Supp. Tables S1-S7 are available online as separate Supporting Information Text Files.

Supp. Table S9 is available online as a separate Supporting Information Excel file.

Supp. Files S1-S4 are available online as separate Supporting Information Text Files.

Supp. Table S8. Gene Exclusion List descriptions

List Name	Number of Genes	Construction Notes
Supp. Table S1 (empiric.txt)	435	Conservative list of genes noted to contribute potentially pathogenic[a] mutations to most exome projects plus *HLA*, *TAS*, *OR* and *MUC* family genes
Supp. Table S2 (ten_in_ten_families.txt)	74	empiric.txt plus genes with ten or more pathogenic mutations[a,b] in ten of 29 families
Supp. Table S3 (ten_in_twenty_families.txt)	36	empiric.txt plus genes with ten or more pathogenic mutations[a,b] in twenty of 29 families
Supp. Table S4 (ten_in_all_families.txt)	14	empiric.txt plus genes with ten or more pathogenic mutations[a,b] in 29 families
Supp. Table S5 (pseudogenes.txt)	1,580 (names)	NCBI list of validated pseudogenes, including known alternate gene names
Supp. Table S6 (ten_in_more_than_3_of_27.txt)	163	
Supp. Table S7 (gene_exclusion_list.txt)	2,157	The UDP working gene exclusion list which includes empiric.txt, pseudogenes.txt, ten_in_all_families.txt and ten_in_more_than_3_of_27.txt. Genes listed in more than one set are listed once.

[a]Frame shifting, stop, nonsense, missense, canonical splice site changing genes

[b]CDPred predicted pathogenicity =< -1

Supp. Table S9 Legend

Variants_with_annotations.xls **contains the combined** ***homozygous_nonref_annotations.xls, heterozygous_nonref_annotations.xls and heterozygous_nonref_annotations_2*** **files in separate tabs)**

These files are derived from the varsifter file format but have been truncated to only show relevant data. For further information about this file format please go to Varsifter wiki:

http://trek.nhgri.nih.gov/wiki/index.php/VarSifter

The Human Gene Mutation Database (HGMD) [Stenson, et al., 2003] was also used to annotate variants found in families sequenced.

Column headers:

A. **Chr**: Chromosome
B. **LeftFlank**: The position to the left of the variant in the Hg18 genome build. For indels, the position to the left of the ambiguous bases.
C. **RightFlank**: The position to the right of the variant in the Hg18 genome build (and indel, as above).
D. **ref_allele**: Reference base at this position (plus strand)
E. **muttype**: **SNP** for single nucleotide variant or **INDEL** for insertion/deletion variant
F. **var_allele**: variant base at this position (plus strand)
G. **type**: Variant type. VarSifter uses the entries in the file to create filter buttons.
H. **Gene_name**: The name of the gene (Refseq gene symbol).
I. **transcript**: The UCSC ID of the displayed transcript.
J. **strand**: The strand of the gene, or "NA".
K. **ref_aa**: Reference amino acid, or -
L. **var_aa**: Variant amino acid, or -
M. **aa_pos**: The position of the amino acid, or 0.
N. **CDPred_score**: The Conserved Domain Prediction score.
O. **RS#**: dbSNP130, or - if none.
P. HGMDids: HGMD Accession number - *Example: CM035497* **(Note: tags are added if genomic variant listed is present at the same genomic coordinate as a variant entry in HGMD, variant may or may not be the same as in HGMD).**
Q. **HGMDdisease:** disease association as listed in HGMD.
R. **HGMDtags:** DM disease mutation, DP disease polymorphism etc
S. **HGMDinGene:** (Y or N) presence or absence of gene in the HGMD database.

Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeysinghe S, Krawczak M, Cooper DN. 2003. Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat 21:577-81.