The American Journal of Human Genetics

Supplemental Data

Incorporating Functional Information in Tests of Excess De Novo Mutational Load

Yu Jiang, Yujun Han, Slavé Petrovski, Kouros Owzar, David B. Goldstein, and Andrew S. Allen

Supplemental Data

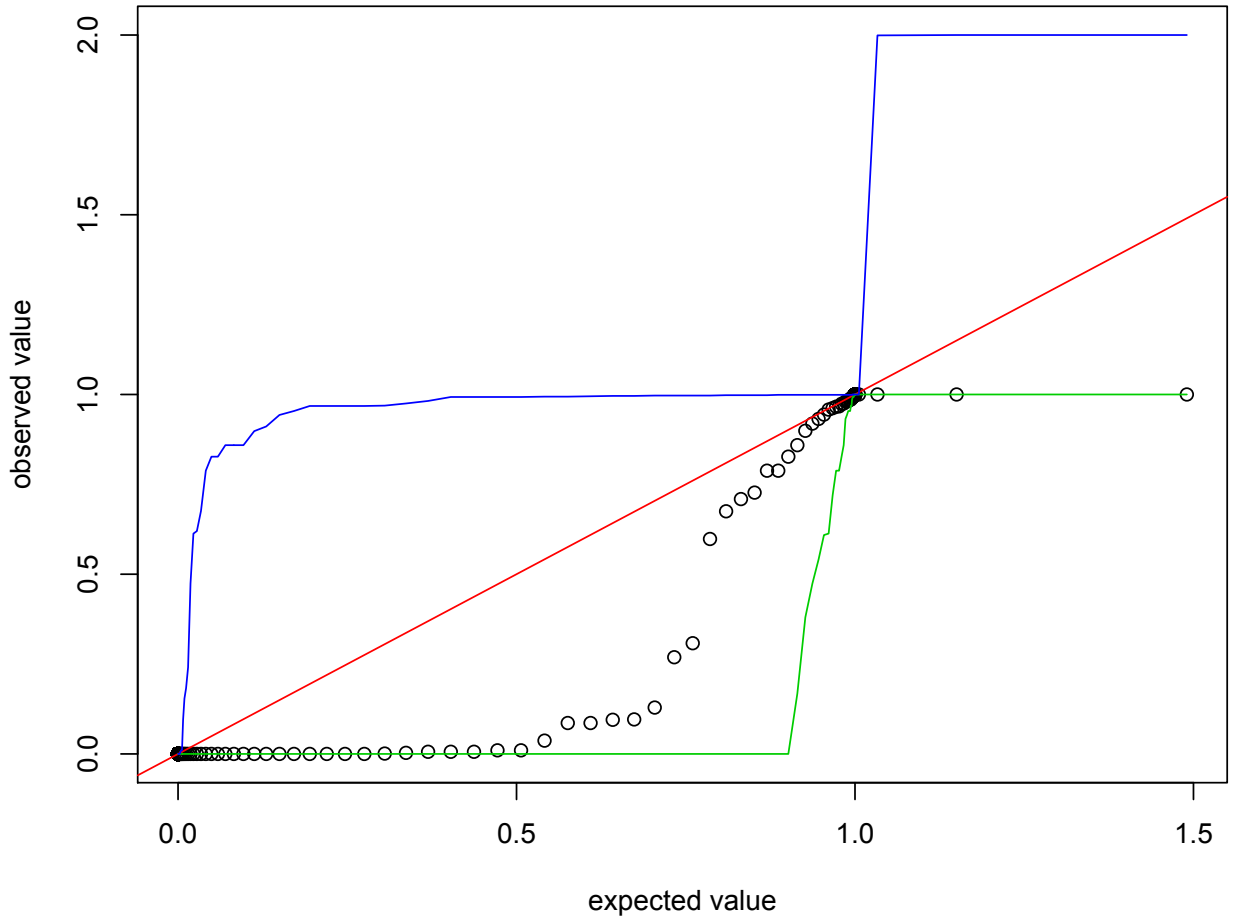

Figure S1: Expected versus observed test statistics for KIRREL3-based simulation. Sample size=150 and 10000 replicates. The blue and green lines denote 95% confidence intervals.

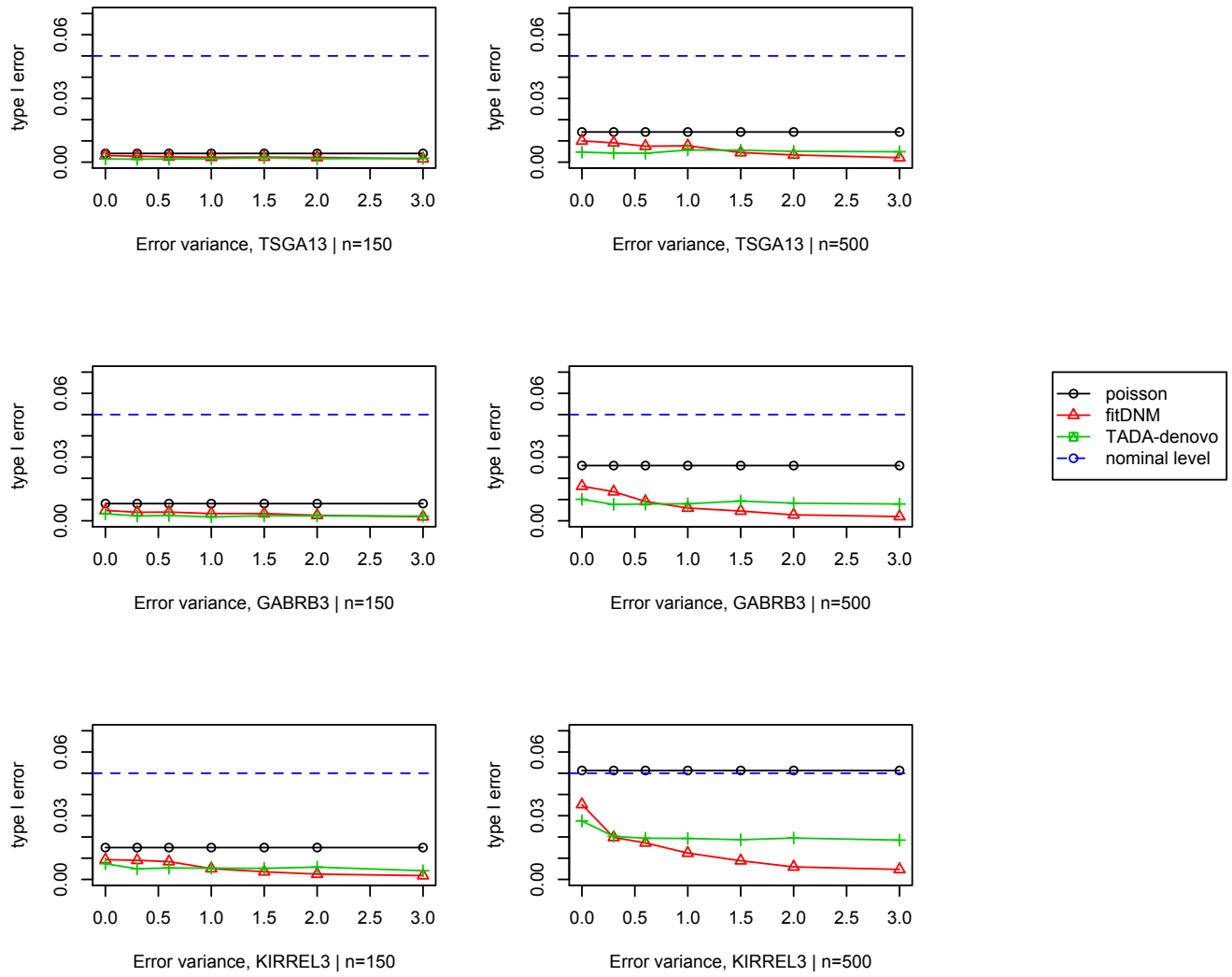

Figure S2: Type I error rates variant deleteriousness is misspecified (alpha=0.05)

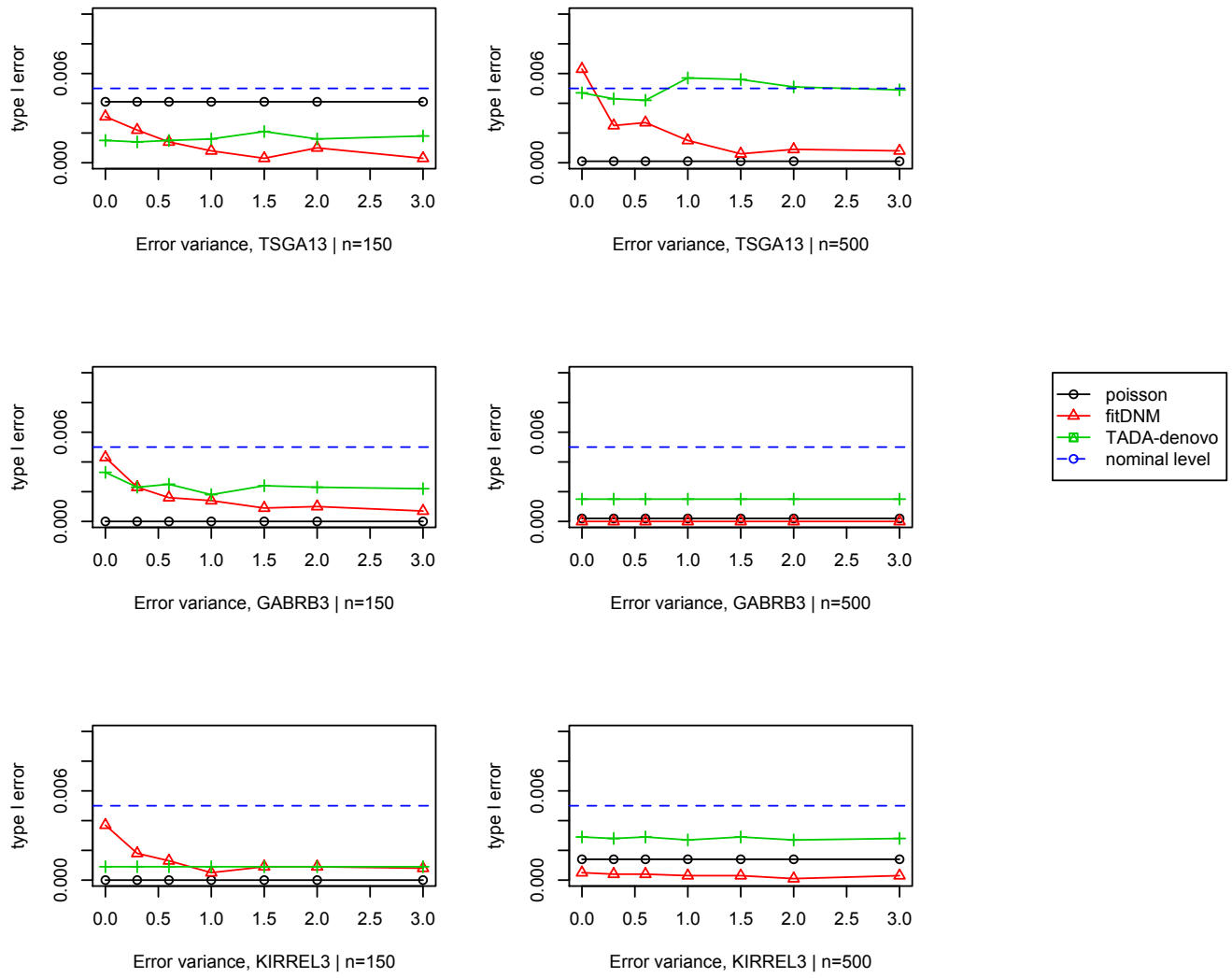

Figure S3: Type I error rates when variant deleteriousness is misspecified (alpha=0.005)

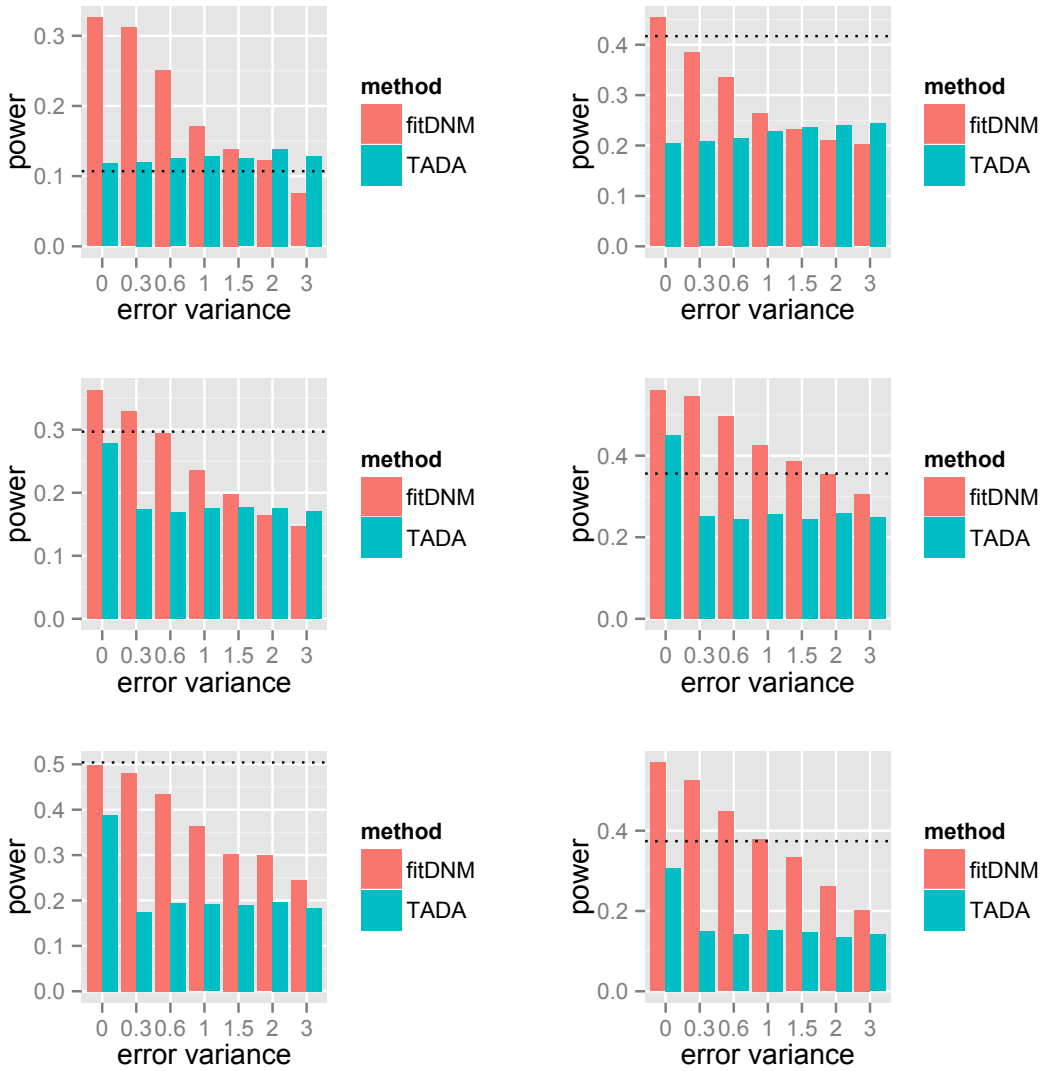

Figure S4: Power when variant deleteriousness is misspecified. The sample size is 150 in the left panels and 500 in the right. Simulations are based on the following genes (top to the bottom): *TSGA13*, *GABRB3*, *KIRREL3*. The dashed horizontal line is the power of Poisson test.

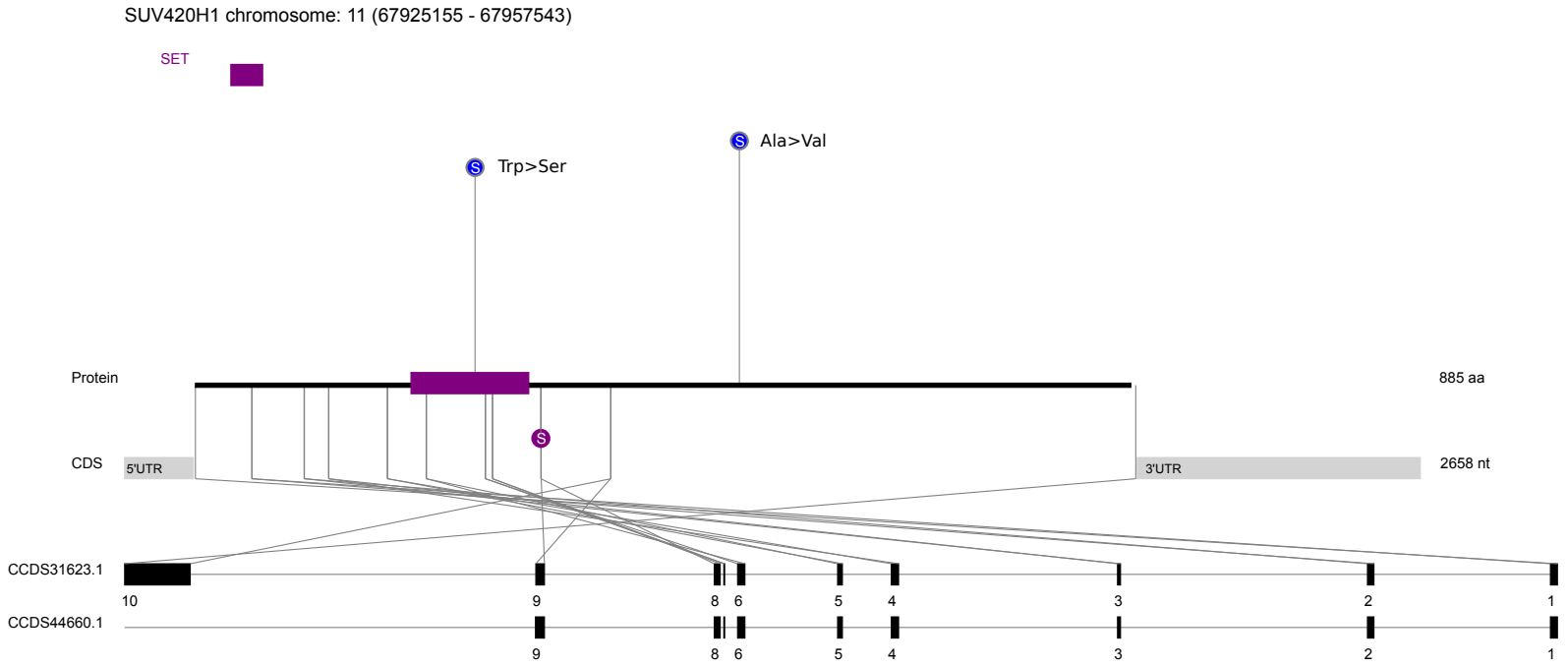

Figure S5: Location of *De novo* mutations in SUV420H1 found in ASD samples.

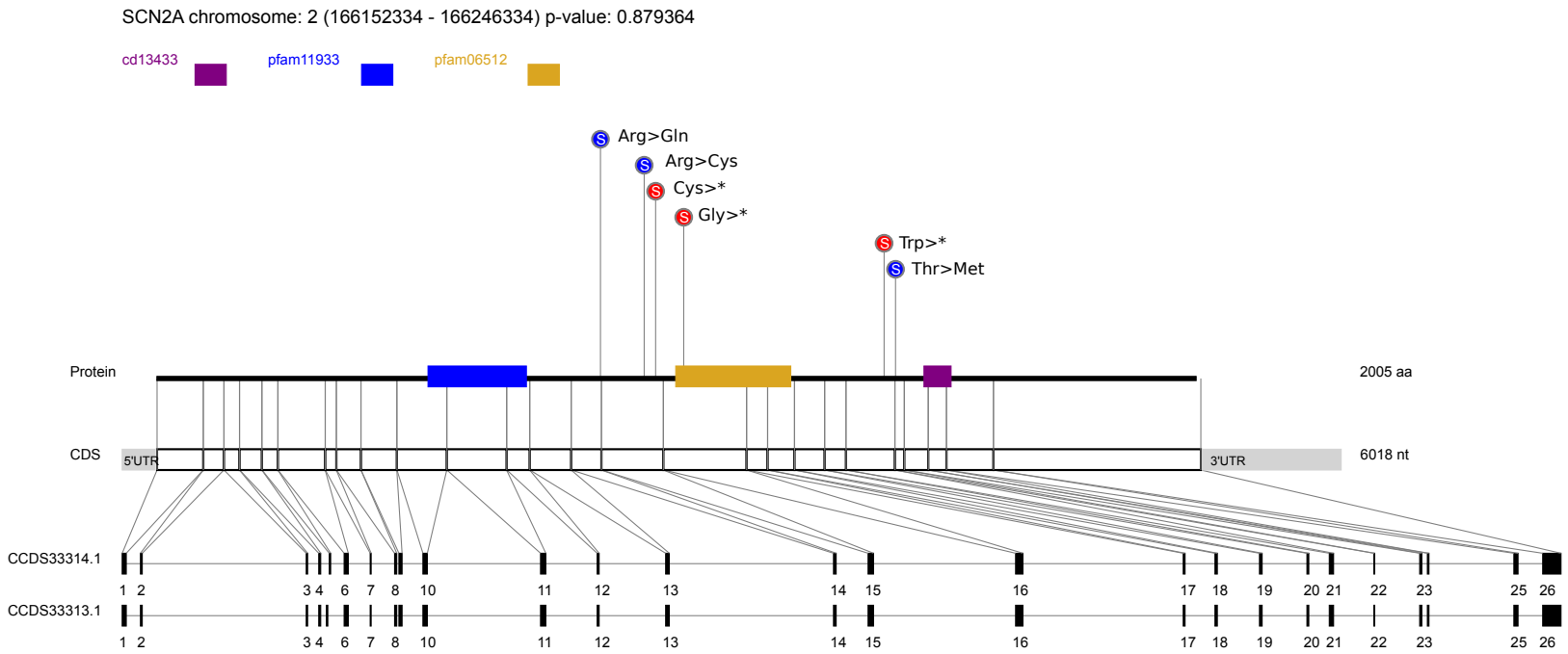

Figure S6: Location of *De novo* mutations in *SCN2A* found in Neurodevelopmental and neuropsychiatric samples. The first mutation (Arg>Gln) occurred twice in EE samples.

Table S1. Correlations between true PolyPhen-2 score and the misspecified Polyphen-2 scores used in analysis.

Gene	Simulation variance σ^2					
	0.3	0.6	1.0	1.5	2.0	3.0
TSGA13	0.878	0.698	0.565	0.462	0.439	0.391
GABRB3	0.915	0.753	0.604	0.507	0.444	0.424
KIRREL3	0.922	0.775	0.635	0.557	0.526	0.464

Table S2. List of *de novo* mutations in disease associated genes from combined analysis

Gene	RVIS	Variant description	Affected transcript	Affected protein	Disease	Polyphen-2
TRIO	0.15%	chr5:g.14388774C>T	NM_007118.2:c.3934C>T	NM_007118.2(TRIO_i001):p.(Arg1312Trp)	ASD	probably damaging
		chr5:g.14390384A>T	NM_007118.2:c.4103A>T	NM_007118.2(TRIO_i001):p.(Asp1368Val)	severeID	probably damaging
		chr5:g.14394220A>T	NM_007118.2:c.4292A>T	NM_007118.2(TRIO_i001):p.(Lys1431Met)	ASD	probably damaging
		chr5:g.14492731C>T	NM_007118.2:c.7688C>T	NM_007118.2(TRIO_i001):p.(Thr2563Met)	severeID	probably damaging
		chr5:g.14508071C>T	NM_007118.2:c.8834C>T	NM_007118.2(TRIO_i001):p.(Thr2945Met)	EE	probably damaging
SUV420H1	11.22 %	chr11:g.67926275G>A	NM_017635.3:c.1538C>T	NM_017635.3(SUV420H1_i001):p.(Ala513Val)	ASD	probably damaging
		chr11:g.67938481C>T	NM_017635.3:c.977+1G>A	Splice donor	ASD	
		chr11:g.67939039C>G	NM_017635.3:c.791G>C	NM_017635.3(SUV420H1_i001):p.(Trp264Ser)	ASD	probably damaging
SCN2A	0.89%	chr2:g.166198975G>A	NM_001040142.1:c.2558G>A	NM_001040142.1(SCN2A_i001):p.(Arg853Gln)	EE (twice)	probably damaging
		chr2:g.166201311C>T	NM_001040142.1:c.2809C>T	NM_001040142.1(SCN2A_i001):p.(Arg937Cys)	severeID	probably damaging
		chr2:g.166201379C>A	NM_001040142.1:c.2877C>A	NM_001040142.1(SCN2A_i001):p.(Cys959*)	ASD	
		chr2:g.166210819G>T	NM_001040142.1:c.3037G>T	NM_001040142.1(SCN2A_i001):p.(Gly1013*)	ASD	
		chr2:g.166231415G>A	NM_001040142.1:c.4193G>A	NM_001040142.1(SCN2A_i001):p.(Trp1398*)	severeID	
		chr2:g.166234111C>T	NM_001040142.1:c.4259C>T	NM_001040142.1(SCN2A_i001):p.(Thr1420Met)	ASD	probably damaging
CDKL5	7.64%	chrX:g.18598064C>T	NM_001037343.1:c.379C>T	NM_001037343.1(CDKL5_i001):p.(His127Tyr)	EE	probably damaging
		chrX:g.18606157G>A	NM_001037343.1:c.638G>A	NM_001037343.1(CDKL5_i001):p.(Gly213Glu)	EE	probably damaging
		chrX:g.18622434C>T	NM_001037343.1:c.1390C>T	NM_001037343.1(CDKL5_i001):p.(Gln464*)	EE	
SCN1A	2.29%	chr2:g.166848071G>A	NM_001165963.1:c.5714C>T	NM_001165963.1(SCN1A_i001):p.(Pro1905Leu)	ASD	probably damaging

		chr2:g.166911147C>T	NM_001165963.1:c.602+1G>A	splice donor variant	EE (twice)	
		chr2:g.166903480G>A	NM_001165963.1:c.1177C>T	NM_001165963.1(SCN1A_i001):p.(Arg393Cys)	EE	probably damaging
		chr2:g.166894356C>T	NM_001165963.1:c.2876G>A	NM_001165963.1(SCN1A_i001):p.(Cys959Tyr)	EE	probably damaging
		chr2:g.166870322G>A	NM_001165963.1:c.3637C>T	NM_001165963.1(SCN1A_i001):p.(Arg1213*)	EE	
		chr2:g.166852575G>T	NM_001165963.1:c.4529C>A	NM_001165963.1(SCN1A_i001):p.(Ala1510Glu)	EE	probably damaging
		chr2:g.166848563C>G	NM_001165963.1:c.5222G>C	NM_001165963.1(SCN1A_i001):p.(Cys1741Ser)	EE	probably damaging
STXBP1	13.64%	chr9:g.130420659G>A	NM_001032221.2:c.175G>A	NM_001032221.2(STXBP1_i001):p.(Glu59Lys)	SevereID	probably damaging
		chr9:g.130422363G>C	NM_001032221.2:c.301G>C	NM_001032221.2(STXBP1_i001):p.(Ala101Pro)	SevereID	possibly damaging
		chr9:g.130425622C>T	NM_001032221.2:c.568C>T	NM_001032221.2(STXBP1_i001):p.(Arg190Trp)	EE	probably damaging
		chr9:g.130428484C>T	NM_001032221.2:c.703C>T	NM_001032221.2(STXBP1_i001):p.(Arg235*)	EE	
		chr9:g.130434370C>T	NM_001032221.2:c.1004C>T	NM_001032221.2(STXBP1_i001):p.(Pro335Leu)	EE	probably damaging
		chr9:g.130438189G>A	NM_001032221.2:c.1217G>A	NM_001032221.2(STXBP1_i001):p.(Arg406His)	EE	probably damaging
		chr9:g.130444768G>A	NM_001032221.2:c.1631G>A	NM_001032221.2(STXBP1_i001):p.(Gly544Asp)	EE	probably damaging
		chr9:g.130444788C>T	NM_001032221.2:c.1651C>T	NM_001032221.2(STXBP1_i001):p.(Arg551Cys)	ASD	probably damaging
GABRB3	17.72%	chr15:g.26806254T>C	NM_000814.4:c.905A>G	NM_000814.4(GABRB3_i001):p.(Tyr302Cys)	EE	probably damaging
		chr15:g.26828484T>C	NM_000814.4:c.539A>G	NM_000814.4(GABRB3_i001):p.(Glu180Gly)	EE	probably damaging
		chr15:g.26828534C>T	NM_000814.4:c.489G>A	NM_000814.4(GABRB3_i001):p.(Met163Ile)	ASD	probably damaging
		chr15:g.26866564C>T	NM_000814.4:c.358G>A	NM_000814.4(GABRB3_i001):p.(Asp120Asn)	EE	probably damaging
		chr15:g.26866594T>C	NM_000814.4:c.328A>G	NM_000814.4(GABRB3_i001):p.(Asn110Asp)	EE	probably damaging

Table S3: Analysis of genes hit by more than one *de novo* mutation in controls

Gene	Sample size	Gene size┼	Calculated loci †	Count of de novos	fitDNM	Poisson	TADA
ADAMTS2 (MIM 604539)	728	3800	3697	2	0.0345	0.00666	1
AGBL5 (MIM 615900)	728	2782	2782	2	0.0277	0.00211	0.0226
AHNAK2 (MIM 103390)	728	17416	17416	2	0.00605	0.0643	0.014
BYSL (MIM 603871)	728	1342	1342	2	0.0206	0.000659	1
EIF4G1 (MIM 600495)	728	4952	4894	2	0.0622	0.00651	0.0451
FOXK2 (MIM 147685)	728	2019	1967	2	0.000246	0.00227	0.0276
GLIS1 (MIM 610378)	728	1895	1895	2	0.0119	0.00144	0.0202
KIF14 (MIM 611279)	728	5063	5063	2	0.0492	0.00349	1
KIF4A‡ (MIM 300521)	388+340	3819	3819	2	0.0381	0.00163	1
KIF4A‡	368+360	3819	3819	2	0.0374	0.00158	1
LRRK1 (MIM 610986)	728	6180	6180	2	0.00149	0.0122	0.0026
*MUC16** (MIM 606154)	728	43860	15106	0	1	1	1
RGS7 (MIM 602617)	728	1611	1611	2	0.0188	0.000561	1
SNRNP200 (MIM 601664)	728	6591	6591	2	0.0879	0.0106	1
SYNE2 (MIM 608442)	728	21229	21229	2	0.00662	0.0565	0.0097
TDRD5 (MIM 614593)	728	3176	3176	2	0.0412	0.00166	1
*TTN** (MIM 188840)	728	115883	45104	1	0.113	0.569	0.2782
UGT2B4 (MIM 600067)	728	1611	1611	2	0.000145	0.000386	0.0110
USP34 (MIM 615295)	728	10961	10961	2	0.00631	0.0168	0.0758

Note: ┼ gene size: size of all exomes (plus splice sites)

† Calculated loci: removed loci with missense mutations which are not annotated by PolyPhen-2.

‡ Gene located in chromosome X, computed twice, assuming all 20 unknown gender samples are females or males.

* Contain 2 de novo mutations inside transcripts, but some de novo mutations fall into regions not annotated by PolyPhen-2