

MAIN PAPER

June 1, 2016

Abstract

Integrating de novo mutations, inherited and case-control variants has been successful in identifying genetic architecture of autism spectrum disorder. Here, this approach is modified to estimate (OR CHANGE TO genetic architecture of SCZ) simultaneously relative risks and the proportion of risk genes using schizophrenia cohorts including 1024 trios, 4954 cases and 6239 controls. We identify 2 autosomal genes at a false discovery rate (FDR) < 0.05 . Increasing FDR to 0.3, 30 genes are determined. These 30 genes are tested on known gene sets. Significant results are observed in the fragile X mental retardation protein (FMRP, $p = 0.02$), haplo-insufficient ($p = 0.02$), the RNA binding proteins Rbfox1/2/3 ($p = 0.03$), the primate accelerated regions ($p = 0.033$), constrained ($p = 0.011$) gene sets.

NOTE

- All these results are in Figure 1, and Table 3. P-values in the abstract are adjusted using the method Benjamini & Hochberg (1995), NOT the method "Bonferroni".
- Author list please? Should I add any people?
- The file below describes the method to obtain p-values for gene sets (Inside Model on GitHub, it would be slightly different because of choosing random gene sets):

`intersect_with_differentGeneSet_2classes.ipynb`

Contents

1	Introduction	2
2	Data and methods	2
2.1	Data	2
2.1.1	Simulated data	2
2.1.2	Schizophrenia data	2
2.1.3	Gene sets	2
2.2	Methods	3
2.2.1	Calculate mutation rates	3
2.2.2	Analyse de novo, transmission and case-control data	3

Source	De novo	De novo control	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	34			
Rauch et al. (2012)		20 (ID)			
Giulio et al. (2016)				4954	6239
Total	1024	54	617	4954	6239

Table 1: De novo, transmitted/non-transmitted and case/control data. De novo trios are from schizophrenia (SCZ), intellectual disability (ID) studies.

3 Results	6
3.1 Simulated data	6
3.2 Schizophrenia data sets	6
3.2.1 Only de novo mutations	6
3.3 Enrichment analyses	8
4 Discussion	11

1 Introduction

2 Data and methods

2.1 Data

2.1.1 Simulated data

2.1.2 Schizophrenia data

These variants were annotated using Plink/Seq as described in Fromer et al. (2014). After that, SnpSift version 4.2 (Cingolani et al., 2012) was used to further annotate these variants using dbnsfp31a (Liu et al., 2015). Variants were groups into different categories. Loss of function (LoF) class comprised of nonsense, splice, and frameshift variants. Missense damaging were defined as missense by Plink/Seq and damaging by results of 7 methods from dbnsfp31a: SIFT, Polyphen2_{HDIV}, Polyphen2_{HVAR}, LRT, PROVEAN, MutationTaster and MutationAssessor.

2.1.3 Gene sets

Human accelerated regions (HARs)

Lists of HARs and primate accelerated regions (PARs) (Lindblad-Toh et al., 2011) were downloaded from

<http://www.broadinstitute.org/scientific-community/science/projects/mammals-models/29-mammals-project-supplementary-info>

on May 11, 2016. The coordinates of these regions were converted to hg19 using Liftover tool (Kent et al., 2002). We used a similar approach as Xu et al. (2015) to obtain genes nearby HARs. Genes in regions flanking 100 kb of the HARs/PARs were used in this study.

Other gene sets

We also test 18 gene sets described in Giulio et al (2016):

- Missense constrained genes from Table 2 of Samocha et al. (2014).

- Loss-of-function intolerant genes (Lek et al., 2015) from ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3/functional_gene_constraint/fordist_cleaned_nonpsych_z_pli_rec_null_data.txt on May 12, 2016.

2.2 Methods

2.2.1 Calculate mutation rates

We used the methodology which was based on trinucleotide context, depth of coverage as described in Fromer et al. (2014) to obtain mutation rates for different classes.

For synonymous mutation rates within the frontal cortex-derived DHS, Takata et al. (2016) showed that there were 23 these mutations in a total of 154 silent mutations observed in controls. Therefore, we multiplied this proportion 23/154 with mutation rates of synonymous mutations to obtain mutation rates for this class.

2.2.2 Analyse de novo, transmission and case-control data

We use the model developed in The Transmission and Disequilibrium Association (TADA) test (He et al., 2013) to describe de novo (x_d) and case (x_1) control (x_0) data as Equation 1:

$$\begin{aligned}x_d &\sim \text{Pois}(2N\mu\gamma_d) \\x_1 &\sim \text{Pois}(qN_1\gamma) \\x_0 &\sim \text{Pois}(qN_0)\end{aligned}\tag{1}$$

in which N_d, N_1, N_0 are sample sizes of trios, cases and controls respectively; γ_d and γ are relative risks for de novo mutations and case-control variants.

At i^{th} gene, two hypotheses: $H_0 : \gamma = 1$ versus $H_1 : \gamma > 1$ are compared

using Bayes Factor:

$$\begin{aligned}
B_i &= \frac{P(x|H_1)}{P(x|H_0)} \\
&= \frac{\prod_{j=1}^K P(x_{ij}|H_1)}{\prod_{j=1}^K P(x_{ij}|H_0)} \\
&= \prod_{j=1}^K \frac{P(x_{ij}|H_1)}{P(x_{ij}|H_0)} \text{ (Independence between categories)} \\
&= \prod_{j=1}^K B_{ij}
\end{aligned} \tag{2}$$

Where B_{ij} is the BF of the gene for j^{th} category:

$$\begin{aligned}
B_{ij} &= \frac{\int P(x_{ij}|\gamma, q)P(q|H_1)P(\gamma|H_1)dq d\gamma}{\int P(x_{ij}|\gamma, q)P(q|H_0)P(\gamma|H_0)dq d\gamma} \\
&\stackrel{\gamma_{H_0}=1}{=} \frac{\int P(x_{ij}|\gamma, q)P(q|H_1)P(\gamma|H_1)dq d\gamma}{\int P(x_{ij}|q)P(q|H_0)dq}
\end{aligned} \tag{3}$$

Or $BF_{ij} = BF_{ij(dn)}BF_{ij(CC)}$

The same as [He et al. \(2013\)](#), gamma distributions are assumed as prior distributions for γ_d and γ as in [4](#).

$$\begin{aligned}
\gamma_d &\sim \text{Gamma}(\bar{\gamma}_d \beta_d, \beta_d) \\
\gamma &\sim \text{Gamma}(\bar{\gamma} \beta, \beta) \\
q &\sim \text{Gamma}(\rho, \nu)
\end{aligned} \tag{4}$$

However, instead of using different ρ_1, ν_1 and ρ_0, ν_0 parameters for H_1 and H_0 ([He et al., 2013](#)); we use simplified parameters as a current TADA version ([De Rubeis et al., 2014](#)), in which $\rho_1 = \rho_0 = \rho$ and $\nu_1 = \nu_0 = \nu$.

To calculate BFs, we need to know hyper parameters in Equation [4](#). Let ϕ_{1j} and ϕ_{0j} be hyperparameters for H_1 and H_0 respectively. Similar to [He et al. \(2013\)](#), we assume a mixture model for all genes, with a probability π for a gene being a risk gene. However, we integrate all categories into the mixture model, and the marginal likelihood as in Eq [5](#).

$$P(x|\phi_1, \phi_0) = \prod_{i=1}^m \left[\pi \prod_{j=1}^K P(x_{ij}|\phi_{1j}) + (1 - \pi) \prod_{j=1}^K P(x_{ij}|\phi_{0j}) \right] \tag{5}$$

To obtain hyperparameters $\phi_{1j} = (\gamma_{j(dn)}, \gamma_j, \beta_{j(dn)}, \beta_j, \rho_j, \nu_j)$, we use a Markov chain Monte Carlo (MCMC) method named Hamiltonian Monte Carlo (HMC) implemented in the **rstan** package ([Carpenter et al., 2015](#); [R Core Team, 2015](#)). However, Equation [5](#) is complex with multiple parameters; therefore, we simplify the Equation to avoid sampling directly $q \sim \text{Gamma}(\rho, \nu)$:

- For de novo data, the same as Equation [1](#).

- For case-control (inheritance) data, we infer ρ, ν from control data.

Approximation method

$$\begin{aligned} P(x_1, x_0|H_j) &= P(x_1, x_1 + x_0|H_j) \\ &= P(x_1|x_1 + x_0, H_j)P(x_1 + x_0|H_j) \end{aligned} \quad (6)$$

- The first part: $P(x_1|x_1 + x_0, H_j)$
Because of $x_1 \sim \text{Pois}(N_1 q \gamma)$ and $x_0 \sim \text{Pois}(N_0 q)$, we assume that x_1 and x_0 are **independent**, we have:
 $x_1|x_1 + x_0, H_j \sim \text{Binomial}(x_1 + x_0, \theta|H_j)$
with $\theta|H_1 = \frac{N_1 \gamma}{N_1 \gamma + N_0}$ and $\theta|H_0 = \frac{N_1}{N_1 + N_0}$
The marginal likelihood is:
 $P(x_1|x_1 + x_0, H_j) = \int P(x_1|x_1 + x_0, \gamma, H_j)P(\gamma|x_1 + x_0, H_j)d\gamma$
- The second part $P(x_1 + x_0|H_j)$ is not used in the estimation process in Equation 5

Change the order of integrals to rely only on relative risks

$$P(x_1, x_0|H_j) = P(x_0|H_j)P(x_1|x_0, H_j) \quad (7)$$

- The first part $P(x_0|H_j)$ is the same as [De Rubeis et al. \(2014\)](#):

$$P(x_0|H_j) = \int P(x_0|q, H_j)P(q|\rho, \nu, H_j)dq = \text{NegBin}(x_0|\rho, \frac{N_0}{\nu + N_0}), j = 0, 1 \quad (8)$$

- The second part:

$$\begin{aligned} P(x_1|H_j, x_0) &= \int P(x_1|q, \gamma)P(q|H_j, x_0)P(\gamma|H_j)dq d\gamma \\ &= \int [P(x_1|q, \gamma)P(q|H_j, x_0)dq] P(\gamma|H_j)d\gamma \\ &= \int \text{NegBin}(x_1|\rho + x_0, \frac{N_0 + \nu}{N_1 \gamma + N_0 + \nu})P(\gamma|H_j)d\gamma \end{aligned} \quad (9)$$

The second line in Equation 9 is because $P(q|H_j, x_0)$ is the posterior probability of q after seeing the data x_0 with $q|H_j, x_0 \sim \text{Gamma}(\rho + x_0, \nu + N_0)$ ([De Rubeis et al., 2014](#)).

In Equation 9

$$\begin{aligned} x_d &\sim \text{Pois}(2N_d \gamma_d) \\ \gamma_d &\sim \text{Gamma}(\bar{\gamma}_d \beta_d, \beta_d) \\ \bar{\gamma}_d &\sim \text{Normal}(15, 10) \\ \beta_d &\sim \text{Normal}(\beta_{ds}, 0.01) \end{aligned} \quad (10)$$

$x \sim \text{Pois}(2N_{dn} \gamma_{dn})$	$\gamma_{dn} \sim \text{Gamma}(\bar{\gamma}_{dn} \beta, \beta)$	$\bar{\gamma}_{dn} \sim \text{Normal}(15, 15)$ $\beta \sim \text{Normal}(1, 0.1)$
$x_1 \sim \text{Pois}(N_1 q \gamma)$	$\gamma \sim \text{Gamma}(\bar{\gamma} \beta, \beta)$ $q \sim \text{Gamma}(\rho, \nu)$	$\bar{\gamma} \sim \text{Gamma}(1, 0.1)$ $\beta \sim \text{Normal}(\beta_0, 0.1)$ $\rho = \text{mean}(x_0), \nu = 200$
$x_0 \sim \text{Pois}(N_0 q)$	$q \sim \text{Gamma}(\rho, \nu)$	$\rho = \text{mean}(x_0), \nu = 200$

(11)

	De novo control	De novo case	Case	Control
LoF	4	114		
Missense	28	628		
Missense damaging	3	100		
Silent	12	227		
Silent FCdDHS	1	50		

Table 2: De novo mutations in trios and unaffected siblings. "Silent FCdDHS" describes for silent mutations within frontal cortex-derived DHS. Missense damaging mutations are missense mutations derived from 7 methods.

<i>Hyperparameter</i>	<i>HPDinterval(Median)ofOnlyDN</i>	<i>HPDinterval(Median)ofDN + Trans + CC</i>
π	(0.017, 0.155) 0.056	
γ_{LoFdn}	(5.722, 32.207) 17.817	
γ_{LoFmd}	(2.003, 5.884) 2.814	
β_{LoFdn}	(0.980, 1.019) 1	
β_{mddn}	(3.480, 3.519) 3.5	

3 Results

3.1 Simulated data

3.2 Schizophrenia data sets

De novo counts for different classes are in Table 3.2.

3.2.1 Only de novo mutations

We used the pipeline to analyse only de novo mutations. Table 3.2.1

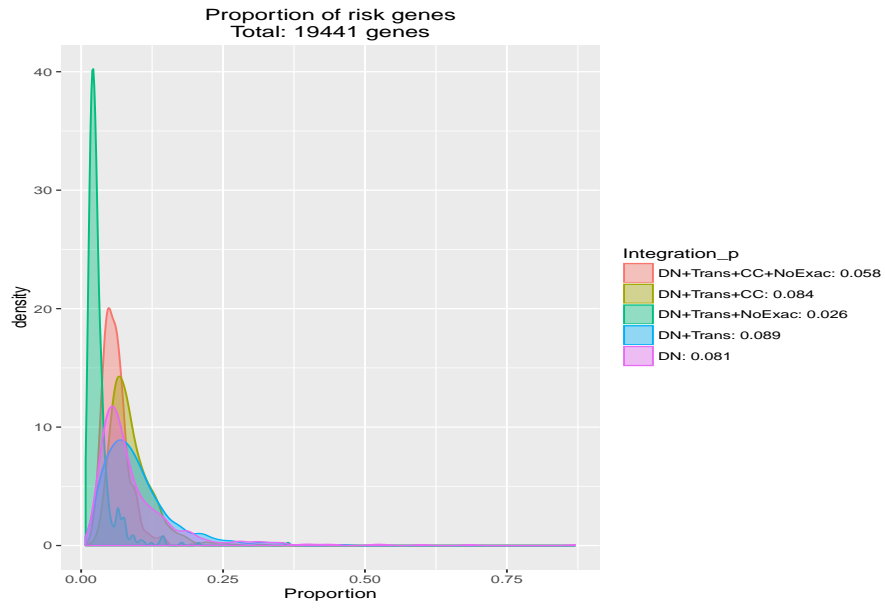


Figure 1: MCMC results of proportion of risk genes for the combination of 2 classes

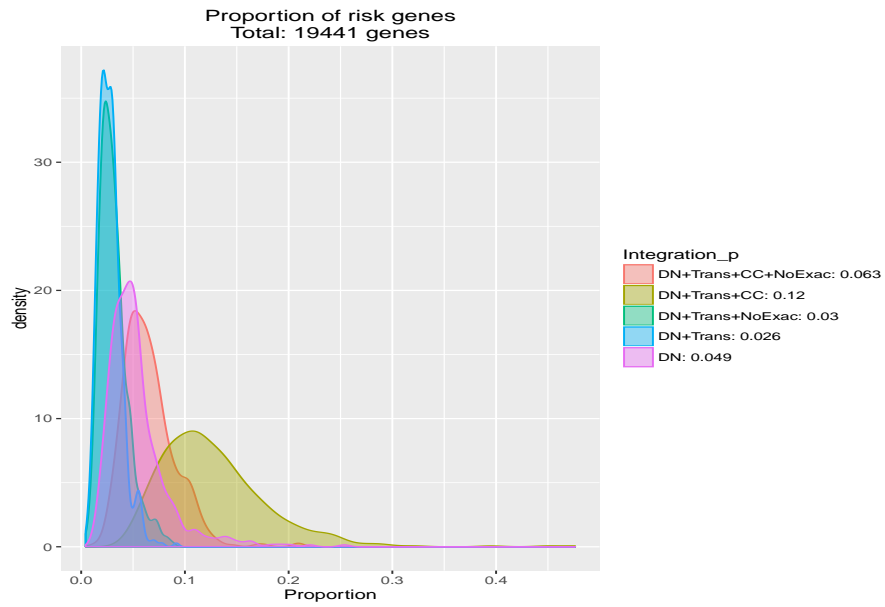
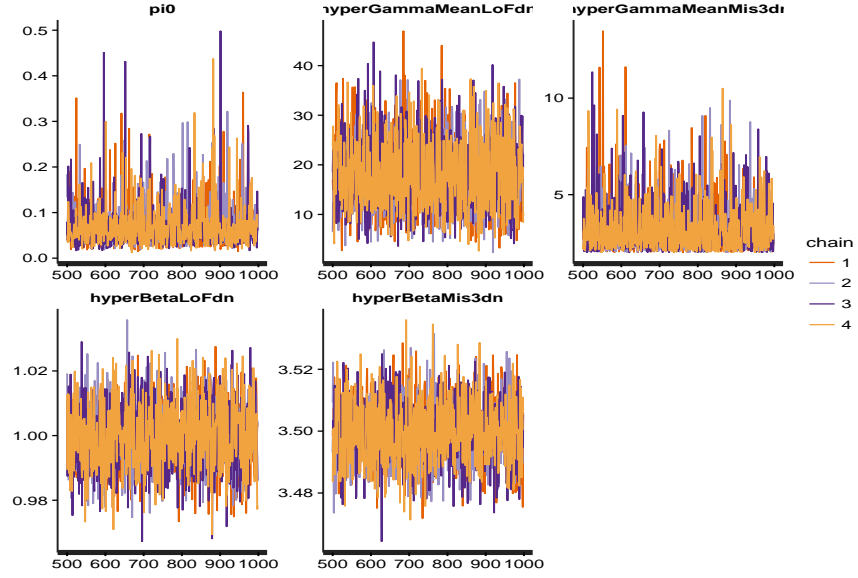


Figure 2: MCMC results of proportion of risk genes for the combination of 3 classes



3.3 Enrichment analyses

We tested the enrichment of the schizophrenia gene set with $FDR < 0.3$ in xx other gene sets. Highest enrichment was observed in the FMRP gene set ($3.99992e-05$) followed by RBFOX2, constrained, RBFOX13 and synaptome ($5.99988e-05$, $7.99984e-05$, 0.0002399952 , 0.0009399812 respectively). We also saw significant results in SNPs and Indel de novo gene set of autism (0.005959881), as well as PSD (0.01101978), and CELF4 (0.01705966).

The results were not significant in CNV de novo gene sets of SCZ, ASD, BD, CHD, EPI, and the SCZ GWAS gene set.

Gene set	TransTop100	TransnoexacTop100	TransCCnoexacTop100	TransCCTop100	Trans(FDR0.3)	Transnoexac(FDR0.3)	TransCCnoexac(FDR0.3)	TransCC(FDR0.3)
FromSZGR.160byLewis.gene	1	0.53	1	1	1	1	1	1
FromSZGR.1974GO _{neurodevelopment.gene}	0.0016	6e-04	0.0016	4e-04	0.24	0.016	0.15	0.25
FromSZGR.coreGeneSet.gene	1	1	1	1	1	1	1	1
FromSZGR.173byNg.gene	0.54	0.042	1	0.55	1	1	1	1
FromSZGR.75genesByCOR.gene	1	0.29	1	1	1	1	1	1
antipsychotics-combined.set.gene	0.22	0.081	0.0052	0.22	1	1	1	0.38
geneInPARs.txt	0.12	0.016	0.016	0.045	0.25	0.17	0.06	0.0062
geneInHARs.txt	0.32	0.15	0.16	0.78	0.27	0.19	0.36	0.2
listMcRae2016.txt	0.0022	0.34	0.0012	0.007	1	1	0.0018	0.12
celf4.txt	0.0014	0.24	2e-04	0.017	0.76	1	0.12	0.0078
constrained.txt	2e-04	0.017	2e-04	2e-04	0.4	0.035	2e-04	0.012
CNV.denovo.gain.asd.txt	0.49	0.026	0.11	0.11	0.46	1	0.23	0.5
CNV.denovo.gain.bd.txt	1	1	1	1	1	1	1	1
CNV.denovo.gain.scz.txt	1	1	1	0.56	1	1	1	1
CNV.denovo.loss.asd.txt	0.38	0.38	0.13	0.55	0.095	0.29	0.57	0.78
CNV.denovo.loss.bd.txt	0.45	0.43	0.43	0.43	1	1	1	1
CNV.denovo.loss.scz.txt	0.2	0.57	0.2	0.19	1	1	0.12	0.21
SNPsINdel.denovo.aut.txt	2e-04	2e-04	2e-04	2e-04	0.18	0.059	0.0024	0.043
SNPsINdel.denovo.chd.txt	0.026	0.11	0.001	0.0066	0.006	0.072	0.014	1
SNPsINdel.denovo.epi.txt	0.059	0.018	0.002	0.19	0.16	1	0.2	1
SNPsINdel.denovo.id.txt	0.034	0.034	0.11	0.34	1	1	1	1
SNPsINdel.denovo.scz.txt	2e-04	2e-04	2e-04	2e-04	2e-04	2e-04	2e-04	2e-04
fmrp.txt	0.0024	0.027	2e-04	2e-04	0.49	0.34	2e-04	0.002
gwas.txt	0.34	0.33	0.06	0.33	0.042	0.026	0.066	0.12
rbfox13.txt	2e-04	0.021	2e-04	2e-04	1	0.69	0.011	0.0042
rbfox2.txt	2e-04	0.015	2e-04	2e-04	0.82	0.65	0.004	0.0068
synptome.txt	0.027	0.015	4e-04	2e-04	0.63	1	0.41	0.11
psd.txt	0.068	0.0018	6e-04	0.0028	0.53	1	0.29	0.32
psd95.txt	0.42	0.41	0.41	0.41	1	1	1	1
pLI09.txt	2e-04	0.0022	2e-04	4e-04	0.86	0.31	0.0026	0.002

Table 3: Test overlapping gene sets with extTADA results for two classes.

Gene set	TransCCnoexacTop100	TransCCTop100	TransTop100	TransnoexacTop100	TransCCnoexac(FDR0.3)	TransCC(FDR0.3)	Trans(FDR0.3)	Transnoexac(FDR0.3)
FromSZGR.160byLewis.gene	1	1	1	1	1	1	1	1
FromSZGR.1974GO _{neurodevelopment.gene}	0.002	0.0012	2e-04	2e-04	0.56	0.22	0.47	1
FromSZGR.coreGeneSet.gene	1	1	1	1	1	1	1	1
FromSZGR.173byNg.gene	1	1	0.55	0.55	1	1	1	1
FromSZGR.75genesByCOR.gene	1	1	1	1	1	1	1	1
antipsychotics-combined.set.gene	0.22	0.024	0.024	0.023	0.14	0.13	1	1
geneInPARs.txt	0.12	0.12	0.27	0.25	0.21	0.24	0.17	0.048
geneInHARs.txt	0.31	0.54	0.54	0.78	1	1	0.19	1
listMcRae2016.txt	0.07	0.068	0.065	0.066	0.036	0.15	1	1
celf4.txt	0.039	0.059	0.34	0.25	0.69	0.29	0.59	1
constrained.txt	2e-04	0.0062	2e-04	4e-04	2e-04	0.002	0.28	0.097
CNV.denovo.gain.asd.txt	0.053	0.19	0.1	0.11	0.39	0.35	1	1
CNV.denovo.gain.bd.txt	1	0.57	0.54	1	1	1	1	1
CNV.denovo.gain.scz.txt	0.2	0.19	0.55	1	1	0.26	1	1
CNV.denovo.loss.asd.txt	0.37	0.23	0.026	0.061	1	0.56	0.048	0.095
CNV.denovo.loss.bd.txt	1	1	0.43	0.44	1	1	1	1
CNV.denovo.loss.scz.txt	0.2	1	0.56	0.2	1	1	1	1
SNPsINdel.denovo.aut.txt	8e-04	0.025	6e-04	2e-04	0.028	0.052	0.063	1
SNPsINdel.denovo.chd.txt	0.32	0.69	0.32	0.11	0.1	0.35	0.08	1
SNPsINdel.denovo.epi.txt	0.064	0.18	0.19	0.065	0.13	0.43	1	1
SNPsINdel.denovo.id.txt	0.032	0.33	0.1	0.0066	0.11	0.068	1	1
SNPsINdel.denovo.scz.txt	2e-04	2e-04	2e-04	2e-04	2e-04	2e-04	2e-04	0.0014
fmrp.txt	2e-04	2e-04	2e-04	2e-04	2e-04	2e-04	0.054	0.12
gwas.txt	1	1	0.33	0.33	1	1	0.026	0.0084
rbfox13.txt	8e-04	0.0058	0.002	0.0018	0.0064	0.02	0.69	1
rbfox2.txt	0.0022	0.011	0.0022	0.004	0.37	0.066	0.65	0.27
synaptome.txt	6e-04	0.0026	0.028	0.0032	0.18	0.014	1	1
psd.txt	2e-04	0.071	0.072	8e-04	0.12	0.24	1	1
psd95.txt	1	0.41	0.42	0.1	1	1	1	1
pLI09.txt	6e-04	2e-04	2e-04	0.0044	0.0014	0.0012	0.31	0.31

Table 4: Test overlapping gene sets with extTADA results for three classes

4 Discussion

References

- B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: a probabilistic programming language. *Journal of Statistical Software*, 2015.
- P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2):80–92, 2012.
- S. De Rubeis, X. He, A. P. Goldberg, C. S. Poultney, K. Samocha, A. E. Cicek, Y. Kou, L. Liu, M. Fromer, S. Walker, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515(7526):209–215, 2014.
- M. Fromer, A. J. Pocklington, D. H. Kavanagh, H. J. Williams, S. Dwyer, P. Gormley, L. Georgieva, E. Rees, P. Palta, D. M. Ruderfer, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*, 506(7487):179–184, 2014.
- S. L. Girard, J. Gauthier, A. Noreau, L. Xiong, S. Zhou, L. Jouan, A. Dionne-Laporte, D. Spiegelman, E. Henrion, O. Diallo, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature genetics*, 43(9):860–863, 2011.
- S. Gulsuner, T. Walsh, A. C. Watts, M. K. Lee, A. M. Thornton, S. Casadei, C. Rippey, H. Shahin, V. L. Nimgaonkar, R. C. Go, et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell*, 154(3):518–529, 2013.
- X. He, S. J. Sanders, L. Liu, S. De Rubeis, E. T. Lim, J. S. Sutcliffe, G. D. Schellenberg, R. A. Gibbs, M. J. Daly, J. D. Buxbaum, et al. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet*, 9(8):e1003671, 2013.

- W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler. The human genome browser at ucsc. *Genome research*, 12(6):996–1006, 2002.
- M. Lek, K. Karczewski, E. Minikel, K. Samocha, E. Banks, T. Fennell, A. O’Donnell-Luria, J. Ware, A. Hill, B. Cummings, et al. Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv*, page 030338, 2015.
- K. Lindblad-Toh, M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 478(7370):476–482, 2011.
- X. Liu, C. Wu, C. Li, and E. Boerwinkle. dbnsfp v3. 0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site snvs. *Human mutation*, 2015.
- S. E. McCarthy, J. Gillis, M. Kramer, J. Lihm, S. Yoon, Y. Berstein, M. Mistry, P. Pavlidis, R. Solomon, E. Ghiban, et al. De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Molecular psychiatry*, 19(6):652, 2014.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL <https://www.R-project.org/>.
- A. Rauch, D. Wieczorek, E. Graf, T. Wieland, S. Ende, T. Schwarzmayer, B. Albrecht, D. Bartholdi, J. Beygo, N. Di Donato, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *The Lancet*, 380(9854):1674–1682, 2012.
- K. E. Samocha, E. B. Robinson, S. J. Sanders, C. Stevens, A. Sabo, L. M. McGrath, J. A. Kosmicki, K. Rehnström, S. Mallick, A. Kirby, et al. A framework for the interpretation of de novo mutation in human disease. *Nature genetics*, 46(9):944–950, 2014.
- A. Takata, I. Ionita-Laza, J. A. Gogos, B. Xu, and M. Karayiorgou. De novo synonymous mutations in regulatory elements contribute to the genetic etiology of autism and schizophrenia. *Neuron*, 89(5):940–947, 2016.

- B. Xu, I. Ionita-Laza, J. L. Roos, B. Boone, S. Woodrick, Y. Sun, S. Levy, J. A. Gogos, and M. Karayiorgou. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nature genetics*, 44(12):1365–1369, 2012.
- K. Xu, E. E. Schadt, K. S. Pollard, P. Roussos, and J. T. Dudley. Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Molecular biology and evolution*, 32(5):1148–1160, 2015.