

A Meta-Analysis of Alzheimer's Disease Brain Transcriptomic Data

Hamel Patel^{a,b,1,*}, Richard J.B. Dobson^{a,b,c,d,e,1} and Stephen J. Newhouse^{a,b,c,d,e,1}

^a*Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK*

^b*NIHR BioResource Centre Maudsley, NIHR Maudsley Biomedical Research Centre (BRC) at South London and Maudsley NHS Foundation Trust (SLaM) and Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, UK*

^c*Health Data Research UK London, University College London, London, UK*

^d*Institute of Health Informatics, University College London, London, UK*

^e*The National Institute for Health Research University College London Hospitals Biomedical Research Centre, University College London, London, UK*

Accepted 11 February 2019

Abstract.

Background: Microarray technologies have identified imbalances in the expression of specific genes and biological pathways in Alzheimer's disease (AD) brains. However, there is a lack of reproducibility across individual AD studies, and many related neurodegenerative and mental health disorders exhibit similar perturbations.

Objective: Meta-analyze publicly available transcriptomic data from multiple brain-related disorders to identify robust transcriptomic changes specific to AD brains.

Methods: Twenty-two AD, eight schizophrenia, five bipolar disorder, four Huntington's disease, two major depressive disorder, and one Parkinson's disease dataset totaling 2,667 samples and mapping to four different brain regions (temporal lobe, frontal lobe, parietal lobe, and cerebellum) were analyzed. Differential expression analysis was performed independently in each dataset, followed by meta-analysis using a combining *p*-value method known as Adaptively Weighted with One-sided Correction.

Results: Meta-analysis identified 323, 435, 1,023, and 828 differentially expressed genes specific to the AD temporal lobe, frontal lobe, parietal lobe, and cerebellum brain regions, respectively. Seven of these genes were consistently perturbed across all AD brain regions with SPCS1 gene expression pattern replicating in RNA-Seq data. A further nineteen genes were perturbed specifically in AD brain regions affected by both plaques and tangles, suggesting possible involvement in AD neuropathology. In addition, biological pathways involved in the "metabolism of proteins" and viral components were significantly enriched across AD brains.

Conclusion: This study identified transcriptomic changes specific to AD brains, which could make a significant contribution toward the understanding of AD disease mechanisms and may also provide new therapeutic targets.

Keywords: Alzheimer's disease, gene expression, human, mental disorders, meta-analysis, microarray analysis, neurodegenerative disorders, neuropathology

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia affecting over 44 million individuals worldwide, and numbers are expected to triple by

¹These authors are joint last authors.

*Correspondence to: Hamel Patel, Department of Biostatistics & Health Informatics, SGDP Centre, IoPPN, De Crespigny Park, Denmark Hill London, SE5 8AF, UK. Tel.: +44 207 848 0969; E-mail: hamel.patel@kcl.ac.uk.

2050 [1]. The hallmark of the disease is characterized by the abnormal brain accumulation of amyloid- β (A β) protein and hyperphosphorylated tau filaments, which forms structures known as plaques and tangles, respectively. The accumulation of these proteins contributes to the loss of connections between neuron synapses, leading to the loss of brain tissue and the disruption of normal cognitive functions.

As AD progresses, the spread of plaques and tangles in the brain usually occurs in a predictable pattern and can begin up to 18 years prior to the onset of clinical symptoms [2]. In the earliest stages of the disease, plaques and tangles form in areas of the brain primarily involved in learning and memory, specifically the hippocampus and entorhinal cortex, both situated in the temporal lobe (TL) region [3]. Next, the frontal lobe (FL), a region involved in voluntary movement, is affected, followed by the parietal lobe (PL), a region involved in processing reading and writing. In the later stage of the disease, the occipital lobe, a region involved in processing information from the eyes, can become affected, followed by the cerebellum (CB), a region which receives information from the sensory systems and the spinal cord to regulate motor movement. Nerve cell death, tissue loss, and atrophy occur throughout the brain as AD progresses, leading to the manifestation of clinical symptoms associated with loss of normal brain function. However, not all brain regions are neuropathologically affected in the same manner. The CB, which only accounts for 10% of the brain but contains over 50% of the brain's total neurons, is often neglected in AD research because it is generally considered to be partially spared from the disease as plaques are only occasionally seen but tangles are generally not reported [4, 5].

The histopathological spread of the disease is well documented, and with the advent of high throughput genomics approaches, we are now able to study the transcriptomic and biological pathways disrupted in AD brains. Microarrays can simultaneously examine thousands of genes, providing an opportunity to identify imbalances in the expression of specific genes and biological pathways. However, microarray reproducibility has always been questionable, with replication of differentially expressed genes (DEGs) very poor [6]. For example, two independent microarray transcriptomic studies performed differential expression analysis in the hippocampus of AD brains. The first study by Miller et al. identified 600 DEGs [7], and a similar study by Hokama et al. identified 1071 DEGs [8]. An overlap of 105 DEGs exist between the two studies; however, after accounting

for multiple testing, no gene was replicated between the two studies. The Miller study consisted of 7 AD and 10 control subjects expression profiled on the Affymetrix platform while the Hakoma study consisted of 31 AD and 32 control subjects expression profiled on the Illumina platform. Replication between the Illumina and Affymetrix platform has been shown to be generally very high [9]; therefore, the lack of replication between the two studies is probably down to a range of other factors including low statistical power, sampling bias, and disease heterogeneity.

Unlike DEGs, replication of the molecular changes at a pathway level are more consistent and have provided insights into the biological processes disturbed in AD. Numerous studies have consistently highlighted disruptions in immune response [10–13], protein transcription/translation [10, 11, 14–17], calcium signaling [10, 18, 19], MAPK signaling [7, 16], various metabolism pathways such as carbohydrates [16], lipids [16, 20], glucose [17, 21, 22], iron [11, 23], chemical synapse [7, 18, 19], and neurotransmitter [11, 18, 19]. However, many of these pathways have also been suggested to be disrupted in other brain-related disorders. For example, disruptions in calcium signaling, MAPK, chemical synapse, and various neurotransmitter pathways have also been implicated in Parkinson's disease (PD) [24, 25]. In addition, glucose metabolism, protein translation, and various neurotransmission pathways have also been suggested to be disrupted in bipolar disorder (BD) [26–29]. Although the biological disruptions involved in AD are steadily being identified, many other neurodegenerative and mental disorders are showing similar perturbations. We are yet to identify robust transcriptomic changes specific to AD brains.

In this study, we combined publicly available microarray gene expression data generated from AD human brain tissue and matched cognitively healthy controls to conduct the most extensive AD transcriptomic microarray meta-analyses known to date. We generate AD expression profiles across the TL, FL, PL, and CB brain regions. We further refine each expression profile by removing perturbations seen in other neurodegenerative and mental disorders (PD, BD, schizophrenia [SCZ], major depressive disorder [MDD], and Huntington's disease [HD]) to decipher specific transcriptomic changes occurring in human AD brains. These AD-specific brain changes may provide new insight and a better understanding of the disease mechanism, which in turn could provide new therapeutic targets for preventing and curing AD.

MATERIALS AND METHODS

Selection of publicly available microarray studies

Publicly available microarray gene expression data was sourced from the Accelerating Medicines Partnership-Alzheimer's Disease AMP-AD (doi:10.7303/syn2580853, doi:10.1038/ng.305, doi:10.1371/journal.pgen.1002707, doi:10.1038/ng.305, doi:10.1038/sdata.2016.89, doi:10.1038/sdata.2018.185) and ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) in June 2016. For a study to be selected for inclusion, the data had to 1) be generated from a neurodegenerative or mental health disorder, 2) be sampled from human brain tissue, 3) have gene expression measured on either the Affymetrix or Illumina microarray platform, 4) contain both diseased and suitably matched healthy controls in the same experimental batch, and 5) contain at least 10 samples from both the diseased and control group.

Microarray gene expression data pre-processing

Data analysis was performed in RStudio (version 0.99.467) using R (version 3.2.2). All data analysis scripts used in this study are available at <https://doi.org/10.5281/zenodo.823256>. In brief, raw Affymetrix microarray gene expression data was “mas5” background corrected using R package “affy” (version 1.42.3) and raw Illumina microarray gene expression data Maximum Likelihood Estimation (MLE) background corrected using R package “MBCB” (version 1.18.0). Studies with samples extracted from multiple tissues were separated into tissue-specific matrices, log₂ transformed and then Robust Spline Normalised (RSN) using R package “lumi” (version 2.16.0).

BRAAK staging is a measure of AD pathology and ranges from I-VI. In general, stages I-II, III-IV and V-VI represent the “low likelihood of AD”, “probable AD” and “definite AD” respectively [30]. To maintain homogeneity within the sample groups and to be able to infer pathological related genetic changes, if BRAAK staging was available, clinical AD samples with BRAAK scores ≤ 3 or clinical control samples with BRAAK scores ≥ 3 were removed from further analysis.

Gender was predicted using the R package “masiR” (version 1.0.1) and used to subset the data into four groups based on diagnosis (case/control) and gender (male/female). Next, probes below the 90th

percentile of the log₂ expression scale in over 80% of samples were deemed “not reliably detected” and were excluded from further analysis to eliminate noise [31] and increase power [32].

Publicly available data is often accompanied by a lack of sample processing information, making it impossible to adjust for known systematic errors introduced when samples are processed in multiple batches, a term often known as “batch effects”. To account for both known and latent variation, batch effects were estimated and removed using the Principal Component Analysis (PCA) and Surrogate Variable Analysis (SVA) using the R package “sva” (version 3.10.0). Gender and diagnosis information were used as covariates in sva when correcting for batch effects. Outlying samples were iteratively identified and removed from each gender and diagnosis group using fundamental network concepts described in [33]. Platform-specific probe ID's were converted to Entrez Gene ID's using the BeadArray corresponding R annotation files (“hgu133plus2.db”, “hgu133a.db”, “hgu133b.db”, “hugene10sttranscriptcluster.db”, “illuminaHumanv4.db”, “illuminaHumanv3.db”) and differential expression analysis was performed within each dataset using the R package “limma” (version 3.20.9).

Finally, study compatibility analysis was investigated through the R package “MetaOmics” (version 0.1.13). This package uses DEGs, co-expression, and enriched biological pathways analysis to generate six quantified measures that are used to generate a PCA plot. The direction of each quality control (QC) measure is juxtaposed on top of the two-dimensional PC subspace using arrows. Datasets in the negative region of the arrows were classed as outliers [34] and were removed from further analysis.

Meta-analysis

Datasets were grouped by the primary cerebral cortex lobes (TL, FL, PL) and the CB. Meta-analysis was performed using a “combining *p*-values” method known as “Adaptively Weighted with One-sided Correction” (AW.OC), implemented through the R package “MetaDE” (version 1.0.5) [34]. A combining *p*-value method was chosen to address the biases introduced from different platforms. AW.OC was chosen as it permits missing information across datasets which are introduced by combining data generated from different microarray platforms and

expression chips. This avoids the need to subset individual datasets to common probes, which essentially allows for the maximum number of genes to be analyzed. Furthermore, the method provides additional information on which dataset is contributing towards the meta-analysis p -value, and has been shown to be amongst the best performing meta-analysis methods for combining p -values for biological associations [35]. The meta-analysis method does not provide an overall directional change for each gene; therefore, the standard error (SE) was calculated from the DE logFC values of each gene across the AW assigned significant datasets and used for standard meta-summary estimate analysis using the R package “rmeta” (version 2.16). This served as the “meta expression” change in downstream analysis where positive values represent a gene being upregulated in AD and negative values as being downregulated in AD. Selecting DEGs based on an arbitrary expression change significantly influences the interpretation of DE results [36]. At least half of differential expression based studies incorporate a fold change cut-off typically between 2–3; however, informative RNAs and expressed transcripts have been shown to have a fold change less than 2 [37], and genes with low fold change have been demonstrated to influence biological effects in signaling cascades and pathways [36]. In addition, gene expression is heavily influenced by tissue, and as this study performs meta-analysis across multiple inter-related tissues within larger brain compartments, we do not employ an arbitrary fold change cut-off to determine if a gene is differentially expressed; however, we do require the gene to be consistently expressed across these tissues. If a gene was significantly DE according to the meta-analysis (FDR adjusted meta p -value ≤ 0.05), but at least one contributing dataset (according to AW.OC weights) had directional logFC discrepancy (i.e., upregulated in one dataset and downregulated in another dataset), the gene was deemed to be discordant and was excluded from further analysis. This ensured we only captured robust, and consistently reproducible expression signatures.

Generation of disease-specific meta-analysis expression profiles

Meta-analysis was performed across all AD datasets, followed by a separate meta-analysis across the non-AD disorder datasets. Using these meta-analysis results we generated three expression profiles: 1) “AD expression profile”, 2) “AD-specific

expression profile”, and 3) “common neurological disorder expression profile”.

The first expression profile, “AD expression profile”, is a direct result of the meta-analysis performed on AD studies, which represents the changes typically observed from an AD and cognitively healthy control study design. The second expression profile, deemed as the “AD-specific expression profile”, is produced by subtracting significantly DEGs found in the non-AD meta-analysis results from the “AD expression profile”. This profile represents transcriptomic changes specifically observed in AD and not in any other neurodegenerative or mental health disorder used in this study. The third expression profile, deemed as the “common neurological disorder expression profile”, represents genes which are significantly DE in all disorders used in this study, including AD.

Replication of significant microarray genes in RNA-Seq data

The genes significantly DE and deemed to be of biological significance in this study were queried in the curated web-based database Agora (data version 9, accessible at <https://agora.ampadportal.org>), which provides expression change of genes in AD based on RNA-Seq of 2100 human brain samples.

Functional and gene set enrichment analysis

Gene set enrichment analysis (GSEA) and Gene Ontology (GO) analysis was conducted using an Over-Representation Analysis (ORA) implemented through the ConsensusPathDB web platform (version 32) [38] in May 2017. ConsensusPathDB incorporates numerous well-known biological pathway databases including BioCarta, KEGG, Reactome, and Wikipathways. The platform performs a hypergeometric test while integrating a background gene list, which in this case is a list of all the genes that pass quality control in this study, compiles results from each database and corrects for multiple testing using the false discovery rate (FDR) [38]. A minimum overlap of the query signature and database was set to 2, and a result was deemed significant if the q -value was ≤ 0.05 .

Network analysis

Protein-protein interaction (PPI) networks were created by uploading the meta-analysis DEG

lists (referred to as seeds in network analysis) along with their meta logFC expression values to NetworkAnalyst's web-based platform <http://www.networkanalyst.ca/faces/home.xhtml> in June 2017. The "Zero-order Network" option was incorporated to allow only seed proteins directly interacting with each other, preventing the well-known "Hairball effect" and allowing for better visualization and interpretation [39]. Sub-modules with a p -value ≤ 0.05 (based on the "InfoMap" algorithm [40]) were considered significant key hubs, and the gene with the most connections within this hub was regarded as the key hub gene.

RESULTS

The AD microarray datasets

We Identified and acquired nine publicly available AD studies from ArrayExpress and AMP-AD, of which seven studies contained samples extracted from differing regions of the brain. The basic characteristics of each study and dataset are provided in Table 1. Separating the nine studies by brain regions resulted in 46 datasets. Here a "dataset" is defined by brain region and study origin. For example, ArrayExpress study E-GEOD-36980 consists of diseased and healthy samples extracted from three different tissues (temporal cortex, hippocampus, and frontal cortex). All samples originating from the same tissue were classified as one dataset; therefore, study E-GEOD-36980 generated three datasets, representing the three different tissues.

The 46 AD datasets contained both AD samples and healthy controls, were assayed using seven different expression chips over two different microarray platforms (Affymetrix and Illumina) and consisted of a total 2,718 samples before QC. Briefly, the MetaOmics analysis identified study syn4552659 as an outlier and was therefore removed from further analysis (see the Supplementary Material), resulting in 1,501 samples (746 AD, 755 controls) in the remaining 22 datasets after QC.

Summary of the AD meta-analysis DEG counts

The AD meta-analysis was performed on the 22 AD datasets and independently identified differentially expressed genes within the TL, FL, PL, and CB brain regions. A summary of the number of datasets in each brain region and the number of significant DEGs identified is provided in Table 2. The complete DE results are provided in Supplementary

Table 1. As mentioned in the methods, due to gene expression being influenced by tissue source and as this study incorporates different brain regions, we do not employ an arbitrary cut-off value to determine genes that are highly or lowly expressed, but primarily focus on genes consistently perturbed. However, we provide the meta expression values in the Supplementary tables and advise readers to consider the expression distribution of all DEGs within each brain region independently, if determining whether a gene is highly/lowly expressed. For instance, the CB meta expression ranges from -0.53 to 0.54 with an interquartile range (Q1–Q3) of -0.1 to 0.1 . In contrast, the PL has a larger meta expression range of -1.5 to 1.35 , with an interquartile range (Q1–Q3) of -0.53 to 0.29 . Therefore, as gene expression distribution varies across brain regions, a sensible cut-off (if one was to be used) for highly and lowly expressed genes may lie at the 1st and 3rd quartiles, respectively, with quartiles calculated per tissue.

The non-AD disorder microarray datasets

Nine non-AD studies were identified and acquired, of which four studies consisted of samples generated from multiple disorders and brain regions. Separating the studies by disease and tissue equated to 21 datasets consisting of 8 SCZ, 6 BD, 4 HD, 2 MDD, and 1 PD dataset with a total of 1,166 samples after QC. The demographics of the non-AD datasets is provided in Table 3.

Summary of non-AD brain disorder meta-analyses DEG counts

A second meta-analysis was performed on all non-AD disorders, and similarly to the AD meta-analysis, datasets were grouped into the TL, FL, PL, and CB brain regions. An overview of the non-AD meta-analysis results are provided in Table 4, and a complete list of DEGs is provided in Supplementary Table 2. SCZ and BD were the only disorders with expression data available across all four brain regions, and the FL brain region was the only region with expression data available from all non-AD disorders identified in this study.

The meta-analysis expression profiles

As described in the methods, three primary expression signatures were derived from the meta-analyses for each of the four brain regions: 1) "AD expression profile", 2) "AD-specific expression profile", and 3)

Table 1
 Characteristics of individual AD studies processed in this meta-analysis

Data repository	Accession details (Publication)	Microarray platform	BeadArray	Tissue source (as stated in the original study publication)	Meta-Analysis brain region mapping	Number of samples after QC	
						AD (M/F)	Control (M/F)
ArrayExpress	E-GEOD-118553	Illumina	HumanHT-12 v4	Entorhinal Cortex	Temporal Lobe	35 (14/21)	21 (12/9)
				Cerebellum	Cerebellum	38 (10/28)	19 (5/14)
				Frontal Cortex	Frontal Lobe	38 (13/25)	22 (11/11)
ArrayExpress	E-GEOD-48350 (76)	Affymetrix	Human Genome U133 Plus 2.0	Temporal Cortex	Temporal Lobe	51 (21/30)	29 (21/8)
				Entorhinal Cortex	Temporal Lobe	11 (6/5)	38 (21/17)
				Hippocampus	Temporal Lobe	15 (8/7)	41 (22/19)
ArrayExpress	E-GEOD-29378 (7)	Illumina	HumanHT-12 v3	Postcentral Gyrus	Parietal Lobe	19 (11/8)	33 (20/13)
				Superior Frontal Gyrus	Frontal Lobe	17 (8/9)	38 (22/16)
				Hippocampus CA1	Temporal Lobe	16 (9/7)	16 (11/5)
ArrayExpress	E-GEOD-36980 (8)	Affymetrix	Human Gene 1.0 ST	Hippocampus CA3	Temporal Lobe	15 (9/6)	16 (11/5)
				Frontal Cortex	Frontal Lobe	14 (7/7)	17 (9/8)
				Hippocampus	Temporal Lobe	7 (3/4)	10 (5/5)
ArrayExpress	E-GEOD-28146 (19)	Affymetrix	Human Genome U133 Plus 2.0	Temporal Cortex	Temporal Lobe	10 (5/5)	19 (8/11)
				Hippocampus CA1	Temporal Lobe	15 (4/11)	8 (5/3)
				Hippocampus	Temporal Lobe	19 (4/11)	9 (6/3)
ArrayExpress	E-GEOD-1297 (77)	Affymetrix	Human Genome U133A	Hippocampus	Temporal Lobe	19 (4/11)	9 (6/3)
				Entorhinal Cortex	Temporal Lobe	10 (4/6)	13 (11/2)
				Hippocampus CA1	Temporal Lobe	10 (4/6)	13 (10/3)
ArrayExpress	E-GEOD-5281 (21)	Affymetrix	Human Genome U133 Plus 2.0	Medial Temporal Gyrus	Temporal Lobe	16 (10/6)	12 (8/4)
				Posterior Cingulate	Parietal Lobe	9 (4/5)	13 (10/3)
				Superior Frontal Gyrus	Frontal Lobe	23 (13/10)	11 (7/4)
AMP	syn3157225 (78)	Illumina	Whole-Genome DASL HT	Temporal Cortex	Temporal Lobe	189 (93/96)	186 (116/70)
				Hippocampus CA1	Temporal Lobe	10 (4/6)	13 (10/3)
				Cerebellum	Cerebellum	169 (87/82)	171 (113/58)

AMP	syn4552659 (179)	Affymetrix	Human Genome U133A	Frontal Pole	Frontal Lobe	25 (6/19)	7 (4/3)
				Precentral Gyrus	Frontal Lobe	20 (5/15)	3 (1/2)
				Inferior Frontal Gyrus	Frontal Lobe	19 (5/14)	4 (1/3)
				Dorsolateral Prefrontal Cortex	Frontal Lobe	19 (4/15)	8 (4/4)
				Superior Parietal Lobule	Parietal Lobe	11 (2/9)	5 (2/3)
				Prefrontal Cortex	Frontal Lobe	23 (7/16)	4 (2/2)
				Parahippocampal Gyrus	Temporal Lobe	18 (5/13)	7 (3/4)
				Hippocampus	Temporal Lobe	20 (5/15)	5 (2/3)
				Inferior Temporal Gyrus	Temporal Lobe	20 (5/15)	6 (3/3)
				Middle Temporal Gyrus	Temporal Lobe	15 (4/11)	7 (4/3)
				Superior Temporal Gyrus	Temporal Lobe	15 (3/12)	8 (4/4)
				Temporal Pole	Temporal Lobe	25 (7/18)	6 (3/3)
				Frontal Pole	Frontal Lobe	26 (8/18)	7 (4/3)
				AMP	syn4552659 (179)	Affymetrix	Human Genome U133B
Inferior Frontal Gyrus	Frontal Lobe	21 (5/16)	5 (2/3)				
Dorsolateral Prefrontal Cortex	Frontal Lobe	20 (5/15)	8 (4/4)				
Superior Parietal Lobule	Parietal Lobe	16 (5/11)	5 (3/2)				
Prefrontal Cortex	Frontal Lobe	23 (7/16)	4 (2/2)				
Parahippocampal Gyrus	Temporal Lobe	19 (7/12)	7 (3/4)				
Hippocampus	Temporal Lobe	22 (6/16)	5 (2/3)				
Inferior Temporal Gyrus	Temporal Lobe	21 (6/15)	7 (4/3)				
Middle Temporal Gyrus	Temporal Lobe	23 (8/15)	7 (4/3)				
Superior Temporal Gyrus	Temporal Lobe	23 (4/19)	8 (4/4)				
Frontal Pole	Frontal Lobe	26 (8/18)	7 (4/3)				

Nine publicly available AD studies were identified and acquired for this study. Separating the studies by tissue resulted in 46 datasets, each containing AD and healthy control samples. The brain tissue in each of the 46 datasets was mapped to their corresponding cerebral cortex (temporal lobe, frontal lobe, or parietal lobe) or the cerebellum. Due to limited phenotypic information in publicly available data, the reported gender was predicted from gene expression if clinical gender was unavailable. M, male; F, female.

Table 2
Summary of AD study meta-analysis DEGs

Brain region	Number of datasets	Number of samples (case/control)	AW.OC Significant DEGs (FDR adjusted $p \leq 0.05$)
Temporal lobe	14	850 (419/431)	323
Frontal lobe	4	180 (92/88)	460
Parietal lobe	2	74 (28/46)	1736
Cerebellum	2	397 (207/190)	867

Twenty-two AD datasets containing a total of 1,501 samples remained in this study after QC. The case/control numbers represent the total number of AD/healthy controls subjects across all datasets within a particular brain region. The number of significant genes was identified through a combining p -value method known as Adaptively Weighted with One-sided Correction (AW.OC).

“common neurological disorder expression profile”. The numbers of significant DEGs in each of the three expression signatures are provided in Table 5.

The DEGs from the “AD expression profile” in the TL brain region were not significantly DE in any other disorder included in this study. Hence, the “AD expression profile” and the “AD-specific expression profile” contained the same 323 genes for the TL brain region. The “AD-specific expression profile” for all four brain regions is provided in Supplementary Table 3.

The “common neurological disorder expression profile” within the four brain regions consisted of very little or no DEGs (except for the PL); hence, the downstream analysis did not yield any statistically significant results of biological relevance. We find little robust evidence of shared biology based on this data analysis and therefore, exclude all results generated from the “common neurological disorder expression profile” from this paper; however, we provide the complete list of significantly DEGs within this profile in Supplementary Table 4.

Common differentially expressed genes across multiple brain regions in AD

AD is known to affect all brain regions through the course of the disease, although not to the same degree, similar transcriptomic changes across all brain regions were deemed disease-specific, while perturbations in a single brain region were considered to be tissue-specific. We were particularly interested in disease-specific transcriptomic changes and therefore decided to focus on genes that were found to be consistently DE across multiple brain regions.

Meta-analysis of the AD datasets identified a total of 2,495 unique genes as significantly DE. The distribution of these genes across the four brain regions is shown in Fig. 1. Forty-two genes were found to

be perturbed across all four brain regions and can be grouped into three sets (Fig. 2). The first group (Gene set 1) are expressed consistently in the same direction across all four brain regions and can be regarded as disease-specific. The second group (Gene set 2) are expressed in the same direction in the TL, FL, and PL, but expression is reversed in the CB brain region, a region suggested to be spared from AD pathology [4, 5]. This expression pattern suggests these genes may be involved in AD pathology. Finally, the third group (Gene set 3) are inconsistently expressed across the four brain regions are most likely tissue-specific or even false-positives.

From the forty-two genes significantly differentially expressed across all brain regions, seven genes were DE in the same direction and belong to the “AD-specific expression profile”, that is, these seven genes (downregulated **NDUFS5**, **SOD1**, **SPCS1** and upregulated **OGT**, **PURA**, **RERE**, **ZFP36L1**) were consistently perturbed in all AD brain regions and not in any other brain region of any other disorder used in this study and can be considered unique to AD brains. The expression of these seven genes across AD brains is shown in Fig. 3.

Differentially expressed genes in brain regions affected by AD histopathology

In AD, the TL, FL, and PL are known to be affected by both plaques and tangles, while the CB brain region is rarely reported to be affected. In addition to identifying genes DE across all brain regions and reversed in the CB brain region, we were also interested in genes perturbed in the TL, FL and PL and not the CB. These genes may also play a role in general AD histopathology and could be new therapeutic targets in preventing or curing AD.

Fifty-five genes were found to be significantly DE in TL, FL and PL but not the CB, of which sixteen

Table 3
 Characteristics of individual non-AD studies included in this meta-analysis

Data repository	ArrayExpress Accession details (Publication)	Microarray Platform	BeadArray	Disorder	Sample source (as stated in the original study publication)	Mapping to brain region	Number of samples after QC	
							AD (M/F)	Control (M/F)
ArrayExpress	E-GEOD-12649 ([43])	Affymetrix	Human Genome U133A	Bipolar Disorder	Prefrontal Cortex	Frontal Lobe	33 (16/17)	34 (25/9)
ArrayExpress	E-GEOD-17612 ([44])	Affymetrix	Human Genome U133 Plus 2.0	Schizophrenia	Prefrontal Cortex	Frontal Lobe	33 (25/8)	32 (24/8)
ArrayExpress	E-GEOD-20168 ([45])	Affymetrix	Human Genome U133A	Schizophrenia	Prefrontal Cortex	Frontal Lobe	27 (18/9)	22 (11/11)
ArrayExpress	E-GEOD-21138 ([46])	Affymetrix	Human Genome U133 Plus 2.0	Parkinson's Disease	Prefrontal Cortex	Frontal Lobe	14 (7/7)	16 (11/5)
ArrayExpress	E-GEOD-21935 ([47])	Affymetrix	Human Genome U133 Plus 2.0	Schizophrenia	Prefrontal Cortex	Frontal Lobe	25 (21/4)	28 (23/5)
ArrayExpress	E-GEOD-35978 ([48])	Affymetrix	Human Gene 1.0 ST	Schizophrenia	Temporal Cortex	Temporal Lobe	22 (12/10)	19 (10/9)
ArrayExpress	E-GEOD-3790 ([49])	Affymetrix	Human Genome U133A	Bipolar Disorder	Cerebellum	Cerebellum	32 (16/16)	46 (29/17)
				Schizophrenia	Cerebellum	Cerebellum	43 (31/12)	46 (29/17)
				Bipolar Disorder	Parietal Lobe	Parietal Lobe	40 (24/16)	45 (32/13)
				Schizophrenia	Parietal Lobe	Parietal Lobe	51 (37/14)	36 (26/10)
				Huntington's Disease	Frontal Lobe	Frontal Lobe	36 (22/14)	27 (19/8)
ArrayExpress	E-GEOD-5388 ([50])	Affymetrix	Human Genome U133A	Huntington's Disease	Cerebellum	Cerebellum	38 (22/16)	27 (16/11)
				Huntington's Disease	Cerebellum	Cerebellum	38 (23/15)	27 (16/11)
				Huntington's Disease	Frontal Lobe	Frontal Lobe	37 (21/16)	29 (19/10)
ArrayExpress	E-GEOD-53987 ([51])	Affymetrix	Human Genome U133 Plus 2.0	Bipolar Disorder	Prefrontal Cortex	Frontal Lobe	30 (16/14)	29 (23/6)
ArrayExpress	E-GEOD-53987 ([51])	Affymetrix	Human Genome U133 Plus 2.0	Bipolar Disorder	Prefrontal Cortex	Frontal Lobe	17 (10/7)	19 (11/8)
				Major Depressive Disorder	Prefrontal Cortex	Frontal Lobe	16 (9/7)	18 (10/8)
				Schizophrenia	Prefrontal Cortex	Frontal Lobe	14 (7/7)	19 (11/8)
				Bipolar Disorder	Hippocampus	Temporal Lobe	18 (11/7)	17 (9/8)
				Major Depressive Disorder	Hippocampus	Temporal Lobe	16 (9/7)	17 (9/8)
Schizophrenia	Hippocampus	Temporal Lobe	15 (9/6)	18 (10/8)				

Nine publicly available non-AD studies were identified and acquired. Separating the studies by tissue resulted in 21 datasets. Each dataset contained both diseased and complimentary healthy controls. The brain tissue in each of the 21 datasets was mapped to their corresponding cerebral cortex (temporal lobe, frontal lobe, or parietal lobe) or the cerebellum. Due to limited phenotypic information in publicly available data, the reported gender was predicted from gene expression if clinical gender was unavailable. M, male; F, female.

Table 4
Summary of non-AD study meta-analysis DEGs

Brain region	Number of BD datasets (case/control)	Number of Schizophrenia datasets (case/control)	Number of HD datasets (case/control)	Number of MDD datasets (case/control)	Number of PD datasets (case/control)	Total number of datasets (case/control)	AW.OC Significant DEGs (FDR adjusted $p \leq 0.05$)
Temporal lobe	1 (18/17)	2 (37/37)	0	1 (16/17)	0	4 (71/71)	51
Frontal lobe	3 (80/82)	4 (99/101)	2 (73/56)	1 (16/18)	1 (14/16)	11 (282/273)	149
Parietal lobe	1 (40/45)	1 (51/36)	0	0	0	2 (91/81)	2611
Cerebellum	1 (32/46)	1 (43/46)	2 (76/54)	0	0	4 (151/146)	177

The table illustrates the non-AD dataset and sample distribution across the four brain regions. The case/control numbers represent the total number of diseased and healthy control subjects within a disease group and brain region. For instance, “3 (80/82)” for BD datasets in the Frontal lobe region indicates three BD datasets with a combined total of 80 BD and 82 complimentary healthy control subjects. The number of significant DEGs was identified through a combining p -value method known as Adaptively Weighted with One-sided Correction (AW.OC). BD, bipolar disease; HD, Huntington’s disease; MDD, major depressive disorder; PD, Parkinson’s disease.

Table 5
Summary of DEGs in each expression signature and brain region

Expression Profile	Cerebellum	Frontal lobe	Parietal lobe	Temporal lobe	Total (unique)
AD	867	460	1736	323	2494
Non-AD	177	149	2611	51	2809
AD-specific	828	435	1023	323	1994
Common	39	25	713	0	755
Total (unique)	1005	584	3642	374	–

The “AD” expression profile represents genes identified as DE in the AD vs control meta-analysis. The “non-AD” expression profile represents genes identified as DE in the non-AD meta-analysis. The “AD-specific” expression profile is a list of genes DE in AD and no other disorder, and the “common” expression profile is a list of genes DE in all mental disorder used in this study. Each expression profile is brain region specific. The “Total (unique)” represents a unique list of the total number of genes identified as significantly DE across brain regions or expression profiles.

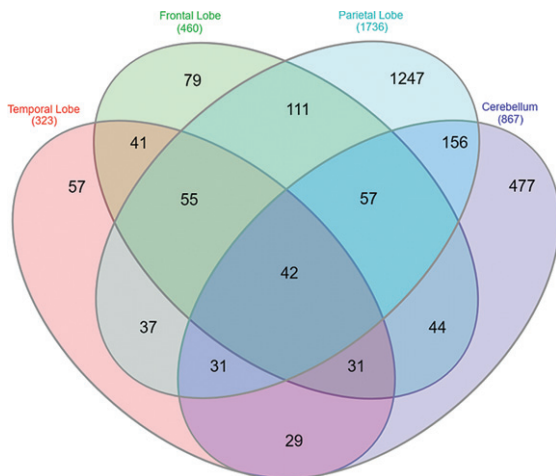


Fig. 1. Overlap of DEGs in the AD expression profile across brain regions. Forty-two genes were observed to be significantly differentially expressed across all four AD brain regions.

were expressed in the same direction and were not DE in the other brain disorders used in this study. Ten of these genes (**ALDOA**, **GABBR1**, **TUBA1A**, **GAPDH**, **DNM3**, **KLC1**, **COX6C**, **ACTG1**, **CLTA**,

SLC25A5) were consistently downregulated, and six genes (**PRNP**, **FDFT1**, **RHOQ**, **B2M**, **SPP1**, **WAC**) were consistently upregulated in AD.

Furthermore, from the forty-two genes identified as significantly DE across all four AD brain regions, ten genes were in consensus in their expression across the TL, FL, and PL brain region but expression is reversed in the CB. Only 3 of these genes (**UBA1**, **EIF4H**, and **CLDND1**) belong to the “AD-specific expression profile”, and all three genes were significantly downregulated in the TL, FL, and PL, but significantly upregulated in the CB brain region (see Gene set 2 in Fig. 2).

Microarray gene expression profiling in RNA-Seq data

The 7 genes (**NDUFS5**, **SOD1**, **SPCS1**, **OGT**, **PURA**, **RERE**, **ZFP36L1**) consistently expressed across all brain regions and the 19 genes (**ALDOA**, **GABBR1**, **TUBA1A**, **GAPDH**, **DNM3**, **KLC1**, **COX6C**, **ACTG1**, **CLTA**, **SLC25A5**, **PRNP**, **FDFT1**, **RHOQ**, **B2M**, **SPP1**, **WAC**, **UBA1**, **EIF4H**, **CLDND1**) consistently expressed in the TL, FL, and PL and not in the CB or reversed in the

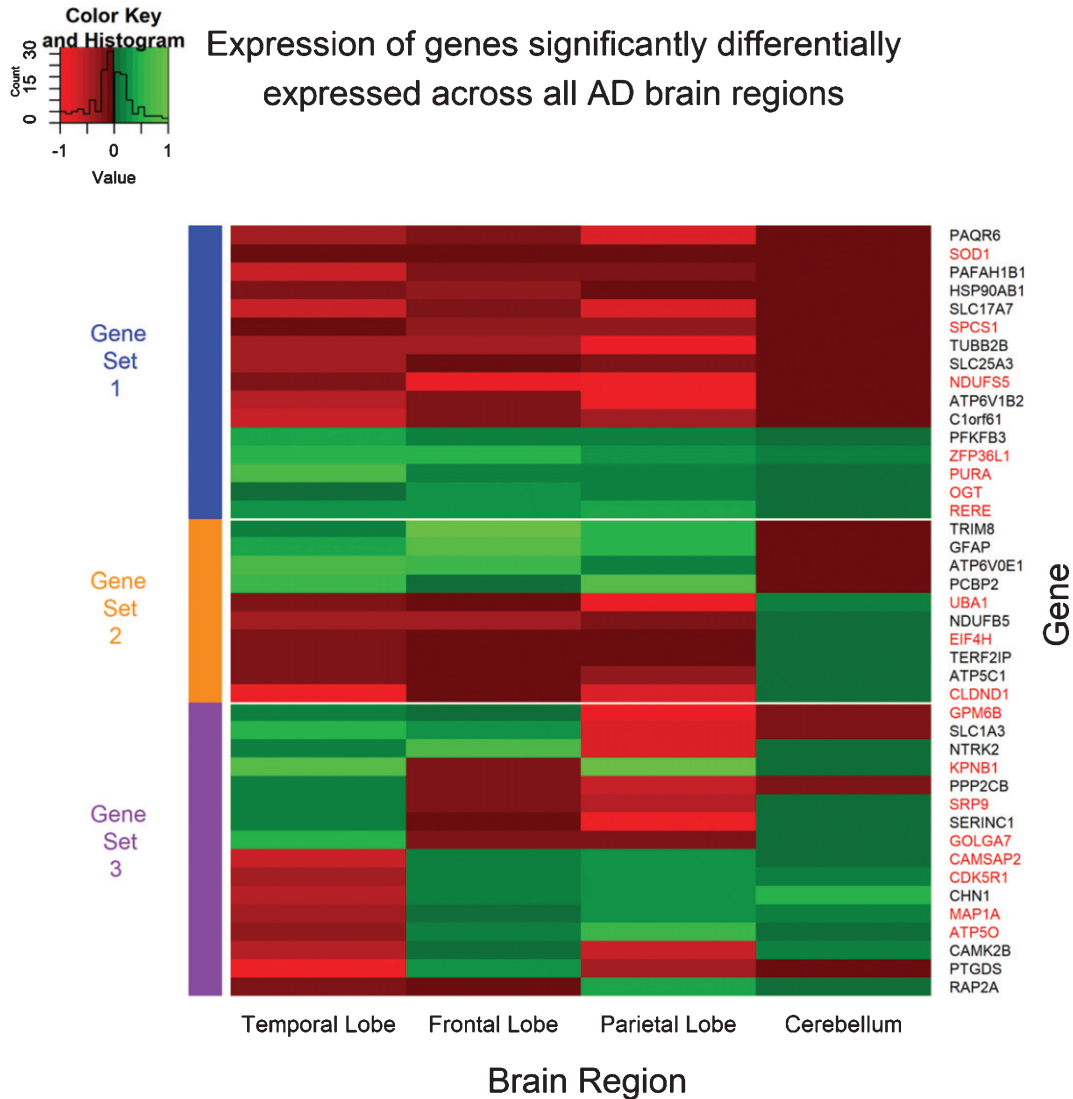


Fig. 2. Expression pattern of genes significantly differentially expressed across all four AD brain regions. The expression values for each gene was obtained from the meta-summary calculations. Red cells represent downregulated genes, and green cells represent upregulated genes. Forty-two genes were observed to be significantly perturbed across all four AD brain regions and can be grouped into three “sets”. Gene set 1 represents genes which are perturbed consistently in the same direction across all AD brain regions and can be considered disease-specific. Gene set 2 represents genes consistent in expression in the temporal lobe, frontal lobe, and parietal lobe brain regions, but reversed in the cerebellum brain region; a region often referred to be free from AD pathology. Finally, Gene set 3 represents genes which are significant DE across all four brain regions, however, directional change is not consistent across the brain regions and may represent tissue-specific genes or even false positive. The gene names highlighted in red are genes perturbed in AD and not in any other disorder used in this study and are deemed “AD-specific”.

CB, were queried in the web-based platform Agora to compare RNA-Seq based expression profiling. The results are provided in Table 6. Agora failed to provide expression profiling for 17/26 genes; however, from the data available, the genes observed to be consistently expressed across all brain regions based on microarray data are relatively mirrored in

RNA-Seq data, specifically genes **SPCS1**, **PURA** and **ZFP36L1**.

RNA-Seq data was available for only 6/19 genes (**DNM3**, **COX6C**, **ACTG1**, **CLTA**, **RHOQ**, and **B2M**) expressed in brain regions affected by hallmark AD pathology (TL, FL, and PL), and were all relatively consistent in directional change across AD

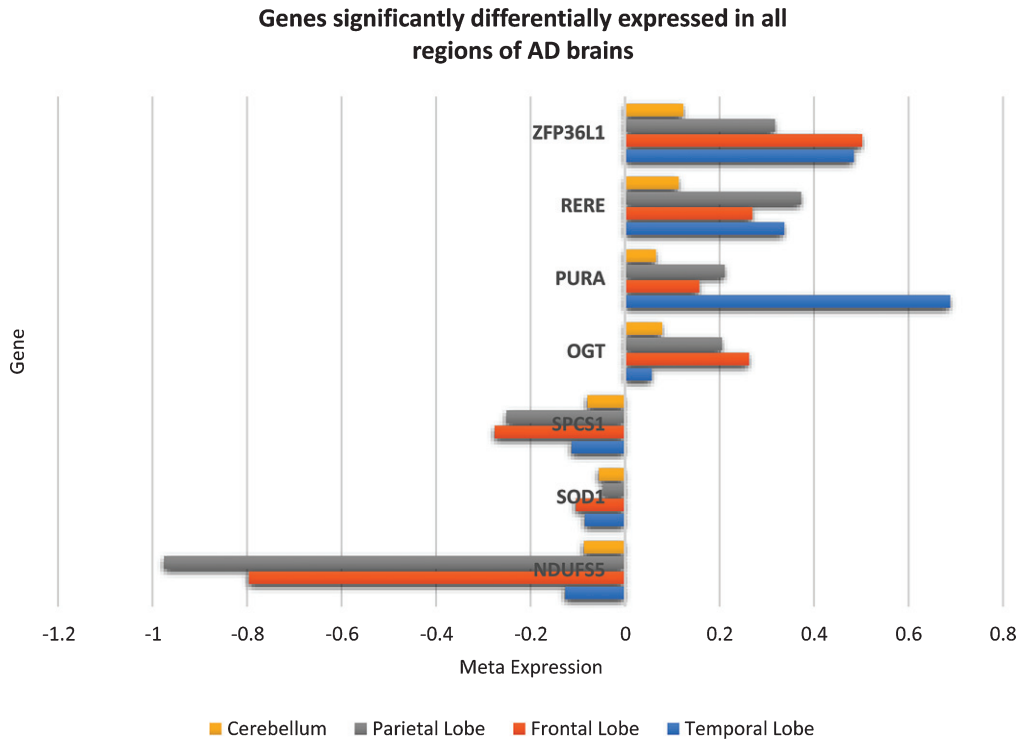


Fig. 3. Seven genes consistently significantly differentially expressed in the same direction in all regions of AD brains but not in schizophrenia, bipolar disorder, Huntington's disease, major depressive disorder, or Parkinson's disease brains. These seven genes can be assumed to be unique to AD brains and may play an important role in disease mechanisms.

brain regions, including the CB, a characteristic undesired by genes which may be associated with hallmark AD pathology.

“AD Expression Profile” functional gene set enrichment and GO analysis

Gene set enrichment analysis of the “AD expression profile” identified 205, 197, 98, and 45 biological pathways significantly enriched in the TL, FL, PL, and CB brain regions, respectively (Supplementary Table 5). There were ten pathways significantly enriched in all four brain regions, of which eight are involved in the “**metabolism of protein**” (specifically the translation process, the most significant being in CB brain region with a q -value = $1.11e-7$), one involved in “**adenosine ribonucleotides de novo biosynthesis**” (TL q -value = 0.007, FL q -value = $7.56e-5$, PL q -value = 0.04, CB q -value = 0.03) and one involved in the “**digestive system**” (TL q -value = 0.02, FL q -value = 0.02, PL q -value = 0.01, CB p -value = 0.02).

When excluding the CB brain region, 42 pathways were significantly enriched in the remaining three brain regions, of which five pathways

obtained an FDR adjusted significance p -value of ≤ 0.01 . The five pathways are “**Alzheimer's disease**” (TL q -value = $6.53e-4$, FL q -value = 0.02, PL q -value = 0.01), “**Electron Transport Chain**” (TL q -value = 0.006, FL q -value = $2.95e-5$, PL q -value = $3.69e-5$), “**Oxidative phosphorylation**” (TL q -value = $1.77e-4$, FL q -value = $4.99e-8$, PL q -value = $4.18e-05$), “**Parkinson's disease**” (TL q -value = $8.57e-4$, FL q -value = $1.59e-6$, PL q -value = $1.77e-6$), and “**Synaptic vesicle cycle**” (TL q -value = $5.19e-4$, FL q -value = $3.82e-7$, PL q -value = $2.03e-4$).

The biological GO analysis identified 384, 417, 216, and 72 biological components as significantly enriched in the TL, FL, PL, and CB brain region respectively (Supplementary Table 6). There were 36 pathways significantly enriched across all four brain regions at a p -value threshold of ≤ 0.05 and nine at an FDR adjusted significant p -value threshold of ≤ 0.01 . These nine processes are “**cellular component biogenesis**” (TL q -value = $1.38e-4$, FL q -value = 0.002, PL q -value = $5.86e-4$, CB q -value = 0.006), “**cellular component organization**” (TL q -value = $1.96e-8$, FL q -value = $1.04e-8$, PL

Table 6
Microarray gene expression compared to RNA-Seq gene expression

Gene	Microarray				RNA-Seq						
	TL	FL	PL	CB	TL			FL			CB
					TCX	STG	PHG	IFG	FP	DLPFC	
NDUFS5	-0.127	-0.796	-0.975	-0.087							
SOD1	-0.085	-0.104	-0.047	-0.055							
SPCS1	-0.113	-0.276	-0.251	-0.079	-0.192		-0.177			-0.124	-0.317
OGT	0.057	0.262	0.205	0.079							
PURA	0.688	0.157	0.211	0.065						0.103	0.318
RERE	0.337	0.270	0.373	0.114							
ZFP36L1	0.484	0.502	0.317	0.123	0.603	0.389	0.619				0.355
ALDOA	-1.389	-0.090	-0.091								
GABBR1	-0.864	-0.116	-0.162								
TUBA1A	-0.728	-0.286	-0.191								
GAPDH	-0.719	-0.095	-0.759								
DNM3	-0.611	-0.154	-0.781		-0.361		-0.119				0.141
KLC1	-0.530	-0.115	-0.056								
COX6C	-0.504	-0.360	-0.651			-0.140	-0.200				-0.220
ACTG1	-0.467	-0.165	-0.326							-0.250	
CLTA	-0.153	-0.097	-0.183		-0.091						-0.112
SLC25A5	-0.147	-0.122	-0.262								
PRNP	0.115	0.168	0.130								
FDFT1	0.158	0.196	0.283								
RHOQ	0.227	0.128	0.330		0.465		0.310			0.190	0.029
B2M	0.556	0.216	0.212		0.430						0.258
SPP1	0.935	0.557	0.316								
WAC	1.111	0.257	0.154								
UBA1	-0.213	-0.090	-0.827	0.134							
EIF4H	-0.201	-0.062	-0.093	0.063							
CLDND1	-1.100	-0.113	-0.734	0.086							

The 7 genes consistently expressed across all brain regions and the 19 genes consistently expressed in the TL, FL and PL and not/reversed in the CB were queried in the web-based platform Agora to compare RNA-Seq expression. Only significantly DEGs are shown. Red cells represent downregulated genes in AD, green cells represent upregulated genes in AD, white cells represent genes not significantly DE, and grey cells are when data is not available. TL, temporal lobe; FL, frontal lobe; PL, parietal lobe; CB, cerebellum; TCX, temporal cortex; STG, superior temporal gyrus; PHG, parahippocampal gyrus; IFG, inferior frontal gyrus; FP, frontal pole; DLPFC, dorsolateral prefrontal cortex.

q-value = 3.35e-5, CB q-value = 0.004), “**inter-species interaction between organisms**” (TL q-value = 1.85e-4, FL q-value = 8.73e-5, PL q-value = 5.59e-5, CB q-value = 0.002), “**multi-organism cellular process**” (TL q-value = 1.12e-4, FL q-value = 4.72e-5, PL q-value = 8.04e-5, CB q-value = 0.002), “**nervous system development**”

(TL q-value = 1.64e-7, FL q-value = 5.90e-14, PL q-value = 3.82e-8, CB q-value = 0.01), “**organonitrogen compound metabolic process**” (TL q-value = 0.002, FL q-value = 1.56e-5, PL q-value = 1.02e-5, CB q-value = 0.002), “**sybiosis, encompassing, mutualism through parasitism**” (TL q-value = 4.04e-4, FL q-value = 1.92e-4, PL q-

value = 3.18×10^{-4} , CB q-value = 0.004), “**translational initiation**” (TL q-value = 0.007, FL q-value = 0.006, PL q-value = 2.41×10^{-4} , CB q-value = 5.24×10^{-6}), and “**viral process**” (TL q-value = 2.82×10^{-4} , FL q-value = 1.17×10^{-4} , PL q-value = 3.18×10^{-4} , CB q-value = 0.002). Excluding the CB brain region resulted in 84 common biological components being significantly enriched across the remaining three brain regions.

“AD-specific expression profile” functional gene set enrichment and GO analysis

Analysis of the “AD-specific expression profile” identified 205, 196, 40, and 42 pathways as significantly enriched in the TL, FL, PL, and CB brain region respectively in the GSEA analysis (Supplementary Table 7). The analysis identified six significantly enriched pathways across all four brain regions, and all are involved in “**metabolism of protein**” (specifically the translation process, with the most significant pathway being in the PL brain region with a q-value = 8.92×10^{-7}). The same six pathways were identified when the CB region was excluded.

The GO analysis identified 384, 344, 36, and 72 significantly enriched biological components for the TL, FL, PL, and CB brain region, respectively. Only four common biological components were significantly enriched across all four brain regions, and all are indicative of interspecies interactions including viral. Excluding the CB identifies only “**neuronal nucleus development**” (TL q-value = 5.35×10^{-5} , FL q-value = 0.007, PL q-value = 0.003) as an additional component being enriched. The complete biological GO analysis results are provided in Supplementary Table 8.

Network analysis hub gene identification

PPI networks were generated for each expression profile and in each of the four brain regions (TL, FL, PL, and CB) to identify genes whose protein product interacts with other protein products from the same expression profile. Genes with more interactions than expected are referred to as hub genes and may be of biological significance.

Temporal lobe hub genes

PPI network analysis was performed on the expression profiles of the TL brain region to identify key hub genes. The “AD expression profile” and the “AD-specific expression profile” both consisted of the same 323 DEGs which represented

282 seed proteins with 716 edges (interactions between proteins). Two significant key hub genes were identified; the downregulated Polyubiquitin-C (**UBC**, p -value = 1.57×10^{-30}) and the upregulated Small Ubiquitin-related Modifier 2 (**SUMO2**, p -value = 3.7×10^{-4}).

Frontal lobe hub genes

The FL “AD expression profile” consisted of 460 DEGs which represented 272 seed proteins and 620 edges. Two significant key hub genes were identified; upregulated Amyloid Precursor Protein (**APP**, p -value = 1.98×10^{-8}) and downregulated Heat Shock Protein 90-alpha (**HSP90AA1**, p -value = 0.003). Using the “AD-specific expression profile” identified the same two key hub genes, with **APP** reaching a significant p -value of 2.11×10^{-9} .

Parietal lobe hub genes

The PL “AD expression profile” consisted of 1,736 DEGs which represented 1,437 seed proteins and 5,720 edges. Similar to the TL and FL, two significant key hub genes were identified; downregulated Cullin-3 (**CUL3**, p -value = 1.84×10^{-10}) and downregulated **UBC** (p -value = 1.84×10^{-10}). Using the “AD-specific expression profile” (1,023 DEGs, 810 seed proteins, and 2,351 edges) identified **UBC** as the only key hub gene, with a more significant p -value of 1.84×10^{-10} . The **CUL3** gene is no longer a significant key hub gene in the network.

Cerebellum hub genes

The CB “AD expression profile” consisted of 867 DEGs which represented 548 seed proteins and 1419 edges. Four significant key hub genes were identified: upregulated **APP** (p -value = 4.24×10^{-26}), downregulated Ribosomal Protein 2 (**RPS2**, p -value = 4.24×10^{-26}), downregulated **SUMO2** (p -value = 4×10^{-5}), and upregulated Glycyl-TRNA Synthetase (**GARS**, p -value = 0.0207). Using the “AD-specific expression profile” for the same brain region identified **APP** (p -value = 3.44×10^{-26}), **RPS2** (p -value = 6.61×10^{-6}), and **SUMO2** (p -value = 3.78×10^{-6}) as the key hub genes only. The **GARS** gene is no longer a key hub gene in the network.

DISCUSSION

In this study, we acquired eighteen publicly available microarray gene expression studies covering

six neurological and mental health disorders; AD, BD, HD, MDD, PD, and SCZ. Data was generated on seven different expression BeadArrays and across two different microarray technologies (Affymetrix and Illumina). The eighteen studies consisted of 3,984 samples extracted from 22 unique brain regions which equated to 67 unique datasets when separating by disorder and tissue. However, due to study and sample outlier analysis, only 43 datasets (22 AD, 6 BD, 4 HD, 2 MDD, 1 PD, and 8 SCZ) totaling 2,667 samples passed QC. We grouped the AD datasets by tissue, into the TL, FL, PL, and CB brain regions to perform the largest microarray AD meta-analysis known to date to our knowledge, which identified 323, 460, 1,736, and 867 significant DEGs, respectively. Furthermore, we incorporated transcriptomic information from other neurological and mental health disorders to subset the initial findings to 323, 435, 1,023, and 828 significant DEGs that were specifically perturbed in the TL, FL, PL, and CB brain regions respectively of AD subjects.

Genes specifically perturbed across AD brain regions

Seven genes (downregulated **NDUFS5**, **SOD1**, **SPCS1** and upregulated **OGT**, **PURA**, **RERE**, **ZFP36L1**) were DE in AD brains and not DE in the other disorders used in this study. We deemed these seven protein-coding genes as “AD-specific”. The expression patterns of three genes (**SPCS1**, **PURA**, and **ZFP36L1**) were relatively mirrored in RNA-Seq data; however, it is important to note the RNA-Seq data does not contain expression profiling for the PL region, and it also contains three specific brain regions within the TL (temporal cortex, superior temporal gyrus, and parahippocampal gyrus) and FL (Inferior frontal gyrus, frontal pole, and dorsolateral prefrontal cortex). Nevertheless, the **SPCS1** gene was observed to be consistently downregulated across all hierarchical AD brain regions available in both the microarray and RNA-Seq data. In addition, based on a network of genomics and epigenomic elements in the region of this genes, in combination with phenotypes, the AMP-AD consortia have nominated **SPCS1** as a druggable target for AD treatment.

Three of the “AD-specific” genes (**NDUFS5**, **SOD1**, and **OGT**) have been previously associated with AD. Down-regulated NADH Dehydrogenase Ubiquinone Fe-S Protein 5 (**NDUFS5**) gene is

part of the human mitochondrial respiratory chain complex; a process suggested to be disrupted in AD in multiple studies [41]. A study investigating blood-based AD biomarkers identified 13 genes, including **NDUFS5**, which was capable of predicting AD with 66% accuracy (67% sensitivity and 75% specificity) in an independent cohort of 118 AD and 118 control subjects [42]. The perturbation in **NDUFS5** expression in the blood and brains of AD subjects suggests this gene may have potential as an AD biomarker and warrants further investigation.

Downregulated Superoxide Dismutase 1 (**SOD1**) gene encodes for copper and zinc ion binding proteins which contribute to the destruction of free superoxide radicals in the body and is also involved in the function of motor neurons [provided by RefSeq, Jul 2008]. Mutations in this gene have been heavily implicated as causes of familial amyotrophic lateral sclerosis (**ALS**) [43] and have also been associated with AD risk [44]. A recent study discovered **SOD1** deficiency in an amyloid precursor protein-overexpressing mouse model accelerated A β oligomerization and also caused tau phosphorylation [45]. They also stated **SOD1** isozymes were significantly decreased in human AD patients, and we can now confirm **SOD1** is significantly under-expressed at the mRNA level in human AD brains as well.

The upregulated O-Linked N-Acetyl Glucosamine Transferase (**OGT**) gene encodes for a glycosyltransferase that links N-acetylglucosamine to serine and threonine residues (O-GlcNAc). O-GlcNAcylation is the post-translational modification of O-GlcNAc and occurs on both neuronal tau and A β PP. Increased brain O-GlcNAcylation has been observed to protect against tau and A β peptide toxicity [46]. A mouse study has demonstrated a deletion of the encoding **OGT** gene causes an increase in tau phosphorylation [47]. In this study, we observe a significant increase in **OGT** gene expression throughout human AD brains, including the CB where tangles are rarely reported, suggesting **OGT** gene is most likely not solely responsible for the formation of tangles.

OGT and O-GlcNAcase (OGA) enzymes facilitate O-GlcNAc cycling, and levels of GlcNAc have also been observed to be increased in the PL of AD brains [48]. Appropriately, OGA inhibitors have been tested for treating AD with promising preliminary results [49], prompting further investigation into targeting **OGT** for AD treatment.

Genes involved in AD histopathology

The CB brain region is known to be free from tau pathology and occasionally free from plaques. We exploited the CB brain region as a secondary control to identify sixteen genes (**ALDOA**, **GABBR1**, **TUBA1A**, **GAPDH**, **DNM3**, **KLC1**, **COX6C**, **ACTG1**, **CLTA**, **SLC25A5**, **PRNP**, **FDFT1**, **RHOQ**, **B2M**, **SPP1**, **WAC**) DE specifically in TL, FL, and PL and not the CB brain region of AD subjects. RNA-Seq data was available for 6 of these genes (**DNM3**, **COX6C**, **ACTG1**, **CLTA**, **RHOQ**, and **B2M**) and all 6 genes failed to replicate expression patterns observed with microarray data. Nevertheless, **DNM3** gene has been previously associated with AD pathology based on proteomic data. **DNM3** gene encodes a member of a family of guanosine triphosphate (GTP)-binding proteins that associate with microtubules and are involved in vesicular transport. A proteomic study identified a module of co-expressed proteins, which included **DNM3**, as negatively correlated with **BRAAK** staging [50]. Although **DNM3** gene expression based on microarray and RNA-Seq data are in disagreement in the CB brain region, a region used in this study to aid in determining whether a gene may be involved with AD pathology, an independent proteomic study demonstrated **DNM3** might indeed be association with AD pathology. This suggests all 6 genes which failed replication in RNA-Seq data may still be associated with AD pathology and require further confirmation.

An additional 9 genes (**GABBR1**, **GAPDH**, **PRPN**, **FDFT1**, **KLC1**, **TUBA1A**, **CLTA**, **COX6C**, and **SLC25A5**), where expression profiling based on RNA-Seq data was unavailable, have also been previously associated with AD, of which four genes (**GABBR1**, **GAPDH**, **PRPN**, and **FDFT1**) have individually been suggested to be involved with the pathogenesis of the disease. **GABBR1** gene encodes a receptor for gamma-aminobutyric acid (GABA), which is the primary inhibitory neurotransmitter in the human central nervous system. As observed in this study, the **GABBR1** gene has been previously reported to be downregulated in AD brains [16]. **GABBR1** receptors are prominent in neuronal soma, where neurofibrillary tangle (NFT) formation is known to accumulate. A study examined the immunohistochemical localization and distribution of **GABBR1** protein in the hippocampus of AD subjects and observed a negative correlation with NFT formation and suggested an increase or stable

expression of **GABBR1** could contribute to neuronal resistance to the disease process [51].

GAPDH gene encodes for a member of the glyceraldehyde-3-phosphate dehydrogenase protein family, which catalyzes an essential step in the carbohydrate metabolism. **GAPDH** has been shown to interact with $A\beta$ PP but not cleaved $A\beta$, and has been proposed to be directly involved in tau aggregation and NFT formation in AD [52–54]. The **PRNP** gene encodes for the prion protein, a membrane glycosylphosphatidylinositol-anchored glycoprotein that tends to aggregate into rod-like structures. Mutations in the **PRNP** gene has been associated with AD and prion protein has also been suggested to be involved in the pathogenesis of AD [55]. **FDFT1** gene encodes a membrane-associated enzyme located at a branch point in the mevalonate pathway, which generates isoprenoids that have been found to be positively correlated with tau pathology [56]. **KLC1** gene encodes for Kinesin Light Chain 1 which transports various cargos such as vesicles, mitochondria, and the Golgi complex along microtubules. An immunoblotting study observed decrease expression of kinesin light chains (KLCs) in the frontal cortex of AD subjects but not in the CB of the same subjects [57]. **TUBA1A** gene encodes for Tubulin Alpha 1a, which has been observed to be perturbed in AD [58], and **CLTA** gene encodes for clathrin Light Chain A, which has been observed to be perturbed in AD as well [59]. **COX6C** and **SLC25A5** gene encodes for products which interact with mitochondria and mitochondrial dysfunction in AD has been suggested on numerous occasions [41, 60, 61].

We identified an additional three AD-specific genes (**UBA1**, **EIF4H**, and **CLDND1**) which were significant DE in all four brain regions. However, the genes were downregulated in the TL FL and PL but upregulated in the CB brain region. Ubiquitin-Like Modifier Activating Enzyme 1 (**UBA1**) encodes for a protein that catalyzes the first step in ubiquitin conjugation to mark cellular proteins for degradation. Eukaryotic Translation Initiation Factor 4H (**EIF4H**) encodes for a translation initiation factors, which functions to stimulate the initiation of protein synthesis at the level of mRNA utilization and Claudin Domain Containing 1 (**CLDND1**) is a transmembrane protein of tight junctions found on endothelial cells [62]. As the CB is the only brain region spared from tangle formation and occasionally from plaque, we suggest these 19 genes (**ALDOA**, **GABBR1**, **TUBA1A**,

GAPDH, DNM3, KLC1, COX6C, ACTG1, CLTA, SLC25A5, PRNP, FDFT1, RHOQ, B2M, SPP1, WAC, UBA1, EIF4H, and CLDND1) could potentially be associated with AD histopathology.

Translation of proteins perturbed specifically in AD brains

Functional gene set enrichment analysis of the “AD expression profile” revealed more pathways were significantly perturbed in the TL, followed by the FL, PL, and CB, which is the general route AD pathology is known to spread through the brain. We originally observed ten biological pathways being enriched across all AD brain regions, which included biological pathways likely to be irrelevant such as the “**digestive system**”. However, when incorporating transcriptomic information from non-AD disorders, we were able to refine the AD expression signature to specific genes perturbed in AD only. This resulted in the enrichment of pathways only involved in the “**metabolism of proteins**”, specifically the translation process which has been previously suggested in be associated with AD on numerous occasions [10, 11, 14–17]. We now suggest this may be a biological process specifically disrupted in AD brains, and not BD, HD, MDD, PD, or SCZ brains.

Previous biological perturbations observed in AD are only associated with the temporal lobe brain region

Previous AD studies have consistently suggested the immune response [10–13], protein transcription/translation regulation [10, 11, 14–17], calcium signaling [10, 18, 19], MAPK signaling [7, 16], chemical synapse [7, 18, 19], neurotransmitter [11, 18, 19], and various metabolism pathways [11, 16, 17, 20–23] are disrupted in AD. We observe the same pathways enriched in our meta-analysis, however, only in the TL brain region, a brain region often heavily investigated in AD. Except for “**metabolism of proteins**”, we did not observe any of these pathways significantly enriched across all of the four brain regions, suggesting these pathways observed to be perturbed in previous studies may be tissue-specific rather than disease-specific.

Interspecies interactions possibly involved in AD

Gene Ontology analysis on the “AD expression profile” identified nine different biological compo-

nents enriched across all four brain regions. However, when we remove genes perturbed in other neurological or mental health disorders, we only observe four biological components as significantly enriched, and all four were indicative of interspecies interactions. AD brains have a prominent inflammatory component which is characteristic of infection, and many microbes have been implicated in AD, notably herpes simplex virus type 1 (HSV1), Chlamydia pneumonia, and several types of spirochete [63]. A very recent study also identified common viral species in normal and aging brains, with an increased human herpesvirus 6A and human herpesvirus 7 in AD brains [64]. Furthermore, A β has been suggested to be an antimicrobial peptide and has been shown to protect against fungal and bacterial infections [65]. Thus, the accumulation of A β may be part of the brains defense mechanism against infections. Although a controversial theory, we also observe a viral component in AD brains, and as a result of this meta-analysis, further suggest this maybe AD-specific and warrants further investigation.

Network analysis identifies AD-specific APP UBC and SUMO2 hub genes

Network analysis identified five (**APP, HSP90AA1, UBC, SUMO2, and RPS2**) significant hub genes specific to AD brain regions. **APP, UBC, and SUMO2** gene appear as hub genes in multiple brain regions. The **APP** gene encodes for a cell surface receptor transmembrane A β PP that is cleaved by secretases to form a number of peptides. Some of these peptides are secreted and can bind to the acetyltransferase complex APBB1/TIP60 to promote transcriptional activation, while others form the protein basis of the amyloid plaques in AD brains. In addition, two of the peptides are antimicrobial peptides, having been shown to have bacteriocidal and antifungal activities [provided by RefSeq, Aug 2014]. Changes in A β PP functions have been suggested to play an essential role in the lack of A β clearance, ultimately leading to the formation of plaques [66].

UBC (ubiquitin-C) gene encodes for a Polyubiquitin-C protein which is part of the ubiquitin-proteasome system (UPS), the primary intracellular protein quality control system in eukaryotic cells. UPS has an immense impact on the amyloidogenic pathway of A β PP processing that generates A β [67]. A recent genome-wide association study identified **UBC** as a novel late-onset AD

gene, and through network analysis also identified **UBC** as a key hub gene. The study validated their findings in a **UBC** *C. elegans* model to discover **UBC** knockout accelerated age-related A β toxicity [68]. We also observe the **UBC** gene being downregulated and as a key hub gene in multiple regions of human AD brains, further providing evidence of its key role in AD.

Small Ubiquitin-Like Modifier 2 (**SUMO2**) gene encodes for a protein that binds to target proteins as part of a post-translational modification system, a process referred to as SUMOylation [69]. However, unlike ubiquitin, which targets proteins for degradation, this protein is involved in a variety of cellular processes, such as nuclear transport, transcriptional regulation, apoptosis, and protein stability [provided by RefSeq, Jul 2008]. Early studies have indicated that the **SUMO** system is likely altered with AD-type pathology, which may impact A β levels and tau aggregation [69]. Genetic studies have supported this theory with a genome-wide association study linking **SUMO**-related genes to late-onset AD [70], with further studies showing that the two natively unfolded proteins, tau and α -synuclein, are sumoylated *in vitro* [71]. We identified **SUMO2** as a significant key hub gene in both the human TL and the CB brain region. However, what makes this discovery interesting is that **SUMO2** is significantly upregulated in the TL, a region where both plaques and tangles can be observed, but significantly downregulated in the CB, where only plaques have been occasionally observed, but tangles never reported. The upregulation of **SUMO2** gene may play a vital role in the formation of tangles, and further investigation into this gene is warranted.

Limitations

Although this study presents novel insights to AD-specific transcriptomic changes in the human brain, limitations to this study must be addressed. Firstly, we meta-analyzed a total of 22 AD and 21 non-AD datasets, and many of these datasets lacked necessary experimental processing or basic phenotypic information such as technical batches, RNA integrity numbers (RIN), age, NFT, clinical gender, or ethnicity, all of which can have confounding effects. To address this, we incorporated recommended best practices to estimate and correct for both known and hidden batch effects using SVA and COMBAT to ensure data is comparable

between experiments and studies. However, this does not guarantee that all technical variation is completely removed.

Secondly, the terminology used to label brain tissue varied across studies, with some reporting a broad region such as the “hippocampus” used in study E-GEOD-48350, while others were particular to the tissue layer, such as “hippocampus CA3” in study E-GEOD-29378. We, therefore, decided to map all brain tissue as mentioned in each dataset publication to their hierarchical cerebral cortex lobe (TL, FL, and PL) and the CB. The mapping procedure was completed using publicly available literature defined knowledge, and we assume tissues within these brain regions are relatively comparable to infer AD-associated histopathological changes.

Thirdly, this study relied on publicly available transcriptomic data, and as previous research has heavily investigated brain regions known to be at the forefront of disease manifestation, this led to unbalanced datasets per brain region in both the AD and non-AD meta-analysis. Subsequently, the AD meta-analysis consisted of 14, 4, 2, and 2 datasets for the TL, FL, PL, and CB brain regions respectively, with the PL brain region consisting of only 74 samples (28 AD and 46 controls) in total. In addition, the non-AD meta-analysis lacked expression signatures from all non-AD diseases across all brain regions (except for FL). Nevertheless, the brain regions most affected by each disorder was captured in this study, suggesting we most likely were able to capture key brain transcriptomic changes relating to each disorder. Furthermore, as AD is known to affect all brain regions, albeit not to the same extent, we focus on transcriptomic changes observed across all brain regions that are also not observed in any brain region of the non-AD subjects, ensuring we capture transcriptomic signatures unique to AD brains.

Fourthly, the advances in next sequencing technologies (RNA-Seq) which are capable of profiling the whole transcriptome, thus not limited by the predefined probes based on known sequencing, would be ideal for disease discoveries. However, AD and mental health studies profiled through RNA-Seq is somewhat limited in the public domain, and those that have published DE results are based on small sample numbers, which would fail our selection criteria, such as [72–75]. In addition, these studies lack the same brain regions and mental health disorders covered in this meta-analysis. Nevertheless, we were able to query our genes of interest in the largest known AD RNA-Seq web-based database (Agora) which

contains DE results from over 2100 human brain samples; however, expression profiling was unavailable for 17/26 genes, and DE on the PL was unavailable. Therefore, this study was unable to validate all findings in RNA-Seq data.

Finally, we assume the non-AD datasets are comparable through meta-analysis, and by identifying common expression signatures that are not associated with individual disease mechanisms may represent false positives or even a general signature for “brain disorder”. Removing this signature from the AD meta-analysis expression profile may result in transcriptomic changes specific to AD brains, revealing more relevant changes to the underlying disease mechanism rather than general diseases. Under this assumption, we observe more relevant and refined biological enrichment results. For example, we originally observed ten biological pathways enriched across all AD brain regions, including biological pathways such as the “**digestive system**”. However, by refining the AD expression signature by removing genes perturbed in non-AD disorders, only pathways involved in the “**metabolism of proteins**” remain, which has been previously suggested in be associated with AD on numerous occasions [10, 11, 14–17]. This observation provides strong evidence of our assumption of incorporating non-AD diseases in this study to infer AD-specific changes as valid.

Conclusion

We present the most extensive human AD brain microarray transcriptomic meta-analysis study to date, incorporating, brain regions both affected and partially spared by AD pathology, and utilize related non-AD disorders to infer AD-specific brain changes. This led to the identification of seven genes specifically perturbed across all AD brain regions and are considered disease-specific, nineteen genes specifically perturbed in AD brains which could play a role in AD neuropathology, and the refinement of GSEA and GO analysis results to identify specific biological pathways and components specific to AD. These AD-specific changes may provide new insights into the disease mechanisms, thus making a significant contribution towards understanding the disease.

ACKNOWLEDGMENTS

This study presents independent research supported by the NIHR BioResource Centre Maudsley at South London and Maudsley NHS Foundation Trust

(SLaM) & Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King’s College London. The views expressed are those of the author(s) and not necessarily those of the NHS, NIHR, Department of Health or King’s College London.

RJBD and SJN are supported by 1. Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust. 2. The National Institute for Health Research University College London Hospitals Biomedical Research Centre.

Authors’ disclosures available online (<https://www.j-alz.com/manuscript-disclosures/18-1085r3>).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <http://dx.doi.org/10.3233/JAD-181085>.

REFERENCES

- [1] Prince M, Albanese E, Guerchet M, Prina M (2014) *World Alzheimer Report 2014. Dementia and risk reduction: An analysis of protective and modifiable factors*. Alzheimer’s Disease International, London, UK.
- [2] Rajan KB, Wilson RS, Weuve J, Barnes LL, Evans DA (2015) Cognitive impairment 18 years before clinical diagnosis of Alzheimer disease dementia. *Neurology* **85**, 898–904.
- [3] Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT (2011) Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med* **1**, 1–23.
- [4] Convit A, De Asis J, De Leon MJ, Tarshish CY, De Santi S, Rusinek H (2000) Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer’s disease. *Neurobiol Aging* **21**, 19–26.
- [5] Jacobs HIL, Hopkins DA, Mayrhofer HC, Bruner E, van Leeuwen FW, Raaijmakers W, Schmahmann JD (2018) The cerebellum in Alzheimer’s disease: Evaluating its role in cognitive decline. *Brain* **141**, 37–47.
- [6] Zhang M, Yao C, Guo Z, Zou J, Zhang L, Xiao H, Wang D, Yang D, Gong X, Zhu J, Li Y, Li X (2008) Apparently low reproducibility of true differential expression discoveries in microarray studies. *Bioinformatics* **24**, 2057–2063.
- [7] Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH (2013) Genes and pathways underlying regional and cell type changes in Alzheimer’s disease. *Genome Med* **5**, 48.

- [8] Hokama M, Oka S, Leon J, Ninomiya T, Honda H, Sasaki K, Iwaki T, Ohara T, Sasaki T, LaFerla FM, Kiyohara Y, Nakabeppu Y (2014) Altered expression of diabetes-related genes in Alzheimer's disease brains: The Hisayama study. *Cereb Cortex* **24**, 2476-2488.
- [9] Barnes M, Freudenberg J, Thompson S, Aronow B, Pavlidis P (2005) Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms. *Nucleic Acids Res* **33**, 5914-5923.
- [10] Sekar S, McDonald J, Cuyugan L, Aldrich J, Kurdoglu A, Adkins J, Serrano G, Beach TG, Craig DW, Valla J, Reiman EM, Liang WS (2015) Alzheimer's disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes. *Neurobiol Aging* **36**, 583-591.
- [11] Li Y, Wu Z, Jin Y, Wu A, Cao M, Sun K, Jia X, Chen M (2012) Analysis of hippocampal gene expression profile of Alzheimer's disease model rats using genome chip bioinformatics. *Neural Regen Res* **7**, 332-340.
- [12] Lambert JC, Grenier-Boley B, Chouraki V, Heath S, Zelenika D, Fievet N, Hannequin D, Pasquier F, Hanon O, Brice A, Epelbaum J, Berr C, Dartigues JF, Tzourio C, Campion D, Lathrop M, Amouyel P (2010) Implication of the immune system in Alzheimer's disease: Evidence from genome-wide pathway analysis. *J Alzheimers Dis* **20**, 1107-1118.
- [13] Chen J, Xie C, Zhao Y, Li Z, Xu P (2016) Gene expression analysis reveals the dysregulation of immune and metabolic pathways in Alzheimer's disease. *Oncotarget* **7**, 1-6.
- [14] Liu T, Ren D, Zhu X, Yin Z, Jin G, Zhao Z, Robinson D, Li X, Wong K, Cui K, Zhao H, Wong STC (2013) Transcriptional signaling pathways inversely regulated in Alzheimer's disease and glioblastoma multiform. *Sci Rep* **3**, 3467.
- [15] Li X, Long J, He T, Belshaw R, Scott J (2015) Integrated genomic approaches identify major pathways and upstream regulators in late onset Alzheimer's disease. *Sci Rep* **5**, 12393.
- [16] Puthiyedth N, Riveros C, Berretta R, Moscato P (2016) Identification of differentially expressed genes through integrated study of Alzheimer's disease affected brain regions. *PLoS One* **11**, 1-29.
- [17] Godoy JA, Rios JA, Zolezzi JM, Braidly N, Inestrosa NC (2014) Signaling pathway cross talk in Alzheimer's disease. *Cell Commun Signal* **12**, 23.
- [18] Ramanan VK, Kim S, Holohan K, Shen L, Nho K, Risacher SL, Foroud TM, Mukherjee S, Crane PK, Aisen S, Petersen RC, Weiner MW, Saykin AJ (2013) Genome-wide pathway analysis of memory impairment in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort implicates gene candidates, canonical pathways, and networks. *Brain Imaging Behav* **6**, 634-648.
- [19] Blalock EM, Buechel HM, Popovic J, Geddes JW, Landfield PW (2011) Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *J Chem Neuroanat* **42**, 118-126.
- [20] Paolo G Di, Kim T (2012) Linking lipids to Alzheimer's disease: Cholesterol and beyond. *Aging (Albany, NY)* **12**, 284-296.
- [21] Liang WS, Reiman EM, Valla J, Dunckley T, Beach TG, Grover A, Niedzielko TL, Schneider LE, Mastroeni D, Caselli R, Kukull W, Morris JC, Hulette CM, Schmechel D, Rogers J, Stephan DA (2008) Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proc Natl Acad Sci U S A* **105**, 4441-4446.
- [22] Ishii K, Sasaki M, Kitagaki H, Yamaji S, Sakamoto S, Matsuda K, Mori E (1997) Reduction of cerebellar glucose metabolism in advanced Alzheimer's disease. *J Nucl Med* **38**, 925-928.
- [23] Oshiro S, Morioka MS, Kikuchi M (2011) Dysregulation of iron metabolism in Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. *Adv Pharmacol Sci* **2011**, 378278.
- [24] Edwards YJK, Beecham GW, Scott WK, Khuri S, Bademci G, Tekin D, Martin ER, Jiang Z, Mash DC, French-Mullen J, Pericak-Vance MA, Tsinoremas N, Vance JM (2011) Identifying consensus disease pathways in Parkinson's disease using an integrative systems biology approach. *PLoS One* **6**, e16917.
- [25] Chandrasekaran S, Bonchev D (2013) A network view on Parkinson's disease. *Comput Struct Biotechnol J* **7**, e201304004.
- [26] Clelland CL, Read LL, Panek LJ, Nadrich RH, Bancroft C, Clelland JD (2013) Utilization of never-medicated bipolar disorder patients towards development and validation of a peripheral biomarker profile. *PLoS One* **8**, 1-11.
- [27] Nurnberger JI, Koller DL, Jung J, Edenberg HJ, Foroud T, Guella I, Vawter MP, Kelsoe JR (2014) Identification of pathways for bipolar disorder. *JAMA Psychiatry* **71**, 657.
- [28] Chen H, Wang N, Zhao X, Ross CA, O'Shea KS, Mcinnis MG (2013) Gene expression alterations in bipolar disorder postmortem brains. *Bipolar Disord* **15**, 177-187.
- [29] Khanzada N, Butler M, Manzardo A (2017) GeneAnalytics pathway analysis and genetic overlap among autism spectrum disorder, bipolar disorder and schizophrenia. *Int J Mol Sci* **18**, 527.
- [30] William R. Markesbery (2010) Neuropathologic alterations in mild cognitive impairment: A review. *J Alzheimers Dis* **19**, 221-228.
- [31] Lazar C, Meganck S, Taminau J, Steenhoff D, Coletta A, Molter C, Weiss-Solis DY, Duque R, Bersini H, Nowé A (2013) Batch effect removal methods for microarray gene expression data integration: A survey. *Brief Bioinform* **14**, 469-490.
- [32] Hackstadt AJ, Hess AM (2009) Filtering for increased power for microarray data analysis. *BMC Bioinformatics* **10**, 11.
- [33] Oldham MC, Langfelder P, Horvath S (2012) Network methods for describing sample relationships in genomic datasets: Application to Huntington's disease. *BMC Syst Biol* **6**, 63.
- [34] Kang DD, Sibille E, Kaminski N, Tseng GC (2012) MetaQC: Objective quality control and inclusion/exclusion criteria for genomic meta-analysis. *Nucleic Acids Res* **40**, 1-14.
- [35] Chang L-C, Lin H-M, Sibille E, Tseng GC (2013) Meta-analysis methods for combining multiple expression profiles: Comparisons, statistical characterization and an application guideline. *BMC Bioinformatics* **14**, 368.
- [36] Dalman MR, Deeter A, Nimishakavi G, Duan Z-H (2012) Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics* **13**(Suppl 2), S11.
- [37] St G, Shtokalo D, Tackett MR, Yang Z, Vyatkin Y, Milos PM, Seilheimer B, McCaffrey TA, Kapranov P (2013) On the importance of small changes in RNA expression. *Methods* **63**, 18-24.
- [38] Kamburov A, Wierling C, Lehrach H, Herwig R (2009) ConsensusPathDB - A database for integrating human

- functional interaction networks. *Nucleic Acids Res* **37**, 623-628.
- [39] Xia J, Benner MJ, Hancock REW (2014) NetworkAnalyst - Integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res* **42**, 167-174.
- [40] Zaki N, Mora A (2015) A comparative analysis of computational approaches and algorithms for protein subcomplex identification. *Sci Rep* **4**, 4262.
- [41] Bonilla E, Tanji K, Hirano M, Vu TH, Dimauro S, Schon EA (1999) Mitochondrial involvement in Alzheimer's disease. *Biochim Biophys Acta* **1410**, 171-182.
- [42] Voyle N, Keohane A, Newhouse S, Lunnon K, Johnston C, Soininen H, Kloszewska I, Mecocci P, Tsolaki M, Vellas B, Lovestone S, Hodges A, Kiddle S, Dobson RJB (2016) A pathway based classification method for analyzing gene expression for Alzheimer's disease diagnosis. *J Alzheimers Dis* **49**, 659-669.
- [43] Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, Donaldson D, Goto J, O'Regan JP, Deng HX, Daniel R, Rosen, Teepu Siddique DP, Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, Donaldson D, Goto J, O'Regan JP, Deng HX, Daniel R, Rosen, Teepu Siddique DP (1993) Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59-62.
- [44] Spisak K, Klimkowicz-Mrowiec A, Pera J, Dziedzic T, Golenia A, Slowik A (2015) rs2070424 of the SOD1 gene is associated with risk of Alzheimer's disease. *Neurol Neurochir Pol* **48**, 342-345.
- [45] Murakami K1, Murata N, Noda Y, Tahara S, Kaneko T, Kinoshita N, Hatsuta H, Murayama S, Barnham KJ, Irie K, Shirasawa T, Shimizu T (2011) SOD1 (copper/zinc superoxide dismutase) deficiency drives amyloid β protein oligomerization and memory loss in mouse model of Alzheimer disease. *J Biol Chem* **286**, 44557-44568.
- [46] Zhu Y, Shan X, Yuzwa SA, Vocado DJ (2014) The emerging link between O-GlcNAc and Alzheimer disease. *J Biol Chem* **289**, 34472-34481.
- [47] O'Donnell N, Zachara NE, Hart GW, Marth JD (2004) Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol Cell Biol* **24**, 1680-1690.
- [48] Förster S, Welleford AS, Triplett JC, Sultana R, Schmitz B, Butterfield DA (2014) Increased O-GlcNAc levels correlate with decreased O-GlcNAcase levels in Alzheimer disease brain. *Biochim Biophys Acta* **1842**, 1333-1339.
- [49] Yuzwa SA, Shan X, Jones BA, Zhao G, Woodward ML, Li X, Zhu Y, McEachern EJ, Silverman MA, Watson NV, Gong CX, Vocado DJ (2014) Pharmacological inhibition of O-GlcNAcase (OGA) prevents cognitive decline and amyloid plaque formation in bigenic tau/APP mutant mice. *Mol Neurodegener* **9**, 42.
- [50] Seyfried NT, Dammer EB, Swarup V, Nandakumar D, Duong DM, Yin L, Deng Q, Nguyen T, Hales CM, Wingo T, Glass J, Gearing M, Thambisetty M, Troncoso JC, Geschwind DH, Lah JJ, Levey AI (2017) A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. *Cell Syst* **4**, 60-72.e4.
- [51] Iwakiri M, Mizukami K, Ikonovic MD, Ishikawa M, Hidaka S, Abrahamson EE, DeKosky ST, Asada T (2005) Changes in hippocampal GABABR1 subunit expression in Alzheimer's patients: Association with Braak staging. *Acta Neuropathol* **109**, 467-474.
- [52] El Kadmiri N, Slassi I, El Moutawakil B, Nadifi S, Tadevosyan A, Hachem A, Soukri A (2014) Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and Alzheimer's disease. *Pathol Biol* **62**, 333-336.
- [53] Butterfield DA, Hardas SS, Lange MLB (2010) Oxidatively modified glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and Alzheimer's disease: Many pathways to neurodegeneration. *J Alzheimers Dis* **20**, 369-393.
- [54] Itakura M, Nakajima H, Kubo T, Semi Y, Kume S, Higashida S, Kaneshige A, Kuwamura M, Harada N, Kita A, Azuma Y-T, Yamaji R, Inui T, Takeuchi T (2015) Glyceraldehyde-3-phosphate dehydrogenase aggregates accelerate amyloid- β amyloidogenesis in Alzheimer disease. *J Biol Chem* **290**, 26072-26087.
- [55] Zhang W, Jiao B, Xiao T, Pan C, Liu X, Zhou L, Tang B, Shen L (2016) Mutational analysis of PRNP in Alzheimer's disease and frontotemporal dementia in China. *Sci Rep* **6**, 38435.
- [56] Pelleieux S, Picard C, Lamarre-Théroux L, Dea D, Leduc V, Tsantrizos YS, Poirier J (2018) Isoprenoids and tau pathology in sporadic Alzheimer's disease. *Neurobiol Aging* **65**, 132-139.
- [57] Morel M, Héraud C, Nicaise C, Suain V, Brion J-P (2012) Levels of kinesin light chain and dynein intermediate chain are reduced in the frontal cortex in Alzheimer's disease: Implications for axoplasmic transport. *Acta Neuropathol* **123**, 71-84.
- [58] Lachen-Montes M, Zelaya MV, Segura V, Fernández-Irigoyen J, Santamaría E (2017) Progressive modulation of the human olfactory bulb transcriptome during Alzheimer's disease evolution: Novel insights into the olfactory signaling across proteinopathies. *Oncotarget* **8**, 69663-69679.
- [59] Nakamura Y, Takeda M, Yoshimi K, Hattori H, Hariguchi S, Kitajima S, Hashimoto S, Nishimura T (1994) Involvement of clathrin light chains in the pathology of Alzheimer's disease. *Acta Neuropathol* **87**, 23-31.
- [60] Swerdlow RH, Khan SM (2004) A "mitochondrial cascade hypothesis" for sporadic Alzheimer's disease. *Med Hypotheses* **63**, 8-20.
- [61] Iwamoto K, Bundo M, Kato T (2005) Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Hum Mol Genet* **14**, 241-253.
- [62] Krause G, Winkler L, Mueller SL, Haseloff RF, Piontek J, Blasig IE (2008) Structure and function of claudins. *Biochim Biophys Acta* **1778**, 631-645.
- [63] Sethi AJ, Wikramanayake RM, Angerer RC, Range RC, Angerer LM (2016) Microbes and Alzheimer's disease. *Science* **335**, 590-593.
- [64] Readhead B, Haure-Mirande JV, Funk CC, Richards MA, Shannon P, Haroutunian V, Sano M, Liang WS, Beckmann ND, Price ND, Reiman EM, Schadt EE, Ehrlich ME, Gandy S, Dudley JT (2018) Multiscale analysis of independent Alzheimer's cohorts finds disruption of molecular, genetic, and clinical networks by human herpesvirus. *Neuron* **99**, 64-82.e7.
- [65] Kumar DKV, Choi SH, Washicosky KJ, Eimer WA, Tucker S, Ghofrani J, Lefkowitz A, McColl G, Goldstein LE, Tanzi RE, Moir RD (2016) Amyloid-beta peptide protects against microbial infection in mouse and worm models of Alzheimer's disease. *Sci Transl Med* **8**, 340ra72-340ra72.

- [66] O'Brien RJ, Wong PC (2011) Amyloid precursor protein processing and Alzheimer's disease. *Annu Rev Neurosci* **34**, 185-204.
- [67] Hong L, Huang H-C, Jiang Z-F (2014) Relationship between amyloid-beta and the ubiquitin-proteasome system in Alzheimer's disease. *Neurol Res* **36**, 276-282.
- [68] Mukherjee S, Russell JC, Carr DT, Burgess JD, Younkin MA, Serie DJ, Boehme KL, Kauwe JSK, Naj AC, Fardo DW, Dickson DW, Montine TJ, Ertekin-Taner N, Kaeberlein MR, Crane PK, Allen M, Serie DJ, Boehme KL, Kauwe JSK, Naj AC, Fardo DW, Dickson DW, Montine TJ, Ertekin-Taner N, Kaeberlein MR, Crane PK (2017) Systems biology approach to late-onset Alzheimer's disease genome-wide association study identifies novel candidate genes validated using brain expression data and *Caenorhabditis elegans* experiments. *Alzheimers Dement* **13**, 1133-1142.
- [69] Lee L, Sakurai M, Matsuzaki S, Arancio O, Fraser P (2013) SUMO and Alzheimer's disease. *Neuromolecular Med* **15**, 720-736.
- [70] Grupe A, Abraham R, Li Y, Rowland C, Hollingworth P, Morgan A, Jehu L, Segurado R, Stone D, Schadt E, Karnoub M, Nowotny P, Tacey K, Catanese J, Sninsky J, Brayne C, Rubinsztein D, Gill M, Lawlor B, Lovestone S, Holmans P, O'Donovan M, Morris JC, Thal L, Goate A, Owen MJ, Williams J (2007) Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. *Hum Mol Genet* **16**, 865-873.
- [71] Dorval V, Fraser PE (2006) Small ubiquitin-like modifier (SUMO) modification of natively unfolded proteins tau and α -synuclein. *J Biol Chem* **281**, 9919-9924.
- [72] Annese A, Manzari C, Lionetti C, Picardi E, Horner DS, Chiara M, Caratozzolo MF, Tullo A, Fosso B, Pesole G, Erchia AMD, D'Erchia AM, Erchia AMD (2018) Whole transcriptome profiling of late-onset Alzheimer's disease patients provides insights into the molecular changes involved in the disease. *Sci Rep* **8**, 1-15.
- [73] Magistri M, Velmeshev D, Makhmutova M, Faghihi MA (2015) Transcriptomics profiling of Alzheimer's disease reveal neurovascular defects, altered amyloid- β homeostasis, and deregulated expression of long noncoding RNAs. *J Alzheimers Dis* **48**, 647-665.
- [74] Bennett JP, Keeney PM (2018) RNA-sequencing reveals similarities and differences in gene expression in vulnerable brain tissues of Alzheimer's and Parkinson's diseases. *J Alzheimers Dis Rep* **2**, 129-137.
- [75] Mills JD, Nalpathamkalam T, Jacobs HIL, Janitz C, Merico D, Hu P, Janitz M (2013) RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. *Neurosci Lett* **536**, 90-95.
- [76] Berchtold NC, Coleman PD, Cribbs DH, Rogers J, Gillen DL, Cotman CW (2013) Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol Aging* **34**, 1653-1661.
- [77] Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR, Landfield PW (2004) Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc Natl Acad Sci U S A* **101**, 2173-2178.
- [78] Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, Carrasquillo MM, Rowley CN, Nair AA, Middha S, Maharjan S, Nguyen T, Ma L, Malphrus KG, Palusak R, Lincoln S, Bisceglia G, Georgescu C, Kouri N, Kolbert CP, Jen J, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Schellenberg GD, Petersen RC, Graff-Radford NR, Dickson DW, Younkin SG, Ertekin-Taner N (2012) Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet* **8**, e1002707.
- [79] Wang M, Roussos P, McKenzie A, Zhou X, Kajiwaru Y, Brennand KJ, De Luca GC, Crary JF, Casaccia P, Buxbaum JD, Ehrlich M, Gandy S, Goate A, Katsel P, Schadt E, Haroutunian V, Zhang B (2016) Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Med* **8**, 1-21.