www.sciencetranslationalmedicine.org/cgi/content/full/4/138/138ra78/DC1

Supplementary Materials for

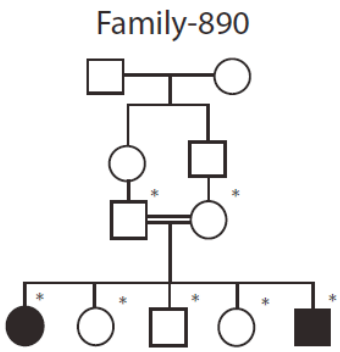
Exome Sequencing Can Improve Diagnosis and Alter Patient Management

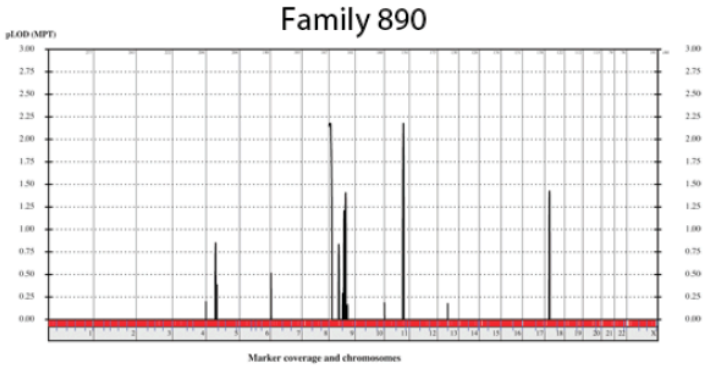
Tracy J. Dixon-Salazar, Jennifer L. Silhavy, Nitin Udpa, Jana Schroth, Stephanie Bielas, Ashleigh E. Schaffer, Jesus Olvera, Vineet Bafna, Maha S. Zaki, Ghada H. Abdel-Salam, Lobna A. Mansour, Laila Selim, Sawsan Abdel-Hadi, Naima Marzouki, Tawfeg Ben-Omran, Nouriya A. Al-Saana, F. Müjgan Sonmez, Figen Celep, Matloob Azam, Kiley J. Hill, Adrienne Collazo, Ali G. Fenstermaker, Gaia Novarino, Naiara Akizu, Kiran V. Garimella, Carrie Sougnez, Carsten Russ, Stacey B. Gabriel, Joseph G. Gleeson*

*To whom correspondence should be addressed. E-mail: jogleeson@ucsd.edu

Published 13 June 2012, *Sci. Transl. Med.* **4**, 138ra78 (2012)
DOI: 10.1126/scitranslmed.3003544

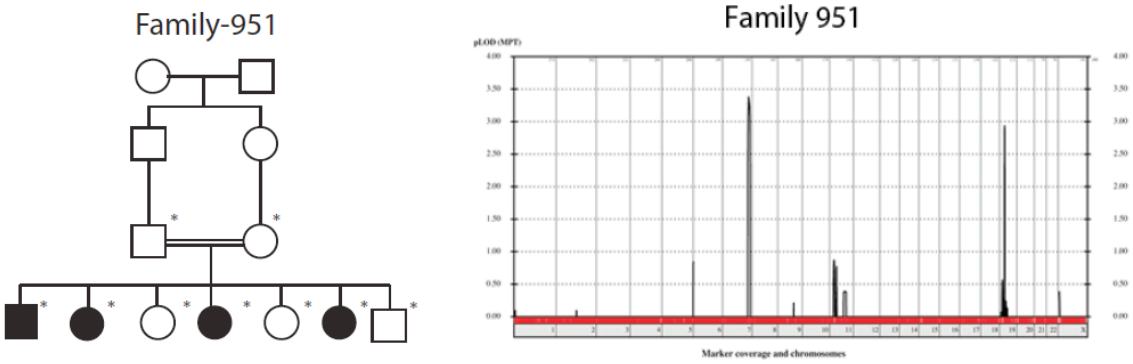
The PDF file includes:





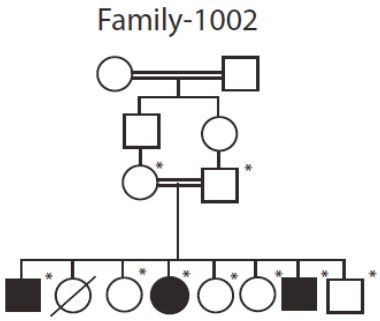
Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
9	2635016	GTTACAA	-	p.G1246fsX1305	DAMAGING	N/A	*VLDLR*	Very low density lipoprotein receptor	Yes	No
9	8487250	T	C	p.T781A	TOLERATED	4.38	*PTPRD*	Receptor-type-tyrosine-protein-phosphatase delta	No	-
9	107371	C	G	p.G250A	TOLERATED	1.21	*FOXD4*	Forkhead box protein D4	-	-
9	107428	G	C	p.P231R	TOLERATED	1.22	*FOXD4*	Forkhead box protein D4	-	-
9	107696	G	A	p.R142C	TOLERATED	1.85	*FOXD4*	Forkhead box protein D4	-	-
9	204804	A	G	p.V198A	TOLERATED	0.55	*C9orf66*	Uncharacterized protein C9orf66	-	-
9	7789653	G	T	p.P28T	TOLERATED	2.27	*C9orf123*	Transmembrane protein C9orf123	-	-

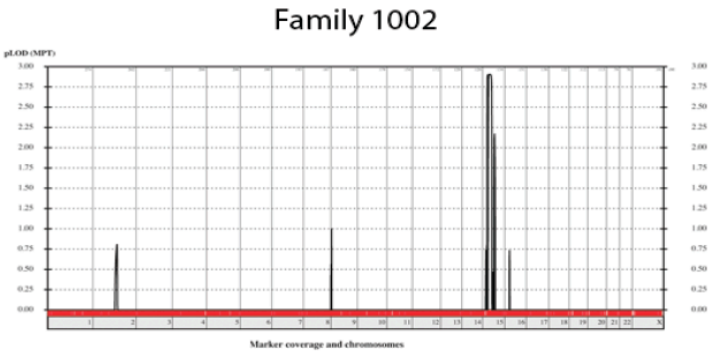
Figure S1. Genetic data for family 890. Pedigree, linkage analysis, and final variant table.



Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
19	12760995	C	T	p.W695*	DAMAGING	4.48	*MAN2B1*	Mannosidase alpha class 2B member 1	Yes	No

Figure S2. Genetic data for family 951. Pedigree, linkage analysis, and final variant table.





Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
15	44877868	G	C	p.A1696G	DAMAGING	5.07	*SPG11*	Spatacsin	Yes	No
15	45258358	T	G	splice	TOLERATED	3.83	*C15orf43*	Uncharacterized protein C15orf43	Yes	Yes
15	45412435	G	A	p.T213M	DAMAGING	4.57	*DUOXA1*	Dual oxidase maturation factor 1	Yes	Yes

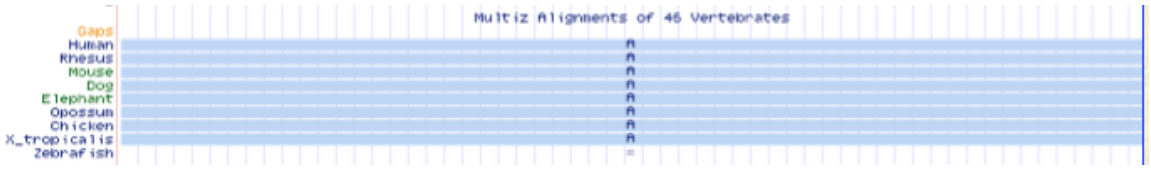
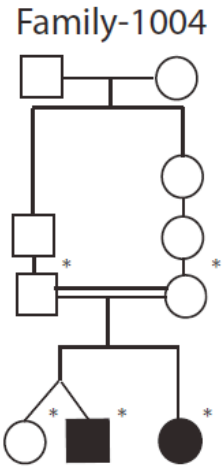
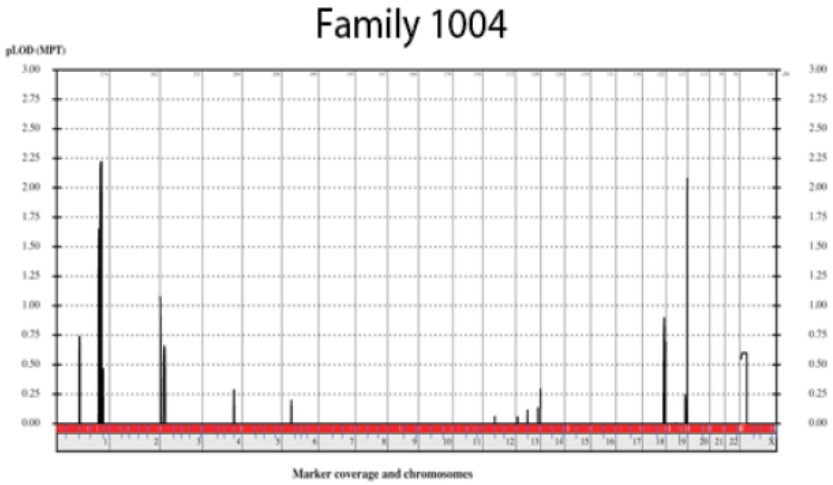


Figure S3. Genetic data for family 1002. Pedigree, linkage analysis, final variant table, and conservation figure for *SPG11* residue 1696. A1696 is conserved in vertebrates





Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
1	228345562	C	T	p.R35C	DAMAGING	3.76	*GJC2*	Gap junction gamma-2	Yes	No

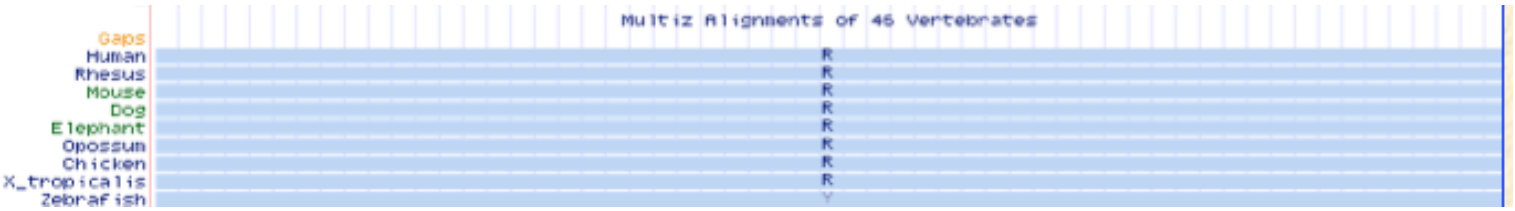
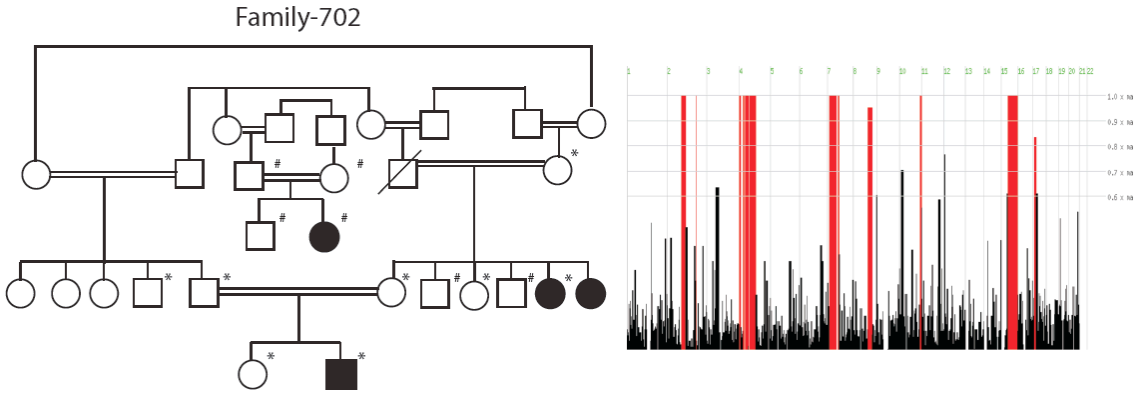
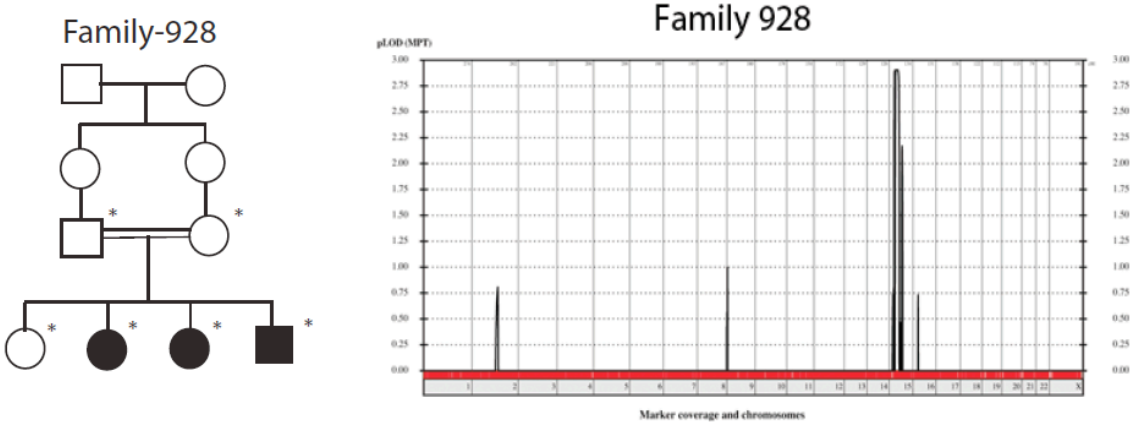


Figure S4. Genetic data for family 1004. Pedigree, linkage analysis, final variant table, and conservation figure for *GJC2* residue 35. R35 is conserved to *Xenopus* in vertebrates.



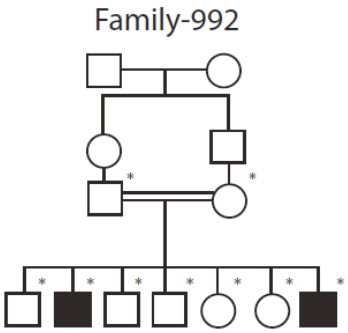
Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
8	100589861	G	T	splice	Damaging	5.4	*VPS13B*	Vacuolar protein sorting-associated protein 13B	Yes	No
7	20795056	C	T	p.Q1195*	Unknown	3.97	*ABCB5*	ATP-binding cassette sub-family B member 5	No	-
7	66274027	A	G	p.Y411C	Tolerated	5.82	*RABGEF1*	Rab5 GDP/GTP exchange factor	-	-
15	77471939	G	A	p.P777L	Damaging	5.54	*SGK269*	NKF3 kinase family member	-	-
15	40751721	C	T	p.S353L	Tolerated	5.5	*BAHD1*	Bromo adjacent homology domain-containing 1	-	-
2	113496452	C	T	p.R729H	Damaging	5.5	*CKAP2L*	Cytoskeleton-associated protein 2-like	-	-
4	55127448	G	A	p.G79D	Damaging	5.5	*PDGFRA*	Alpha-type platelet-derived growth factor receptor	-	-
4	57314612	C	G	p.S141C	Tolerated	5.48	*PAICS*	Multifunctional protein ADE2	-	-
2	25042941	G	A	p.R1099C	Damaging	5.2	*ADCY3*	Adenylate cyclase type 3	-	-
6	7395350	G	A	p.R114Q	Unknown	5.2	*RIOK1*	RIO kinase 1	-	-
1	237058743	G	A	p.R1164H	Tolerated	5.14	*MTR*	Methionine synthase	-	-
4	87684148	G	T	p.W1255C	Tolerated	5.1	*PTPN13*	Tyrosine-protein phosphatase non-receptor type 13	-	-
15	75897537	T	C	p.K211R	Tolerated	5.04	*SNUPN*	Snurportin-1	-	-
4	71472164	T	C	p.L354P	Tolerated	5.03	*AMBN*	Ameloblastin Precursor	-	-
7	33976935	C	T	p.P85L	Tolerated	4.83	*BMPER*	BMP-binding endothelial regulator protein	-	-
1	236645609	C	T	p.S93F	Tolerated	4.76	*EDARADD*	Ectodysplasin-A receptor-associated adapter	-	-
15	67457698	A	G	p.I65V	Tolerated	4.6	*SMAD3*	Mothers against decapentaplegic homolog 3	-	-
15	74174064	G	A	p.R83Q	Tolerated	4.57	*TBC1D21*	TBC1 domain family member 21	-	-
4	944254	G	A	p.V80I	Tolerated	4.55	*TMEM175*	Transmembrane protein 175	-	-
2	105961794	C	T	p.A222V	Tolerated	4.47	*C2orf49*	Ashwin	-	-

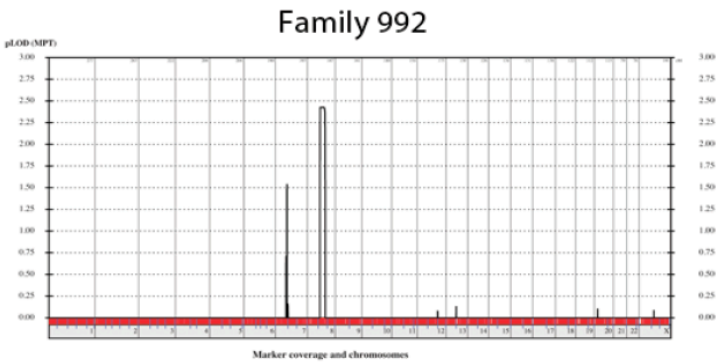
Figure S5. Genetic data for family 702. Pedigree, homozygosity analysis, and final variant table.



Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
2	202572665	CAG	C	p.144fs	DAMAGING	N/A	*ALS2*	Alsin	Yes	No
2	201515700	C	T	splice	N/A	4.02	*AOX1*	Aldehyde oxidase	No	-
2	203765756	T	C	p.I75V	TOLERATED	4.35	*WDR12*	Ribosome biogenesis protein WDR12	No	-
2	203846817	A	T	p.Y571F	TOLERATED	4.12	*ALS2CR 8*	Amyotropic lateral sclerosis 2 region gene 8	No	-
2	219029339	C	T	p.R199Q	DAMAGING	4.51	*CXCR1*	Chemokine CXC motif receptor 1	No	-
2	219507166	G	A	p.A1358V	TOLERATED	3.57	*ZNF142*	Zinc finger protein 142	No	-
20	60887356	G	A	p.T3126I	TOLERATED	4.01	*LAMA5*	Laminin subunit alpha-5	No	-

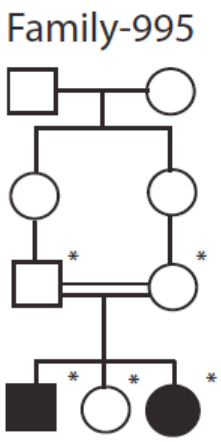
Figure S6. Genetic data for family 928. Pedigree, linkage analysis, and final variant table.

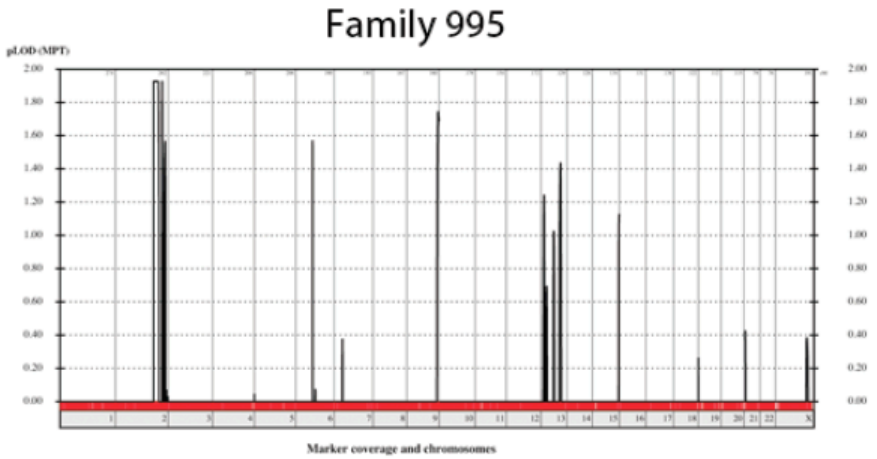




Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
8	1001233409	G	T	p.E222*	DAMAGING	4.88	*VSP13B*	Vacuolar protein sorting associated 13B	Yes	No
8	94793190	A	T	p.S239C	DAMAGING	5.49	*TMEM67*	Meckelin	No	-
8	94784818	G	C	splice	TOLERATED	5.24	*TMEM67*	Meckelin	No	-
8	98656966	G	T	p.A78S	TOLERATED	5.01	*MTDH*	Protein LYRIC, metadherin	No	-

Figure S7. Genetic data for family 992. Pedigree, linkage analysis, and final variant table.





Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
2	202626188	C	A	p.G177C	DAMAGING	5.47	*ALS2*	Alsin	Yes	No
2	196748512	T	C	p.I1854V	TOLERATED	5.68	*DNAH7*	Dyein heavy chain 7, axonemal	No	-
2	213921685	T	C	p.N93S	TOLERATED	4.74	*IKZF2*	Zinc finger protein Helios	No	-
2	214160817	G	A	p.D56N	TOLERATED	5.1	*SPAG16*	Sperm-associated antigen 16 protein	No	-
2	214354811	G	A	p.S356N	DAMAGING	4.78	*SPAG16*	Sperm-associated antigen 16 protein	No	-
2	28472253	T	C	p.D161G	NOT SCORED	4.26	*GPX6*	Glutathione peroxidase 6	No	-
2	29012712	T	C	p.M81V	TOLERATED	3.02	*OR2W1*	Olfactory receptor 2W1	No	-
2	32024395	C	T	p.R2704H	TOLERATED	3.62	*TNXB*	Putative tenascin-XA	No	-
2	33243751	G	C	p.V67L	TOLERATED	4.85	*RPS18*	40S ribosomal protein S18	No	-
15	75117912	G	T	p.R516M	TOLERATED	3.77	*LMAN1L*	Protein ERGIC-53-like precursor	No	-

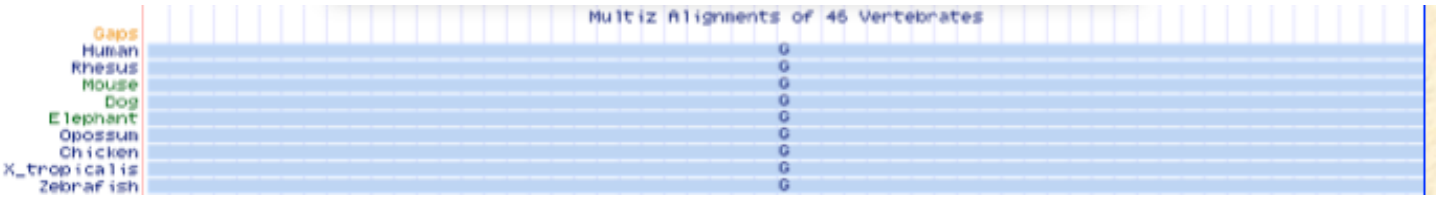
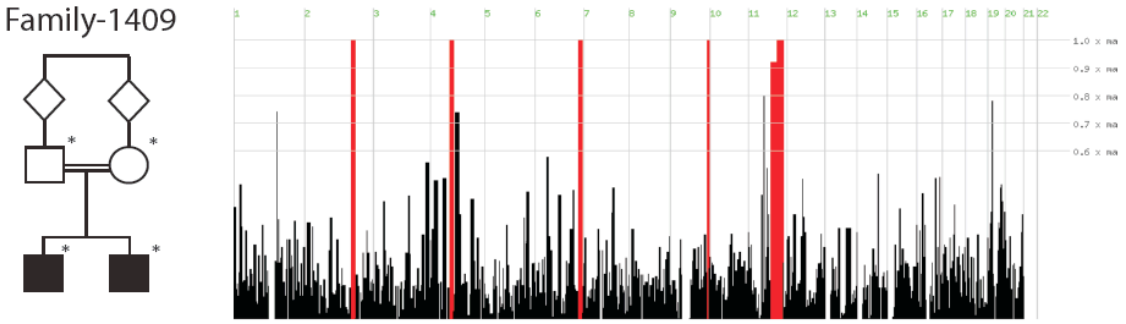
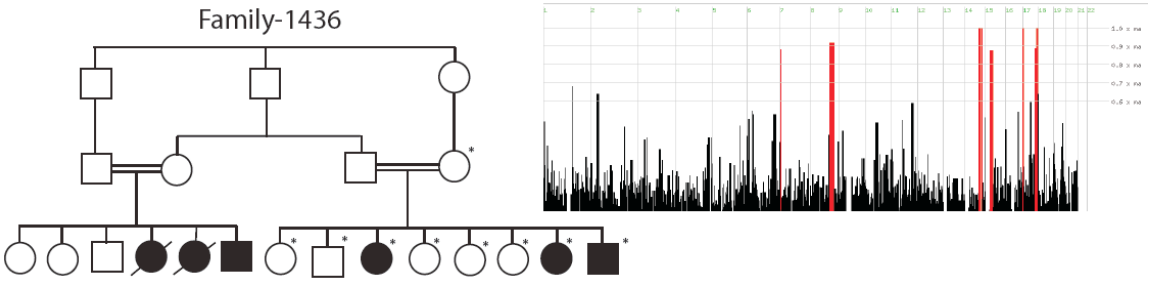


Figure S8. Genetic data for family 995. Pedigree, linkage analysis, final variant table, and conservation figure for *ALS2* residue 177. G177 is conserved in vertebrates.



Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
9	136218932	G	A	p.Q273*	UNKNOWN	5.04	*SURF1*	Surfeit locus protein 1	Yes	No
4	40103817	G	A	p.E118K	TOLERATED	5.73	*N4BP2*	NEDD4 binding protein 2	No	-
4	40123790	T	G	p.S1353R	TOLERATED	5.61	*N4BP2*	NEDD4 binding protein 2	No	-
17	1551223	C	T	p.G284S	TOLERATED	4.99	*RILP*	RAB interacting lysosomal protein	No	-
9	136305530	C	T	p.P618A	DAMAGING	4.77	*ADAMTS13*	A disintergrin and metalloproteinase motif 13	-	-
X	23731321	T	C	splice	TOLERATED	4.26	*ACOT9*	Acyl-coenzyme A thioesterase 9	No	-
X	16859867	C	T	p.P390L	TOLERATED	4.21	*TXLNG*	Gamma-taxillin	-	-
6	161152819	C	T	p.A494V	TOLERATED	4.12	*PLG*	Plasminogen precursor	-	-
6	159188474	G	A	p.P472L	TOLERATED	3.83	*EZR*	Ezrin	-	-
11	65318751	G	A	p.A438V	DAMAGING	3.64	*LTBP3*	Latent transforming growth factor beta 3	-	-
9	133936490	A	G	p.N743D	TOLERATED	3.62	*LAMC3*	Laminin subunit gamma 3	-	-

Figure S9. Genetic data for family 1409. Pedigree, homozygosity analysis, and final variant table.



Chr	Position	RefNT	AltNT	AA Change	Prediction	Conserv	Gene	Gene Description	Segregates in Family	Present in 200 Controls
14	77743795	C	T	p.G726E	DAMAGING	5.12	*POMT2*	Protein O-mannosyl-transferase 2	Yes	No
8	99142224	C	G	p.Q169E	TOLERATED	4.86	*POP1*	Ribonuclease P/MRP protein subunit POP1	No	-
14	77275486	G	A	p.P189S	TOLERATED	3.52	*ANGEL1*	Protein angel homolog 1	No	-
17	74208560	C	T	splice	N/A	N/A	*RNF157*	RING finger protein 157	No	-

Figure S10. Genetic data for family 1436. Pedigree, homozygosity analysis, and final variant table.

	Dominant or de novo	Recessive without consanguinity	Recessive with consanguinity
Total variants	26K +/- 2K	26K +/- 2K	26K +/- 2K
Exonic and splice variants	19K +/- 3K	19K +/- 3K	19K +/- 3K
Non-synonymous +Splice + INDELs	10K +/- 2K	10K +/- 2K	10K +/- 2K
Variant either not present or rare (<1:300) in dbSNP	400 +/- 200	400 +/- 200	400 +/- 200
Variants not present in our in-house database (unique)	150 +/- 50	150 +/- 50	150 +/- 50
Compatible with inheritance model	Heterozygous 85 +/- 30	Compound Heterozygous 10 +/- 5	Homozygous 5 +/- 4
Segregates in family	40 +/- 20	5 +/- 5	1 +/- 2

Exome sequencing generally produces 26,000 +/- 2000 (26K +/- 5K) variants across the genome from a single individual, irrespective of the mode of disease. Filtering those that do not occur in exons or at potential splice sites leaves 19K +/- 3K, then filtering for those that are not non-synonymous, at splice sites, or produce insertion/deletion polymorphisms (INDELs) leaves 10K +/- 2K. Further filtering of variants reported in the public databases at rates over that expected from Hardy-Weinberg equilibrium (i.e. >1:300) leaves 400 +/- 200. Filtering non-unique variants from our in-house variant database leaves 150 +/- 50 variants to consider. At this point, the mode of inheritance becomes paramount in isolating the disease-causing gene. Specifically, a recessive disease from a consanguineous family should result from a homozygous mutation that segregates in the family as a recessive trait. On average 5 +/- 4 variants will meet this criterion and only 1 +/- 2 variants will actually segregate in the family. For these reasons, identifying disease-causing genetic mutations is more robust in consanguineous families with recessive disease (data compiled from published literature and personal experience).

Table S1. Estimated variants to be considered as causative from WES are greatly reduced in recessive disease with documented consanguinity.

Family	Exome coverage	Total variants	#Coding, Splice variants	#Non-synonymous variants	#Homozygous recessive variants	#Variants absent or rare in dbSNP	#Variants not present in-house database	#Variants in linkage intervals	Type of variants	#Variants conserved across evolution	#Predicted damaging variants
890	94.7%	21,693	14,451	7,757	2,961	79	33	7	1 deletion 6 missense	1	1
951	96.9%	27,112	20,607	10,610	3,374	91	26	10	1 nonsense 9 missense	2	2
1002	97.4%	26,119	17,553	10,286	3,371	96	37	8	1 splice 2 missense	2	1
1004	97.1%	26,932	18,653	10,809	3,197	83	17	5	5 missense	3	2
702	97.2%	27,454	18,795	11,515	3,522	112	37	21	1 nonsense 2 splice 18 missense	10	5
928	97.2%	27,423	19,853	10,598	3,123	78	15	7	1 nonsense 1 splice 5 missense	1	1
992	97.2%	27,155	17,814	10,892	3,287	75	19	4	1 nonsense 1 splice 2 missense	4	1
995	96.9%	27,748	19,605	10,983	3,310	101	15	10	10 missense	3	1
1409	97.2%	26,426	18,683	10,356	3,236	85	26	11	1 nonsense 1 splice 9 missense	3	1
1436	95.4%	25,864	23,334	9,661	3,022	78	18	4	1 splice 3 missense	2	1

On average, the total number of variants identified per family was 26K. Approximately 19K of those were found in coding and splice regions. Approximately 10K of the 19K were found to be non-synonymous missense, splice or deletion/insertion polymorphisms. The remaining ~10K variants could then be further filtered down by the number of homozygous variants, the number absent or rare in dbSNP or in our in-house databases, and the number found in linkage intervals. The remaining variants were then prioritized based on type of mutation, conservation of the affected amino acids or base pairs across species, and predicted damage to the protein.

Table S2. Number of variants identified in each family at each step of the variant filtering and prioritization pipeline.