Supplementary Material for:

Common variants in *MS4A4*/*MS4A6E*, *CD2AP*, *CD33*, and *EPHA1* are associated with late-onset Alzheimer's disease.

Adam C Naj[1,115], Gyungah Jun[2-4,115], Gary W Beecham[1,5], Li-San Wang[6], Badri Narayan Vardarajan [3], Jacqueline Buros[3], Paul J Gallins[1], Joseph D Buxbaum[7-9], Gail P Jarvik[10,11], Paul K Crane[12], Eric B Larson[13], Thomas D Bird[14], Bradley F Boeve[15], Neill R Graff-Radford[16,17], Philip L De Jager[18,19], Denis Evans[20], Julie A Schneider[21,22], Minerva M Carrasquillo[16], Nilufer Ertekin-Taner[16,17], Steven G Younkin[16], Carlos Cruchaga[23], John SK Kauwe[24], Petra Nowotny[23], Patricia Kramer[25,26], John Hardy[27], Matthew J Huentelman[28], Amanda J Myers[29], Michael M Barmada[30], F. Yesim Demirci[30], Clinton T Baldwin[3], Robert C Green[3,31,32], Ekaterina Rogaeva[33], Peter St George-Hyslop[33,34], Steven E Arnold[35], Robert Barber[36], Thomas Beach[37], Eileen H Bigio[38], James D Bowen[39], Adam Boxer[40], James R Burke[41], Nigel J Cairns[42], Chris S Carlson[43], Regina M Carney[44], Steven L Carroll[45], Helena C Chui[46], David G Clark[47], Jason Corneveaux[28], Carl W Cotman[48], Jeffrey L Cummings[49], Charles DeCarli[50], Steven T DeKosky[51], Ramon Diaz-Arrastia[52], Malcolm Dick[48], Dennis W Dickson[16], William G Ellis[53], Kelley M Faber[54], Kenneth B Fallon[45], Martin R Farlow[55], Steven Ferris[56], Matthew P Frosch[57], Douglas R Galasko[58], Mary Ganguli[59], Marla Gearing[60,61], Daniel H Geschwind[62], Bernardino Ghetti[63], John R Gilbert[1,5], Sid Gilman[64], Bruno Giordani[65], Jonathan D Glass[66], John H Growdon[67], Ronald L Hamilton[68], Lindy E Harrell[47], Elizabeth Head[69], Lawrence S Honig[70], Christine M Hulette [71], Bradley T Hyman[67], Gregory A Jicha[72], Lee-Way Jin[53], Nancy Johnson[73], Jason Karlawish[74], Anna Karydas[40], Jeffrey A Kaye[26,75], Ronald Kim[76], Edward H Koo[58], Neil W Kowall[31, 77], James J Lah[66], Allan I Levey[66], Andrew P Lieberman[78], Oscar L Lopez[79], Wendy J Mack[80], Daniel C Marson[47], Frank Martiniuk[81], Deborah C Mash[82], Eliezer Masliah[58,83], Wayne C McCormick[12], Susan M McCurry[84], Andrew N McDavid[43], Ann C McKee[31,77], Marsel Mesulam[85,86], Bruce L Miller[40], Carol A Miller[87], Joshua W Miller[53], Joseph E Parisi[88,89], Daniel P Perl[90], Elaine Peskind[91], Ronald C Petersen[15], Wayne W Poon[48], Joseph F Quinn[26], Ruchita A Rajbhandary[1], Murray Raskind[91], Barry Reisberg[56,92], John M Ringman[49], Erik D Roberson[47], Roger N Rosenberg[52], Mary Sano[8], Lon S Schneider[46,93], William Seeley[40], Michael L Shelanski[94], Michael A Slifer[1,5], Charles D Smith[72], Joshua A Sonnen[95], Salvatore Spina[63], Robert A Stern[31], Rudolph E Tanzi[67], John Q Trojanowski[6], Juan C Troncoso[96], Vivianna M Van Deerlin[6], Harry V Vinters[49,97], Jean Paul Vonsattel[98], Sandra Weintraub[85,86], Kathleen A Welsh-Bohmer[41,]Adam C Naj[1,115], Gyungah Jun[2-4,115], Gary W Beecham[1,5], Li-San Wang[6], Badri Narayan Vardarajan [3], Jacqueline Buros[3], Paul J Gallins[1], Joseph D Buxbaum[7-9], Gail P Jarvik[10,11], Paul K Crane[12], Eric B Larson[13], Thomas D Bird[14], Bradley F Boeve[15], Neill R Graff-Radford[16,17], Philip L De Jager[18,19], Denis Evans[20], Julie A Schneider[21,22], Minerva M Carrasquillo[16], Nilufer Ertekin-Taner[16,17], Steven G Younkin[16], Carlos Cruchaga[23], John SK Kauwe[24], Petra Nowotny[23], Patricia Kramer[25,26], John Hardy[27], Matthew J Huentelman[28], Amanda J Myers[29], Michael M Barmada[30], F. Yesim Demirci[30], Clinton T Baldwin[3], Robert C Green[3,31,32], Ekaterina Rogaeva[33], Peter St George-Hyslop[33,34], Steven E Arnold[35], Robert Barber[36], Thomas Beach[37], Eileen H Bigio[38], James D Bowen[39], Adam Boxer[40], James R Burke[41], Nigel J Cairns[42], Chris S Carlson[43], Regina M Carney[44], Steven L Carroll[45], Helena C Chui[46], David G Clark[47], Jason Corneveaux[28], Carl W Cotman[48], Jeffrey L Cummings[49], Charles DeCarli[50], Steven T DeKosky[51], Ramon Diaz-Arrastia[52], Malcolm Dick[48], Dennis W Dickson[16], William G Ellis[53], Kelley M Faber[54], Kenneth B Fallon[45], Martin R Farlow[55], Steven Ferris[56], Matthew P Frosch[57], Douglas R Galasko[58], Mary

Nature Genetics: doi.10.1038/ng.801

Ganguli[59], Marla Gearing[60,61], Daniel H Geschwind[62], Bernardino Ghetti[63], John R Gilbert[1,5], Sid Gilman[64], Bruno Giordani[65], Jonathan D Glass[66], John H Growdon[67], Ronald L Hamilton[68], Lindy E Harrell[47], Elizabeth Head[69], Lawrence S Honig[70], Christine M Hulette [71], Bradley T Hyman[67], Gregory A Jicha[72], Lee-Way Jin[53], Nancy Johnson[73], Jason Karlawish[74], Anna Karydas[40], Jeffrey A Kaye[26,75], Ronald Kim[76], Edward H Koo[58], Neil W Kowall[31, 77], James J Lah[66], Allan I Levey[66], Andrew P Lieberman[78], Oscar L Lopez[79], Wendy J Mack[80], Daniel C Marson[47], Frank Martiniuk[81], Deborah C Mash[82], Eliezer Masliah[58,83], Wayne C McCormick[12], Susan M McCurry[84], Andrew N McDavid[43], Ann C McKee[31,77], Marsel Mesulam[85,86], Bruce L Miller[40], Carol A Miller[87], Joshua W Miller[53], Joseph E Parisi[88,89], Daniel P Perl[90], Elaine Peskind[91], Ronald C Petersen[15], Wayne W Poon[48], Joseph F Quinn[26], Ruchita A Rajbhandary[1], Murray Raskind[91], Barry Reisberg[56,92], John M Ringman[49], Erik D Roberson[47], Roger N Rosenberg[52], Mary Sano[8], Lon S Schneider[46,93], William Seeley[40], Michael L Shelanski[94], Michael A Slifer[1,5], Charles D Smith[72], Joshua A Sonnen[95], Salvatore Spina[63], Robert A Stern[31], Rudolph E Tanzi[67], John Q Trojanowski[6], Juan C Troncoso[96], Vivianna M Van Deerlin[6], Harry V Vinters[49,97], Jean Paul Vonsattel[98], Sandra Weintraub[85,86], Kathleen A Welsh-Bohmer[41, 99], Jennifer Williamson[70], Randall L Woltjer[100], Laura B Cantwell[6], Beth A Dombroski[6], Duane Beekly[101], Kathryn L Lunetta[2], Eden R Martin[1,5], M. Ilyas Kamboh[30,79], Andrew J Saykin[54,102], Eric M Reiman[28,103-105], David A Bennett[22,106], John C Morris[42,107], Thomas J Montine[95], Alison M Goate[23], Deborah Blacker[108,109], Debby W Tsuang[91], Hakon Hakonarson[110], Walter A Kukull[111], Tatiana M Foroud[54], Jonathan L Haines[112,113], Richard Mayeux[70,114], Margaret A Pericak-Vance[1,5], Lindsay A Farrer[2-4,31,32], Gerard D Schellenberg[6]

*A full list of author affiliations appears at the end of the paper.

Nature Genetics: doi.10.1038/ng.801

Supplementary Table 1. Sample size and descriptive statistics by dataset in the Discovery and Replication sample in the ADGC.

Cohort	Cases (N)	Cases					Controls (N)	Controls			
		Autopsied (N, %)	Female (N, %)	Age at onset (mean ± SD)	Age at exam (mean ± SD)	*APOE* ε2/ε3/ε4 (allele %)		Autopsied (N, %)	Female (N, %)	Age at exam (mean ± SD)	*APOE* ε2/ε3/ε4 (allele %)
Discovery (Stage 1)											
ACT	566	70 (12%)	357 (63%)	83.90 (4.8)	84.72 (4.9)	0.05/0.69/0.26	1696	155 (9%)	947 (56%)	81.08 (6.0)	0.08/0.81/0.11
ADC1	1566	1566 (100%)	850 (54%)	72.47 (7.1)	81.61 (7.0)	0.03/0.55/0.42	515	114 (22%)	305 (59%)	75.00 (8.0)	0.08/0.76/0.16
ADC2	738	195 (26%)	377 (51%)	73.19 (7.1)	80.06 (7.2)	0.03/0.57/0.39	160	0 (0%)	110 (69%)	75.68 (7.9)	0.09/0.75/0.16
ADNI	268	0 (0%)	113 (42%)	75.30 (7.2)	77.96 (6.5)	0.03/0.55/0.42	173	0 (0%)	70 (40%)	78.6 (5.5)	0.08/0.79/0.14
GenADA	669	9 (1%)	380 (57%)	74.59 (6.2)	80.36 (6.2)	0.04/0.58/0.38	713	0 (0%)	456 (64%)	74.21 (7.0)	0.08/0.79/0.13
UM/VU/MSSM	1186	409 (34%)	764 (64%)	74.06 (7.8)	77.48 (6.9)	0.03/0.61/0.36	1135	136 (12%)	696 (61%)	74.00 (8.3)	0.08/0.80/0.12
MIRAGE	509	0 (0%)	324 (64%)	71.16 (6.5)	75.97 (6.6)	0.04/0.60/0.36	753	0 (0%)	440 (58%)	72.04 (7.2)	0.06/0.72/0.23
NIA-LOAD	1811	492 (27%)	1176 (65%)	73.57 (6.7)	82.49 (7.1)	0.02/0.51/0.46	1575	50 (3%)	947 (60%)	73.99 (8.5)	0.07/0.73/0.20
OHSU	132	132 (100%)	81 (61%)	86.10 (5.5)	90.40 (5.2)	0.07/0.70/0.23	153	153 (100%)	84 (55%)	83.86 (7.6)	0.10/0.82/0.08
TGEN2	864	864 (100%)	633 (73%)	74.91 (7.2)	82.00 (7.6)	0.04/0.57/0.40	493	493 (100%)	186 (38%)	80.19 (8.7)	0.10/0.79/0.11
TOTAL	**8309**	**3737 (45%)**	**5055 (61%)**	--	--	--	**7366**	**1101 (15%)**	**4241 (58%)**	--	--
Replication (Stage 2)											
ADC3	897	527 (59%)	490 (55%)	75.00 (8.5)	80.51 (8.9)	0.04/0.59/0.37	588	4 (1%)	371 (63%)	75.30 (9.8)	0.08/0.78/0.14
MAYO	728	221 (30%)	419 (58%)	ND	73.89 (4.9)	0.02/0.56/0.42	1173	216 (18%)	601 (51%)	73.30 (4.4)	0.08/0.77/0.15
ROSMAP	296	291 (98%)	208 (70%)	85.59 (6.3)	89.83 (5.7)	0.05/0.75/0.20	776	0 (0%)	559 (72%)	82.03 (7.0)	0.10/0.81/0.10
UP	1271	277 (22%)	802 (63%)	72.91 (6.4)	77.38 (6.3)	0.03/0.63/0.34	841	2 (0.2%)	533 (63%)	75.37 (6.1)	0.09/0.81/0.10
WU	339	0 (0%)	194 (57%)	ND	74.23 (8.0)	0.06/0.63/0.32	187	0 (0%)	113 (60%)	76.85 (8.4)	0.07/0.78/0.15
TOTAL	**3531**	**1039 (29%)**	**2113 (60%)**	--	--	--	**3565**	**220 (6%)**	**2177 (61%)**	--	--
Discovery + Replication	**11840**	**4776 (40%)**	**7168 (61%)**	--	--	--	**10931**	**1321 (12%)**	**6418 (59%)**	--	--

Nature Genetics: doi.10.1038/ng.801

Supplementary Table 2. Description of Public Availability of Discovery and Replication Datasets for Analysis[a].

Cohort	Cases (N)	Controls (N)	Available through request on dbGAP (Yes/No)?	Study Website	Study Contact (Study E-mail)
ACT	566	1696	No	(ACT) http://www.grouphealthresearch.org/capabilities/clinic/clin_std.html#act (eMERGE) https://www.mc.vanderbilt.edu/victr/dcc/projects/acc/index.php/Main_Page	Eric B. Larson, MD, MPH, MACP (larson.e@ghc.org)
ADC (1, 2, & 3)	3201	1263	No	(NACC) https://www.alz.washington.edu/ (NCRAD) http://ncrad.iu.edu/	NACC: Walter Kukull, PhD (naccmail@u.washington.edu) NCRAD: Tatiana Foroud, PhD (alzstudy@iupui.edu)
ADNI	268	173	No	http://adni.loni.ucla.edu/ (Online Application)	Andrew J. Saykin, PsyD (asaykin@iupui.edu)
GenADA	669	713	Yes	http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000219.v1.p1	dbGaP: JAAMHDAC@mail.nih.gov
UM/VU/MSSM	1186	1135	No	http://www.hihg.org/ (Contact Directly)	Margaret A. Pericak-Vance, PhD (mpericak@med.miami.edu)
MIRAGE	509	753	No	http://www.bumc.bu.edu/genetics/research/alzheimers-disease/	Lindsay A. Farrer, PhD (farrer@bu.edu)
NIA-LOAD	1811	1575	Yes	http://www.niageneticsinitiative.org/	Richard Mayeux, MD, MSc (rpm2@columbia.edu)
OHSU	132	153	No	http://www.ohsu.edu/xd/research/centers-institutes/neurology/alzheimers/research/data-tissue/biomarkers-genetics.cfm	Patricia L. Kramer (kramer@ohsu.edu)
TGEN2	864	493	No	http://www.tgen.org/research/index.cfm?pageid=1065	Eric Reiman, MD (eric_reiman@tgen.org)
MAYO	728	1173	No	http://mayoresearch.mayo.edu/mayo/research/alzheimers_center/	Steven G. Younkin, MD, PhD (younkin.steven@mayo.edu)
ROSMAP	296	776	Yes	https://www.radc.rush.edu/res/ext/docs/Rush_ADCC_Data_Sharing.html	David Bennett, MD (dbennett@rush.edu) Data Sharing: Gregory Klein (Gregory_Klein@rush.edu)
UP	1271	841	No	http://www.adrc.pitt.edu/neurocore.asp (Contact Directly)	M. Ilyas Kamboh, PhD (kamboh@pitt.edu)
WU	339	187	No	http://www.niageneticsdata.org/ (Online Application)	Alison M. Goate, DPhil (goatea@psychiatry.wustl.edu)
ADGC Discovery + Replication	**11840**	**10931**		**(Main Study) http://alois.med.upenn.edu/adgc/** **(Dataset Access) http://www.niageneticsdata.org/**	**Gerard Schellenberg (gerardsc@mail.med.upenn.edu)** **Data Sharing: Li-San Wang (lswang@mail.med.penn.edu)**

[a] After publication, we will make available to qualified investigators final results from both the meta-analysis and the joint analysis via the ADGC (http://alois.med.upenn.edu/adgc/about/overview.html) and NIAGADS (http://niageneticsdata.org/) web sites. Genotype and phenotype data will be placed in dbGaP for datasets where consent forms allow.

Nature Genetics: doi.10.1038/ng.801

Supplementary Table 3. Genotyping platform and quality control by dataset in the Discovery and Replication sample in the ADGC.

Cohort		Genotyping platform	Number of SNPs with call rate threshold	Number of Individuals with completeness threshold (cases/controls)	Principal components used to adjust for population substructure	Allelic Concordance Rate[a]	Number of SNPs in the final set	Inflation Factor
Discovery								
	ACT	Illumina 660	536,993	566/1696	1st, 2nd, 3rd	0.97	2,442,125	0.993
	ADC1	Illumina 660	534,380	1566/515	1st, 2nd, 3rd	0.95	2,440,272	1.015
	ADC2	Illumina 660	527,149	738/160	1st, 2nd, 3rd	0.97	2,446,428	1.005
	ADNI	Illumina 610	548,414	268/173	1st, 2nd	0.97	2,448,857	0.999
	GSK	Affymetrix 500	442,833	669/713	1st, 2nd	0.91	2,325,107	1.008
	UM/VU/MSSM	Illumina 550 Illumina 1M Illumina 1M-Duo Affy 6.0 Illumina 610-Quad	1,477,026	1186/1135	1st, 2nd, 3rd, 4th	0.96	2,437,465	1.024
	MIRAGE	Illumina 610 Illumina 330	562,414	509/753	1st, 2nd, 3rd	0.97	2,389,722	1.015
	OHSU	Illumina 610	558,930	132/153	1st, 2nd	0.97	2,437,824	1.012
	NIA LOAD	Illumina 370	331,230	1811/1575	1st, 2nd	0.96	2,391,156	1.036
	TGEN2	Affymetrix 1M	658,617	864/493	1st, 2nd, 3rd	0.93	2,479,052	1.015
	Meta-Analysis	--	--	**8309/7366**	--	--	**2,312,972**	**1.027**
	Joint Analysis	--	--		--	--		**1.052**
Replication								
	ADC3	Illumina OmniExpress	661,363	897/588	1st, 2nd, 3rd	0.97	--	--
	MAYO	Affymetrix 6.0	309,603	728/1173	1st, 2nd, 3rd, 4th	0.95	--	--
	ROSMAP	Illumina 1M	635,774	296/776	1st, 2nd	0.97	--	--
	UP	Illumina Omni1-quad	738,049	1271/841	1st, 2nd, 3rd	0.96	--	--
	WU	Illumina 660	546,354	339/187	1st, 2nd	0.97	--	--
	Meta-Analysis	--	--	**3531/3565**	--	--	--	--
	Joint Analysis	--	--		--	--	--	--
Discovery + Replication								
	Meta-Analysis	--	--	**11840/10931**	--	--	--	--

[a] Allelic concordance rate for imputation was calculated using the *mask* option to the MACH program. Values reported are the average concordance rate among all genotyped SNPs on chromosome 19, which showed the worst imputation quality of all chromosomes.

Nature Genetics: doi.10.1038/ng.801

Supplementary Table 4. Samples excluded through sample quality control measures by dataset in the ADGC.

Cohort	Low call rate	Gender mismatch	Relatedness	Not Caucasian	Age < 60	Control with MMSE < 26	Unclear phenotype	Duplicates	Final Sample (% Original Sample)
Discovery									
ACT	0	2	28	215	0	168	111	6	2,262 (81.0%)
ADC1	26	24	51	290	296	32	56	43	2,081 (71.8%)
ADC2	23	9	4	5	4	3	23	203	898 (76.6%)
ADNI	67	0	2	58	17	0	195	39	441 (53.8%)
GSK	0	0	17	31	150	8	2	1	1,382 (86.9%)
UM/VU/MSSM	0	0	0	0	0	0	0	149	2,321 (94%)
MIRAGE	0	0	0	0	227	0	13	0	1,262 (84%)
OHSU	20	31	43	0	0	0	254	73	285 (40.4%)
NIA LOAD	0	0	0	705	643	0	486	0	3,386 (64.9%)
TGEN2	79	0	1	27	0	0	1	161	1,357 (83.5%)
Replication									
ADC3	44	31	27	391	18	1	7	101	1,485 (70.5%)
MAYO	41	0	25	37	0	0	0	95	1,901 (90.6%)
ROSMAP	13	0	0	0	0	0	515	109	1,072 (62.7%)
UP	103	47	74	14	8	79	5	20	2,112 (85.8%)
WU	0	0	0	3	22	0	20	102	526 (78.2%)
Discovery + Replication	416	144	272	1,776	1,385	291	1,688	1,102	22,771 (76.3%)

Nature Genetics: doi.10.1038/ng.801

Supplementary Table 8. Summary for LOAD associations with chromosome 19 SNPs in the region encompassing *NKPD1* and *EXOC3L2* with $P < 1\times10^{-4}$. Results are presented for the basic model and the extended model, which includes adjustment for age, gender, and dosage of the *APOE* ε4 allele, in the combined Discovery and Replication ADGC cohorts.

Chr	SNP	Gene	MA	MAF	Basic Model				Extended Model			
					Meta-Analysis OR (95% CI)	Meta-Analysis $P^{\ddagger}$	Joint Analysis OR (95% CI)	Joint Analysis $P^{\ddagger}$	Meta-Analysis OR (95% CI)	Meta-Analysis $P^{\ddagger}$	Joint Analysis OR (95% CI)	Joint Analysis $P^{\ddagger}$
19	rs17643262	*NKPD1*	A	0.10	1.33 (1.25-1.42)	5.1E-18	1.34 (1.24-1.45)	2.7E-14	1.04 (0.97-1.12)	0.30	1.02 (0.94-1.11)	0.67
19	rs1114832	*NKPD1*	T	0.11	1.33 (1.25-1.41)	3.2E-18	1.34 (1.24-1.44)	2.4E-14	1.05 (0.98-1.13)	0.18	1.03 (0.94-1.12)	0.56
19	rs1114831	*NKPD1*	A	0.10	1.33 (1.25-1.42)	4.1E-18	1.34 (1.24-1.45)	2.9E-14	1.04 (0.97-1.12)	0.30	1.02 (0.93-1.11)	0.69
19	rs1048699	*NKPD1*	T	0.10	1.33 (1.25-1.42)	6.2E-18	1.34 (1.24-1.44)	5.2E-14	1.04 (0.97-1.12)	0.31	1.01 (0.93-1.11)	0.74
19	rs10416371	*NKPD1*	C	0.46	1.15 (1.10-1.21)	1.8E-10	1.17 (1.11-1.23)	5.5E-09	1.02 (0.97-1.07)	0.50	1.04 (0.98-1.10)	0.22
19	rs606757	*EXOC3L2*	C	0.161	1.17 (1.10-1.25)	9.81E-07	1.17 (1.10-1.24)	6.39E-07	1.03 (0.96-1.11)	0.37	1.03 (0.96-1.10)	0.42
19	rs582747	*EXOC3L2*	G	0.18	1.19 (1.12-1.26)	2.7E-09	1.18 (1.10-1.26)	1.2E-06	1.03 (0.97-1.10)	0.29	1.01 (0.93-1.08)	0.86
19	rs605003	*EXOC3L2*	G	0.17	1.19 (1.12-1.26)	2.0E-09	1.17 (1.10-1.25)	1.2E-06	1.03 (0.97-1.10)	0.29	1.01 (0.94-1.08)	0.86
19	rs2627641	*EXOC3L2*	G	0.17	1.18 (1.12-1.25)	1.9E-09	1.17 (1.10-1.25)	1.4E-06	1.03 (0.97-1.10)	0.31	1.01 (0.94-1.08)	0.87
19	rs597668	*EXOC3L2*	C	0.17	1.18 (1.12-1.25)	1.5E-09	1.17 (1.10-1.25)	1.4E-06	1.03 (0.97-1.10)	0.30	1.01 (0.94-1.08)	0.87
19	rs7249082	*EXOC3L2*	G	0.17	1.19 (1.12-1.25)	1.1E-09	1.17 (1.10-1.25)	1.0E-06	1.03 (0.97-1.10)	0.28	1.01 (0.94-1.08)	0.85
19	rs10415983	*EXOC3L2*	T	0.18	1.19 (1.13-1.26)	1.3E-09	1.18 (1.10-1.26)	8.6E-07	1.04 (0.98-1.10)	0.25	1.01 (0.94-1.09)	0.75

Chr: chromosome number; SNP: SNP rs id; MA: minor allele; MAF: weighted-average minor allele frequency; OR: odds ratio; 95% CI: 95% confidence interval; *P*: P-value.

Nature Genetics: doi.10.1038/ng.801

Supplementary Table 9. Estimates of Power (Post Hoc) to detect observed associations and estimates of the population attributable fraction among the most significantly associated SNPs in nine genomic regions from Stage 1 discovery analysis. Odds ratios used to estimate PAFs were derived from GEE models incorporating all ten SNPs, the first model without adjustment for dosage of the *APOE* ε4 allele, the second including adjustment for *APOE* ε4 allelic dosage.

SNP	CH:MB	Nearest Gene	MA	MAF	Power						Population Attributable Fraction (PAF)[c]					
					Power[a] in ADGC Discovery (Stage 1)				Detectable OR at 80% Power‡		Unadjusted for *APOE* ε4 dose			Adjusted for *APOE* ε4 dose		
					OR_M	$Power_M$	OR_J	$Power_J$	OR_P	OR_D	OR	MAF_c	PAF	OR	MAF_c	PAF
rs6701713	1:207.8	*CR1*	A	0.20	1.18	84%	1.19	84%	0.87	1.17	1.19	0.19	0.03	1.19	0.19	0.03
rs7561528	2:127.9	*BIN1*	A	0.35	1.18	98%	1.18	98%	0.91	1.16	1.18	0.33	0.06	1.20	0.33	0.06
rs9349407	6:47.5	*CD2AP*	C	0.27	1.14	61%	1.14	61%	0.88	1.18	1.14	0.26	0.03	1.15	0.26	0.04
rs11767557	7:143.1	*EPHA1*	C	0.19	0.85	83%	0.84	83%	0.86	1.18	0.85	0.2	0.04	0.85	0.2	0.03
rs1532278	8:27.5	*CLU*	T	0.36	0.90	83%	0.89	83%	0.91	1.21	0.89	0.37	0.04	0.89	0.37	0.04
rs4938933	11:60.0	*MS4A4A*	C	0.39	0.88	100%	0.87	100%	0.99	1.16	0.88	0.41	0.05	0.89	0.41	0.05
rs561655	11:85.8	*PICALM*	G	0.34	0.88	98%	0.88	98%	0.95	1.17	0.88	0.35	0.05	0.91	0.35	0.03
rs3752246	19:1.1	*ABCA7*	G	0.19	1.16	50%	1.15	50%	0.83	1.21	1.15	0.18	0.03	1.14	0.18	0.02
rs3865444	19:51.7	*CD33*	A	0.30	0.88	94%	0.88	94%	0.91	1.17	0.89	0.32	0.04	0.89	0.32	0.04
									Combined PAF[d] (All SNPs)				0.31			0.50

CH:MB, chromosome:position (in megabasepairs); MA, minor allele; MAF, minor allele frequency; # SNPs, the number of SNPs for which $P \leq 1\times10^{-6}$ in meta-analysis from the combined analysis in Stage 1+2; OR_M, odds ratio in meta-analysis; $Power_M$, Power to detect observed OR_M; OR_J, odds ratio in joint analysis; $Power_J$, Power to detect observed OR_J.

[a] Power was calculated under an additive model with prevalence of 0.13 (US prevalence of AD over age 65) and significance level of $P<10^{-6}$ given sample size, marker frequency, and effect size for each SNP in Stage 1 (Online Methods ref. 76). We assumed that the frequency of marker and disease variant is similar and LD (r^2) between them is 0.8.
[b] Odds ratio (OR) thresholds were calculated for protective (OR_P) or deleterious (OR_D) alleles to have at least 80% power given sample size and marker frequency.
[c] Population Attributable Fraction was estimated using Odds Ratios (ORs) taken from a multivariate GEE model incorporating all SNPs so that each OR is adjusted for the effects of the other nine SNPs (and in the adjusted model, each SNP is adjusted for the other nine SNPs and dosage of the *APOE* ε4 allele)

Nature Genetics: doi.10.1038/ng.801

$$PAF = \frac{(AF_{ctrls} * (OR - 1))}{(AF_{ctrls} * (OR - 1)) + 1}$$

where AF_{ctrls} is the risk allele frequency in controls.

[d] Combined Population Attributable Risk was estimated using a formula original derived by Bruzzi et al. (Online Methods ref. 76)):

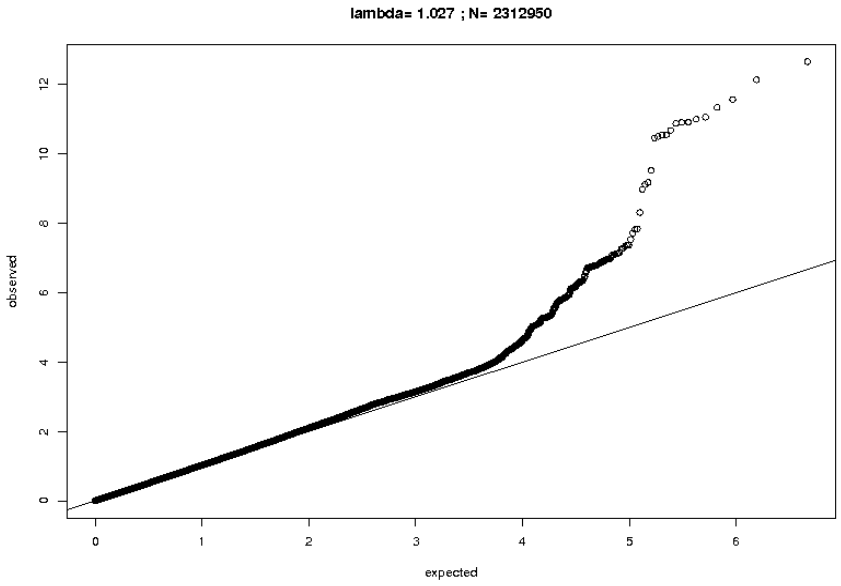
$$Combined\ PAF = 1 - \prod_{all\ \mathrm{SNPs}} (1 - PAF_i)$$

where PAF_i is the population attributable fraction for SNP i.

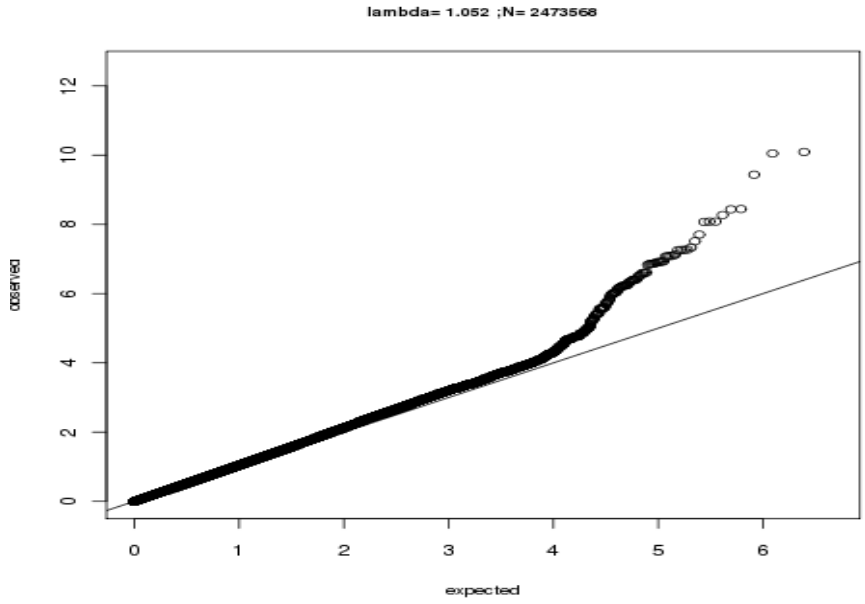
Nature Genetics: doi.10.1038/ng.801

Supplementary Figure 1. Quantile-quantile plots of observed (y-axis) vs. expected (x-axis) *P*-values for LOAD from the ADGC Discovery sample (Stage 1) under the basic model using (**A**) meta-analysis and (**B**) joint analysis approaches after filtering out SNPs with $P < 10^{-13}$ that reside in the *APOE* region on chromosome 19.

A



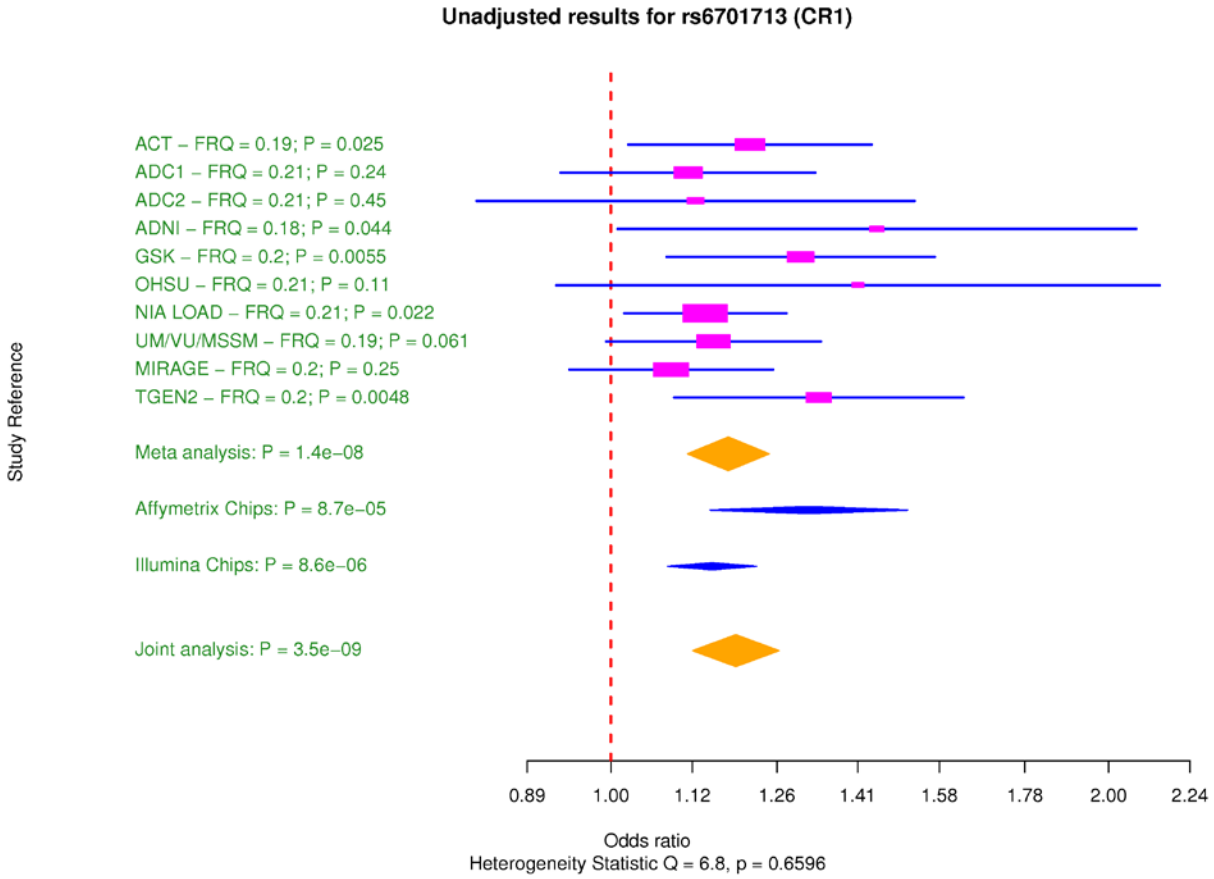
B



Nature Genetics: doi.10.1038/ng.801

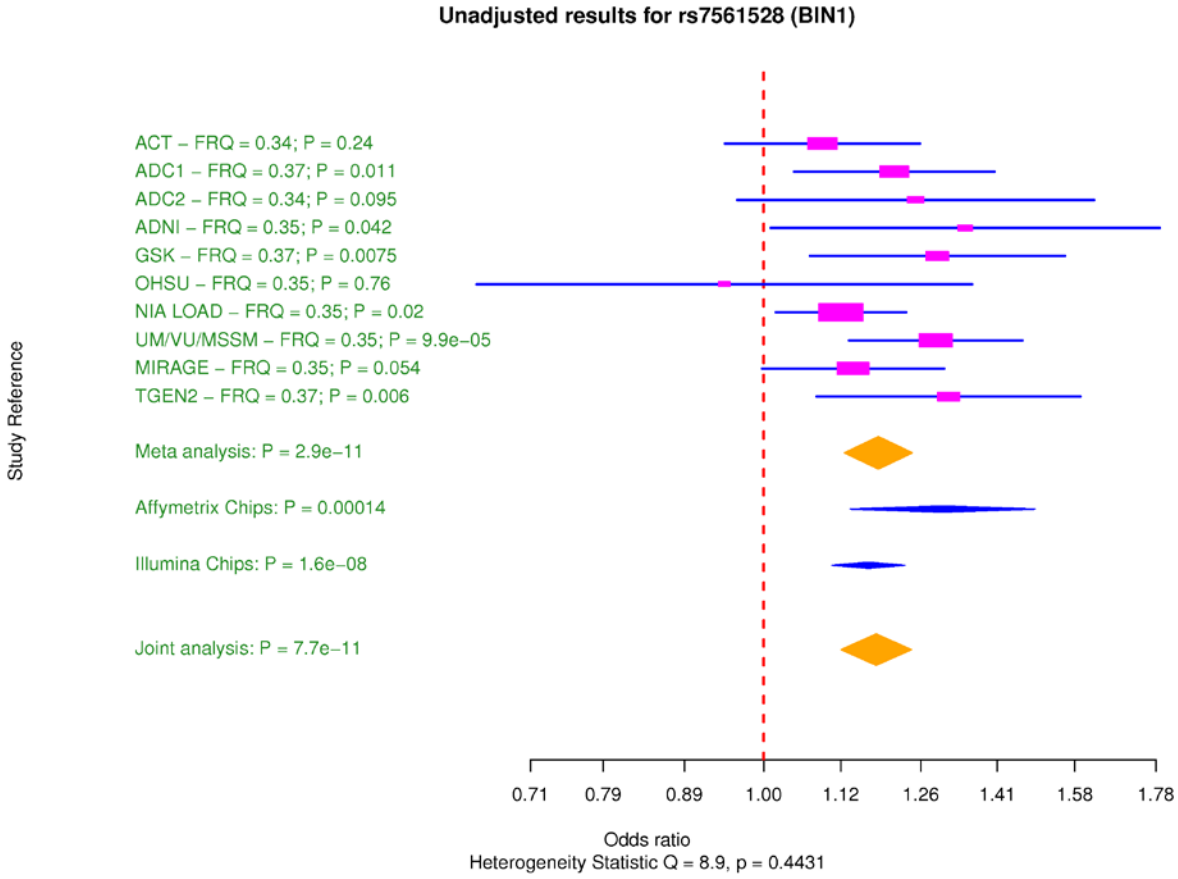
Supplementary Figure 2. Forest plots of SNPs with the most statistically significant associations in nine genomic regions from Stage 1 discovery analysis under the basic model of covariate adjustment (population substructure only in meta-analysis; site and population substructure in joint analysis). SNPs depicted include (**A**) rs6701713 in the chromosome 1 *CR1* region, (**B**) rs7561528 in the chromosome 2 *BIN1* region, (**C**) rs9349407 in the chromosome 6 *CD2AP* region, (**D**) rs11767557 in the chromosome 7 *EPHA1* region, (**E**) rs1532278 in the chromosome 8 *CLU* region, (**F**) rs4938933 in the chromosome 11 *MS4A* cluster, (**G**) rs561655 in the chromosome 11 *PICALM* region, (**H**) rs3752246 in the chromosome 19 *ABCA7* region, and (**J**) rs3865444 in the chromosome 19 *CD33* region.

A



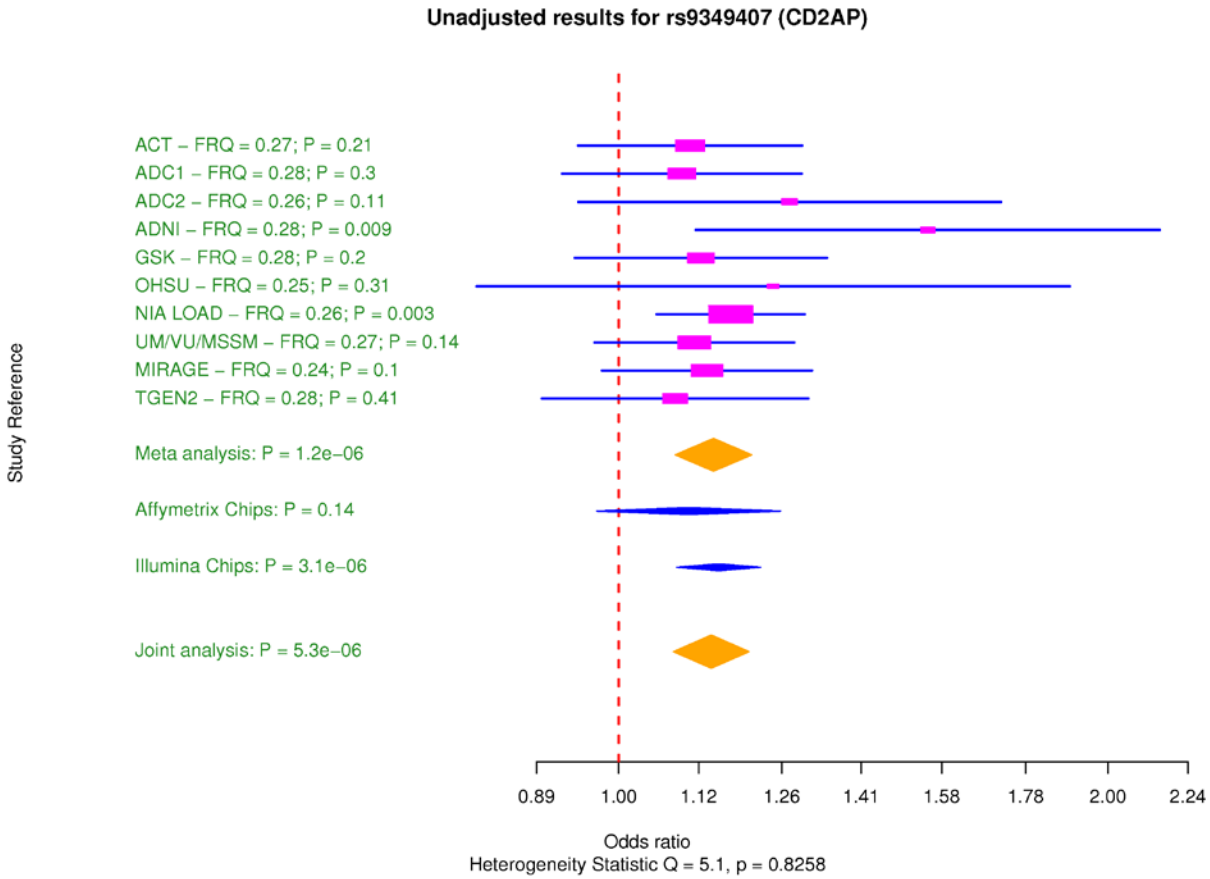
Nature Genetics: doi.10.1038/ng.801

B



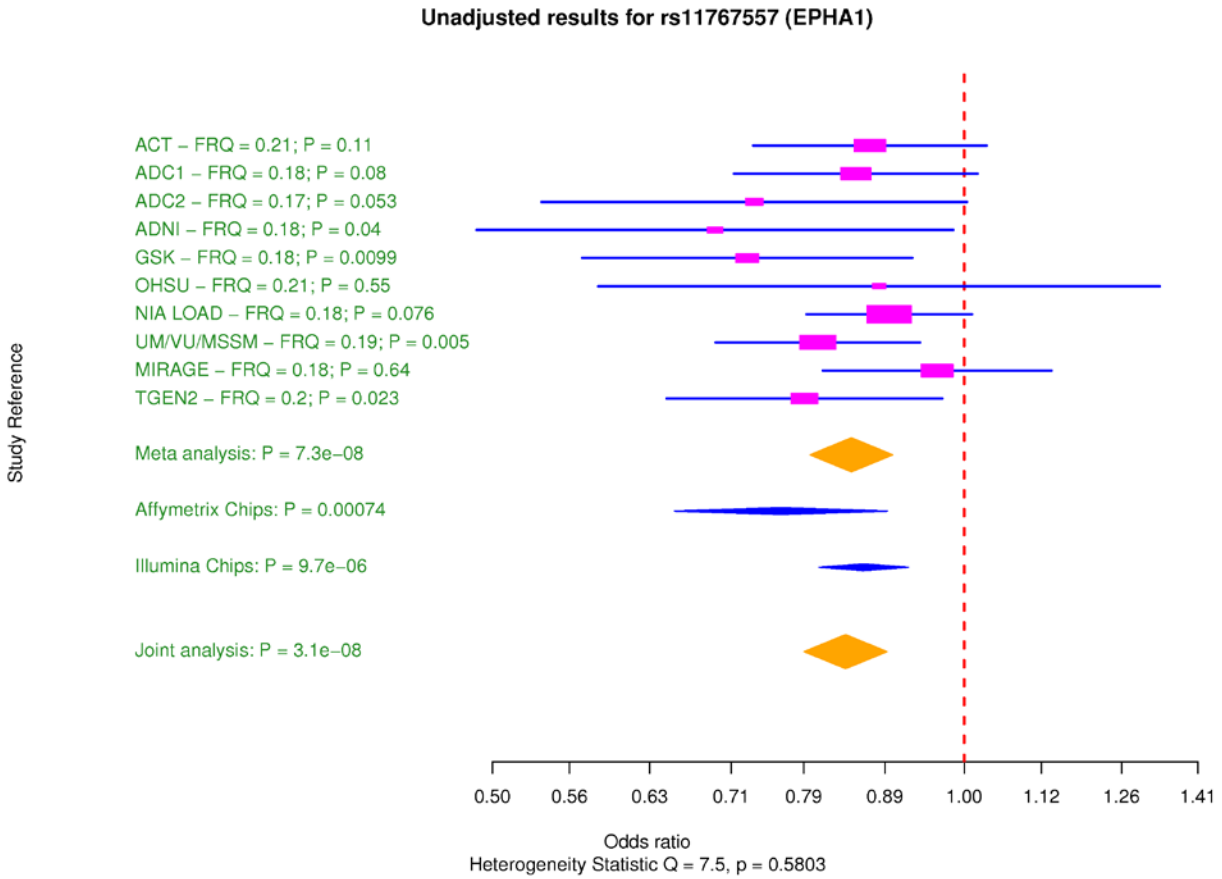
Nature Genetics: doi.10.1038/ng.801

C



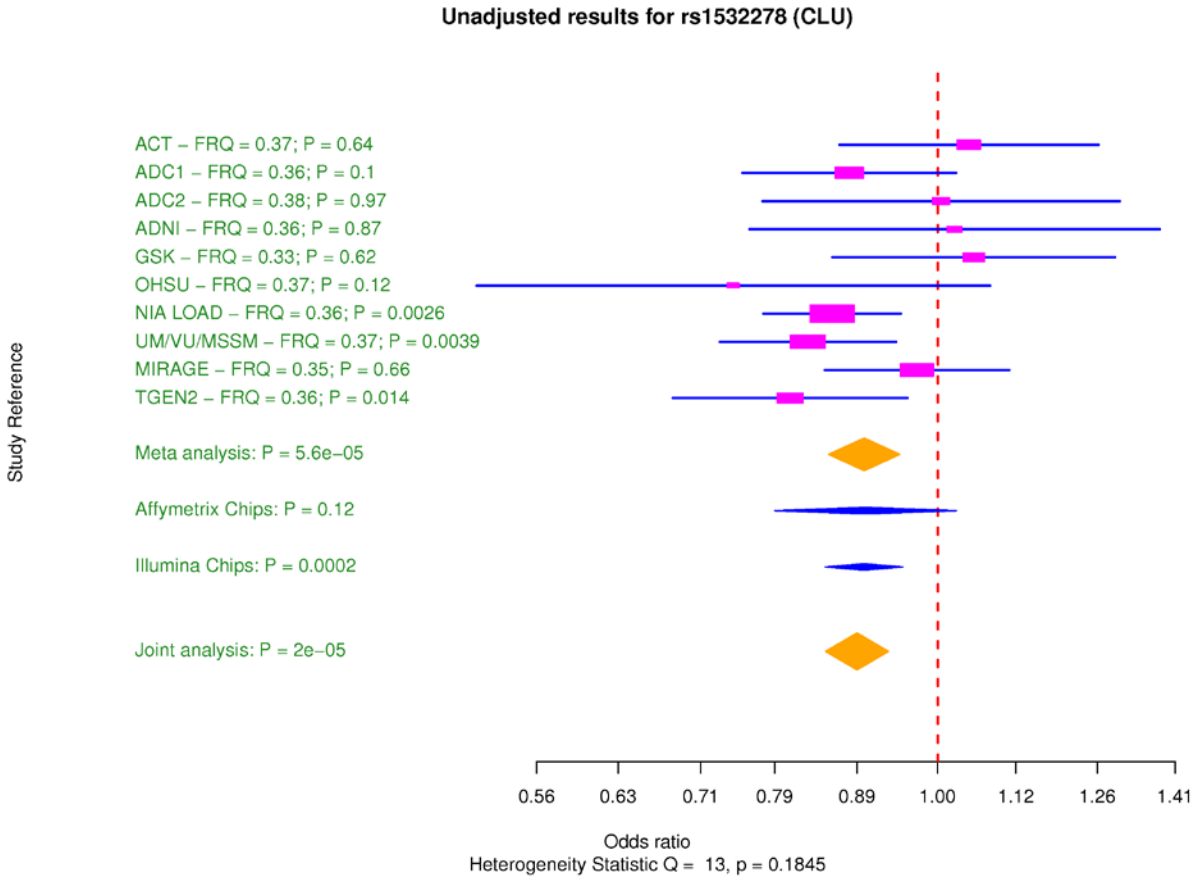
Nature Genetics: doi.10.1038/ng.801

D



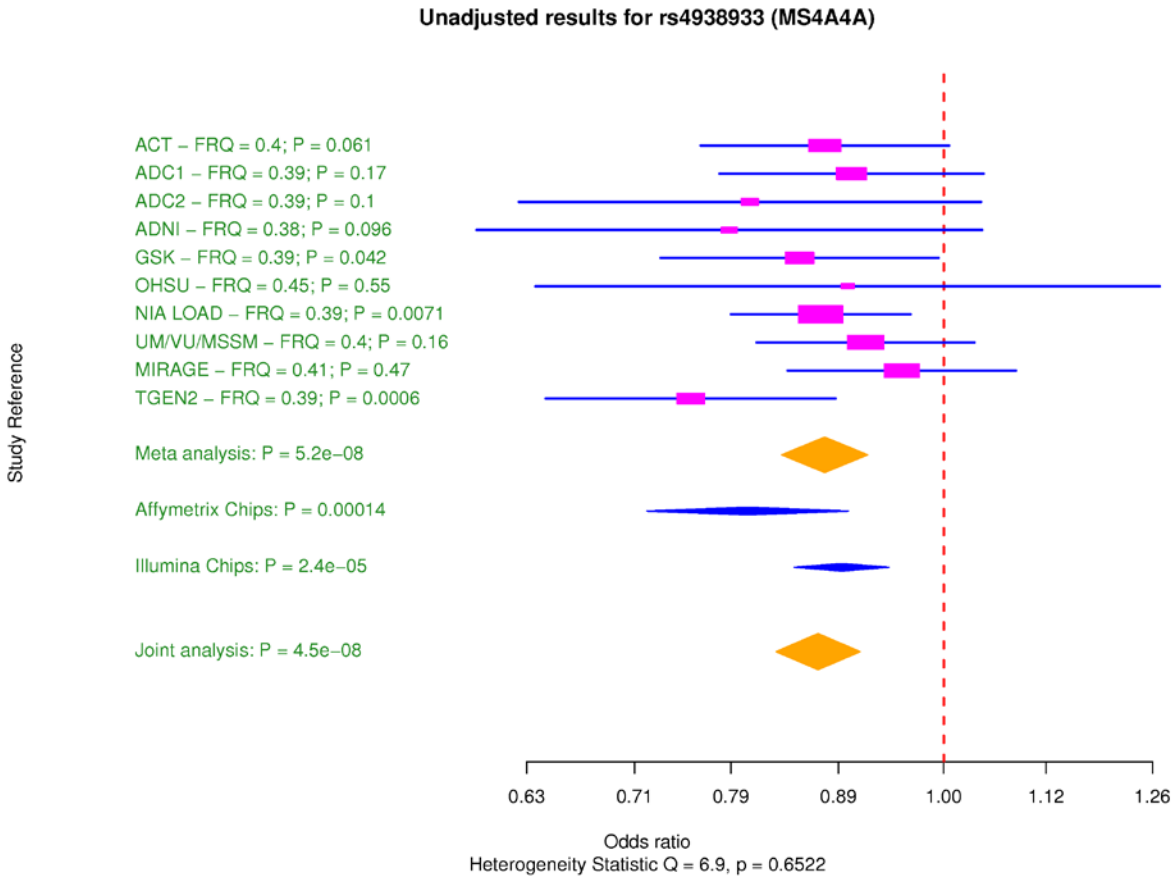
Nature Genetics: doi.10.1038/ng.801

E



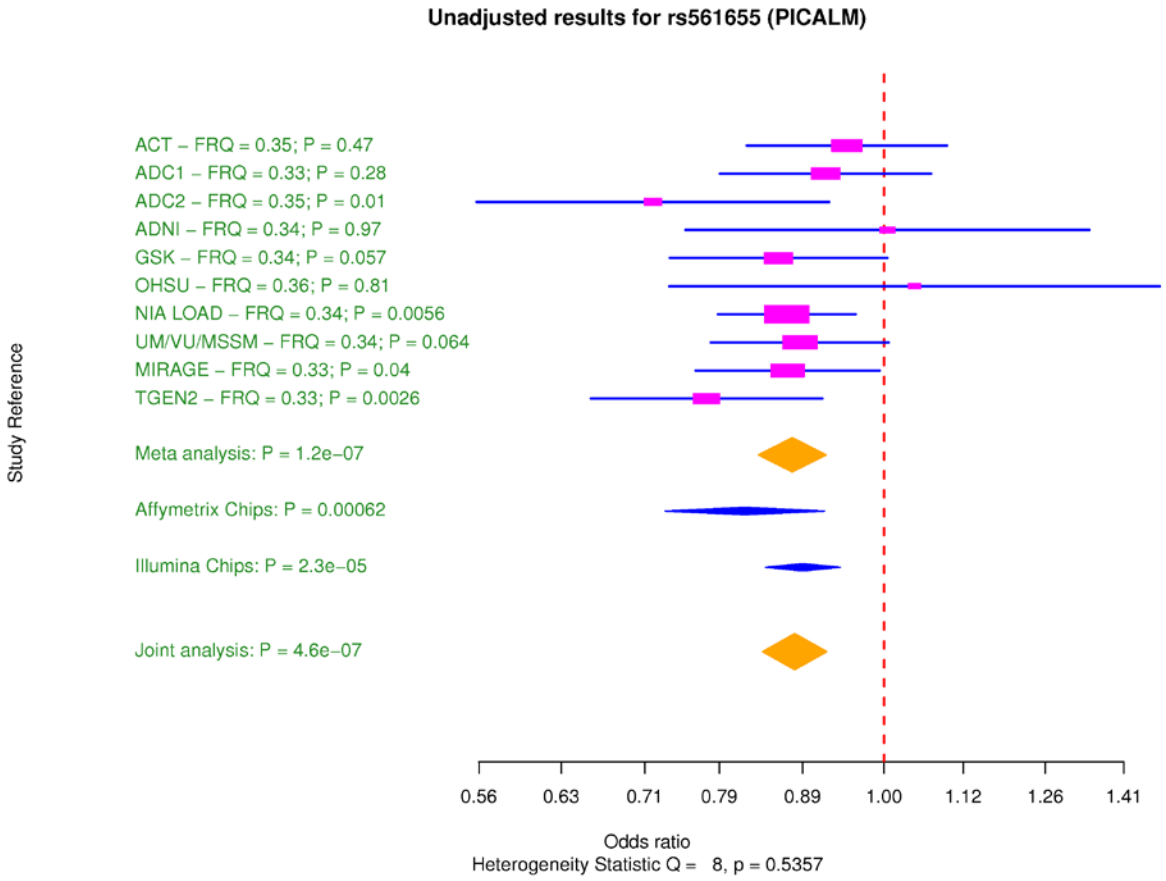
Nature Genetics: doi.10.1038/ng.801

F



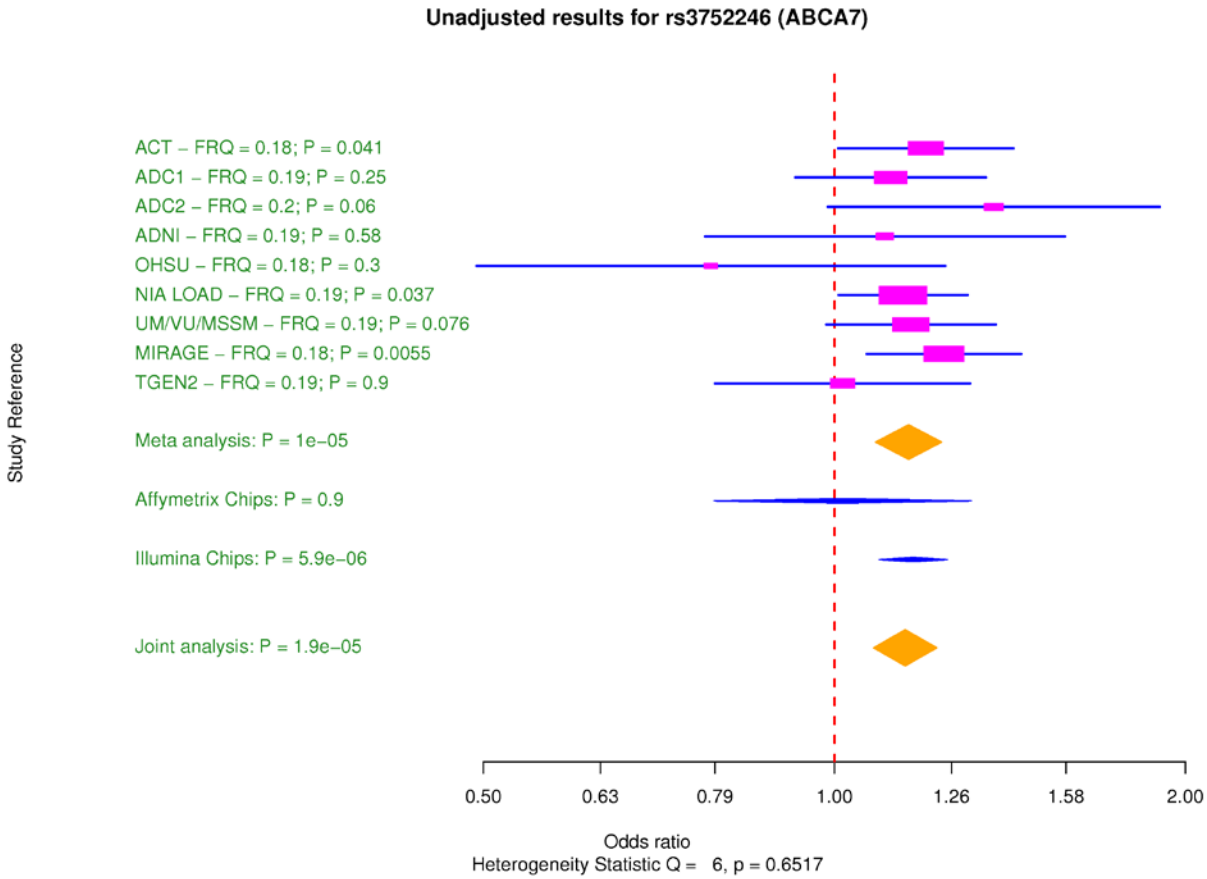
Nature Genetics: doi.10.1038/ng.801

G



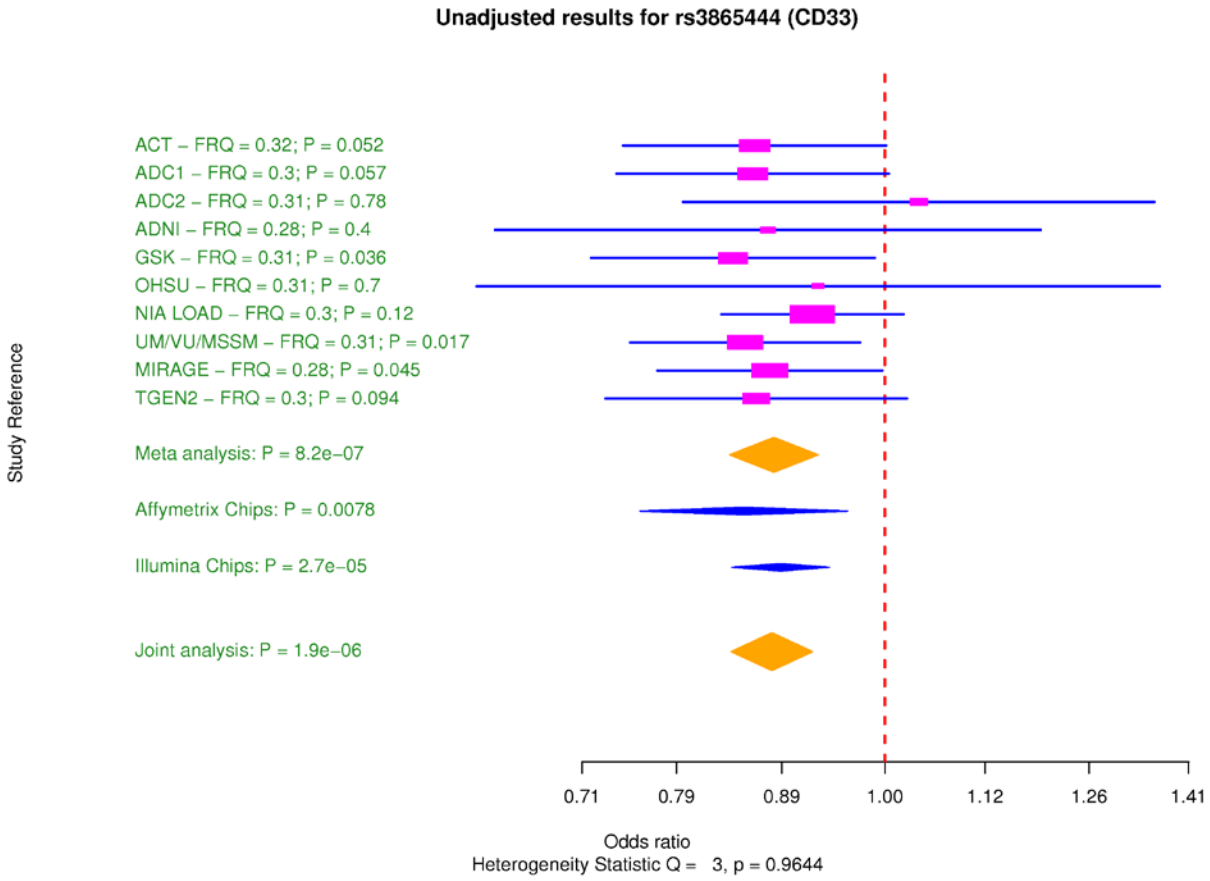
Nature Genetics: doi.10.1038/ng.801

H



Nature Genetics: doi.10.1038/ng.801

J



Nature Genetics: doi.10.1038/ng.801

Supplementary Note Dataset Descriptions

Subjects

The Discovery (Stage 1) dataset comprises subjects from the Adult Changes in Thought (ACT)/ Electronic Medical Records and Genetics (eMERGE) study, the National Institute on Aging (NIA) Alzheimer Disease Centers (ADCs), the Alzheimer Disease Neuroimaging Initiative (ADNI) Study, the Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimers Disease (GenADA) Study, the University of Miami/Vanderbilt University/Mt. Sinai School of Medicine (UM/VU/MSSM), the MIRAGE Study, Oregon Health and Science University (OHSU), the NIA-LOAD Study, and the Translational Genomics Research Institute series 2 (TGEN2) dataset. The Replication (Stage 2) dataset includes subjects from the Mayo Clinic, the Rush University Religious Orders Study/Memory and Aging Project (ROSMAP), the University of Pittsburgh (UP), and Washington University (WU). Detailed descriptions of the ascertainment and evaluation of subjects in the ADC, ADNI, UM/VU/MSSM, MIRAGE, and NIA-LOAD cohorts have been provided elsewhere[1]; brief descriptions included here note any differences between data used in this study and data used in the previously published ADGC study[1]. As noted in the main text, we restricted analyses to individuals of European ancestry because there were an insufficient number of subjects from other ethnic groups to obtain meaningful results. All data from members of other ethnic groups examined in our prior publication were not evaluated here. We describe here in more detail the novel discovery and replication cohorts including those from the ACT/eMERGE Study, the GenADA Study, TGEN2, the Mayo Clinic, ROS/MAP, UP, and WU. All subjects were recruited under protocols approved by the appropriate Institutional Review Boards.

Nature Genetics: doi.10.1038/ng.801

Discovery Dataset (Stage 1)

The NIA ADC Samples (ADC): The NIA ADC cohort included subjects ascertained and evaluated by the clinical and neuropathology cores of the 29 NIA-funded ADCs. Data collection is coordinated by the National Alzheimer's Coordinating Center (NACC). NACC coordinates collection of phenotype data from the 29 ADCs, cleans all data, coordinates implementation of definitions of AD cases and controls, and coordinates collection of samples. The ADC cohort consists of 2,288 autopsy-confirmed and 913 clinically-confirmed AD cases, and 519 cognitively normal elders (CNEs) with complete neuropathology data who were older than 60 years at age of death, and 744 living CNEs evaluated using the Uniform dataset (UDS) protocol[2,3] who were documented to not have mild cognitive impairment (MCI) and were between 60 and 100 years of age at assessment.

Based on the data collected by NACC, the ADGC Neuropathology Core Leaders Subcommittee derived inclusion and exclusion criteria for AD and control samples. All autopsied subjects were age $\geq$ 60 years at death. AD cases were demented according to DSM-IV criteria or Clinical Dementia Rating (CDR) $\geq$ 1.

Neuropathologic stratification of cases followed NIA/Reagan criteria explicitly, or used a similar approach when NIA/Reagan criteria[4] were coded as not done, missing, or unknown. Cases were intermediate or high likelihood by NIA/Reagan criteria with moderate to frequent amyloid plaques[5]and neurofibrillary tangle (NFT) Braak stage of III-VI[6,7]. Persons with Down's syndrome, non-AD tauopathies and synucleinopathies were excluded. All autopsied controls had a clinical evaluation within two years of death. Controls did not meet DSM-IV criteria for dementia, did not have a diagnosis of mild cognitive impairment (MCI), and had a CDR of 0, if

Nature Genetics: doi.10.1038/ng.801

performed. Controls were did not meet or were low-likelihood AD by NIA/Reagan criteria, had sparse or no amyloid plaques, and a Braak NFT stage of 0 – II.

ADCs sent frozen tissue from autopsied subjects and DNA samples from some autopsied subjects and from living subjects to the ADCs to the National Cell Repository for Alzheimer's Disease (NCRAD). DNA was prepared by NCRAD for genotyping and sent to the genotyping site at Children's Hospital of Philadelphia. ADC samples were genotyped and analyzed in separate batches.

While most neuropathologically- and clinically-characterized cases and CNEs were included in two waves included in the Discovery dataset (ADC1 and ADC2), a third wave of clinically-identified living cases and CNEs (ADC3) were incorporated into the replication dataset, and are described in more detail below. ADC1 and ADC2 contributed 2,304 AD cases (1,761 autopsy-confirmed; 543 clinically-confirmed) and 675 CNEs (515 autopsy-confirmed; 160 clinically-confirmed), of which 1,595 autopsied-confirmed AD cases and 132 CNEs were analyzed in our previous study[1].

Oregon Health and Science University (OHSU): The OHSU dataset includes 132 autopsy-confirmed AD cases and 153 deceased controls that were evaluated for dementia within 12 months prior to death (age at death > 65 years), which are a subset of the 193 cases and 451 controls examined in our previous study[1] meeting more stringent QC criteria in this study. Subjects were recruited from aging research cohorts at 10 NIA-funded ADCs, and did not overlap other samples assembled by the ADGC. A more extensive description of control samples can be found elsewhere[8].

The ADNI Study (ADNI): ADNI is a longitudinal, multi-site observational study including AD, mild cognitive impairment (MCI), and elderly individuals with normal cognition assessing

Nature Genetics: doi.10.1038/ng.801

clinical and cognitive measures, MRI and PET scans (FDG and 11C PIB) and blood and CNS biomarkers. For this study, ADNI contributed data on 268 AD cases with MRI confirmation of AD diagnosis and 173 healthy controls with AD-free status confirmed as of most recent follow-up. AD subjects were between the ages of 55–90, had an MMSE score of 20–26 inclusive, met NINCDS/ADRDA criteria for probable AD[9], and had an MRI consistent with the diagnosis of AD. Control subjects had MMSE scores between 28 and 30 and a Clinical Dementia Rating of 0 without symptoms of depression, MCI or other dementia and no current use of psychoactive medications. According to the ADNI protocol, subjects were ascertained at regular intervals over 3 years, but for the purpose of our analysis we only used the final ascertainment status to classify case-control status. Additional details of the study design are available elsewhere[1,10,11].

The MIRAGE Study (MIRAGE): The MIRAGE study is a family-based genetic epidemiological study of AD that enrolled AD cases and unaffected sibling controls at 17 clinical centers in the United States, Canada, Germany, and Greece (details elsewhere[12]), and contributed 1,262 subjects (509 AD cases and 753 CNEs), a subset of the 559 cases and 788 controls that were incorporated into our prior study[1] which met more stringent QC criteria for this study. Briefly, families were ascertained through a proband meeting the NINCDS-ADRDA criteria for definite or probable AD. Unaffected sibling controls were verified as cognitively healthy based on a Modified Telephone Interview of Cognitive Status score ≥ 86[13].

The NIA LOAD Family Study (NIA-LOAD): The NIA LOAD Family Study[14] recruited families with two or more affected siblings with LOAD and unrelated, CNEs similar in age and ethnic background. A total of 1,819 cases and 1,969 CNEs from 1,802 families were recruited through the NIA-LOAD study, NCRAD, and the University of Kentucky and included for analysis, of which a subset of 985 cases and 881 controls were used in the previous study[1]. One

Nature Genetics: doi.10.1038/ng.801

case per family was selected after determining the individual with the strictest diagnosis (definite > probable > possible LOAD). If there were multiple individuals with the strictest diagnosis, then the individual with the earliest age of onset was selected. The controls included only those samples that were neurologically evaluated to be normal and were not related to a study participant.

University of Miami/Vanderbilt University/Mt. Sinai School of Medicine (UM/VU/MSSM): The UM/VU/MSSM dataset contains 1,186 cases and 1,135 CNEs (new and previously published)[15-18] ascertained at the University of Miami, Vanderbilt University and Mt. Sinai School of Medicine, including 409 autopsy-confirmed cases and 136 controls, primarily from the Mt. Sinai School of Medicine[19]. An additional 16 cases were included and 34 controls excluded from the data analyzed in the prior study[1]. Each affected individual met NINCDS-ADRDA criteria for probably or definite AD with age at onset greater than 60 years as determined from specific probe questions within the clinical history provided by a reliable family informant or from documentation of significant cognitive impairment in the medical record. Cognitively healthy controls were unrelated individuals from the same catchment areas and frequency matched by age and gender, and had a documented MMSE or 3MS score in the normal range. Cases and controls had similar demographics: both had ages-at-onset/ages-at-exam of 74 (± 8 standard deviations), and cases were 63% female, and controls were 61% female.

The ACT/eMERGE Studies (ACT): The ACT cohort is an urban and suburban elderly population from a stable HMO that includes 2,581 cognitively intact subjects age ≥ 65 who were enrolled between 1994 and 1998[20,21]. An additional 811 subjects were enrolled in 2000-2002 using the same methods except oversampling clinics with more minorities. More recently, a Continuous Enrollment strategy was initiated in which new subjects are contacted, screened and

Nature Genetics: doi.10.1038/ng.801

enrolled to keep 2000 active at-risk person-years accruing in each calendar year. This resulted in an enrollment of 4,146 participants as of May 2009. All clinical data are reviewed at a consensus conference. Dementia onset is assigned half way between the prior biennial and the exam that diagnosed dementia. Enrollment for eMERGE Study began in 2007. A waiver of consent was obtained from the IRB to enroll deceased ACT participants. In total, ACT/eMERGE contributed data on 566 individuals with probable or possible AD (70 with autopsy-confirmation) and on 1,696 CNEs (155 with autopsy-confirmation) who were included in analyses.

The GenADA Study: Data from the GenADA cohort that were analyzed included 669 AD cases and 713 CNEs ascertained from nine memory referral clinics in Canada between 2002 and 2005. Patients and CNEs were of Caucasian ancestry from Northern Europe. All patients with AD satisfied NINCDS-ADRDA and DSM-IV criteria for probable AD with Global Deterioration Scale scores of 3-7. CNEs had MMSE test scores higher than 25 (mean 29.2 ± 1.1), a Mattis Dementia Rating Scale score of ≥ 136, a Clock Test without error, and no impairments on seven instrumental activities of daily living questions from the Duke Older American Resources and Services Procedures test. Data were collected under an academic-industrial grant from Glaxo-Smith-Kline, Canada by Principal Investigator P. St George-Hyslop. Detailed characteristics of this cohort have been described previously[22].

The TGEN2 Study: Among the TGEN2 data analyzed were 864 clinically- and neuropathologically-characterized brain donors, and 493 CNEs without dementia or significant AD pathology. Of these cases and CNEs, 667 were genotyped as a part of the TGEN1 series[23]. Samples were obtained from twenty-one different National Institute on Aging-supported AD Center brain banks and from the Miami Brain Bank as previously described[23-25]. Additional

Nature Genetics: doi.10.1038/ng.801

individual samples from other brain banks in the United States, United Kingdom, and the Netherlands were also obtained in the same manner. The criteria for inclusion were as follows: self-defined ethnicity of European descent, neuropathologically confirmed AD or neuropathology present at levels consistent with status as a control, and age of death greater than 65. Autopsy diagnosis was performed by board certified neuropathologists and was based on the presence or absence of the characterization of probable or possible AD. Where it was possible, Braak and Braak staging and/or CERAD classification were employed. Samples derived from subjects with a clinical history of stroke, cerebrovascular disease, comorbidity with any other known neurological disease, or with the neuropathological finding of Lewy bodies were excluded.

Replication Dataset (Stage 2)

The ADC3 dataset became available after analyses of the discovery dataset were completed. The ADC3 dataset contains 897 clinically-identified living cases (527 with autopsy-confirmation) and 588 CNEs (4 with autopsy-confirmation) who were genotyped between July and August 2010, and were ascertained similarly to ADC1 and ADC2 as previously described. No ADC3 data were examined in the previous publication of the ADGC[1].

Mayo Clinic: All 728 cases and 1,173 controls consisted of Caucasian subjects from the United States ascertained at the Mayo Clinic. All subjects were diagnosed by a neurologist at the Mayo Clinic in Jacksonville, Florida or Rochester, Minnesota. The neurologist confirmed a Clinical Dementia Rating score of 0 for all controls; cases had diagnoses of possible or probable AD made according to NINCDS-ADRDA criteria[9]. Autopsy-confirmed samples (221 cases, 216 CNEs) came from the brain bank at the Mayo Clinic in Jacksonville, FL and were evaluated by a single neuropathologist. In clinically-identified cases, the diagnosis of definite AD was made

Nature Genetics: doi.10.1038/ng.801

according to NINCDS-ADRDA criteria[9]. All AD brains analyzed in the study had a Braak score of 4.0 or greater. Brains employed as controls had a Braak score of 2.5 or lower but often had brain pathology unrelated to AD and pathological diagnoses that included vascular dementia, frontotemporal dementia, dementia with Lewy bodies, multi-system atrophy, amyotrophic lateral sclerosis, and progressive supranuclear palsy.

The ROS/MAP Studies: ROS/MAP are two community-based cohort studies. The ROS has been on-going since 1993, with a rolling admission. Through July of 2010, 1,139 older nuns, priests, and brothers from across the United States initially free of dementia who agreed to annual clinical evaluation and brain donation at the time of death completed their baseline evaluation. The MAP has been on-going since 1997, also with a rolling admission. Through July of 2010, 1,356 older persons from across northeastern Illinois initially free of dementia who agreed to annual clinical evaluation and organ donation at the time of death completed their baseline evaluation. Details of the clinical and neuropathologic evaluations have been previously reported[26-29]. A total of 1,072 persons passed genotyping QC. Of these, 296 met clinical criteria for AD at the time of their last clinical evaluation or time of death and met neuropathologic criteria for AD for those on whom neuropathologic data were available, and 776 were without dementia or MCI at the time of their last clinical evaluation or time of death and did not meet neuropathologic criteria for AD for those on whom neuropathologic data were available.

University of Pittsburgh (UP): The University of Pittsburgh dataset contains 1,271 Caucasian AD cases (of which 277 were autopsy-confirmed) recruited by the University of Pittsburgh Alzheimer's Disease Research Center, and 841 Caucasian, CNEs ages 60 and older (2 were

Nature Genetics: doi.10.1038/ng.801

autopsy-confirmed). All AD cases met NINCDS/ADRDA criteria for probable or definite AD. Additional details of the cohort used for GWAS have been previously published[30].

Washington University (WU): A European American LOAD case-control dataset consisting of 339 cases and 187 healthy elderly controls was used in analyses for this study. Participants were recruited as part of a longitudinal study of healthy aging and dementia. Diagnosis of dementia etiology was made in accordance with standard criteria and methods[3]. Severity of dementia was assessed using the Clinical Dementia Rating scale.

Functional Implications of LOAD Susceptibility Genes

The work presented here and in previous studies demonstrates there are 10 known LOAD susceptibility genes (*APOE, CR1, CLU, PICALM, BIN1, EPHA1, MS4A, CD33, CD2AP*, and *ABCA7*). The functional information available for these genes and their products suggests a number of avenues for future research.

ABCA7 has demonstrated expression in the brain in murine models[31], and appears to demonstrate variable expression profiles in differentiated monocytes in response to the addition of low density lipoproteins[32], performing a role in lipid regulation by mediating the cellular efflux of phospholipids to apolipoproteins[31]. Lipid levels are known to be associated with Alzheimer disease[33], supporting a lipid-mediated role for *ABCA7* in LOAD risk.

MS4A4A encodes Membrane-Spanning 4-domains subfamily A member 4A and is a member of the membrane-spanning 4A gene family, the members of which are localized to a cluster on 11q12. Little is known about the function of this gene, which was first cloned in 2001[34,35] along with other members of the *MS4A* family due to homology with *CD20* (*MS4A1*). *CD20* is expressed only in B lymphocytes[36,37], and plays regulatory roles in B lymphocyte signal transduction and cell cycle transitions[38]. *MS4A4A* may be involved in signal transduction as a

Nature Genetics: doi.10.1038/ng.801

component of a multimeric receptor complex[34], and has been shown to be expressed in brain[35]. CD33 encodes myeloid cell surface antigen and is known to involve in apoptosis during immune response[39].

The gene *CD2AP* encodes the CD2-associated protein, which is instrumental in actin cytoskeleton regulation[40], and is known to bind to the cytoplasmic domain of nephrin[41], a transmembrane protein that is part of the glomerular slit diaphragm[42]. It has been directly implicated in focal segmental glomerulosclerosis[43], and may contribute to genetic susceptibility to multiple diseases of kidney function[44-46]. This gene may influence neurodegenerative diseases, and LOAD in particular, through its role in endocytosis, which is necessary for synaptic vesicle reuptake during synaptic transmission and following damage to the neuron, when delivery of lipids is essential to the survival and repair of neuronal injury[47]. Defects in endosomal trafficking have been tied to a number of neurodegenerative diseases, and aberrant endosomal function is one of the initial pathological features of Alzheimer disease, preceding even β-amyloid plaque deposition and the appearance of neurofibrillary tangles[48].

EPHA1, which encodes ephrin type-A receptor 1, is the first identified member of the ephrin (EPH) receptor subfamily of the protein-tyrosine kinase family and is expressed primarily in epithelial tissues, and is implicated in tumorigenicity and cell spreading[49]. Studies have suggested that EPH and EPH-related receptors may be involved in nervous system development (reviewed in Palmer and Klein[50]). *EPHA1* is expressed in brain, and has been shown to be involved in brain tumor development, where its expression is down-regulated in glioblastoma biopsied tissue relative to normal brain tissue[51]. While evidence of a role for *EPHA1* in neurodegenerative processes has not yet been observed, products of other Ephrin genes have demonstrated connections to neurodegeneration. Ephrin-A2, an EphA1 ligand, forms a complex

Nature Genetics: doi.10.1038/ng.801

with the *ADAM10* metalloprotease, a member of a disintegrin and metalloprotease family which proteolyses of the amyloid precursor protein[52]; previous studies examining variants in *ADAM10* have observed statistically significant associations with LOAD[53], but in a recent study, this association did not replicate[54].

Nature Genetics: doi.10.1038/ng.801

References

1. Jun, G. et al. Meta-analysis Confirms CR1, CLU, and PICALM as Alzheimer Disease Risk Loci and Reveals Interactions With APOE Genotypes. *Arch Neurol* (2010).
2. Beekly, D.L. et al. The National Alzheimer's Coordinating Center (NACC) database: the Uniform Data Set. *Alzheimer Dis Assoc Disord* 21, 249-58 (2007).
3. Morris, J.C. et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Dis Assoc Disord* 20, 210-6 (2006).
4. Hyman, B.T. & Trojanowski, J.Q. Consensus recommendations for the postmortem diagnosis of Alzheimer disease from the National Institute on Aging and the Reagan Institute Working Group on diagnostic criteria for the neuropathological assessment of Alzheimer disease. *J Neuropathol Exp Neurol* 56, 1095-7 (1997).
5. Mirra, S.S., Hart, M.N. & Terry, R.D. Making the diagnosis of Alzheimer's disease. A primer for practicing pathologists. *Arch Pathol Lab Med* 117, 132-44 (1993).
6. Nagy, Z. et al. Assessment of the pathological stages of Alzheimer's disease in thin paraffin sections: a comparative study. *Dement Geriatr Cogn Disord* 9, 140-4 (1998).
7. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* 82, 239-59 (1991).
8. Kramer, P.L. et al. Alzheimer disease pathology in cognitively healthy elderly: A genome-wide study. *Neurobiol Aging* (2010).
9. McKhann, G. et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34, 939-44 (1984).
10. Petersen, R.C. et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* 74, 201-9 (2010).
11. Saykin, A.J. et al. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimers Dement* 6, 265-73 (2010).
12. Green, R.C. et al. Risk of dementia among white and African American relatives of patients with Alzheimer disease. *Jama* 287, 329-36 (2002).
13. Roccaforte, W.H., Burke, W.J., Bayer, B.L. & Wengel, S.P. Validation of a telephone version of the mini-mental state examination. *J Am Geriatr Soc* 40, 697-702 (1992).
14. Lee, J.H., Cheng, R., Graff-Radford, N., Foroud, T. & Mayeux, R. Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci. *Arch Neurol* 65, 1518-26 (2008).
15. Scott, W.K. et al. Complete genomic screen in Parkinson disease: evidence for multiple genes. *Jama* 286, 2239-44 (2001).
16. Beecham, G.W. et al. Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. *Am J Hum Genet* 84, 35-43 (2009).
17. Edwards, T.L. et al. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. *Ann Hum Genet* 74, 97-109 (2010).

Nature Genetics: doi.10.1038/ng.801

18. Naj, A.C. et al. Dementia Revealed: Novel Chromosome 6 Locus for Late-Onset Alzheimer Disease Provides Genetic Evidence for Folate-Pathway Abnormalities. *PLoS Genet* 6(2010).
19. Haroutunian, V. et al. Regional distribution of neuritic plaques in the nondemented elderly and subjects with very mild Alzheimer disease. *Arch Neurol* 55, 1185-91 (1998).
20. Kukull, W.A. et al. Dementia and Alzheimer disease incidence: a prospective cohort study. *Arch Neurol* 59, 1737-46 (2002).
21. Larson, E.B. et al. Exercise is associated with reduced risk for incident dementia among persons 65 years of age and older. *Ann Intern Med* 144, 73-81 (2006).
22. Li, H. et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 65, 45-53 (2008).
23. Reiman, E.M. et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* 54, 713-20 (2007).
24. Caselli, R.J. et al. Cognitive domain decline in healthy apolipoprotein E epsilon4 homozygotes before the diagnosis of mild cognitive impairment. *Arch Neurol* 64, 1306-11 (2007).
25. Webster, J.A. et al. Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet* 84, 445-58 (2009).
26. Bennett, D.A. et al. Natural history of mild cognitive impairment in older persons. *Neurology* 59, 198-205 (2002).
27. Bennett, D.A., Schneider, J.A., Bienias, J.L., Evans, D.A. & Wilson, R.S. Mild cognitive impairment is related to Alzheimer disease pathology and cerebral infarctions. *Neurology* 64, 834-41 (2005).
28. Bennett, D.A. et al. The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. *Neuroepidemiology* 25, 163-75 (2005).
29. Schneider, J.A., Arvanitakis, Z., Bang, W. & Bennett, D.A. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology* 69, 2197-204 (2007).
30. Kamboh, M.I. et al. Association of CLU and PICALM variants with Alzheimer's disease. *Neurobiol Aging* (2010).
31. Wang, N. et al. ATP-binding cassette transporter A7 (ABCA7) binds apolipoprotein A-I and mediates cellular phospholipid but not cholesterol efflux. *J Biol Chem* 278, 42906-12 (2003).
32. Kaminski, W.E. et al. Identification of a novel human sterol-sensitive ATP-binding cassette transporter (ABCA7). *Biochem Biophys Res Commun* 273, 532-8 (2000).
33. Reitz, C., Tang, M.X., Luchsinger, J. & Mayeux, R. Relation of plasma lipids to Alzheimer disease and vascular dementia. *Arch Neurol* 61, 705-14 (2004).
34. Liang, Y., Buckley, T.R., Tu, L., Langdon, S.D. & Tedder, T.F. Structural organization of the human MS4A gene cluster on Chromosome 11q12. *Immunogenetics* 53, 357-68 (2001).
35. Ishibashi, K., Suzuki, M., Sasaki, S. & Imai, M. Identification of a new multigene four-transmembrane family (MS4A) related to CD20, HTm4 and beta subunit of the high-affinity IgE receptor. *Gene* 264, 87-93 (2001).
36. Stashenko, P., Nadler, L.M., Hardy, R. & Schlossman, S.F. Characterization of a human B lymphocyte-specific antigen. *J Immunol* 125, 1678-85 (1980).

Nature Genetics: doi.10.1038/ng.801

37. Tedder, T.F. et al. Cloning of a complementary DNA encoding a new mouse B lymphocyte differentiation antigen, homologous to the human B1 (CD20) antigen, and localization of the gene to chromosome 19. *J Immunol* 141, 4388-94 (1988).
38. Tedder, T.F. & Engel, P. CD20: a regulator of cell-cycle progression of B lymphocytes. *Immunol Today* 15, 450-4 (1994).
39. Vitale, C. et al. Surface expression and function of p75/AIRM-1 or CD33 in acute myeloid leukemias: engagement of CD33 induces apoptosis of leukemic cells. *Proc Natl Acad Sci U S A* 98, 5764-9 (2001).
40. Bruck, S. et al. Identification of a novel inhibitory actin-capping protein binding motif in CD2-associated protein. *J Biol Chem* 281, 19196-203 (2006).
41. Li, C., Ruotsalainen, V., Tryggvason, K., Shaw, A.S. & Miner, J.H. CD2AP is expressed with nephrin in developing podocytes and is found widely in mature kidney and elsewhere. *Am J Physiol Renal Physiol* 279, F785-92 (2000).
42. Tryggvason, K. Nephrin: role in normal kidney and in disease. *Adv Nephrol Necker Hosp* 31, 221-34 (2001).
43. Lowik, M.M. et al. Focal segmental glomerulosclerosis in a patient homozygous for a CD2AP mutation. *Kidney Int* 72, 1198-203 (2007).
44. Gigante, M. et al. CD2AP mutations are associated with sporadic nephrotic syndrome and focal segmental glomerulosclerosis (FSGS). *Nephrol Dial Transplant* 24, 1858-64 (2009).
45. Lehtonen, S. et al. In vivo interaction of the adapter protein CD2-associated protein with the type 2 polycystic kidney disease protein, polycystin-2. *J Biol Chem* 275, 32888-93 (2000).
46. Wolf, G. & Stahl, R.A. CD2-associated protein and glomerular disease. *Lancet* 362, 1746-8 (2003).
47. Vance, J.E., Campenot, R.B. & Vance, D.E. The synthesis and transport of lipids for axonal growth and nerve regeneration. *Biochim Biophys Acta* 1486, 84-96 (2000).
48. Keating, D.J., Chen, C. & Pritchard, M.A. Alzheimer's disease and endocytic dysfunction: clues from the Down syndrome-related proteins, DSCR1 and ITSN1. *Ageing Res Rev* 5, 388-401 (2006).
49. Yamazaki, T. et al. EphA1 interacts with integrin-linked kinase and regulates cell morphology and motility. *J Cell Sci* 122, 243-55 (2009).
50. Palmer, A. & Klein, R. Multiple roles of ephrins in morphogenesis, neuronal networking, and brain function. *Genes Dev* 17, 1429-50 (2003).
51. Hafner, C. et al. Differential gene expression of Eph receptors and ephrins in benign human tissues and cancers. *Clin Chem* 50, 490-9 (2004).
52. Lammich, S. et al. Constitutive and regulated alpha-secretase cleavage of Alzheimer's amyloid precursor protein by a disintegrin metalloprotease. *Proc Natl Acad Sci U S A* 96, 3922-7 (1999).
53. Kim, M. et al. Potential late-onset Alzheimer's disease-associated mutations in the ADAM10 gene attenuate {alpha}-secretase activity. *Hum Mol Genet* 18, 3987-96 (2009).
54. Cai, G. et al. Evidence against a role for rare ADAM10 mutations in sporadic Alzheimer Disease. *Neurobiol Aging* (2010).

Nature Genetics: doi.10.1038/ng.801