Sex Influence on APOE Genotype Expression

STAT 530 Final Report

Jing Huang	Mingyu Chen	Tianqi Wu	Tianyu Ao
jingh7@illinois.edu	mingyuc4@illinois.edu	twu38@illinois.edu	tianyua2@illinois.edu

1 Introduction

1.1 background

Alzheimer's disease(AD) is one of the most serious diseases affecting the health of the elderly in the world [1]. It is a chronic neurological disease and 60% -70% of dementia is caused by it [2].

AD is characterized by pathological tissues in the brain surrounded by be-ta amyloid plaques [1]. The cause of AD is still controversial in the industry, but the current mainstream hypothesis is caused by the accumulation of be-ta amyloid variants. Recent years it is reported that a large amount of the risk is believed to be inheritary with many genes involved [3].

The apolipoprotein E (APOE) genotype is one of the protein factors that includes three alleles (e2, e3, e4) that are located on chromosome 19q3.2.Along with female sex chromosome, it is one of the genetic risk factors for dementia and AD [4]. Previous researches show APOE4 carriers display more widespread brain glucose hypometabolism [5] and a more severe phenotype in regions vulnerable to AD pathology, including parietal lobe, temporal lobe, and cingulate areas [6].

1.2 Motivation

Although APOE4 genotype and sex are well-established risk factors for AD, the effect of APOE4 genotype can be different for male and female. Previous researches on cross-sectional studies showed that APOE4 imposes a higher risk for AD in women than in men [13-15]. In experiments of diagnostic groups, they also found a pattern of effects whereby the adverse effect of APOE4 on brain structure/function manifests in men but not until clinical disease stages in women [16].

Combined with previous studies, we found out that limited researches have been conducted to study the potential direct relationship between female and male on APOE4. By studying relationship between them, better medicine interventions can be developed in a targeted manner. Hopefully, Our study could provide valuable indications to delay or prevent the progression of AD.

1.3 Research Directions

Inspired by Yuan's latest research on interaction of APOE genotype and sex [7], we aim to conduct our own research based on RNA sequencing(RNA-Seq) data set they provided. In particular, we want to find the difference of effect of APOE4 for male and female by performing differential gene expression analysis, enrichment analysis, gene ontology analysis and clustering.

2 Materials and Methods

2.1 Data Source

Data comes from paper by Yuan et al. [7]. It is available from GEO under accession number GSE140205. Their experiment is on female and male humanized APOE3 targeted replacement (hAPOE3) and APOE4 targeted replacement

(hAPOE4) homozygous mice. For simplicity purpose, we will denote them as APOE3 and APOE4 in the following illustration. In total, there are 4 groups and 5 animals in each group. Mice with APOE3 genotype targeted can be seen as control group and mice with APOE4 genotype targeted can be seen as treatment group.

2.2 Data Preprocessing

2.2.1 Get Gene Counts from Raw .SRA data

First, we download the SRA files from GEO database using SRA Toolkit and original SRA files are in paired-end format. Then, we convert those files to FASTQ files. With the mouse index downloaded from Ensemble v96 mouse index, we use Kallisto (version 0.46.2) to perform pseudoalignment. No quality control is used in this step. After that, R (version 3.6.3) package biomaRt (version 2.42.1) and tximport (version 1.14.2) are used to get annotations and gene counts.

2.2.2 Low-expressed Genes Filtering

We also checked the integrity of the data, and found no missing values. Then we filtered the data by keeping genes that are expressed enough in either group. We categorize genes as low-expressed in a group if that gene has cpm values >=1 in less than 2 samples from that group. Genes that are low-expressed in all 4 groups were dropped out. After filtering, 15983 out of 36441 genes were left. We then ran a TMM normalization using edgeR (version 3.28.1) on our data.

2.2.3 Choice of Transformation

The distribution of counts of expression data is shown in Fig.1. Since it is apparently right-skewed, we can apply transformation to our data to make it distribute more like a normal distribution before doing statistical analysis. We pick log2 transformation and nature transformation and see their behaviors. From the figure, both transformation display bell-shaped histograms, while log2 transformation attains larger variation In fact, most packages in R (including edgeR) use log2 transformation to normalize expression data. Hence, we can just follow the default setting of the functions they provided.

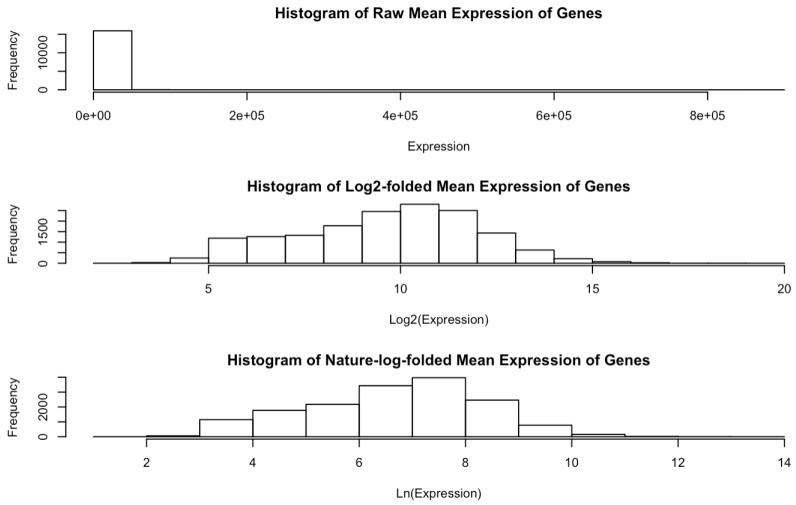

Figure 1: Distribution of Counts for Genes

2.3 Analyses

2.3.1 Differential Gene Expression Analysis

There are four experimental groups: female mice with APOE3 targeted, male mice with APOE3 targeted, female mice with APOE4 targeted and male mice with APOE4 targeted. For the simplicity of following illustration, we denote the four groups as APOE3.F, APOE3.M, APOE4.F, APOE4.M.

The differential gene expression analysis is done using edgeR (version 3.28.1) with quasi-likelihood negative binomial generalized log-linear model (glmQL) [11]. The design matrix in the analysis is: ~0+groups, where groups are specified above and 0 is included to drop the intercept column. Also, we decide to use tagwise dispersion since it provides better results.

We further define contrasts to specify the tests interested. There are three tests in total: APOE3vs4.F (The effect of APOE for female group), APOE3vs4.M (The effect the APOE for male group), APOE3vs4.FvsM (The difference of effect of APOE for female vs. male groups). Note that the last term can be seen as interaction term. In specific, we use function glmQLFit() to fit genewise negative binomial generalized linear models and glmQLFtest() to conduct quasi-likelihood tests. Toptags() is used to extract the most differentially expressed genes and p.value is adjusted using false discovery rate with threshold 0.05.

2.3.2 Enrichment Analysis

After extracting the most differentially expressed genes, we use DAVID 6.8 [18] to perform enrichment analysis.

2.3.3 Gene Ontology Analysis

We use package org.Mm.eg.db (version 3.10.0) to map the gene ensemble ids to Entrez Gene identifiers. Then, goana() and topGO() from package limma (version 3.42.2) are used to find over-representation of gene ontology (GO) terms. Finally, the results of GO terms and p.value are uploaded to REVIGO [17] to generate the visualizations.

2.3.4 Clustering

Hierarchical clustering and uniform manifold approximation and projection (UMAP) [12] are also performed to check the ability of separating data with the differentially expressed genes identified. Visualizations are done using package pheatmap (version 1.0.12) and umap (version 0.2.5.0).

3 Results

3.1 Differential Gene Expression Analysis

3.1.1 Heteroscedasticity exists in males/females expression level

To intuitively display the similarity of samples, we did a MDS (Multidimensional Scaling) plot. There are many commonly used distance metrics like euclidean distance, but here we choose the Pearson metric, because it can show the relative relationship between genes' expression level of samples, regardless of their absolute value [20].

Fig.2 shows correlation between the first dimension and the effect of APOE4. Also, on this plot, males' samples are more scattered than females, which suggests different dispersion level in two sex groups.

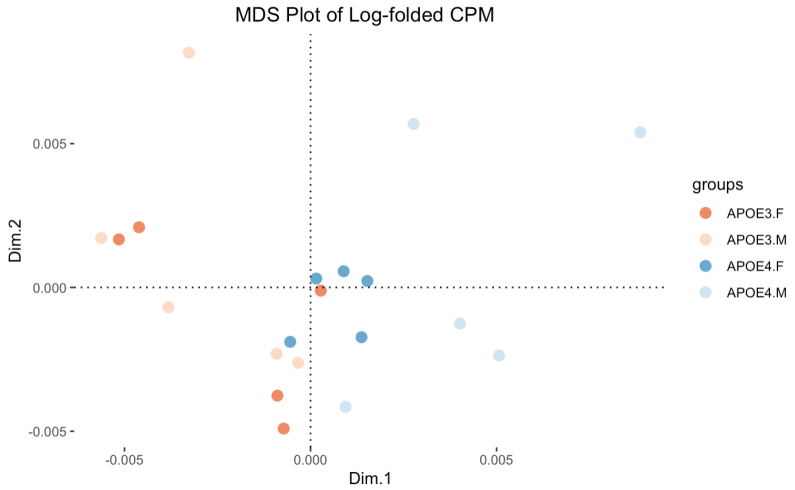

Figure 2: MDS Plot of Log-folded CPM

We are interested in finding the differentially expressed genes under two influencing factors: genotype(APOE3/APOE4) and sex(male/female). We choose to fit a quasi-likelihood negative binomial generalized log-linear model (glmQL) on the pre-processed data, and take a likelihood ratio test to find genes with highest p-values. More specifically, We want to know whether or not APOE4 influences gene expression level differently in males and females. To verify the existence of such distinction, we divide data into males and females, each includes 10 samples, run simple GLM with design matrix ~APOE separately, and then take a look at the coefficients. Fig.3 shows that, for most of the genes, the effect of APOE4 is similar in males and females. However, for some of the genes, the coefficients deviate largely. Hence, it is reasonable to assume the existence of interaction between APOE4's effect and sex type.

Figure 3: Distribution of Counts for Genes in Genders

3.1.2 APOE4 causes different expression of Ica1, Serpina3n and Oscar in both males and females

We ran genewise negative binomial GLMs with Quasi-likelihood tests on males and females separately, with FDR adjusted rate <= 0.05. There are 225 DEGs for males and 264 DEGs for females. Table 1,2 show top 6 DEGs separately.

	logFC	p.value	Gene Name
ENSMUSG00000046952	5.761883	1.1373×10^{-39}	Gm5815
ENSMUSG00000021091	-1.367038	1.609×10^{-38}	Ica1
ENSMUSG00000062995	1.921515	1.140×10^{-37}	Serpina3n
ENSMUSG00000054594	3.614955	2.045×10^{-37}	Oscar
ENSMUSG00000096768	-5.582101	1.777×10^{-36}	Gm47283
ENSMUSG00000038608	3.971723	1.091×10^{-18}	Dock10

Table 1: Top 6 Differentially Expressed Genes in APOE4-male and APOE3-male

We picked top 3 functional genes to present their expression level between APOE3 and APOE4 samples. Note that these three genes are most differentially expressed in both males and females.

	logFC	p.value	Name
ENSMUSG00000046952	7.7680787	2.591×10^{-47}	Gm5815
ENSMUSG00000062995	2.1170827	8.500×10^{-45}	Serpina3n
ENSMUSG00000021091	-1.3544659	6.834×10^{-38}	Ica1
ENSMUSG00000054594	3.4115980	1.356×10^{-34}	Oscar
ENSMUSG00000096768	-4.3660795	1.303×10^{-25}	Gm47283
ENSMUSG00000110275	-3.7975726	1.962×10^{-20}	Gm5905

Table 2: Top 6 Differentially Expressed Genes in APOE4-female and APOE3-female

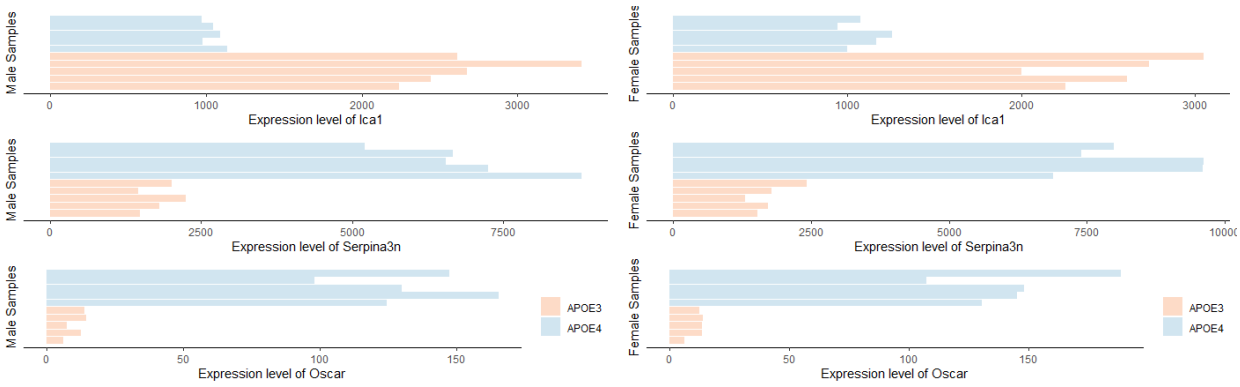

Figure 4: Differential Expression in Males and Females

Table.1 and 2 give exact logFC and p.value of top DEGs. Fig.4 shows the specific gene expression levels of APOE3 and APOE4 in males and females of three crucial genes we identified. Despite nature of pseudogene, these are top 3 most differentially expressed genes with highest p-values. Both females and males show an down-regulation of Ica1 (Islet cell autoantigen 1), an up-regulation of Serpina3n(Serine protease inhibitor A3N), and an up-regulation of Oscar under APOE4 compared to APOE3, with similar quantities.

We further investigated how these three genes might influence the Alzheimer's disease. Ica1 encodes a protein with an arfaptin homology domain that is found both in the cytosol and as membrane-bound form on the Golgi complex and immature secretory granules [22]. This protein is believed to be an autoantigen in insulin-dependent diabetes mellitus and primary Sjogren's syndrome. At the same time, we also know that excessive high blood sugar can damage tissues and organs, including the brain, and may cause Alzheimer's disease [23].

Serpina3n is a Protein Coding gene and it is implicated in nutritionally mediated hypothalamic inflammation in mice [24]. Mice lacking SerpinA3N develop more neuropathic mechanical allodynia than wild-type (WT) mice, and exogenous delivery of SerpinA3N attenuated mechanical allodynia in WT mice. Therefore, this gene has a strong correlation with brain lesions [25].

Oscar is the osteoclast associated receptor. Patients with AD have excessive amyloid plaques in the brain, and the pathology may extend to peripheral organs and cause skeletal amyloid deposition, which would enhance receptor activator nuclear factor-kappa B ligand signaling and lead to greater osteoclast activities[24]. Patients with osteoporosis may have Vitamin D deficiency or lower levels of Vitamin D binding protein, which protects against amyloid aggregation, thus linking Vitamin D deficiency and AD or osteoporosis and AD [26].

3.1.3 Enrichment Analysis gives more signals for females than males

We also performed gene functional annotation enrichment analysis on two sets of DEGs specified in males and females. Fig.5 and Fig.6 show the top enriched cluster extracted from DAVID. The result for males seems to be less intuitive, compared to the one for females, where we can identify an important gene family, the GPCR (family 3). The first explanation is that females react to APOE4 through GPCR while males do not. Another reason could be that there is larger dispersion in males' data, resulting more noises.

For females, we found that GPCR family 3 (or family C) is the top gene set enriched in all of the genes selected. G-protein-coupled receptors (GPCRs) represent one of the largest gene families in the animal genome, and many

diseases that endanger human health are inseparable with them. We also found previous study suggesting that these gene could play central roles on human AD [27].

20 Cluster(s) Download File

Annotation Cluster 1	Enrichment Score: 9.87			Count	P_Value	Benjamini
INTERPRO	GPCR, family 3, nine cysteines domain	RT		15	2.4E-12	7.5E-10
INTERPRO	GPCR, family 3	RT		15	3.6E-12	5.6E-10
INTERPRO	GPCR, family 3, C-terminal	RT		15	5.8E-12	6.0E-10
INTERPRO	GPCR, family 3, vomeronasal receptor, type 2	RT		14	7.3E-12	5.6E-10
INTERPRO	Extracellular ligand-binding receptor	RT		15	3.3E-11	2.0E-9
UP_KEYWORDS	G-protein coupled receptor	RT		37	6.1E-9	5.1E-7
UP_KEYWORDS	Transducer	RT		37	1.5E-8	8.2E-7
GOTERM_MF_DIRECT	G-protein coupled receptor activity	RT		37	9.8E-8	2.0E-5

Figure 5: Annotation clustering from DAVID for females

22 Cluster(s) Download File

Annotation Cluster 1	Enrichment Score: 2.01			Count	P_Value	Benjamini
UP_SEQ_FEATURE	domain:NACHT	RT		6	3.8E-6	1.6E-3
INTERPRO	NACHT nucleoside triphosphatase	RT		6	4.4E-6	7.3E-4
UP_SEQ_FEATURE	domain:DAPIN	RT		4	1.1E-3	2.0E-1
INTERPRO	DAPIN domain	RT		4	1.2E-3	1.2E-1
SMART	SM01289	RT		4	1.4E-3	1.2E-1
INTERPRO	Death-like domain	RT		5	5.0E-3	3.4E-1
UP_SEQ_FEATURE	repeat:LRR 7	RT		6	6.8E-3	6.2E-1
UP_SEQ_FEATURE	repeat:LRR 6	RT		6	1.5E-2	7.9E-1
INTERPRO	Leucine-rich repeat	RT		7	1.5E-2	6.5E-1
UP_SEQ_FEATURE	repeat:LRR 1	RT		7	1.7E-2	7.7E-1
UP_SEQ_FEATURE	repeat:LRR 2	RT		7	1.7E-2	7.7E-1
UP_SEQ_FEATURE	repeat:LRR 8	RT		5	1.8E-2	7.3E-1
UP_SEQ_FEATURE	repeat:LRR 5	RT		6	2.1E-2	7.2E-1
UP_KEYWORDS	Leucine-rich repeat	RT		7	2.3E-2	6.3E-1
UP_SEQ_FEATURE	repeat:LRR 4	RT		6	2.9E-2	7.8E-1
UP_SEQ_FEATURE	nucleotide phosphate-binding region:ATP	RT		14	2.9E-2	7.4E-1
UP_SEQ_FEATURE	repeat:LRR 3	RT		6	4.2E-2	8.1E-1
UP_SEQ_FEATURE	repeat:LRR 9	RT		4	6.5E-2	9.1E-1
INTERPRO	P-loop containing nucleoside triphosphate hydrolase	RT		11	1.7E-1	9.9E-1
UP_SEQ_FEATURE	repeat:LRR 10	RT		3	1.8E-1	9.9E-1
UP_KEYWORDS	Inflammatory response	RT		3	3.7E-1	9.3E-1
GOTERM_BP_DIRECT	inflammatory response	RT		4	4.9E-1	1.0E0

Figure 6: Annotation clustering from DAVID for males

3.1.4 Males and females react differently to APOE4 in gene-expression level

From what we observed in Section 3.1.1, there is potential difference between reactions of males and females to APOE4. Here, we tested to find specific genes that are influenced by APOE4 differently in males and females, using Quasi-likelihood test with FDR adjusted p-value <= 0.05. There are 67 genes that treated APOE4 with different expression variation. Genes with highest p-values are Ret Finger Protein-Like 4 (Rfpl4), Dock10(Dedicator Of Cytokinesis 10), Sall4(Spalt Like Transcription Factor 4) and Olfr907(olfactory receptor 907).

In female groups, for Rfpl4, Dock10, Sall4 express, the expression level of APOE3 and APOE4 are in a consistent level while for Olfr907, APOE3 expression is significantly higher than APOE4. In male groups, for all Rfpl4, Dock10, Sall4 and Olfr907, APOE4 all have the higher expression level than APOE3.

Rfpl4 is a RING finger-like protein with a B30.2 domain. This level of mRNA expression is maintained in all growing oocytes, through to the completion of meiosis, and declines during early embryonic cleavage [29]. Rfpl4 may regulate cell cycle progression through meiosis in germ cells by holding in check the machinery responsible for the progression of mitosis, which is related to part of the funtion of APOE allele family [30].

Dock10 is identified as an interleukin-4 (IL4)-inducible gene in chronic lymphocytic leukemias (CLLs) [31]. It is distributed in the cytoplasm and nucleus of CLL cells, and IL4 increased its expression in both cellular compartments and could represent a point of convergence for IL4 signalling and small Rho GTPase function in B cells[31].

Sall4 is a transcription factor encoded by a member of the Spalt-like (SALL) gene family. It plays an essential role in maintaining the pluripotent and self-renewal properties of embryonic stem cells (ESCs) [32].

There is an opposite regulation of Olfr907 in males and females. This gene is one of the Olfactory receptors interacting with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell [33]. It belongs to the same family of GPCR arising from single coding-exon genes[33].

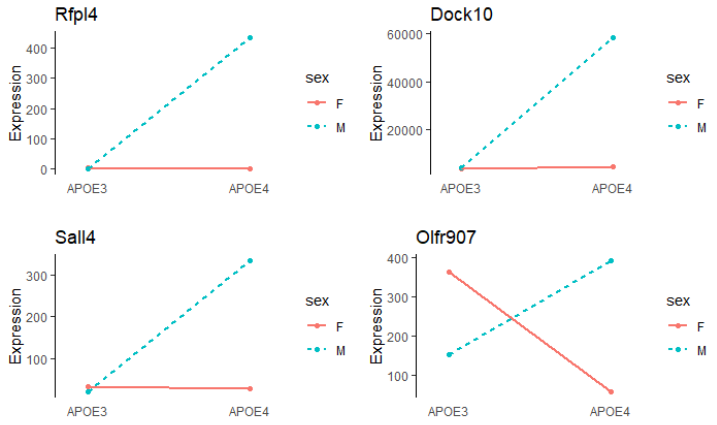

Figure 7: Differential Expression of APOE4 over APOE3 in Males/Females

3.2 Further Analysis

3.2.1 Gene ontology(GO) analysis

We decide to perform gene ontology analysis on highly expressed genes for the test (APOE3 vs. APOE4 in females) for the same reason as enrichment analysis. Fig.8 presents the resulting scatterplot from online tool REVIGO[17]. It shows the cluster representatives (i.e. terms remaining after the redundancy reduction) in a two dimensional space derived by applying multidimensional scaling to a matrix of the GO terms' semantic similarities. The Bubble colors indicate the user-provided p-values (legend is in upper right-hand corner); the sizes indicate the frequencies of the GO terms in the underlying GOA database (larger bubbles represent more general terms) [28]. The most expressed genes are mainly concentrated in protein signaling pathway and tissue developments.

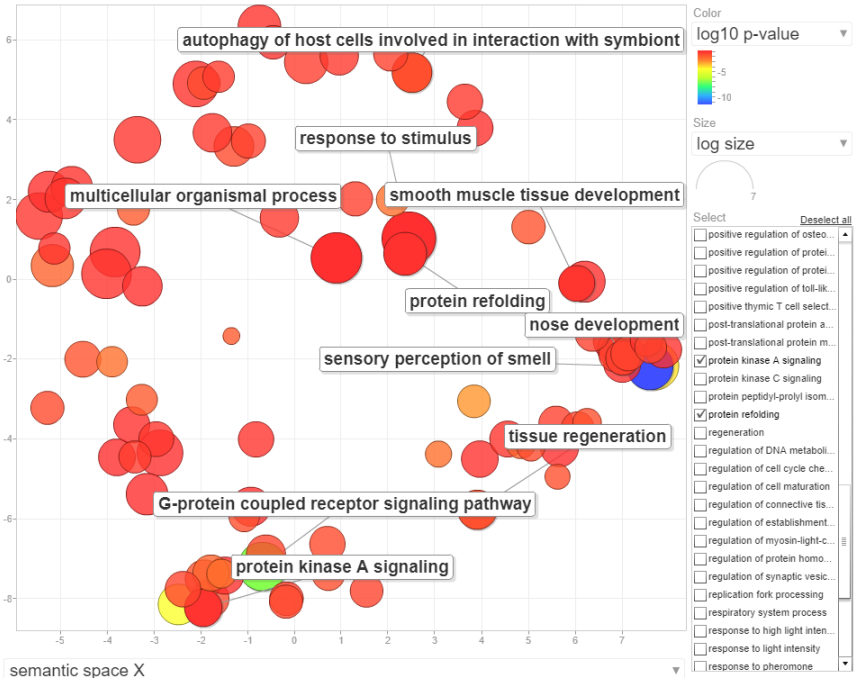

Figure 8: REVIGO clustering result

3.2.2 Hierarchical clustering and UMAP

After identifying the highly expressed genes for the three tests illustrated above, we decide to take the union of those gene sets to perform hierarchical clustering. Fig.9 and Fig.10 show the clustering results on 440 selected genes and it is clear that genes are separated into four groups. Despite several misclassifications, the performance is relatively acceptable. If we cut the tree at four clusters, the classification accuracy would be 85%. Two of samples from group APOE3.F a remisclassified as APOE3.M and one of the sample from group APOE4.M is misclassified as APOE4.F. The model has no problem distinguishing between APOE3 and APOE4.

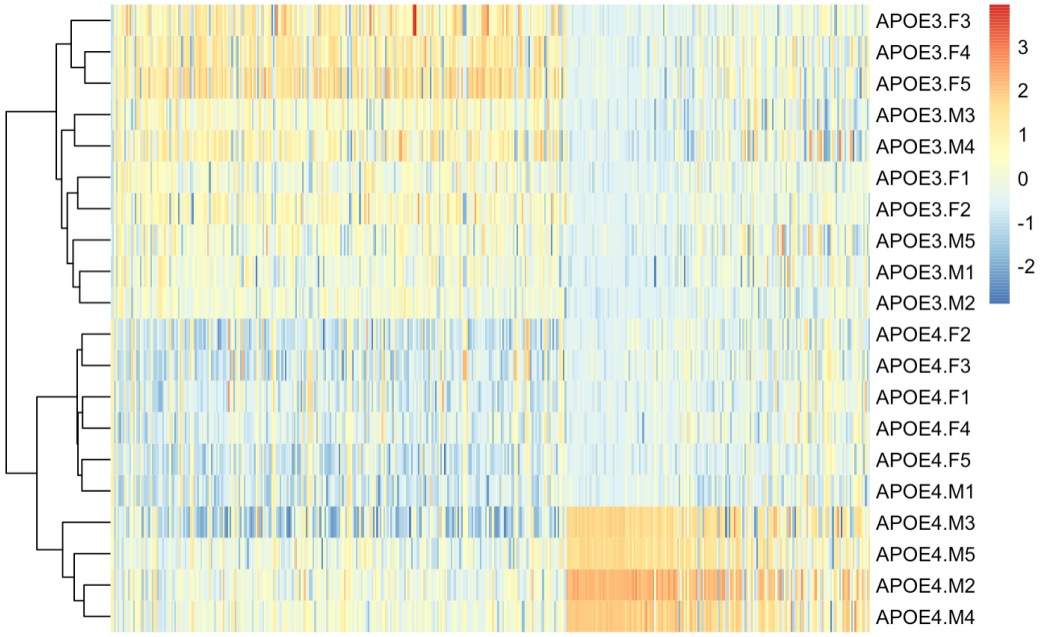

Figure 9: Hierarchical Clustering

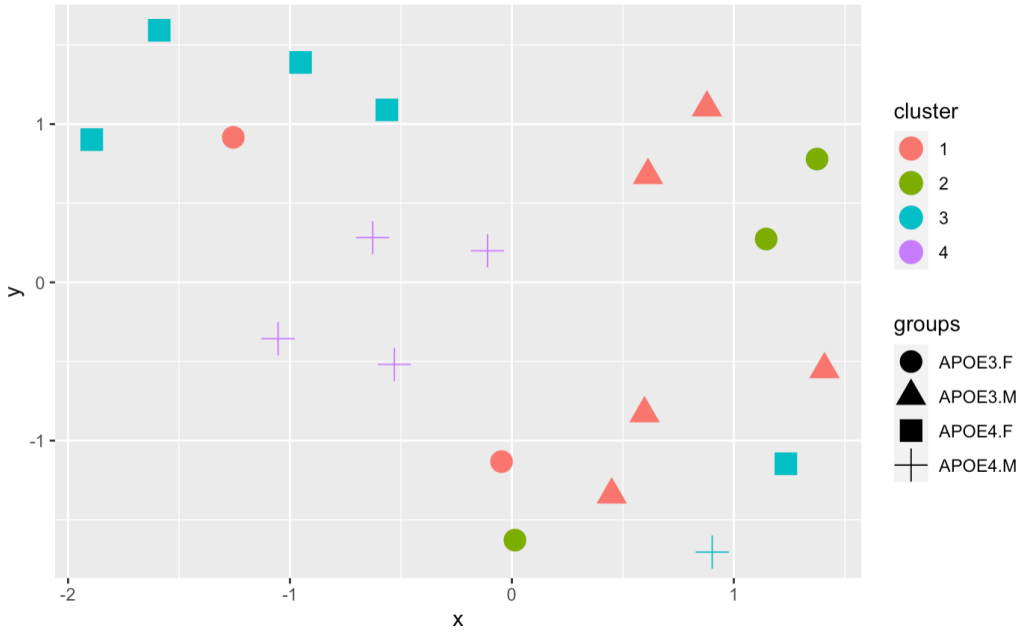

Figure 10: UMAP

4 Discussion

4.1 Summary

To investigate the effect of sex and APOE genotype, we performed a differential gene expression analysis following gene set enrichment analysis and gene ontology on 16 months old APOE3 and APOE4 male and female mice. Our results show large changes of genes that are related to insulin level, brain lesions and osteoclast activities in both males and females. Our results also indicate that the influence of hAPOE4 is different for some genes in males and females. In general, females has a clearer reaction to APOE4 than males. We found strong activity of GPCR family 3 in females, but less in males. And for females, the DEGs are most involved in protein signaling pathway and tissue developments.

4.2 Future insights

There is much discussion under the statistical methods we used. First, the order of our preprocessing steps is uncertain. We did twice normalizations, first used for filtering data, and the second after filtering, all before estimating dispersion. However, we could also choose not to do the second normalization. The order of these processing might heavily influence our results of differential expression analysis, according to Lin et al.(2016) [20]. Also, different normalization methods, distance matrix, and even packages we used could lead to different result, and yet these are barely discussed in our analysis.

Data dispersion is also an interesting topic because different dispersion estimation sometimes give different results. In our data, males' samples have larger dispersion than females, and in our report we used moderated gene-wise or tag-wise dispersion for analysis, which estimates and corrects dispersion for each gene independently. But it could also be a good choice to do a sex-wise dispersion, or sex-wise + trended-wise dispersion for males and females separately. We didn't do that because we are unsure if this is statistically valid. But it is worth investigating.

Also, analysing interaction effect with glm is sensitive to data processing and models. For easy interpretation, we modeled regression on 4 groups directly, not on sex/APOE factor plus a factorial interaction term. By using this ANOVA-like glm, we are assuming that our four groups are independently behaving, and the effects on them are different. To answer our main interests we performed three likelihood tests independently. The drawback of doing this is that it might not be convincing to join and compare the results of these tests, because the p-values are calculated and adjusted under different scenarios. Hence, for the gene set enrichment analysis and gene ontology analysis, we cannot perform an overall analysis, instead we had to do these works respectively, or only picked one of our three tests to visualize, which lacks general interpretation. More discussion about different statistical methods for differential gene expression analysis can be found in Lin et al.(2016) [20].

How should we choose the top differentially expressed genes? In our report, we sort on the p-values, and we could also choose genes based on their influential sizes. By doing this, we are assuming "the more the better" and are capturing large signals, but there may be small but stable signals which are identically important. Since we only have data of cohort experiments, it is hard to measure which signal is stable and which isn't. The situation might be changed if we can use high-frequency expression data one day.

Understanding the triggers and flows behind expression variation are the next step we should go. We have recognized GPCR family 3 as the most differentially expressed gene set under APOE4 in both males and females, but yet we know little about its working manner. There could be, for instance, genes acting as intermediary between GPCR and Alzheimer's Disease, and maybe we can detect them with statistical methods. In addition, what we found for the third interest are individual genes that react differently to APOE4 in males and females. However, it's not clear how and why this distinction exists. For example, look at the differences for Rfpl4, Dock10 and Sall4 displayed in Fig.7, we might think that these genes are not influenced by APOE4 in females, however, they could also possibly be suppressed by other bio-processes acting in females uniquely, or maybe APOE3 has influences on males.

5 References

1. Hardy, John A., and Higgins, Gerald A. "Alzheimer's disease: the amyloid cascade hypothesis." Science, vol. 256, no. 5054, 1992, pp. 184-186.

2. World Health Organization. "Dementia Fact sheet". Dec 2017, https://www.who.int/news-room/fact-sheets/detail/dementia. Accessed 15 May. 2020.

3. Ballard C., et al. "Alzheimer's disease." Lancet, vol. 377 (9770): 1019–31.

4. Lu, Po H., et al. "Apolipoprotein E Genotype Is Associated with Temporal and Hippocampal Atrophy Rates in Healthy Elderly Adults: A Tensor-Based Morphometry Study1." Journal of Alzheimer's Disease, vol. 23, no. 3, 2011, pp. 433–442., doi:10.3233/jad-2010-101398.

5. Mosconi, L., et al. "Brain Metabolic Decreases Related to the Dose of the ApoE e4 Allele in Alzheimer's Disease." Journal of Neurology, Neurosurgery & Psychiatry, vol. 75, no. 3, 2004, pp. 370–376., doi:10.1136/jnnp.2003.014993.

6. Mosconi, L., et al. "MCI Conversion to Dementia and the APOE Genotype: A Prediction Study with FDG-PET." Neurology, vol. 63, no. 12, 2004, pp. 2332–2340., doi:10.1212/01.wnl.0000147469.18313.3b.

7. Shang, Yuan, et al. "Evidence in Support of Chromosomal Sex Influencing Plasma Based Metabolome vs APOE Genotype Influencing Brain Metabolome Profile in Humanized APOE Male and Female Mice." Plos One, vol. 15, no. 1, 2020, doi:10.1371/journal.pone.0225392.

8. Huang, Da Wei, et al. "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources." Nature Protocols, vol. 4, no. 1, 2008, pp. 44–57., doi:10.1038/nprot.2008.211.

9. Robinson, Mark D, and Alicia Oshlack. "A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data." Genome Biology, vol. 11, no. 3, 2010, doi:10.1186/gb-2010-11-3-r25.

10. Chen, Yidong, and Paul S. Meltzer. "Gene Expression Analysis via Multidimensional Scaling." Current Protocols in Bioinformatics, vol. 10, no. 1, 2005, doi:10.1002/0471250953.bi0711s10.

11. Lun, Aaron T. L., et al. "It's DE-Licious: A Recipe for Differential Expression Analyses of RNA-Seq Experiments Using Quasi-Likelihood Methods in EdgeR." Methods in Molecular Biology Statistical Genomics, 2016, pp. 391–416., doi:10.1007/978-1-4939-3578-9_19.

12. Mcinnes, Leland, et al. "UMAP: Uniform Manifold Approximation and Projection." Journal of Open Source Software, vol. 3, no. 29, Feb. 2018, p. 861., doi:10.21105/joss.00861.

13. Bretsky, P M, et al. "Evidence for an Interaction between Apolipoprotein E Genotype, Gender, and Alzheimer Disease." Alzheimer Disease & Associated Disorders, vol. 13, no. 4, 1999, pp. 216–221., doi:10.1097/00002093-199910000-00007.

14. Payami, H, et al. "Alzheimer's Disease, Apolipoprotein E4, and Gender." JAMA, vol. 271, no. 17, 1994, pp. 1316-1317.

15. Poirier, J, et al. "Apolipoprotein E Polymorphism and Alzheimer's Disease." Lancet, vol. 342, no. 8873, 1993, pp. 697-699.

16. Sundermann, Erin E et al. "Sex differences in the association between apolipoprotein E 4 allele and Alzheimer's disease markers." Alzheimer's & dementia (Amsterdam, Netherlands) vol. 10 438-447. 7 Jul. 2018, doi:10.1016/j.dadm.2018.06.004

17. Supek, Fran, et al. "REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms." PLoS ONE, vol. 6, no. 7, 2011, doi:10.1371/journal.pone.0021800.

18. Huang, Da Wei, et al. "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources." Nature Protocols, vol. 4, no. 1, 2008, pp. 44–57., doi:10.1038/nprot.2008.211.

19. Castaldi, Peter J et al. "Screening for interaction effects in gene expression data." PloS one vol. 12,3 e0173847. 16 Mar. 2017, doi:10.1371/journal.pone.0173847

20. Lin, Yanzhu, et al. "Comparison of Normalization and Differential Expression Analyses Using RNA-Seq Data from 726 Individual Drosophila Melanogaster." BMC Genomics, vol. 17, no. 1, 2016, doi:10.1186/s12864-015-2353-z.

21. Mathé, Ewy, and Sean Davis, eds. Statistical Genomics: Methods and Protocols. Vol. 1418. Methods in Molecular Biology. New York, NY: Springer New York, 2016.

22. Mally, M. I., et al. "ICA69 is expressed equally in the human endocrine and exocrine pancreas." Diabetologia 39.4 (1996): 474-480.

23. Nguyen, Cuong Q., and Ammon B. Peck. "Unraveling the pathophysiology of Sjogren syndrome-associated dry eye disease." The ocular surface 7.1 (2009): 11-27.

24. Sergi, Domenico, et al. "SerpinA3N Is a Novel Hypothalamic Gene Upregulated by a High-Fat Diet and Leptin in Mice." Genes & Nutrition, vol. 13, no. 1, 2018, doi:10.1186/s12263-018-0619-1.

25. Vicuña, Lucas, et al. "The Serine Protease Inhibitor SerpinA3N Attenuates Neuropathic Pain by Inhibiting T Cell–Derived Leukocyte Elastase." Nature Medicine, vol. 21, no. 5, 2015, pp. 518–523., doi:10.1038/nm.3852.

26. Chen, Yu-Hung, and Raymond Y Lo. “Alzheimer’s disease and osteoporosis.” Ci ji yi xue za zhi = Tzu-chi medical journal vol. 29, no. 3, 2017, pp. 138-142. doi:10.4103/tcmj.tcmj_54_17

27. Dal Prà, Ilaria et al. “Family C G-Protein-Coupled Receptors in Alzheimer’s Disease and Therapeutic Implications.” Frontiers in pharmacology, vol. 10, 1282. 28 Oct. 2019, doi:10.3389/fphar.2019.01282

28. Supek, Fran et al. “REVIGO summarizes and visualizes long lists of gene ontology terms.” PloS one, vol. 6,7 (2011): e21800. doi:10.1371/journal.pone.0021800

29. Suzumori, Nobuhiro, et al. "RFPL4 interacts with oocyte proteins of the ubiquitin-proteasome degradation pathway." Proceedings of the National Academy of Sciences 100.2 (2003): 550-555.

30. Hennebold, Jon D. "Characterization of the ovarian transcriptome through the use of differential analysis of gene expression methodologies." Human reproduction update 10.3 (2004): 227-239.

31. Yelo, Estefanía, et al. "Dock10, a novel CZH protein selectively induced by interleukin-4 in human B lymphocytes." Molecular immunology 45.12 (2008): 3411-3418.

32. Tatetsu, Hiro, et al. "SALL4, the missing link between stem cells, development and cancer." Gene 584.2 (2016): 111-119.

33. Touhara, Kazushige. "Odor discrimination by G protein-coupled olfactory receptors." Microscopy research and technique 58.3 (2002): 135-141.