LETTER

doi:10.1038/nature13394

Genome sequencing identifies major causes of severe intellectual disability

Christian Gilissen[1]*, Jayne Y. Hehir-Kwa[1]*, Djie Tjwan Thung[1], Maartje van de Vorst[1], Bregje W. M. van Bon[1], Marjolein H. Willemsen[1], Michael Kwint[1], Irene M. Janssen[1], Alexander Hoischen[1], Annette Schenck[1], Richard Leach[2], Robert Klein[2], Rick Tearle[2], Tan Bo[1,3], Rolph Pfundt[1], Helger G. Yntema[1], Bert B. A. de Vries[1], Tjitske Kleefstra[1], Han G. Brunner[1,4]*, Lisenka E. L. M. Vissers[1]* & Joris A. Veltman[1,4]*

Supplementary Table 1: Detailed cohort description of 50 trios studied by WGS

Level of ID	Number of patients
IQ <30	31
IQ 30-50	19
IQ 50-70	0

Gender	
Male	26
Female	24

Age Groups	
<10 yrs	26
10-20 yrs	8
>20 yrs	16

Sibship size	
1	4
2	27
3	17
≥4	2

Multiple Congenital Anomalies	
0	27
1	17
2	6

Short Stature	
Yes	15
No	35

Microcephaly	
Yes	18
No	32

Macrocephaly	
Yes	3
No	47

Epilepsy	
Yes	23
No	27

Abnormalities on brain imaging	
Yes	17
No	19
Not assessed	14

Cardiac malformations	
Yes	1
No	49

Abnormalities of the urogenital system	
Yes	11
No	39

Extended Data Table 2 | *De novo* SNVs of potential clinical relevance identified using WGS

Trio	Gene	Protein effect	Mutation type	PhyloP‡	Gene Classification§
1	*NGFR*	p.(Cys122Arg)	Missense	4.97	-
2	*GFPT2*	p.(Thr680Ser)	Missense	6.02	-
6	*WWP2*	p.(Gly10Gly) †	Synonymous	-0.12	-
7	*TBR1*	p.(Gln373Arg)	Missense	3.51	Known
9	*WDR45*	p.(Cys344Alafs*67)	Frameshift		Known
13	*SMC1A*	p.(Asn788Lysfs*10)	Frameshift		Known
15	*SPTAN1*	p.(Glu91Lys)	Missense	5.69	Known
17	*ASUN*	p.(Gln99*)	Nonsense		-
21	*ALG13*	p.(Asn107Ser)	Missense	1.34	Known‖
21	*RAI1*	p.(Gln88*)	Nonsense		Known
22	*MED13L*	p.(Asp860Gly)	Missense	4.75	Candidate
24	*BRD3*	p.(Phe334Ser)	Missense	4.48	-
25	*SATB2*	p.(Gln310delinsHisCysLysAlaThr)	Insertion		Known
26	*PPP2R5D*	p.(Trp207Arg)	Missense	5.13	Candidate
27	*KCNA1*	p.(Thr371Ile)	Missense	5.69	Known
28	*SCN2A*	p.(Gln1521*)	Nonsense		Known
30	*MAST1*	p.(Pro1177Arg)	Missense	5.28	-
34	*APPL2*	p.(Ser329*)	Nonsense		-
41	*NACC1*	p.(Arg468Cys)	Missense	3.51	-
43	*POGZ*	p.(Arg1001*)	Nonsense		Candidate
46	*TBR1*	p.Thr532Argfs*144	Frameshift		Known
49	*KANSL2*	p.(Gly151Gly) †	Synonymous	1.58	Candidate

A dash indicates genes that have not yet been implicated in ID, but fulfil the criteria for diagnostic reporting of a pathogenic variant (that is, a possible cause for ID).
† Predicted effect on splicing.
‡ PhyloP score for nonsense and frameshift mutations is not provided as this are deleterious mutations regardless of their evolutionary conservation.
§ 'Known' refers to known ID gene whereas 'Candidate' refers to a gene that is listed on the candidate ID gene list.
‖ Since the inclusion of this patient in this study, the same *de novo* mutation in *ALG13* has been described elsewhere[16]. This may suggest that this mutation, despite its low conservation and the identification of a nonsense mutation in *RAI1*, may also contribute to the disease phenotype in this patient. See also Supplementary Table 8 legend.

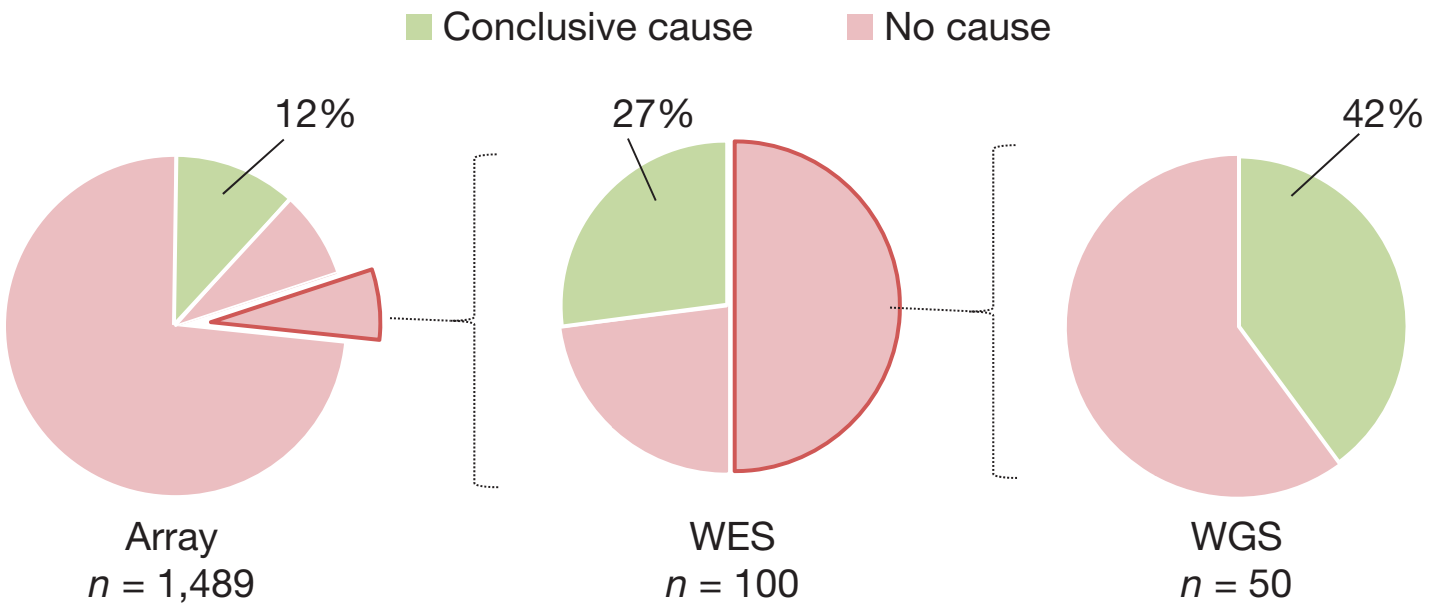


Figure 1 | Study design and diagnostic yield in patients with severe ID per technology. Diagnostic yield for patients with severe ID (IQ < 50), specified by technology: genomic microarrays, WES and WGS. Percentages indicate the number of patients in whom a conclusive cause was identified using the specified technique. Brackets indicate the group of patients in whom no genetic cause was identified and whose DNA was subsequently analysed using the next technology. WES data are updated with permission from ref. 6 (see Supplementary Methods).

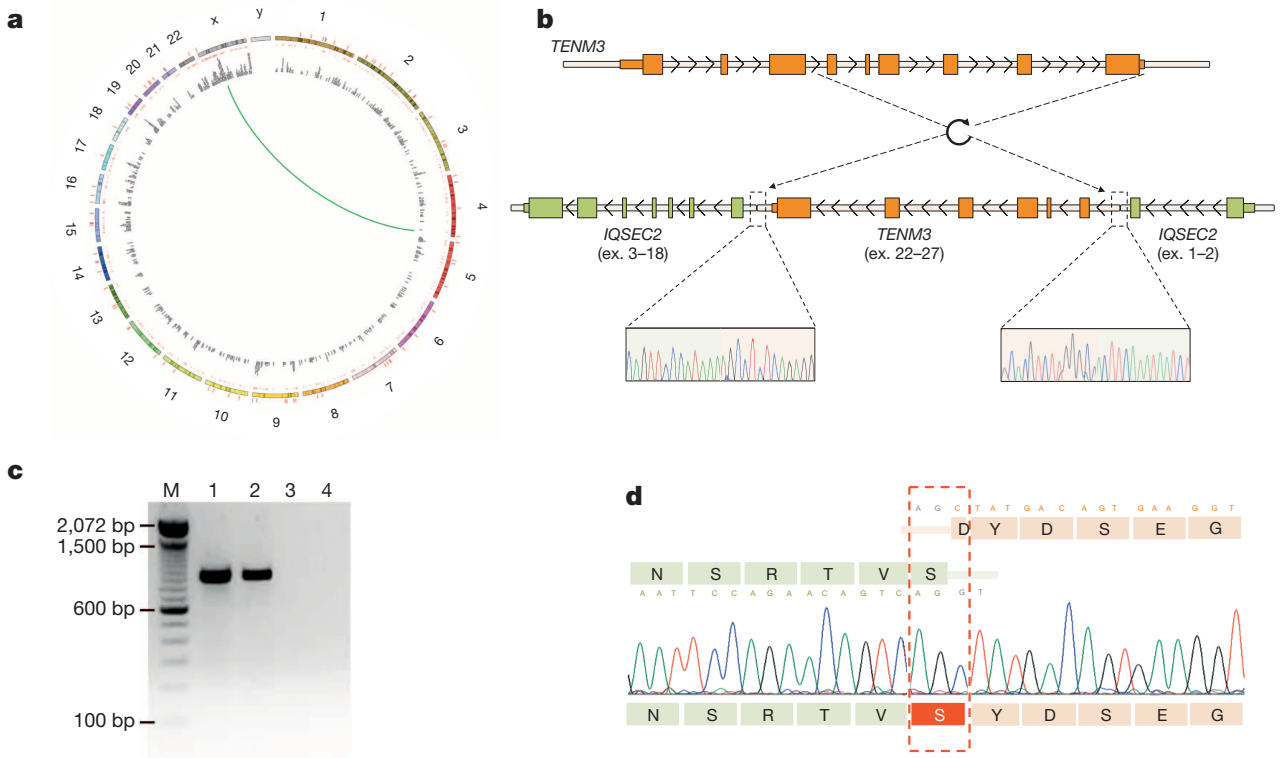


Figure 2 | Detected duplication of a chromosome 4 region into the X-chromosomal *IQSEC2* gene. **a**–**d**, Graphical representation of a *de novo* duplication–insertion event in patient 31. **a**, Circos plot with chromosome numbers and *de novo* mutations in the outer shell. Red bars represent genome-wide potential *de novo* SNVs, whereas blue lines represent potential *de novo* CNVs/structural variants. Inner shell represents the location of known ID genes (red marks) with the respective gene names. Green line illustrates a duplication event on chromosome 4, which is inserted into chromosome X. **b**, Details for inserted duplication event on chromosome X. The last six exons of *TENM3* are inserted in inverted orientation into intron 2 of *IQSEC2*, predicted to result in an in-frame *IQSEC2-TENM3* fusion gene. ex., exon. **c**, **d**, PCR (**c**) on and Sanger sequencing (**d**) of complementary DNA junction fragment in patient 31. Lanes in **c** represent the following: M, 100 bp marker; 1, cDNA of patient with cyclohexamide treatment; 2, cDNA of patient without cyclohexamide treatment; 3, control cDNA with cyclohexamide treatment; 4, control cDNA without cyclohexamide treatment. Our data verify the presence of a fusion gene in patient 31 that is suggested to escape nonsense-mediated decay.

Table 1 | Diagnostic yield by WGS for a pre-screened cohort of 50 ID trios

Genetic cause	Number of patients
Total positive diagnosis	21
Dominant *de novo*	20
Autosomal SNV	11
Autosomal CNV	5
X-linked SNV	2
X-linked CNV	2
Recessive	1
Homozygous	0
Compound heterozygous	1
X-linked	0
Candidate ID genes	8
No diagnosis	21

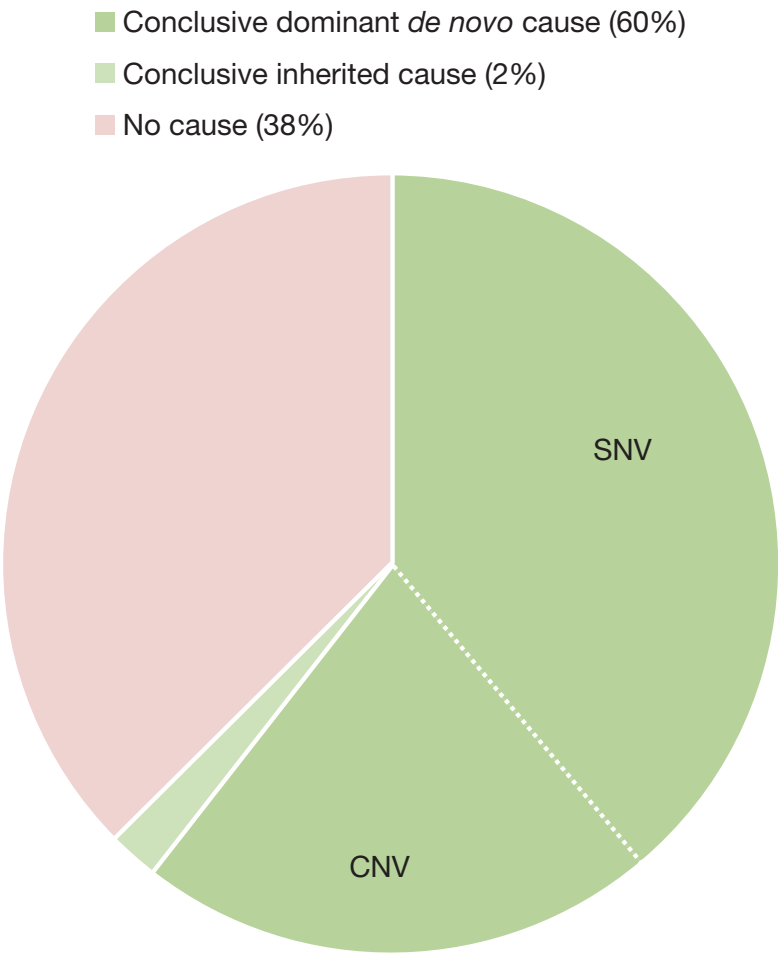


Figure 3 | Pie chart showing role of *de novo* mutations in severe ID. Contribution of genetic causes to severe ID on the basis of the cumulative estimates provided per technology. Our data indicate that *de novo* mutations are a major cause of severe ID. Note, small variants include SNVs and insertion/deletion events whereas large variants include structural variants and CNVs (>500 bp).

METHODS SUMMARY

Patients were selected to have severe ID (IQ < 50) and negative results on diagnostic genomic microarrays and exome sequencing[6] (Fig. 1). WGS was performed by Complete Genomics as previously described[8,29]. *De novo* SNVs were identified using Complete Genomics' cgatools 'calldiff' program. CNVs and structural variants were reported by Complete Genomics on the basis of read-depth deviations and discordant read pairs, respectively. *De novo* CNVs and structural variants were then identified by excluding variants with minimal evidence or overlapping with CNVs and structural variants identified in the parents or control data sets. All variants were annotated using an in-house analysis pipeline and subsequently prioritized for validation based on their confidence level (low/medium/high) and location in the genome (coding/non-coding). High-confidence candidate *de novo* mutations in non-coding variants in known ID genes were prioritized on the basis of evolutionary conservation and overlap with ENCODE chromatin state segments and transcription-factor-binding sites[23]. Statistical overrepresentation of mutations in known and candidate ID gene lists was calculated using Fisher's exact test based on RefSeq genes. Enrichments for loss-of-function CNV events were calculated using the exact Poisson test. To clinically interpret (*de novo*) mutations, each variant (both CNV and SNV) was assessed for mutation impact as well as functional relevance to ID according to diagnostic protocols for variant interpretation[6,24–27]. The diagnostic yield of WGS in an unbiased cohort was calculated based on cumulative estimates of diagnostic yield per technology (genomic microarray, WES and WGS).

METHODS

Patient selection. Patients were selected to have severe ID (IQ < 50) and negative results on diagnostic genomic microarrays and exome sequencing[6] (Fig. 1 and Supplementary Methods).

Whole genome sequencing. WGS was performed by Complete Genomics as previously described[8]. Sequence reads were mapped to the reference genome (GRCh37) and variants were called by local *de novo* assembly according to the methods previously described[29].

Identification of *de novo* small variants. *De novo* SNVs were identified using Complete Genomics' cgatools 'calldiff' program. On the basis of the rank order of the two confidence scores of a *de novo* mutation, we binned the variants in three groups: low confidence (at least one score < 0), medium confidence (both scores ≥ 0 but at least one < 5) and high confidence (both scores ≥ 5) (Supplementary Methods).

Identification of X-linked, recessive and compound heterozygous SNVs. Maternally inherited X-linked variants (in male patients), homozygous variants and compound heterozygous variant pairs were identified using the Complete Genomics' cgatools 'listvariants' and 'testvariants' programs to select variants according to their respective segregation. Compound heterozygous variants affecting the same gene were identified using RefSeq gene annotation (Supplementary Methods).

Identification of *de novo* CNVs and structural variants. CNVs were reported by Complete Genomics on the basis of read-depth deviations across 2 kb windows. Structural variants were reported by Complete Genomics based on discordant read pairs. *De novo* CNVs/structural variants were then identified by excluding variants with minimal evidence or overlapping with CNVs/structural variants identified in the parents or control data sets (Supplementary Methods).

Generation of lists for known and candidate ID genes. To prioritize and for subsequent interpretation of *de novo* variants for each patient individually, two gene lists were generated, one containing known ID genes (defined by five or more patients with ID having a mutation in the respective gene) and one containing candidate ID genes (defined by at least one but less than five patients with ID (or a related phenotype) showing a mutation in the respective gene) (Supplementary Methods).

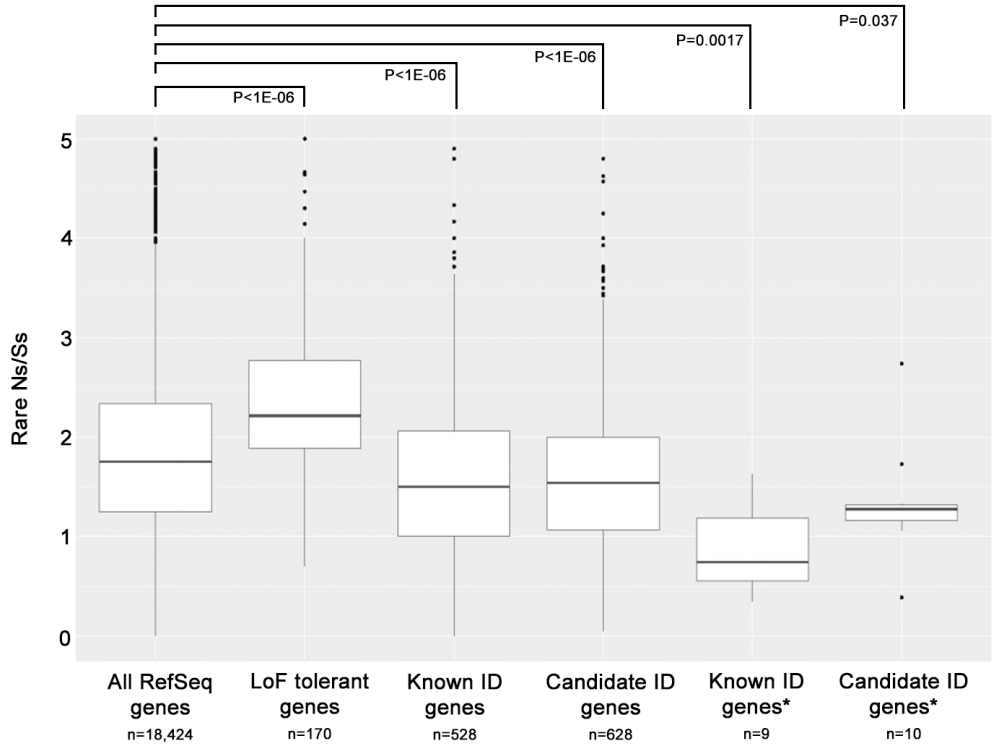
Prioritization of clinically relevant SNVs and CNVs or structural variants. All SNVs were annotated using an in-house analysis pipeline. Variants were prioritized for validation in two distinct ways: (1) medium and high-confidence *de novo* SNVs and *de novo* CNVs/structural variants affecting coding regions and/or canonical splice sites; and (2) all potential *de novo* variants within known ID genes, irrespective of confidence level. Interpretation of coding *de novo* variants was performed as described previously[6]. High-confidence candidate *de novo* mutations in non-coding variants were prioritized on the basis of evolutionary conservation and overlap with ENCODE chromatin state segments and transcription-factor-binding sites (Supplementary Methods)[23].

Clinical interpretation of mutations. To clinically interpret (*de novo*) mutations, each *de novo* mutation (both CNV and SNV) was assessed for mutation impact as well as functional relevance to ID according to diagnostic protocols for variant interpretation[24–27] that are used in our accredited diagnostic laboratory for genetic analysis (accredited to the 'CCKL Code of Practice', which is based on EN/ISO 15189 (2003), registration numbers R114/R115, accreditation numbers 095/103) (Supplementary Methods).

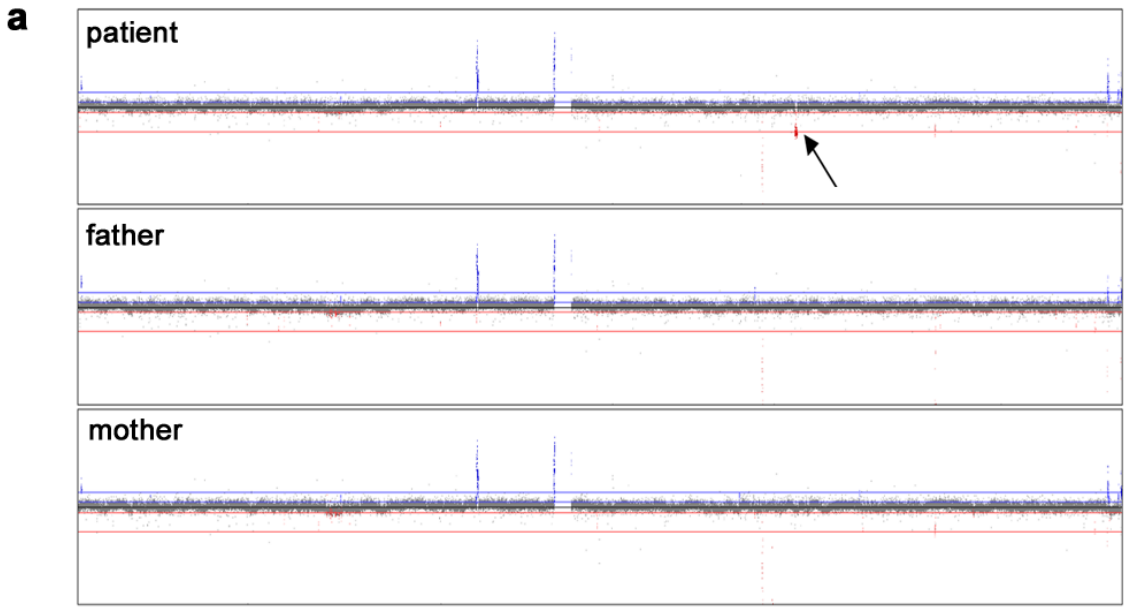
Statistical analysis. Overrepresentation of mutations in gene lists was calculated using Fisher's exact test based on the total coding size of all RefSeq genes and coding size of the genes from the respective gene list. Overrepresentation of loss-of-function mutations was calculated using Fisher's exact test based on published control cohorts. Enrichments for loss-of-function CNV events were calculated using the exact Poisson test. Enrichment for known ID genes was calculated using Fisher's exact test and odds ratios were calculated to compare the frequency of exonic CNVs in ID and control cohorts, respectively (Supplementary Methods).

Calculation of diagnostic yield. Our in-house phenotypic database contains 1,489 patients with severe ID who have all had a diagnostic genomic microarray in the time period 2003–2013. In 173 (11.6%) of these patients, a *de novo* CNV was identified as a cause of ID. Subsequently, 100 array-negative patients were subjected to WES, which resulted in a *de novo* cause for ID in 27% of patients. Of all WES-negative patients, 50 were selected for this WGS study, in which 42% obtained a conclusive genetic cause. Cumulative estimates were subsequently determined using the diagnostic yield per technology (Supplementary Methods).

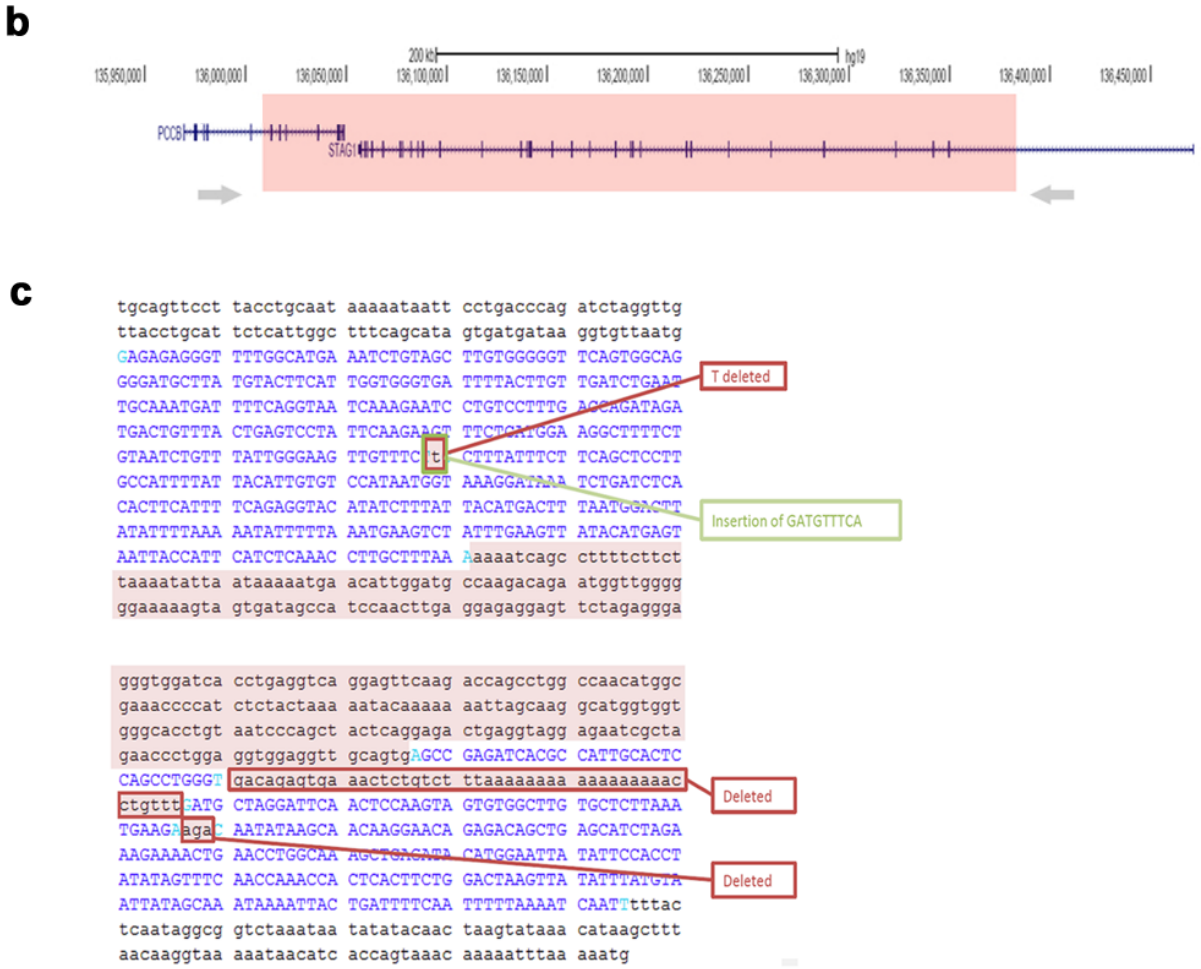
29. Carnevali, P. *et al.* Computational techniques for human genome resequencing using mated gapped reads. *J. Comput. Biol.* **19,** 279–292 (2012).
30. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335,** 823–828 (2012).



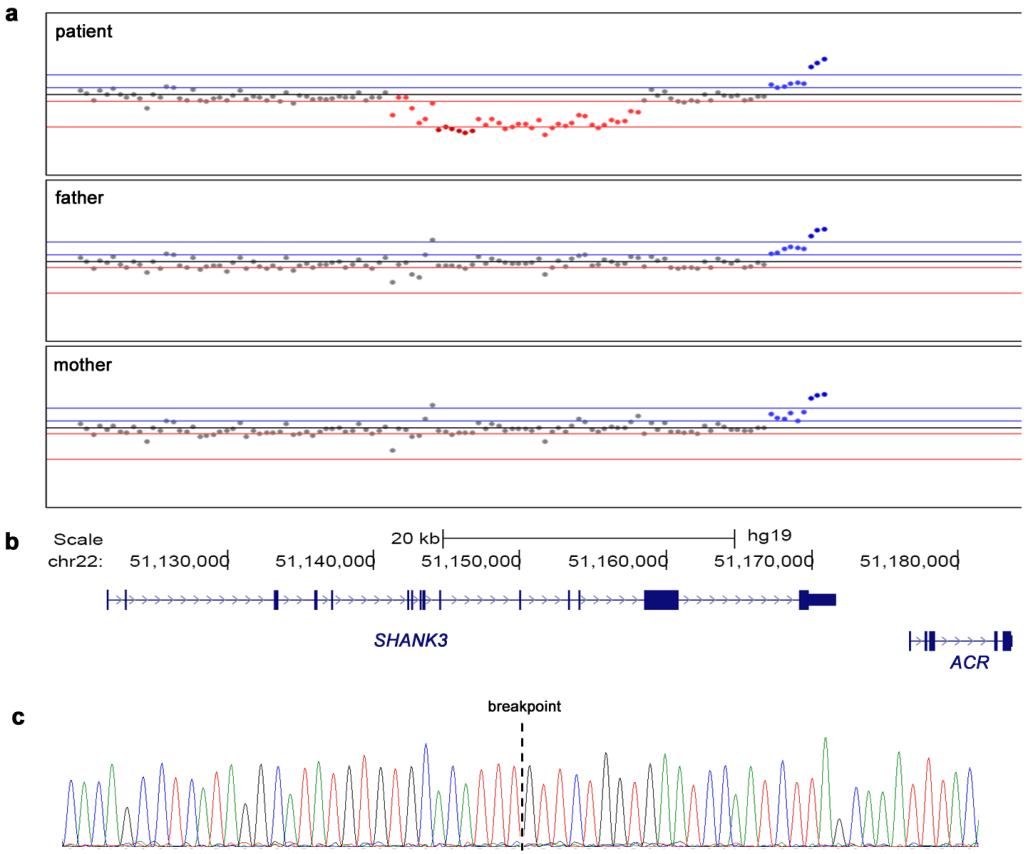
Extended Data Figure 1 | Boxplots of rare missense burden in different gene sets. Boxplots showing the difference in tolerance for rare missense variation in the general population. The vertical axis shows the distribution for each gene set of the number of rare (<1% in NHLBI Exome Sequencing Project) missense variants divided by the number of rare synonymous variants. From left to right the following gene sets are depicted: all 18,424 RefSeq genes, 170 loss-of-function tolerant genes from ref. 30, all 528 known ID genes (Supplementary Table 10), all 628 candidate ID genes (Supplementary Table 11), 9 known ID genes in which *de novo* mutations were identified in this study (Supplementary Table 8), and 10 candidate ID genes in which *de novo* mutations were identified in this study (Supplementary Table 8).



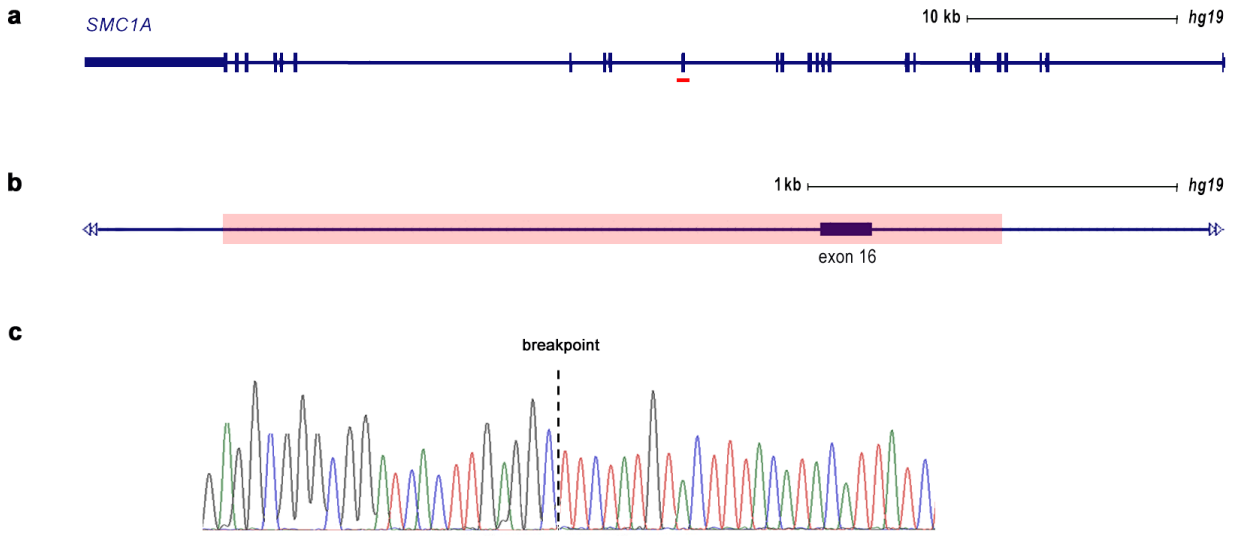
Extended Data Figure 2 | Structural variant involving *STAG1* (patient 40). **a**–**c**, CNV identified using WGS in patient 40, including the *STAG1* gene. **a**, Chromosome 3 profile (log_2 test over reference (T/R) ratios) based on read-depth information for patient, father and mother. Black arrow points towards the *de novo* event in patient 40. **b**, Genic contents of deletion. Grey arrows show primers used to amplify the junction fragment. **c**, Details on the proximal and distal breakpoints, showing the 'fragmented' sequence at both ends. Breakpoints are provided in Extended Data Table 1.



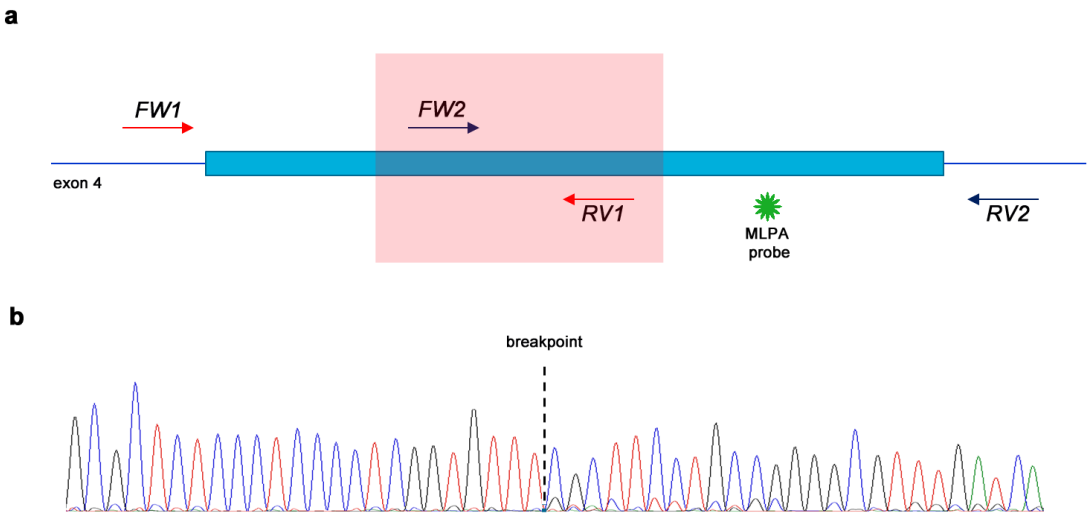
Extended Data Figure 2 | Structural variant involving *STAG1* (patient 40). **a**–**c**, CNV identified using WGS in patient 40, including the *STAG1* gene. **a**, Chromosome 3 profile (log_2 test over reference (T/R) ratios) based on read-depth information for patient, father and mother. Black arrow points towards the *de novo* event in patient 40. **b**, Genic contents of deletion. Grey arrows show primers used to amplify the junction fragment. **c**, Details on the proximal and distal breakpoints, showing the 'fragmented' sequence at both ends. Breakpoints are provided in Extended Data Table 1.



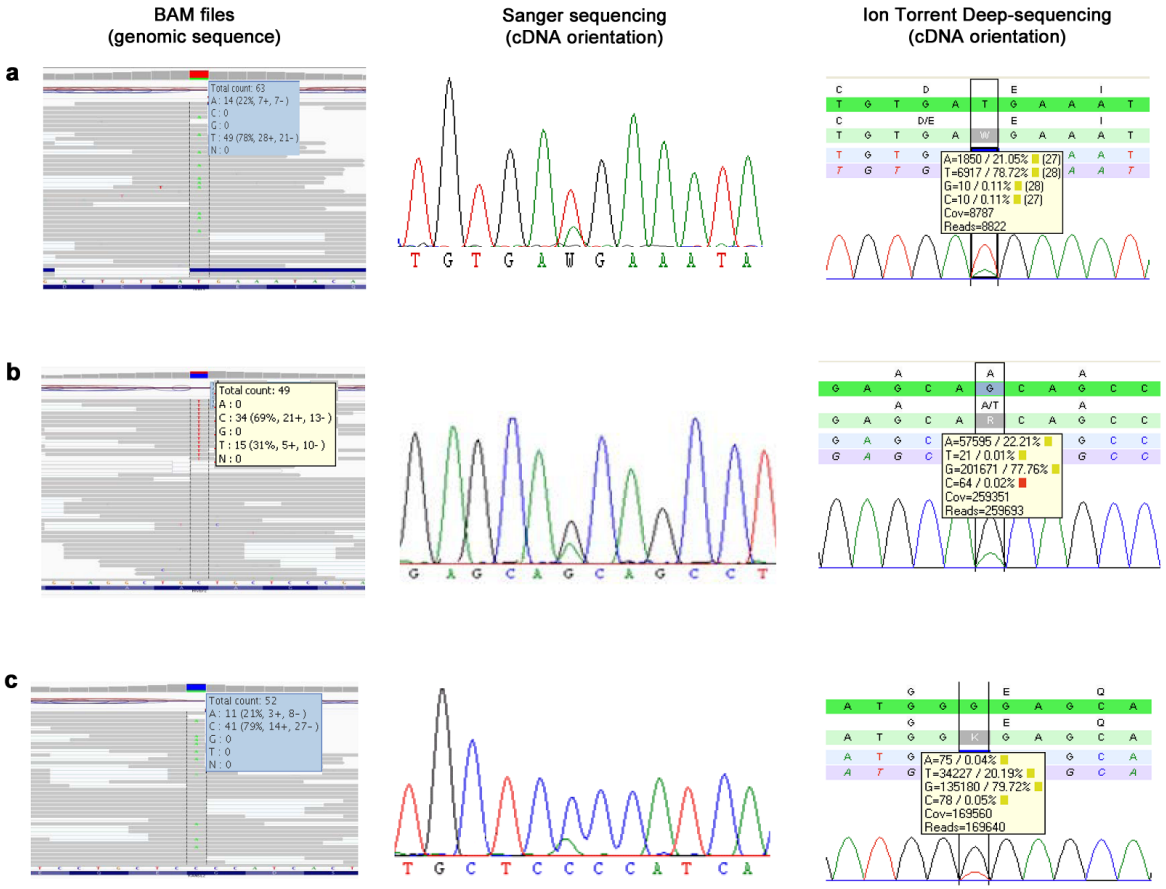
Extended Data Figure 3 | Structural variant involving *SHANK3* (patient 5). **a**–**c**, CNV identified using WGS in patient 5, including the *SHANK3* gene. **a**, Detail of chromosome 22 profile (log_2 T/R ratios) based on read-depth information for patient, father and mother. Red dots in top panel show ratios indicating the *de novo* deletion in patient 5. **b**, Genic content of the deletion. **c**, Sanger validation for the junction fragment. Dotted vertical line indicates the breakpoint with sequence on the left side originating from sequence proximal to *SHANK3* and on the right side sequence that originates from sequence distal to *ACR*. Breakpoints are provided in Extended Data Table 1.



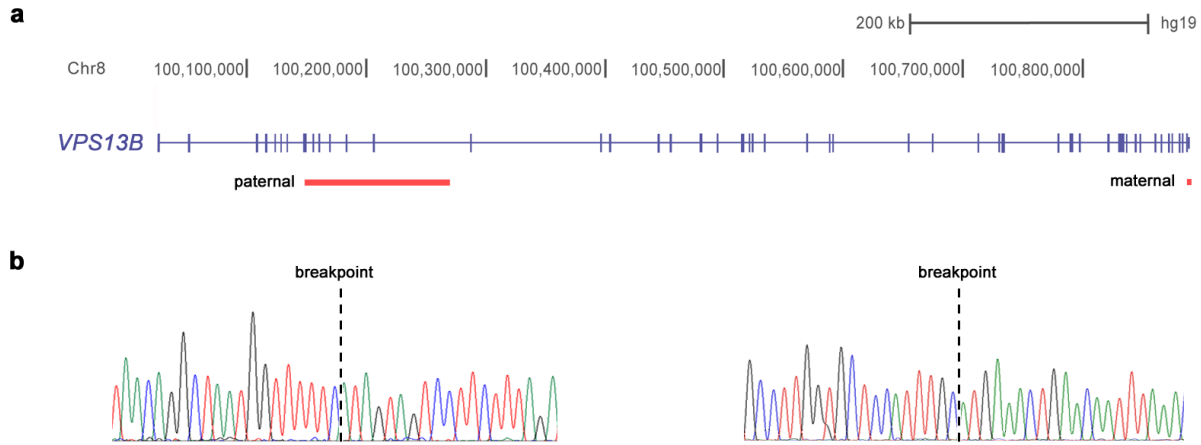
Extended Data Figure 4 | Single-exon deletion involving *SMC1A* (patient 48). **a**, Schematic depiction of the deletion identified in patient 48 involving a single exon of *SMC1A*. Pink horizontal bar highlights the exon that was deleted in the patient. **b**, Details at the genomic level of the deletion including exon 16, with Sanger sequence validation of the breakpoints. Junction is indicated by a black vertical dotted line. Breakpoints are provided in Extended Data Table 1.



Extended Data Figure 5 | Intra-exonic deletion involving *MECP2* (patient 18). **a**, Schematic depiction of the deletion identified in patient 18, which is located within exon 4 of *MECP2*. Initial Sanger sequencing in a diagnostic setting could not validate the deletion as the primers used to amplify exon 4 removed the primer-binding sites (FW2 and RV1 respectively). Multiplex ligation probe amplification (MLPA) analysis for CNV detection showed normal results as the MLPA primer-binding sites were located just outside of the deleted region. **b**, Combining primers FW1 and RV2 amplified the junction fragment, clearly showing the deletion within exon 4. Of note, the background underneath the Sanger sequence is derived from the wild-type allele. Breakpoints are provided in Extended Data Table 1.



Extended Data Figure 6 | Confirmation of mosaic mutations in *PIAS1*, *HIVEP2* and *KANSL2*. **a**–**c**, Approaches used to confirm the presence of mosaic mutations in *PIAS1* (**a**), *HIVEP2* (**b**) and *KANSL2* (**c**). Images and read-depth information showing the base counts in the BAM files (left) indicated that the variants/wild-type allele were not in a 50%/50% distribution. Sanger sequencing (middle) then confirmed the variant to be present in the patient, and absent in the parents (data from parents not shown), again indicating that the mutation allele is underrepresented. Guided by these two observations, amplicon-based deep sequencing using Ion Torrent subsequently confirmed the mosaic state of the mutations (right). On the basis of deep sequencing, percentages of mosaicism for *PIAS1*, *HIVEP2* and *KANSL2* were estimated at 21%, 22% and 20%, respectively.



Extended Data Figure 7 | Compound heterozygous structural variation affecting *VPS13B* (patient 12). **a**, **b**, CNVs of *VPS13B* identified using WGS in patient 12. **a**, Schematic representation of *VPS13B*, with vertical bars indicating coding exons. In patient 12 two deletions were identified, one ~122 kb in size which was inherited from his father, and another ~2 kb in size, which was inherited from his mother and consisted only of a single exon. **b**, Both CNV junction fragments were subsequently validated using Sanger sequencing. Left, junction fragment from the paternally inherited deletion. Right, junction fragment from the maternally inherited deletion. Breakpoints are provided in Extended Data Table 1.

Extended Data Table 1 | Large variants of potential clinical relevance identified using WGS and probability of exonic CNVs occurring in affected and control individuals for these loci

Trio	Type*	Genomic characterization	Size (kb)	CN	Origin	Genes affected	Affected (n=7,743)	Controls (n=4,056)	OR	CI	P-value#
5	CNV	chr22(GRCh37):g.51121756-51187704del	66	1	*De novo*	***SHANK3***; *ACR*	41	4	5.4	1.9-15.0	0.0013
18	SV	chrX(GRCh37):g.153295929-153296514del	0.6	1	*De novo*	***MECP2***§	41	0	22.6	1.4-367.6	0.04
31	CNV complex	chr4(GRCh37):g.183693432-183756173dup	62	3	*De novo*	*TENM3*	5	0	0.6	0.2-2.4	1.0
		Insertion point: chrX(GRCh37):g.53318362-53318363	-	-		***IQSEC2***	28	0	7.6	1.0-56.1	0.092
37	CNV complex	chr3(GRCh37):g.48532000-49156000dup†	624	3	*De novo*	n=22‖	22	0	23.6	1.4-388.7	0.033
		chr3(GRCh37):g.49298000-49848000dup†	550	3		n=20‖					
		chr3(GRCh37):g.49849505-49870969del	21	1		n=2‖					
		chr3(GRCh37):g.49872000-49958000dup†	86	3		n=4‖					
40	CNV	chr3(GRCh37):g.136003159delinsGATGTTTCA	-								
		chr3(GRCh37):g.136003363-136385607del	382	1	*De novo*	***STAG1***; *PCCB*	9¶	0	10.0	0.6-171.1	0.26
		chr3(GRCh37):g.136385640-136385685del	-								
		chr3(GRCh37):g.136385737-136385739del	-								
48	SV	chrX(GRCh37):g.53424894-53427008del	2.1	1	*De novo*	***SMC1A***§	33	0	17.3	1.1-282.9	0.075
49	SV	chr1(GRCh37):g.40247181-40256104dup	9	3	*De novo*	*BMP8B*§	2	0	2.6	0.1-54.6	1.0
50	CNV	chr16(GRCh37):g.29567295-30177916‡	611	1	*De novo*	n=29‖	31	4	4.1	1.4-11.5	0.0093
12	SV	chr8(GRCh37):g.100887349-100889133del	1.7	1	*Maternal*	***VPS13B***§☆	5	0	5.8	0.3-104.2	0.80
	CNV	chr8(GRCh37):g.100147792-100270123del	122	1	*Paternal*	***VPS13B***☆	0	0			

Genes highlighted in bold are either listed as known ID genes or candidate ID genes. Please note that all patients had 250K SNP microarrays. Re-evaluation of these data showed that for all but one CNV the number of probes within the region was insufficient, either because of the small genomic size of the CNV, or due to uneven genome-wide probe spacing leaving fewer probes than required for the hidden Markov model algorithms to be identified.

* Primary method used to identify the rearrangement (see also Supplementary Methods).

† Not assessed at base-pair level due to complexity of CNV event including an inversion, duplication and deletion.

‡ Not assessed at base-pair level as the CNV event, involving a known microdeletion syndrome region, is mediated by low-copy repeats.

§ Single exon.

‖ Number of genes affected rather than individual gene names are provided due to the large number of genes.

¶ Observed 13 times in Decipher.

Corrected for multiple testing using Benjamini–Hochberg with a false discovery rate (FDR) of 0.1.

☆ *VPS13B* recessive ID gene.

Table 1: Re-evaluated molecular diagnosis in 100 patients with severe ID after publication of diagnostic exome sequencing in patients with severe ID.

Patient ID in de Ligt et al	Molecular diagnosis	Based on [gene] (inheritance)	Re-evalutated molecular diagnosis	Based on [gene] (inheritance)	Change due to:
4	Highly likely	*PDHA1 (X-linked)*	NO	-	X-linked variant does not segregate
5	Possibly	*PHIP (de novo)*	Highly likely	*PHIP (de novo)*	In house 2nd case identified through diagnostic WES in patient with severe ID
6	Possibly	*PSMA7 (de novo)*	Highly likely	*EHMT1* deletion *(de novo)*	CNV calling on WES data identified deletion of *EHMT1* (*de novo*)
14	NO		Highly likely	*DDX3X (de novo)*	Bioinformatic re-analysis of WES data identified de novo mutation in DDX3X; A second patient was subsequently identified by diagnostic WES in a patient with ID
18	Highly likely	*ARHGEF9 (X-linked)*	NO		X-linked variant does not segregate
40	Possibly	*WAC/ MIB1 (de novo)*	Highly likely	*WAC (de novo)*	In house 2nd case identified through diagnostic WES in patient with severe ID
60	NO		Highly likely	*TCF4 (de novo)*	Bioinformatic re-analysis identified a *de novo* mutation in *TCF4*
72	Possibly	*PPP2R5D (de novo)*	Highly likely	*PPP2R5D (de novo)*	2nd case identified in house in patient with severe ID
74	Possibly	*KIF5C (de novo)*	Highly likely	*KIF5C (de novo)*	Literature reports additional patient with *de novo* mutation in *KIF5C*
77	NO		Highly likely	*STXBP1 (de novo)*	Bioinformatic re-analysis identified a *de novo mutation STXBP1*
85	NO		Highly likely	*SCN2A (de novo)*	Bioinformatic re-analysis identified a *de novo* mutation in *SCN2A*
87	Possibly	*RAPGEF1 (de novo)*	Highly likely	*MECP2 (de novo)*	Bioinformatic re-analysis identified *MECP2 de novo* mutation
91	Possibly	*EEF1A2 (de novo)*	Highly likely	*EEF1A2 (de novo)*	2nd patient in literature with exact same *de novo* mutation
92	Possibly	*MYT1L (de novo)*	Highly likely	*MYT1L (de novo)*	Overlapping small CNV containing *MYT1L* described in literature
96	NO		Highly likely	*ADNP (de novo)*	Bioinformatic re-analysis of WES data identified a *de novo*

Table 2: Cross-references for patient numbers used in de Ligt et al. and the current study

Trio ID current study	Trio ID de Ligt et al. 2012
1	1
2	4
3	8
4	10
5	12
6	17
7	18
8	19
9	20
10	21
11	23
12	24
13	26
14	27
15	28
16	30
17	33
18	34
19	35
20	36

Trio ID current study	Trio ID de Ligt et al. 2012
21	37
22	38
23	44
24	45
25	46
26	47
27	49
28	50
29	51
30	53
31	55
32	56
33	57
34	65
35	67
36	68
37	71
38	73
39	75
40	76

Trio ID current study	Trio ID de Ligt et al. 2012
41	78
42	80
43	82
44	86
45	94
46	95
47	97
48	98
49	99
50	100

Table 3: De novo pathogenic and non-pathogenic variants previously detected using WES used to optimize de novo variant detection

Trio ID	*Gene*	*Genomic position (GRCh37)*	Protein level	Variant in WGS data	Identified as *de novo* mutation?
A: Known pathogenic mutations in patients with severe ID identified using WES					
Pos. control 1	***TCF4***	Chr18(GRCh37):g.52896230C>T	p.(Arg576Gln)	+	*n.a.*
Pos. control 2	***SCN2A***	Chr2(GRCh37):g.166231415G>A	p.(Trp1398*)	+	*n.a*
Pos. control 3	***GRIN2B***	Chr12(GRCh37):g.13764781G>A	p.(Pro553Leu)	+	*n.a*
Pos. control 4	***GRIN2A***	Chr16(GRCh37):g.9928084G>C	p.(Pro522Arg)	+	*n.a*
Pos. control 5	***PDHA1***	ChrX(GRCh37):g.19369435delinsAGA	p.(Pro110Arg*fs**71)	+	*n.a*
Pos. control 6	***GATAD2B***	Chr1(GRCH37):g.153785737G>A	p.(Gln470*)	+	*n.a*
Pos. control 7	***CTNNB1***	Chr3(GRCh37):g.41275106-41275109del	p.(Ser425Thr*fs**11)	+	*n.a*
Pos. control 8	***LRP2***	Chr2(GRCh37):g.170009333del	p.(Gly4146Glu*fs**2)	+	*n.a*
B: De novo mutations identified using WES[2] that are not considered to be the cause of ID					
2	***TRIO***	Chr5(GRCh37):g.14390384A>T	p.(Aps1368Val)	+	+
3	***C15orf40***	Chr15(GRCh37):g.83677401C>T	p.(Ala89Thr)	+	+
3	***RB1***	Chr13(GRCh37):g.48921977T>C	p.(Tyr173His)	+	+
4	***CNGA3***	Chr2(GRCh37):g.99012638G>T	p.(Trp335Cys)	+	+
4	***CXXC11***	Chr2(GRCh37):g.242814942G>A	p.(Gly412Asp)	+	-
7	***CLRN3***	Chr10(GRCh37):g.129690984A>T	p.(Ile22Asn)	+	+
7	***OSBPL9***	Chr1(GRCh37):g.52231512C>T	p.(Pro276Leu)	+	+
10	***MFAP3***	Chr5(GRCh37):g.153429439A>G	p.(Ser53Gly)	+	+
13	***FHDC1***	Chr4(GRCh37):g.153884006G>C	p.(Leu297Phe)	+	+
21	***ALG13***	ChrX(GRCh37):g.110928268A>G	p.(Asn107Ser)	+	+
21	***KRT32***	Chr17(GRCh37):g.39623313C>T	p.(Glu89Lys)	+	+
22	***HADHA***	Chr2(GRCh37):g.26457124C>T	p.(=)	+	+
23	***ZNF831***	Chr20(GRCh37):g.57829608G>A	p.(Arg1615His)	+	+
29	***USP8***	Chr15(GRCh37):g.50790871G>A	p.(=)	+	+
30	***NEK1***	Chr4(GRCh37):g.170359295T>G	p.(Lys901Asn)	+	-*
36	***ATP7B***	Chr13(GRCh37):g.52536008G>A	p.(=)	+	+
39	***EFS***	Chr14(GRCh37):g.23829243G>A	p.(=)	+	+
40	***ZFYVE16***	Chr5(GRCh37):g.79733643C>T	p.(Ala380Val)	+	+
43	***C20orf26***	Chr20(GRCh37):g.20243639A>C	p.(Thr790Pro)	+	+
44	***RBL2***	Chr16(GRCh37):g.53472982G>T	p.(Val99Phe)	+	+
47	***ABCC8***	Chr11(GRCh37):g.17483205G>A	p.(=)	+	+

*+: Identified; -: not identified; n.a.: not assessed. *Variant was detected as somatic variant in the patient after deep-sequencing (see also Figure 1). Trio ID refer to numbers used in this study.*

Table 4: Classification of genes according to the frequency of reported pathogenic mutations

	Total genes	Known ID gene	Candidate ID gene
HGMD with phenotype ID or MR	368	347	21
PubMed search 'WES+ID'	154	102	52
Diagnostic gene list	494	407	87
Large-scale sequencing studies with overlapping phenotypes	462	77	385
Total unique genes		**528**	**628**

Table 5: Overview of de novo mutations in healthy individuals

	Number of controls (c) or siblings (s)	Total	LoF	Missense	Synonymous
Rauch *et al.*	20 (c)	24	2	15	7
O'Roak *et al.*	50 (s)	29	1	19	9
Sanders *et al.**	200 (s)	125	5	82	38
Iossifov *et al.***	343 (s)	288	12	203	73
Gulsuner *et al.*	84 (c)	67	12	36	19
Xu *et al.****	34 (c)	16	0	11	5
Total	**731**	**549**	**32**	**366**	**151**

*Putative LoF includes nonsense, frameshift and canonical splice site mutations based on gene annotation; * study did not consider indels; ** not all de novo mutations were validated by Sanger sequencing; ***excluded splice site variants if not in canonical di-nucleotide of splice site. For some studies the numbers deviate from the numbers given in the main publication, numbers considered here were retrieved from de novo mutation overviews from supplementary tables.*

Table 6: De novo mutations in dominant ID genes in control individuals

chr	genomic position	Mutation type	Ref	Mut	Gene	Protein change	PhyloP	known or candidate ID gene
chr12	51090925	Missense	C	T	*DIP2B*	p.Thr672Met	4.27	candidate
chr17	29684348	Missense	C	T	*NF1*	p.Ala2623Val	5.50	known
chr5	176715871	Missense	C	G	*NSD1*	p.Thr2068Ser	6.08	known
chr10	123260408	Missense	G	A	*FGFR2*	p.Ala498Val	6.22	known
chr17	17697260	inframe insertion	-	AGT	*RAI1*	p.Gln333_Tyr334insVal	-	known
chr12	13716876	frameshift	CG	N	*GRIN2B*	p.Arg1099fs	-	known

Table 7: Calculation of de novo mutations as cause of severe ID

	Diagnostic yield in cohort	Estimated cumulative percentage
Genomic microarray		
(n=1,489 patients with severe ID)		
De novo dominant cause by CNV (n=173)	11.60%	11.60%
Whole Exome Sequencing		
(n=100 array negative patients)		
De novo dominant cause by SNV (n=25)	25.00%	22.10%
De novo dominant cause by CNV (n=1)	1.00%	0.88%
Recessive cause by SNV (n=1)	1.00%	0.88%
Whole Genome Sequencing		
(n=50 WES negative patients)		
De novo dominant cause by SNV (n=13)	26.00%	16.77%
De novo dominant cause by CNV (n=7)	14.00%	8.39%
Recessive cause by SNV (n=1)	2.00%	1.29%

Supplementary Table 2: Average genome sequencing statistics per individual

Genome	Average
Fully called genome fraction	**0.98**
No-called called genome fraction	*0.02*
Reference genome fraction of bases with coverage >= 40x	*0.92*
Total number of called single nucleotide substitution variants	**3,509,927**
Total number of called insertion variants	**384,487**
Total number of called deletion variants	**375,832**
Total number of complex single nucleotide variants	**176,009**
Total CNV segment count	**276**

Exome*	Average
Fully called exome fraction	**0.99**
No-called exome fraction	*0.01*
Exome fraction of bases with coverage >= 40x	*0.94*
Total number of single nucleotide substitution variants	**22,186**
Synonymous SNV loci	10,744
Non-synonymous SNV loci	10,081
Missense	*9,966*
Nonsense	*84*
Disruption stop codon	*12*
Disruption initiation codon	*20*
Block substitutions	*77*
Total number of insertion variants	**590**
Total number of deletion variants	**484**
Total number of complex single nucleotide variants	**1,243**

*Sequencing statistics as supplied by Complete Genomics. Genome fractions are based on the GRCh37 reference genome. Number of variants represent unfiltered total calls as made by Complete Genomics. Functional annotation of SNVs is based on RefSeq annotation by Complete Genomics. *'Exome' as defined by Complete Genomics.*

Supplementary Table 3: Comparison of variants between WES and WGS identified in regions targeted by exome enrichment kit

Overlap for WGS and WES	Average (150 samples)	SD
Total number of variants	**16,964**	**1,716**
Insertions/Deletions	*843*	*110*
In regions with high WES coverage	*1,392*	*1,018*
In regions with low WES coverage	*1,219*	*319*
Variants in coding sequence	**10,812**	**975**

Whole Genome Sequencing only	Average (150 samples)	SD
Total number of variants	**28,028**	**2,187**
Insertions/Deletions	*2,579*	*168*
In regions with high WES coverage	*1,450*	*835*
In regions with low WES coverage	*10,480*	*1,633*
Variants in coding sequence	**11,333**	**1,186**

Whole Exome Sequencing only	Average (150 samples)	SD
Total number of variants	**1,213**	**123**
Insertions/Deletions	*94*	*18*
In regions with high WES coverage	*357*	*92*
In regions with low WES coverage	*60*	*18*
Variants in coding sequence	**622**	**64**

Comparison between variants called by genome sequencing and exome sequencing within the targets of the Agilent 50Mb exome enrichment kit (v2). Variants called exclusively by genome sequencing tended to lie in regions with very low exome coverage, whereas variants called exclusively by exome sequencing tended to lie in regions of very high exome coverage. This latter is likely due to mapping issues in repeat regions. Note, regions of very high exome coverage are defined as regions of at least three times the median exome coverage. Regions of very low exome coverage are regions with less than one third of the median exome coverage.

Supplementary Table 4: Overview of putative de novo SNVs detected through WGS

	High confidence				Medium confidence				Low confidence			
	Sum	**Min**	**Max**	**Median**	**Sum**	**Min**	**Max**	**Median**	**Sum**	**Min**	**Max**	**Median**
Total	**4,081**	**47**	**111**	**82**	**5,034**	**68**	**142**	**98.5**	**62,622**	**909**	**1,500**	**1,279**
Exonic	**80**	**0**	**5**	**1**	**82**	**0**	**5**	**1**	**670**	**6**	**23**	**14**
Missense	*51*	*0*	*1*	*1*	*49*	*0*	*0*	*1*	*360*	*5*	*0*	*10*
Synonymous	*21*	*0*	*2*	*0*	*21*	*0*	*2*	*0*	*189*	*1*	*10*	*3*
Nonsense	*8*	*0*	*2*	*0*	*3*	*0*	*1*	*0*	*20*	*0*	*3*	*0*
In-frame insertions	*0*	*0*	*0*	*0*	*1*	*0*	*1*	*0*	*5*	*0*	*1*	*0*
In-frame deletions	*0*	*0*	*0*	*0*	*1*	*0*	*1*	*0*	*8*	*0*	*1*	*0*
In-frame complex	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*1*	*0*
Out-of-frame insertions	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*67*	*0*	*5*	*1*
Out-of-frame deletions	*0*	*0*	*0*	*0*	*4*	*0*	*1*	*0*	*18*	*0*	*2*	*0*
Out-of-frame complex	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*2*	*0*	*1*	*0*
Splice sites	**6**	**0**	**1**	**0**	**6**	**0**	**1**	**0**	**72**	**0**	**5**	**1**
SA canonical	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*4*	*0*	*1*	*0*
SD canonical	*0*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	*4*	*0*	*1*	*0*
SA	*4*	*0*	*1*	*0*	*3*	*0*	*1*	*0*	*50*	*0*	*4*	*1*
SD	*1*	*0*	*1*	*0*	*1*	*0*	*0*	*0*	*14*	*0*	*3*	*0*
Intronic	**1,382**	**13**	**42**	**26.5**	**1,558**	**17**	**57**	**30.5**	**19,202**	**288**	**504**	**376**
of known ID genes	*34*	*0*	*4*	*0*	*n.a*	*n.a*	*n.a*	*n.a*	*n.a*	*n.a*	*n.a*	*n.a*
UTR	**193**	**0**	**11**	**3.5**	**206**	**0**	**13**	**4**	**2525**	**7**	**82**	**61.5**
of known ID genes	*6*	*0*	*2*	*0*	*n.a*	*n.a*	*n.a*	*n.a*	*n.a*	*n.a*	*n.a*	*n.a*
MicroRNA	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**2**	**0**	**1**	**0**
Non-genic	**2,422**	**31**	**69**	**48**	**3,184**	**41**	**90**	**63**	**40,151**	**590**	**964**	**813.5**

Statistics of de novo mutations identified across 50 trios. De novo calls were categorized in three confidences (see supplementary methods). Top row column header indicates the confidence category; second row column header indicates for each category the total number, the minimum number, the maximum number and the median of variants identified across all 50 trios. Rows indicate functional categorization of variants based on their location in the genome. Splice sites are divided into splice acceptor (SA) and splice acceptor (SD) canonical, being the first two bases into the intron as well as the complete splice acceptor being 20 bp within the intron and the complete splice

Supplementary Table 5: Validation rates of autosomal candidate de novo SNVs grouped by genomic location and confidence level

	Confidence			Total
	Low	Medium	High	
Coding	4/139	17/82	59/80	**80/301** (27%)
Splice sites*	0/15	1/6	4/6	**5/27** (19%)
Intronic	*n.a.*	*n.a.*	42/46	**42/46** (91%)
UTR	*n.a.*	*n.a.*	3/7	**3/7** (42%)
Promoter	*n.a.*	*n.a.*	1/1	**1/1** (100%)
Total	**4/154** (3%)	**18/88** (20%)	**109/140** (78%)	**131/382** (34%)

*n.a. = not assessed. *includes canonical splice sites.*

Supplementary Table 6: Distribution of de novo SNVs in coding sequence identified through WGS

	Number of patients (WGS)
Mutation Negative	**9**
0 mutations	*9*
Mutation Positive	**41**
1 mutation	*15*
2 mutations	*14*
3 mutations	*8*
4 mutations	*3*
5 mutations	*1*

Supplementary Table 7: De novo substitution rates and type of mutations identified through WGS

	Number of mutations WGS
Synonymous	**19**
Predicted effect on splicing	4
No predicted effect on splicing	15
Non-synonymous	**65**
Nonsense	10
Missense	49
Insertion/deletion	5
Canonical splice site	1
De novo* mutation rate	**1.58**

**Calculation of the de novo mutation rate is based on substitutions only.*

Supplementary Table 8: Coding de novo SNVs identified in 50 patients with severe ID

Trio	Gene	Genomic annotation	cDNA	Protein	Mutation type	predicted effect on splicing?[#]	Gene status	Functional evaluation	Nature of mutation	Prediction software	Genomic Location	Expression pattern	Enrichment for gene ontology	Analysis of Mouse phenotypes	Diagnostic interpretation of individual variants
1	*NGFR*	Chr17(GRCh37):g.47583816T>C	NM_002507.3:c.364T>C	p.Cys122Arg	Missense	-		F		P	C	B	G	M	Candidate
2	*GFPT2*	Chr5(GRCh37):g.179728574G>C	NM_005110.2:c.2039C>G	p.Thr680Ser	Missense	-		F		P	C	B	G		Candidate
2	*TRIO*	Chr5(GRCh37):g.14390384A>T	NM_007118.2:c.4103A>T	p.Asp1368Val	Missense	-	K[#]	F				B	G	M	NO*
3	*C15orf40*	Chr15(GRCh37):g.83677401C>T	NM_001160115.1:c.265G>A	p.Ala89Thr	Missense	-				P					NO
3	*RB1*	Chr13(GRCh37):g.48921977T>C	NM_000321.2:c.517T>C	p.Tyr173His	Missense	-				P		B	G	M	NO
3	*SSPO*	Chr7(GRCh37):g.149487581G>T	NM_198455:c.4895G>T	p.Gly1632Val	Missense	-		F							NO
3	*USP42*	Chr7(GRCh37):g.6196670C>G	NM_032172.2:c.3927C>G	p.Leu1309Leu	Synonymous	-						u		M	NO
4	*CNGA3*	Chr2(GRCh37):g.99012638G>T	NM_001298.2:c.1005G>T	p.Trp335Cys	Missense	-				P	C	B	G	M	NO
4	*CXXC11*	Chr2(GRCh37):g.242814942G>A	NM_173821.2:c.1235G>A	p.Gly412Asp	Missense	-				P		B			NO
4	*LPPR3*	Chr19(GRCh37):g.815253C>T	NM_024888.1:c.336G>A	p.Ala112Ala	Synonymous	no		F				u			NO
6	*CDC5L*	Chr6(GRCh37):g.44392245C>T	NM_001253.2:c.1494C>T	p.Ala498Ala	Synonymous	no						B			NO
6	*POP1*	Chr8(GRCh37):g.99142350C>T	NM_001145860.1:c.631C>T	p.Arg211Trp	Missense	-		F		P					NO
6	*WWP2*	Chr16(GRCh37):g.69820943A>T	NM_007014.3:c.30A>T	p.Gly10Gly	Synonymous	possibly		F	D			B	G		Candidate
7	*CLRN3*	Chr10(GRCh37):g.129690984A>T	NM_152311.3:c.65T>A	p.Ile22Asn	Missense	-				P					NO
7	*OSBPL9*	Chr1(GRCh37):g.52231512C>T	NM_148909.3:c.827C>T	p.Pro276Leu	Missense	-				P	C	B			NO
7	*OTX2*	Chr14(GRCh37):g.57268586G>T	NM_021728.2:c.761C>A	p.Ser254*	Nonsense	-		F	D			B	G	M	NO**
7	*TBR1*	Chr2(GRCh37):g.162275551A>G	NM_006593.2:c.1118A>G	p.Gln373Arg	Missense	-	K	F		P	C	B	G	M	Known
7	*YTHDC1*	Chr4(GRCh37):g.69203353T>C	NM_001031732.2:c.396A>G	p.Lys132Lys	Synonymous	no						B			NO
8	*CNOT1*	Chr16(GRCh37):g.58564225C>A	NM_016284.3:c.6204G>T	p.Leu2068Leu	Synonymous	no	K[#]					B	G		NO*

9	*CARD8*	Chr19(GRCh37):g.48734215G>A	NM_001184900.1:c.591C>T	p.Leu197Leu	Synonymous	no						B			NO
9	*ELP2*	Chr18(GRCh37):g.33713240C>T	NM_018255.1:c.178C>T	p.Arg60*	Nonsense	-	K[#]	F	D			B			NO***
9	*WDR45*	ChrX(GRCh37):g.48932518del	NM_007075.3:c.1030del	p.Cys344Alafs*67	Frameshift	-	K	F	D			B			Known
10	*LIPI*	Chr21(GRCh37):g.15535845C>T	NM_198996.2:c.965-1G>A	-	Splice site	possibly							G	M	NO
10	*MFAP3*	Chr5(GRCh37):g.153429439A>G	NM_005927.4:c.157A>G	p.Ser53Gly	Missense	-						B			NO
13	*FHDC1*	Chr4(GRCh37):g.153884006G>C	NM_033393.2:c.891G>C	p.Leu297Phe	Missense	-		F		P		B			NO
13	*FHOD1*	Chr16(GRCh37):g.67281195G>A	NM_013241.2:c.119C>T	p.Ala40Val	Missense	-					C				NO
13	*SMC1A*	ChrX(GRCh37):g.53430554del	NM_006306.2:c.2364del	p.Asn788Lysfs*10	Frameshift	-	K	F	D			B	G	M	Known
13	*ZNF566*	Chr19(GRCh37):g.36940271C>T	NM_001145345.1:c.865G>A	p.Gly289Arg	Missense	-				P	C	B			NO
14	*GPR52*	Chr1(GRCh37):g.174417259T>G	NM_005684.4:c.10T>G	p.Ser4Ala	Missense	-						B			NO
15	*DSP*	Chr6(GRCh37):g.7583973C>T	NM_004415.2:c.6478C>T	p.Arg2160*	Nonsense	-			D					M	NO
15	*PTPN21*	Chr14(GRCh37):g.88945321C>T	NM_007039.3:c.2454G>A	p.Pro818Pro	Synonymous	no									NO
15	*SPTAN1*	Chr9(GRCh37):g.131331084G>A	NM_001130438.2:c.271G>A	p.Glu91Lys	Missense	-	K	F		P	C	B	G	M	Known
17	*ASUN*	Chr12(GRCh37):g.27087494G>A	NM_018164.2:c.295C>T	p.Gln99*	Nonsense	-		F	D			B			Candidate
17	*PI4K2B*	Chr4(GRCh37):g.25236025G>C	NM_018323.3:c.240G>C	p.Glu80Asp	Missense	-									NO
19	*TBKBP1*	Chr17(GRCh37):g.45786784A>C	NM_014726.2:c.1685A>C	p.His562Pro	Missense	-						B			NO
20	*SON*	Chr21(GRCh37):g.34923418_34923419del	NM_138927.1:c.1881_1882del	p.Val629Alafs*56	Frameshift	-		F	D			B		u	NO
21	*ALG13*	ChrX(GRCh37):g.110928268A>G	NM_001099922.2:c.320A>G	p.Asn107Ser	Missense	-	K[#]	F		P		B	G	M	Known****
21	*CBLB*	Chr3(GRCh37):g.105421227G>A	NM_170662.3:c.1670C>T	p.Pro557Leu	Missense	-				P	C	B		M	NO
21	*KRT32*	Chr17(GRCh37):g.39623313C>T	NM_002278.3:c.265G>A	p.Glu89Lys	Missense	-				P					NO
21	*RAI1*	Chr17(GRCh37):g.17696524C>T	NM_030665.3:c.262C>T	p.Gln88*	Nonsense	-	K	F	D			B	G	M	Known
22	*HADHA*	Chr2(GRCh37):g.26457124C>T	NM_000182.4:c.414G>A	p.Val138Val	Synonymous	no						B	G	M	NO
22	*MED13L*	Chr12(GRCh37):g.116435026T>C	NM_015335.4:c.2579A>G	p.Asp860Gly	Missense	-	K[#]	F		P	C	B			Known
23	*POLQ*	Chr3(GRCh37):g.121208286C>T	NM_199420.3:c.3492G>A	p.Glu1164Glu	Synonymous	no									NO
23	*ZNF831*	Chr20(GRCh37):g.57829608G>A	NM_178457.1:c.4844G>A	p.Arg1615His	Missense	-				P		u			NO
24	*BRD3*	Chr9(GRCh37):g.136913290A>G	NM_007371.3:c.1001T>C	p.Phe334Ser	Missense	-		F		P	C	B			Candidate
24	*SEZ6*	Chr17(GRCh37):g.27286180C>T	NM_178860.4:c.1970G>A	p.Arg657Gln	Missense	possibly		F				B	G	M	NO
24	*TNIP3*	Chr4(GRCh37):g.122078277A>G	NM_001128843.1:c.335T>C	p.Leu112Pro	Missense	-									NO

25	*SATB2*	Chr2(GRCh37):g.200213667_200213668insGTTGCCTTACAA	NM_001172517.1:c.929_930insTTGTAAGGCAAC	p.Gln310delinsHisCysLys AlaThr	Insertion	no	K	F	D		C	B	G	M	Known
26	*ELMO2*	Chr20(GRCh37):g.45022193C>A	NM_133171.3:c.167G>T	p.Gly56Val	Missense	-				P	C	B			NO
26	*PPP2R5D*	Chr6(GRCh37):g.42975030T>A	NM_006245.2:c.619T>A	p.Trp207Arg	Missense	-	K[#]	F		P	C	B	G	M	Known
27	*CATSPERB*	Chr14(GRCh37):g.92136238G>T	NM_024764.2:c.1207C>A	p.Arg403Arg	Synonymous	possibly									NO
27	*KCNA1*	Chr12(GRCh37):g.5021656C>T	NM_000217.2:c.1112C>T	p.Thr371Ile	Missense	-	K	F		P	C	B	G	M	Known
27	*SPACA7*	Chr13(GRCh37):g.113030776C>T	NM_145248.4:c.77C>T	p.Pro26Leu	Missense	-						u			NO
28	*DOK4*	Chr16(GRCh37):g.57513380C>A	NM_018110.3:c.40G>T	p.Val14Leu	Missense	-		F			C	B	G		NO
28	*SCN2A*	Chr2(GRCh37):g.166243265C>T	NM_001040142.1:c.4561C>T	p.Gln1521*	Nonsense	-	K	F	D			B	G	M	Known
29	*USP8*	Chr15(GRCh37):g.50790871G>A	NM_001128611.1:c.3117G>A	p.Gln1039Gln	Synonymous	no						B			NO
30	*MAST1*	Chr19(GRCh37):g.12984501C>G	NM_014975.2:c.3530C>G	p.Pro1177Arg	Missense	-		F		P	C	B	G		Candidate
30	*NEK1*	Chr4(GRCh37):g.170359295T>G	NM_001199397.1:c.2703A>C	p.Lys901Asn	Missense	-						B		M	NO
32	*PNLIPRP1*	Chr10(GRCh37):g.118354347C>T	NM_006229.2:c.436C>T	p.Gln146*	Nonsense	-			D				G		NO
33	*CDC34*	Chr19(GRCh37):g.541442G>A	NM_004359.1:c.601G>A	p.Glu201Lys	Missense	-		F				B			NO
33	*FAM46B*	Chr1(GRCh37):g.27339042G>A	NM_052943.3:c.120C>T	p.Ala40Ala	Synonymous	no		-							NO
34	*APPL2*	Chr12(GRCh37):g.105591609G>C	NM_018171.3:c.986C>G	p.Ser329*	Nonsense	-		F	D			B	G	M	Candidate
34	*SMAD6*	Chr15(GRCh37):g.66995638G>A	NM_005585.4:c.42G>A	p.Trp14*	Nonsense	-			D					M	NO
34	*ZNF423*	Chr16(GRCh37):g.49670125T>C	NM_015069.3:c.2938A>G	p.Thr980Ala	Missense			F				B			NO
36	*ATP7B*	Chr13(GRCh37):g.52536008G>A	NM_000053.3:c.1911C>T	p.Asn637Asn	Synonymous	no						B	G	M	NO
36	*CEP170B*	Chr14(GRCh37):g.105352391G>C	NM_001112726.2:c.1959G>C	p.Pro653Pro	Synonymous	no						u			NO
37	*PIAS1*	Chr15(GRCh37):g.68468014T>A	NM_016166.1:c.1209T>A	p.Asp403Glu	Missense	-	K[#]	F				u	G		NO*
38	*GIPC2*	Chr1(GRCh37):g.78511993G>T	NM_017655.4:c.215G>T	p.Gly72Val	Missense	-									NO
39	*EFS*	Chr14(GRCh37):g.23829243G>A	NM_005864.2:c.444C>T	p.Tyr148Tyr	Synonymous	no						B			NO
40	*ZFYVE16*	Chr5(GRCh37):g.79733643C>T	NM_014733.3:c.1139C>T	p.Ala380Val	Missense	no						B	G		NO
41	*CLSTN3*	Chr12(GRCh37):g.7295875C>T	NM_014718.3:c.1815C>T	p.Gly605Gly	Synonymous	no		F				B			NO
41	*NACC1*	Chr19(GRCh37):g.13249038C>T	NM_052876.2:c.1402C>T	p.Arg468Cys	Missense	-		F			C	B			Candidate
42	*HIVEP2*	Chr6(GRCh37):g.143092683C>T	NM_006734.3:c.3193G>A	p.Ala1065Thr	Missense	-	K[#]	F				B			NO*
43	*C20orf26*	Chr20(GRCh37):g.20243639A>C	NM_015585.3:c.2368A>C	p.Thr790Pro	Missense	-				P	C	B			NO
43	*EYA4*	Chr6(GRCh37):g.133777699C>G	NM_004100.4:c.283C>G	p.Leu95Val	Missense	-	K[#]				C			M	NO*

43	*POGZ*	Chr1(GRCh37):g.151378510G>A	NM_015100.3:c.3001C>T	p.Arg1001*	Nonsense	-	K#	F	D			B				Known
44	*RBL2*	Chr16(GRCh37):g.53472982G>T	NM_005611.3:c.295G>T	p.Val99Phe	Missense	-				P		B			M	NO
46	*CREBL2*	Chr12(GRCh37):g.12788868G>C	NM_001310.2:c.173G>C	p.Arg58Pro	Missense	-				P		B	G			NO
46	*TBR1*	Chr2(GRCh37):g.162280277_162280283dup	NM_006593.2:c.1588_1594dup	p.Thr532Argfs*144	Frameshift	-	K	F	D			B	G		M	Known
47	*ABCC8*	Chr11(GRCh37):g.17483205G>A	NM_000352.3:c.747C>T	p.Ala249Ala	Synonymous	no						B	G			NO
48	*EPHB1*	Chr3(GRCh37):g.134514481T>G	NM_004441.4:c.8T>G	p.Leu3Arg	Missense	-		F				B	G		M	NO
48	*UCP1*	Chr4(GRCh37):g.141489817C>T	NM_021833.4:c.67G>A	p.Ala23Thr	Missense	-							G		M	NO
49	*ERV3-1*	Chr7(GRCh37):g.64452368A>C	NM_001007253.3:c.1037T>G	p.Ile346Ser	Missense	-						B				NO
49	*KANSL2*	Chr12(GRCh37):g.49072911C>A	NM_017822.3:c.453G>T	p.Gly151Gly	Synonymous	possibly	K#	F	D		C	u				Known

*K: Known ID gene; K#: Candidate ID gene; F: Functional link to ID; D: deleterious mutation; P: Predicted pathogenic mutation; C: Conserved genomic location (not assessed for deleterious mutations); B: Brain expressed; G: Annotated with GO BP terms that are enriched in known ID genes; M: Annotated with mouse phenotypes that are enriched in known ID genes.*Despite being a mutation in a candidate ID gene, this mutation is not considered the cause of ID as the mutation impact on protein function is not (predicted) pathogenic. **: the OTX2 mutation in this patient is not diagnostically reported as cause of ID due to the classical phenotype for OTX2 mutations not being observed (ocular anomalies and pituitary deficiency). Alternatively, the phenotypic consequences of the OTX2 mutation in this patient may be atypic and superseded by the de novo TBR1 mutation. ***ELP2 mutations are known for autosomal recessive ID.[77] In this patient, no second (inherited) pathogenic mutation was encountered, suggesting that this mutation creates a carrier ship rather than a cause for disease. This explanation is further supported by the fact that the patient has a pathogenic mutation in WDR45, fully explaining the patients phenotype. ****: Additional ALG13 mutations have been reported in patients with ID since selection of this patient cohort.[46,78]; In retrospect, this mutation may have clinical consequences. Nonetheless, a nonsense mutation in RAI1 has been identified in the same patient during this study, which seemingly explains the full phenotype of the patient. Alternatively, the phenotype due to the ALG13 mutation may be merged into the RAI1 phenotype, making it impossible to distinguish which part of the phenotype is caused by which mutation. This phenomenon, that two (de novo) pathogenic mutations are identified in a single patient has been described before.[79]*

Supplementary Table 9: Coding de novo mutation rates of published studies

Study	Number of individuals	Patients/ Controls	DIsease*	# *de novo* SNVs excl indels	# *de novo* SNVs incl indels	Coding mutation rate SNVs excl indels	Coding mutation rate SNVs incl indels
This study	50	Patients	ID	79	84	1.58	1.68
Rauch et al. [32]	51	Patients	ID	72	85	1.41	1.67
O'Roak et al.[33]	189	Patients	ASD	225	242	1.19	1.28
Epi4K[46]	264	Patients	EE	309	329	1.17	1.25
Jiang et al.[34]	32	Patients	ASD	36	38	1.13	1.19
Rauch et al.[32]	20	Controls	-	22	23	1.10	1.15
Neale et al.[17]	175	Patients	ASD	161	167	0.92	0.95
Gulsuner et al.[51]	105	Patients	SCHZ	96	103	0.91	0.98
Iossifov et al.[31]	343	Patients	ASD	305	305	0.89	0.89
Iossifov et al.[31]	343	Controls	-	294	294	0.86	0.86
Gulsuner et al.[51]	84	Controls	-	66	67	0.79	0.80
de Ligt et al.[2]	100	Patients	ID	72	79	0.72	0.79
Sanders et al.[30]	238	Patients	ASD	164	167	0.69	0.70
Xu et al.[53]	53	Patients	SCHZ	36	40	0.68	0.75
Sanders et al.[30]	200	Controls	-	125	125	0.63	0.63
Xu et al.[53]	22	Controls	-	8	8	0.36	0.36

**: ID: Intellectual Disability; EE: Epileptic Encephalopathies; SCHZ: Schizophrenia; ASD: Autism Spectrum Disorder*

Supplementary Table 10: List of 528 known ID genes used for prioritization of variants

Gene										
ABCC9	*ASXL1*	*CHD7*	*ERCC5*	*GPC3*	*KIF11*	*MTR*	*PAFAH1B1*	*PYCR1*	*SLC35C1*	*TNK2*
ABCD1	*ATP1A2*	*CHKB*	*ERCC6*	*GPHN*	*KIF7*	*MTRR*	*PAK3*	*RAB18*	*SLC4A4*	*TOR1A*
ABCD4	*ATP2A2*	*CHRND*	*ERCC8*	*GPR56*	*KIRREL3*	*MUT*	*PANK2*	*RAB27A*	*SLC6A3*	*TPK1*
ABHD5	*ATP6V0A2*	*CLCN7*	*ESCO2*	*GRIA3*	*KIT*	*MVK*	*PARP1*	*RAB3GAP1*	*SLC6A8*	*TPO*
ACAD9	*ATP7A*	*CLN8*	*ETFB*	*GRIN2A*	*KNCQ2*	*MYCN*	*PAX6*	*RAB3GAP2*	*SLC7A7*	*TRAPPC9*
ACOX1	*ATR*	*CNTNAP2*	*ETHE1*	*GRIN2B*	*KRAS*	*MYH9*	*PC*	*RAF1*	*SLC9A6*	*TREX1*
ACSF3	*ATRX*	*COG8*	*EXOSC3*	*GRM1*	*L1CAM*	*MYO1E*	*PCDH19*	*RAI1*	*SMAD4*	*TRIM32*
ACTB	*ATXN1*	*COL4A1*	*EXT1*	*GSS*	*L2HGDH*	*MYO3A*	*PCNT*	*RARS2*	*SMARCA2*	*TRPM6*
ACTG1	*AUH*	*COL4A2*	*FAM123B*	*GUSB*	*LAMA2*	*MYT1L*	*PDE4D*	*RASGEF1B*	*SMARCA4*	*TSC1*
ACVR1	*AUTS2*	*COLEC11*	*FANCB*	*HAX1*	*LAMC3*	*NAGA*	*PDHA1*	*RBFOX1*	*SMARCB1*	*TSC2*
ADAR	*B3GALTL*	*COQ2*	*FANCD2*	*HCCS*	*LAMP2*	*NAGLU*	*PEPD*	*RFT1*	*SMARCE1*	*TSPAN7*
ADCK3	*BBS1*	*CREBBP*	*FBN1*	*HCFC1*	*LARGE*	*NBN*	*PEX1*	*RFX3*	*SMC1A*	*TTC8*
ADSL	*BBS10*	*CTNNB1*	*FGD1*	*HDAC4*	*LCT*	*NDE1*	*PEX10*	*RNASEH2A*	*SMOC1*	*TTN*
AFF2	*BBS12*	*CUBN*	*FGFR2*	*HDAC8*	*LIG4*	*NDP*	*PEX13*	*RNASEH2B*	*SMPD1*	*TUBA1A*
AGA	*BBS2*	*CUL3*	*FGFR3*	*HESX1*	*LRP1*	*NDUFS1*	*PEX26*	*RNASEH2C*	*SMS*	*TUBB2B*
AGPAT2	*BBS4*	*CUL4B*	*FH*	*HLCS*	*LRP2*	*NDUFS2*	*PEX5*	*ROGDI*	*SNAP29*	*TUSC3*
AGTR2	*BBS5*	*CYB5R3*	*FKRP*	*HMGA2*	*MAN2B1*	*NDUFS4*	*PEX7*	*RP2*	*SOS1*	*TWIST1*
AHCY	*BBS7*	*D2HGDH*	*FKTN*	*HNRNPU*	*MANBA*	*NDUFS8*	*PGAP2*	*RPGRIP1L*	*SOX10*	*UBE2A*
AHI1	*BBS9*	*DARS2*	*FLG*	*HOXA1*	*MAP2K1*	*NDUFV1*	*PGK1*	*RPS6KA3*	*SOX2*	*UBE3A*
AK1	*BCKDHA*	*DBT*	*FLNA*	*HPD*	*MAP2K2*	*NEDD4L*	*PHF21A*	*RTEL1*	*SOX3*	*UBR1*
AKT3	*BCKDHB*	*DCX*	*FMR1*	*HPRT1*	*MAPT*	*NEU1*	*PHF6*	*RUNX1*	*SOX5*	*UPB1*
ALDH18A1	*BCOR*	*DDHD2*	*FOXG1*	*HRAS*	*MAT1A*	*NF1*	*PHF8*	*RYR1*	*SPG7*	*UPF3B*
ALDH3A2	*BCS1L*	*DHCR24*	*FOXP1*	*HSD17B10*	*MBD5*	*NFIA*	*PHGDH*	*RYR2*	*SPRED1*	*VLDLR*
ALDH5A1	*BLM*	*DHCR7*	*FOXP2*	*HUWE1*	*MBTPS2*	*NFIX*	*PIGV*	*SALL1*	*SPTAN1*	*VPS13B*
ALG1	*BRAF*	*DISC1*	*FRAS1*	*IDS*	*MCCC1*	*NHS*	*PLCB1*	*SATB2*	*SRCAP*	*VPS39*
ALG12	*BSCL2*	*DKC1*	*FTO*	*IDUA*	*MCCC2*	*NIPBL*	*PLP1*	*SBF1*	*SRD5A3*	*WDR11*
ALG3	*BUB1B*	*DLD*	*FTSJ1*	*IFT172*	*MCOLN1*	*NKX2-1*	*PMM2*	*SCN1A*	*SRGAP3*	*WDR19*
ALG6	*C5ORF42*	*DLG3*	*FUCA1*	*IGF1*	*MCPH1*	*NLGN1*	*PNKP*	*SCN2A*	*STRA6*	*WDR45*
AMT	*C7orf11*	*DMD*	*GABRA1*	*IKBKG*	*MECP2*	*NLGN4X*	*PNP*	*SCN8A*	*STS*	*WDR62*
ANK2	*CA2*	*DNAH5*	*GABRB3*	*IL1RAPL1*	*mecp2e1*	*NLRP3*	*POLG*	*SCO2*	*STXBP1*	*XPA*
ANK3	*CACNA1A*	*DNMT3B*	*GALE*	*INPP5E*	*MED12*	*NOG*	*POLR3A*	*SDHA*	*SUCLA2*	*YWHAE*
ANKH	*CACNA1C*	*DOCK8*	*GALT*	*IQSEC2*	*MEF2C*	*NOTCH3*	*POLR3B*	*SERAC1*	*SUOX*	*ZDHHC9*
ANKRD11	*CASK*	*DPAGT1*	*GAMT*	*IRS1*	*MGAT2*	*NPHP1*	*POMGNT1*	*SETBP1*	*SURF1*	*ZEB2*
ANO5	*CBS*	*DPM1*	*GATM*	*ISPD*	*MID1*	*NR0B1*	*POMT1*	*SHANK2*	*SYN1*	*ZFHX4*
ANTXR1	*CC2D2A*	*DPYD*	*GCH1*	*ITGB3*	*MKKS*	*NR4A2*	*POMT2*	*SHANK3*	*SYNE1*	*ZFYVE26*
AP1S2	*CCBE1*	*DYM*	*GDI1*	*ITPR1*	*MLH1*	*NRXN1*	*PORCN*	*SHH*	*SYNGAP1*	*ZIC2*
AP3B1	*CCDC22*	*DYNC1H1*	*GFAP*	*KANK1*	*MLL*	*nrxn1b*	*POU1F1*	*SHOC2*	*SYP*	*ZNF41*
APAF1	*CCDC39*	*DYRK1A*	*GJC2*	*KANSL1*	*MLL2*	*NSD1*	*POU3F4*	*SHOX*	*TAT*	*ZNF674*
APTX	*CDH15*	*EFTUD2*	*GK*	*KAT6B*	*MLL3*	*NSDHL*	*PPM1D*	*SHROOM4*	*TBC1D24*	
ARFGEF2	*CDK5RAP2*	*EHMT1*	*GLDC*	*KCNA1*	*MLYCD*	*NSUN2*	*PPOX*	*SIL1*	*TBR1*	
ARHGEF9	*CDKL5*	*EIF2AK3*	*GLI2*	*KCNJ10*	*MMAA*	*NTRK1*	*PRODH*	*SIX3*	*TCF4*	
ARID1A	*CDON*	*EIF4G1*	*GLI3*	*KCNJ11*	*MMACHC*	*OBSL1*	*PRPS1*	*SKI*	*TFAP2A*	
ARID1B	*CELSR1*	*ELOVL4*	*GLRA1*	*KCNQ2*	*MMADHC*	*OCA2*	*PSEN1*	*SLC12A6*	*TGFBR1*	

ARL6	*CENPJ*	*EP300*	*GMPPB*	*KCNQ3*	*MOCS1*	*OCLN*	*PTCH1*	*SLC16A2*	*TGFBR2*
ARSE	*CEP152*	*EPHA5*	*GNAS*	*KCNT1*	*MOCS2*	*OCRL*	*PTCHD1*	*SLC17A5*	*TGIF1*
ARX	*CEP290*	*EPHB2*	*GNPAT*	*KCTD7*	*MPDU1*	*OFD1*	*PTDSS1*	*SLC25A15*	*THRB*
ASL	*CEP41*	*ERBB4*	*GNRHR*	*KDM5C*	*MPDZ*	*OPHN1*	*PTEN*	*SLC26A9*	*TIMM8A*
ASPA	*CHAT*	*ERCC2*	*GNS*	*KDM6A*	*MSH6*	*ORC1*	*PTPN11*	*SLC2A1*	*TMEM237*
ASPM	*CHD2*	*ERCC3*	*GP1BB*	*KIAA1279*	*MTMR2*	*OTC*	*PVRL1*	*SLC33A1*	*TMEM67*

Supplementary Table 11: List of 628 Candidate ID genes used for prioritization of variants

Gene									
ABCA2	*C16orf62*	*DDX11*	*FGD3*	*KIAA0317*	*NDUFA1*	*PPP2R2C*	*SLC25A22*	*TMLHE*	*ZNF480*
ABCG4	*C17orf53*	*DDX20*	*FGF22*	*KIAA1033*	*NDUFA11*	*PPP2R5D*	*SLC25A39*	*TMPRSS12*	*ZNF526*
ABI3BP	*C1QTNF6*	*DDX3X*	*FLVCR1*	*KIAA1109*	*NDUFA12*	*PQBP1*	*SLC25A5*	*TNKS2*	*ZNF565*
ACACB	*C6orf174*	*DDX50*	*FMN2*	*KIAA1324L*	*NDUFS3*	*PRDM12*	*SLC31A1*	*TNPO2*	*ZNF592*
ACBD6	*C7orf43*	*DEAF1*	*FREM3*	*KIAA1462*	*NDUFS7*	*PRDX6*	*SLC35A2*	*TNR*	*ZNF673*
ACO2	*C9orf86*	*DENR*	*FRMPD4*	*KIAA2018*	*NFASC*	*PRKCA*	*SLC4A10*	*TOP1*	*ZNF711*
ACSL4	*CA8*	*DGCR14*	*FRY*	*KIAA2022*	*NGEF*	*PRKCB*	*SLC4A8*	*TRAK1*	*ZNF81*
ADAT3	*CACNA1E*	*DGCR2*	*GABRB1*	*KIF14*	*NISCH*	*PRKRA*	*SLC6A1*	*TRAPPC11*	
ADCY7	*CACNA1G*	*DGKH*	*GAD1*	*KIF1A*	*NLGN2*	*PRMT10*	*SLC6A13*	*TRIM29*	
ADK	*CACNG2*	*DHDDS*	*GAS2*	*KIF5C*	*NLRC5*	*PROX2*	*SMARCC2*	*TRIM8*	
ADNP	*CAMK2G*	*DHFR*	*GATAD2B*	*KLF12*	*NOTUM*	*PRPF39*	*SMC3*	*TRIO*	
ADRA2B	*CAP1*	*DHRS4L1*	*GCSH*	*KLF8*	*NPAS4*	*PRSS12*	*SMG9*	*TRIP12*	
AGPAT3	*CARKD*	*DHTKD1*	*GIMAP8*	*KLHL11*	*NPRL2*	*PSMA7*	*SMURF1*	*TRMT1*	
AIFM1	*CASP2*	*DIAPH3*	*GNAO1*	*KPNA1*	*NR2F1*	*PSMG4*	*SNIP1*	*TRMT10A*	
AIMP1	*CC2D1A*	*DIP2B*	*GOLGA3*	*KRT80*	*NRN1*	*PTGR1*	*SNX3*	*TRPC5*	
AKR1C4	*CCDC137*	*DIP2C*	*GON4L*	*LAMA1*	*NTNG1*	*PTPRK*	*SOBP*	*TRPM3*	
ALG2	*CCDC18*	*DLG1*	*GPD2*	*LARP7*	*NUAK1*	*PTPRM*	*SP7*	*TRPM5*	
ALG9	*CCDC78*	*DLG4*	*GPR115*	*LINS*	*NUP54*	*PTPRR*	*SPAG17*	*TRPM7*	
ALS2CL	*CCDC8*	*DLL1*	*GPR153*	*LRPPRC*	*NXF5*	*PTPRT*	*SPATA13*	*TRRAP*	
ALX1	*CCDC84*	*DLX3*	*GPR84*	*LTN1*	*OMG*	*PUS1*	*SPATA5*	*TSPAN17*	
AMY2B	*CCNT1*	*DMPK*	*GPS1*	*LZTR1*	*OPLAH*	*PWWP2A*	*SRBD1*	*TSPYL5*	
ANKRD12	*CDC42BPB*	*DNAH17*	*GRB14*	*MADD*	*OPRL1*	*RAB2A*	*SRPR*	*TTI2*	
ANO10	*CDK11A*	*DNAH7*	*GRIA1*	*MAGEL2*	*OR10S1*	*RAB33B*	*SRPX2*	*TUBGCP6*	
AP2A2	*CDK5R1*	*DNAJB9*	*GRIA2*	*MAGT1*	*OR5M1*	*RAB39B*	*SSBP3*	*UBE3B*	
AP3B2	*CDK6*	*DNAJC19*	*GRIK2*	*MAN1B1*	*OSBPL5*	*RAB40AL*	*ST3GAL3*	*UBE3C*	
AP3M1	*CDKL3*	*DNAJC6*	*GRIN1*	*MAOA*	*PACS1*	*RAB5C*	*ST3GAL5*	*UBN2*	
AP4B1	*CDS2*	*DNM1*	*GRM5*	*MAPK10*	*PACS2*	*RAD21*	*ST3GAL6*	*UBQLN1*	
AP4E1	*CELF2*	*DNMT3A*	*GRM7*	*MAPK8IP1*	*PAFAH1B3*	*RAD21L1*	*ST5*	*UBR3*	
AP4S1	*CEP135*	*DOCK9*	*GSPT2*	*MBTPS1*	*PALLD*	*RALGAPB*	*STAG1*	*UBR5*	
APH1A	*CHAMP1*	*DPP3*	*GTF2H5*	*MCAM*	*PAX1*	*RALGDS*	*STAP2*	*UBR7*	
ARHGAP30	*CHD1*	*DPP7*	*GTPBP8*	*MCM3AP*	*PAX5*	*RALGPS1*	*STIL*	*UBTF*	
ARHGEF10	*CHD3*	*DRD4*	*H2AFV*	*MED13*	*PBRM1*	*RAPGEF1*	*STK36*	*UGGT1*	
ARHGEF6	*CHD6*	*DST*	*HARS*	*MED13L*	*PCDH11X*	*RARG*	*STT3A*	*UNC13C*	
ARIH1	*CHL1*	*DUS1L*	*HDGFRP2*	*MED17*	*PCDH18*	*RARS*	*STT3B*	*UNC80*	
ARL13B	*CHRNA7*	*DUSP15*	*HDLBP*	*MED23*	*PCDHA13*	*RASGRP1*	*STX1B*	*UPF2*	
ARMC9	*CIC*	*EBAG9*	*HDX*	*MEGF11*	*PCDHB13*	*RASIP1*	*STXBP3*	*URB2*	
ASAH2	*CIT*	*EDA2R*	*HECTD1*	*MEOX2*	*PCDHB4*	*RB1CC1*	*SUPT16H*	*USP15*	
ASB1	*CLIC2*	*EEF1A2*	*HERC2*	*METTL14*	*PCNX*	*RBM10*	*SV2B*	*USP46*	
ASCC3	*CLK2*	*EEF1B2*	*HIPK3*	*MGAT4C*	*PCOLCE*	*RBM28*	*SVIL*	*UTP14C*	
ASCL1	*CNKSR1*	*EFHC2*	*HIST1H1E*	*MIB1*	*PDCD1*	*RBMS3*	*SYNCRIP*	*VCX3A*	
ASH1L	*CNKSR2*	*EFR3A*	*HIST1H2AE*	*MKLN1*	*PDIA6*	*RECK*	*SYNRG*	*VPS37A*	
ASNS	*CNOT1*	*EIF2C1*	*HIST1H2AG*	*MLL4*	*PDIK1L*	*RELN*	*SYT14*	*WAC*	
ATP10D	*CNOT3*	*EIF2S3*	*HIST3H3*	*MLL5*	*PDSS1*	*REPS2*	*SZT2*	*WDFY3*	
ATP12A	*CNOT4*	*ELK1*	*HIVEP2*	*MMP27*	*PDSS2*	*REST*	*TAAR2*	*WDR13*	

ATP1B1	*CNTN5*	*ELP2*	*HTR7*	*MMP8*	*PECR*	*RGMA*	*TAF1*	*WDR45L*
ATP2B4	*COG1*	*EMILIN3*	*IER3IP1*	*MOB4*	*PEX11B*	*RGS14*	*TAF2*	*WDR62*
ATP6AP2	*COG6*	*EMX2*	*IFT81*	*MPHOSPH8*	*PGRMC1*	*RGS7*	*TAF7L*	*WHSC1L1*
ATP8A2	*COG7*	*ENTPD1*	*IGBP1*	*MPP6*	*PHACTR1*	*RIMS1*	*TANC2*	*XPO1*
B3GNT4	*COL25A1*	*EPB41L1*	*inpp4a*	*MRPS22*	*PHF19*	*RIMS2*	*TBC1D14*	*XPO5*
B4GALT1	*COL4A3BP*	*EPC2*	*IQGAP2*	*MRS2*	*PHF2*	*RIOK3*	*TBC1D7*	*XPR1*
B4GALT7	*COLEC12*	*ERLIN2*	*IQSEC1*	*MTF1*	*PHIP*	*RMND1*	*TBCE*	*XYLT1*
BCAP31	*COQ5*	*ESAM*	*ITGA5*	*MTMR12*	*PIAS1*	*RNF38*	*TBL1XR1*	*YWHAG*
BCORL1	*COX15*	*EXOC6B*	*JAM3*	*MTOR*	*PIGL*	*RREB1*	*TBX18*	*YY1*
BEST3	*CRADD*	*EYA4*	*JARID2*	*MYH10*	*PIGN*	*RRP1B*	*TCF7L1*	*ZBTB40*
BIRC6	*CRBN*	*FAAH2*	*KANSL2*	*MYO5A*	*PIGO*	*RTN4RL1*	*TECPR2*	*ZBTB41*
BMP1	*CRK*	*FAM102A*	*KATNAL2*	*MYO7B*	*PIGT*	*RUNX1T1*	*TECR*	*ZC3H12B*
BRAT1	*CRTAC1*	*FAM116B*	*KCNB1*	*MYOF*	*PIK3R2*	*RUVBL1*	*TFDP1*	*ZC3H14*
BRD4	*CSDE1*	*FAM129B*	*KCNH1*	*MYT1*	*PIK3R3*	*RXFP1*	*TGM3*	*ZCCHC12*
BRPF1	*CSNK1E*	*FAM13C*	*KCNK12*	*N6AMT1*	*PIWIL4*	*S100G*	*THBS1*	*ZCCHC8*
BRSK2	*CTCF*	*FAM45A*	*KCNK9*	*NAA10*	*PLA1A*	*SAP30BP*	*THOC2*	*ZDHHC15*
BRWD1	*CTDP1*	*FAM63B*	*KDM1A*	*NAA40*	*PLCL2*	*SC5DL*	*THOC6*	*ZFHX3*
BRWD3	*CTTNBP2*	*FAM8A1*	*KDM1B*	*NACA*	*PLP2*	*SCAF4*	*TLK2*	*ZMYM3*
BTN1A1	*CTTNBP2NL*	*FAM91A1*	*KDM2B*	*NALCN*	*PNPLA7*	*SCP2*	*TMC4*	*ZMYM6*
C10orf11	*CUL5*	*FASN*	*KDM5A*	*NAPRT1*	*POC1A*	*SDCBP2*	*TMCO1*	*ZNF238*
C11orf46	*CUX2*	*FASTKD5*	*KDM5B*	*NAV2*	*POGZ*	*SEMA4G*	*TMEM135*	*ZNF292*
C12orf57	*CYP4F3*	*FAT1*	*KDM6B*	*NBEA*	*POLR2A*	*SETD5*	*TMEM165*	*ZNF385B*
C12orf65	*CYTH1*	*FBXW9*	*KIAA0100*	*NCAPD2*	*POLR2M*	*SFPQ*	*TMEM231*	*ZNF407*
C15orf38	*DBR1*	*FCRL6*	*KIAA0182*	*NCKAP1*	*POLRMT*	*SLC16A3*	*TMEM41A*	*ZNF44*
C15orf62	*DCAF4*	*FETUB*	*KIAA0232*	*NDST1*	*PPP1R15B*	*SLC1A2*	*TMEM85*	*ZNF451*

Supplementary Table 12: Summary of de novo non-coding SNVs identified in 528 known ID genes

Non-coding component	Number of mutations
Promoter	1
5'UTR	2
Intron	38
Non-canonical splice site	1
3'UTR	1

Supplementary Table 13: De novo SNVs in non-coding sequence identified in patients with severe ID

Trio	Gene	Location	Genomic annotation	cDNA	PhyloP	Predicted effect on splicing?	ENCODE annotation
1	***LAMA2***	Intron	Chr6(GRCh37):g.129635764C>T	NM_000426.3:c.3412-36C>T	-0.59	no	
6	***MCOLN1***	Intron	Chr19(GRCh37):g.7597165G>C	NM_020533.2:c.1576-1244G>C	-0.44	no	
7	***CNTNAP2***	Intron	Chr7(GRCh37):g.145833308G>T	NM_014141.5:c.97+19243G>T	0.21	no	
8	***GRM1***	Intron	Chr6(GRCh37):g.146511626C>T	NM_000838.3:c.950+30893C>T	-0.36	no	
8	***HDAC4***	Intron	Chr2(GRCh37):g.239978216T>C	NM_006037.3:c.2989-1687A>G	-0.92	no	
9	***LRP2***	Splice site	Chr2(GRCh37):g.170070389T>C	NM_004525.2:c.5827-9A>G	-0.02	no	
10	***LARGE***	5'UTR	Chr22(GRCh37):g.34159094C>A	NM_004737.4:c.-82-1549G>T	-0.42	no	
12	***MAP2K1***	Intron	Chr15(GRCh37):g.66775036A>G	NM_002755.3:c.693+819A>G	-0.20	no	
13	***STXBP1***	Intron	Chr9(GRCh37):g.130417532G>A	NM_003165.3:c.169+1457G>A	-0.40	no	
13	***AUH***	Intron	Chr9(GRCh37):g.94078475C>T	NM_001698.2:c.505+9125G>A	1.57	no	
14	***ACVR1***	Intron	Chr2(GRCh37):g.158654234G>A	NM_001105.4:c.67+1705C>T	-0.60	no	
16	***PNKP***	Intron	Chr19(GRCh37):g.50367525C>T	NM_007254.3:c.579-32G>A	-0.70	possibly	
17	***MCPH1***	Intron	Chr8(GRCh37):g.6438621C>A	NM_024596.3:c.2215-40354C>A	-0.44	no	
17	***GRM1***	Intron	Chr6(GRCh37):g.146468131A>G	NM_000838.3:c.701-12353A>G	0.04	no	
18	***ALG12***	Intron	Chr22(GRCh37):g.50298304C>T	NM_024105.3:c.993-150G>A	-2.86	no	
18	***GRIN2A***	Intron	Chr16(GRCh37):g.10168818T>C	NM_000833.3:c.414+105037A>G	0.04	no	
19	***VPS13B***	Intron	Chr8(GRCh37):g.100785479A>G	NM_017890.4:c.7323-3524A>G	0.37	no	
20	***FOXP1***	Intron	Chr3(GRCh37):g.71222718G>A	NM_001244810.1:c.180+24635C>T	0.12	no	
22	***NRXN1***	Intron	Chr2(GRCh37):g.51178159T>C	NM_001135659.1:c.872-25066A>G	0.12	no	
22	***NRXN1***	Intron	Chr2(GRCh37):g.51141214T>C	NM_001135659.1:c.931+7793A>G	-1.51	no	
22	***PVRL1***	Intron	Chr11(GRCh37):g.119583361G>T	NM_002855.4:c.79+15824C>A	-1.07	no	
24	***SPRED1***	Intron	Chr15(GRCh37):g.38603586G>A	NM_152594.2:c.208-10856G>A	-0.44	no	
25	***ALDH18A1***	5'UTR	Chr10(GRCh37):g.97416435_97416438del	NM_002860.3:c.-113_-110del	-	no	
27	***MMADHC***	Promoter	Chr2(GRCh37):g.150489713A>G	-	0.24	-	Weak Promoter; Heterochrom/lo;

							Repetitive/CNV; WeakEnhancer
28	***BBS9***	Intron	Chr7(GRCh37):g.33191509G>A	NM_198428.2:c.113-804G>A	-0.44	no	
28	***DPYD***	Intron	Chr1(GRCh37):g.97588236C>T	NM_000110.3:c.2623-24048G>A	0.93	no	
33	***GNPAT***	Intron	Chr1(GRCh37):g.231397861A>G	NM_014236.3:c.439-608A>G	-1.17	no	
33	***SMARCA4***	Intron	Chr19(GRCh37):g.11086322G>A	NM_001128849.1:c.-31-8475G>A	-2.30	no	
36	***CCBE1***	Intron	Chr18(GRCh37):g.57124433G>T	NM_133459.3:c.554-2250C>A	-0.60	no	
36	***NF1***	Intron	Chr17(GRCh37):g.29533557T>C	NM_001042492.2:c.1392+168T>C	0.36	no	
38	***SETBP1***	Intron	Chr18(GRCh37):g.42569791G>A	NM_015559.2:c.4000+36486G>A	-0.04	no	
38	***SCN8A***	Intron	Chr12(GRCh37):g.52189411C>T	NM_014191.3:c.4795+986C>T	-0.76	no	
41	***NRXN1***	Intron	Chr2(GRCh37):g.50558625A>G	NM_001135659.1:c.3485-94517T>C	-0.28	no	
41	***MEF2C***	Intron	Chr5(GRCh37):g.88123620C>T	NM_001193347.1:c.-142-3873G>A	-1.89	no	
43	***BBS9***	Intron	Chr7(GRCh37):g.33549086C>T	NM_198428.2:c.2298+3829C>T	1.66	possibly	
43	***PEPD***	Intron	Chr19(GRCh37):g.33980994G>A	NM_000285.3:c.442-31C>T	-0.36	no	
45	***DLG3***	Intron	ChrX(GRCh37):g.69718156G>A	NM_021120.3:c.1820-214G>A	-0.12	no	
48	***NRXN1***	Intron	Chr2(GRCh37):g.51078981C>T	NM_001135659.1:c.931+70026G>A	-1.10	no	
49	***CEP290***	Intron	Chr12(GRCh37):g.88447045C>T	NM_025114.3:c.7129+384G>A	0.21	no	
49	***ARID1B***	Intron	Chr6(GRCh37):g.157276718A>G	NM_020732.3:c.2037+20008A>G	1.50	no	
50	***CCBE1***	3'UTR	Chr18(GRCh37):g.57102290C>T	NM_133459.3:c.*850G>A	0.26	miRNA binding?	
50	***PLCB1***	Intron	Chr20(GRCh37):g.8327174C>T	NM_015192.2:c.178-24855C>T	0.22	no	
50	***SHROOM4***	Intron	ChrX(GRCh37):g.50462314T>C	NM_020717.3:c.118-23377A>G	0.53	no	

Confirmed non-coding de novo mutations. Effect on splicing was determined using Alamut software that integrates a number of prediction methods for splice signal detection as well as exonic splicing enhancer (ESE) binding site detection. Encode annotation was based on Chromatin state segments of nine human cell types (Broad ChromHMM) and transcription factor binding sites (Txn Factor ChIP).[41]

Supplementary Table 14: Phenotypic comparison of patients with mutations in known and candidate ID genes

Trio	Phenotype	*Mutation type*	*Gene*	Disease association	References
5	Severe ID, signs of autism, aggressive behavior, 2 café au lait spots, increased pain threshold.	*CNV*	*SHANK3* [OMIM 606232]	Phelan-McDermid syndrome is characterized by neonatal hypotonia, global developmental delay, absent to severely delayed speech, and normal to accelerated growth. Behavior characteristics include mouthing or chewing non-food items, decreased perception of pain, and autistic-like affect. Phelan-McDermid syndrome can be caused by a heterozygous contiguous gene deletion at chromosome 22q13 or by mutation in the *SHANK3* gene.	[80,81]
7	Developmental delay. No major dysmorphisms.	*SNV*	*TBR1* [OMIM 604616]	TBR1, a transcription factor involved in early cortical development, is a strong candidate for the intellectual disability phenotype seen in patients with 2q24.3 deletions. *De novo* mutations in *TBR1* have also been reported in individuals with autism.	[49,82]
9	ID from infancy. Regression at adult age. She showed parkinsonism and dystonia.	*SNV*	*WDR45* [OMIM 300526]	Neurodegeneration with brain iron accumulation (NBIA) is a group of genetic disorders characterized by abnormal iron deposition in the basal ganglia. *De novo* mutations in *WDR45*, encoding a beta-propeller scaffold protein with a putative role in autophagy, cause a distinctive NBIA phenotype. The clinical features include early-onset global developmental delay and further neurological deterioration (parkinsonism, dystonia, and dementia developing by early adulthood).	[83,84]

12	ID without speech, autism. Hypotonia. Recurrent infection. Sleep disturbances, delayed puberty and obesitas. Short stature and microcephaly. Deep set eyes, hypertelorism, large ears, large nose, short philtrum, full lips. Kyphosis, narrow hands with tapering fingers, partial cutaneous syndactyly of 2nd and 3rd toes and sandal gaps. No ophthalmologic anomalies. Of note, Cohen syndrome was part of the differential diagnosis. *VPS13B* was sequenced but with normal results.	*CNV*	*VPS13B* [OMIM 216550]	Cohen syndrome is an autosomal recessive disorder characterized by a characteristic facial appearance, failure to thrive, microcephaly, progressive retinochoroidal dystrophy and high myopia, neutropenia with recurrent infections, a cheerful disposition, joint hypermobility, truncal obesity in puberty, early-onset hypotonia and moderate-profound ID.	[85]
13	Severe ID with epilepsy. No speech. Short stature, microcephaly and enlarged ventricles and hypertrophy of the cerebellar vermis. Hypotonia, cleft palate, cataract, scoliosis and stereotypic movements. Dysmorphic features included a low frontal-temporal hairline, deep-set eyes, a flat midface, long narrow ears, a long nose with a high bridge, a short philtrum, an everted lower lip and a prominent jaw. In addition, she had scoliosis, short narrow hands, tapered fingers and several contractures.	*SNV*	*SMC1A* [OMIM 610759]	Cornelia de Lange syndrome CdLS is a clinically and genetically heterogeneous developmental disorder. Clinical features include growth retardation, intellectual disability, limb defects, arched eyebrows, synophrys thin lips and other systemic involvement. Mutations in *SMC1A* are associated with a milder phenotype with moderate neurocognitive impairment and a paucity of major structural defects. Recently, one loss-of-function mutation was described in a female patient with developmental delay, microcephaly, congenital diaphragmatic hernia, and generalized epilepsy. Dysmorphic features included round face with arched eyebrows, short nose, smoothed philtrum, prominent jaw, clinodactyly of the 5^{th} fingers and camptodactyly of the 5^{th} finger on the right.	[78,86-88]
15	Neonatal feeding problems, hip dysplasia. ID (IQ 50) and autism. Epicanthic folds, straight eyebrows, posteriorly rotated ears and downturned corners of the mouth. MRI of brain is not available. No epilepsy.	*SNV*	*SPTAN1* [OMIM 613417]	Epileptic encephalopathy, early infantile, 5 (EIEE5). *SPTAN1* mutations have been reported in patients with severe ID, infantile spasms with hypsarrhythmia, hypomyelnation and atrophy of various regions of the brain, including the cerebellum and brainstem. Mutations have also been reported in a patients with non-syndromic ID.	[89]

26	ID (dyshormonic profile; performal IQ 50, verbal 90), hypotonia, no facial dysmorphisms, progressive scoliosis, hip dysplasia, fatigue problems.	*SNV*	*PPP2R5D* [OMIM 601646]	Recently, we published a patient with a de novo mutation in PPP2R5D was identified. The patient presented with ID, absence of speech, no facial dysmorphisms, short stature, microcephaly, cataract, anxiety, autistic features and mood swings. The product of this gene belongs to the phosphatase 2A regulatory subunit B family. Protein phosphatase 2A is one of the four major Ser/Thr phosphatases, and it is implicated in the negative control of cell growth and division.	[32]
27	Severe ID. Epilepsy (tonic-clonic seizures), myokymia. Normal height, head circumference -2 SD. Broad based walking pattern. Narrow long face, prominent supraorbital ridges, long nose and philtrum, thin upper lip, hypodontia, long fingers and toes and narrow thorax.	*SNV*	*KCNA1* [OMIM 160120]	Mutations in *KCNA1* have been reported in patients with episodic ataxia and myokymia, without ID. In one report a family with persistent cerebellar dysfunction, cerebellar atrophy, and cognitive delay has been described. All affected family members had myokymia and epilepsy. Additional features included postural abnormalities, episodic stiffness and weakness.	[94]
28	ID. No major dysmorphisms. No epilepsy. Sleep disturbances and obsessive behaviour.	*SNV*	*SCN2A* [OMIM 182390]	*SCN2A* mutations have been reported in patients with early-onset infantile epileptic encephalopathy and also in patients with severe ID and auto-aggressive behavior but without epilepsy.	[32,45]
31	Severe ID, microcephaly, epilepsy, progressive spasticity, small hands and feet, poor vision.	*CNV*	*TENM3 [OMIM 610083]* *IQSEC2 [OMIM 309530]*	*IQSEC2* is a non-syndromic X-linked ID gene in which carrier females show normal or mild learning difficulties. Male patients, show phenotypic overlap with pathogenic mutations in *MECP2, FOXG1, CDKL5* and *MEF2C*, but may also manifest seizures, autistic-like behavior, psychiatric problems and delayed language skills. In most male patients the mutation occurred *de* novo (rather than maternally inherited). One other female patient with severe infantile spasms, profound global developmental arrest, hypsarrhythmia and severe ID, is reported to have a balanced translocation that disrupts *IQSEC2.*	[45,95-97]
40	Moderate ID, features of autism, anxiety, mood instability and aggressive outbursts. No evident dysmorphic features.	*CNV*	*STAG1* [OMIM 604358]	A *de novo* loss of function mutation was previously reported as a possible cause for ID in a patient with normal growth and head circumference.	[45]

43	Moderate ID,left hydronephrosis , right dysplastic kidney, severe hypermetropia, Horner syndrome, microcephaly .	*SNV*	*POGZ* [OMIM 614787]	Mutations in *POGZ* have been identified in patients with autism.	47,48
46	Severe hypotonia, severe motor delay and speech delay. Deficient gyration frontal region, mild periventricular white matter abnormalities	*SNV*	*TBR1* [OMIM 604616]	*TBR1*, a transcription factor involved in early cortical development, is a strong candidate for the intellectual disability phenotype seen in patients with 2q24.3 deletions. *De novo* mutations in *TBR1* have also been reported in individuals with autism.	49,82
48	Severe ID, severe epilepsy, Rett-like phenotype, motor handicap, severe feeding difficulties, progressive scoliosis	*CNV*	*SMC1A* [OMIM 610759]	Cornelia de Lange syndrome CdLS is a clinically and genetically heterogeneous developmental disorder. Clinical features include growth retardation, intellectual disability, limb defects, arched eyebrows, synophrys thin lips and other systemic involvement. Mutations in *SMC1A* are associated with a milder phenotype with moderate neurocognitive impairment and a paucity of major structural defects. Recently, one loss-of-function mutation was described in a female patient with developmental delay, microcephaly, congenital diaphragmatic hernia, and generalized epilepsy. Dysmorphic features included round face with arched eyebrows, short nose, smoothed philtrum, prominent jaw, clinodactyly of the 5^{th} fingers and camptodactyly of the 5^{th} finger on the right.	78,86-88
50	Profound ID, microcephaly, short stature, vaginal atresia, unilateral renal agenesis, cortical blindness, evident dysmorphic features, small hands and feet	*CNV*	Proximal 16p11.2 deletion [OMIM 611913]	Patients with the typical 600 kb 16p11 deletion can have borderline-mild ID, psychiatric disorders (80%), obesity (50%) and seizures (25%). This locus is known for its reduced penetrance (estimated at to be 47%), with the majority of CNVs occurring *de novo* (70% of cases). Eight percent of patients show additional variants expected to contribute to the phenotype. Such variant may also be expected in this patient.	98-100

Supplementary Table 15: Molecular diagnosis per patient after WGS

Trio	Molecular diagnosis?*	Based on [*gene*] *(inheritance)*	Mutation type
1	Possible	*NGFR (de novo)*	Autosomal SNV
2	Possible	*GFPT2 (de novo)*	Autosomal SNV
3	NO		
4	NO		
5	Highly likely	*SHANK3 (de novo)*	Autosomal CNV
6	Possible	*WWP2 (de novo)*	Autosomal SNV
7	Highly Likely	*TBR1 (de novo)*	Autosomal SNV
8	NO		
9	Highly likely	*WDR45 (de novo)*	X-linked SNV
10	NO		
11	NO		
12	Highly likely	*VPS13B (inherited)*	Autsomal CNV
13	Highly likely	*SMC1A (de novo)*	X-linked SNV
14	NO		
15	Highly likely	*SPTAN (de novo)*	Autosomal SNV
16	NO		
17	Possible	*ASUN (de novo)*	Autosomal SNV
18	Highly likely	*MECP2 (de novo)*	X-linked CNV
19	NO		
20	NO		
21	Highly likely	*RAI1 (de novo)*	Autosomal SNV
22	Highly likely	*MED13L (de novo)*	Autosomal SNV
23	NO		
24	Possible	*BRD3 (de novo)*	Autosomal SNV
25	Highly likely	*SATB2 (de novo)*	Autosomal SNV
26	Highly likely	*PPP2R5D (de novo)*	Autosomal SNV
27	Highly likely	*KCNA1 (de novo)*	Autosomal SNV
28	Highly likely	*SCN2A (de novo)*	Autosomal SNV
29	NO		
30	Possible	*MAST1 (de novo)*	Autosomal SNV
31	Highly likely	*IQSEC2 (de novo)*	Autosomal CNV/ X-linked
32	NO		
33	NO		
34	Possible	*APPL2 (de novo)*	Autosomal SNV
35	NO		
36	NO		
37	Highly likely	multiple (*de novo*)*	Autosomal CNV
38	NO		
39	NO		
40	Highly likely	*STAG1 (de novo)*	Autosomal CNV
41	Possible	*NACC1 (de novo)*	Autosomal SNV

42	NO		
43	Highly likely	*POGZ (de novo)*	Autosomal SNV
44	NO		
45	NO		
46	Highly likely	*TBR1 (de novo)*	Autosomal SNV
47	NO		
48	Highly likely	*SMC1A (de novo)*	X-linked CNV
49	Highly likely	*KANSL2 (de novo)*	Autosomal SNV
50	Highly likely	16p11.2 microdeletion syndrome *(de novo)*	Autosomal CNV

**Of note, a possible diagnosis was made in 8 patients who showed severe de novo SNVs in genes with functional support for ID-related disorders (Extended Data Table 2). Interestingly these 8 genes showed a significantly lower tolerance for normal variation similar to our set of known and candidate ID genes (P=0.003, Extended Data Figure 1). Nonetheless, conclusive diagnoses for these 8 patients cannot be made based on single observations and are thus not included in the calculated diagnostic yields. **48 genes are affected by the rearrangement – phenotype may be the effect of a single dosis sensitive gene in this CNV or it may reflect a contiguous gene syndrome, with multiple genes contributing.*