**Supplementary Material:** Machine learning methods applied to genotyping data capture interactions between single nucleotide variants in late onset Alzheimer’s disease

[Supplementary Figure 1 2](#_Toc95139436)

[Supplementary Figure 2 3](#_Toc95139437)

[Supplementary Figure 3 4](#_Toc95139438)

[Supplementary Figure 4 5](#_Toc95139439)

[Supplementary Figure 5 6](#_Toc95139440)

[Supplementary Figure 6 7](#_Toc95139441)

[Supplementary Table 1 8](#_Toc95140892)

[Supplementary Table 2 12](#_Toc95140893)

[Supplementary Table 3 21](#_Toc95140894)

[Supplementary Table 4 22](#_Toc95140895)

[Supplementary Table 5 23](#_Toc95140896)

[Supplementary Table 6 24](#_Toc95140897)

Supplementary Figure 1

Chart, box and whisker chart

Description automatically generated

**Supplementary Figure 1.** Distribution of the selected samples in UKB over sex in **a** and age in **b**. In **c** and **d**, distribution of the SNVs in AD predictors over chromosomes, andover the different genomic regions respectively.

Supplementary Figure

Diagram, schematic

Description automatically generated

**Supplementary Figure 2.** 10-fold cross-validation on the training set split for hyperparameter optimization and overfitting correction step. Python 3.7.6 with Scikit-learn v0.22.1 module was used to build the ML models. A train/test split was applied to have 80% of samples for training and 20% of samples for testing. Random undersampling was used to balance cases and controls, as oversampling strategy was prone to overfitting. Hyperparameter selection was performed on the training set through a 10-fold cross-validation. Two metrics such as AUC-ROC and f-score were considered for determining the best hyperparameter configuration for each model and to correct overfitting. For each combination of hyperparameters, the mean and the standard deviation was obtained across the 10 folds on the training and validation set. In addition, 199 random SNVs not present in the 145 AD related SNVs were used as reference predictors. These reference SNVs were in genes from DisGeNet “curated gene-disease associations” dataset but not reported in the “curated variant-disease associations” dataset using the same AD categories (*Supplementary Table 1, page 3*). Models were evaluated with the AD predictors and reference predictors independently. The median of AUC-ROC (a, b, c) and f-score (d, e, f) was around 0.5 in validation sets of models with reference predictors, and higher than 0.7 in validation sets of models with AD predictors. With these differences, we confirmed that the good performance of the ML models on the validation set using the selected AD SNVs as predictors (from 0.7 to 0.9) is not due to an overfitting effect caused by the intrinsic properties of genomic variants in the selected genomic regions. The list of SNVs used as AD predictors (145) and reference predictors (199) is provided in *Supplementary Table 2*. The python scripts used in this step are available in GitHub48. In order to choose the optimum hyperparameter configuration, the parameters were selected based on: a) higher mean scores in the validation sets with the two evaluation metrics, b) lower standard deviation of the evaluation metrics across the 10 cross-validation folds, c) lower mean differences in the evaluation metrics when comparing train and validation sets in order to correct for overfitting. The selected hyperparameters are listed in *Supplementary Table 1, page 4*. The final model was trained on the original train set (80%) and tested on the test set (20% of samples).

Supplementary Figure

Chart

Description automatically generated with medium confidence

**Supplementary Figure 3.** In **a** AUC-ROC obtained on the testing set with the three ML models. In **b**, **c**, **d** FI score for the 145 SNVs in AD predictors obtained in GB, ET and RF respectively, and the cut-off used to prioritize the SNVs.

Supplementary Figure

Chart, box and whisker chart

Description automatically generated

**Supplementary Figure 4**. AUC-ROC obtained in the 10-fold cross-validation with different hyperparameter configurations to compare performances of the ML methods with different predictors. The metrics obtained in the 10-fold cross-validation step with different hyperparameter configurations were used to compare the performance of the models, as comparing models with different predictors and a single hyperparameter configuration may produce biased results. In **a b** and **c** comparison of the performances using the two APOEe4 SNVs (rs429358 and rs7412) and the set of 145 SNVs in AD predictors for RF, ET, and GB models respectively. The distribution of AUC-ROC performances was higher in models using the AD predictors with respect to models only using the two APOE SNVs. The statistical significance was assessed applying a Wilcoxon test on the distribution of values in validation sets (blue boxes).

Supplementary Figure

Diagram

Description automatically generated

**Supplementary Figure 5**. Genomic profiles of correctly classified samples in ET defined by the 20 prioritized SNVs. Same structure as in *Figure 2*.

Supplementary Figure

Diagram

Description automatically generated

**Supplementary Figure 6**. Genomic profiles of correctly classified samples in RF defined by the 15 prioritized SNVs. Same structure as in *Figure 2*.

Supplementary Table

|  |  |
| --- | --- |
| **ICD10 Coding** | **Meaning** |
| F00 | F00 Dementia in Alzheimer's disease |
| F001 | F00.1 Dementia in Alzheimer's disease with late onset |
| F002 | F00.2 Dementia in Alzheimer's disease, atypical or mixed type |
| F009 | F00.9 Dementia in Alzheimer's disease, unspecified |
| G30 | G30 Alzheimer's disease |
| G301 | G30.1 Alzheimer's disease with late onset |
| G308 | G30.8 Other Alzheimer's disease |
| G309 | G30.9 Alzheimer's disease, unspecified |

**Supplementary Table1, page 1** CD-10 codes used to select AD cases in UK Biobank.



**Supplementary Table1, page 2** Criteria to categorize AD and controls in ADNI3 based on the annual reports. The fields probable or possible dementia due to Alzheimer's Disease (column DXAPP) and dementia due to Alzheimer's Disease (column DXDDUE) were used to categorize AD cases. Individuals with cognitive normal evaluation (column DIAGNOSIS), a lack of AD indicators in columns DXAPP, DXDDUE, and a lack of mild cognitive impairment (MCI) due to Alzheimer's Disease (column DXMDUE) were categorized as controls. Using these filters, 13 AD and 126 controls were selected from ADNI3.

|  |  |
| --- | --- |
| **Disease ID** | **Disease Name** |
| C0002395 | Alzheimer's Disease |
| C1851958 | Lewy Body Variant of Alzheimer Disease |
| C1843013 | Alzheimer disease, familial, type 3 |
| C1843014 | Alzheimer Disease, Familial, 3, with Spastic Paraparesis and Unusual Plaques |
| C1843015 | Alzheimer Disease, Familial, 3, with Spastic Paraparesis and Apraxia |
| C0276496 | Familial Alzheimer Disease (FAD) |
| C0494463 | Alzheimer Disease, Late Onset |
| C0750900 | Alzheimer's Disease, Focal Onset |
| C1837149 | Alzheimer Disease 9 |

**Supplementary Table1, page 3** Disease IDs in MedGen concept ID format (column Disease ID) used to select Alzheimer’s disease related variants in DisGeNet database.

|  |  |
| --- | --- |
| **ML method** | **Parameters** |
| Gradient Boosted Decision Trees (GB) | * n\_estimators = 100 * learning\_rate = 0.0001 * subsample = 0.7 * max\_depth = 9 * loss = 'deviance' |
| Extremely Randomized Trees (ET) | * n\_estimators = 100 * min\_samples\_split = 8 * min\_samples\_leaf = 1 * max\_depth = 7 |
| Random Forest (RF) | * n\_estimators = 60 * min\_samples\_split = 5 * min\_samples\_leaf = 2 * max\_depth = 7 |

**Supplementary Table1, page 4** Parameters selected in the three ML methods. Hyperparameter selection was performed on the training set through a 10-fold cross-validation. Parameters not listed in the table were used as default in Scikit-learn module.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **% LOAD** | **% Cntrl** | **Fisher pval** |
| **TP** | **C1** | 27.83 | 1.21 | 2.69E-168 |
| **C2** | 3.24 | 0.00 | 1.04E-39 |
| **C3** | 12.62 | 0.00 | 1.44E-159 |
| **TN** | **C1** | 0.49 | 24.42 | 9.97E-52 |
| **C2** | 0.00 | 17.50 | 3.60E-69 |

**Supplementary Table1, page 5** Percentage of AD and controls with genomic profiles in clusters captured by GB models (*Figure 2*). The percentages are calculated over the total AD n = 618 and Controls n= 62392 without any missing value in the prioritized SNVs. The differences between AD and controls are measured with a Fisher test.



**Supplementary Table1, page 6** Information on the eQTLs obtained from GTEx portal on the 6 SNVs commonly prioritized by the three ML models.



**Supplementary Table1, page 7** Information on the sQTLs obtained from GTEx portal on the 6 SNVs commonly prioritized by the three ML models.

Supplementary Table

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **dbSNP ID** | **Chr** | **hg19 Postition** | **Ref** | **Alt** | **Region** | **Gene Symbol** | **DisGeNet AD SNV** |
| rs1038026 | 19 | 45405062 | A | G | intronic | TOMM40 | YES |
| rs10405693 | 19 | 45326664 | C | T | intergenic | BCAM;PVRL2 | YES |
| rs10415983 | 19 | 45711598 | C | T | intergenic | BLOC1S3;EXOC3L2 | YES |
| rs10498633 | 14 | 92926952 | G | T | intronic | SLC24A4 | YES |
| rs10792832 | 11 | 85867875 | G | A | intergenic | PICALM;EED | YES |
| rs10850408 | 12 | 115380393 | C | T | intergenic | TBX3;MED13L | YES |
| rs10948363 | 6 | 47487762 | A | G | intronic | CD2AP | YES |
| rs11136000 | 8 | 27464519 | C | T | intronic | CLU | YES |
| rs11257238 | 10 | 11717397 | T | C | intergenic | USP6NL;ECHDC3 | YES |
| rs115786578 | 20 | 32737601 | G | A | intergenic | EIF2S2;ASIP | YES |
| rs1160985 | 19 | 45403412 | C | T | intronic | TOMM40 | YES |
| rs11610206 | 12 | 47639526 | T | C | intergenic | PCED1B;LOC105369747 | YES |
| rs11667640 | 19 | 45379791 | C | T | intronic | PVRL2 | YES |
| rs11669338 | 19 | 45382984 | T | G | intronic | PVRL2 | YES |
| rs11673139 | 19 | 45383037 | A | T | intronic | PVRL2 | YES |
| rs116932763 | 7 | 149549009 | G | A | intronic | ZNF862 | YES |
| rs11767557 | 7 | 143109139 | T | C | ncRNA\_intronic | EPHA1-AS1 | YES |
| rs11770757 | 7 | 133747946 | G | A | intronic | EXOC4 | YES |
| rs11771145 | 7 | 143110762 | G | A | ncRNA\_intronic | EPHA1-AS1 | YES |
| rs117964204 | 17 | 48692082 | C | T | intronic | CACNA1G | YES |
| rs117983694 | 11 | 85396880 | A | C | exonic | CCDC89 | YES |
| rs118071777 | 7 | 88244249 | G | A | intergenic | STEAP4;ZNF804B | YES |
| rs118152978 | 8 | 127353538 | A | G | intergenic | LOC101927657;FAM84B | YES |
| rs11824773 | 11 | 60076940 | G | C | downstream | MS4A4A | YES |
| rs12041233 | 1 | 37752707 | G | A | intergenic | MIR4255;LINC01137 | YES |
| rs12334143 | 7 | 85183016 | T | C | intergenic | LINC00972;GRM3 | YES |
| rs12361953 | 11 | 24611130 | C | G | intronic | LUZP2 | YES |
| rs12610605 | 19 | 45370838 | G | A | intronic | PVRL2 | YES |
| rs12787412 | 11 | 85703093 | C | T | intronic | PICALM | YES |
| rs12805520 | 11 | 85630411 | C | T | exonic | CCDC83 | YES |
| rs12989701 | 2 | 127887985 | C | A | intergenic | BIN1;CYP27C1 | YES |
| rs1385600 | 11 | 77936166 | A | G | exonic | GAB2 | YES |
| rs1466662 | 4 | 155347393 | A | T | intronic | DCHS2 | YES |
| rs1476679 | 7 | 100004446 | T | C | intronic | ZCWPW1 | YES |
| rs1531517 | 19 | 45242173 | G | A | intergenic | CEACAM16;BCL3 | YES |
| rs1532278 | 8 | 27466315 | C | T | intronic | CLU | YES |
| rs1562990 | 11 | 60023087 | A | C | intergenic | MS4A4E;MS4A4A | YES |
| rs157580 | 19 | 45395266 | A | G | intronic | TOMM40 | YES |
| rs157582 | 19 | 45396219 | C | T | intronic | TOMM40 | YES |
| rs157590 | 19 | 45398716 | A | C | intronic | TOMM40 | YES |
| rs1582763 | 11 | 60021948 | G | A | intergenic | MS4A4E;MS4A4A | YES |
| rs16979595 | 19 | 45477381 | G | A | intronic | CLPTM1 | YES |
| rs17125944 | 14 | 53400629 | T | C | intronic | FERMT2 | YES |
| rs17817600 | 11 | 85677471 | A | G | intronic | PICALM | YES |
| rs1800795 | 7 | 22766645 | G | C | ncRNA\_intronic | LOC541472 | YES |
| rs190982 | 5 | 88223420 | A | G | ncRNA\_intronic | MEF2C-AS1 | YES |
| rs1925690 | 6 | 87867063 | C | T | intronic | ZNF292 | YES |
| rs2043948 | 14 | 75073048 | C | T | intronic | LTBP2 | YES |
| rs2075650 | 19 | 45395619 | A | G | intronic | TOMM40 | YES |
| rs2228467 | 3 | 42906116 | T | C | exonic | ACKR2 | YES |
| rs2279590 | 8 | 27456253 | C | T | intronic | CLU | YES |
| rs2373115 | 11 | 78091150 | C | A | intronic | GAB2 | YES |
| rs2421847 | 1 | 171557600 | A | G | exonic | PRRC2C | YES |
| rs2718058 | 7 | 37841534 | A | G | intergenic | GPR141;NME8 | YES |
| rs283813 | 19 | 45389174 | T | A | intronic | PVRL2 | YES |
| rs28834970 | 8 | 27195121 | T | C | intronic | PTK2B | YES |
| rs2965101 | 19 | 45237812 | T | C | intergenic | CEACAM16;BCL3 | YES |
| rs2965109 | 19 | 45225345 | C | T | intergenic | CEACAM16;BCL3 | YES |
| rs34342646 | 19 | 45388130 | G | A | intronic | PVRL2 | YES |
| rs34404554 | 19 | 45395909 | C | G | intronic | TOMM40 | YES |
| rs35349669 | 2 | 234068476 | C | T | intronic | INPP5D | YES |
| rs3745150 | 19 | 45385759 | G | C | intronic | PVRL2 | YES |
| rs3748140 | 8 | 8999019 | C | T | exonic | PPP1R3B | YES |
| rs3752246 | 19 | 1056492 | C | G | exonic | ABCA7 | YES |
| rs3760627 | 19 | 45457180 | T | C | upstream | CLPTM1 | YES |
| rs3763849 | 11 | 60193191 | C | T | intergenic | MS4A14;MS4A5 | YES |
| rs3764650 | 19 | 1046520 | T | G | intronic | ABCA7 | YES |
| rs3781834 | 11 | 121445940 | A | G | intronic | SORL1 | YES |
| rs3794318 | 12 | 8758543 | A | G | intronic | AICDA | YES |
| rs3818361 | 1 | 207784968 | G | A | intronic | CR1 | YES |
| rs3851179 | 11 | 85868640 | C | T | intergenic | PICALM;EED | YES |
| rs3865444 | 19 | 51727962 | C | A | upstream | CD33 | YES |
| rs387976 | 19 | 45379060 | A | C | intronic | PVRL2 | YES |
| rs3931397 | 4 | 149079497 | G | T | intronic | NR3C2 | YES |
| rs4038129 | 2 | 17774746 | A | G | intronic | VSNL1 | YES |
| rs4038131 | 2 | 17775032 | A | G | intronic | VSNL1 | YES |
| rs405509 | 19 | 45408836 | G | T | upstream | APOE | YES |
| rs405697 | 19 | 45404691 | G | A | intronic | TOMM40 | YES |
| rs4147929 | 19 | 1063443 | G | A | intronic | ABCA7 | YES |
| rs416041 | 19 | 45370854 | A | G | intronic | PVRL2 | YES |
| rs429358 | 19 | 45411941 | T | C | exonic | APOE | YES |
| rs439401 | 19 | 45414451 | C | T | intergenic | APOE;APOC1 | YES |
| rs440277 | 19 | 45361224 | G | A | intronic | PVRL2 | YES |
| rs4420638 | 19 | 45422946 | A | G | downstream | APOC1 | YES |
| rs4557697 | 8 | 63843631 | A | G | intronic | NKAIN3 | YES |
| rs4663098 | 2 | 127873035 | C | T | intergenic | BIN1;CYP27C1 | YES |
| rs4676049 | 2 | 109635257 | C | T | intergenic | EDAR;SH3RF3-AS1 | YES |
| rs471470 | 11 | 85831541 | A | C | intergenic | PICALM;EED | YES |
| rs4803770 | 19 | 45427353 | C | G | intergenic | APOC1;APOC1P1 | YES |
| rs483082 | 19 | 45416178 | G | T | intergenic | APOE;APOC1 | YES |
| rs4844610 | 1 | 207802552 | C | A | intronic | CR1 | YES |
| rs4938933 | 11 | 60034429 | T | C | intergenic | MS4A4E;MS4A4A | YES |
| rs519113 | 19 | 45376284 | C | G | intronic | PVRL2 | YES |
| rs519825 | 19 | 45366779 | T | C | intronic | PVRL2 | YES |
| rs536841 | 11 | 85787824 | T | C | intergenic | PICALM;EED | YES |
| rs541458 | 11 | 85788351 | T | C | intergenic | PICALM;EED | YES |
| rs56039743 | 1 | 26620806 | T | C | exonic | UBXN11 | YES |
| rs561655 | 11 | 85800279 | A | G | intergenic | PICALM;EED | YES |
| rs569214 | 8 | 27487790 | G | T | intergenic | CLU;SCARA3 | YES |
| rs59007384 | 19 | 45396665 | G | T | intronic | TOMM40 | YES |
| rs6088727 | 20 | 33713639 | A | G | intronic | EDEM2 | YES |
| rs610932 | 11 | 59939307 | G | T | UTR3 | MS4A6A | YES |
| rs611267 | 11 | 60005268 | G | A | intronic | MS4A4E | YES |
| rs61510607 | 7 | 155224715 | T | C | intergenic | LOC100286906;EN2 | YES |
| rs61812598 | 1 | 154420087 | G | A | intronic | IL6R | YES |
| rs624290 | 9 | 3928115 | T | C | intronic | GLIS3 | YES |
| rs639012 | 11 | 85681583 | G | A | intronic | PICALM | YES |
| rs6431223 | 2 | 127895487 | G | A | intergenic | BIN1;CYP27C1 | YES |
| rs6448799 | 4 | 11630049 | C | T | intergenic | HS3ST1;LINC02360 | YES |
| rs6701713 | 1 | 207786289 | G | A | intronic | CR1 | YES |
| rs6834555 | 4 | 10062326 | A | G | intergenic | SLC2A9;WDR1 | YES |
| rs6857 | 19 | 45392254 | C | T | UTR3 | PVRL2 | YES |
| rs6859 | 19 | 45382034 | G | A | UTR3 | PVRL2 | YES |
| rs690705 | 13 | 34654918 | A | G | intergenic | RFC3;LINC02343 | YES |
| rs6922617 | 6 | 41336101 | G | A | intergenic | NCR2;LINC01276 | YES |
| rs7009219 | 8 | 53214265 | C | T | intronic | ST18 | YES |
| rs7011581 | 8 | 9000465 | G | A | intronic | PPP1R3B | YES |
| rs7081208 | 10 | 13991865 | G | A | intronic | FRMD4A | YES |
| rs71352238 | 19 | 45394336 | T | C | upstream | TOMM40 | YES |
| rs714948 | 19 | 45165912 | C | A | UTR3 | PVR | YES |
| rs7225151 | 17 | 5137047 | G | A | ncRNA\_intronic | LOC100130950 | YES |
| rs7259620 | 19 | 45407788 | G | A | downstream | TOMM40 | YES |
| rs7274581 | 20 | 55018260 | T | C | intronic | CASS4 | YES |
| rs72807343 | 5 | 179238261 | C | T | intronic | SQSTM1 | YES |
| rs7295246 | 12 | 43967677 | T | G | intergenic | ADAMTS20;PUS7L | YES |
| rs7412 | 19 | 45412079 | C | T | exonic | APOE | YES |
| rs741780 | 19 | 45404431 | T | C | intronic | TOMM40 | YES |
| rs744373 | 2 | 127894615 | A | G | intergenic | BIN1;CYP27C1 | YES |
| rs74495807 | 7 | 34088535 | C | T | intronic | BMPER | YES |
| rs749005 | 6 | 6283666 | G | T | intronic | F13A1 | YES |
| rs753129 | 4 | 56668431 | A | G | intergenic | NMU;LOC644145 | YES |
| rs7561528 | 2 | 127889637 | G | A | intergenic | BIN1;CYP27C1 | YES |
| rs75617873 | 22 | 44526105 | A | C | intronic | PARVB | YES |
| rs760136 | 19 | 45403858 | A | G | intronic | TOMM40 | YES |
| rs769449 | 19 | 45410002 | G | A | intronic | APOE | YES |
| rs7818382 | 8 | 96054000 | C | T | intronic | NDUFAF6 | YES |
| rs7920721 | 10 | 11720308 | A | G | intergenic | USP6NL;ECHDC3 | YES |
| rs7933202 | 11 | 59936926 | A | C | intergenic | MS4A2;MS4A6A | YES |
| rs79335261 | 12 | 56497641 | G | T | upstream;downstream | PA2G4;ERBB3 | YES |
| rs8035452 | 15 | 51040798 | T | C | intronic | SPPL2A | YES |
| rs8105340 | 19 | 45367777 | T | C | intronic | PVRL2 | YES |
| rs8106922 | 19 | 45401666 | A | G | intronic | TOMM40 | YES |
| rs9302457 | 16 | 11059837 | G | A | intronic | CLEC16A | YES |
| rs9331888 | 8 | 27468862 | C | G | intronic | CLU | YES |
| rs9331896 | 8 | 27467686 | T | C | intronic | CLU | YES |
| Affx-13926609 | 17 | 43881790 | C | T | intronic | CRHR1;LINC02210-CRHR1 | NO |
| Affx-18180474 | 21 | 48022230 | G | C | exonic | S100B | NO |
| Affx-28240779 | 6 | 18222102 | - | TT | intronic | KDM1B | NO |
| Affx-5731786 | 11 | 67200819 | C | T | exonic | RPS6KB2 | NO |
| rs10081503 | 8 | 4639285 | A | G | intronic | CSMD1 | NO |
| rs10088378 | 8 | 3265590 | T | C | exonic | CSMD1 | NO |
| rs10116775 | 9 | 112406611 | G | A | intronic | PALM2 | NO |
| rs10134607 | 14 | 93129070 | G | A | intronic | RIN3 | NO |
| rs10141024 | 14 | 94391699 | A | G | exonic | FAM181A | NO |
| rs10170021 | 2 | 170295503 | C | T | intergenic | LRP2;BBS5 | NO |
| rs1021108 | 8 | 3488764 | T | C | intronic | CSMD1 | NO |
| rs10262722 | 7 | 29540205 | T | G | intronic | CHN2 | NO |
| rs10408984 | 19 | 15298270 | G | A | intronic | NOTCH3 | NO |
| rs10415811 | 19 | 11177408 | T | C | intergenic | SMARCA4;LDLR | NO |
| rs10420355 | 19 | 15252881 | A | G | intergenic | ILVBL;NOTCH3 | NO |
| rs10422888 | 19 | 45453763 | A | G | downstream | APOC2;APOC4-APOC2 | NO |
| rs1043996 | 19 | 15295134 | A | G | exonic | NOTCH3 | NO |
| rs1044009 | 19 | 15271771 | A | G | exonic | NOTCH3 | NO |
| rs1045510 | 8 | 27255128 | T | C | exonic | PTK2B | NO |
| rs1045511 | 8 | 27255146 | G | A | exonic | PTK2B | NO |
| rs10467864 | 14 | 93006902 | T | G | intronic | RIN3 | NO |
| rs10501087 | 11 | 27670108 | T | C | ncRNA\_intronic | BDNF-AS | NO |
| rs1054080 | 14 | 104095701 | C | A | UTR5 | KLC1 | NO |
| rs1057128 | 11 | 2797237 | G | A | exonic | KCNQ1 | NO |
| rs1065489 | 1 | 196709774 | G | T | exonic | CFH | NO |
| rs10750931 | 11 | 60059810 | A | G | exonic | MS4A4A | NO |
| rs10767076 | 11 | 23530869 | G | A | intergenic | MIR8054;LUZP2 | NO |
| rs10785450 | 12 | 43981503 | T | C | intergenic | ADAMTS20;PUS7L | NO |
| rs10790434 | 11 | 121191782 | C | G | intergenic | SC5D;SORL1 | NO |
| rs10834419 | 11 | 24690002 | A | C | intronic | LUZP2 | NO |
| rs11028503 | 11 | 25211011 | G | C | intergenic | LUZP2;ANO3 | NO |
| rs11028630 | 11 | 25376205 | A | G | intergenic | LUZP2;ANO3 | NO |
| rs11085016 | 19 | 3372514 | G | A | intronic | NFIC | NO |
| rs11085020 | 19 | 3446107 | T | C | intronic | NFIC | NO |
| rs11102924 | 1 | 115867271 | A | G | intronic | NGF | NO |
| rs11234568 | 11 | 85876222 | A | G | intergenic | PICALM;EED | NO |
| rs112369693 | 1 | 207676604 | T | C | intronic | CR1 | NO |
| rs11593273 | 10 | 13812143 | A | G | intronic | FRMD4A | NO |
| rs11603136 | 11 | 85874322 | A | C | intergenic | PICALM;EED | NO |
| rs11627027 | 14 | 94370417 | C | T | downstream | FAM181A-AS1 | NO |
| rs11669705 | 19 | 3423339 | G | T | intronic | NFIC | NO |
| rs11671513 | 19 | 3429937 | G | A | intronic | NFIC | NO |
| rs11702494 | 21 | 46876306 | G | A | exonic | COL18A1 | NO |
| rs11736797 | 4 | 183778788 | T | C | intergenic | TENM3;DCTD | NO |
| rs12304486 | 12 | 10290338 | A | G | intergenic | CLEC7A;OLR1 | NO |
| rs1235382 | 19 | 45019507 | T | C | intronic | CEACAM20 | NO |
| rs12419105 | 11 | 24843020 | A | G | intronic | LUZP2 | NO |
| rs12620324 | 2 | 227950021 | C | T | intronic | COL4A4 | NO |
| rs12673065 | 7 | 33923564 | A | G | intergenic | BBS9;BMPER | NO |
| rs12981961 | 19 | 45009977 | C | T | downstream | CEACAM20 | NO |
| rs12982299 | 19 | 53967873 | A | C | intergenic | ZNF765-ZNF761;ZNF813 | NO |
| rs12983082 | 19 | 11216561 | A | C | intronic | LDLR | NO |
| rs13002844 | 2 | 40072092 | G | T | intergenic | THUMPD2;SLC8A1-AS1 | NO |
| rs1405655 | 19 | 50882619 | T | C | intronic | NR1H2 | NO |
| rs1427477 | 7 | 34021434 | C | T | intronic | BMPER | NO |
| rs1449701 | 11 | 110574690 | G | A | intronic | ARHGAP20 | NO |
| rs1467558 | 11 | 35229673 | C | T | exonic | CD44 | NO |
| rs16867576 | 5 | 88746331 | A | G | ncRNA\_intronic | MEF2C-AS1 | NO |
| rs17005935 | 2 | 70835618 | C | T | intergenic | TGFA;ADD2 | NO |
| rs17215782 | 10 | 84974074 | T | C | intergenic | NRG3;LOC105378397 | NO |
| rs17259045 | 1 | 207782707 | A | G | exonic | CR1 | NO |
| rs1727757 | 19 | 45064435 | G | A | intergenic | CEACAM22P;IGSF23 | NO |
| rs17368528 | 1 | 9324213 | C | T | exonic | H6PD | NO |
| rs17848190 | 2 | 170029800 | T | C | intronic | LRP2 | NO |
| rs179486 | 11 | 2590497 | G | A | intronic | KCNQ1 | NO |
| rs1799794 | 14 | 104179267 | T | C | UTR5 | XRCC3 | NO |
| rs1799898 | 19 | 11227554 | C | T | exonic | LDLR | NO |
| rs1799969 | 19 | 10394792 | G | A | exonic | ICAM1 | NO |
| rs1805096 | 1 | 66102257 | G | A | exonic | LEPR | NO |
| rs1871045 | 19 | 45326768 | C | T | intergenic | BCAM;PVRL2 | NO |
| rs1880560 | 7 | 143198905 | G | T | ncRNA\_intronic | EPHA1-AS1 | NO |
| rs1884082 | 14 | 95078677 | T | G | upstream | SERPINA3 | NO |
| rs1950902 | 14 | 64882380 | G | A | exonic | MTHFD1 | NO |
| rs1997606 | 21 | 30257560 | G | C | exonic | N6AMT1 | NO |
| rs201925102 | 19 | 3372704 | - | T | intronic | NFIC | NO |
| rs2028373 | 12 | 8757481 | A | G | exonic | AICDA | NO |
| rs203717 | 19 | 45139812 | C | T | UTR3 | IGSF23 | NO |
| rs2045405 | 4 | 183313353 | T | C | intronic | TENM3 | NO |
| rs2071430 | 21 | 42798065 | G | T | UTR5 | MX1 | NO |
| rs2074977 | 19 | 3434028 | A | C | intronic | NFIC | NO |
| rs2103262 | 19 | 45029894 | T | C | intronic | CEACAM20 | NO |
| rs2116898 | 19 | 11241810 | G | A | intronic | LDLR | NO |
| rs216616 | 2 | 80804156 | A | G | intronic | CTNNA2 | NO |
| rs2237214 | 6 | 16405185 | C | T | intronic | ATXN1 | NO |
| rs2243191 | 1 | 207015957 | C | T | exonic | IL19 | NO |
| rs2248407 | 11 | 77937800 | G | A | exonic | GAB2 | NO |
| rs2253563 | 19 | 44983567 | C | G | exonic | ZNF180 | NO |
| rs2272522 | 3 | 361508 | C | T | exonic | CHL1 | NO |
| rs2292245 | 3 | 62189189 | G | A | exonic | PTPRG | NO |
| rs2303973 | 11 | 14264916 | A | G | exonic | SPON1 | NO |
| rs2365314 | 12 | 60217809 | G | A | intergenic | SLC16A7;FAM19A2 | NO |
| rs2392252 | 7 | 33684775 | C | T | intergenic | BBS9;BMPER | NO |
| rs2402374 | 14 | 94389292 | C | T | ncRNA\_intronic | FAM181A-AS1 | NO |
| rs2510044 | 11 | 77909014 | G | A | exonic | USP35 | NO |
| rs2511188 | 11 | 77920930 | A | G | exonic | USP35 | NO |
| rs2569559 | 19 | 11214533 | T | G | intronic | LDLR | NO |
| rs2584922 | 2 | 24665131 | C | T | intergenic | ITSN2;NCOA1 | NO |
| rs2617101 | 8 | 4462550 | G | C | intronic | CSMD1 | NO |
| rs2631433 | 11 | 24749376 | G | T | intronic | LUZP2 | NO |
| rs2644620 | 1 | 156861168 | T | C | intergenic | NTRK1;PEAR1 | NO |
| rs2700753 | 8 | 5886134 | A | G | intergenic | CSMD1;LOC100287015 | NO |
| rs2738446 | 19 | 11227326 | C | G | intronic | LDLR | NO |
| rs2738458 | 19 | 11237296 | T | C | intronic | LDLR | NO |
| rs2740953 | 8 | 3868458 | G | A | intronic | CSMD1 | NO |
| rs2782643 | 1 | 43886494 | C | T | exonic | SZT2 | NO |
| rs281437 | 19 | 10397238 | C | T | UTR3 | ICAM1 | NO |
| rs2883927 | 1 | 240608299 | T | G | intronic | FMN2 | NO |
| rs290186 | 11 | 85395652 | A | G | UTR3 | CCDC89 | NO |
| rs2942164 | 17 | 43721283 | G | C | ncRNA\_intronic | LINC02210 | NO |
| rs2965160 | 19 | 45196135 | A | G | intergenic | CEACAM19;CEACAM16 | NO |
| rs2975359 | 8 | 3728020 | T | C | intronic | CSMD1 | NO |
| rs3130311 | 6 | 32217367 | A | G | intergenic | NOTCH4;LOC101929163 | NO |
| rs322143 | 19 | 11425097 | T | C | intronic | TSPAN16 | NO |
| rs34257217 | 20 | 10326528 | G | A | intergenic | SNAP25;MKKS | NO |
| rs34389960 | 8 | 4660974 | G | A | intronic | CSMD1 | NO |
| rs350751 | 2 | 52870277 | T | C | intergenic | LOC730100;MIR4431 | NO |
| rs35494829 | 3 | 67426281 | T | C | exonic | SUCLG2 | NO |
| rs35817328 | 19 | 45201027 | A | G | intergenic | CEACAM19;CEACAM16 | NO |
| rs35891370 | 19 | 45519100 | A | G | intronic | RELB | NO |
| rs3729640 | 19 | 45381917 | C | T | UTR3 | PVRL2 | NO |
| rs3742717 | 14 | 93118668 | C | T | exonic | RIN3 | NO |
| rs3752109 | 19 | 59082605 | C | T | exonic | MZF1 | NO |
| rs3764645 | 19 | 1042809 | A | G | exonic | ABCA7 | NO |
| rs3786507 | 19 | 45548255 | A | G | intronic | CLASRP | NO |
| rs3787283 | 20 | 10284418 | A | G | intronic | SNAP25 | NO |
| rs3816921 | 11 | 12246233 | G | C | exonic | MICAL2 | NO |
| rs3820198 | 1 | 53792651 | A | C | exonic | LRP8 | NO |
| rs3829946 | 14 | 95651589 | T | C | UTR3 | CLMN | NO |
| rs3852856 | 19 | 45361574 | G | A | intronic | PVRL2 | NO |
| rs3856806 | 3 | 12475557 | C | T | exonic | PPARG | NO |
| rs3859564 | 19 | 3403004 | G | A | intronic | NFIC | NO |
| rs3862611 | 11 | 121831524 | G | A | intergenic | SORL1;MIR100HG | NO |
| rs3925162 | 19 | 45639716 | A | G | intronic | PPP1R37 | NO |
| rs4438709 | 4 | 5089676 | A | G | intronic | STK32B | NO |
| rs4601775 | 11 | 23784536 | G | T | intergenic | MIR8054;LUZP2 | NO |
| rs466448 | 21 | 27544108 | A | G | upstream | APP | NO |
| rs469390 | 21 | 42817930 | A | G | exonic | MX1 | NO |
| rs4806928 | 19 | 3397431 | A | G | intronic | NFIC | NO |
| rs4861500 | 4 | 183160086 | C | A | intergenic | MIR1305;TENM3 | NO |
| rs4882709 | 12 | 129101576 | C | A | intronic | TMEM132C | NO |
| rs4903233 | 14 | 74963994 | G | C | downstream | LTBP2 | NO |
| rs4937373 | 11 | 110643147 | T | C | intergenic | ARHGAP20;C11orf53 | NO |
| rs4944550 | 11 | 85685574 | C | T | intronic | PICALM | NO |
| rs537160 | 6 | 31916400 | G | A | intronic | CFB | NO |
| rs58180668 | 15 | 87387247 | G | A | intronic | AGBL1 | NO |
| rs6006622 | 22 | 44424108 | T | C | intronic | PARVB | NO |
| rs60106607 | 7 | 29245302 | A | T | ncRNA\_intronic | LOC102724484 | NO |
| rs6032826 | 20 | 10203817 | A | G | intronic | SNAP25 | NO |
| rs60718666 | 19 | 3393376 | A | G | intronic | NFIC | NO |
| rs62000962 | 21 | 46910210 | G | A | exonic | COL18A1 | NO |
| rs62171294 | 2 | 170267086 | C | G | intergenic | LRP2;BBS5 | NO |
| rs622657 | 7 | 103391003 | A | G | intronic | RELN | NO |
| rs62472729 | 7 | 143116061 | G | C | ncRNA\_intronic | EPHA1-AS1 | NO |
| rs62620047 | 3 | 61975382 | T | C | exonic | PTPRG | NO |
| rs6484069 | 11 | 24728006 | G | A | intronic | LUZP2 | NO |
| rs6558702 | 8 | 2909992 | A | G | exonic | CSMD1 | NO |
| rs667595 | 8 | 2832139 | G | C | exonic | CSMD1 | NO |
| rs667859 | 8 | 2820745 | G | C | exonic | CSMD1 | NO |
| rs67149430 | 19 | 53997179 | T | C | UTR3 | ZNF813 | NO |
| rs6996296 | 8 | 4565946 | T | C | intronic | CSMD1 | NO |
| rs7072125 | 10 | 11684042 | G | A | intergenic | USP6NL;ECHDC3 | NO |
| rs7109593 | 11 | 24722990 | A | G | intronic | LUZP2 | NO |
| rs7144273 | 14 | 92920371 | T | C | exonic | SLC24A4 | NO |
| rs7144561 | 14 | 94376460 | T | C | ncRNA\_intronic | FAM181A-AS1 | NO |
| rs715284 | 2 | 81663984 | G | A | intergenic | CTNNA2;LINC01815 | NO |
| rs7155685 | 14 | 94388642 | C | T | ncRNA\_exonic | FAM181A-AS1 | NO |
| rs716066 | 11 | 121216567 | G | A | intergenic | SC5D;SORL1 | NO |
| rs7217331 | 17 | 47603719 | G | A | ncRNA\_intronic | LOC100288866 | NO |
| rs7247952 | 19 | 35246250 | C | T | intergenic | ZNF181;ZNF599 | NO |
| rs7259278 | 19 | 11227466 | G | T | intronic | LDLR | NO |
| rs7260180 | 19 | 45028169 | T | C | exonic | CEACAM20 | NO |
| rs72624064 | 8 | 3892598 | A | G | intronic | CSMD1 | NO |
| rs7325541 | 13 | 106408200 | G | A | ncRNA\_intronic | LINC00343 | NO |
| rs73350477 | 14 | 100194128 | T | C | downstream | CYP46A1 | NO |
| rs7523654 | 1 | 115861265 | C | T | intronic | NGF | NO |
| rs7681423 | 4 | 155542248 | C | T | intergenic | FGG;LRAT | NO |
| rs77166016 | 19 | 3371522 | T | C | intronic | NFIC | NO |
| rs7810666 | 7 | 29410168 | C | T | intronic | CHN2 | NO |
| rs78367292 | 8 | 4640437 | T | C | intronic | CSMD1 | NO |
| rs78455237 | 7 | 33911970 | A | C | intergenic | BBS9;BMPER | NO |
| rs79216212 | 7 | 16756053 | T | C | intergenic | BZW2;TSPAN13 | NO |
| rs7939652 | 11 | 23739454 | A | C | intergenic | MIR8054;LUZP2 | NO |
| rs79724424 | 14 | 94390867 | A | G | ncRNA\_intronic | FAM181A-AS1 | NO |
| rs8012955 | 14 | 94380907 | T | C | ncRNA\_intronic | FAM181A-AS1 | NO |
| rs8100718 | 19 | 45016116 | A | G | exonic | CEACAM20 | NO |
| rs8104483 | 19 | 45372354 | T | G | intronic | PVRL2 | NO |
| rs846859 | 19 | 45129772 | A | G | intronic | IGSF23 | NO |
| rs896412 | 19 | 11444131 | C | G | intronic | RAB3D | NO |
| rs902321 | 3 | 67411166 | A | G | exonic | SUCLG2 | NO |
| rs9302654 | 16 | 54009545 | C | T | intronic | FTO | NO |
| rs930461 | 19 | 45043259 | T | A | ncRNA\_exonic | CEACAM22P | NO |
| rs9314348 | 8 | 27345426 | T | C | intergenic | CHRNA2;EPHX2 | NO |
| rs941650 | 14 | 92909073 | C | T | exonic | SLC24A4 | NO |
| rs944019 | 13 | 20917309 | G | A | intergenic | GJB6;CRYL1 | NO |
| rs94517 | 19 | 11407315 | C | T | intronic | TSPAN16 | NO |
| rs9644382 | 8 | 4913711 | G | A | intergenic | CSMD1;LOC100287015 | NO |
| rs9704394 | 11 | 13997399 | G | A | intronic | SPON1 | NO |
| rs974680 | 21 | 27011112 | G | A | upstream | JAM2 | NO |
| rs9989716 | 19 | 1007109 | A | G | intronic | GRIN3B | NO |

**Supplementary Table 2.**SNVs from DisGenet used as AD predictors in this study. We report chromosome (Chr), position (hg19 Postition), reference allele (Ref), alternative allele (Alt), genetic region (Region) and gene ID (Gene Symbol). SNVs used as reference predictors are also included (column “DisGeNet AD SNV” = NO).

Supplementary Table



**Supplementary Table 3.** Comparison of allele frequencies (AF) in UKB, ADNI3, in general population (gnomAD), and in UKB official reports to discard the presence of major biases in the data. The AF of the 14 SNVs used in the tests of pairwise interactions (*Figure 3*) was contrasted across datasets. When comparing differences in AF between UKB (column “AF UKB”) and ADNI (column “AF ADNI”), there were 5 SNVs with statistically significant differences in AF between both datasets (column “Fisher pVal AF” < 0.01). However, when comparing the AF in AD and controls separately, there were no significative differences between both datasets (columns “Fisher pVal AF.ad” and “Fisher pVal AF.cntrl”). Therefore, part of the differences observed when comparing the full datasets are probably due to the different proportion of AD and controls in UKB and ADNI. All the SNVs in the table passed the Hardy-Weinberg Equilibrium test (p-value > 10E-4) in both datasets (columns “HW.pVal UKB” and “HW.pVal ADNI”). The zygosity or proportion of the reference and alternative allele pairs reported in the UKB samples selected in this study was highly similar with the one reported in the UKB genomic search resource49 (columns “F.xx UKB Official”), excluding a major bias originated in sampling or pre-processing steps.

Supplementary Table

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Performances on the training set** | | | | | | | |
|  | **Accuracy** | **AUC-ROC** | **F Score** | **Sensitivity** | **Specificity** | **PPV** | **NPV** |
| **GB** | 0.84 ± 0.12 | 0.90 ± 0.15 | 0.83 ± 0.13 | 0.80 ± 0.13 | 0.87 ± 0.12 | 0.86 ± 0.13 | 0.82 ± 0.11 |
| **ET** | 0.79 ± 0.09 | 0.89 ± 0.08 | 0.78 ± 0.09 | 0.76 ± 0.10 | 0.83 ± 0.08 | 0.81 ± 0.08 | 0.78 ± 0.08 |
| **RF** | 0.71 ± 0.05 | 0.81 ± 0.06 | 0.69 ± 0.05 | 0.65 ± 0.06 | 0.76 ± 0.04 | 0.73 ± 0.05 | 0.69 ± 0.05 |
| **Performances on the test set** | | | | | | | |
|  | **Accuracy** | **AUC-ROC** | **F Score** | **Sensitivity** | **Specificity** | **PPV** | **NPV** |
| **GB** | 0.76 ± 0.12 | 0.83 ± 0.15 | 0.75 ± 0.13 | 0.71 ± 0.12 | 0.81 ± 0.12 | 0.79 ± 0.13 | 0.74 ± 0.11 |
| **ET** | 0.73 ± 0.05 | 0.82 ± 0.04 | 0.71 ± 0.05 | 0.68 ± 0.05 | 0.78 ± 0.05 | 0.75 ± 0.05 | 0.71 ± 0.05 |
| **RF** | 0.70 ± 0.03 | 0.75 ± 0.05 | 0.68 ± 0.04 | 0.63 ± 0.04 | 0.77 ± 0.05 | 0.73 ± 0.05 | 0.68 ± 0.03 |

**Supplementary Table 4.**Nested cross-validation (CV) to evaluate overfitting of ML models. In order to estimate the generalized performance of the ML models and to measure overfitting, nested CV with k=5 folds in the outer loop was applied. Hyperparameter selection in the inner loop through a grid-search and a 10-fold CV was used, and the best configuration was obtained based on AUC-ROC and f-score metrics in the validation set. The hyperparameters selected for the final models were obtained using the same selection procedure adopted in the inner CV. The average values and standard deviation obtained in both training and test set in the nested CV are provided. Slight differences between training and test performances in all the metrics indicate the lack of considerable overfitting in the machine learning models. In this regard, differences between training and test set fall inside one standard deviation from the corresponding estimated values.

Supplementary Table







**Supplementary Table 5.**SNVs prioritized by GB, ET and RF models (total of 9, 20 and 15, respectively).  We report SNV ID (SNV), gene ID (Gene), genetic region (Region), chromosome (Chr), position (hg19 Position), allele frequency in Alzheimer’s Disease (AF AD) and allele frequency in control group (AF Control) in testing and training sets (AF test AD, AF test Control, AF train AD, AF train Control) and the corresponding statistical metrics (LOG2 AF AD/Cntrl, FI and Fisher p-value).

Supplementary Table

|  |  |  |
| --- | --- | --- |
| **SNVs** | **UKB HMP** | **ADNI HMP** |
| rs1160985\_rs405509 | 2.29E-137 | 0.560606906 |
| rs405509\_rs769449 | 4.19E-68 | 0.030960105 |
| rs71352238\_rs405509 | 4.93E-67 | 0.107087216 |
| rs157580\_rs405509 | 1.26E-65 | 0.425607113 |
| rs2075650\_rs405509 | 2.20E-64 | 0.095557151 |
| rs405509\_rs439401 | 1.03E-44 | 0.003556346 |
| rs6857\_rs405509 | 7.91E-36 | 0.027156421 |
| rs405509\_rs429358 | 1.85E-31 | 0.050175991 |
| rs59007384\_rs405509 | 6.21E-23 | 0.01801069 |
| rs405509\_rs7412 | 1.18E-22 | 0.999693465 |
| rs157582\_rs405509 | 3.17E-21 | 0.020315139 |
| rs405509\_rs4420638 | 2.35E-16 | 0.004391476 |
| rs157582\_rs1160985 | 1.27E-06 | 0.002237019 |
| rs59007384\_rs1160985 | 5.81E-06 | 0.002139991 |
| rs59007384\_rs7412 | 0.000494401 | 0.999253877 |
| rs1160985\_rs439401 | 0.001028552 | 0.001800618 |
| rs157582\_rs7412 | 0.002368163 | 0.999522227 |
| rs1160985\_rs4420638 | 0.002723365 | 0.00021184 |
| rs7561528\_rs439401 | 0.008204021 | 0.239664455 |
| rs157582\_rs4420638 | 0.015417011 | 0.983591603 |
| rs7561528\_rs157580 | 0.024285469 | 0.016728113 |
| rs59007384\_rs4420638 | 0.024479659 | 0.995244944 |
| rs71352238\_rs157580 | 0.031010608 | 0.026539527 |
| rs157580\_rs2075650 | 0.044776731 | 0.025769427 |
| rs71352238\_rs1160985 | 0.05967868 | 0.002372651 |
| rs157580\_rs157582 | 0.070847303 | 0.029143402 |
| rs157580\_rs59007384 | 0.072012863 | 0.025553556 |
| rs6857\_rs157580 | 0.07607867 | 0.046063806 |
| rs2075650\_rs1160985 | 0.082633995 | 0.002319076 |
| rs7561528\_rs6857 | 0.082669372 | 0.742136259 |
| rs157582\_rs439401 | 0.087556469 | 0.255748508 |
| rs6857\_rs157582 | 0.093820988 | 0.972750872 |
| rs6857\_rs769449 | 0.097469545 | 0.831514667 |
| rs2075650\_rs439401 | 0.101222764 | 0.51844485 |
| rs429358\_rs4420638 | 0.101947977 | 0.998492405 |
| rs157582\_rs769449 | 0.109783631 | 0.643311566 |
| rs2075650\_rs7412 | 0.113614937 | 0.935963621 |
| rs71352238\_rs439401 | 0.125035527 | 0.223088528 |
| rs157580\_rs439401 | 0.133178462 | 0.031286379 |
| rs6857\_rs7412 | 0.139174288 | 0.985355805 |
| rs157582\_rs59007384 | 0.144969246 | 0.985995437 |
| rs7561528\_rs769449 | 0.145717995 | 0.183436153 |
| rs769449\_rs4420638 | 0.16582046 | 0.901403807 |
| rs7561528\_rs4420638 | 0.167446201 | 0.064971325 |
| rs6857\_rs1160985 | 0.190002844 | 0.003089704 |
| rs7561528\_rs429358 | 0.196336355 | 0.037883375 |
| rs71352238\_rs157582 | 0.216360064 | 0.592254407 |
| rs7561528\_rs59007384 | 0.223433009 | 0.824247014 |
| rs1160985\_rs7412 | 0.232861439 | 0.999898157 |
| rs2075650\_rs157582 | 0.262085551 | 0.557750172 |
| rs7561528\_rs71352238 | 0.262380682 | 0.748051804 |
| rs157580\_rs1160985 | 0.321264294 | 0.884838864 |
| rs6857\_rs2075650 | 0.337036681 | 0.736094361 |
| rs157582\_rs429358 | 0.343319092 | 0.993933251 |
| rs7561528\_rs2075650 | 0.376335722 | 0.63371156 |
| rs157580\_rs7412 | 0.380029365 | 0.99987313 |
| rs7412\_rs439401 | 0.388808821 | 0.999254011 |
| rs71352238\_rs7412 | 0.422939273 | 0.935964544 |
| rs6857\_rs439401 | 0.437921788 | 0.838836483 |
| rs7561528\_rs1160985 | 0.466443336 | 0.072612673 |
| rs59007384\_rs769449 | 0.484899306 | 0.754131706 |
| rs7561528\_rs157582 | 0.556916541 | 0.835424215 |
| rs157580\_rs769449 | 0.737324676 | 0.01255412 |
| rs59007384\_rs439401 | 0.759580568 | 0.557479368 |
| rs769449\_rs439401 | 0.838844659 | 0.855362147 |
| rs157580\_rs4420638 | 0.953647105 | 0.010409405 |
| rs59007384\_rs429358 | 0.958540097 | 0.995693919 |
| rs769449\_rs429358 | 0.967606966 | 0.950973052 |
| rs7412\_rs4420638 | 0.970958988 | 0.988440351 |
| rs2075650\_rs59007384 | 0.971862958 | 0.50201013 |
| rs7561528\_rs7412 | 0.97617084 | 0.999754317 |
| rs71352238\_rs2075650 | 0.993881488 | 0.656180779 |
| rs6857\_rs59007384 | 0.994381795 | 0.966990023 |
| rs71352238\_rs59007384 | 0.997976272 | 0.261259969 |
| rs71352238\_rs4420638 | 0.998717769 | 0.82039613 |
| rs71352238\_rs769449 | 0.998812849 | 0.944273953 |
| rs2075650\_rs769449 | 0.999386355 | 0.9646197 |
| rs6857\_rs71352238 | 0.999416169 | 0.580637875 |
| rs6857\_rs429358 | 0.999614465 | 0.999119268 |
| rs769449\_rs7412 | 0.999819219 | 0.916857447 |
| rs2075650\_rs429358 | 0.99985407 | 0.941925422 |
| rs6857\_rs4420638 | 0.999929831 | 0.991329488 |
| rs71352238\_rs429358 | 0.999931686 | 0.872571519 |
| rs429358\_rs439401 | 0.999987665 | 0.715038946 |
| rs2075650\_rs4420638 | 0.999988249 | 0.963603767 |
| rs157580\_rs429358 | 0.999989858 | 0.087219709 |
| rs1160985\_rs429358 | 0.999999891 | 0.004603853 |
| rs429358\_rs7412 | 0.999999908 | 0.988440511 |
| rs1160985\_rs769449 | 0.999999933 | 0.000841927 |
| rs439401\_rs4420638 | 0.999999986 | 0.532560915 |
| rs7561528\_rs405509 | 1 | 0.018935427 |
|  |  |  |

**Supplementary Table 6.** Pairwise test of interactions of the 14 SNVs prioritized by any of the three ML models and commonly present in UKB and ADNI3 arrays. Comparisons were made between the full glm built with two SNVs considering the individual effect and interactions to predict AD and controls, and the reduced glm only considering the individual effect of the two SNVs to predict both conditions. The differences between the full and reduced glm were measured with an anova and the corresponding chi-squared test p-value was obtained for each comparison. In the case of UKB, for each pairwise set of variants the test was applied 1000 times, randomly selecting 500 AD and 500 controls in each iteration. In ADNI 100 iterations were applied subsampling 10 AD and 10 controls. The asymptotically exact harmonic mean p-value (HMP) was used to summarize all the p-values obtained across the iterations and to correct for multiple comparisons. A cut-off HMP < 0.01 was used to consider an interaction as statistically significant. Interactions with rs1160985 or rs405509 are highlighted in red. These two SNVs: 1) had high FI scores in the ML models, 2) had low AF differences between AD and controls, 3) were involved in interaction patterns of the genomic profiles obtained with ML approaches.