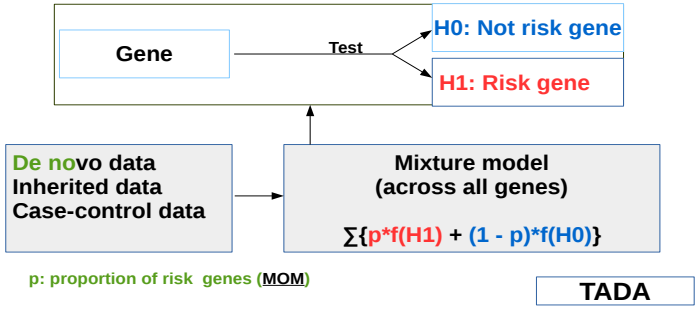
May, 2015

May 2, 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.

Main strategy

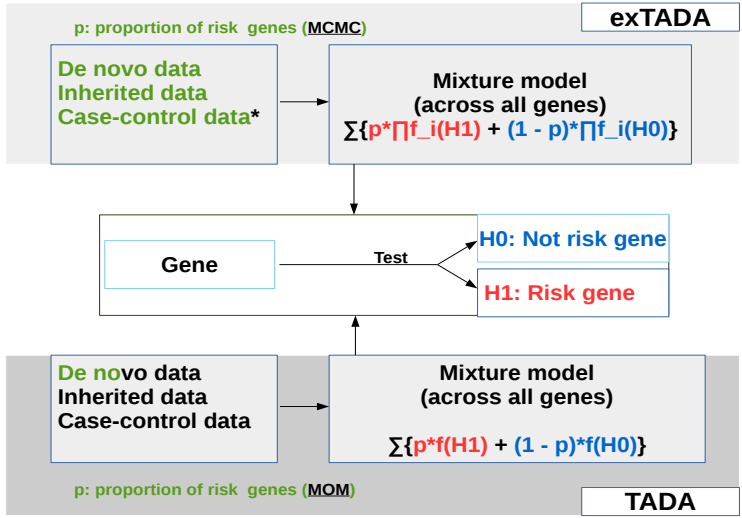
Original TADA

- Use only LoF de novo mutations => the proportion of risk genes (π) **
- 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 π**
- Do not use known risk genes (it can be used, but not necessary).

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).**

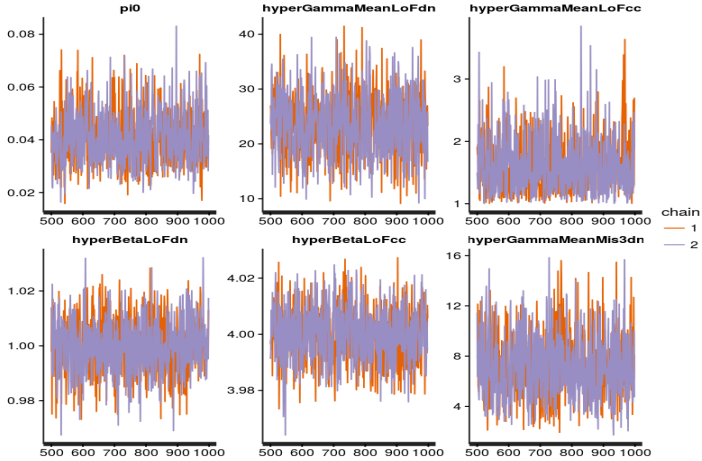
Model: **Internal product**

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

[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).

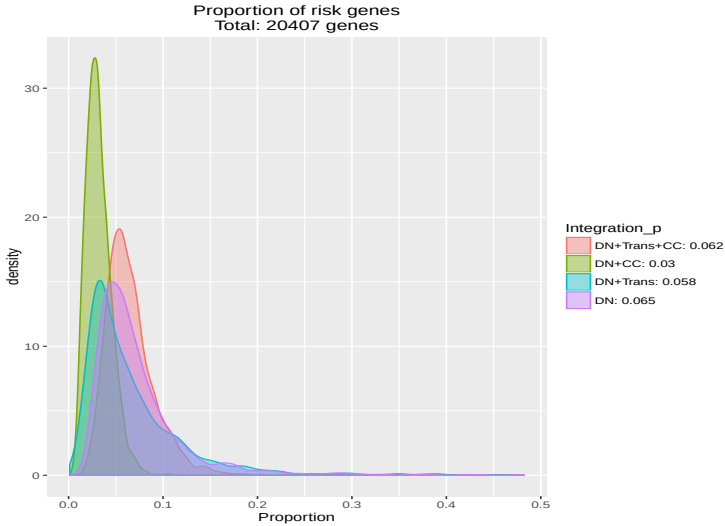
How many risk genes

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).

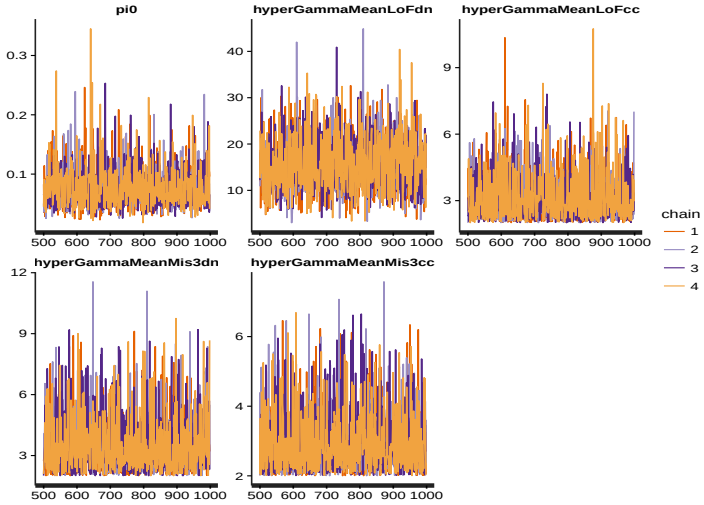
How many risk genes from exTADA?

Singleton data + NOT private.



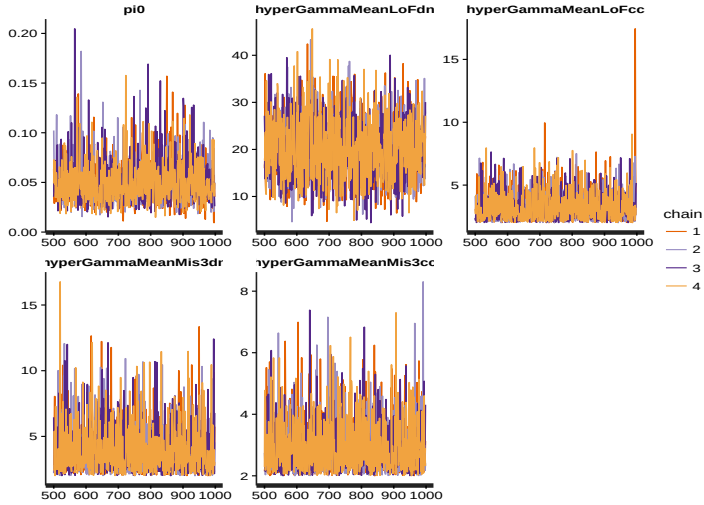
Singleton

DN + Trans + Case/Control



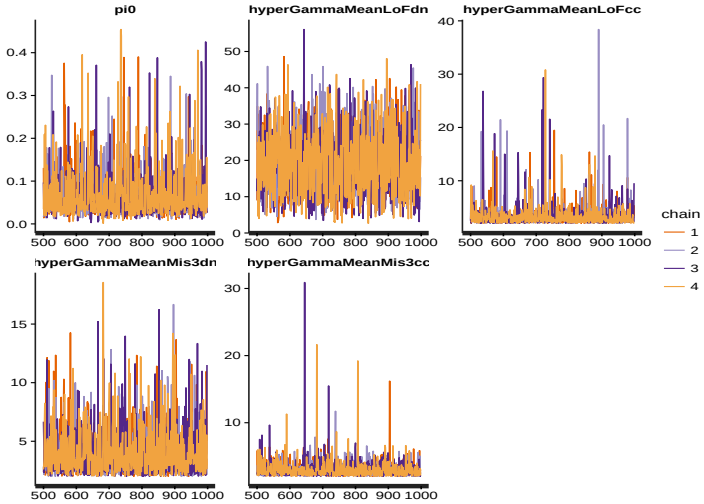
Singleton

DN + Case/Control



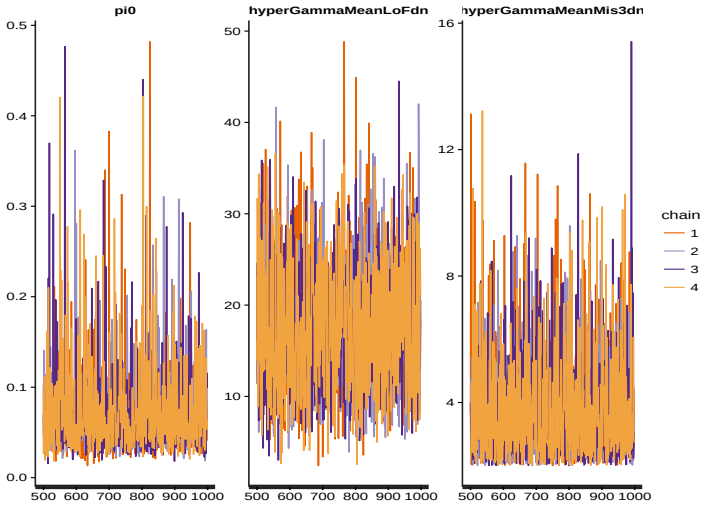
Singleton

DN + Trans

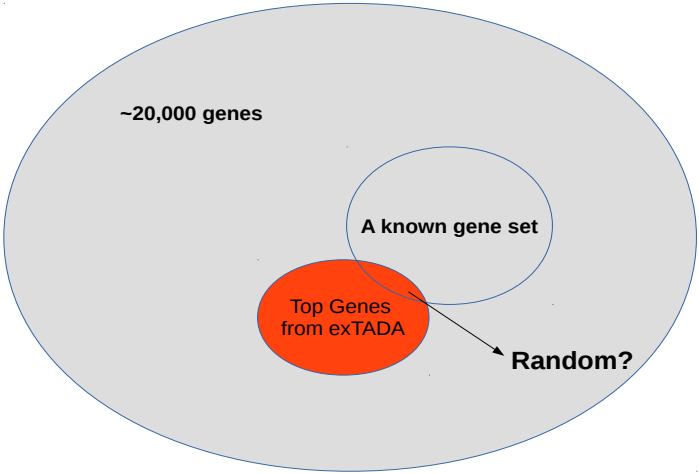


Singleton

DN



Intersect with known gene sets



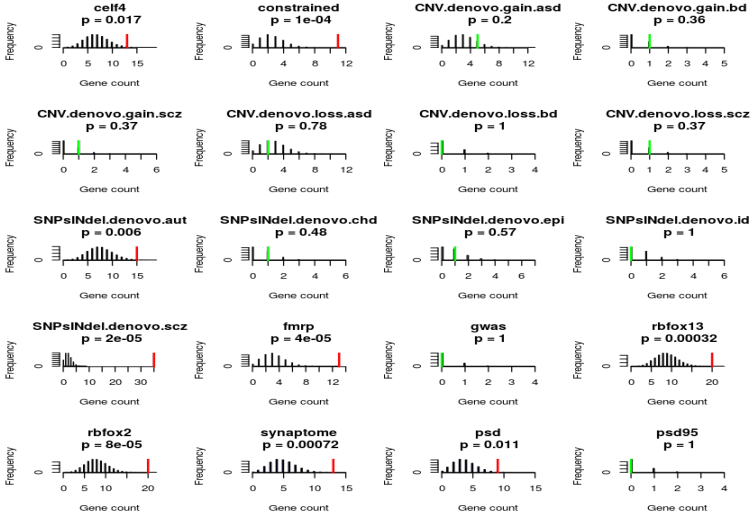
Intersect with known gene sets

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 $vM = c(M_1, M_2, ..M_K)$

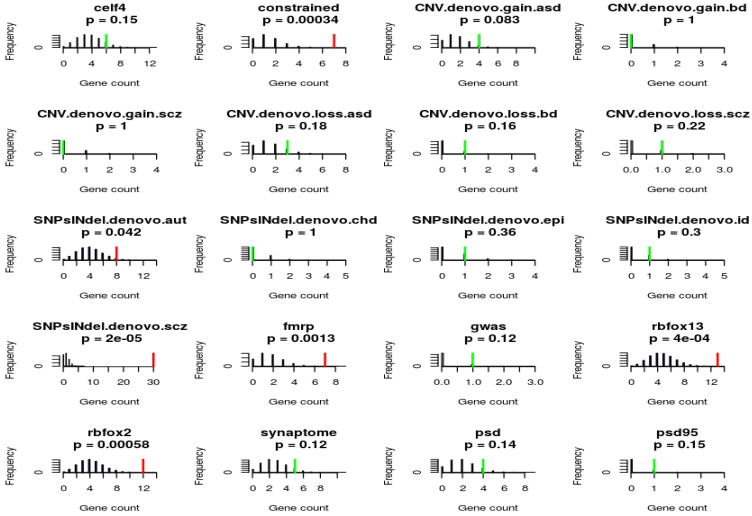
FDR < 0.3

DN + Trans + CC



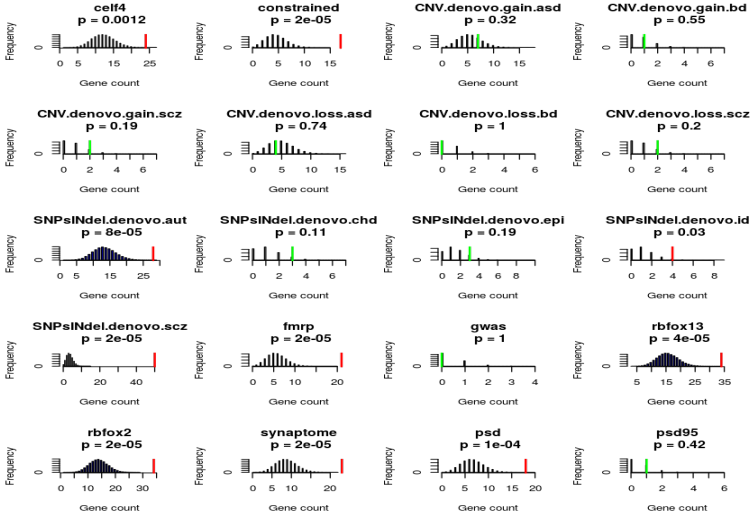
FDR < 0.3

DN + Trans



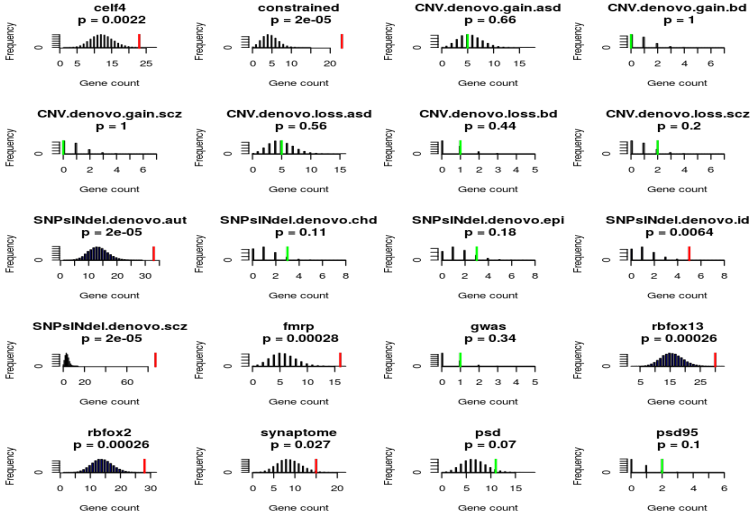
Top 100 genes

DN + Trans + CC



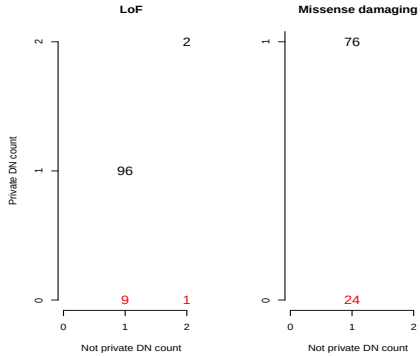
Top 100 Genes

DN + Trans



Not EXAC

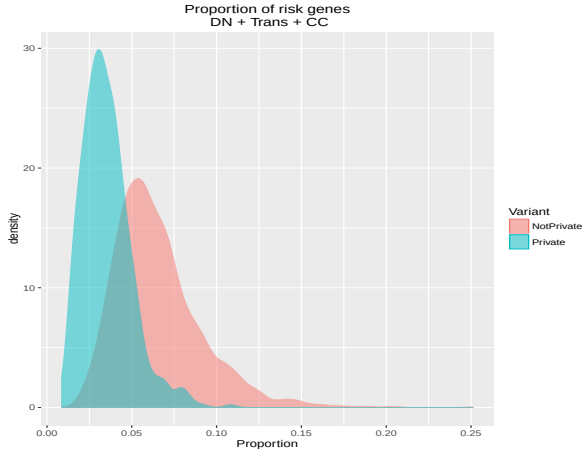
- Remove all variants in EXAC (= Private variants) => lose 1 LoF double-hit gene.
- Use the same mutation rates. [2]



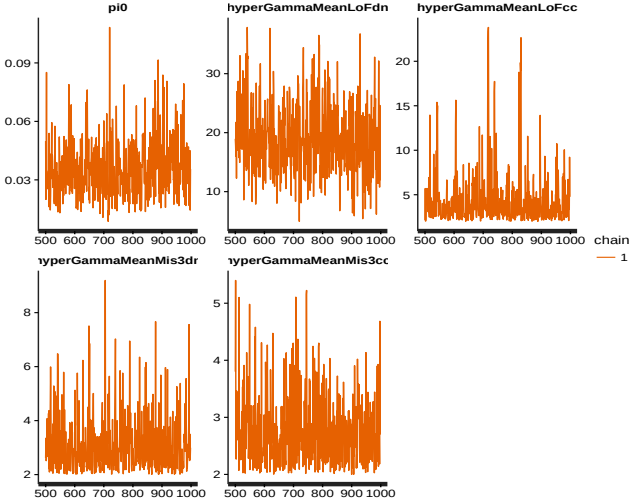
[2] Re-calculating mutation rates by removing all Exac variants

Not EXAC

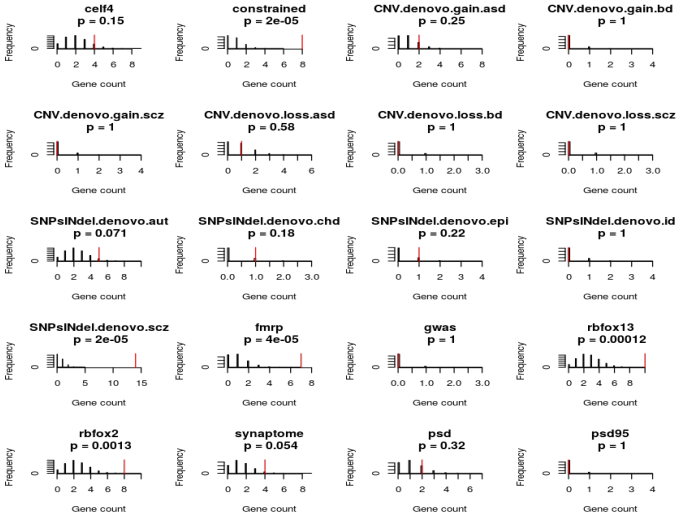
Proportion of risk genes not high



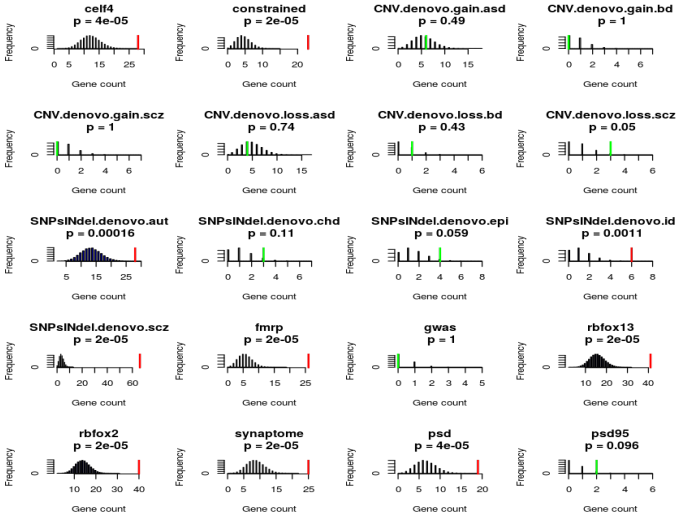
Not EXAC



FDR < 0.3, NonExac

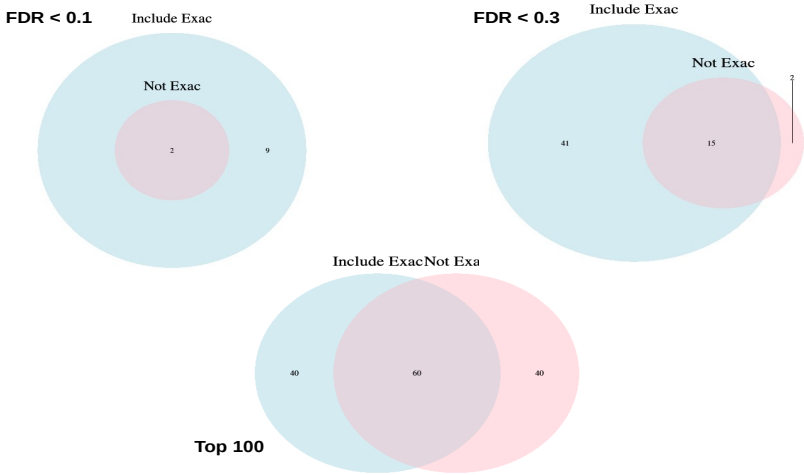


Top 100, NonExac



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	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

Working on:

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

Should test other gene sets?

THANK YOU!!!!!