Neurobiology of Aging 74 (2019) 225–233

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

Hippocampal transcriptome profiling combined with protein-protein interaction analysis elucidates Alzheimer's disease pathways and genes

Jeroen G.J. van Rooij [a,b,*], Lieke H.H. Meeter [a], Shami Melhem [a], Diana A.T. Nijholt [a], Tsz Hang Wong [a], Netherlands Brain Bank[c], Annemieke Rozemuller [d], Andre G. Uitterlinden [b], Joyce G. van Meurs [b,1], John C. van Swieten [a,1]

[a] Department of Neurology, Erasmus Medical Center, Rotterdam, the Netherlands
[b] Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands
[c] Netherlands Institute for Neuroscience, Amsterdam, the Netherlands
[d] Department of Pathology, VU University Medical Center, Amsterdam, the Netherlands

ARTICLE INFO

Article history:
Received 18 December 2017
Received in revised form 12 October 2018
Accepted 24 October 2018
Available online 29 October 2018

Keywords:
Alzheimer's disease
RNA sequencing
Protein-protein interactions
Hippocampus

ABSTRACT

Knowledge about the molecular mechanisms driving Alzheimer's disease (AD) is still limited. To learn more about AD biology, we performed whole transcriptome sequencing on the hippocampus of 20 AD cases and 10 age- and sex-matched cognitively healthy controls. We observed 2716 differentially expressed genes, of which 48% replicated in a second data set of 84 AD cases and 33 controls. We used an integrative network-based approach for combining transcriptomic and protein-protein interaction data to find differentially expressed gene modules that may reflect key processes in AD biology. A total of 735 differentially expressed genes were clustered into 33 modules, of which 82% replicated in a second data set, highlighting the robustness of this approach. These 27 modules were enriched for signal transduction, transport, response to stimulus, and several organic and cellular metabolic pathways. Ten modules interacted with previously described AD genes. Our study indicates that analyzing RNA-expression data based on annotated gene modules is more robust than on individual genes. We provide a comprehensive overview of the biological processes involved in AD, and the detected differentially expressed gene modules may provide a molecular basis for future research into mechanisms underlying AD.

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder hallmarked by progressive loss of memory, currently affecting over 40 million individuals worldwide (Prince et al., 2013; Scheltens et al., 2016). Previous studies have shown neurodegenerative changes in the hippocampus 15–20 years before symptom onset (Boyle et al., 2013; Karran et al., 2011; Murray et al., 2011). The main pathological features are amyloid plaques and tau tangles throughout the brain (Braak and Braak, 1995; Holtzman et al., 2016; Jellinger, 2008; Selkoe and Hardy, 2016; Thal et al., 2014; Tomiyama, 2010). Multiple AD-associated genetic loci have been identified, although their pathophysiological mechanisms remain largely unknown (Bekris et al., 2010; Lambert et al., 2013; Van Cauwenberghe et al., 2016).

Transcriptomic studies on postmortem AD brain tissue have been performed to further our understanding of AD biology (Kavanagh et al., 2013; Sutherland et al., 2011). Most of these studies report differentially expressed genes and pathways in brain tissue of AD cases compared with controls (Ashburner et al., 2000; Gene Ontology, 2015; Ogata et al., 1999). Most of these studies report a decrease in synaptic transmission, mitochondrial function, and cytoskeleton biology. By contrast, an increase is often reported in immune response, inflammation, and apoptosis in AD cases (Liang et al., 2008; Ray and Zhang, 2010; Sekar et al., 2015; Twine et al., 2011). Recently, network-based analysis are used to provide more extensive and robust insights in these data, for example, based on protein-protein interaction (PPI) data (Chi et al., 2016; C. Humphries et al., 2015; C.E. Humphries et al., 2015; Kong et al., 2014, 2015). The largest among these studies investigated gene

* Corresponding author at: Department of Neurology, Erasmus Medical Center, Room Ee-2238, Wytemaweg 80, 3015 CN, Rotterdam, the Netherlands. Tel.: +31107043822; fax: +31107044721.
E-mail address: j.vanrooij@erasmusmc.nl (J.G.J. van Rooij).
[1] These authors contributed equally.

https://doi.org/10.1016/j.neurobiolaging.2018.10.023

expression (measured with RNA arrays) in more than a thousand brain samples, spread across 19 regions in 125 individuals (Wang et al., 2016). By performing gene coexpression analysis on AD cases of varying severity and nondemented controls, they identified dysregulated gene modules and pathways. The study concluded that some of those originated from early disease stages and might reflect causal mechanisms, but also highlighted to use of gene modules rather than individual genes. In this study, these modules are based on only coexpression type PPI data.

The goal of our study was to compare whole transcriptome sequencing of 20 AD cases with 10 age- and sex-matched cognitively healthy controls. We aim to identify differentially expressed genes and cluster these into functional gene modules using PPI data. We aim to replicate these differentially expressed genes, gene modules, and functions in a second independent RNA sequencing data set (van der Brug et al., 2017) to determine the robustness of replication based on gene modules compared with individual genes.

2. Materials and methods

2.1. Data generation

Hippocampus samples were selected from the Netherlands Brain Bank for 20 AD cases (Braak and Braak, 1995; Mirra et al., 1991) and matched for age and gender with brains from 10 nondemented cognitively healthy controls (Table 1). The dentate gyrus and cornu amonis were macrodissected from the hippocampus tissue, and total RNA was isolated using the manufacturer's protocol (Qiagen AllPrep RNA isolation, Cat No. 80224). Sequencing was performed after poly-A selection and TruSeq library prep at the Human Genomics facility (www.glimdna.org) on a HiSeq2000 at 2 × 50 bp. Data were processed per sample using trim-o-matic (v0.33), STAR (v2.3.0) (Bolger et al., 2014; Dobin et al., 2013), picard (v1.90), and fastQC (v0.11.3). Transcript quantification was performed using featureCounts (v1.4.3) against all 57,820 gene features in GENCODE (version date; 2013-12-05) (Harrow et al., 2012; Liao et al., 2014). For replication, data set GSE95587 was downloaded from the Gene Expression Omnibus. This data set contained raw RNA-seq counts of the fusiform gyrus for 84 AD cases and 33 controls and was processed in parallel to the discovery data set in all subsequent steps (van der Brug et al., 2017).

2.2. Data analysis

Counts were normalized using the edgeR (v3.8.6) trimmed mean of M-values method to counts per million values, and all low-abundant features were omitted (<1 counts per million in 75% of samples). Principal components (PCs) were calculated using "prcomp" in R, and then plotted to visually identify sample outliers.

Table 1
Study sample characteristics

Characteristic	Controls	Cases	Cases_QC
Number	10	20	18
Gender (% male)	50%	30%	44%
Age (±SD)	76 ± 12	75 ± 7	75 ± 7
Braak	1.5 ± 1.3	5.5 ± 0.5*	5.6 ± 0.5*
Amyloid	0.9 ± 1.1	2.9 ± 0.3*	2.9 ± 0.3*
pmd	551 ± 297	348 ± 108*	329 ± 98*
pH	6.6 ± 0.3	6.3 ± 0.3*	6.3 ± 0.3*
Brain weight	1319 ± 240	1045 ± 119*	1035 ± 113*
apoe (32/33/44)	4/6/0	1/9/10*	1/8/9*

An asterisk denotes statistically significant difference compared with controls.
All values represent means with standard deviations unless otherwise indicated.
"Cases_QC" indicates metrics after removing two outlier cases.

Statistical analysis was performed per gene using the exacTest function in edgeR, correcting for age, gender, and the first 2 PCs (McCarthy et al., 2012; Robinson et al., 2010). We combined FDR-corrected p-values and log fold changes to calculate a differential expression score; $\frac{-\log_{10} p_{(FDR)}}{10} * \sqrt{\frac{logFC * logFC}{3}}$. Genes with a differentially expressed score ≥0.10 are considered differentially expressed genes (DE genes) and retained for further analysis. All steps were performed identically and separately for the discovery and replication data sets Table 2.

2.3. Protein-protein interaction clustering

For all DE genes, we extracted experimental, coexpression, and database interactions scored ≥500 from STRING v10 (von Mering et al., 2003). This network was imported to Cytoscape (v3.4.0) and subjected to the Markov Clustering Algorithm (MCL) to identify gene modules (Morris et al., 2011; Smoot et al., 2011). In short, MCL clusters graphical data to determine groups of genes (modules) with more interactions within the module than to the rest of the network (Enright et al., 2002). This clustering method revolves around one main parameter which determines the module sizes; the inflation factor. We optimized the inflation factor to retain modules between 10–100 genes to allow for subsequent gene set enrichment analysis (Subramanian et al., 2005). Each gene can only be assigned to a single module. Modules smaller than 10 genes are excluded. All steps are performed separately for the discovery and replication data set.

2.4. Functional annotation of modules

For each identified gene module, enrichments for gene ontology biological processes (GOBPs) were performed using WebGestalt (v27-1-17) (Gene Ontology, 2001; Ogata et al., 1999). For GOBP enrichment the "noRedundant" terms were used. All enrichments were FDR (Benjamini-Hochberg) corrected, using a threshold of $p < 0.05$ for statistical significance. Only the first three enriched GOBP terms were extracted for each gene module. All three GOBP terms for all gene modules were then pooled and divided into shared common ancestor terms, denoted as GOBP branches (Ashburner et al., 2000; Carbon et al., 2009). Therefore, each gene module can be annotated with three GOBP terms and their respective GOBP branch. Modules from discovery and replication are divided in to the same GOBP branches. They can thus be enriched for the same GOBP term or enriched for different GOBP terms that are closely related by sharing a common ancestor term.

2.5. Replication of DE genes and modules

DE genes and gene modules were generated separately for the discovery and replication data sets using the exact same methodology. Replication of discovery modules is assessed by the number of overlapping genes and overlapping GOBP terms within the replication modules. Different degrees of robustness of overlap between our data and the replication data set were classified. Category (1) a gene module overlaps in genes and in GOBP term(s) with a gene module from the replication data set. Category (2) a gene module overlaps in GOBP term(s), but not in genes with a replication module. Category (3) a gene module overlaps in genes with a replication module, but not in GOBP term(s). When a module from discovery shares a parent GOBP term with a replication module this was also considered replication.

Table 2
Overview of all three gene ontology biological processes of the discovery modules

Discovery module	Number of genes	GOBP branch	GOBP term	Replication module	AD genes
M1	90	2	G-protein–coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	R1	APP
		2	Phospholipase C-activating G-protein–coupled receptor signaling pathway	R1	
		2	Neuropeptide signaling pathway	R1, R7, R35	
M2	52	9	Multicellular organismal signaling	R10	MAPT, ABCA7
		3	Divalent inorganic cation transport	R10	
		4	Regulation of transmembrane transport	R10, R18	
M3	39	1	DNA-templated transcription, initiation	R2	
		8	Response to type I interferon	-	
		8	Response to interferon-gamma	-	
M4	35	1	Macromolecule deacylation	R3	
		1	Histone modification	R3	
		4	Regulation of chromatin organization	-	
M5	32	4	Potassium ion transport	R18	
		4	Regulation of transmembrane transport	R10, R18	
		6	Protein oligomerization	R18	
M6	31	4	Regulation of small GTPase mediated signal transduction	R5	
		2	Ras protein signal transduction	R5	
		4	Regulation of cell morphogenesis	-	
M7	30	2	Glutamate receptor signaling pathway	R35	PTK2B
		4	Modulation of synaptic transmission	-	
		9	Neuromuscular process	-	
M8	29	5	mRNA processing	R17	CELF1
		1	Peptidyl-threonine modification	R15	
		5	RNA splicing	R17	
M9	27	3	Receptor-mediated endocytosis	-	BIN1, HLA-DRB1,
		9	Microtubule-based movement	-	HLA-DRB5, PICALM
		4	Regulation of response to biotic stimulus	R12	
M10	26	2	Integrin-mediated signaling pathway	R4	
		6	Extracellular structure organization	R4	
		9	Cell-substrate adhesion	-	
M11	21	1	Peptidyl-tyrosine modification	R19, R25	
		6	Axon development	-	
		1	Ephrin receptor signaling pathway	R25	
M12	22	1	DNA replication	R36	
		1	DNA repair	R22, R36	
		1	DNA recombination	R22, R36	
M13	20	-	-	-	
		-	-	-	
		-	-	-	
M14	19	3	Exocytosis	R27	CLU
		9	Actin filament-based movement	R34	
		6	Actin filament organization	R37	
M15	19	5	Dephosphorylation	-	
		5	RNA splicing	R17	
		9	Meiotic cell cycle	-	
M16	18	3	Synaptic vesicle cycle	-	
		3	Exocytosis	R27	
		3	Neurotransmitter transport	-	
M17	17	8	Stress-activated protein kinase signaling cascade	-	
		4	Positive regulation of MAPK cascade	-	
		4	Positive regulation of kinase activity	R9	
M18	18	2	I-kappaB kinase/NF-kappaB signaling	R12, R23	
		8	Cellular response to biotic stimulus	R23	
		9	Type I interferon production	-	
M19	17	1	Translational elongation	-	
		1	Mitochondrial translation	-	
		6	Macromolecular complex disassembly	-	
M20	17	3	Inorganic anion transport	R7	
		3	Anion transmembrane transport	R7	
		2	Gamma-aminobutyric acid signaling pathway	R7	
M21	15	3	Transition metal ion transport	R24	
		3	Hydrogen transport	R18	
		7	Autophagy	-	
M22	15	6	NADH dehydrogenase complex assembly	-	
		6	Mitochondrial respiratory chain complex assembly	-	
		9	Mitochondrial respiratory chain complex I biogenesis	-	
M23	14	7	Multicellular organism metabolic process	R4	
		6	Extracellular structure organization	R4	
		-	-	-	

(*continued on next page*)

Table 2 (*continued*)

Discovery module	Number of genes	GOBP branch	GOBP term	Replication module	AD genes
M24	13	7	Glycerolipid metabolic process	-	
		7	Lipid modification	-	
		5	Phospholipid metabolic process	-	
M25	13	1	Peptidyl-serine modification	R15, R19, R25	
		-	-	-	
		-	-	-	
M26	12	1	Protein ubiquitination involved in ubiquitin-dependent protein catabolic process	-	
		1	Protein polyubiquitination	-	
		7	Amine metabolic process	R20	
M27	12	9	Neuron projection guidance	R34	
		-	-	-	
		-	-	-	
M28	11	5	Organophosphate catabolic process	-	
		1	Carbohydrate derivative catabolic process	R13	
		5	Aromatic compound catabolic process	-	
M29	11	-	-	-	
		-	-	-	
		-	-	-	
M30	10	5	Pyruvate metabolic process	R32	
		7	Small molecule catabolic process	-	
		7	Carbohydrate catabolic process	-	
M31	10	-	-	-	
		-	-	-	
		-	-	-	
M32	10	3	Mitochondrial transport	R8	
		5	Nucleoside monophosphate metabolic process	R13	
		3	Hydrogen transport	R18	
M33	10	6	Chromatin remodeling	R3	
		9	Protein acylation	R3	
		-	-	-	

Per module, the number of genes is shown. For each term, the name and respective GOBP branch is shown. The column "replication module" indicates which replication module also had this GOBP term. The last column indicates interaction of discovery module with known AD genes.
Key: GOBP, gene ontology biological process; AD, Alzheimer's disease.

2.6. Mapping known AD genes

We selected a list of 27 known AD risk genes, compiled from known AD GWAS loci and Mendelian causal genes (Lambert et al., 2013; Van Cauwenberghe et al., 2016). All experimental and database interactions between these 27 AD genes and the genes in discovery modules were extracted from STRING, using a cutoff of ≥500. An AD gene was considered to interact with a discovery module when it interacted with at least two of the genes in that module.

3. Results

3.1. Study sample characteristics

The demographic data of the AD group did not differ from the control group, as shown in Table 1. As expected, mean brain weight, Braak and CERAD stages and postmortem delay differed significantly between AD cases and controls. On average, 48,772,000 reads were sequenced per sample. All sequencing quality and alignment QC metrics were similar between groups. Two outliers were identified by PCs, driven by high expression of *TTR*. This gene is specifically expressed in the choroid plexus, which was confirmed using routine staining and both cases were excluded. The replication data set GSE95587 consisted of fusiform gyrus from 84 AD cases and 33 controls and is described elsewhere (van der Brug et al., 2017).

3.2. Differentially expressed genes and modules

A total of 2716 genes was differentially expressed in the discovery data set (DE score ≥0.1), as shown in Fig. 1. Examination of

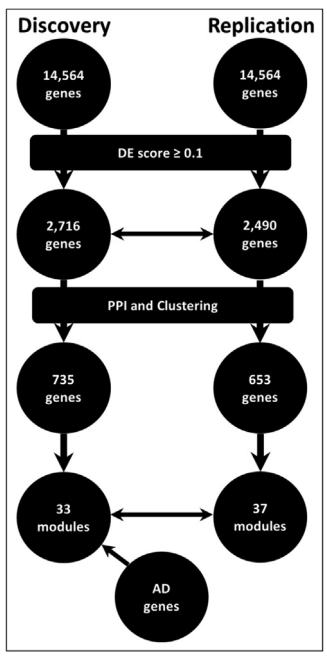

Fig. 1. Flowchart of data analysis. Discovery and replication data set are analyzed and differentially expressed genes are determined. An interaction network is constructed for each data set, which is then clustered in gene modules. These modules are compared directly on overlapping genes and on enriched gene ontology biological processes. Interactions of modules identified in discovery with known AD genes are also investigated. Abbreviations: AD, Alzheimer's disease.

known interactions between these DE genes showed that 1610 DE genes shared one or more interaction(s). Using this interaction network, we clustered 735 discovery DE genes into 33 discovery gene modules. The expression table and gene-module assignments can be found in supplemental Table 1. In the replication data set, 2490 DE genes were identified. A total of 1311 DE genes from the discovery data set (48%) replicated in the replication data set, as shown in Fig. 2. From the interaction network of replication DE genes, 653 DE genes were clustered into 37 replication modules.

3.3. Functional annotation and replication of modules

Gene set enrichment analysis of each module resulted in three significantly enriched gene ontology biological processes per module in discovery and replication. These enriched GOBP terms were pooled across all discovery and replication modules and assigned to eight main GOBP branches: "Organic substance metabolic process", "Signal Transduction", "Transport", "Regulation of biological process", "Cellular metabolic process", "Cellular component organization", "Other metabolic processes," and

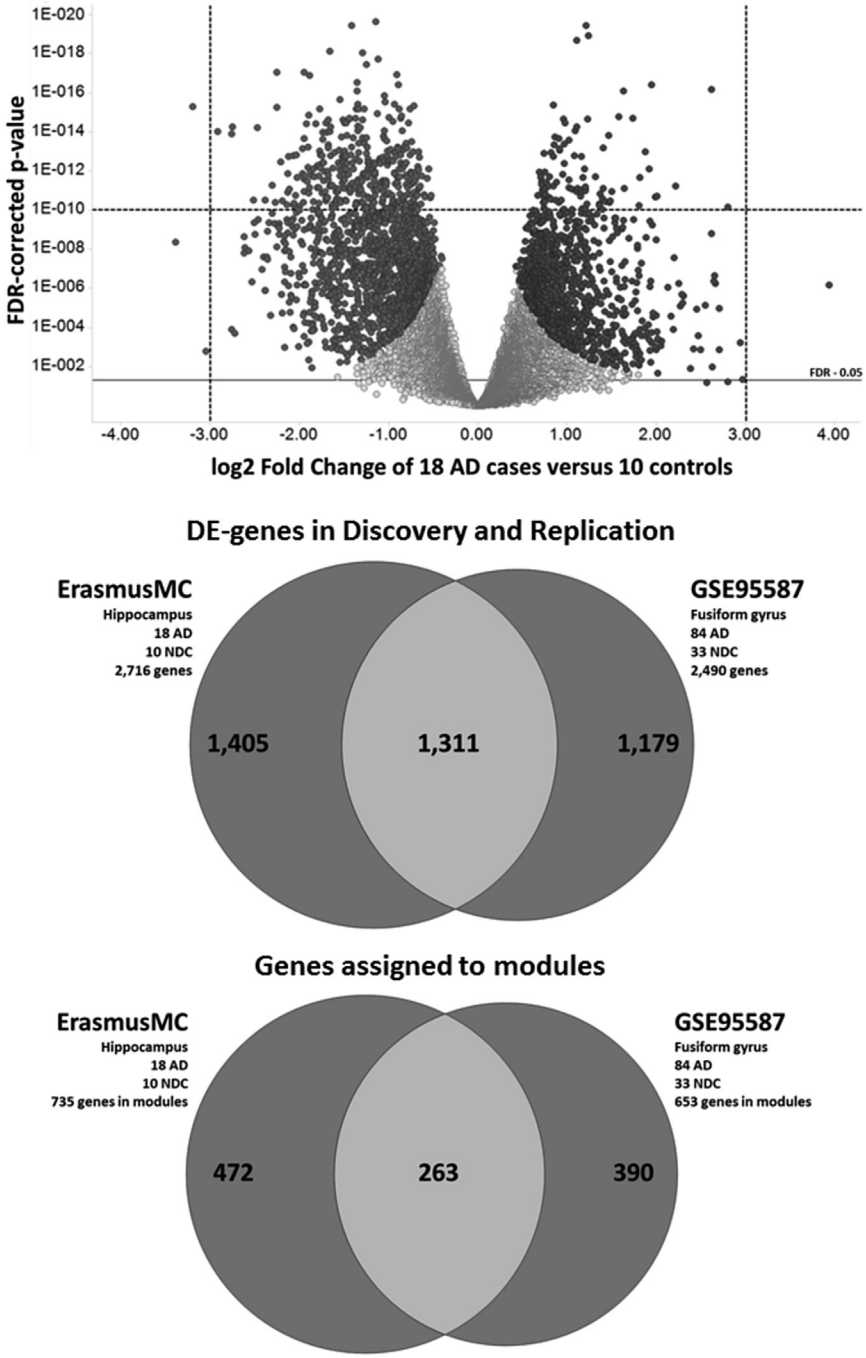

Fig. 2. Volcano plot of 14,564 analyzed protein-coding genes. Each dot is a gene, those dark-gray pass the 0.1 DE score threshold. Upper score limits (set to maximum of 1) are displayed by dotted lines. The solid line displays the default FDR-corrected ≥0.05 threshold. The Venn diagram displays the number of overlapping DE genes between the discovery and replication cohorts. Abbreviations: DE, differentially expressed.

"Response to stimulus". The remaining terms are grouped under a 9th branch: "Other biological processes". Table 2 shows the three GOBP terms for all discovery modules, their respective branches and category of overlap with the replication modules. Further details about these branches and overlap can be found in supplementary Fig. 2.

Combined across all 33 differentially expressed gene modules in the discovery data set, we identified 84 GOBP terms (at maximum three per gene module, see Table 2). For 19 of the 33 discovery modules, the discovery module overlaps both in genes and GOBP term with a replication module (overlap category 1), as shown in Fig. 3. Another eight gene modules overlap a GOBP term with a replication module, but do not overlap in genes (overlap category 2). Five modules overlapped in genes but did not overlap in GOBP term with the same replication module (overlap category 3). A single module did not overlap in either genes or GOBP term with the replication modules. This result brings the replication results of gene modules with the replication data set at 73% when based on overlapping genes (category 1 and 3) compared with 82% based on overlapping GOBP term(s) (category 1 and 2).

3.4. Interaction with AD genes

Of 27 known AD risk genes, 25 were expressed in the brain tissue that was studied. Three genes (11%) showed a DE score of ≥0.1: *CD2AP* (score 0.18), *MEF2C* (−0.29), and *PTK2B* (−0.50); none of these were assigned to a module. Only *MEF2C* and *PTK2B* are replicated with a DE score of −0.39 and −0.13, respectively. Ten AD genes interacted at least twice with a discovery module: *ABCA7*, *APP*, *BIN1*, *CELF1*, *CLU*, *HLA-DRB1*, *HLA-DRB5*, *MAPT*, *PICALM*, and *PTK2B*, as shown in Table 2. Six AD genes interacted only once with a discovery module: *APOE*, *CD2AP*, *INPP5D*, *MEF2C*, *PSEN1*, and *PSEN2*. Nine AD genes did not interact with any discovery module: *CASS4*, *CD33*, *CR1*, *FERMT2*, *MS4A6A*, *RIN3*, *SLC24A4*, *SORL1*, and *ZCWPW1*.

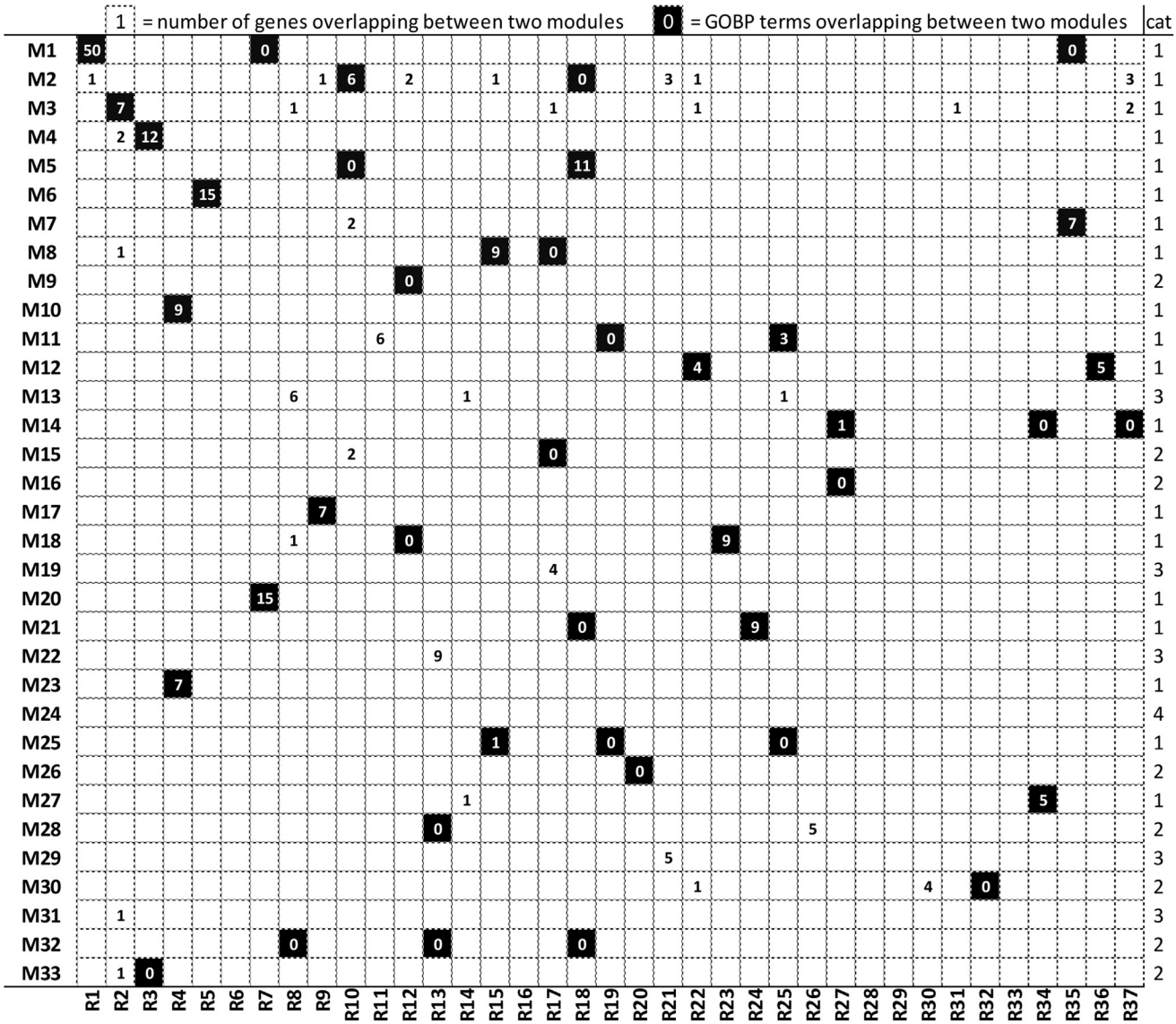

Fig. 3. Overlap between discovery and replication modules. Each module of discovery is shown horizontally, and the replication modules are shown vertically. The numbers shown indicate the overlapping number of genes between two modules. Intersections marked in black indicate modules that share a gene ontology biological process. The last column indicates the category of overlap for each discovery module (1: overlap in genes and GOBP term, 2: overlap in GOBP term, but not in genes, 3: overlap in genes, but not in GOBP term). Abbreviations: GOBP, gene ontology biological process.

4. Discussion

Our study identified 2716 differentially expressed genes (DE genes) in hippocampus of 18 AD cases compared with 10 age- and sex-matched nondemented controls. Of these 2716 DE genes, 735 were clustered in 33 gene modules based on PPI data. These 33 gene modules were assigned 84 gene ontology biological processes (GOBP terms, at maximum three for each gene module) which together comprise nine main GOBP branches. All nine branches were frequently observed in previous AD studies.

4.1. Replication by gene modules and GOBP terms is more robust and identifies the most central AD changes

Replication of our results in an independent data set (GSE95587, fusiform gyrus of 84 AD cases and 33 controls, [van der Brug et al., 2017]) was based on different categories of overlap, reflecting the robustness of these overlapping processes in the underlying pathophysiology of AD. The finding that most of our gene modules falls into category 1 ($n = 19$) indicates that the combined approach of GOBP annotated and PPI clustered gene modules identifies the most robust changes in AD gene expression. The gene modules in category 2 ($n = 8$) and category 3 ($n = 5$) might reflect some variability of gene expression between hippocampus in our study and fusiform gyrus of AD brains in the replication study.

The present comparative study supports the idea that the overlapping data sets based on gene modules or GOBP term per module is more robust than based on overlapping genes only, as the overlap of all DE genes (48%) can be improved by categorization into gene modules (72%), and even more by overlap based on GOBP terms (82%).

4.2. GOBP branches represent common AD pathways

The nine main GOBP branches are previously observed in literature of AD expression studies (Chi et al., 2016; C.E. Humphries et al., 2015; Liang et al., 2008; Ray and Zhang, 2010; Sekar et al., 2015; Twine et al., 2011; Wang et al., 2016). These GOBP branches can be found in detail, containing all module annotations in supplemental Table 2.

GOBP branch 1, named "organic substance metabolic process", consists of metabolic processes such as DNA replication and repair and RNA translation and post-translational modifications. These metabolic processes underlie many other biological processes and are dysregulated in AD cases as a response to the various disease-related changes in the AD hippocampus (C.E. Humphries et al., 2015; Liang et al., 2008; Sekar et al., 2015; Twine et al., 2011; Wang et al., 2016). The second GOBP branch, called "signal transduction", consists of six gene modules that represent the same distinct neurotransmitter signaling pathways in both the discovery and replication data sets (all six gene module are in overlap category 1). These results indicate a broad dysfunction of synaptic transmission in the AD brain. These are likely the result of neuronal degeneration in AD hippocampus and are often found dysregulated in AD literature (Chi et al., 2016; C.E. Humphries et al., 2015; Liang et al., 2008; Ray and Zhang, 2010; Sekar et al., 2015; Wang et al., 2016). GOBP branch 3, enriched for "transport", mostly represents ion transport GOBP terms, as shown in supplemental Fig. 2. Many modules in this GOBP branch are involved in energy production, which is often described as dysfunctional in previous AD studies (C.E. Humphries et al., 2015; Wang et al., 2016). These results are likely caused by neuronal degradation, and thus reduced energy consumption, although activation of glial cells might also influence this process (Sekar et al., 2015). GOBP branch 4 "regulation of biological processes" is largely complementary to the other GOBP branches. It contains modules annotated to both an executive biological process, for example, "transmembrane transport" and its regulative process, "regulation of transmembrane transport". Six of its seven gene modules fall into overlap category 1, indicating a robust dysfunction in this GOBP branch.

These first four GOBP branches are the largest and therefore underlie changes in AD pathophysiology that stand out the most in our study. Of the 33 identified gene modules in our study, 23 are involved in these four GOBP branches. With 17 gene modules in overlap category 1, this indicates that these four GOBP branches are among the most robust changes in AD pathophysiology. Given the functions of these central GOBP branches, we conclude that organic substance metabolic processes, neurotransmitter signaling, energy transport, and regulation of biological processes are main dysfunctional pathways in AD pathophysiology.

Of the remaining GOBP branch 5 (including RNA splicing and dephosphorylation), branch 6 (incl axon development), and branch 7 (other metabolic processes), gene modules overlap mostly on in category 2 and category 3 with the replication data set. These three GOBP branches do not contain any unique gene modules and are likely not as robustly involved in AD as the other GOBP branches. GOBP branch 8, "response to stimulus," is the smallest GOBP branch, indicating a response to neurodegeneration resulting in inflammation and glial cell activation which has often been observed in previous studies (C.E. Humphries et al., 2015; Sekar et al., 2015; Wang et al., 2016). All three gene modules in GOBP branch 8 overlapped in category 1 with the replication data set, suggesting that this small GOBP branch represents a robust change to AD pathophysiology. The biological processes of GOBP branch 9, including "neuromuscular process", "actin-filament based movement," and "neuron projection guidance might also be robust changes in AD, but are represented by only a small number of gene modules in our data, possibly because of the late stage of the disease in our samples.

4.3. Interactions with AD genes

Of 27 AD genetic risk factor genes, only three were differentially expressed in our data set, and two replicated (*MEF2C* and *PTK2B*). Several discovery modules interacted with these AD genes, suggesting a degree of overlap in biological function. *HLA-DRB1*, *HLA-DRB5*, *BIN1*, and *PICALM* interact with M9 and might be involved in endocytosis and/or microtubule-based movement (Baig et al., 2010; Zhou et al., 2014). *ABCA7* and *MAPT* interact with a gene module involved in ion transport and signaling (M2). *APP* interacts with M1 and is involved in signal transduction (Cheng et al., 2014; Cirrito et al., 2008). *PTK2B* is differentially expressed in both discovery and replication and interacts with modules involved in receptor signaling and protein modification (M7, M10, and M25) (Beecham et al., 2014; Han et al., 2017). *CELF1* interacts with genes involved in RNA processing and protein modification (M8) and *CLU* interacts genes involved in exocytosis and actin-based filament organization (M14). These interactions suggest roles of these genes also in later stages of AD and do not represent the typical associations of these genes in a causal inference (Lambert et al., 2013; Van Cauwenberghe et al., 2016).

4.4. Limitations of this study

This study holds several limitations. First, PPI networks comprise existing databases, which generate bias to well-known genes and biological processes (Gillis et al., 2014; Schaefer et al., 2015; von Mering et al., 2003). Indeed, of the 2716 DE genes identified in discovery, only 1610 held an interaction in the STRING

database, and some relevant genes might have been excluded as a result.

An important issue in using PPI data for your network analysis is that there are no clear guidelines on what to use for the interaction score cutoff, Markov clustering inflation factor threshold, or on the proper functional annotation of modules. Nevertheless, some consensus is emerging and these most commonly used parameters were also applied in this study. These parameters are (1) prioritizing or limiting to experimental interactions types, or not using text-mining based types because this minimizes bias of the results (Szklarczyk et al., 2017; von Mering et al., 2003); (2) optimizing the MCL inflation factor to generate modules of 10–100 genes (Subramanian et al., 2005; van Dongen and Abreu-Goodger, 2012); 3. replication, preferably on a functional annotation level as gene ontology (Ashburner et al., 2000; Gene Ontology, 2015).

To optimize clustering of the gene modules, additional metrics of the generated PPI network could be included, for example, the direction of effect, or weighting PPI interactions. This study was designed as a cross-sectional case-control analysis, and many of the observed differences might be caused by neurodegeneration. Our sample size of 18 cases and 10 controls is not optimal to robustly detect all deviations in AD, and some genes/GOBP terms might have been missed.

4.5. Conclusions

Our method provides a comprehensive and complete overview of dysregulation based on GOBPs in AD. We show that the PPI and MCL clustering approach identifies functional gene modules which replicate in other data sets. Where individual genes might differ between studies, overall GOBP terms are preserved and can be identified in this manner. Replication based on gene module GOBP terms was more robust than based on individual genes (82% vs 48%).

Disclosure statement

This work was supported by the Joint Programming Initiative Neurodegenerative Diseases (JPND) and ZonMW (grant number [733051022]).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.neurobiolaging.2018.10.023.

References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 25–29.

Baig, S., Joseph, S.A., Tayler, H., Abraham, R., Owen, M.J., Williams, J., Kehoe, P.G., Love, S., 2010. Distribution and expression of picalm in Alzheimer disease. J. Neuropathol. Exp. Neurol. 69, 1071–1077.

Beecham, G.W., Hamilton, K., Naj, A.C., Martin, E.R., Huentelman, M., Myers, A.J., Corneveaux, J.J., Hardy, J., Vonsattel, J.P., Younkin, S.G., Bennett, D.A., De Jager, P.L., Larson, E.B., Crane, P.K., Kamboh, M.I., Kofler, J.K., Mash, D.C., Duque, L., Gilbert, J.R., Gwirtsman, H., Buxbaum, J.D., Kramer, P., Dickson, D.W., Farrer, L.A., Frosch, M.P., Ghetti, B., Haines, J.L., Hyman, B.T., Kukull, W.A., Mayeux, R.P., Pericak-Vance, M.A., Schneider, J.A., Trojanowski, J.Q., Reiman, E.M., Alzheimer's Disease Genetics, C., Schellenberg, G.D., Montine, T.J., 2014. Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. Plos Genet. 10, e1004606.

Bekris, L.M., Yu, C.E., Bird, T.D., Tsuang, D.W., 2010. Genetics of Alzheimer disease. J. Geriatr. Psychiatry Neurol. 23, 213–227.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Boyle, P.A., Wilson, R.S., Yu, L., Barr, A.M., Honer, W.G., Schneider, J.A., Bennett, D.A., 2013. Much of late life cognitive decline is not due to common neurodegenerative pathologies. Ann. Neurol. 74, 478–489.

Braak, H., Braak, E., 1995. Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol. Aging 16, 271–278 discussion 8-84.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Ami, G.O.H., Web Presence Working, G., 2009. AmiGO: online access to ontology and annotation data. Bioinformatics 25, 288–289.

Cheng, X., Wu, J., Geng, M., Xiong, J., 2014. Role of synaptic activity in the regulation of amyloid beta levels in Alzheimer's disease. Neurobiol. Aging 35, 1217–1232.

Chi, L.M., Wang, X., Nan, G.X., 2016. In silico analyses for molecular genetic mechanism and candidate genes in patients with Alzheimer's disease. Acta Neurol. Belg. 116, 543–547.

Cirrito, J.R., Kang, J.E., Lee, J., Stewart, F.R., Verges, D.K., Silverio, L.M., Bu, G., Mennerick, S., Holtzman, D.M., 2008. Endocytosis is required for synaptic activity-dependent release of amyloid-beta in vivo. Neuron 58, 42–51.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Enright, A.J., Van Dongen, S., Ouzounis, C.A., 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584.

Gene Ontology, C., 2001. Creating the gene ontology resource: design and implementation. Genome Res. 11, 1425–1433.

Gene Ontology, C., 2015. Gene ontology consortium: going forward. Nucleic Acids Res. 43, D1049–D1056.

Gillis, J., Ballouz, S., Pavlidis, P., 2014. Bias tradeoffs in the creation and analysis of protein-protein interaction networks. J. Proteomics 100, 44–54.

Han, Z., Huang, H., Gao, Y., Huang, Q., 2017. Functional annotation of Alzheimer's disease associated loci revealed by GWASs. PLoS One 12, e0179677.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., Steward, C., Harte, R., Lin, M., Howald, C., Tanzer, A., Derrien, T., Chrast, J., Walters, N., Balasubramanian, S., Pei, B., Tress, M., Rodriguez, J.M., Ezkurdia, I., van Baren, J., Brent, M., Haussler, D., Kellis, M., Valencia, A., Reymond, A., Gerstein, M., Guigo, R., Hubbard, T.J., 2012. GENCODE: the reference human genome annotation for the ENCODE Project. Genome Res. 22, 1760–1774.

Holtzman, D.M., Carrillo, M.C., Hendrix, J.A., Bain, L.J., Catafau, A.M., Gault, L.M., Goedert, M., Mandelkow, E., Mandelkow, E.M., Miller, D.S., Ostrowitzki, S., Polydoro, M., Smith, S., Wittmann, M., Hutton, M., 2016. Tau: from research to clinical development. Alzheimers Dement 12, 1033–1039.

Humphries, C., Kohli, M.A., Whitehead, P., Mash, D.C., Pericak-Vance, M.A., Gilbert, J., 2015a. Alzheimer disease (AD) specific transcription, DNA methylation and splicing in twenty AD associated loci. Mol. Cell Neurosci 67, 37–45.

Humphries, C.E., Kohli, M.A., Nathanson, L., Whitehead, P., Beecham, G., Martin, E., Mash, D.C., Pericak-Vance, M.A., Gilbert, J., 2015b. Integrated whole transcriptome and DNA methylation analysis identifies gene networks specific to late-onset Alzheimer's disease. J. Alzheimers Dis. 44, 977–987.

Jellinger, K.A., 2008. Neuropathological aspects of Alzheimer disease, Parkinson disease and frontotemporal dementia. Neurodegener Dis. 5, 118–121.

Karran, E., Mercken, M., De Strooper, B., 2011. The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. Nat. Rev. Drug Discov. 10, 698–712.

Kavanagh, T., Mills, J.D., Kim, W.S., Halliday, G.M., Janitz, M., 2013. Pathway analysis of the human brain transcriptome in disease. J. Mol. Neurosci. 51, 28–36.

Kong, W., Mou, X., Zhang, N., Zeng, W., Li, S., Yang, Y., 2015. The construction of common and specific significance subnetworks of Alzheimer's disease from multiple brain regions. Biomed. Res. Int. 2015, 394260.

Kong, W., Zhang, J., Mou, X., Yang, Y., 2014. Integrating gene expression and protein interaction data for signaling pathway prediction of Alzheimer's disease. Comput. Math. Methods Med. 2014, 340758.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., Russo, G., Thorton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C., Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M.L., Ruiz, A., Bihoreau, M.T., Choi, S.H., Reitz, C., Pasquier, F., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O.L., De Jager, P.L., Deramecourt, V., Johnston, J.A., Evans, D., Lovestone, S., Letenneur, L., Moron, F.J., Rubinsztein, D.C., Eiriksdottir, G., Sleegers, K., Goate, A.M., Fievet, N., Huentelman, M.W., Gill, M., Brown, K., Kamboh, M.I., Keller, L., Barberger-Gateau, P., McGuiness, B., Larson, E.B., Green, R., Myers, A.J., Dufouil, C., Todd, S., Wallon, D., Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J., Lleo, A., Bayer, A., Tsuang, D.W., Yu, L., Tsolaki, M., Bossu, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S., Sanchez-Garcia, F., Fox, N.C., Hardy, J., Deniz Naranjo, M.C., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., European Alzheimer's Disease, I., Genetic, Environmental Risk in Alzheimer's, D., Alzheimer's Disease Genetic, C., Cohorts for, H., Aging Research in Genomic, E., Moebus, S., Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., Nacmias, B., Gilbert, J.R., Mayhaus, M., Lannefelt, L., Hakonarson, H., Pichler, S., Carrasquillo, M.M., Ingelsson, M., Beekly, D., Alvarez, V., Zou, F., Valladares, O., Younkin, S.G., Coto, E., Hamilton-Nelson, K.L., Gu, W., Razquin, C., Pastor, P., Mateo, I., Owen, M.J., Faber, K.M., Jonsson, P.V., Combarros, O., O'Donovan, M.C., Cantwell, L.B., Soininen, H., Blacker, D., Mead, S., Mosley Jr., T.H., Bennett, D.A., Harris, T.B., Fratiglioni, L., Holmes, C., de Bruijn, R.F., Passmore, P., Montine, T.J.,

Bettens, K., Rotter, J.I., Brice, A., Morgan, K., Foroud, T.M., Kukull, W.A., Hannequin, D., Powell, J.F., Nalls, M.A., Ritchie, K., Lunetta, K.L., Kauwe, J.S., Boerwinkle, E., Riemenschneider, M., Boada, M., Hiltuenen, M., Martin, E.R., Schmidt, R., Rujescu, D., Wang, L.S., Dartigues, J.F., Mayeux, R., Tzourio, C., Hofman, A., Nothen, M.M., Graff, C., Psaty, B.M., Jones, L., Haines, J.L., Holmans, P.A., Lathrop, M., Pericak-Vance, M.A., Launer, L.J., Farrer, L.A., van Duijn, C.M., Van Broeckhoven, C., Moskvina, V., Seshadri, S., Williams, J., Schellenberg, G.D., Amouyel, P., 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. 45, 1452–1458.

Liang, W.S., Dunckley, T., Beach, T.G., Grover, A., Mastroeni, D., Ramsey, K., Caselli, R.J., Kukull, W.A., McKeel, D., Morris, J.C., Hulette, C.M., Schmechel, D., Reiman, E.M., Rogers, J., Stephan, D.A., 2008. Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. Physiol. Genomics 33, 240–256.

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.

McCarthy, D.J., Chen, Y., Smyth, G.K., 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40, 4288–4297.

Mirra, S.S., Heyman, A., McKeel, D., Sumi, S.M., Crain, B.J., Brownlee, L.M., Vogel, F.S., Hughes, J.P., van Belle, G., Berg, L., 1991. The consortium to establish a registry for Alzheimer's disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 41, 479–486.

Morris, J.H., Apeltsin, L., Newman, A.M., Baumbach, J., Wittkop, T., Su, G., Bader, G.D., Ferrin, T.E., 2011. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics 12, 436.

Murray, M.E., Graff-Radford, N.R., Ross, O.A., Petersen, R.C., Duara, R., Dickson, D.W., 2011. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. Lancet Neurol. 10, 785–796.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M., 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 27, 29–34.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., Ferri, C.P., 2013. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 9, 63–75.e2.

Ray, M., Zhang, W., 2010. Analysis of Alzheimer's disease severity across brain regions by topological analysis of gene co-expression networks. BMC Syst. Biol. 4, 136.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Schaefer, M.H., Serrano, L., Andrade-Navarro, M.A., 2015. Correcting for the study bias associated with protein-protein interaction measurements reveals differences between protein degree distributions from different cancer types. Front Genet. 6, 260.

Scheltens, P., Blennow, K., Breteler, M.M., de Strooper, B., Frisoni, G.B., Salloway, S., Van der Flier, W.M., 2016. Alzheimer's disease. Lancet 388, 505–517.

Sekar, S., McDonald, J., Cuyugan, L., Aldrich, J., Kurdoglu, A., Adkins, J., Serrano, G., Beach, T.G., Craig, D.W., Valla, J., Reiman, E.M., Liang, W.S., 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.

Selkoe, D.J., Hardy, J., 2016. The amyloid hypothesis of Alzheimer's disease at 25 years. EMBO Mol. Med. 8, 595–608.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Ideker, T., 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27, 431–432.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. 102, 15545–15550.

Sutherland, G.T., Janitz, M., Kril, J.J., 2011. Understanding the pathogenesis of Alzheimer's disease: will RNA-Seq realize the promise of transcriptomics? J. Neurochem. 116, 937–946.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., Jensen, L.J., von Mering, C., 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 45, D362–D368.

Thal, D.R., Attems, J., Ewers, M., 2014. Spreading of amyloid, tau, and microvascular pathology in Alzheimer's disease: findings from neuropathological and neuroimaging studies. J. Alzheimers Dis. 42 (Suppl 4), S421–S429.

Tomiyama, T., 2010. Involvement of beta-amyloid in the etiology of Alzheimer's disease. Brain Nerve 62, 691–699.

Twine, N.A., Janitz, K., Wilkins, M.R., Janitz, M., 2011. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. PLoS One 6, e16266.

Van Cauwenberghe, C., Van Broeckhoven, C., Sleegers, K., 2016. The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet. Med. 18, 421–430.

van der Brug, H., Huntly, M., Cao, Y., 2017. Heterogeneity in neurodegenerative disease (GSE95587). NCBI GEO.

van Dongen, S., Abreu-Goodger, C., 2012. Using MCL to extract clusters from networks. Methods Mol. Biol. 804, 281–295.

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., Snel, B., 2003. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 31, 258–261.

Wang, M., Roussos, P., McKenzie, A., Zhou, X., Kajiwara, Y., Brennand, K.J., De Luca, G.C., Crary, J.F., Casaccia, P., Buxbaum, J.D., 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, 104.

Zhou, Y., Hayashi, I., Wong, J., Tugusheva, K., Renger, J.J., Zerbinatti, C., 2014. Intracellular clusterin interacts with brain isoforms of the bridging integrator 1 and with the microtubule-associated protein Tau in Alzheimer's disease. PLoS One 9, e103187.