Supplementary Material

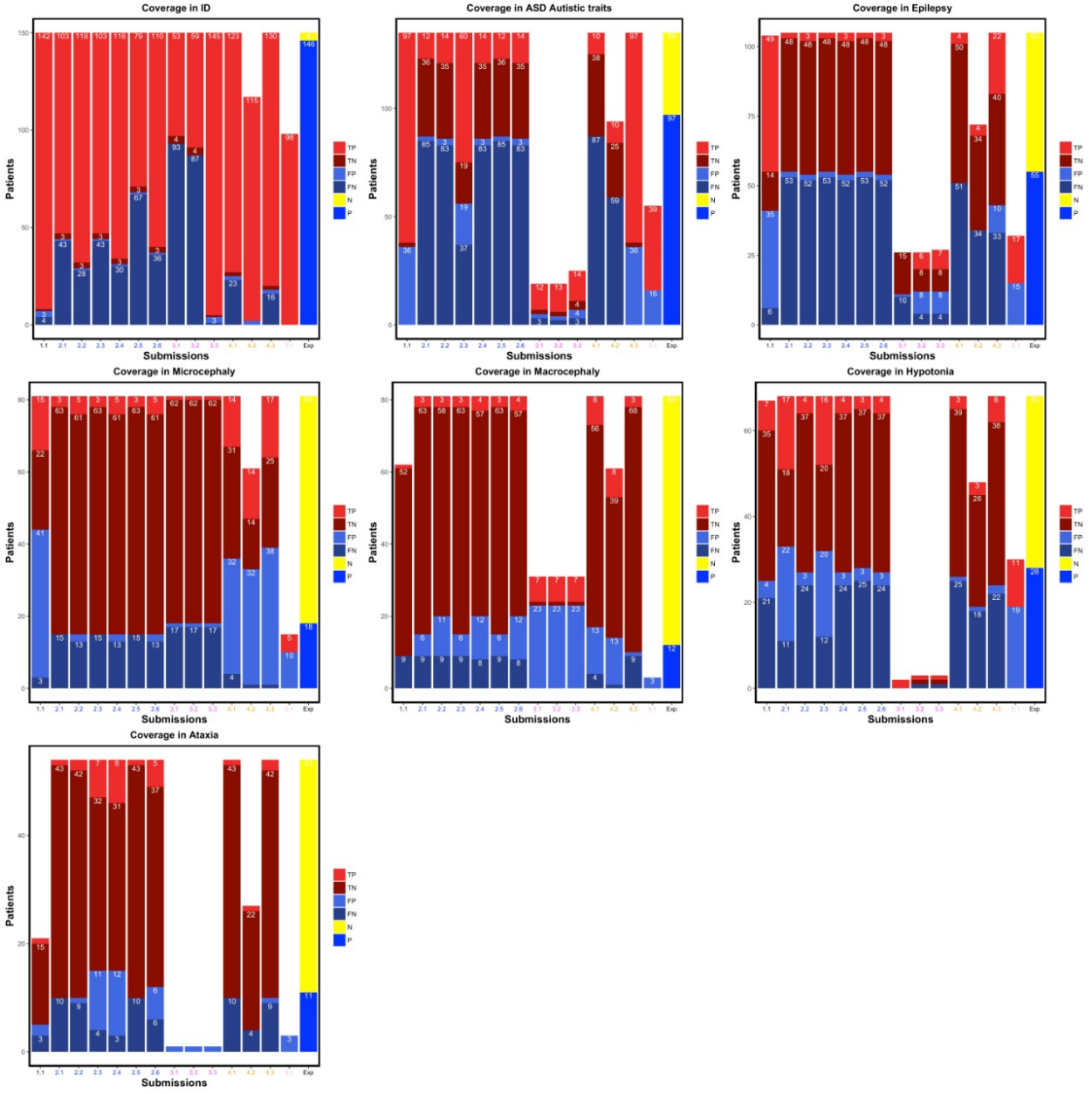
Assessment of patient clinical descriptions and pathogenic variants from gene panel sequences in the CAGI-5 intellectual disability challenge

Marco Carraro[1,], Alexander Miguel Monzon[1,*], Luigi Chiricosta[1], Francesco Reggiani[1,2], Maria Cristina Aspromonte[3], Mariagrazia Bellini[3, 4], Kymberleigh Pagel[5], Yuxiang Jiang[5], Predrag Radivojac[5], Kunal Kundu [6,7], Lipika R. Pal[6], Yizhou Yin[6, 7], Ivan Limongelli[8],Gaia Andreoletti[6, 9], John Moult[6, 9], Stephen J. Wilson[10], Panagiotis Katsonis[10], Olivier Lichtarge[10], Jingqi Chen[11], Yaqiong Wang[11], Zhiqiang Hu[11], Steven E. Brenner[11], Carlo Ferrari[2], Alessandra Murgia[3, 4], Silvio C.E. Tosatto[1,12,*], Emanuela Leonardi[3, 4,*]*

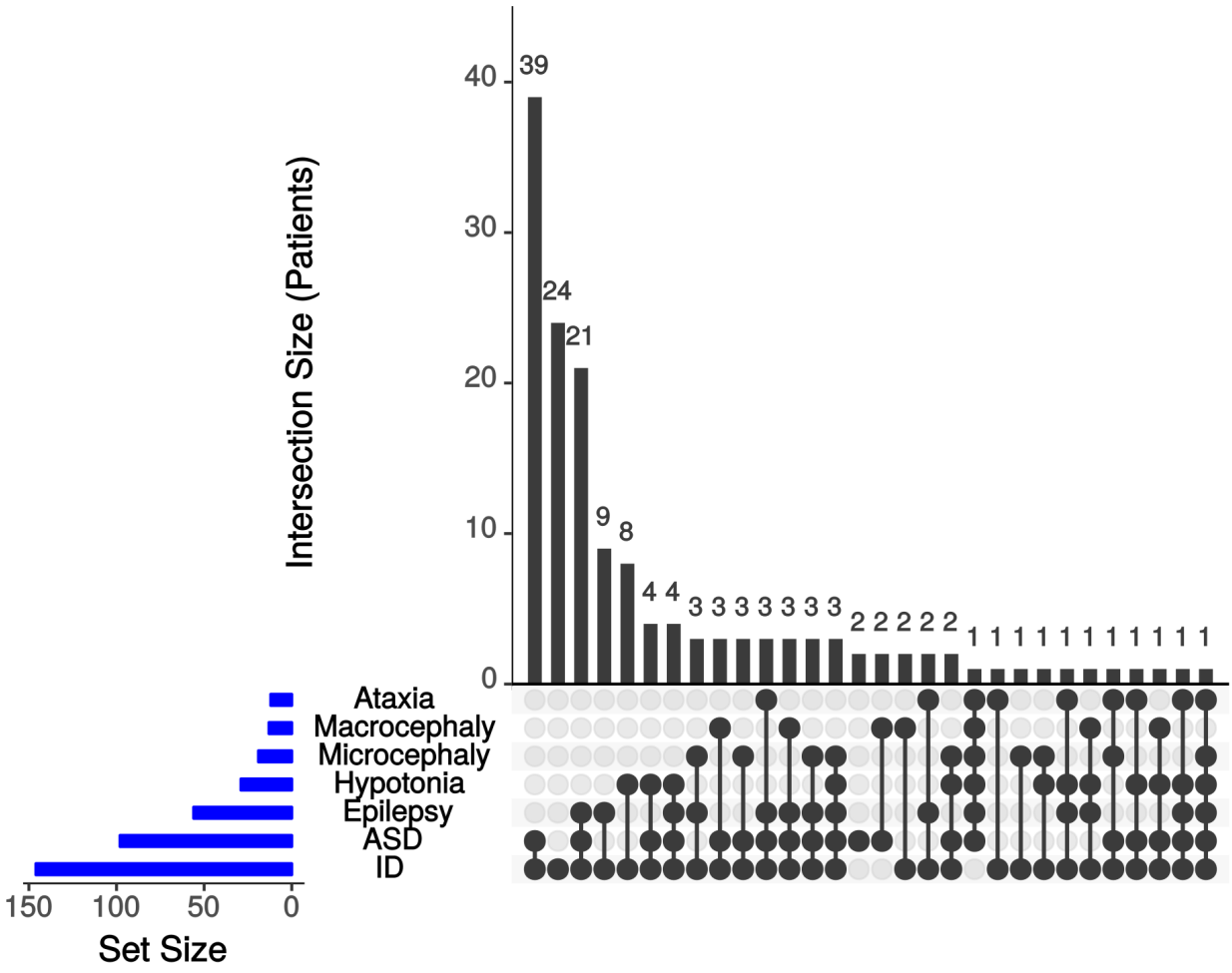
[1] *Department of Biomedical Sciences, University of Padua, Padua, Italy*
[2] *Department of Information Engineering, University of Padua, Padua, Italy*
[3] *Department of Woman and Child Health, University of Padua, Padua, Italy*
[4] *Fondazione Istituto di Ricerca Pediatrica (IRP), Città della Speranza, Padova, Italy*
[5] *Khoury College of Computer and Information Sciences, Northeastern University, 440, Huntington Avenue, Boston, MA 02115, USA*
[6] *Institute for Bioscience and Biotechnology Research, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850, USA*
[7] *Computational Biology, Bioinformatics and Genomics, Biological Sciences Graduate Program, University of Maryland, College Park, MD 20742, USA*
[8] *enGenome srl, via Ferrata 5, Pavia, Italy.*
[9] *Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA*
[10] *Baylor College of Medicine, Department of Molecular and Human Genetics, Houston, TX 77030, USA*
[11] *Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA.*
[12] *CNR Institute of Neuroscience, Padua, Italy*
[*] *Contributed equally*

Corresponding authors: Silvio Tosatto <silvio.tosatto@unipd.it>,
Emanuela Leonardi <emanuela.leonardi@unipd.it>

Supplementary Figures



Suppl. Figure S1. Patient submission coverage by phenotype. Each bar represents the number of patients for whom a probability value was submitted by the predictor. Each color represents true positive (TP), true negative (TN), false positive (FP) and false negative (FN) predicted patient phenotypes. In addition, blue (P) and yellow (N) bars indicate the experimentally identified positive and negative cases, respectively.



Suppl. Figure S2: Matrix layout for all intersections of seven phenotypes, sorted by size. Dark circles in the matrix indicate phenotypes that are part of the intersection. Gray bars indicate the amount of patients in each intersection and blue bars are the amount of patients with the phenotype positively identified in the whole data set.

Supplementary Tables

Phenotypic traits	Description
Intellectual disability (ID)	Intellectual disability is characterized by significant limitations in both intellectual functioning and in adaptive behavior, which covers many everyday social and practical skills. This disability originates during development. Intellectual functioning (also called intelligence) refers to general mental capacity, such as learning, reasoning, problem solving, and so on. Adaptive behavior is the collection of conceptual, social, and practical skills that are learned and performed by people in their everyday lives.
Autism spectrum disorder (ASD)	This term can be used to refer to autism spectrum disorder as a phenotypic feature that can be a component of a disease. A disorder beginning in childhood, it is marked by the presence of markedly abnormal or impaired development in social interaction and communication and a markedly restricted repertoire of activity and interest. Manifestations of the disorder vary greatly depending on the developmental level and chronological age of the individual (DSM-IV). Autism spectrum disorders range from a severe form, called autistic disorder, to a milder form, Asperger syndrome.
Epilepsy	Seizures are an intermittent abnormality of the central nervous system due to a sudden, excessive, disorderly discharge of cerebral neurons characterized clinically by some combination of disturbance of sensation, loss of consciousness, impairment of psychic function, or convulsive movements. The term epilepsy is used to describe chronic, recurrent seizures.
Microcephaly	Occipitofrontal (head) circumference less than -3 standard deviations compared to appropriate, age matched, normal standards (Potter 1978). Alternatively, decreased size of the cranium.
Macrocephaly	Occipitofrontal (head) circumference greater than the 97th centile compared to appropriate, age matched, sex-matched normal standards. Alternatively, an apparently increased size of the cranium.
Hypotonia	Muscular hypotonia (abnormally low muscle tone) manifesting in infancy.
Ataxic gait	A type of ataxia characterized by impairment of the ability to coordinate movements required for normal walking. Gait ataxia is characterized by a wide-based staggering gait with a tendency to fall.

Suppl. Table S1. Summary of the seven diseases classes in the CAGI-5 intellectual disability challenge.

CAGI-ID	Name	Sex	Class of variant	Chr:POS:REF:ALT	Exonic Function	AAChange	Mode of inheritance	Inheritance	Number of Groups	Number of predictions	ID of Groups	Predictions ID
135	MR2409	F	Contributing factor	chr22:51117766:T:G	nonsynonymous	SHANK3:NM_033517:exon7:c.T795G:p.H265Q	AD	paternal	4	11	1, 2, 4, 5	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.1, 4.2, 4.3, 5.1
97	MR2250	M	Contributing factor	chr14:21863113:G:C	nonsynonymous	CHD8:NM_001170629:exon29:c.C5348G:p.A1783G	AD	paternal	3	9	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.1, 4.3, 5.1
125	MR2368	M	Contributing factor	chr3:11078549:G:A	nonsynonymous	SLC6A1:NM_003042:exon16:c.G1697A:p.R566H	AD	paternal	3	9	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.1, 4.3, 5.1
83	MR2189	M	Contributing factor	chr17:84702701:C:G	nonsynonymous	MYH10:NM_001256012:exon30:c.G3838C:p.E1280Q	AD	paternal	3	8	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.3, 5.1
17	MR1635	M	Contributing factor	chr11:70653140:C:T	unknown	SHANK2:NM_012309.3:c.G1484A:p.E544K	AD	materna	3	8	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.3, 5.1
112	MR2337	F	Contributing factor	chr11:70653124:C:T	nonsynonymous	SHANK2:NM_012309.4:c.G1646A:p.R549H	AD	girl adopted	2	8	2, 4	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.1, 4.3
6	MR1289	M	Contributing factor	chr14:21860898:C:T	nonsynonymous	CHD8:NM_001170629:exon33:c.G6539A:p.R2180H	AD	maternal	1	6	2	2.1, 2.2, 2.3, 2.4, 2.5, 2.6
47	MR2033	M	Contributing factor	chr7:146829601:G:A	nonsynonymous	CNTNAP2:NM_014141:exon8:c.G1348A:p.G450S	AR, AD	maternal	3	6	1, 3, 4	1.1, 3.1, 3.2, 3.3, 4.1, 4.3

34	MR1980	F	Contributing factor	chr7:103243828:C:A	nonsynonymous	RELN:NM_005045:exon24:c.G3256T:p.V1086F	AR, AD	paternal, paternal grandmother	1	6	2	2.1, 2.2, 2.3, 2.4, 2.5, 2.6
94	MR2241	M	Contributing factor	chr7:103130201:C:T	nonsynonymous	RELN:NM_005045:exon60:c.G9751A:p.E3251K	AR, AD	paternal	1	6	2	2.1, 2.2, 2.3, 2.4, 2.5, 2.6
127	MR2375	F	Contributing factor	chr11:70319321:G:A	nonsynonymous	SHANK2:NM_133266:exon11:c.C3439T:p.P1147S	AD	n.d.	3	5	3, 4, 5	3.1, 3.2, 3.3, 4.3, 5.1
12	MR1543	M	Contributing factor	chr7:148112649:A:C	nonsynonymous	CNTNAP2:NM_014141:exon24:c.A3937C:p.N1313H	AR, AD	maternal	2	4	3, 4	3.1, 3.2, 3.3, 4.3
17	MR1635	M	Contributing factor	chr5:14471497:C:T	nonsynonymous	TRIO:NM_007118:exon38:c.C5834T:p.S1945L	AD	maternal	1	1	1	1,1
105	MR2276	M	Disease causing	chrX:76909661:T:C	nonsynonymous	ATRX:NM_138270:exon13:c.A4130G:p.N1377S	XLD, XLR	maternal; X-inactivation: balanced	5	14	1, 2, 3, 4, 5	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1, 3.2, 3.3, 4.1, 4.2, 4.3, 5.1
140	MR414	F	Disease causing	chrX:153296777:G:A	stop gain	MECP2:NM_001110792:exon3:c.C538T:p.R180X	XLD	n.d.	4	12	2, 3, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.2, 3.3, 4.1, 4.2, 4.3, 5.1
79	MR2166	M	Disease causing	chr9:140728837:G:C	nonsynonymous	EHMT1:NM_024757:exon26:c.G3577C:p.G1193R	AD	de novo	4	11	1, 2, 3, 5	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1, 3.2, 3.3, 5.1
142	MR602	F	Disease causing	chrX:41401980:G:A	stop gain	CASK:NM_003688:exon22:c.C2119T:p.Q707X	XL	de novo	3	10	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.1, 4.2, 4.3, 5.1
64	MR2113	M	Disease causing	chr12:116445337:C:T	nonsynonymous	MED13L:NM_015335:exon11:c.G2117A:p.G706E	AD	De novo	3	10	1, 2, 3	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1, 3.2, 3.3
90	MR2233	M	Disease causing	chr6:33411228:C:T	stop gain	SYNGAP1:NM_006772:exon15:c.C2899T:p.R967X	AD	de novo	4	10	2, 3, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.2, 3.3, 4.3, 5.1
69	MR2	F	Disease causing	chr6:157528165:G:T	stop gain	ARID1B:NM_001346813:exon2	AD	de novo	3	9	2, 3, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.2, 3.3, 5.1

	127					0:c.G6010T:p.E2004X						
72	MR2140	M	Disease causing	chrX:122460015:G:A	nonsynonymous	GRIA3:NM_000828:exon4:c.G647A:p.R216Q	XL	maternal, X-inactivation: mutated allele 30%	4	9	1, 2, 4, 5	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.3, 5.1
35	MR1985	M	Disease causing	chr2:200213882:G:A	stop gain	SATB2:NM_001172509:exon7:c.C715T:p.R239X	AD	de novo	3	9	2, 3, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.2, 3.3, 5.1
104	MR2274	M	Disease causing	chr18:42531498:AAGAGC:A	frameshift deletion	SETBP1:NM_015559:exon4:c.2194_2198del:p.R732fs	AD	de novo	3	9	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.1, 4.3, 5.1
23	MR1749	M	Disease causing	chr22:51160432:GA:G	frameshift deletion	SHANK3:NM_033517:exon21:c.4130delA:p.E1377fs	AD	de novo	3	9	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.1, 4.2, 5.1
106	MR2278	M	Disease causing	chr12:13761626:T:G	nonsynonymous	GRIN2B:NM_000834:exon9:c.A1921C:p.I641L	AD	de novo	3	8	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.3, 5.1
89	MR2230	F	Disease causing	chr22:51153476:G:A	splicing	SHANK3:NM_033517:exon19:c.2223+1G>A	AD	de novo	3	8	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.3, 5.1
150	MR984	M	Disease causing	chr5:14390392:C:T	nonsynonymous	TRIO:NM_007118:exon26:c.C4111T:p.H1371Y	AD	de novo	3	8	1, 2, 5	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 5.1
96	MR2243	M	Disease causing	chr9:140657209:GA:G	frameshift deletion	EHMT1:NM_024757:exon10:c.1585delA:p.S529fs	AD	de novo	2	7	2, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 5.1
41	MR2019	M	Disease causing	chr12:13724822:C:T	nonsynonymous	GRIN2B:NM_000834:exon10:c.G2087A:p.R696H	AD	de novo	2	7	2, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 5.1
21	MR1730	F	Disease causing	chrX:67273488:C:T	nonsynonymous	OPHN1:NM_002547:exon22:c.G2323A:p.V775M	XLR	De novo	4	6	1, 3, 4, 5	1.1, 3.1, 3.2, 3.3, 4.3, 5.1

32	MR1974	M	Disease causing	chrX:154490151:A:C	nonsynonymous	RAB39B:NM_171998:exon2:c.T579G:p.F193L	XLR	maternal, mother affected	3	3	1, 4, 5	1.1, 4.3, 5.1
64	MR2113	M	Disease causing	chr1:155449342:T:C	nonsynonymous	ASH1L:NM_018489:exon3:c.A3319G:p.I1107V	AD	de novo	1	2	4	4.1, 4.3
87	MR2222	M	Disease causing	chr21:38858777:G:C	nonsynonymous	DYRK1A:NM_101395:exon7:c.G525C:p.K175N	AD	de novo	2	2	1, 5	1.1, 5.1
113	MR2338	F	Disease causing	chr16:89346136:CAG:C	frameshift deletion	ANKRD11:NM_013275:exon9:c.6812_6813del:p.P2271fs	AD	n.d.	1	1	5	5,1
47	MR2033	M	Disease causing	chr16:89345974:CCTTCGGGG:C	frameshift deletion	ANKRD11:NM_013275:exon9:c.6968_6975del:p.A2323fs	AD	de novo	1	1	5	5,1
102	MR2271	M	Disease causing	chr22:51159718:C:T	stop gain	SHANK3:NM_033517:exon21:c.C3415T:p.R1139X	AD	de novo	1	1	5	5,1
78	MR2165	M	Disease causing	chr5:14394159:C:T	stop gain	TRIO:NM_007118:exon28:c.C4231T:p.R1411X	AD	maternal, mother affected	1	1	5	5,1
31	MR1970	F	Disease causing	chr22:51159830:A:TTC	frameshift delins	SHANK3:NM_033517:exon21:c.3527delinsTTC:p.D1176fs	AD	de novo	0	0		
48	MR2039	M	Putative	chr16:37885561:C:T	nonsynonymous	CREBBP:NM_004380:exon26:c.G4393A:p.G1465R	AD	n.d.	5	14	1, 2, 3, 4, 5	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1, 3.2, 3.3, 4.1, 4.2, 4.3, 5.1
109	MR2322	M	Putative	chrX:76764055:T:A	nonsynonymous	ATRX:NM_000489:exon35:c.A7253T:p.Y2418F	XLD, XLR	n.d.	4	13	2, 3, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1, 3.2, 3.3, 4.1, 4.2, 4.3, 5.1
103	MR2272	M	Putative	chr10:89690828:G:A	nonsynonymous	PTEN:NM_000314:exon4:c.G235A:p.A79T	AD	maternal	3	12	2, 3, 4	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1, 3.2, 3.3, 4.1, 4.2, 4.3

24	MR1769	M	Putative	chr3:71026867:A:C	nonsynonymous	FOXP1:NM_032682:exon16:c.T1355G:p.I452S	AD	paternal	3	8	1, 2, 5	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 5.1
24	MR1769	M	Putative	chr7:146829502:G:T	nonsynonymous	CNTNAP2:NM_014141:exon8:c.G1249T:p.D417Y	AR, AD	maternal	1	3	3	3.1, 3.2, 3.3
114	MR2340	M	Putative	chr2:166165900:C:T	nonsynonymous	SCN2A:NM_021007:exon6:c.C644T:p.A215V	AD	maternal, familial epilepsy	3	8	1, 2, 4	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.3
40	MR2007	M	Putative	chr11:70644598:G:A	nonsynonymous	SHANK2:NM_012309:exon13:c.C1727T:p.P576Q	AD	n.d.	3	8	2, 4, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.3, 5.1
30	MR196	M	Putative	chrX:41448842:A:G	nonsynonymous	CASK:NM_003688:exon13:c.T1159C:p.Y387H	XL	n.d.	3	5	1, 4, 5	1.1, 4.1, 4.2, 4.3, 5.1
99	MR2264	M	Putative	chr16:89349967:T:C	nonsynonymous	ANKRD11:NM_013275:exon9:c.A2983G:p.K995E	AD	n.d.	2	7	1, 2	1.1, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6
73	MR2141	M	Putative	chr14:21876977:G:A	nonsynonymous	CHD8:NM_001170629:exon11:c.C2372T:p.P791L	AD	Not in the mother	2	7	2, 4	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 4.3
130	MR2389	F	Putative	chr14:21882498:T:C	nonsynonymous	CHD8:NM_001170629:exon8:c.A2104G:p.K702E	AD	n.d.	2	7	2, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 5.1
127	MR2375	F	Putative	chr11:684897:C:T	splicing	DEAF1:NM_021008:exon6:c.870+1G>A	AR, AD	n.d.	2	7	2, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 5.1
5	MR1192	F	Putative	chr12:13720096:C:G	nonsynonymous	GRIN2B:NM_000834:exon12:c.G2461C:p.V821L	AD	n.d.	2	7	2, 5	2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 5.1
33	MR1975	M	Putative	chr18:44595922:C:T	nonsynonymous	KATNAL2:NM_031303:exon10:c.C743T:p.A248V	AD	maternal, gene with low penetrance	1	6	2	2.1, 2.2, 2.3, 2.4, 2.5, 2.6
126	MR2374	F	Putative	chr7:148080864:C:T	nonsynonymous	CNTNAP2:NM_014141:exon22:c.C3599T:p.S1200L	AR, AD	n.d.	2	4	3, 4	3.1, 3.2, 3.3, 4.3

					SNV							
116	MR2344	M	Putative	chrX:53964467:A:G	nonsynonymous SNV	PHF8:NM_001184897:exon22:c.T2794C:p.C932R	XLR	maternal, X-inactivation: mutated allele 70%	1	3	4	4.1, 4.2, 4.3
56	MR2053	F	Putative	chr2:171702114:C:T	nonsynonymous SNV	GAD1:NM_000817:exon8:c.C850T:p.L284F	AR	paternal	0	0		
56	MR2053	F	Putative	chr2:171678594:T:C	splice region	GAD1:NM_013445.3:c.83-3T>C	AR	maternal	0	0		

Table S2. Causative experimentally identified variants and groups predictions

Submission	ID				ASD				Epilepsy				Microcephaly				Macrocephaly				Hypotonia				Ataxia			
	AUC	**MCC**	**ACC**	**F1**	**AUC**	**MCC**	**ACC**	**F1**	**AUC**	**MCC**	**ACC**	**F1**	**AUC**	**MCC**	**ACC**	**F1**	**AUC**	**MCC**	**ACC**	**F1**	**AUC**	**MCC**	**ACC**	**F1**	**AUC**	**MCC**	**ACC**	**F1**
1.1	0.57	0.20	0.95	0.98	0.51	**0.20**	**0.73**	**0.84**	0.53	**0.24**	**0.61**	**0.71**	0.57	0.16	0.46	0.41	0.64	0.27	0.86	0.15	0.49	0.20	0.63	0.36	0.46	0.08	0.78	0.14
2.1	0.70	0.16	0.71	0.82	0.55	0.10	0.36	0.22	0.38	-0.01	0.48	0.07	0.51	**0.37**	**0.81**	0.29	0.55	0.18	0.81	0.29	0.52	0.06	0.51	**0.51**	0.61	0.27	0.81	0.17
2.2	0.75	0.22	0.81	0.89	0.5	0.09	0.36	0.25	0.41	0.03	0.49	0.10	0.55	0.36	**0.81**	0.40	0.50	0.09	0.75	0.23	0.47	0.11	0.60	0.23	0.66	0.28	0.81	0.29
2.3	0.71	0.16	0.71	0.82	**0.56**	0.11	0.59	0.68	0.39	-0.01	0.48	0.07	0.49	**0.37**	**0.81**	0.29	0.56	0.18	0.81	0.29	0.51	0.07	0.53	0.50	0.66	0.33	0.72	0.48
2.4	**0.78**	0.21	0.79	0.88	0.49	0.09	0.36	0.25	0.41	0.03	0.49	0.10	0.48	0.36	**0.81**	0.40	0.56	0.14	0.75	0.29	0.49	0.11	0.60	0.23	**0.72**	**0.37**	0.72	**0.52**
2.5	0.64	0.09	0.55	0.70	0.55	0.10	0.36	0.22	0.40	-0.01	0.48	0.07	0.53	**0.37**	**0.81**	0.29	0.57	0.18	0.81	0.29	0.49	0.06	0.59	0.18	0.56	0.27	**0.81**	0.17
2.6	0.74	0.18	0.75	0.86	0.46	0.09	0.36	0.25	0.41	0.03	0.49	0.10	0.49	0.36	**0.81**	0.40	0.55	0.14	0.75	0.29	0.5	0.11	0.60	0.23	0.66	0.32	0.78	0.45
3.1	0.51	0.12	0.38	0.53	0.52	0.10	0.36	0.22	0.43	-0.10	0.47	0	0.50	0.11	0.78	0.10	0.62	0.18	0.65	0.33	0.54	0.21	0.62	0.13	0.49	-0.07	0.78	0
3.2	0.55	0.13	0.42	0.58	0.52	0.12	0.36	0.23	0.43	-0.07	0.46	0.17	0.50	0.11	0.78	0.10	0.62	0.18	0.65	0.33	0.52	0.15	0.60	0.07	0.49	-0.07	0.78	0

3.3	0.68	**0.34**	**0.97**	**0.99**	0.48	0.05	0.36	0.24	0.44	-0.05	0.47	0.20	0.50	0.11	0.78	0.10	0.62	0.18	0.65	0.33	0.52	0.15	0.60	0.07	0.49	-0.07	0.78	0
4.1	0.61	0.15	0.83	0.91	**0.56**	0.18	0.36	0.19	0.54	0.19	0.51	0.14	0.57	0.23	0.56	0.44	**0.70**	**0.39**	0.79	**0.48**	0.51	0.17	0.62	0.19	0.45	0.27	0.81	0.17
4.2	0.61	0.11	0.78	0.87	**0.56**	0.18	0.36	0.19	0.53	0.19	0.51	0.14	0.56	0.23	0.56	0.44	0.69	**0.39**	0.79	**0.48**	0.52	0.17	0.62	0.19	0.47	0.27	0.81	0.17
4.3	0.68	0.19	0.88	0.94	**0.56**	**0.20**	**0.73**	**0.84**	**0.56**	0.22	0.59	0.51	**0.63**	0.30	0.52	**0.47**	0.67	**0.39**	**0.88**	0.38	**0.56**	**0.25**	**0.65**	0.33	0.46	0.28	0.81	0.29
5.1	0.84	0.23	0.68	0.80	0.49	-0.02	0.45	0.51	0.50	0.01	0.50	0.39	0.56	0.13	0.72	0.30	0.48	-0.08	0.81	0	0.46	-0.08	0.47	0.38	0.47	-0.12	0.74	0

Suppl. Table S3. Summary of performance measures for all submissions and phenotypes.

Submission	nC	nCV	nC-CV	nC1	nC2	nC3	nC4	nC5	nC6	nC7
1.1	49	16	6	3	22	14	2	5	2	1
2.1	21	37	3	2	8	1	6	1	1	2
2.2	24	37	7	3	7	1	4	2	1	6
2.3	25	37	4	2	11	2	5	3	0	2
2.4	23	37	7	3	7	1	4	1	0	7
2.5	18	37	3	2	7	1	2	2	1	3
2.6	24	37	7	3	7	1	4	2	1	6
3.1	16	12	2	3	7	0	0	3	0	3
3.2	18	16	3	3	7	0	0	4	0	4
3.3	26	16	3	3	10	1	3	3	1	5
4.1	19	16	1	3	7	1	2	2	1	3
4.2	17	10	1	3	7	1	2	2	1	1
4.3	46	29	11	3	19	11	7	4	1	1
5.1	18	35	4	3	8	2	2	1	0	2

Suppl. Table S4: Summary of phenotype and variant prediction for all patients. nC is the number of patients where their phenotypic trait/s was/were correctly predicted. nC1, nC2, nC3, nC4, nC5, nC6 and nC7 are similar to nC but considering the number of phenotypic traits (from 1 to 7) provided for that patients by Padua NDD Lab. nCV is the number of patients for whom variants were correctly predicted. nC-Cv mean the number of patients for whom their phenotype/s and variant/s were well predicted.

Performance measures

All quantitative measures used for the determination of a ranking between the submissions are listed:

Accuracy (ACC):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Matthew Correlation Coefficient (MCC):

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

F_1 score (F_1):

$$F_1 = 2\frac{TPR \times PPV}{TPR + PPV}$$

Late submission assessment: Group 5 (enGenome)

Group 5 performed well in predicting most of the causative, putative and contributing factor variants indicated by the Padua NDD Lab. This group predicted most of causative variants (22 out of 25) compared to other groups, and a total amount of 36 out of 56 well predicted variants.
For the ID phenotype, group 5 achieved the highest AUC value (0.84). However, as was explained before this phenotype is highly biased to true positive values and the AUC didn't reflect the real performance. For the other phenotypes the performance of this method was poor, obtaining an overall rank of 9 among all the submissions (Suppl. Table S3). Moreover, group 5 was one of the least accurate in predicting the correct combination of patient phenotypes, correctly predicting just 7 patients where the Padua NDD Lab provided at least three phenotypic traits (Suppl. Table S3).

Prediction methodology

VCF files have been annotated and interpreted by the EVAI software v0.3.1[1]. EVAI uses various omics databases to annotate variants at genomic region, domain, gene, and disease level. Publicly available resources have been used for this purpose. EVAI generates a set of evidences to support variant pathogenicity assessment, according to ACMG/AMP guidelines[2] used to assign a five-tier class. EVAI classifies and scores variants according to every possible condition associated to the corresponding genes as reported in resources such as MedGen and ClinVar. Variability in disease-based classification can be related, for instance, to variant type or information about the gene (e.g. loss-of-function is a known mechanism for a certain disease). The expected incidence for a certain disease is taken into account as well while evaluating variant allele frequency in population databases. To identify gene-phenotype association, EVAI generated a list of annotated and ACMG-classified variants, prioritized by pathogenicity score (score >= 3). Variants present in more than 8 cases (about 5% allele frequency) and low quality INDELS (Quality by Depth < 4) were filtered out. Moreover, among variants with pathogenicity score >= 0, we retained those absent in the general populations (ExAC, 1TGP, ESP), that were unique for each sample, coding and predicted damaging (without conflict) by at least one in-silico functional predictors integrated in EVAI (PaPI, PolyPhen-2, SIFT, DANN). We then excluded variants that did not match the expected inheritance pattern reported by MedGen and OMIM. Finally, each selected variant has been associated to one or more target phenotypes according to OMIM.

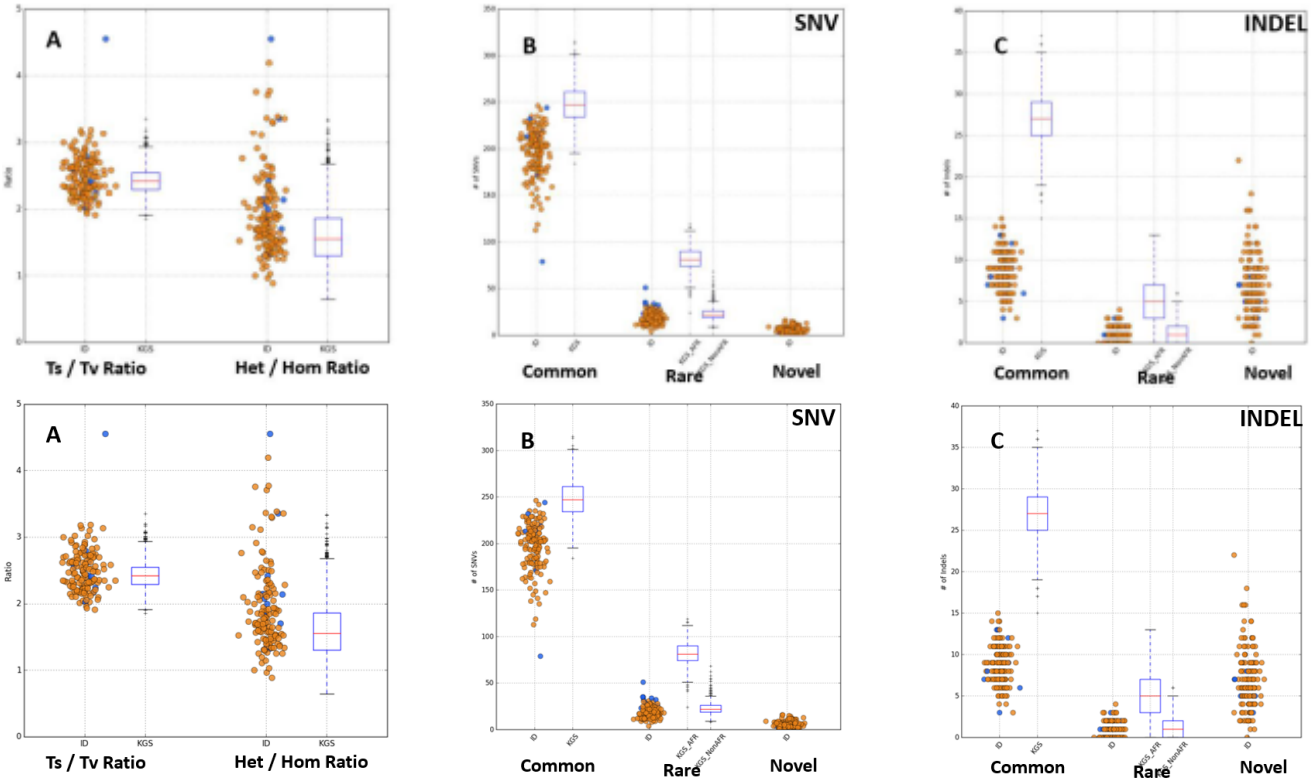
[1] Nicora G, Limongelli I, Gambelli P, Memmi M, Malovini A, Mazzanti A, Napolitano C, Priori S, Bellazzi R. 2018. CardioVAI: An automatic implementation of ACMG-AMP variant interpretation guidelines in the diagnosis of cardiovascular diseases. Hum Mutat 39:1835–1846.
[2] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, et al. 2015. 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. Genet Med 17:405–424.

Additional description of prediction methodology

Group 2: Moult Lab

QC Analysis

We performed routine QC analysis to see if there were obvious outlier samples. Figure 1 shows the results of the QC analysis. Variants flagged as PASS in the VCF file and with a genotype quality (GQ) greater than 30 were considered for the QC analysis. The transition/transversion ratio (Ts/Tv) and heterozygous/homozygous ratio across all the samples were comparable to 1000 Genomes dataset for the genomic regions captured for sequencing in the challenge dataset. Comparison of common, rare, and novel variant counts across samples was also performed. We based this analysis on the 1000 Genomes[3] database, with a variant considered novel if it was not present in 1000 Genomes, and considered low frequency if present in 1000 Genomes with a MAF less than 5%. Other 1000 Genomes variants were considered common. We generated two comparison plots – one for SNVs and the other for INDELs. We observed that the rare SNV and INDEL count distribution is comparable to 1000 Genomes samples, while the common SNV and INDEL count distributions are low compared to the variant count distribution of the 1000 Genomes. The novel SNV count varied from zero to 16 across samples and the novel INDEL count varied from zero to 22.



[3] Exome Aggregation Consortium (ExAC), Cambridge, MA (URL:http://exac.broadinstitute.org).

Figure 1: Comparison of variant calling quality for 150 challenge samples versus 2,504 1000 Genomes samples across the 74 genes in the intellectual disability panel. ID: Intellectual Disability Challenge Samples, KGS: 1000 Genomes samples, KGS_AFR: African samples in 1000Genomes, KGS_NonAFR: Non-African samples in 1000 Genomes. Figure 1A shows the distribution of Transition vs. Transversion (Ts/Tv) and Heterozygous SNVs vs. Homozygous SNVs (Het/Hom). By both measures, ID and KGS data are similar. Figure 1B shows the distribution of common, rare and novel SNV types. There are fewer common variants compared to 1000 genome data, but an approximately similar distribution of rare variants. Figure 1C shows the distribution of common, rare and novel Indels.

Building the Gene-Phenotype list

All the genes annotated in the bed files provided as part of the challenge were extracted for building the gene-phenotype list. These 74 genes were mapped to one or more of the seven phenotypes (intellectual disability, autism spectrum disorder, epilepsy, microcephaly, macrocephaly and ataxic gait) using two independent approaches generating two different gene-phenotype mapped files. The first approach used the OMIM database and Genetic Home Reference databases to map the phenotypes to the genes. The second approach using the Human Phenotype Ontology in addition to OMIM to map the phenotypes to the genes. The variant prioritization procedure was performed on each of these phenotypes lists.

Putative Causative Variant Search Method

The putative causative variant searches considered only rare or novel variants flagged as PASS in the VCF files. We considered a variant as rare if it was reported in ExAC with a MAF less than or equal to 1% and considered it novel if it was not reported in ExAC. Indels in low complexity regions (LCR) were excluded from the analysis, based on the LCR dataset pre-computed for the human genome by Heng Li[4]. A strand bias filter was used to remove variants whose alternate allele was present only on one strand of the reads mapped to the variant position. Search criteria were applied consecutively, starting with criteria deemed most reliable for finding causative variants and progressing to those considered less reliable. Two criteria were used: Variant quality and Variant impact. Five levels of variant quality were defined –

1. Variant with GQ>=30, variant not shared with any other sample,
2. Variant with GQ>=20, variant shared with one or no samples,
3. Variant with GQ>=20 and read depth (DP)>=30 and variant not shared with any other sample,
4. Variant with GQ>=10, read depth (DP)>=30 and variant not shared with any other sample, and
5. Variant with GQ>=30, variant shared with five or fewer samples.

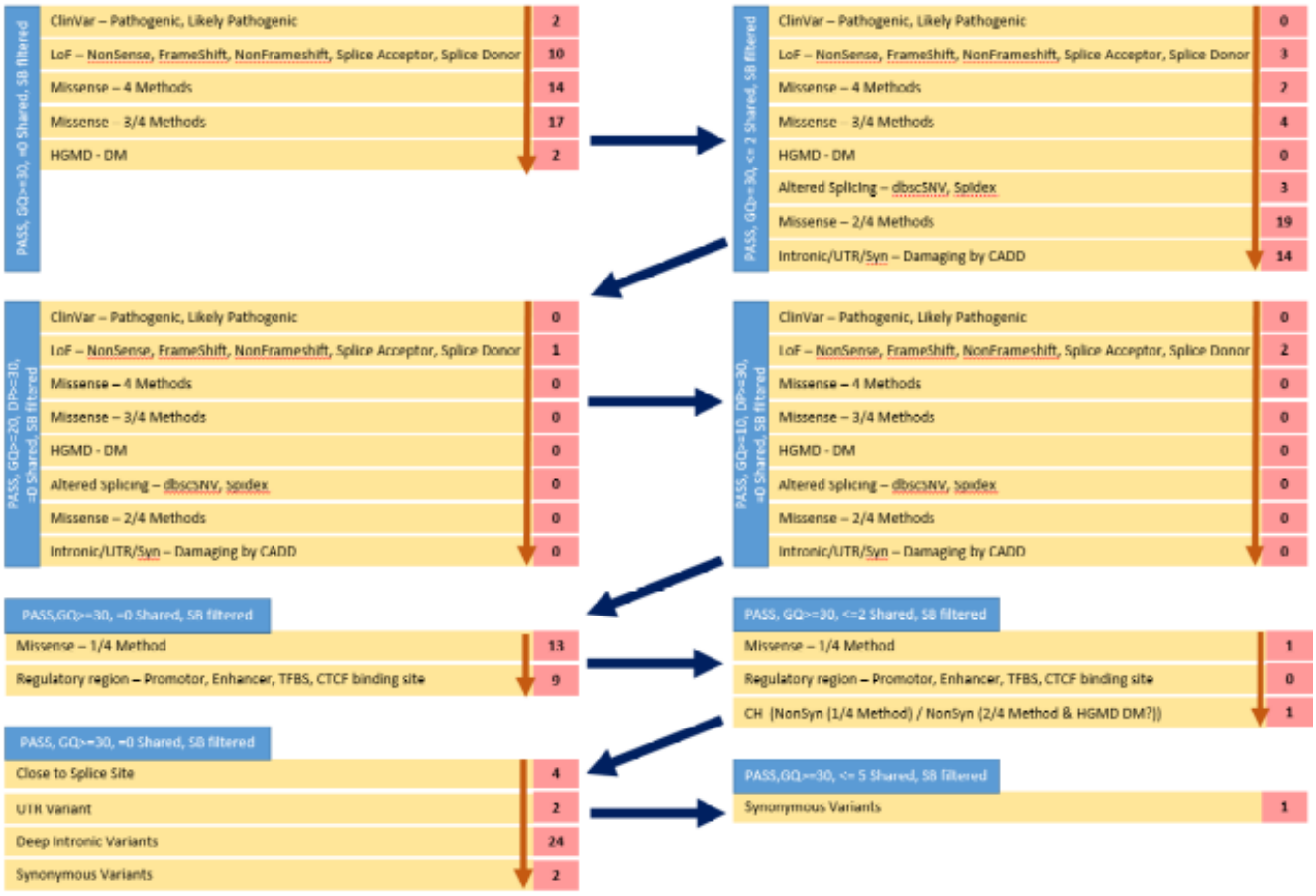
13 types of variant impact were defined based on the variant impact severity –

1. Variant reported in ClinVar with Pathogenic or Likely Pathogenic clinical significance.
2. Loss of Function variant (Nonsense, Frameshift, Non-Frameshift, Direct splicing acceptor/donor)

[4] Li H. Toward better understanding of artifacts in variant calling from high coverage samples. Bioinformatics.2014; 30:2843–2851.

3. Missense predicted damaging by four out of four methods.
4. Missense predicted damaging by three methods.
5. Variant reported in HGMD with DM status.
6. Variant predicted to alter splicing by either dbscSNV or SPIDEX.
7. Missense predicted damaging by only two methods.
8. Intronic, UTR or synonymous variant predicted damaging by CADD.
9. Missense predicted damaging by only one method.
10. Variant in a regulatory site (Promoter, Enhancer, TFBS or CTCF binding site).
11. Intronic variant
12. UTR variant
13. Synonymous variant.

The variant quality and impact filters were applied in a sequential manner to each sample as shown in Figure 2. If a putative causative variant was found in a sample for a specific combination of variant quality and impact filter, the search was terminated and that assignment was taken as final for that sample. Variants identified were further filtered for inheritance model associated with the gene. Variants were assigned to autosomal dominant, autosomal recessive, compound heterozygous, or X-linked models according to the available information for the gene concerned in OMIM and Genetic Home Reference database. Since the genotypes of the variants were not phased, we assumed a gene to follow a compound heterozygous model if two or more heterozygous variants were observed in that gene, but assigned a lower probability to such cases (see later) In practice, there was only one such case. Using this procedure we were able to assign 150 patients to one or more of the seven phenotypes, although a substantial fraction of assignments are of very low confidence (Figure 2).


PASS, GQ>=30, =0 Shared, SB filtered
ClinVar – Pathogenic, Likely Pathogenic 2
LoF – NonSense, FrameShift, NonFrameshift, Splice Acceptor, Splice Donor 10
Missense – 4 Methods 14
Missense – 3/4 Methods 17
HGMD - DM 2
PASS, GQ>= 30, <= 2 Shared, SB filtered
ClinVar – Pathogenic, Likely Pathogenic 0
LoF – NonSense, FrameShift, NonFrameshift, Splice Acceptor, Splice Donor 3
Missense – 4 Methods 7
Missense – 3/4 Methods 4
HGMD - DM 0
Altered Splicing – dbscSNV, Spidex 3
Missense – 2/4 Methods 19
Intronic/UTR/Syn – Damaging by CADD 14
PASS, GQ>=20, DP>=30, =0 Shared, SB filtered
ClinVar – Pathogenic, Likely Pathogenic 0
LoF – NonSense, FrameShift, NonFrameshift, Splice Acceptor, Splice Donor 1
Missense – 4 Methods 0
Missense – 3/4 Methods 0
HGMD - DM 0
Altered Splicing – dbscSNV, Spidex 0
Missense – 2/4 Methods 0
Intronic/UTR/Syn – Damaging by CADD 0
PASS, GQ>=10, DP>=30, =0 Shared, SB filtered
ClinVar – Pathogenic, Likely Pathogenic 0
LoF – NonSense, FrameShift, NonFrameshift, Splice Acceptor, Splice Donor 2
Missense – 4 Methods 0
Missense – 3/4 Methods 0
HGMD - DM 0
Altered Splicing – dbscSNV, Spidex 0
Missense – 2/4 Methods 0
Intronic/UTR/Syn – Damaging by CADD 0
PASS,GQ>=30, =0 Shared, SB filtered
Missense – 1/4 Method 13
Regulatory region – Promotor, Enhancer, TFBS, CTCF binding site 9
PASS, GQ>=30, <=2 Shared, SB filtered
Missense – 1/4 Method 1
Regulatory region – Promotor, Enhancer, TFBS, CTCF binding site 0
CH (NonSyn (1/4 Method) / NonSyn (2/4 Method & HGMD DM?)) 1
PASS, GQ>=30, =0 Shared, SB filtered
Close to Splice Site 4
UTR Variant 2
Deep Intronic Variants 24
Synonymous Variants 2
PASS,GQ>=30, <= 5 Shared, SB filtered
Synonymous Variants 1

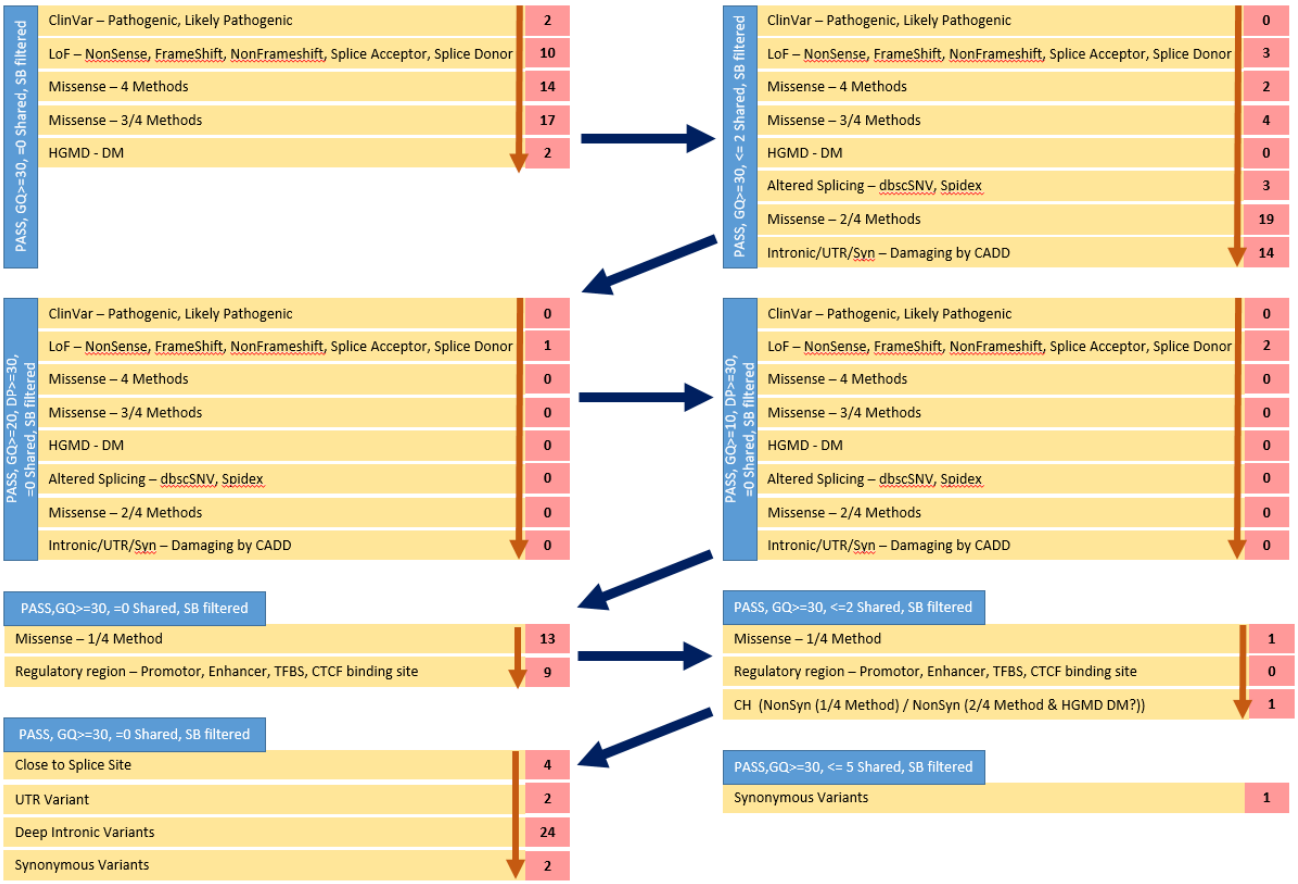


Figure 2. Variant prioritization strategy. The blue boxes show the variant quality filter and the yellow boxes show the variant impact filters. The arrows show the sequence in which the combination of quality and impact filters were applied to each sample. The numbers next to the impact filters represent the number of samples in which putative causative variants were found at that filtering stage.

Probability score computation

Developing a proper model for assigning a probability of variant causing a disease phenotype in principle possible, but requires substantial analysis to properly calibrate. For this reason, we used a number of ad hoc procedures. An exception was for missense variants, where we assigned the probability of disease using the extent of consensus among the four missense analysis methods, previously calibrated from HGMD data and a control set of inter-species variants. Other variant types were subjectively assigned probabilities as shown in Table 1. For autosomal recessive situations, we assumed both contributing variants must have high impact, and so used the product of the corresponding probabilities. We further assumed that the occurrence of homozygous cases increased the likelihood of a correct disease assignment, and added 0.2 to probabilities in these situations. A smaller increment was used for compound heterozygous situations since these have a significant probability of occurring by chance and in some cases may be on the same copy of the

gene. Ad hoc probabilities of a correct variant call were also assigned to each variant (Table 2). A prior probability of each phenotype was computed from the 93 high confidence prediction set.
The final probability for the 93 high confidence sample set was taken as the product of the call and impact probabilities and was scaled to match the prior probability of the predicted phenotypes. For the remaining 57 sample sets, the prior probability of the phenotypes
We made six submissions based on the two different gene-phenotype lists and different combination of probabilities.

- Submission 1: Gene-Phenotype files based on OMIM, probabilities for 93 high confidence sample set are based on the product of the call and impact probabilities and further scaled to match the prior probabilities of the each phenotypes. The remaining 57 samples were assigned prior probabilities of individual phenotypes ignoring the weak genetic signal.
- Submission 2: Gene-Phenotype files based on OMIM+HPO, probabilities for 93 high confidence sample set are based on the product of the call and impact probabilities and further scaled to match the prior probabilities of the each phenotypes. The remaining 57 samples were assigned prior probabilities of individual phenotypes ignoring the weak genetic signal.
- Submission 3: Gene-Phenotype files based on OMIM, probabilities for 93 high confidence sample set are based on the product of the call and impact probabilities and further scaled to match the prior probabilities of the each phenotypes. For 22 low confidence samples where the prioritized variant were either missense predicted damaging by ¼ methods or regulatory variants, the predicted phenotype's prior probabilities were increased compared to prior probabilities of other non-predicted phenotypes. The remaining 35 samples were assigned prior probabilities of individual phenotypes ignoring the weak genetic signal.
- Submission 4: Gene-Phenotype files based on OMIM+HPO, probabilities for 93 high confidence sample set are based on the product of the call and impact probabilities and further scaled to match the prior probabilities of the each phenotypes. For 22 low confidence samples where the prioritized variant were either missense predicted damaging by ¼ methods or regulatory variants, the predicted phenotype's prior probabilities were increased compared to prior probabilities of other non-predicted phenotypes. The remaining 35 samples were assigned prior probabilities of individual phenotypes ignoring the weak genetic signal.
- Submission 5: Gene-Phenotype files based on OMIM, probabilities for all sample set are based on the product of the call and impact probabilities and further scaled to match the prior probabilities of the each phenotypes.
- Submission 6: Gene-Phenotype files based on OMIM+HPO, probabilities for all sample set are based on the product of the call and impact probabilities and further scaled to match the prior probabilities of the each phenotypes.

Variant Type	Probability Score
Reported in ClinVar as pathogenic	1
Reported in HGMD as DM	0.7
NonSyn – Predicted damaging by 4/4 tools	0.8
NonSyn – Predicted damaging by 3/4 tools	0.7
NonSyn – Predicted damaging by 2/4 tools	0.5
NonSyn – Predicted damaging by 1/4 tools	0.25
NonSyn - Not predicted damaging by any tools	0.15
Nonsense	1
Frameshift / Non-Frameshift	1
Direct Splice Site – Donor/Acceptor	1
Variant predicted to affect splicing	0.7
UTR/Intronic/Syn predicted damaging by CADD	0.55
Variant at Regulatory Site	
Variant close to Splice Donor site	0.2
Variant close to Splice Acceptor site	0.2
UTR Variant	0.05
Intronic Variant	0.05
Synonymous Variant	0.05

Table 1. Probability assignments for each variant type.

Variant Type	Probability Score
PASS, GQ>=30, 0 shared	1
PASS, GQ>=30, 1 shared	0.9
PASS, GQ >= 20, DP>30, 0 shared	0.7
PASS, GQ >= 10, DP>30, 0 shared	0.6
PASS, GQ>=30, 5 shared	0.5

Table 2. Probability of a correct variant call based the variant quality filters. GQ=Genotype Quality, DP=Read Depth, # shared indicates considering variant for prioritization that was shared in <=# samples.

Group 4: Brenner Lab

The methodology of CHESS (v0.1)

Public data used in CHESS

We downloaded and processed the variant frequency data from GNOMAD v2.0.2[5], pre-calculated variant deleterious scores by REVEL[6], and the clinical evidence data from ClinVar[7] downloaded on 2017-10-02.

The scoring scheme

Firstly, for each case, we first collect the phenotype terms of the case and calculate the **phenotype matching scores for all the genes** in the human genome that have relevant information with the phenotypes or have functional relationships with the genes that are known to be associated with the phenotypes. We use **Phenolyzer**[8] to calculate these gene-phenotype scores.

Secondly, we take the pre-called variants from the case exome as the input data, and annotate the data with VEP[9], GNOMAD variant frequency data, ClinVar evidence, and the pre-calculated REVEL scores. To reduce the computing burden, we exclude common variants (variants with MAF >= 5%) from our analysis, given the notion that the diseases of interest are rare mendelian diseases. Severest annotation by VEP are used for each variant, and thus variants with no annotation that they are protein-altering are excluded from the analysis.

Thirdly, we calculate all possible events based on both dominant and recessive models, preferably with data from the parents available. For cases where family information is not available, single variant events for the dominant model and compound heterozygous events for the recessive model are estimated, with false positive events expected. This can be adjusted on a case-by-case basis.

Fourthly, we systematically score all the remaining variants using the following equation:

$$V_{score} = \log \frac{e^{qual+w} + e^{del+w} + e^{pheno+w}}{3}$$

Where:

- Qual is 0-1, variant quality (transformed GQ value)
- del is 0-1, variant deleterious value, a sum of two components:
    - Impact annotation: 0.25 for missense/inframe, 0.5 for frameshift
    - REVEL score
- pheno – 0-1, phenotype match score of the gene (Phenolyzer)

[5] Lek, M., et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536(7616):285-291.

[6] Ioannidis, N.M., et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet 2016;99(4):877-885.

[7] Landrum, M.J., et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 2016;44(D1):D862-868.

[8] Yang, H., Robinson, P.N. and Wang, K. Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. Nat Methods 2015;12(9):841-843.

[9] McLaren, W., et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17(1):122.

- w– a supplement weight to help weigh down benign/common variants, with two components:
    - Default benign = 0; benign = -1 if annotated as benign in clinVar with >=2 stars
    - Rareness = (0.01 - MAF) / 0.05
- Please note that log averaged exponential privileges the larger value, so that we could catch a variant that has one feature that stands out.

Lastly, we calculate a score for each possible event in each gene, by combining the scores of each involved variant.

Special adjustments of CHESS v0.1 made for the CAGI intellectual disability panel challenge:

- Since no family info is available, the single gene dominant events are estimated and may not be real, which will give a high false positive rate. The scores for these events were adjusted by dividing the scores for such single variants by 2 as the event scores.
- Similarly, the scores for "compound heterozygous" events were calculated as the sum of two variants involved divided by 3 (or 2, in less stringent version).
- Since the challenge is to identify the disease from a total of seven diseases, a score was calculated for each event for each disease in each proband.
- **The stringent model:**
    - Variant called in more than 30% of the samples were excluded from the analysis.
    - Only top hit for each disease was reported, and events scoring lower than 0.5 would not be reported.
    - **The medium stringent model**:
        - Variant called in more than 30% of the samples were excluded from the analysis.
        - Only top hit for each disease was reported.
    - **The less stringent model:**
        - Variant called in more than 50% of the samples were excluded from the analysis.
        - The scores for "compound heterozygous" events were calculated as the sum of two variants involved divided by 2.
        - Top 2 hits for each disease was reported, and the probability (score) used for each prediction was the mean of the top 2 hits.