Supplementary Online Content

Tammimies K, Marshall CR, Walker S, et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *JAMA*. doi:10.1001/jama.2015.10078.

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods

Clinical examination and morphology classification

Families were offered participation in the study after diagnosis of autism spectrum disorder (ASD) by a multidisciplinary team lead by a developmental paediatrician (C.V., V.C., S.L., T.D.). No exclusion was made based on syndromic features or a known or suspected syndrome. The parents/guardians of all patients provided written informed consent. All probands (age range 1.5 to 16 years) recruited for the study were examined by a clinical geneticist and underwent detailed assessments for the presence of birth defects and minor physical anomalies. The proband's antenatal and perinatal histories, medical record and family history were reviewed including radiology and EEG reports. Other screens for birth defects were arranged based on a standard physical examination of the proband, which included a cardiovascular examination (e.g. echocardiogram for a proband with a murmur consistent with a ventricular septal defect). A detailed morphologic examination was performed on all cases by an experienced dysmorphologist (B.A.F). Each body part was assessed for relative size position and/or shape, including the cranium, facies, eyes, nose, philtrum, mouth, oropharynx, ears, neck, thorax, pelvis, back, limbs, genitalia, skin, hair and teeth.[1] Whenever possible, both parents were examined in the same way. Clinical photographs were taken. The proband's ASD was classified as being of regressive onset if there was loss of expressive language skills, either isolated or in conjunction with deterioration of social interaction skills.

All clinical data was entered into a relational database. Each measurement was converted into age, and if available sex-specific, centile.[1] Anomalies from the physical examination were standardized using London Dysmorphology Database codes (n=683) and assigned into three groups: 1) minor physical anomalies that occur in <5% of the population (e.g. single palmar crease, overfolded ear helix); 2) measurement anomalies beyond two standard deviations from the mean (e.g. ocular hypertelorism, large ears) and; 3) descriptive traits which occur in more than 4% of the population which are often familial and which cannot be easily measured (deep-set eyes, broad forehead) (as described by Miles and colleagues[2]). Each proband was assigned a 'Minor Physical Anomaly' (MPA) score by giving one point for each embryologically unrelated MPA, measurement abnormality, or descriptive abnormality that was not present in a parent (e.g. one point would be assigned for a head circumference > +2 standard deviation (SD) with frontal bossing). When parents were not available for examination, we attempted to review photographs. Using the MPA score each proband was classified into one of three groups: 'essential' (MPA score of 0-3), 'equivocal' (MPA score of 4 or 5), or 'complex' (MPA score greater than or equal to 6).

Based on the chart review and the results of radiologic imaging, each proband was assigned a Major Congenital Anomaly (MCA) score with 2 points for each embryologically unrelated birth defect outside the brain. Those who had a brain MRI were given a Brain Anomaly score, with 2 points for each embryologically unrelated structural brain abnormality. Each proband was also classified using a combination of the dysmorphology examination and the presence or absence of birth defects (total morphology score = MPA score + MCA score +/- brain anomaly score). This classification was done as for the MPA only scores.

Microarray analysis

Genomic DNA extracted from peripheral blood or lymphoblast-derived cell lines was used for genotyping at The Centre for Applied Genomic (TCAG) with one of the following high resolution microarrays: Affymetrix 6.0 (N=115) (Affymetrix, Santa Clara, CA), Illumina Omni2.5M-quad (N=55) (Illumina Inc, San Diego, CA), Illumina 1M (N=38), Agilent 1M (N=14) (Agilent Technologies, Santa Clara, CA), Affymetrix Cytoscan HD (N=17) or Illumina 1MDuo (N=5). Quality control and copy number variant (CNV) detection were performed as previously described.[3-8] In brief, a combination of different algorithms was used for the CNV calling depending on the microarray platform used. These included Chromosome Analysis Suite software (Affymetrix, Santa Clara, CA), QuantiSNP[9], PennCNV[10], iPattern[11], Nexus[12] (BioDiscovery Inc., CA), DNA Analytics v.4.0.85 (Agilent Technologies, Santa Clara, CA), DNAcopy[13] and Partek[14] (Partek Inc., St. Louis, MO). We restricted the analysis to CNVs that were detected by at least by two algorithms with 50% reciprocal overlap, spanning a minimum of 5 consecutive microarray probes and size larger than 20 kb. Data from population based control cohorts[15] matched by microarray platform were used to remove all common CNVs (>0.1% frequency).

A total of 150 probands received results from clinical microarray analysis, out of which 125 were performed using oligonucleotide arrays that were used for the analysis. For these 125 samples, DNA extracted from peripheral blood was used to perform array comparative genomic hybridization (aCGH) analysis using the 4x44K microarray platform (Agilent Technologies, Santa Clara, CA) with the custom design described previously[16]; or a custom designed 4X180K oligonucleotide microarray platform (Oxford Gene Technology (OGT), Oxford, UK).

Microarray experiments were performed according to the manufacturer's instructions. Briefly, DNA from the proband and pooled same-sex reference DNA (Promega, Madison, WI) were labeled with Cy3-dCTP and Cy5-dCTP, respectively and were hybridized to the array slide according to the manufacturer's protocol (OGT). The arrays were scanned using the Agilent G2505B microarray scanner. Data analysis was performed using the Agilent Feature Extraction software (10.7.11) and the DNA Analytics version 4.0 (Agilent Technologies) for the 4x44K array, or the CytoSure Interpret Software version 3.4.3 (OGT) for the 4x180K array.

Identified CNVs were classified according to American College of Medical Genetics and Genomics (ACMG)[17] guidelines and pathogenic and variant-of-unknown-significance – likely pathogenic were considered as significant.

Whole-exome sequencing

Exome library preparation and sequencing

DNA from blood or lymphoblast-derived cell lines (100 ng) was used for the exome library preparation with the Ion AmpliSeq Exome Kit (Life Technologies) according to the manufacturer's recommendations. The Ion PI Template OT2 200 Kit v3 was used for preparing the templated Ion PI™ Ion Sphere™ particles. The sequencing was performed using the Ion PI Sequencing 200 Kit v3 and Ion PI Chip v2 in the Ion Proton™ semiconductor sequencing system. Library preparation and sequencing was performed at The Centre for Applied Genomics (TCAG), Toronto, Canada.

Variant calling

Alignment and variant calling of sequencing data were performed using Torrent Suite (v4.0) on the Ion Proton Server. Reads were aligned to the GRCh37/hg19 Ion Torrent reference sequence using TMap software. The Torrent Variant Caller (TVC) 4.0 was used for variant discovery and genotype calling using Ion Proton standard low stringency settings. To obtain more sensitive detection of putative *de novo* variants in the probands, a hotspot bed file was created using TVC for each trio, by combining variants from the proband, mother and father's vcf files. This trio specific hotspot bed file was used in a second phase of variant calling and each sample of the trio was analysed again with custom settings to lower the stringency (details available on request). The secondary variant calling using the hotspot mode enables reference calls to be generated for all positions in the hotspot file. After the secondary variant calling, the newly generated vcf files were processed using an in-house script to remove alleles that were reported in the alternate field by the variant caller but not called as variants. Some reference calls were corrected using an intermediate file containing filtered out variants generated by TVC.

Variant annotation

Variant calls were annotated using a custom pipeline that creates a merged file for each trio. Annotation of the variants was based on Annovar (August 2013)[18] and RefSeq gene models (downloaded from UCSC 2013 February 12). Population frequency for each variant was based the 1000 Genomes Project, the Exome Variant Server hosted by NHLBI (NHLBI GO Exome Sequencing Project) [19] and 69 Complete Genomics public genomes[20]. Additional information about the putative pathogenicity and disease association of the variants were annotated based on different databases as listed below:

Annotation database versions:
Annovar databases for SIFT, PolyPhen2 HDIV, MutationAssessor: ljb23 (compiled 2014 February 22).
dbSNP: version 138.
Clinvar: downloaded 2014 July.
Cosmic: version 68.
HGMD: licensed commercial version downloaded March 2014.
RefSeq: RefGene table from UCSC
OMIM: morbidmap downloaded 2013 October 17.
CGD, HPO, MGI/MPO: downloaded and processed 2014 March 28.
Annovar database for CADD: caddgt10 (compiled 2014 March 10).
Annovar database for PhastCons: mce46way (downloaded from UCSC 2013 November 8),
Annovar database for SegDup: downloaded from UCSC 2011 October 25.
PhyloP: placental mammals: downloaded from UCSC 2013 February 22; 100 vertebrates: downloaded from 2013 September 16.
Repeat: downloaded from UCSC 2013 October 17.

PFAM: downloaded from UCSC 2013 October 21.

De novo variant detection

After the secondary variant calling, we restricted the discovery of *de novo* variants to exonic or splice site positions where an alternative allele was present in the proband and homozygous reference genotype was called in the parents. We further removed positions with coverage less than ten reads in the offspring and both parents. After these steps, five probands had an excess (>2 SD from the mean putative *de novo* variants found in the probands) of putative *de novo* variants and these were excluded from further analysis in this study.

High confidence putative *de novo* variants were selected based on the following criteria: 1) Alternative allele supported by reads mapping to both the forward and reverse strands; 2) The fraction of sequence reads supporting the alternative allele between 0.3 and 0.7 (≥0.9 for X and Y-linked variants in male subjects); 3) Less than 5% of the reads in the parents matching the alternative allele called in the proband; 4) Coverage depth at the position in the proband ≥10% of the sequence depth in the parents; 5) Variant call not overlapping known regions of segmental duplication; 6) Variant frequency < 0.01 in the databases; and 7) Variant not found in the 200 parents sequenced within this study. For short insertions/deletions (indels), we removed variants where an overlapping call (either SNV or indel) was made in the parents within a 5 bp window. For deletions, average read depth of the deletion position and equal size window in flanking regions was extracted and used to calculate an average read depth ratio for the variant. Regions with this ratio <0.9 in the parents were subsequently excluded.

Out of the 95 probands, 68.4% had a putative *de novo* variant/s. In these 65 probands, we detected 158 putative *de novo* SNVs and nine *de novo* indels ranging from 1 to 10 variants per individual. All putative *de novo* variants were subject to secondary validation using PCR and Sanger sequencing. Our validation rate was 94.6% for SNVs and 62.5% for indels. As the majority of DNA samples used for WES was extracted from LCLs, additional confirmation of the validated *de novo* variants in blood-derived DNA was performed. Only 59.6% (66/114 variants in 48 probands) of variants were confirmed in the blood-derived DNA, in contrast to what has been previously reported[21]. In total, 79 variants were confirmed to be *de novo* in blood derived DNA and used for the statistical analysis (eTable 5). Seventeen variants were confirmed to be *de novo* in LCL DNA, however no other source of DNA was available for the confirmation experiments (eTable 5).

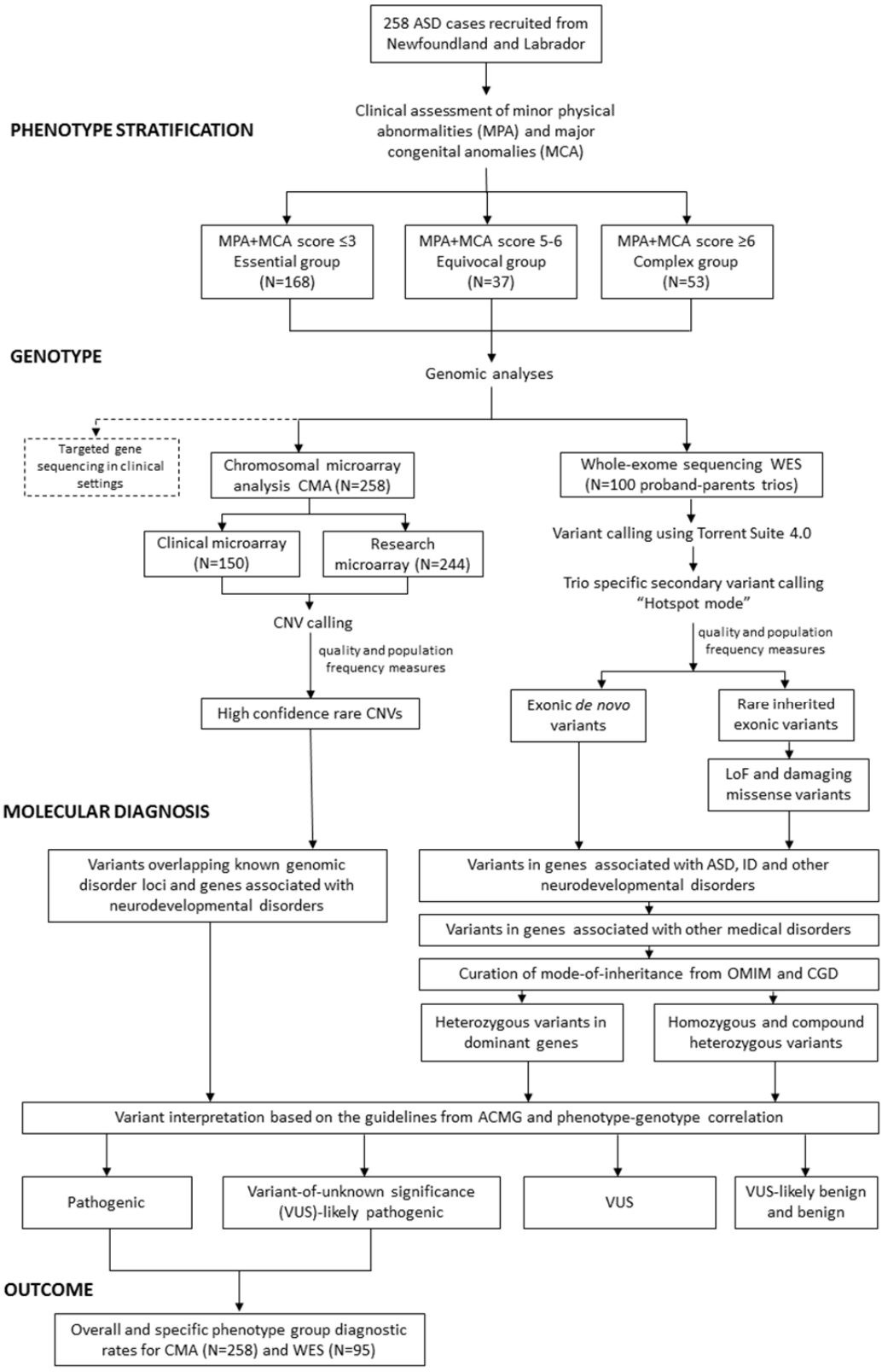
Variant prioritization and molecular diagnosis

All variants present in >1% of the population based on public and internal (based on the 190 parents in this study) databases were removed and analysis was restricted to the remaining rare variants. Further prioritization was made based on the predicted pathogenicity of the variants; all *de novo*, Loss-of-function (i.e. stop-gain, frameshift, essential splice site alterations or missense near splice-site), damaging missense variants (missense predicted damaging by at least 2 of these 5 criteria: SIFT[22] ≤ 0.05, PolyPhen2[23] ≥ 0.453, Mutation Assessor[24] ≥ 2, vertebrate PhyloP[25] ≥ 2, CADD[26] (Phred score ≥ 15) and variants previously reported in disease variant databases (Clinvar and HGMD) were chosen for further evaluation.

The genes affected by the prioritized variants were compared against a list of known ASD and ID genes[4] and genes implicated in neurodevelopmental/behavioral phenotypes in human or mouse from the human phenotype ontology (HPO)[27]. We further curated a list of genes recently implicated in ASD in two large scale sequencing studies[28,29], including 33 genes highlighted in the Autism Sequencing Consortium[28] and additional genes with at least three loss-of-function and/or *de novo* missense variants seen in cases but not in controls (including unaffected siblings). The mode of inheritance for the genes was curated from the Online Mendelian Inheritance in Man (OMIM) database and Clinical Genomics Database (CGD)[30]. All the newly described genes based on *de novo* mutations found in ASD cases were presumed to have autosomal dominant inheritance. Last, we evaluated loss-of-function variants or previously described mutations in other disease causing genes found in Online Mendelian Inheritance in Man and Clinical Genomics Database for medically actionable variants (including the list of genes provided by ACMG) and partial diagnosis based on the phenotype description. Based on all the curated information and filtering, the variants were grouped into four categories 1) heterozygous variants in dominant ASD associated genes, 2) homozygous and compound heterozygous variants in recessive ASD associated genes (including X-linked hemizygous variants), 3) heterozygous variants in dominant genes associated with other disorders and 4) homozygous and compound heterozygous variants in recessive genes associated with other disorders. Variants in these categories were then classified as clinically significant (pathogenic or variant-of-uncertain significance (VUS)-likely pathogenic) or VUS according to the guidelines from the American College of Medical Genetics and Genomics (ACMG).[31,32] We took into consideration the knowledge of penetrance, segregation in the family, disease-causing mechanism, and potential pathogenicity of the previously described mutations.

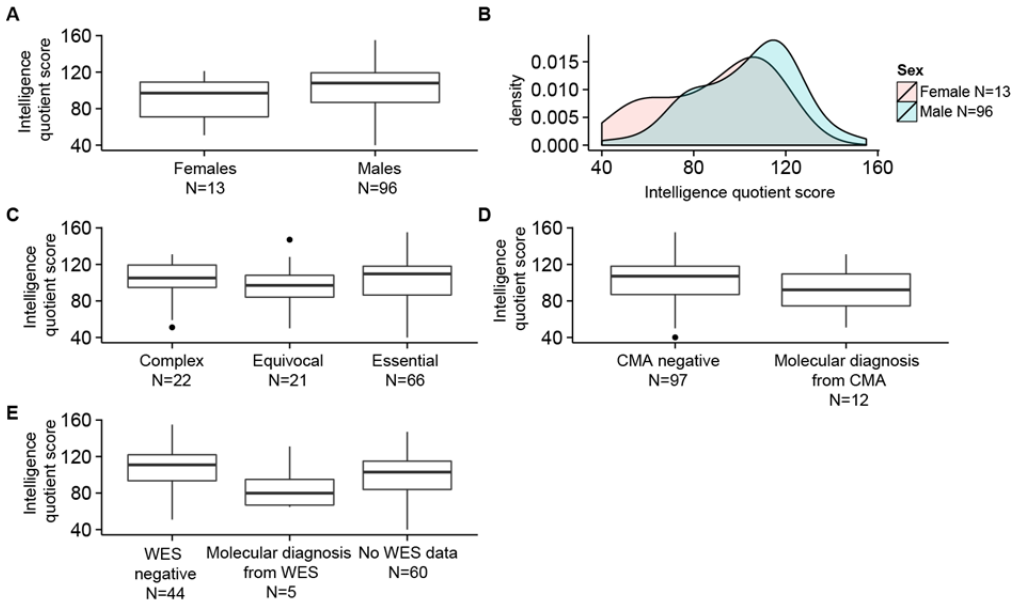
Power calculation for the statistical tests

We have computed the statistical power for the three group comparison test using estimation used for chi-squared test using conventional large effect size (pwr package in R program) followed by pairwise power calculations for Fisher exact test based on the observed proportions (statmod package in R program). Chi-squred test statistics were highly similar to the Fisher exact test resuls and therefore it is reasonable to use same power estimation. All the three group comparisons had estimated power >0.9. Majority of the pairwise comparisons had estimated power >0.5, expect the the comparisons between equivocal and complex groups that had power <0.1.

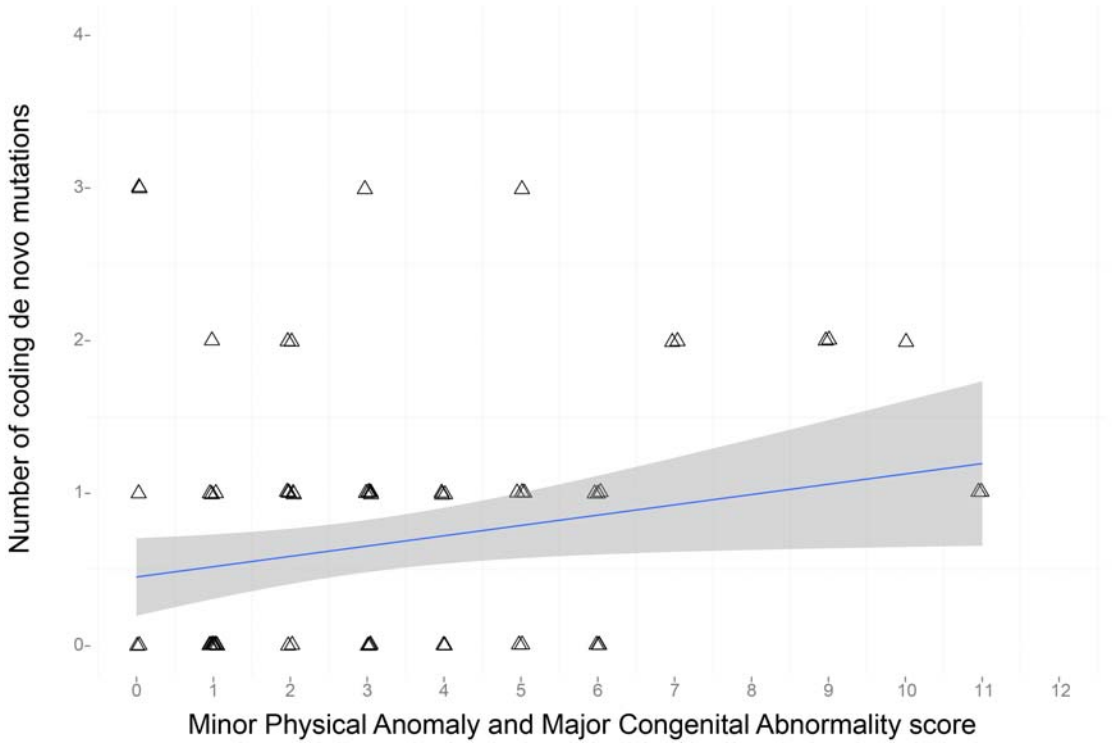


eFigure 1 Project workflow.

Summary of the phenotypic groups, genomic methods used in the study and overview of the molecular diagnosis criteria. ASD, Autism Spectrum Disorder; OMIM, Online Mendelian Inheritance in Man; CGD, Clinical Genomics Database; ACMG, American College of Medical Genetics.



eFigure 2. Intelligence quotient (IQ) scores for 109 probands. A) IQ scores in females and males. B) Density distribution of the IQ scores in females and males. C) IQ scores in probands in the three morphological groups based on the minor physical anomaly and major congenital abnormality score (Table 1 and 3). D) IQ scores for probands that received molecular diagnosis from chromosomal microarray analyses compared with probands that did not receive diagnosis from CMA. E) IQ scores for probands that received molecular diagnosis from whole exome sequencing compared with probands that did not receive molecular diagnosis or did not receive whole-exome sequencing testing.



eFigure 3. Correlation between the phenotype score based on presence of minor physical anomalies and major congenital abnormalities and number of *de novo* mutations affecting coding sequence from 89 individuals. Only *de novo* variants (eTable 5) confirmed in blood DNA were used for the analysis. The correlation was tested using Spearman correlation test (r=0.202, p= 0.029). Correlation is demonstrated in the figure using linear model fitting of the data (blue line) and 95% confidence interval (grey shading). Overlapping data points have been shifted by 0.04 on the X-axis and 0.01 Y-axis.

eTables

Note: The study codes assigned to the cases were unrelated to any identifying information from them.

eTable 1. Whole-exome sequencing mean base depth (mean number of sequences representing one nucleotide) and coverage statistics for 95 probands used for the analysis.

Case	mean base depth on target	mean base depth on targeted CDS	number of targeted-CDS bases at 20X	% of targeted-CDS bases at 20X	% of CDS bases at 1X	number of SNVs on target	number of insertions on target	number of deletions on target	number of rare SNVs on targeted-CDS	number of rare insertions on targeted-CDS	number of rare deletions on targeted-CDS
3-0005-000	111.8	116.9	29958433	0.926	0.959	48307	1458	1859	1141	218	270
3-0025-000	125.8	130.2	30052231	0.928	0.961	48280	1459	2031	1126	225	328
3-0027-000	75.74	77.8	27153539	0.839	0.951	46037	1455	1915	1109	273	299
3-0030-000	124.7	131.4	30220020	0.934	0.960	48522	1470	1926	1092	215	261
3-0031-000	126.3	133.6	30178008	0.932	0.962	48422	1332	2002	1109	145	384
3-0039-000	87.38	91.0	29008342	0.896	0.958	47746	1326	1972	1186	171	400
3-0053-000	114.7	119.4	29908912	0.924	0.960	47815	1384	1757	1057	163	231
3-0055-000	124.6	130.5	30126027	0.931	0.961	48086	1279	1972	1114	153	326
3-0058-000	105	109.3	29148846	0.901	0.958	47184	1475	1942	1132	247	369
3-0059-000	127.3	133.2	30091265	0.930	0.960	47806	1336	2174	1086	142	495
3-0070-000	123.9	129.7	30113372	0.930	0.959	47652	1326	1936	1061	120	368
3-0075-000	130.6	135.7	30512549	0.943	0.962	49121	1529	2355	1110	219	365
3-0077-000	95.34	99.4	28757274	0.888	0.958	47084	1569	1817	1221	286	307
3-0080-000	129.8	136.5	30685323	0.948	0.963	48766	1432	1868	1048	166	316
3-0094-000	102.7	107.0	29469308	0.910	0.960	47813	1547	1779	1102	260	340
3-0095-000	107.6	111.8	29669301	0.917	0.960	47094	1540	1936	1100	273	302
3-0096-000	38.66	118.7	29870978	0.923	0.961	47882	1362	1784	1087	165	288
3-0097-000	108.1	112.1	29344373	0.907	0.958	47785	1420	1848	1100	204	334
3-0101-000	103.7	106.7	28961767	0.895	0.958	47058	1416	1721	1012	178	228
3-0103-000	107	111.6	29778053	0.920	0.961	47682	1431	1742	1099	215	252
3-0109-000	105.6	109.3	29579969	0.914	0.960	47209	1295	1944	1008	152	311
3-0111-000	122.3	128.2	30425354	0.940	0.962	48399	1379	2141	1119	158	391
3-0113-000	121.6	126.6	29788055	0.920	0.959	47891	1279	1950	1037	135	260

Case	mean base depth on target	mean base depth on targeted CDS	number of targeted-CDS bases at 20X	% of targeted-CDS bases at 20X	% of CDS bases at 1X	number of SNVs on target	number of insertions on target	number of deletions on target	number of rare SNVs on targeted-CDS	number of rare insertions on targeted-CDS	number of rare deletions on targeted-CDS
3-0134-000	119.3	123.8	29712770	0.918	0.960	47860	1546	1811	1103	273	277
3-0138-000	111.5	116.4	29496334	0.911	0.959	47737	1457	1828	1069	242	318
3-0140-000	124.6	129.9	29854924	0.922	0.959	48327	1360	1814	1012	158	301
3-0145-000	69.79	72.1	28002711	0.865	0.957	46834	1438	1843	1335	236	376
3-0151-000	119.1	123.6	29963255	0.926	0.959	48604	1416	2202	1056	169	433
3-0155-000	104.5	109.4	29363307	0.907	0.957	47181	1433	1763	1030	177	327
3-0161-000	118.3	125.2	30301487	0.936	0.961	48616	1369	1971	1038	153	385
3-0163-000	111.8	116.7	29883827	0.923	0.960	48507	1391	1891	1074	172	334
3-0178-000	108.2	113.7	29578448	0.914	0.959	48118	1289	1693	968	125	280
3-0180-000	122	128.0	30399266	0.939	0.961	48740	1447	1992	1076	193	302
3-0183-000	118.6	123.4	29904898	0.924	0.960	47795	1539	2030	1110	264	335
3-0186-000	113.4	118.1	29530254	0.912	0.960	47952	1455	1966	1120	188	333
3-0191-000	71.94	74.4	27325342	0.844	0.956	45825	1271	1481	924	170	180
3-0192-000	122.6	128.4	29854180	0.922	0.959	47919	1219	1836	1050	102	273
3-0198-000	76.82	78.2	26119717	0.807	0.951	44733	1370	1565	1011	225	225
3-0208-000	122.6	128.8	29898619	0.924	0.960	47404	1330	2049	970	150	361
3-0209-000	59	61.8	25070520	0.775	0.947	44813	1340	2383	1101	248	428
3-0211-000	76.28	79.6	28208660	0.872	0.957	46744	1353	1696	1046	229	255
3-0215-000	137.9	143.7	30279801	0.936	0.960	48104	1410	1877	1048	171	265
3-0223-000	111.4	115.9	29400529	0.908	0.960	47328	1308	1937	1061	159	305
3-0227-000	91.17	95.3	28821781	0.890	0.958	47449	1441	2195	1095	255	316
3-0229-000	129.2	134.9	29791032	0.920	0.959	47790	1247	1941	949	111	308
3-0241-000	58.78	61.0	25513101	0.788	0.951	44453	1605	2365	1227	347	468
3-0242-000	52.05	53.5	24961250	0.771	0.954	44229	1360	1548	1099	257	299
3-0244-000	130.4	136.5	30278992	0.936	0.961	48012	1414	1786	1060	159	246
3-0253-000	142.2	149.4	28561941	0.882	0.959	47693	1920	2244	1414	441	466
3-0254-000	96.77	100.4	29114828	0.900	0.958	47058	1275	2033	1013	156	313
3-0258-000	91.49	95.3	28628164	0.884	0.957	46699	1551	1808	1160	284	367
3-0261-000	73.87	75.0	26224714	0.810	0.949	44828	1278	1658	1104	197	291
3-0262-000	99.99	104.2	28547668	0.882	0.957	47145	1469	1671	1153	257	246
3-0270-000	117.8	121.9	29121884	0.900	0.957	47071	1510	1708	1090	239	283
3-0284-000	122.4	126.6	29122337	0.900	0.957	47761	1204	1836	956	121	243
3-0294-000	85.83	89.6	28959048	0.895	0.958	47106	1404	1769	1156	232	271

Case	mean base depth on target	mean base depth on targeted CDS	number of targeted-CDS bases at 20X	% of targeted-CDS bases at 20X	% of CDS bases at 1X	number of SNVs on target	number of insertions on target	number of deletions on target	number of rare SNVs on targeted-CDS	number of rare insertions on targeted-CDS	number of rare deletions on targeted-CDS
3-0295-000	117.2	123.5	29754879	0.919	0.959	47650	1420	1909	1109	205	317
3-0305-000	135.4	141.1	30400496	0.939	0.961	48839	1502	1968	1070	205	329
3-0307-000	119	123.5	30038184	0.928	0.960	48869	1582	2101	1186	266	379
3-0309-000	108.5	113.4	28714744	0.887	0.957	46877	1310	1938	1082	193	387
3-0312-000	90.13	93.1	28060938	0.867	0.957	46684	1526	1699	987	259	217
3-0314-000	114.6	119.9	29773878	0.920	0.959	48394	1329	1873	1049	128	267
3-0318-000	75.83	78.2	26834663	0.829	0.951	45436	1185	1476	823	117	216
3-0319-000	119.6	124.4	29724486	0.918	0.959	47956	1567	1966	1236	254	364
3-0321-000	101.6	106.3	29282714	0.905	0.959	46997	1386	1807	1036	207	298
3-0323-000	124.1	129.0	29990832	0.927	0.960	48044	1474	1781	1029	200	286
3-0324-000	108.9	113.9	29696295	0.917	0.960	48115	1672	1886	1179	300	290
3-0325-000	126.1	131.6	30189854	0.933	0.961	47949	1371	1913	1072	158	307
3-0326-000	100.4	104.8	28583595	0.883	0.958	47497	1540	1996	1144	261	437
3-0329-000	108.5	112.5	29618194	0.915	0.959	48580	1329	1925	1104	149	280
3-0330-000	102.6	106.9	29487142	0.911	0.959	47742	1423	1804	1099	213	310
3-0332-000	103.4	108.5	28974260	0.895	0.958	47660	1535	1647	1128	260	255
3-0336-000	117.1	122.8	30045564	0.928	0.960	47745	1368	1746	1045	173	302
3-0338-000	114.6	120.3	30352626	0.938	0.962	48681	1455	2042	1282	191	363
3-0346-000	124.1	130.5	29670628	0.917	0.959	47723	1366	1987	1119	192	320
3-0359-000	115.5	119.3	29787666	0.920	0.960	48341	1496	2078	1177	218	395
3-0368-000	117.8	121.2	29827736	0.922	0.961	48581	1299	1949	1000	132	306
3-0371-000	122.3	127.3	30006918	0.927	0.960	48004	1374	1710	1023	169	274
3-0385-000	108.8	114.1	29922635	0.924	0.960	51913	1876	2184	2427	389	418
3-0391-000	114.5	119.8	29649978	0.916	0.960	48250	1380	1789	997	159	264
3-0393-000	129.1	135.6	30714129	0.949	0.962	48780	1330	1949	1106	123	331
3-0398-000	121.7	126.9	30212672	0.933	0.960	49018	1444	2413	1265	182	464
3-0402-000	124.9	130.7	30178202	0.932	0.962	49787	1458	2327	1256	199	495
3-0405-000	112.4	117.5	29654413	0.916	0.959	47499	1369	1755	1013	162	237
3-0409-000	113.9	118.6	29925839	0.925	0.960	48376	1437	1799	1075	194	288
3-0410-000	97.07	101.4	29865055	0.923	0.961	48111	1294	2566	1296	139	725
3-0419-000	115.3	120.2	30186067	0.933	0.961	47941	1324	1703	1064	129	237
3-0459-000	125.7	132.7	30551292	0.944	0.962	48722	1507	1771	1122	208	311
3-0465-000	101.6	107.0	30115929	0.930	0.963	47910	1326	3300	1253	171	1081

Case	mean base depth on target	mean base depth on targeted CDS	number of targeted-CDS bases at 20X	% of targeted-CDS bases at 20X	% of CDS bases at 1X	number of SNVs on target	number of insertions on target	number of deletions on target	number of rare SNVs on targeted-CDS	number of rare insertions on targeted-CDS	number of rare deletions on targeted-CDS
3-0469-000	127.8	133.9	30254490	0.935	0.960	49089	1422	2245	1120	161	435
3-0477-000	119	125.2	29712576	0.918	0.959	47823	1255	2165	1078	124	449
3-0503-000	126.6	133.8	30460763	0.941	0.962	48657	1429	1837	1182	185	262
3-0504-000	111.1	115.5	29471606	0.911	0.960	47661	1345	1743	998	191	292
3-0516-000	102.9	106.6	29481025	0.911	0.959	47393	1458	1899	1077	220	393
3-0519-000	121	126.8	29886902	0.923	0.960	47989	1460	2240	1154	212	390

CDS, coding DNA sequence; SNV, single nucleotide variant.

eTable 2. Description of 12 probands with syndromes diagnosed prior to chromosomal microarray and whole-exome sequencing in the clinical assessment and/or through targeted sequencing.

Case	Sex	Suspected syndrome	Diagnosis method	Molecular diagnosis mutation details[a]	Diagnosis related to Autism spectrum disorder
3-0032-000[b]	Female	Down syndrome (MIM:190685)	Clinical diagnosis	details eTable3	Yes
3-0110-001	Male	Bardet-Biedl syndrome (MIM: 605231)	Targeted sequencing	*BBS6* p.F94fsX103/F94fsX103	*Yes*
3-0102-000	Male	Fetal alcohol syndrome	Clinical diagnosis	-	*Yes*
3-0111-000[b]	Male	Unspecified overgrowth syndrome	Clinical diagnosis	details eTable4	Yes
3-0119-000	Male	Neurofibromatosis 1 (MIM: 162200)	Clinical diagnosis	-	Yes
3-0197-000	Female	Cowden syndrome (MIM: 158350)	Targeted sequencing	*PTEN* p.Gly165Val/+ (*de novo*)	Yes
3-0400-000	Male	Fetal alcohol syndrome	Clinical diagnosis	-	Yes
3-0477-000[b]	Male	Smith-Magenis syndrome (MIM: 182290)	Clinical diagnosis	details eTable3	Yes
3-0484-000	Male	Autosomal Dominant Polycystic kidney disease (MIM: 173900)	Clinical diagnosis	Targeted sequencing negative	No
3-0503-000	Female	Osteogenesis Imperfecta (MIM: 615220)	Targeted sequencing	*WNT1 p.Gln96Profs*54/Gln96Profs*54*	Potentially
3-0506-000	Female	Sotos syndrome (MIM: 117550)	Clinical diagnosis	-	Yes
3-0520-000	Male	Ocular motor apraxia, Cogan type (MIM: 257550)	Clinical diagnosis	-	Potentially

[a] Mutation identified from clinical targeted sequencing; [b] Molecular diagnosis resolved in this study; "-" indicates that no targeted sequencing was perform or results not received. MIM, Mendelian Inheritance in Man.

eTable 3. Molecular diagnoses in 24 probands identified from chromosomal microarray analysis

Case	Sex	Chr	Start[a]	End	Size (bp)	Type of copy number variant	Platform[b]	Mendelian Inheritance in Man	Syndrome / genes	Inheritance
3-0002-000	Male	7	154144653	154213814	69162	loss	AFFY6	126141	*DPP6*	*de novo*
3-0021-000	Female	15	22770422	28526905	5145962	gain	CytoscanHD	105830	Angelman syndrome	maternal
3-0032-000	Female	21	-	-	-	gain	AFFY6	190685	Trisomy 21	*de novo*
3-0039-000	Female	22	47939539	51234455	3294917	loss	AFFY6	606232	22q13 deletion syndrome	paternal, translocation
3-0049-001	Male	6	156743463	158569886	1826424	loss	ILMN1M	614556 (614562)	*ARID1B*	*de novo*
3-0066-000	Male	9	119431463	119530849	99387	loss	ILMN1M	612856; 602290	*ASTN2/TRIM32*	maternal
3-0067-000	Male	16	89273092	89518181	245090	loss	AFFY6	611192 (148050)	*ANKRD11 (KBG SYNDROME)*	*de novo*
3-0075-000	Male	X	-	-	-	gain	CytoscanHD		Klinefelter syndrome (XX,Y)	*de novo*
3-0095-000	Male	4	-	-	-	UPD	ILMN1M		Paternal isodisomy chr4	*de novo*
3-0097-000	Female	X	169922	6008799	5838878	loss	AFFY6	300427	*NLGN4*	*de novo*
3-0133-000	Female	16	29652799	30199507	546709	gain	AGLT1M	611913	16p11.2 microduplication syndrome	maternal
3-0134-000	Male	6	3154826	3224209	69384	loss	ILMN1M	615101;612850	*TUBB2A, TUBB2B*	paternal

Case	Sex	Chr	Start[a]	End	Size (bp)	Type of copy number variant	Platform[b]	Mendelian Inheritance in Man	Syndrome / genes	Inheritance
3-0152-000	Male	22	22297934	23258994	961061	loss	AFFY6	611867	22q11.2 distal deletion syndrome (critical region)	NA
3-0251-000	Male	17	34437128	36244358	1807231	gain	AFFY6	614526	17q12 duplication syndrome	paternal
3-0269-000	Male	16	29567296	30321320	754025	loss	CytoscanHD	613604	16p11.2-p12.2 microdeletion syndrome	maternal
3-0313-000	Male	16	3420386	5121367	1700981	gain	ILMN2.5M	613458	16p13.3 microduplication	not maternal, father unavailable
3-0338-000	Male	9	138924023	139929435	1005412	loss	AFFY6	190198	> 50 genes affected including *NOTCH1*	*de novo*
3-0360-000	Male	20	62920	1340439	1277520	loss	AFFY6	611663	> 25 genes affected including *TBC1D20*	NA
3-0368-000	Male	2	50568338	51057138	488801	loss	ILMN2.5M	600565	*NRXN1*	*de novo*
3-0391-000	Male	22	50690823	51197850	507028	loss	CytoscanHD	606230 (606232)	*SHANK3* (Phelan-Mcdermid syndrome)	*de novo*
3-0392-000	Female	22	50040382	51234455	1194074	loss	AFFY6	606232	22q13 deletion syndrome	*de novo*
3-0437-000	Male	22	18649190	21915521	3266332	gain	CytoscanHD	608363	22q11 duplication syndrome	paternal
3-0477-000	Male	17	16757132	20340342	3583211	loss	ILMN2.5M	182290	Smith-Magenis Syndrome	*de novo*
3-0523-000	Male	7	73799887	75034321	1234435	gain	CytoscanHD	609757	7q11.23 duplication syndrome	maternal

[a] Genomic coordinates from Hg19 build.[b] Genotyping platform used for the CNV calling. AFFY6, Affymetrix 6.0 (Affymetrix, Santa Clara, CA); ILMN2.5M, Illumina Omni2.5M-quad (Illumina Inc, San Diego, CA); ILM1MIllumina 1M; AGILT1M, Agilent 1M (Agilent Technologies, Santa Clara, CA); CytoscanHD, Affymetrix Cytoscan HD. Chr, chromosome; NA denotes that inheritance testing could not be performed.

eTable 4. Molecular diagnoses (nine variants) identified in eight probands from whole-exome sequencing of 95 probands.

Case	Sex	Gene symbol	Mendelian Inheritance in MAN	Chr	Position[a]	Reference allele	Alternative allele	Transcript ID	Mutation	inheritance	1000 genomes[b]	NHLBI Exome server[c]	Evidence for pathogenicity and additional comments
3-0027-000	Male	*SLITRK5*	609680	13	88329338	G	A	NM_015567	c.G1695A p.W565X	maternal	0	0	Mutations identified in ASD cases, knockout mouse phenotype consistent with neurodevelopmental disorder. Autosomal dominant inheritance pattern.
3-0075-000	Male	*EDA*	300451	X	68836277	T	A	NM_001005609	c.T125A p.L42H	maternal	0	0	Mutations seen in cases with ectodermal dysplasia and behavioral phenotypes. The proband has features of ectodermal dysplasia.
3-0095-000	Male	*WAC*	615049	10	28879727	GCAAGCAACA	-	NM_100486	c.576_585delGCAAGCAACA p.Q192fs*31	*de novo*	0	0	Mutations identified in ASD and ID cases. Autosomal dominant inheritance pattern
		FAT4	612411	4	126320038	A	G	NM_024582	hom c.A5275G p.I1759V	paternal	0.0009	0.002461	Mutations identified in cases with Van Maldergem syndrome 2, features of which seen in the proband
3-0111-000	Male	*BBS1*	209901	11	66293652	T	G	NM_024649	hom c.T1169G p.M390R	inherited	0.0014	0.002002	Known mutation identified in Bardet Biedl syndrome cases

Case	Sex	Gene symbol	Mendelian Inheritance in MAN	Chr	Position[a]	Reference allele	Alternative allele	Transcript ID	Mutation	inheritance	1000 genomes[b]	NHLBI Exome server[c]	Evidence for pathogenicity and additional comments
3-0140-000	Male	*ASH1L*	607999	1	155324288	G	A	NM_018489	c.C7189T p.R2397X	*de novo*	0	0	Mutations identified in ASD and ID cases. Autosomal dominant inheritance pattern.
3-0211-000	Male	*FGFR2*	176943	10	123258119	T	C	NM_023029	c.A1295G p.D432G, near splice site	paternal	0	0	Mutations identified in cases with craniosynostosis associated syndromes with variable phenotypes including ASD and ID
3-0261-000	Male	*SCN2A*	182390	2	166237202	G	C	NM_001040143	c.G4409C p.G1470A	*de novo*	0	0	Mutations identified in ASD, ID and epilepsy cases.
3-0459-000	Male	*TCF12*	600480	15	57535740	-	G	NM_207037	c.1106_1107insG p.P369fs*20	maternal	0	0	Mutations identified in cases with craniosynostosis with variable phenotypes including ASD and ID

[a] Genomic coordinates from Hg19 build, [b] Allele frequency from the1000 genomes project, [c] Allele frequency from NHLBI exome variant server obtained January 2013. Chr, chromosome; ASD, autism spectrum disorder; ID, intellectual disability

eTable 5. 96 *de novo* variants identified and verified in 55 probands from the 95 trios analyzed by whole-exome sequencing.

Case	Chr	Position[a]	Reference allele	Alternative allele	Zygosity	Gene symbol	transcript ID	exon number	coding change	protein change	Effect
3-0005-000	10	104129511	G	A	ref-alt	*GBF1*	NM_001199378	exon25	c.G3194A	p.R1065Q	nonsynonymous
3-0025-000	3	119109669	C	T	ref-alt	*ARHGAP31*	NM_020754	exon7	c.C720T	p.A240A	synonymous
3-0027-000	11	26587227	T	C	ref-alt	*MUC15*	NM_001135092	exon3	c.A260G	p.E87G	nonsynonymous
3-0027-000	19	39025386	C	T	ref-alt	*RYR1*	NM_001042723	exon78	c.C11271T	p.L3757L	synonymous
3-0031-000	11	55761926	A	G	ref-alt	*OR5F1*	NM_003697	exon1	c.T176C	p.M59T	nonsynonymous
3-0031-000	17	61559007	CTGGGAAGGGTCGATG	-	ref-alt	*ACE*	NM_000789	exon7	c.1026_1041del	p.342_347del	frameshift deletion
3-0053-000	1	244587309	C	T	ref-alt	*ADSS*	NM_001126	exon6	c.G527A	p.R176Q	nonsynonymous
3-0058-000	2	47612316	A	G	ref-alt	*EPCAM*	NM_002354	exon8	c.A870G	p.R290R	synonymous
3-0058-000	16	68389650	A	G	ref-alt	*PRMT7*	NM_001290018	exon17	c.A1675G	p.R559G	nonsynonymous
3-0059-000	19	50156162	T	C	ref-alt	*SCAF1*	NM_021228	exon7	c.T2516C	p.V839A	nonsynonymous
3-0070-000	1	248309147	G	C	ref-alt	*OR2M5*	NM_001004690	exon1	c.G698C	p.G233A	nonsynonymous
3-0075-000	11	6644563	G	A	ref-alt	*DCHS1*	NM_003737	exon21	c.C8344T	p.R2782W	nonsynonymous
3-0075-000	22	32112105	T	C	ref-alt	*PRR14L*	NM_173566	exon4	c.A1720G	p.S574G	nonsynonymous
3-0080-000	18	77895382	T	G	ref-alt	*ADNP2*	NM_014913	exon4	c.T2086G	p.C696G	nonsynonymous

Case	Chr	Position[a]	Reference allele	Alternative allele	Zygosity	Gene symbol	transcript ID	exon number	coding change	protein change	Effect
3-0080-000	X	55513637	G	A	hom	*USP51*	NM_201286	exon2	c.C1736T	p.S579F	nonsynonymous
3-0094-000	2	239040083	G	A	ref-alt	*ESPNL*	NM_194312	exon9	c.G2728A	p.A910T	nonsynonymous
3-0095-000	10	28879727	GCAAGCAACA	-	ref-alt	*WAC*	NM_100486	exon6	c.576_585del	p.192_195del	frameshift deletion
3-0096-000	9	86518307	C	T[b]	ref-alt	*KIF27*	NM_017576	exon4	c.G1126A	p.G376S	nonsynonymous
3-0096-000	9	129458577	C	T[b]	ref-alt	*LMX1B*	NM_002316	exon8	c.C1035T	p.N345N	Synonymous
3-0101-000	6	99771430	G	A	ref-alt	*FAXC*	NM_032511	exon4	c.C713T	p.T238M	Nonsynonymous
3-0109-000	10	28149756	T	A[b]	ref-alt	*ARMC4*	NM_001290021	exon12	c.A1394T	p.H465L	Nonsynonymous
3-0109-000	10	115973247	G	A[b]	ref-alt	*TDRD1*	NM_198795	exon15	c.G1974A	p.T658T	Synonymous
3-0109-000	5	35740319	G	A[b]	ref-alt	*SPEF2*	NM_024867	exon23	c.G3280A	p.D1094N	nonsynonymous
3-0109-000	9	97333814	G	A[b]	ref-alt	*FBP2*	NM_003837	exon4	c.C497T	p.A166V	nonsynonymous
3-0111-000	8	43157164	T	C	ref-alt	*POTEA*	NM_001005365	exon5	c.T744C	p.S248S	synonymous
3-0134-000	5	151166277	G	A	ref-alt	*G3BP1*	NM_198395	exon2	c.95+1G>A		splice site
3-0140-000	1	155324288	G	A	ref-alt	*ASH1L*	NM_018489	exon16	c.C7189T	p.R2397X	stopgain
3-0140-000	10	48414275	T	C	ref-alt	*GDF2*	NM_016204	exon2	c.A593G	p.Q198R	nonsynonymous
3-0140-000	3	47376220	G	A	ref-alt	*KLHL18*	NM_025010	exon6	c.G809A	p.R270Q	nonsynonymous

Case	Chr	Position[a]	Reference allele	Alternative allele	Zygosity	Gene symbol	transcript ID	exon number	coding change	protein change	Effect
3-0161-000	13	26967554	A	G	ref-alt	*CDK8*	NM_001260	exon7	c.A697G	p.I233V	nonsynonymous
3-0161-000	19	42813829	G	C	ref-alt	*PRR19*	NM_199285	exon2	c.G93C	p.K31N	nonsynonymous
3-0161-000	3	25671897	G	A	ref-alt	*TOP2B*	NM_001068	exon12	c.C1454T	p.A485V	nonsynonymous
3-0163-000	16	86585881	G	C	ref-alt	*MTHFSD*	NM_001159378	exon3	c.C141G	p.A47A	synonymous
3-0163-000	5	66400357	G	C	ref-alt	*MAST4*	NM_015183	exon9	c.G743C	p.R248P	nonsynonymous
3-0178-000	4	3317988	G	A	ref-alt	*RGS12*	NM_002926	exon2	c.G91A	p.G31S	nonsynonymous
3-0180-000	10	121579002	C	G	ref-alt	*INPP5F*	NM_001243194	exon1	c.C7G	p.H3D	nonsynonymous
3-0186-000	1	150636154	T	C	ref-alt	*GOLPH3L*	NM_018178	exon3	c.A269G	p.Y90C	nonsynonymous
3-0191-000	22	40660882	C	G	ref-alt	*TNRC6B*	NM_015088	exon5	c.C648G	p.T216T	synonymous
3-0209-000	11	760129	T	C	ref-alt	*TALDO1*	NM_006755	exon4	c.T337C	p.F113L	nonsynonymous
3-0209-000	12	9317793	C	G	ref-alt	*PZP*	NM_002864	exon19	c.G2429C	p.R810P	nonsynonymous
3-0209-000	22	38131444	C	G	ref-alt	*TRIOBP*	NM_001039141	exon9	c.C5101G	p.L1701V	nonsynonymous
3-0209-000	22	38131445	T	G	ref-alt	*TRIOBP*	NM_001039141	exon9	c.T5102G	p.L1701R	nonsynonymous
3-0215-000	1	220835359	G	A	ref-alt	*MARK1*	NM_001286124	exon18	c.G2242A	p.A748T	nonsynonymous
3-0215-000	10	116233698	C	T	ref-alt	*ABLIM1*	NM_006720	exon5	c.G195A	p.S65S	synonymous

Case	Chr	Position[a]	Reference allele	Alternative allele	Zygosity	Gene symbol	transcript ID	exon number	coding change	protein change	Effect
3-0223-000	19	11213417	G	A	ref-alt	*LDLR*	NM_001195803	exon3	c.G268A	p.D90N	nonsynonymous
3-0227-000	14	75151253	G	-[b]	ref-alt	*AREL1*	NM_001039479	exon4	c.147delC	p.Y49X	stopgain
3-0227-000	15	79750388	A	T[b]	ref-alt	*KIAA1024*	NM_015206	exon2	c.A1899T	p.K633N	nonsynonymous
3-0227-000	16	47614206	C	G[b]	ref-alt	*PHKB*	NM_001031835	exon9	c.C690G	p.S230R	nonsynonymous
3-0227-000	6	26199346	C	T[b]	ref-alt	*HIST1H2AD*	NM_021065	exon1	c.G126A	p.E42E	synonymous
3-0227-000	6	43018958	G	A[b]	ref-alt	*CUL7*	NM_014780	exon4	c.C981T	p.P327P	synonymous
3-0253-000	10	114170349	G	A	ref-alt	*ACSL5*	NM_016234	exon9	c.G920A	p.S307N	nonsynonymous
3-0254-000	17	42932002	C	T	ref-alt	*EFTUD2*	NM_004247	exon22	c.G2181A	p.L727L	synonymous
3-0254-000	2	242611640	C	T	ref-alt	*ATG4B*	NM_013325	exon13	c.C1143T	p.F381F	synonymous
3-0261-000	2	166237202	G	C	ref-alt	*SCN2A*	NM_001040143	exon23	c.G4409C	p.G1470A	nonsynonymous
3-0262-000	22	42463262	G	A	ref-alt	*NAGA*	NM_000262	exon4	c.C357T	p.Y119Y	synonymous
3-0262-000	5	79465157	C	T	ref-alt	*SERINC5*	*NM_001174071*	exon7	c.763+1G>A		splice site
3-0270-000	6	27100214	C	A	ref-alt	*HIST1H2BJ*	NM_021058	exon1	c.G316T	p.E106X	stopgain
3-0270-000	6	158094104	T	C	ref-alt	*ZDHHC14*	NM_153746	exon9	c.T1372C	p.S458P	nonsynonymous
3-0270-000	7	7514248	G	C	ref-alt	*COL28A1*	NM_001037763	exon15	c.C1286G	p.S429X	Stopgain

Case	Chr	Position[a]	Reference allele	Alternative allele	Zygosity	Gene symbol	transcript ID	exon number	coding change	protein change	Effect
3-0294-000	5	64892877	C	T	ref-alt	*TRIM23*	*NM_001656*	exon9	c.1309+1G>A		splice site
3-0309-000	3	122864390	T	C	ref-alt	*PDIA5*	NM_006810	exon12	c.T932C	p.M311T	nonsynonymous
3-0309-000	3	155615774	G	T	ref-alt	*GMPS*	NM_003875	exon3	c.G268T	p.D90Y	nonsynonymous
3-0312-000	2	106471675	G	A	ref-alt	*NCK2*	NM_001004720	exon2	c.G156A	p.P52P	synonymous
3-0314-000	11	107501449	A	C[b]	ref-alt	*ELMOD1*	NM_001130037	exon4	c.A187C	p.S63R	nonsynonymous
3-0314-000	14	22133323	G	T[b]	ref-alt	*OR4E2*	NM_001001912	exon1	c.G27T	p.V9V	synonymous
3-0314-000	6	170632250	C	G[b]	ref-alt	*FAM120B*	NM_001286379	exon3	c.C1854G	p.T618T	synonymous
3-0318-000	9	35101724	G	A	ref-alt	*STOML2*	NM_001287033	exon4	c.C289T	p.L97L	synonymous
3-0325-000	2	168103257	T	C	ref-alt	*XIRP2*	NM_001199144	exon7	c.T4689C	p.V1563V	synonymous
3-0329-000	1	154143940	C	T	ref-alt	*TPM3*	NM_001043351	exon5	c.G480A	p.E160E	synonymous
3-0330-000	22	23230296	G	C[b]	ref-alt	*IGLL5*	NM_001178126	exon1	c.G63C	p.R21S	nonsynonymous
3-0330-000	4	87705691	C	A[b]	ref-alt	*PTPN13*	NM_080684	exon35	c.C5525A	p.P1842Q	nonsynonymous
3-0330-000	6	117248533	C	A[b]	ref-alt	*RFX6*	NM_173560	exon17	c.C2229A	p.S743R	nonsynonymous
3-0338-000	19	7553881	C	A	ref-alt	*PEX11G*	NM_080662	exon1	c.A16T	p.S6C	nonsynonymous

Case	Chr	Position[a]	Reference allele	Alternative allele	Zygosity	Gene symbol	transcript ID	exon number	coding change	protein change	Effect
3-0338-000	20	5933079	C	T	ref-alt	*MCM8*	NM_032485	exon3	c.C158T	p.P53L	nonsynonymous
3-0368-000	22	20050958	G	A	ref-alt	*TANGO2*	NM_001283148	exon8	c.G701A	p.R234Q	nonsynonymous
3-0371-000	2	119739065	G	A	ref-alt	*MARCO*	NM_006770	exon9	c.G847A	p.G283R	nonsynonymous
3-0385-000	17	67136812	T	G	ref-alt	*ABCA6*	NM_080284	exon2	c.A33C	p.Q11H	nonsynonymous
3-0391-000	4	88412805	G	A	ref-alt	*SPARCL1*	NM_001291976	exon6	c.C881T	p.T294M	nonsynonymous
3-0393-000	3	63996422	C	T	ref-alt	*PSMD6*	NM_014814	exon8	c.G1092A	p.W364X	stopgain
3-0393-000	7	82584377	C	T	ref-alt	*PCLO*	NM_033026	exon5	c.G5892A	p.T1964T	Synonymous
3-0402-000	11	9111335	C	T	ref-alt	*SCUBE2*	NM_020974	exon2	c.G175A	p.D59N	nonsynonymous
3-0402-000	19	54973833	G	A	ref-alt	*LENG9*	NM_198988	exon1	c.C943T	p.P315S	nonsynonymous
3-0409-000	15	63824890	A	G	ref-alt	*USP3*	NM_006537	exon2	c.A136G	p.S46G	nonsynonymous
3-0409-000	2	15417138	G	A	ref-alt	*NBAS*	NM_015909	exon43	c.C5226T	p.H1742H	synonymous
3-0409-000	2	15417139	T	A	ref-alt	*NBAS*	NM_015909	exon43	c.A5225T	p.H1742L	nonsynonymous
3-0409-000	2	191524448	G	A	ref-alt	*NAB1*	NM_005966	exon4	c.G546A	p.A182A	synonymous
3-0410-000	17	39551867	A	G	ref-alt	*KRT31*	NM_002277	exon4	c.T597C	p.N199N	synonymous
3-0410-000	18	13019104	A	G	ref-alt	*CEP192*	NM_032142	exon9	c.A949G	p.S317G	nonsynonymous

Case	Chr	Position[a]	Reference allele	Alternative allele	Zygosity	Gene symbol	transcript ID	exon number	coding change	protein change	Effect
3-0410-000	8	42287662	T	C	ref-alt	*SLC20A2*	NM_006749	exon9	c.A1629G	p.G543G	synonymous
3-0465-000	15	60806943	CTC	-	ref-alt	*RORA*	NM_134261	exon4	c.294_296del	p.98_99del	nonframeshift deletion
3-0469-000	3	150916649	A	G	ref-alt	*GPR171*	NM_013308	exon3	c.T525C	p.N175N	synonymous
3-0503-000	5	115323514	G	C	ref-alt	*AQPEP*	NM_173800	exon4	c.G983C	p.R328P	nonsynonymous
3-0503-000	7	154876164	C	A	ref-alt	*HTR5A*	NM_024012	exon2	c.C1041A	p.S347R	nonsynonymous
3-0504-000	12	12288208	C	T	ref-alt	*LRP6*	NM_002336	exon17	c.G3634A	p.G1212S	nonsynonymous
3-0504-000	8	95774047	G	A	ref-alt	*DPY19L4*	NM_181787	exon8	c.G855A	p.V285V	synonymous

[a] Genomic coordinates from Hg19 build. Chr, chromosome. [b] Variant confirmed only in lympoblastoid cell line derived DNA.

eTable 6. Incidental and medically actionable findings from whole-exome sequencing identified in eight probands.

Case	Sex	Gene symbol	MIM	Chromosome	Position[a]	Reference allele	Alternative allele	Transcript ID	Mutation	Inheritance	1000 genomes[b]	NHLBI exome server[c]	Evidence for pathogenicity
3-0080-000	Male	*FBN1*	134797	15	48704816	G	A	NM_000138	c.C8176T p.R2726W	paternal	0.0005	0.001078	Seen in cases
3-0094-000	Male	*COCH*	603196	14	31346846	C	T	NM_004086	c.C151T p.P51S	maternal	0	0	Seen in cases
3-0096-000, 3-0469-000	Male, Female	*CACNA1S*	114208	1	201020165	T	A	NM_000069	c.A4060T p.T1354S	paternal, paternal	0.0009	0.003691	Seen in cases
3-0097-000	Female	*SDHA*	600857	5	236714	GG	-	NM_001294332	c.1288_1288del p.430_430del	maternal	0	0	Novel mutation. Loss-of-function mutations have been implicated in paragangliomas and gastrointestinal stromal tumors
3-0209-000	Male	*SDHB*	185470	1	17349179	C	T	NM_003000	c.G689A p.R230H	paternal	0	0	Seen in cases
3-319-000	Female	*CPT1A*	600528	11	68548130	G	A	NM_001031847	hom c.C1436T p.P479L	inherited	0	0	seen in cases
3-0229-000	Male	*MYPN*	608517	10	69933995	C	T	NM_001256267	c.C2146T p.Q716X	maternal	0	0	Novel mutation. Loss-of-function have been reported in cases with familial restrictive cardiomyopathy.

[a] Genomic coordinates from Hg19 build, [b] Allele frequency found in 1000 genomes project, [c] Allele frequency in the NHLBI exome server.

eTable 7. Phenotype information for 30 probands who received molecular diagnosis either from chromosomal microarray and/or whole-exome sequencing

Case	Sex	Clinical classification[a]	Description of birth defect/s and MRI findings	Description of dysmorphologies	Additional phenotype information	Molecular diagnosis from WES[b]	Molecular diagnosis from CMA[c]
3-0002-000	M	Essential	-	None	At diagnosis almost no speech. At age 4yrs, 100 single words.	N/A	YES
3-0021-000	F	Essential	-	Smooth philtrum, overbite/dental crowding/high palate and low set ears	Persistently abnormal EEG with potential epileptogenic discharges but no clinical seizures	N/A	YES
3-0027-000	M	Equivocal	Bicuspid aortic valve	Brachcephaly with tall forehead, two frontal cowlicks, long downslanting palpebral fissures, minor pectus excavatum deformity, and significant large and small joint laxity	Normal intelligence (IQ=95). Language regression at 13 months. Clinical examination was consistent with Ehler-Danlos syndrome type III	YES	NO
3-0032-000	F	Complex	Atrioventricular canal defect	bilateral 5th finger clindactaly, flat facies, Brachycephaly, upslanting palpebral fissures and short stature	Clinical examination indicated Down syndrome	N/A	YES
3-0039-000	F	Complex	Submucous cleft palate, single umbilical artery and enlarged cisterna magna	Ear contour abnormality, macrocephaly, low frontal hairline, small hands with distally tapered fingers	-	NO	YES
3-0049-001	M	Complex	Bilateral hearing loss with hearing aids from age 4 years, bilateral cryptorchidism,congenital left optic nerve pit, complete agenesis corpus callosum and loss of white matter occipital lobes	Coarse facial features, macrocephaly, short stature, low ears, wide spaced teeth, thin upper lip and smooth philtrum	Not testable for IQ measures	N/A	YES
3-0066-000	M	Equivocal	-	Macrocephaly , malar hypoplasia/long midface, overfolded superior helices, deep-set nails	Medicated for aggression and obsessive thoughts. Hyperactive with a short attention span. Normal intelligence (IQ=108)	N/A	YES
3-0067-000	M	Complex	Large cisterna magna	Long palpebral fissures, smooth philtrum with thin upper lip, wide mouth and extremely pointed chin	-	N/A	YES

Case	Sex	Clinical classification[a]	Description of birth defect/s and MRI findings	Description of dysmorphologies	Additional phenotype information	Molecular diagnosis from WES[b]	Molecular diagnosis from CMA[c]
3-0075-000	M	Complex	Congenitally absent upper lateral incisors and left external ear malformation	Wide spaced teeth and clinodactyly l finger	High/normal intelligence (IQ=131). At 4yr5mth: average/low average speech and language skills	YES	YES
3-0095-000	M	Complex	Glandular hypospadias, retrocerebellar arachnoid cyst, papilledema (diagnosed with idiopathic intracranial hypertension)	Midface hypoplasia, supernumerary tooth (canine), macrocephaly, large ears, long palpebral fissures, hypertelorism, small hands and brachydactyly	Mild intellectual disability (IQ=67). Obesity, significant central hypotonia in infancy	YES	YES
3-0097-000	F	Essential	-	Prominent ears, telecanthus with long palpebral fissures and wide mouth	Normal intelligence (IQ=114)	NO	YES
3-0111-000	M	Complex	Post axial polydacty (4 limbs)	Large outstanding ears with ear lobe creases, bitemporal narrowing, inverted nipples	Global developmental delay (no formal IQ testing done), Asthma, Hypothyroidism, Impaired fasting blood sugar	YES	NO
3-0133-000	F	Essential	-	At 13 years, height 5th centile, weight < 5th centile, decreased upper segment/lower segment (0.875) and smooth philtrum	-	N/A	YES
3-0134-000	M	Essential	-	Ocular hypertelorism and wide mouth	Normal/low intelligence (IQ=82)	NO	YES
3-0140-000	M	Essential	-	5th finger clindodactly, brachycephaly and petite facies	Normal/low intelligence (IQ=80)	YES	NO
3-0152-000	M	Essential	-	Downslanting palpebral fissures and hypertelorism.	-	N/A	YES
3-0211-000	M	Equivocal	Multicystic L temporal lobe mass (0.8x1.2x0.9cm)	Dolicocephaly with parietal bossing, frontal cowlick, deep set eyes, single palmar crease, 5th finger clinodactyly	-	YES	NO
3-0251-000	M	Essential	-	Smooth philtrum, unilateral single palmar crease and large hypopigmented macule	-	N/A	YES
3-0261-000	M	Essential	-	Ocular hypertelorism with long palpebral fissures, and flat nailbeds with deepset nails.	Mild intellectual disability (IQ=65), abnormal EEGs at age 7-8 years showing multifocal epileptiform discharges	YES	NO

Case	Sex	Clinical classification[a]	Description of birth defect/s and MRI findings	Description of dysmorphologies	Additional phenotype information	Molecular diagnosis from WES[b]	Molecular diagnosis from CMA[c]
3-0269-000	M	Complex	Umbilical hernia	Hypertelorism, hypoplastic midface with anteverted nares, smooth philtrum and wide mouth, posteriorly rotated ears and clinodactyly of 5th fingers	Not testable for IQ measures.	N/A	YES
3-0313-000	M	Complex	Drusen left optic nerve	Small mouth, abnormal eyebrows, asymmetric palpebral fissures, malar hypoplasia with square nose, dystrophic nails, dimples over scapulae.	Normal intelligence (IQ=120). Immobile soft palate/velopharyngeal insufficiency, speech dyspraxia	N/A	YES
3-0338-000	M	Complex	Oculomotor apraxia, atrial septal defect and small fibrous subaortic ridge	Frontal/parietal bossing with brachycephaly, ocular hypertelorism, midface hypoplasia, large low-set ears and bilateral 5th finger clinodactly	-	NO	YES
3-0360-000	M	Equivocal	-	Double hair whorl, prominent philtrum, ear asymmetry, bilateral 5th finger clinodadactly, short stature and small hands	Normal intelligence (IQ=86)	N/A	YES
3-0368-000	M	Complex	Small patent ductus arteriosus, nodular heterotopia frontal lobe	Macrocephaly, frontal cowlick, long palpebral fissures, midface hypoplasia and wide mouth	Normal intelligence (IQ=98)	NO	YES
3-0391-000	M	Complex	Complex pineal gland cyst (7mm)	Double whorl, brachycephaly, upslanting palpebral fissures, short smooth philtrum and thin upper lip, deep set fingers and toe nails	Moderate Intellectual disability (IQ=59)	NO	YES
3-0392-000	F	Complex	Bilateral vesicoureteral reflux	Macrocephaly, double crown and low hairlines, short prominent forehead and glabella, long palpebral fissures, hypertelorism, short smooth philtrum, large fleshy ears, attached lobules cross bite, cubitus valgus and large feet	Moderate Intellectual disability (IQ=51)	N/A	YES
3-0437-000	M	Equivocal	Agenesis corpus callosum	Macrocephaly, low hairlines with thick coarse hair, ear asymmetry with hypoplasia left superior crus, and wide spaced incisors with misplaced canines	Normal Intelligence (IQ=108)	N/A	YES

Case	Sex	Clinical classification[a]	Description of birth defect/s and MRI findings	Description of dysmorphologies	Additional phenotype information	Molecular diagnosis from WES[b]	Molecular diagnosis from CMA[c]
3-0459-000	M	Complex	Focal cortical dysplasia involving the left frontal lobe.	Slightly brachycephalic with a double hair whorl, deeply grooved philtrum, accessory nipple (below left nipple), clinodactyly 5th fingers and small feet	Global developmental delay, IQ not testable.	YES	NO
3-0477-000	M	Complex	-	Macrocephaly, prominent forehead, brachycephalic, short philtrum, abnormal ears, thick superior helix and webbed toes	Normal intelligence (IQ=84)	NO	YES
3-0523-000	M	Equivocal	Strabismus, missing kidney and hydrocele/hernia	Short stature	Normal/low intelligence (IQ=77)	N/A	YES

[a] Morphological categorization based on scoring of minor physical anomalies and major congenital abnormalities. Essential (scores 1-3), Equivocal (scores 4-5) and Complex (scores ≥6)

[b] Molecular diagnosis received from Whole exome sequencing (WES). YES, mutation details in eTable4; NO, no findings in this study; N/A, no WES performed in this study. [c] Molecular diagnosis received Chromosomal microarray analysis (CMA). YES, mutation details in eTable3; NO, no findings in this study. IQ, intelligence quotient.

eReferences

1. Hall JG, Allanson JE, Gripp KW, Slavotinek AM. *Handbook of Physical Measurements.* Vol 2nd edition: Oxford University Press; 2007.
2. Miles JH, Hillman RE. Value of a clinical morphology examination in autism. *Am J Med Genet.* Apr 10 2000;91(4):245-253.
3. Gazzellone MJ, Zhou X, Lionel AC, et al. Copy number variation in Han Chinese individuals with autism spectrum disorder. *Journal of neurodevelopmental disorders.* 2014;6(1):34.
4. Pinto D, Delaby E, Merico D, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American journal of human genetics.* May 1 2014;94(5):677-694.
5. Pinto D, Pagnamenta AT, Klei L, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature.* Jul 15 2010;466(7304):368-372.
6. Prasad A, Merico D, Thiruvahindrapuram B, et al. A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3 (Bethesda).* Dec 2012;2(12):1665-1685.
7. Sato D, Lionel AC, Leblond CS, et al. SHANK1 Deletions in Males with Autism Spectrum Disorder. *American journal of human genetics.* May 4 2012;90(5):879-887.
8. Lionel AC, Crosbie J, Barbosa N, et al. Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci Transl Med.* Aug 10 2011;3(95):95ra75.
9. Colella S, Yau C, Taylor JM, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* 2007;35(6):2013-2025.
10. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research.* Nov 2007;17(11):1665-1674.
11. Pinto D, Darvishi K, Shi X, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol.* Jun 2011;29(6):512-520.
12. Darvishi K. Application of Nexus copy number software for CNV detection and analysis. *Curr Protoc Hum Genet.* Apr 2010;Chapter 4:Unit 4 14 11-28.
13. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics.* Mar 15 2007;23(6):657-663.
14. Downey T. Analysis of a multifactor microarray study using Partek genomics solution. *Methods Enzymol.* 2006;411:256-270.
15. Lionel AC, Tammimies K, Vaags AK, et al. Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. *Human molecular genetics.* May 15 2014;23(10):2752-2768.
16. Baldwin EL, Lee JY, Blake DM, et al. Enhanced detection of clinically relevant genomic imbalances using a targeted plus whole genome oligonucleotide microarray. *Genetics in medicine : official journal of the American College of Medical Genetics.* Jun 2008;10(6):415-429.
17. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genetics in medicine : official journal of the American College of Medical Genetics.* Jul 2011;13(7):680-685.
18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* Sep 2010;38(16):e164.
19. Exome VS. esp6500si version http://evs.gs.washington.edu/EVS/. Accessed January 22, 2013.
20. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science.* Jan 1 2010;327(5961):78-81.
21. Schafer CM, Campbell NG, Cai G, et al. Whole exome sequencing reveals minimal differences between cell line and whole blood derived DNA. *Genomics.* Oct 2013;102(4):270-277.

22. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome research.* May 2001;11(5):863-874.
23. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods.* Apr 2010;7(4):248-249.
24. Reva B, Antipin Y, Sander C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome biology.* 2007;8(11):R232.
25. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research.* Jan 2010;20(1):110-121.
26. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics.* Mar 2014;46(3):310-315.
27. Kohler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic acids research.* Jan 2014;42(Database issue):D966-974.
28. De Rubeis S, He X, Goldberg AP, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature.* Nov 13 2014;515(7526):209-215.
29. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature.* Nov 13 2014;515(7526):216-221.
30. Solomon BD, Nguyen AD, Bear KA, Wolfsberg TG. Clinical genomic database. *Proceedings of the National Academy of Sciences of the United States of America.* Jun 11 2013;110(24):9851-9855.
31. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genetics in medicine : official journal of the American College of Medical Genetics.* Apr 2008;10(4):294-300.
32. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics.* Mar 5 2015.