

# **Transcriptome-Wide Association Studies**

Bridging the gap between genome, transcriptome and disease

---

Supervisor: Prof. Paolo Provero  
Candidate: Federico Marotta

Università degli Studi di Torino  
Dipartimento di Biotecnologie Molecolari e Scienze per la Salute

Tesi di Laurea, 18 luglio 2018

# Nature or Nurture: that is the question

## Complex diseases:

As opposed to Mendelian diseases, complex ones cannot be explained by a mutation in a single gene

Do they have a genetic basis at all?

# Nature or Nurture: that is the question

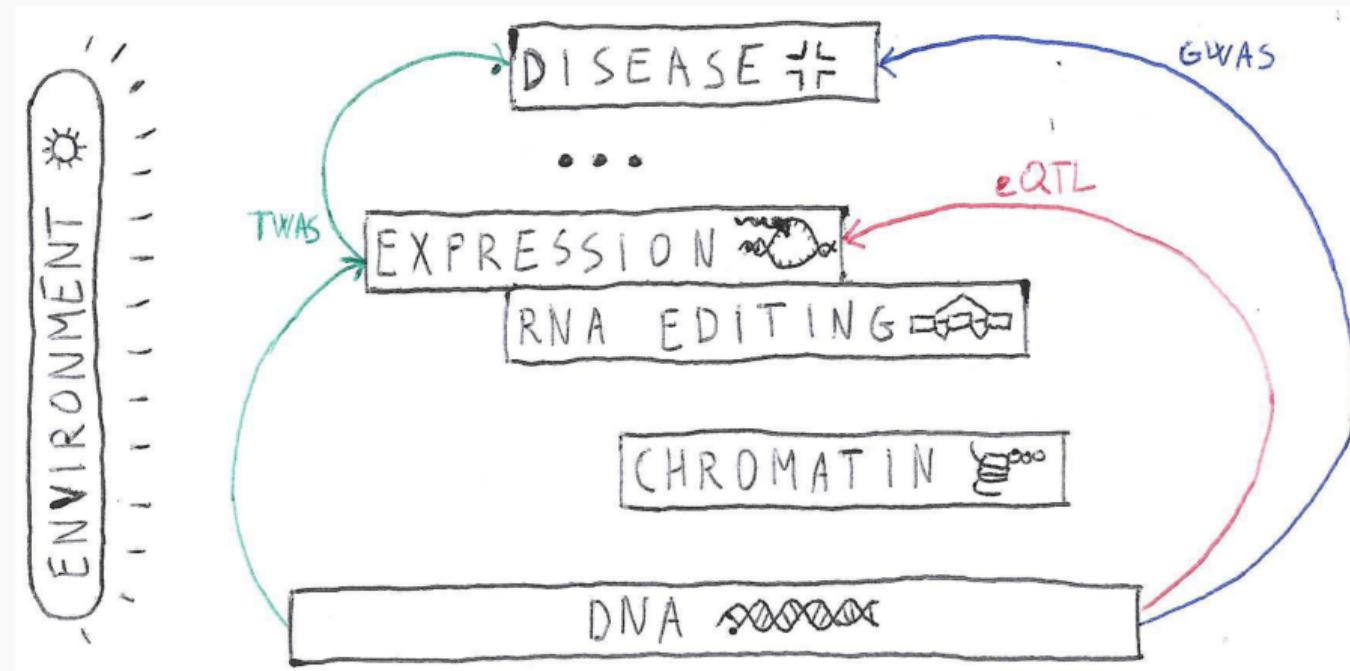
## Complex diseases:

As opposed to Mendelian diseases, complex ones cannot be explained by a mutation in a single gene

Do they have a genetic basis at all?

- Initially, there were **linkage studies** and **candidate-gene association studies**
- Gradually, the focus moved towards **populations** and **whole genomes**
- **GWAS**, especially if combined with eQTL mapping and functional annotations, have found many risk SNPs

# Gene expression, the missing link



**Figure 1:** TWAS focus on the genetic component of expression

# A gene-based association method

---

Eric R. Gamazon et al. 'A gene-based association method for mapping traits using reference transcriptome data'. In: *Nat. Genet.* 47.9 (Sept. 2015), pp. 1091–1098. DOI: [10.1038/ng.3367](https://doi.org/10.1038/ng.3367)

# Training on reference transcriptome data sets

The genetically regulated component of expression must be imputed

$$T = w_1 X_1 + w_2 X_2 + \dots$$

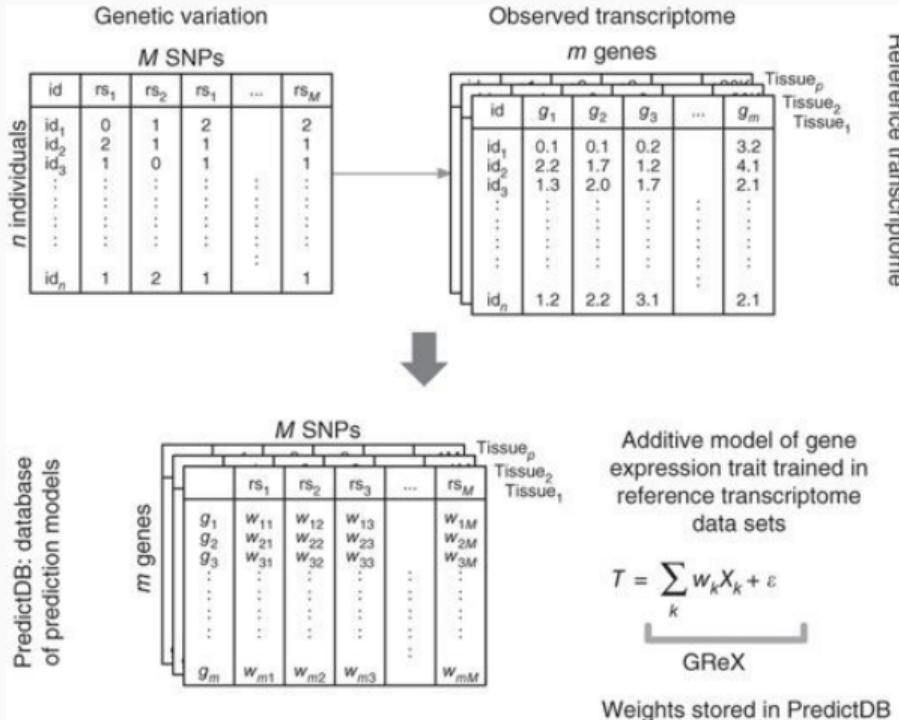
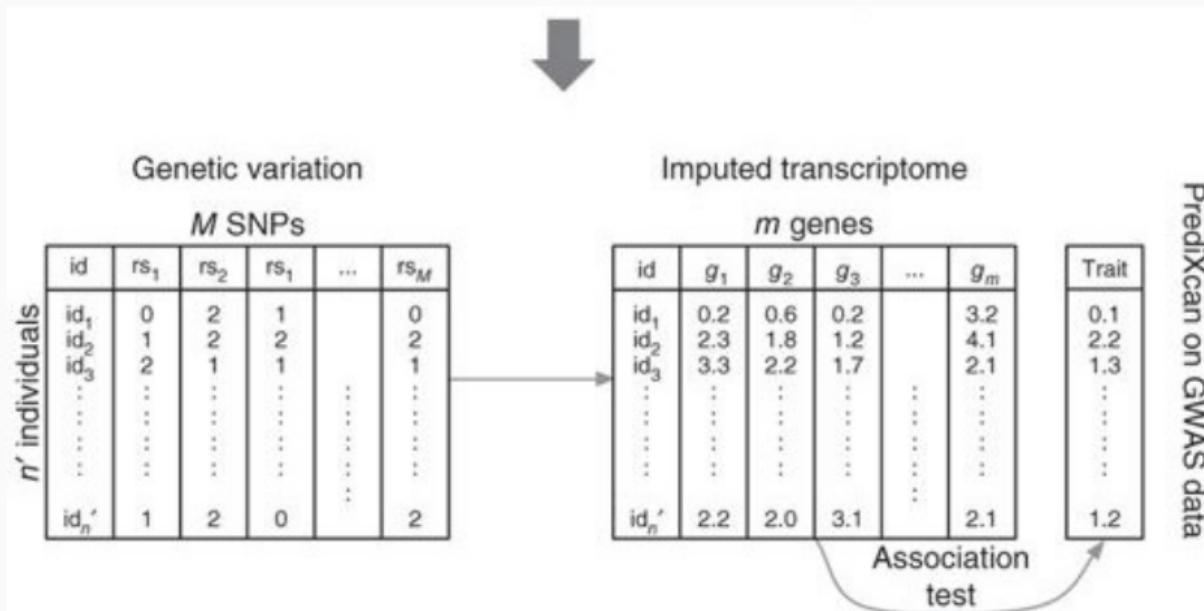


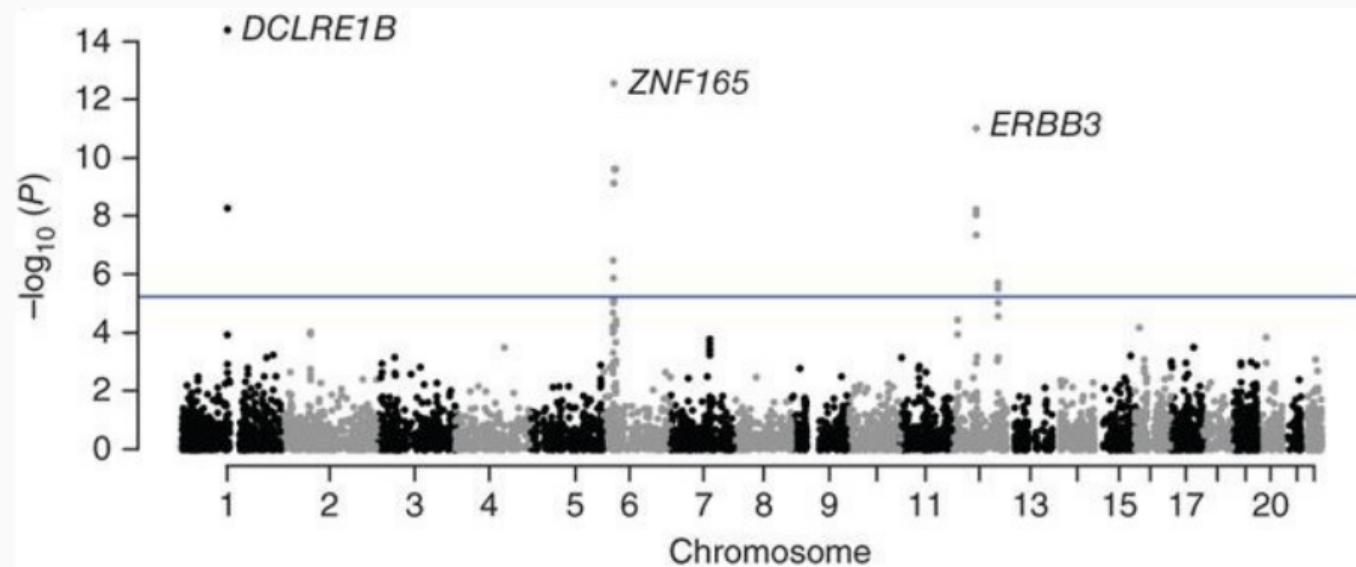
Figure 2: Training of the SNP-expression model 3/12

# Association of the imputed expression to the phenotype



**Figure 3:** The actual TWAS

## Results in T1D



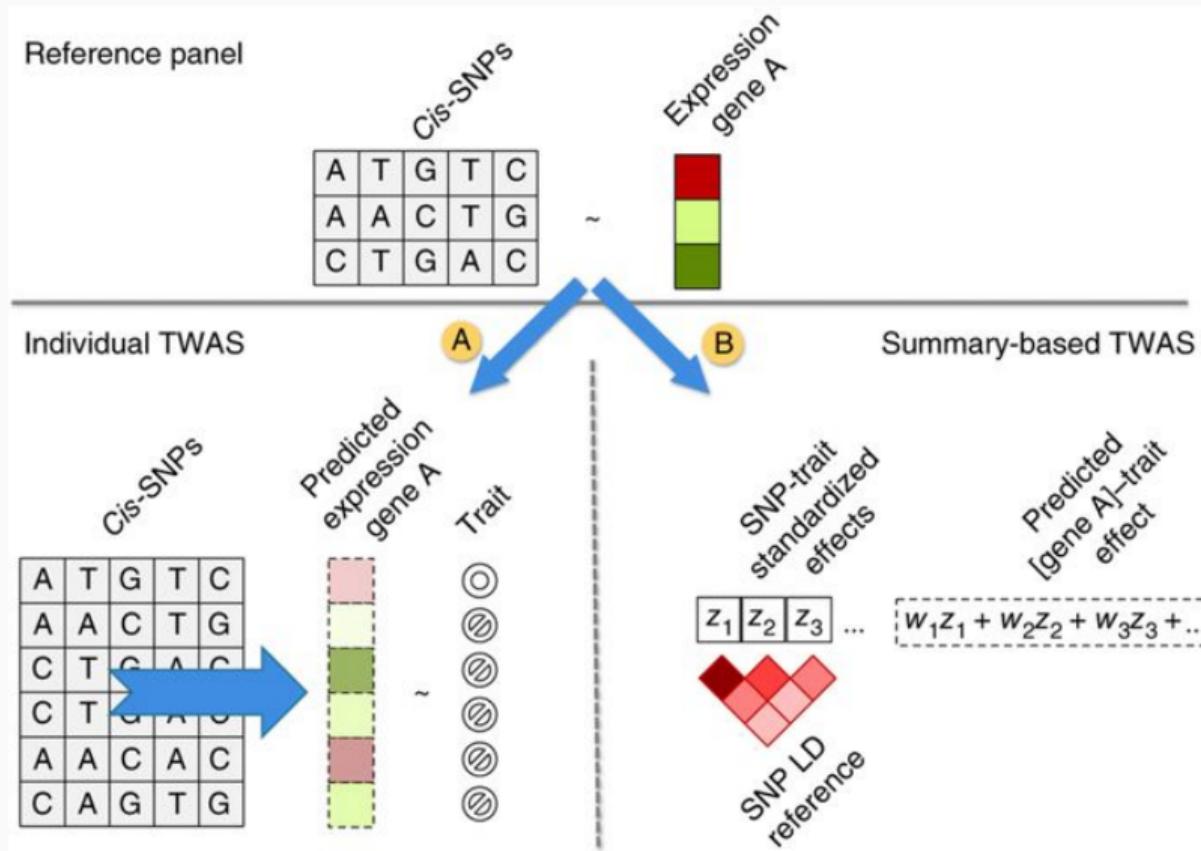
**Figure 4:** Manhattan plot of P-values for gene-disease associations

# Integrative approaches

---

Alexander Gusev et al. 'Integrative approaches for large-scale transcriptome-wide association studies'. In: *Nat. Genet.* 48.3 (Mar. 2016), pp. 245–252. DOI: [10.1038/ng.3506](https://doi.org/10.1038/ng.3506)

# From individual-level to summary-based TWAS



## Application to obesity GWAS

---

The estimation of SNP weights was performed on ~3000 individuals

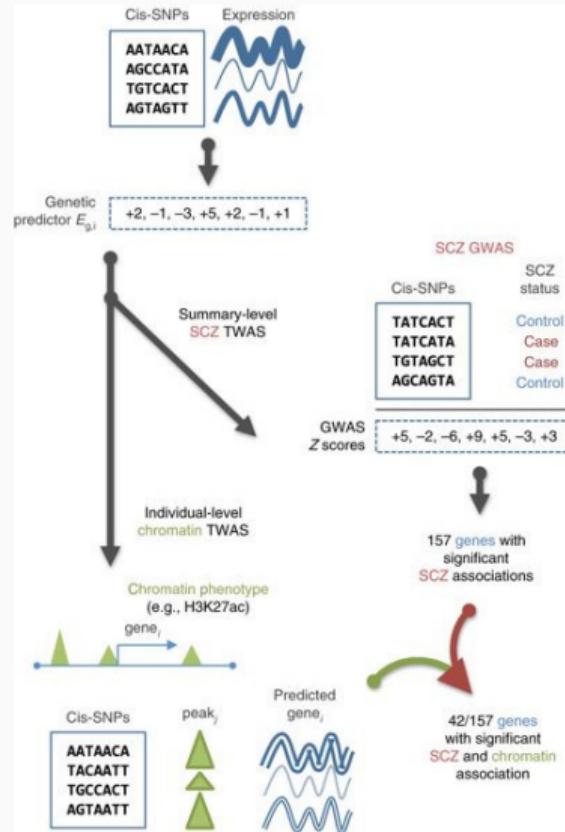
- When applied to a small-cohort GWAS, this approach found some genes that were only reported in a later large-cohort GWAS

# Moving beyond genetic variants alone

---

Alexander Gusev et al. ‘Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights’. In: *Nat. Genet.* 50.4 (Apr. 2018), pp. 538–548. DOI: [10.1038/s41588-018-0092-1](https://doi.org/10.1038/s41588-018-0092-1)

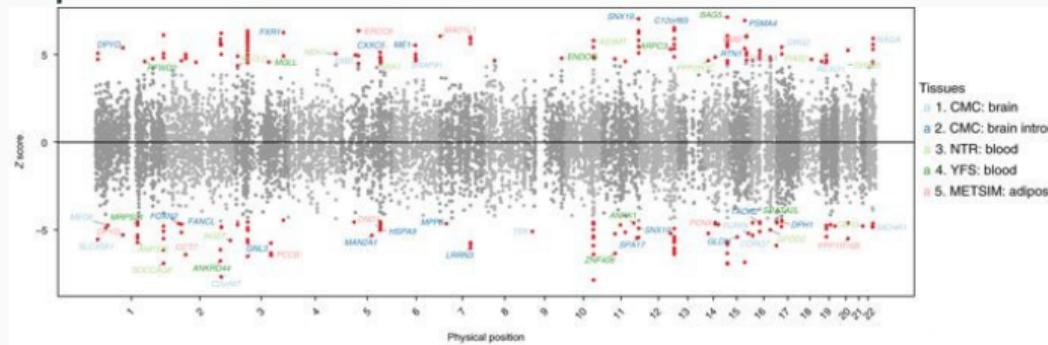
# Integrating many types of data



- Association does not imply mechanism
- Gene expression is not the only link

# A threefold pipe line

- Schizophrenia TWAS



- Chromatin TWAS

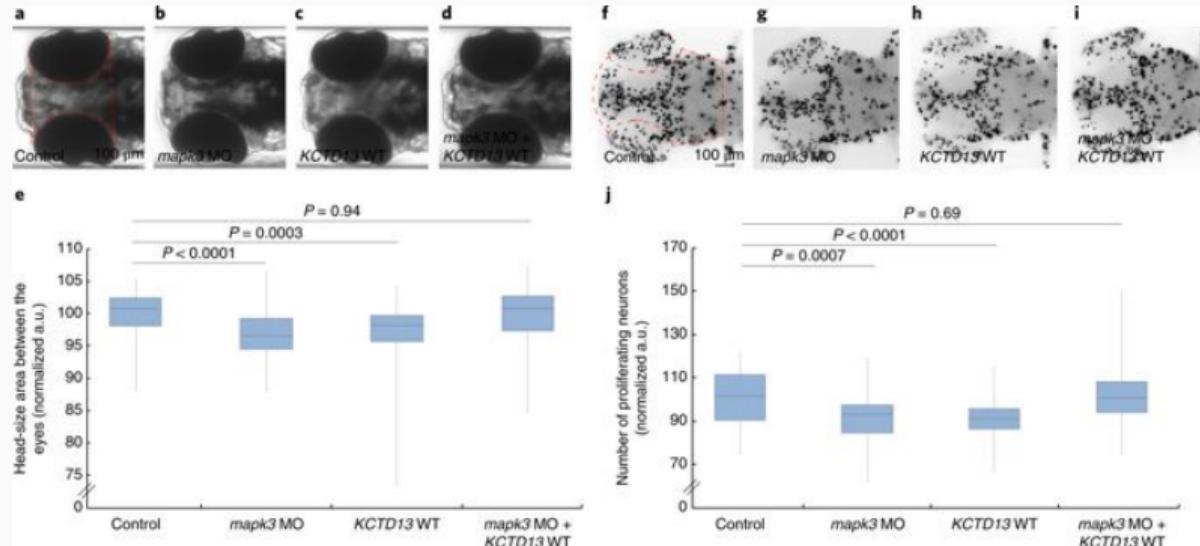
- Of the 157 SCZ-associated genes, 42 were also associated to a chromatin peak

- Schizophrenia 'Spliceome'-WAS

- 46 splicing events in the brain associated to disease

# Biological example: *MAPK3*

- *MAPK3* and *KCTD13* are coregulated
- *KCTD13* decreases proliferation, *MAPK3* is a functional trigger



**Figure 5:** Zebrafish over-expressing *KCTD13*, with or without inhibition of *MAPK3*

# Conclusions

---

TWAS address some of the limitations of classical association studies:

- Increased interpretability and druggability
- No need to assemble large cohorts

But have limitations of their own:

- If expression is not involved, TWAS do not work
- Pleiotropy cannot be modeled

## Future perspectives

---

- An association study may be performed with any intermediate molecular phenotype

## Future perspectives

---

- An association study may be performed with any intermediate molecular phenotype
- And in any population

## Future perspectives

---

- An association study may be performed with any intermediate molecular phenotype
- And in any population
- Epistasis and dominance should be taken into account

## Future perspectives

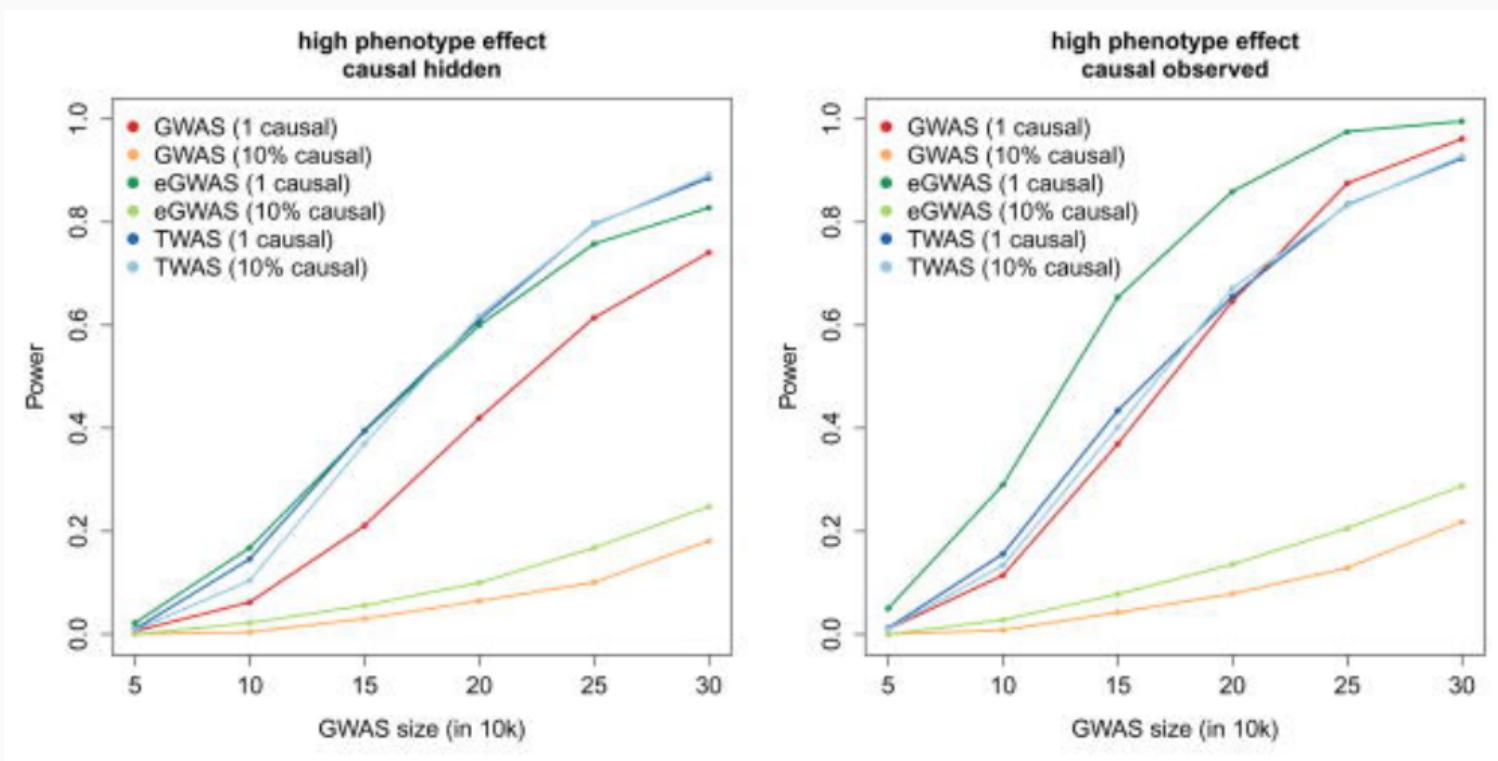
---

- An association study may be performed with any intermediate molecular phenotype
- And in any population
- Epistasis and dominance should be taken into account
- A ‘network approach’ could help to further explain the results

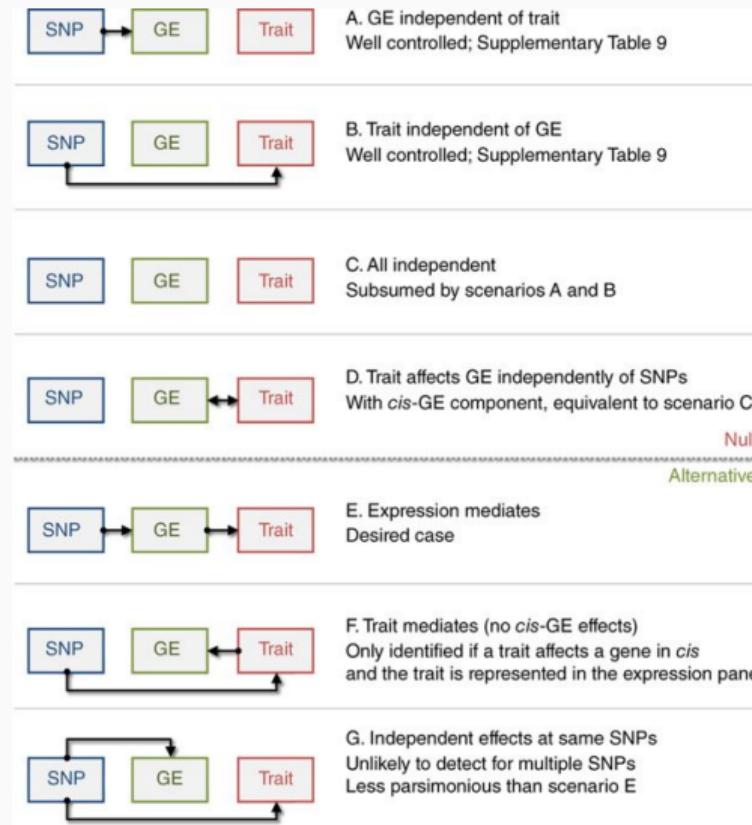
## References

-  Eric R. Gamazon et al. 'A gene-based association method for mapping traits using reference transcriptome data'. In: *Nat. Genet.* (Sept. 2015).
-  Alexander Gusev et al. 'Integrative approaches for large-scale transcriptome-wide association studies'. In: *Nat. Genet.* (Mar. 2016).
-  Alexander Gusev et al. 'Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights'. In: *Nat. Genet.* (Apr. 2018).

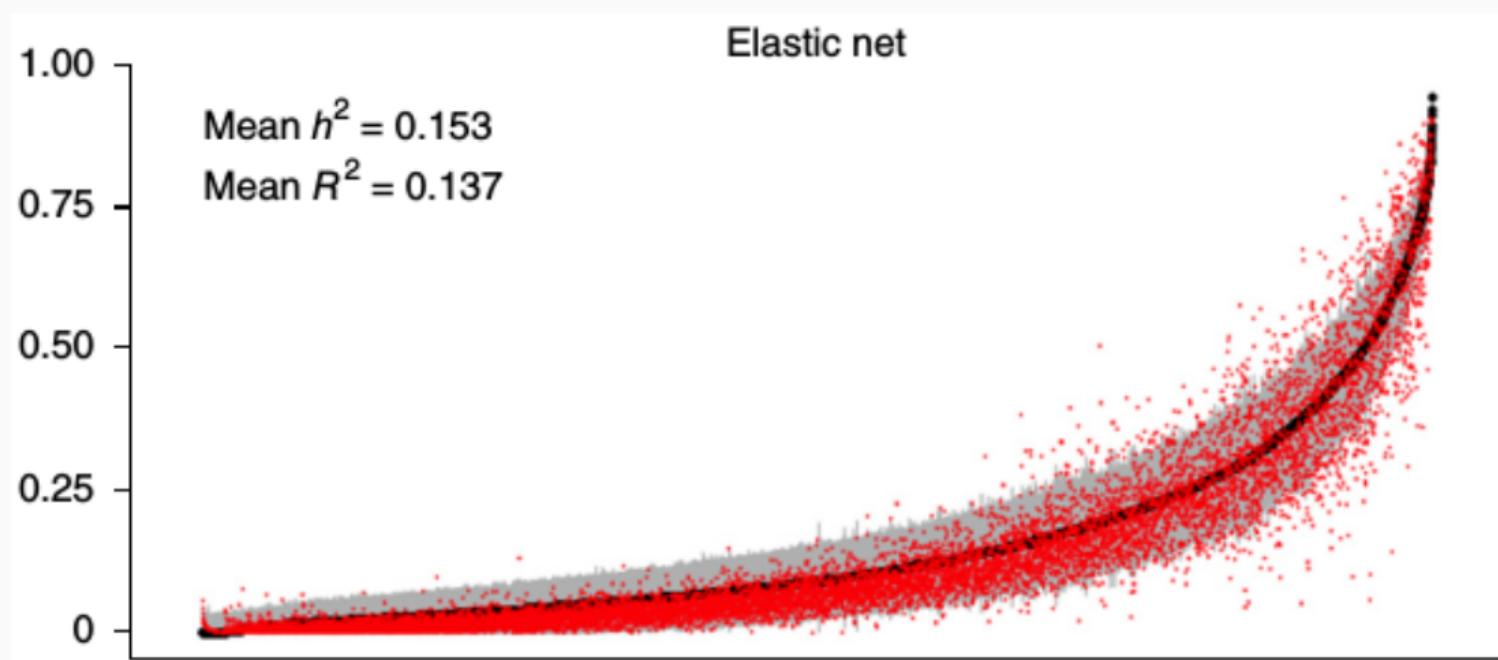
# Comparison of TWAS with classical approaches



# Causality models



# Heritability and prediction of gene expression



# Different levels of phenotype

