May, 2015

August 25, 2016

Meeting

- extTADA pipeline.
- Apply extTADA to schizophrenia data.

May, 2015

extTADA

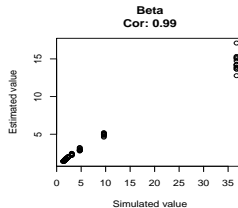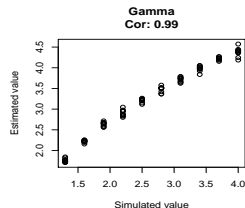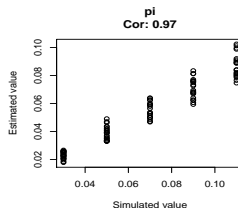Extended
TADA

Schizophrenia
data

Extended TADA

Steps

- De novo mutations: the same as original TADA (Or using binomial distribution).

- **Inherited/Case-control**:
    1. Only use a non heterogeneous population (obtain by using LM/GLM).
    2. Use an approximate model in the estimation process.

- **Estimate all parameters using a MCMC method from Equation 1 (known risk genes are not necessary).**

$$P(x|\phi_1, \phi_0) = \prod_{i=1}^{m} \left[\pi \prod_{j=1}^{K_1} f_{1_{DN_j}} \prod_{h=1}^{K_2} f_{1_{CC_h}} + (1 - \pi) \prod_{j=1}^{K_1} f_{0_{DN_j}} \prod_{h=1}^{K_2} f_{0_{CC_h}} \right] \quad (1)$$

May, 2015

extTADA

Extended
TADA

Schizophrenia
data

CC correlation

Use original TADA model to simulate => use approximate CC model estimate parameters.

May, 2015

extTADA

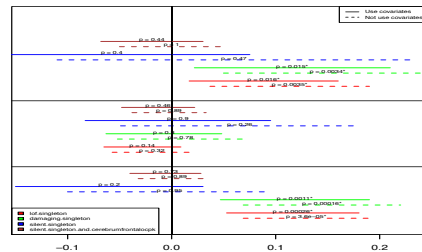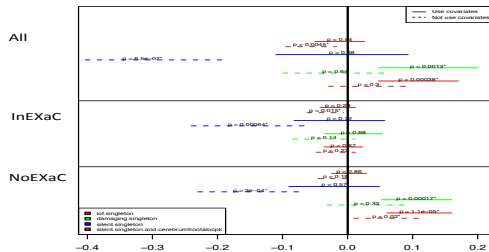Extended
TADA

Schizophrenia
data

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

May, 2015

extTADA

Extended
TADA

Schizophrenia
data

SCZ Case control

May, 2015
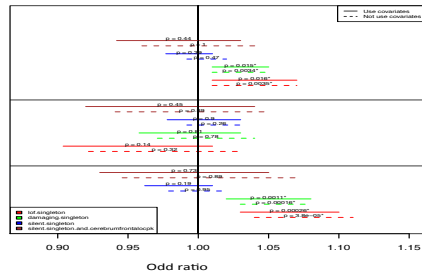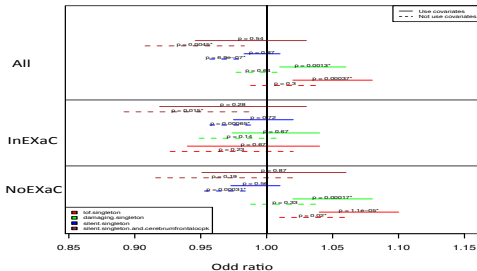
extTADA

Extended
TADA

Schizophrenia
data

SCZ de novo

Mutation	dnControl	dnCase	lOR	uOR	odds.ratio	pvalue
lof	43	111	1.34	2.87	1.94	0.000302
missense	334	612	1.45	2.15	1.77	7.04e-09
silent	134	227	0.994	1.62	1.27	0.0552
damaging missense (md)	31	100	1.6	3.83	2.44	1.21e-05
silentCerebellumocPk.narrowPeak	14	30	0.788	3.18	1.55	0.216
silentCerebrumfrontalocPk.narrowPeak (nr)	14	50	1.42	5.19	2.63	0.00109
silentFrontalcortexocPk.narrowPeak	21	55	1.13	3.37	1.92	0.0122

May, 2015

extTADA

Extended
TADA

Schizophrenia
data

Estimate parameters for only SCZ CC

md DN: missense damaging de novo

Proportion of risk genes
Total: 19441 genes

May, 2015

extTADA

Extended
TADA

Schizophrenia
data

Estimate parameters for only SCZ CC

May, 2015

extTADA

Extended
TADA

**Schizophrenia
data**
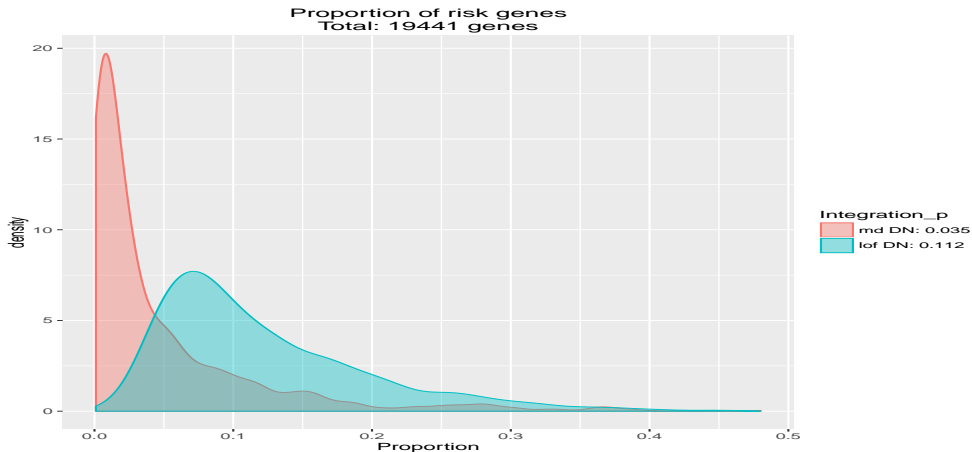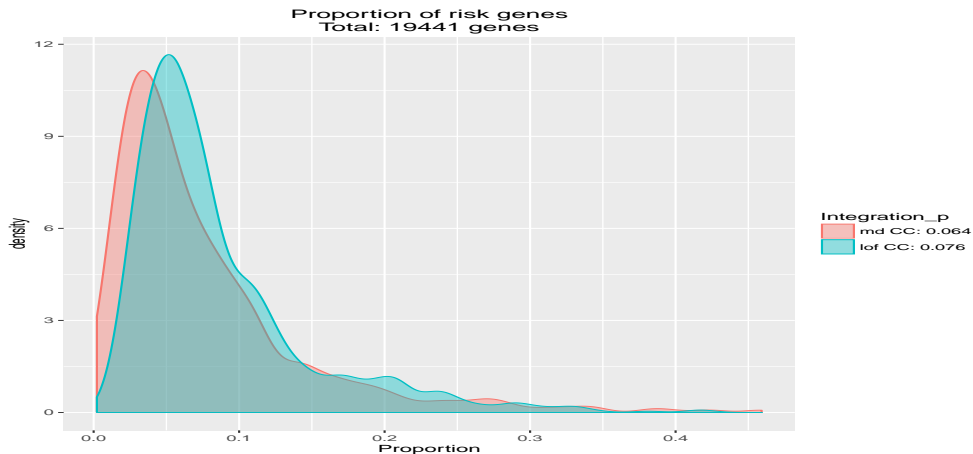
Estimate simultaneously parameters for SCZ DN and CC

May, 2015

extTADA

Extended
TADA

Schizophrenia
data
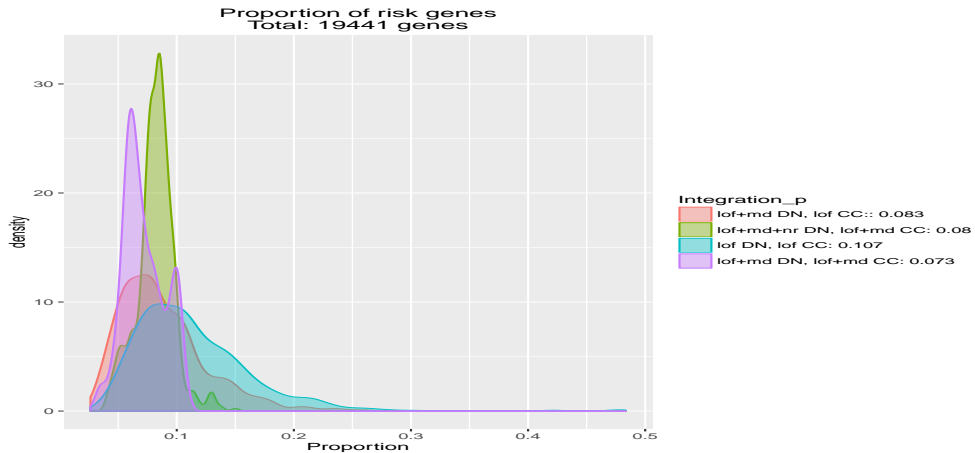
How many risk genes

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).
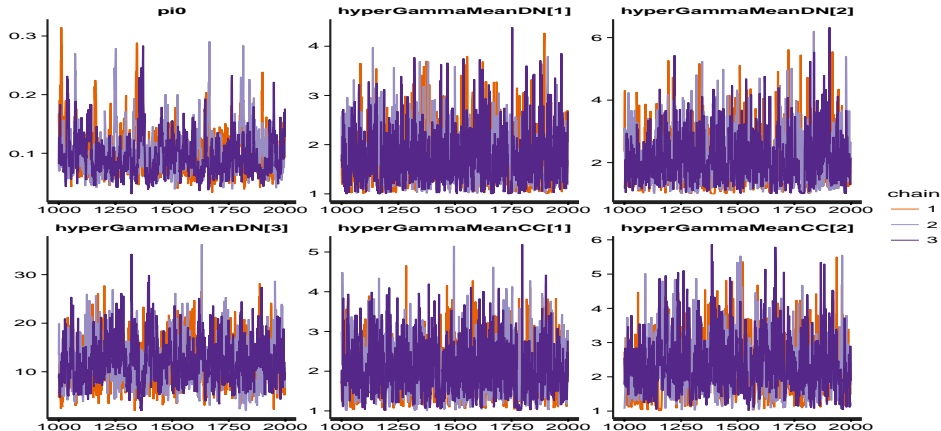
Current study (extTADA)

- $\sim$ 2000 genes.

May, 2015

extTADA

Extended
TADA

Schizophrenia
data

MCMC results for DN + CC

Top genes

Top genes from extTADA

GENE	FDR
SETD1A	0.0545668538593245
TAF13	0.0621595450057018
DARC	0.124379091145705
RB1CC1	0.166929375517358
PRRC2A	0.209294278939901
SHANK1	0.239767227415426
STXBP2	0.262757710600909
LPHN2	0.280827362530608
HSPA8	0.296109953864339

May, 2015

extTADA

Extended
TADA

Schizophrenia
data

Gene sets

- Choose top 100, 500, 1000 genes from extTADA results.
- Intersect with known gene sets.

Gene set	nGene	Top 100	Top 500	Top 1000
SNPsINdel.denovo.aut	2789	4.00E-05	1.00E-05	1.00E-05
SNPsINdel.denovo.scz	737	1.00E-05	1.00E-05	1.00E-05
celf4	2461	4.60E-04	1.00E-05	1.00E-05
constrained	936	1.00E-05	1.00E-05	1.00E-05
denovo.aut	2789	1.00E-05	1.00E-05	1.00E-05
denovo.scz	737	1.00E-05	1.00E-05	1.00E-05
fmrp	1194	1.00E-05	1.00E-05	1.00E-05
listLoFtolerantEXAC	2969	1.00E-05	1.00E-05	1.00E-05
listMcRae2016	74	1.00E-05	1.00E-05	1.00E-05
newListFromSupTable2	45	1.00E-05	1.00E-05	1.00E-05
nmdarc	74	5.00E-04	1.23E-01	1.38E-02
pLI09	3231	1.00E-05	1.00E-05	1.00E-05
rbfox13	3226	6.20E-04	1.00E-05	1.00E-05
rbfox2	2892	6.30E-04	1.00E-05	1.00E-05
synaptome	1812	2.01E-03	1.90E-04	6.00E-05

p-values were calculated by sampling 100,000 times.