Weighted burden analysis of exome-sequenced late onset Alzheimer's cases and controls provides further evidence for involvement of *PSEN1* and demonstrates protective role for variants in tyrosine phosphatase genes

David Curtis
d.curtis@ucl.ac.uk
00 44 7973 906 143

UCL Genetics Institute, UCL, Darwin Building, Gower Street, London WC1E 6BT.
Centre for Psychiatry, Queen Mary University of Lonodn, Charterhouse Square, London EC1M 6BQ.

Keywords
LOAD; PSEN1; tyrosine phosphatase; TIAF1; C1R; SCG3.

Summary

Previous studies have implicated common and rare genetic variants as risk factors for late onset Alzheimer's disease (LOAD). Here, weighted burden analysis was applied to over 10,000 exome sequenced subjects from the Alzheimer's Disease Sequencing Project. Analyses were carried out to investigate whether rare genetic variants predicted to have a functional effect were more commonly seen in cases or in controls. Confirmatory results were obtained for *TREM2*, *ABCA7* and *SORL1*. Additional support was provided for *PSEN1* ($p = 0.0002$), which previously had been only weakly implicated in LOAD. There was suggestive evidence that functional variants in *C1R* and *EXOC5* might increase risk and that variants in *TIAF1*, *GFRAL*, *SCG3* and *GADD45G* might have a protective effect. Overall, there was strong evidence ($p = 4 \times 10^{-7}$) that variants in tyrosine phosphatase genes reduce the risk of developing LOAD.

Introduction

A recent genetic meta-analysis of Late Onset Alzheimer's Disease (LOAD) to study the effects of common variants implicated Aβ, tau, immunity and lipid processing and reported a total of 25 genome-wise significant loci, including variants at or near *TREM2*, *ABCA7* and *SORL1*(Kunkle et al. 2019). Analyses of risk genes and pathways showed enrichment for rare variants which remained to be identified. The Alzheimer's Disease Sequencing Project (ADSP) has the aim of identifying rare genetic variants which increase or decrease the risk of Late Onset Alzheimer's Disease (LOAD) and a full description of its membership, support and methodologies is provided on its website (https://www.niagads.org/adsp/content/about) (Beecham et al. 2017). A number of analyses utilising this dataset have been published to date. One study incorporating this dataset with others identified 19 loss of function variants of *SORL1* in cases as opposed to 1 in controls and a collapsing test of loss of function ultra-rare variants highlighted other genes including *GRID2IP*, *WDR76* and *GRN* (Raghavan et al. 2018). A subsequent study used variant-based and gene-based tests of association and followed up significant or suggestive results in other samples to implicate three novel genes, *IGHG3*, *AC099552.4* and *ZNF655* as well as novel and predicted functional genetic variants in genes previously associated with Alzheimer's disease (AD). In the ADSP samples alone, five genes were exome-wide significant: *ABCA7*, *TREM2*, *CBLC*, *OPRL1* and *GAS2L2*. Additionally, three individual variants were exome-wide significant: rs75932628 in *TREM2* ($p = 4.8 \times 10^{-12}$), rs2405442 in *PILRA* ($p = 1.7 \times 10^{-7}$), and a novel variant at 7:154,988,675 *AC099552.4* ($p = 1.2 \times 10^{-7}$).

Another study focussed on a panel of 35 known dementia genes using variants thought likely to be consequential based on ClinVar, literature review and predicted effect (Blue et al. 2018). Using gene-based analyses in the ADSP case-control sample, *TREM2* was significant at $p = 1.8 \times 10^{-11}$ and *APOE* was significant at $p = 0.0069$ while *ARSA*, *CSF1R*, *PSEN1* and *MAPT* were significant at $p < 0.05$. In another study, 95 dementia associated genes were examined and it was reported that overall there was an excess of deleterious rare coding variants. Additionally, 10 cases and no controls carried in rs149307620, a missense variant in *NOTCH3*, and 4 cases and no controls carried rs104894002, a high-impact variant in *TREM2* which in homozygous form causes Nasu-Hakola disease.

A limitation of these previous studies is that they focus on specific genes and/or types of variant. We have previously described a method for a carrying out a weighted burden analysis of all variants within a gene, assigning more weights to variants which are rare and to variants predicted to have a functional effect (Curtis 2012; Curtis & UK10K Consortium 2016). This method was applied to the ADSP sample in the hope that it would more completely capture the effect of variants influencing susceptibility to LOAD.

Methods

Phenotype information and variant calls based on whole exome sequencing using the GRCh37 assembly for the Alzheimer's Disease Sequencing Project (ADSP) Distribution were downloaded from dbGaP (phs000572.v7.p4.c1-c6). As described previously, participants were at least 60 years old and of were mostly of European-American ancestry though with a small number of Caribbean Hispanic ancestry and all cases met NINCDS-ADRDA criteria for possible, probable or definite AD based on clinical assessment, or had presence of AD (moderate or high likelihood) upon neuropathology examination (Beecham et al. 2017). The cases were chosen whose AD risk score suggested that their disease was not well explained by age, sex or APOE status. Whole exome sequencing was carried out using standard methods as described previously (Bis et al. 2018) and on the ADSP website: https://www.niagads.org/adsp/content/sequencing-pipelines. The downloaded VCF files for single nucleotide variants and indels were merged into a single file and the PrevAD field was used to define phenotype, yielding 4,600 cases and 6,199 controls.

Each variant was annotated using VEP, PolyPhen and SIFT (McLaren et al. 2016; Adzhubei et al. 2013; Kumar et al. 2009). The same analytic process was used has had previously been applied to an exome sequenced sample of schizophrenia cases and controls (Curtis et al. 2018). Variants were weighted according to their functional annotation using the default weights provided with the GENEVARASSOC program, which was used to generate input files for weighted burden analysis by SCOREASSOC (Curtis 2012; Curtis 2016). For example, a weight of 5 was assigned for a synonymous variant, 10 for a non-synonymous variant and 20 for a stop gained variant. Additionally 10 was added to the weight if the PolyPhen annotation was possibly or probably damaging and also if the SIFT annotation was deleterious. The full set of weights is shown in Supplementary Table 1, copied from (Curtis et al. 2018). Since we reasoned that the effects of any common variants would have been well characterised by previous studies we excluded variants with MAF>0.01 in either cases or controls. SCOREASSOC weights rare variants more strongly than common ones but restricting attention to rare variants meant that this frequency-based weighting had little effect in practice. Variants were also excluded if they did not have a PASS in the information field or if there were more than 10% of genotypes missing or of low quality in either cases or controls or if the

heterozygote count was smaller than both homozygote counts in both cohorts. SCOREASSOC produces a score for each subject consisting of the sum of the weights for each variant that the subject carries. It then performs a t test to compare the average scores between cases and controls. Results are expressed as signed log p (SLP) which is the logarithm base 10 of the p value from this t test and is positive if the scores are higher in cases and negative if they are higher in controls.

As in the schizophrenia study, pathway analysis was performed using PATHWAYASSOC, which applies the same weighted burden analysis to all the variants seen within a set of genes rather than a single gene, which is easily done by simply summing the gene-wise scores for each subject for all the genes in the set (Curtis 2016). This was carried out for the 1454 "all GO gene sets, gene symbols" pathways downloaded from the Molecular Signatures Database at http://www.broadinstitute.org/gsea/msigdb/collections.jsp (Subramanian et al. 2005).

Results

The weighted burden tests evaluated 1,199,503 variants in 18,357 autosomal genes. Figure 1 shows the QQ plot of the observed SLPs against the distribution expected under the null hypothesis. It can be seen that for almost all genes the results conform well to the SLP=eSLP line. There is one marked outlier, consisting of *TREM2* with an SLP of 14.07, and a few other genes with more extremely positive and negative SLPs than would be expected under the null hypothesis. Table 1 shows results for all genes with an absolute value for the SLP of 3 (equivalent to p=0.001) or greater. 30 genes meet this criterion, while one would expect 18 by chance. Thus, some of those listed may represent those in which variants can affect susceptibility to LOAD.

Of genes with SLPs above 3, a number have previously been implicated by sequencing studies, consisting of *TREM2*, *ABCA7*, *SORL1* and *PSEN1* (Guerreiro et al. 2013; Blue et al. 2018; Patel et al. 2019; Kim 2018; Raghavan et al. 2018). Additionally, it has been reported that the R subcomponent of complement component 1, the product of *C1R*, is enriched amyloid plaques from patients with Alzheimer's disease compared to amyloid plaques from subjects without (Xiong et al. 2019). The gene which codes for the receptor for complement component 1, *CR1*, has been implicated as an LOAD gene through studies of common variants and of copy number variation (Kunkle et al. 2019; Brouwers et al. 2012). Although there is no direct evidence to implicate *EXOC5* in LOAD, in a mouse model cell-specific deletion of *Exoc5* results in progressive hearing loss associated with loss of hair cells and spiral ganglion neurons (Lee et al. 2018).

The gene with the most strongly negative SLP, *GFRAL*, codes for a member of the glial-derived neurotropic factor receptor α family which is a receptor for growth differentiation factor 15. This is implicated in biological processes including energy homeostasis and body weight regulation (Baek & Eling 2019). The fact that *TIAF1* produces an SLP of -3.74 is of interest because of a report that its product can self-aggregate and that polymerised TIAF1 physically interacts with amyloid fibrils, putatively promoting plaque formation (Lee et al. 2010). Although *SCG3*, which codes for secretogranin III, has not previously been associated with LOAD, secretogranin I is also known as chromogranin B and in LOAD patients about 10 to 20% of the amyloid-immunoreactive plaques have been reported to contain either chromogranin A, secretogranin I or secretoneurin (Marksteiner et al. 2002). There is a report that *GADD45G* gene expression is modulated by nuclear spheres, which are highly toxic aggregates in cell nuclei and which are found extensively LOAD neurons, with amyloid precursor protein playing a central role in their generation (Loosse et al. 2016; Kolbe et al. 2016).

SOX14 codes for a transcription factor which is required for normal neuronal development in the cerebellum but there seems to be little evidence for its involvement in LOAD (Prekop et al. 2018). *CACNA1A* has been reported to be downregulated in familial Alzheimer's disease due to the PS1-E280A mutation but has not been implicated in the pathogenesis of LOAD (Sepulveda-Falla et al. 2014).

In view of these results, the output files of a small number of other genes were examined. There was no difference in the weighted burden scores between cases and controls for *APP* (SLP=0.11) and there were no valid variants for *PSEN2* or *APOE.* Although *SGC3*, the gene for secretogranin III, had an increased burden of variants among controls there was no suggestion that the genes for secretogranin I and secretogranin II, named *CHGB* and *SCG2*, were associated with LOAD.

In general large numbers of variants which were very rare contributed to the observed SLPs and as each of these might occur in only one or two subjects one could not make any inference about their individual effect on risk. Table 2 shows the genotype counts for variants in the above-named genes of possible interest for those variants with a minor allele count of at least 10 and an odds ratio over 2, or, for the genes with SLP<3, less than 0.5. The variants in *TREM2* have been reported in a previous analysis this dataset (Song et al. 2017). Likewise, deleterious mutations in *ABCA7* are known to be risk factors for both early and late onset Alzheimer's disease (De Roeck et al. 2017). The variants in *SORL1* at 11:121348842 (rs140888526), 11:121421313 (rs148430425) and 11:121421361 (rs146438170) have not previously been implicated in LOAD susceptibility and although the gene-based SLP of 4.42 suggests that overall there is an excess of functionally significant variants among cases the numbers are too small to draw any definitive conclusions about any of these variants individually. The three qualifying variants listed for *C1R* are not predicted to impact on protein function. The variant in *EXOC5* at 14:57714397 (rs199912312) is predicted to be deleterious and probably damaging and is seen in 14 cases and only 3 controls but has not been reported to be associated with any phenotype. The variant in *TIAF1* at 17:27401061 (rs73986791) is not predicted to affect protein function and has not previously been reported. The two qualifying variants in *SCG3* are both synonymous so seem unlikely to have any functional effect.

In the gene set analysis using PATHWAYASSOC the highest SLP achieved was 2.84 for the PROTEIN_MATURATION set, which given that 1454 sets were tested is well within chance expectation. Examination of the output suggested that this was driven largely by the inclusion of *PSEN1* in this set and no other gene in it had an individual SLP higher than 1.3 (p=0.05). Table 3 lists the gene sets with SLP of 3 or less. The most negative is PROTEIN_TYROSINE_PHOSPHATASE_ACTIVITY with SLP=-6.48. This result is of interest because it indicates that controls are more likely to carry functionally significant variants than cases, suggesting that impairment of this activity might be protective against LOAD. There is an extensive literature on the role of tyrosine phosphatase in Alzheimer pathology and genetic reduction of striatal-enriched tyrosine phosphatase has been shown to reverse cognitive and cellular deficits in mouse model of AD (Hendriks et al. 2013; Zhang et al. 2010). Recent studies have shown that dephosphorylated Tau triggers a pro-inflammatory profile in microglia and that bergenin, a tyrosine phosphatase inhibitor, has neuroprotective effects in a rat model of AD (Perea et al. 2018; Barai et al. 2019). Table 4 lists the genes within this set which individually have SLP of -1.3 (equivalent to p=0.05) or less. It can be seen that the genes for protein tyrosine phosphatase, receptor types R, S, T and U all have SLPs of approximately -2. Most of the other gene sets which produced strongly negative SLPs did so because

they also contained tyrosine phosphatase genes. For those which did not, inspection of the outputs did not reveal any genes which seemed likely to be related to AD susceptibility.

Discussion

TREM2, *ABCA7* and *SORL1*, the three genes with the highest SLPs, are well-established susceptibility genes for LOAD, which inspires some confidence in the validity of the approach. Although a previous analysis of this dataset did find some support for the involvement of *PSEN1*, only specific variants were included and the results were significant only at $p<0.05$. The current analysis includes all rare variants and yields an SLP of 3.67, equivalent to $p=0.0002$. This provides stronger evidence and also suggests that a wider range of *PSEN1* variants may have an effect on LOAD risk. It is unclear whether the results for *C1R* and *EXOC5* reflect any influence on LOAD susceptibility or whether they have occurred simply by chance. Aside from the risk variants previously reported for *TREM2* and *ABCA7*, the current study is unable to implicate specific variants rather than genes.

Perhaps more interesting are the results for genes in which the SLP is strongly negative, implying that variants may protect against the development of LOAD. Perhaps the most plausible result is for *TIAF1* because of the claim that TIAF1 self-aggregates and leads to apoptosis, Aβ formation and plaque formation (Lee et al. 2010). If this is so then variants which interfered with this process would be expected to reduce LOAD risk. In terms of other individual genes, there is a possibility that variants in *GFRAL*, *SCG3* and/or *GADD45G* might have a protective effect but additional studies would be needed to settle this question.

Probably the most striking result is the evidence for a protective effect from variants in the set of genes encoding tyrosine phosphatases, which produces an SLP of 6.43 (equivalent to $p = 4 \times 10^{-7}$). Given the established role of tyrosine phosphatase processes in the development of AD this result is eminently plausible but has not been described previously. In particular, there is a suggestion that variants in the protein tyrosine phosphatases with receptor types R, S, T and U may be protective.

Analysing all variants together in a single joint analysis makes good use of the available data while avoiding problems due to multiple testing issues. The results obtained serve to highlight genes rather than individual variants, which tend to be too rare for individual inferences to be made. The results reported here serve to reinforce some known aspects of LOAD biology and to suggest additional leads which might profitably be pursued.

Acknowledgments

The author wishes to acknowledge the staff supporting the High Performance Computing Cluster, Computer Science Department, University College London.

The Alzheimer's Disease Sequencing Project (ADSP) is comprised of two Alzheimer's Disease (AD) genetics consortia and three National Human Genome Research Institute (NHGRI) funded Large Scale Sequencing and Analysis Centers (LSAC). The two AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institute of Health (NIH) institutes and other foreign governmental and non-governmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Schellenberg, Farrer, Pericak-Vance,

Mayeux, and Haines); U01AG049505 to Dr. Seshadri; U01AG049506 to Dr. Boerwinkle; U01AG049507 to Dr. Wijsman; and U01AG049508 to Dr. Goate and the Discovery Extension Phase analysis is supported through U01AG052411 to Dr. Goate, U01AG052410 to Dr. Pericak-Vance and U01 AG052409 to Drs. Seshadri and Fornage. Data generation and harmonization in the Follow-up Phases is supported by U54AG052427 (to Drs. Schellenberg and Wang).

The ADGC cohorts include: Adult Changes in Thought (ACT), the Alzheimer's Disease Centers (ADC), the Chicago Health and Aging Project (CHAP), the Memory and Aging Project (MAP), Mayo Clinic (MAYO), Mayo Parkinson's Disease controls, University of Miami, the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE), the National Cell Repository for Alzheimer's Disease (NCRAD), the National Institute on Aging Late Onset Alzheimer's Disease Family Study (NIA-LOAD), the Religious Orders Study (ROS), the Texas Alzheimer's Research and Care Consortium (TARC), Vanderbilt University/Case Western Reserve University (VAN/CWRU), the Washington Heights-Inwood Columbia Aging Project (WHICAP) and the Washington University Sequencing Project (WUSP), the Columbia University Hispanic- Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA), the University of Toronto (UT), and Genetic Differences (GD).

The CHARGE cohorts are supported in part by National Heart, Lung, and Blood Institute (NHLBI) infrastructure grant HL105756 (Psaty), RC2HL102419 (Boerwinkle) and the neurology working group is supported by the National Institute on Aging (NIA) R01 grant AG033193. The CHARGE cohorts participating in the ADSP include the following: Austrian Stroke Prevention Study (ASPS), ASPS-Family study, and the Prospective Dementia Registry-Austria (ASPS/PRODEM-Aus), the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Erasmus Rucphen Family Study (ERF), the Framingham Heart Study (FHS), and the Rotterdam Study (RS). ASPS is funded by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180 and the Medical University of Graz. The ASPS-Fam is funded by the Austrian Science Fund (FWF) project I904),the EU Joint Programme - Neurodegenerative Disease Research (JPND) in frame of the BRIDGET project (Austria, Ministry of Science) and the Medical University of Graz and the Steiermärkische Krankenanstalten Gesellschaft. PRODEM-Austria is supported by the Austrian Research Promotion agency (FFG) (Project No. 827462) and by the Austrian National Bank (Anniversary Fund, project 15435. ARIC research is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). Neurocognitive data in ARIC is collected by U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, 2U01HL096917 from the NIH (NHLBI, NINDS, NIA and NIDCD), and with previous brain MRI examinations funded by R01-HL70825 from the NHLBI. CHS research was supported by contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114 from the NHLBI with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629, R01AG15928, and R01AG20098 from the NIA. FHS research is supported by NHLBI contracts N01-HC-25195 and HHSN268201500001I. This study was also supported by additional

grants from the NIA (R01s AG054076, AG049607 and AG033040 and NINDS (R01 NS017950). The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by a joint grant from the Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organization for 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. Genetic data sets are also supported by the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project 050-060-810. All studies are grateful to their participants, faculty and staff. The content of these manuscripts is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the U.S. Department of Health and Human Services.

The four LSACs are: the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), The American Genome Center at the Uniformed Services University of the Health Sciences (U01AG057659), and the Washington University Genome Institute (U54HG003079).

Biological samples and associated phenotypic data used in primary data analyses were stored at Study Investigators institutions, and at the National Cell Repository for Alzheimer's Disease (NCRAD, U24AG021886) at Indiana University funded by NIA. Associated Phenotypic Data used in primary and secondary data analyses were provided by Study Investigators, the NIA funded Alzheimer's Disease Centers (ADCs), and the National Alzheimer's Coordinating Center (NACC, U01AG016976) and the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS, U24AG041689) at the University of Pennsylvania, funded by NIA, and at the Database for Genotypes and Phenotypes (dbGaP) funded by NIH. This research was supported in part by the Intramural Research Program of the National Institutes of health, National Library of Medicine. Contributors to the Genetic Analysis Data included Study Investigators on projects that were individually funded by NIA, and other NIH institutes, and by private U.S. organizations, or foreign governmental or nongovernmental organizations.

Conflict of interest statement

The author declares no conflict of interest.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

The data used was downloaded from dbGaP. Scripts, supporting files, interim files and full results will be uploaded to the NIAGADS site on acceptance.

References

Adzhubei, I., Jordan, D.M. & Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, 7 Unit7.20.

Baek, S.J. & Eling, T., 2019. Growth differentiation factor 15 (GDF15): A survival protein with therapeutic potential in metabolic diseases. *Pharmacology & Therapeutics*.

Barai, P. et al., 2019. Neuroprotective effects of bergenin in Alzheimer's disease: Investigation through molecular docking, in vitro and in vivo studies. *Behavioural Brain Research*, 356, pp.18–40.

Beecham, G.W. et al., 2017. The Alzheimer's Disease Sequencing Project: Study design and sample selection. *Neurology Genetics*, 3(5), p.e194.

Bis, J.C. et al., 2018. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. *Molecular Psychiatry*, p.1.

Blue, E.E. et al., 2018. Genetic Variation in Genes Underlying Diverse Dementias May Explain a Small Proportion of Cases in the Alzheimer's Disease Sequencing Project. *Dementia and geriatric cognitive disorders*, 45(1–2), pp.1–17.

Brouwers, N. et al., 2012. Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Molecular Psychiatry*, 17(2), pp.223–233.

Curtis, D., 2012. A rapid method for combined analysis of common and rare variants at the level of a region, gene, or pathway. *Adv Appl Bioinform Chem*, 5, pp.1–9.

Curtis, D., 2016. Pathway analysis of whole exome sequence data provides further support for the involvement of histone modification in the aetiology of schizophrenia. *Psychiatric Genetics*, 26, pp.223–7.

Curtis, D. et al., 2018. Weighted Burden Analysis of Exome-Sequenced Case-Control Sample Implicates Synaptic Genes in Schizophrenia Aetiology. *Behavior Genetics*, p.203521.

Curtis, D. & UK10K Consortium, 2016. Practical experience of the application of a weighted burden test to whole exome sequence data for obesity and schizophrenia. *Ann Hum Genet*, 80, pp.38–49.

Guerreiro, R. et al., 2013. *TREM2* Variants in Alzheimer's Disease. *New England Journal of Medicine*, 368(2), pp.117–127.

Hendriks, W.J.A.J. et al., 2013. Protein tyrosine phosphatases in health and disease. *FEBS Journal*, 280(2), pp.708–730.

Kim, J.H., 2018. Genetics of Alzheimer's Disease. *Dementia and Neurocognitive Disorders*, 17(4), p.131.

Kolbe, K. et al., 2016. Extensive nuclear sphere generation in the human Alzheimer's brain. *Neurobiology of Aging*, 48, pp.103–113.

Kumar, P., Henikoff, S. & Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(8), pp.1073–1081.

Kunkle, B.W. et al., 2019. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature Genetics*, 51(3), pp.414–430.

Lee, B. et al., 2018. Exocyst Complex Member EXOC5 Is Required for Survival of Hair Cells and Spiral Ganglion Neurons and Maintenance of Hearing. *Molecular Neurobiology*, 55(8), pp.6518–6532.

Lee, M.-H. et al., 2010. TGF-β induces TIAF1 self-aggregation via type II receptor-independent signaling that leads to generation of amyloid β plaques in Alzheimer's disease. *Cell death & disease*, 1(12), p.e110.

Loosse, C. et al., 2016. Nuclear spheres modulate the expression of BEST1 and GADD45G. *Cellular Signalling*, 28(1), pp.100–109.

Marksteiner, J. et al., 2002. Synaptic Proteins in Alzheimer's Disease. *Journal of Molecular Neuroscience*, 18(1–2), pp.53–64.

McLaren, W. et al., 2016. The Ensembl Variant Effect Predictor. *Genome biology*, 17(1), p.122.

Patel, D. et al., 2019. Association of Rare Coding Mutations With Alzheimer Disease and Other Dementias Among Adults of European Ancestry. *JAMA Network Open*, 2(3), p.e191350.

Perea, J.R., Ávila, J. & Bolós, M., 2018. Dephosphorylated rather than hyperphosphorylated Tau triggers a pro-inflammatory profile in microglia through the p38 MAPK pathway. *Experimental Neurology*, 310, pp.14–21.

Prekop, H.-T. et al., 2018. Sox14 Is Required for a Specific Subset of Cerebello–Olivary Projections. *Journal of Neuroscience*, 38(44), pp.9539–9550.

Raghavan, N.S. et al., 2018. Whole-exome sequencing in 20,197 persons for rare variants in Alzheimer's disease. *Annals of Clinical and Translational Neurology*, 5(7), pp.832–842.

Raha-Chowdhury, R. et al., 2019. Choroid Plexus Acts as Gatekeeper for TREM2, Abnormal Accumulation of ApoE, and Fibrillary Tau in Alzheimer's Disease and in Down Syndrome Dementia. *Journal of Alzheimer's Disease*, pp.1–19.

De Roeck, A. et al., 2017. Deleterious ABCA7 mutations and transcript rescue mechanisms in early onset Alzheimer's disease. *Acta Neuropathologica*, 134(3), pp.475–487.

De Roeck, A., Van Broeckhoven, C. & Sleegers, K., 2019. The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics. *Acta Neuropathologica*.

Sepulveda-Falla, D. et al., 2014. Familial Alzheimer's disease–associated presenilin-1 alters cerebellar activity and calcium homeostasis. *Journal of Clinical Investigation*, 124(4), pp.1552–1567.

Song, W. et al., 2017. Alzheimer's disease-associated TREM2 variants exhibit either decreased or increased ligand-dependent activation. *Alzheimer's & Dementia*, 13(4), pp.381–387.

Subramanian, A. et al., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43), pp.15545–15550.

Xiong, F., Ge, W. & Ma, C., 2019. Quantitative proteomics reveals distinct composition of amyloid plaques in Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 15(3), pp.429–440.

Zhang, Y. et al., 2010. Genetic reduction of striatal-enriched tyrosine phosphatase (STEP) reverses cognitive and cellular deficits in an Alzheimer's disease mouse model. *Proceedings of the National Academy of Sciences*, 107(44), pp.19014–19019.

Zhong, L. et al., 2019. Soluble TREM2 ameliorates pathological phenotypes by modulating microglial functions in an Alzheimer's disease model. *Nature Communications*, 10(1), p.1365.

Table 1

List of genes with absolute value for SLP of 3 (equivalent to p = 0.001) or more.

Gene symbol	SLP	Gene name
TREM2	14.07	Triggering receptor expressed on myeloid cells 2
ABCA7	5.8	ATP binding cassette subfamily A member 7
SORL1	4.42	Sortilin related receptor 1
KLHL35	3.73	Kelch like family member 35
PSEN1	3.67	Presenilin 1
CCDC87	3.64	Coiled-coil domain containing 87
WNT7A	3.63	Wnt family member 7A
C1R	3.52	Complement component 1, R subcomponent
EXOC5	3.49	Exocyst complex component 5
OXCT2	3.41	3-oxoacid CoA-transferase 2
ARHGAP8	3.26	Rho GTPase activating protein 8
PRPF38A	3.06	Pre-mRNA processing factor 38A
OSBPL8	3.01	Oxysterol binding protein like 8
CACNA1A	-3	Calcium voltage-gated channel subunit alpha1 A
SOX14	-3	SRY-box 14
GADD45G	-3.03	Growth arrest and DNA damage inducible gamma
LIMD1	-3.09	LIM domains containing 1
FHDC1	-3.11	FH2 domain containing 1
SCG3	-3.13	Secretogranin III
TCERG1L	-3.14	Transcription elongation regulator 1 Like
OAS3	-3.4	2'-5'-oligoadenylate synthetase 3
AIPL1	-3.49	Aryl hydrocarbon receptor interacting protein like 1
METTL14	-3.56	Methyltransferase like 14
PISD	-3.65	Phosphatidylserine decarboxylase
TBC1D23	-3.66	TBC1 domain family member 23
CFAP46	-3.74	Cilia and flagella associated protein 46
TIAF1	-3.74	TGFB1-induced anti-apoptotic factor 1
MUS81	-4.24	MUS81 structure-specific endonuclease subunit
IDNK	-4.57	IDNK, Gluconokinase
GFRAL	-5.06	GDNF family receptor alpha like

Table 2

List of variants in genes of interest with total minor allele count of at least 10 and odds ratio greater than 2 or, for genes with negative SLP, less than 0.5.

Gene	DNA change		Controls			Cases		SIFT/Polyphen/VEP annotation	Amino acid change
		AA	AB	BB	AA	AB	BB		
TREM2	6:41127543:G:A	6141	3	0	4562	10	0	deleterious(0.01) possibly_damaging(0.814) missense_variant	157|H/Y|
TREM2	6:41127613:C:A	5682	10	0	4331	16	0	synonymous_variant	
TREM2	6:41129133:C:T	6186	11	0	4576	23	0	tolerated(0.28) probably_damaging(0.999) missense_variant	87|D/N|
TREM2	6:41129252:C:T	6134	30	0	4482	86	1	deleterious(0.01) probably_damaging(0.984) missense_variant	47|R/H|
ABCA7	19:1043350:C:T	6195	4	0	4591	9	0	synonymous_variant	
ABCA7	19:1047507:AGGAGCAG:A	6091	20	0	4500	46	0	frameshift_variant	708-710|EEQ/X
ABCA7	19:1049360:G:A	6192	5	0	4585	13	0	tolerated(0.15) probably_damaging(0.944) missense_variant	826|G/R|
ABCA7	19:1051944:G:A	6086	12	0	4531	2	0	deleterious(0) probably_damaging(0.999) missense_variant	989|R/H|
ABCA7	19:1054255:G:A	5983	4	0	4480	7	0	stop_gained	1214|W/*
ABCA7	19:1056918:G:A	6154	9	0	4543	18	0	synonymous_variant	
SORL1	11:121348842:G:A	6187	6	0	4587	10	0	tolerated(0.13) probably_damaging(0.993) missense_variant	140|D/N|
SORL1	11:121421313:G:A	6196	2	0	4592	8	0	deleterious(0.01) probably_damaging(0.999)	734|D/N|

								missense_variant	
SORL1	11:121421361:G:A	6192	5	0	4586	10	0	tolerated(0.51) benign(0.089) missense_variant	750|V/I|
C1R	12:7188502:G:A	6181	14	0	4594	3	0	synonymous_variant	
C1R	12:7241211:T:G	6193	6	0	4590	10	0	tolerated(1) benign(0.014) missense_variant	359|I/L|
C1R	12:7244253:G:A	6187	3	0	4589	7	0	tolerated(0.69) benign(0.004) missense_variant	9|P/L|
EXOC5	14:57714397:G:T	6189	3	0	4579	14	0	deleterious(0.01) probably_damaging(0.99) missense_variant	21|L/I|
TIAF1	17:27401061:C:T	6177	14	1	4591	3	0	tolerated_low_confidence(0.22) benign(0.033) missense_variant	53|V/M|
SCG3	15:51980575:T:G	6183	12	0	4597	3	0	synonymous_variant	
SCG3	15:51984514:A:G	6161	12	0	4594	3	0	synonymous_variant	

Table 3

List of gene sets achieving SLP of -3 (equivalent to p = 0.001) or less.

Gene set	SLP
CELL_CELL_SIGNALING	-3.05
INORGANIC_CATION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	-3.07
MEIOSIS_I	-3.22
DEPHOSPHORYLATION	-3.55
PROTEIN_AMINO_ACID_DEPHOSPHORYLATION	-3.63
TRANSMEMBRANE_RECEPTOR_PROTEIN_PHOSPHATASE_ACTIVITY	-4.60
PHOSPHORIC_MONOESTER_HYDROLASE_ACTIVITY	-4.65
PHOSPHOPROTEIN_PHOSPHATASE_ACTIVITY	-5.18
PROTEIN_TYROSINE_PHOSPHATASE_ACTIVITY	-6.43

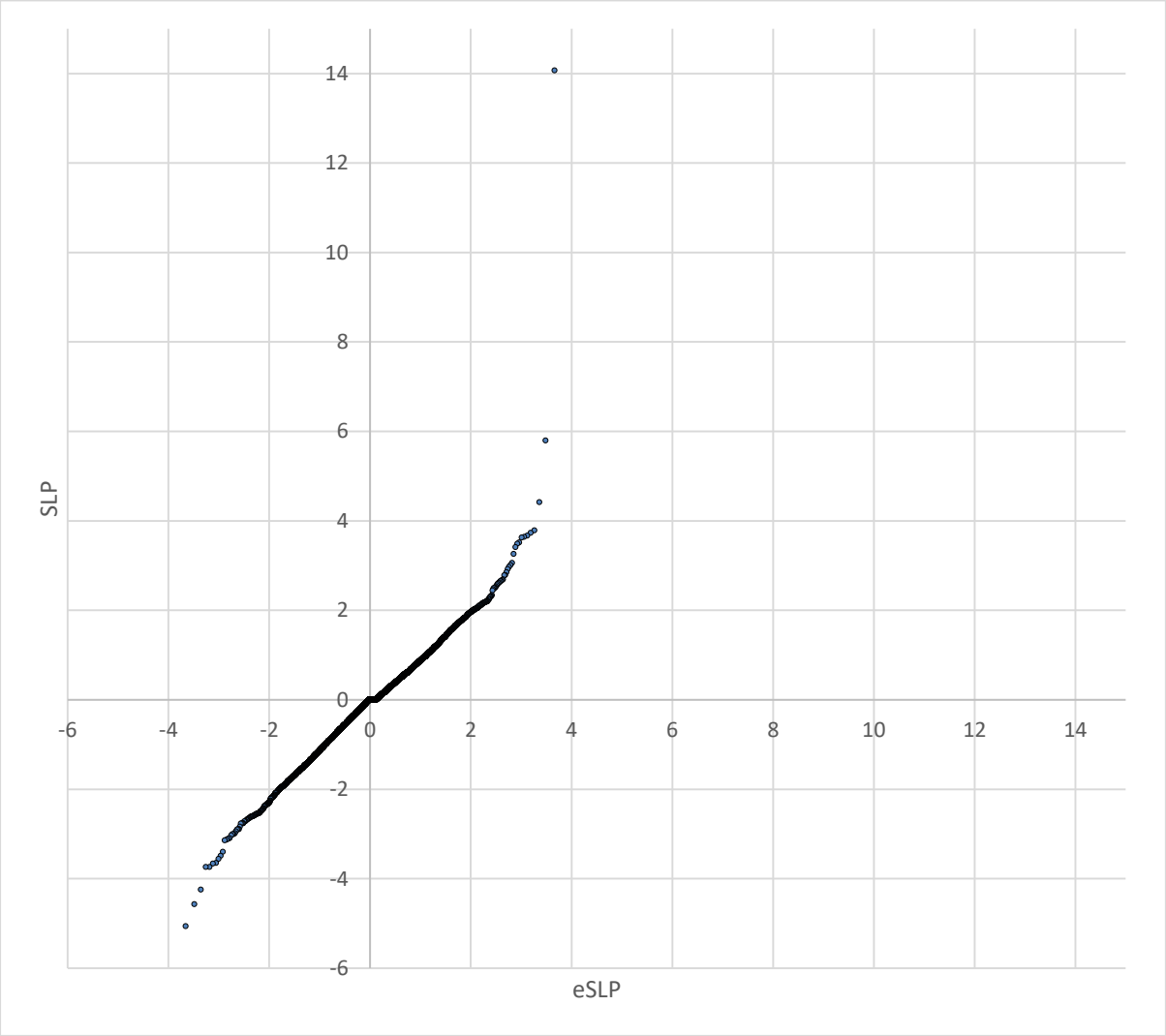
Table 4

Genes in the set PROTEIN_TYROSINE_PHOSPHATASE_ACTIVITY with SLPs of -1.3 (equivalent to p=0.05) or less.

Gene symbol	SLP	Gene name
PTPRS	-2.36	Protein tyrosine phosphatase, receptor type S
DUSP6	-2.33	Dual specificity phosphatase 6
PTPRU	-2.09	Protein tyrosine phosphatase, receptor type U
PTPRT	-2.08	Protein tyrosine phosphatase, receptor type T
PTPRR	-1.98	Protein tyrosine phosphatase, receptor type R
PTPN12	-1.71	Protein tyrosine phosphatase, non-receptor type 12
PTP4A3	-1.37	Protein tyrosine phosphatase type IVA, member 3
PTPN22	-1.34	Protein tyrosine phosphatase, non-receptor type 22
PTPN1	-1.31	Protein tyrosine phosphatase, non-receptor type 1

Figure 1

QQ plot of observed versus expected gene-wise SLP.



Supplementary table S1

The table shows the weight accorded to each type of variant as annotated by VEP (McLaren et al. 2016). 10 was added to this weight if the variant was annotated by Polyphen as possibly or probably damaging and 10 was added if SIFT annotated it as deleterious (Adzhubei et al. 2013; Kumar et al. 2009).

VEP annotation	Weight
intergenic_variant	1
feature_truncation	3
regulatory_region_variant	3
feature_elongation	3
regulatory_region_amplification	3
regulatory_region_ablation	3
TF_binding_site_variant	3
TFBS_amplification	3
TFBS_ablation	3
downstream_gene_variant	3
upstream_gene_variant	3
non_coding_transcript_variant	3
NMD_transcript_variant	3
intron_variant	3
non_coding_transcript_exon_variant	3
3_prime_UTR_variant	5
5_prime_UTR_variant	5
mature_miRNA_variant	5
coding_sequence_variant	5
synonymous_variant	5
stop_retained_variant	5
incomplete_terminal_codon_variant	5
splice_region_variant	5
protein_altering_variant	10
missense_variant	10
inframe_deletion	15
inframe_insertion	15
transcript_amplification	15
start_lost	15
stop_lost	15
frameshift_variant	20
stop_gained	20
splice_donor_variant	20
splice_acceptor_variant	20
transcript_ablation	20