The American Journal of Human Genetics, Volume *94*

Supplemental Data

Phevor Combines Multiple Biomedical Ontologies for Accurate Identification of Disease-Causing Alleles in Single Individuals and Small Nuclear Families

Marc V. Singleton, Stephen L. Guthery, Karl V. Voelkerding, Karin Chen, Brett Kennedy, Rebecca L. Margraf, Jacob Durtschi, Karen Eilbeck, Martin G. Reese, Lynn B. Jorde, Chad D. Huff, and Mark Yandell

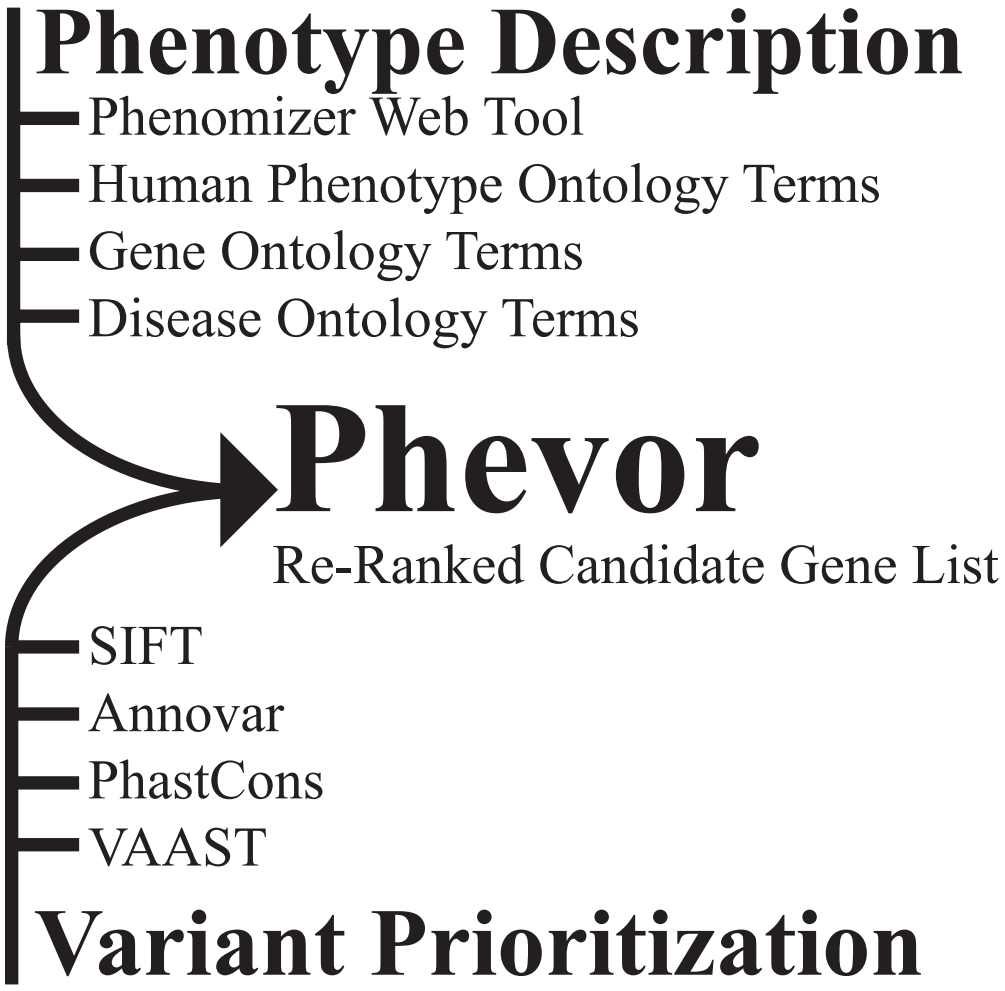


Figure S1. Inputs to Phevor. Phevor re-ranks the outputs of variant prioritization tools in light of phenotype and gene function information. The inputs to Phevor are individual variant scores from tools like SIFT and PhastCons, candidate gene lists as returned by Annovar, or prioritized gene lists such as VAAST output files. These are used together with a list of terms or their IDs describing the patient phenotype, gene functions, etc. drawn from the Human Phenotype Ontology (HPO), the Disease Ontology (DO), the Mammalian Phenotype Ontology (MPO), or the Gene Ontology (GO). Mixtures of terms from more than one ontology are permitted, as are OMIM disease terms. Users may also employ the online tool Phenomizer to describe a patient phenotype and to assemble a list of candidate-genes.

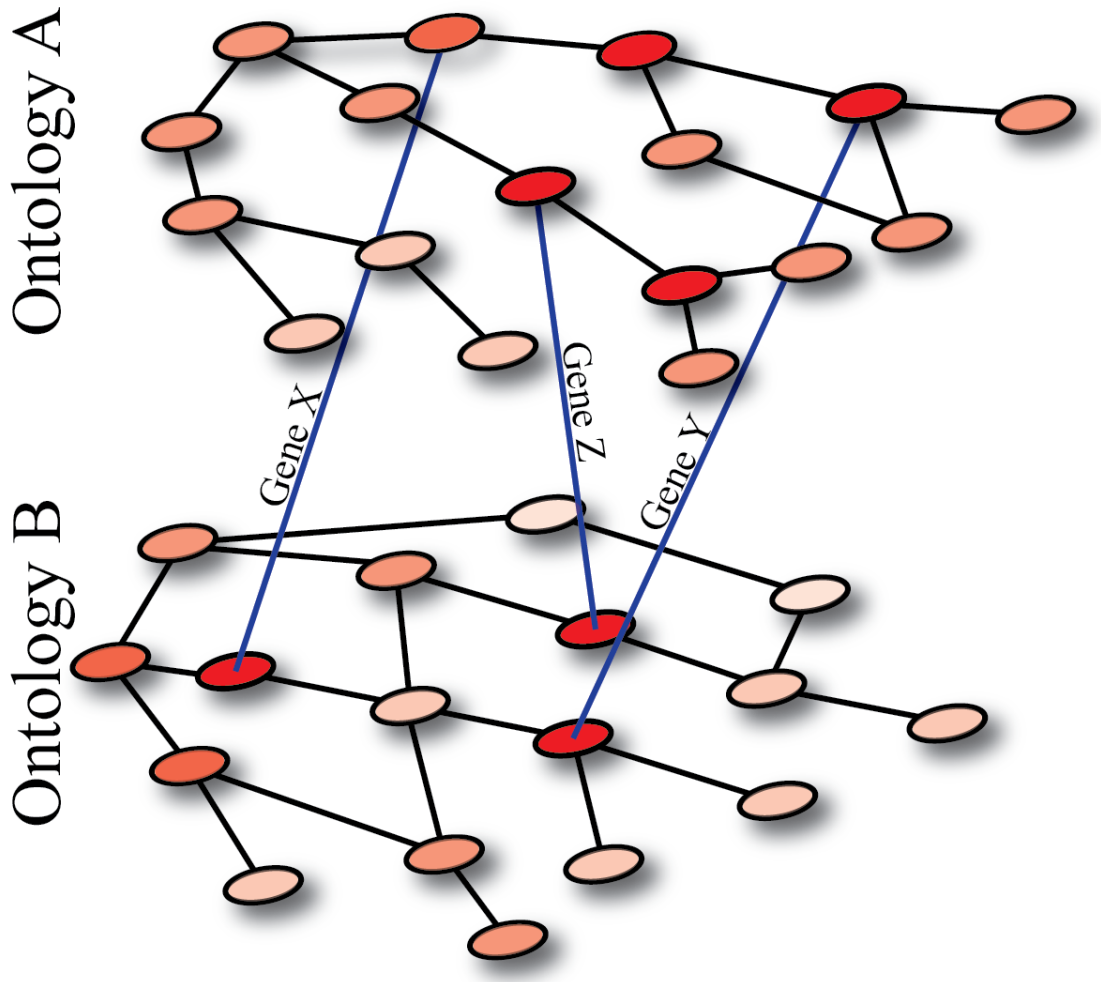


Figure S2. Combining Ontologies. Phevor relates different ontologies via their common gene annotations. Shown above are two generic ontologies A and B. Circles denote terms, or 'nodes', with edges denoting relationships between terms. For purposes of illustration, assume that each edge is directed, with the root of both ontologies lying at the top left-most node of the graph. The blue lines connecting the two ontologies represent three different genes X, Y and Z that are annotated to both ontologies. Phevor uses genes that have been annotated to two or more ontologies to relate terms in ontology A to those in B and vice versa. This cross-ontology linking procedure allows Phevor to combine knowledge from different ontological domains, e.g. phenotype information from HPO and gene function, process and location information from GO.

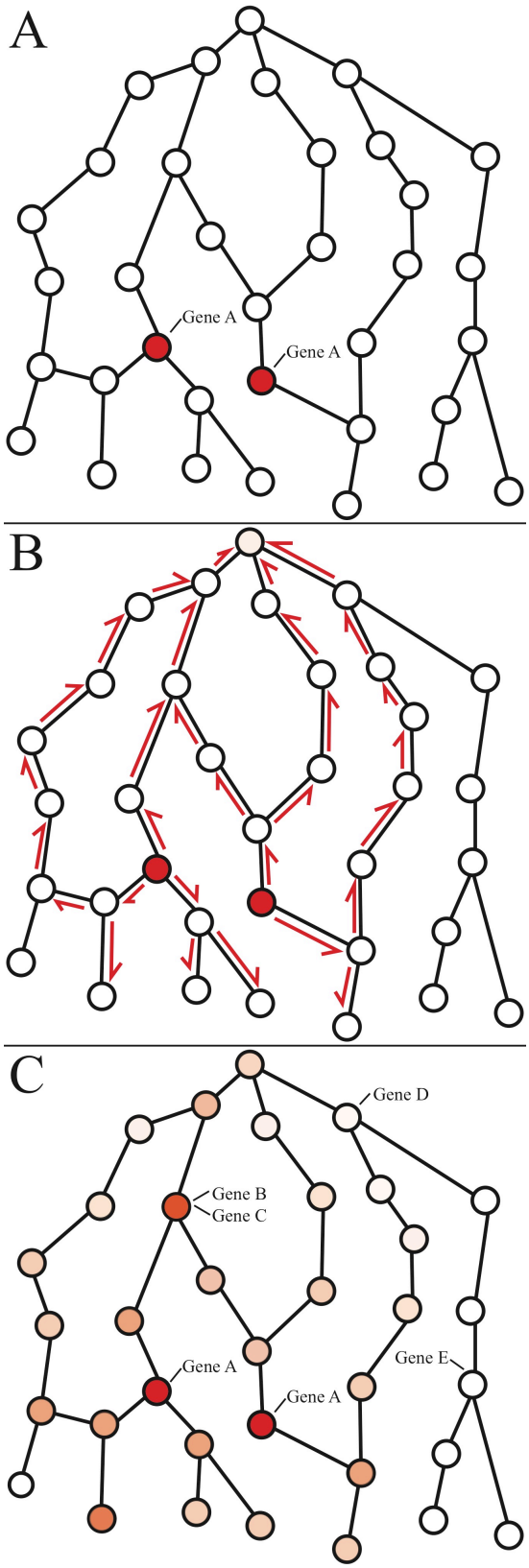
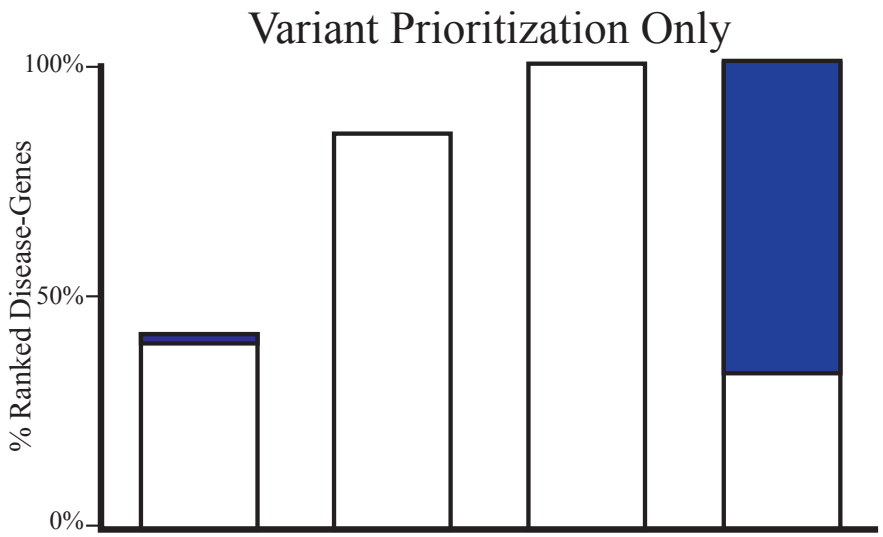


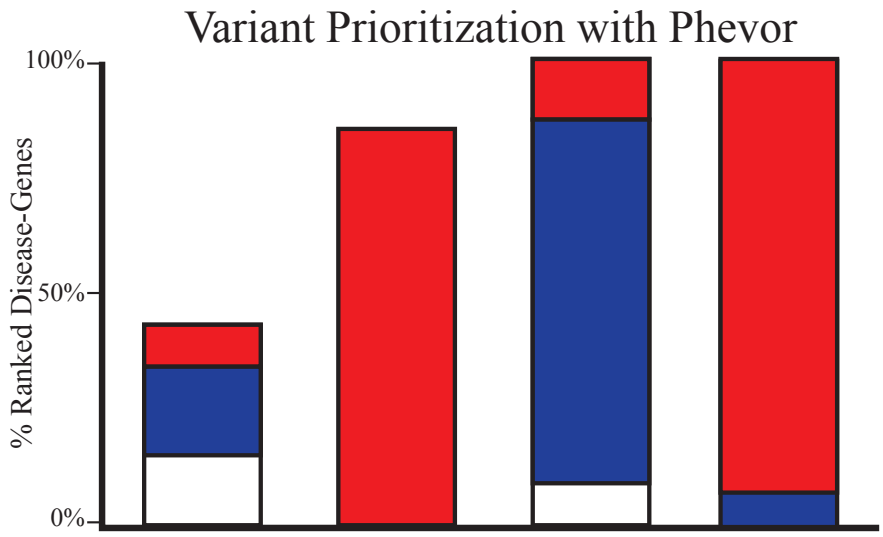
Figure S3. Ontological propagation. Starting from a user-provided set of terms (nodes), supplemented by the cross-ontology linking procedure illustrated in Figure S2, Phevor next propagates this information across each ontology. **Panel A** shows a hypothetical ontology, with two user-provided terms (nodes) shown in red. In this example, gene A has previously been annotated to both of these terms. This information is propagated across the ontology as illustrated in **Panel B.** First, these two 'seed nodes' are assigned a value of 1, and each time an edge is crossed to a neighboring node, the current value of the previous node is divided by 2. **Panel C** illustrates the end result of the propagation process, with node colors corresponding to the magnitudes of their propagation scores, with red representing nodes with the greatest scores, white nodes with scores near zero. Note that nodes located at intersecting threads of propagation, far from the original seeds can attain high values, even exceeding those of the starting seed nodes. The phenomenon is illustrated by the darker red nodes in **Panel C,** in which propagation has identified two additional gene-candidates, B and C not associated with the original seed nodes, but annotated to nodes with high propagation scores.

A



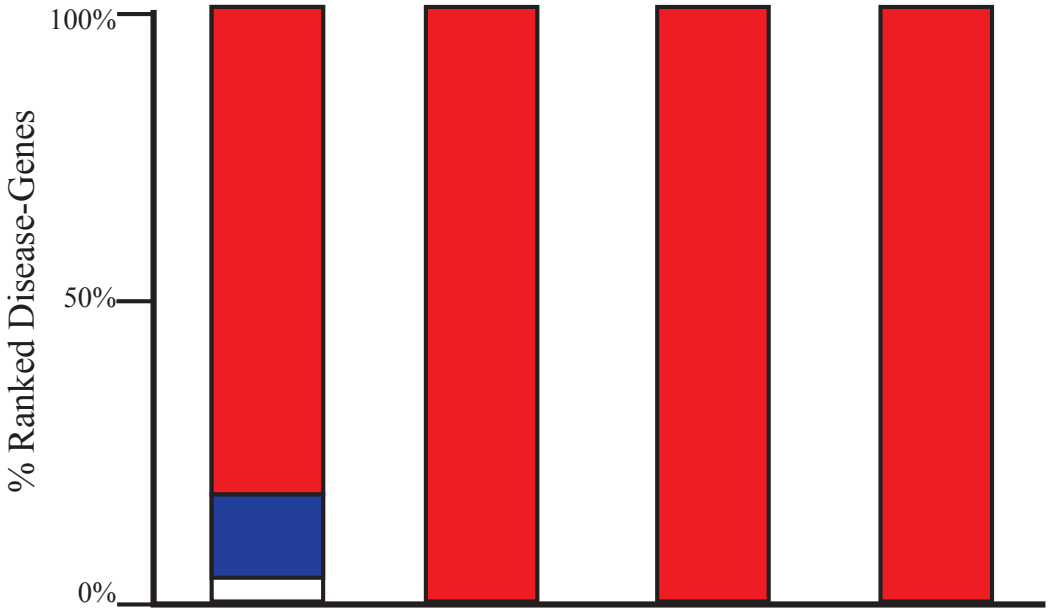
	SIFT	Annovar	PhastCons	VAAST
Top 10 Rank	0 ± 0%	0 ± 0%	0 ± 0%	0 ± 0%
Top 100 Rank	2 ± 1%	0 ± 0%	0 ± 0%	67 ± 5%
Ranked	42%	85%	100%	100%
Average Rank	11,161 ± 958	3,645 ± 873	4,305 ± 127	116 ± 9

B



	SIFT	Annovar	PhastCons	VAAST
Top 10 Rank	9 ± 3%	85 ± 4%	13 ± 4%	93 ± 3%
Top 100 Rank	28 ± 4%	85 ± 4%	91 ± 3%	100 ± 0%
Ranked	43%	85%	100%	100%
Average Rank	6,445 ± 492	1,628 ± 399	160 ± 82	4 ± 0.5

Figure S4. Variant Prioritization for Known Disease Genes (dominant). Performance comparisons of four different variant prioritization tools before (top panel), and after (bottom panel) post-processing them with Phevor. A single copy of a known disease-causing allele was randomly selected from HGMD and spiked into a single target exome at the reported genomic location; hence these results model simple, dominant diseases. This process was repeated 100 times for 100 different, randomly selected known disease genes in order to determine margins of error. Bar charts show the percentage of time the disease gene was ranked among the top ten candidates genome-wide (red), or among the top 100 candidates (blue), with white denoting a rank greater than 100 in the candidate list. For the Phevor analyses shown in the bottom panel, each tool's output files were fed to Phevor along with phenotype report containing the HPO terms annotated to each disease gene. The table below the bar charts summarizes this information in more detail. Some bars do not reach 100% due to false negatives, *i.e.* the tool is unable to prioritize the disease-causing allele. Damaging alleles predicted to be benign were placed at the midpoint of the list 22,107 annotated human genes.



Phenotype Specificity	Phenomizer Based Search	HPO Terms Using Phevor	GO Terms Using Phevor	DO Terms Using Phevor
Top 10 Rank	82 ± 4%	100 ± 1%	100 ± 0%	100 ± 0%
Top 100 Rank	96 ± 2%	100 ± 0%	100 ± 0%	100 ± 0%
Average Rank	67 ± 83	0.8 ± 0.1	0.9 ± 0.1	0.6 ± 0.1

Figure S5. Phevor Accuracy using different sources to describe patient phenotype and gene function. Comparison of methods to generate the initial list of ontology terms to be fed into Phevor. Two copies of a known disease-causing allele were randomly selected from HGMD and spiked into a single target exome at the reported genomic location; hence these results model simple, recessive diseases. This process was repeated 100 times for 100 different, randomly selected known disease genes to determine margins of error. Bar charts show the percentage of time the disease gene was ranked among the top ten candidates genome-wide (red), or among the top 100 candidates (blue), with white denoting a rank greater than 100 in the candidate list. For reasons of economy, only VAAST results are shown. Each Phevor analyses were fed the VAAST results along with a term list consisting of ontology terms annotated to each disease gene as indicated. For the Phenomizer comparison, The HPO terms associated with that gene were used to query Phenomizer and the resulting report containing a list of candidate disease-genes was fed directly to Phevor. The table below the bar charts summarizes this information in more detail.

Table S1. Phenotype terms used to create Figure 1 and Figure S4

Disease-Gene	Exomiser Rank	Annovar Rank	Annovar+Phevor Rank	OMIM Entry	Disease Name
AARS2	842	88	0	OMIM:614096	COMBINED OXIDATIVE PHOSPHORYLATION DEFICIENCY 8
ACVR1	11000	88	1	OMIM:135100	FIBRODYSPLASIA OSSIFICANS PROGRESSIVA
ALDH7A1	11000	11000	11000	OMIM:266100	EPILEPSY, PYRIDOXINE-DEPENDENT
ARHGAP31	81	88	0	OMIM:100300	ADAMS-OLIVER SYNDROME
ATP1A2	2681	88	1	OMIM:602481	MIGRAINE, FAMILIAL HEMIPLEGIC, 2
ATP6V1B1	785	11000	11000	OMIM:267300	RENAL TUBULAR ACIDOSIS, DISTAL, WITH PROGRESSIVE NERVE DEAFNESS
BUB1B	2207	88	0	OMIM:257300	MOSAIC VARIEGATED ANEUPLOIDY SYNDROME 1
C1S	2091	11000	11000	OMIM:613783	COMPLEMENT COMPONENT C1S DEFICIENCY
C5	132	88	1	OMIM:609536	COMPLEMENT COMPONENT 5 DEFICIENCY
CDKN1C	4	88	1	OMIM:130650	BECKWITH-WIEDEMANN SYNDROME
CHD7	1	88	1	OMIM:214800	CHARGE SYNDROME
CNGB1	2	88	0	OMIM:613767	RETINITIS PIGMENTOSA 45
COL10A1	2	88	1	OMIM:156500	METAPHYSEAL CHONDRODYSPLASIA, SCHMID TYPE
CST3	2625	11000	11000	OMIM:611953	MACULAR DEGENERATION, AGE-RELATED, 11
DLX3	2963	88	0	OMIM:104510	AMELOGENESIS IMPERFECTA, TYPE IV
DNAJC19	99	88	0	OMIM:610198	3-@METHYLGLUTACONIC ACIDURIA, TYPE V
DRD2	11000	11000	11000	OMIM:159900	MYOCLONIC DYSTONIA
EGR2	2	88	0	OMIM:145900	HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS
ENO3	11000	11000	11000	OMIM:612932	GLYCOGEN STORAGE DISEASE XIII
ETHE1	135	88	1	OMIM:602473	ENCEPHALOPATHY, ETHYLMALONIC
FOXC1	1	88	1	OMIM:601631	IRIDOGONIODYSGENESIS, TYPE 1
FOXP3	1	88	1	OMIM:304790	IMMUNODYSREGULATION, POLYENDOCRINOPATHY, AND ENTEROPATHY, X-LINKED
GCH1	25	11000	11000	OMIM:128230	DYSTONIA, DOPA-RESPONSIVE
GCM2	1	88	0	OMIM:146200	HYPOPARATHYROIDISM, FAMILIAL ISOLATED
GDF3	3120	88	0	OMIM:613704	MICROPHTHALMIA, ISOLATED 7
GJA8	1	88	0	OMIM:116200	CATARACT, ZONULAR PULVERULENT 1
GLRX5	11000	11000	11000	OMIM:205950	ANEMIA, CONGENITAL SIDEROBLASTIC, B6-NONRESPONSIVE
GNAS	2793	88	0	OMIM:174800	MCCUNE-ALBRIGHT SYNDROME
GRXCR1	16	11000	11000	OMIM:613285	DEAFNESS, AUTOSOMAL RECESSIVE 25
GTF2H5	825	88	0	OMIM:601675	TRICHOTHIODYSTROPHY, PHOTOSENSITIVE
HBA1	791	11000	11000	OMIM:140700	HEINZ BODY ANEMIAS
HESX1	17	88	0	OMIM:182230	SEPTOOPTIC DYSPLASIA
HSPB3	11000	11000	11000	OMIM:613376	NEURONOPATHY, DISTAL HEREDITARY MOTOR, TYPE IIC
IFT122	1706	88	1	OMIM:218330	CRANIOECTODERMAL DYSPLASIA
IFT140	27	88	1	OMIM:266920	MAINZER-SALDINO SYNDROME
IFT80	1	11000	11000	OMIM:611263	ASPHYXIATING THORACIC DYSTROPHY 2
IL1RAPL1	174	88	0	OMIM:300143	MENTAL RETARDATION, X-LINKED 21
IMPDH1	3615	88	0	OMIM:180105	RETINITIS PIGMENTOSA 10
INPP5E	1	88	1	OMIM:213300	JOUBERT SYNDROME 1
KLK4	10	11000	11000	OMIM:204700	AMELOGENESIS IMPERFECTA, PIGMENTED HYPOMATURATION TYPE
KRIT1	2840	88	1	OMIM:116860	CEREBRAL CAVERNOUS MALFORMATIONS
KRT14	2	11000	11000	OMIM:131760	EPIDERMOLYSIS BULLOSA SIMPLEX, DOWLING-MEARA TYPE
LOXHD1	96	11000	11000	OMIM:613079	DEAFNESS, AUTOSOMAL RECESSIVE 77
MLC1	11000	88	0	OMIM:604004	MEGALENCEPHALIC LEUKOENCEPHALOPATHY WITH SUBCORTICAL CYSTS 1
MOGS	1184	11000	11000	OMIM:606056	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE IIB
MTMR2	102	88	1	OMIM:601382	CHARCOT-MARIE-TOOTH DISEASE, TYPE 4B1
MYH7	11000	88	1	OMIM:160500	MYOPATHY, DISTAL, 1
NDUFAF4	11000	88	1	OMIM:252010	MITOCHONDRIAL COMPLEX I DEFICIENCY
NEUROD1	72	11000	11000	OMIM:606394	MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 6
NOP10	25	11000	11000	OMIM:224230	DYSKERATOSIS CONGENITA, AUTOSOMAL RECESSIVE, 1
NOTCH2	5	88	1	OMIM:610205	ALAGILLE SYNDROME 2
NPHP4	2935	88	0	OMIM:606966	NEPHRONOPHTHISIS 4
NPHS2	11000	88	0	OMIM:600995	NEPHROTIC SYNDROME, TYPE 2
NRAS	2643	88	0	OMIM:188470	THYROID CARCINOMA, FOLLICULAR
OPN1MW	20	11000	11000	OMIM:303700	BLUE CONE MONOCHROMACY
OTX2	1093	88	0	OMIM:613986	PITUITARY HORMONE DEFICIENCY, COMBINED, 6
PEX11B	2234	88	1	OMIM:614920	PEROXISOME BIOGENESIS DISORDER 14B
PITX1	1	11000	11000	OMIM:119800	CLUBFOOT, CONGENITAL, WITH OR WITHOUT DEFICIENCY OF LONG BONES AND/ORMIRROR-IMAGE POLYDACTYLY
PLP1	7	88	1	OMIM:312920	SPASTIC PARAPLEGIA 2, X-LINKED
PROP1	11000	88	0	OMIM:262600	PITUITARY HORMONE DEFICIENCY, COMBINED, 2
PROS1	14	88	0	OMIM:614514	THROMBOPHILIA DUE TO PROTEIN S DEFICIENCY, AUTOSOMAL RECESSIVE
PRPH2	2	88	1	OMIM:608161	MACULAR DYSTROPHY, VITELLIFORM, ADULT-ONSET
RAB3GAP1	215	88	1	OMIM:600118	WARBURG MICRO SYNDROME 1
RBM28	886	11000	11000	OMIM:612079	ALOPECIA, NEUROLOGIC DEFECTS, AND ENDOCRINOPATHY SYNDROME
RFX5	28	88	0	OMIM:209920	BARE LYMPHOCYTE SYNDROME, TYPE II
ROR2	1	88	0	OMIM:268310	ROBINOW SYNDROME, AUTOSOMAL RECESSIVE
RPGRIP1	132	88	0	OMIM:608194	CONE-ROD DYSTROPHY 13
RPS10	11000	88	0	OMIM:613308	DIAMOND-BLACKFAN ANEMIA 9
RSPH4A	2026	11000	11000	OMIM:612649	CILIARY DYSKINESIA, PRIMARY, 11
SCN2A	27	88	1	OMIM:607745	SEIZURES, BENIGN FAMILIAL INFANTILE, 3
SCN3B	2172	11000	11000	OMIM:613120	BRUGADA SYNDROME 7
SERPINA6	4892	11000	11000	OMIM:611489	CORTICOSTEROID-BINDING GLOBULIN DEFICIENCY
SH3BP2	2023	88	0	OMIM:118400	CHERUBISM
SHANK3	2454	88	0	OMIM:606232	CHROMOSOME 22Q13.3 DELETION SYNDROME
SLC22A5	11000	88	0	OMIM:212140	CARNITINE DEFICIENCY, SYSTEMIC PRIMARY
SLC35C1	2512	88	1	OMIM:266265	CONGENITAL DISORDER OF GLYCOSYLATION, TYPE IIC
SLC7A7	35	88	0	OMIM:222700	LYSINURIC PROTEIN INTOLERANCE
SLCO1B1	11000	11000	11000	OMIM:237450	HYPERBILIRUBINEMIA, ROTOR TYPE
SMAD4	4	88	0	OMIM:174900	JUVENILE POLYPOSIS SYNDROME
SMARCA4	3555	88	0	OMIM:614609	MENTAL RETARDATION, AUTOSOMAL DOMINANT 16
SMOC2	94	88	1	OMIM:125400	DENTIN DYSPLASIA, TYPE I
SOD1	1728	88	0	OMIM:105400	AMYOTROPHIC LATERAL SCLEROSIS 1
SOS1	2306	88	0	OMIM:135300	FIBROMATOSIS, GINGIVAL, 1
SPTLC1	185	88	1	OMIM:162400	NEUROPATHY, HEREDITARY SENSORY AND AUTONOMIC, TYPE IA
SQSTM1	136	88	1	OMIM:602080	PAGET DISEASE OF BONE
SRD5A3	100	88	0	OMIM:612713	KAHRIZI SYNDROME
TCN2	82	88	0	OMIM:275350	TRANSCOBALAMIN II DEFICIENCY
TGFBR1	2344	88	0	OMIM:609192	LOEYS-DIETZ SYNDROME, TYPE 1A
TGM5	818	11000	11000	OMIM:609796	PEELING SKIN SYNDROME, ACRAL TYPE
TJP2	5233	11000	11000	OMIM:607748	HYPERCHOLANEMIA, FAMILIAL
TMEM216	25	88	0	OMIM:608091	JOUBERT SYNDROME 2
TP63	7	88	0	OMIM:106260	ANKYLOBLEPHARON-ECTODERMAL DEFECTS-CLEFT LIP/PALATE
TTN	1	88	0	OMIM:600334	TIBIAL MUSCULAR DYSTROPHY, TARDIVE
UBR1	1697	88	0	OMIM:243800	JOHANSON-BLIZZARD SYNDROME
WDR11	11000	88	0	OMIM:614858	HYPOGONADOTROPIC HYPOGONADISM 14 WITH OR WITHOUT ANOSMIA
WIPF1	114	88	0	OMIM:614493	WISKOTT-ALDRICH SYNDROME 2
WNT7A	10	88	0	OMIM:276820	ULNA AND FIBULA, ABSENCE OF, WITH SEVERE LIMB DEFICIENCY
XPC	8	88	1	OMIM:278720	XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP C
ZFYVE26	1068	88	1	OMIM:270700	SPASTIC PARAPLEGIA 15, AUTOSOMAL RECESSIVE
ZMPSTE24	77	88	1	OMIM:275210	RESTRICTIVE DERMOPATHY, LETHAL