Transcriptome-wide association study of schizophrenia and chromatin phenotypes

Alexander (Sasha) Gusev

Harvard TH Chan School of Public Health
Challenges of studying complex disease

- **Goal**: specific mechanistic hypotheses.

- Schizophrenia is highly polygenic, with up to 70% of 1MB regions harboring an association.

- For an associated SNP, causal mechanism often not through nearest gene [...] or mediated by CNV / TFBS / etc

- Relevant molecular features (RNA-seq/ChIP-seq in brain) are difficult to collect.

Methods for molecular + GWAS data:

**Integrative approaches for large-scale transcriptome-wide association studies**

Alexander Gusev¹-³, Arthur Ko⁴,⁵, Huwenbo Shi⁶, Gaurav Bhatia¹-³, Wonil Chung¹, Brenda W J H Penninx⁷, Rick Jansen⁷, Eco J C de Geus⁸, Dorret I Boomsma⁸, Fred A Wright⁹, Patrick F Sullivan¹⁰-¹², Elina Nikkola⁴, Marcus Alvarez⁴, Mete Civelek¹³, Aldons J Lusis⁴,¹³, Terho Lehtimäki¹⁴, Emma Raitoharju¹⁴, Mika Kähönen¹⁵, Ilkka Seppälä¹⁴, Olli T Raitakari¹⁶,¹⁷, Johanna Kuusisto¹⁸, Markku Laakso¹⁸, Alkes L Price¹-³, Päivi Pajukanta⁴,⁵ & Bogdan Pasaniuc⁴,⁶,¹⁹

**Gusev et al. 2016 Nat. Genet.**
Outline

1. **TWAS**: [cis] molecular features can be predicted into un-assayed samples

2. TWAS to find *expression* associated with schizophrenia

3. TWAS to find *expression* associated with ChIP-Seq *chromatin activity*
1. **TWAS**: [cis] molecular features can be predicted into un-assayed samples

2. TWAS to find expression associated with schizophrenia

3. TWAS to find expression associated with ChIP-Seq chromatin activity
TWAS using individual-level data

expression reference

<table>
<thead>
<tr>
<th>cis-SNPs</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>T</td>
</tr>
</tbody>
</table>

Linear genetic predictor

\[ \beta_1 \beta_2 \beta_3 \beta_4 \beta_5 \]

60-80% prediction accuracy

individual GWAS data

<table>
<thead>
<tr>
<th>cis-SNPs</th>
<th>Predicted Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>A</td>
</tr>
</tbody>
</table>

Gamazon et al. 2015 Nat Genet; Gusev et al. 2016 Nat Genet
TWAS using summary-level data

**expression reference**

<table>
<thead>
<tr>
<th>cis-SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A T G T C</td>
</tr>
<tr>
<td>A A C T G</td>
</tr>
<tr>
<td>C T G A C</td>
</tr>
</tbody>
</table>

Linear genetic predictor

\[ \beta_1 \beta_2 \beta_3 \beta_4 \beta_5 \]

60-80% prediction accuracy

**summary GWAS data**

SNP-trait association

Predicted Gene-Trait association

\[ R^2 > 0.9 \text{ w/ individual-level prediction} \]

Gusev et al. 2016 Nat Genet; Barbeira et al. biorxiv
Outline

1. TWAS: [cis] molecular features can be predicted into un-assayed samples

2. TWAS to find expression associated with schizophrenia

3. TWAS to find expression associated with ChIP-Seq chromatin activity
TWAS for schizophrenia

4 Expression Reference Panels

<table>
<thead>
<tr>
<th>A</th>
<th>T</th>
<th>G</th>
<th>T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>C</td>
<td>T</td>
<td>G</td>
<td>A</td>
<td>C</td>
</tr>
</tbody>
</table>

CommonMind Consortium (CMC)¹: 500 brain RNA-seq alternative splice variants²

Netherlands Twin Registry (NTR)³: 1,200 blood array
METabolic Syndrome In Men (METSIM): 600 adipose RNA-seq
Young Finns Study (YFS): 1,300 blood array

GWAS: Psychiatric Genetics Consortium
~80,000 SCZ samples
[PGC 2014 Nature]

TWAS for schizophrenia

157 gene associations (45 in novel loci)
80 splice-variant associations
conditioning on expression explains all genome-wide significant effect at known loci
polygenic effects across thousands of genes

polygenic score using predicted expression, evaluated in independent SCZ cohort

\( R^2 \) w/ SCZ status

P-value threshold

brain

0.00
0.01
0.02
0.03
0.04
0.05

CMC/brain genes

P-value threshold

liability−scale R2

0.00
0.01
0.02
0.03
0.04
0.05
more associations with brain and splicing

polygenic score using predicted expression, evaluated in independent SCZ cohort

\[ R^2 \text{ w/ SCZ status} \]

all predicted cis-expression explain 26% of total SCZ SNP-heritability (upper bound on mediated effect)
Outline

1. TWAS: [cis] molecular features can be predicted into un-assayed samples

2. TWAS to find expression associated with schizophrenia

3. TWAS to find expression associated with ChIP-Seq chromatin activity
Chromatin variation is under genetic control

[YRI] Grubert et al. 2015 Cell:
Chromatin variation and expression under shared genetic control,

[CEU] Waszak et al. 2015 Cell:
Local chromatin-expression interactions form modules.
connecting genes to chromatin peaks with top hits

for every gene:
expression - QTLs

for every peak
chromatin - QTLs
TWAS for expression-chromatinin associations

cis-SNPs

<table>
<thead>
<tr>
<th>A</th>
<th>T</th>
<th>G</th>
<th>T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>C</td>
<td>T</td>
<td>G</td>
<td>A</td>
<td>C</td>
</tr>
</tbody>
</table>

Expression

\[ \text{expression reference} \]

linear genetic predictor

\[ \beta_1 \beta_2 \beta_3 \beta_4 \beta_5 \]

individual epigenomic data

Predicted Expression

chromatin activity

\[ \sim \]
TWAS identifies 10x more gene/chromatin associations than eQTL overlap approach.

**chromatin-TWAS**

- **H3K4ME1**
- **H3K4ME3**
- **H3K27AC**
- **DHS**

**overlapping top hits**

- **H3K4ME1**
- **H3K4ME3**
- **H3K27AC**
- **DHS**

The plots show the number of significant associations as a function of distance to the gene TSS.
Validating TWAS gene-chromatin associations

**Discovery:**
TWAS gene–chromatin associations

brain/blood

*reference*

expression
Validating TWAS gene-chromatin associations

Discovery:
TWAS gene–chromatin associations

Replication:
measured LCL RNA-seq expression in chromatin samples

brain/blood
reference
target
expression
expression
Chromatin variation is highly associated with expression at predicted gene-peak associations.

Replication $R^2 \approx$ total expression cis-$h_g^2$

ChIP-seq and expression from Waszak et al. 2015 Cell
Enrichment for chromatin associations with SCZ TWAS genes

157 SCZ TWAS genes
42 with chromatin TWAS associations
4x background ($P=1\times10^{-11}$)

significant enrichment for SCZ splice variants with chromatin associations
TWAS provides mechanistic hypotheses

N=80,000

108 SCZ loci
TWAS provides mechanistic hypotheses

- N=80,000

108 SCZ loci

- 26/108 have TWAS splicing association
- 48/108 have TWAS splicing or gene association (76% not nearest gene)

- N=3,500
TWAS provides mechanistic hypotheses

- N=80,000
- 108 SCZ loci
  - 26/108 have TWAS splicing association
  - 48/108 have TWAS splicing or gene association (76% not nearest gene)
  - 34/48 also have chromatin association (85% not in the promoter)

- N=3,500
- N=150
Conclusion / Future Work

• TWAS implicates specific genes and their regulators.

• Epigenetic mediators can be a common disease mechanism

• Yields zozens of mechanistic hypotheses for schizophrenia loci (following up with experimental validation)
Nick Mancuso  Gregory Crawford  Psychiatric Genomics Consortium
Hilary Finucane  Nicholas Katsanis  CommonMind Consortium
Yakir Reshef  Patrick F Sullivan  PsychENCODE
Lingyun Song  Bogdan Pasaniuc  1000 Genomes
Alexias Safi  Alkes L Price
Edwin Oh

Steve McCarroll
Ben Neale
Mick O’Donovan
Roel Ophoff

recruiting in new lab at Dana Farber Cancer Institute & Harvard Medical School

pre-print on biorxiv

gusevlab.org