An integrated approach to identify sex-specific genes, transcription factors, and pathways in Alzheimer's disease

Adolfo López-Cerdán[1,2†], Zoraida Andreu[3†], Marta R. Hidalgo[1], Antonio Porlan[1], Macarena Pozo-Morales[1], María de la Iglesia-Vayá[2], Akiko Mikozami[3], Franca R. Guerini[4], Francisco García-García[1]

[1] Computational Biomedicine Laboratory, Principe Felipe Research Center (CIPF), 46012, Valencia, Spain

[2] Biomedical Imaging Unit FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana, 46012, Valencia, Spain

[3] Foundation Valencian Institute of Oncology (FIVO), 46009, Valencia, Spain

[4] Oral Health/Brain Health/Total health (OBT) Research Center, Faculty of Dental Science, Kyushu University, Fukuoka, Japan.

[5] IRCCS Fondazione Don Carlo Gnocchi ONLUS, 20148 Milan, Italy.

ABSTRACT

Background

Age is one of the bigger risk factors for developing Alzheimer's disease (AD). Still, in the last decade, several experimental studies and reviews document the role of sex in several features of AD. Understanding the impact of sex in specific molecular mechanisms in this disease is critical for tailored interventions.

Methods

To explore sex differences impact disease (SDID) in AD, we conduct a systematic review during 2021 to select transcriptomic studies of AD with sex information data in public repositories in the 2002-2021 period. We focus on the main brain regions affected by AD; cortex (CT) and hippocampus (HP). Then, a differential expression analysis for each study and two tissue-specific meta-analyses were performed. Last, focusing on the cortex, where significant SDID were detected, we conducted a

functional characterization: protein-protein network interaction (PPI), Over-Representation Analysis (ORA) to explore biological processes and pathways, and a VIPER analysis to estimate transcription factor activity.

Results

A group of eight studies in CT and five in HP were selected from Gene Expression Omnibus (GEO) repository, for tissue-specific meta-analyses. We detected 389 significant SDID genes in CT. In general, females show more affected genes than males and the genes were grouped in six subsets according to the expression profile in females and males with AD. Only subset I (repressed genes in affected females) displayed significant results at functional profiling. Females.AD compared to males. AD showed a more significant impairment in biological processes related to the regulation and organization of synapsis, the pathways more impaired those linked to neurotransmitters (Glutamate and GABA) and several pathways related to protein folding, Aβ aggregation, and accumulation. This could explain, in part, why the cognitive decline is more pronounced in females.AD. Finally, we detected 23 TFs having different activation patterns by sex, some of them associated with AD for the first time. All results generated during this study are readily available through an open web resource.

Conclusion

Our meta-analyses indicate differences in several mechanisms involved in AD between females and males. Such sex differences represent the base of new hypotheses and could importantly impact precision medicine, and improve diagnosis and clinical outcomes in AD for females and males.

Keywords: Alzheimer's disease; Cognitive function; Meta-analysis; Sex difference impact disease, Transcriptomics.

BACKGROUND

Alzheimer's disease (AD) is a neurodegenerative disease and the most common form of dementia in the elderly population. The latest update of WHO (2023) indicates that more than 55 million persons suffer from dementia worldwide and AD represents 60-70% of the cases. The incidence of AD is rising fast, (about 10 million new cases each year), and over 115 million AD cases are estimated by 2050 [Weller and Budson, 2018; Lane et al., 2018; Mohamed et al., 2019]. Nowadays, AD is incurable and terminal, being one of the leading causes of dependency, disability, and mortality [Lane et al., 2018; Swarbrick et al., 2019].

In people with AD, the brain suffers significant neuronal death. The neuropathology is mainly related to the accumulation of beta-amyloid protein plaques and neurofibrillary tangles of Tau protein [PMID: 18371174; PMID: 22265587]. The excess of plaques and tangles generate toxicity processes that disrupt the vital actions that keep neurons healthy: communication, metabolism, and repair [PMID: 33809300]. The limbic system is the first to be attacked, particularly the hippocampus area which is key for the formation of new memories and the learning process. After, the cerebral cortex suffers a deterioration, resulting in the inability to control emotional outbursts and daily tasks. Finally, the brainstem is the last region to be damaged in advanced stages and causes organ failure. Several subtypes of AD based on their progress and symptoms by the trajectory of tau deposition have been identified as described in PMID: 33927414. The disease is clinically characterized mainly by memory loss but presents other symptoms, both cognitive and behavior-related.

The causes of AD still are unknown, genetic factors seem to determine early onset AD (EOAD). While the interaction among risk and environmental factors is responsible for the late or sporadic onset (LOAD) encompassing 95% of cases [PMID: 31556570]. Several studies point to the female sex as an essential risk factor in neurodegenerative diseases and to develop late or sporadic onset AD [PMID: 9343467; PMID: 34623401; PMID: 30298180]. About two-thirds of patients with AD are female PMID: 32157811 and sex-related differences are evidenced in patterns of disease manifestation, in the rates of cognitive decline and brain atrophy, suggesting that sex is a crucial variable in disease heterogeneity [PMID: 26735615; PMID: 29189302; PMID: 29324460; PMID: 21606896; PMID: 29985474; PMID: 21356315; PMID: 32280740]. Nevertheless, knowledge about the molecular mechanisms for sex-biased in AD is diffuse. In this scenario, multi-omics analysis and datasets from human AD samples and animal models offer an excellent work platform to study molecular changes and pathways related to sex differences in AD. The inclusion of the sex variable in AD research is a need to improve precision medicine strategies and a faster advance in diagnosis and treatment.

To investigate sex differences in this devastating disease, the present study provides two independent meta-analyses taking into account sex information from human patients with AD. One meta-analysis has been performed in the cortex (CT, 8 studies) and the other one in the hippocampus (HP, 5 studies), two of the main brain regions affected by AD pathogenesis. Our in silico approach has

revealed differential gene subsets among females and males with AD. At a functional level, these clusters are involved in relevant functions related to neurotransmission regulation, synapses and pre-synapses, calcium signaling, pH reduction, MAPK activity, ubiquitination, and protein folding, mainly. Additionally, females.AD compared to males.AD present significantly increased several transcription factors with different activity patterns: ADNP, HMGN3, IRF3, KLF5, KLF9, MAZ, MBD3, MYNN, PRDM14, SIX5, and ZNF207 (activation pattern) and GTF2B, HOXB13, NANOG, NME2, PCGF2, SNAI2, ZBTB7A, ZC3H8, ZHX1, and ZHX2 (repression pattern). Two TFs, CEBPZ and TERFS displayed divergent patterns by sex. The results are freely available in the Metafun-AD web tool and may be the starting point of future studies.

METHODS

All bioinformatics and statistical analysis were performed using R software.

Study Search and Selection

Available datasets were collected from the Gene Expression Omnibus (GEO) [https://doi.org/10.1093/nar/gks1193] public repository. A systematic search of all published studies in public repositories (2002-2021) was conducted during 2021, following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [PMID: 19621072]. Keywords employed in the search were "Alzheimer", "Alzheimer's Disease", and "AD". We applied the following inclusion criteria:

- Transcriptomics studies on *Homo sapiens*
- Control and AD-affected subjects included
- Sex, disease status, age, and brain region variables registered
- RNA extracted directly from post-mortem brain tissues (no cell lines or cultures)
- Brain tissues from either CT or HP

- Sample size > 3 for case and control groups in both sexes

Finally, we retrieve normalized gene expression data of 8 microarray AD datasets (GSE118553, GSE1297, GSE132903, GSE15222, GSE29378, GSE37263, GSE48350, GSE5281, and GSE84422) and the raw counts' matrix of the GSE125583 RNAseq study from the GEO repository. The download was performed using the GEOquery R package [https://doi.org/10.1093/bioinformatics/btm254] for the microarray studies and manually for the RNAseq study.

Individual Transcriptomics Analysis

For each selected study, we perform an individual transcriptomics analysis comprising 2 steps: preprocessing and differential expression analysis.

Data preprocessing included the standardization of the terminology for the clinical variables in each study, the homogenization of gene annotation, and exploratory data analysis. For the microarray datasets, the normalization methods performed by the original authors were assessed, transforming data matrices by applying $\log_2$ when necessary. All probe sets were annotated to HUGO gene symbols [PMID: 33152070] using the annotation provided by each microarray platform. When dealing with duplicated probe-to-symbol mappings, the median of their expression values was calculated. For the RNAseq dataset, the count matrix was preprocessed using the edgeR R package [https://doi.org/10.12688/f1000research.8987.2] and transformed using the *Voom* function included in the limma R package [PMID: 25605792]. The exploratory analysis included unsupervised clustering and PCA to detect expression patterns between samples and genes and the presence of batch effects in each study.

Differential expression analyses were performed using the limma R package to detect sex-based differentially expressed genes. To achieve this goal, the following comparison was applied:

(AD.female - Control.female) - (AD.male - Control.male)

This comparison allows the detection of genes having a sex-differential impact on the disease (SDID). Genes with a Log_2 Fold Change (LFC) greater than 0 show either a higher increase or a lesser decrease in expression in females when comparing the effect of the disease between sexes. We will

refer to these genes as "Positive SDID" genes. On the contrary, genes with an LFC lower than 0 have a higher increase or a lesser decrease in expression in males when comparing the effect of the disease between sexes. We will refer to these genes as "Negative SDID" genes.

To get a better understanding of the sex-based differential behavior detected with the previous comparison, we performed two additional comparisons: A case vs control comparison performed only in females (AD.female - Control.female), that evaluates the impact of the disease in females (IDF); and another one performed only in males (AD.male - Control.male), that informs us about the impact of the disease in males (IDM).

These comparisons were applied separately to all samples from cortical tissues (CT) and hippocampus (HP). The patients' age was included in the limma linear model as a blocking factor to reduce its impact on the results. P-values were corrected using the Benjamini-Hochberg procedure and considered significant below a threshold of 0.05.

Gene Expression Meta-Analysis

Differential gene expression results were integrated into a single meta-analysis for each brain region (CT and HP). Meta-analyses were implemented with the R package metafor, under the DerSimonian & Laird random-effects model, considering individual study heterogeneity. This model considers the variability of individual studies by increasing the weights of studies with less variability when computing meta-analysis results. Thus, the most robust functions between studies are highlighted.

P-values, corrected p-values, LFC, LFC standard error (SE), and LFC 95% confidence intervals (CI) were calculated for each evaluated gene. Genes with corrected p-values lower than 0.05 were considered significant, and both funnel and forest plots were computed for each. These representations were evaluated to assess for possible biased results, where the LFC represents the effect size of a gene, and the SE of the LFC serves as a study precision measure []. Sensitivity analysis (leave-one-out cross-validation) was conducted for each significant gene to verify alterations in the results due to the inclusion of any study. The Open Targets platform (release 22.09) was used to explore the associations of the significant genes to AD.

Sex-based Functional Signature in the CT

Gene meta-analysis of CT data revealed sets of genes having a significantly differential expression pattern between male and female AD patients. Several analyses were carried out to identify the functional implications of these differences.

Over-Representation Analysis (ORA) through clusterProfiler and ReactomePA R packages was first used to determine the biological functions and pathways overrepresented in all significant gene subsets. P-values and corrected p-values were calculated for each GO (Gene Ontology) term from the “Biological Processes” GO and each Reactome pathway. Every function and pathway with a corrected p-value < 0.05 was labeled as over-represented in each gene set.

Protein-protein interaction (PPI) networks were then calculated using the STRING web tool for each subset of genes. The total number of edges was examined, and PPI enrichment was assessed using the default parameters for each network.

Finally, a VIPER analysis with human regulons obtained from the DoRothEA R package was performed to estimate transcription factor activity. Regulons with a confidence level of A, B, C, or D were selected, excluding those with less than 25 genes (n = 217). The p-values were corrected with the Benjamini & Hochberg method. Normalized enrichment scores (NES) were calculated by VIPER as a measure of relative transcription factor activity.

Metafun-AD Web Tool

All data and results generated in the different steps of the meta-analysis are available in the Metafun-AD web tool, which is freely accessible to any user and allows the confirmation of the results described in this manuscript and the exploration of other results of interest. The front end was developed using the Quarto, and the interactive graphics used in this web resource have been implemented with plotly.

This easy-to-use resource is divided into 7 sections: 1) summary of analysis results in each phase. Then, for each of the studies, the detailed results of the 2) exploratory analysis and 3) differential expression. 4) The gene meta-analysis results of the four different performed meta-analyses. The

user can interact with the web tool through graphics and tables and search for specific information for a gene or function. Finally, sections 5-7) provide the detailed tables and figures corresponding to the results of the 3 functional profiling methods (ORA, PPI, and VIPER).

RESULTS

In this study, we aimed to investigate the existence of sex-based differences in AD using a systematic review and two meta-analyses of transcriptomic studies. The studies included information on the sex of the patients and were obtained from the Gene Expression Omnibus (GEO) repository [23]. We conducted one meta-analysis for each of the two primary brain regions affected by AD pathogenesis - CT (8 studies) and HP (5 studies). Subsequently, we explored the biological implications of the CT meta-analysis results by utilizing three distinct functional profiling methods, including over-representation analysis (ORA), protein-protein interaction (PPI) network construction, and transcription factor activity analysis.

Systematic Review and Study Selection

In our search for studies on AD, we identified 76 non-duplicated entries, 40 of which (52%) included both male and female patients. After applying inclusion and exclusion criteria (Methods, Figure 2), we selected 14 studies for comparison, but 4 were excluded after exploratory analysis. Thus, we analyzed 10 studies comprising 2508 samples (909 controls and 1599 AD cases) from the CT and HP brain regions (Table 1). Figure 3 presents the sex distribution by study and region, with an overall proportion of 44% males and 56% females. The median age of the participants was 85. More information on the selected studies and their clinicopathological features is available in Table 1 and Figure 2.

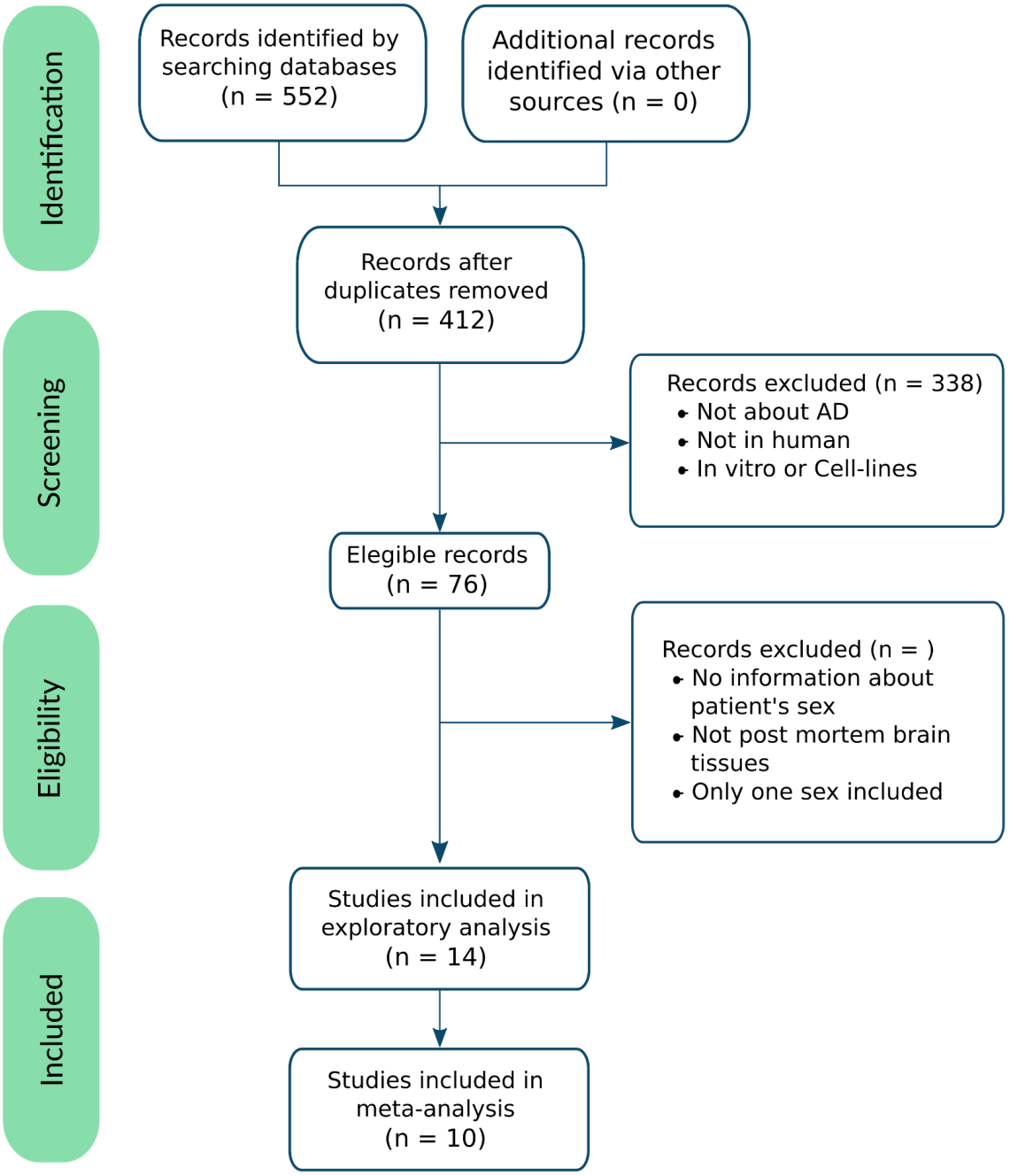

Figure 1. The flow of information through the distinct phases of the systematic review following PRISMA statement guidelines.

Table 1. Studies selected after the systematic review.

Study	Platform	Brain area	Publication
GSE118553	Illumina HumanHT-12 V4.0 expression beadchip	CT	
GSE125583	Illumina HiSeq 2500	CT	
GSE1297	Affymetrix Human Genome U133A Array	HP	
GSE132903	Illumina HumanHT-12 V4.0 expression beadchip	CT	
GSE15222	Sentrix HumanRef-8 Expression BeadChip	CT	
GSE29378	Illumina HumanHT-12 V3.0 expression beadchip	HP	
GSE37263	Affymetrix Human Exon 1.0 ST Array	CT	-
GSE48350	Affymetrix Human Genome U133 Plus 2.0 Array	CT, HP	
GSE5281	Affymetrix Human Genome U133 Plus 2.0 Array	CT, HP	
GSE84422	Affymetrix Human Genome U133A/U133B Arrays	CT, HP	

Brain areas: Cortex (CT) and Hippocampus (HP)

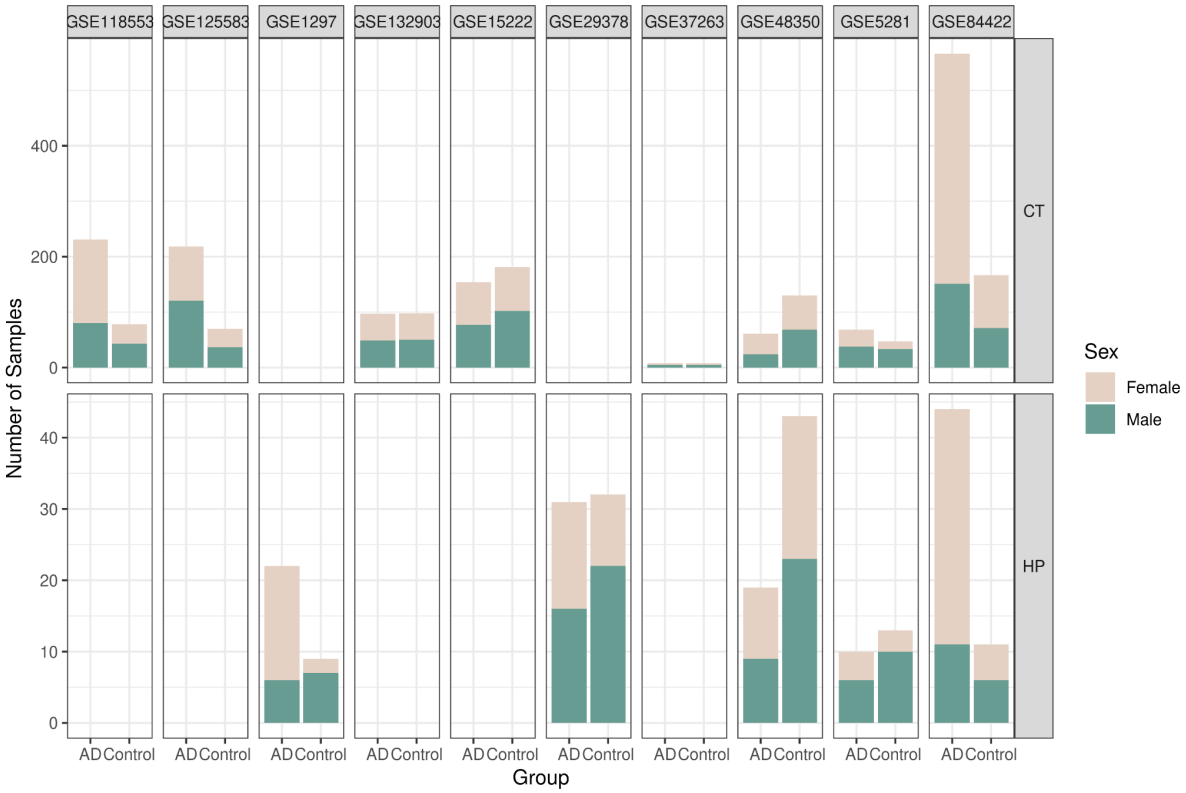

Figure 2. The number of samples per study, divided by sex, study, and experimental group (AD –

Alzheimer's disease).

Individual Transcriptomic Analysis

We carried out exploratory and processing steps on the datasets to ensure their comparability in subsequent analyses. We applied $\log_2$ transformation to studies GSE1297, GSE5281, GSE15222, GSE29378, and GSE48350 to homogenize magnitude order and filtered out samples from regions different to CT and HP from all studies. The studies GSE44768, GSE44770, GSE44771 and GSE33000 were excluded from the selection for presenting incompatibility due the way to process them.

The differential expression results for each study provided a variable number of significant genes across contrasts and studies (Table 2). In the case of SDID contrast, only GSE15222, GSE29378, and GSE84422 reported significant genes.

Table 2. Summary of differential gene expression analysis by brain region and contrast.

	Significant Genes IDF			**Significant Genes IDM**			**Significant Genes SDID**		
	LFC > 0	LFC < 0	Total	LFC > 0	LFC < 0	Total	LFC > 0	LFC < 0	Total
GSE118553 (CT)	387	291	678	759	401	1160	0	0	0
GSE125583 (CT)	4461	4230	8691	3366	3482	6948	0	0	0
GSE1297 (HP)	2	0	2	0	0	0	0	0	0
GSE132903 (CT)	5925	4782	10707	4643	4473	9116	0	0	0
GSE15222 (CT)	5042	3699	8741	2054	1628	3682	267	148	415
GSE29378 (HP)	63	21	84	347	193	540	0	1	1
GSE37263 (CT)	0	0	0	0	0	0	0	0	0
GSE48350 (CT)	28	34	62	2	6	8	0	0	0
GSE48350 (HP)	760	421	1181	15	1	16	0	0	0
GSE5281 (CT)	718	2307	3025	618	3815	4433	0	0	0
GSE5281 (HP)	96	70	166	138	111	249	0	0	0
GSE84422 (CT)	109	235	344	355	490	845	174	89	263

GSE84422 (HP)	0	0	0	0	0	0	0	0	0

CT = cortex tissue, HP = hippocampus tissue, IDF = Impact of the disease in females, IDM = Impact of the disease in males, and SDID = sex-differential impact of the disease.

Gene-Expression Meta-analyses reveal sex differences in Cortex

For each different comparison (IDF, IDM, and SDID), we performed a gene expression meta-analysis in the HP and CT, analyzing 8 studies for CT and 5 studies for HP. SDID contrast was employed to determine the set of genes with sex-impacted expression patterns in AD. Then, we performed both the IDF and the IDM contrasts to get into more detail about the impact of sex in this subset of genes.

SDID contrast revealed no significant genes in HP, so this region was excluded from further functional analyses. However, SDID detected 389 significant genes with a sex-differential expression in CT. Also, IDF and IDM comparisons revealed 3763 and 1876 differentially expressed genes in AD in females and males, respectively (Supplementary Table 1).

We divided the 389 SDID significant genes set in different subsets according to their significance and value of LFC on the three comparisons (Figures 3a and 3b). The six resultant subsets of genes were:

I) More repressed genes in affected females (179 genes)

II) More expressed genes in affected females (57 genes)

III) More repressed genes in affected males (14 genes)

IV) More expressed genes in affected males (6 genes)

V) Increased in AD females related to AD males (69 genes)

VI) Decreased in AD females related to AD males (64 genes)

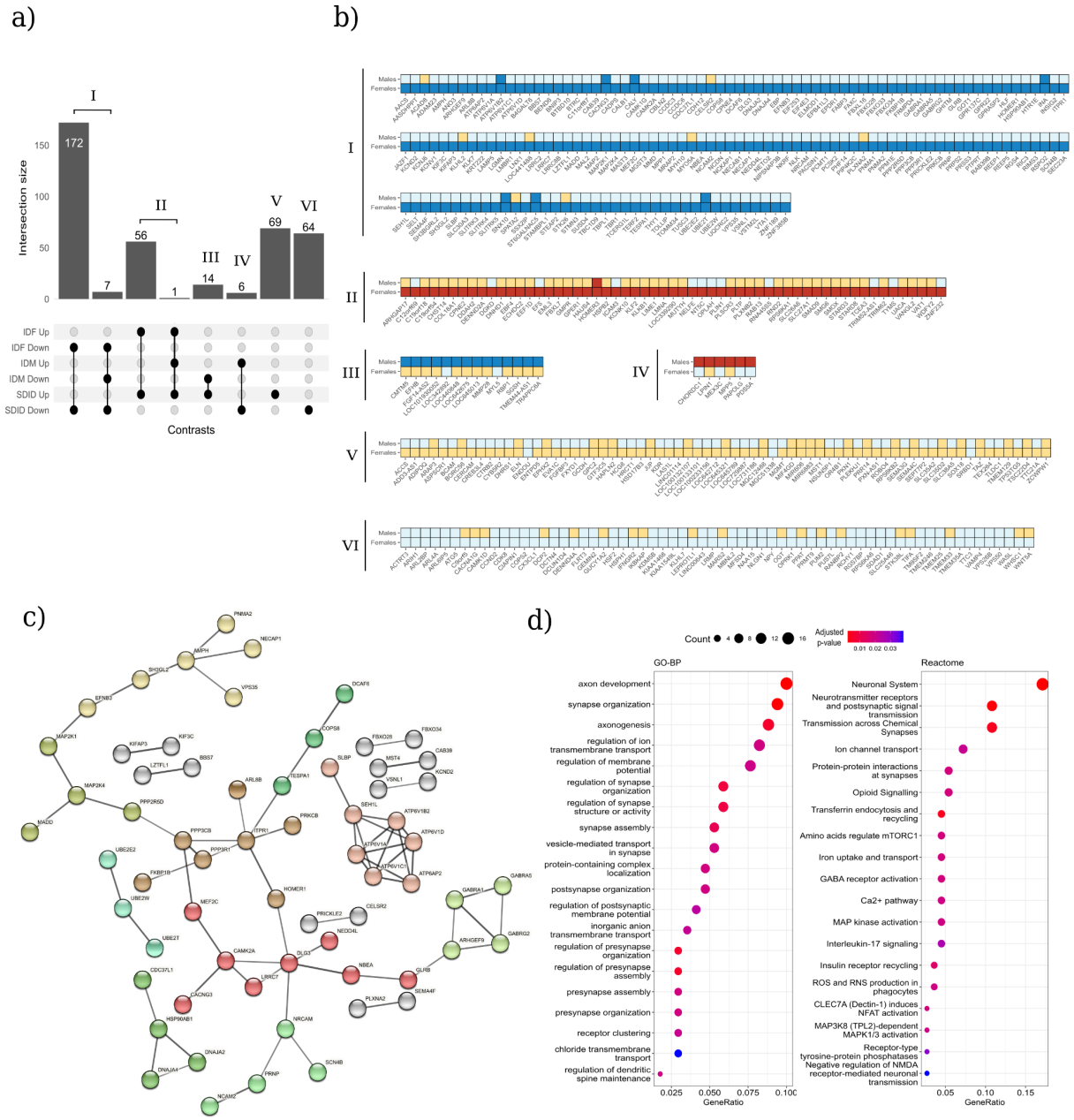

Figure 3. Significant sex-impacted genes detected in CT. **A**) Upset plot of the intersections between the differentially expressed genes in IDF, IFM, and SDID. Only the intersections containing SDID significant genes are shown. **B**) Tile plot of every gene composing the six resulting subsets. The direction of expression and significance are reflected, for females and males, with the following color pattern: blue indicates underexpression, red indicates overexpression, and the darker colors indicate statistical significance (adjusted p-value < 0.05). **C**) PPI network calculated from the gene subset I, showing only network edges with an interaction score greater than 0.7. Network nodes are colored according to cluster belonging (Table 3). **D**) Dot Plots summarizing the significant GO biological processes (left) and Reactome pathways (right) detected by ORA in the gene subset I.

Functional profiling of significant genes in the CT

We functionally characterized the gene subsets described above through ORA and the calculation of PPI networks. However, we only obtained significant results by both methods in the subset of genes repressed in females (subset I).

ORA results on gene subset I revealed a significant overrepresentation of 34 GO terms (BP ontology) and 20 Reactome pathways (Figure 3d, Supplementary Table 3). Most of these significant GO terms are related to various processes involved in the synaptic organization (GO:0050808, GO:0099174, GO:0050807, GO:0099173), assembly (GO:1905606, GO:0099054, GO:0051963), and vesicle-mediated transport (GO:0099003), as well as regulation of membrane potential (GO:0042391, GO:0060078, GO:1904062), axonogenesis (GO:0061564, GO:0007409), and cytoskeleton-dependent intracellular transport (GO:0030705), which play essential roles in the development and function of neurons. Regarding the Reactome pathways, most of them are involved in the neuronal system (R-HSA-112316) as well, such as Neurotransmitter receptors and postsynaptic signal transmission (R-HSA-112314), GABA receptor activation (R-HSA-977443), Ca2+ pathway (R-HSA-4086398), or Negative regulation of NMDA receptor-mediated neuronal transmission (R-HSA-9617324).

The PPI network calculated (Figure 3c) showed significantly more interactions between genes than expected for a random set of genes of the same size and degree distribution, having a PPI enrichment p-value of 1.27 x 10-11. Gene clusters were computed using the MCL clustering method included in the STRING web tool. As a result, 10 clusters composed of at least 3 genes were elucidated, most functionally related to neurological processes such as synapsis (Table 3). Functional enrichments in each cluster were provided by STRING (Supplementary Table 2).

Table 3. Gene clusters extracted from the PPI network calculated from the gene subset I (Significantly repressed genes in affected females).

Cluster	Genes	Relevant Significant Functions

Cluster 1 (8 genes; red)	CACNG3, CAMK2A, DLG3, GLRB, LRRC7, MEF2C, NBEA, NEDD4L	GO:1900449: Regulation of glutamate receptor signaling pathway R-HSA-112314: Neurotransmitter receptors and postsynaptic signal transmission
Cluster 2 (7 genes; salmon)	ATP6AP2, ATP6V1A, ATP6V1B2, ATP6V1C1, ATP6V1D, SEH1L, SLBP	GO:0051452: Intracellular pH reduction R-HSA-9639288: Amino acids regulate mTORC1
Cluster 3 (7 genes; brown)	ARL8B, FKBP1B, HOMER1, ITPR1, PPP3CB, PPP3R1, PRKCB	GO:0019722: Calcium-mediated signaling R-HSA-5607763: CLEC7A (Dectin-1) induces NFAT activation
Cluster 4 (6 genes; pearl)	AMPH, EFNB3, NECAP1, PNMA2, SH3GL2, VPS35	GO:0098793: Presynapse R-HSA-8856828: Clathrin-mediated endocytosis
Cluster 5 (4 genes; pista green)	MADD, MAP2K1, MAP2K4, PPP2R5D	GO:0000187: Activation of MAPK activity R-HSA-450294: MAP kinase activation
Cluster 6 (4 genes)	ARHGEF9, GABRA1, GABRA5, GABRG2	GO:0051932: Synaptic transmission, GABAergic R-HSA-977443: GABA receptor activation
Cluster 7 (4 genes; olive)	CDC37L1, DNAJA2, DNAJA4, HSP90AB1	GO:0006457: Protein folding R-HSA-3371497: HSP90 chaperone cycle for steroid hormone receptors (SHR)
Cluster 8 (4 genes; mint)	NCAM2, NRCAM, PRNP, SCN4B	- R-HSA-445095: Interaction between L1 and Ankyrins
Cluster 9 (3 genes; green)	COPS8, DCAF6, TESPA1	GO:0010387: COP9 signalosome assembly -
Cluster 10 (3 genes; turquoise)	UBE2E2, UBE2T, UBE2W	GO:0070979: Protein K11-linked ubiquitination -

Transcription Factors Activity in the CT

The TF activity analysis of the consensus profiles calculated from the three meta-analyses (IDF, IDM, and SDID) reported 23 TFs having sex-impacted activation patterns (adjusted p-value < 0.05 in SDID). All these significant TFs were split into three groups according to their activation profiles: Activated, Repressed and Divergent (Figure 4).

We labeled as Activated the 11 significant TFs with an NES > 0 in the three contrasts. These TFs are

predicted to be activated in both female and male AD cases but significantly increased in female cases compared to male cases. Among these TFs are ADNP, HMGN3, IRF3, KLF5, KLF9, MAZ, MBD3, MYNN, PRDM14, SIX5, and ZNF207.

The 10 TFs with NES < 0 in the three contrasts were grouped as Repressed. As opposed to the Activated TFs, these are predicted to be repressed in female and male cases. However, the repression in female cases is significantly greater than in male cases. The Repressed TFs are GTF2B, HOXB13, NANOG, NME2, PCGF2, SNAI2, ZBTB7A, ZC3H8, ZHX1, and ZHX2.

Lastly, we grouped as Divergent the last 2 remaining TFs. These TFs showed NES values with different signs in the three comparisons. In the case of CEBPZ, results indicate activation in female cases and repression in male cases. Regarding TERF2, it is activated in male cases and repressed in female cases.

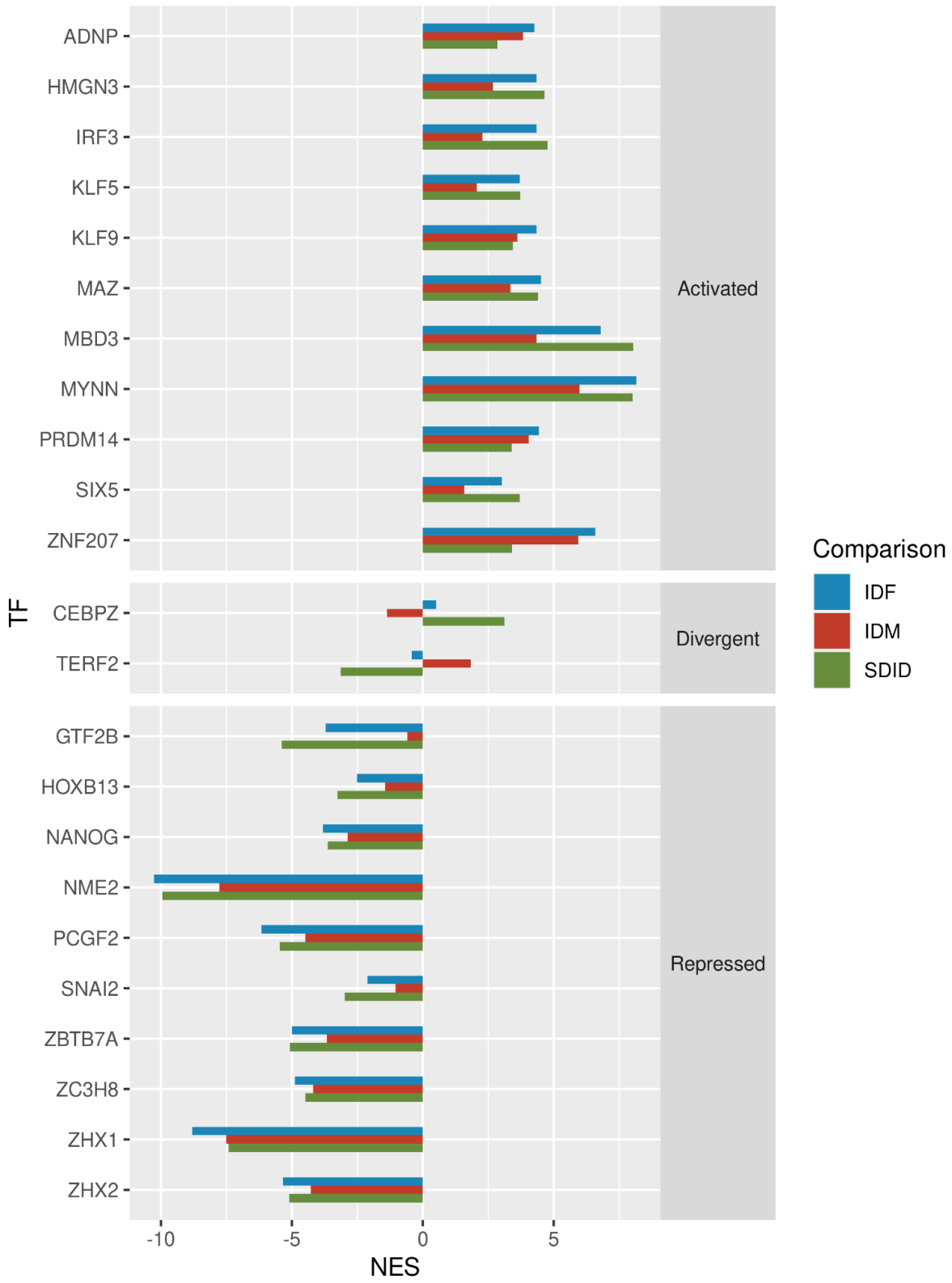

Figure 4. TFs with significantly altered activity calculated in the SDID consensus profile (adjusted p-value < 0.05). Activation values are measured as normalized enrichment scores (NES). TFs labeled as "Activated"

display positive values of NES in all comparisons. TFs labeled "Repressed" display negative values of NES in all comparisons. "Divergent" TFs show qualitative differences in their activation across comparisons.

DISCUSSION

AD is manifested by a progressive memory loss and other cognitive features. Their etiology and pathogenesis are complex. In the last decades, many omics studies have focused on exploring the influence of sex on neurodegenerative disease. Although there are some discrepancies, extensive literature collects the influence of sex on risk factors, prevalence and incidence, clinical manifestation, and treatment response [PMID: 29907423, PMID: 33896621]. Therefore, evaluating the impact of sex differences in AD is critical to improve prevention, diagnosis, and outcomes in males and females. We investigated sex differences in molecular mechanisms of AD by performing two specific meta-analyses in CT and HP; as the most involved brain areas in AD. While CT is the main region involved in cognitive functions, HP takes part in mental processes related to memory and those related to the production and regulation of emotional states and spatial navigation.

In general, the results showed that AD.females have a higher number of affected genes than AD.males. We haven't found significant SDID in HP, but 389 SDID significant genes were found in CT. This result suggests sex bias may mainly affect cognitive function. Therefore, this work was focused on CT and further on the set of genes repressed in AD.females (Set I) that showed significant consequences at functional profiling. Consistent with our findings, the multi-omics review performed by Lei Guo and colleagues points out that only 20 % of selected omic studies of sex differences in AD show male-specific changes and molecular-level changes are mostly attributed to women PMID: 33896621. Moreover, the significant SDID in AD.female were detected in CT, the main region in the brain involved in cognitive function. This could be one of the potential reasons why females present a faster decline and cognitive deterioration than males [PMID: 24434111].

Focused on SDID, the resulting network of the Set I genes in females.AD was divided into 10 clusters whose significantly altered processes mainly affect signal neurotransmission, Aβ aggregation, and deposition with important consequences in cognitive impairment. Regarding neurotransmission, females.AD displayed dysregulation of neurotransmitters such as glutamate and GABA, and synaptic transmission, including presynapse and postsynaptic synapsis. Glutamate is the most important excitatory neurotransmitter in the central nervous system (CNS), especially relevant for memory and its recovery. Additionally, glutamate is considered the main mediator of sensory, motor, cognitive, and emotional information [PMID: 34516293]. Conversely, GABA, widely distributed in the neurons of the cortex, acts as an inhibitory messenger, thereby stopping the action of excitatory neurotransmitters. GABA contributes to motor control, and vision and regulates anxiety among other cortical functions [PMID: 33554649]. As the main neurotransmitters are affected, the synaptic circuit (presynapse, postsynaptic, and transmission) is also significantly disturbed in females.AD, being the synaptic loss

the best correlate of cognitive deficits in AD [PMID: 32428558]. Interaction between L1 and Ankyrins proteins also appeared significantly altered in females.AD, several components of this family of scaffold proteins are involved in neurodegenerative diseases including AD by modulating neuronal excitability and neuronal connectivity via ion channels [PMID: 33485190; PMID 33239761].

The accumulation of misfolded proteins in the human brain is critical in many neurodegenerative diseases, including AD [PMID: 32479244]. The balance in the mechanisms that mediate protein folding, elimination, trafficking, Aβ aggregation, and deposition is key in AD. Females.AD presented dysregulation in pH and Calcium mediated signaling, both calcium and pH are closely involved in synapsis and neurotoxicity PMID: 15261480. Brain acidosis is associated with aging and might affect pathophysiological processes such as Aβ aggregation or inflammation in AD [PMID: 33578193]. In the brain, calcium could promote the accumulation of Aβ plaques and neurofibrillary tangles [PMID: 30059692] and is key for synaptic activity and memory, a process closely related to signal transduction pathways as MAPK kinase activation [PMID: 19419557] that also is significantly altered in females.AD. Females.AD also presented alterations in amino acids regulation of mTORC1 that have been related to autophagy inhibition, clearance, and abnormal protein formation in human neurons and mice [PMID: 28148781, PMID: 35672148]. The function of hps90 is regulated by steroid receptors, among other molecules, being key for their signal-transducing function and their participation in the folding, stabilization, and trafficking of several proteins [PMID: 9521088]. The imbalance in hsp90 and other chaperones can contribute to Tauopathies, disease onset, and progression [PMID: 29311797, PMID: 30128052]. Moreover, imbalance and abnormalities in endocytic processes happen earliest in AD. In this sense, clathrin-mediated endocytosis is dysregulated in females.AD, being one of the vias importantly involved in the internalization of amyloid precursor protein (APP) and the Aβ generation [PMID: 28655199]. The Aβ generation and cognitive decline are also associated with NFAT activation [PMID: 19828810] which we found altered and related to dectin-1 in AD.females. COP9 signalosome assembly is involved in ubiquitin-mediated proteolysis [PMID: 14647295]. In parasites, the disruption of COP9 signalosome causes dysregulation of the ubiquitin-proteasome pathway, therefore protein degradation and cell death [PMID: 32960936]. And no less important is the dysregulation of Protein K11-linked ubiquitination involved in protein degradation but also in inflammation [PMID: 31241013 PMID: 31241013 PMID: 34017102, PMID: 35343654].

We found 23 TFs with SDID, taking account of the tree meta-analyses, 11 of them displayed an activation pattern in both, females.AD and males.AD, but the increase in activation is more accused in female.AD and represent potential targets. The aberrant expression of ADNP has been linked to neural developmental disorders and proposed as a novel marker for the onset of frontotemporal dementia [PMID: 32533114, PMID: 24489906]. HMN3 has been related to Huntington's disease [PMID: 34111223]. IRF3 with other molecules is involved in the progression of AD and cognitive impairment [PMID: 36964897]. KFL5 and KFL9 might be critical in AD progression and development

[PMID: 35883144, PMID: 31545826]. MAZ, are involved in amyloidosis [PMID: 9819419], MBD3 induces neurotoxicity in a mouse model [PMID: 33174148], and MYNN present a great diagnostic value in patients with AD [PMID: 32309439]. PRMD14 has been related to pluripotency, motor neurons, and brain germinoma, and SIX5 keeps a relation with abnormal tau expression [PMID: 11726559].

We also found TFs with a repression pattern in females.AD and males.AD but significantly greater repressed in females. Among them, GTF2B has already been identified as a dysregulated transcription factor in AD [PMID: 33406707], while HOXB13 has been associated with prostate cancer risk and is the first time related to AD [PMID: 34911933, PMID: 31260412]. Nanog is a pluripotency stem cell marker with an important role in stem cell therapies for neurodegenerative diseases [PMID: 32655369, PMID: 27789410]. In the case of NME2, there are indications about NME proteins in cancer progression [PMID: 29058704], but never about their implication in AD or neural disorders. Regarding PCFG2, increased frequency has been reported at proteomic levels [PMID: 34182925]. We also found several type zinc finger TFs related to cancer progression and neural development and associated for the first time with AD in our study. The EMT-transcription factor SNAI2 is important for neural crest specification, neural crest cell migration, and survival during neural crest development [PMID: 31862304]. This TF has therapeutic implications in neuroblastoma cells [PMID: 31862304]. Studies in hepato-carcinoma cells showed that ZBTB7A may regulate neural development [PMID: 20336405]. Zc3h8 has been described as a repressor of inflammation in zebrafish [PMID: 29871925]. ZHX1 and ZHX2 participate in the development and progression of several types of cancer and the low expression of both has an impact on a poor prognosis in chronic lymphocytic leukemia [PMID: 33541423].

Related to transcription factors with a divergent pattern expression in females and males, CEBPZ activated in females.AD and repressed in males.AD, belongs to a family of TFs relevant in immune response control and inflammation [PMID: 8530141]. Other components of this family have been associated with neuroinflammation, an early event in AD [PMID: 35393391; PMID: 35353567]. While TERF2 repressed in females.AD and activated in male.AD has been related to senescence in AD [PMID: 36444490; PMID: 33107640]. In AD patients, the expression of TRF2 next to TRF1 has been positively correlated with age and Tau protein in blood serum [PMID: 30761946].

Strengths and limitations

Exploring sexual bias in AD is the greatest significance to improve clinical outcomes in females and males. For this challenge, we profit from an in silico approach by using computational models that represent a powerful tool to evaluate and integrate data. Moreover, the sample size increases with an augmentation of the number of studies integrated into the meta-analysis therefore, more subtle effects can be detected. All together allows greater consensus and statistical power in the obtained results.

Additionally, our in silico analysis is based on FAIR data (Findable, Accessible, Interoperable, Reusable), which we believe is of special relevance. Indeed, we believe it is vital and key to share and reuse research data legally in new research to improve advances.

Nowadays, cumulative evidence points to the sex influence in different disorders, including AD. However, still, it is difficult to find the segregation of data by sex in research studies, not even in those carried out to explore diagnostic or prognostic factors. For example, we excluded 44 (53%) studies from our systematic review due to the absence of sex information. We highlight the need to include sex information in research studies and databases, as they are relevant to health.

Finally, point out that we have analyzed AD regardless of early or late onset AD (EOAD, LOAD). Although there is evidence that EOAD and LOAD show differential clinical features and could be 2 different forms of a single entity it's not clear [PMID: 26546285].

Perspectives and significance

The results obtained in this study contribute to a better understanding of the impact of sex differences in molecular and functional levels in AD. The genes, biological processes, pathways, and TFs identified in SDID represent a source of targets valid for personalized clinical outcomes and the springboard for future research so as comparisons with other works.

Conclusions

Our results point out that sex differences have more impact on females.AD and particularly in their cognitive abilities. We have mainly identified impairments in the main neurotransmitters (Glutamate and GABA) and different mechanisms related to the generation and deposition of Aβ plaques. We have also identified genes and TFs that could be novel options for guiding new therapeutic strategies. More studies that consider sex dimension are required for a better interpretation of the results and to avoid masking essential differences that could impact tailored interventions. Last, underline the relevance of sharing data and work in open platforms for scientific progress.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data used for the analyses described in this work are publicly available at GEO (16). The accession numbers of the GEO datasets downloaded are GSE118553, GSE125583, GSE1297, GSE132903, GSE15222, GSE29378, GSE37263, GSE48350, GSE5281, and GSE84422.

Competing interests

The authors declare that they have no competing interests

Funding

This research was supported by and partially funded by the Institute of Health Carlos III (project IMPaCT-Data, exp. IMP/00019), co-funded by the European Union, European Regional Development Fund (ERDF, “A way to make Europe”), PID2021-124430OA-I00 funded by MCIN/AEI/10.13039/501100011033/ FEDER, UE (“A way to make Europe”).

Authors contributions

ALC analyzed the data; FGG designed and supervised the bioinformatics analysis; ALC designed and implemented the web tool; ALC, ZA, AM, FRG and FGG wrote the manuscript; ALC, ZA, AP, MPM, AM, FRG, MRH and FGG helped in the interpretation of the results; all authors writing-review and editing; FGG conceived the work. All authors read and approved the final manuscript.

Acknowledgments

The authors thank the Principe Felipe Research Center (CIPF) for providing access to the cluster, co-funded by European Regional Development Funds (FEDER) in Valencian Community 2014-2020.