Supplemental methods

Exome capture

The exonic portion of the patients DNA was captured through hybridization with exon-targetting oligonucleotides using the Nimblegen HD2 sequence capture array (http://www.nimblegen.com/). These arrays tile oligonucleotides from approximately 180,000 exons of 18,673 protein-coding genes and 551 micro-RNAs and comprise 34.0 Mb of genomic sequence. In brief, genomic DNA is nebulized for 1 minute using 45psi of pressure. The sheared DNA fragments are cleaned with the DNA Clean & Concentrator-25 Kit (Zymo Research) and a fragment size distribution ranging from 300 bp to 500 bp verified via Bioanalyzer (Agilent). Following end-polishing of the genomic fragments, GS FLX Titanium adaptors are ligated to the sheared genomic fragments. These ligated fragments are then hybridized to the 2.1 M exome array within Maui hybridization stations; the array is then washed to remove unbound DNA, array-bound fragments are subsequently eluted from the array within Nimblegen elution chambers. The captured fragments then undergo PCR amplification using primers targeting the Nimblegen linkers to generate sufficient DNA template for downstream applications. The capture efficiency is evaluated via quantitative PCR using 4 QC control *loci* that determine the degree of capture success. Finally, the amount of captured DNA (after elution) for sequencing is determined by spectrophotometry, and the sample concentration adjusted to the correct volume and concentration for sequencing.

Roche 454 Sequencing

Captured DNA samples were subjected to standard sample preparation procedures for 454 GS FLX sequencing. In brief, the DNA-sequencing library was prepared prior to amplification by emulsion PCR (emPCR). The emPCR followed the steps described in the Roche 454 emPCR Method Manual. The sample was then loaded in a single 2-lane gasket PicoTiterPlate device and sequenced in a GS FLX

system with the Titanium Roche/454 protocols. The 454 pyrosequencing data were collected after a 10-hour run on the GS FLX system.

Next-generation sequencing data were initially processed using the gsMapper software package (Roche Inc.) supplied with the GS FLX instrument. High quality sequencing reads were aligned to the human genome reference sequence hg18 (UCSC). Variants with respect to the hg18 reference sequence were identified using the standard Roche 454 software.

Variant identification, validation and annotation

The Roche 454 gsMapper software produces a variant file (454HCDiff) file, which was used to identify variants in the sample vs reference sequence. For each variant the following annotations were determined and assigned: 1) Novelty, based on existence of the same variant (by position and nucleotide) in the dbSNP database (dbSNP build 129). 2) Depth of coverage (derived from the Roche software output). 3) Quality score (derived from the Roche software output). 4) Amino acid physio-chemical properties thought to be important in the determination of protein structure (for both the reference and variant amino acids of protein coding variants); e.g. charge, polarity, and size (standard amino acid property tables used). 5) Class of change (synonymous, non-synonymous, stop codon etc.). 6) Phylogenetic conservation based on UCSC PhastCons scores (providing a measure of the functional importance of the residue at that position in the protein; more highly conserved residues inferred as being more important to the function of the protein; a score of 0.9 used as highly conserved). 7) Genic or genomic location (e.g. intronic, intergenic); based on comparison with the reference gene models from EntrezGene. 8) Zygosity (based on the number of reads that differ from the reference; 100% is defined as being homozygous, between 99% and 81% inclusive as probably homozygous, between 20% and 80% inclusive as heterozygous; lower than 20% were initially categorised as likely sequencing or assembly errors). 9) Effects on splice sites (Gene Splicer tool run on the reference and variant containing

DNA sequence in-house, output parsed). 10) Polyphen score, prediction, and effect (algorithm uses structural and sequence information to predict impact of substitution on the structure and function of a protein; run in house). 11) PDB structures for this protein or a related protein (derived from PolyPhen output). 12) Online Mendelian Inheritance in Man disease association(s) for the gene containing the variant (identified using the OMIM disease to gene mapping tables from NCBI, and presented as disease names which are OMIM links). 13) Protein annotation including protein ID, protein function, and description (obtained from RefSeq). 14) Gene annotation including chromosomal location, gene name, unique identifiers, and gene function. 15) Links to expression profiles derived from the GEO compendium (based on protein ID to expression profile mapping provided by NCBI).

Capillary sequencing validation

Total DNA was extracted from peripheral blood leukocytes using a commercially available DNA isolation kits (Gentra Puregene, Qiagen) according to the manufacturer's protocols. Sequence specific oligonucleotide primers were designed against 15 interesting variant targets using Primer 3 v 0.4 (http://frodo.wi.mit.edu/primer3/) with a minimum 50 base flanking sequence. Targets were amplified using Gotaq mastermix (Promega, Madison) (Primers and reaction conditions are available upon request). Sequencing reactions were performed using the BigDye Terminator Cycle Sequencing kit (version 3.1), purified using Millipore Montage 96-well plates on a Beckman FX and analyzed on an ABI3730XL automated DNA sequencer with data collection software v3.0 (Applied Biosystems, Foster City, CA, USA). Sequence was aligned to ensure correct genomic position. Then control and case traces were analyzed using Mutation Surveyor version 3.25 and the relevant Genebank sequences for each gene were used as the reference with mutations reported according to Human Genome Variation Society (HGVS) conventions. This work was carried out in the research laboratory of one of our researchers.