Universiteit Leiden

Master Computer Science

Linking genetic variants to proteomics in Gyrus Temporalis Medialis shows disrupted correlation and expression patterns between individuals with Alzheimer's Disease and non-demented controls

Name: Gerard Bouland
Student ID: s2378310
Date: 30/06/2020

Specialisation: Bioinformatics

1st supervisor: Katy Wolstencroft (LIACS)
2nd supervisor: Marcel Reinders (TU Delft)

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science
Leiden University

Niels Bohrweg 1

2333 CA Leiden

The Netherlands

Abstract

Alzheimer's Disease (AD) is a progressive disease characterized by loss of cognitive functions and autonomy, eventually leading to death. It is estimated that up to 80% of the risk for AD is determined by genetics of which 30% can attributed to the ε4 allele of *APOE* gene. GWASs are performed to identify additional risk factors. Understanding the mechanistic pathways through which these genetic risk factors are associated with AD is difficult. Here, we investigated how genetic variants influence AD associated aberrant behavior of proteins in Gyrus Temporalis Medialis (GTM). By investigating protein quantitative trait loci (pQTLs) we are looking at a molecular level that is directly involved in the biological processes that are supposedly affected.

5,861 variant have been identified associated with aberrant expression of 153 proteins. The identified pQTL variants were compared to previously identified pQTLs and expression quantitative trait loci (eQTLs) in various brain regions. Eleven variants have been identified, associated with aberrant expression of proteins while also associated with AD risk. Additionally, we revealed that when individuals were grouped on homozygous genotypes of known AD risk variants, within those groups, distinct protein correlation structures exist. Finally, emerging, and disrupted protein co-expression networks associated with distinct biological processes were identified in individuals diagnosed with AD compared to non-demented (ND) controls.

We showed that, if we want to understand genetic control on the brain proteome and its interaction with the brain transcriptome, region specific pQTL and eQTL studies are required. Furthermore, with the differential correlation analysis with respect to the genotypes of AD risk variants, we show that this approach is a promising addition to GWAS- and QTL-studies. Finally, with the emerged and disrupted protein co-expression networks we presented a set of proteins that might be associated with neurodegenerative consequences of AD and is associated with a dysregulation of metabolic processes and, alternative signal transduction that initiates a collaboration between proteins that is associated with a detrimental immune response.

In conclusion, in this study, we linked genetic variants to proteomics in the GTM and revealed association between AD status and altered protein expression and correlation.

Abbreviations

AD, Alzheimer's Disease

CHC, Cognitive healthy centenarians

ND, Non-demented

eQTL, expression Quantitative Trait Loci

pQTL, protein Quantitative Trait Loci

GWAS, Genome-wide association study

MAF, Minor allele frequency

LD, Linkage Disequilibrium

TSS, Transcription start site

GTM, Gyrus Temporalis Medialis

Terms

In this report, ***eQTL*** is used to describe a variant in consolidation with its respective mRNA transcript. ***eQTL variant*** is used when solely the variant is indicated.

Similar to *eQTL, pQTL* is used to describe a variant in consolidation with the associating protein. ***pQTL variant*** is used to indicate solely the variant.

Braak stage is a pathological assessment of AD progression and describes the spread of AD related neurofibrillary-tangles and hyperphosphorylated tau protein in different brain regions[1]

Symbols

β – BETA: The regression coefficient between $\log_2$ intensity of protein and variant. Describes the linear additive genetic effect of the minor allele relative to the major allele on protein expression.

OR – Odds ratio: Association between AD status (ND = 0, AD = 1) and a variant's genotypes.

The genetic linear additive association of variants with protein expression is expressed in regression coefficient: beta (β). This is different from the association between variants and phenotype (AD vs control), in this case, the relationship is expressed in odds-ratios (OR).

r – Correlation coefficient: Pearson's correlation coefficient

r^2 – r squared: Squared Pearson's correlation coefficient, used as threshold measure as it captures negative and positive co-expression.

Δr – Difference in correlation: The absolute difference of two Pearson's correlation coefficients.

Range: min = 0, max = 2

ΔZ – Z score difference: Difference in z-scores that also follows a normal distribution.

Contents

1. Introduction

Alzheimer's Disease (AD) is a progressive disease characterized by loss of cognitive functions and autonomy, eventually leading to death[2]. It is estimated that up to 80% of the risk of AD is determined by genetics[3] of which 30% can attributed to the ε4 allele of *APOE* gene. The genetic risk and functional consequences of the ε4 allele of *APOE* have extensively been investigated[4,5,6]. Additionally, many genome wide association studies (GWASs) have been performed in order to identify additional genetic risk factors and to understand their role in AD etiology[7,8,9,10,11]. These GWASs have identified >40 genetic variants that modify the risk of AD, however, understanding the mechanistic pathways through which these genetic factors are associated with AD is difficult. Many risk variants are located in non-coding and intergenic regions[12]. As such, for many risk variants it is investigated whether they are also expression quantitative trait loci (eQTL), *i.e.* variants that are associated with differential expression of a messenger RNA transcript (mRNA). When risk variants are also eQTLs, RNA transcripts are often considered synonymous for proteins: while mRNAs code for proteins, their expression levels are not always correlated with protein expression levels[13,14,15] and eQTLs are often not protein QTLs (pQTLs)[16]. Proteins perform functions in the majority of all biological domains and are involved in many biological processes. Understanding how AD variants associate with protein expression might explain or elucidate the risk associated with the respective variants, as the biological processes in which the proteins are involved might be altered or disturbed. By investigating pQTLs we are looking at a molecular level that is directly involved in the biological processes that are supposedly affected. Concurrently, consequences of variants are not always reflected in differential expression. For instance, a missense variant might alter the amino acid sequence of protein and alter its function without changing its abundance. Here, we hypothesize that downstream consequences of disease-associated variants may be uncovered by genotype-specific co-expression networks. Under the assumption that co-expressed proteins are functionally related, these genotype specific co-expression networks might indicate an activation or deactivation of disease associated biological pathways. We tested this hypothesis on fifteen previously discovered AD risk variants[11] and investigated supposedly downstream consequences of the respective variants. While genotype specific co-expression networks might elucidate the genetic risk associated with AD, investigating co-expression differences associated with disease state can uncover distinct molecular signatures associated with AD, which was also investigated in this study. The target tissue, subject in this study is the Gyrus Temporalis Medialis (GTM), the GTM is involved with episodic memory[17] and is part of the temporal lobe, a region visibly affected by AD related neurofibrillary-tangles in a late stage of disease progression[1]. The cohort, subject in this study is comprised of controls, and individuals representing the extreme ends of the AD spectrum. Namely, 1) individuals that have been diagnosed with AD at relatively early age, and, on the other end, 2) cognitively healthy centenarians (CHC). A previous study showed that with extreme phenotypes, the effect sizes of AD associated variants identified in GWASs nearly doubled[18]. In this study, it is expected that due to the inclusion of

the extreme ends of the AD spectrum, we increase power and are able to uncover genetic control of proteins associated with AD risk that would otherwise go unnoticed.

Here, we investigated genetic control of variants on protein expression by means of a large-scale pQTL analysis, and, whether there are pQTL variants that are also associated with AD risk. Genetic control of AD associated risk variants on co-expression between proteins in the GTM was also investigated. Furthermore, we investigated protein co-expression networks specifically present and absent in individuals diagnosed with AD compared to non-demented (ND) controls.

1.1 Analysis workflow

This study consists of three major analyses that are complemented with additional smaller analyses. The first analysis (Fig. 1a) involves a large scale pQTL analysis (See Methods *pQTL identification*). For every measured protein in the GTM, we tested whether their abundance was associated with certain genetic variants. Each protein was tested against variants located 250 kilo base pairs (Kbp) up- or downstream the protein's transcription start site (TSS). The pQTLs identified in this study were compared with previously identified brain pQTLs[16] and eQTLs from twelve brain regions from GTEx[19]. For all identified pQTLs it was tested whether the variant was associated with increased genetic risk for AD, this was described in a separate subsequent results section. Additionally, a polygenic risk score (PRS) was constructed of pQTLs and tested on its predictive value of AD risk (See *Appendix*). The second major analysis (Fig. 1b) investigated whether fifteen known AD risk variants[11] were associated with distinct protein correlations structures. This was determined by means of differential correlation of pairs of proteins between individuals homozygous for the AD-risk allele and those homozygous for the non-risk allele. When one of the homozygous alleles was not present in the population, the differential correlation was determined between the homozygous allele that was present and heterozygous allele. (See Methods *Differential correlation*). In the third major analysis we investigated differential correlation of proteins between individuals diagnosed with AD and ND controls (See methods *Differential correlation with respect to phenotype status*). For this analysis, we defined two classes of differential correlation. The first class (Fig 1c) includes protein pairs that are specifically co-expressed in AD individuals, and not in ND controls. The second class (Fig 1d) includes pairs of proteins specifically co-expressed in ND controls, and not correlated in AD individuals. In addition, associations of proteins with neuropathological features of AD were tested in the context of differential correlation. Lastly, differentially correlated proteins involved in co-expression networks were subject in pathway analyses to identify the aberrant biological pathways.

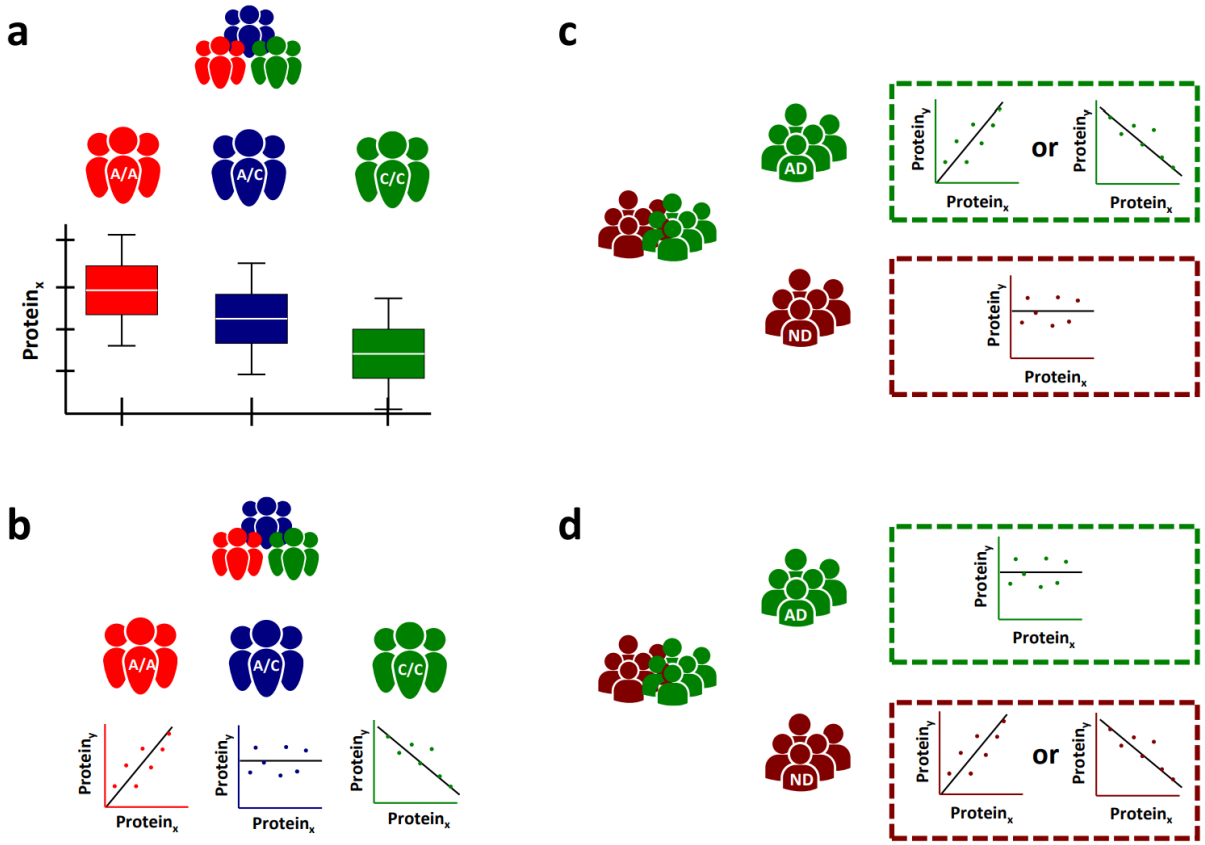

Figure 1 Schematic overview of the four major analysis that have been performed. **a)** Schematic representation of a pQTL. Individuals are divided based on the genotypes of an arbitrary variant and when the expression of a protein is significantly linearly associated with the genotype, a pQTL is identified. **b)** Schematic overview of differential correlation between proteins with respect to genotype. Again, individuals are dived based on the genotypes of an arbitrary variant and pairs of proteins are identified that are differentially correlated ($P_{FDR} \leq 0.05$) with respect to this variant. Note, three groups are presented in the schematic overview, differential correlation is only calculated between two groups. **c)** Schematic overview of identifying differential co-expression network specifically present in AD individuals but absent in control. The scheme represents the analysis that identifies a single pair of proteins that is differentially correlated between the respective groups. **d)** Schematic overview of identifying differential co-expression network specifically present in controls but absent in AD individuals. The scheme represents the analysis that identifies a single pair of proteins that is differentially correlated between the respective groups.

2. Results

2.1 Demographics

After quality control and pre-processing of the genetic data, 6,607 individuals were subject in the analyses. For the protein expression data from the GTM, 190 individuals remained. The number of individuals that were present in both datasets was 140 individuals (Table 1). The mean age of individuals in the genetic data was 68.4 (σ = 15.8) and consisted 53.7% of females. The mean age of the 190 individuals of which protein data was available was 86.8 (σ = 13.8) and consisted 73.7% of females.

The individuals of the intersection had a mean age of 91.0 (σ = 14.2) and the percentage of females was 74.3%.

Table 1: Population characteristics

	Genetic data			Protein data			intersection		
Number of individuals	6607			190			140		
Females (%)	3549 (53.7)			140 (73.7)			104 (74.3)		
Age (σ)	68.4 (15.8)			86.8 (13.8)			91.0 (14.2)		
Diagnosis[a]	AD	ND	CHC	AD	ND	CHC	AD	ND	CHC
N	2416	3848	343	88	53	49	67	27	46

[a](AD = AD cases, ND = Non-demented controls, CHC = cognitively healthy centenarians)

2.2. Identified pQTL variants associated with abundance of 153 proteins in Gyrus Temporalis Medialis

In this pQTL analysis (Fig. 1a), 140 individuals of which both genetic and protein data was available were subject. (See Table 1: *intersection*). Of the 3,556 proteins that were available, 3,427 proteins were tested on association with genetic variants. 129 proteins were not tested, as the gene of the respective protein was not located on chromosome 1-22 or the protein was not present in the Ensembl genome browser[20]. In total, 5,861 pQTL variants were significantly associated with the abundance of 153 proteins ($P_{FDR} \leq 0.05$, Fig. 2a). As the mapping window for pQTL discovery was set on 250 Kilo base pairs (Kbp) down- and upstream around the TSS of the respective proteins, all variants are considered *cis*-pQTLs. Of the 5,861 pQTL variants discovered in the GTM, 2,265 (39%) pQTL variants were also previously found in the dorsolateral prefrontal cortex in 144 cognitively healthy individuals[16] (See methods *Summary statistics pQTL study*). Next, we compared the regression coefficients (β) of the linear pQTL associations of the two studies. Overall, the correlation between pQTL effect sizes of matching pQTLs was r = 0.39 ($P = 1.37 \times 10^{-83}$, Fig. s1a). For the pQTL direction effects that were in the same direction (N = 1,591) the Pearson's correlation coefficient between estimates was r = 0.92 ($P < 5.00 \times 10^{-100}$, Fig. s1b). For pQTLs with opposite directional effects (N = 674) the Pearson's correlation coefficient was r = -0.93 ($P < 5.00 \times 10^{-100}$, Fig. s1c). Of 153 proteins for which we identified pQTL variants, 60 unique proteins were associated with the shared pQTLs. Next, we tested whether the 5,861 pQTL variants were also eQTL variants. For this, the pQTLs were clumped ($R^2 \geq 0.001$, MAF ≥ 0.05) on P-values of the pQTL association. This resulted in 154 independent pQTL variants associated with 145 proteins. Eight proteins were lost as the minor allele frequency (MAF) of the associated pQTL variants were lower than 5% in the 1,000 Genomes Project reference panel[21]. Of the 154 pQTL variants, 53 (34%) variants were also an eQTL variant in varying brain tissues available in the GTEx Portal[19] (N

= 12). The most significant eQTL was rs8012 with *GCDH* in cerebellum ($P = 1.10 \times 10^{-52}$, NES = 0.93, Fig. 2c). rs8012 is moderate pQTL variant for *GCDH* in the GTM ($P = 4.70 \times 10^{-06}$, $\beta = 0.29$, Fig. 2b). Fourteen pQTLs variants were eQTL variants in all twelve brain tissues (Fig. 2d). Most eQTL variants were observed in the cortex (N = 45) and frontal cortex Brodmann area 9 (N = 38). Figure 2d shows all pQTL variants that are also a significant eQTL variant in any of the twelve brain regions. The effect sizes of the matching pQTLs and eQTLs were significantly correlated in all brain regions ($P \leq 6.55 \times 10^{-3}$, $r \geq 0.50$). The strongest correlation of effect sizes was observed with eQTL from caudate basal ganglia ($P = 3.50 \times 10^{-5}$, r = 0.64, Fig. s2c). The weakest correlation was observed with effect sizes of eQTLs in the cerebellum ($P = 2.77 \times 10^{-5}$, r = 0.50, Fig. s2d).

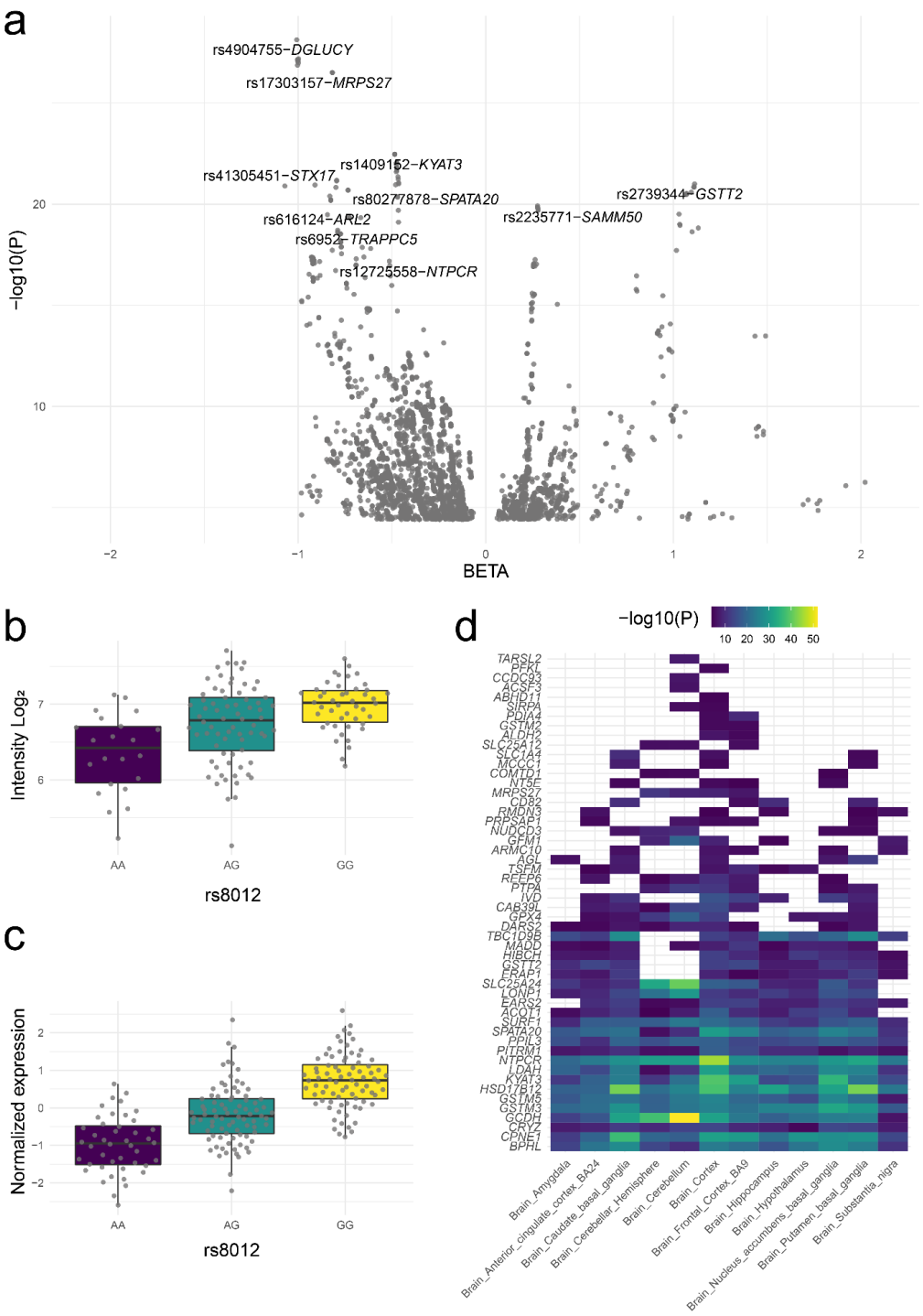

Figure 2 Overview of significant pQTLs and comparison of pQTLs with GTEx eQTLs. **a)** Volcano plot of identified pQTLs ($P_{FDR} \leq 0.05$). X-axis represents the effect size of the pQTL association, and the y-axis represents -$\log_{10}$ P-value of the pQTL association. Labels are shown for pQTLs $P \leq 1 \times 10^{-15}$. **b)** Boxplots of GCDH protein $\log_2$ intensity grouped on genotypes of rs8012. X-axis represent the genotypes of rs8012, and the y-axis represents the $\log_2$ intensity of GCDH. **c)** Boxplots of normalized expression of GCDH in Cerebellum grouped on genotypes of rs8012 (GTEx). X-axis represent the genotypes of rs8012, and y-axis represent the normalized expresion of the GCDH transcript from GTEx. **d)** Heatmap of all pQTL variants that are also an eQTL variant, squares are rendered for significant eQTLs and the color represents the -$\log_{10}$ P-value. The x-axis of the heatmap are the tested brain tissues and the y-axis the protein/gene names.

2.3. pQTLs associated with *APOE* abundance also associated with increased Alzheimer's Disease risk

To identify pQTL variants associated with AD risk, we started with the pQTL set identified in the previous analysis (N pQTLs = 5,861, N proteins = 153). Next, we filtered the pQTLs based on the proteins, and selected pQTLs of which the proteins were differentially expressed in AD individuals. For this, we performed a differential protein expression analysis between ND controls and AD individuals. 2,196 proteins were found differentially expressed in AD individuals ($P_{FDR} \leq 0.05$, Fig. 3b). 1,033 proteins were upregulated, and 1,163 proteins were downregulated in AD individuals. 97 proteins were found differentially expressed and also to associate with pQTL variants (N pQTLs = 3,543, N proteins = 97, Fig. 3a). These 3,543 pQTLs variants were tested on association with AD status (see Methods *Testing pQTL variants on association with AD risk*, Fig. 3a). Eleven variants were associated with AD risk ($P_{FDR} \leq 0.05$). Of these, six variants were a pQTL for *APOE* and five variants were a pQTL for *SIRPA.* The *APOE* protein was upregulated in AD individuals ($P_{FDR} = 9.22 \times 10^{-6}$, logFC = 0.22, Fig. 3c). All six variants (rs429358, rs6857, rs769449, rs12972970, rs12972156 and rs34342646) had strong association with AD risk, ($P \leq 1.80 \times 10^{-112}$, OR ≥ 2.8), with rs429358 being the strongest ($P = 1.90 \times 10^{-171}$, OR=3.56). Of these six variants, rs6857 was the strongest associated with *APOE* abundance ($P = 2.75 \times 10^{-6}$, $\beta = 0.25$, Fig. s3a). The Linkage Disequilibrium (LD) between rs429358 and rs6857 is $R^2 = 0.76$ (Fig. s3c). Rs429358 is a missense variant for *APOE*, rs6857 is a 3'-URT variant for *NECTIN2*, rs769449 is a non-coding exon variant, the remaining variants (rs12972970, rs12972156, rs34342646) are located in introns. In GTEx, none of the six pQTL variants are an eQTL variant for *APOE* in any tissue or brain region. *SIRPA* was downregulated in AD individuals ($P_{FDR} = 2.51 \times 10^{-6}$, logFC = -0.20, Fig. 3d). *SIRPA* was also associated with five pQTL variants that are also associated with AD risk (rs6081094, rs2024867, rs2024868, rs6081105, rs754829). All variants were in LD ($R^2 \geq 0.94$), as such, had identical ORs for AD risk (OR = 1.16, $P \leq 1.30 \times 10^{-4}$) and all variants are intron variants for *SIRPA*. The strongest pQTL variant was rs6081094 ($P = 1.43 \times 10^{-5}$, $\beta = 0.18$).

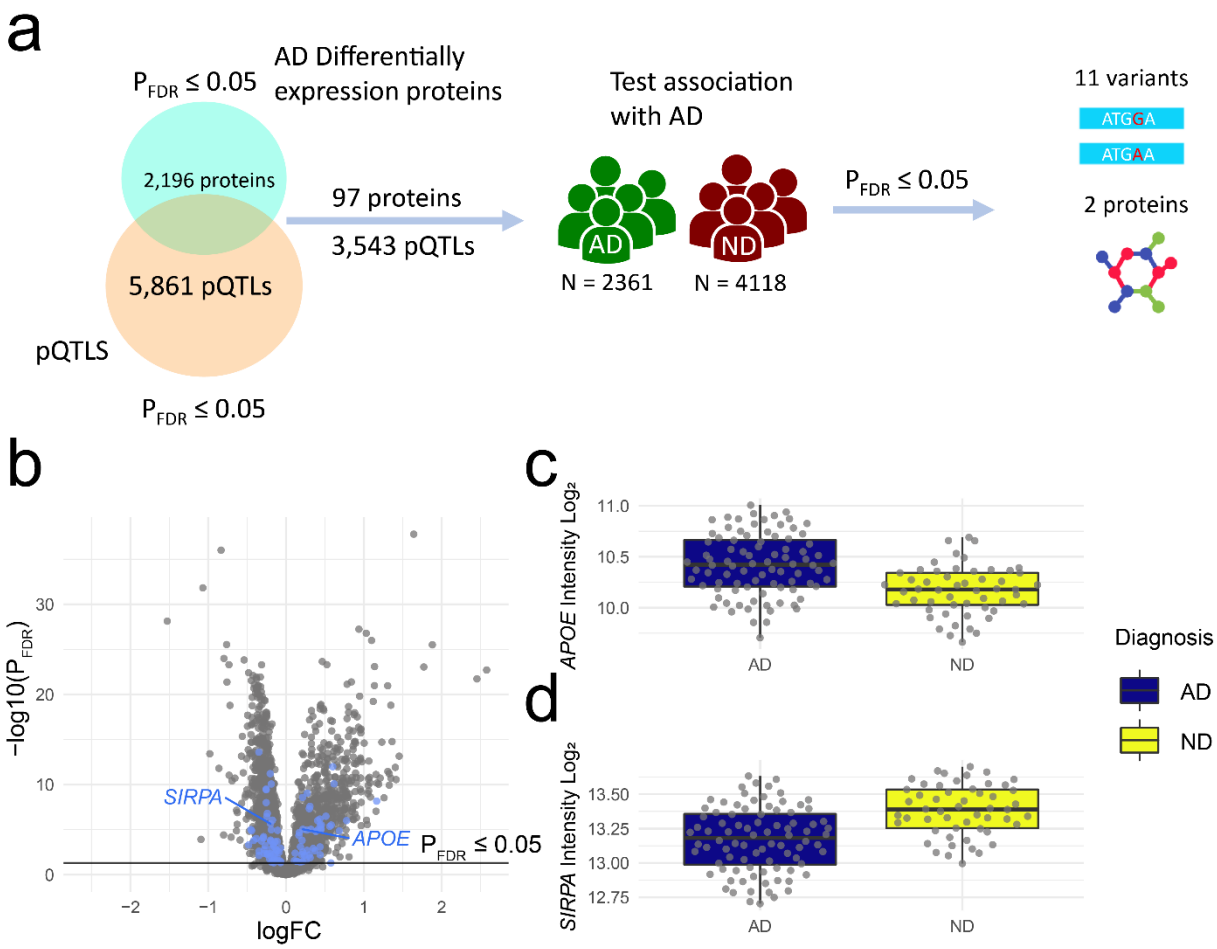

Figure 3 Overview of analysis workflow and differentially expressed proteins. **a)** Schematic overview of the analysis workflow that was used to identify pQTLs that are also associated with AD risk. This is purely a schematic overview, as such, the sizes of the circles in the Venn diagram are not representative for the data that that is described. **b)** Volcano plot of differentially expressed proteins. Blue dots indicate differentially expressed proteins for which also significant pQTLs were found. Labels are added as APOE and SIRPA were associated with pQTL variants that are also associated with AD risk. **c)** Boxplot of APOE, the x-axis represents the diagnosis of the individuals and the y-axis represents the Log_2 normalized intensity of APOE. **d)** Boxplot of SIRPA, the x-axis represents the diagnosis of the individuals and the y-axis represents the Log_2 normalized intensity of SIRPA.

2.4. Rs9381040 associated with 67 pairs of differentially correlated proteins

Here, we analyzed known AD risk variants[11] and whether individuals with different genotypes of the respective variants have distinct proteins correlation structures (Fig. 1b). The individuals subject in this analysis were the individuals for which genetic data and protein expression data was available (N = 140, See Table 1: *intersection*). Of the 41 known AD risk variants, after stringent selection, fifteen remained (See methods *Differential correlation with respect to AD variants genotype*). Thirteen variants remained with all three genotypes present, and two variants with only the homozygous genotype and heterozygous genotype present (Table S2). 156 pairs of proteins ($P_{FDR} \leq 0.05$, Fig 4a) were found to be differentially correlated. Most differentially correlated proteins were found between the homozygous genotypes of rs9381040 (N pairs = 67, N proteins = 112), an intergenic variant located near *TREML2* gene. Of the 67

differentially correlated proteins pairs, *DDX17* formed a pair with sixteen other proteins (Fig 4b), as such was found differentially correlated with sixteen proteins between homozygous rs9381040-C (N_{CC} = 78) and rs9381040-T (N_{TT} = 11). *DDX17* was most differentially correlated with *PKLR* ($\Delta r = 0.98$, $P_{FDR} = 1.55 \times 10^{-3}$, Fig. 4c) and *CRYM* ($\Delta r = 1.23$, $P_{FDR} = 7.41 \times 10^{-3}$, Fig. 4d). The correlation between *DDX17* and *PKLR* in individuals with the risk allele was $r_{CC} = 0.00$ and correlation in individuals with the protective allele was $r_{TT} = 0.98$. For the heterozygous allele, the correlation was $r_{CT} = 0.42$. The correlation between *DDX17* and *CRYM* in individuals with the risk allele was $r_{CC} = -0.28$ and correlation in individuals with the protective allele was $r_{TT} = 0.96$. For the heterozygous allele, the correlation was $r_{CT} = 0.16$. Functional enrichment analysis of the proteins differentially correlated with *DDX17* showed enrichment for, among other terms, *GO cellular component*: cytosol (N = 14, $P_{FDR} = 1.90 \times 10^{-4}$) and *PFAM Protein Domains*: Synuclein (N = 3, 2.06×10^{-7}). All members of synuclein family (*SNCA*, *SNCB* and *SNCG*, $\Delta r \geq 1.15$, $P_{FDR} \leq 2.98 \times 10^{-2}$) were differentially correlated with *DDX17* with respect to the genotypes of rs9381040. Next, we investigated the differentially correlated proteins with respect to the genotypes of rs11218343, an intron variant for the *SORL1* gene. Sixteen pairs of proteins, comprised of 31 unique proteins, were found differentially correlated between homozygous individuals (N_{TT} = 128) and heterozygous individuals (N_{TC} = 12). *E.g*. in the homozygous individuals *PRDX1* and *MRPS27* had a correlation coefficient of r = -0.26, while in the heterozygous individuals the correlation coefficient was r = -0.98 ($\Delta r = 0.72$, $P_{FDR} = 6.02 \times 10^{-3}$). Nineteen of the 31 proteins were involved in the *GO biological process:* localization (N = 19, $P_{FDR} = 7.50 \times 10^{-3}$), eight proteins were involved in *GO cellular component*: dendrite (N = 8, $P_{FDR} = 2.00 \times 10^{-4}$) and neuron projection (N = 10, $P_{FDR} = 2.80 \times 10^{-4}$). Additionally, we identified two proteins (*NUB1-TUBB*) that were differentially correlated with respect to the genotypes of four different variants, rs9381040 ($\Delta r = 0.34$, $P = 2.84 \times 10^{-9}$), rs7920721 ($\Delta r = 0.50$, $P = 1.24 \times 10^{-12}$), rs3740688 ($\Delta r = 0.93$ $P = 3.61 \times 10^{-13}$), rs4311 ($\Delta r = 0.40$, $P = 2.31 \times 10^{-9}$). The degree of overlap between the high-correlation alleles was relatively low (Fig s4).

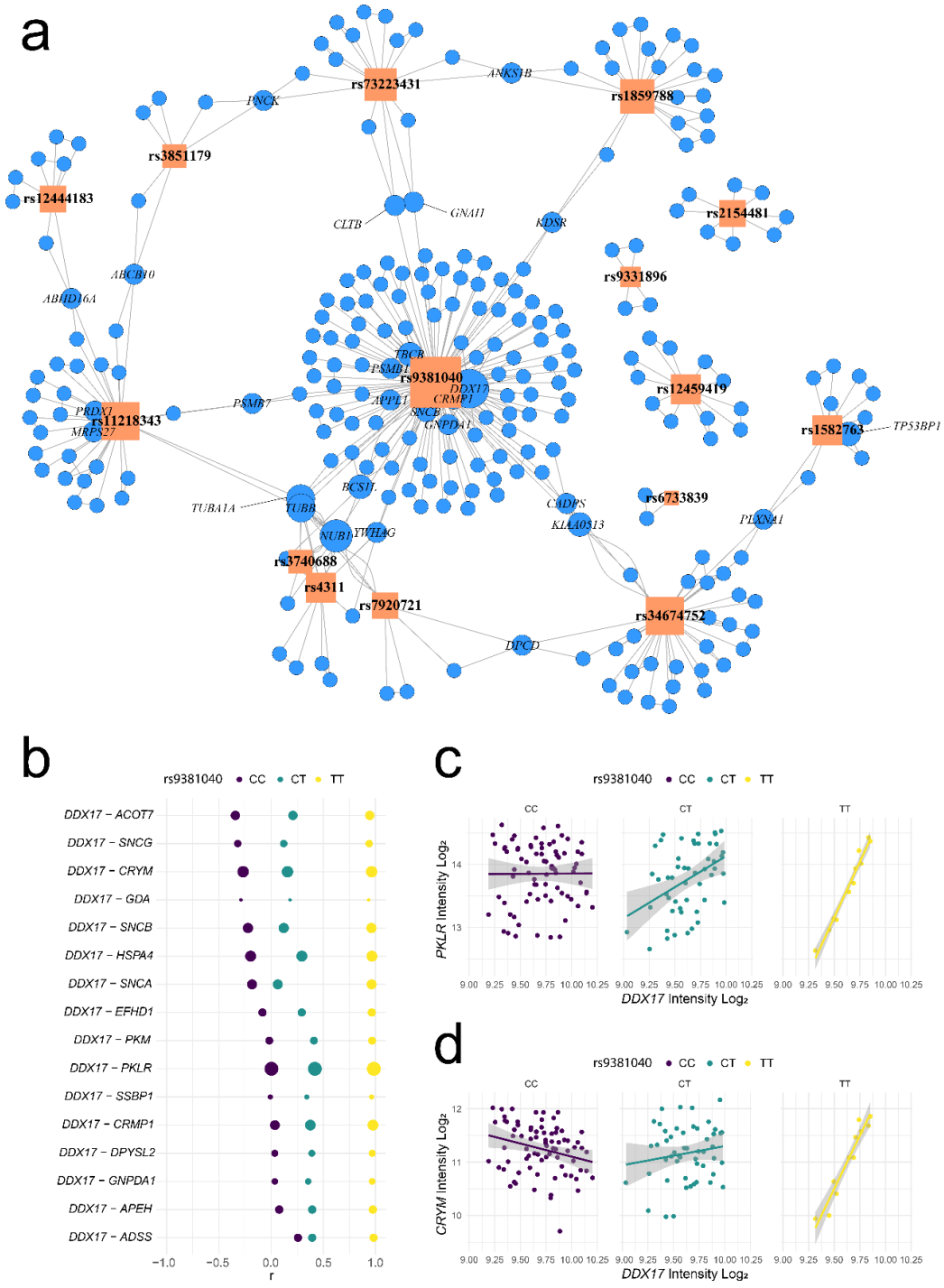

Figure 4 Overview of differentially correlated proteins with respect to AD variants. **a)** Network graph of differentially correlated proteins and their respective variants. Blue nodes are proteins, an edge between two proteins means that they are differentially correlated with respect to the variant they are connected with (orange square). Size of nodes is determined by the degree and labels are shown for nodes with degree ≥ 4. **b)** All proteins DDX17 is differentially correlated with, with respect to rs9381040. The x-axis represents the Pearson's correlation coefficient between the protein pairs for individuals with the respective genotypes. The y-axis represents each protein pair and the color of dots indicates the different genotypes. The size of the dot is determined by the -$\log_{10}$ p-value **c)** Scatter plot of the intensities of DDX17 (x-axis) and PKLR (y-axis), individuals are separated on rs9381040 genotype and colored accordingly. For each genotype, a linear regression line is drawn. **d)** Scatter plot of the intensities of DDX17 (x-axis) and CRYM (y-axis), individuals are separated on rs9381040 genotype and colored accordingly. For each genotype, a linear regression line is drawn.

2.5. AD specific co-expression network associated with signal transduction and immune system

Here, we performed a differential correlation analysis with respect to AD status. Specifically, we investigated proteins that were co-expressed in AD individuals and not in ND controls (See methods *Differential correlation with respect to phenotype status,* Fig. 1c). The individuals subject in this analysis were the individuals for which protein data was available (N=190, See Table 1: *Protein data*). CHC with Braak stage ≥ 4 were exclude from this analysis (N=13). The remaining CHC were considered ND controls. As a result, 177 individuals remained, 88 individuals diagnosed with AD and 89 ND controls. Pairwise correlations between all available proteins were separately calculated for the AD and ND groups. In total, 345 pairs of proteins were identified that were significantly more correlated in the AD group ($r^2 \geq 0.65$) compared to the ND group ($r^2 \leq 0.20$, $P_{FDR} \leq 0.001$). The 345 pairs of proteins were comprised of 151 unique proteins. The most differentially correlated protein pairs were *HSPB1-GNA13* ($\Delta r = 0.85$, $P_{FDR} = 3.37 \times 10^{-12}$, Fig. 5a) and *TLN1-EEF1A1* ($\Delta r = 0.93$, $P_{FDR} = 3.87 \times 10^{-11}$, Fig. 5a). *TLN1* was found differentially correlated with the most proteins (N = 26). *HSPB1* and *ITGB1* were both found differentially correlated with 25 proteins. These three proteins shared thirteen proteins with which they were differentially correlated (Fig. 5b). Although 345 proteins pairs were found significantly differentially correlated, pairwise associations between all 151 proteins in the ND individuals were relatively low with a median r^2 of 0.03 (Fig. 5c) compared to the association in the AD individuals where all protein pairs had a median r^2 of 0.41 (Fig. 5d) Additionally, the first principal component of the 151 proteins in all individuals which explained 97% of the variance was associated with the Braak stage ($r = 0.75$, $P = 5.10 \times 10^{-33}$) and age ($r = -0.28$, $P = 1.91 \times 10^{-4}$). Of the 345 differentially correlated protein pairs, 316 pairs also showed an interaction with Braak stage ($P_{FDR} \leq 0.05$) as the linear association between proteins changed with increasing Braak stage (Fig s6). Pathway analysis with String (v11)[22] revealed enrichment for, among others, synaptic vesicle cycle (N = 13, $P = 9.83 \times 10^{-13}$), immune system (N = 29, $P_{FDR} = 3.24 \times 10^{-6}$), innate immune system (N = 18, $P_{FDR} = 9.17 \times 10^{-5}$) and signal transduction (N = 38, $P_{FDR} = 8.39 \times 10^{-8}$) for KEGG pathways. Signaling by receptor tyrosine kinases (N = 25, $P = 2.25 \times 10^{-12}$) and L1CAM interactions (N = 13, $P = 3.48 \times 10^{-09}$) were among enriched REACTOME pathways. Four proteins were also associated with pQTL variants. Rs2854248, rs17134970, rs7960152 and rs950798 were found to be associated ($P_{FDR} \leq 0.05$) with the abundance of *H2BFS*, *PFKP*, *ANKS1B* and *PRKCG*, respectively.

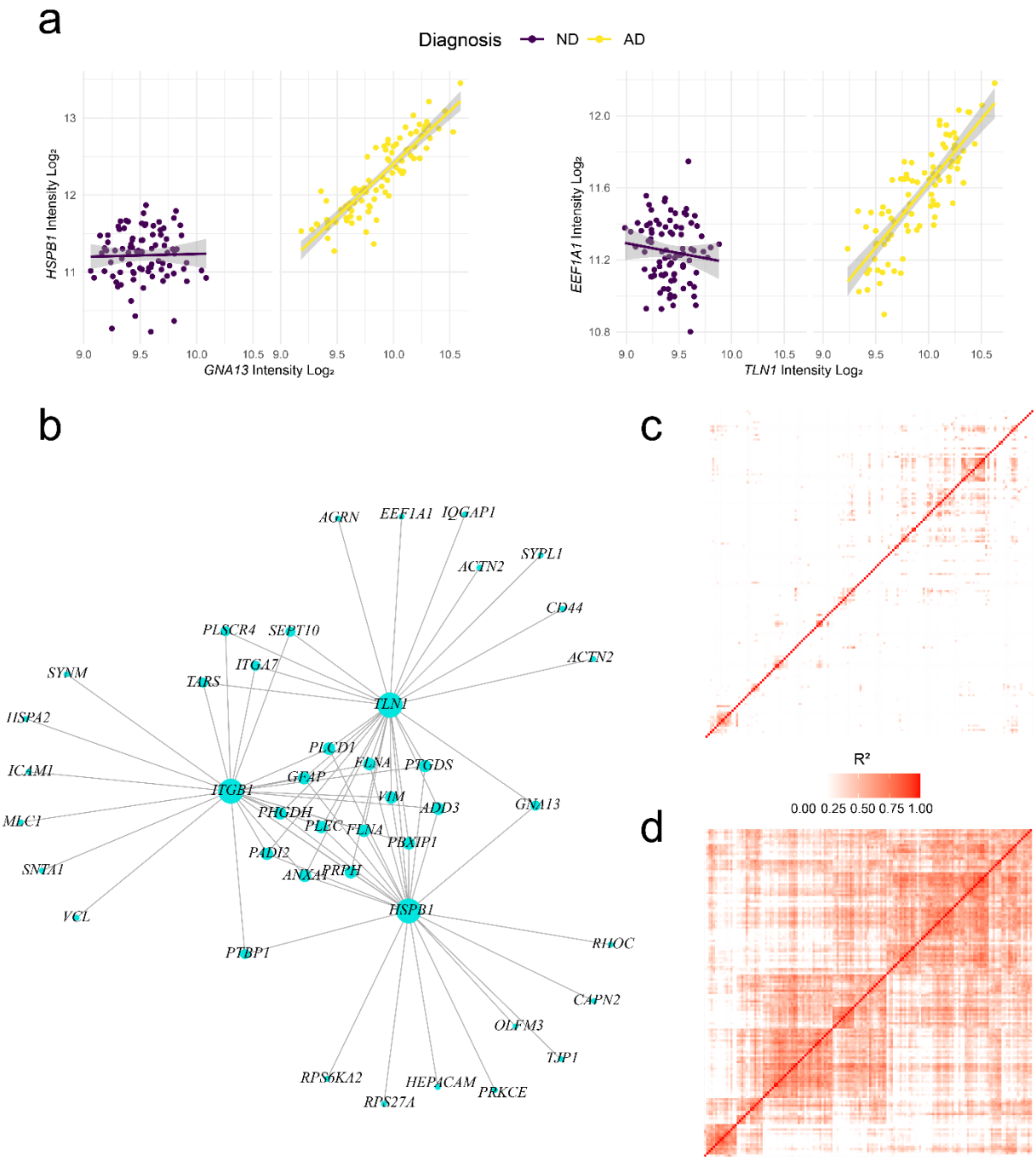

Figure 5 Overview of proteins specifically correlated in AD individuals. **a)** Left figure, log_2 intensity (x-axis) of GNA13 and HSPB1 (y-axis), with individuals divided in controls (ND, Purple) and AD individuals (yellow). Right, log_2 intensity (x-axis) of TLN1 and EEF1A1 (y-axis), with individuals divided in controls (ND, Purple) and AD individuals (yellow). **b)** network graph of all the proteins that are differentially correlated with ITGB1, HSPB1 and TLN1, including those respective proteins. **c)** Pairwise squared Pearson's correlation of the 151 unique proteins in the controls. **d)** Pairwise squared Pearson's correlation of the 151 unique proteins in the AD individuals.

2.6. Co-expression between proteins involved in metabolism disrupted in AD individuals

Here, a similar analysis, with the same individuals was performed as described in the previous section. However, proteins were investigated that were co-expressed in the ND control group and not correlated in the AD group (Fig. 1d). In total, 178 differentially correlated pairs were identified involving 111 unique proteins. The most significant differentially correlated protein pairs were *OXR1-SCRN1* ($\Delta r = 1.03$, $P_{FDR} = 1.23 \times 10^{-12}$, Fig. 6a) and *OXR1-ST13* ($\Delta r = 0.72$, $P_{FDR} = 2.58 \times 10^{-11}$, Fig. 6a). *FKBP4* was co-expressed with 24 proteins in the ND group while not in the AD group. *OXR1* was also co-expressed with 24 proteins in the ND group, which was also not observed in the AD group (Fig. 6c). Of the 178 differentially correlated protein pairs, 159 pairs also showed an interaction with Braak stage ($P_{FDR} \leq 0.05$) as the strength of linear association between proteins decreased with increasing Braak stage (Fig s7). The first principal component (97% of variance) of the 111 proteins of all 177 individuals was correlated with Braak stage ($r = 0.65$, $P = 2.48 \times 10^{-22}$) and age ($r = -0.16$, $P = 0.03$). The 111 proteins were enriched, among other terms, for metabolism pathway of REACTOME ($N = 35$, $P = 1.14 \times 10^{-07}$) and KEGG ($N = 32$, $P = 1.34 \times 10^{-07}$). For five proteins, pQTL variants were identified. Rs13027631, rs77916722, rs11967589, rs3118634 and rs9920103 for *TUBA1A*, *CD47*, *TUBB*, *PTPA* and *IDH3A*, respectively.

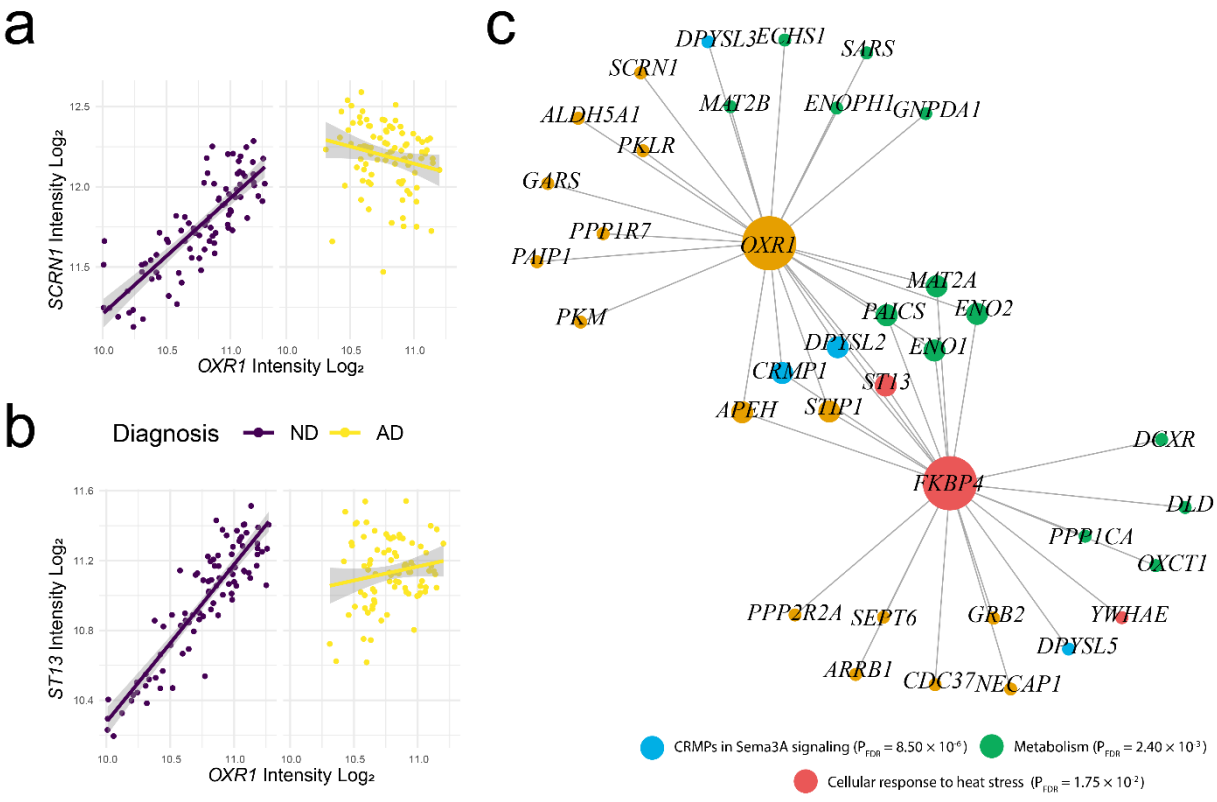

Figure 6 Overview of proteins specifically correlated in control. **a)** Log_2 intensity (x-axis) of OXR1 and SCRN1 (y-axis), with individuals divided in controls (ND, Purple) and AD individuals (yellow). **b)** Log_2 intensity (x-axis) of OXR1 and ST13 (y-axis), with individuals divided in controls (ND, Purple) and AD individuals (yellow). **c)** Network graph of all the proteins that are differentially correlated with OXR1 and FKBP4, including those respective proteins. Colors highlight the enriched

REACTOME pathway the respective proteins belong to. If a protein was involved in multiple pathways, then, the node is colored according to the most significant pathway.

3. Discussion

In the current study, we performed a large scale pQTL analysis in the GTM. We identified 153 proteins that show aberrant expression associated with 5861 variants. Of these variants, eleven were associated with increased AD risk. Among them, the well-known rs429358 variant which is associated with a 3x increase of AD risk[11]. Here, we show that rs429358 is also associated with increased expression of *APOE* in the GTM We also performed a differential correlation analysis of proteins with respect to fifteen known AD variants and showed that for these variants the correlation structure between certain proteins is distinct in individuals with different genotypes. In addition, we showed that between controls and individuals diagnosed with AD also distinct correlations structures between proteins exist. As such, we have identified disrupted co-expression networks in AD individuals and also identified the emergence of new co-expression networks in AD individuals.

In this study, we compared the identified pQTLs with previously identified pQTLs in the dorsolateral prefrontal cortex[16]. Robins et al,[16] included only cognitively healthy individuals, in contrast, our population consisted of ND controls, CHC and individuals diagnosed with AD. CHC as well as AD patients are likely enriched with genetic variants that are involved in the maintenance/disruption of their cognitive health[23], as such, our population includes more extreme phenotypes with respect to the AD spectrum. As a consequence, different pQTLs variants might have been identified. As both studies had divergent designs, different analyses have also been performed with different settings. In addition, brain regions have distinct protein expression profiles[24]. Despite all these the differences, of the 5861 pQTLs that we have identified in GTM, 2265 (37%) were also found in dorsolateral prefrontal cortex.

Additionally, we investigated whether pQTL variant in GTM are also eQTL variants in various brain regions (cerebellum, cortex, nucleus accumbens, caudate, cerebellar hemisphere, frontal cortex (BA9), hypothalamus, putamen, hippocampus, anterior cingulate cortex (BA24), amygdala and substantia nigra). Of an independent set of 154 pQTL variants, 53 (34.4%) variants were also an eQTL variant for the transcript of the respective protein. This corroborates the finding that genetic influence on mRNA and protein expression is distinct[16]. Most overlap between pQTL variants and eQTL variants was found with the cortex. As the temporal lobe is part of the cortex, it is not surprising that most overlap was found with this region. A recent study[25] observed decreasing correlation between mRNAs and their protein counterparts associated with increasing age in brains of *Nothobranchius furzeri*. The decrease in correlation is mainly attributed to decreased activity of post-transcriptional mechanisms. As such, age might also affect the degree of overlap found between pQTLs and eQTLs.

In this study, we identified 2196 differentially expressed proteins in AD cases. For 97 differentially expressed proteins, significant pQTLs were found. In an independent population, we tested

whether these pQTL variants can also be directly linked to AD risk. The reasoning being, that a pQTL variant associated with the expression of a protein with the same directional effect as was observed with differentially expression analysis might also, in part, explain genetic AD risk. This was only true for eleven pQTL variants, associated with two proteins (*APOE* and *SIRPA*). As AD is known for its genetic and clinical heterogeneity[26,27], identifying evidence for AD associated genetic control on the expression of proteins with relatively small and independent populations is difficult. Nevertheless, to the best of our knowledge, this is the first study in which rs429358, a variant associated with 3x increased risk for AD[11], has been associated with aberrant expression of *APOE* in any brain tissue. Interestingly, rs429358 was not found significantly associated with *APOE* in dorsolateral prefrontal cortex[16] (P = 0.51). In addition, rs429358 is also not a significant eQTL in any brain tissue for the *APOE* transcript. Finally, it should be noted that *SIRPA* was downregulated in AD individuals, while the AD risk genotype of the identified pQTLs were associated with increased expression of *SIRPA.*

DDX17 (DEAD-Box Helicase 17) is thought to play central role in functional consequences of rs9381040, as it was differentially correlated with sixteen protein. *DDX17* acts as a transcriptional coregulator for various target genes[28], is involved in the splicing machinery and splicosome[29] and plays a role in the regulation of alternative splicing[30,31]. *DDX17* has also been found to be involved innate immunity as it acts on viral infections by facilitating microRNA (miRNA) processing[32]. *DDX17* is a master regulator for the estrogen signaling pathway and plays a major role in the androgen signaling pathway[28]. The androgen signaling pathway has been linked to protective actions against neurodegenerative diseases as androgens have been found to negatively regulate β-amyloid accumulation[33]. In addition, androgen is tightly connected with the innate and adaptive immune system as it plays a role in regulating inflammatory response, development and activation of B cells and T cells[34]. Additionally, a central role for *TREML2* (Triggering Receptor Expressed On Myeloid Cells Like 2) also seems likely as rs9381040 located near this gene, *TREML2* was not measured in this study. *TREML2* is located on the human *TREM* gene cluster[35] together with *TREM1, TREM2, TREML1* and *TREM3.* In response to inflammatory factors, *TREML2* increases the expression of neutrophils and macrophages associated with immune responses[36]. As *TREML2* is known to initiate immune response, and *DDX17* and androgen are also tightly connected with immune responses, we hypothesize that the protective consequences of rs9381040 results from altered behavior of androgen and *DDX17* and is potentially initiated through immune related responses by *TREML2.* However, how *TREML2* and *DDX17* are functionally associated in relation to protective action against AD should be further elucidated.

In this study we identified 345 pairs proteins that were not correlated in ND controls and co-expressed in individuals with AD. *TLN1* and *ITGB1* were found differentially correlated with 26 and 25 proteins, respectively. As such, it is expected that these proteins may play important roles in the maintenance and disruption of protein networks in AD individuals. The protein *TLN1* (Talin 1) is located

between cells in the extracellular matrix and is involved in cell-cell adhesion[37]. *TLN1* is involved in integrin signaling and acts as adaptor protein to promote integrin-mediated signal transduction[38]. *TLN1* is of interest in cancer and hematological disease research as previous studies have shown that a dysregulation of *TLN1* is associated with cell disease states involving spreading, migration, and cell survival[37]. *ITGB1 (*Integrin Subunit Beta 1*)* is a receptor protein for *IL1B*[39] (Interleukine-1 beta) which is known to be involved in neuroinflammation and AD[40]. *ITGB1* also plays an important role in immune response[41]. Among the enriched pathways for the AD specific co-expression network, were, among many others, the REACTOME pathways immune system (N = 29, $P_{FDR} = 3.24\times 10^{-6}$) and innate immune system (N = 18, $P_{FDR} = 9.17 \times 10^{-5}$). Additionally, many proteins involved in signal transduction were found (N = 38, $P_{FDR} = 8.39 \times 10^{-8}$). Altogether, our results suggest that the AD specific co-expression network might be caused due to alternative signal transduction, and, as a consequence might initiate alternative immune response associated with AD pathology. As such, investigating the signal transduction proteins presented in this study in more detail (e.g. *in vitro* and *in vivo*), might result in more insight in regard to immune related responses and AD.

Besides emerging co-expression networks in AD, we also observed disrupted co-expression networks in AD individuals. Central proteins herein were *FKBP4* and *OXR1*, as both proteins were differentially correlated with 24 proteins. *FKBP4/FKBP52* (FKBP Prolyl Isomerase 4) is part of the *FKBP* (FK506 binding protein) immunophilins family. Members of the FKBP family are known to regulate tau and amyloid beta[42]. Additionally, *FKBP4/FKBP52* is involved in the translocation of glucocorticoid receptors into the nucleus and glucocorticoids are associated with an upregulation of tau and amyloid beta[43]. *OXR1* (Oxidation Resistance 1) was found differentially correlated with 24 proteins. In mice, Oxr1 is found to play an important role in controlling oxidative stress resistance of neuronal cells[44]. In absence of Oxr1, mice showed cerebellar neurodegeneration, while, with an overexpression, the neurons in mice were less susceptible to exogenous stress. Interestingly, *OXR1* was not found differentially expressed in AD individuals (logFC = 0.09, $P_{FDR} = 0.08$). Oxidative stress is known to increase with age and is known to be involved in neurodegeneration in AD[45]. Additionally, of the 24 proteins that were differentially correlated with *OXR1*, eleven were involved in metabolism.

A strength of this study is the inclusion of individuals with extreme phenotypes, such as individuals with increased genetic risk for AD and AD escapers; as such, power and effect size are increased for AD specific variants[18]. Also, it should be noted that an AD specific signature may be underlying some of our pQTL results, specifically in regard to variants associated with AD risk. Furthermore, pQTL variants were compared to synonymous eQTL variants in twelve brain regions, due to the brain region wide comparison we revealed shared genetic control of transcripts between investigated brain regions. In this study, we investigated differential correlation of proteins between individuals with different genotypes of AD associated variants. Although this is an unorthodox approach of investigating genetic consequences, a similar approach, including software was proposed and

described in 2015[46]. Adoption and widespread use never occurred for the approach. Our study shows promising results and might present novel research avenues for investigating genetic consequences of AD risk variants. However, the approach is not without its flaws. For instance, using this approach on a proteome- and genome-wide level is not feasible, as the number of tests is depended on the number of pair-wise tests between proteins times the number of variants. As such, a hypothesis driven approach is required, as was done in this study. Additionally, differential correlation can only be calculated between two groups, in general when investigating variants, three groups are present. Lastly, biological intuition in regard to this approach requires additional thought, as adoption lacks, biological interpretations of corresponding results are also virtually absent. Therefore, *in silico* results generated with this approach are ideally complemented with *in vitro* experiments, which could develop a better intuition of results generated with a differential correlated analysis.

4. Conclusion

In this study, we linked genetic variants to proteomics in the GTM and revealed association between AD status and altered protein expression and correlation.

We presented evidence of genetic control on protein expression in the GTM and were able to link eleven variants, associated with aberrant expression of *APOE* and *SIRPA*, with AD risk. Among these variants, was a pQTL for *APOE*, that is also part of the notorious ε4 allele of *APOE* (rs429358).

We showed that, if we want to understand genetic control on the brain proteome and its interaction with the brain transcriptome, region specific pQTL and eQTL studies are required, taking into account age and ideally with harmonized analyses, a consistent population and corresponding brain regions of eQTL and pQTL analyses.

Furthermore, with the differential correlation analysis with respect to the genotypes of AD risk variants, we show that this approach is a promising addition to GWAS- and QTL-studies, as it highlighted potential proteins functionally involved in downstream consequences of disease associated risk variants. In contrast to GWAS- and QTL-studies that point towards the closest gene or associated eQTLs/pQTLs as functional culprit, this approach expands on that and considers disrupted correlation structures associated with the respective variant. As such, the central regulatory protein *DDX17* is hypothesized to play in important role in the functional consequences of rs9381040.

Finally, with a differential correlation analysis with respect to AD status we identified the emergence of protein networks functionally enriched for signal transduction and the immune system, among other functions. A disruption of a protein network was observed that was functionally enriched for metabolism, among other functions. In conclusion, we presented a set of proteins that might be associated with neurodegenerative consequences of AD and is associated with a dysregulation of

metabolic processes and, alternative signal transduction that initiates a collaboration between proteins that is associated with a detrimental immune response.

5. Materials and Methods

5.1. Gyrus Temporalis Medialis Proteomics data

The proteomics data from the GTM was generated with the sequential window acquisition of all theoretical mass spectra (SWATH- MS) method with a data independent acquisition (DIA) approach. MaxQuant software[47] was used for spectrum annotation and relative protein quantification. The Uniprot human reference proteome[48] was used as reference. In total, 237 individuals were included, and 4,829 proteins were measured. 102 individuals were diagnosed with AD, 62 individuals were CHC and 73 ND controls were included.

5.2. Amsterdam Genetic data (AGD)

Individuals indicated as AD individuals were clinically diagnosed probable AD patients from the Amsterdam Dementia Cohort[49] (N=2668) or pathologically confirmed AD patients from the Netherlands Brain Bank[50] (N=436). The population consist of ***1)*** 1779 individuals aged 55-58 years from the Longitudinal Aging Study Amsterdam[51] (LASA), ***2)*** 1206 individuals with subjective cognitive decline that visited the memory clinic of the Alzheimer center Amsterdam and were labelled cognitively normal after extensive examination, ***3)*** 40 healthy individuals from the Netherlands Brain Bank, ***4)*** 201 individuals from the twin study[52] and ***5)*** 444 individuals from the 100-plus Study cohort[23]. The 100-plus Study cohort consist of Dutch-speaking individuals who have provided official evidence for being 100 years and older and self-reported to be cognitively healthy. The self-reported cognitive health was confirmed by the respective family members and their partners. For this study, a total of N=358 CHC and N=86 partners of centenarian's children.

The Medical Ethics Committee of the Amsterdam UMC (METC) approved all studies. All participants and/or their legal representatives provided written informed consent for participation in clinical and genetic studies.

5.3 Summary statistics pQTL study

The pQTL summary statistics of Robins et al.[16] were obtained from http://brainqtl.org. The pQTLs were identified in 144 healthy individuals that were originally subject in the ROSMAP study. The population consisted of 63.1% females and the median age was 86.5 with minimum of 67.4 and maximum of 102.7. The protein expression data is from the dorsolateral prefrontal cortex. Proteins were tested against variants that reside within 50 (Kbp) up- and downstream of the TSS of the respective proteins. In the available data, pQTL summary statistics of 7,901 proteins and 2,599,383 variants were presented. In total, 4,199,577 pQTLs were present. Proteins were defined by their Uniprot accession IDs and variants were defined by their GRCh37/hg19 genomic coordinate. P-values were corrected for multiple tests with

Bonferroni and FDR separately. With Bonferroni, 2,955 ($P_{BONF} \leq 0.05$) significant pQTLs were identified and, with FDR 28,211 ($P_{FDR} \leq 0.05$) significant pQTLs were identified.

5.4. eQTLs from GTEx

The eQTL data was accessed through the Application programming interface (API) of GTEx[19]. In total, twelve brain regions were tested. Each brain region had varying numbers of individuals available with genotype and RNA-seq data (Table S1). The total population is comprised 395 individuals of which 72% was male. GTEx does not report exact ages for the individuals but has specified the following age ranges, 20-29 (N = 8), 30-39 (N = 10), 40-49 (N = 36), 50-59 (N = 119), 60-69 (N = 200) and 70-79 (N = 22).

The eQTL statistics received from GTEx contains a normalized effect size (NES), which is the slope of the linear regression and it is the effect to the alternative allele (ALT) relative to the reference allele (REF) according to human genome reference GRCh38/hg38. The data also contains nominal p-values of the eQTL association and a p-value threshold. The p-value threshold is determined by $P_{FDR} \leq 0.05$ but is translated to a nominal p-value. Variants are defined by their reference SNP identification number (rs IDs). Transcripts are defined by their gene symbol and an Ensembl transcript ID.

5.5. Gyrus Temporalis Medialis Proteomics Quality Control and Pre-processing

Quality control was performed on a sample basis and protein basis separately. First, samples with more than 34% of low-quality ($Q \geq 0.01$) peptides were excluded from the analyses (N = 35). After removing low-quality samples, a reference peptide intensity distribution was calculated of the remaining samples by averaging all peptide intensity distributions. Distances between every individual peptide intensity distribution and the reference distribution was calculated with Kolmogorov–Smirnov test. Sample distributions with a greater distance (*D*) than 0.04 from the reference distribution were excluded from the analyses (N = 1). For replicates, lower quality samples were determined with a paired t-test on the quality measures. The lower quality samples were removed, as such, eleven replicates were excluded. The proteomics data was generated in bottom-up fashion, meaning, that peptides were measured and used to estimate the expression of their respective protein. When peptides that make up a single protein had a low-quality in more than 10% of the samples, the respective protein was excluded. Proteins were represented by the sum of intensities of their respective peptides. Finally, the protein intensities were $\log_2$ transformed to ensure normality of the protein intensity distributions. The final proteomics dataset consists of 3556 proteins and 190 individuals.

Next, batch effects were removed during the pre-processing step, as these effects have no biological meaning. First, age, sex, Braak stage I-VI, post-mortem delay (PMD), *APOE* genotype ($\log_2$ Polygenic Risk Score) and batch were tested on association with the variation in the proteomics data. This was done with the R-package variancePartition[53,54]. VariancePartition utilizes a mixed linear model to determine the percentage of variation that is associated to variables. Among the tested variables

substantial proportions of the variation was explained by age, Braak stage and batch. The `combat` function from the R-package sva[55] was used to remove the variation associated with batch from the protein intensity data.

5.6. Amsterdam Genetic data processing

Genetic variants were determined with standard genotyping and imputation methods, additionally, established quality control methods were applied. Genotyping of individuals was performed using Illumina Global Screening Array (GSAsharedCUSTOM_20018389_A2). High-quality genotyping in all individuals was used (individual call rate > 98%, variant call rate > 98%), individuals with sex mismatches were excluded and departure from Hardy–Weinberg equilibrium was considered significant at $P < 1\times10^{-6}$. Genotypes were prepared for imputation using provided scripts (HRC-1000G-check-bim.pl)[56], which compares variant ID, strand and allele frequencies to the Haplotype Reference Panel (HRC v1.1, April 2016)[57]. All autosomal variants were submitted to the Sanger imputation server (https://imputation.sanger.ac.uk). The server uses MACH to phase data. Imputation to the reference panel (HRC v1.1, April 2016) was performed with PBWT. 3,670 population subjects and 3,106 AD cases passed quality control. Prior to analysis, we excluded individuals of non-European ancestry based on 1000Genomes clustering and individuals with a family relation based on identity-by-descent > 0.2. This led to the exclusion of 205 population controls and 152 AD cases with non-European ancestry and 217 population controls and 100 AD cases with a family relation, leaving 4,191 population subjects and 2,416 AD cases for the analyses (total sample size = 6607).

5.7. pQTL identification

Genetic variants associated with protein expression were identified with Plink (v2.00a2LM)[58]. For the association, we used linear models with genotype dosages as predictors for protein expression, assuming additive genetic effects. Analyses were corrected for population substructure using the first five principal components (See supplements *pQTL linear regression model*). Resulting effect-sizes (β) were calculated with the minor allele relative to the major allele in the investigated population. Association P-values were corrected for multiple tests with False Discovery Rate (FDR) and significance was assumed when $P_{FDR} \leq 0.05$. The analyses were restricted to variants with a MAF higher than 5% and variants located 250 Kbp down- and upstream of the TSS of the respective proteins. Four window sizes were tested (50 Kbp, 250 Kbp, 500 Kbp and 1 Mb, Fig. s5). With 250 Kbp, we reduced to total number of tests, while still capturing most of pQTLs. Genomic locations of the TSSs were acquired with biomaRt (v2.42.0)[59,60]. The retrieved genomic locations of the TSSs were for genomic build GRCh38/hg38. The liftOver R-package (v1.10.0)[61,62] was used to lift over the genomic coordinates to build GRCh37/hg19, as the genotype files were based on this genomic build. An independent set of pQTLs in linkage equilibrium was derived using LD-based clumping. Variants located near each other are often in linkage disequilibrium, which means that they are correlated. Variants that are correlated have similar associations with the same protein. With clumping, only the strongest associating variant remains within

a certain window. We clumped all variants for each protein individually on $R^2 \geq 0.001$ and MAF ≥ 0.05. Clumping was done with Plink (v1.90b4.6)[58] European individuals from the 1,000 Genomes Project reference panel[21] was used to calculate the linkage disequilibrium between variants.

5.8. Testing pQTL variants on association with AD risk

To identify pQTL variants that are also associated with AD risk, we first performed a differential expression analysis on protein intensities between ND controls and AD cases. This analysis was performed on the *Gyrus Temporalis Medialis Proteomics data,* on 141 individuals, 88 individuals diagnosed with AD and 53 ND controls. CHC (N = 53) were excluded from this analysis. Differentially expressed proteins were identified with a moderate t-test[63] using the R-package limma[64]. P-values were FDR corrected for multiple tests. Significance was assumed $P_{FDR} \leq 0.05$.

Next, for every significant differentially expressed protein, it was checked whether the previous pQTL analysis yielded a significant association with a genetic variant. When a differentially expressed protein was also associated with a pQTL variant, the respective pQTL variant was tested on association with AD risk. This was done with the Amsterdam Genetic data (AGD) population (N = 6,479) comprising of 2,361 AD cases and 4,118 ND controls. To have an independent population, individuals from AGD that were used to identify pQTLs were excluded (N = 128). As such, there was no overlap of individuals between the population that was used for the differential expression analysis and the population that was used for the genetic association tests. To test the association of pQTL variants with AD status, a logistic regression model in R (v3.6.3) was used with AD status as discrete outcome variable (ND = 0, AD = 1) and the pQTL variant's genotypes as predictor variable. The model was adjusted for population substructure (principal component 1-5). P-values were adjusted for multiple tests with FDR and significant association was assumed with $P_{FDR} \leq 0.05$.

5.9. Temporalis gyrus medius and dorsolateral prefrontal cortex pQTL comparison

For all the significant pQTLs we checked whether they previously have been found in the dorsolateral prefrontal cortex. For this, we used the pQTL summary statistics from Robins et al [16] made available on http://brainqtl.org. Proteins were matched on UniProt[48] accession IDs and variants were matched on genomic locations. From both studies, only pQTLs were selected that were FDR significant at $P_{FDR} \leq 0.05$. Of matching pQTLs, the directional effect of the pQTLs were compared by means of calculating the Pearson's correlation coefficient between effect sizes in R (v3.6.3)[65] with the `cor.test` function.

5.10. pQTL and eQTL comparison

For all the significant pQTLs we checked whether they were also an eQTL variant. This was done with eQTL data from twelve brain tissues (Table S1) from GTEx (v8)[19]. Brain region wide and brain region specific associations of pQTLs from the GTM with eQTLs were investigated. Every gene - variant pair was queried from the GTEx API for the twelve brain tissues. For this, the `get_eQTL_bulk` function was used from the R-package CONQUER (v1.0)[66], which requires tissue ID, gene symbol and RS ID to

be supplied. The P-value thresholds supplied by GTEx were used to determine significance of the tested eQTLs. The directional effects of pQTLs with their synonymous eQTLs were compared by calculating the Pearson's correlation coefficient between effect sizes with the `cor.test` function.

5.11. Differential correlation

We investigated whether correlation between proteins is subject to change when comparing two distinct groups. Differential correlation with respect to phenotype status was investigated, and differential correlation with respect to AD variants genotype was investigated. As both analyses require separate approaches, both are separately discussed in subsequent sections. However, the statistical procedure to determine differential correlation is shared, as such, that is discussed here.

First, Pearson's correlation between a pair proteins is calculated separately for the groups of interest. Assuming two groups *x* and *y* results in two correlation coefficients r_x for group x and r_y for group y. Next, the correlation coefficients are translated to z-scores. This is done with the Fisher z-transformation[67] (Eq. 1).

$$z = atanh(r) = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \quad (1)$$

Then, the difference between z-scores z_x and z_y can be calculated with equation 2. Where $var(r)$ is calculated by $\frac{1}{n-3}$. Here, n is the sample size of the respective groups.

$$\Delta z = \frac{(z_x - z_y)}{\sqrt{var(r_x) + var(r_y)}} \quad (2)$$

As Δz is normally distributed, a two-sided P-value for the differential correlation between each pair of proteins can be calculated.

5.12. Differential correlation with respect to AD variants genotype

In this analysis, the intersect of individuals from AGD and GTM proteomics were subject, 67 AD individuals, 27 ND controls and 46 CHC. With these individuals we performed differential correlation analysis of proteins with respect to known AD variants. As starting point 41 variants[11] that are known risk variants for AD risk were considered. From the 41 variants we selected variants where each available genotype was at least present in 10 individuals. The minimal number of individuals was set to N = 10, as this reduces disparity between population sizes. Additionally, small population sizes increase the chance on false positives. When all three genotypes were present, the differential correlation was calculated between the two homozygous genotypes. In this case, the correlation between proteins with the heterozygous genotype is only reported. When only two genotypes were present, the differential correlation was determined between the homozygous genotype and heterozygous genotype. For the variants, the differential correlation method was implemented in R (v3.6.3)[65]. The p-values were FDR

corrected for total number of tests (94,811,850, i.e. differential correlation was calculated for all possible pairs of proteins, fifteen times). Significance was assumed at $P_{FDR} \leq 0.05$.

5.13. Differential correlation with respect to phenotype status

The individuals subject in this analysis were the individuals for which GTM proteomics data was available. CHC with Braak stage ≥ 4 were exclude from this analysis. The remaining CHC (Braak stage ≤ 3) were considered ND controls. Altogether, this analysis involved 177 individuals, 88 AD individuals, and 89 ND controls. All 3,556 proteins measured in the GTM were pairwise tested for differential correlation in ND controls and AD individuals. Here, the `ddcorAll` function of the R-package DGCA[68] (v1.0.2) was used, which calculates differential correlation as described in the previous section (See methods *Differential correlation*). P-values were adjusted for multiple tests with FDR. Two classes of differential correlation were defined that intuitively have different biological meanings. The first class includes protein pairs that are specifically co-expressed in AD individuals, and not in controls ($r^2 \leq 0.20$ in controls, $r^2 > 0.65$ in AD individuals). As such, this reveals co-expression between proteins that emerges in AD individuals. The second class includes pairs of proteins specifically co-expressed in controls, and not correlated in AD individuals ($r^2 > 0.65$ in controls, $r^2 \leq 0.20$ in AD individuals). This reveals co-expression between proteins that gets disrupted in AD individuals. Here, R-squared was used as threshold measure as it captures co-expression in positive and negative direction simultaneously.

5.14. Braak interaction models and principal component analysis

The interaction of Braak stage with the linear associations between the differentially correlated proteins pairs were tested. For this, a linear regression model was utilized with the $\log_2$ intensity of one protein (P_y) as outcome variable and the $\log_2$ intensity of the other protein (P_x) and Braak stage (0-6) as predictor variables and an added interaction term between the predictor protein (P_x) and Braak stage (Eq. 3).

$$P_y = \beta_0 + \beta_1 P_x + \beta_2 Braak + \beta_3 P_x Braak \quad (3)$$

Where, β_i are the regression coefficients. Identified protein networks were tested on association with Braak stage and age. For this, principal component analyses were performed with the `prcomp` function on the intensities of the respective proteins. The first principal component was tested on association with Braak stage and age by calculating the Pearson's correlation coefficient with the `cor.test` function. Pathway enrichments were performed on the webserver of String (v11)[22].

References

1. Braak H, Alafuzoff I, Arzberger T, Kretzschmar H, Tredici K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol*. 2006;112(4):389-404. doi:10.1007/s00401-006-0127-z

2. Thies W, Bleiler L. 2012 Alzheimer's disease facts and figures. *Alzheimer's Dement*. 2012;8(2):131-168. doi:10.1016/j.jalz.2012.02.001

3. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63(2):168-174. doi:10.1001/archpsyc.63.2.168

4. Kim J, Basak JM, Holtzman DM. The Role of Apolipoprotein E in Alzheimer's Disease. *Neuron*. 2009;63(3):287-303. doi:10.1016/j.neuron.2009.06.026

5. Riedel BC, Thompson PM, Brinton RD. Age, APOE and sex: Triad of risk of Alzheimer's disease. *J Steroid Biochem Mol Biol*. 2016;160:134-147. doi:10.1016/j.jsbmb.2016.03.012

6. Blacker D, Haines JL, Rodes L, et al. ApoE-4 and age at onset of Alzheimer's disease: The NIMH genetics initiative. *Neurology*. 1997;48(1):139-147. doi:10.1212/WNL.48.1.139

7. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1088-1093. doi:10.1038/ng.440

8. Lambert JC, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1094-1099. doi:10.1038/ng.439

9. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA - J Am Med Assoc*. 2010;303(18):1832-1840. doi:10.1001/jama.2010.574

10. Desikan RS, Schork AJ, Wang Y, et al. Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. *Circulation*. 2015;131(23):2061-2069. doi:10.1161/CIRCULATIONAHA.115.015489

11. Rojas I de, Moreno-Grau S, Tesi N, et al. Common variants in Alzheimer's disease: Novel association of six genetic variants with AD and risk stratification by polygenic risk scores. *medRxiv*. January 2020:19012021. doi:10.1101/19012021

12. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science (80-)*. 2012;337(6099):1190-1195. doi:10.1126/science.1222794

13. Guo Y, Xiao P, Lei S, et al. How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta Biochim Biophys Sin (Shanghai)*. 2008;40(5):426-436. doi:10.1111/j.1745-7270.2008.00418.x

14. De Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C. Global signatures of protein and mRNA expression levels. *Mol Biosyst*. 2009;5(12):1512-1526. doi:10.1039/b908315d

15. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012;13(4):227-232. doi:10.1038/nrg3185

16. Robins C, Wingo AP, Fan W, et al. Genetic control of the human brain proteome. *bioRxiv*. November 2019:816652. doi:10.1101/816652

17. Patel A, Fowler JB. *Neuroanatomy, Temporal Lobe*. StatPearls Publishing; 2019. http://www.ncbi.nlm.nih.gov/pubmed/30137797. Accessed June 16, 2020.

18. Tesi N, van der Lee SJ, Hulsman M, et al. Centenarian controls increase variant effect sizes by an average twofold in an extreme case–extreme control analysis of Alzheimer's disease. *Eur J Hum Genet*. 2019;27(2):244-253. doi:10.1038/s41431-018-0273-5

19. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653

20. Hunt SE, Mclaren W, Gil L. Ensembl variation resources. *Database*. 2018;2018:1-12. doi:10.1093/database/bay119

21. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

22. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613. doi:10.1093/nar/gky1131

23. Holstege H, Beker N, Dijkstra T, et al. The 100-plus Study of cognitively healthy centenarians: rationale, design and cohort description. *Eur J Epidemiol*. 2018;33(12):1229-1249. doi:10.1007/s10654-018-0451-3

24. Xu J, Patassini S, Rustogi N, et al. Regional protein expression in human Alzheimer's brain correlates with disease severity. *Commun Biol*. 2019;2(1). doi:10.1038/s42003-018-0254-9

25. Kelmer Sacramento E, Kirkpatrick JM, Mazzetto M, et al. Reduced proteasome activity in the aging brain results in ribosome stoichiometry loss and aggregation. *Mol Syst Biol*. 2020;16(6). doi:10.15252/msb.20209596

26. Nacmias B, Bagnoli S, Piaceri I, Sorbi S. Genetic Heterogeneity of Alzheimer's Disease: Embracing Research Partnerships. *J Alzheimer's Dis*. 2018;62(3):903-911. doi:10.3233/JAD-170570

27. De Waal H, Stam CJ, Blankenstein MA, Pijnenburg YAL, Scheltens P, Van Der Flier WM. EEG abnormalities in early and late onset Alzheimer's disease: Understanding heterogeneity. *J Neurol Neurosurg Psychiatry*. 2011;82(1):67-71. doi:10.1136/jnnp.2010.216432

28. Samaan S, Tranchevent LO-C, Dardenne E, et al. The Ddx5 and Ddx17 RNA helicases are cornerstones in the complex regulatory array of steroid hormone-signaling pathways. doi:10.1093/nar/gkt1216

29. Liu Z-R. p68 RNA Helicase Is an Essential Human Splicing Factor That Acts at the U1 snRNA-5' Splice Site Duplex. *Mol Cell Biol*. 2002;22(15):5443-5450. doi:10.1128/mcb.22.15.5443-5450.2002

30. Dardenne E, Pierredon S, Driouch K, et al. Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. *Nat Struct Mol Biol*. 2012;19(11):1139-1146. doi:10.1038/nsmb.2390

31. Germann S, Gratadou L, Zonta E, et al. Dual role of the ddx5/ddx17 RNA helicases in the control of the pro-migratory NFAT5 transcription factor. *Oncogene*. 2012;31(42):4536-4549. doi:10.1038/onc.2011.618

32. Moy RH, Cole BS, Yasunaga A, et al. Stem-loop recognition by DDX17 facilitates miRNA processing and antiviral defense. *Cell*. 2014;158(4):764-777. doi:10.1016/j.cell.2014.06.023

33. Pike CJ, Nguyen TV V., Ramsden M, Yao M, Murphy MP, Rosario ER. Androgen cell signaling pathways involved in neuroprotective actions. *Horm Behav*. 2008;53(5):693-705. doi:10.1016/j.yhbeh.2007.11.006

34. Lai JJ, Lai KP, Zeng W, Chuang KH, Altuwaijri S, Chang C. Androgen receptor influences on body defense system via modulation of innate and adaptive immune systems: Lessons from conditional AR knockout mice. *Am J Pathol*. 2012;181(5):1504-1512. doi:10.1016/j.ajpath.2012.07.008

35. Allcock RJN, Barrow AD, Forbes S, Beck S, Trowsdale J. The human TREM gene cluster at

6p21.1 encodes both activating and inhibitory single IgV domain receptors and includes NKp44. *Eur J Immunol*. 2003;33(2):567-577. doi:10.1002/immu.200310033

36. Klesney-Tait J, Turnbull IR, Colonna M. The TREM receptor family and signal integration. *Nat Immunol*. 2006;7(12):1266-1273. doi:10.1038/ni1411

37. Wei X, Sun Y, Wu Y, et al. Downregulation of Talin-1 expression associates with increased proliferation and migration of vascular smooth muscle cells in aortic dissection. *BMC Cardiovasc Disord*. 2017;17(1):162. doi:10.1186/s12872-017-0588-0

38. Manevich E, Grabovsky V, Feigelson SW, Alon R. Talin 1 and paxillin facilitate distinct steps in rapid VLA-4-mediated adhesion strengthening to vascular cell adhesion molecule. *J Biol Chem*. 2007;282(35):25338-25348. doi:10.1074/jbc.M700089200

39. Ma G, Liu M, Du K, et al. Differential Expression of mRNAs in the Brain Tissues of Patients with Alzheimer's Disease Based on GEO Expression Profile and Its Clinical Significance. 2019. doi:10.1155/2019/8179145

40. White CS, Lawrence CB, Brough D, Rivers-Auty J. Inflammasomes as therapeutic targets for Alzheimer's disease. *Brain Pathol*. 2017;27(2):223-234. doi:10.1111/bpa.12478

41. Li Z, Xiong ZZ, Manor LC, Cao H, Li T. Integrative computational evaluation of genetic markers for Alzheimer's disease. *Saudi J Biol Sci*. 2018;25(5):996-1002. doi:10.1016/j.sjbs.2018.05.019

42. Cao W, Konsolaki M. FKBP immunophilins and Alzheimer's disease: A chaperoned affair. doi:10.1007/s12038

43. Blair LJ, Baker JD, Sabbagh JJ, Dickey CA. The emerging role of peptidyl-prolyl isomerase chaperones in tau oligomerization, amyloid processing, and Alzheimer's disease. *J Neurochem*. 2015;133(1):1-13. doi:10.1111/jnc.13033

44. Oliver PL, Finelli MJ, Edwards B, et al. Oxr1 Is Essential for Protection against Oxidative Stress-Induced Neurodegeneration. Orr HT, ed. *PLoS Genet*. 2011;7(10):e1002338. doi:10.1371/journal.pgen.1002338

45. Tönnies E, Trushina E. Oxidative Stress, Synaptic Dysfunction, and Alzheimer's Disease. *J Alzheimer's Dis*. 2017;57(4):1105-1121. doi:10.3233/JAD-161088

46. Lareau CA, White BC, Montgomery CG, McKinney BA. DcVar: A method for identifying common variants that modulate differential correlation structures in gene expression data. *Front Genet*. 2015;6(OCT). doi:10.3389/fgene.2015.00312

47. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367-1372. doi:10.1038/nbt.1511

48. D506-D515. UniProt: a worldwide hub of protein knowledge The UniProt Consortium. *Nucleic Acids Res*. 2019;47. doi:10.1093/nar/gky1049

49. Van Der Flier WM, Scheltens P. Amsterdam dementia cohort: Performing research to optimize care. *J Alzheimer's Dis*. 2018;62(3):1091-1111. doi:10.3233/JAD-170850

50. Rademaker MC, de Lange GM, Palmen SJMC. The Netherlands Brain Bank for Psychiatry. In: *Handbook of Clinical Neurology*. Vol 150. Elsevier B.V.; 2018:3-16. doi:10.1016/B978-0-444-63639-3.00001-3

51. Huisman M, Poppelaars J, Van Der Horst M, et al. Cohort Profile: The Longitudinal Aging Study Amsterdam How did the study come about? *Int J Epidemiol*. 2011;40:868-876.

doi:10.1093/ije/dyq219

52. Willemsen G, De Geus EJC, Bartels M, et al. The Netherlands twin register biobank: A resource for genetic epidemiological studies. *Twin Res Hum Genet*. 2010;13(3):231-245. doi:10.1375/twin.13.3.231

53. Hoffman GE, Schadt EE. variancePartition: Interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics*. 2016;17(1):483. doi:10.1186/s12859-016-1323-z

54. Hoffman GE, Roussos P. dream: Powerful differential expression analysis for repeated measures designs. *bioRxiv*. January 2018:432567. doi:10.1101/432567

55. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD, Kelso J. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinforma Appl NOTE*. 2012;28(6):882-883. doi:10.1093/bioinformatics/bts034

56. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287. doi:10.1038/ng.3656

57. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

58. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7. doi:10.1186/s13742-015-0047-8

59. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat Protoc*. 2009;4(8):1184-1191. doi:10.1038/nprot.2009.97

60. Durinck S, Moreau Y, Kasprzyk A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinforma Appl NOTE*. 2005;21(16):3439-3440. doi:10.1093/bioinformatics/bti525

61. Bioconductor - liftOver. https://www.bioconductor.org/packages/release/workflows/html/liftOver.html. Accessed April 27, 2020.

62. Hinrichs AS. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res*. 2006;34(90001):D590-D598. doi:10.1093/nar/gkj144

63. Yu L, Gulati P, Fernandez S, Pennell M, Kirschner L, Jarjoura D. Fully moderated T-statistic for small sample size gene expression arrays. *Stat Appl Genet Mol Biol*. 2011;10(1). doi:10.2202/1544-6115.1701

64. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7). doi:10.1093/nar/gkv007

65. R Core Team. R: A Language and Environment for Statistical Computing. 2020. https://www.r-project.org/.

66. Bouland GA, Beulens JW, Nap J, et al. Understanding functional consequences of type 2 diabetes risk loci using the universal data integration and visualization R package CONQUER. *bioRxiv*. March 2020:2020.03.27.011627. doi:10.1101/2020.03.27.011627

67. Fisher RA. Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika*. 1915;10(4):507. doi:10.2307/2331838

68. McKenzie AT, Katsyv I, Song WM, Wang M, Zhang B. DGCA: A comprehensive R package for

Differential Gene Correlation Analysis. *BMC Syst Biol*. 2016;10(1). doi:10.1186/s12918-016-0349-1

Supplements

pQTL linear regression model

For identifying significant pQTLs, the generalized linear model (GLM) from Plink (v2.00a2LM)[58] is used. Which is the primary association analysis method in Plink for quantitative phenotypes. The model applied on our data is as follows:

$$P_y = \beta_0 G + \beta_1 PC_1 + \cdots + \beta_5 PC_5 + e$$

Where:

P_y is the $\log_2$ intensity for the individuals of respective protein (quantitative phenotype).

G are the dosages for the individuals of the respective variant that is tested.

PC_i are the principal components for the individuals of the population substructure.

e is an error term that gets minimized with least squares minimization.

Table s1: Genotyped and RNAseq sample sizes from GTEx for all twelve investigated brain regions

Tissue	# RNASeq and Genotyped samples	# RNASeq Samples
Cerebellum	209	241
Cortex	205	255
Nucleus accumbens (basal ganglia)	202	246
Caudate (basal ganglia)	194	246
Cerebellar Hemisphere	175	215
Frontal Cortex (BA9)	175	209
Hypothalamus	170	202
Putamen (basal ganglia)	170	205
Hippocampus	165	197
Anterior cingulate cortex (BA24)	147	176
Amygdala	129	152
Substantia nigra	114	139

Table S2: AD risk variants subject in differential correlation analysis

RS ID	Chromosome	Genomic location	Closest Gene	Genotypes		
rs6733839	2	127892810	*BIN1*	C/C 43	C/T 71	T/T 26
rs9381040	6	41154650	*TREML2*	C/C 78	C/T 51	T/T 11
rs1859788	7	99971834	*PILRA*	A/A 13	A/G 58	G/G 69

rs73223431	8	27219987	*PTK2B*	C/C 43	C/T 81	T/T 16
rs9331896	8	27467686	*CLU*	C/C 18	C/T 64	58 T/T
rs34674752*	8	145154222	*SHARPIN*	A/A 0	G/A 11	G/G 129
rs7920721	10	11720308	*ECHDC3*	A/A 60	A/G 56	G/G 24
rs3740688	11	47380340	*SPI1*	G/G 22	G/T 68	T/T 50
rs1582763	11	60021948	*MS4A4A*	A/A 18	G/A 77	G/G 45
rs3851179	11	85868640	*PICALM*	C/C 53	T/C 66	T/T 21
rs11218343*	11	121435587	*SORL1*	C/C 0	T/C 12	T/T 128
rs12444183	16	81773209	*PLCG2*	A/A 20	A/G 65	G/G 55
rs4311	17	61560763	*ACE*	C/C 34	T/C 75	T/T 31
rs12459419	19	51728477	*CD33*	C/C 68	C/T 58	T/T 14
rs2154481	21	27473875	*APP*	C/C 35	C/T 64	T/T 41

*** = variant of which two genotypes were present in population**

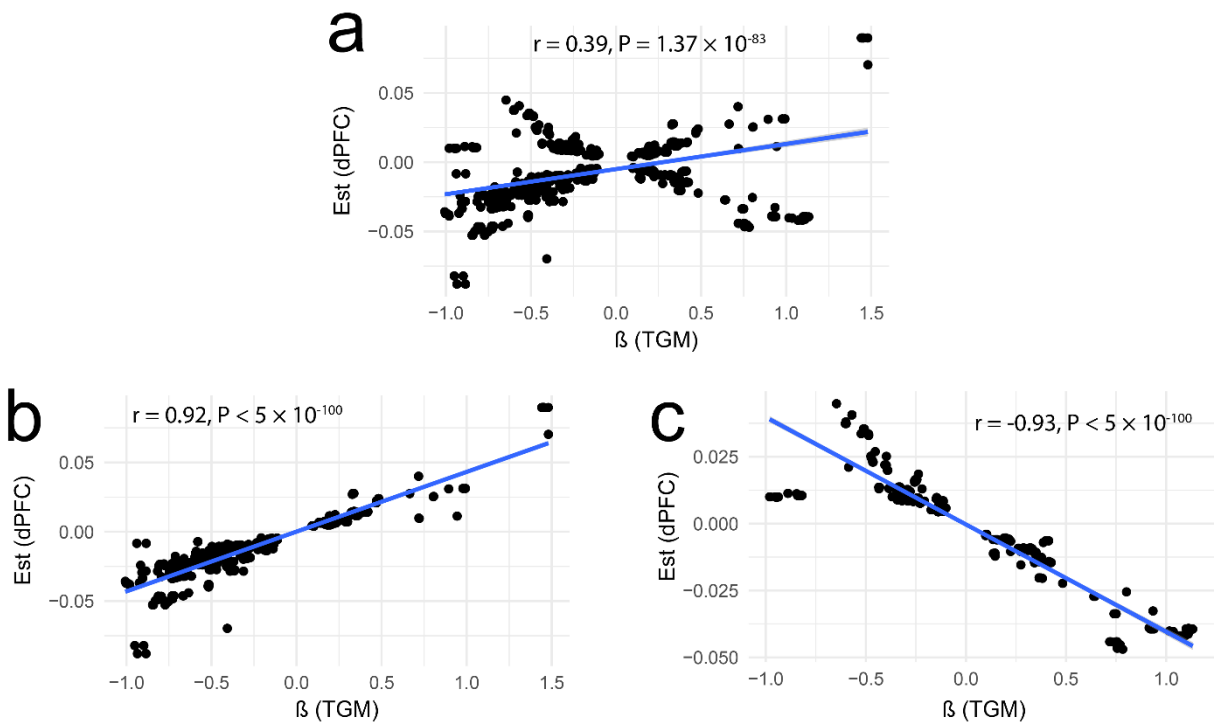

Figure s1 Estimate comparison of pQTL studies, in all sub-figures the x-axis represents the estimates of this current study and y-axis represents the estimates from [16]. a) Estimates of all matching pQTLs. b) Estimates of all matching pQTLs where the directional effects were the same. c) Estimates of all matching pQTLs where the directional effects opposite.

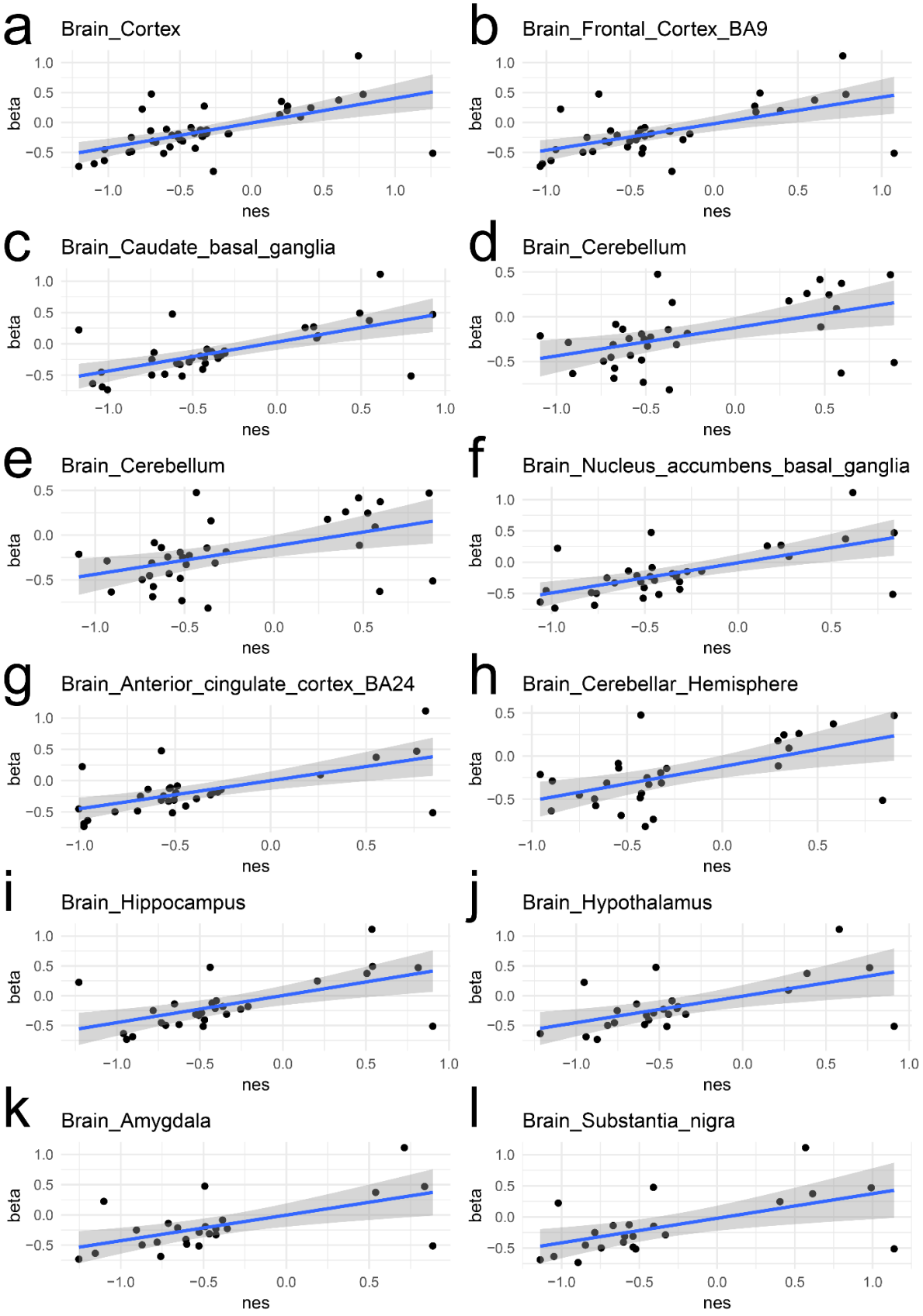

Figure s2 Estimate comparison of pQTLs versus the eQTL NESs from GTEx for all investigated brain regions. X-axes represent the NESs from GTEx for a particular eQTL – eGene pair. The y-axes represent the betas of the pQTL – protein pair synonymous for the eQTL – eGene pair.

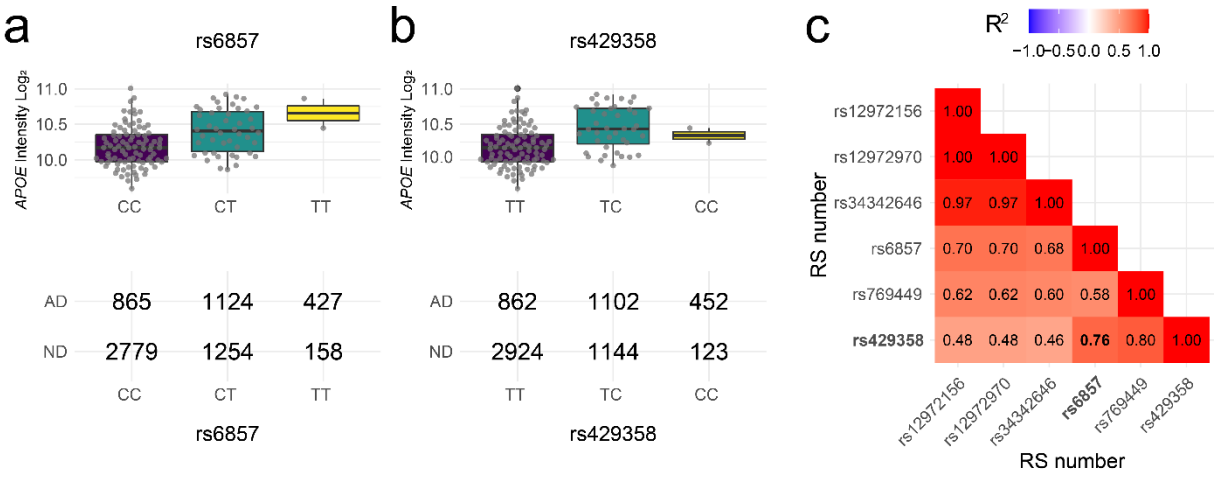

Figure s3 Overview of APOE associated pQTL variants. a) Boxplot of rs6857 genotypes versus APOE intensity, x-axis represent the genotypes, y-axis represents the log_2 normalized intensity of APOE. b) Boxplot of rs429358 genotypes versus APOE intensity, x-axis represent the genotypes, y-axis represents the log_2 normalized intensity of APOE. c) LD correlation between the six pQTL variants associated with APOE.

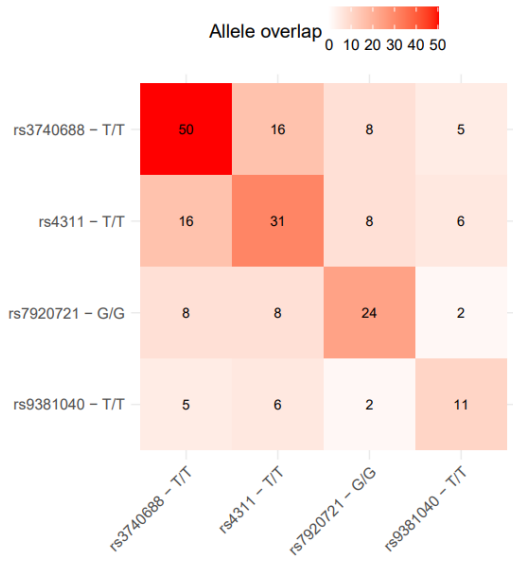

Figure s4 Overlap of individuals between the high-correlation alleles, the diagonal is the number of individuals carrying the single allele of the respective variant.

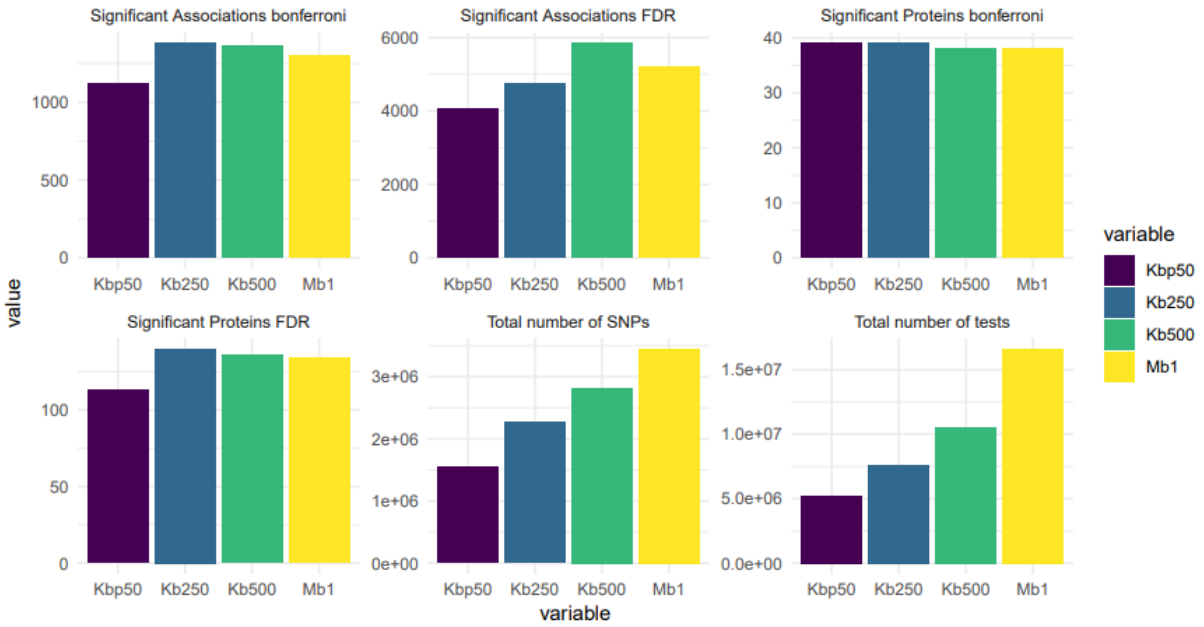

Figure s5 Overview of pQTL variant mapping window. X-axes represent the mapping windows of 50 Kbp, 250 Kbp, 500 Kbp and 1 Mb. The y-axes represent the count of the respective statistic that is shown. The title above each plot is the respective statistic.

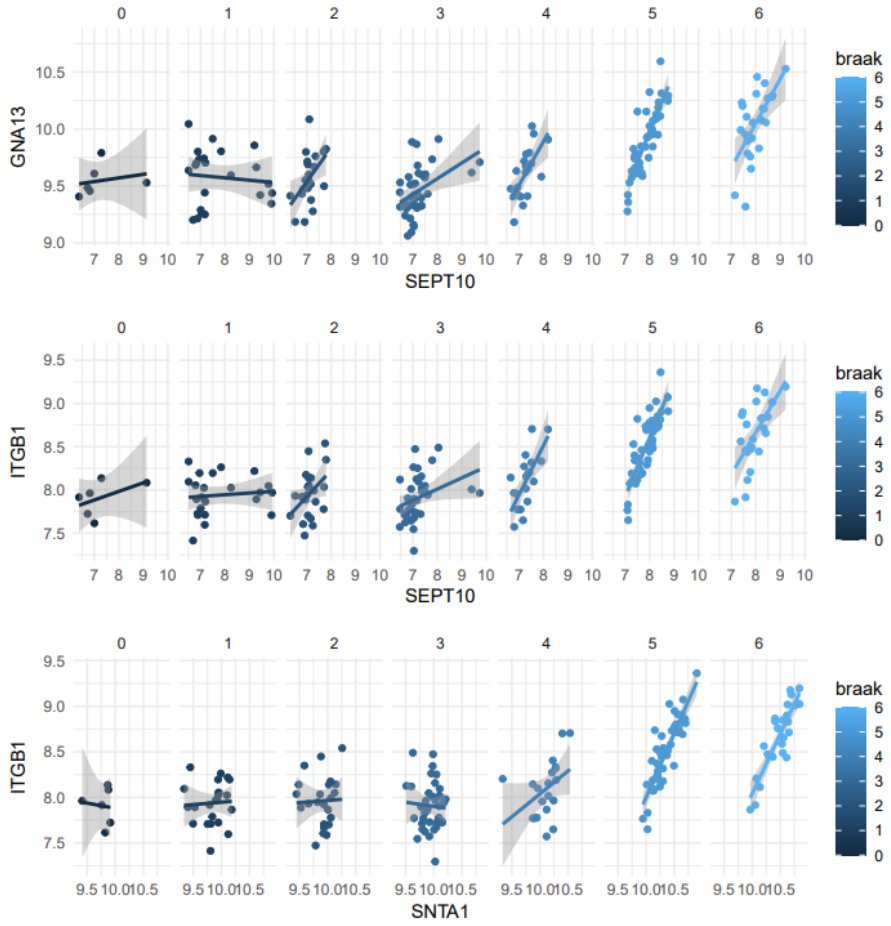

Figure s6 Top three protein pairs of which the linear interaction had an interaction with the Braak stage. The pairs were also significantly differentially correlated between AD individuals and controls (controls = $r^2 \leq 0.2$, AD = $r^2 \geq 0.65$). The x-axis

represents the log_2 intensity of the respective protein and the y-axis also represents the log_2 intensity of the respective protein. Individuals are divided on their Braak stage. The Braak stage is also represented by the color of the dots.

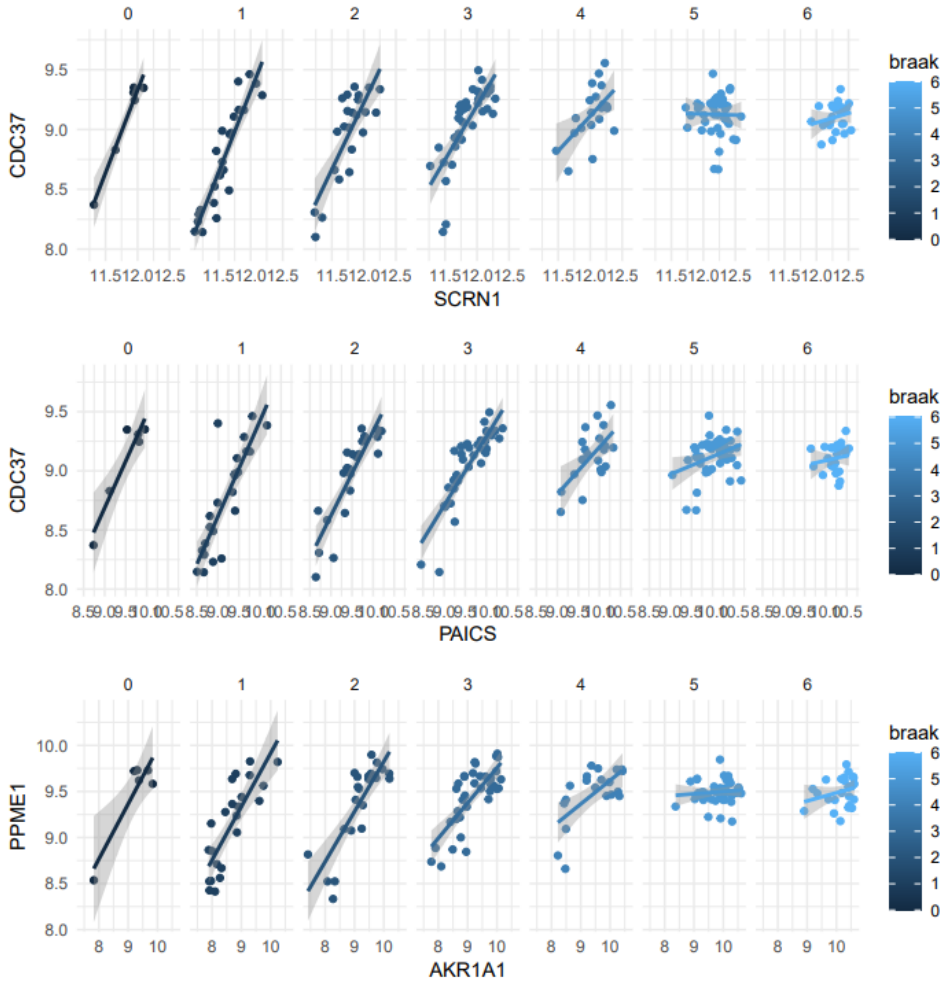

Figure s7 Top three protein pairs of which the linear interaction had an interaction with the Braak stage. The pairs were also significantly differentially correlated between control and AD individuals (controls = $r^2 \geq 0.65$, AD = $r^2 \leq 0.2$). The x-axis represents the log_2 intensity of the respective protein and the y-axis also represents the log_2 intensity of the respective protein. Individuals are divided on their Braak stage. The Braak stage is also represented by the color of the dots.

6. Appendix

6.1. Polygenic risk score analysis

6.1.1. Background

Polygenic risk scores (PRSs) are composite scores for an individual to determine their increased or decreased genetic risk for a disease. PRSs are defined as the sum of risk alleles of disease associated variants, where each variant is weighted by their individual risk. PRSs can be used to identify individuals with increased risk for diseases but can also be used for precision and personalized medicine. Here, we constructed a PRS based on pQTLs that are associated with AD. The pQTLs that were selected to construct the PRS were associated with AD in two ways. 1) The respective protein had to be significantly differentially expressed in AD individuals, and 2) The variant itself had to be associated with AD risk. As our PRSs have associated proteins, they can be functionally analyzed and put into context with AD pathology and potentially reveal targets for precision medicine.

6.1.2. Materials

This analysis uses the same data as described in the main report (See *Proteomics data* and *Genetic data*). However, an additional dataset is used, which contains the results of an AD meta-GWAS[11]. It should be noted that our cohort / genetic data was part of the meta-GWAS. For the meta-GWAS, summary statistics from three studies were combined. Spanish case-control study (GR@ACE/DEGESCO study, N = 12,386), case-control study of International Genomics of Alzheimer project (IGAP, N = 82,771) and UK Biobank(UKB) AD-by-proxy case-control study (N = 314,278). In total, risk for AD was determined for 16,358,696 variants.

6.1.3. Methods

pQTL selection

First, we started with the pQTLs identified in this study (See *Identified pQTL variants associated with abundance of 153 proteins in Gyrus Temporalis Medialis*). Here, 5,861 significant pQTLs were identified. Next, we selected variants (Fig. A1.1) from the meta-GWAS ($P \leq 0.005$) that were also a significant pQTL variant ($P_{FDR} \leq 0.05$). In the next step, we filtered the remaining pQTLs on whether the respective protein was also differentially expressed in AD individuals (Fig. a1.2). This was done with the differential expression analysis described in *Testing pQTL variants on association with AD risk.* In total, 236 pQTL variants were used for PRS construction.

PRS construction

An individuals` PRS is constructed was follows:

$$PRS = \sum_{p}^{P} D_p \beta_p$$

Where, *P* is the set of 236 pQTL variants and *D* is the predicted genotype dosage of variant *p*, which are determined by imputation (See Methods: *Amsterdam Genetic data processing*). *β* is the weight assigned to variant *p*. This weight was derived from the effect size from the meta-GWAS of the respective variant.

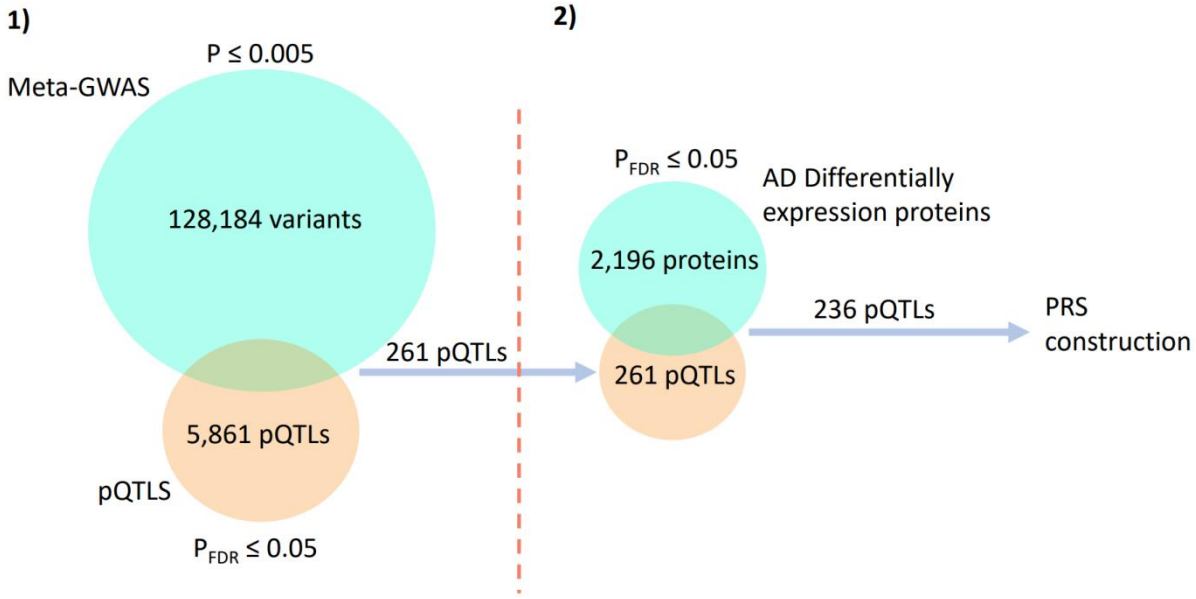

Figure A1 Workflow for pQTLs selection to construct the PRS with. **1)** Significant pQTLs that are also associated with AD risk according to the meta GWAS. **2)** pQTLs of which the respective protein was also differentially expressed in AD individuals.

Validation

To validate the PRS, we calculated PRSs for all individuals for which genetic data was available. As such, validation was performed on 6607 individuals for which genotypes were available. 2,416 individuals diagnosed with AD and 4,191 controls. These individuals were also part of the meta-GWAS, as such, it is not a proper validation as it is not an independent set of individuals. The performance of the PRSs was tested with a logistic regression model, with AD status as outcome variable (ND = 0, AD = 1) and the PRS as predictor variable. Two separate PRSs were constructed for each individual, one including *APOE* variants and one without the *APOE* variants.

6.1.4. Results

The 236 pQTLs were associated with eight proteins (*PITRM1*, *MADD*, *ACOT1*, *EARS*, *RHOT2*, *TRAP2*, *APOE* and *PLIN3*). The variants that were associated with *APOE* were also strongest associated with AD according to the meta-GWAS ($P \leq 2.06 \times 10^{-287}$, $OR \geq 2.35$). The PRS without *APOE* variants was not significantly associated with AD status ($P = 0.07$, $OR = 1.01$). The PRS that included *APOE* variants, was as expected significantly associated with AD status ($P = 1.36 \times 10^{-80}$, $OR = 1.12$).

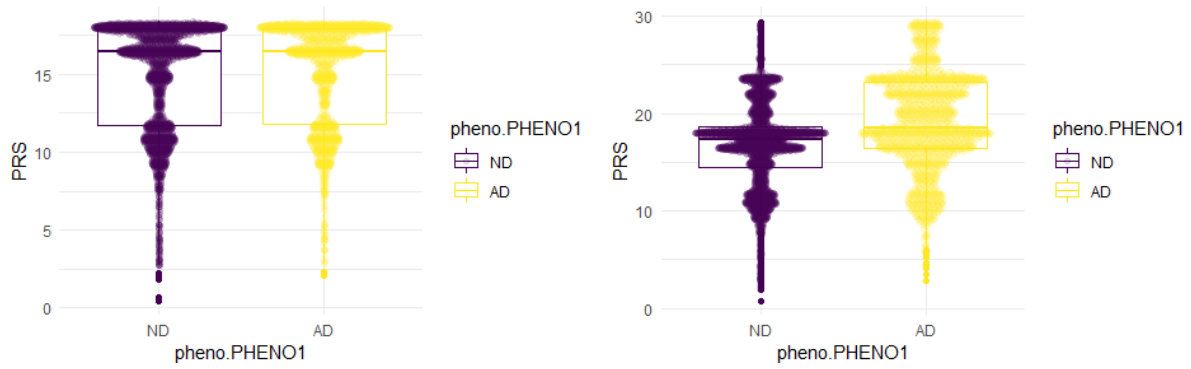

Figure A2 Boxplots with violin plot overlay of the PRSs. Every dot is an individual and the x-axis represent the diagnosis (ND controls and AD). the y-axis represents the PRS of the respective individual.

6.1.5. Discussion

Here, we tested whether a PRS constructed with pQTLs that are associated with AD have an added value on top of the *APOE* variant. Our results show the PRS without *APOE* is borderline not significant. Additionally, the *PRS* including the *APOE* variants showed less association with AD compared to *APOE* variants on their own. Altogether, constructing a PRS with pQTLs does not improve the predictive value of PRSs with this specific approach. Finally, an independent population should have been used for performance testing and validation, unfortunately, this was not feasible for this study.