**Supplementary Table 1.** Descriptive statistics for all samples.

| | Stage 1 | | | | Stage 2 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | GERAD 2 | | | | | | | | | | AD-IG | deCODE |
| | GERAD 1 | EADI1 | ADNI | TGEN1 | Total GERAD2 Sample | MRC | ART | Belgium | Bonn | Caerphilly | UCL-PRION | Laser | Greece | Munich | | |
| **Genotyping Platform** | Illumina 610,550 & 300 | Illumina 610 | Illumina 610 | Affymetrix 500K | Sequenom | | | | | | | | | | Illumina 550 & 610 | Illumina 300, CNV370 |
| **AD Cases** | | | | | | | | | | | | | | | | |
| n | 3941§ | 2025§ | 151§ | 571§ | 3262 | 292§ | 628§ | 1078§ | 347§ | 51§ | 92§ | 42§ | 404§ | 328§ | 709§ | 925§ |
| % Female | 62.7 | 66.0 | 47.0 | 52.0 | 64.4 | 63.5 | 61.3 | 66.2 | 79.3 | 0 | 57.1 | 69.0 | 64.6 | 66.8 | 56.1 | 65.6 |
| Age at onset, Mean | 73.2 | 68.3 | 73.5 | N/A | 72.9 | 75.7 | 70.6‡ | 74.9 | 70.3 | N/A | 61.2 | N/A | 69.0‡ | 70.5 | 69.5 | N/A |
| Age at Interview/ascertainment, Mean | 78.6 | 73.7 | 76.6 | 81.0 | 77.7 | 81.1 | 78.4† | 78.6 | 76.2 | N/A | N/A | 79.3 | 76.7 | 73.2 | 72.8 | N/A |
| Age at death, Mean | 80.4* | N/A | N/A | N/A | 81.6 | N/A | 81.6† | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| **Controls** | | | | | | | | | | | | | | | | |
| n | 7848#,††,‡‡ | 5328# | 177# | 332**,†† | 3320 | 45# | 375#,** | 593# | 896** | 0 | 0 | 0 | 147‡‡ | 858# | 971‡‡ | 612** |
| % Female | 65.8 | 61 | 44.6 | 63 | 56.1 | 65.0 | 61.4 | 57.4 | 66.4 | N/A | N/A | N/A | 53.1 | 39.3 | 48 | 60.6 |
| Age at Interview/ascertainment, Mean | 55.6 | 73.8 | 78.0 | 80 | 73.7 | 76.4 | 75.3† | 73.5 | 79.5 | N/A | N/A | N/A | 73.2 | 66.0 | 47.9 | N/A |
| Age at death, Mean | 80.4* | N/A | N/A | N/A | 76.7 | N/A | 76.7† | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |

| | Stage 3 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EADI2 | | | Mayo2 | | | | | | | CHARGE | | | |
| | Finland | Italy | Spain | Jacksonville | Rochester | Autopsy | NCRAD | Norway | Poland | ART | CHS | FHS | Rotterdam | AGES |
| **Genotyping Platform** | Sequenom | | | TaqMan® | | | | | | | Illumina CNV370 | Affymetrix 500+50K Gene Focused Panel | Illumina 550 | Illumina CNV370 |
| **AD Cases** | | | | | | | | | | | | | | |
| n | 563§ | 1460§ | 728§ | 849¶ | 587¶ | 580§ | 702¶ | 345¶ | 479§ | 626§ | 93¶ | 52§ | 171§ | 78¶ |
| % Female | 68.0 | 68.0 | 57.0 | 62.0 | 60.6 | 58.5 | 64.8 | 69.9 | 66.2 | 55.2 | 53 | 81 | 75 | 50 |
| Age at onset, Mean | 71.3 | 73.8 | 72.5 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Age at Interview/ascertainment, Mean | N/A | 76.6 | 75.3 | 77.8 | 80.2 | N/A | 75.2 | 80.2 | 76.7 | 75.8 | 80 | 87 | 84 | 81 |
| Age at death, Mean | N/A | N/A | N/A | N/A | N/A | 81.1 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| **Controls** | | | | | | | | | | | | | | |
| n | 529c | 1262# | 829# | 1303** | 2390** | 355†† | 209** | 553‡‡,# | 182** | 742#,** | 2429** | 2091# | 5700** | 2684# |
| % Female | 58.0 | 55.0 | 62.0 | 57.3 | 53.7 | 42.5 | 61.7 | 59.9 | 76.9 | 49.9 | 62 | 57 | 59 | 58 |
| Age at Interview/ascertainment, Mean | 69.0 | 72.3 | 76.9 | 79.3 | 78.3 | N/A | 78.3 | 75.4 | 73.0 | 76.3 | 75 | 76 | 69 | 76 |
| Age at death, Mean | N/A | N/A | N/A | N/A | N/A | 75.8 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| **Incidence studies** | | | | | | | | | | | | | | |
| Cohort at risk | | | | | | | | | | | 2429 | 806 | 5700 | |
| % Female | | | | | | | | | | | 62 | 60 | 59 | |
| Age at start | | | | | | | | | | | 75 | 82 | 69 | |
| Incident AD cases | | | | | | | | | | | 435§ | 76§ | 462§ | |

* Data only available for a proportion of the sample † Age at interview not available for 438 AD cases and 104 controls. Age at death is provided for these subjects where available. ‡ Age at onset data only available for less than 75% of the sample. § diagnosed according to NINCDS-ADRDA, DSM or CERAD Criteria for probable AD or definite AD. ¶ diagnosed according to NINCDS-ADRDA, or DSM Criteria for possible or probable AD. # Screened for dementia using the MMSE. ** Control screened for dementia using the modified MMSE, TICS-M, Geriatric Mental Schedule, Cognitive Performance Scale, SIDAM or Clinical Dementia Rating Scale. †† Neuropathological conifrmed controls according to CERAD criteria or Braak and Braak Staging. ‡‡ Unscreened population controls.

**Supplementary Table 2**. Results for SNPs with P≤1x10$^{-5}$ in Stage 1

| SNP | Closest Gene | CHR | BP | A1 | A2 | Stage 1 Datasets | | | | | | | | Stage 1 | | Stage 1 Sample Size | | SNP selected for Stage 2? | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GERAD1 | | EADI1 | | ADNI | | TGEN1 | | | | | | | |
| | | | | | | OR | P | OR | P | OR | P | OR | P | OR | P | Cases | Controls | | |
| rs4844579 | *C4BPA* | 1 | 207,377,891 | T | C | 1.12 | 3.6E-03 | 1.17 | 5.8E-04 | 1.28 | 2.0E-01 | 1.18 | 1.6E-01 | 1.14 | 1.9E-06 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as it is located at the *CR1* locus |
| rs6673080 | *C4BPA* | 1 | 207,390,204 | T | C | 1.11 | 4.8E-03 | 1.17 | 6.5E-04 | 1.27 | 2.2E-01 | 1.20 | 1.2E-01 | 1.14 | 2.5E-06 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as it is located at the *CR1* locus |
| rs6540433 | *CR2* | 1 | 207,653,395 | C | A | 1.12 | 1.8E-03 | 1.18 | 8.0E-04 | 1.45 | 1.0E-01 | 1.39 | 2.6E-02 | 1.16 | 5.1E-07 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *CR1* locus |
| rs4310446 | *CR1* | 1 | 207,676,604 | C | T | 1.10 | 1.0E-02 | 1.19 | 2.1E-04 | 1.44 | 7.8E-02 | 1.25 | 9.4E-02 | 1.14 | 1.8E-06 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *CR1* locus |
| rs3818361 | *CR1* | 1 | 207,784,968 | A | G | 1.16 | 3.7E-05 | 1.28 | 8.5E-08 | 1.58 | 2.4E-02 | 1.20 | 1.4E-01 | 1.21 | 3.2E-12 | 6,688 | 13,251 | Yes | SNP is most significant at the *CR1* locus. |
| rs6701713 | *CR1* | 1 | 207,786,289 | A | G | 1.17 | 8.7E-06 | 1.25 | 1.1E-06 | 1.58 | 2.4E-02 | 1.20 | 1.4E-01 | 1.20 | 6.0E-12 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *CR1* locus |
| rs1408077 | *CR1* | 1 | 207,804,141 | A | C | 1.17 | 8.3E-06 | 1.25 | 1.7E-06 | 1.66 | 1.5E-02 | 1.20 | 1.4E-01 | 1.20 | 6.4E-12 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *CR1* locus |
| rs744373 | *BIN1* | 2 | 127,894,615 | G | A | 1.16 | 1.4E-06 | 1.15 | 5.7E-04 | 1.13 | 4.4E-01 | 1.35 | 1.7E-02 | 1.17 | 1.5E-10 | 6,688 | 12,716 | Yes | SNP is most significant at the *BIN1* locus. |
| rs11136000 | *CLU* | 8 | 27,464,519 | T | C | 0.84 | 1.4E-09 | 0.81 | 5.2E-08 | 0.93 | 6.4E-01 | 0.80 | 1.0E-01 | 0.83 | 2.0E-16 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *CLU* locus |
| rs10761558 | *CDK1* | 10 | 62,523,470 | A | G | 1.26 | 1.6E-05 | 1.12 | 4.4E-03 | 1.12 | 5.1E-01 | 1.07 | 5.4E-01 | 1.16 | 1.0E-06 | 6,080 | 7,062 | Yes | |
| rs667897 | *MS4A6A* | 11 | 59,936,979 | G | A | 0.89 | 2.5E-05 | 0.94 | 7.0E-02 | 0.96 | 7.8E-01 | 0.76 | 6.4E-03 | 0.90 | 8.7E-07 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as a proxy SNP rs610932 ($r^2$=0.88) was genotyped in Stage 2 |
| rs610932 | *MS4A6A* | 11 | 59,939,307 | T | G | 0.87 | 1.5E-06 | 0.93 | 4.6E-02 | 0.88 | 4.5E-01 | 0.74 | 2.8E-03 | 0.88 | 1.8E-08 | 6,688 | 13,251 | Yes | |
| rs662196 | *MS4A6A* | 11 | 59,942,757 | C | T | 0.88 | 5.2E-06 | 0.95 | 1.5E-01 | 0.92 | 6.2E-01 | 0.75 | 3.9E-03 | 0.89 | 2.5E-07 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as a proxy SNP rs610932 ($r^2$=0.934) was genotyped in Stage 2 |
| rs583791 | *MS4A6A* | 11 | 59,947,252 | C | T | 0.88 | 5.3E-06 | 0.95 | 1.6E-01 | 0.92 | 6.2E-01 | 0.75 | 3.9E-03 | 0.90 | 7.4E-07 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as a proxy SNP rs610932 ($r^2$=0.934) was genotyped in Stage 2 |
| rs7926344 | *MS4A6A* | 11 | 59,962,166 | A | G | 0.88 | 7.9E-05 | 0.95 | 1.4E-01 | 0.91 | 5.8E-01 | 0.72 | 1.3E-03 | 0.90 | 2.4E-06 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as a proxy SNP rs610932 ($r^2$=0.782) was genotyped in Stage 2 |
| rs670139 | *MS4A4E* | 11 | 59,971,795 | T | G | 1.13 | 8.7E-05 | 1.06 | 1.2E-01 | 1.08 | 6.6E-01 | 1.26 | 2.3E-02 | 1.11 | 1.0E-05 | 6,080 | 11,863 | Yes | SNP was selected as a proxy SNP for rs676309 ($r^2$=1) |
| rs7929589 | *MS4A4E* | 11 | 59,975,078 | T | C | 0.88 | 8.3E-05 | 0.95 | 1.7E-01 | 0.89 | 4.7E-01 | 0.72 | 1.1E-03 | 0.91 | 6.0E-06 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as a proxy SNP rs610932 ($r^2$=0.721) was genotyped in Stage 2 |
| rs676309 | *MS4A4E* | 11 | 60,001,573 | C | T | 1.14 | 6.3E-06 | 1.04 | 2.1E-01 | 1.07 | 6.7E-01 | 1.27 | 2.0E-02 | 1.11 | 2.0E-06 | 6,688 | 13,685 | No | Not possible to design a multiplex assay including this SNP; a proxy SNP rs670139 ($r^2$=1) was genotyped in Stage 2 |
| rs1562990 | *MS4A4A* | 11 | 60,023,087 | C | A | 0.88 | 1.0E-05 | 0.96 | 2.4E-01 | 0.96 | 7.8E-01 | 0.71 | 6.7E-04 | 0.90 | 1.4E-06 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as a proxy SNP rs610932 ($r^2$=0.618) was genotyped in Stage 2 |
| rs677909 | *PICALM* | 11 | 85,757,589 | C | T | 0.88 | 1.6E-05 | 0.88 | 3.0E-03 | 1.02 | 9.0E-01 | 0.91 | 3.5E-01 | 0.88 | 1.4E-07 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *PICALM* locus |
| rs536841 | *PICALM* | 11 | 85,787,824 | C | T | 0.86 | 3.5E-06 | 0.88 | 3.7E-03 | 0.98 | 9.3E-01 | 0.90 | 3.4E-01 | 0.87 | 3.6E-08 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as it is located at the *PICALM* locus |
| rs541458 | *PICALM* | 11 | 85,788,351 | C | T | 0.87 | 2.3E-06 | 0.88 | 3.5E-03 | 1.00 | 9.9E-01 | 0.90 | 3.4E-01 | 0.87 | 1.9E-08 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *PICALM* locus |
| rs1237999 | *PICALM* | 11 | 85,815,030 | G | A | 0.87 | 1.1E-06 | 0.94 | 1.1E-01 | 1.03 | 8.4E-01 | 0.91 | 3.6E-01 | 0.90 | 1.7E-06 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *PICALM* locus |
| rs543293 | *PICALM* | 11 | 85,820,077 | A | G | 0.86 | 6.9E-07 | 0.93 | 9.1E-02 | 0.92 | 6.4E-01 | 0.88 | 2.2E-01 | 0.89 | 2.7E-07 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *PICALM* locus |
| rs659023 | *PICALM* | 11 | 85,824,859 | A | G | 0.86 | 3.6E-06 | 0.93 | 7.1E-02 | 1.07 | 7.0E-01 | 0.93 | 4.5E-01 | 0.89 | 3.1E-06 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as it is located at the *PICALM* locus |
| rs7941541 | *PICALM* | 11 | 85,858,538 | G | A | 0.84 | 2.1E-07 | 0.91 | 2.4E-02 | 0.90 | 5.4E-01 | 0.84 | 1.0E-01 | 0.87 | 9.0E-09 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as it is located at the *PICALM* locus |
| rs3851179 | *PICALM* | 11 | 85,868,640 | T | C | 0.85 | 1.9E-08 | 0.92 | 3.1E-02 | 0.83 | 2.8E-01 | 0.82 | 6.3E-02 | 0.87 | 7.2E-10 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *PICALM* locus |
| rs10501927 | *CNTN5* | 11 | 99,757,729 | G | T | 1.16 | 2.3E-05 | 1.10 | 1.2E-02 | 1.38 | 1.0E-01 | 1.00 | 1.0E+00 | 1.14 | 8.0E-07 | 6,688 | 13,251 | Yes | |
| rs3809278 | *CUX2* | 12 | 111,725,185 | A | C | 1.12 | 5.4E-03 | 1.23 | 1.0E-04 | 1.34 | 2.3E-01 | 1.23 | 1.5E-01 | 1.17 | 1.1E-06 | 6,688 | 13,685 | Yes | |
| rs739565 | *C16orf88* | 16 | 19,716,505 | A | G | 1.07 | 3.1E-02 | 1.19 | 9.9E-06 | 1.01 | 9.5E-01 | 1.38 | 3.6E-02 | 1.11 | 3.9E-06 | 6,688 | 13,251 | Yes | |
| rs1858973 | *IQCK* | 16 | 19,743,649 | C | T | 0.91 | 1.5E-02 | 0.80 | 1.6E-05 | 0.92 | 6.8E-01 | 0.73 | 1.4E-02 | 0.86 | 7.2E-07 | 6,688 | 13,251 | Yes | |
| rs4782279 | *IQCK* | 16 | 19,759,007 | C | A | 0.92 | 2.1E-02 | 0.80 | 5.4E-06 | 1.03 | 9.0E-01 | 0.73 | 1.2E-02 | 0.87 | 7.7E-07 | 6,688 | 13,251 | Yes | |
| rs9931167 | *IQCK* | 16 | 19,792,598 | T | C | 0.93 | 8.0E-02 | 0.80 | 2.3E-05 | 0.92 | 7.0E-01 | 0.71 | 5.8E-03 | 0.87 | 6.0E-06 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as a proxy SNP rs4782279 ($r^2$=0.945) was genotyped in Stage 2 |
| rs7191155 | *IQCK* | 16 | 19,800,213 | C | T | 0.91 | 1.1E-02 | 0.80 | 2.4E-05 | 0.92 | 6.8E-01 | 0.71 | 5.9E-03 | 0.86 | 3.6E-07 | 6,688 | 13,251 | Yes | |
| rs3764650 | *ABCA7* | 19 | 1,046,520 | G | T | 1.24 | 4.1E-05 | 1.21 | 4.0E-03 | 1.01 | 9.7E-01 | N/A | N/A | 1.22 | 2.6E-07 | 5,509 | 11,531 | Yes | |
| rs7255066 | *PVR* | 19 | 45,146,103 | C | T | 0.88 | 7.4E-05 | 0.90 | 3.4E-03 | 0.70 | 2.9E-02 | 1.35 | 1.4E-01 | 0.89 | 5.9E-07 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs2927488 | *CEACAM16* | 19 | 45,231,478 | A | G | 0.91 | 4.7E-03 | 0.82 | 6.5E-06 | 1.00 | 9.9E-01 | 1.14 | 4.6E-01 | 0.88 | 1.8E-06 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs2965101 | *BCL3* | 19 | 45,237,812 | C | T | 0.86 | 4.6E-07 | 0.78 | 1.9E-09 | 1.00 | 9.9E-01 | 1.11 | 4.9E-01 | 0.84 | 2.8E-13 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs2927438 | *BCL3* | 19 | 45,242,107 | A | G | 1.25 | 3.0E-11 | 1.21 | 2.9E-05 | 1.25 | 2.2E-01 | N/A | N/A | 1.23 | 2.3E-15 | 6,117 | 13,353 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs4803750 | *BCL3* | 19 | 45,247,627 | G | A | 0.77 | 2.8E-05 | 0.73 | 5.7E-04 | 0.83 | 6.3E-01 | N/A | N/A | 0.75 | 5.9E-08 | 5,509 | 11,965 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs8103315 | *BCL3* | 19 | 45,254,168 | A | C | 1.24 | 2.9E-07 | 1.28 | 4.1E-06 | 1.63 | 2.8E-02 | N/A | N/A | 1.26 | 8.5E-13 | 5,509 | 11,965 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs1871045 | *BCAM* | 19 | 45,326,768 | T | C | 0.89 | 1.2E-04 | 0.89 | 2.6E-03 | 1.10 | 5.5E-01 | 0.86 | 1.5E-01 | 0.89 | 7.6E-07 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs10402271 | *BCAM* | 19 | 45,329,214 | G | T | 1.36 | 1.5E-26 | 1.36 | 8.5E-15 | 0.94 | 6.9E-01 | 1.22 | 5.5E-02 | 1.34 | 7.2E-39 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs1871047 | *PVRL2* | 19 | 45,351,746 | G | A | 0.85 | 1.3E-08 | 0.82 | 2.0E-06 | 0.94 | 6.9E-01 | 0.84 | 2.6E-01 | 0.84 | 9.8E-14 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs377702 | *PVRL2* | 19 | 45,362,667 | A | G | 1.20 | 8.4E-11 | 1.10 | 1.2E-02 | 1.07 | 6.9E-01 | N/A | N/A | 1.16 | 1.8E-11 | 6,117 | 13,353 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs12610605 | *PVRL2* | 19 | 45,370,838 | A | G | 0.81 | 7.7E-07 | 0.81 | 6.8E-05 | 0.99 | 9.7E-01 | N/A | N/A | 0.81 | 3.3E-10 | 5,509 | 11,965 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs6859 | *PVRL2* | 19 | 45,382,034 | A | G | 1.46 | 7.0E-41 | 1.31 | 3.8E-13 | 1.39 | 3.9E-02 | N/A | N/A | 1.40 | 1.3E-52 | 6,117 | 13,353 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs157580 | *TOMM40* | 19 | 45,395,266 | G | A | 0.63 | 9.6E-54 | 0.62 | 5.1E-33 | 0.46 | 1.7E-05 | N/A | N/A | 0.62 | 3.7E-90 | 6,117 | 13,353 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs2075650 | *TOMM40* | 19 | 45,395,619 | G | A | 2.53 | 1.8E-157 | 3.17 | 1.2E-130 | 3.07 | 6.3E-08 | N/A | N/A | 2.72 | 7.8E-266 | 6,117 | 13,353 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs8106922 | *TOMM40* | 19 | 45,401,666 | G | A | 0.68 | 5.4E-39 | 0.66 | 6.5E-27 | 0.79 | 1.7E-01 | N/A | N/A | 0.67 | 2.3E-65 | 6,117 | 13,353 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs405509 | *APOE* | 19 | 45,408,836 | G | T | 0.70 | 4.9E-37 | 1.38 | 5.3E-18 | 0.72 | 5.4E-02 | N/A | N/A | 0.71 | 6.4E-53 | 6,117 | 13,353 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs439401 | *APOE* | 19 | 45,414,451 | T | C | 0.72 | 2.7E-23 | 0.63 | 5.1E-31 | 0.47 | 1.9E-05 | N/A | N/A | 0.68 | 1.4E-46 | 5,509 | 11,965 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs5167 | *APOC4* | 19 | 45,448,465 | G | T | 1.17 | 3.4E-08 | 1.09 | 1.9E-02 | 1.29 | 1.4E-01 | 1.11 | 5.8E-01 | 1.14 | 2.8E-09 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs3760627 | *CLPTM1* | 19 | 45,457,180 | C | T | 1.15 | 4.0E-07 | 1.12 | 2.8E-03 | 1.25 | 1.7E-01 | 1.13 | 5.0E-01 | 1.14 | 2.0E-09 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs7257916 | *CLPTM1* | 19 | 45,482,884 | C | T | 0.89 | 7.8E-05 | 1.10 | 9.2E-03 | 0.85 | 3.1E-01 | 0.95 | 7.3E-01 | 0.89 | 1.3E-06 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs8111069 | *CLPTM1* | 19 | 45,483,438 | C | A | 1.16 | 6.1E-06 | 1.11 | 4.6E-03 | 1.23 | 2.4E-01 | 1.06 | 6.6E-01 | 1.14 | 6.3E-08 | 6,080 | 12,297 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs1114832 | *LRRC68* | 19 | 45,636,201 | T | C | 1.31 | 2.2E-09 | 1.12 | 4.6E-02 | 1.20 | 4.9E-01 | 2.18 | 4.7E-06 | 1.26 | 2.1E-11 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs10425074 | *LRRC68* | 19 | 45,640,124 | G | A | 1.15 | 1.9E-05 | 1.09 | 1.0E-01 | 0.91 | 6.0E-01 | 1.46 | 1.3E-03 | 1.14 | 8.1E-07 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs1048699 | *LRRC68* | 19 | 45,650,386 | T | C | 1.32 | 1.5E-09 | 1.17 | 9.4E-03 | 1.12 | 6.8E-01 | 2.20 | 6.0E-06 | 1.29 | 8.4E-13 | 6,688 | 13,685 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs2627641 | *EXOC3L2* | 19 | 45,708,758 | G | A | 1.15 | 3.9E-04 | 1.18 | 1.1E-03 | 1.29 | 2.4E-01 | N/A | N/A | 1.17 | 7.6E-07 | 5,509 | 11,965 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |
| rs597668 | *EXOC3L2* | 19 | 45,708,888 | C | T | 1.16 | 5.2E-04 | 1.19 | 1.0E-03 | 1.29 | 2.4E-01 | N/A | N/A | 1.17 | 1.3E-07 | 6,117 | 13,353 | No | SNP was not selected for Stage 2 as it is located at the *APOE* locus |

**Supplementary Table 3**. Stage 2, 3 and meta-analysis results for the 12 SNPs tested in Stage 2

| SNP | Closest Gene | CHR | BP | A1 | A2 | Stage 2 | | Stage 2 Sample Size | | SNP significant at Bonferroni adjusted level (P<0.0042)? | Stage 3 | | Stage 3 Sample Size | | SNP significant at Bonferroni adjusted level (P<0.0167)? | Meta-analysis of all data | | | | | Meta-analysis Sample Size | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | OR | P | Cases | Controls | | OR | P | Cases | Controls | | OR | 95% CI | P | Cochran's Q Test P | $I^2$ | Cases | Controls |
| rs3764650 | *ABCA7* | 19 | 1046520 | G | T | 1.28 | 1.9E-05 | 4896 | 4903 | Yes | 1.22 | 2.9E-07 | 7176 | 17754 | Yes | 1.23 | 1.18-1.30 | 4.5E-17 | 0.80 | 0 | 17683 | 34269 |
| rs744373 | *BIN1* | 2 | 127894615 | G | A | 1.17 | 3.8E-05 | 4896 | 4903 | Yes | - | - | - | - | - | 1.17 | 1.12-1.21 | 2.6E-14 | 0.59 | 0 | 11584 | 17619 |
| rs670139 | *MS4A4E* | 11 | 59971795 | T | G | 1.11 | 1.1E-03 | 4896 | 4903 | Yes | 1.06 | 3.2E-03 | 8224 | 21194 | Yes | 1.09 | 1.06-1.12 | 1.4E-09 | 0.55 | 0 | 19262 | 38024 |
| rs3818361 | *CR1* | 1 | 207784968 | A | G | 1.14 | 1.4E-03 | 4896 | 4903 | Yes | - | - | - | - | - | 1.18 | 1.13-1.24 | 3.7E-14 | 0.26 | 22.3% | 11584 | 18154 |
| rs610932 | *MS4A6A* | 11 | 59939307 | T | G | 0.90 | 1.6E-03 | 4896 | 4903 | Yes | 0.91 | 2.1E-05 | 7312 | 19874 | Yes | 0.90 | 0.87-0.92 | 1.8E-14 | 0.17 | 29.7% | 18990 | 38080 |
| rs7191155 | *IQCK* | 16 | 19800213 | C | T | 0.94 | 1.5E-01 | 4896 | 4903 | No | - | - | - | - | - | - | - | - | - | - | - | - |
| rs4782279 | *IQCK* | 16 | 19759007 | C | A | 0.95 | 1.9E-01 | 4896 | 4903 | No | - | - | - | - | - | - | - | - | - | - | - | - |
| rs1858973 | *IQCK* | 16 | 19743649 | C | T | 0.95 | 2.1E-01 | 4896 | 4903 | No | - | - | - | - | - | - | - | - | - | - | - | - |
| rs739565 | *C16orf88* | 16 | 19716505 | A | G | 1.02 | 5.5E-01 | 4896 | 4903 | No | - | - | - | - | - | - | - | - | - | - | - | - |
| rs10501927 | *CNTN5* | 11 | 99757729 | G | T | 1.02 | 6.9E-01 | 4896 | 4903 | No | - | - | - | - | - | - | - | - | - | - | - | - |
| rs10761558 | *CDK1* | 10 | 62523470 | A | G | 1.00 | 9.0E-01 | 4187 | 3932 | No | - | - | - | - | - | - | - | - | - | - | - | - |
| rs3809278 | *CUX2* | 12 | 111725185 | A | C | 1.00 | 9.9E-01 | 4896 | 4903 | No | - | - | - | - | - | - | - | - | - | - | - | - |

**Supplementary Table 4a**. SNPxSNP interaction P-values

| **SNP** | rs744373 (*BIN1*) | rs11136000 (*CLU*) | rs610932 (*MS4A*) | rs3851179 (*PICALM*) | rs3764650 (*ABCA7*) | rs429358 (*APOE*) |
|---|---|---|---|---|---|---|
| rs3818361 (*CR1*) | 0.6607 | 0.4892 | 0.4572 | 0.1942 | 0.9913 | 0.9367 |
| rs744373 (*BIN1*) | | 0.9979 | 0.9780 | 0.9331 | 0.4509 | 0.1270 |
| rs11136000 (*CLU*) | | | 0.4100 | 0.0613 | 0.5545 | 0.6737 |
| rs610932 (*MS4A*) | | | | 0.2474 | 0.5479 | 0.5909 |
| rs3851179 (*PICALM*) | | | | | 0.4491 | 0.7350 |
| rs3764650 (*ABCA7*) | | | | | | 0.6242 |

NB: Data calculated from GERAD1 sample.

**Supplementary Table 4b**. Logistic regression analyses with and without adjustment for the presence of at least one *APOE* e4 allele. Note that only samples with *APOE* genotypes were included in the analysis

| SNP | Datset | Unadjusted for APOE | | | Adjusted for APOE | | |
|---|---|---|---|---|---|---|---|
| | | OR | 95% CI | P | OR | 95% CI | P |
| rs610932 | GERAD1 | 0.86 | 0.79-0.92 | 8.0E-05 | 0.86 | 0.79-0.93 | 3.3E-04 |
| | GERAD2 | 0.95 | 0.88-1.03 | 1.8E-01 | 0.95 | 0.87-1.03 | 1.8E-01 |
| | EADI1 | 0.93 | 0.86-1.00 | 4.6E-02 | 0.94 | 0.86-1.02 | 1.0E-01 |
| | EADI2 | 0.90 | 0.83-0.97 | 7.7E-03 | 0.89 | 0.82-0.97 | 6.6E-03 |
| rs670139 | GERAD1 | 1.16 | 1.06-1.27 | 9.5E-04 | 1.16 | 1.05-1.27 | 2.3E-03 |
| | GERAD2 | 1.11 | 1.03-1.20 | 7.7E-03 | 1.12 | 1.03-1.22 | 6.8E-03 |
| | EADI1 | 1.06 | 0.99-1.14 | 1.2E-01 | 1.05 | 0.97-1.13 | 2.2E-01 |
| | EADI2 | 1.02 | 0.94-1.11 | 5.9E-01 | 1.02 | 0.94-1.12 | 6.4E-01 |
| rs3764650 | GERAD1 | 1.15 | 0.99-1.33 | 6.6E-02 | 1.11 | 0.95-1.29 | 2.0E-01 |
| | GERAD2 | 1.31 | 1.14-1.50 | 1.1E-04 | 1.30 | 1.12-1.50 | 4.0E-04 |
| | EADI1 | 1.21 | 1.08-1.37 | 1.0E-03 | 1.20 | 1.08-1.32 | 2.7E-03 |
| | EADI2 | 1.31 | 1.15-1.50 | 8.7E-05 | 1.34 | 1.18-1.52 | 7.0E-06 |

**Supplementary Table 5**. Analysis of rs3764650, rs670139 and rs610932 in published expression quantitative trait loci (eQTL) datasets.

| | Stranger et al. 2005 (ref 49) | | | | Gibbs et al. 2010 (ref 50) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Gene | Probe | Lymphoblastoid Cell Line | | Gene | Probe | Cerebellum | | Frontal Cortex | | PONS | | Temporal Cortex | |
| SNP | | | Beta | P-Value | | | Beta | P-Value | Beta | P-Value | Beta | P-Value | Beta | P-Value |
| rs3764650 | ***ABCA7*** | GI_15451836-I | -0.009600 | 0.568 | ***ABCA7*** | ILMN_1660203 | - | - | - | - | - | - | -2.904 | 0.3897 |
| | | GI_15451837-A | 0.037000 | 0.533 | ***ABCA7*** | ILMN_1729894 | - | - | - | - | - | - | 0.997 | 0.7575 |
| rs670139 | ***MS4A6A*** | GI_23238233-A | 0.003962 | 0.9149 | ***MS4A6A*** | ILMN_1721035 | 0.1022 | 0.0789 | 0.02031 | 0.6744 | 0.1073 | 0.1113 | 0.1345 | 0.01461 |
| | ***MS4A2*** | GI_23397640-S | -0.01104 | 0.3985 | ***MS4A6A*** | ILMN_1797731 | 0.09749 | 0.06666 | 0.009008 | 0.836 | 0.09604 | 0.1494 | 0.09399 | 0.06527 |
| | ***MS4A7*** | GI_23110999-S | 0.03997 | 0.6868 | ***MS4A2*** | ILMN_1806721 | 0.005871 | 0.6822 | 0.01349 | 0.3531 | -0.00509 | 0.6967 | -0.02791 | 0.06277 |
| | ***MS4A3*** | GI_23110992-S | 0.002893 | 0.8109 | ***MS4A7*** | ILMN_1743932 | - | - | 0.03082 | 0.4309 | 0.2313 | 0.7232 | 0.06794 | 0.09215 |
| | ***MS4A4A*** | GI_23110994-A | -0.01535 | 0.3443 | ***MS4A3*** | ILMN_1726552 | - | - | - | - | - | - | -0.00115 | 0.9245 |
| | ***PLAC1L*** | GI_42476002-S | -0.0078 | 0.4747 | ***MS4A3*** | ILMN_1751625 | - | - | - | - | - | - | -0.014 | 0.4099 |
| | ***MS4A6E*** | GI_23110998-S | -0.03553 | 0.5326 | ***MS4A4A*** | ILMN_1741712 | - | - | - | - | - | - | -0.00201 | 0.9129 |
| | | | | | ***PLAC1L*** | ILMN_1757220 | - | - | - | - | - | - | -0.01391 | 0.2999 |
| rs610932 | ***MS4A6A*** | GI_23238233-A | -0.02596 | 0.4978 | ***MS4A6A*** | ILMN_1721035 | -0.1298 | 0.01917 | -0.06277 | 0.1781 | -0.1068 | 0.101 | -0.1229 | 0.02279 |
| | ***MS4A2*** | GI_23397640-S | -0.00338 | 0.8032 | ***MS4A6A*** | ILMN_1797731 | -0.1239 | 0.01431 | -0.05795 | 0.1669 | -0.09928 | 0.1232 | -0.1002 | 0.04446 |
| | ***MS4A7*** | GI_23110999-S | -0.01259 | 0.9024 | ***MS4A2*** | ILMN_1806721 | 0.004206 | 0.7592 | -0.00184 | 0.8959 | 0.01337 | 0.2885 | 0.02492 | 0.0899 |
| | ***MS4A3*** | GI_23110992-S | 0.009299 | 0.4564 | ***MS4A7*** | ILMN_1743932 | - | - | -0.03136 | 0.4072 | -0.02838 | 0.6533 | -0.05286 | 0.1814 |
| | ***MS4A4A*** | GI_23110994-A | -0.01123 | 0.5048 | ***MS4A3*** | ILMN_1726552 | - | - | - | - | - | - | 0.01008 | 0.3962 |
| | ***PLAC1L*** | GI_42476002-S | 0.007844 | 0.4874 | ***MS4A3*** | ILMN_1751625 | - | - | - | - | - | - | 0.01783 | 0.283 |
| | ***MS4A6E*** | GI_23110998-S | -0.05092 | 0.3869 | ***MS4A4A*** | ILMN_1741712 | - | - | - | - | - | - | 0.01617 | 0.3682 |
| | | | | | ***PLAC1L*** | ILMN_1757220 | - | - | - | - | - | - | 0.02278 | 0.0816 |

**Supplementary Note for "Common variants in ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease.".**

## Sample Description

### Stage 1 Samples

*GERAD1*

The GERAD1[1] sample comprised up to 3941 AD cases and 7848 controls. This sample included 4113 cases and 1602 elderly screened controls genotyped at the Sanger Institute on the Illumina 610-quad chip, referred to collectively hereafter as the 610 group. These samples were recruited by the Medical Research Council (MRC) Genetic Resource for AD (Cardiff University; Institute of Psychiatry, London; Cambridge University; Trinity College Dublin), the Alzheimer's Research Trust (ART) Collaboration (University of Nottingham; University of Manchester; University of Southampton; University of Bristol; Queen's University Belfast; the Oxford Project to Investigate Memory and Ageing (OPTIMA), Oxford University); Washington University, St Louis, United States; MRC PRION Unit, University College London; London and the South East Region AD project (LASER-AD), University College London; Competence Network of Dementia (CND) and Department of Psychiatry, University of Bonn, Germany and the National Institute of Mental Health (NIMH) AD Genetics Initiative. These data were combined with data from 844 AD cases and 1255 elderly screened controls ascertained by the Mayo Clinic, Jacksonville, Florida; Mayo Clinic, Rochester, Minnesota; and the Mayo Brain Bank (Mayo1 dataset). All AD cases met criteria for either probable (NINCDS-ADRDA[2], DSM-IV) or definite (CERAD[3]) AD. All elderly controls were screened for dementia using the MMSE or ADAS-cog, were determined to be free from dementia at neuropathological examination or had a Braak score of 2.5 or lower. A total of 6,825 unscreened population controls were included GERAD1. These were drawn from large existing cohorts with available GWAS data, including the 1958 British Birth Cohort (1958BC) (http://www.b58cgene.sgul.ac.uk), NINDS funded neurogenetics collection at Coriell Cell Repositories (Coriell) (see http://ccr.coriell.org/), the KORA F4 Study[4], Heinz Nixdorf Recall Study[5,6] and ALS Controls. (NB: KORA samples were also included in the German Alzheimer's disease Integrated Genome Research

Network (AD-IG) GWAS, therefore for SNPs that were carried forward to Stage 2, KORA samples were removed from the GERAD1 analysis).

*EADI1*

The EADI1 sample (2025 AD cases and 5328 controls) have been described in detail previously[7]. Briefly, AD cases were ascertained by neurologists from Bordeaux, Dijon, Lille, Montpellier, Paris, Rouen, and were identified as French Caucasian[8]. Clinical diagnosis of probable AD was established according to DSM-III-R and NINCDS-ADRDA criteria[2]. Controls were selected from the 3C Study[9]. The 3C Study is a population-based, prospective study of the relationship between vascular factors and dementia. It has been carried out in three French cities: Bordeaux (southwest France), Montpellier (south France) and Dijon (central eastern France). A sample of non-institutionalised, over-65 subjects was randomly selected from the electoral rolls of each city. Between January 1999 and March 2001, 9686 subjects meeting the inclusion criteria agreed to participate. Following recruitment, 392 subjects withdrew from the study. Thus, 9294 subjects were finally included in the study (2104 in Bordeaux, 4931 in Dijon and 2259 in Montpellier). At the baseline clinical examination, blood samples were obtained from 8414 individuals who were representative of the source population. Trained psychologists administered a battery of neuropsychological tests, including the MMSE. All participants in Bordeaux and Montpellier were also examined by a neurologist at baseline. All control participants were followed for 4 years and did not develop dementia during this time period.

*ADNI*

Following quality control (QC) filters applied in this study, the Alzheimer's Disease Neuroimaging Initiative (ADNI; see www.loni.ucla.edu/ADNI) sample included 151 AD cases and 177 controls. These samples have been described in detail elsewhere[10]. ADNI is a multi-site observational study including AD, mild cognitive impairment (MCI), and elderly individuals with normal cognition assessing clinical and cognitive measures, MRI and PET scans and blood and CNS biomarkers. AD cases were between the ages of 55–90, with an MMSE score of

20–26 inclusive and meeting NINCDS-ADRDA criteria for probable AD[2], and having an MRI consistent with the diagnosis of AD. Control individuals were screened for dementia using the MMSE, adopting a cut off of 27 or above.

*TGEN*

The Translational Genomics Research Institute (TGEN) GWA study included 861 AD cases and 550 controls. Following QC applied in this study 571 AD cases and 332 controls were included in subsequent analyses. These samples have been described previously[11]. Briefly, the sample comprised two neuropathological cohorts of brain doners (Cases n=458; Controls n=274) and a 'clinical cohort' (Cases n=113; Controls n=58). All participants were at least 65 years old at the time of their death or last clinical assessment. For the two neuropathological cohorts, brain tissue for DNA extraction, neuropathological diagnoses and data were supplied by investigators from 20 of the National Institute on Aging (NIA)-sponsored Alzheimer's Disease Centers (ADCs) (in accordance with agreements with the NIA, the ADCs, and the National Alzheimer's Coordinating Center) and from the Netherlands Brain Bank. For the clinical cohort, DNA extracted from blood, clinical diagnoses and data from subjects assessed in Rochester, MN were supplied by investigators from the Mayo Clinic. Neuropathological AD cases satisfied clinical and neuropathological criteria for LOAD. Brain donor controls did not have significant cognitive impairment or significant neuropathological features of AD. Clinical cases satisfied NINCS-ADRDA criteria for probable AD. Clinical controls did not have clinically significant cognitive impairment.

Stage 1 included a total of up to 6688 AD cases and 13685 controls. All AD cases were diagnosed according NINCDS-ADRDA[2], DSM-IV or CERAD[3] criteria for either probable or definite AD. AD cases were predominantly female (62.4%). The mean age at disease onset and ascertainment in AD cases were 71.6 and 77.3 years, respectively. Stage 1 included a total of 7915 aged (≥60 years), screened controls (59.9% female; mean age at collection, 74.5 years) and 5770 population based, unscreened controls from the GERAD1 study (50.8% female, mean age at collection 48.6 years).

## Stage 2 Samples

Stage 2 included individual genotyping of the GERAD2 sample (3262 cases and 3320 controls) and *in silico* replication in the deCODE and AD-IG GWAS datasets (925 cases and 612 controls; 709 cases and 971 controls respectively).

*GERAD2*

The GERAD2 sample comprised 3262 AD cases and 3320 controls. These samples were ascertained by the Medical Research Council (MRC) Genetic Resource for AD (Cardiff University; Institute of Psychiatry, London; Cambridge University; Trinity College Dublin), the Alzheimer's Research Trust (ART) Collaboration (University of Nottingham; University of Manchester; University of Southampton; University of Bristol; Queen's University Belfast), Washington University, St Louis, United States; MRC PRION Unit, University College London; London and the South East Region AD project (LASER-AD), University College London; Competence Network of Dementia (CND) and Department of Psychiatry, University of Bonn, Germany and the National Institute of Mental Health (NIMH)AD Genetics Initiative, Aristotle University of Thessaloniki; the Caerphilly Prospective Study; the University of Munich; and a Belgian sample derived from a prospective clinical study at the Memory Clinic and Department of Neurology, ZNA Middelheim, Antwerpen[12]. All AD cases met criteria for either probable (NINCDS-ADRDA[2], DSM-IV) or definite (CERAD[3]) AD. Control subjects were aged (>60 years of age) and predominantly screened for dementia (95.5%).

Control subjects from the MRC Genetic Resource for AD, Queen's University Belfast (ART collaboration) and Belgium were screened for cognitive decline using the MMSE[13] or ADAS-Cog. Controls ascertained by the University of Bristol and University of Nottingham, as part of the ART collaboration, were neuropathologically assessed and were dementia-free according to CERAD criteria[3]. The control group from Munich was a population-based random sample from Munich, Germany. Individuals were screened for dementia and other neuropsychiatric disorders using a comprehensive interview including the SCID[14]. Additionally, the Family History Assessment Module was conducted to exclude psychiatric disorders including dementias among first-degree relatives.

A neurological examination was also conducted to exclude subjects with current CNS impairment. Individuals older than 60 years were screened for cognitive impairment using the Mini Mental Status Examination[13]. The control subjects from the University of Bonn were recruited within the German Study on Aging, Cognition and Dementia (AgeCoDe). Cognitive impairment was ruled out in those individuals with the Structured Interview for Diagnosis of Dementia of Alzheimer type, Multi-infarct Dementia and Dementia of other Aetiology according to DSM-IV and ICD-10 (SIDAM)[15], which includes a cognitive battery. All control subjects performed within the normal age, sex and education adjusted norms on this cognitive battery[16]. Greek controls were unrelated carers of AD patients or recruited from the Greek blood donation service.

*German Alzheimer's disease Integrated Genome Research Network GWAS*

This study included 709 AD cases and 971 controls of German extraction from the Alzheimer's disease Integrated Genome Research Network (AD-IG) GWA study, which has been previously reported in part[17]. All patients were recruited by specialists at the outpatient clinic of the Technische Universität München. AD cases were diagnosed according to NINCDS-ADRDA[2] criteria for probable AD. Cognitive performance was assessed using standard neuropsychological tests, such as the Cambridge Cognitive Examination[18] or a test endorsed by the Consortium to Establish a Registry for Alzheimer's disease[19] which includes the Mini Mental State Examination[13]. Controls were drawn from two population-based cohorts: the PopGen Biobank, run by the Universitätsklinikum Schleswig-Holstein and the KORA F4 Study[4].

*deCODE*

The deCODE sample comprised 925 AD cases and 612 controls. AD cases were enrolled through the Memory Clinic at Landspitali University Hospital, to which all Icelanders suffering from cognitive decline are referred. Additional individuals were selected for enrolment based on an encrypted list of 3,188 patients with cognitive impairment compiled from Icelandic hospitals and nursing homes, or based on phenotype information obtained through the Resident Assessment Instrument (RAI). Individuals diagnosed with definite,

probable or possible AD according to NINCDS-ADRDA criteria were included in the study (N = 823). Individuals recruited based on RAI data met ICD-10 criteria for Alzheimer's disease (N = 102).

Controls were characterized based on phenotype information from RAI, more specifically the Minimum Data Set for Nursing Homes (MDS-NH)[20] and Home Care (MDS-HC)[21]. Individuals with a score of zero on the Cognitive Performance Scale (CPS)[22] at age 85 or older (N = 612) were used as cognitively intact controls.

All samples were collected through studies approved by the Data Protection commission and the National Bioethics Committee of Iceland. All participating individuals, or their guardians, gave their informed consent before blood samples were drawn, and all sample identifiers were encrypted in accordance with the regulations of the Icelandic Data Protection Committee.

Stage 2 included a total of up to 4896 AD cases and 4903 controls. All cases were diagnosed according NINCDS-ADRDA[2], DSM-IV or CERAD[3] criteria for either possible, probable or definite AD. AD cases were predominantly female (63.4%). The mean age at disease onset and ascertainment in AD cases were 72.3 and 76.8 years, respectively. The stage 2 control group (55.1% female, mean age at ascertainment 70.0 years) were predominantly aged (≥60 years) and screened for dementia (77.2%).

**Stage 3 Sample**

*EADI2*

EADI2 case-control samples were obtained from centres in Finland (1 centre)[23], Spain (3 centres)[24,25] and Italy (10 centres)[26-35]. Clinical diagnoses of probable AD were all established according to the DSM-III-R and NINCDS-ADRDA criteria[2]. Controls were defined as subjects without DMS-III-R dementia criteria and with integrity of their cognitive functions (MMS>25). Written informed consent was obtained as described above, and the study protocols for all populations were reviewed and approved by the appropriate Institutional review boards of each country.

*Mayo2*

The Mayo2 case-control series consisted of Caucasian subjects from the United States ascertained at the Mayo Clinic Jacksonville, Mayo Clinic Rochester, and in the Mayo Clinic autopsy-confirmed sample. Additional Caucasian subjects from the United States were obtained through the National Cell Repository for Alzheimer's Disease (NCRAD), and European Caucasian subjects were obtained from Norway[36], Poland[37], and from six research institutes in the United Kingdom that are part of the Alzheimer's Research Trust Network (ART). AD cases ascertained at the Mayo Clinic Jacksonville, Mayo Clinic Rochester and NCRAD were diagnosed according to NINCDS-ADRDA criteria for possible or probable AD. Controls had a Clinical Dementia Rating[38] scale score of 0. Cases from the Mayo autopsy series were diagnosed according to NINCDS-ADRDA criteria for definite AD and had a Braak stage score of 4 or greater. Brains employed as controls had a Braak score of 2.5 or lower and were free from AD pathology at autopsy. AD cases ascertained in Norway were diagnosed according to NINCDS-ADRDA criteria for possible or probable AD. Controls were determined to be cognitively intact using a brief clinical interview and did not have a first degree relative with dementia. A proportion were screened for cognitive impairment using the MMSE[13]. AD cases in the Polish cohort were diagnosed with probable AD according to NINCDS-ADRDA criteria for AD. Polish controls were screened for cognitive impairment and did not show symptoms of dementia. Although the ART samples used in this follow-up do not overlap with those employed in Stage 1 of the study (genotyped as part of the GERAD1 GWAS[1] the same subject/sample ascertainment methodology was followed. The ART series included here are from Bristol, University of Leeds, Manchester, Nottingham, Oxford and Southampton. The Mayo2 cohort comprised 880 AD cases and 1332 controls genotyped as part of the GWAS study reported by Carrasquillo and colleagues[39] which were included in Stage 1 of this study. These individuals were only genotyped and used in the analysis of rs670139 as this SNP was not genotyped as part of the GWAS and these data were not included in Stage 1 of this study. Approval was obtained from the ethics committee or institutional review board of each institution responsible for the ascertainment and collection

of samples. Written informed consent was obtained for all individuals that participated in this study. Samples used in this study do not overlap with those included in the Harold et al. 1 publication.

*CHARGE*

The CHARGE[40] dataset analyzed here includes four large, prospective, community-based cohort studies that have genome-wide association data coupled with extensive data on multiple phenotypes[41]: the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Rotterdam Study, and the Age, Gene/Environment Susceptibility – Reykjavik Study (AGES-RS). A neurology working-group arrived at a consensus on phenotype harmonization, covariate selection and analytic plans for within-study analyses and meta-analysis of results[42]. Consent procedures, examination and surveillance components, data security, genotyping protocols and study design at each study were approved by a local Institutional Review Board. (NB: Stage 1 of the CHARGE GWA study reported by Seshadri and colleagues included data from the Mayo[39] and the TGEN[11] GWA studies. These data overlap with samples used in Stage 1 of this study and were removed from analyses of the CHARGE dataset in this study).

Clinical characteristics of all samples can be found in Supplementary Table 1. An overview of the study design is shown in Figure 1 of the manuscript.

All individuals included in these analyses have provided consent to take part in genetic association studies. We have obtained ethical approval to use these samples to search for susceptibility genes for Alzheimer's disease (MREC 04/09/030; Amendment 2 and 4; approved 27 July 2007).

**Genotyping and association analysis: GWA datasets**

*GERAD1*

Individuals were genotyped using either the Illumina 610-quad, the HumanHap550 or the HumanHap300 array. QC and analysis has been described in detail elsewhere[1]. Briefly, 529205 autosomal SNPs were tested for association with Alzheimer's disease using logistic regression assuming an additive model.

Covariates were included in the logistic regression analysis to allow for geographical region and chip. The first four principal components extracted from an EIGENSTRAT analysis were also included as covariates.

*EADI1*

Individuals were genotyped using the Illumina 610-quad array. QC and analysis has been described in detail elsewhere[7]. Briefly, 537029 autosomal SNPs were tested for association with Alzheimer's disease using logistic regression assuming an additive model, including age, sex and principal components to account for possible population stratification as covariates.

*ADNI*

Individuals were genotyped using the Illumina 610-quad array. No quality control had been performed on the publicly available ADNI dataset, therefore the data were subjected to QC-filtering prior to analysis by logistic regression. This included retaining individuals with missing genotype rates < 0.01, with mean autosomal heterozygosity between 0.32 and 0.34, and with mean X-chromosome heterozygosity either <0.02 for males, or between 0.25 and 0.40 for females. 523539 SNPs with a minor allele frequency > 0.01, a missing data rate <0.03 and a Hardy-Weinberg $P>1 \times 10^{-4}$ were retained in the study. These SNPs were tested for association with Alzheimer's disease using logistic regression assuming an additive model.

*TGEN*

Individuals were genotyped using the Affymetrix 500K array. Although some quality control had been performed on the publicly available TGEN data, additional filters were applied. We removed 172 individuals with missing genotype rates > 0.03. We also applied a filter based on mean autosomal heterozygosity, excluding 302 individuals with values above or below empirically determined thresholds. All individuals passing these QC filters were examined for potential genetic relatedness by calculating identity-by-descent (IBD) estimates for all possible pairs of individuals in PLINK, and removing one of each pair with an IBD estimate >0.125 (the level expected for first cousins). As

a result, 1 individual was excluded. We also sought to detect non-European ancestry. To this end, TGEN genotype data was merged with genotypes at the same SNPs from 210 unrelated European (CEU), Asian (CHB and JPT) and Yoruban (YRI) samples from the HapMap project. Subsequent to removing SNPs in extensive regions of LD (Chr 5:44–51.5 Mb; Chr 6: 25–33.5 Mb; Chr 8: 8–12 Mb; Chr 11: 45–57 Mb), we further excluded SNPs if any pair within a 50-SNP window had $r^2 > 0.2$. Genome-wide average identity-by-state (IBS) distance was calculated in PLINK between each pair of individuals in the resulting dataset. The resulting matrix of IBS distances was used as input for classical multidimensional scaling (MDS). When the first two dimensions were extracted and plotted against each other, three clusters were observed as corresponding to the European, Asian and Yoruban samples. Four samples appeared to be ethnic outliers from the European cluster and were excluded from further analysis. We assessed population structure within the data using principal components analysis as implemented in EIGENSTRAT to infer continuous axes of genetic variation. Eigenvectors were calculated based on the previously described LD-pruned subset. The EIGENSTRAT program also identifies genetic outliers, which are defined as individuals whose ancestry is at least 6 standard deviations from the mean on one of the top ten axes of variation. As a result of this analysis, 29 outliers were identified and excluded. SNPs with a minor allele frequency < 0.01, a missing data rate >0.03 and a Hardy-Weinberg $P<1 \times 10^{-4}$ were excluded. Following QC, 571 Alzheimer's disease cases, 332 controls and 301243 SNPs were included in the analysis. As there is little overlap between the Affymetrix 500K array and the Illumina 610 array, unobserved genotypes were imputed with MACH v.1.0, using haplotypes released from initial low coverage sequencing of 112 European ancestry samples in the 1000 genomes project (ftp://ftp.sanger.ac.uk/pub/1000genomes/REL-0908/LowCov/) as a reference sample. Imputation generated data for >8.2 million SNPs. These were subsequently filtered to exclude SNPs with MAF<0.01 or RSQR<0.3. SNPs not present on the Illumina 610 array were also excluded. 457509 autosomal SNPs were tested for association with Alzheimer's disease using logistic regression assuming an additive model. A covariate was included to distinguish between

country of origin, *i.e.* USA or the Netherlands. The first principal component from the EIGENSTRAT analysis was also included as a covariate.

*AD-IG*

Genotyping was performed by Illumina (San Diego, CA, USA) using their Sentrix HumanHap550 Genotyping BeadChip. Eighteen individuals with missing genotype rates > 0.3 were removed. All individuals passing this QC filter were examined for potential sex misclassification in PLINK. Seventeen individuals with differences in reported and estimated sex on the X-chromosome were excluded. Genome-wide average identity-by-state (IBS) distance was calculated in PLINK between each pair of individuals in the resulting dataset and removing one of each pair with an IBS estimated distance >0.985 (the level expected for identical individuals and monozygotic twins). As a result, 21 individuals were excluded. The resulting matrix of IBS distances was used as input for classical multidimensional scaling (MDS) to assess population structure[43]. When the first four dimensions were extracted and plotted against each other only one cluster without any outliers was observed in accordance with the origin and ethnic background of the German sample. To account further for any hidden population stratification the first two dimensions from the MDS approach were used as covariates in the logistic regression analysis[44]. SNPs with a minor allele frequency < 0.01, a call rate <0.8 and a Hardy-Weinberg $P < 1 \times 10^{-3}$ were excluded. Following QC, 709 Alzheimer's disease cases, 971 controls and 521102 SNPs were included in the analysis. SNPs were tested for association with Alzheimer's disease using logistic regression assuming an additive model. Age and sex were also included as covariates, along with the first two components from the MDS analysis.

*deCODE*

Individuals were genotyped using Illumina HumanHap300, Illumina HumanHap300-duo or Illumina HumanCNV370-duo BeadChips. Samples with yield below 98%, a higher-quality duplicate in the data set or evidence of non-European ancestry based on results from STRUCTURE[45] were excluded. SNPs were deemed unusable if they had yield below 95%, HWE $P < 1 \times 10^{-6}$ or an allele

frequency difference between chips with $P < 1 \times 10^{-6}$. For genotyped SNPs, analysis was carried out using a previously-described likelihood procedure[45]. Imputation was performed using IMPUTE[46] with the HapMap CEU samples as a training set or, for rs10761558, using an IMPUTE-like algorithm developed at deCODE and a long-range-phased[47] Icelandic training set typed using Illumina Human1M BeadChips. For analysis of imputed genotype probabilities, the likelihood method in SNPTEST was used. All results were corrected for relatedness and possible population stratification using genomic control[48]. The inflation factor was 1.13.

*CHARGE*

For analysis of prevalent events in the four cohorts, SNPs were tested for association with Alzheimer's disease using logistic regression assuming an additive model. For the analysis of incident events in the cohorts, participants who were free of dementia entered the analysis at the time of the DNA sample collection and were followed until the development of incident AD; participants were censored at death, at the time of their last follow-up examination or health status update when they were known to be free of clinical dementia, and when they developed dementia due to an alternate cause. Cox proportional hazards models were used to calculate hazard ratios with corresponding 95% confidence intervals after ensuring that assumptions of proportionality of hazards were met. In the CHS, FHS, and Rotterdam studies, controls contributed one set of person-years to the prevalent analysis and a second, non-overlapping set of person-years to the incident analyses. Under the martingale property of Cox models, the two analyses are independent and their independence was confirmed in simulation studies. Primary analyses were adjusted for age and sex and any evidence of population stratification. An inverse variance-weighted meta-analysis combined results from seven discrete sources: incident AD in the CHS, FHS, and Rotterdam cohorts, prevalent AD in the AGES, CHS, FHS, and Rotterdam cohorts. Note that in stage 1 of their GWA study, Seshadri *et al*.[40] meta-analyzed data from the CHARGE dataset, plus data from the Mayo sample from Carrasquillo *et al*.[39] (which also forms part of GERAD1), plus data from the TGEN sample. However, only the CHARGE summary statistics are included from this

group to prevent any overlap. Also note that as the CHARGE data was generated using multiple platforms, imputation had been performed to bridge any gaps.[40]

**Genotyping and association analysis: Non-GWAS Samples**

GERAD2

Genotyping was performed using the MassARRAY and iPlexGOLD systems (Sequenom™) according to manufacturer's recommendations. All assays were initially optimized by genotyping DNA from 30 CEPH parent-offspring trios (Utah residents with ancestry from northern and western Europe: CEU), also genotyped by the HapMap project. All plates for genotyping contained a mixture of cases, controls, blanks, and CEU samples. All Sequenom cluster-plots were visually inspected and double-genotyping was performed for every assay. Genotypes were called blind to sample identity, affected status, and blind to the other rater. Assays were only considered suitable for analysis if genotypes of CEU individuals were concordant with those in HapMap, where available. Genotypes from controls were tested for departure from Hardy-Weinberg equilibrium (HWE); rs10501927 alone showed nominally significant evidence of departure from HWE ($P$=0.03). GERAD2 data were analyzed by logistic regression assuming an additive model including covariates to distinguish between (i) the UK sample (ii) the Belgium sample (iii) the Bonn sample (iv) the Munich sample and (v) the Greece sample.

EADI2

EADI2 genotyping was performed using Sequenom assays. The primer and probe sequences for the genotyping assays are available upon request. In order to avoid any genotyping bias, cases and controls were randomly mixed when genotyping, and laboratory personnel were blinded to case/control status. Genotyping success rate was at least 95%, and no departure from Hardy-Weinberg equilibrium was observed for the markers. Statistical analyses was performed in each country (Finland, Italy and Spain) under an additive genetic model using logistic regression taking account of age, sex and disease status using SAS software release 9.1 (SAS Institute, Cary, NC). Inverse variance-weighted meta-analysis was used to combine results from the three cohorts.

*MAYO2*

All genotyping was performed at the Mayo Clinic in Jacksonville using TaqMan® SNP Genotyping Assays in an ABI PRISM® 7900HT Sequence Detection System with 384-Well Block Module from Applied Biosystems, California, USA. The genotype data was analyzed using the SDS software version 2.2.3 (Applied Biosystems, California, USA). The Mayo2 data data were analyzed by logistic regression assuming an additive model including covariates to distinguish between (i) the US sample (ii) the UK sample (iii) the Norweigen sample and (iv) the Polish sample.

## References


1. Harold, D. et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **41**, 1088-93 (2009).
2. 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).
3. Mirra, S.S. et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479-86 (1991).
4. Wichmann, H.E., Gieger, C. & Illig, T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**, S26-30 (2005).
5. Birnbaum, S. et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet* **41**, 473-7 (2009).
6. Hillmer, A.M. et al. Susceptibility variants for male-pattern baldness on chromosome 20p11. *Nat Genet* **40**, 1279-81 (2008).
7. Lambert, J.C. et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* (2009).
8. Dreses-Werringloer, U. et al. A polymorphism in CALHM1 influences Ca2+ homeostasis, Abeta levels, and Alzheimer's disease risk. *Cell* **133**, 1149-61 (2008).
9. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology* **22**, 316-25 (2003).
10. Potkin, S.G. et al. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. *PLoS One* **4**, e6501 (2009).
11. Reiman, E.M. et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* **54**, 713-20 (2007).
12. Brouwers, N. et al. Genetic variability in progranulin contributes to risk for clinically diagnosed Alzheimer disease. *Neurology* **71**, 656-64 (2008).
13. Folstein, M.F., Folstein, S.E. & McHugh, P.R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-98 (1975).
14. First, M., Spitzer, R., Gibbon, M. & Williams, J. *Structured Clinical Interview for DSM-IV Axis I Disorders - Patient Edition. 1995.*, (New York: Biometrics Research Department, New York State Psychiatric Institute., 1995).
15. Zaudig, M. et al. *Strukturiertes Interview für die Diagnose einer Demenz vom Alzheimer Typ, der Multiinfarkt- (oder vaskulären) Demenz und Demenzen anderer Ätiologie nach DSM-III-R, DSM-IV und ICD-10*, (Huber, Bern, 1996).
16. Luck, T., Zaudig, M., Wiese, B. & Riedel-Heller, S.G. Age- and Education-Specific Reference Values for the Cognitive Test Section According to the New CASMIN Educational Classification. *Zeitschrift Fur Gerontopsychologie und Psychiatrie* **20**, 31-38 (2007).
17. Feulner, T.M. et al. Examination of the current top candidate genes for AD in a genome-wide association study. *Mol Psychiatry* **15**, 756-66 (2009).

18. Roth, M. et al. CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *Br J Psychiatry* **149**, 698-709 (1986).
19. Welsh, K.A. et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology* **44**, 609-14 (1994).
20. Morris, J.N. et al. Designing the national resident assessment instrument for nursing homes. *Gerontologist* **30**, 293-307 (1990).
21. Morris, J.N. et al. Comprehensive clinical assessment in community setting: applicability of the MDS-HC. *J Am Geriatr Soc* **45**, 1017-24 (1997).
22. Morris, J.N. et al. MDS Cognitive Performance Scale. *J Gerontol* **49**, M174-82 (1994).
23. Viswanathan, J. et al. An association study between granulin gene polymorphisms and Alzheimer's disease in Finnish population. *Am J Med Genet B Neuropsychiatr Genet* **150B**, 747-50 (2009).
24. Alvarez, V. et al. Mitochondrial transcription factor A (TFAM) gene variation and risk of late-onset Alzheimer's disease. *J Alzheimers Dis* **13**, 275-80 (2008).
25. Infante, J. et al. Synergistic effect of two oxidative stress-related genes (heme oxygenase-1 and GSK3beta) on the risk of Parkinson's disease. *Eur J Neurol* **17**, 760-2 (2009).
26. Arosio, B. et al. Interleukin-10 and interleukin-6 gene polymorphisms as risk factors for Alzheimer's disease. *Neurobiol Aging* **25**, 1009-15 (2004).
27. Bossu, P. et al. Interleukin 18 gene polymorphisms predict risk and outcome of Alzheimer's disease. *J Neurol Neurosurg Psychiatry* **78**, 807-11 (2007).
28. Colacicco, A.M. et al. Alpha-2-macroglobulin gene, oxidized low-density lipoprotein receptor-1 locus, and sporadic Alzheimer's disease. *Neurobiol Aging* **30**, 1518-20 (2009).
29. Bosco, P. et al. The CDC2 I-G-T haplotype associated with the APOE epsilon4 allele increases the risk of sporadic Alzheimer's disease in Sicily. *Neurosci Lett* **419**, 195-8 (2007).
30. Kauwe, J.S. et al. Variation in MAPT is associated with cerebrospinal fluid tau levels in the presence of amyloid-beta deposition. *Proc Natl Acad Sci U S A* **105**, 8050-4 (2008).
31. Mancuso, M. et al. Lack of association between mtDNA haplogroups and Alzheimer's disease in Tuscany. *Neurol Sci* **28**, 142-7 (2007).
32. Nacmias, B. et al. Implication of GAB2 gene polymorphism in Italian patients with Alzheimer's disease. *J Alzheimers Dis* **16**, 513-5 (2009).
33. Ravaglia, G. et al. Blood inflammatory markers and risk of dementia: The Conselice Study of Brain Aging. *Neurobiol Aging* **28**, 1810-20 (2007).
34. Piccardi, M. et al. Alzheimer's disease: case-control association study of polymorphisms in ACHE, CHAT, and BCHE genes in a Sardinian sample. *Am J Med Genet B Neuropsychiatr Genet* **144B**, 895-9 (2007).
35. Seripa, D. et al. Association analysis of GRIN2B, encoding N-methyl-D-aspartate receptor 2B subunit, and Alzheimer's disease. *Dement Geriatr Cogn Disord* **25**, 287-92 (2008).
36. Sando, S.B. et al. Risk-reducing effect of education in Alzheimer's disease. *Int J Geriatr Psychiatry* **23**, 1156-62 (2008).

37. Klimkowicz-Mrowiec, A. et al. Interleukin-1 gene -511 CT polymorphism and the risk of Alzheimer's disease in a Polish population. *Dement Geriatr Cogn Disord* **28**, 461-4 (2009).
38. Berg, L. Clinical Dementia Rating (CDR). *Psychopharmacol Bull* **24**, 637-9 (1988).
39. Carrasquillo, M.M. et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat Genet* **41**, 192-8 (2009).
40. Seshadri, S. et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303**, 1832-40 (2010).
41. Psaty, B.M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* **2**, 73-80 (2009).
42. Ikram, M.A. et al. Genomewide association studies of stroke. *N Engl J Med* **360**, 1718-28 (2009).
43. Wang, D. et al. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. *BMC Proc* **3 Suppl 7**, S109 (2009).
44. Li, Q. & Yu, K. Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. *Genet Epidemiol* **32**, 215-26 (2008).
45. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59 (2000).
46. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
47. Kong, A. et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet* **40**, 1068-75 (2008).
48. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
49. Stranger, B.E. et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet 1*, **e78** (2005).
50. Gibbs, J.R. et al. Abundant quantitative trait Loci exist for DNA methylation and gene expression in human brain. *PLoS Genet 6*, **e1000952** (2010).

**Complete Acknowledgements**

We thank the individuals and families who took part in this research. We acknowledge use of data from the GERAD1 and EADI1 GWAS conducted by Harold and colleagues (2009) and Lambert and Colleagues (2009), respectively. Please see the original publications for full acknowledgements relating to these studies. The following acknowledgements refer to independent replication samples genotyped as part of this study: Cardiff University was supported by the Wellcome Trust, Medical Research Council (MRC), Alzheimer's Research Trust (ART) and the Welsh Assembly Government (WAG). The MRC and Mercer's Institute for Research on Ageing supported the Trinity College group. ART supported sample collections at the following institutions; Institute of Psychiatry, the South West Dementia Bank and the Universities of Cambridge, Nottingham, Manchester and Belfast. The Belfast group acknowledges support from the Alzheimer's Society, Ulster Garden Villages, Northern Ireland Research and Development Office (NI R&D) and the Royal College of Physicians–Dunhill Medical Trust. The Institute of Psychiatry, Cambridge University and Trinity College, Dublin received support from the MRC. The South West Dementia Brain Bank acknowledges support from Bristol Research into Alzheimer's and Care of the Elderly (BRACE). Washington University was funded by US National Institutes of Health (NIH) grants, the Barnes Jewish Foundation and the Charles and Joanne Knight Alzheimer's Research Initiative. Patient recruitment from the MRC Prion Unit, University College London (UCL) was supported by the UCL Hospital/UCL Biomedical Centre. LASER-AD was funded by Lundbeck. The Bonn group was supported by the German Federal Ministry of Education and Research (BMBF), Competence Network Dementia and Competence Network Degenerative Dementia, and by the Alfried Krupp von Bohlen und Halbach-Stiftung. Work on this sample was supported in part by the IRP of the National Institute on Aging (NIA), NIH, Department of Health and Human Services; Z01 AG000950-06. The Antwerp site was supported by the VIB Genetic Service Facility, the Biobank of the Institute Born-Bunge, the Special Research Fund of the University of Antwerp, the Fund for Scientific Research-Flanders, the Foundation for Alzheimer Research, a Methusalem Excellence grant of the Flemish government and the Interuniversity Attraction Poles program P6/43 of the Belgian Federal Science Policy Office. K.S. is a postdoctoral fellow and K.B. a PhD fellow (Fund for Scientific Research-Flanders). The Caerphilly Prospective study was undertaken by the former MRC Epidemiology Unit (South Wales) and was funded by the Medical Research Council. The Mayo2 group were supported by grants from the US National Institutes of Health, NIA (R01 AG18023, NRG-R, SGY); Mayo Alzheimer’s Disease Research Center, (P50 AG16574, RCP, DWD, NRG-R, SGY); Mayo Alzheimer’s Disease Patient Registry, (U01 AG06576, RCP); US National Institute on Aging (AG25711, AG17216, AG03949, DWD); and National Institute

of Neurological Disorders and Stroke (NINDS; R01 NS057567-03, 1RC2 NS070276-01, P50 NS 072187-01, ZKW). Samples from the National Cell Repository for Alzheimer's Disease (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the National Institute on Aging (NIA), were used in this study. This project was also generously supported by the Robert and Clarice Smith Postdoctoral Fellowship (MMC); Robert and Clarice Smith and Abigail Van Buren Alzheimer's Disease Research Program (RCP, DWD, NRG-R; SGY); the Palumbo Professorship in Alzheimer's disease Research (SGY); the Mayo Clinic Florida Research Committee CR and Collaborative programs (ZKW); and by the Carl Edward Bolch, Jr. and Susan Bass Bolch Gift (ZKW).

We acknowledge the use of existing and new genotype data from the European Alzheimer's Disease Initiative (EADI). We thank Dr. Anne Boland (Centre National de Génotypage) for her technical help in preparing the DNA samples for analyses of EADI data. The EADI group were supported by the National Foundation for Alzheimer's disease and related disorders, the Institut Pasteur de Lille and the Centre National de Génotypage (CNG). The Three-City Study was performed as part of a collaboration between the Institut National de la Santé et de la Recherche Médicale (Inserm), the Victor Segalen Bordeaux II University and Sanofi-Synthélabo. The Fondation pour la Recherche Médicale funded the preparation and initiation of the study. The 3C Study was also funded by the Caisse Nationale Maladie des Travailleurs Salariés, Direction Générale de la Santé, MGEN, Institut de la Longévité, Agence Française de Sécurité Sanitaire des Produits de Santé, the Aquitaine and Bourgogne Regional Councils, Fondation de France and the joint French Ministry of Research/INSERM "Cohortes et collections de données biologiques" programme. Lille Génopôle received an unconditional grant from Eisai. The EADI2 cohort includes samples ascertained in Finland, Spain and Italy. Financial support for the Finish sample collection was provided by the Health Research Council of the Academy of Finland, EVO grant 5772708 of Kuopio University Hospital, and the Nordic Centre of Excellence in Neurodegeneration. Italian sample collections: the Bologna site (FL) obtained funds from the Italian Ministry of research and University as well as Carimonte Foundation. The Florence site was supported by a grant from the Italian ministry of Health (RFPS-2006-7-334858). The Milan site was supported by a grant from the "fondazione Monzino" and the Italian Ministry of Health (PS39). We thank the expert contribution of Mr. Carmelo Romano. Spanish sample collection: the Madrid site (MB) was supported by grants of the Ministerio de Educación y Ciencia and the Ministerio de Sanidad y Consumo (Instituto de Salud Carlos III), and an institutional grant of the Fundación Ramón Areces to the CBMSO. We thank I. Sastre and Dr. A Martínez-García for the preparation and control of the DNA collection, and Drs. P. Gil and P. Coria for their cooperation in the cases/controls recruitment. We are grateful to the Asociación de Familiares de

Alzheimer de Madrid (AFAL) for continuous encouragement and help. We acknowledge use of existing genotype data from the Alzheimer's disease Neuroimaging Initiative (AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., and Wyeth, as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. The ADNI group were also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation. The German Alzheimer's disease Integrated Genome Research Network (AD-IG) GWA study was funded by the German National Genome Research Network and the German Ministry for Education and Research (Grant Number 01GS0465MR). We acknowledge the use of existing genotype data from the Aging Gene-Environment Susceptibility- Reykjavik Study, this study has been funded by NIA contract N01-AG-12100 with contributions from NEI, NIDCD and NHLBI, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). We acknowledge the use of existing genotype data from the Cardiovascular Health Study (CHS), the research reported in this article was supported by contract numbers N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, grant numbers U01 HL080295 and R01 HL087652 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke and grants AG15928 and AG20098 from the National Institute on Aging. DNA handling and genotyping was supported in part by National Center for Research Resources grant M01RR00425 to the Cedars-Sinai General Clinical Research Center Genotyping core, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant DK063491 to the Southern California Diabetes Endocrinology Research Center, and the Cedars-Sinai Board of Governors' Chair in Medical Genetics (JIR). A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. We acknowledge the use of existing genotype data from the Framingham Heart Study of the National Heart Lung and Blood Institute (NHLBI) of the National Institutes of Health and Boston University School of Medicine. This work was

supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. This study was also supported by grants from the National Institute of Neurological Disorders and Stroke (NS17950) and the National Institute of Aging (AG08122, AG16495, AG033193 and AG031287, AG033040 and P30AG013846). We acknowledge the use of existing genotype data from the Rotterdam Study. The GWA database of the Rotterdam Study was funded through the Netherlands Organisation of Scientific Research NWO (nr. 175.010.2005.011). This study was further supported by the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. The Rotterdam Study is supported by the Erasmus Medical Center and Erasmus University, Rotterdam; the Netherlands organization for scientific research (NWO), the Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. We also thank Heike Kölsch we established the AD DNA bank for the University Bonn Samples used in this study. None of the study funders had a role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.