Supplementary information for:

EXOME SEQUENCING IN MULTIPLEX AUTISM FAMILIES SUGGESTS A MAJOR ROLE FOR HETEROZYGOUS TRUNCATING MUTATIONS

Claudio Toma, Bàrbara Torrico, Amaia Hervás, Rafael Valdés-Mas, Alba Tristán-Noguero, Vanesa Padillo, Marta Maristany, Marta Salgado, Concepción Arenas, Xose S. Puente, Mònica Bayés, Bru Cormand[†]

[†] To whom correspondence should be addressed. Email: bcormand@ub.edu

This file includes:

1. CLINICAL DATA

The affected individuals met DSM-IV criteria for autism, Asperger disorder or pervasive developmental disorder not otherwise specified (PDD-NOS), assessed using ADI-R (Autism-Diagnostic Interview-Revised) and ADOS-G (Autism Diagnostic Observation Schedule-Generic) diagnostic instruments.[1,2] The study was approved by the relevant ethical committee and written informed consent was obtained from all parents. Karyotype and Fragile X tests were previously performed and resulted normal in all families of this study.

CNV STUDY

The array includes 1.9 million non-polymorphic probes and 750,000 SNP probes. Data were analysed with the Affymetrix Chromosome Analysis Suite software considering at least 10 altered probes spanning a minimum of 30kb for deletions/duplications. The CNV data from the CytoScan® HD Array were filtered using the Database of Genomic Variants (http://projects.tcag.ca/variation) in order to exclude all polymorphic structural variants from the study and were also matched with data of CNVs previously found in autism.[3] We considered those structural variants spanning exonic regions of genes and present in both affected probands of a family or present at least in two out of three probands (in the case of family SJD_50, with three affected siblings). Experimental validation was performed by quantitative PCR (qPCR) using two independent probes for each candidate deletion/duplication and one reference probe in the *FOXP2* gene. Experimental details are available upon request.

2. STATISTICAL PROCEDURES

Multiple linear regression analysis

The relation between Nonverbal Intelligence Quotient (NVIQ) and different type of genetic variants (truncating=TRU, non-synonymous damaging changes=NSD, non-synonymous benign changes=NSB and synonymous variants=SYN) was assessed by the following multiple linear regression model[4]:

$$NVIQ = \beta_0 + \beta_1 * TRU + \beta_2 * NSD + \beta_3 * NSB + \beta_4 * SYN + \varepsilon,$$

where ε is the residual term, i.e. the random error component. The adequacy of the model requires that errors are normally and independently distributed random variables with constant variance, and that the relationship between the dependent variable (NVIQ) and the regressors (TRU, NSD, NSB and SYN) is linear. The Shapiro-Wilks test of normality (P = 0.962) ensured the normality assumption. For the other assumptions we considered the plot of studentized errors versus the corresponding predicted values.[4] As observed in Supplementary Figure 8, there is no indication of violation of the assumptions in this plot. The method of least squares was used to estimate the regression coefficients β_j.[4] The hypothesis $H_0: \beta_j = 0$ against $H_1: \beta_j \neq 0$ are helpful in determining the significance of each of the regressors in the model. With our data, we obtained the following estimations and significances:

$$\hat{\beta}_1 = -11.805, P = 0.007;\ \hat{\beta}_2 = -1.206, P = 0.385;$$

$$\hat{\beta}_3 = -1.229, P = 0.484;\ \hat{\beta}_4 = -0.154, P = 0.825.$$

Thus, only truncating variants contribute significantly to NVIQ.

Bonferroni correction in multiple linear regression analysis

It is not common to use corrected tests in multiple regression analysis as the analysis becomes overly conservative. [5] However, if we wish to identify the regressors that are actually related to the

response, a Bonferroni adjustment can be carried out. In this case, the adjusted level of significance α^* is calculated as $\frac{\alpha}{k}$, where α is the nominal level, usually 0.05, and k is the number of regressors. With our data $\alpha^* = \frac{0.05}{4} = 0.0125$.

Simulation study for the multiple linear regression analysis

The sample size is always important in order to have a reasonable probability of obtaining parameter estimates that are statistically significant. When the sample size is small, as in our multiple linear regression analysis (n = 21), a simulation study can be performed to reinforce the accuracy of the results.

By Bootstrap resampling,[6] fifty new simulated analyses with a sample size of 100 individuals each were computed. For each analysis we considered the multiple linear regression model as follows:

$$NVIQ = \beta_0 + \beta_1 * TRU + \beta_2 * NSD + \beta_3 * NSB + \beta_4 * SYN + \varepsilon,$$

and the corresponding analyses with the Bonferroni correction were obtained. Supplementary Table 7 shows the results. Hundred percent of the analyses show that truncating variants contribute significantly to NVIQ. For all simulated analyses, truncating variants correlated negatively with NVIQ (mean±standard deviation = –0.517 ± 0.087) and explained a similar percentage of NVIQ variance in our ASD sample (mean±standard deviation = 27.096 ± 4.934%). For non-synonymous damaging changes (NSD), non-synonymous benign changes (NSB) and synonymous variants (SYN) the results were not significant, confirming those obtained with the original data.

Multiple logistic regression analysis

Let p be the probability that a dichotomous response Y is 1. In our study, Y is defined by $Y = 1$ if the non-synonymous variants were transmitted to both affected siblings in a family and $Y = 0$ if those variants were not transmitted from parents. We are interested in the possible relationship between Y

and the continuous SIFT (S) and Polyphen (P) scores. Thus, we modeled the response *Y* using the multiple logistic regression model[7] as

$$p = \frac{\exp(\beta_0+\beta_1*S+\beta_2*P)}{1+\exp(\beta_0+\beta_1*S+\beta_2*P)}.$$

This model can be written in an equivalent linear form by,

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * S + \beta_2 * P\,.$$

Logistic regression does not require assumptions of linearity, homoscedasticity or normality, and the only real limitation is that the response variable must be binary. The regression coefficients are usually estimated using maximum likelihood estimation.[7] A likelihood-ratio test is used to assess the contribution of regressors.[7,8] With our data we obtained the following estimations and significances:

$$\hat{\beta}_1 = 0.530, P = 0.122;\ \hat{\beta}_2 = 0.051, P = 0.811.$$

Thus, neither SIFT nor Poliphen scores contribute significantly to the variant status. With respect to the required sample size, it is known that the power for logistic regression analysis is based not only on the total sample size, but also on the balance between the outcomes. To have a reasonable power level, the number of observations in the smallest outcome group must be greater or equal to 10 times the number of regressors.[9,10] It has been described that some problems in the parameters estimation can appear for samples with size smaller than 200.[11] In our study, the total sample size is 621, and the outcomes are quite balanced (363 and 258, respectively). Furthermore, the number of observations in the smallest outcome group, 258, is greater than 10*2.

4. SUPPLEMENTARY REFERENCES

1. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994; **24:** 659-685.

2. Lord C, Risi S, Lambrecht L, Cook EH, Jr., Leventhal BL, DiLavore PC *et al.* The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 2000; **30:** 205-223.

3. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466:** 368-372.

4. Montgomery DC, Peck EA. *Introduction to linear regression analysis.* John Wiley & Sons, INC., New York, 1992.

5. Mundfrom DJ, Perrett JJ, Schaffer J, Piccone A. Bonferroni adjustments in tests for regression coefficients. *Multiple Linear Regression Viewpoints* 2006; **32:** 1–6.

6. Efron B, Tibshirani RJ. An introduction to the Bootstrap. Chapman & Hall, 1994.

7. Hosmer DW, Lemeshow, S. *Applied Logistic Regression*. John Wiley & Sons, 2004.

8. Menard, S W. *Applied Logistic Regression*. SAGE, 2002.

9. Bull SB, Mak C, Greenwood CMT. A modified score function estimator for multinomial logistic regression in small samples. *Computational Statistics and Data Analysis* 2002; **39**: 57–74.

10. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 1996; **49**: 1373–1379.

11. DeMaris A. Explained variance in logistic regression. A Monte Carlo study of proposed measures. *Sociological Methods & Research* 2002; **31**: 27–74.

5. SUPPLEMENTARY TABLES

Supplementary Table 1. Statistics for the callability of each sample under study in %. Callability is defined as the percentage of the exome covered by at least 10 reads with a mapping quality score greater than 30, and with a base quality Phred-score greater than 20. The first column lists the family ID (Fam_ID); the following columns list each single individual.

Fam_ID	Father (.1)	Mother (.2)	Sib1 (.3)	Sib2 (.4)	Sib3 (.5)
MT_28	87.02	86.59	86.45	88.09	
SJD_49	83.91	79.14	81.44	82.80	
MT_151	88.30	88.00	88.60	89.10	
SJD_10	87.81	86.96	89.31	89.25	
MT_69	80.80	75.89	75.25	80.01	
MT_109	87.78	88.42	85.91	83.80	
SJD_50	85.90	86.28	85.80	84.17	85.13
SJD_34	77.70	73.64	75.76	75.70	
MT_160	78.03	78.25	79.22	79.34	
MT_76	87.59	84.56	89.93	78.82	

Supplementary Table 2. Rare variants[#] shared by the affected sibs in each of the ten multiplex families.

FAM_ID	Origin	Gene	ENSG	AA change	Chr	Position[†]	Ref	Alt
MT_109	M	*TCTEX1D1*	ENSG00000152760	R125Q	1	67242971	G	A
MT_109	M	*RPS6KC1*	ENSG00000136643	E903Q	1	213415526	G	C
MT_109	F	*SCN1A*	ENSG00000144285	Q1166R	2	166872170	A	G
MT_109	F	*IQSEC1*	ENSG00000144711	R737H	3	12955034	G	A
MT_109	M	*GABRA4*	ENSG00000109158	N383S	4	46930759	A	G
MT_109	M	*BTNL9*	ENSG00000165810	fs indel	5	180477255	ACC	AC
MT_109	F	*BAT2*	ENSG00000204469	H1754R	6	31603009	A	G
MT_109	F	*ZNF311*	ENSG00000197935	fs indel	6	28962882-886	TTCTT	-
MT_109	M	*TSC1*	ENSG00000165699	A186T	9	135797313	G	A
MT_109	M	*OBP2B*	ENSG00000171102	V31M	9	136083971	G	A
MT_109	M	*MUPCDH*	ENSG00000099834	R410W	11	619539	C	T
MT_109	M	*PHLDB1*	ENSG00000019144	R1242Q	11	118518737	G	A
MT_109	F	*SPATA19*	ENSG00000166118	R137*	11	133712408	C	T
MT_109	F	*ADAMTS20*	NA	C446Y	12	43860485	G	A
MT_109	M	*RBM26*	ENSG00000139746	Q810E	13	79916801	C	G
MT_109	M	*CYFIP1*	ENSG00000068793	R866T	15	22969287	G	C
MT_109	M	*CTU2*	ENSG00000174177	R344W	16	88780568	C	T
MT_109	M	*ENO3*	ENSG00000108515	R253C	17	4858791	C	T
MT_109	F	*AZI1*	ENSG00000141577	K541N	17	79171545	G	C
MT_109	F	*CYP4F2*	ENSG00000186115	T152M	19	16003189	C	T
MT_109	F	*ZNF440*	ENSG00000171295	R412*	19	11943225	C	T
MT_109	F	*CABIN1*	ENSG00000099991	R490C	22	24456455	C	T
MT_109	M	*BEND2*	ENSG00000177324	R604G	X	18192321	A	G
MT_151	M	*HTR1D*	ENSG00000179546	I159F	1	23520238	A	T
MT_151	M	*IVL*	ENSG00000163207	fs indel	1	152883252	G	-
MT_151	M	*SULT6B1*	ENSG00000138068	M242I	2	37398632	G	C
MT_151	M	*SLC11A1*	ENSG00000018280	V101M	2	219249897	G	A
MT_151	F	*PRICKLE2*	ENSG00000163637	D807N	3	64084843	G	A
MT_151	F	*GRAMD1C*	ENSG00000178075	R413*	3	113652385	C	T
MT_151	M	*ANK2*	ENSG00000145362	R3906Q	4	114294463	G	A
MT_151	F	*TIGD4*	ENSG00000169989	S382Y	4	153691012	C	A
MT_151	F	*LAP3*	ENSG00000002549	fs indel	4	17606229-232	TGGG	-
MT_151	M	*ZNF498*	ENSG00000197037	A114T	7	99217569	G	A
MT_151	M	*TRIM4*	ENSG00000146833	S477*	7	99489859	C	G
MT_151	F	*GSDMD*	ENSG00000104518	Y206*	8	144642836	C	A
MT_151	F	*EPS8L2*	ENSG00000177106	M62L	11	720080	A	T
MT_151	M	*COG3*	ENSG00000136152	R101I	13	46050463	G	T
MT_151	M	*C16orf71*	ENSG00000166246	S66L	16	4787868	C	T
MT_151	F	*NF1*	ENSG00000196712	R2637Q	17	29684327	G	A
MT_151	M	*DRP2*	ENSG00000102385	E432*	X	100503119	G	T
MT_160	F	*TCHH*	ENSG00000159450	E1323D	1	152081724	G	C
MT_160	M	*POU2F1*	ENSG00000143190	A687V	1	167384806	C	T
MT_160	M	*ADD2*	ENSG00000075340	T461M	2	70905837	C	T
MT_160	M	*ITGAV*	ENSG00000138448	G307R	2	187505657	G	A
MT_160	M	*FAM171B*	ENSG00000144369	S81C	2	187604958	C	G
MT_160	M	*FETUB*	ENSG00000090512	C254S	3	186369071	G	C
MT_160	F	*KIAA0226*	ENSG00000145016	R684Q	3	197409416	G	A
MT_160	F	*NDST3*	ENSG00000164100	I621N	4	119154209	T	A
MT_160	M	*C4orf41*	ENSG00000168538	Q352R	4	184601362	A	G
MT_160	M	*BAI3*	ENSG00000135298	P562L	6	69685183	C	T
MT_160	M	*MYO1G*	ENSG00000136286	P775L	7	45005293	C	T
MT_160	M	*AKAP9*	ENSG00000127914	Q432R	7	91630490	A	G
MT_160	M	*DOCK4*	ENSG00000128512	M1I	7	111428810	G	A
MT_160	M	*C8orf74*	ENSG00000171060	fs indel	8	10557940-941	-	C
MT_160	M	*YWHAZ*	ENSG00000164924	fs indel	8	101936203-204	-	T

MT_160	M	*KANK1*	ENSG00000107104	Q1263H	9	742297	G	C
MT_160	M	*KDM4C*	ENSG00000107077	Y201C	9	6849673	A	G
MT_160	M	*DNHD1*	ENSG00000179532	R2555W	11	6578188	C	T
MT_160	F	*IGHMBP2*	ENSG00000132740	G513D	11	68701932	G	A
MT_160	M	*MMP8*	ENSG00000118113	H227Y	11	102589250	C	T
MT_160	M	*DYNC2H1*	ENSG00000187240	M3762V	11	103175330	A	G
MT_160	F	*GLIPR1*	ENSG00000139278	R209*	12	75892479	C	T
MT_160	F	*AMZ2*	ENSG00000196704	G257R	17	66251859	G	A
MT_28	M	*LCT*	ENSG00000115850	F71C	2	136594528	T	G
MT_28	F	*RABL2A*	ENSG00000144134	fs indel	2	114392641	G	-
MT_28	F	*PET112L*	ENSG00000059691	S378I	4	152622545	G	T
MT_28	M	*FAM200A*	ENSG00000221909	T107A	7	99145712	A	G
MT_28	F	*TJP2*	ENSG00000119139	G958S	9	71863039	G	A
MT_28	M	*IFIT1*	ENSG00000185745	fs indel	10	91162622-623	-	A
MT_28	F	*TRIM29*	ENSG00000137699	K185E	11	120008187	A	G
MT_28	F	*PRICKLE1*	ENSG00000139174	Q789H	12	42853740	G	C
MT_28	M	*ZNF140*	ENSG00000196387	T376M	12	133682990	C	T
MT_28	F	*AQP4*	ENSG00000171885	D305G	18	24436233	A	G
MT_28	M	*DYM*	ENSG00000141627	A307V	18	46812830	C	T
MT_28	M	*ZNF532*	ENSG00000074657	E643K	18	56587446	G	A
MT_28	F	*SIPA1L3*	ENSG00000105738	E830K	19	38610142	G	A
MT_28	F	*NAPSA*	ENSG00000131400	R38C	19	50865542	C	T
MT_28	M	*GPR112*	ENSG00000156920	T2022N	X	135431930	C	A
MT_69	M	*SRBD1*	ENSG00000068784	L496R	2	45780792	T	G
MT_69	F	*HK2*	ENSG00000159399	D246N	2	75101437	G	A
MT_69	F	*TRPM8*	ENSG00000144481	A211T	2	234851324	G	A
MT_69	F	*DNAH1*	ENSG00000114841	L3295V	3	52425336	C	G
MT_69	M	*RFT1*	ENSG00000163933	P138T	3	53156434	C	A
MT_69	F	*CCDC36*	ENSG00000173421	S52*	3	49274079	C	G
MT_69	F	*PDS5A*	ENSG00000121892	I23T	4	39978130	T	C
MT_69	F	*RBM47*	ENSG00000163694	A451V	4	40434858	C	T
MT_69	F	*KLHL8*	ENSG00000145332	S345N	4	88099691	G	A
MT_69	M	*ARSJ*	ENSG00000180801	N306K	4	114824312	C	G
MT_69	F	*STOX2*	ENSG00000173320	P236H	4	184930698	C	A
MT_69	F	*JMY*	ENSG00000152409	R514W	5	78595988	C	T
MT_69	F	*C8orf45*	ENSG00000178460	fs indel	8	67813550	G	-
MT_69	F	*PTPRD*	ENSG00000153707	T1100M	9	8484233	C	T
MT_69	F	*VWA2*	ENSG00000165816	R629W	10	116049011	C	T
MT_69	F	*TMEM9B*	ENSG00000175348	R178L	11	8969931	G	T
MT_69	M	*ROBO3*	ENSG00000154134	G296E	11	124740181	G	A
MT_69	M	*TSPAN32*	ENSG00000064201	fs indel	11	2339146-147	TC	-
MT_69	M	*FAM55B*	ENSG00000204361	W529*	11	114577559	G	A
MT_69	M	*UBE3B*	ENSG00000151148	L148F	12	109924375	C	T
MT_69	M	*HVCN1*	ENSG00000122986	D130N	12	111093062	G	A
MT_69	M	*LRP10*	ENSG00000197324	A459T	14	23345532	G	A
MT_69	F	*SPTBN5*	ENSG00000137877	Q2419*	15	42154996	C	T
MT_69	F	*CCDC102B*	ENSG00000150636	A59G	18	66504176	C	G
MT_69	F	*ZNF154*	ENSG00000179909	L53P	19	58216223	T	C
MT_69	F	*TFIP11*	ENSG00000100109	fs indel	22	26890166-169	AGAT	-
MT_69	M	*RGAG1*	ENSG00000243978	P656L	X	109695812	C	T
MT_76	M	*VANGL1*	ENSG00000173218	R207H	1	116206697	G	A
MT_76	F	*FASTKD2*	ENSG00000118246	W26R	2	207631493	T	C
MT_76	F	*RPN1*	ENSG00000163902	V443A	3	128344444	T	C
MT_76	M	*EPHB3*	ENSG00000182580	T419S	3	184294872	A	T
MT_76	F	*CCDC127*	ENSG00000164366	I257T	5	205425	T	C
MT_76	M	*EGFLAM*	ENSG00000164318	fs indel	5	38412639-642	AAGT	-
MT_76	M	*POLK*	ENSG00000122008	fs indel	5	74882867	A	-
MT_76	F	*MMRN2*	ENSG00000173269	Q479*	10	88703106	C	T
MT_76	F	*CASC4*	ENSG00000166734	N67K	15	44581428	T	G
MT_76	M	*RASL12*	ENSG00000103710	M138V	15	65350778	A	G
MT_76	M	*SV2B*	ENSG00000185518	F581C	15	91832784	T	G
MT_76	F	*CHST4*	ENSG00000140835	K237T	16	71571290	A	C

MT_76	M	*DSG1*	ENSG00000134760	C669R	18	28926066	T	C
MT_76	M	*LRRC8E*	ENSG00000171017	fs indel	19	7965639-640	CA	-
SJD_10	F	*MIIP*	ENSG00000116691	E126A	1	12082414	A	C
SJD_10	F	*UBR4*	ENSG00000127481	A3891S	1	19439148	G	T
SJD_10	F	*IFT172*	ENSG00000138002	A783P	2	27684231	G	C
SJD_10	M	*GPD2*	ENSG00000115159	T547I	2	157427677	C	T
SJD_10	M	*ZAK*	ENSG00000091436	S302T	2	174081896	G	C
SJD_10	F	*CD14*	ENSG00000170458	E58K	5	140012397	G	A
SJD_10	F	*IRAK1BP1*	ENSG00000146243	I238M	6	79607982	A	G
SJD_10	F	*WASF1*	ENSG00000112290	R307H	6	110423393	G	A
SJD_10	F	*C6orf118*	ENSG00000112539	A212T	6	165715177	G	A
SJD_10	F	*ABHD11*	ENSG00000106077	P284L	7	73150986	C	T
SJD_10	F	*ADCY8*	ENSG00000155897	W532*	8	131921998	G	A
SJD_10	M	*HIATL1*	ENSG00000148110	Splicing indel	9	97207469	A	-
SJD_10	M	*HOGA1*	ENSG00000241935	R108H	10	99361725	G	A
SJD_10	M	*NLRP14*	ENSG00000158077	G856D	11	7079615	G	A
SJD_10	M	*ASAM*	ENSG00000166250	D186E	11	122953914	C	G
SJD_10	M	*NAT10*	ENSG00000135372	fs indel	11	34167721-22	-	A
SJD_10	F	*IFT81*	ENSG00000122970	R87H	12	110566766	G	A
SJD_10	F	*ABCB9*	ENSG00000150967	V140M	12	123444365	G	A
SJD_10	M	*GPR84*	ENSG00000139572	fs indel	12	54756741	G	-
SJD_10	M	*CLEC14A*	ENSG00000176435	R141Q	14	38724806	G	A
SJD_10	F	*CHGA*	ENSG00000100604	I440T	14	93401174	T	C
SJD_10	M	*PREX1*	ENSG00000124126	Y732C	20	47271842	A	G
SJD_10	M	*TUBA8*	ENSG00000183785	W346*	22	18609783	G	A
SJD_34	F	*PAX7*	ENSG00000009709	H65Y	1	18960904	C	T
SJD_34	F	*ATP1A4*	ENSG00000132681	K682*	1	160143953	A	T
SJD_34	M	*SNTG2*	ENSG00000172554	L372S	2	1271174	T	C
SJD_34	F	*FLNB*	ENSG00000136068	A1971S	3	58134022	G	T
SJD_34	M	*SHOX2*	ENSG00000168779	D156N	3	157820628	G	A
SJD_34	M	*GRSF1*	ENSG00000132463	H450R	4	71691057	A	G
SJD_34	M	*MAML3*	ENSG00000196782	R433W	4	140811293	C	T
SJD_34	M	*TMEM106B*	ENSG00000106460	D86N	7	12258122	G	A
SJD_34	M	*SOSTDC1*	ENSG00000171243	Y142C	7	16502441	A	G
SJD_34	M	*ZMIZ2*	ENSG00000122515	R573Q	7	44802861	G	A
SJD_34	F	*C10orf122*	ENSG00000175018	I21N	10	127350536	T	A
SJD_34	F	*HRASLS5*	ENSG00000168004	N232H	11	63233635	A	C
SJD_34	F	*KCTD21*	ENSG00000188997	D154G	11	77885347	A	G
SJD_34	M	*ZNF259*	ENSG00000109917	N287D	11	116655126	A	G
SJD_34	F	*ARHGEF12*	ENSG00000196914	V94I	11	120291542	G	A
SJD_34	M	*TECTA*	ENSG00000109927	Q1518*	11	121028796	C	T
SJD_34	F	*KRT84*	ENSG00000161849	R260W	12	52776827	C	T
SJD_34	F	*MAX*	ENSG00000125952	T127I	14	65544567	C	T
SJD_34	F	*CCDC105*	ENSG00000160994	R241Q	19	15131319	G	A
SJD_34	F	*ZNF347*	ENSG00000197937	fs indel	19	53644208-209	AA	-
SJD_49	M	*TMEM198*	ENSG00000188760	R359Q	2	220414569	G	A
SJD_49	M	*PAX3*	ENSG00000135903	T241S	2	223096868	A	T
SJD_49	M	*RUFY3*	ENSG00000018189	M201V	4	71634283	A	G
SJD_49	M	*PJA2*	ENSG00000198961	S91C	5	108714917	A	T
SJD_49	M	*KCNQ5*	ENSG00000185760	V568I	6	73900363	G	A
SJD_49	M	*MTO1*	ENSG00000135297	T679M	6	74207618	C	T
SJD_49	F	*NRG1*	ENSG00000157168	H419D	8	32617875	C	G
SJD_49	M	*CSMD3*	ENSG00000164796	S1270R	8	113599372	A	C
SJD_49	M	*KCNQ3*	ENSG00000184156	T655M	8	133142164	C	T
SJD_49	F	*STIM1*	ENSG00000167323	Q256R	11	4091409	A	G
SJD_49	F	*SCUBE2*	ENSG00000175356	D960H	11	9043479	G	C
SJD_49	M	*P2RX4*	ENSG00000135124	R298Q	12	121666863	G	A
SJD_49	F	*RXFP2*	ENSG00000133105	I417M	13	32366048	C	G
SJD_49	M	*GJB6*	ENSG00000121742	fs indel	13	20796930-931	-	T
SJD_49	F	*MYH6*	ENSG00000197616	V496L	14	23869560	G	C
SJD_49	M	*ZKSCAN2*	ENSG00000155592	N184S	16	25266562	A	G
SJD_49	M	*STX10*	ENSG00000104915	V234L	19	13255274	G	T

SJD_49	M	*KEAP1*	ENSG00000079999	fs indel	19	10610629-633	TGCCC	-
SJD_49	M	*PLTP*	ENSG00000100979	T190M	20	44536520	C	T
SJD_49	M	*SMTN*	ENSG00000183963	H526Y	22	31487777	C	T
SJD_49	M	*KDM5C*	ENSG00000126012	T418A	X	53240828	A	G
SJD_49	M	*ALAS2*	ENSG00000158578	P520L	X	55039960	C	T
SJD_49	M	*FAAH2*	ENSG00000165591	L14W	X	57313299	T	G
SJD_50[&]	M	*METTL13*	ENSG00000010165	R441Q	1	171759604	G	A
SJD_50[&]	M	*TEDDM1*	ENSG00000203730	C83R	1	182369374	T	C
SJD_50[&]	M	*GALNTL6*	ENSG00000174473	F48C	4	173150811	T	G
SJD_50[&]	F	*GLCCI1*	ENSG00000106415	V494I	7	8126004	G	A
SJD_50[&]	M	*ZC3H3*	ENSG00000014164	R663Q	8	144550669	G	A
SJD_50[&]	F	*FBXO22*	ENSG00000167196	fs indel	15	76225440-443	ATAA	-
SJD_50[&]	M	*ZFHX3*	ENSG00000140836	H3611Y	16	72821344	C	T
SJD_50[&]	F	*XRCC1*	ENSG00000073050	F173L	19	44057635	C	G
SJD_50[&]	M	*LRRN4*	ENSG00000125872	G386V	20	6022734	G	T
SJD_50[&]	M	*RBL1*	ENSG00000080839	R421H	20	35684650	G	A
SJD_50[&]	M	*MAP7D2*	ENSG00000184368	G735D	X	20029039	G	A
SJD_50[§]	F	*FNDC7*	ENSG00000143107	R206W	1	109264962	C	T
SJD_50[§]	M	*LAMC1*	ENSG00000135862	T891I	1	183094556	C	T
SJD_50[§]	F	*NCF2*	ENSG00000116701	L70S	1	183556078	T	C
SJD_50[§]	F	*IGFN1*	ENSG00000163395	T3362M	1	201190758	C	T
SJD_50[§]	M	*HYAL1*	ENSG00000114378	D57N	3	50340219	G	A
SJD_50[§]	F	*MRPS22*	ENSG00000175110	M106I	3	139065865	G	A
SJD_50[§]	M	*HES1*	ENSG00000114315	A15V	3	193854213	C	T
SJD_50[§]	M	*CCDC96*	ENSG00000173013	R376L	4	7043539	G	T
SJD_50[§]	M	*LMTK2*	ENSG00000164715	P506L	7	97821294	C	T
SJD_50[§]	F	*SLC45A4*	ENSG00000022567	V527M	8	142228160	G	A
SJD_50[§]	M	*NDOR1*	ENSG00000188566	Y102H	9	140107044	T	C
SJD_50[§]	M	*DCHS1*	ENSG00000166341	R335Q	11	6661841	G	A
SJD_50[§]	M	*GNB5*	ENSG00000069966	D223E	15	52427912	C	G
SJD_50[§]	M	*ISG20*	ENSG00000172183	M159T	15	89198692	T	C
SJD_50[§]	M	*SLC4A1*	ENSG00000004939	L440V	17	42335140	C	G
SJD_50[§]	M	*FAM104A*	ENSG00000133193	R202*	17	71205688	C	T
SJD_50[§]	F	*DNM2*	ENSG00000079805	K419T	19	10906796	A	C
SJD_50[§]	M	*KLK1*	ENSG00000167748	L4R	19	51326994	T	G
SJD_50[§]	M	*ASXL1*	ENSG00000171456	A611T	20	31022346	G	A
SJD_50[§]	M	*LAMA5*	ENSG00000130702	S1640L	20	60902604	C	T
SJD_50[§]	M	*OGFR*	ENSG00000060491	D453N	20	61444324	G	A
SJD_50[§]	M	*PPIL2*	ENSG00000100023	D416Y	22	22048931	G	T
SJD_50[§]	F	*DDX17*	ENSG00000100201	R74C	22	38895486	C	T
SJD_50[§]	M	*CELSR1*	ENSG00000075275	Q1659E	22	46805736	C	G

Abbreviations: F, variant inherited from the father; M, variant inherited from the mother; ENSG, Ensembl Gene ID; Ref, reference allele; Alt, alternative allele found in the patient; AA, amino acid; Chr, chromosome; fs, frame shift; indel, insertion/deletion

[#] Rare variants validated by Sanger sequencing that are predicted to be pathogenic. They include missense changes, truncating changes (nonsense mutations and frameshift indels) and variants altering canonical splice sites or start codons.

[†]Positions are indicated according to the GRCh37/hg19 assembly of the UCSC Genome Browser (genome.ucsc.edu)

[&] Multiplex family with three probands: variant found in all three affected individuals

[§] Multiplex family with three probands: variant found in two out of three affected individuals

Supplementary Table 3. Genes included in the category "Cell junction" found overrepresented in the Gene Ontology (GO) "Cellular component" analysis.

Cellular component	GO number	P-value (adj P)	Gene
Cell junction	30054	0.0002 (0.043)	*ASAM* (adipocyte-specific adhesion molecule)
			CYFIP1 (cytoplasmic FMR1 interacting protein 1)
			DNM2 (dynamin 2)
			DSG1 (desmoglein 1)
			EGFLAM (EGF-like, fibronectin type III and laminin G domains)
			FLNB (filamin B, beta)
			KEAP1 (kelch-like ECH-associated protein 1)
			GABRA4 (gamma-aminobutyric acid (GABA) A receptor, alpha 4)
			GJB6 (gap junction protein, beta 6, 30kDa)
			MYH6 (myosin, heavy chain 6, cardiac muscle, alpha)
			P2RX4 (purinergic receptor P2X, ligand-gated ion channel, 4)
			PJA2 (praja ring finger 2)
			SCN1A (sodium channel, voltage-gated, type I, alpha subunit)
			SV2B (synaptic vesicle glycoprotein 2B)
			TJP2 (tight junction protein 2, zona occludens 2)

Supplementary Table 4. Categories found overrepresented in the Gene Ontology (GO) "Biological process" analysis.

Biological process	GO number	P-value (adj P)
Tissue development	9888	2.4 e-05 (0.03529)
Morphogenesis of embryonic epithelium	16331	3.5 e-05 (0.03529)
Neural tube development	21915	7.7 e-05 (0.044)
Cilium assembly	42384	1.0 e-04 (0.044)
Regulation of action potential	1508	9.0 e-05 (0.044)

Supplementary Table 5. Disease and disorders found overrepresented in the "Top Bio function" by Ingenuity Pathway analysis (IPA).

Diseases and Disorders	Number of genes included	P-value range
Connective Tissue Disorders	17	7.7e-05 – 3.3e-02
Developmental Disorder	25	7.7e-05 – 3.3e-02
Skeletal and Muscular Disorders	21	7.7e-05 – 3.3e-02
Inflammatory Response	3	6.6e-04 – 2.1e-02
Neurological Disease	36	2.5e-03 – 3.3e-02

Supplementary Table 6. Rare disrupting mutations found in parents and not transmitted to affected probands. Father's ID is identified as ".1", whereas mother's ID is identified as ".2".

Individual_ID	Chr	Position[1]	Gene	AA change	Ref	Alt
SJD_49.2	14	24033289	*AP1G2*	E353*	G	T
SJD_49.2	6	130467193-194	*SAMD3*	fs indel	-	C
SJD_49.1	8	87060992	*PSKH2*	W286*	G	A
SJD_34.1	12	29617570	*OVCH1*	fs indel	A	-
SJD_10.2	12	52790723	*KRT82*	E338*	G	T
SJD_10.1	18	28648000-001	*DSC2*	fs indel	-	CT
SJD_10.1	18	7024387	*LAMA1*	W827*	G	A
SJD_10.1	8	55538067	*RP1*	S542*	C	G
MT_76.2	1	10494017-018	*APITD1*	fs indel	AG	-
MT_76.2	19	10659709-712	*ATG4D*	fs indel	AGGT	-
MT_76.2	6	43643079	*MRPS18A*	W176* †	G	A
MT_69.2	3	15628093	*HACL1*	L133*	T	G
MT_28.2	3	3139647	*IL5RA*	K206*	A	T
MT_28.2	4	152609849	*PET112*	fs indel	C	-
MT_28.1	6	127608693-694	*RNF146*	fs indel	GA	-
MT_160.1	10	98995039	*ARHGAP19*	R407*	C	T
MT_160.1	10	88417868	*OPN4*	fs indel	G	-
MT_151.1	2	220039534-535	*CNPPD1*	fs indel	AG	-

Abbreviations: Chr, chromosome; AA, amino acid; Ref: reference allele; Alt: alternative allele; fs, frameshift; indel, insertion/deletion

[1]Positions are indicated according to the GRCh37/hg19 assembly of the UCSC Genome Browser (genome.ucsc.edu)

†Rare variants already described in dbSNP 137

Supplementary Table 7. Simulation study for the multiple linear regression analysis

Simulation	β_1	β_2	β_3	β_4	*r*	%
1	**0,0000****	0,0248	0,0386	0,9285	-0,520	27,008
2	**0,0000****	0,0195	0,0206	0,8532	-0,416	17,337
3	**0,0000****	0,0282	0,0703	0,4871	-0,502	25,244
4	**0,0000****	0,0414	0,0430	0,9512	-0,555	30,762
5	**0,0000****	0,0281	0,0730	0,7511	-0,527	27,819
6	**0,0000****	*0,0067**	0,0573	0,7731	-0,550	30,228
7	**0,0000****	*0,0088**	0,0256	0,6943	-0,526	27,716
8	**0,0000****	0,0173	0,3133	0,6739	-0,479	22,916
9	**0,0000****	0,0161	0,1563	0,9597	-0,490	24,033
10	**0,0000****	*0,0108**	0,1680	0,2668	-0,528	27,876
11	**0,0000****	0,0134	0,0174	0,8375	-0,509	29,459
12	**0,0000****	*0,0096**	0,1753	0,8482	-0,543	28,051
13	**0,0000****	0,1101	0,1033	0,1380	-0,530	33,081
14	**0,0000****	*0,0026**	0,5488	0,2706	-0,575	35,021
15	**0,0000****	0,0716	0,3642	0,1094	-0,592	25,778
16	**0,0000****	0,0775	0,1754	0,1804	-0,508	14,529
17	**0,0000****	0,1491	*0,0047*	0,8937	-0,381	26,974
18	**0,0000****	*0,0036**	0,1767	0,6958	-0,519	29,801
19	**0,0000****	0,0163	0,2328	0,2408	-0,546	30,447
20	**0,0000****	*0,0028**	0,2392	0,7387	-0,552	33,231
21	**0,0000****	0,1195	0,3797	0,2361	-0,576	33,231
22	**0,0000****	0,0675	0,1774	0,5998	-0,581	33,701
23	**0,0000****	0,0688	0,0242	0,6791	-0,515	26,481
24	**0,0000****	0,0314	*0,0042*	0,8626	-0,586	34,381
25	**0,0000****	0,1800	0,2622	0,2861	-0,456	20,834
26	**0,0000****	*0,0019**	0,1443	0,4033	-0,430	18,506
27	**0,0000****	*0,0037**	0,0909	0,8301	-0,508	25,783
28	**0,0000****	0,0165	0,1697	0,4585	-0,518	26,820
29	**0,0000****	0,0325	0,0182	0,8068	-0,519	26,887
30	**0,0000****	*0,0116**	0,0250	0,8437	-0,557	30,980
31	**0,0000****	0,0170	0,0317	0,8816	-0,477	22,761
32	**0,0000****	0,0259	0,0948	0,6757	-0,533	28,425
33	**0,0000****	0,0129	*0,0048*	0,7554	-0,455	20,726
34	**0,0000****	0,0168	0,1757	0,5562	-0,513	26,311
35	**0,0000****	0,0696	0,0355	0,3998	-0,586	34,327
36	**0,0000****	0,1610	*0,0115*	0,9873	-0,569	32,399
37	**0,0000****	0,0424	0,1702	0,3629	-0,515	26,509
38	**0,0000****	*0,0063**	0,1016	0,6489	-0,456	20,799
39	**0,0000****	0,0285	0,1217	0,9828	-0,450	20,287
40	**0,0000****	0,0510	0,0941	0,2577	-0,524	27,475
41	**0,0000****	0,0450	0,0209	0,7007	-0,536	28,781
42	**0,0000****	0,0948	*0,0100*	0,7280	-0,508	25,852

43	**0,0000****	0,0269	0,4189	0,4633	-0,607	36,902
44	**0,0000****	0,1602	*0,0065*	0,6455	-0,495	24,510
45	**0,0000****	0,1370	0,1699	0,5255	-0,548	30,073
46	**0,0000****	*0,0038**	0,5211	0,4385	-0,517	26,777
47	**0,0000****	*0,0031**	0,0191	0,8846	-0,495	24,551
48	**0,0000****	0,0309	0,0126	0,5643	-0,520	27,048
49	**0,0000****	0,0283	0,0746	0,6720	-0,411	16,904
50	**0,0000****	0,6161	0,0600	0,4459	-0,534	28,476

For the simulation study, the first four columns indicate the significance of any regression coefficient, β_j, in the multiple linear regression equation:

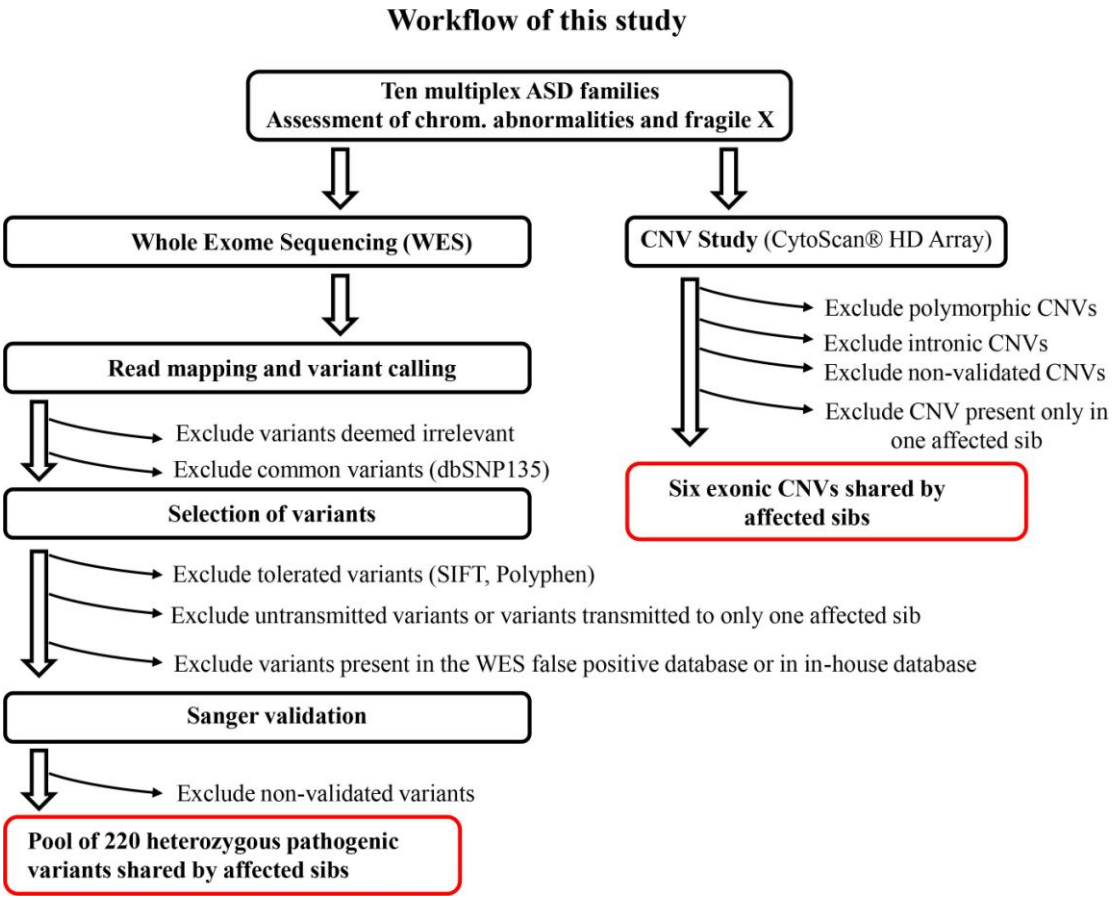
$$NVIQ = \beta_0 + \beta_1 * TRU + \beta_2 * NSD + \beta_3 * NSB + \beta_4 * SYN + \varepsilon.$$

This model assesses the relation between the nonverbal Intelligence Quotient (NVIQ) and the different categories of genetic variants: truncating variants (TRU), non-synonymous damaging changes (NSD), non-synonymous benign changes (NSB) and synonymous variants (SYN).

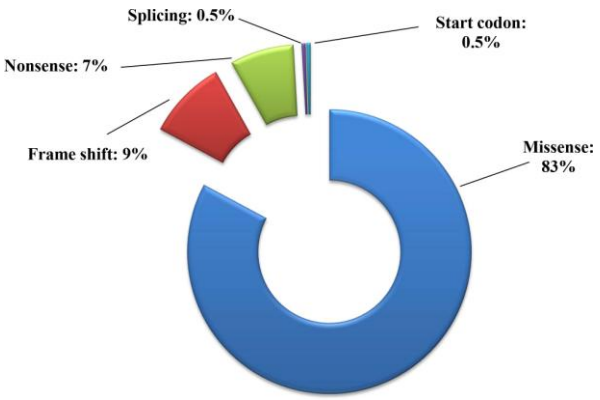
In *italics**, the *p*-values indicating significance using the Bonferroni correction (adjusted significance level $\alpha^* = \frac{0.05}{4} = 0.0125$), considering each simulation as an individual model.

In **black****, the *p*-values indicating significance after multiple test correction (adjusted significance level $\alpha^{**} = \frac{0.125}{50} = 0.00025$). The last two columns show, respectively, the correlation coefficient (*r*) between nonverbal Intelligence Quotient (NVIQ) with truncating (TRU) and the variance (in %) explained by NVIQ in our ASD sample (r^2).

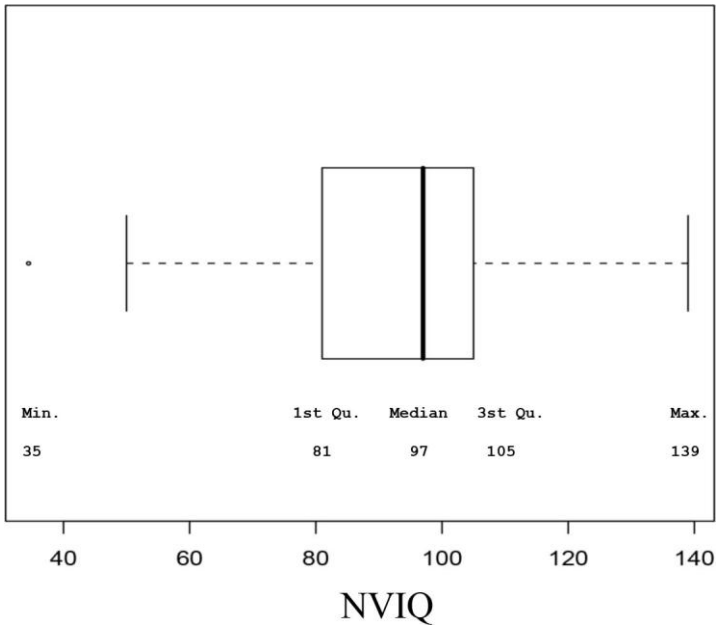
6. SUPPLEMENTARY FIGURES



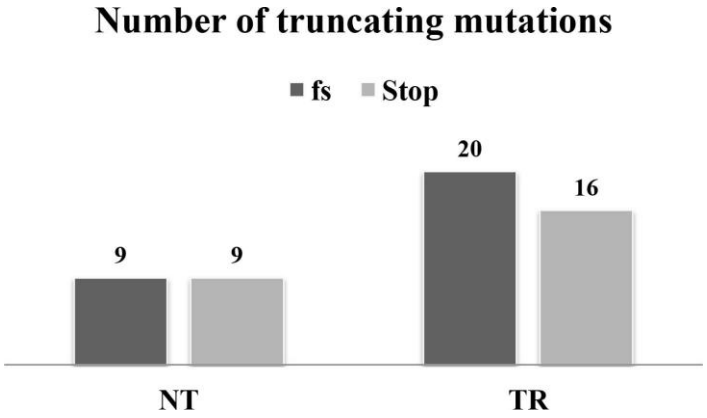
Supplementary Figure 1. Overview of the workflow of this study where different filters and prioritization criteria were applied to select the final pool of genetic rare variants and Copy Number Variants (CNVs). WES: Whole-Exome Sequencing.



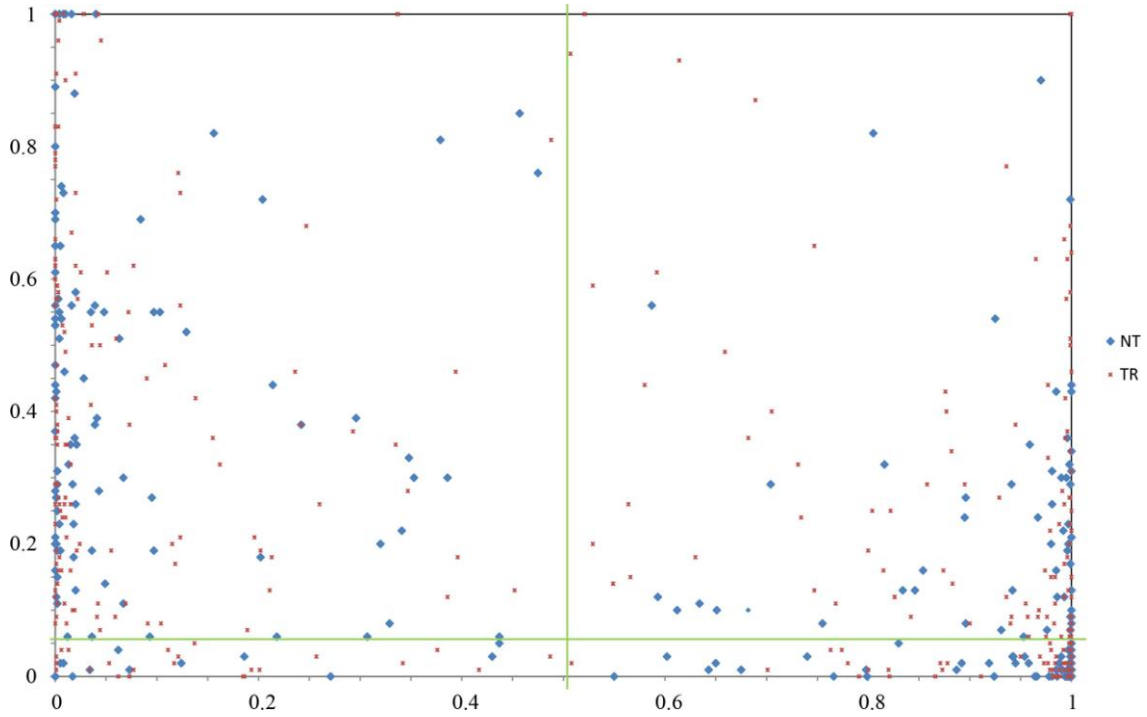
Supplementary Figure 2. Classification of all the predicted pathogenic rare variants validated in our study (n=220).



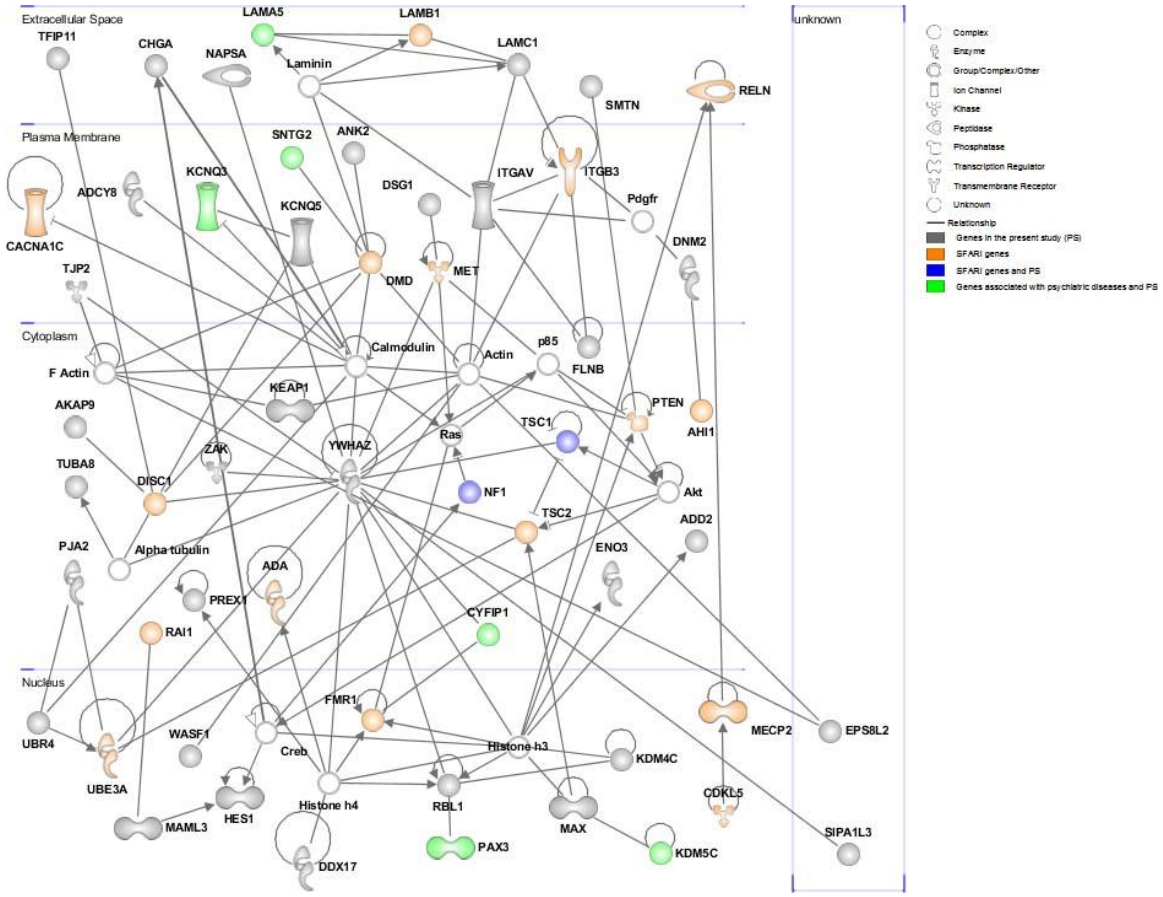
Supplementary Figure 3. Boxplot of Non-Verbal Intelligence Quotient (NVIQ) in our sample of 21 ASD probands. The values are distributed along the scale with an NVIQ mean of 91.57.



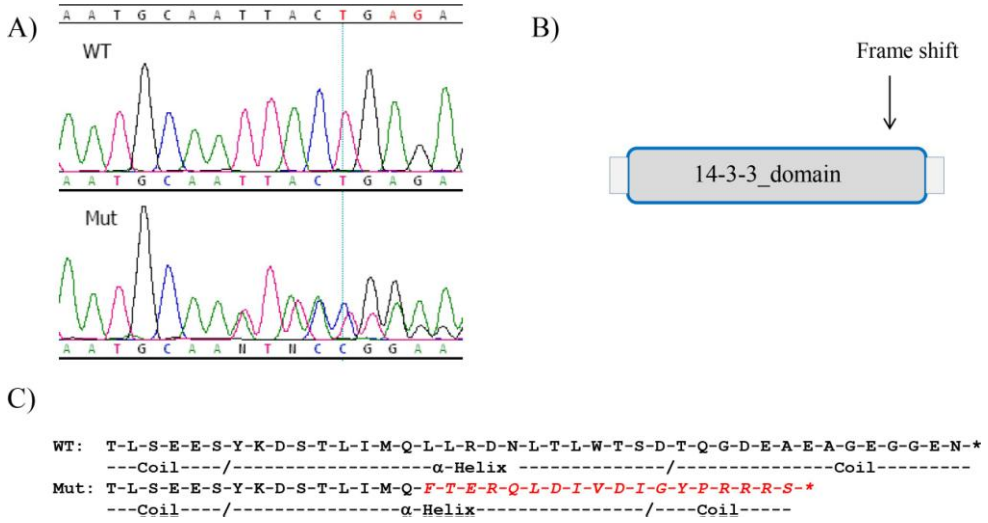
Supplementary Figure 4. Comparison of the total number of truncating mutations that were not transmitted to any of the sibs (NT) versus those transmitted to two siblings (TR). The number of transmitted truncating mutations is significantly higher than those not transmitted (Fisher exact test, $P = 0.015$). Only frame-shift (fs) and nonsense (Stop) mutations were considered. The comparison was made relative to the total number of rare variants (non-synonymous and synonymous) in all families (NT variants: 477; TR variants: 530).



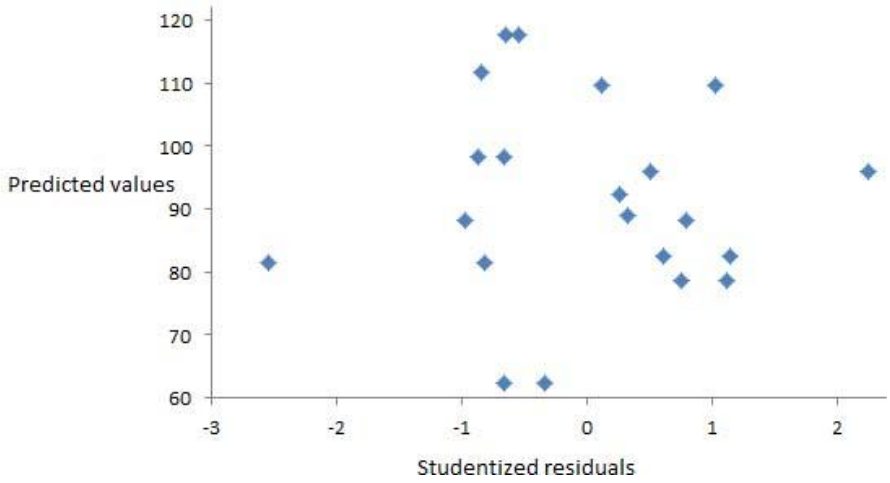
Supplementary Figure 5. The graph shows all the missense rare variants transmitted to two affected probands (TR) and those not transmitted (NT) plotted against the Polyphen score (X axis) and the SIFT score (Y axis). The thresholds that separate pathogenic variants (Polyphen: >0.5, SIFT: <0.05) and non-pathogenic variants are indicated with green lines.



Supplementary Figure 6. Protein-protein interaction analysis (IPA) including all the genes found mutated in our study (220 variants) and 43 genes with strong evidence for their involvement in autism according to the SFARI database (categories: S, 2, 3). Only direct interactions among proteins were considered. Proteins in grey or in green represent genes identified in our study, with green indicating a previous association with autism, epilepsy or schizophrenia. Proteins depicted in orange are those found involved in syndromic or idiopathic cases of autism (SFARI genes), and in blue those found also in our study. Proteins depicted in white are those not present in our study nor in the SFARI list. Up-regulator effects are represented by outward pointing arrows; Down-regulator effects are represented by outward ticks. Circular arrows indicate homotypic interactions.



Supplementary Figure 7. (A) The mutation found in the *YWHAZ* gene in family MT_160 is a nucleotide insertion (the reverse sequence is shown) that produces a frame shift starting at codon 220 of the protein, (B) located in the 14-3-3 domain (SMART prediction). (C) The mutation is predicted to create a premature stop codon, leading to a truncated polypeptide of 237 amino acids.



Supplementary Figure 8. Plot of studentized errors versus predicted values for the multiple linear regression analysis given by the following equation:

$$NVIQ = \beta_0 + \beta_1 * TRU + \beta_2 * NSD + \beta_3 * NSB + \beta_4 * SYN + \varepsilon.$$

This model assesses the relation between the Nonverbal Intelligence Quotient (NVIQ) and the different categories of genetic variants: truncating variants (TRU), non-synonymous damaging changes (NSD), non-synonymous benign changes (NSB) and synonymous variants (SYN). There is no indication of violation of the assumptions of model adequacy.