The American Journal of Human Genetics, Volume *105*

Supplemental Data

A Rare Variant Nonparametric Linkage Method for Nuclear and Extended Pedigrees with Application to Late-Onset Alzheimer Disease via WGS Data

Linhai Zhao, Zongxiao He, Di Zhang, Gao T. Wang, Alan E. Renton, Badri N. Vardarajan, Michael Nothnagel, Alison M. Goate, Richard Mayeux, and Suzanne M. Leal

Supplemental Material

Supplemental Figures and Tables

Figure S1. Pedigree structures used for evaluation of intra-familial locus heterogeneity

Pedigrees with intra-familial locus heterogeneity were simulated by generating genotypes on extended families with two branches with three of the four children in the last generation being affected (panel A) and analyzing pedigrees with all children affected (panel B), to mimic intra intra-familial locus heterogeneity

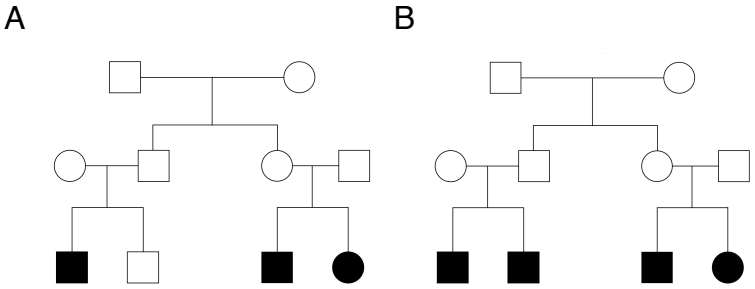
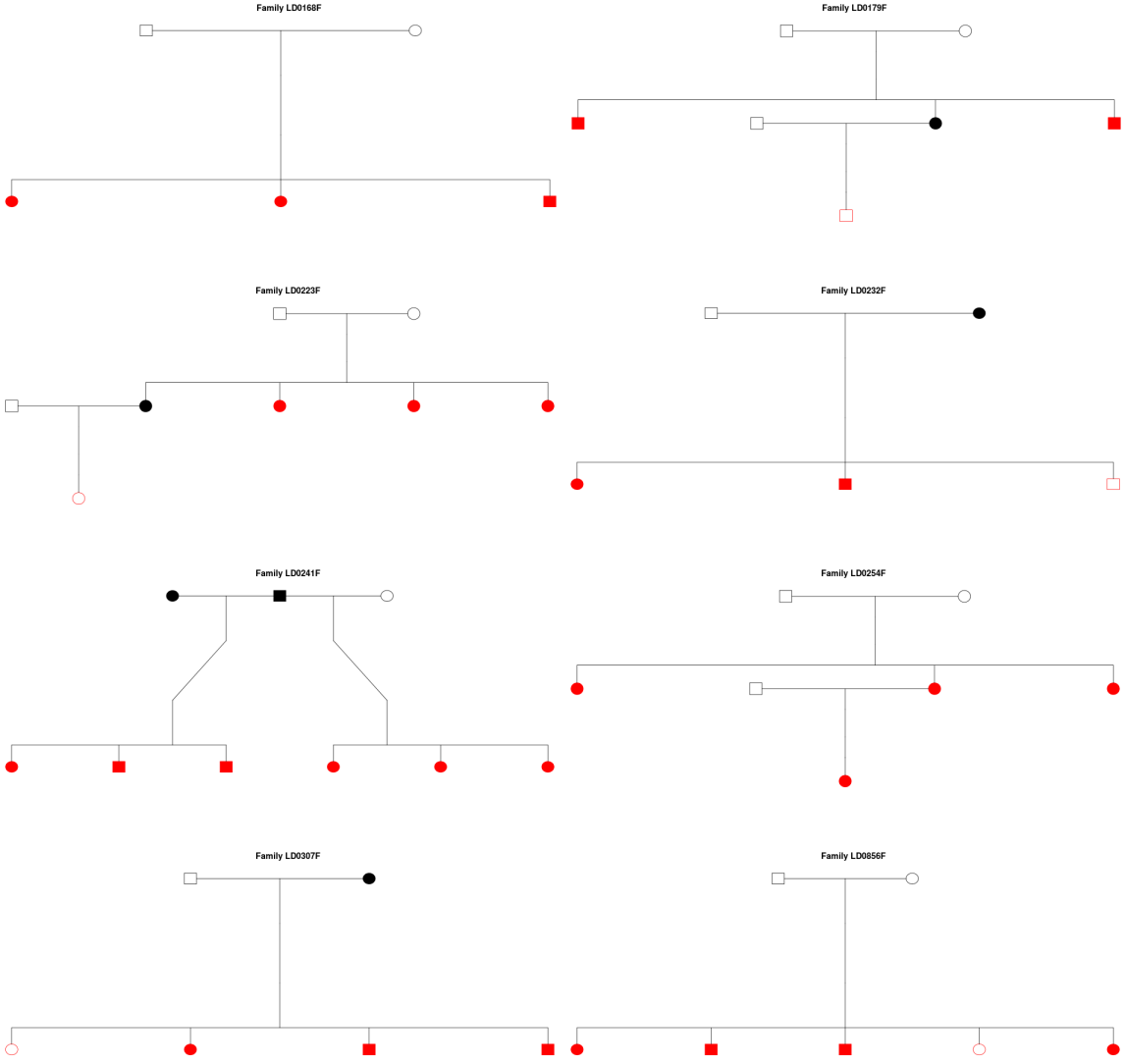


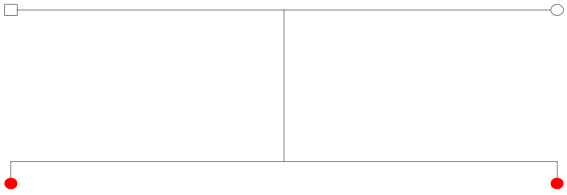
Figure S2. Pedigrees included in the analysis from the Alzheimer's Disease Sequencing Project

The 107 pedigrees (42 European pedigrees and 65 Hispanic pedigrees) which were analyzed. Squares represent males and circles females. Filled symbols are individuals affected Alzheimer's disease and open symbols represent unaffected family members. Those individuals shown in red have whole genome sequence data available, while those in black do not have available genotype data.

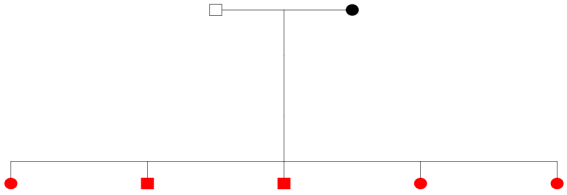
1. European pedigrees



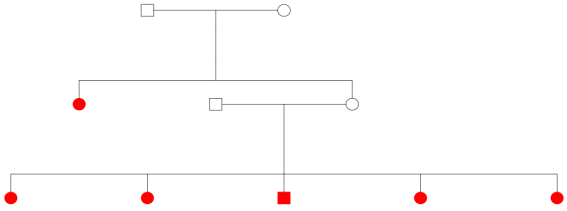
Family LD0949F



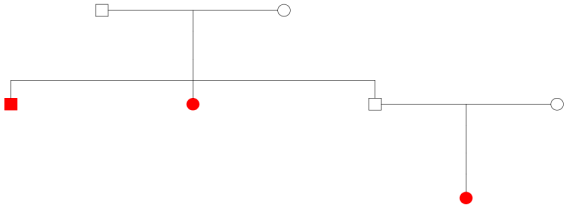
Family LD1012F



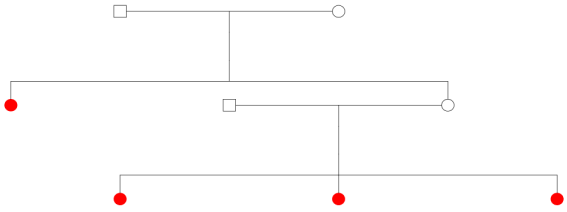
Family LD1223F



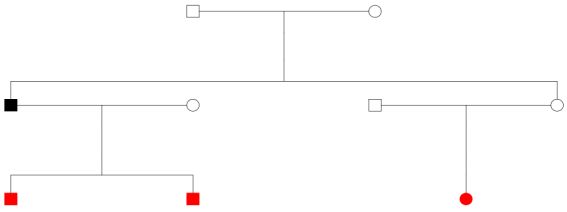
Family LD1260F



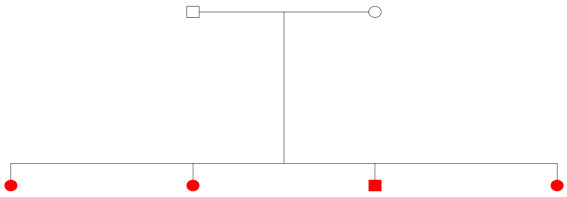
Family LD1265F



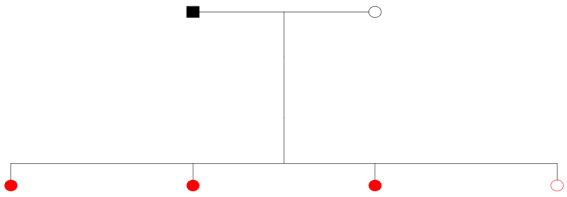
Family LD1315F



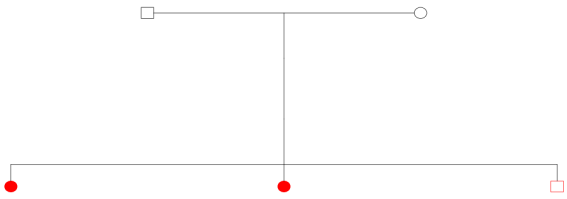
Family LD1329F



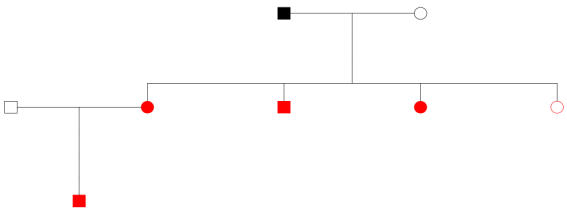
Family LD1579F



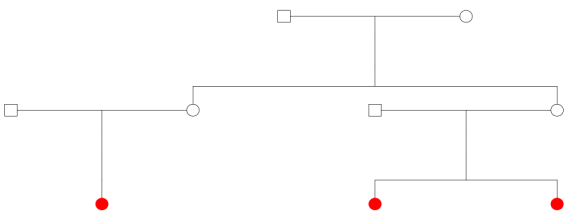
Family NC0049F



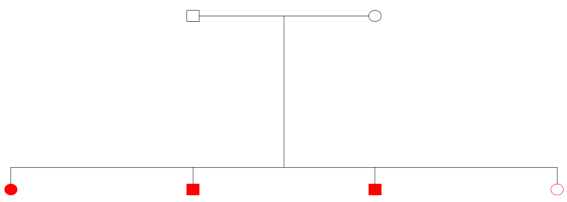
Family NC0131F

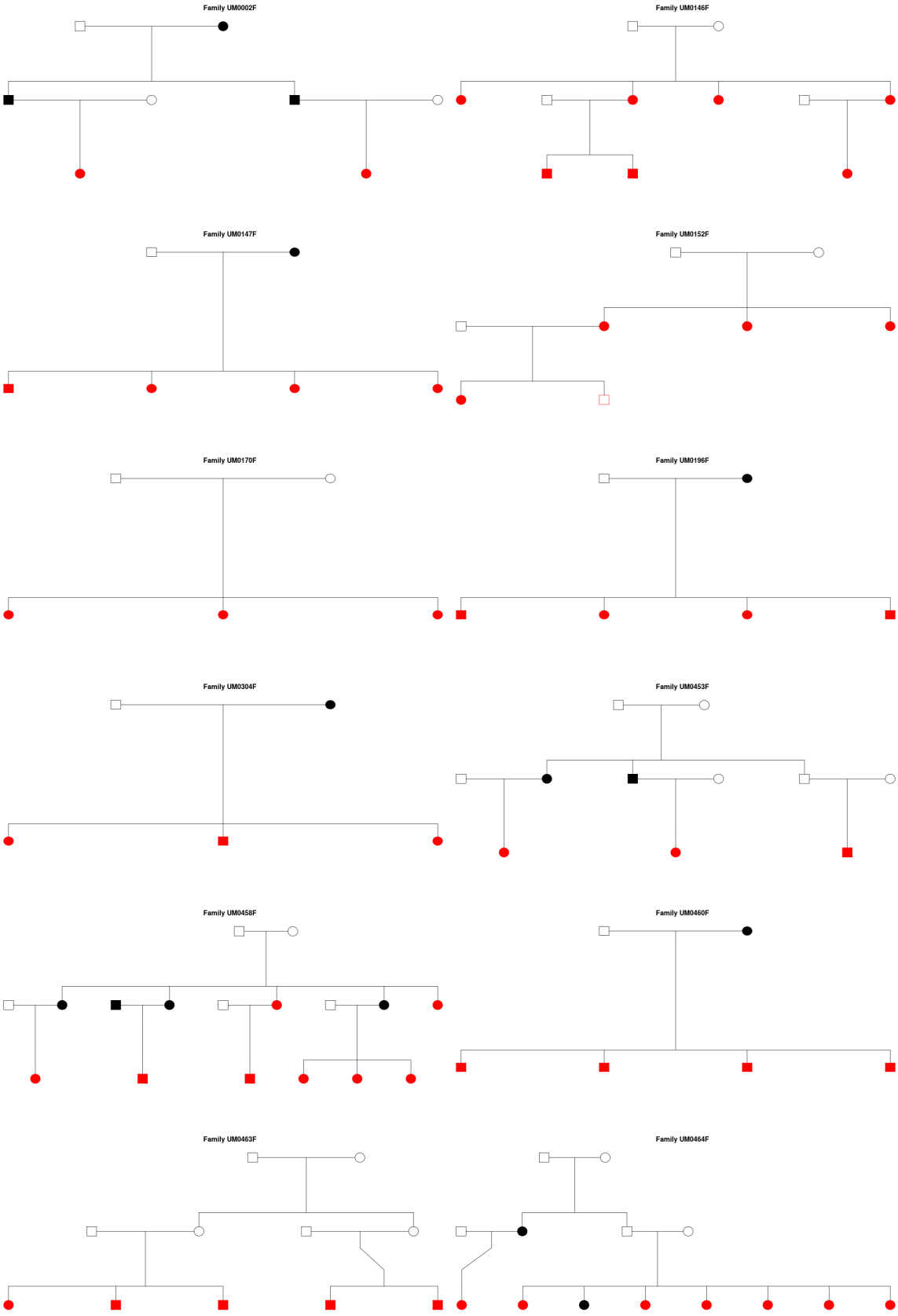


Family NC0205F

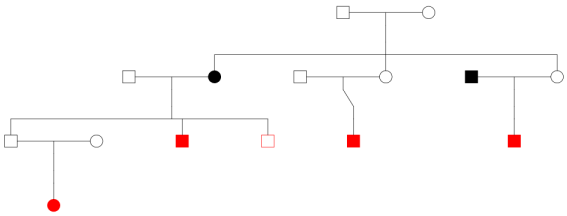


Family NC0302F

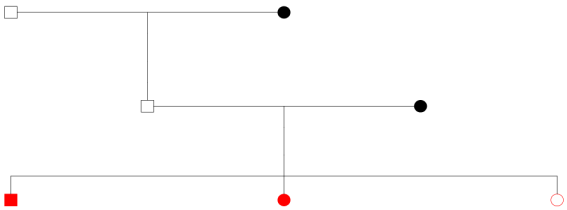



Family UM0002F
Family UM0146F
Family UM0147F
Family UM0152F
Family UM0170F
Family UM0196F
Family UM0304F
Family UM0453F
Family UM0458F
Family UM0460F
Family UM0463F
Family UM0464F

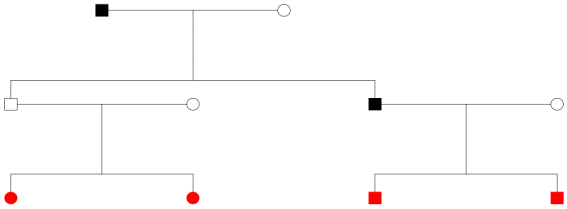
Family UP0001F



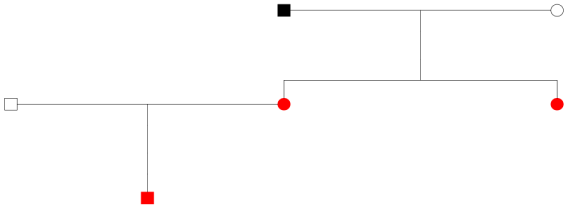
Family UP0002F



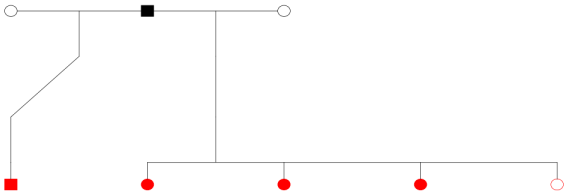
Family UP0002F-2



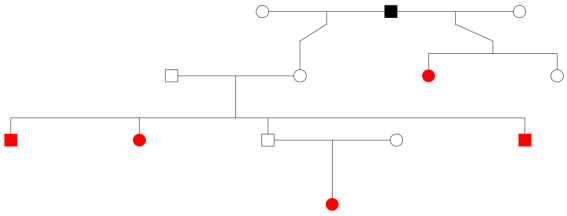
Family UP0003F



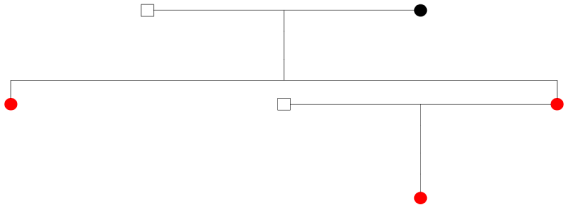
Family UP0004F



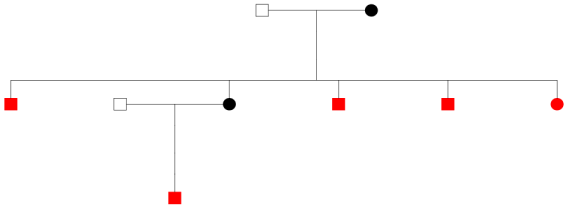
Family UP0005F



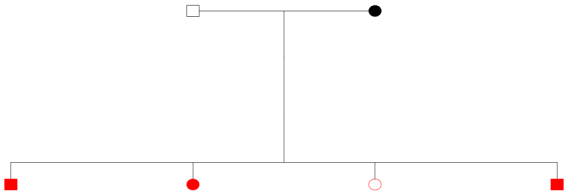
Family UP0006F



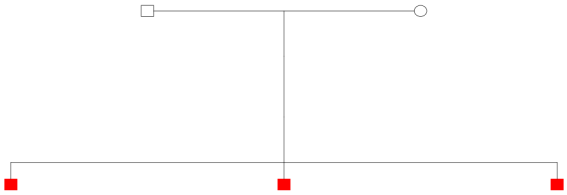
Family UP0007F

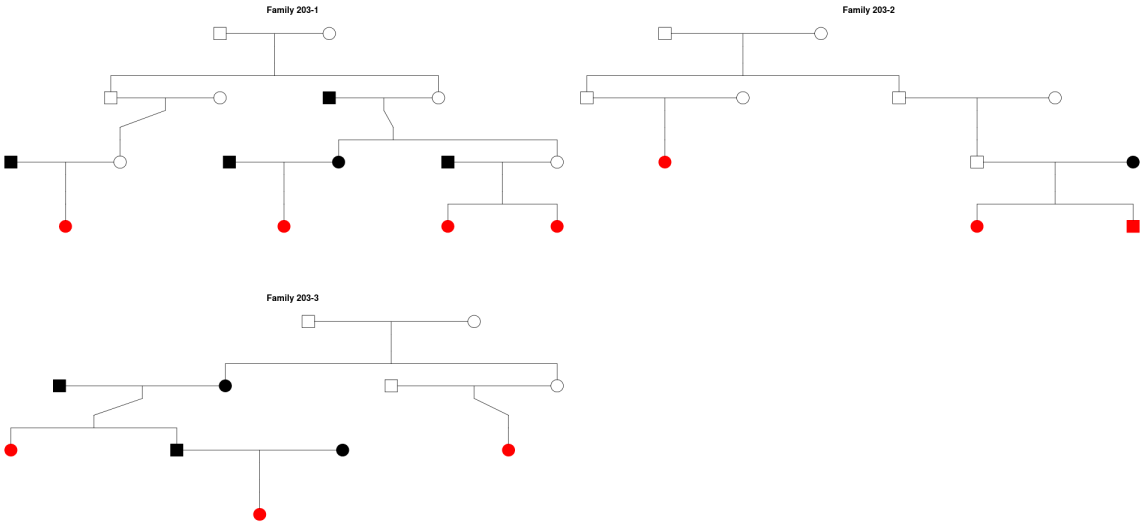


Family UP0008F

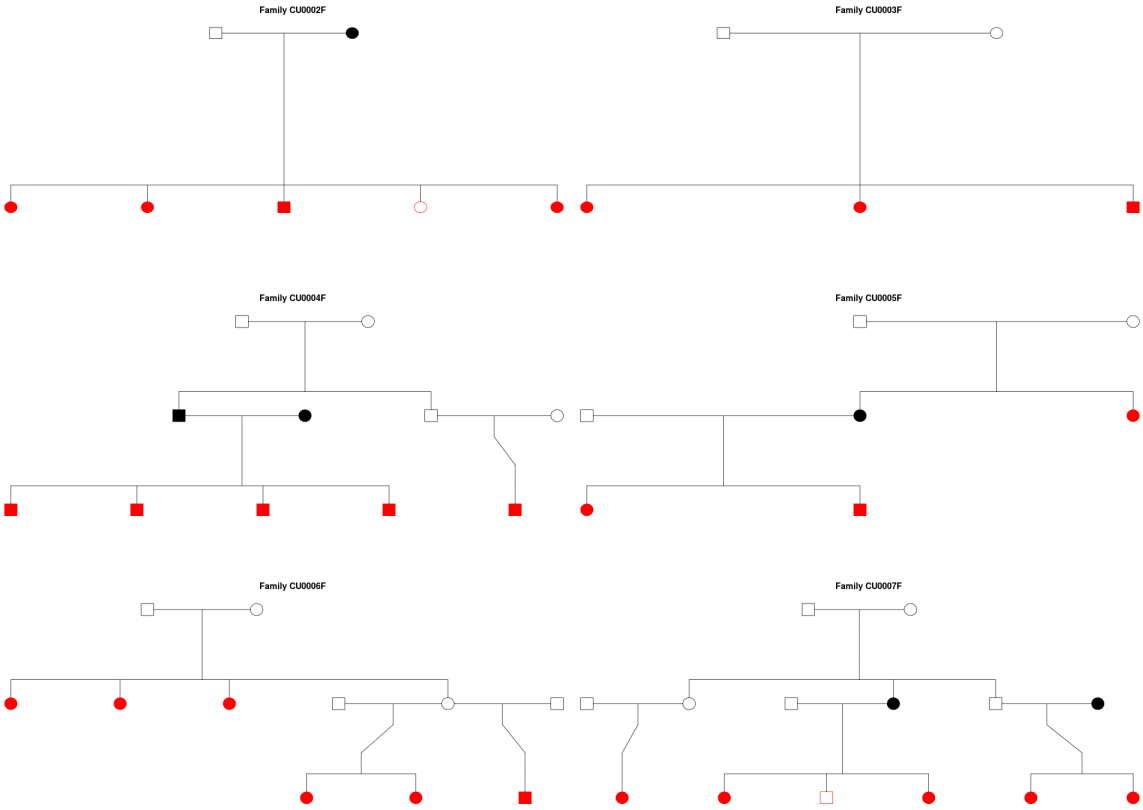


Family VU0072F

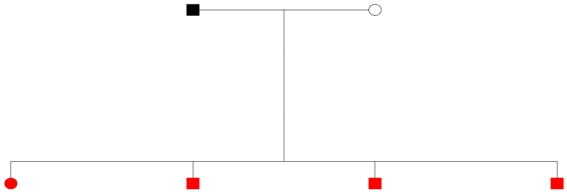




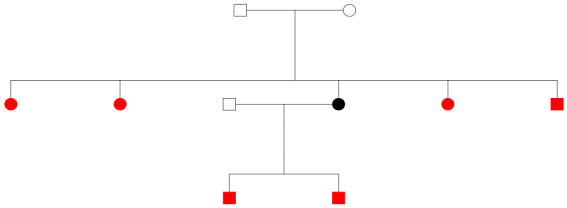
2. Hispanic Pedigrees



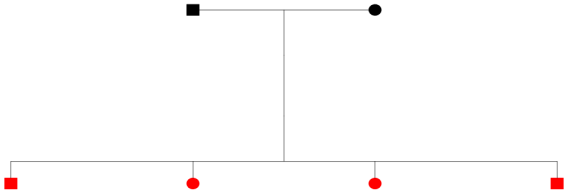
Family CU0008F



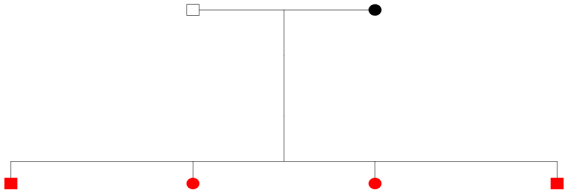
Family CU0009F



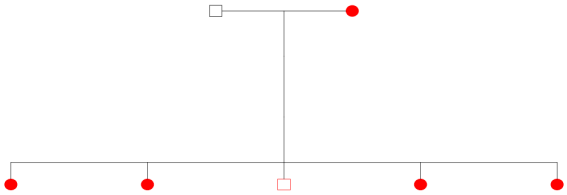
Family CU0010F



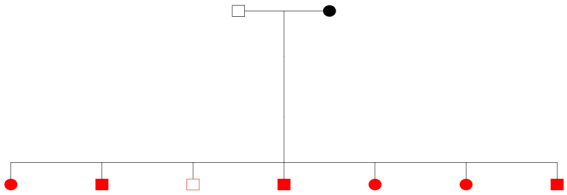
Family CU0012F



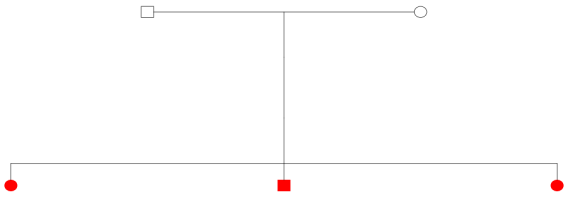
Family CU0013F



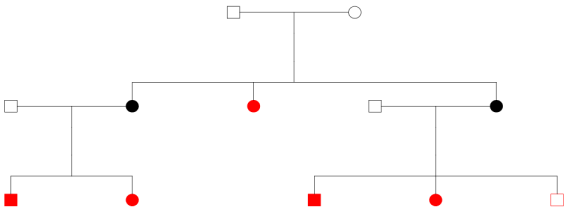
Family CU0014F



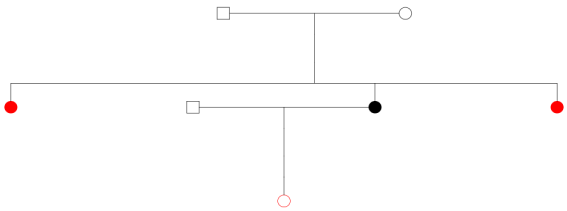
Family CU0015F



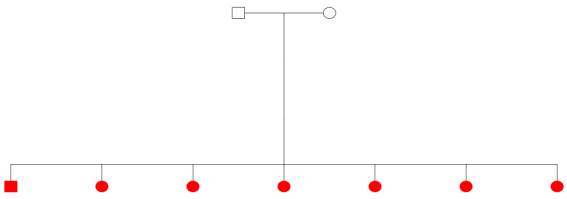
Family CU0016F



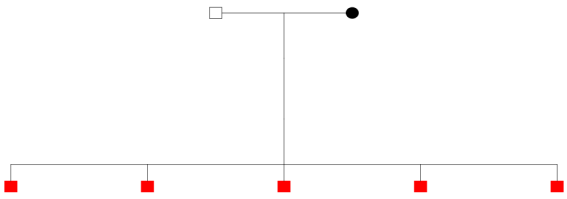
Family CU0017F



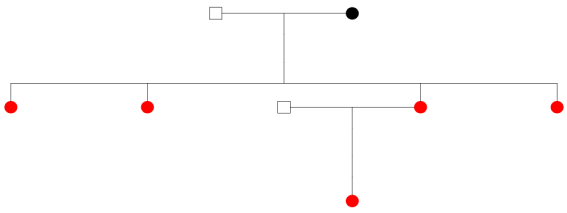
Family CU0018F



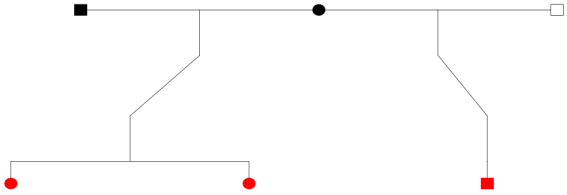
Family CU0019F



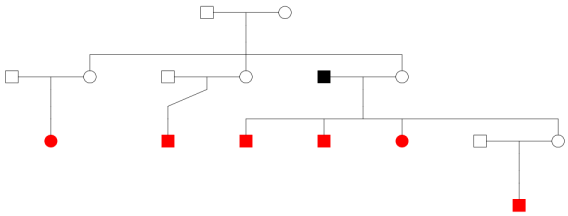
Family CU0020F



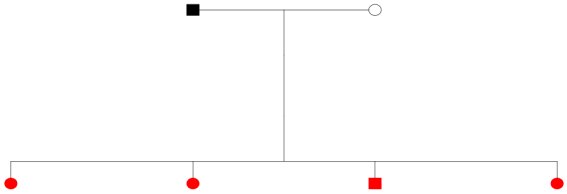
Family CU0022F



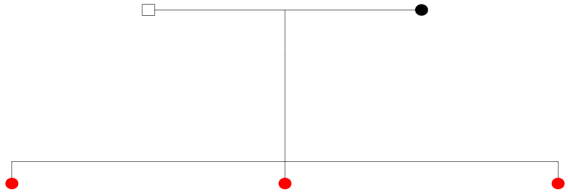
Family CU0023F



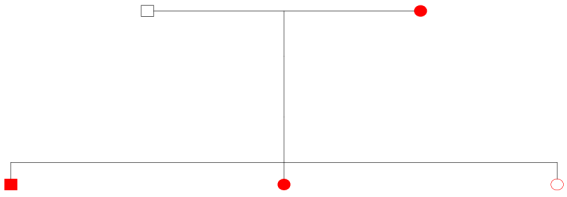
Family CU0024F



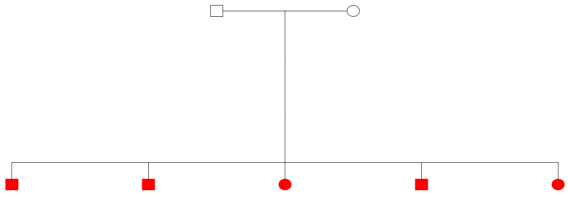
Family CU0025F



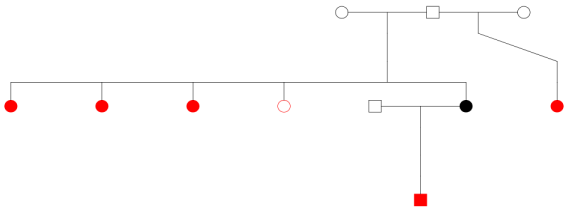
Family CU0026F



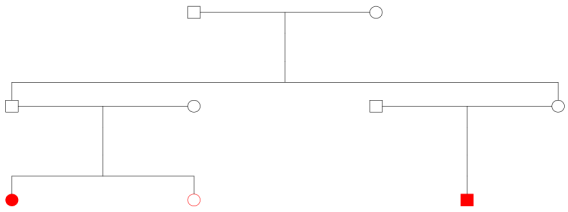
Family CU0029F



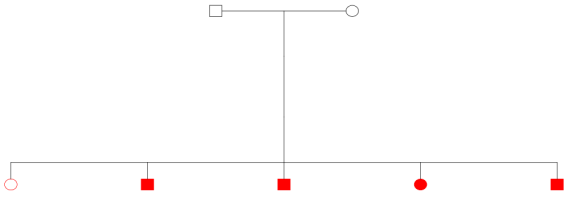
Family CU0030F



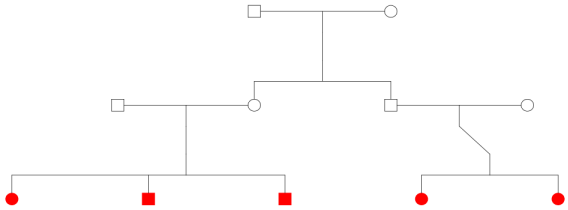
Family CU0033F



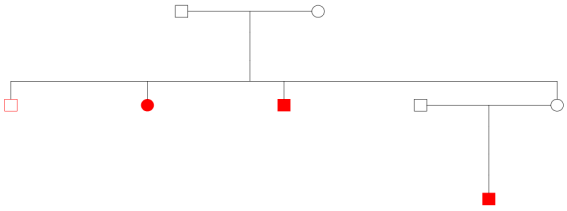
Family CU0035F



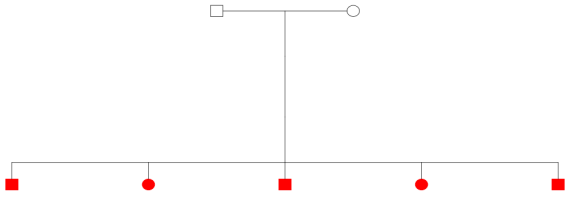
Family CU0036F



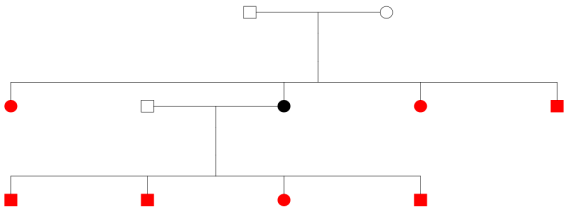
Family CU0037F



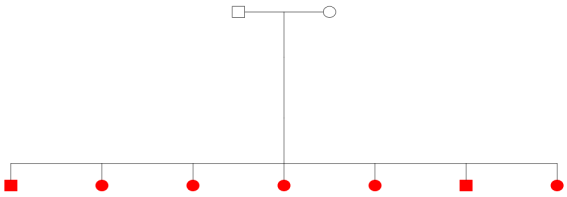
Family CU0038F



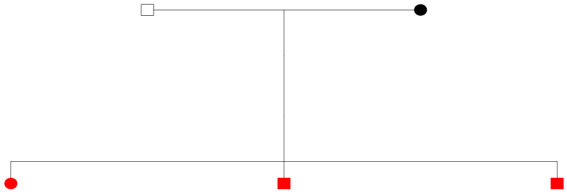
Family CU0039F



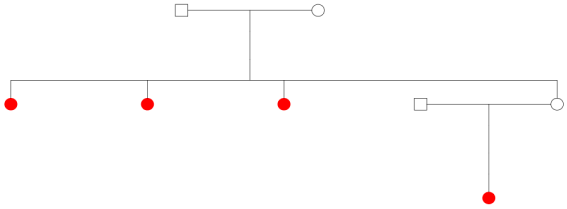
Family CU0040F



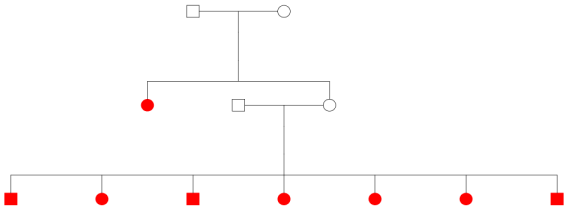
Family CU0041F



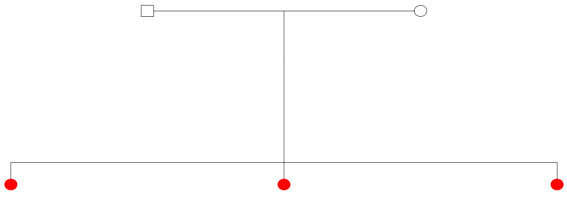
Family CU0043F



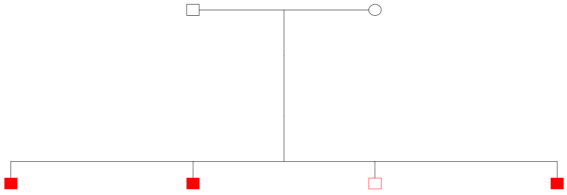
Family CU0044F



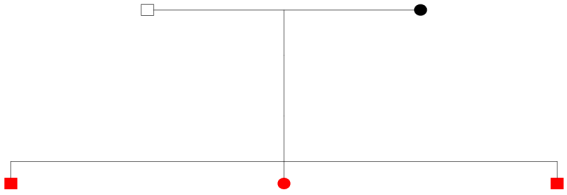
Family CU0044F-2



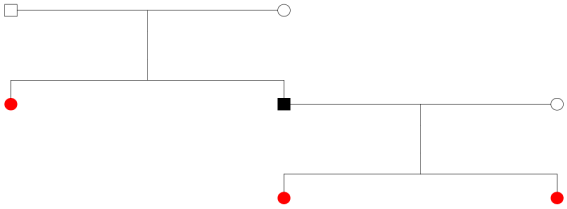
Family CU0045F



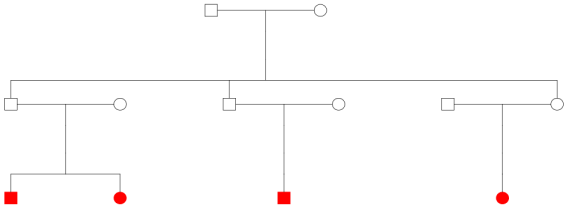
Family CU0046F



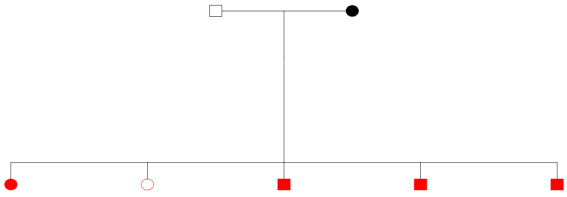
Family CU0047F



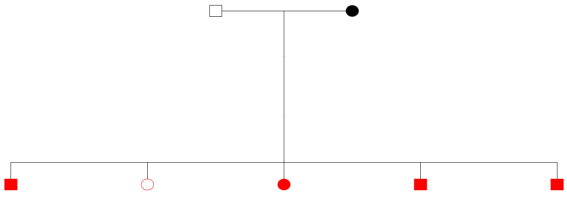
Family CU0048F



Family CU0049F



Family CU0050F



Family CU0052F

Family CU0053F

Family CU0055F

Family CU0057F

Family CU0058F

Family CU0059F

Family CU0060F

Family CU0064F

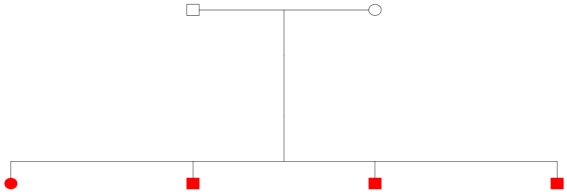
Family CU0065F

Family CU0067F

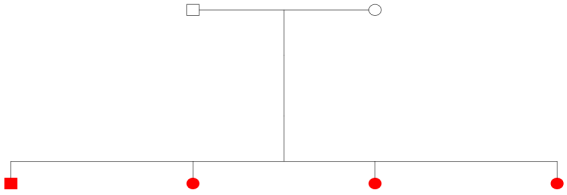
Family CU0068F

Family CU0070F

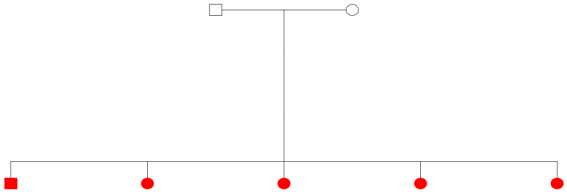
Family CU0071F



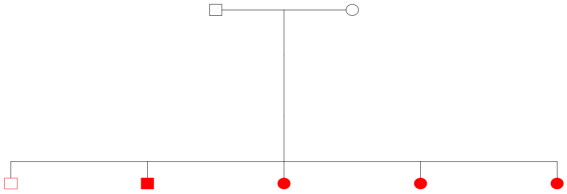
Family CU0073F



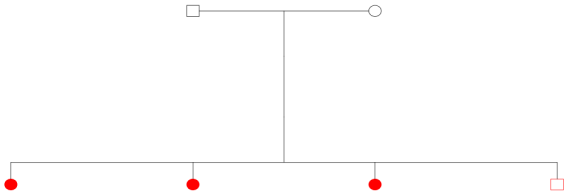
Family CU0075F



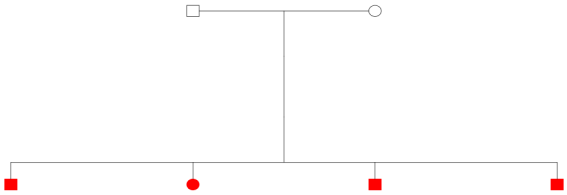
Family CU0076F



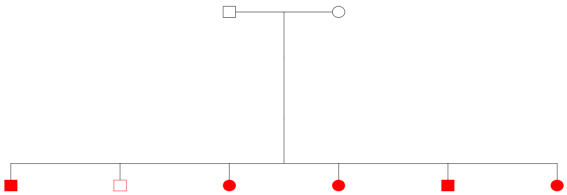
Family CU0078F



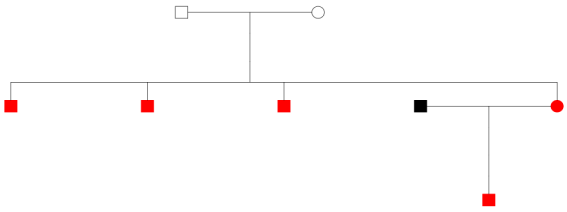
Family CU0079F



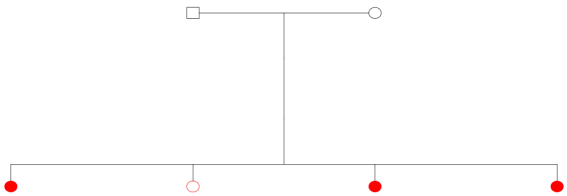
Family CU0081F



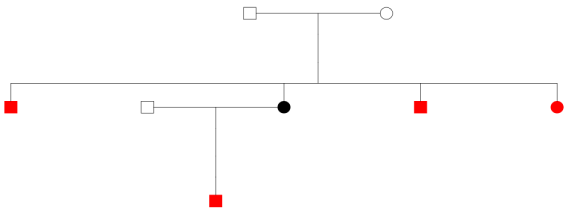
Family CU0082F



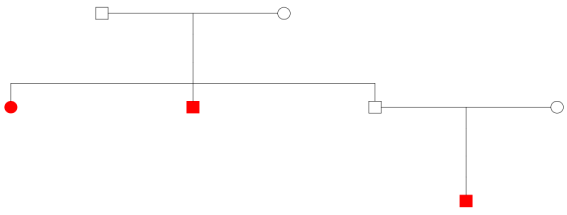
Family CU0083F



Family CU0032F



Family CU0042F



Family CU0051F

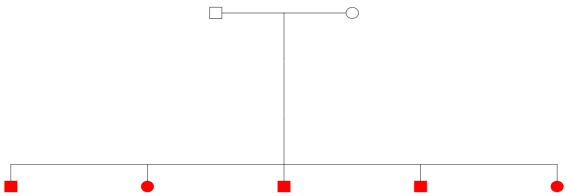


Figure S3. QQ plots for CHP-NPL$_{Pairs}$ under the null hypothesis of no linkage

One-thousand replicates of exomes were generated under the null hypothesis of no linkage for 2,000 affected sib-pairs without missing genotype date (panel A); 2,000 affected sib-pairs with founders missing all exome data (panel B); 300 nuclear families with three affected siblings without missing genotype data (panel C); 300 nuclear families with three affected siblings with founders missing all exome data (panel D); 100 extended families without missing genotype data (panel E); and 100 extended families with founders missing all exome data (panel F); and analyzed using CHP-NPL$_{Pairs}$ obtaining analytical p-values.

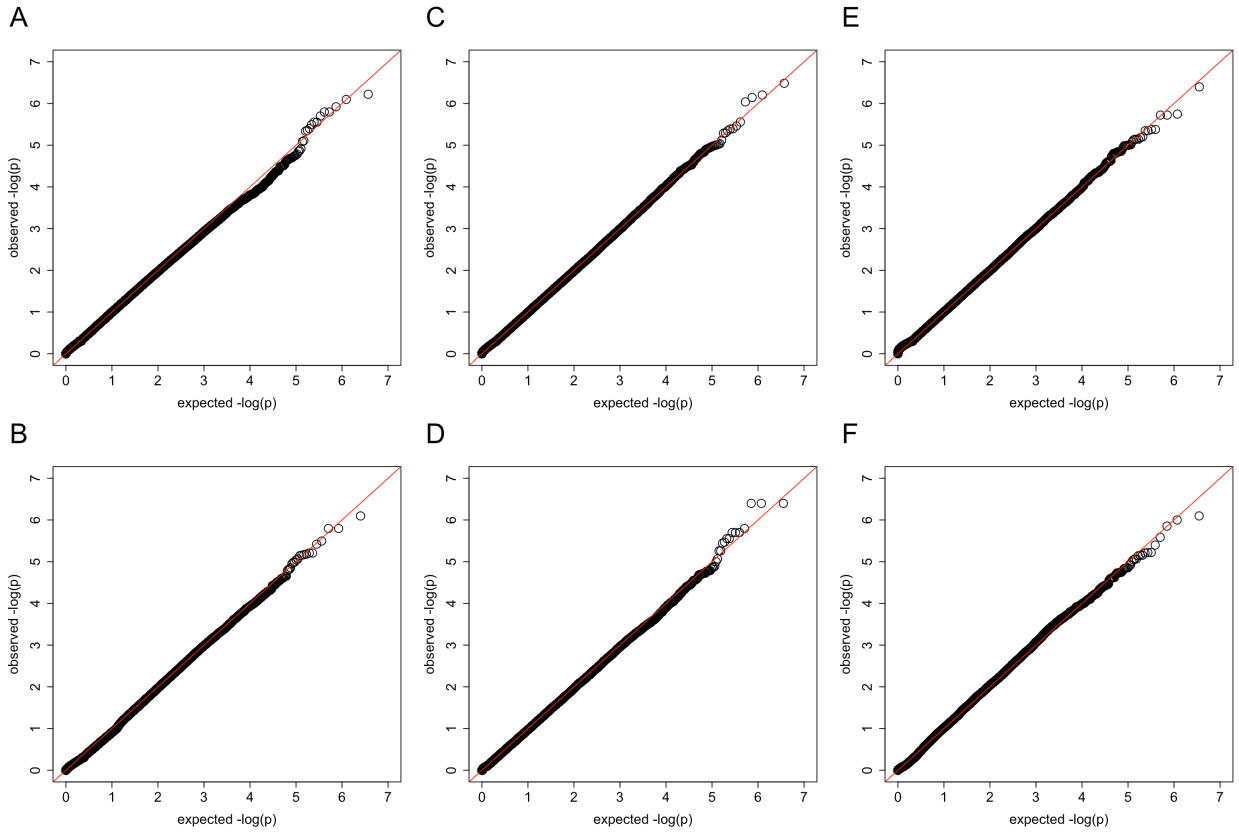


Figure S4. QQ plots for CHP-NPL$_{All}$ under the null hypothesis of no linkage

One-thousand replicates of exomes were generated under the null hypothesis of no linkage for 2,000 affected sib-pairs without missing genotype date (panel A); 2,000 affected sib-pairs with founders missing all exome data (panel B); 300 nuclear families with three affected siblings without missing genotype data (panel C); 300 nuclear families with three affected siblings with founders missing all exome data (panel D); 100 extended families without missing genotype data (panel E); and 100 extended families with founders missing all exome data (panel F); and analyzed using CHP-NPL$_{All}$ obtaining analytical p-values.

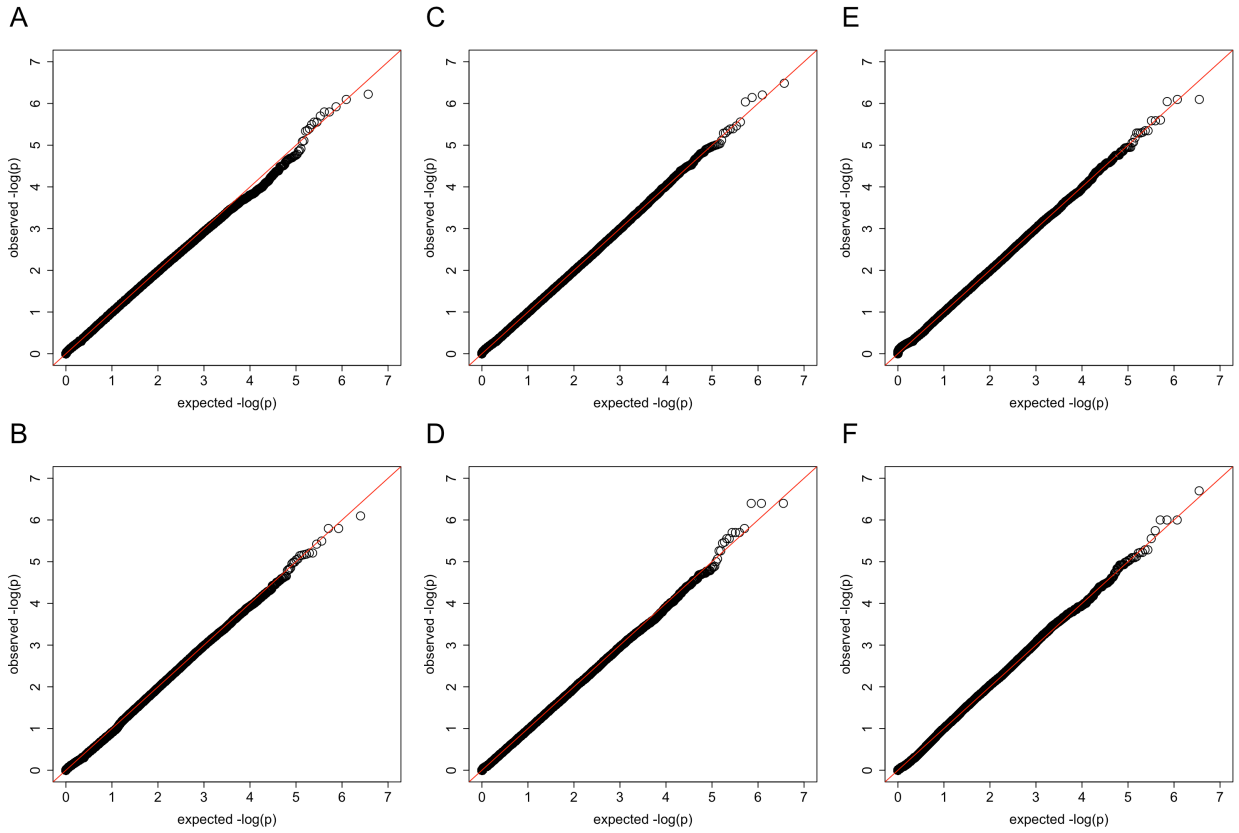


Figure S5. QQ plots for RV-NPL$_{Pairs}$ under the null hypothesis of no linkage

One-thousand replicates of exomes were generated under the null hypothesis of no linkage for 2,000 affected sib-pairs without missing genotype date (panel A); 2,000 affected sib-pairs with founders missing all exome data (panel B); 300 nuclear families with three affected siblings without missing genotype data (panel C); 300 nuclear families with three affected siblings with founders missing all exome data (panel D); 100 extended families without missing genotype data (panel E); and 100 extended families with founders missing all exome data (panel F); and analyzed using RV-NPL$_{Pairs}$ obtaining empirical p-values using 1,000,000 permutations. The observed plateau is due to the number of permutations performed.

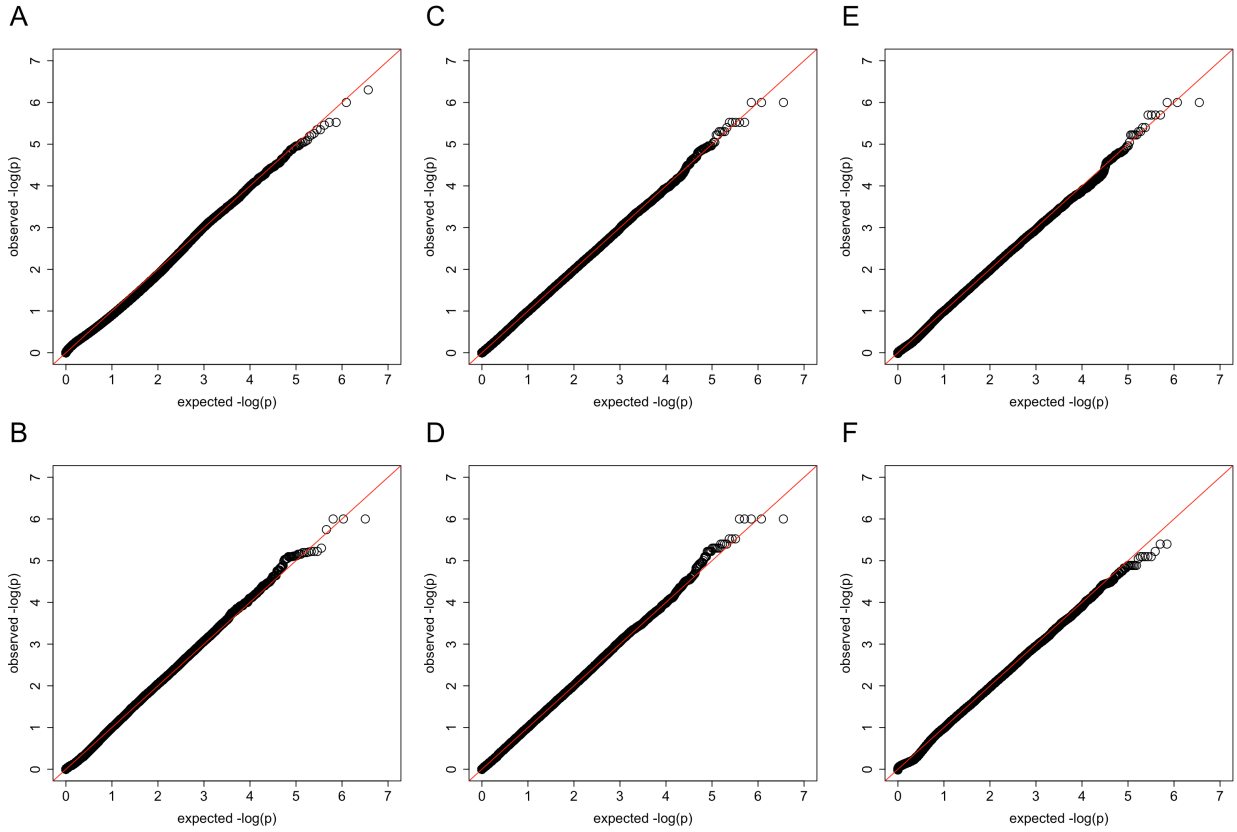


Figure S6. QQ plots for RV-NPL$_{All}$ under the null hypothesis of no linkage

One-thousand replicates of exomes were generated under the null hypothesis of no linkage for 2,000 affected sib-pairs without missing genotype date (panel A); 2,000 affected sib-pairs with founders missing all exome data (panel B); 300 nuclear families with three affected siblings without missing genotype data (panel C); 300 nuclear families with three affected siblings with founders missing all exome data (panel D); 100 extended families without missing genotype data (panel E); and 100 extended families with founders missing all exome data (panel F); and analyzed using RV-NPL$_{All}$ obtaining empirical p-values using 1,000,000 permutations. The observed plateau is due to the number of permutations performed.

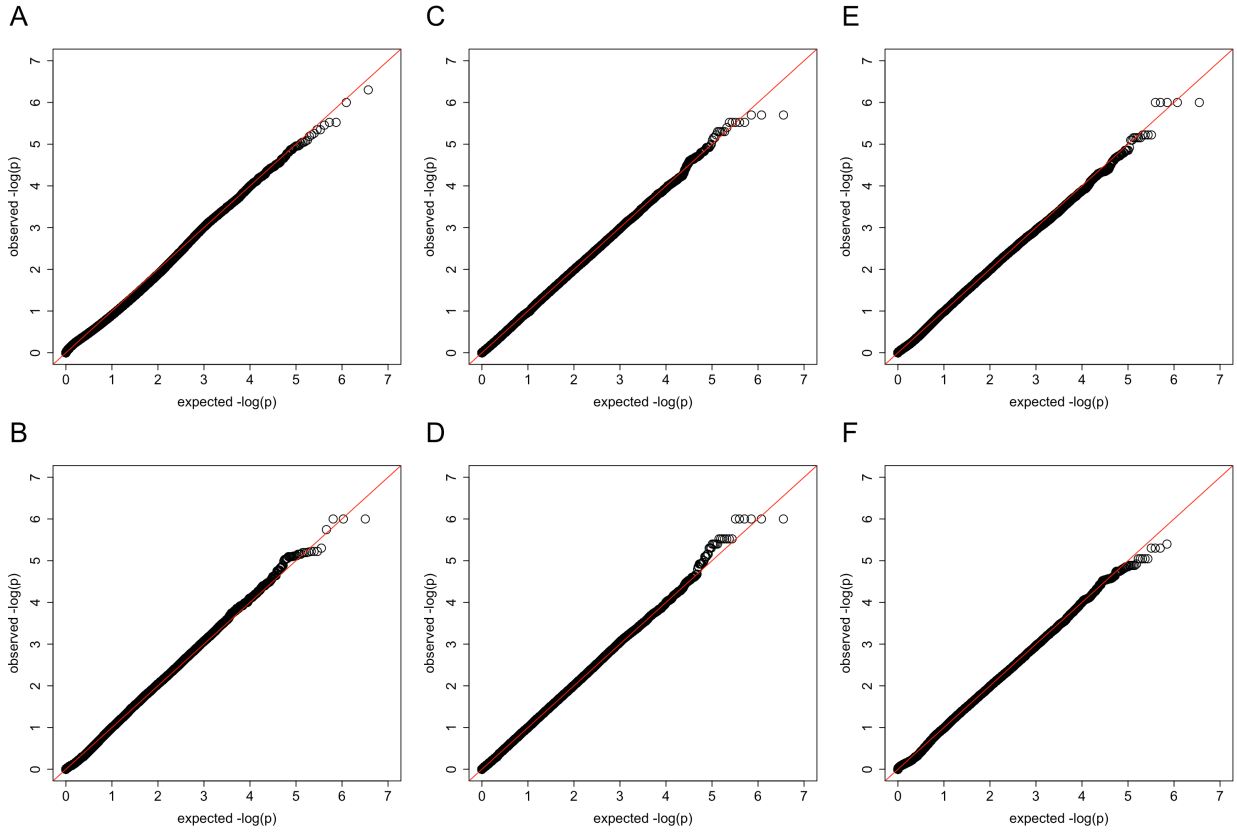


Figure S7. Power comparison for NPL_{Pairs} for nuclear families with three affected siblings

Genotypes were generated for 300 nuclear families with three affected siblings conditional on affection status assuming a multiplicative model in which each causal variant within a gene region has an OR of 5.0. Analysis was performed using RV-NPL_{Pairs}, CHP-NPL_{Pairs}, and Multipoint-NPL_{Pairs}: with 100%, 75% and 50% of the variant being causal and the remaining non-causal (OR=1.0) (panel A); with only causal nonsynonymous (NS) variants as well as with causal nonsynonymous (NS) and non-causal synonymous (S) variants (panel B); with 0%, 10%, 30%, and 50% of the founders missing all genotype data (panel C); and with no heterogeneity (NH), i.e. 300 linked families as well as with locus heterogeneity (H), i.e., 300 linked and 150 unlinked families (panel D).

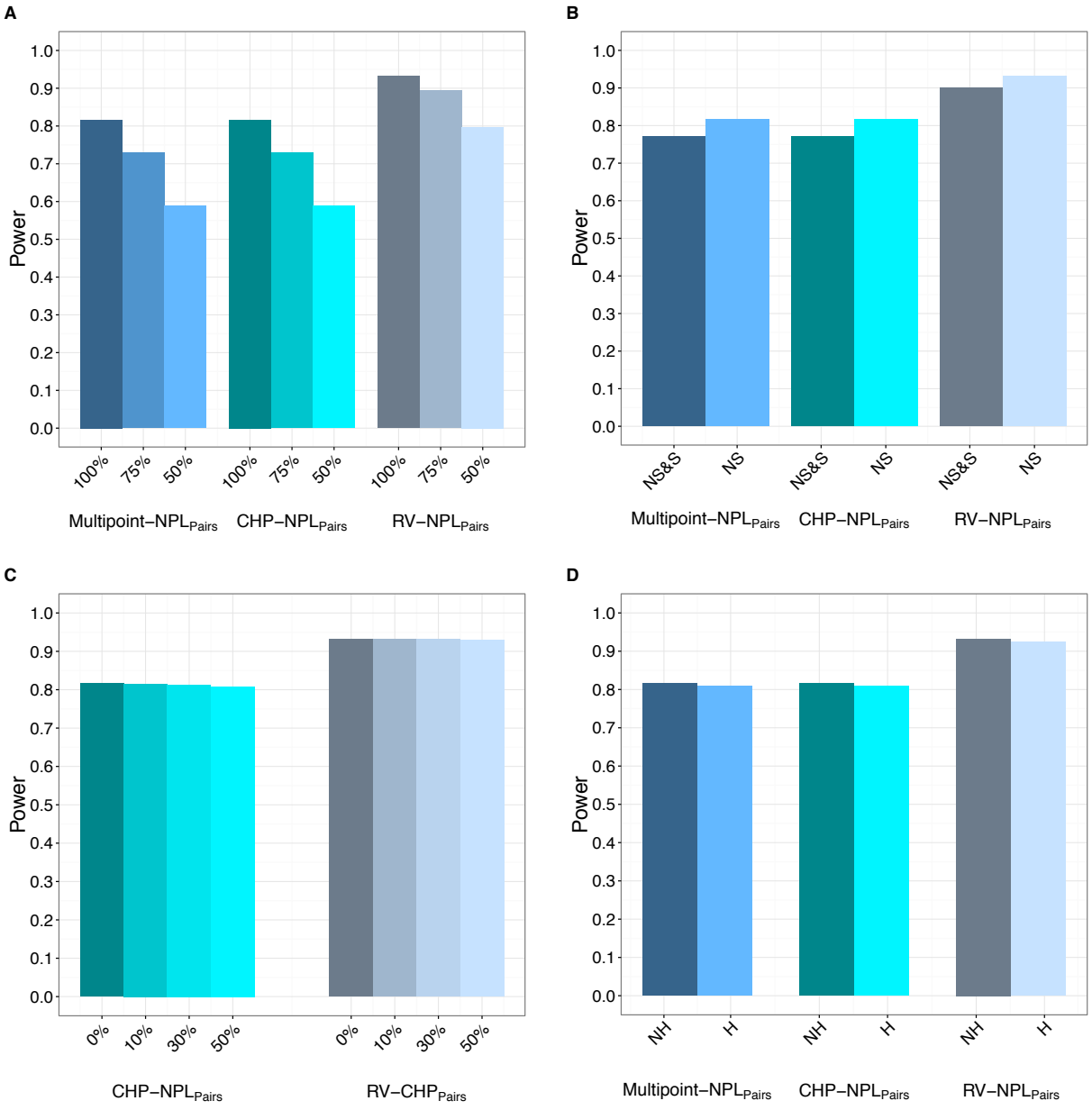

A
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100%
75%
50%
100%
75%
50%
100%
75%
50%
Multipoint–NPL$_{Pairs}$
CHP–NPL$_{Pairs}$
RV–NPL$_{Pairs}$
B
Power
NS&S
NS
NS&S
NS
NS&S
NS
Multipoint–NPL$_{Pairs}$
CHP–NPL$_{Pairs}$
RV–NPL$_{Pairs}$
C
Power
0%
10%
30%
50%
0%
10%
30%
50%
CHP–NPL$_{Pairs}$
RV–CHP$_{Pairs}$
D
Power
NH
H
NH
H
NH
H
Multipoint–NPL$_{Pairs}$
CHP–NPL$_{Pairs}$
RV–NPL$_{Pairs}$

Figure S8. Power comparison for NPL_{Pairs} on affected sibpairs

Genotypes were generated for 2,000 nuclear families with affected sibpairs conditional on affection status assuming a multiplicative model in which each causal variant within a gene region has an OR of 5.0. Analysis was performed using RV-NPL_{Pairs}, CHP-NPL_{Pairs}, and Multipoint-NPL_{Pairs}: with 100%, 75% and 50% of the variant being causal and the remaining non-causal (OR=1.0) (panel A); with only causal nonsynonymous (NS) variants as well as with causal nonsynonymous (NS) and non-causal synonymous (S) variants (panel B); with 0%, 10%, 30%, and 50% of the founders missing all genotype data (panel C); and with no heterogeneity (NH), i.e. 2,000 linked families as well as with locus heterogeneity (H), i.e., 2,000 linked and 1,000 unlinked families (panel D).

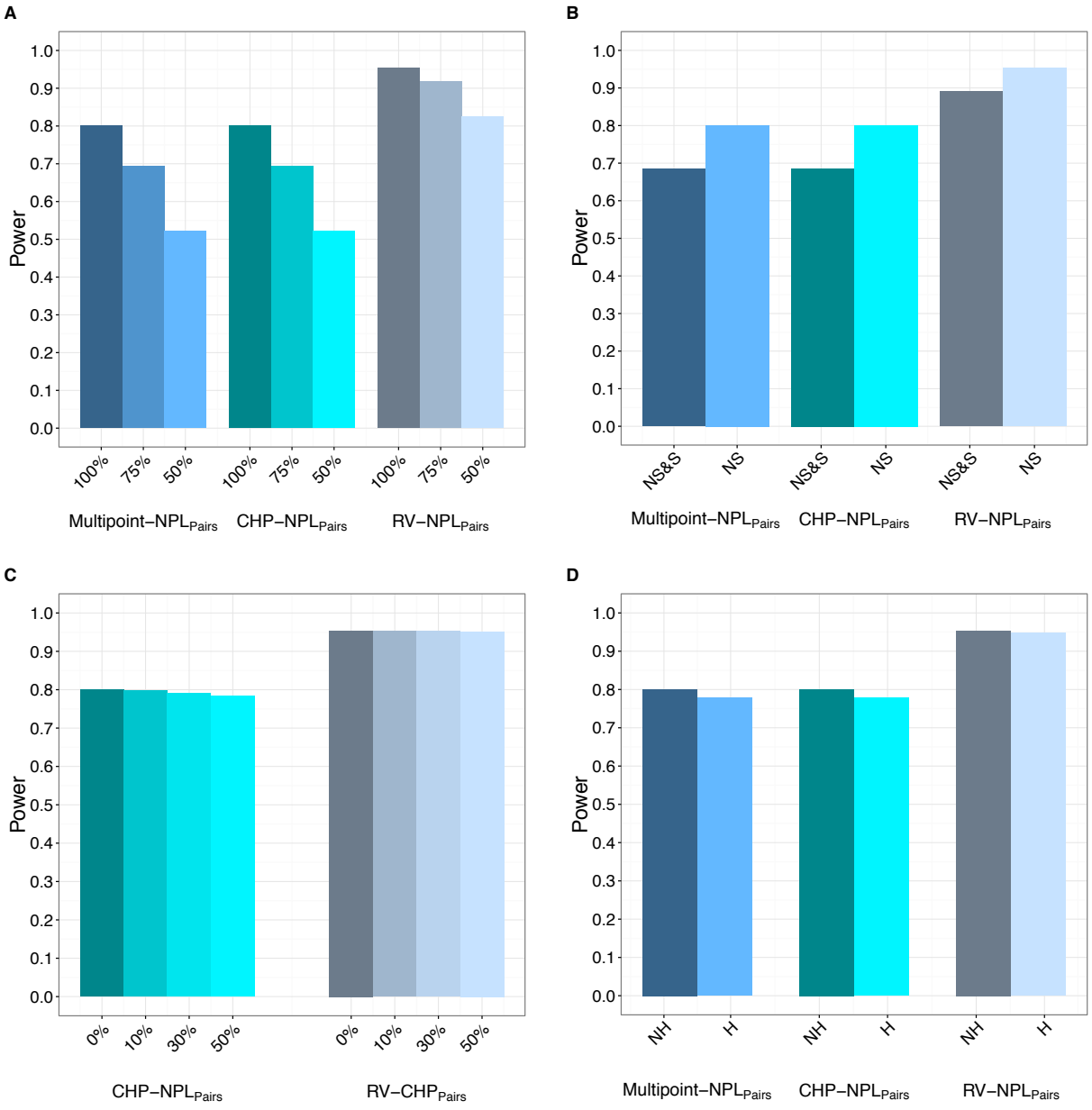

A
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100%
75%
50%
100%
75%
50%
100%
75%
50%
Multipoint–NPL$_{Pairs}$
CHP–NPL$_{Pairs}$
RV–NPL$_{Pairs}$
B
Power
NS&S
NS
NS&S
NS
NS&S
NS
Multipoint–NPL$_{Pairs}$
CHP–NPL$_{Pairs}$
RV–NPL$_{Pairs}$
C
Power
0%
10%
30%
50%
0%
10%
30%
50%
CHP–NPL$_{Pairs}$
RV–CHP$_{Pairs}$
D
Power
NH
H
NH
H
NH
H
Multipoint–NPL$_{Pairs}$
CHP–NPL$_{Pairs}$
RV–NPL$_{Pairs}$

Figure S9. Power comparison for NPL_{All} on extended families

Genotypes were generated for 100 extended families conditional on affection status assuming a multiplicative model in which each causal variant within a gene region has an OR of 5.0. Analysis was performed using RV-NPL_{All}, CHP-NPL_{All}, and Multipoint-NPL_{All}: with 100%, 75% and 50% of the variant being causal and the remaining non-causal (OR=1.0) (panel A); with only causal nonsynonymous (NS) variants as well as with causal nonsynonymous (NS) and non-causal synonymous (S) variants (panel B); with 0%, 10%, 30%, and 50% of the founders missing all genotype data (panel C); and with no heterogeneity (NH), i.e. 100 linked families as well as with locus heterogeneity (H), i.e., 100 linked and 50 unlinked families (panel D).

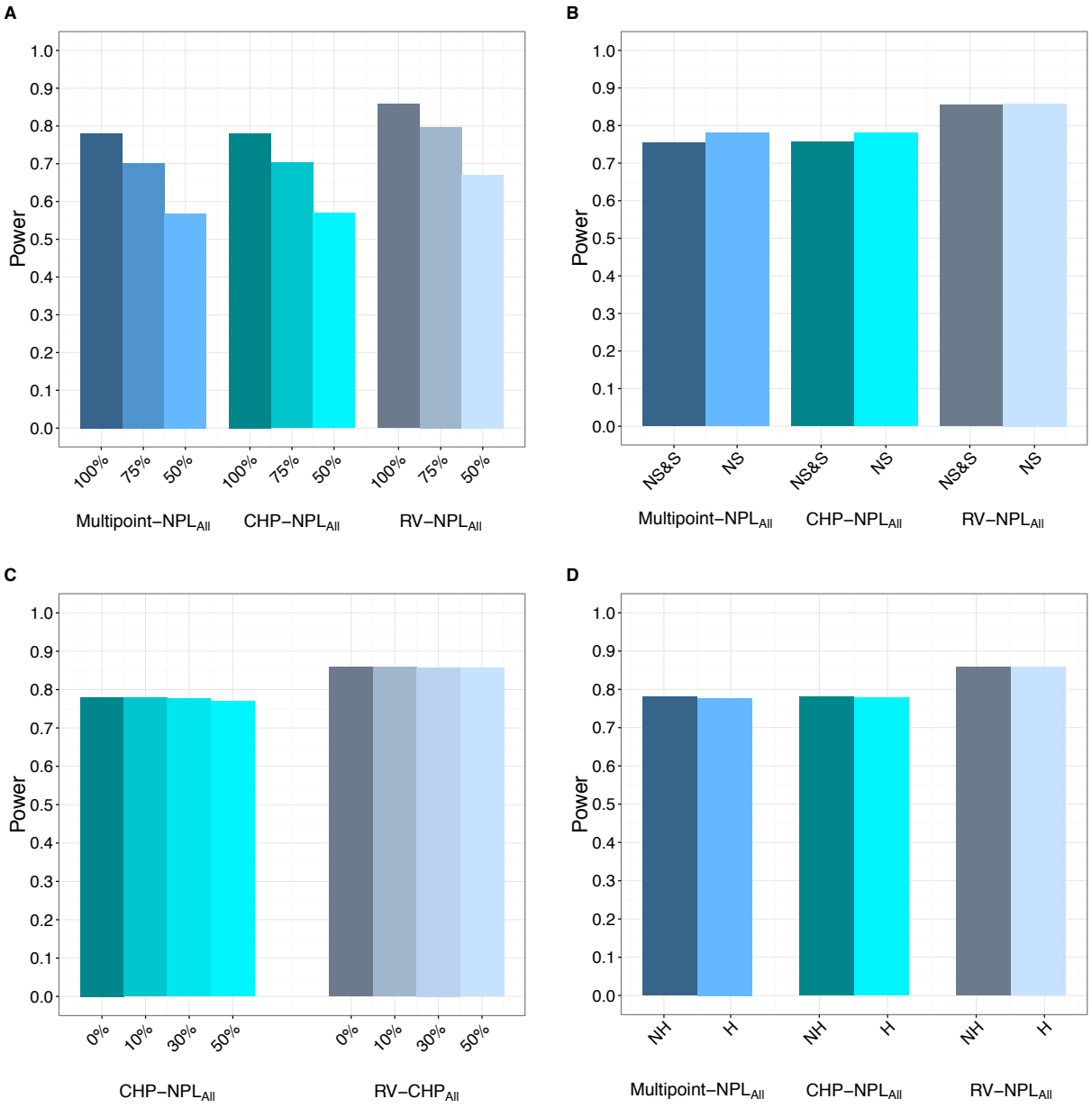

A
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100%
75%
50%
100%
75%
50%
100%
75%
50%
Multipoint–NPL$_{All}$
CHP–NPL$_{All}$
RV–NPL$_{All}$
B
Power
NS&S
NS
NS&S
NS
NS&S
NS
Multipoint–NPL$_{All}$
CHP–NPL$_{All}$
RV–NPL$_{All}$
C
Power
0%
10%
30%
50%
0%
10%
30%
50%
CHP–NPL$_{All}$
RV–CHP$_{All}$
D
Power
NH
H
NH
H
NH
H
Multipoint–NPL$_{All}$
CHP–NPL$_{All}$
RV–NPL$_{All}$

Figure S10. Power comparison for NPL_{All} on nuclear families with three affected siblings

Genotypes were generated for 300 nuclear families with three affected siblings conditional on affection status assuming a multiplicative model in which each causal variant within a gene region has an OR of 5.0. Analysis was performed using RV-NPL_{All}, CHP-NPL_{All}, and Multipoint-NPL_{All}: with 100%, 75% and 50% of the variant being causal and the remaining non-causal (OR=1.0) (panel A); with only causal nonsynonymous (NS) variants as well as with causal nonsynonymous (NS) and non-causal synonymous (S) variants (panel B); with 0%, 10%, 30%, and 50% of the founders missing all genotype data (panel C); and with no heterogeneity (NH), i.e. 100 linked families as well as with locus heterogeneity (H), i.e., 300 linked and 150 unlinked families (panel D).

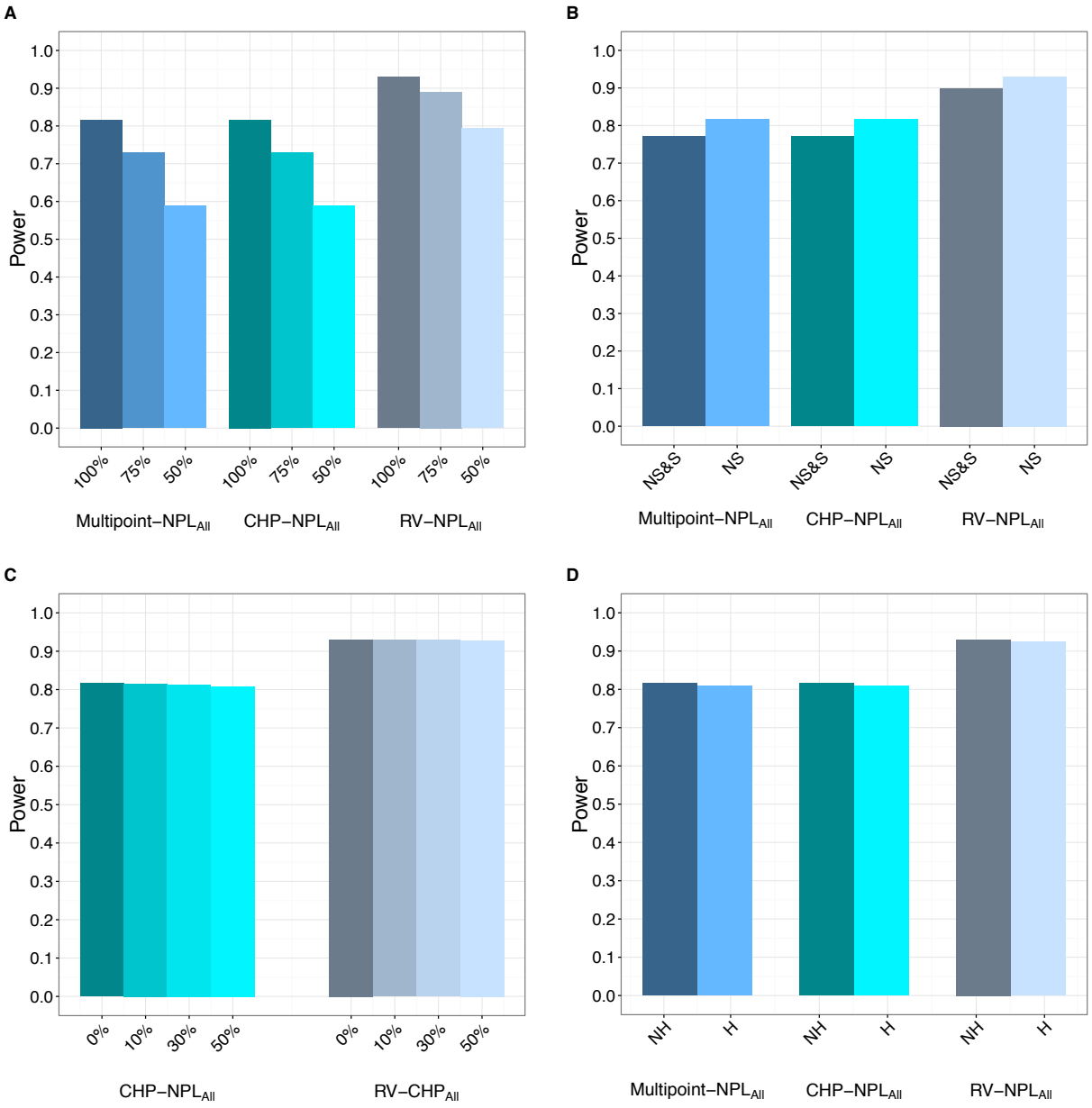

A
Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100%
75%
50%
100%
75%
50%
100%
75%
50%
Multipoint–NPL$_{All}$
CHP–NPL$_{All}$
RV–NPL$_{All}$
B
Power
NS&S
NS
NS&S
NS
NS&S
NS
Multipoint–NPL$_{All}$
CHP–NPL$_{All}$
RV–NPL$_{All}$
C
Power
0%
10%
30%
50%
0%
10%
30%
50%
CHP–NPL$_{All}$
RV–CHP$_{All}$
D
Power
NH
H
NH
H
NH
H
Multipoint–NPL$_{All}$
CHP–NPL$_{All}$
RV–NPL$_{All}$

Table S1: The Ethnicities of Alzheimer's disease families included in the analysis

Ethnicity	Number of Families	Family IDs
Dominican	62	CU0002F, CU0003F[a], CU0004F[^], CU0005F[a], CU0006F[*a], CU0007F, CU0008F, CU0009F[c], CU0010F, CU0012F, CU0013F[a], CU0014F, CU0015F, CU0016F[a], CU0017F[^], CU0018F, CU0019F[^a], CU0020F, CU0022F, CU0023F[*], CU0024F, CU0025F, CU0026F, CU0029F, CU0030F[*a], CU0033F, CU0035F[^], CU0036F, CU0037F, CU0038F[*], CU0039F[a], CU0040F[a], CU0041F[a], CU0043F[c], CU0044F[a], CU0045F[c], CU0046F, CU0047F, CU0048F[^], CU0049F[*], CU0050F, CU0052F, CU0053F, CU0055F[^], CU0057F, CU0058F, CU0059F, CU0060F[a], CU0064F[a], CU0065F, CU0067F[*a], CU0068F[^], CU0070F[^a], CU0071F[a], CU0073F[^a], CU0075F[^], CU0076F[*], CU0078F[a], CU0079F[^], CU0081F[a], CU0082F, CU0083F[a]
European Descent	41	LD0168F, LD0179F, LD0223F, LD0232F, LD0241F[d], LD0254F, LD0307F, LD0856F, LD0949F, LD1012F[d], LD1223F, LD1260F, LD1265F, LD1315F[b], LD1329F[*], LD1579F[d], NC0049F, NC0131F, NC0205F[^], NC0302F, UM0002F, UM0146F[^b], UM0147F[d], UM0152F, UM0170F, UM0196F[d], UM0304F, UM0453F, UM0458F, UM0460F, UM0463F[^b], UM0464F, UP0001F, UP0002F, UP0003F, UP0004F[d], UP0005F, UP0006F, UP0007F, UP0008F, VU0072F
Puerto Rican	3	CU0032F, CU0042F, CU0051F
Dutch Isolate	1	203[d]

[^]Pedigrees with excess RV sharing for gene *PSMF1*.
*Pedigrees with excess RV sharing for gene *PTPN21*.
[a]Pedigrees with excess RV sharing for gene *ABCA7*; [b]Pedigrees with excess RV sharing for gene *ACE*;
[c]Pedigrees with excess RV sharing for gene *EPHA1*; [d]Pedigrees with excess RV sharing for gene *SORL1*.

Table S2. Type I error rate of CHP-NPL and RV-NPL at α-level of 0.05 and 0.005

		Nuclear pedigree with two affected siblings			Nuclear pedigree with three affected siblings			Extended pedigree		
	α-level	5.0×10^{-2}	5.0×10^{-3}	1.5×10^{-5}	5×10^{-2}	5.0×10^{-3}	1.5×10^{-5}	5.0×10^{-2}	5.0×10^{-3}	1.5×10^{-5}
No missing genotype	CHP-NPL_{Pairs}	4.8×10^{-2}	4.5×10^{-3}	1.0×10^{-5}	4.9×10^{-2}	4.8×10^{-3}	1.5×10^{-5}	5.0×10^{-2}	5.0×10^{-3}	1.5×10^{-5}
	CHP-NPL_{All}	4.8×10^{-2}	4.5×10^{-3}	1.0×10^{-5}	4.9×10^{-2}	4.8×10^{-3}	1.5×10^{-5}	5.0×10^{-2}	4.9×10^{-3}	1.3×10^{-5}
	RV-NPL_{Pairs}	4.6×10^{-2}	4.3×10^{-3}	1.4×10^{-5}	4.9×10^{-2}	4.8×10^{-3}	1.7×10^{-5}	4.9×10^{-2}	4.8×10^{-3}	1.7×10^{-5}
	RV-NPL_{All}	4.6×10^{-2}	4.3×10^{-3}	1.4×10^{-5}	4.9×10^{-2}	4.9×10^{-3}	1.6×10^{-5}	4.9×10^{-2}	4.9×10^{-3}	1.0×10^{-5}
All founders missing genotype	CHP-NPL_{Pairs}	4.6×10^{-2}	4.8×10^{-3}	1.4×10^{-5}	4.6×10^{-2}	4.5×10^{-3}	1.0×10^{-5}	5.1×10^{-2}	5.3×10^{-3}	1.4×10^{-5}
	CHP-NPL_{All}	4.6×10^{-2}	4.8×10^{-3}	1.4×10^{-5}	4.6×10^{-2}	4.5×10^{-3}	1.0×10^{-5}	5.1×10^{-2}	5.2×10^{-3}	1.7×10^{-5}
	RV-NPL_{Pairs}	5.1×10^{-2}	5.2×10^{-3}	1.7×10^{-5}	5.1×10^{-2}	5.2×10^{-3}	1.7×10^{-5}	4.9×10^{-2}	4.6×10^{-3}	1.5×10^{-5}
	RV-NPL_{All}	5.1×10^{-2}	5.2×10^{-3}	1.7×10^{-5}	5.1×10^{-2}	5.2×10^{-3}	1.6×10^{-5}	5.0×10^{-2}	4.8×10^{-3}	1.5×10^{-5}

Exome-wide type I error was evaluated using data generated for 1000 exomes and analyzing each gene. Three different values for α-level are shown here: $5.0\text{x}10^{-2}$, $5.0\text{x}10^{-3}$ and 1.5×10^{-5}. Type I error rate was calculated by dividing the total number of genes with a p-value equal or smaller than the α-level value by the number of genes analyzed across all 1000 generated exomes.

Table S3. Power comparison of NPL_{Pairs} and NPL_{All} in intra-familial locus heterogeneity

	RV-NPL_{Pairs}	RV-NPL_{All}	$Z_{All} > Z_{Pairs}$[a]
Without intra-familial locus heterogeneity	0.6410	0.6411	69.08%
With intra-familial locus heterogeneity	0.2997	0.2870	38.52%

Power was compared between RV-NPL_{Pairs} and RV-NPL_{All} in extended families with and without intra-familial locus heterogeneity.

[a]Proportion of total genes that have Z-scores of RV-NPL_{All} higher than that of RV-NPL_{Pairs}

Table S4: Bioinformatic evaluation and frequencies of analyzed rare variants within *PSMF1*

dbSNP rsID	rs751905514*^	rs35236223*^	rs148476395*	rs146300768^	rs146612629	rs79465651*	rs148156083*^	rs758812434*
hg19 position	20:1106192	20:1106214	20:1115798	20:1115864	20:1115870	20:1143797	20:1145081	20:1145111
Reference Allele	A	G	A	C	T	T	G	G
Alternate Allele	G	A	G	T	A	C	A	A
cDNA change	c.181A>G	c.203G>A	c.400A>G	c.466C>T	c.472T>A	c.575T>C	c.725G>A	c.755G>A
ACC	p.Asn61Asp	p.Arg68Gln	p.Ile134Val	p.Arg156Trp	p.Phe158Ile	p.Val192Ala	p.Arg242His	p.Ser252Asn
MAF[a]	$7.22x10^{-6}$	$3.61x10^{-5}$	$6.90x10^{-5}$	$4.08x10^{-4}$	$2.78x10^{-4}$	$5.61x10^{-3}$	$1.49x10^{-3}$	$3.66x10^{-5}$
MAF (NFE)[b]	$1.58x10^{-5}$	$3.16x10^{-5}$	$1.07x10^{-4}$	$5.53x10^{-5}$	$3.95x10^{-4}$	$7.26x10^{-4}$	$2.50x10^{-3}$	0
MAF (AMR)[c]	0	$2.91x10^{-5}$	$5.96x10^{-5}$	$3.20x10^{-4}$	$5.23x10^{-4}$	$3.31x10^{-3}$	$9.30x10^{-4}$	0
GERP score	4.93	4.93	4.12	2.07	5.03	5.26	5.11	4.12
PhyloP score	3.37	6.37	2.02	0.27	0.81	4.56	8.44	3.37
CADD score[d]	18.7	34.0	6.0	23.9	22.6	12.7	28.4	12.1
FATHMM	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated
MutationTaster	disease causing	disease causing	disease causing	polymorphism	disease causing	polymorphism	disease causing	polymorphism
Polyphen-2 HVAR	possibly damaging	probably damaging	benign	probably damaging	benign	benign	benign	benign
PROVEAN	neutral	deleterious	neutral	deleterious	neutral	neutral	deleterious	neutral
SIFT	tolerated	damaging	tolerated	damaging	tolerated	tolerated	tolerated	tolerated
LRT	deleterious	deleterious	deleterious	neutral	neutral	neutral	deleterious	deleterious

Abbreviations: ACC, amino acid change; MAF, minor allele frequency; NFE, Non-Finnish European; AMR, Latino; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant.

[a]MAFs are from gnomAD (Genome Aggregation Database) combining all populations; [b]MAFs are from gnomAD NFE population; [c]MAFs are from gnomAD AMR population; [d]Scaled CADD score.

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

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

Table S5: Bioinformatic evaluation and frequencies of analyzed rare variants within *PTPN21*

dbSNP rsID	rs141951135*^	rs150736820*^	rs143571855	rs3825676*^	rs149927113	rs138752198*	rs146847601*^
hg19 position	14:88935348	14:88935351	14:88945312	14:88945407	14:88945485	14:88974290	14:89016641
Reference Allele	G	G	G	C	C	T	C
Alternate Allele	A	A	C	G	G	C	A
cDNA change	c.3308C>T	c.3305C>T	c.2463C>G	c.2368G>C	c.2290G>C	c.425A>G	c.121G>T
ACC	p.Pro1103Leu	p.Pro1102Leu	p.Asp821Glu	p.Gly790Arg	p.Val764Leu	p.Gln142Arg	p.Val41Leu
MAF[a]	$5.41x10^{-5}$	$1.61x10^{-3}$	$1.95x10^{-3}$	$1.84x10^{-2}$	$2.17x10^{-4}$	$3.58x10^{-4}$	$4.94x10^{-4}$
MAF (NFE)[b]	$3.58x10^{-5}$	$2.73x10^{-3}$	$1.85x10^{-5}$	$1.95x10^{-2}$	0	$6.00x10^{-4}$	$7.90x10^{-6}$
MAF (AMR)[c]	$2.08x10^{-4}$	$1.16x10^{-4}$	$7.87x10^{-4}$	$3.86x10^{-3}$	$3.74x10^{-5}$	$2.06x10^{-4}$	$1.16x10^{-4}$
GERP score	5.90	5.90	-6.17	4.66	-2.01	5.36	5.50
PhyloP score	9.48	3.71	-1.29	5.10	0.88	2.42	7.60
CADD score[d]	34.0	22.9	0.04	19.8	0.1	7.9	23.6
FATHMM	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated	tolerated
MutationTaster	disease_causing	disease causing	polymorphism	disease causing	polymorphism	disease causing	disease_causing
Polyphen-2 HVAR	probably damaging	benign	benign	probably damaging	benign	benign	probably damaging
PROVEAN	deleterious	deleterious	neutral	neutral	neutral	neutral	neutral
SIFT	damaging	damaging	tolerated	damaging	tolerated	tolerated	tolerated
LRT	deleterious	deleterious	deleterious	deleterious	neutral	neutral	deleterious

Abbreviations: ACC, amino acid change; MAF, minor allele frequency; NFE, Non-Finnish European; AMR, Latino; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant.

[a]MAFs are from gnomAD (Genome Aggregation Database) combining all populations; [b]MAFs are from gnomAD NFE population; [c]MAFs are from gnomAD AMR population; [d]Scaled CADD score.

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

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

Table S6: Bioinformatic evaluation and frequencies of analyzed rare variants within *ABCA7*

dbSNP rsID	rs146597357	rs151054304	NA*^	rs138055574	rs76282929^	rs149949633*^	rs111940546*
hg19 position	19:1041922	19:1042353	19:1043175	19:1044672	19:1048898	19:1048950	19:1051209
Reference Allele	C	C	A	G	G	G	G
Alternate Allele	A	T	G	A	C	A	T
cDNA change	c.253C>A	c.455C>T	c.715A>G	c.1144G>A	c.2274G>C	c.2326G>A	c.2740G>T
ACC	p.Leu85Met	p.Pro152Leu	p.Asn239Asp	p.Gly382Ser	p.Gln758His	p.Gly776Arg	p.Ala914Ser
MAF[a]	$3.08x10^{-4}$	$1.42x10^{-3}$	.	$5.34x10^{-5}$	$4.28x10^{-3}$	$1.13x10^{-4}$	$1.86x10^{-4}$
MAF (NFE)[b]	$4.95x10^{-5}$	$2.44x10^{-5}$	.	$2.38x10^{-5}$	$6.58x10^{-5}$	$1.95x10^{-4}$	0
MAF (AMR)[c]	$2.41x10^{-4}$	$7.63x10^{-4}$	.	0	$1.59x10^{-3}$	$8.91x10^{-5}$	$1.75x10^{-4}$
GERP score	1.68	2.06	3.04	2.83	3.99	3.99	3.4
PhyloP score	0.83	-0.85	2.17	0.73	-5.34	6.44	1.39
CADD score[d]	13.4	7.4	20.4	8.8	25.3	28.6	8.16
FATHMM	damaging	damaging	damaging	damaging	damaging	tolerated	tolerated
MutationTaster	polymorphism	polymorphism	polymorphism	polymorphism	polymorphism	disease_causing	polymorphism
Polyphen-2 HVAR	benign	benign	probably damaging	benign	probably damaging	probably damaging	benign
PROVEAN	neutral	neutral	deleterious	neutral	deleterious	deleterious	neutral
SIFT	tolerated	tolerated	damaging	tolerated	damaging	damaging	tolerated
LRT	.	.	.	.	.	.	.

Abbreviations: ACC, amino acid change; MAF, minor allele frequency; NFE, Non-Finnish European; AMR, Latino; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant.

[a]MAFs are from gnomAD (Genome Aggregation Database) combining all populations; [b]MAFs are from gnomAD NFE population; [c]MAFs are from gnomAD AMR population; [d]Scaled CADD score.

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

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

Table S6: Bioinformatic evaluation and frequencies of analyzed rare variants within ABCA7 (continued)

dbSNP rsID	rs947668738*	rs114614802^	rs369849959	rs184590335*^	rs73505232*^	rs114782266^
hg19 position	19:1053401	19:1054324	19:1056127	19:1057919	19:1058635	19:1059056
Reference Allele	G	G	G	C	C	G
Alternate Allele	C	A	A	T	T	A
cDNA change	c.3294G>C	c.3710G>A	c.4301G>A	c.4886C>T	c.5168C>T	c.5435G>A
ACC	p.Glu1098Asp	p.Arg1237His	p.Arg1434His	p.Ser1629Leu	p.Ser1723Leu	p.Arg1812His
MAF[a]	$6.37x10^{-5}$	$2.38x10^{-3}$	$2.52x10^{-5}$	$1.29x10^{-3}$	$1.21x10^{-2}$	$1.06x10^{-2}$
MAF (NFE)[b]	0	$2.49x10^{-5}$	$1.60x10^{-5}$	0	$1.74x10^{-4}$	$6.41x10^{-3}$
MAF (AMR)[c]	$1.19x10^{-3}$	$1.03x10^{-3}$	$8.76x10^{-5}$	$9.47x10^{-3}$	$5.01x10^{-3}$	$5.35x10^{-3}$
GERP score	1.25	3.64	-2.23	4.22	4.23	0.81
PhyloP score	2.42	0.36	-0.98	7.64	2.03	4.26
CADD score[d]	22.9	32.0	2.8	35.0	33.0	21.8
FATHMM	tolerated	damaging	damaging	damaging	damaging	damaging
MutationTaster	polymorphism	polymorphism	polymorphism	disease_causing	polymorphism	polymorphism
Polyphen-2 HVAR	benign	probably damaging	benign	benign	benign	benign
PROVEAN	neutral	deleterious	neutral	deleterious	deleterious	deleterious
SIFT	tolerated	damaging	tolerated	damaging	damaging	damaging
LRT	.	.	.	.	.	.

Abbreviations: ACC, amino acid change; MAF, minor allele frequency; NFE, Non-Finnish European; AMR, Latino; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant.

[a]MAFs are from gnomAD (Genome Aggregation Database) combining all populations; [b]MAFs are from gnomAD NFE population; [c]MAFs are from gnomAD AMR population; [d]Scaled CADD score.

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

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

Table S7: Bioinformatic evaluation and frequencies of analyzed rare variants within *ACE*

dbSNP rsID	rs148943954*^	rs3730043*^	rs765069550
hg19 position	17:61560846	17:61568577	17:61574683
Reference Allele	C	C	C
Alternate Allele	G	T	T
cDNA change	c.1513C>G	c.2747C>T	c.3877C>T
ACC	p.Pro505Ala	p.Thr916Met	p.His1293Tyr
MAF[a]	$5.37x10^{-4}$	$4.02x10^{-3}$	$2.26x10^{-5}$
MAF (NFE)[b]	$1.24x10^{-4}$	$6.50x10^{-3}$	$4.38x10^{-5}$
MAF (AMR)[c]	$7.34x10^{-4}$	$1.55x10^{-3}$	0
GERP score	4.90	4.25	-0.28
PhyloP score	3.27	2.39	0.93
CADD score[d]	25.4	28.8	15.1
FATHMM	damaging	tolerated	tolerated
MutationTaster	disease_causing	disease_causing	polymorphism
Polyphen-2 HVAR	possibly damaging	probably damaging	benign
PROVEAN	deleterious	deleterious	neutral
SIFT	damaging	damaging	damaging
LRT	deleterious	deleterious	neutral

Abbreviations: ACC, amino acid change; MAF, minor allele frequency; NFE, Non-Finnish European; AMR, Latino; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant.

[a]MAFs are from gnomAD (Genome Aggregation Database) combining all populations; [b]MAFs are from gnomAD NFE population; [c]MAFs are from gnomAD AMR population; [d]Scaled CADD score.

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

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

Table S8: Bioinformatic evaluation and frequencies of analyzed rare variants within *EPHA1*

dbSNP rsID	rs139482378*^	rs139711610*^
hg19 position	7:143088584	7:143091417
Reference Allele	C	C
Alternate Allele	T	T
cDNA change	c.2897G>A	c.2372G>A
ACC	Arg966His	p.Arg791His
MAF[a]	$6.01x10^{-4}$	$3.26x10^{-4}$
MAF (NFE)[b]	$1.12x10^{-3}$	$1.55x10^{-5}$
MAF (AMR)[c]	$3.67x10^{-4}$	$3.11x10^{-4}$
GERP score	5.24	4.67
PhyloP score	2.51	7.56
CADD score[d]	35.0	35.9
FATHMM	tolerated	damaging
MutationTaster	disease_causing	disease_causing
Polyphen-2 HVAR	probably damaging	probably damaging
PROVEAN	deleterious	deleterious
SIFT	damaging	damaging
LRT	deleterious	deleterious

Abbreviations: ACC, amino acid change; MAF, minor allele frequency; NFE, Non-Finnish European; AMR, Latino; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant.

[a]MAFs are from gnomAD (Genome Aggregation Database) combining all populations; [b]MAFs are from gnomAD NFE population; [c]MAFs are from gnomAD AMR population; [d]Scaled CADD score.

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

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

Table S9: Bioinformatic evaluation and frequencies of analyzed rare variants within *SORL1*

dbSNP rsID	rs1051430452*^	rs150609294*^	rs139710266*^	rs62617129	rs62622819	rs140327834*^	rs142884576*^
hg19 position	11:121360768	11:121384931	11:121384991	11:121444958	11:121485599	11:121495816	11:121498300
Reference Allele	A	A	A	A	T	A	C
Alternate Allele	G	C	G	G	A	T	T
cDNA change	c.707A>G	c.1112A>C	c.1172A>G	c.3346A>G	c.5439T>A	c.6194A>T	c.6401C>T
ACC	p.Asp236Gly	p.Asn371Thr	p.Tyr391Cys	p.Ile1116Val	p.His1813Gln	p.Asp2065Val	p.Thr2134Met
MAF[a]	$3.98x10^{-6}$	$1.37x10^{-3}$	$3.18x10^{-5}$	$5.31x10^{-3}$	$4.99x10^{-3}$	$2.54x10^{-3}$	$3.29x10^{-4}$
MAF (NFE)[b]	$8.79x10^{-6}$	$2.17x10^{-3}$	$4.40x10^{-5}$	$8.25x10^{-3}$	$8.97x10^{-3}$	$4.10x10^{-3}$	$5.89x10^{-4}$
MAF (AMR)[c]	0	$1.41x10^{-4}$	0	$2.65x10^{-3}$	$1.89x10^{-3}$	$1.53x10^{-3}$	$5.64x10^{-5}$
GERP score	5.68	5.66	5.56	-5.57	-8.35	5.32	5.74
PhyloP score	8.73	9.24	9.24	-0.74	-1.34	8.64	2.63
CADD score[d]	33.0	24.1	25.0	0.05	9.4	25.5	23.9
FATHMM	tolerated	tolerated	tolerated	damaging	tolerated	tolerated	damaging
MutationTaster	disease_causing	disease_causing	disease_causing	polymorphism	disease_causing	disease_causing	disease_causing
Polyphen-2 HVAR	probably damaging	possibly damaging	probably damaging	benign	benign	probably damaging	benign
PROVEAN	deleterious	deleterious	deleterious	neutral	neutral	deleterious	neutral
SIFT	damaging	damaging	tolerated	tolerated	tolerated	tolerated	damaging
LRT	deleterious	deleterious	deleterious	neutral	neutral	deleterious	neutral

Abbreviations: ACC, amino acid change; MAF, minor allele frequency; NFE, Non-Finnish European; AMR, Latino; CADD, Combined Annotation Dependent Depletion; FATHMM, Functional Analysis through Hidden Markov Models; PROVEAN, Protein Variation Effect Analyzer; SIFT, Sorting Intolerant From Tolerant.

[a]MAFs are from gnomAD (Genome Aggregation Database) combining all populations; [b]MAFs are from gnomAD NFE population; [c]MAFs are from gnomAD AMR population; [d]Scaled CADD score.

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

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

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.