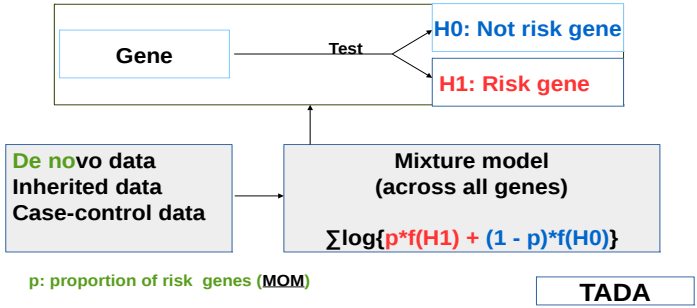
May, 2015

May 3, 2016

Meeting

- 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.

TADA

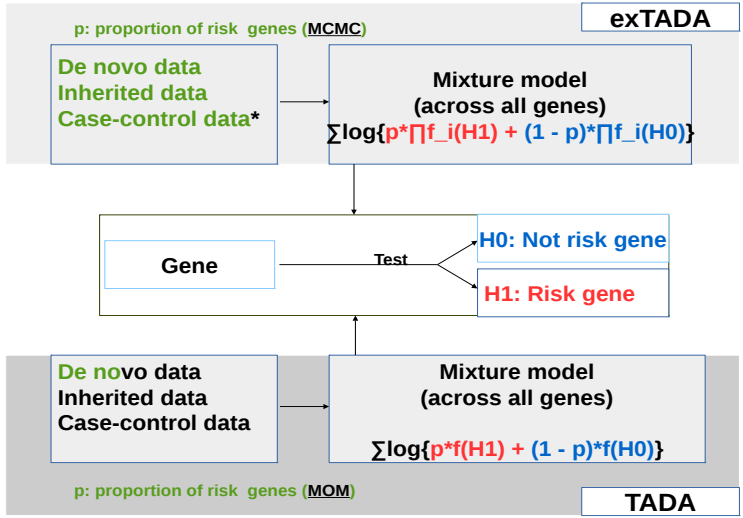


PLoS Genet. 2013;9(8):e1003671. doi: 10.1371/journal.pgen.1003671. Epub 2013 Aug 15.

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

He X[1], 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.

ExTADA



Extended TADA

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).**

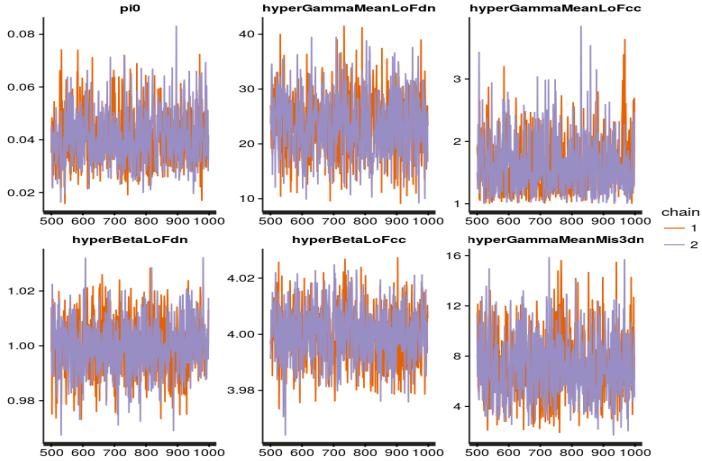
Model: **Internal product**

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

[1] Idea of changing case-control model is from Xin He

Autism Data

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



Schizophrenia data

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

Methods

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).

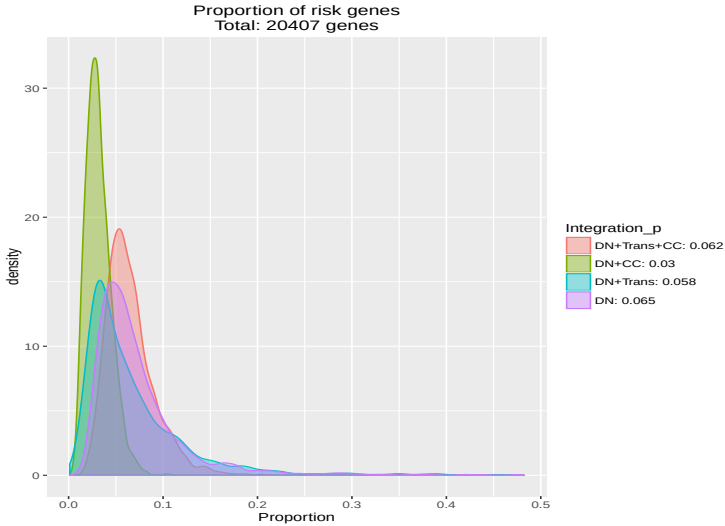
How many risk genes

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).

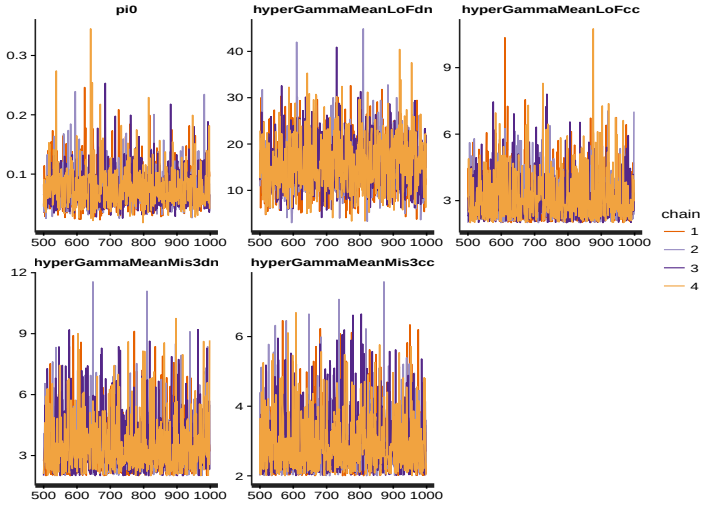
How many risk genes from exTADA?

Singleton data + NOT private.

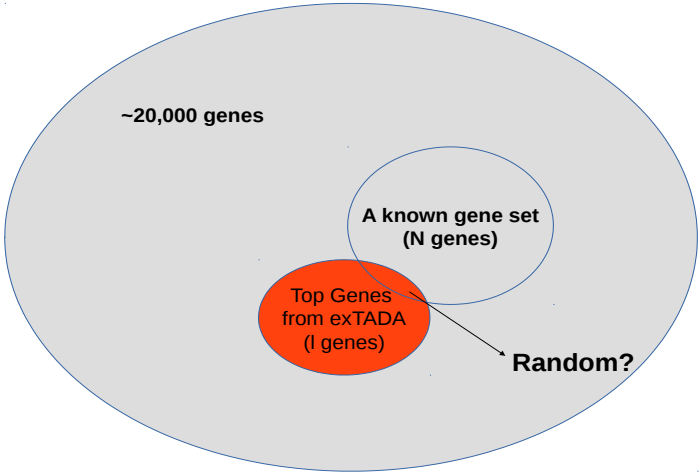


Singleton

DN + Trans + Case/Control



Intersect with known gene sets



Top genes from exTADA

1. Calculate Bayes factor ($\frac{p(data|H_1)}{p(data|H_0)}$) for each gene.
2. Use Bayes factors to calculate false discovery rates (FDRs).
3. Arrange genes by FDRs.
4. **Top Genes**: use a threshold for FDR or **top genes with smallest FDRs**.

	FDR < 0.1	FDR < 0.3
DN+Trans	5	30
DN + Trans + CC	11	56

Known gene sets: the same gene sets of Giulio et al.

=> Can we see similar results between exTATA (DN + Trans) and Giulio et al (CC).

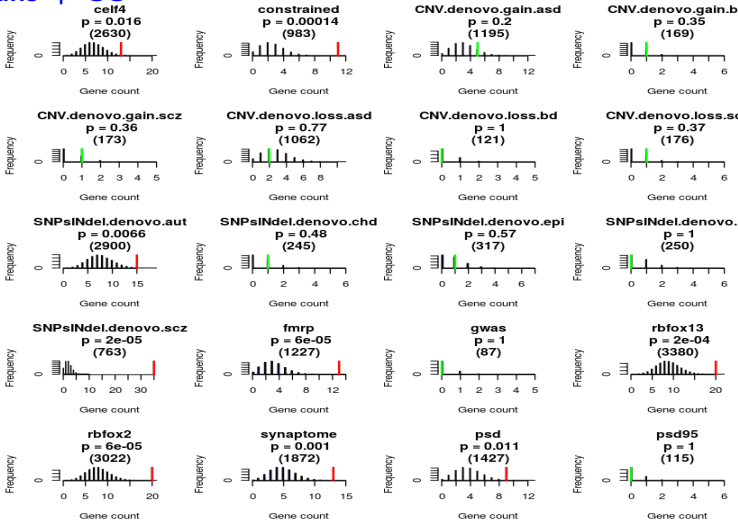
Intersect with known gene sets

Calculate p value for each gene set (N genes)

- Count the number of genes overlapping between the *lGene* 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 *lGene* genes and the random gene set, M_i.
  - pValue $= (length(vM[vM >= nG]) + 1)/(K + 1)$
    with $vM = c(M_1, M_2, ..M_K)$

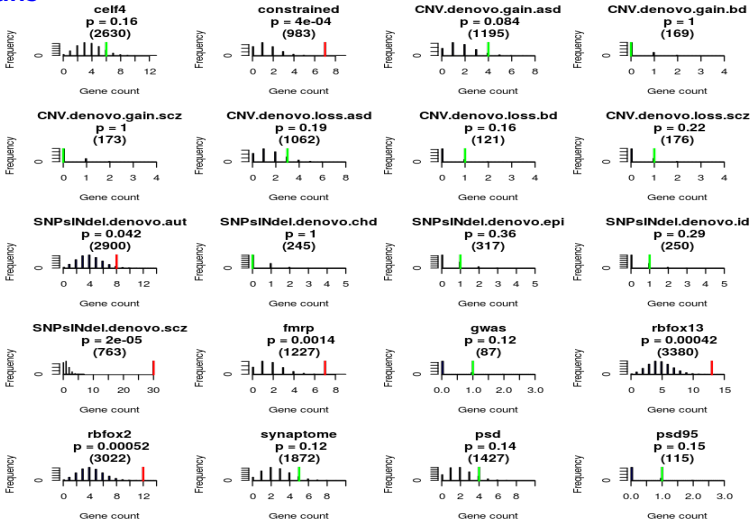
FDR < 0.3 (56 genes)

DN + Trans + CC



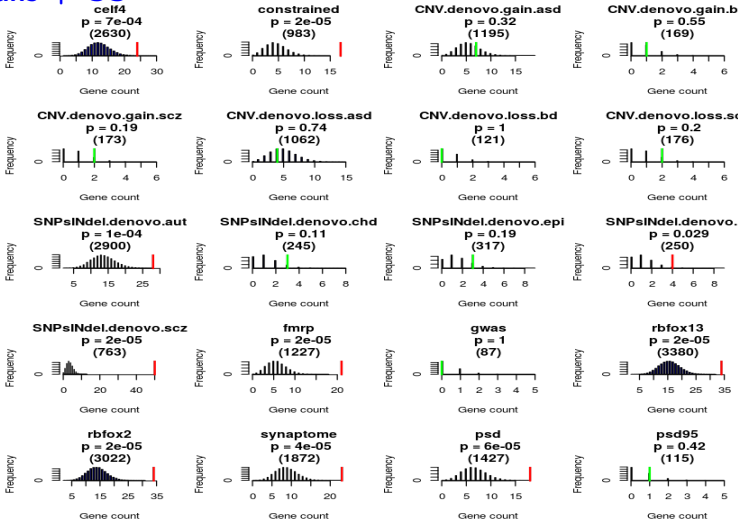
DN + Trans

FDR < 0.3 (30 genes)



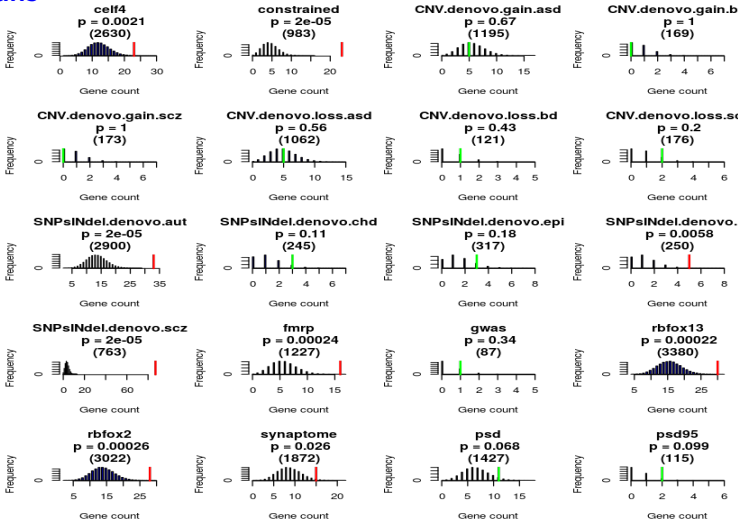
Top 100 genes (FDR < 0.38)

DN + Trans + CC



Top 100 Genes (FDR < 0.44)

DN + Trans



Not EXAC

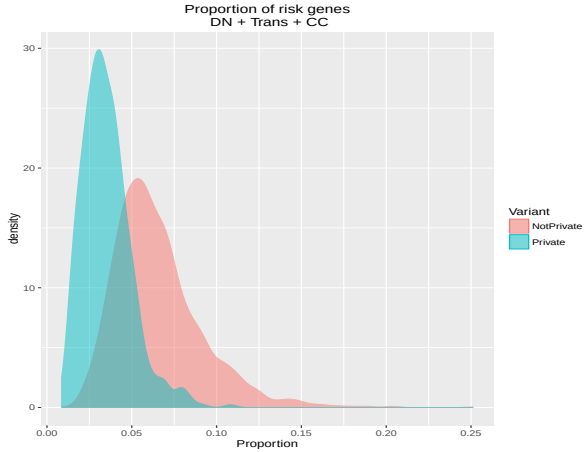
- Remove all variants in EXAC (= Private variants) => lose 1 LoF double-hit gene.
- Use the same mutation rates. [2]

Variant	Private	Non-private
LoF	108	98
Missense damaging	100	76

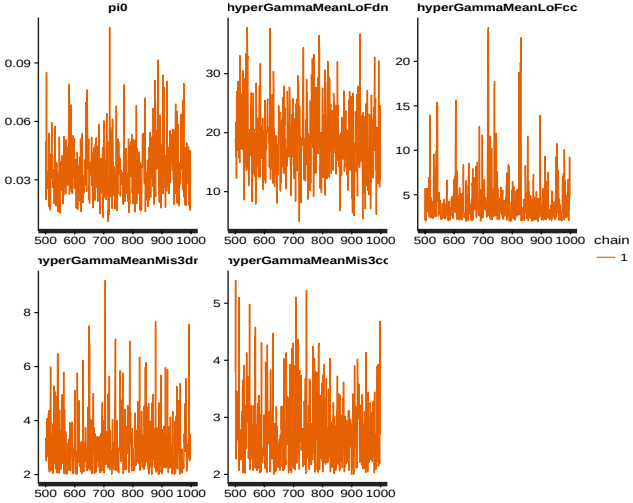
[2] Re-calculating mutation rates by removing all Exac variants

Not EXAC

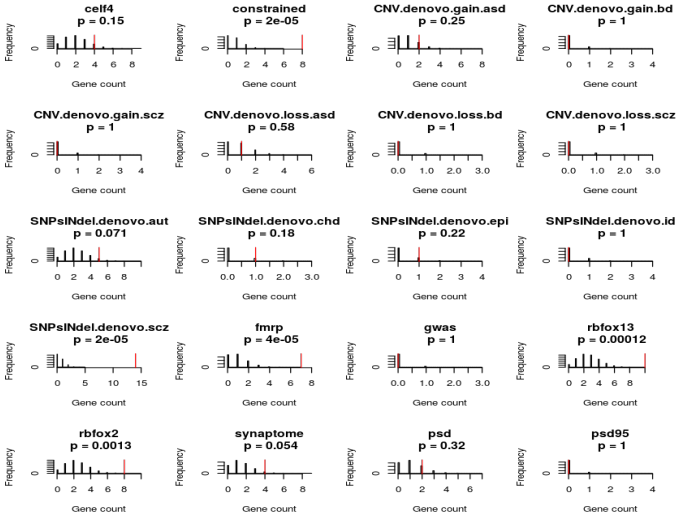
Proportion of risk genes not high



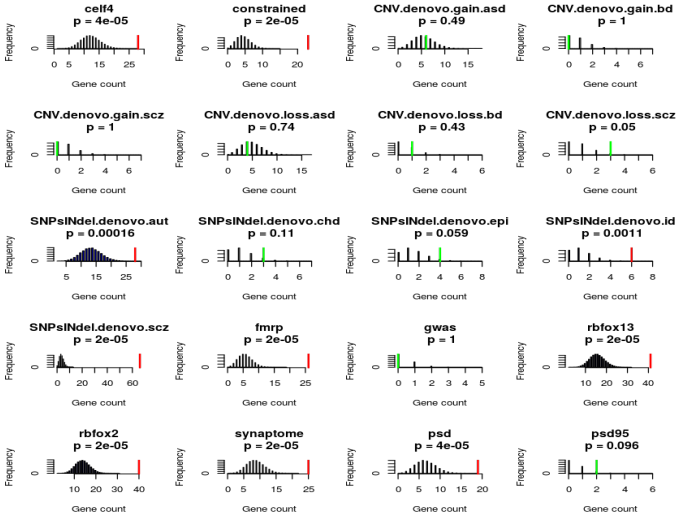
Not EXAC



FDR < 0.3, NonExac

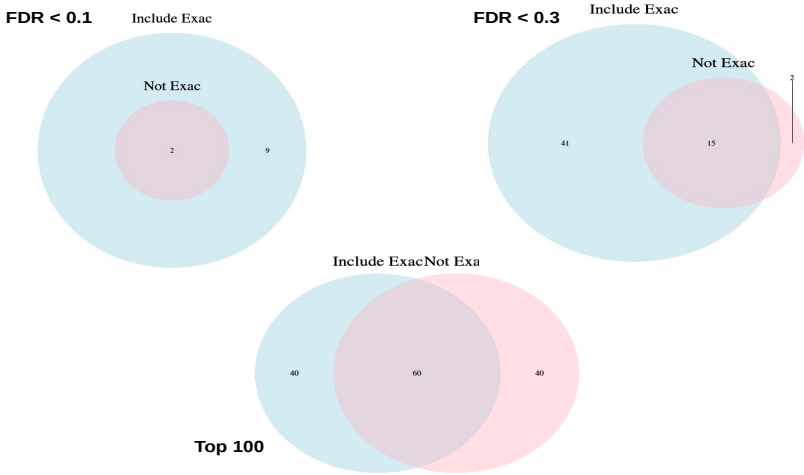


Top 100, NonExac



Private and Not private

Overlapping genes with different thresholds:



Overlapping genes

Overlapping genes: FDR < 0.1
Both LoF de novos of TAF13 are in Exac variants.

Gene	NotPrivate	Private
ADCY6	Yes	No
BLNK	Yes	No
EPHA5	Yes	No
HEATR2	Yes	No
MARK4	Yes	No
MPO	Yes	No
PRRC2A	Yes	No
ROBO1	Yes	No
TAF13	Yes	No
RB1CC1	Yes	Yes
SETD1A	Yes	Yes

Working on:

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

Should test other gene sets?

THANK YOU!!!!!