Supplementary Table

	dbSNP ID	MAF[1]	Blossum	Polyphen	SIFT	Panther	Pmut	SNAP
Reported disease causing mutations and nonsynonymous SNPs with MAF in dbSNP								
p.R120H	rs114476330	0.009	-	1	0	-4.11417	0.5381	N
p.E143del[*, #]	-	0.002[×]	-	-	-	-	-	-
p.L148P[#]	rs139763321	-	-3	0.999	0.01	-4.59230	0.0733	NN
p.R159Q[#]	-	-	1	1	0	-2.67823	0.4965	NN
p.R157W	rs35873579	0.035	-3	1	0	-4.63087	0.8827	NN
p.R157Q	rs35051736	0.012	1	0.617	0.19	-1.64275	0.4156	NN
p.K178R	rs146404747	0.001	2	0.025	0.45	-1.33609	0.0816	N
p.D202H	rs114579367	0.007	-1	0.981	0.03	-2.71613	0.0847	N
p.E206K	rs115260488	0.001	1	1	0.20	-1.83365	0.4336	N
p.M245I	rs114930663	0.002	1	0.92	0	-2.54376	0.1471	N
p.T248R	rs16999131	0.015	-1	1	0.09	-2.38001	0.4019	N
p.C303S	rs76747058	0.025	-1	0.128	0.36	-1.45217	0.0616	N
p.E322K[#]	-	-	1	1	0	-2.19785	0.2993	NN
p.A332T	rs116804918	0.009	-	1	0.03	-3.10125	0.1251	N
p.R344H	rs116548533	0.007	-1	0.128	0.03	-1.45217	0.0616	N
p.R367Q	rs142282494	0.001	1	0.027	0.60	-1.72945	0.5621	N
p.M374T[%]	rs6022990	0.075[%]	-1	0.607	0	-3.09497	0.5060	NN
p.C380Y	rs150006710	0.001	-2	1	0.01	-4.89778	0.9316	NN
p.R396W[#]	rs114368325	0.001	-3	1	0	-5.49258	0.9135	NN
p.L409S[*, #]*	rs6068812	0.003	-2	0.999	0.01	-3.98574	0.1234	NN
p.R439H	rs141152573	0.001	-	1	0	-3.94494	0.6497	NN
p.V457I	rs112596218	0.002	3	0.003	1	0.60917	0.0367	N
p.A510V	rs116065115	0.011	-	0.838	0.14	-1.54737	0.4924	N
Total MAF:		**0.212**						
Total deleterious MAF:		**0.140**						
Other nonsynonymous SNPs in dbSNP								
p.E105K	rs147642444	-[2]	1	0.01	0.71	-1.15838	0.2905	N
p.L129M	rs149806586	-	2	0.976	0.09	-1.97344	0.0766	N
p.L148P	rs139763321	-	-3	0.999	0.01	-4.5923	0.0733	NN
p.V158A	rs139655790	-	-	0.022	0.02	-2.72874	0.1611	N
p.L207M	rs149235939	-	2	0.999	0.30	-2.45364	0.0825	N
p.K209R	rs138489641	-	2	1	0.43	-1.97704	0.0353	N
p.R396Q[#]	rs143934667	-	1	1	0	-3.50209	0.5003	NN
p.Y407N	rs140189382	-	-2	1	0	-4.26812	0.2150	NN
p.R481C	rs143523685	-	-3	1	0	-4.12556	0.7764	NN
p.R505Q	rs146980218	-	1	1	0.14	-2.88802	0.5593	N
p.P25A	rs140851407	-	-1	0.049	0.09	0.21306	0.0828	N
p.P126S	rs148084028	-	-1	1	0	-3.76463	0.0430	N
p.E153K	rs185120393	-	1	-	0.92	-0.73103	0.5131	N
p.E258D	rs190860407	-	2	0.994	0.15	-2.71834	0.0436	N
p.P375L	rs189801930	-	-3	1	0	-4.40742	0.5923	NN
p.M495V	rs77167734	-	1	0.001	1	-0.45058	0.4063	N
Reported artificial CYP24A1 mutations								
p.L148F[#]	Artificial	-	0	0.989	0.2	-2.84807	0.0922	NN
p.I131F[#]	Artificial	-	0	1	0.15	-3.42230	0.1951	N
p.A326G[#]	Artificial	-	0	0	0.23	-0.98296	0.1583	N
Reported splice-site variants		**BDGP score**			**NatGene2 NN score**			
c.732+1G>A		0.94>0			0.742>0			
c.733-2A>G		0.9>0			0.852>0			

[1] MAF = Minor Allele Frequency, as reported in dbSNP (March 2012).

[2] MAF= - ; not reported MAF in dbSNP.

[%] SNP p.M374T has an unusual high MAF (see text)

[×]Not reported in dbSNP, but in publication (45)

*Decreased CYP24A1 enzyme activity; reported in this paper, measured in fibroblasts.

[#] Enzyme activity *in vitro* decreased, measured in expressed mutants.

Gray Highlight: Likely deleterious variant predicted by at least 4 out of 6 prediction programs and/or decreased CYP24A1 enzyme activity.

Deleterious ranking per program: **BLOSSUM**: Negative score is deleterious; **Polyphen**: > 0.85 is deleterious; **SIFT**: ≤0.01 is deleterious;

Panther: ≤ -3 is deleterious; **Pmut**: >0.5 is deleterious; **SNAP**: NN, Non-neutral, deleterious (N=neutral).

*This paper, measured in fibroblasts

[a] In vitro activity, measured in expressed mutant

Supplementary Methods

SNP arrays

For SNP genotyping, genomic DNA was run on a Human 1M-Duo DNA Analysis BeadChip and the data analyzed using the GenomeStudio software (both from Illumina, San Diego, CA).

Missense Variant Prediction Tools

The effect of missense variations on protein function was evaluated using the mutation prediction programs POLYPHEN, PANTHER and PMUT.

POLYPHEN

(http://genetics.bwh.harvard.edu/pph/; POLYmorphism PHENotyping) predicts the effect of an amino acid substitution on the structure and function of a protein. POLYPHEN predictions are based on empirical rules that are applied to the sequence, as well as phylogenetic and known structural information that characterize the substitution. The Position-Specific Independent Counts (PSIC) is calculated for the two different alleles and the score for wild type and variant mapping to the known 3D structure.[1]

PANTHER

(http://www.pantherdb.org/; Protein ANalysis THrough Evolutionary Relationships) estimates the likelihood of a non-synonymous variant to cause loss of function of the protein. The output, the subPSEC (substitution position-specific evolutionary conservation), is the negative logarithm of the probability ratio of the wild-type and mutant amino acids at a particular position based on a library. This library contains over 5,000 protein families and 30,000 subfamilies, each represented by a multiple sequence alignment and Hidden Markov Model. PANTHER subPSEC scores are continuous from 0 to −10. A value of 0 is interpreted as a functionally neutral variant; the more negative the subPSEC value, the more deleterious the substitution. The cutoff value suggested is −3. [2-4]

PMUT

(http://mmb2.pcb.ub.es:8080/PMut/) uses neural networks that have been trained with a large database of disease-associated and neutral variants to predict the impact of a given amino acid substitution. The output gives a neural network (NN) value between 0 and 1 (the higher this value, the more deleterious the variant) and a confidence value between 0 and 9 (the higher this value, the more reliable the NN) [5]

SIFT

(http://sift.jcvi.org/)

Scale-invariant feature transform (SIFT) predicts whether an amino acid substitution affects protein **function**. SIFT prediction is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences, collected through PSI-BLAST. SIFT can be applied to naturally occurring nonsynonymous polymorphisms or laboratory-induced missense mutations.[6]

BLOSSUM

Blosum62 (ftp://ftp.ncbi.nih.gov/blast/matrices/BLOSUM62; **BLO**cks of Amino Acid **SU**bstitution **M**atrix) is a substitution matrix for pairwise protein sequence alignments. You will encounter Blosum62 in a number of bioinformatics applications that align protein sequences or analyze the homology between sequences. It contains similarity scores for all permutations of two amino acids, assigning higher (better) scores to similar amino acids. Scores within a BLOSUM are log-odds scores that measure, in an alignment, the logarithm for the ratio of the likelihood of two amino acids appearing with a biological sense and the likelihood of the same amino acids appearing by chance. A positive score is given to the more likely substitutions while a negative score is given to the less likely substitutions

SNAP

SNAP (screening for non-acceptable polymorphisms; http://cubic.bioc.columbia.edu/services/snap/) predicts the functional effects of single amino acid substitutions. Single Nucleotide Polymorphisms (SNPs) represent a very large portion of all genetic variations. SNPs found in the coding regions of genes are often non-synonymous, changing a single amino acid in the encoded protein sequence.[7] SNPs are either "neutral" in the sense that the resulting point-mutated protein is not functionally discernible from the wild-type, or they are "non-neutral" in that the mutant and wild-type differ in function. The ability to identify non-neutral substitutions in an ocean of SNPs could significantly aid targeting disease causing detrimental mutations, as well as SNPs that increase the fitness of particular phenotypes.

PREDICTION SOFTWARES FOR SPLICE-SITE MUTATIONS

The effect of splice site variations was also evaluated, using different analysis programs, including the splice site prediction tool from the Berkeley Drosophila Genome Project (**BDGP**) web site (http://www.fruitfly.org/seq_tools/splice.html). This is based on a generalized Hidden Markov Model to predict the strength of the possible splice site, using a neural network that has been trained by a set of 793 unrelated human genes Berkeley Drosophila Genome Project (BDGP) web site).[8] Another tool used was NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/), a service producing neural network predictions of splice sites in human, C. elegans and A. thaliana DNA.[9]

References

1. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002;**30**(17):3894–3900.

2. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003;**13**(9):2129–2141.

3. Thomas PD, Kejariwal A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci U S A*. 2004;**101**(43):15398–15403.

4. Thomas PD, Kejariwal A, Guo N, Mi H, Campbell MJ, et al. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res*. 2006;**34**(Web Server issue):W645–650.

5. Ferrer-Costa C, Gelpi JL, Zamakola L, Parraga I, de la Cruz X, et al. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics*. 2005;**21**(14):3176–3178.

6. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc. 2009*;**4**(7):1073-81.

7. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res. 2007;**35**(11):3823–3835

8. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997;**4**(3):311–323.

9. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. *Nucleic Acids Res*. 1996; **24**(17):3439-3452.