Extended

Mode

Autism dat

Schizophren

data

.

Methods

Results:includeEXAC

Risk gene Intersect v

gene sets

Intersect wi

May, 2015

May 2, 2016

Autism dat

Schizophre

Data Methods Results:includeEXAC Risk genes

Intersect with gene sets
Results:NotEXAC
Risk genes
Intersect with

- Review exTADA (Transmission And De novo Association) model.
- Test exTADA on autism data.
- Apply exTADA to schizophrenia data.
  - Estimate the proportion of risk genes.
  - Test results on gene sets.

Extende TADA Model

Autism dat

Schizophre

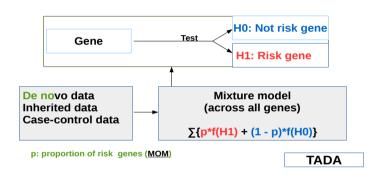
Data

Methods
Results:includeEXAC

Intersect with gene sets

Results:NotEXA

Risk genes Intersect with gene sets



PLoS Genet. 2013;9(8):e1003671. doi: 10.1371/journal.pgen.1003671. Epub 2013 Aug 15.

Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes.

He X1, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS, Schellenberg GD, Gibbs RA, Daly MJ, Buxbaum JD, State MW, Devlin B, Roeder K.

### **Original TADA**

- Use only LoF de novo mutations => the proportion of risk genes  $(\pi)$  \*\*
- Use known risk genes AND \*\* to infer other information of other categories.

#### **Extended TADA**

- Estimate simultaneously all parameters of all annotations (e.g., LoF, missense damaging) including  $\pi$
- Do not use known risk genes (it can be used, but not necessary).

## **ExTADA**

Review TADA

Extend TADA Model

Autism data

Schizophrer

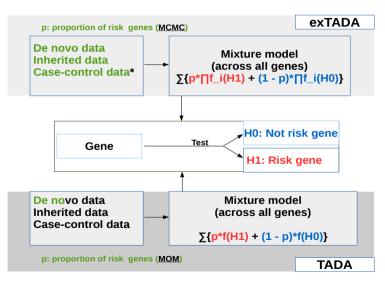
Data

Methods
Results:includeEXAC

Risk genes
Intersect with

gene sets Results:NotEXA

Risk genes
Intersect with



Model

Results:includeEXAC

#### Main work

- De novo mutations: the same as original TADA.
- Inherited/Case-control: use an approximate model in the estimation process 1.
- Estimate all parameters using a MCMC method (known risk genes are not necessary).

Model: Internal product

$$P(\mathbf{x} | \textit{parameters}) = \prod_{i=1}^{m} \left[ \pi P(\mathbf{x}_{i_{LOF}} | H_1) P(\mathbf{x}_{i_{mis3}} | H_1) + (1 - \pi) P(\mathbf{x}_{i_{LOF}} | H_0) P(\mathbf{x}_{i_{mis3}} | H_0) \right]$$



<sup>&</sup>lt;sup>1</sup>Idea of changing case-control model is from Xin He

#### Autism data

Schizophre

Data Mathods

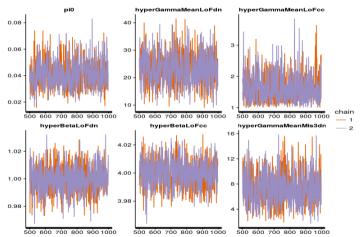
Methods
Results:includeEXAC
Risk genes

Intersect with gene sets
Results:NotEXA

Risk genes
Intersect wit

## Autism Data

Use **non-information priors** => similar results as TADA (based on known risk genes)



# Schizophrenia data

Review TADA

Extended TADA Model

Autism data

data Data Methods

Methods
Results:includeEXAC
Risk genes
Intersect with
gene sets
Results: NotEXAC

Results:NotEXA Risk genes Intersect with gene sets Sample sizes from different studies.

| Source                 | De novo | Non/Transmitted | Case | Control |
|------------------------|---------|-----------------|------|---------|
| Fromer et al. (2014)   | 617     | 617             |      |         |
| Girard et al. (2011)   | 14      |                 |      |         |
| Gulsuner et al. (2013) | 105     |                 |      |         |
| McCarthy et al. (2014) | 57      |                 |      |         |
| Xu et al. (2012)       | 231     |                 |      |         |
| Giulio et al. (2016)   |         |                 | 4954 | 6239    |
| Total                  | 1024    | 617             | 4954 | 6239    |

Autism data

Schizophrenia
data
Data
Methods
Results:includeEXAC
Risk genes
Intersect with
gene sets

#### Focused on:

• De novo (DN) + Transmitted (Trans) + Case-Control (CC).

#### Also tested:

- De novo (DN) + Transmitted (Trans).
- De novo (DN) + Case-Control (CC).
- De novo (DN).

Categories: LoF and missense damaging (7 methods from Giulio). Private (Not in Exac) or Non-private (include Exac).

TADA Model

Autism dat

Schizophre

Data
Methods
Results:includeEXAC
Risk genes
Intersect with
gene sets
Results:NotEXAC
Risk genes

Previous studies (The unseen species problem):

- Xu et al (2012): **868 genes** based on 231 parent-proband trios and 34 unaffected trios.
- Fromer et al (2014): 4000 to 12000 genes based on 623 schizophrenia trios (use LoF and NS mutations).

### How many risk genes from exTADA?

Singleton data + NOT private.

Review TADA

Extende TADA Model

Autism dat

Schizophre

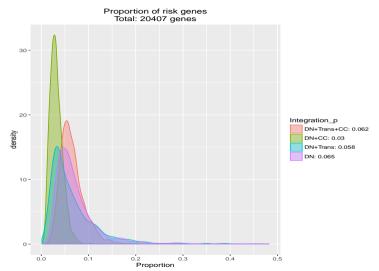
Data

Results:includeEXAC

Risk genes Intersect with

gene sets Results:NotEXA

Risk genes
Intersect with
gene sets



# Singleton

Review TADA

Extendary
TADA

Model

Autism dat

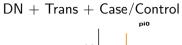
data

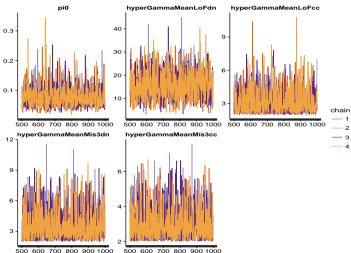
Methods Results:includeEXAC

Risk genes Intersect wi

gene sets Results:NotEX

Risk genes Intersect with gene sets





Extend TADA Model

Autism dat

Schizophreni data

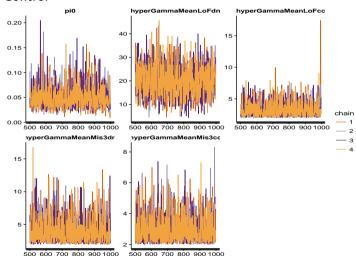
Data
Methods
Results:includeEXAC

Risk genes

gene sets
Results:NotEX

Risk genes
Intersect with

### DN + Case/Control



Autism dat

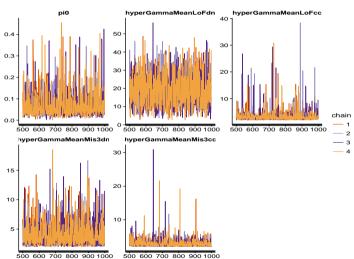
Schizophren

Data Mothods

Results:includeEXAC Risk genes

gene sets

Results:NotEX/ Risk genes DN + Trans



Autism dat

Schizophrer

Data

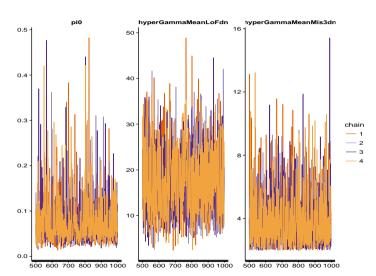
Results:includeEXAC

Risk genes

gene sets

Results:NotEX

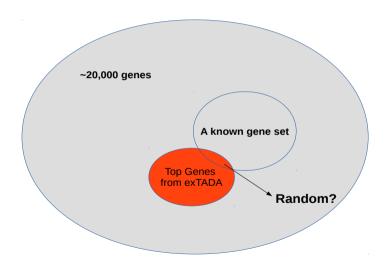
Intersect wit



Results:includeEXAC

Intersect with

## Intersect with known gene sets



## Calculate p value for each gene set (N genes)

- Count the number of genes overlapping between the *IGene* genes and the gene set, nG.
- For i from 1 to K (times)
  - Randomly choose a set of N genes from all genes (>20000 genes).
  - Count the number of genes overlapping between the *IGene* genes and the random gene set,  $M_i$ .
  - pValue = (length(vM[vM >= nG]) + 1)/(K + 1)with  $vM = c(M_1, M_2, ...M_K)$

Extende TADA Model

Autism data

#### Schizophren

Data Methods

Results:includeEXAC Risk genes

Intersect with gene sets

Results:NotEXA

Risk genes

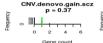
### DN + Trans + CC









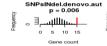




Gene count





























Extended TADA Model

Autism data

#### Schizophren

Data

Methods
Results:includeEXAC

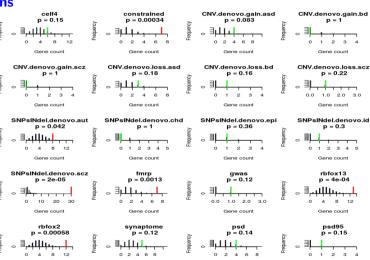
Risk genes Intersect with

gene sets
Results:NotEXA

Risk genes
Intersect with

## **DN** + Trans

Gene count



Gene count

Gene count

Gene count

#### Raviou TADA

Extende TADA Model

Autism data

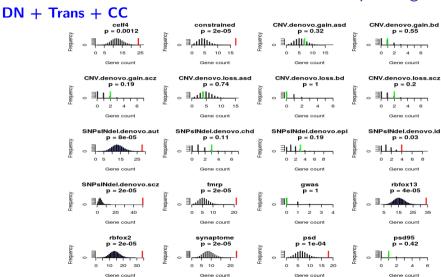
### Schizophren

Data Methods

Results:includeEXAC
Risk genes
Intersect with

Intersect with gene sets Results:NotEXAC

Risk genes
Intersect with





Autism data

#### Schizophren

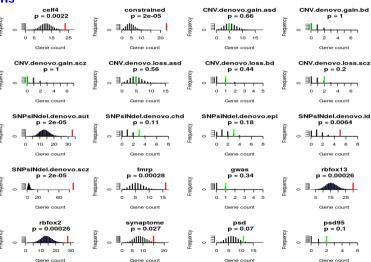
Data Methods

Results:includeEXAC Risk genes

Intersect with gene sets Results:NotEXAC

Risk genes Intersect with gene sets

### DN + Trans

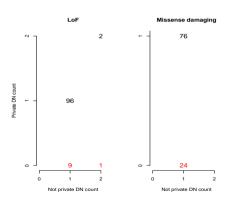


Results:includeFXAC

Results: NotEXAC

 Remove all variants in EXAC (= Private variants) => lose 1 LoF double-hit gene.

Use the same mutation rates. <sup>2</sup>



<sup>&</sup>lt;sup>2</sup>Re-calculating mutation rates by removing all Exac variants ←□ → ←② → ←② → ←② → ◆② → ◆② ←



## Not EXAC

Review TADA

Extende TADA Model

Autism data

Schizophrer

data

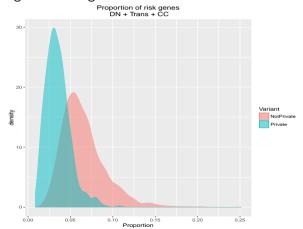
Methods Results:includeEXAC

Risk genes Intersect with

Results:NotEXA
Risk genes

Intersect with

## Proportion of risk genes not high



## Not EXAC

Review TADA

Extended

TADA

Autism data

Schizophreni data

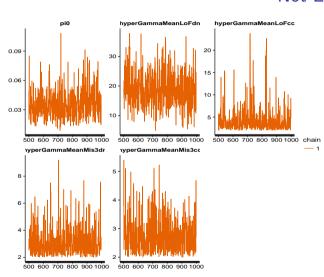
Data

Results:includeEXAC

Intersect with gene sets

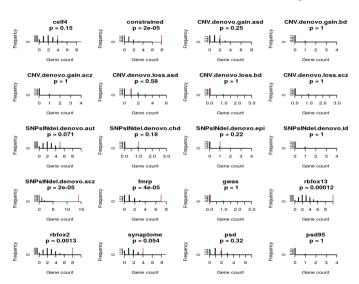
Results:NotEXA

Intersect wit



Results:includeFXAC

Intersect with gene sets



#### Daudani TADA

Extende TADA

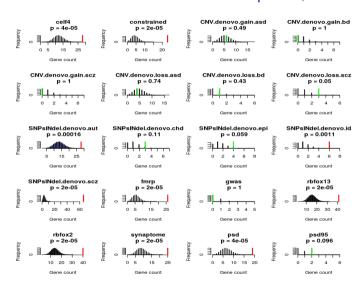
Autism data

#### Schizophren

data Data Methods Results:includeEXAC Risk genes

Intersect with gene sets
Results:NotEXA

Intersect with



Extende TADA

Autism dat

Schizophre

data

Data Methods

Results:includeEXAC

Intersect with gene sets

Results:NotEX/ Risk genes

Intersect with gene sets

# Private and Not private

Overlapping genes with different thresholds:



Intersect with gene sets
Results:NotEXA

Intersect with gene sets

Overlapping genes: FDR < 0.1

Both LoF de novos of TAF13 are in Exac variants.

| Gene   | NotPrivate | Private |  |
|--------|------------|---------|--|
| ADCY6  | 1          | 0       |  |
| BLNK   | 1          | 0       |  |
| EPHA5  | 1          | 0       |  |
| HEATR2 | 1          | 0       |  |
| MARK4  | 1          | 0       |  |
| MPO    | 1          | 0       |  |
| PRRC2A | 1          | 0       |  |
| ROBO1  | 1          | 0       |  |
| TAF13  | 1          | 0       |  |
| RB1CC1 | 1          | 1       |  |
| SETD1A | 1          | 1       |  |

Autism dat

data Data Mathada

Results:includeEXAC
Risk genes
Intersect with

Intersect with gene sets
Results:NotEXA

Intersect with gene sets

## Working on:

- Simulation data.
- Private variants with new mutation rates.

Should test other gene sets?

THANK YOU!!!!!