The American Journal of Human Genetics, Volume *107*

Supplemental Data

Interpretable Clinical Genomics with a Likelihood Ratio Paradigm

Peter N. Robinson, Vida Ravanmehr, Julius O.B. Jacobsen, Daniel Danis, Xingmin Aaron Zhang, Leigh C. Carmody, Michael A. Gargano, Courtney L. Thaxton, UNC Biocuration Core, Guy Karlebach, Justin Reese, Manuel Holtgrewe, Sebastian Köhler, Julie A. McMurry, Melissa A. Haendel, and Damian Smedley

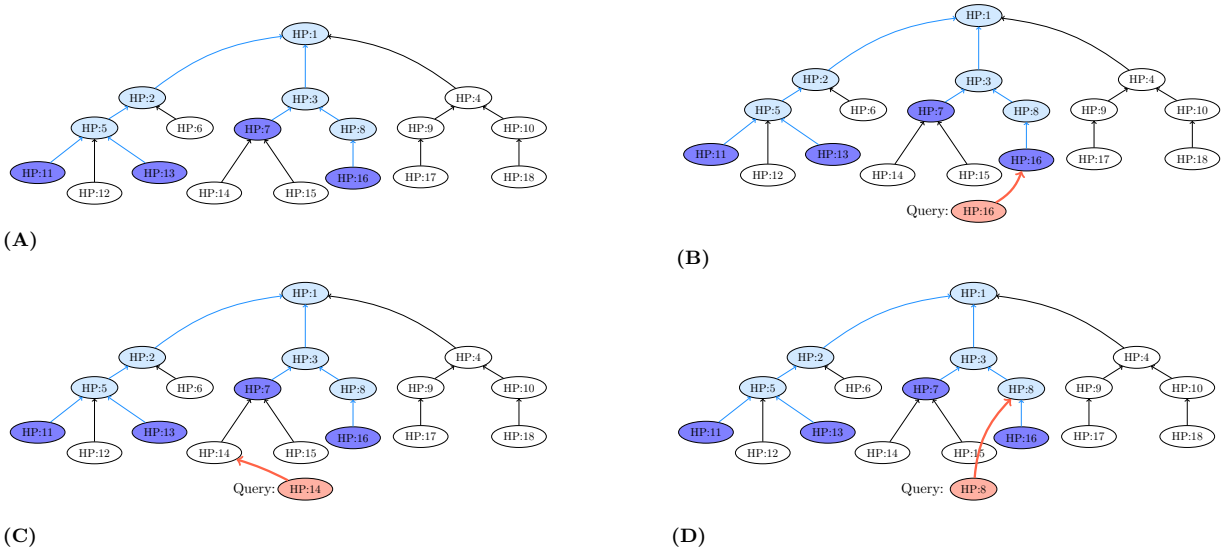
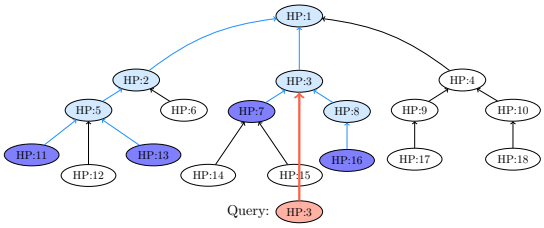
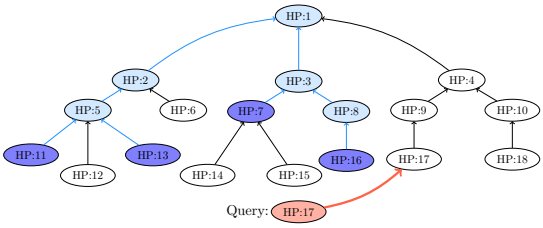


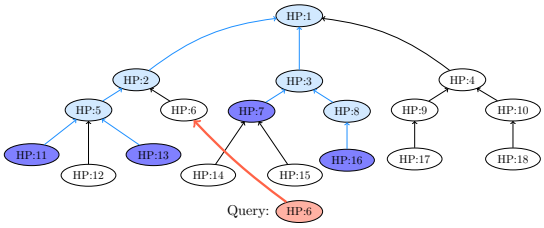
Figure S1. Calculating the likelihood ratios for phenotypes. **(A)** We will explain how the likelihood ratios (LR) for phenotypes are calculated using the example ontology shown here. The ontology contains 17 terms. For a certain disease, which we will call $\mathcal{D}$, four of the terms are directly annotated (HP:7, HP:11, HP:13, and HP:16, shown in dark blue). Because of the propagation of annotations, each of the ancestors of these terms is implicitly annotated to $\mathcal{D}$ as well (the terms are shown in light blue, and the edges encoding this inheritance are also shown in light blue). For instance, if HP:15 refers to `Nuclear cataract`, HP:8 refers to `Cataract`, and HP:3 refers to `Abnormal lens morphology`, then if we annotate disease $\mathcal{D}$ to `Nuclear cataract`, then we are also stating that the disease is characterized by `Cataract` and by `Abnormal lens morphology`. The term HP:1 is the root of this ontology (comparable to `Phenotypic abnormality` in the full HPO); **(B)** In this case a query term matches one of the directly annotated terms exactly. Then probability of observing HP:16 in an individual with $\mathcal{D}$ is simply the frequency of HP:16 in $\mathcal{D}$, or $P(h_{16}|\mathcal{D}) = f_{16}^{\mathcal{D}}$; **(C)** In this case, the query (HP:14) term matches a descendent of HP:7. HP:14 is not itself annotated to $\mathcal{D}$. In this case, we assume that the direct annotation (HP:7) is equally likely to correspond to an of its k subterms. If we assume that all individuals with disease $\mathcal{D}$ have the phenotypic feature represented by HP:7, then the frequency is 100%, i.e., $f_7^{\mathcal{D}} = 1.0$. We therefore divide this frequency by k. In this case, HP:7 has two descendents and $k = 2$, and therefore $P(h_{14}|\mathcal{D}) = \frac{f_7^{\mathcal{D}}}{2} = 0.5$; **(D)** Here, the query is to HP:8, an ancestor of a term that is directly annotated to $\mathcal{D}$. Because of annotation propagation, the probability of observing HP:8 in individuals with this disease is equivalent to the probability of observing HP:16, viz., $P(h_8|\mathcal{D}) = f_{16}^{\mathcal{D}}$.



(E)



(F)



(G)

Figure S1. Calculating the likelihood ratios for phenotypes (continued). **(E)** HP:3 is an ancestor of two terms used to annotate $\mathcal{D}$. Here the maximum probability of HP:7 and HP:16 is taken, i.e., $P(h_3|\mathcal{D}) = \max(f_7^{\mathcal{D}}, f_{16}^{\mathcal{D}})$; **(F and G)** In this case, the query term is not directly annotated in the disease and is not a subclass of a disease term, nor is a disease term a subclass of the query term. Following the graph, the query term and some disease annotation have a common ancestor. This common ancestor can be a root term (**F**) or a non-root term (**G**). If their common ancestor is at the root, then the query does not affect an organ that is affected by the disease. An arbitrary small likelihood ratio of $\frac{1}{100}$ is assigned in this case. If there is a common ancestor below the root (h_{ca}), then the query term affects the same organ as the disease annotation without being a closely matched feature. In this case, we model the probability as being related to the overall frequency of the feature in the HPO corpus, but set the probability to be a minimum of $\frac{1}{100}$ to avoid an overly large influence of very rare features.

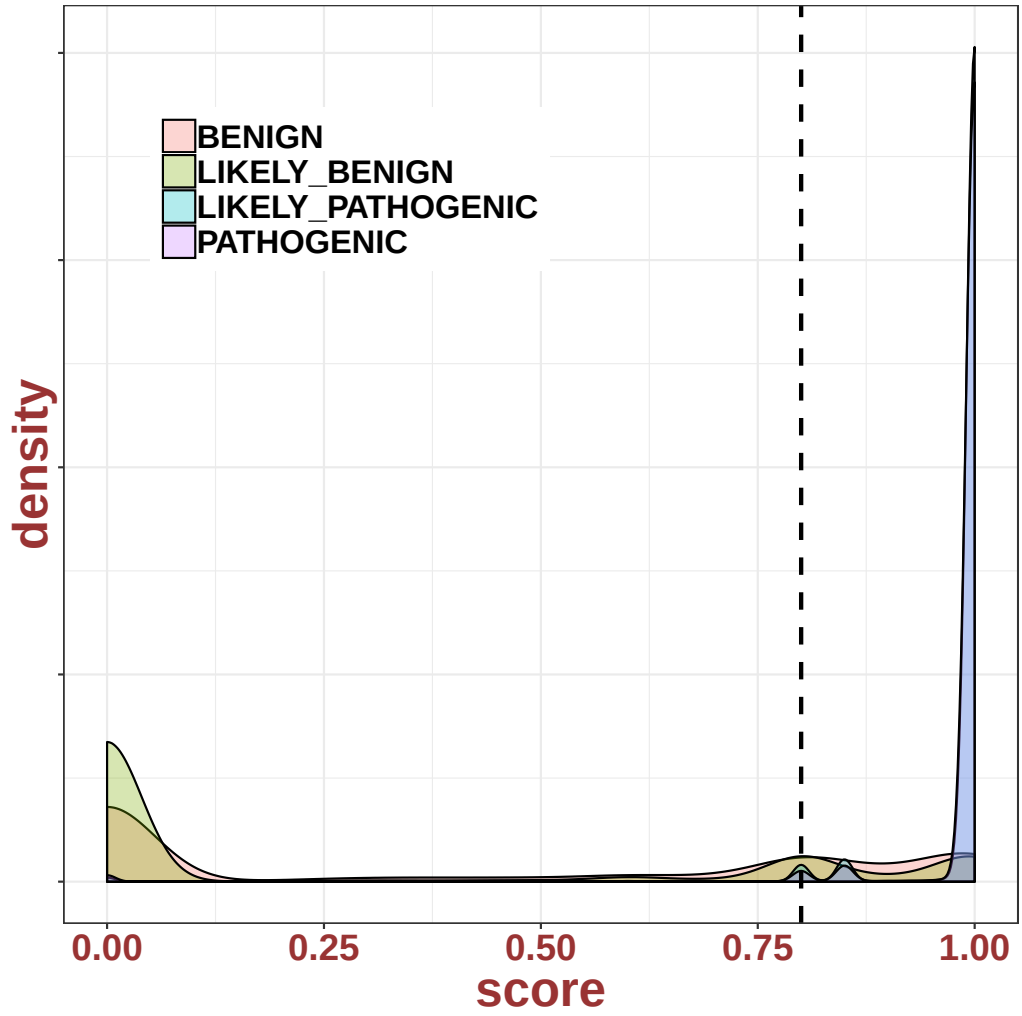


Figure S2. The Exomiser predicted pathogenicity score was calculated for each variant in ClinVar whose genomic position was precisely specified as nucleotide positions (these tend to be single-nucleotide variants or variants encompassing a small number of nucleotides rather than structural variants). A total of 160,714 such variants were available for analysis in the Exomiser data distribution version 12.1.0. The were 16,499 `benign` variants (10.3%), 64,123 `likely benign` variants (39.9%), 27,830 `likely pathogenic` variants (17.3%), and 52,262 `pathogenic` variants (32.5%). For the purpose of this analysis, the category `likely benign or benign` was assigned to `likely benign`, and `likely pathogenic or pathogenic` was assigned to `likely pathogenic`. In this work, a threshold pathogenicity score of 0.8 was chosen. The percentages of variants with an Exomiser score of at least 0.8 was: `benign`: 36.1%, `likely benign`: 26.5%, `likely pathogenic`: 99.3%, and `pathogenic`: 98.9%. The analysis was performed using the hg19 data. Similar results were obtained for hg38.

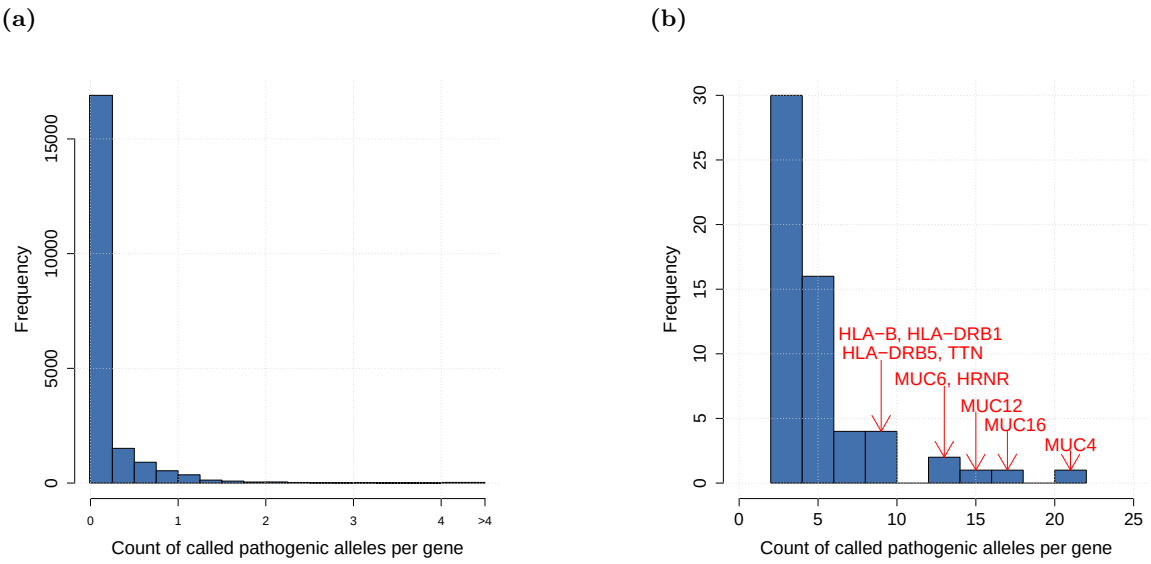


Figure S3. Frequencies of called pathogenic variants per gene. The frequencies of variants whose predicted pathogenicity score was 0.8 or higher was summed for each of 20,632 protein-coding genes and the count (frequency) of genes is plotted. Data are derived from the hg19 gnomAD dataset. Similar results were obtained for hg38. **(a)** An overview of the entire distribution. **(b)** Counts are shown for the 59 genes with counts above 3. Gene symbols are shown for all genes with counts above 8.

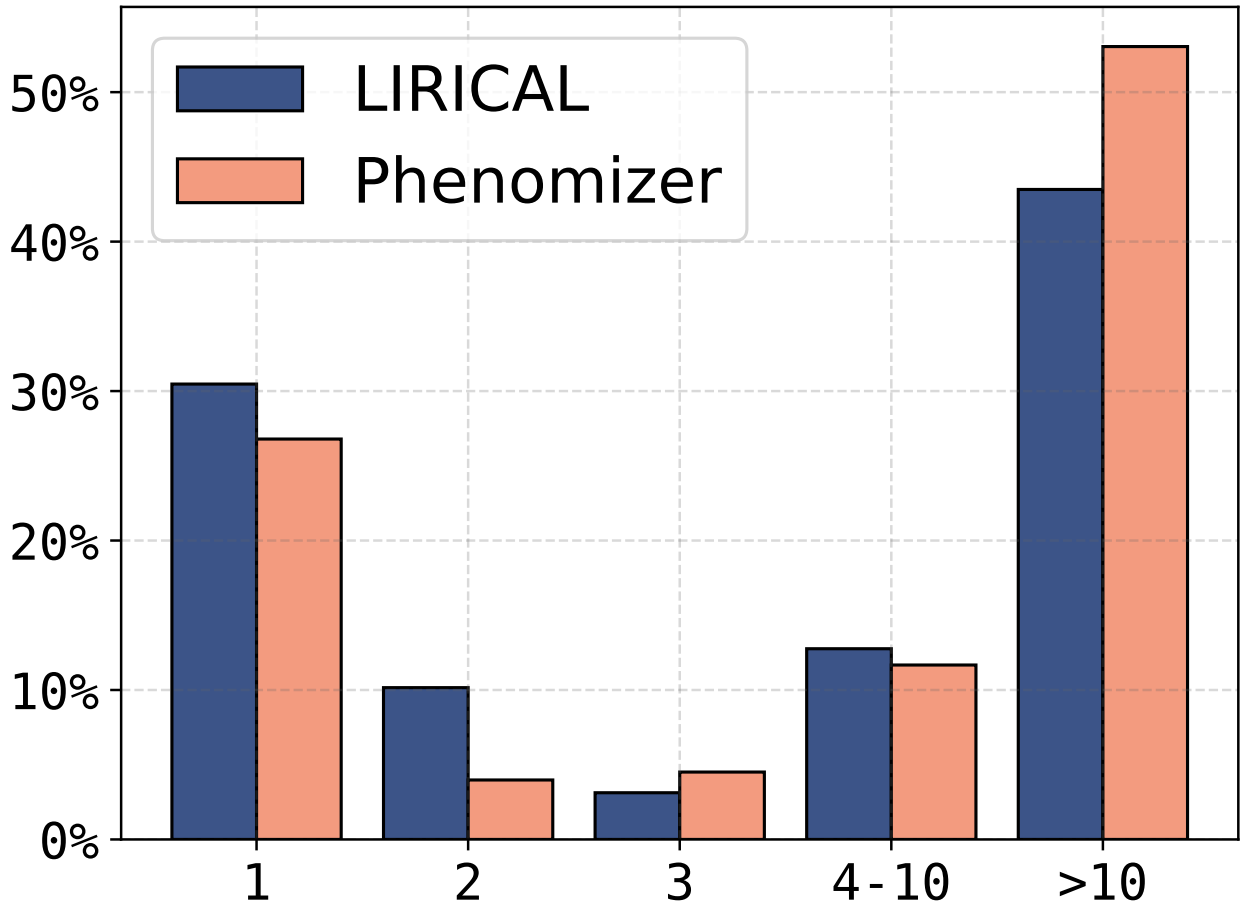


Figure S4. Comparison of LIRICAL and Phenomizer. The performance of LIRICAL (phenotype-only mode) was compared with that of Phenomizer [1] on the dataset of 384 Phenopackets (Table S2). For this analysis, the genetic information was not used, because Phenomizer is not able to use genetic information. The percentage of cases in which the true diagnosis was placed at a given rank is shown on the Y axis. The X axis shows the ranks or rank groups. LIRICAL placed a total of 43.7% of cases in the top 3 ranks, and Phenomizer placed a total of 35.3% of cases in the top 3 ranks.

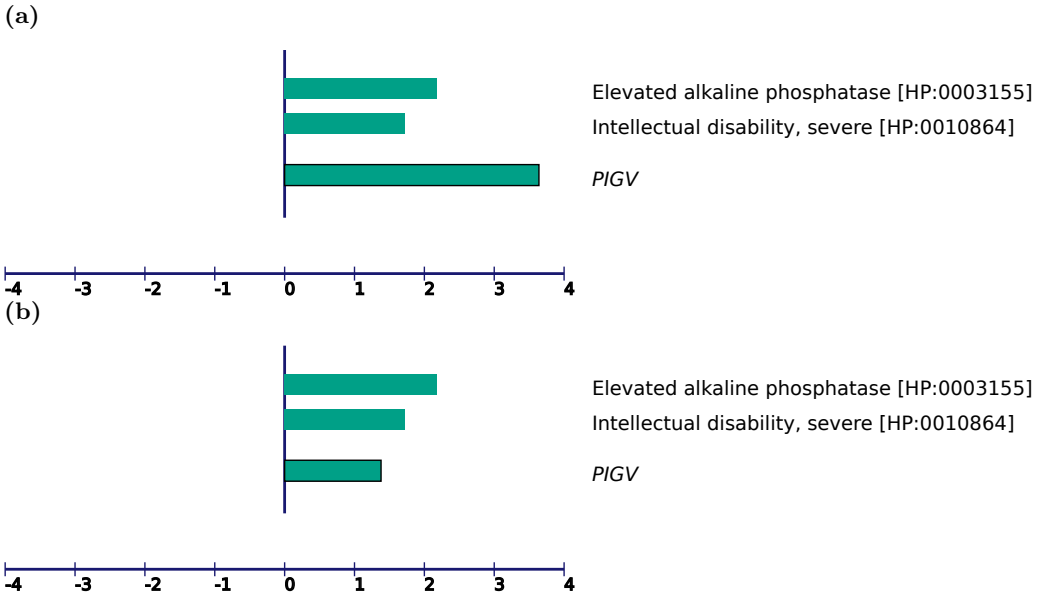


Figure S5. Ranking of an autosomal recessive disease with one pathogenic allele. Current exome and genome technologies can miss variants in highly GC-rich exons or can fail to detect structural variants. This may lead to only one of the expected two pathogenic alleles being identified for an autosomal recessive candidate disease. In this example, we show a simulated case of Hyperphosphatasia with mental retardation syndrome 1 (OMIM:239300) with two typical features. LIRICAL does not apply a hard filter to such cases but instead employs a flexible genotype likelihood ratio score. (a) Simulation with two pathogenic alleles; (b) Simulation in which one of the two alleles was removed. The LR for PIGV is lower but still contributory and the correct diagnosis remained in rank #1. The variants are chr1:27121140C>G (NM_001202554.1:c.615C>G, NP_001189483.1:p.(Asn205Lys)) and chr1:27121379A>G (NM_001202554.1:c.854A>G, NP_001189483.1:p.(Tyr285Cys)). Chromosomal coordinates are according to hg19.

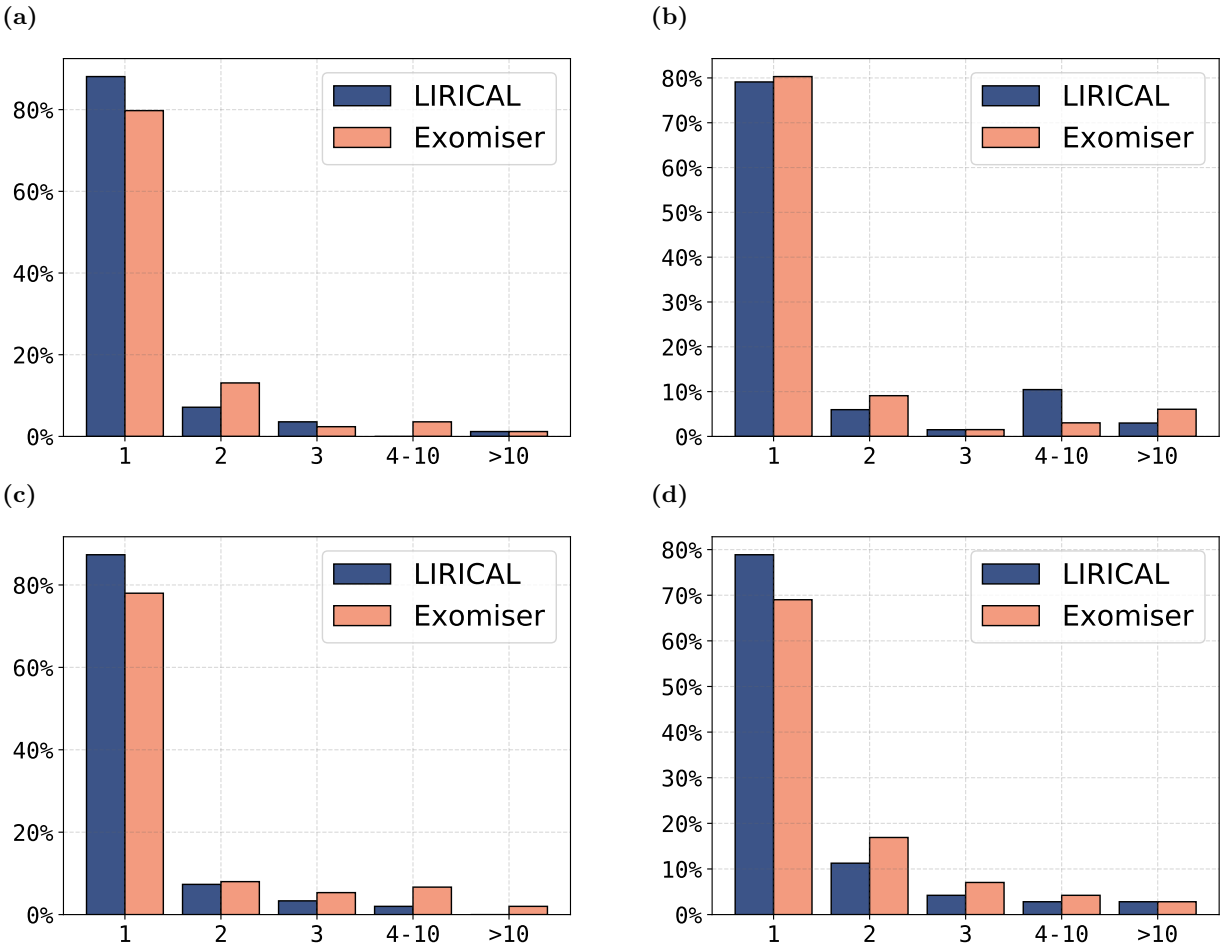


Figure S6. Performance of LIRICAL and Exomiser according to mode of inheritance and ClinVar status. The evaluation shown in Figure 2 of the main manuscript was repeated for subsets of the data. **(a)** Autosomal dominant diseases with disease-associated variant listed as pathogenic in ClinVar ($n = 84$); **(b)** Autosomal dominant diseases without variant listed as pathogenic in ClinVar ($n = 67$); **(c)** Autosomal recessive diseases with at least one disease-associated variant listed as pathogenic in ClinVar ($n = 150$); **(d)** Autosomal recessive diseases without variant listed as pathogenic in ClinVar ($n = 71$).

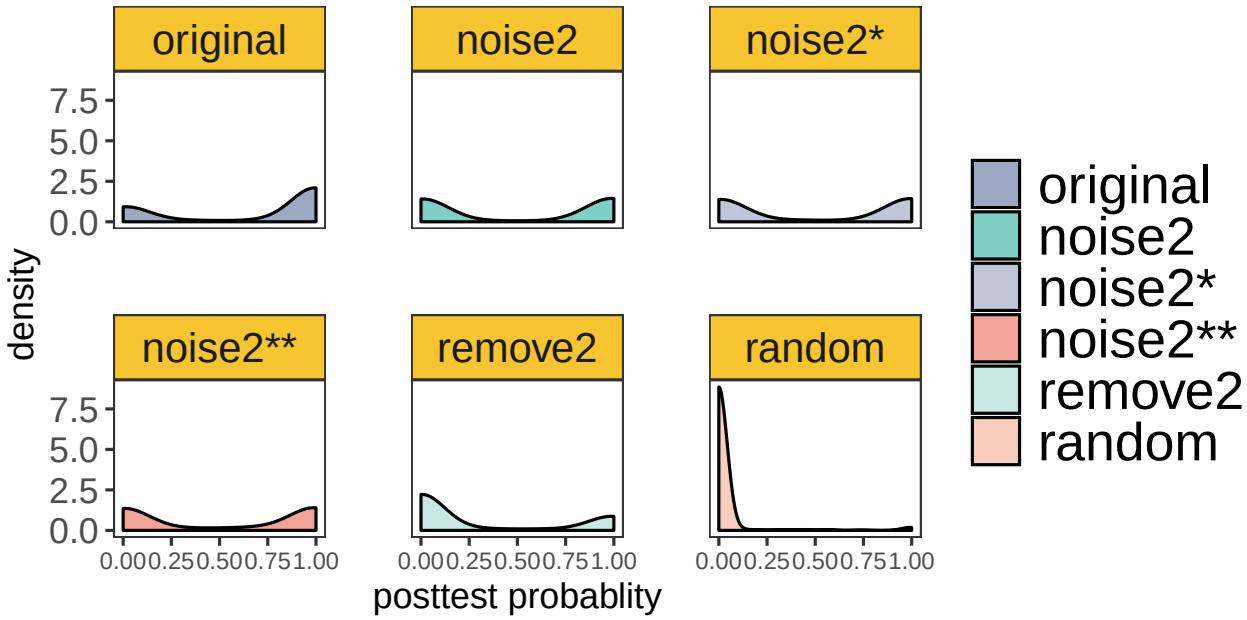


Figure S7. LIRICAL's posttest probability estimates. The post-test probability of the correct diagnosis was calculated for each of the 384 phenopacket case reports (Original). The mean post-test probability (pp) of the original data was 67.4%. Five procedures were applied to add noise to this data (Supplemental Table S3). Results for the original data are shown as `original`. `noise2`: two random HPO terms were added (mean pp: 50.8%); `noise2*`: two random HPO terms were added and original terms were replaced by parent terms (mean pp: 50.3%); `noise2**`: two random HPO terms were added and original terms were replaced by grandparent terms (mean pp:(mean pp: 50.3%); `remove2`: All pathogenic alleles were removed (mean pp: 29.4%); `random`: All HPO terms were replaced by random terms (mean pp: 2.9%).

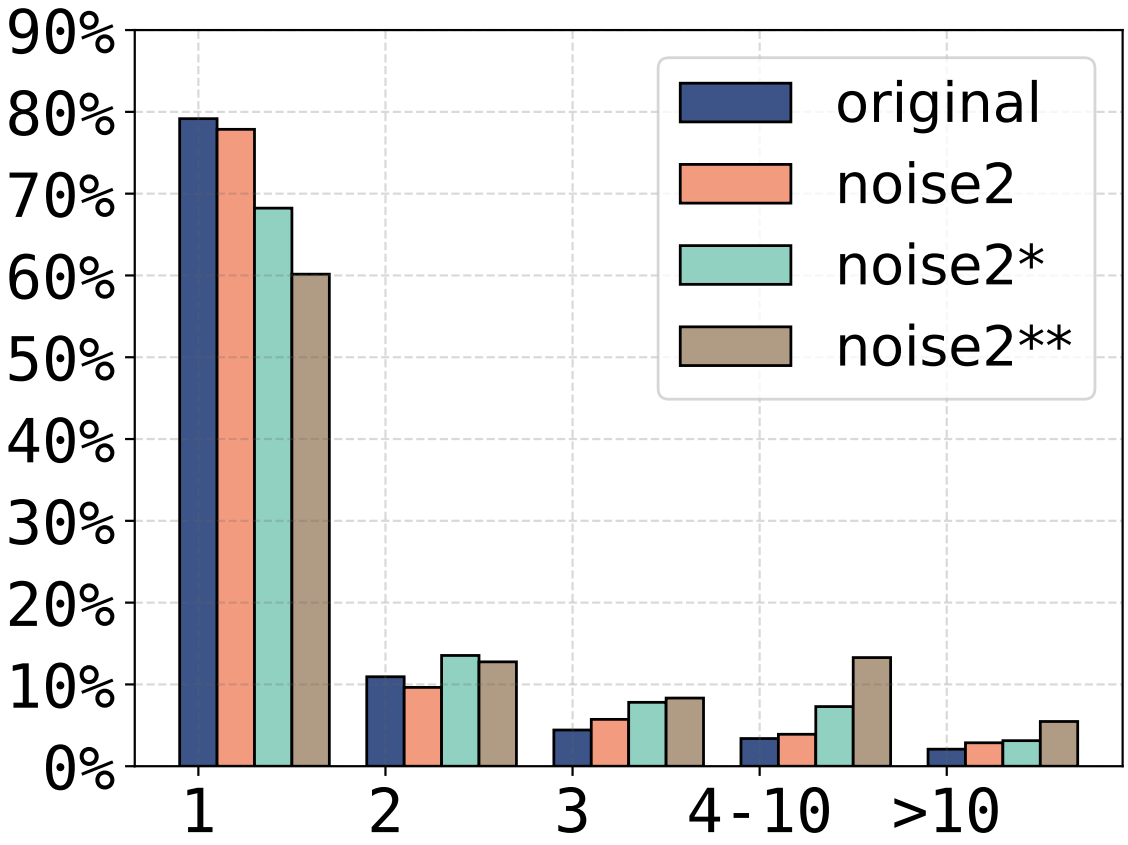
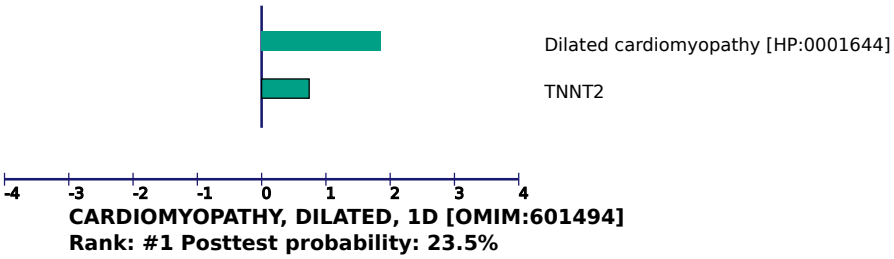
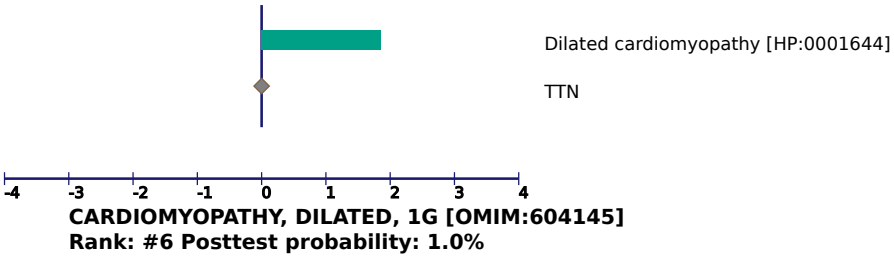


Figure S8. LIRICAL disease ranks. The ability of LIRICAL to predict the correct disease was assessed with 384 case reports (Table S2). This is the same simulation as presented in Fig. 2 of the main manuscript, but the rank is recorded for diseases instead of for disease genes. This is a harder prediction task because many genes are associated with multiple Mendelian diseases. Four tests were performed: `original`: unaltered data from the case reports. `noise2`: Two random ("noise") HPO terms are added to each case; `noise2*`: Original terms are replaced by a parent term and two noise terms are added; `noise2**`: Original terms are replaced by a grandparent term and two noise terms are added. The X axis shows the rank assigned by LIRICAL to the correct disease gene. The Y axis shows the percentage of cases in which the given rank was achieved.

(a)



(b)



(c)

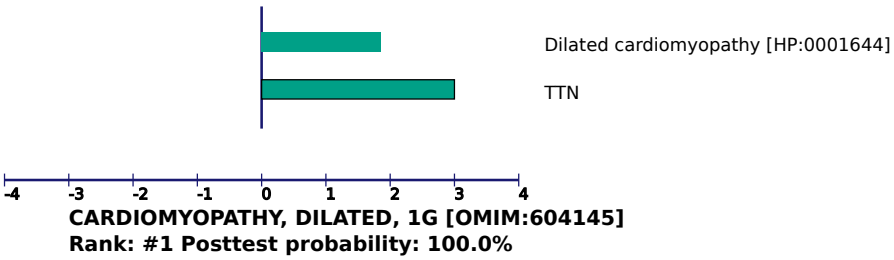


Figure S9. LIRICAL's treatment of ClinVar pathogenic variants. In this example, we simulate a patient with a rare (0.02% maximum population frequency) variant in *TTN*, NM_001267550.2(TTN):c.18295C>T [2], who is noted to have `Dilated cardiomyopathy` (HP:0001644). The variant is listed as having Uncertain significance in ClinVar (VCV000263438.2). **(a)** The candidate placed in rank 1 is a false positive, Dilated cardiomyopathy 1D (OMIM:601494) related to a variant in the *TNNT2* gene (NM_000364.3: c.683T>C, p.(Ile228Thr) that was present in the control VCF file. This variant is listed in ClinVar as having uncertain clinical significance (VCV000181604.2). **(b)** The correct candidate is placed at rank 6, Dilated cardiomyopathy 1G (OMIM:604145). The *TTN* mutation is scored with a likelihood ratio of just 2.70 in favor of OMIM:604145 because of the high background frequency of variants in this gene ($\lambda^{\mathcal{B}} = 9.4564$), despite the near maximal raw pathogenicity score of Exomiser (0.997). **(c)** In a separate simulation, the *TTN* variant NM_001267550.2:c.2926T>C (p.Trp976Arg) was spiked into the same control VCF file. This variant is listed in ClinVar as likely pathogenic (VCV000012651.3), and for this reason is (heuristically) assigned a likelihood ratio of 1000 by LIRICAL. The candidate is now correctly ranked in first place.

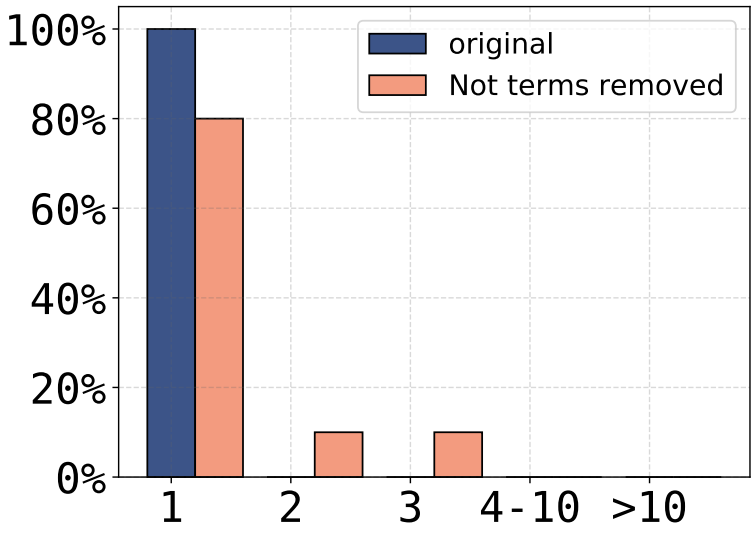
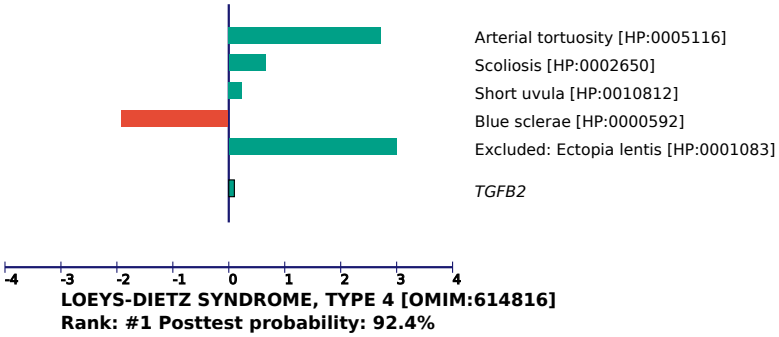
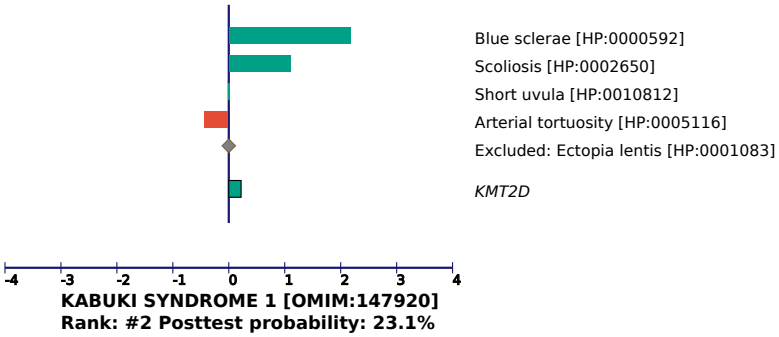


Figure S10. Negated annotations. LIRICAL was run with ten cases with a negated ("not") annotation deemed important for the differential diagnosis. For instance, Loeys-Dietz syndrome 4 is annotated not to have `Ectopia lentis`. Although the overall performance was good even without the negated annotations, in two of the ten cases, including the negated annotation boosted the rank of the correct candidate disease from 2 or 3 to 1. The X axis shows the rank assigned by LIRICAL to the correct disease gene. The Y axis shows the percentage of cases in which the given rank was achieved.

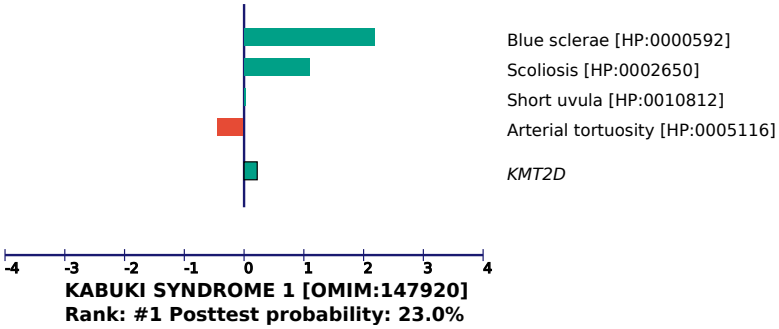
(a)



(b)



(c)



(d)

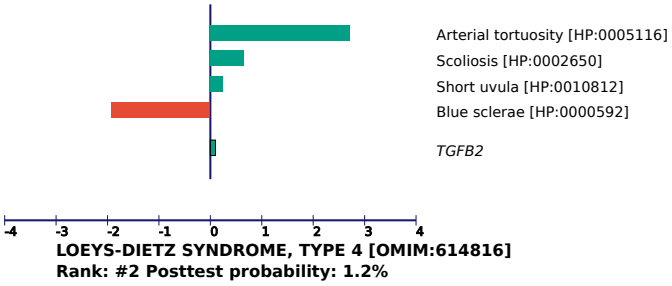
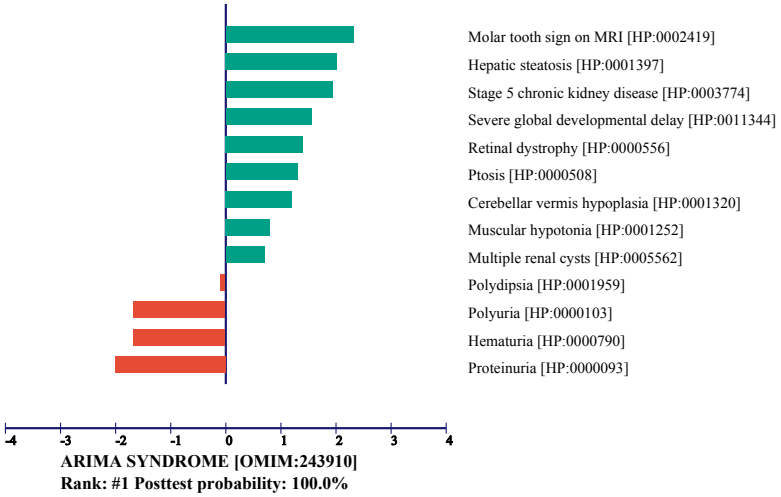


Figure S11. Ranking of candidate diseases with and without excluded features. In this example, panels **(a)** and **(b)** were run using a negated query term, `Ectopia lentis`, that had been excluded by examination of a hypothetical proband. Ranks 1 and 2 are shown. The correct diagnosis, Loeys-Dietz syndrome 4, has a posttest probability of 92.4%. In panels **(c)** and **(d)**, the excluded term was omitted and the correct diagnosis was placed in rank 2.

(a)



(b)

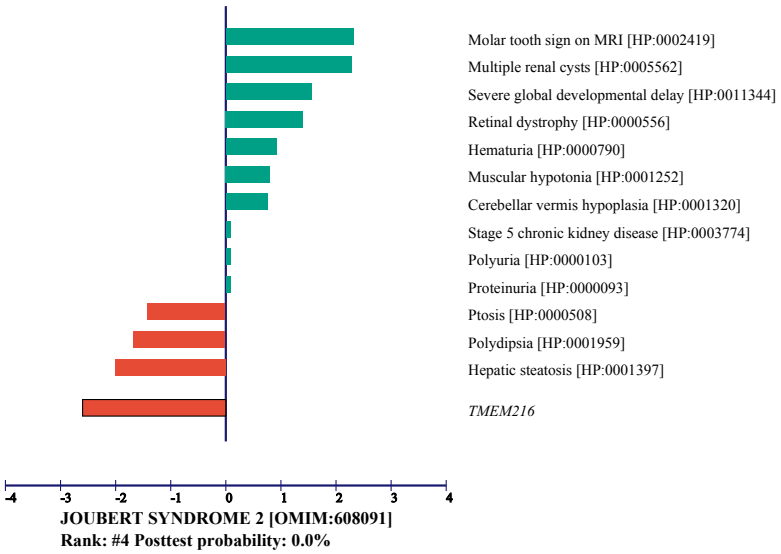


Figure S12. Assessment of diseases based only on clinical criteria. In this example, a case of Arima syndrome is simulated based on case 1 in a report on the clinicopathological features of the renal disease in Arima syndrome [3]. Arima syndrome shares many phenotypic features with Joubert syndrome. **(a)** In the simulated case using the control VCF file (without spiking in any pathogenic variant), Arima syndrome was correctly ranked in first place. **(b)** A type of Joubert syndrome was ranked in fourth place. No pathogenic alleles were identified in the causative gene *TMEM216*, which reduced the likelihood ratio (red bar corresponding to *TMEM216*).

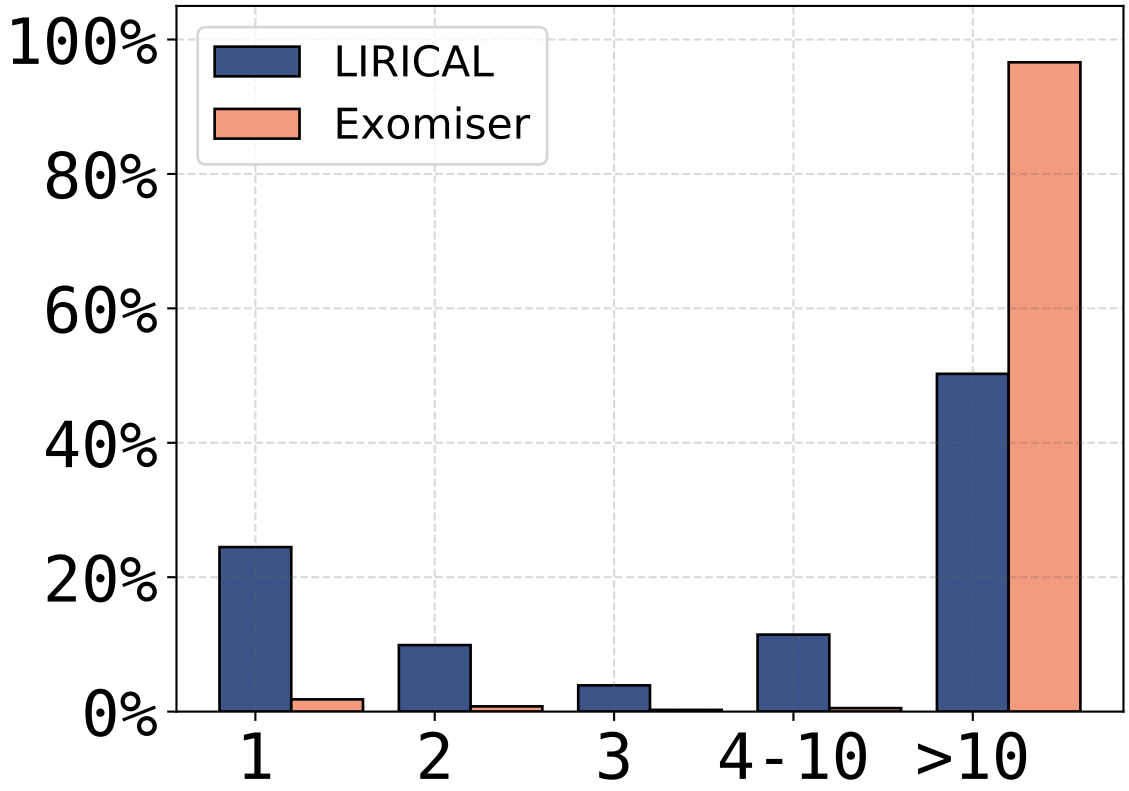


Figure S13. Rankings with all pathogenic alleles removed. Performance of LIRICAL (blue) and Exomiser (orange) on 384 case reports from which all pathogenic alleles have been removed from the VCF file. LIRICAL placed the correct candidate in the first ten ranks in 49.7% of cases, while Exomiser placed 4 of 384 candidates in rank 1 and failed to rank any of the other candidates. The X axis shows the rank assigned by LIRICAL or Exomiser to the correct disease gene. The Y axis shows the percentage of cases in which the given rank was achieved.

Gene	Frequency	Associated disease
TTN (7273)	9.46	CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 9 (OMIM:613765)
HLA-DRB1 (3123)	9.29	SARCOIDOSIS, SUSCEPTIBILITY TO, 1 (OMIM:181000)
KRT18 (3875)	7.25	CIRRHOSIS, FAMILIAL (OMIM:215600)
FLG (2312)	5.98	DERMATITIS, ATOPIC, 2 (OMIM:605803)
NEB (4703)	5.29	NEMALINE MYOPATHY 2 (OMIM:256030)
MUC5B (727897)	4.99	PULMONARY FIBROSIS, IDIOPATHIC (OMIM:178500)
HLA-DQB1 (3119)	4.47	CELIAC DISEASE, SUSCEPTIBILITY TO, 1 (OMIM:212750)
SYNE2 (23224)	3.94	EMERY-DREIFUSS MUSCULAR DYSTROPHY 5, AUTOSOMAL DOMINANT (OMIM:612999)
SYN2 (6854)	3.71	SCHIZOPHRENIA (OMIM:181500)
RP1L1 (94137)	3.53	OCCULT MACULAR DYSTROPHY (OMIM:613587)
DSPP (1834)	3.38	DEAFNESS, AUTOSOMAL DOMINANT 39, WITH DENTINOGENESIS IMPERFECTA 1 (OMIM:605594)
FSIP2 (401024)	3.14	SPERMATOGENIC FAILURE 34 (OMIM:618153)
SCARF2 (91179)	3.11	VAN DEN ENDE-GUPTA SYNDROME (OMIM:600920)
ARMC9 (80210)	3.04	JOUBERT SYNDROME 30 (OMIM:617622)
DNAH11 (8701)	3.00	CILIARY DYSKINESIA, PRIMARY, 7 (OMIM:611884)
KMT2C (58508)	2.96	KLEEFSTRA SYNDROME 2 (OMIM:617768)
HLA-DQA1 (3117)	2.96	CELIAC DISEASE, SUSCEPTIBILITY TO, 1 (OMIM:212750)
EYS (346007)	2.87	RETINITIS PIGMENTOSA 25 (OMIM:602772)
HPS4 (89781)	2.73	HERMANSKY-PUDLAK SYNDROME 4 (OMIM:614073)
ALMS1 (7840)	2.54	ALSTROM SYNDROME (OMIM:203800)
FAT2 (2196)	2.47	SPINOCEREBELLAR ATAXIA 45 (OMIM:617769)
PIEZO1 (9780)	2.41	LYMPHATIC MALFORMATION 6 (OMIM:616843)
DST (667)	2.41	EPIDERMOLYSIS BULLOSA SIMPLEX, AUTOSOMAL RECESSIVE 2 (OMIM:615425)
ACAN (176)	2.37	SPONDYLOEPIMETAPHYSEAL DYSPLASIA, AGGRECAN TYPE (OMIM:612813)
HNF1A (6927)	2.37	DIABETES MELLITUS, INSULIN-DEPENDENT, 20 (OMIM:612520)
TNXB (7148)	2.35	VESICOURETERAL REFLUX 8 (OMIM:615963)
TRIOBP (11078)	2.33	DEAFNESS, AUTOSOMAL RECESSIVE 28 (OMIM:609823)
ISCU (23479)	2.29	MYOPATHY WITH LACTIC ACIDOSIS, HEREDITARY (OMIM:255125)
SON (6651)	2.21	ZTTK SYNDROME (OMIM:617140)
ADGRV1 (84059)	2.19	USHER SYNDROME, TYPE IIC (OMIM:605472)
TREH (11181)	2.16	TREHALASE DEFICIENCY (OMIM:612119)
SERPINA1 (5265)	2.11	ALPHA-1-ANTITRYPSIN DEFICIENCY (OMIM:613490)
FRRS1L (23732)	2.02	EPILEPTIC ENCEPHALOPATHY, EARLY INFANTILE, 37 (OMIM:616981)
FRG1 (2483)	2.01	FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY 1 (OMIM:158900)
CTU2 (348180)	1.95	MICROCEPHALY, FACIAL DYSMORPHISM, RENAL AGENESIS, AND AMBIGUOUS GENITALIA SYNDROME (OMIM:618142)
KRT13 (3860)	1.79	WHITE SPONGE NEVUS 2 (OMIM:615785)
STXBP2 (6813)	1.79	HEMOPHAGOCYTIC LYMPHOHISTIOCYTOSIS, FAMILIAL, 5 (OMIM:613101)
GEMIN4 (50628)	1.75	NEURODEVELOPMENTAL DISORDER WITH MICROCEPHALY, CATARACTS, AND RENAL ABNORMALITIES (OMIM:617913)
DUOX2 (50506)	1.75	THYROID DYSHORMONOGENESIS 6 (OMIM:607200)
A2ML1 (144568)	1.75	OTITIS MEDIA, SUSCEPTIBILITY TO (OMIM:166760)
APOL1 (8542)	1.71	FOCAL SEGMENTAL GLOMERULOSCLEROSIS 4, SUSCEPTIBILITY TO (OMIM:612551)
MYO5B (4645)	1.71	DIARRHEA 2, WITH MICROVILLUS ATROPHY (OMIM:251850)
TMEM216 (51259)	1.70	JOUBERT SYNDROME 2 (OMIM:608091)
LTBP4 (8425)	1.69	CUTIS LAXA, AUTOSOMAL RECESSIVE, TYPE IC (OMIM:613177)
PCLO (27445)	1.64	PONTOCEREBELLAR HYPOPLASIA, TYPE 3 (OMIM:608027)
KIZ (55857)	1.64	RETINITIS PIGMENTOSA 69 (OMIM:615780)
VCAN (1462)	1.61	WAGNER VITREORETINOPATHY (OMIM:143200)
VPS13B (157680)	1.61	COHEN SYNDROME (OMIM:216550)
RAI1 (10743)	1.60	SMITH-MAGENIS SYNDROME (OMIM:182290)
VWA3B (200403)	1.60	SPINOCEREBELLAR ATAXIA, AUTOSOMAL RECESSIVE 22 (OMIM:616948)
DHFR (1719)	1.58	MEGALOBLASTIC ANEMIA DUE TO DIHYDROFOLATE REDUCTASE DEFICIENCY (OMIM:613839)

Table S1. The 50 Mendelian disease-associated genes with the highest sum of population frequencies of called pathogenic variants.

Table S2. Phenopackets analyzed in this work.

Disease	Gene	Proband	n. HPO terms	Publication
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 9	84	PMID:27087320
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	6-2	5	PMID:27040691
Ectodermal Dysplasia 1, Hypohidrotic, X-Linked	EDA	proband	9	PMID:18702659
Deafness, Autosomal Recessive 7	TMC1	935-IV:1	2	PMID:18616530
Osteogenesis Imperfecta, Type Xiv	TMEM38B	family2-patient2	9	PMID:26911354
Cutis Laxa, Autosomal Recessive, Type Iic	ATP6V1E1	Family 5 - IV:2	4	PMID:27023906
Codas syndrome	LONP1	Proband	13	PMID:28148925
Thrombocythemia 2	MPL	FT2:VI:3	1	PMID:19036112
Parkinson Disease 23, Autosomal Recessive, Early Onset	VPS13C	VPS13C case	11	PMID:28862745
Nemaline Myopathy 4	TPM2	1A	5	PMID:23378224
Noonan syndrome 3	KRAS	Patient 2	14	PMID:17056636
Spinocerebellar Ataxia, Autosomal Recessive 20	SNX14	IV-1	18	PMID:30473892
Cleidocranial Dysplasia	RUNX2	Family-A-II1	19	PMID:31548836
Epileptic Encephalopathy, Early Infantile, 28	WWOX	Patient 1	18	PMID:27495153
Congenital Disorder Of Glycosylation, Type Iip	TMEM199	F1-II2	11	PMID:26833330
Loeys-Dietz syndrome 1	TGFBR1	patient	18	PMID:30701076
Ataxia-Pancytopenia syndrome	SAMD9L	P5	2	PMID:29217778
Branchiooculofacial syndrome	TFAP2A	10-year-old girl	13	PMID:20461149
Lowe Oculocerebrorenal syndrome	OCRL	Patient 1	8	PMID:29300302
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	Fam1Pat1	7	PMID:29379883
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 7	88	PMID:27087320
Cerebral Creatine Deficiency syndrome 1	SLC6A8	proband	7	PMID:30400883
Epileptic Encephalopathy, Early Infantile, 14	KCNT1	Patient-1	5	PMID:24029078
Cutis Laxa, Autosomal Recessive, Type Iid	ATP6V1A	PV:1	11	PMID:28065471
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	index	4	PMID:28566166
Cohen syndrome	VPS13B	proposita	18	PMID:29149870
Combined Oxidative Phosphorylation Deficiency 30	TRMT10C	Subject 1	18	PMID:27132592
Cutis Laxa, Autosomal Recessive, Type Iic	ATP6V1E1	PII:1	13	PMID:28065471
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	3-1	15	PMID:27040691
Nijmegen Breakage syndrome	NBN	12-year-old girl	18	PMID:24044622
Microcephaly 6, Primary, Autosomal Recessive	CENPJ	IV-5	7	PMID:16900296
Spondylocostal Dysostosis 1, Autosomal Recessive	DLL3	II.6	6	PMID:15200511
Microcephaly 3, Primary, Autosomal Recessive	CDK5RAP2	patient	6	PMID:23726037
Aarskog-Scott syndrome	FGD1	II-1	10	PMID:23211637
Bardet-Biedl syndrome 4	BBS4	4-year-old female patient	10	PMID:25533820
Muscular Dystrophy, Limb-Girdle, Type 2z	POGLUT1	Patient II.1	13	PMID:27807076
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 7	38	PMID:29330883

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Mental Retardation, Autosomal Dominant 42	GNB1	proband	10	PMID:29174093
Ataxia-Pancytopenia syndrome	SAMD9L	UB085	12	PMID:29146883
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 8	37	PMID:29330883
Cornelia De Lange syndrome 1	NIPBL	Patient 1	14	PMID:25447906
Tietz syndrome	MITF	family 815	6	PMID:10851256
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 6	13	PMID:29330883
Papillon-Lefevre syndrome	CTSC	Case 1P1	6	PMID:23311634
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 3	35	PMID:29330883
Townes-Brocks syndrome	SALL1	VMFS	23	PMID:29110636
Retinitis Pigmentosa 18	PRPF3	020001-II:4	4	PMID:27886254
Ataxia-Pancytopenia syndrome	SAMD9L	UB081	7	PMID:29146883
Bernard-Soulier syndrome	GP1BA	Patient 3	10	PMID:26044173
Ehlers-Danlos syndrome, Classic Type	COL5A1	AN-002501	9	PMID:23587214
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 10	6	PMID:27087320
Retinitis Pigmentosa With Or Without Skeletal Anomalies	CWC27	II-4	11	PMID:28285769
Metabolic Encephalomyopathic Crises, Recurrent, With Rhabdomyolysis, Cardiac Arrhythmias, And Neurodegeneration	TANGO2	Subject 5	21	PMID:26805781
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 8	93	PMID:27087320
Hajdu-Cheney syndrome	NOTCH2	proband	12	PMID:23566664
Retinitis Pigmentosa 11	PRPF31	IV:3	5	PMID:30099644
Intellectual Developmental Disorder With Dysmorphic Facies And Ptosis	BRPF1	Individual 11/Family F	11	PMID:27939639
Treacher Collins syndrome 2	POLR1D	family 1:patient	4	PMID:24603435
Amyloidosis, Finnish Type	GSN	III:5	6	PMID:26915616
Legius syndrome	SPRED1	P62	2	PMID:28150585
Neuropathy, Hereditary Sensory And Autonomic, Type Iib	RETREG1	F2:IV:1	8	PMID:30643655
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	Fam2Pat1	7	PMID:29379883
Myhre syndrome	SMAD4	patient	18	PMID:24715504
Thrombocytopenia 3	FYB1	IV:5	4	PMID:25516138
Homocystinuria Due To Cystathionine Beta-Synthase Deficiency	CBS	patient	4	PMID:8755636
Albinism, Oculocutaneous, Type Iii	TYRP1	patient 2	3	PMID:21739261
Rett syndrome, Congenital Variant	FOXG1	Patient 2	11	PMID:28851325
Emery-Dreifuss Muscular Dystrophy 3, Autosomal Recessive	LMNA	II3	12	PMID:23313286
Spastic Ataxia 8, Autosomal Recessive, With Hypomyelinating Leukodystrophy	NKX6-2	IV-6	4	PMID:28575651
Oliver-Mcfarlane syndrome	PNPLA6	18 year-old female	17	PMID:30097146
Metabolic Encephalomyopathic Crises, Recurrent, With Rhabdomyolysis, Cardiac Arrhythmias, And Neurodegeneration	TANGO2	F1:II.2	23	PMID:26805782

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Ehlers-Danlos syndrome, Classic-Like, 2	AEBP1	AN-006205	23	PMID:30759870
Gm1-Gangliosidosis, Type Iii	GLB1	KT	6	PMID:1907800
Hyperoxaluria, Primary, Type Ii	GRHPR	patient	11	PMID:28569194
Bethlem Myopathy 1	COL6A1	II.1	21	PMID:30808312
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	Patient 2	8	PMID:28148688
Ehlers-Danlos syndrome, Musculocontractural Type 1	CHST14	3-year old boy	16	PMID:30249733
Stankiewicz-Isidor syndrome	PSMD12	Subject 2	30	PMID:28132691
Marfan syndrome	FBN1	B15	7	PMID:11175294
Nemaline Myopathy 3	ACTA1	Patient 5	10	PMID:30517146
Fanconi Anemia, Complementation Group C	FANCC	proband	7	PMID:22701786
Autoimmune Polyendocrine syndrome, Type I, With Or Without Reversiblemetaphyseal Dysplasia	AIRE	V-1	10	PMID:28540407
Noonan syndrome 6	NRAS	case 1	15	PMID:26467218
Mental Retardation, Autosomal Recessive 38	HERC2	Pedigree 1A,VIII:8	9	PMID:23243086
Marfan syndrome	FBN1	Patient 2	11	PMID:30101859
Retinitis Pigmentosa With Or Without Skeletal Anomalies	CWC27	3:II-1	2	PMID:28285769
Cockayne syndrome B	ERCC6	index	18	PMID:30113454
Neuropathy, Hereditary Sensory And Autonomic, Type Iia	WNK1	Patient	13	PMID:16636245
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	A-II-1	28	PMID:27040692
Elliptocytosis 2	SPTA1	proband	10	PMID:29484404
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	II-4	5	PMID:29678961
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	B-IV-6	16	PMID:27040692
Homocystinuria Due To Cystathionine Beta-Synthase Deficiency	CBS	III:3	12	PMID:26667307
Ataxia-Pancytopenia syndrome	SAMD9L	UB049	7	PMID:29146883
Waardenburg syndrome, Type 3	PAX3	proposita	8	PMID:12949970
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	D:IV-1	23	PMID:28132690
Osteogenesis Imperfecta, Type Xi	FKBP10	proband	9	PMID:29801479
Retinitis Pigmentosa 27	NRL	II:2	4	PMID:28106895
Cutis Laxa, Autosomal Recessive, Type Ia	FBLN5	4-year-old Burmese girl	12	PMID:24962763
Rubinstein-Taybi syndrome 2	EP300	11	26	PMID:29506490
Amelogenesis Imperfecta, Type Ij	ACP4	Family 1-IV:3	2	PMID:28513613
Osteogenesis Imperfecta, Type Viii	P3H1	proband	4	PMID:27864101
Cornelia De Lange syndrome 3	SMC3	patient 1	23	PMID:28781842
3-@methylglutaconic Aciduria With Deafness, Encephalopathy, And Leigh-Likesyndrome	SERAC1	proband	23	PMID:31251474
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 2	84	PMID:27087320
Geleophysic Dysplasia 1	ADAMTSL2	patient	16	PMID:27057656

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Parkinson Disease 23, Autosomal Recessive, Early Onset	VPS13C	Family B, II-1	5	PMID:26942284
Spastic Ataxia 8, Autosomal Recessive, With Hypomyelinating Leukodystrophy	NKX6-2	Patient 4 II-1	10	PMID:28969374
Myopathy, Distal, Tateyama Type	CAV3	I1	16	PMID:18930476
Ataxia-Pancytopenia syndrome	SAMD9L	III-1	10	PMID:28202457
Stankiewicz-Isidor syndrome	PSMD12	Subject 3	12	PMID:28132691
Arthrogryposis, Distal, Type 2a	MYH3	proband	13	PMID:28584669
Polymicrogyria With Seizures	RTTN	Patient 3	10	PMID:29883675
Cutis Laxa, Autosomal Recessive, Type Iid	ATP6V1A	PIV:1	19	PMID:28065471
Glycogen Storage Disease Vi	PYGL	2-year 5-month old child	14	PMID:28984260
Polyarteritis Nodosa, Childhood-Onset	ADA2	patient 1	13	PMID:28830446
Bardet-Biedl syndrome 1	BBS1	IV-5/family A	7	PMID:23559858
Arthrogryposis, Distal, With Impaired Proprioception And Touch	PIEZO2	Patient	12	PMID:27974811
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	2-1	12	PMID:27040691
Severe Combined Immunodeficiency, Autosomal Recessive, T Cell-Negative,b Cell-Negative, Nk Cell-Negative, Due To Adenosine Deaminase Deficiency	ADA	Patient	6	PMID:1680289
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	Patient 2	16	PMID:30103036
Spastic Ataxia 8, Autosomal Recessive, With Hypomyelinating Leukodystrophy	NKX6-2	Patient 36-16DG1123	5	PMID:28940097
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 5	31	PMID:29330883
Structural Heart Defects And Renal Anomalies syndrome	TMEM260	1-II-1	23	PMID:28318500
Cone-Rod Dystrophy 2	CRX	IV:5	4	PMID:30095615
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	B-IV-4	24	PMID:27040692
Smith-Lemli-Opitz syndrome	DHCR7	patient	13	PMID:28503313
Congenital Disorder Of Glycosylation, Type Il	ALG9	IV:5	16	PMID:26453364
Nephrotic syndrome, Type 1	NPHS1	patient 1	9	PMID:28392951
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 3	96	PMID:27087320
Acromesomelic Dysplasia, Maroteaux Type	NPR2	IV-2/family-A	10	PMID:25959430
Ayme-Gripp syndrome	MAF	patient CSA108.01	1	PMID:28482824
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	SAL-399-073	7	PMID:27320912
Geleophysic Dysplasia 2	FBN1	Family 1, Patient 1	14	PMID:29191498
Robinow syndrome, Autosomal Recessive	ROR2	Patient 1	20	PMID:24932600
Parkinson Disease 23, Autosomal Recessive, Early Onset	VPS13C	Family C, II-1	4	PMID:26942284
Wiedemann-Steiner syndrome	KMT2A	P1	16	PMID:25186178

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Diarrhea 8, Secretory Sodium, Congenital	SLC9A3	Patient 9	9	PMID:26358773
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	Index	7	PMID:28321562
Retinitis Pigmentosa With Or Without Skeletal Anomalies	CWC27	4:II-3	14	PMID:28285769
Spastic Paraplegia 7, Autosomal Recessive	SPG7	II-3	13	PMID:17646629
Hyaline Fibromatosis syndrome	ANTXR2	II-3	13	PMID:30050362
Cleidocranial Dysplasia	RUNX2	Family-B-II1	19	PMID:31548836
Heterotaxy, Visceral, 1, X-Linked	ZIC3	III-1	12	PMID:9354794
Autoimmune Lymphoproliferative syndrome	FASLG	patient	14	PMID:22857792
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	E:II-1	20	PMID:28132690
Muenke syndrome	FGFR3	Proband 27	5	PMID:26740388
Congenital Disorder Of Glycosylation, Type Iip	TMEM199	Patient 1	7	PMID:29321044
Marfan syndrome	FBN1	Patient 1	4	PMID:30101859
Mental Retardation, Autosomal Dominant 7	DYRK1A	Patient 2	19	PMID:26922654
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	Patient 1	50	PMID:28331220
Van Den Ende-Gupta syndrome	SCARF2	proband	17	PMID:29378527
Bartter syndrome, Type 4a	BSND	family-A-III3	9	PMID:18776122
Loeys-Dietz syndrome 3	SMAD3	54-year old woman	2	PMID:28286188
Holoprosencephaly 5	ZIC2	proband	3	PMID:30855487
Epidermolysis Bullosa, Junctional, Herlitz Type	LAMC2	patient	5	PMID:24533970
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 6	78	PMID:27087320
Apert syndrome	FGFR2	Patient 1	14	PMID:23546041
Stankiewicz-Isidor syndrome	PSMD12	Subject 4	21	PMID:28132691
Myasthenic syndrome, Congenital, 8	AGRN	Patient 3/Kinship 2	15	PMID:24951643
Donohue syndrome	INSR	ISR1	14	PMID:24498630
Cornelia De Lange syndrome 1	NIPBL	Patient 2	10	PMID:25447906
Microcephaly 5, Primary, Autosomal Recessive	ASPM	patient	10	PMID:29644084
Hypothyroidism, Thyroidal Or Athyroidal, With Spiky Hair And Cleftpalate	FOXE1	patient	7	PMID:24219130
Fanconi Anemia, Complementation Group I	FANCI	NCI-309-1	9	PMID:26590883
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	1-1	20	PMID:27040691
Camurati-Engelmann Disease	TGFB1	patient	13	PMID:30034812
Bernard-Soulier syndrome	GP1BA	73 year old male	5	PMID:9233564
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	AII-1	14	PMID:28132690
Galloway-Mowat syndrome 4	TP53RK	II-1	10	PMID:30053862
Leukocyte Adhesion Deficiency, Type I	ITGB2	P1	4	PMID:26497373
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	Family B-IV:1	7	PMID:27153400
Ataxia-Pancytopenia syndrome	SAMD9L	II-4	13	PMID:28202457
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	8-1	14	PMID:27040691
Trichothiodystrophy 3, Photosensitive	GTF2H5	male infant	27	PMID:30359777

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Deafness, Autosomal Recessive 15	GIPC3	Ahv-14:23	1	PMID:29605370
Galactosemia	GALT	FKT118	7	PMID:25681079
Vici syndrome	EPG5	18-month son	15	PMID:29983806
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	B:II-2	28	PMID:28132690
Sick Sinus syndrome 2, Autosomal Dominant	HCN4	family A/II:1	3	PMID:25145518
Charcot-Marie-Tooth Disease, Demyelinating, Type 1c	LITAF	Proband	14	PMID:19541485
Chudley-Mccullough syndrome	GPSM2	case 1	13	PMID:27180139
Schinzel-Giedion Midface Retraction syndrome	SETBP1	proposita	26	PMID:29333303
Orofaciodigital syndrome V	DDX59	Patient 1	11	PMID:29127725
Ventricular Tachycardia, Catecholaminergic Polymorphic, 1, With Orwithout Atrial Dysfunction And/or Dilated Cardiomyopathy	RYR2	proband	6	PMID:30296944
Long Qt syndrome 15	CALM2	Case 1	4	PMID:27374306
Cleidocranial Dysplasia	RUNX2	Family-D-II1	19	PMID:31548836
Renal Cysts And Diabetes syndrome	HNF1B	patient	6	PMID:29491316
Ataxia-Pancytopenia syndrome	SAMD9L	II-4	6	PMID:27259050
Acromicric Dysplasia	FBN1	patient	17	PMID:27834076
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 2	22	PMID:29330883
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	II-4	16	PMID:27275012
Intellectual Developmental Disorder With Dysmorphic Facies, Seizures, And Distal Limb Anomalies	OTUD6B	proband	14	PMID:30364145
Spastic Ataxia 8, Autosomal Recessive, With Hypomyelinating Leukodystrophy	NKX6-2	Patient 3 II-3	9	PMID:28969374
Fibrodysplasia Ossificans Progressiva	ACVR1	patient	10	PMID:29482508
Neurodegeneration With Brain Iron Accumulation 1	PANK2	Family I patient I	7	PMID:28821231
Al Kaissi syndrome	CDK10	F1-II:1	20	PMID:28886341
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 5	92	PMID:27087320
Hypotrichosis, Congenital, With Juvenile Macular Dystrophy	CDH3	Patient	14	PMID:28061825
Epileptic Encephalopathy, Early Infantile, 4	STXBP1	P1	6	PMID:29896790
Myopathy, Centronuclear, 1	DNM2	Patient 1	12	PMID:24465259
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	C-II-1	26	PMID:27040692
Apert syndrome	FGFR2	Patient 2	16	PMID:23546041
Kufor-Rakeb syndrome	ATP13A2	Case 1	12	PMID:30746398
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	4-2	4	PMID:27040691
Ichthyosis, Congenital, Autosomal Recessive 11	ST14	patient	7	PMID:18445049
Alzheimer Disease 4	PSEN2	proband	3	PMID:30104866
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	Patient 3	28	PMID:28148688
Congenital Disorder Of Glycosylation, Type Iip	TMEM199	F2-II2	17	PMID:26833330

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	4-1	10	PMID:27040691
Tuberous Sclerosis 2	TSC2	III-1	4	PMID:8825048
Osteogenesis Imperfecta, Type Ix	PPIB	second fetus	5	PMID:28242392
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	Tun66275	3	PMID:27320912
Chitayat syndrome	ERF	proband	17	PMID:30569521
Charge syndrome	CHD7	Patient A III-2	14	PMID:17661815
Cholestasis, Progressive Familial Intrahepatic, 4	TJP2	proband	17	PMID:30658709
Congenital Disorder Of Glycosylation, Type Iip	TMEM199	Patient 3	6	PMID:29321044
Osteogenesis Imperfecta, Type Xii	SP7	II:5	16	PMID:29382611
Ectodermal Dysplasia 9, Hair/nail Type	HOXC13	IV-1	7	PMID:28403827
Diamond-Blackfan Anemia 1	RPS19	patient	5	PMID:27732904
Spinal Muscular Atrophy With Progressive Myoclonic Epilepsy	ASAH1	patient	13	PMID:31213928
Cutis Laxa, Autosomal Recessive, Type Iib	PYCR1	Patient 4	16	PMID:21487760
Intellectual Developmental Disorder With Dysmorphic Facies And Behavioral Abnormalities	FBXO11	Individual 1	26	PMID:30057029
Nemaline Myopathy 1	TPM3	II.2	20	PMID:24239060
Skraban-Deardorff syndrome	WDR26	Individual 1, PPMD01P, GEA055P	53	PMID:28686853
Stankiewicz-Isidor syndrome	PSMD12	Subject 1	34	PMID:28132691
Myasthenic syndrome, Congenital, 9, Associated With Acetylcholinereceptor Deficiency	MUSK	patient	17	PMID:23326516
Neurodevelopmental Disorder With Progressive Microcephaly, Spasticity, And Brain Anomalies	PLAA	Family A-IV:6	22	PMID:28413018
Peutz-Jeghers syndrome	STK11	20-year-old woman	3	PMID:15200509
Structural Heart Defects And Renal Anomalies syndrome	TMEM260	2-II-4	19	PMID:28318500
Spherocytosis, Type 4	SLC4A1	c.1432-2A¿T	3	PMID:23255290
Spastic Ataxia 8, Autosomal Recessive, With Hypomyelinating Leukodystrophy	NKX6-2	III-1	19	PMID:28575651
Multiple Endocrine Neoplasia, Type I	MEN1	III-3	15	PMID:26239674
Hyper-Ige Recurrent Infection syndrome, Autosomal Dominant	STAT3	12 year old girl	12	PMID:20149460
Stickler syndrome, Type Ii	COL11A1	proband	9	PMID:28971234
Congenital Disorder Of Glycosylation, Type Iip	TMEM199	Patient 2	7	PMID:29321044
Spastic Paraplegia 45, Autosomal Recessive	NT5C2	II.3	13	PMID:28327087
Werner syndrome	WRN	48-year-old male	18	PMID:30891318
Alagille syndrome 1	JAG1	Proband	18	PMID:30046498
Corneal Dystrophy, Fuchs Endothelial, 4	SLC4A11	Patient 1	2	PMID:25007886
Parkinson Disease 23, Autosomal Recessive, Early Onset	VPS13C	Family A, V-2	18	PMID:26942284
Congenital Disorder Of Glycosylation, Type Iip	TMEM199	F3-II1	12	PMID:26833330
Epidermolysis Bullosa, Junctional, Herlitz Type	LAMA3	Proband	2	PMID:20881434
Mucolipidosis Ii Alpha/beta	GNPTAB	proband	14	PMID:30208878
Combined Oxidative Phosphorylation Deficiency 30	TRMT10C	Subject 2	15	PMID:27132592

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Stickler syndrome, Type I	COL2A1	proband	19	PMID:28841907
Myotonia Congenita, Autosomal Dominant	CLCN1	man	7	PMID:30243293
Spondyloepimetaphyseal Dysplasia With Joint Laxity, Type 1, With Orwithout Fractures	B3GALT6	P7/F6	29	PMID:23664117
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	6-1	13	PMID:27040691
Tuberous Sclerosis 1	TSC1	II:3/Family 2	7	PMID:18830229
Weill-Marchesani syndrome 1	ADAMTS10	18-year-old woman	9	PMID:25469541
Megalocornea	CHRDL1	III-1	4	PMID:24073597
Hyperuricemic Nephropathy, Familial Juvenile, 1	UMOD	proband	6	PMID:15673476
Cutis Laxa, Autosomal Recessive, Type Iid	ATP6V1A	PIII:1	14	PMID:28065471
Spastic Ataxia 8, Autosomal Recessive, With Hypomyelinating Leukodystrophy	NKX6-2	F6,II-2	24	PMID:29388673
Tuberous Sclerosis 1	TSC1	patient 6	7	PMID:29196670
Retinitis Pigmentosa 78	ARHGEF18	Individual 1	8	PMID:28132693
Amelogenesis Imperfecta, Type Ia	LAMB3	proband	2	PMID:27220909
Bardet-Biedl syndrome 5	BBS5	II:2	6	PMID:30850397
Bleeding Disorder, Platelet-Type, 17	GFI1B	II:6	5	PMID:30655368
Nemaline Myopathy 7	CFL2	Patient 1	12	PMID:22560515
Neurodevelopmental Disorder With Progressive Microcephaly, Spasticity, And Brain Anomalies	PLAA	A-VI3	17	PMID:28007986
Bardet-Biedl syndrome 2	BBS2	II:2	9	PMID:26078953
Neurofibromatosis, Type I	NF1	0548	8	PMID:9101303
Gapo syndrome	ANTXR1	14 year old brother	11	PMID:27587992
Charcot-Marie-Tooth Disease, Axonal, Type 2a2	MFN2	patient	11	PMID:26956144
Platelet Disorder, Familial, With Associated Myeloid Malignancy	RUNX1	Pedigree I, V:2	3	PMID:28181366
Trichohepatoenteric syndrome 1	TTC37	index	17	PMID:28292286
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	Patient 1	30	PMID:30103036
Glutaric Acidemia I	GCDH	Patient 5	5	PMID:27672653
Choreoacanthocytosis	VPS13A	Patient-2	9	PMID:28446873
Ataxia-Pancytopenia syndrome	SAMD9L	P7	2	PMID:29217778
Albinism, Oculocutaneous, Type Ii	OCA2	B	4	PMID:29050284
Cockayne syndrome A	ERCC8	Patient A	7	PMID:30200888
Pseudoachondroplasia	COMP	patient	17	PMID:23562786
Galactosialidosis	CTSA	BAB3767	13	PMID:24769197
Neurodevelopmental Disorder With Progressive Microcephaly, Spasticity, And Brain Anomalies	PLAA	Family D-Case VIII-1	14	PMID:28413018
Cardiomyopathy, Dilated, 1g	TTN	JK109	4	PMID:11846417
Joubert syndrome 30	ARMC9	UW132-3	5	PMID:28625504
Dyskeratosis Congenita, Autosomal Dominant 3	TINF2	proband	12	PMID:29742735
Temtamy Preaxial Brachydactyly syndrome	CHSY1	IV-1	16	PMID:24269551
Krabbe Disease	GALC	child	6	PMID:26567009
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 4	16	PMID:29330883
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	SAL-584-005	4	PMID:27320912

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Ataxia, Early-Onset, With Oculomotor Apraxia And Hypoalbuminemia	APTX	V-3	4	PMID:28652255
Retinitis Pigmentosa 78	ARHGEF18	Individual 2	8	PMID:28132693
Hypochondroplasia	FGFR3	VI-5	9	PMID:30681580
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	5-1	18	PMID:27040691
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	BII-1	26	PMID:28132690
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	Family A-V:2	8	PMID:27153400
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 1	88	PMID:27087320
Larsen syndrome	FLNB	patient	12	PMID:18322662
Muckle-Wells syndrome	NLRP3	proband	9	PMID:27435956
Leukocyte Adhesion Deficiency, Type Iii	FERMT3	index	4	PMID:31068971
Cardiofaciocutaneous syndrome 1	BRAF	CFC16	16	PMID:16474404
Ataxia-Pancytopenia syndrome	SAMD9L	UB612	3	PMID:29146883
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	Patient 1	10	PMID:28148688
Metabolic Encephalomyopathic Crises, Recurrent, With Rhabdomyolysis, Cardiac Arrhythmias, And Neurodegeneration	TANGO2	Subject 6	17	PMID:26805781
Boucher-Neuhauser syndrome	PNPLA6	II.2	8	PMID:29749493
Nail-Patella syndrome	LMX1B	index	6	PMID:30881852
Neurodegeneration With Brain Iron Accumulation 2b	PLA2G6	family II patient II	8	PMID:28821231
Osteogenesis Imperfecta, Type Xv	WNT1	proband	11	PMID:30012084
Spastic Paraplegia 10, Autosomal Dominant	KIF5A	proband	12	PMID:30057544
Palmoplantar Keratoderma, Epidermolytic	KRT9	III:4	3	PMID:18477167
Cerebral Dysgenesis, Neuropathy, Ichthyosis, And Palmoplantar Keratodermasyndrome	SNAP29	The patient	19	PMID:29051910
Metabolic Encephalomyopathic Crises, Recurrent, With Rhabdomyolysis, Cardiac Arrhythmias, And Neurodegeneration	TANGO2	Suject 1	20	PMID:26805781
Spastic Ataxia 8, Autosomal Recessive, With Hypomyelinating Leukodystrophy	NKX6-2	Patient 1 II-1	13	PMID:28969374
Hyperekplexia, Hereditary 1	GLRA1	proband	5	PMID:24969041
Rett syndrome, Congenital Variant	FOXG1	Patient 4	9	PMID:28851325
Loeys-Dietz syndrome 4	TGFB2	proposita	15	PMID:25163805
Smith-Magenis syndrome	RAI1	SMS324	25	PMID:20932317
Metabolic Encephalomyopathic Crises, Recurrent, With Rhabdomyolysis, Cardiac Arrhythmias, And Neurodegeneration	TANGO2	Subject 4	16	PMID:26805781
Parkinson Disease 15, Autosomal Recessive Early-Onset	FBXO7	ANK-07	7	PMID:25085748
Alpha-Thalassemia/mental Retardation syndrome, X-Linked	ATRX	Proband	9	PMID:28371217
Rett syndrome, Congenital Variant	FOXG1	Patient 1	12	PMID:28851325
Smith-Kingsmore syndrome	MTOR	index	9	PMID:27753196
Trichorhinophalangeal syndrome, Type I	TRPS1	girl	4	PMID:23691375

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Holoprosencephaly 4	TGIF1	male proband	7	PMID:16962354
Candidiasis, Familial, 2	CARD9	Patient	8	PMID:26044242
Megaloblastic Anemia 1	AMN	III:1	3	PMID:26040326
Desmosterolosis	DHCR24	proband	34	PMID:29175559
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	Family C-IV:13	3	PMID:27153400
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	1-2	8	PMID:27040691
Larsen syndrome	FLNB	19	12	PMID:16801345
Toe Syndactyly, Telecanthus, And Anogenital And Renal Malformations	CCNQ	Case 2	14	PMID:18297069
Loeys-Dietz syndrome 2	TGFBR2	Patient 4	15	PMID:30101859
Poikiloderma With Neutropenia	USB1	patient	7	PMID:27247962
Neuropathy, Hereditary, With Liability To Pressure Palsies	PMP22	Proband	7	PMID:29078790
Pierpont syndrome	TBL1XR1	seven year old male	29	PMID:28687524
Hemophagocytic Lymphohistiocytosis, Familial, 2	PRF1	8-year-old boy	6	PMID:28468610
Jervell And Lange-Nielsen syndrome 1	KCNQ1	family III-IV-5	4	PMID:29037160
Niemann-Pick Disease, Type C1	NPC1	The proband	14	PMID:27900365
Spherocytosis, Type 5	EPB42	proposita	5	PMID:7803799
Cutis Laxa, Autosomal Recessive, Type Iic	ATP6V1E1	PI:1	13	PMID:28065471
Multiple Endocrine Neoplasia, Type Iia	RET	DM patient	3	PMID:24331334
Polymicrogyria, Symmetric Or Asymmetric	TUBB2B	proband	18	PMID:28966590
Ataxia-Pancytopenia syndrome	SAMD9L	IV-1	5	PMID:27259050
Metabolic Encephalomyopathic Crises, Recurrent, With Rhabdomyolysis, Cardiac Arrhythmias, And Neurodegeneration	TANGO2	Subject 2	33	PMID:26805781
Myasthenic syndrome, Congenital, 22	PREPL	proband	18	PMID:29483676
Gaucher Disease, Perinatal Lethal	GBA	boy weighing 1690 g	7	PMID:15967693
Kabuki syndrome 2	KMT2D	3 month old boy	21	PMID:30509212
Charge syndrome	CHD7	B III-3	14	PMID:17661815
Mental Retardation, Autosomal Recessive 18	MED23	IV.8	7	PMID:30847200
Citrullinemia, Classic	ASS1	5	8	PMID:23099195
Long Qt syndrome 14	CALM1	Case 2	4	PMID:27374306
Nance-Horan syndrome	NHS	III:1	9	PMID:30642278
Palmoplantar Keratoderma, Punctate Type Ia	AAGAB	family 1:proband	4	PMID:28239884
Mental Retardation, Autosomal Dominant 21	CTCF	proband	28	PMID:28619046
Ventricular Tachycardia, Catecholaminergic Polymorphic, 3	TECRL	Patient 1	8	PMID:27861123
Immunoskeletal Dysplasia With Neurodevelopmental Abnormalities	EXTL3	C:II-1	12	PMID:28132690
Parkinson Disease 7, Autosomal Recessive Early-Onset	PARK7	proband	13	PMID:27460976
Cockayne syndrome A	ERCC8	Patient C	5	PMID:30200888

Continued on next page

Table S2 – *Continued from previous page*

Disease	**Gene**	**Proband**	**n. HPO terms**	**Publication**
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	II-3	21	PMID:27275012
Myopathy, Myofibrillar, 3	MYOT	patient	5	PMID:19458539
Codas syndrome	LONP1	Patient 1	8	PMID:25808063
Rett syndrome, Congenital Variant	FOXG1	Patient 3	9	PMID:28851325
Retinitis Pigmentosa With Or Without Skeletal Anomalies	CWC27	1:II-3	12	PMID:28285769
Cockayne syndrome B	ERCC6	Patient B	5	PMID:30200888
Mucolipidosis Iv	MCOLN1	6 year old boy	8	PMID:28620732
Chediak-Higashi syndrome	LYST	patient	14	PMID:28183707
Marfan syndrome	FBN1	Patient 3	4	PMID:30101859
Congenital Disorder Of Glycosylation, Type Iih	COG8	proband	27	PMID:30690882
Pseudoachondroplasia	COMP	II-1	9	PMID:27330822
Polymicrogyria, Bilateral Frontoparietal	ADGRG1	Family A, II:2	4	PMID:29707406
Dyggve-Melchior-Clausen Disease	DYM	Patient 2	7	PMID:24300288
Arthrogryposis, Distal, Type 9	FBN2	IV:7	6	PMID:30147916
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 9	21	PMID:29330883
Smith-Magenis syndrome	RAI1	SMS335	15	PMID:20932317
Inclusion Body Myopathy With Early-Onset Paget Disease With Or Withoutfrontotemporal Dementia 1	VCP	II-3	11	PMID:19208399
Neurodevelopmental Disorder With Or Without Anomalies Of The Brain, Eye, Or Heart	RERE	Subject 4	87	PMID:27087320
Bleeding Disorder, Platelet-Type, 15	ACTN1	proband	5	PMID:24069336
Encephalopathy, Neonatal Severe, With Lactic Acidosis And Brain Abnormalities	LIPT2	P1	16	PMID:28757203
Cleidocranial Dysplasia	RUNX2	III:3	10	PMID:24966961
Congenital Disorder Of Glycosylation, Type Iic	SLC35C1	Proband 1	20	PMID:24403049
Rubinstein-Taybi syndrome 2	EP300	38	26	PMID:29506490
Craniofrontonasal syndrome	EFNB1	3269	17	PMID:23335590
Brugada syndrome 1	SCN5A	proband	3	PMID:31590245
Amyotrophic Lateral Sclerosis 1	SOD1	patient	7	PMID:30236613
Spastic Paraplegia 76, Autosomal Recessive	CAPN1	R-III:1	5	PMID:27320912
Hypotonia, Infantile, With Psychomotor Retardation And Characteristic Facies 3	TBCK	D-II-1	27	PMID:27040692
Ehlers-Danlos syndrome, Classic Type, 2	COL5A2	patient	8	PMID:27656288
Birt-Hogg-Dube syndrome	FLCN	253	2	PMID:96481
Diarrhea 3, Secretory Sodium, Congenital, With Or Without Other Congenitalanomalies	SPINT2	two-month-old male	13	PMID:29575628
Robinow syndrome, Autosomal Recessive	ROR2	Patient 2	19	PMID:24932600

`original`	unaltered data from case report
`noise2`	two "random" HPO terms added
`noise2`*	like noise2 but the original terms were replaced by a randomly chosen parent term
`noise2`**	like noise2 but the original terms were replaced by a randomly chosen grandparent term
`allele`$^{-2}$	remove all pathogenic alleles (i.e., remove one allele for dominant and two for recessive). Otherwise do not change the data
`allele`$^{-2,**}$	remove all pathogenic alleles (i.e., remove one allele for dominant and two for recessive), replace all terms with a parent term and then add two noise terms
`terms-randomized`	replace all HPO terms by "random" terms
`biallelic`	limit the case reports to those describing autosomal recessive (biallelic) diseases
`biallelic`$^{-1}$	same as `biallelic` but one of two pathogenic alleles is removed
`not`	10 cases in which a negated ("not") finding is important to the differential diagnosis (Table S4)
`not*`	same as `not`, but all negated terms are removed

Table S3. Approaches to add noise to the case report data (Phenopackets).

Correct diagnosis	Differential diagnosis	Differentiating feature
Loeys-Dietz syndrome 4 [OMIM:614816]	Marfan syndrome [OMIM:154700]	Ectopia lentis [HP:0001083]
Tietz albinism-deafness syndrome [OMIM:103500]	Waardenburg syndrome, type 2A [OMIM:193510]	Heterochromia iridis [HP:0001100]
Hypochondroplasia [OMIM:146000]	Achondroplasia [OMIM:100800]	Trident hand [HP:0004060]
Osteogenesis imperfecta, type XII [OMIM:613849]	Osteogenesis imperfecta, type IV [OMIM:166220]	Dentinogenesis imperfecta [HP:0000703]
Spinal muscular atrophy with progressive myoclonic epilepsy [OMIM:159950]	Spinal and bulbar muscular atrophy of Kennedy [OMIM:313200]	Elevated serum creatine kinase [HP:0003236]
Myotonia congenita, dominant [OMIM:160800]	Myotonic dystrophy 1 [OMIM:160900]	Muscle weakness [HP:0001324]
Trichorhinophalangeal syndrome, type I [OMIM:190350]	Trichorhinophalangeal syndrome, type II [OMIM:150230]	Intellectual disability [HP:0001249]
GM1-gangliosidosis, type III [OMIM:230650]	GM1-gangliosidosis, type I [OMIM:230500]	Cherry red spot of the macula [HP:0010729]
Megalocornea 1, X-linked [OMIM:309300]	Glaucoma 3, primary congenital, A [OMIM:231300]	Abnormal intraocular pressure [HP:0012632]
Ectodermal dysplasia 9, hair/nail type [OMIM:614931]	Ectodermal dysplasia 1, hypohidrotic, X-linked [OMIM:305100]	Abnormality of the dentition [HP:0000164]

Table S4. Pairs of diseases whose differential diagnosis is defined in part by the absence of the phenotypic abnormality listed in the third column. For instance, Loeys-Dietz syndrome 4 is noted not to be characterized by ectopia lentis, while the phenotypically similar disease Marfan syndrome is [4]. In each case, the disease in the first column is explicitly annotated not to have the phenotype in question, and the disease in the second column is annotated to have the feature. These ten cases are included in the 384 case reports (Phenopackets) analyzed in this work.

Tool	First published	**VCF**	**HPO**	**Web**	**Shell**	**Assemblies**	**Last update**
eXtasy [5]	2013	✓	✓	✓	✓	hg19	2013
Exomiser [6, 7, 8]	2014	✓	✓	×	✓	hg19, hg38	2019
Phen-Gen [9]	2014	✓	✓	✓	✓	hg19	2014
PhenoVar [10]	2014	✓(a)	✓	✓	×	hg19	2017
BierApp [11]	2014	✓	✓	✗(no access)	×	hg19	2016
wANNOVAR [12]	2015	✓	✓	✓	×	hg19, hg38	2019
OVA [13]	2015	✓	✓	✓	×	hg19	2015
OMIM Explorer [14]	2016	✓	✓	✗(no access)	×	hg19	2016
QueryOR [15]	2017	✓	✓	✗(no access)	×	hg19	2016
GenIO [16]	2018	✓	✓	✓	×	hg19	2017
AMELIE/Phrank [17]	2019	✓	✓	✗(b)	×	hg19	2019
Phenoxome [18]	2019	✓	✓	✓	×	hg19	2019
DeepPVP [19]	2019	✓	✓	×	✗(c)	hg19	2019
MutationDistiller [20]	2019	✓	✓	✓	×	hg19	2019
PhenoPro [21]	2019	✓	✓	✗(no access)	×	hg19	2019

Table S5. Other tools for phenotype-driven exome/genome analysis. Symbols: ✓ The tool has the capability denoted in the column. × The tool does not have the capability denoted in the column. ✗ The publication describes the capability in question but it was not functional during the period of time this manuscript was being prepared (Sep.-Dec., 2019). Additional comments: (a) Requires registration, which is not working; (b) Web version of AMELIE not accepting jobs (attempted various times, October–December, 2020); (c) Install instructions failed on dependencies or docker file; (no access): Web server could not be accessed on multiple occasions.

References

[1] Sebastian Köhler, Marcel H Schulz, Peter Krawitz, Sebastian Bauer, Sandra Dölken, Claus E Ott, Christine Mundlos, Denise Horn, Stefan Mundlos, and Peter N Robinson. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *The American Journal of Human Genetics*, 85(4):457–464, 2009.

[2] Roberta Roncarati, Chiara Viviani Anselmi, Peter Krawitz, Giovanna Lattanzi, Yskert von Kodolitsch, Andreas Perrot, Elisa di Pasquale, Laura Papa, Paola Portararo, Marta Columbaro, Alberto Forni, Giuseppe Faggian, Gianluigi Condorelli, and Peter N Robinson. Doubly heterozygous LMNA and TTN mutations revealed by exome sequencing in a severe form of dilated cardiomyopathy. *European journal of human genetics : EJHG*, 21:1105–1111, October 2013.

[3] Satoko Kumada, Masaharu Hayashi, Kunimasa Arima, Hiroshi Nakayama, Kenji Sugai, Masayuki Sasaki, Kiyoko Kurata, and Michio Nagata. Renal disease in Arima syndrome is nephronophthisis as in other Joubert-related Cerebello-oculo-renal syndromes. *American journal of medical genetics. Part A*, 131:71–76, November 2004.

[4] Yskert von Kodolitsch and Peter N Robinson. Marfan syndrome: an update of genetics, medical and surgical management. *Heart (British Cardiac Society)*, 93:755–760, June 2007.

[5] Alejandro Sifrim, Dusan Popovic, Leon-Charles Tranchevent, Amin Ardeshirdavani, Ryo Sakai, Peter Konings, Joris R Vermeesch, Jan Aerts, Bart De Moor, and Yves Moreau. eXtasy: variant prioritization by genomic data fusion. *Nature methods*, 10:1083–1084, November 2013.

[6] Peter N Robinson, Sebastian Köhler, Anika Oellrich, Sanger Mouse Genetics Project, Kai Wang, Christopher J Mungall, Suzanna E Lewis, Nicole Washington, Sebastian Bauer, Dominik Seelow, Peter Krawitz, Christian Gilissen, Melissa Haendel, and Damian Smedley. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome research*, 24:340–348, February 2014.

[7] Damian Smedley, Julius OB Jacobsen, Marten Jäger, Sebastian Köhler, Manuel Holtgrewe, Max Schubach, Enrico Siragusa, Tomasz Zemojtel, Orion J Buske, Nicole L Washington, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nature Protocols*, 10(12):2004, 2015.

[8] Damian Smedley, Max Schubach, Julius O B Jacobsen, Sebastian Köhler, Tomasz Zemojtel, Malte Spielmann, Marten Jäger, Harry Hochheiser, Nicole L Washington, Julie A McMurry, Melissa A Haendel, Christopher J Mungall, Suzanna E Lewis, Tudor Groza, Giorgio Valentini, and Peter N Robinson. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in Mendelian disease. *American journal of human genetics*, 99:595–606, September 2016.

[9] Asif Javed, Saloni Agrawal, and Pauline C Ng. Phen-Gen: combining phenotype and genotype to analyze rare disorders. *Nature methods*, 11:935–937, September 2014.

[10] Yannis J Trakadis, Caroline Buote, Jean-François Therriault, Pierre-Étienne Jacques, Hugo Larochelle, and Sébastien Lévesque. PhenoVar: a phenotype-driven approach in clinical genomics for the diagnosis of polymalformative syndromes. *BMC medical genomics*, 7:22, May 2014.

[11] Alejandro Alemán, Francisco Garcia-Garcia, Francisco Salavert, Ignacio Medina, and Joaquín Dopazo. A web-based interactive framework to assist in the prioritization of disease candidate genes in whole-exome sequencing studies. *Nucleic acids research*, 42:W88–W93, July 2014.

[12] Hui Yang and Kai Wang. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature protocols*, 10:1556–1566, October 2015.

[13] Agne Antanaviciute, Christopher M Watson, Sally M Harrison, Carolina Lascelles, Laura Crinnion, Alexander F Markham, David T Bonthron, and Ian M Carr. OVA: integrating molecular and physical phenotype data from multiple biomedical domain ontologies with variant filtering for enhanced variant prioritization. *Bioinformatics (Oxford, England)*, 31:3822–3829, December 2015.

[14] Regis A James, Ian M Campbell, Edward S Chen, Philip M Boone, Mitchell A Rao, Matthew N Bainbridge, James R Lupski, Yaping Yang, Christine M Eng, Jennifer E Posey, and Chad A Shaw. A visual and curatorial approach to clinical variant prioritization and disease gene discovery in genome-wide diagnostics. *Genome medicine*, 8:13, February 2016.

[15] Loris Bertoldi, Claudio Forcato, Nicola Vitulo, Giovanni Birolo, Fabio De Pascale, Erika Feltrin, Riccardo Schiavon, Franca Anglani, Susanna Negrisolo, Alessandra Zanetti, Francesca D'Avanzo, Rosella Tomanin, Georgine Faulkner, Alessandro Vezzi, and Giorgio Valle. QueryOR: a comprehensive web platform for genetic variant analysis and prioritization. *BMC bioinformatics*, 18:225, April 2017.

[16] Daniel Koile, Marta Cordoba, Maximiliano de Sousa Serro, Marcelo Andres Kauffman, and Patricio Yankilevich. GenIO: a phenotype-genotype analysis web server for clinical genomics of rare diseases. *BMC bioinformatics*, 19:25, January 2018.

[17] Karthik A Jagadeesh, Johannes Birgmeier, Harendra Guturu, Cole A Deisseroth, Aaron M Wenger, Jonathan A Bernstein, and Gill Bejerano. Phrank measures phenotype sets similarity to greatly improve Mendelian diagnostic disease prioritization. *Genetics in medicine : official journal of the American College of Medical Genetics*, 21:464–470, February 2019.

[18] Chao Wu, Batsal Devkota, Perry Evans, Xiaonan Zhao, Samuel W Baker, Rojeen Niazi, Kajia Cao, Michael A Gonzalez, Pushkala Jayaraman, Laura K Conlin, Bryan L Krock, Matthew A Deardorff, Nancy B Spinner, Ian D Krantz, Avni B Santani, Ahmad N Abou Tayoun, and Mahdi Sarmady. Rapid and accurate interpretation of clinical exomes using Phenoxome: a computational phenotype-driven approach. *European journal of human genetics : EJHG*, 27:612–620, April 2019.

[19] Imane Boudellioua, Maxat Kulmanov, Paul N Schofield, Georgios V Gkoutos, and Robert Hoehndorf. DeepPVP: phenotype-based prioritization of causative variants using deep learning. *BMC bioinformatics*, 20:65, February 2019.

[20] Daniela Hombach, Markus Schuelke, Ellen Knierim, Nadja Ehmke, Jana Marie Schwarz, Björn Fischer-Zirnsak, and Dominik Seelow. MutationDistiller: user-driven identification of pathogenic DNA variants. *Nucleic acids research*, 47:W114–W120, July 2019.

[21] Zixiu Li, Feng Zhang, Yukai Wang, Yue Qiu, Yang Wu, Yulan Lu, Lin Yang, William J Qu, Huijun Wang, Wenhao Zhou, and Weidong Tian. PhenoPro: a novel toolkit for assisting in the diagnosis of Mendelian disease. *Bioinformatics (Oxford, England)*, 35:3559–3566, October 2019.