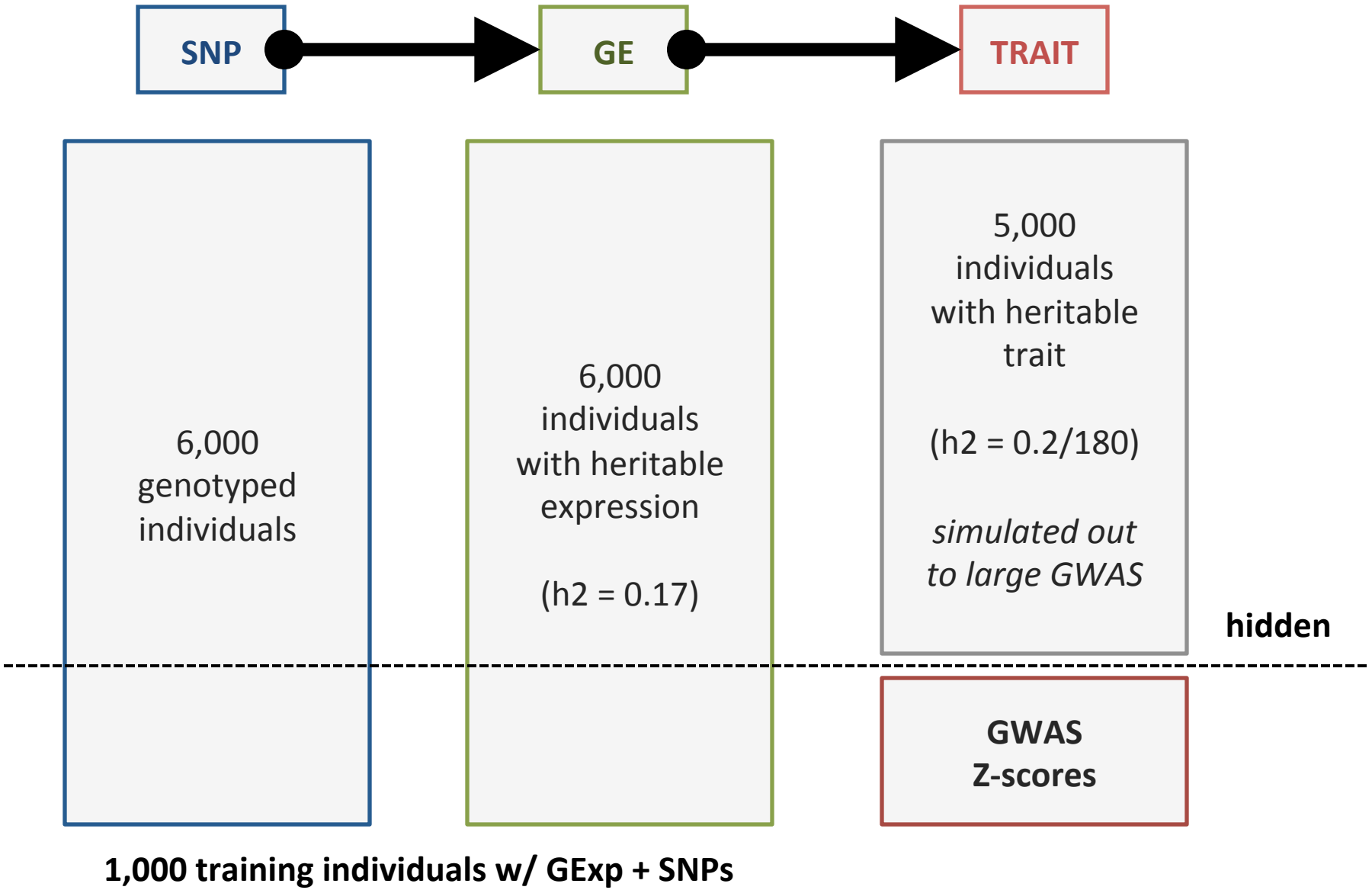
Individual-based
TWAS Z-score



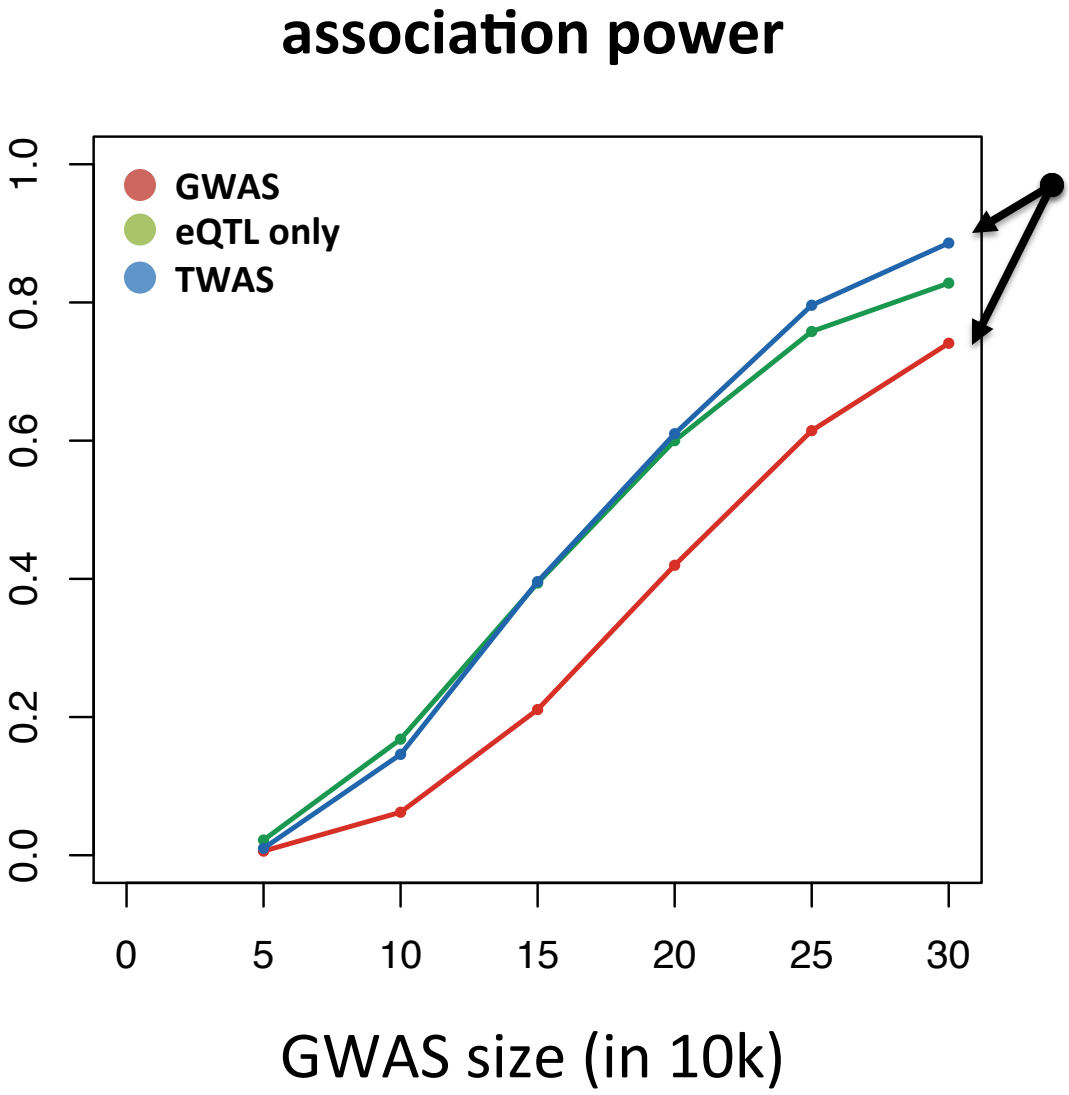
Summary-based
TWAS Z-score

Nearly identical associations (correlation 0.998)

Power: Simulated expression & GWAS study



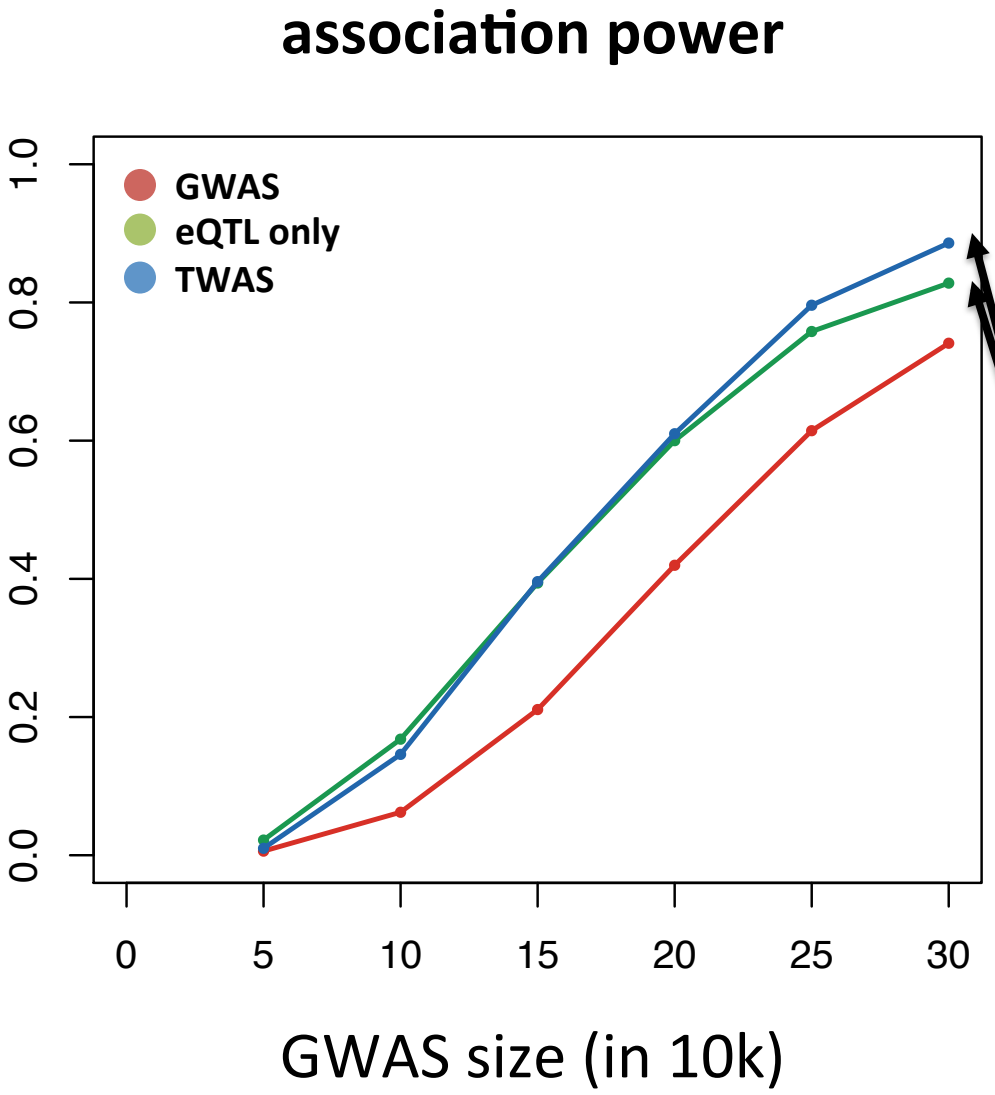
Power to detect heritable genic effect



1 causal variant:

More power than GWAS due to lower testing burden.

Power to detect heritable genic effect

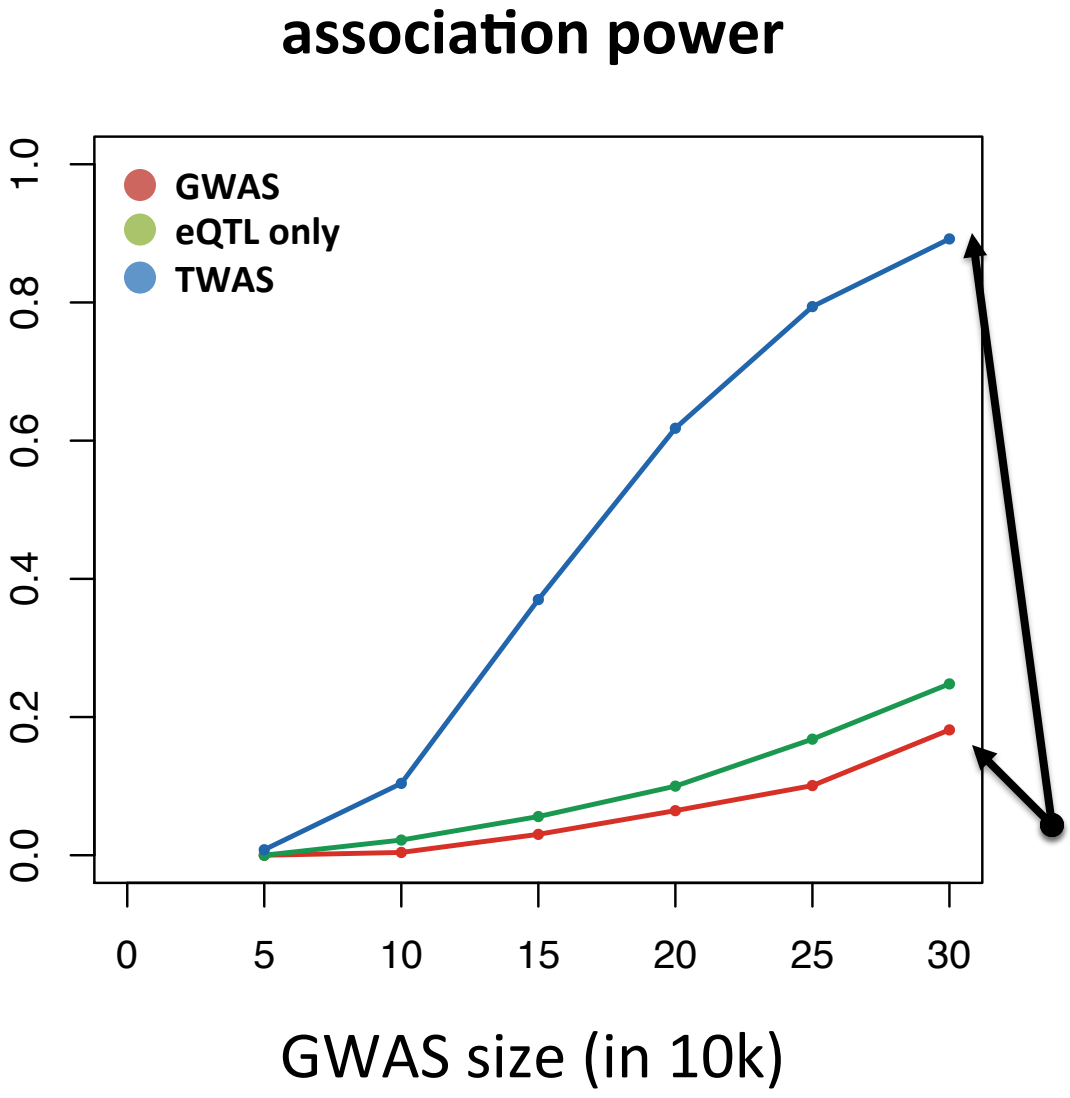


1 causal variant:

More power than GWAS due to lower testing burden.

Comparable to eQTL with one causal variant.

Power to detect heritable genic effect



1 causal variant:

More power than GWAS due to lower testing burden.

Comparable to eQTL with one causal variant.

10% causal:

Much more power than GWAS/eQTL with multiple effects.

Outline

1. Imputing gene expression & individual TWAS
2. Summary-based TWAS
3. TWAS on real phenotypes

Novel TWAS associations predict GWAS associations in larger studies

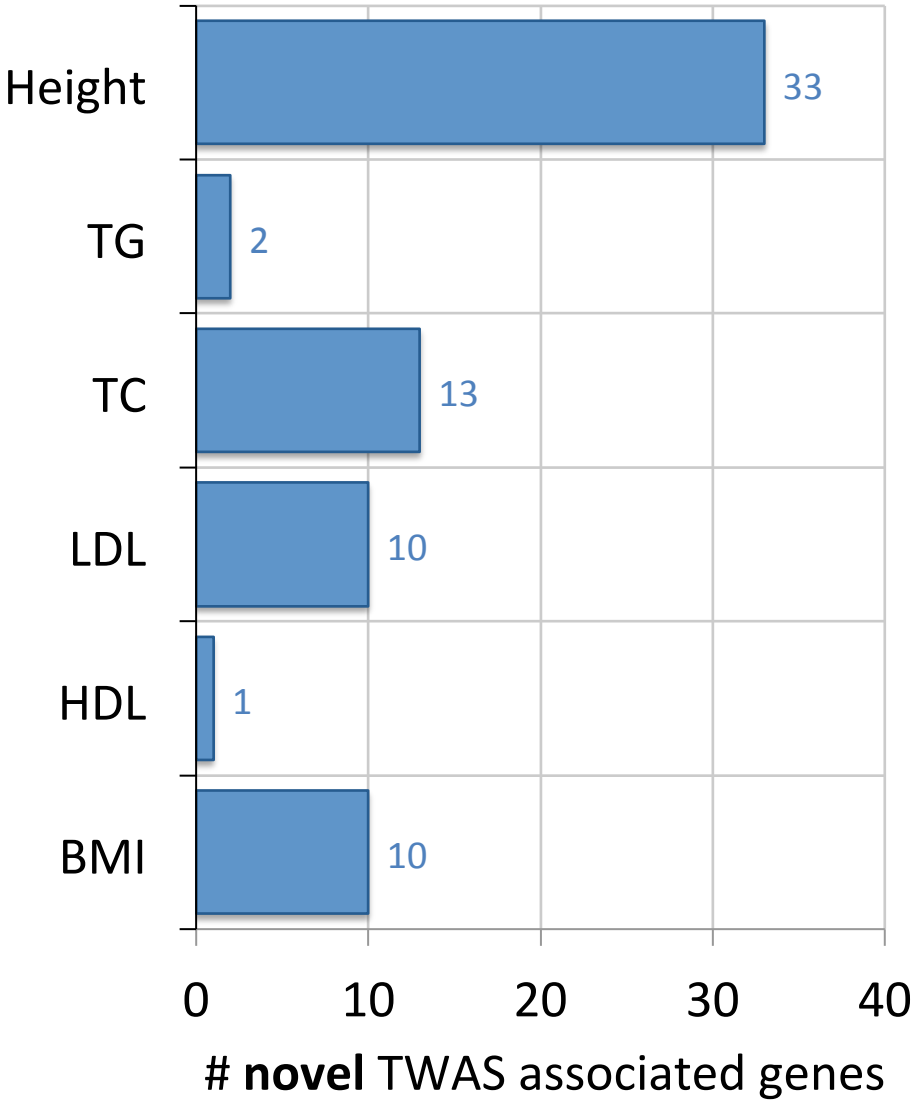
1. TWAS In 2010 lipid GWAS[a] (100,000 samples) discovers **25** genes not overlapping known hits.

2. In 2013 lipid GWAS[b] (190,000 samples): **19**/25 overlap genome-wide significant SNP

3. Highly significant compared to background (hypergeometric $P=1x10^{-24}$)

(a) Teslovitch et al. 2010 Nature; (b) Willer et al. 2013 Nat Gen

69 novel TWAS genes in four current GWAS



- No nearby (1MB) significant SNP association.
- 54 are strong indicators of allelic heterogeneity (by permutation test).
- 46 replicate using external expression panels
- 927 gene associations at known loci

Wood et al. 2014 Nat Gen; Willer et al. 2013 Nat Gen; Locke et al. 2015 Nature

TWAS identifies relevant genes at known loci

1. YFS study has expression & height
2. Run TWAS at 697 known height loci
3. Correlate 3 expression risk scores with height:

	R^2 with height:
Nearest gene to lead GWAS SNP	0.038
Gene with **best eQTL** from lead GWAS SNP	0.031
Most significant **TWAS gene**	0.054

TWAS genes most strongly correlated with trait ($P=2x10^{-4}$ in a joint model with other scores)

Conclusions

- Full expression-trait association (TWAS), not just top eQTL, can be accurately imputed with individual **or summary data**.
- TWAS genes predictive of future GWAS associations, implicating causal genetic effects.
- Pinpointing specific biological units for follow-up.

- *Informative of causality?*

Acknowledgements

Arthur Ko

Huwenbo Shi

Gaurav Bhatia

Alkes L. Price

Paivi Pajukanta

Bogdan Pasaniuc

NTR:

BWJH Penninx, R Jansen, EJC de Gues, DI Boomsma, FA Wright, PF Sullivan

METSIM + YFS:

E Nikkola, M Alvarez, M Civelek, AJ Lusis, T Lehtimaki, E Raitoharku, M Kahonen, I Seppala, OT Raitakari, J kuusisto, M Laakso

Also see poster: **1018F** "*Enhanced methods for gene expression imputation from genetic variation data*". **N. Mancuso.**