

# Transcriptome imputation enables drug repositioning for bipolar disorder

Dr Kévin Vervier, PhD

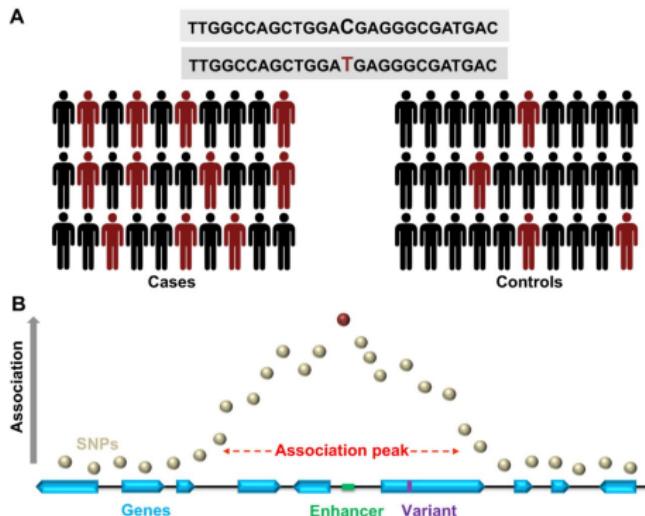
Department of Psychiatry  
University of Iowa Hospitals and Clinics

April 12, 2018



# Genetic variation and diseases

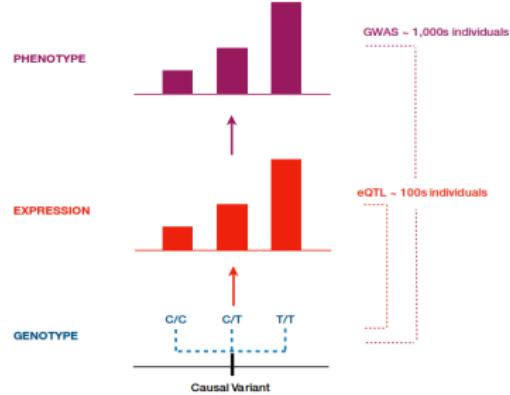
- Polygenic traits (e.g., psychiatric disorders)
- Genome-wide association study (GWAS)
- Each SNP has a small effect on explaining the trait



Source: Wangler, 2017

# Mutations, gene expression, disease

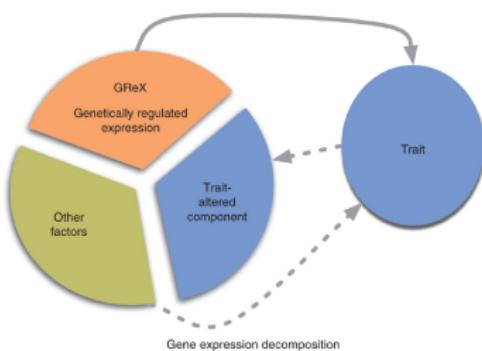
- Gene expression measures how active a gene is.
- Mutations can dysregulate expression → diseases.



- Other factors can impact measured expression (e.g. medication).

# Transcriptome-wide association studies

- TWAS [impute](#) gene expression by statistically combining variants.

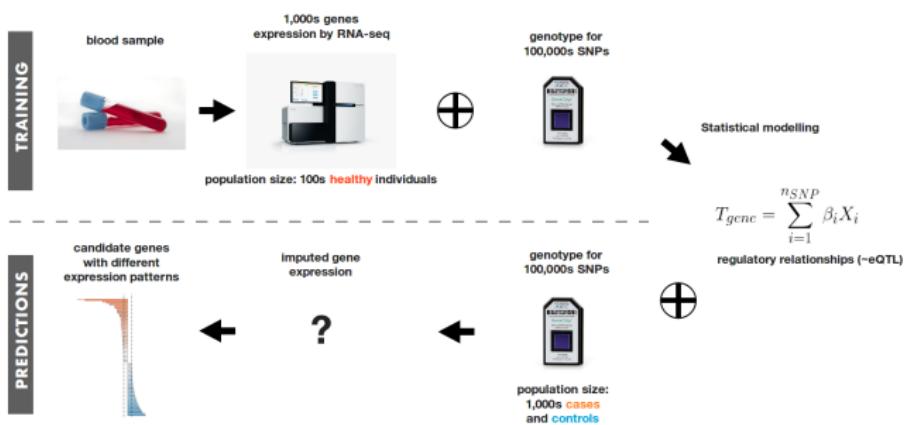


Adapted from Gamazon, 2015

- [Gene-level study](#) detects stronger association, and is easier to interpret/validate.
- Short and Long-range INfluencers of Gene Expression Regulation [SLINGER](#) (Vervier & Michaelson, 2016)

# Predictive modeling in TWAS

- Training: construct **associations** between genotype and gene expression.
- Predictions: given a **new** clinical cohort, use imputed gene expression for **disease genes identification**.

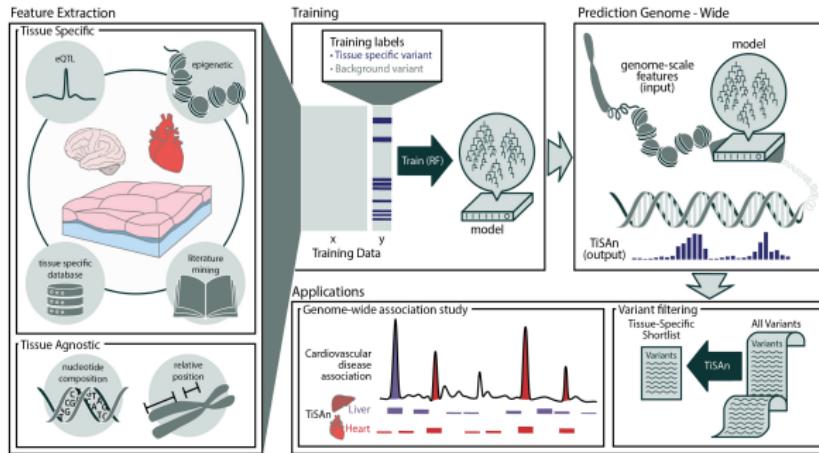


# TWAS for brain diseases: limitations

- TWAS models are trained using large [whole-blood](#) expression datasets.
- Tissue-specific eQTLs in Gene-Tissue Expression ([GTEx](#)) database.
- Small [training set](#) available for brain tissues → non-robust brain-specific models.
- Is there a way to [transfer](#) SLINGER models to brain expression prediction?

# TiSAn: tissue-specific annotation for variants

- Estimate how likely a mutation will impact different tissues.
- Machine learning and data integration of various omics sources .

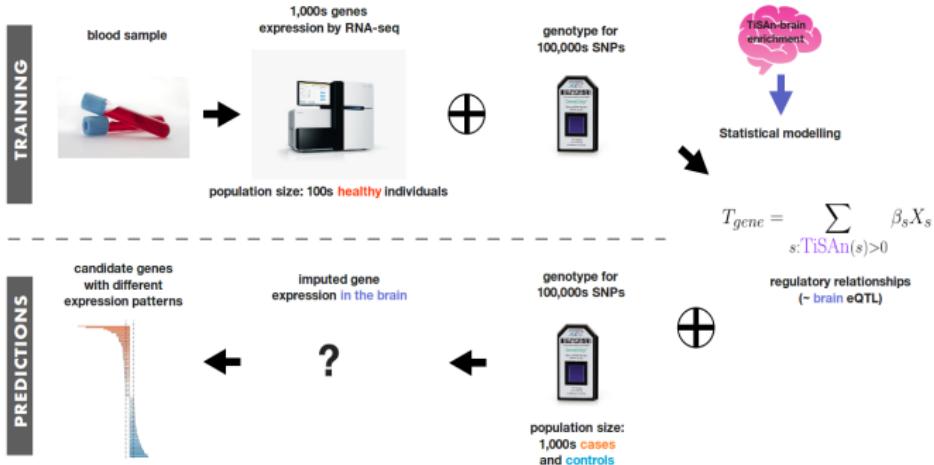


Source: Vervier & Michaelson, 2018

([github.com/kevinVervier/TiSAn](https://github.com/kevinVervier/TiSAn))

# Brain-related statistical prior in genotype

- train SLINGER models using only SNPs with enrichment in brain



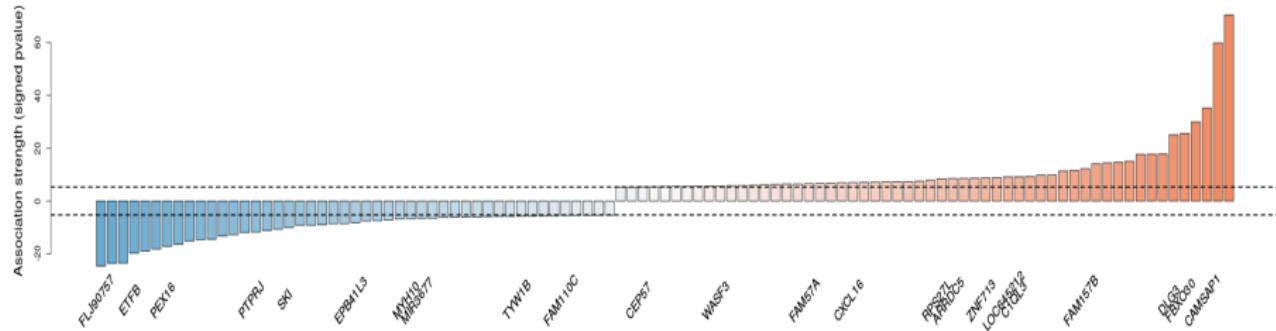
- Transfer Learning from whole-blood expression:

$$T_{total} = T_{shared} + T_{WB} + T_{other}$$

- ideally, using TiSAn-brain as a filter should remove  $T_{WB}$

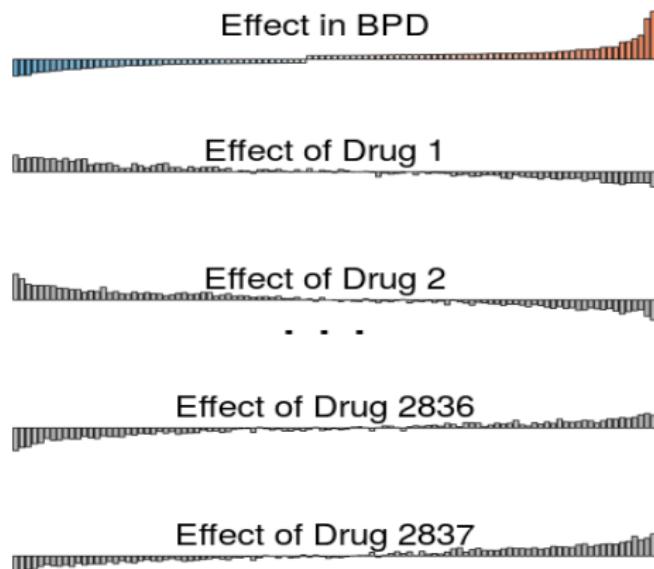
# WTCCC Bipolar disorder (BPD) case-control cohort

- Genotypes of 1,868 BPD individuals and 2,938 controls  
→ inferred transcriptome with brain-SLINGER
- Find **103 differentially expressed** genes using brain models
- Imputed **down**(47)/**up**(56) regulation:



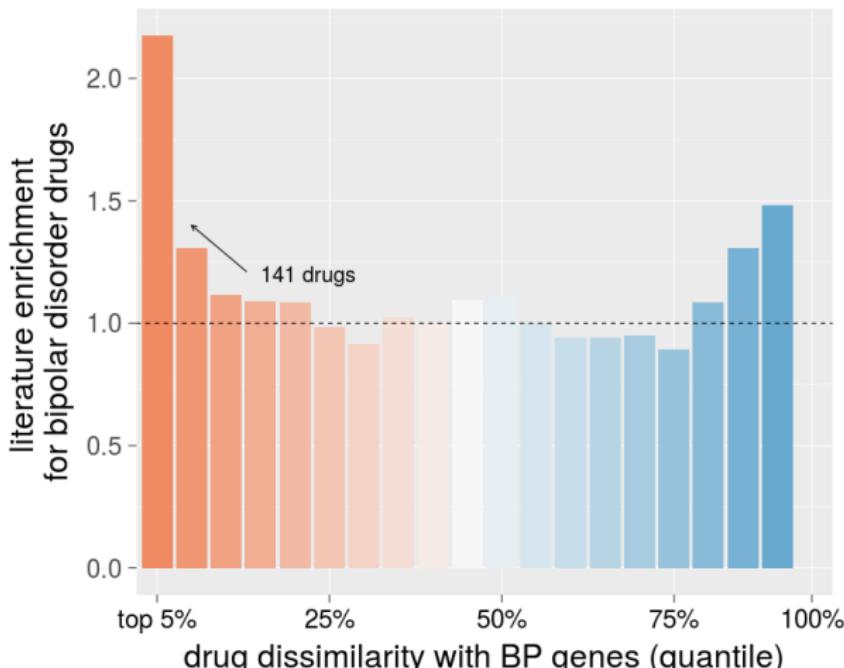
# Pharmacogenomics analysis

- Library of Integrated Cellular Signatures ([LINCS](#), Broad Institute).
- [Drug-induced change](#) in gene expression in cell lines.
- Candidate treatment profiles [rescue](#) BPD genes profile.



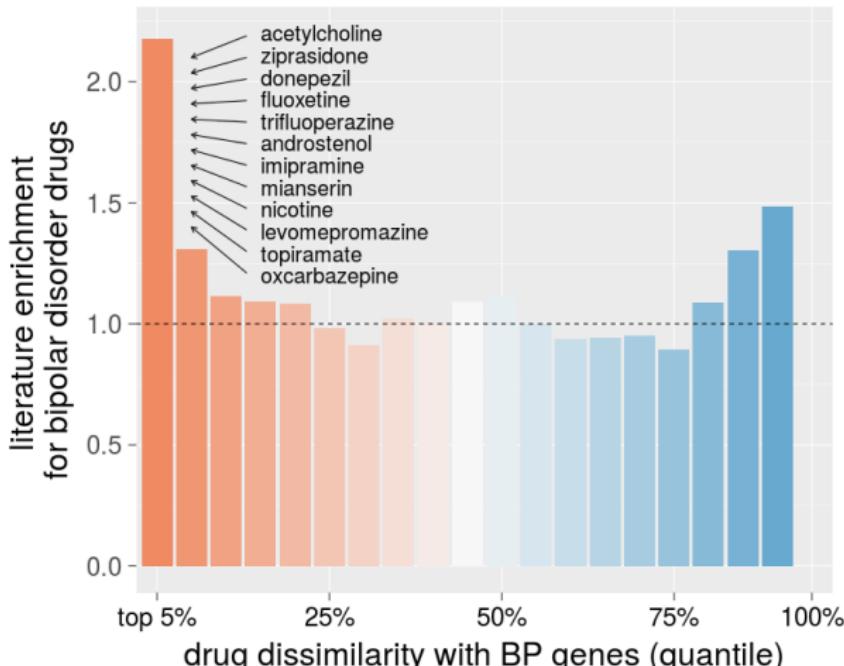
# Pharmacogenomics analysis

- Library of Integrated Cellular Signatures (LINCS, Broad Institute)
- Dissimilarity measure between drug profiles and BPD genes
- Literature enrichment for known BPD-related drugs (PubMed)



# Pharmacogenomics analysis

- Library of Integrated Cellular Signatures (LINCS, Broad Institute)
- Dissimilarity measure between drug profiles and BPD genes
- Literature enrichment for known BPD-related drugs (PubMed)



# Future directions

- From genetic variation to candidate treatments using computer science.
- Application to other disorders (e.g., eating disorder).
- Data and software are available online (Github: kevinVervier).

## SCIENTIFIC REPORTS

OPEN

**SLINGER: large-scale learning for predicting gene expression**

Kévin Vervier & Jacob J. Michaelson

**TiSAN: Estimating Tissue Specific Effects of Coding and Noncoding Variants**

Kévin Vervier<sup>1</sup> and Jacob J Michaelson<sup>1,2</sup>

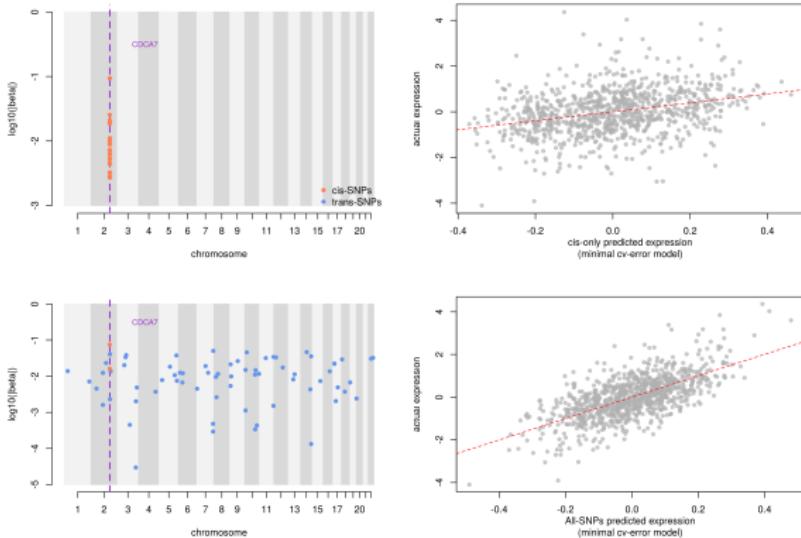
# Acknowledgements

- This work is supported by the NIH (MH105527).
- Dr Jacob Michaelson lab at UIHC.



# SLINGER: gain in accuracy

- For 4,133 genes, we improved by at least 10% its model accuracy.
- Performance are reported for *CDCA7*, found to be significantly associated with Type 1 diabetes.



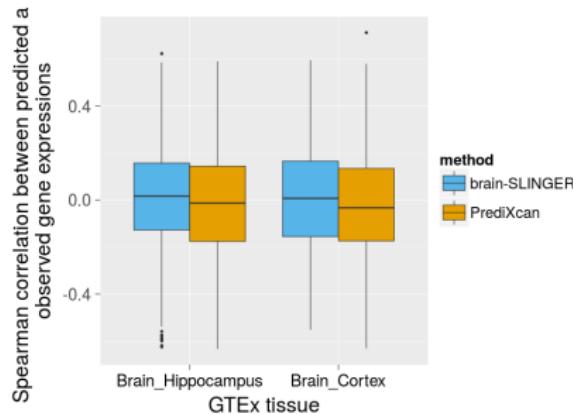
Source: Vervier *et al.*, 2016 (Fig.1)

## Predictable genes enrichment

- 1,337 gene models estimable only with brain-SLINGER.
- GO term enrichment in neurological system process ( $P = 4 \times 10^{-5}$ ), signal transducer activity ( $P = 3 \times 10^{-5}$ ) and receptor activity ( $P = 7 \times 10^{-6}$ ).
- 3,549 genes predicted in SLINGER were lost in the process, but no GO term enrichment.
- SLINGER models for lost genes are less accurate than average models ( $P < 2 \times 10^{-16}$ , average  $R^2$  values: 0.075 versus 0.11).
- brain-SLINGER models are significantly enriched in *trans* signal, supporting tissue-specificity (Liu, 2017)

# Comparison with models trained on brain expression

- PrediXcan provides models trained on GTEx brain samples.
- on ~ 2,000 genes, we found no difference between PRX and brain-SLINGER for 8/10 regions.
- for [2 GTEx regions](#), we found significantly higher accuracy in brain-SLINGER



# GTEx expression in multiple brain regions

- for each GTEx brain region, infer gene expression using SLINGER and brain-SLINGER.
- get correlation with measured GTEx expression
- compare genes with higher correlation in SLINGER to the ones better predicted by brain-SLINGER.

GTEx tissue	GO term	corrected pvalue
Anterior cingulate cortex (BA24)	regulation of axogenesis	5.13E-03
Caudate basal ganglia	regulation of histone acetylation	1.77E-02
Cerebellar hemisphere	muscle contraction	4.56E-03
Cerebellar	cell migration	1.15E-03
Cortex	transmembrane receptor	3.42E-02
Frontal cortex	cell differentiation	1.59E-02
Hippocampus	neuron differentiation	1.64E-02
Hypothalamus	positive regulation of neurogenesis	8.84E-06
Putamen basal ganglia	calcium channel activity	4.10E-04

**Table :** GO term enrichment for the genes with higher correlation with brain-models.