Supplementary material.

Appendix e-1. Logistic Mixed model for Caribbean Hispanics (EFIGA study).

For full description of genotype array, quality control, imputation methods for the EFIGA study see Tosto et al. [1]. Population ancestral component were estimated employing the ADMIXTURE software [2]. Methods are reported in details elsewhere [1]. Briefly, we conducted supervised admixture analyses using African Yoruba (YRI), Whites with European Ancestry (CEU) from the 1000G project as surrogates for European and African ancestral populations respectively; eight Surui, 21 Maya, 14 Karitiana, 14 Pima and seven Colombian individuals from the Human Genome Diversity Project (HGDP) were used as surrogates for Native American ancestry [3]. We used ~70,000 autosomal SNPs that were I) genotyped in all three data sets (Caribbean Hispanics, 1000G and HGDP); II) common (i.e. MAF >5 %) and III) in linkage equilibrium. Global ancestral estimation for each subject were then used for kinship matrix estimation using the REAP (Relatedness Estimation in Admixed Populations) software [4] which is specifically designed to estimates autosomal kinship coefficients and identity-by-descent sharing probabilities using SNP genotype data in samples with admixed ancestry. The resulting kinship matrix was then used as random effect in the mixed model run by the GEMMA software [5]. Pipeline of this multi-step process is presented in **Figure e-1**.

Appendix e-2. Logistic mixed models codes (LOAD=late onset Alzheimer's disease; FID=family ID).

MODEL1:

LOAD ~ SEX + AGE + GRS + (1|FID), data, family = binomial(link=logit)

MODEL 2:

LOAD ~ SEX + AGE + APOE-e4 + GRS + (1|FID), data, family = binomial(link=logit)

MODEL 3:

LOAD ~ SEX + AGE + APOE-e4 + GRS + GRS*APOE-e4 + (1|FID), data, family = binomial(link=logit)

Table e-1. Single nucleotide polymorphisms (SNP) included in the genetic risk score (GRS) for NIA-LOAD families. Additional information provided: chromosomal (CHR) and base pair (BP) location, minor allele (A1) and major allele (A2), odds ratio (OR) derived from Lambert et al. [6]

SNP	CHR	BP	GENE	A1	A2	OR	MAF
rs6656401	1	207,692,049	*CR1*	A	G	1.18	0.197
rs6733839	2	127,892,810	*BIN1*	T	C	1.22	0.409
rs35349669	2	234,068,476	*INPP5D*	T	C	1.08	0.488
rs190982	5	88,223,420	*MEF2C*	G	A	0.93	0.408
rs9271192	6	32,578,530	*HLA-DRB5/1*	C	A	1.11	0.276
rs10948363	6	47,487,762	*CD2AP*	G	A	1.10	0.266
rs2718058	7	37,841,534	*NME8*	G	A	0.93	0.373
rs1476679	7	100,004,446	*ZCWPW1*	C	T	0.91	0.287
rs11771145	7	143,110,762	*EPHA1*	A	G	0.90	0.338
rs28834970	8	27,195,121	*PTK2B*	C	T	1.10	0.366
rs9331896	8	27,467,686	*CLU*	C	T	0.86	0.379
rs10838725	11	47,557,871	*CELF1*	C	T	1.08	0.316
rs983392	11	59,923,508	*MS4A6A*	G	A	0.90	0.403
rs10792832	11	85,867,875	*PICALM*	A	G	0.87	0.358
rs11218343	11	121,435,587	*SORL1*	C	T	0.77	0.039
rs17125944	14	53,400,629	*FERMT2*	C	T	1.14	0.092
rs10498633	14	92,926,952	*RIN-SLC24A4*	T	G	0.91	0.217
rs8093731*	18	29,088,958	*DSG2*	T	C	0.73	0.017
rs4147929	19	1,063,443	*ABCA7*	A	G	1.15	0.19
rs3865444*	19	51,727,962	*CD33*	A	C	0.94	0.307
rs7274581	20	55,018,260	*CASS4*	C	T	0.88	0.083

*these two SNP were found genome-wide significant in the IGAP discovery sample but not replicated in the final meta-analysis; therefore, they were excluded in the "conservative GRS" construct.

Table e-2. Single nucleotide polymorphisms (SNP) included in the genetic risk score (GRS) for EFIGA families.

SNP	CHR	BP	GENE	A1	A2	MAF	OR	SE	p-value
rs2796259	1	207,876,068	*CR1*	G	T	0.596	0.84	0.06	3.37E-03
rs56404717	2	127,922,465	*BIN1*	G	T	0.782	0.75	0.07	2.77E-05
rs141658619	2	233,992,565	*INPP5D*	G	A	0.986	0.54	0.24	1.10E-02
rs117958293	5	88,299,791	*MEF2C*	G	A	0.957	0.67	0.15	8.38E-03
rs687308	6	32,567,256	*HLA-DRB5/1*	C	T	0.875	0.82	0.09	2.62E-02
rs34746412	6	47,450,542	CD2AP	T	A	0.976	0.52	0.20	9.75E-04
rs59976014	7	37,937,114	*NME8*	C	T	0.978	1.82	0.20	3.29E-03
rs56307653	7	100,119,315	*ZCWPW1*	T	C	0.879	1.29	0.09	4.04E-03
rs182263486	7	143,082,748	*EPHA1*	T	C	0.971	0.46	0.19	4.41E-05
rs185236056	8	27,123,254	*PTK2B*	C	G	0.972	1.91	0.18	2.17E-04
rs142885341	8	27,414,325	*CLU*	C	T	0.983	0.49	0.24	2.47E-03
rs11605348	11	47,606,483	*CELF1*	G	A	0.761	0.86	0.07	2.54E-02
rs188448548	11	59,840,510	*MS4A6A*	C	T	0.975	1.81	0.20	3.17E-03
rs116136578	11	85,597,801	*PICALM*	A	T	0.988	2.71	0.27	2.62E-04
rs12280714	11	121,416,106	*SORL1*	C	T	0.989	2.54	0.28	9.93E-04
rs113575650	14	53,356,419	*FERMT2*	G	A	0.982	2.36	0.23	1.70E-04
rs1742703	14	92,704,484	*RIN-SLC24A4*	A	T	0.864	1.48	0.08	3.91E-06
rs7241860*	18	28,990,898	*DSG2*	G	C	0.968	1.81	0.17	6.91E-04
rs61242726	19	1,068,873	*ABCA7*	G	A	0.78	1.29	0.07	2.13E-04
rs11882065*	19	51,704,565	*CD33*	T	C	0.987	2.65	0.27	3.36E-04
rs6069777	20	55,130,979	*CASS4*	A	T	0.63	0.80	0.06	3.45E-04

*these two SNP were excluded in the “conservative GRS” construct.

OR: Odds ratio; SE=standard error

Figure e-1. Schematic representation of methods applied to the EFIGA cohort described in Appendix e-1 (large squares describe the statistics; small grey-shaped squares indicate the software applied).

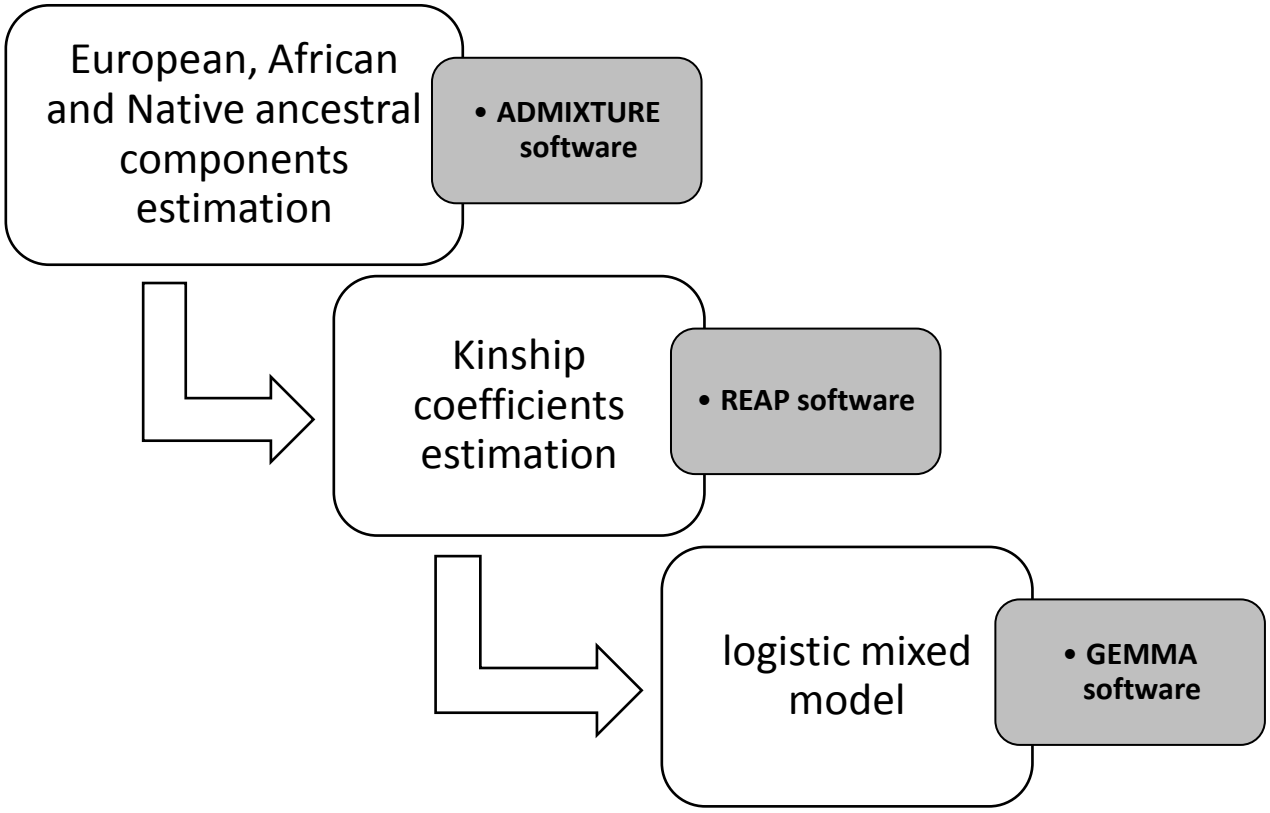
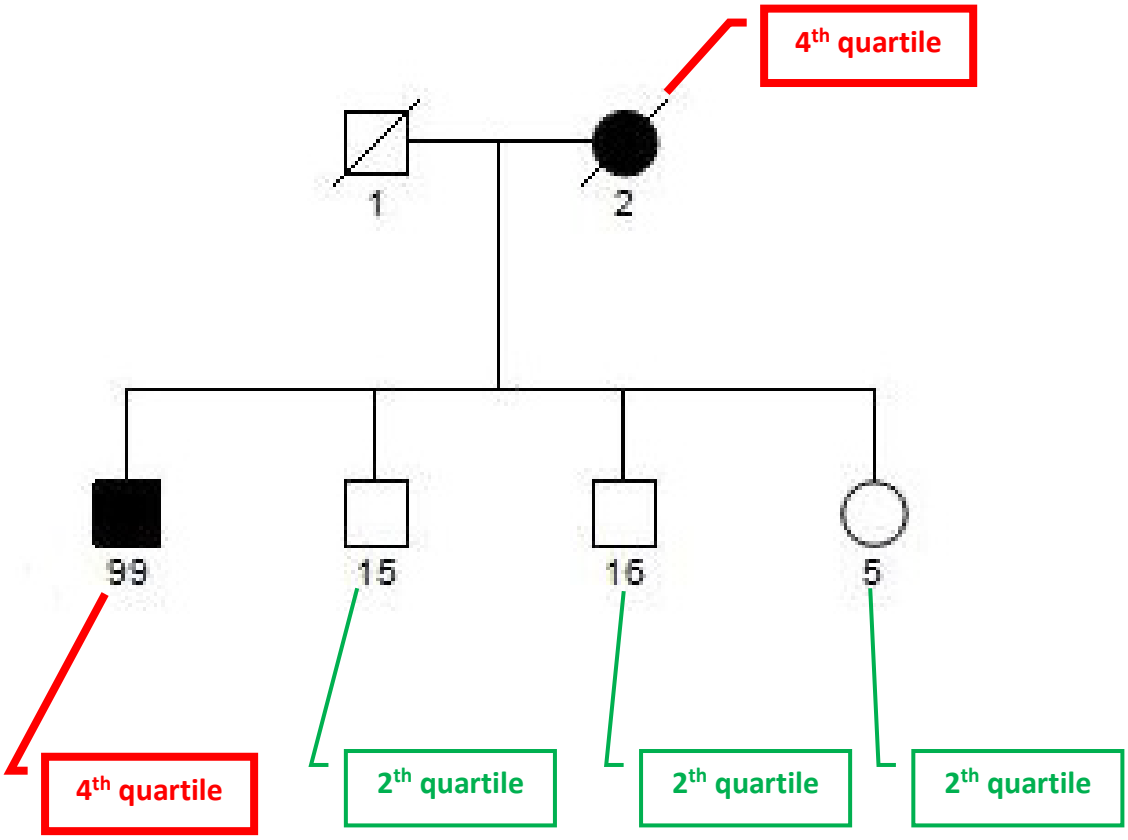
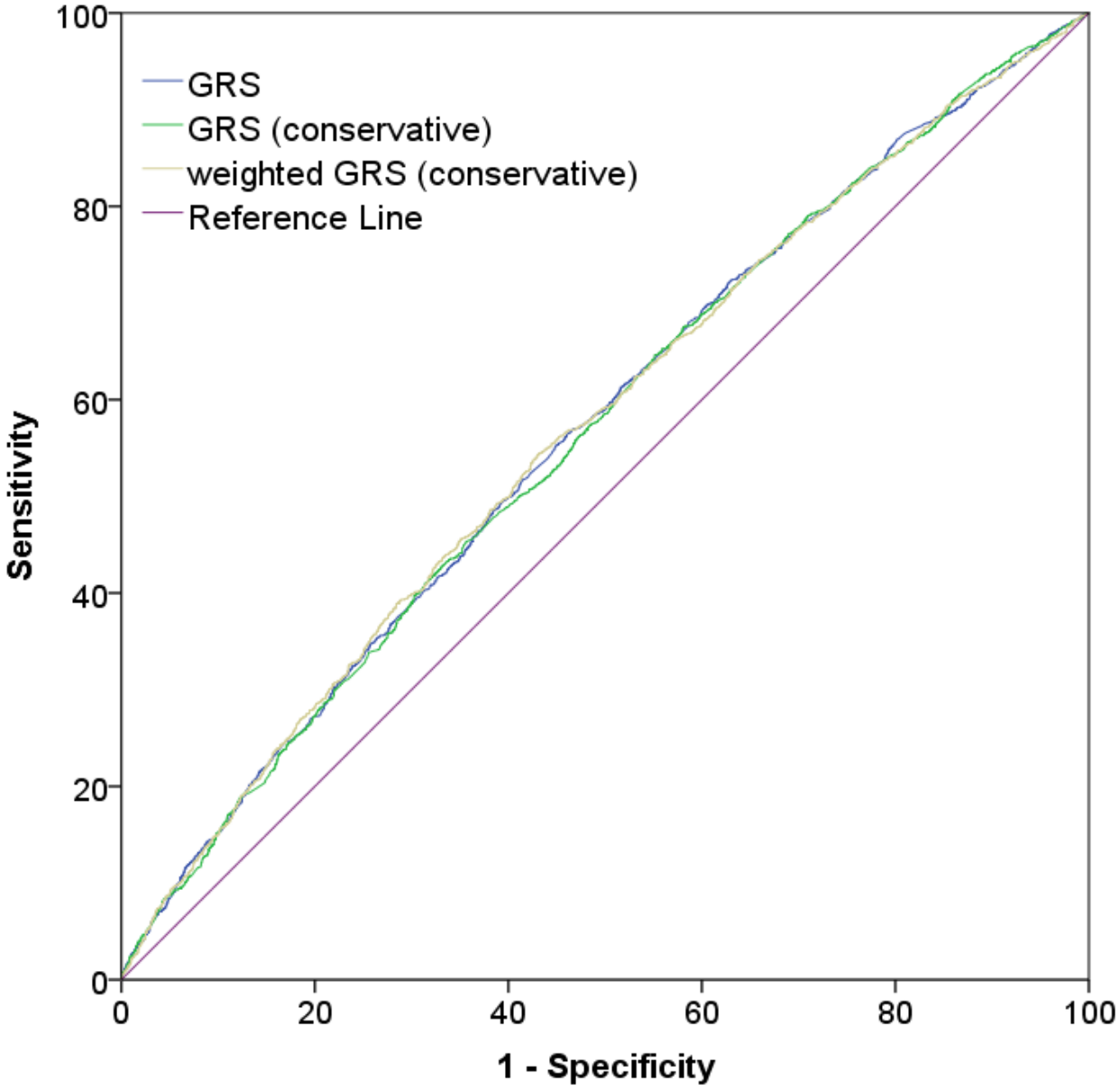


Figure e-2. an example of an EFIGA family pedigree included in the analyses; we reported the affection status (black = affected; white= unaffected; square= male; circle= female) and the corresponding GRS score (for clarity, the GRS was binned into quartiles. 2nd quartile labeled with green color, 4th quartile labeled with red color to designate riskier scores).



NB: no clinical information nor GWAS are available for subject "1"

Figure e-3. Clustered ROC curve for the NIA-LOAD dataset.



REFERENCES.

1. Tosto G, Fu H, Vardarajan BN, et al. F-box/LRR-repeat protein 7 is genetically associated with Alzheimer's disease. Annals of clinical and translational neurology. Wiley Online Library; 2015;2(8):810–820.

2. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome research. Cold Spring Harbor Lab; 2009;19(9):1655–1664.

3. Li JZ, Absher DM, Tang H, et al. Worldwide human relationships inferred from genome-wide patterns of variation. science. American Association for the Advancement of Science; 2008;319(5866):1100–1104.

4. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. Estimating kinship in admixed populations. Am J Hum Genet. 2012;91(1):122–138.

5. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nature genetics. Nature Publishing Group; 2012;44(7):821–824.

6. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45(12):1452–1458.