The American Journal of Human Genetics, Volume *100*

Supplemental Data

The Rare-Variant Generalized Disequilibrium Test for Association Analysis of Nuclear and Extended Pedigrees with Application to Alzheimer Disease WGS Data

Zongxiao He, Di Zhang, Alan E. Renton, Biao Li, Linhai Zhao, Gao T. Wang, Alison M. Goate, Richard Mayeux, and Suzanne M. Leal

Supplemental Acknowledgements

The Alzheimer's Disease Sequencing Project (ADSP) is comprised of two Alzheimer's Disease (AD) genetics consortia and three National Human Genome Research Institute (NHGRI) funded Large Scale Sequencing and Analysis Centers (LSAC). The two AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institute of Health (NIH) institutes and other foreign governmental and non-governmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Schellenberg, Farrer, Pericak-Vance, Mayeux, and Haines); RF1AG015473 to Dr. Mayeux; U01AG049505 to Dr. Seshadri; U01AG049506 to Dr. Boerwinkle; U01AG049507 to Dr. Wijsman; and U01AG049508 to Dr. Goate.

The ADGC cohorts include: Adult Changes in Thought (ACT), the Alzheimer's Disease Centers (ADC), the Chicago Health and Aging Project (CHAP), the Memory and Aging Project (MAP), Mayo Clinic (MAYO), Mayo Parkinson's Disease controls, University of Miami, the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE), the National Cell Repository for Alzheimer's Disease (NCRAD), the National Institute on Aging Late Onset Alzheimer's Disease Family Study (NIA-LOAD), the Religious Orders Study (ROS), the Texas Alzheimer's Research and Care Consortium (TARC), Vanderbilt University/Case Western Reserve University (VAN/CWRU), the Washington Heights-Inwood Columbia Aging Project (WHICAP) and the Washington University Sequencing Project (WUSP), the Columbia University Hispanic- Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA), the University of Toronto (UT), and Genetic Differences (GD**)**.

The CHARGE cohorts, with funding provided by 5RC2HL102419 and HL105756, include the following: Atherosclerosis Risk in Communities (ARIC) Study which is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), Austrian Stroke Prevention Study (ASPS), Cardiovascular Health Study (CHS), Erasmus Rucphen Family Study (ERF), Framingham Heart Study (FHS), and Rotterdam Study (RS).

The three LSACs are: the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), and the Washington University Genome Institute (U54HG003079).

Biological samples and associated phenotypic data used in primary data analyses were stored at Study Investigators institutions, and at the National Cell Repository for Alzheimer's Disease (NCRAD, U24AG021886) at Indiana University funded by NIA. Associated Phenotypic Data used

in primary and secondary data analyses were provided by Study Investigators, the NIA funded Alzheimer's Disease Centers (ADCs), and the National Alzheimer's Coordinating Center (NACC, U01AG016976) and the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS, U24AG041689) at the University of Pennsylvania, funded by NIA, and at the Database for Genotypes and Phenotypes (dbGaP) funded by NIH. Contributors to the Genetic Analysis Data included Study Investigators on projects that were individually funded by NIA, and other NIH institutes, and by private U.S. organizations, or foreign governmental or nongovernmental organizations.

Supplemental Figures and Tables

Figure S1. QQ plot of the $-\log_{10}$ p-values for data simulated under the null hypothesis of no association.

P-values for the RV-GDT were obtained empirically using 100,000 permutations. Confidence intervals are highlighted in gray. Four different types of family data were investigated: 1,000 discordant sib-pairs (Figure 1 pedigree structure A), 1,000 affected sib-pairs (Figure 1 pedigree structure B), 1,000 extended pedigrees (Figure 1 pedigree structure C), and mixed family types including 500 discordant sib-pairs, 250 affected sib-pairs and 250 extended pedigrees. Panel A: genotypes simulated using ExAC Non-Finnish European variant information; Panel B: genotypes simulated using ExAC Non-Finnish European variant information with ~50% of the founders missing their genotype data. Panel C: genotypes were simulated for an admixed population (20% Non-Finnish Europeans and 80% African/African American); Panel D: genotypes were simulated for a population with substructure (50% Non-Finnish European families and 50% African/African American families).

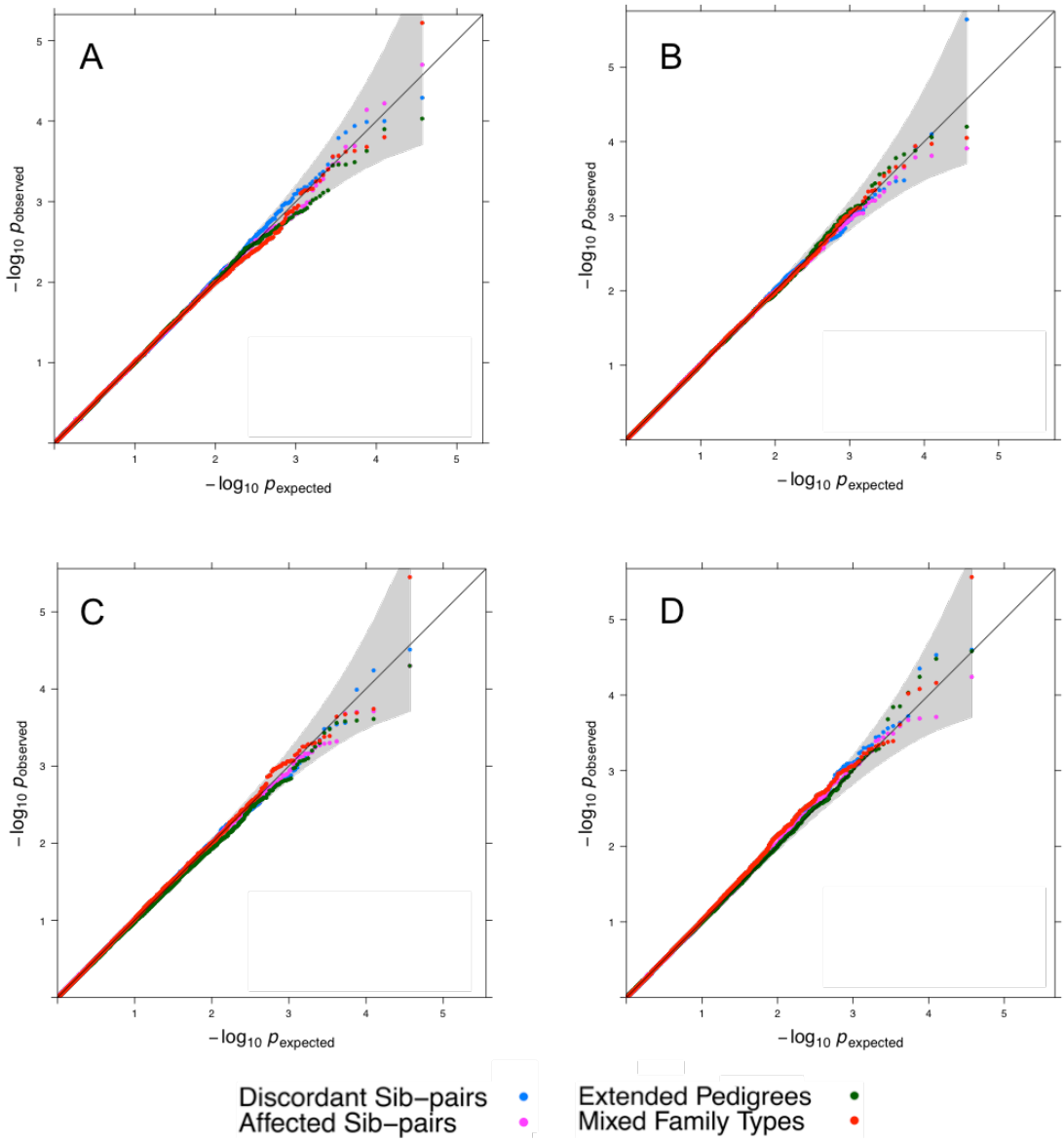
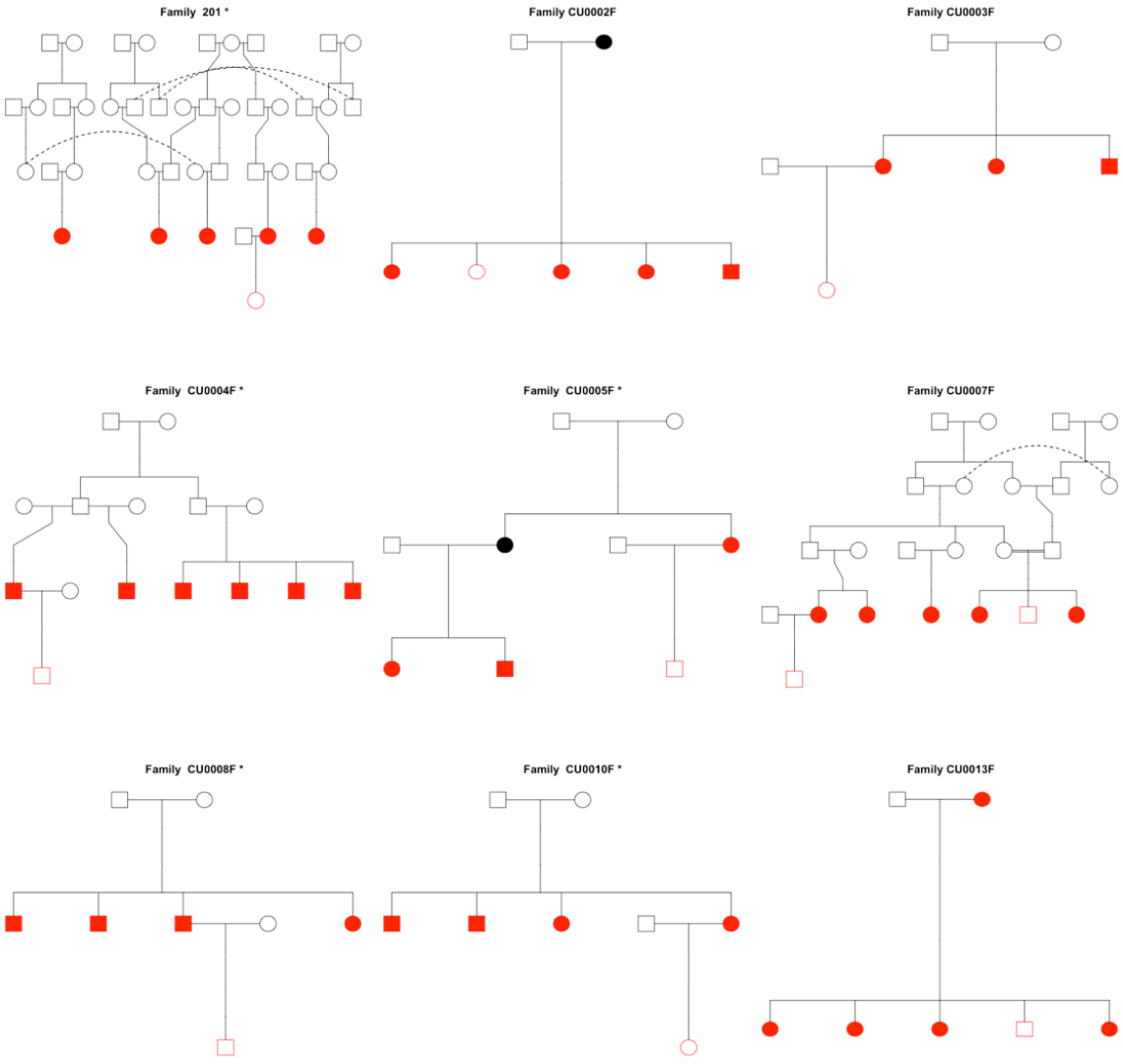
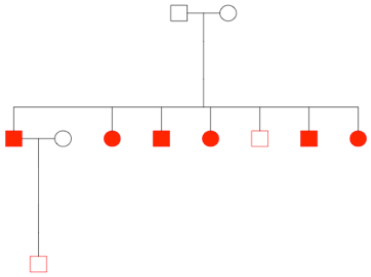

A
B
C
D
$-\log_{10} p_{observed}$
$-\log_{10} p_{expected}$
Discordant Sib−pairs
Affected Sib−pairs
Extended Pedigrees
Mixed Family Types

Figure S2. Eighty-one Alzheimer's disease pedigrees included in the analysis from Alzheimer's Disease Sequencing Project.

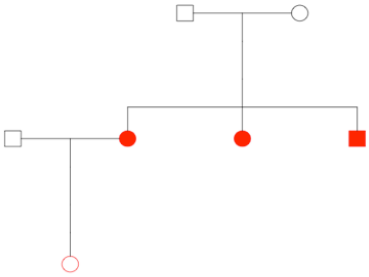
The dataset includes 338 individuals with Alzheimer's disease of which 316 have whole genome sequence data and 494 unaffected individuals of which 98 have whole genome sequence data. Filled squares and circles represent males and females with Alzheimer's disease, respectively, while family members with open squares and circles represent unaffected individuals. Individuals represented in red have whole genome sequence data available, while those in black do not have genotype data. There are 25 families that cannot be analyzed by the FBAT software, and these families are marked with asterisks (*).



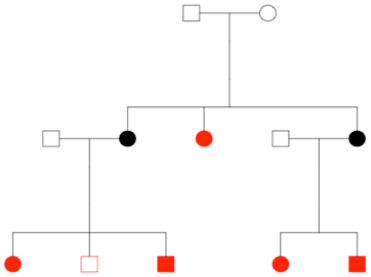
Family CU0014F



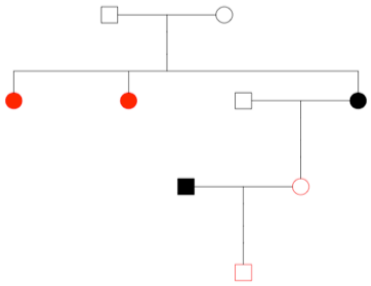
Family CU0015F



Family CU0016F

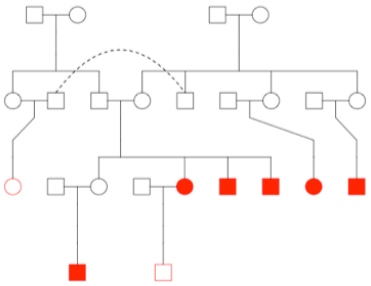


Family CU0017F *

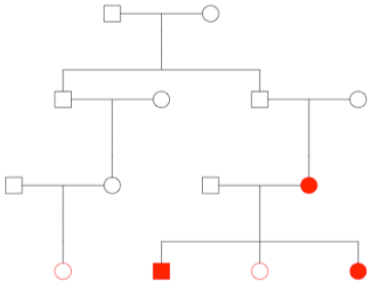


Family CU0021F

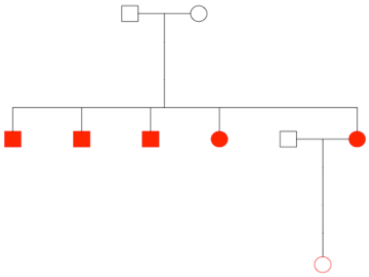
Family CU0023F



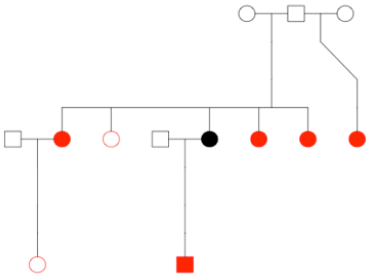
Family CU0026F



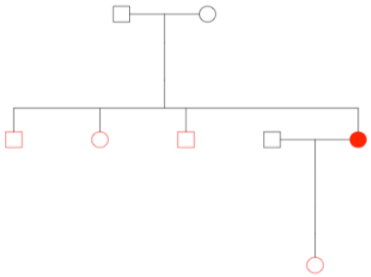
Family CU0029F *



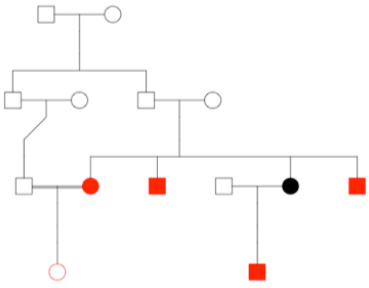
Family CU0030F



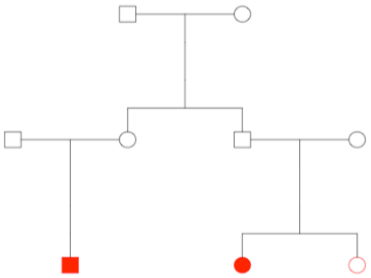
Family CU0031F



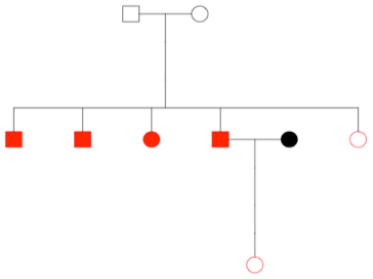
Family CU0032F



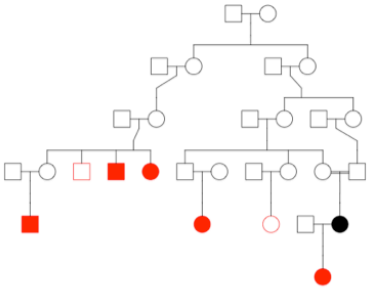
Family CU0033F



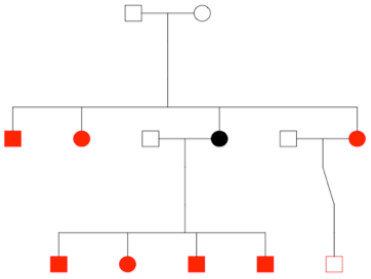
Family CU0035F



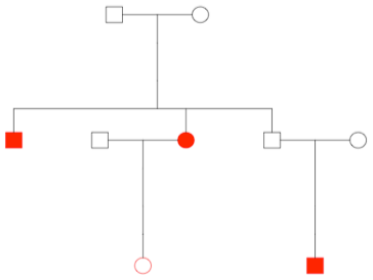
Family CU0037F



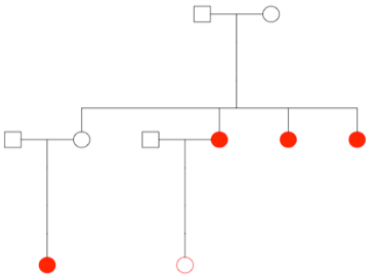
Family CU0039F *



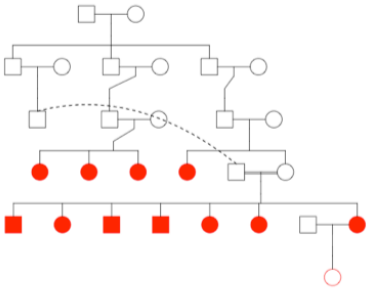
Family CU0042F *



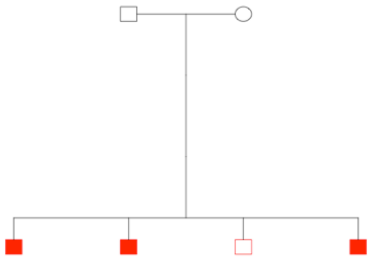
Family CU0043F



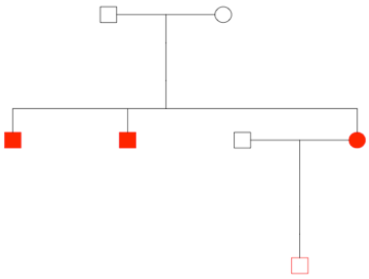
Family CU0044F



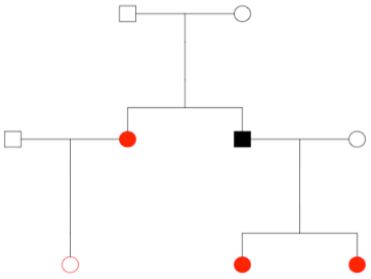
Family CU0045F



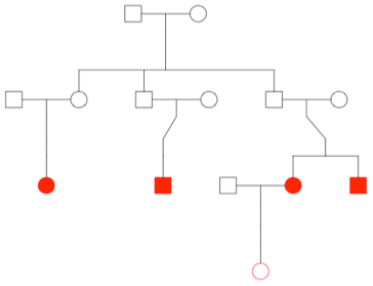
Family CU0046F



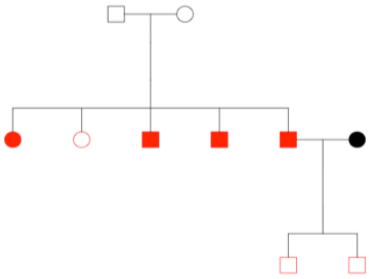
Family CU0047F *



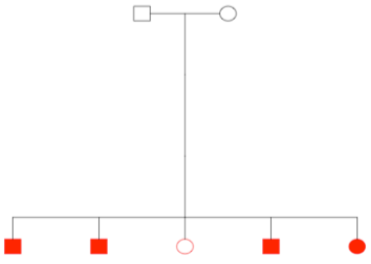
Family CU0048F *



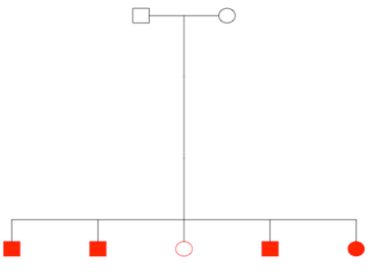
Family CU0049F



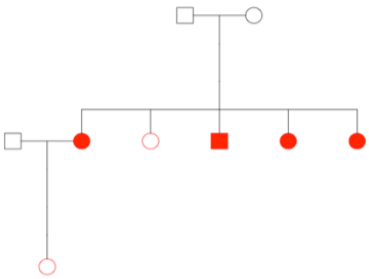
Family CU0050F



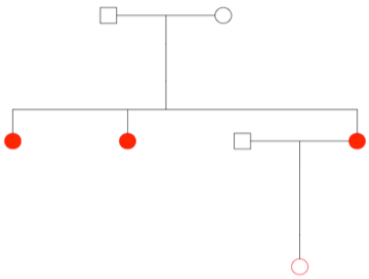
Family CU0052F



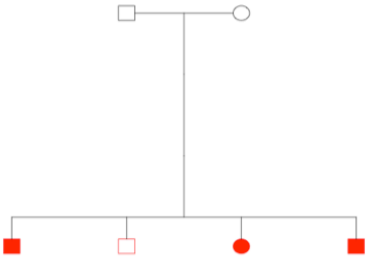
Family CU0053F



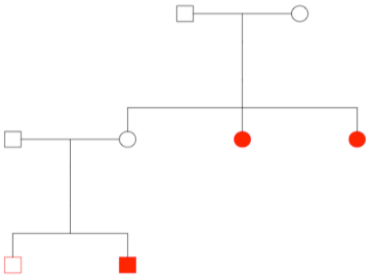
Family CU0057F



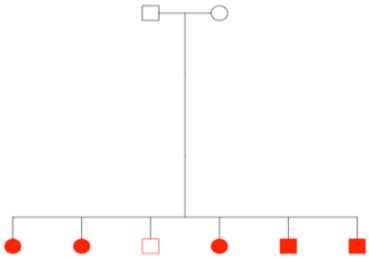
Family CU0058F



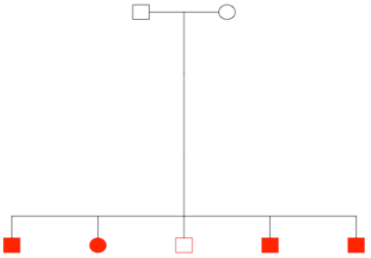
Family CU0060F



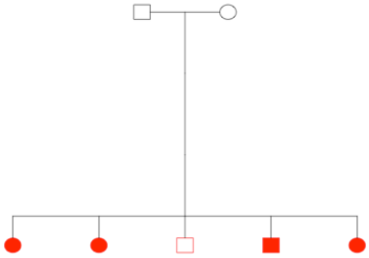
Family CU0064F



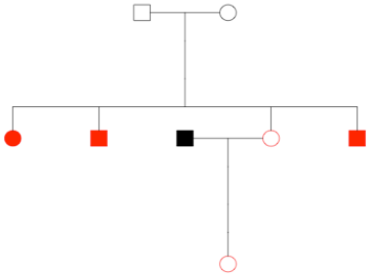
Family CU0065F



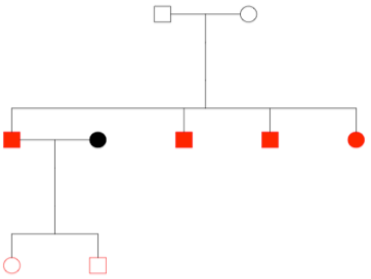
Family CU0067F



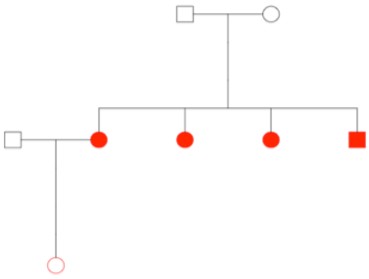
Family CU0068F



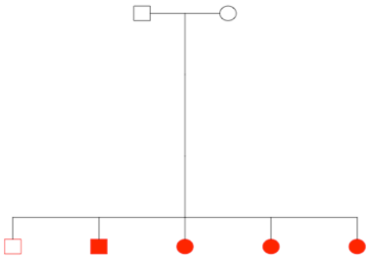
Family CU0071F *

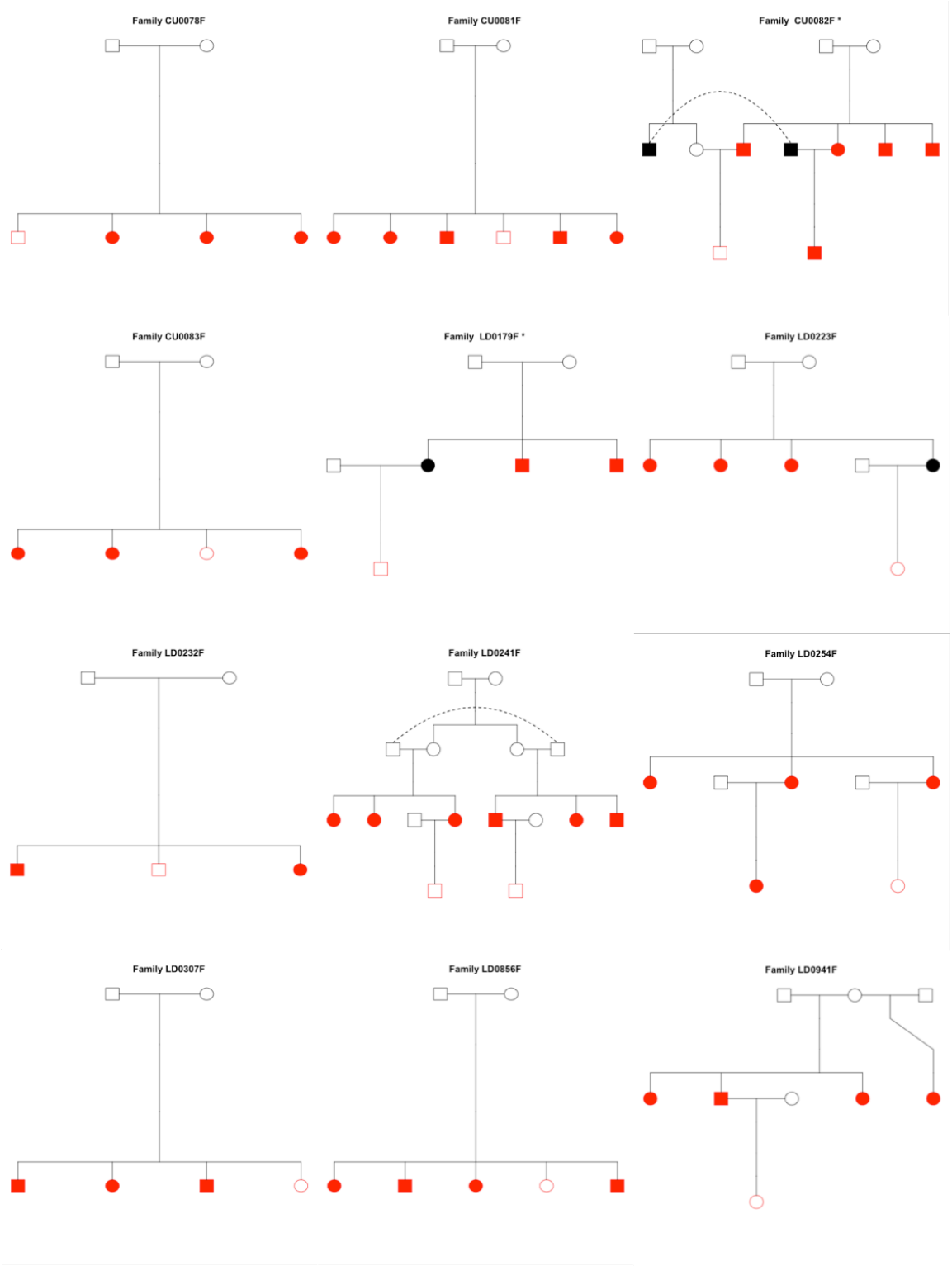


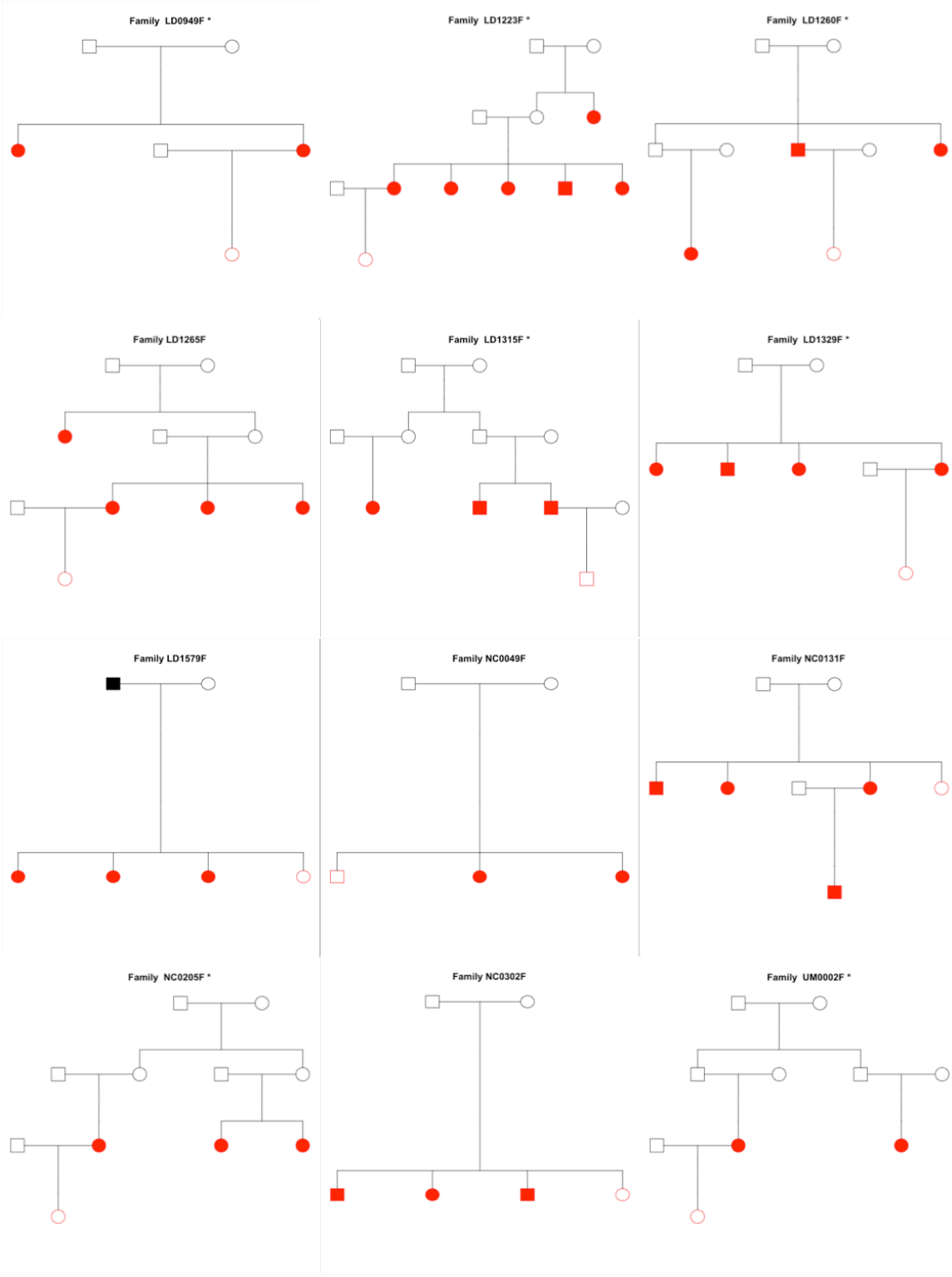
Family CU0073F *



Family CU0076F




Family CU0078F
Family CU0081F
Family CU0082F *
Family CU0083F
Family LD0179F *
Family LD0223F
Family LD0232F
Family LD0241F
Family LD0254F
Family LD0307F
Family LD0856F
Family LD0941F


Family LD0949F *
Family LD1223F *
Family LD1260F *
Family LD1265F
Family LD1315F *
Family LD1329F *
Family LD1579F
Family NC0049F
Family NC0131F
Family NC0205F *
Family NC0302F
Family UM0002F *

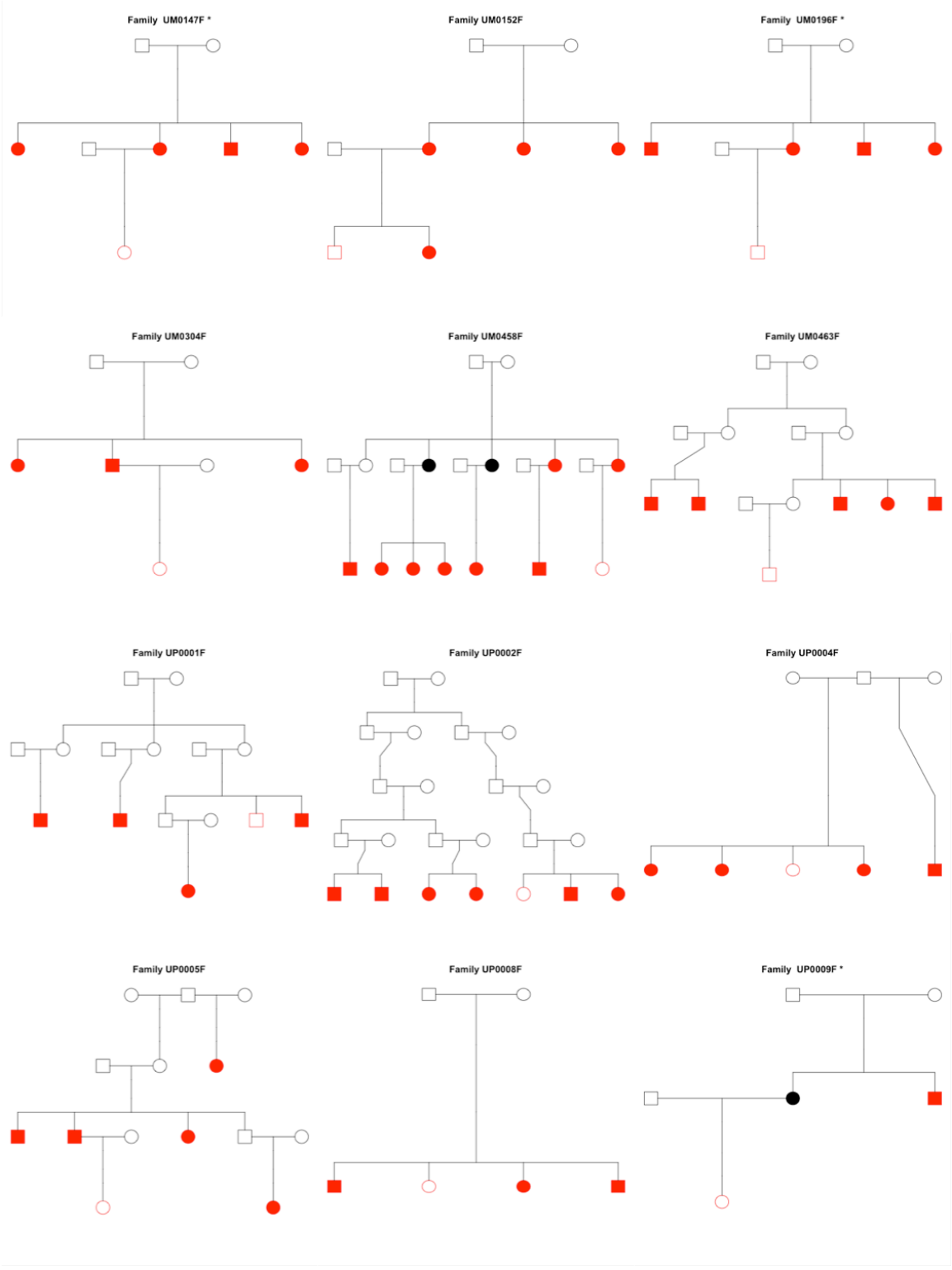

Family UM0147F *
Family UM0152F
Family UM0196F *
Family UM0304F
Family UM0458F
Family UM0463F
Family UP0001F
Family UP0002F
Family UP0004F
Family UP0005F
Family UP0008F
Family UP0009F *

Table S1. Type I error rate for RareIBD at α levels of 0.05

	Discordant Sib-pair	Affected Sib-pair	Extended Pedigree	Mixed Family Types
RareIBD	0.064	0.033	0.465	0.094
	Each Founder has a probability of 50% to be missing all of their genotype data			
RareIBD	0.064	0.046	0.296	0.091
	80% African and 20% European population admixture			
RareIBD	0.059	0.028	0.395	0.086
	50% African and 50% European families			
RareIBD	0.063	0.030	0.476	0.089

One-thousand families for each pedigree structure shown in Figure 1 and mixed pedigree structures were simulated. Genotype data were generated for all autosomal genes across the genome with an OR = 1.0, and type I error rate was defined as the proportion of genes with a p-value less than 0.05. Variant information for 17,987 autosomal genes from ExAC Non-Finnish European were used to generate family data when each founder has a probability (0% or 50%) to be missing their genotype data. Variant information for 17,873 autosomal genes that are present in both ExAC Non-Finnish European and African/African American populations was used to generate family data with population admixture and substructure.

Table S2. The Ethnicities of Alzheimer's disease families included in the analysis

Ethnicity	Number of Families	Family IDs
Dominican	46	CU0002F, CU0003F, CU0004F, CU0005F, CU0007F, CU0008F, CU0010F, CU0013F, CU0014F, CU0015F, CU0016F, CU0017F, CU0021F, CU0023F, CU0026F, CU0029F, CU0030F, CU0031F, CU0033F, CU0035F, CU0037F, CU0039F, CU0043F, CU0044F, CU0045F, CU0046F, CU0047F, CU0048F, CU0049F, CU0050F, CU0052F, CU0053F, CU0057F, CU0058F, CU0060F, CU0064F, CU0065F, CU0067F, CU0068F, CU0071F, CU0073F, CU0076F, CU0078F, CU0081F, CU0082F, CU0083F
European Descent	31	LD0179F, LD0223F, LD0232F, LD0241F, LD0254F, LD0307F, LD0856F, LD0949F, LD1223F, LD1260F, LD1265F, LD1315F, LD1329F, LD1579F, NC0049F, NC0131F, NC0205F, NC0302F, UM0002F, UM0147F, UM0152F, UM0196F, UM0304F, UM0458F, UM0463F, UP0001F, UP0002F, UP0004F, UP0005F, UP0008F, UP0009F
Puerto Rican	2	CU0032F, CU0042F
African American	1	LD0941F
Dutch Isolate	1	201

Table S3. Bioinformatic evaluation and frequencies of analyzed rare variants within *MARCH10*

dbSNP rsID	rs146326363*	rs116835087*^	rs147046907*	rs60472825	rs138015683	rs374880698*^	rs141415486*	rs78457484
hg19 Position	17:60782924	17:60813470	17:60813550	17:60813944	17:60813982	17:60814417	17:60837208	17:60837337
Reference allele	G	C	G	G	T	A	C	G
Alternate allele	C	T	A	A	C	C	T	A
cDNA change	c.2347C>G	c.1759G>A	c.1679C>T	c.1285C>T	c.1247A>G	c.812T>G	c.370G>A	c.241C>T
Amino acid change	p.Gln783Glu	p.Gly587Ser	p.Thr560Ile	p.His429Tyr	p.Asn416Ser	p.Phe271Cys	p.Glu124Lys	p.Pro81Ser
ExAC all MAF	0.0005	0.0037	0.0068	0.0056	0.0010	1.63E-05	0.0011	0.0072
Number of alternative alleles - AD pedigree members (n=316)	2	4	10	19	1	3	1	13
Number of alternative alleles - unaffected pedigree members (n=98)	1	0	0	4	1	0	0	0
GERP score	4.24	4.4	3.32	3	-0.978	2.61	4.79	0.166
PhyloP score	2.254	1.601	1.565	0.095	-0.474	1.492	2.695	0.711
CADD score, scaled	19.6	22.4	6.022	12.46	4.402	22.4	15.74	9.87
FATHMM	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated
MutationTaster	disease-causing	disease-causing	polymorphism	polymorphism	polymorphism	disease-causing	disease-causing	polymorphism automatic
Polyphen-2 HVAR	benign	possibly damaging	benign	benign	Benign	probably damaging	benign	possibly damaging
PROVEAN	neutral	neutral	neutral	neutral	neutral	deleterious	neutral	neutral
SIFT	tolerated	damaging	tolerated	damaging	tolerated	damaging	tolerated	damaging

Abbreviations are as follows: ExAC, Exome Aggregation Consortium; MAF, minor allele frequency; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant. Conservation scores and bioinformatics results as compiled by dbNSFP v.2.9.

*Variant is deemed as conserved nucleotide (both GERP and PhyloP scores > 1).

^Variant is deemed damaging by at least three of six bioinformatics tools (variant with CADD scaled score >15 is deemed to be deleterious).

Table S4. Bioinformatic Evaluation and Frequencies of Rare Missense Variants within *AMBN*

dbSNP rsID	rs143940501*^	rs146167261*^	rs115723025	rs113506649	rs139319140*^	rs150017698*^	NA	rs76503327
hg19 Position	4:71469167	4:71471905	4:71471958	4:71471985	4:71472001	4:71472053	4:71472146	4:71472235
Reference allele	C	G	G	C	G	A	C	G
Alternate allele	T	A	A	A	A	C	T	A
cDNA change	c.743C>T	c.802G>A	c.855G>A	c.882C>A	c.898G>A	c.950A>C	c.1043C>T	c.1132G>A
Amino acid change	p.Ala248Val	p.Gly268Arg	p.Met285Ile	p.His294Gln	p.Gly300Ser	p.Glu317Ala	p.Ala348Val	p.Val378Ile
ExAC all MAF	0.0007	0.0026	0.0103	0.0053	0.0004	0.0004	NA	0.0068
Number of alternative alleles - AD pedigree members (n=316)	2	4	14	4	1	0	3	13
Number of alternative alleles - unaffected pedigree members (n=98)	0	0	1	0	0	1	0	2
GERP score	5.07	5.79	2.73	-1.87	4.95	4.66	1.98	4.07
PhyloP score	2.394	3.499	0.453	-1.137	2.596	3.297	0.204	0.908
CADD score, scaled	24.6	26.3	10.97	0.023	26.1	19.31	0.002	10.97
FATHMM	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated
MutationTaster	disease-causing	disease-causing	polymorphism	polymorphism	disease-causing	disease-causing	polymorphism	polymorphism
Polyphen-2 HVAR	possibly damaging	probably damaging	benign	benign	probably damaging	probably damaging	benign	benign
PROVEAN	neutral	deleterious	neutral	neutral	deleterious	deleterious	neutral	neutral
SIFT	damaging	damaging	damaging	tolerated	damaging	tolerated	tolerated	tolerated

Abbreviations are as follows: NA, Not Available; ExAC, Exome Aggregation Consortium; MAF, minor allele frequency; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant. Conservation scores and bioinformatics results as compiled by dbNSFP v.2.9.

*Variant is deemed as conserved nucleotide (both GERP and PhyloP scores > 1).

^Variant is deemed damaging by at least three of six bioinformatics tools (variant with CADD scaled score >15 is deemed to be deleterious).

Table S5. Bioinformatic Evaluation and Frequencies of Rare Missense Variants within *TCOF1*

dbSNP rsID	rs56180593^	rs142477153*^	rs181203524	rs144327167*^	rs189476787^	rs75181211^	rs114326915	rs201458471	rs75583421^
hg19 Position	5:149740732	5:149747428	5:149748403	5:149753894	5:149754226	5:149754229	5:149754325	5:149754948	5:149755362
Reference allele	C	A	C	G	C	C	C	T	G
Alternate allele	T	G	T	A	T	T	T	C	A
cDNA change	c.122C>T	c.326A>G	c.503C>T	c.797G>A	c.899C>T	c.902C>T	c.998C>T	c.1304T>C	c.1552G>A
Amino acid change	p.Ala41Val	p.Asn109Ser	p.Thr168Met	p.Ser266Asn	p.Pro300Leu	p.Ala301Val	p.Ser333Leu	p.Met435Thr	p.Val518Ile
ExAC all MAF	0.0023	0.0001	0.0004	0.0034	0.0002	0.0035	0.0038	0.0001	0.0077
Number of alternative alleles - AD pedigree members (n=316)	4	1	7	2	3	13	2	3	6
Number of alternative alleles - unaffected pedigree members (n=98)	0	0	0	0	0	2	1	1	1
GERP score	2.98	4.33	-1.6	2.98	1.5	2.63	-0.0656	1.02	1.02
PhyloP score	0.602	2.068	-0.037	4.082	0.661	0.386	1.055	-0.049	0.48
CADD score, scaled	22.3	23.2	7.967	22.9	12.91	16.34	6.996	0.009	18.7
FATHMM	damaging	damaging	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated
MutationTaster	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism
Polyphen-2 HVAR	benign	probably damaging	probably damaging	probably damaging	probably damaging	benign	benign	benign	possibly damaging
PROVEAN	neutral	deleterious	neutral	neutral	deleterious	deleterious	deleterious	neutral	neutral
SIFT	damaging	damaging	damaging	damaging	damaging	damaging	damaging	tolerated	damaging

Abbreviations are as follows: NA, Not Available; ExAC, Exome Aggregation Consortium; MAF, minor allele frequency; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant. Conservation scores and bioinformatics results as compiled by dbNSFP v.2.9.

*Variant is deemed as conserved nucleotide (both GERP and PhyloP scores > 1).

^Variant is deemed damaging by at least three of six bioinformatics tools (variant with CADD scaled score >15 is deemed to be deleterious).

Table S5. Bioinformatic Evaluation and Frequencies of Rare Missense Variants within *TCOF1* (continued)

rs201843484*^	rs73270846^	rs149384321	rs376431510	rs150515843^	rs114689020	NA^	rs200424706^	rs144793475*	rs55980697^
5:149755712	5:149756088	5:149756091	5:149759108	5:149759198	5:149759201	5:149759213	5:149771660	5:149772288	5:149776067
C	C	A	G	C	C	C	T	A	A
T	T	G	A	T	T	G	G	T	G
c.1730C>T	c.2014C>T	c.2017A>G	c.2441G>A	c.2531C>T	c.2534C>T	c.2546C>G	c.3207T>G	c.3304A>T	c.3773A>G
p.Ala577Val	p.Pro672Ser	p.Thr673Ala	p.Gly814Glu	p.Pro844Leu	p.Ser845Leu	p.Ala849Gly	p.Ser1069Arg	p.Thr1102Ser	p.Lys1258Arg
0.0003	0.0040	0.0001	4.08E-05	0.0015	0.0088	NA	0.0018	0.0001	0.0104
1	2	0	18	2	1	3	2	1	2
0	0	1	3	0	0	1	0	0	0
3.92	4.38	-10.7	1.61	2.82	0.389	0.811	-8.13	3.19	4.98
2.275	0.147	-4.381	0.338	0.68	-0.568	-0.368	-1.384	1.026	0.612
22.5	22.8	0.001	7.651	23.2	9.89	22	20.7	13.47	22.2
tolerated	tolerated	tolerated	tolerated	damaging	tolerated	tolerated	tolerated	tolerated	tolerated
polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	disease-causing
possibly damaging	possibly damaging	benign	benign	benign	benign	probably damaging	probably damaging	benign	probably damaging
deleterious	deleterious	neutral	deleterious	deleterious	neutral	deleterious	deleterious	neutral	neutral
damaging	tolerated	tolerated	tolerated	damaging	damaging	damaging	damaging	damaging	tolerated

Table S6. Bioinformatic Evaluation and Frequencies of Rare Missense Variants within *AXIN1*

dbSNP rsID	rs34015754*	rs117208012*	rs367788267*	rs141148118	rs149849071	rs146947903*	rs116350678*	rs140151215*
hg19 Position	16:338189	16:347063	16:347143	16:347930	16:347957	16:348021	16:348233	16:396221
Reference allele	C	C	G	C	C	G	C	G
Alternate allele	T	T	A	T	T	C	T	C
cDNA change	c.2522G>A	c.1948G>A	c.1868C>T	c.1576G>A	c.1549G>A	c.1485C>G	c.1273G>A	c.805C>G
Amino acid change	p.Arg841Gln	p.Gly650Ser	p.Ser623Leu	p.Ala526Thr	p.Val517Ile	p.Asp495Glu	p.Gly425Ser	p.Gln269Glu
ExAC all MAF	0.0085	0.0168	0.0004	0.0007	0.0016	0.0098	0.0075	0.0009
Number of alternative alleles - AD pedigree members (n=316)	2	3	1	3	1	2	19	3
Number of alternative alleles - unaffected pedigree members (n=98)	0	0	0	0	0	0	1	0
GERP score	4.43	2.63	4.98	-5.51	-10.1	3.79	3.31	5.1
PhyloP score	4.824	1.19	2.46	-1.489	-2.359	1.208	2.732	7.611
CADD score, scaled	13.68	8.312	16.57	0.018	0.001	10.82	13.84	23.6
FATHMM	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated
MutationTaster	disease-causing	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	disease-causing
Polyphen-2 HVAR	possibly damaging	benign	benign	benign	benign	benign	benign	benign
PROVEAN	neutral	neutral	neutral	neutral	neutral	neutral	neutral	neutral
SIFT	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated

Abbreviations are as follows: ExAC, Exome Aggregation Consortium; MAF, minor allele frequency; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant. Conservation scores and bioinformatics results as compiled by dbNSFP v.2.9.

*Variant is deemed as conserved nucleotide (both GERP and PhyloP scores > 1).

Table S7. Bioinformatic Evaluation and Frequencies of Rare Missense Variants within *TNK1*

dbSNP rsID	rs201180891*^	rs61730812*^	NA*^	rs56093628^	rs80015268*	rs141588799
hg19 Position	17:7286889	17:7287832	17:7287859	17:7291869	17:7291943	17:7292117
Reference allele	T	C	T	C	G	G
Alternate allele	G	A	A	G	C	A
cDNA change	c.380T>G	c.896C>A	c.923T>A	c.1637C>G	c.1711G>C	c.1802G>A
Amino acid change	p.Phe127Cys	p.Ala299Asp	p.Met308Lys	p.Ser546Cys	p.Gly571Arg	p.Ser601Asn
ExAC all MAF	3.42E-05	0.0114	NA	0.0022	0.0027	0.0005
Number of alternative alleles - AD pedigree members (n=316)	4	8	2	4	1	2
Number of alternative alleles - unaffected pedigree members (n=98)	0	0	0	0	0	0
GERP score	1.99	4.03	5.3	4.18	3.27	1.89
PhyloP score	5.177	2.355	6.029	0.375	3.52	0.736
CADD score, scaled	26.2	22.1	22.1	20.4	9.985	5.874
FATHMM	damaging	damaging	tolerated	tolerated	tolerated	tolerated
MutationTaster	disease-causing	disease-causing	disease-causing	polymorphism	polymorphism	polymorphism
Polyphen-2 HVAR	probably damaging	possibly damaging	probably damaging	possibly damaging	benign	benign
PROVEAN	deleterious	neutral	deleterious	neutral	neutral	neutral
SIFT	damaging	tolerated	damaging	damaging	damaging	damaging

Abbreviations are as follows: NA, Not Available; ExAC, Exome Aggregation Consortium; MAF, minor allele frequency; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant. Conservation scores and bioinformatics results as compiled by dbNSFP v.2.9.

*Variant is deemed as conserved nucleotide (both GERP and PhyloP scores > 1).

^Variant is deemed damaging by at least three of six bioinformatics tools (variant with CADD scaled score >15 is deemed to be deleterious).