**Identification of the co-expression pattern for differentially expressed genes in both hippocampus and olfactory bulb of mice**

Guangxin Chen

[chen6751@umn.edu](mailto:chen6751@umn.edu) 5509431

University of Minnesota

05/14/2021

**Introduction**

Alzheimer’s disease (AD) is a progressive neurodegenerative disorder marked with cognitive dysfunction and memory impairment, which is the most common cause of dementia. In the past studies, many brain regions were identified to be highly related to this progressive disease. For example, hippocampus is found primarily related to memory, and it is one of the first regions in the brain that suffers damages in Alzheimer’s disease (Lane et al., 2018). Also, emerging evidence show that olfactory function loss can make a prediction for the start of cognitive decline, while the relationships of olfactory deficits and neurodegeneration is not clear (Murphy, 2019). In molecular level, the change of gene expression contributes to AD progression. Traditional section methods like immunohistochemistry can help researchers identify the protein expression preserving the spatial structure of brain regions, while the cost of antibody would be high if we have wide types of proteins to study. Methods like Western blot and next-generation sequencing are alternative ways for studying gene expression while having the disadvantage for losing the anatomical information. Spatial Transcriptomics (ST) is a novel method that can detect the gene expression while preserving the spatial information for brain, thus is widely used for finding differentially expressed genes among different regions. This project is designed to find the co-expression pattern for differentially expressed genes in both hippocampus and olfactory bulb based on the paper of Navarro et al., which used ST to identify these genes in mice.

Gene co-expression describes the result which a cluster of genes are all upregulated or downregulated in certain condition. The identified co-expression genes are sometimes called modules and have related common features. For example, in the paper of Chen et al., they identified co-expression network of plaque-induced genes and oligodendrocyte genes using ST from transgenic mice based on Weighted correlation network analysis (WGCNA). Also, they used WGCNA analyzed the gradual co-expression of plaque-induced genes according to different phases of AD and visualized the results using Circos plots (Chen el al., 2020). The primary idea about the project was trying to identify the gradual co-expression of oligodendrocyte genes using WGCNA, while this was found hard to achieve due to lack of source files. However, datasets in the paper of Navarro el al. are complete, and the co-expression analysis is missing. So, identifying the co-expression pattern for different types of mice and comparing the results is worthwhile to try as a complement for their research.

**Data Preparation and Methods**

Four types of mice were used in the research conducted by Navarro et al.: there were two transgenic types 3xAD and 3xPB, with their corresponding control WT (wild-type C57BL/6J) and PB (transgenic Polb+/-). The 3xAD mouse has triple transgenic APP, PS1, and MAPT while the 3xPB mouse has a quadrupole transgenic Polb+/-, APP, PS1, and MAPT. The researchers created the 3xPB type to validate their hypothesis that the progression rate of AD is faster due to the loss of DNA repair (loss of function of DNA polymerase β), and found it did have more AD features than 3xAD mouse (Hou el al., 2018). In total, 48 tissue sections were analyzed using ST, in which the expression of 16484 genes in hippocampus and 15107 genes in olfactory bulb were counted and then clustered using factor analysis. The differentially expressed genes were identified in both hippocampus and olfactory bulb, and the authors used multiple analysis methods like hierarchically clustered heatmaps and individual gene plot identifying genes that showed substantial gene expression distinctions between different mouse model and clear spatial patterns. There were totally 55 significantly differential expressed genes found in hippocampus and 73 genes in that of olfactory bulb. While the number of such genes described in the provided supplementary files is 20 in hippocampus and 23 for olfactory bulb.

In this project, the counted and unselected gene expression datasets of four types of mouse in both hippocampus and olfactory bulb were used as the source files. There were around 5 to 6 such file for each category, and I just randomly chose one of them from each. The differentially expressed genes in both regions were selected based on the supplementary files provided by the authors. As described above, the number of such genes is 20 in hippocampus and 23 for olfactory bulb. The number of choosen genes in olfactory bulb was then truncated to 22 because the lack of gene *Fam32* in the provided source file. Those genes were clustered using WGCNA package in R and their co-expression profiles were analyzed.

WGCNA is abbreviation for weighted correlation network analysis, it focuses on cluster of genes or modules rather than individual gene identification. This method uses the interaction pattern between genes and constructs a co-expression network based on that. Then module-based analysis is performed on the network, which can help us to find interesting modules and identify module preservation from different data. To construct the network, input file with gene expression profiles is converted to adjacency matrix, which encodes how a pair of nodes is connected. Due to the intrinsic nature of gene expression network, it is constructed as a weighted signed network based on the equation:

Where is the correlation coefficient for two genes and , is the soft thresholding power. The soft thresholding with power function was used because network results are highly robust with respect to the choice of the power β (Zhang et al., 2005). Empirically, many co-expression networks based on expression data from a single tissue exhibit scale free topology. Thus, the exact value of needs to be determined based on each provided input file, and the lowest value for the network to exhibit such character is often chose as the value for .

The connectivity of each gene is defined as the row sum of adjacency matrix:

For weighted network, it indicates the sum of connection strengths to other nodes(genes). We can conclude that a certain gene is highly co-expressed with others of its connectivity is high. The mean score of connectivity is then the average of intramodular connective score of all genes in the given module.

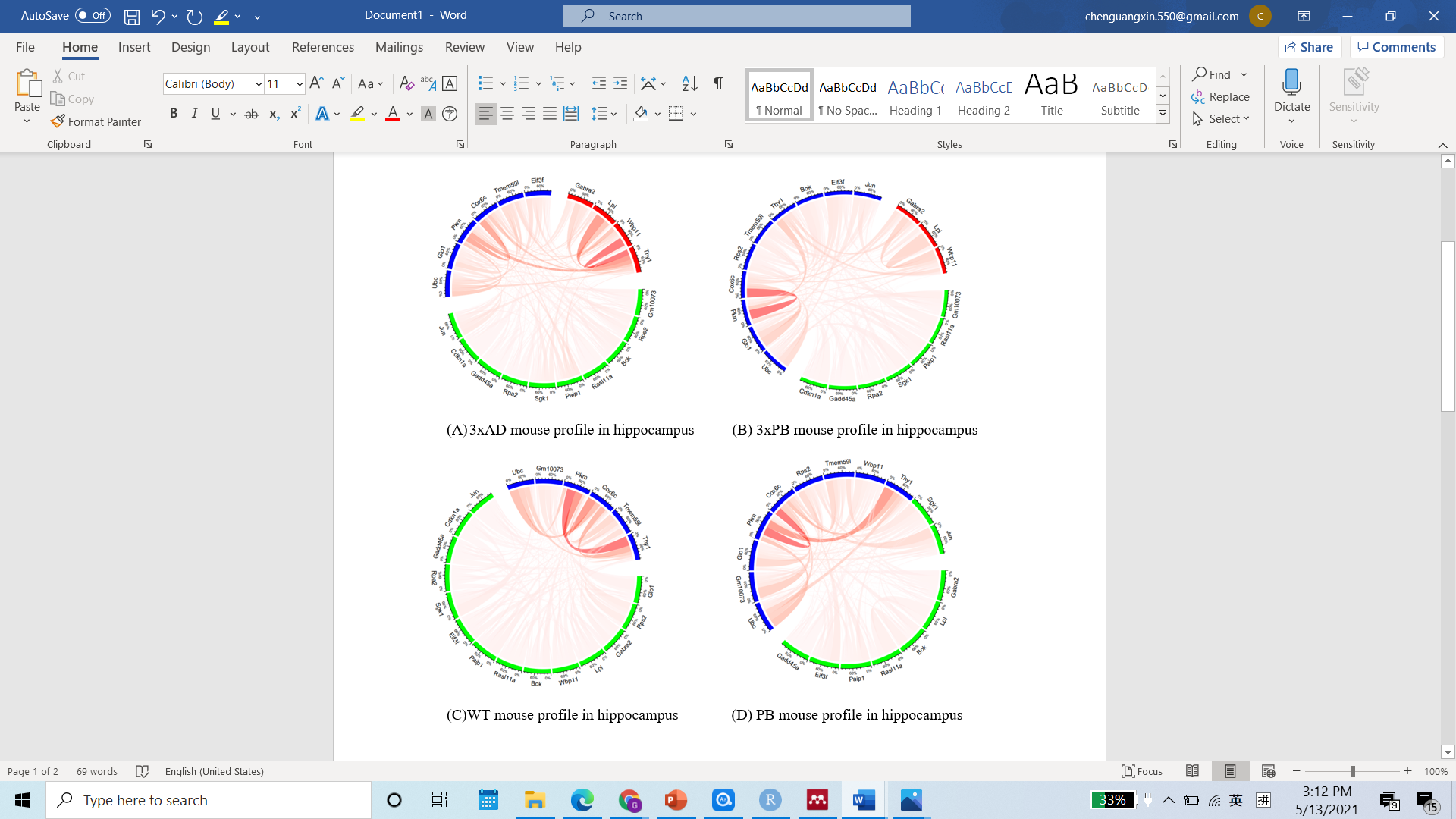
The modules of genes are defined from a hierarchical cluster tree as branches. In WGCNA, modules can be labeled as different integers or colors. And the package uses a dynamic hybrid branch cutting method for branch choosing. The idea about module eigengene is developed as a summary of the expression profiles in a module. It is the most highly connected intramodular hub gene and is defined by the singular value decomposition of gene expression data. The resulted adjacent matrix and module can be exported and visualized using many methods and software, and I chose the Circos plot due to its intuitive on finding co-expression pattern.

**Experiment Design**

The datasets of ST profile were downloaded from the source data provided by Navarro el al., the names of selected differential expressed genes were found from the supplementary tables. After the input file was loaded to R, it was subsetted to contain only the selected genes. Then the soft power was calculated by WGCNA’s ‘‘pickSoftThreshold’’ function to construct the adjacency matrix. As stated above, the threshold is picked based on the distribution of input data, thus the of scale free topology model fit for the data is plotted against soft threshold to find the best value for . The modules were then identified using the “blockwiseModules” function with found soft power, and the network was constructed using “adjacency” to find the adjacency matrix. The resulting matrix was then visualized using Circos plots based on R package circlize. This pipeline of analysis was performed for each type of mouse in both gene expression profile at hippocampus and olfactory bulb, so there were eight Circos plots in total.

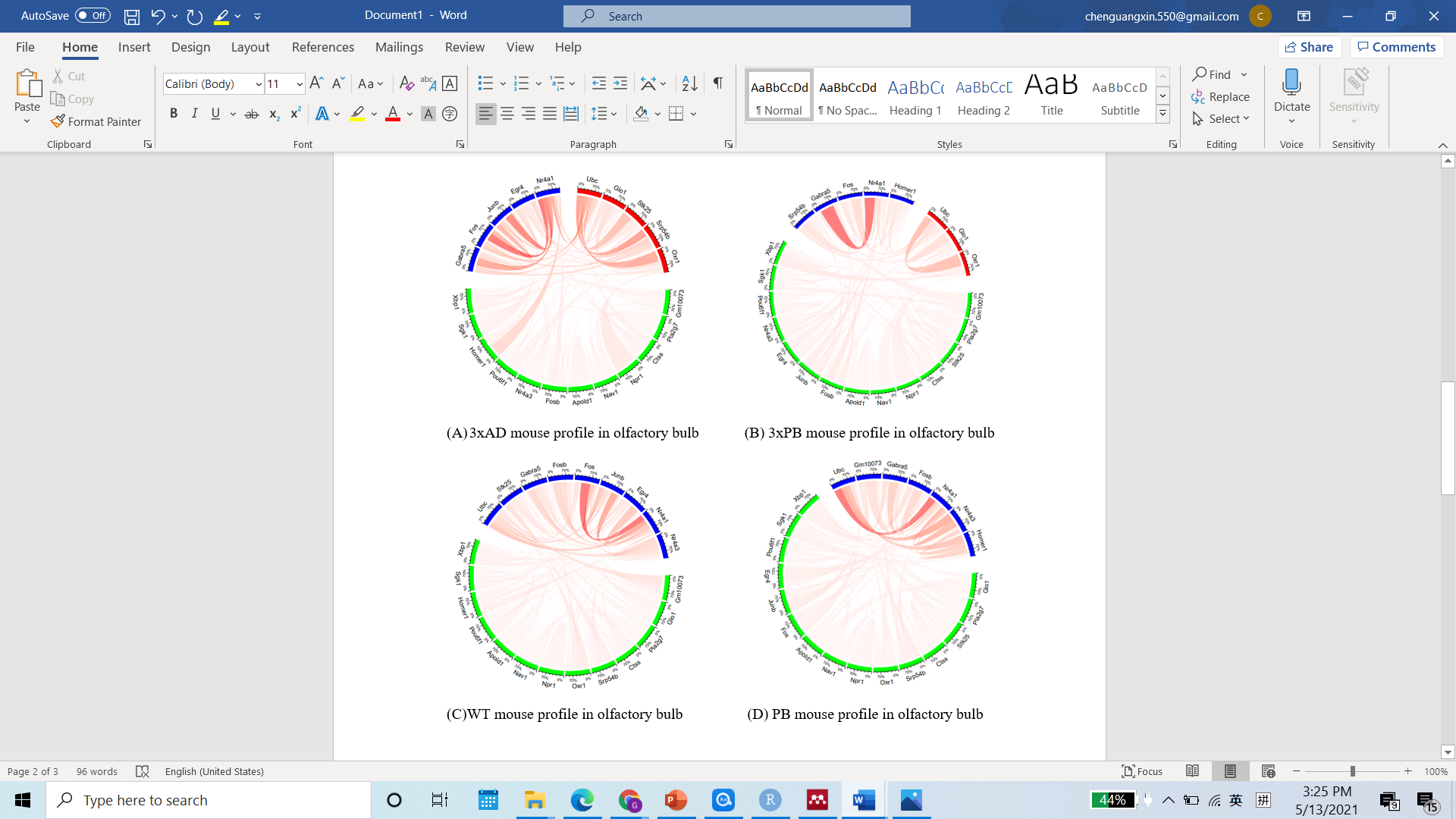
**Results**

The gene co-expression patterns in hippocampus were visualized using Circos plots, genes in the same module based on analysis of WGCNA were plotted together in same color. The links between each gene were plotted in red and the transparency of links indicating their connectivity strength. Generally, AD transgenic mice tended to have three expression modules while their control usually have two modules. The intermodular co-expression was more significant than intramodular expression in all plots. Specifically, for differential expressed genes in hippocampus, the transgenic mouse type 3xAD and 3xPB tended to have similar co-expression pattern. The top five co-expression gene pairs in 3xAD were: *Wbp11* and *Thy1*, *Lpl1* and *Thy1*, *Pkm* and *Thy1*, *Pkm* and *Cox6c*, *Pkm* and *Ubc*. And the top five co-expression pair in 3xPB were: *Pkm* and *Cox6c*, *Pkm* and *Thy1*, *Cox6c* and *Ubc*, *Cox6c* and *Glo1*, *Lpl* and *Wbp11*. While in WT, the top five pair were: *Pkm* and *Thy1*, *Pkm* and *Cox6c*, *Thy1* and *Cox6c*, *Cox6c* and *Ubc*, *Pkm* and *Ubc*. For PB mouse, such pairs were: *Pkm* and *Cox6c*, *Pkm* and *Thy1*, *Pkm* and *Ubc*, *Thy1* and *Cox6c*, *Cox6c* and *Gm10073* (Figure 1). It could be found that for differential expressed genes in hippocampus, genes *Pkm* and *Thy1* were crucial in the co-expression network, genes *Cox6c* and *Ubc* were also identified to be important in the co-expression pattern.



**Figure 1:** (A-D) The co-expression profile of selected differential expressed genes for four types of mouse in hippocampus. The transparency of links indicates the strength of connectivity, genes that are more strongly connected have darker color. Genes are clustered to different groups based on the cluster of WGCNA and those genes are plotted together.

Similarly, for differential expressed genes in olfactory bulb, regular genes pairs were also identified which were quite different with that of hippocampus. The top five co-expression gene pairs for 3xAD in olfactory bulb were: *Fos* and *Nr4a1*, *Gabra5* and *Nr4a1*, *Junb* and *Nr4a1*, *Ubc* and *Srp54b*, *Ubc* and *Ocr1*. In 3xPB, these pairs were: *Gabra5* and *Nr4a1*, *Ubc* and *Ocr1*, *Fos* and *Ubc*, *Ubc* and *Glo1*, *Fos* and *Nr4a1*. As for the WT mouse, those co-expression pairs were: *Fos* and *Nr4a1*, *Junb* and *Nr4a1*, *Egr4* and *Nr4a1*, *Ubc* and *Nr4a1*, *Ubc* and *Nr4a3*. And pairs for PB mouse were: *Ubc* and *Nr4a1*, *Ubc* and *Nr4a3*, *Ubc* and *Homer1*, *Nr4a1* and *Nr4a3*, *Nr4a1* and *Homer*. It could be found that genes *Nr4a1* and *Ubc* were crucial for all those co-expression pattern and the gene *Fos* was highly co-expressed with other genes in both types of AD transgenic mouse. The difference between major co-expressed genes in hippocampus and olfactory bulb indicating a differential co-expression pattern in both regions.



**Figure 2:** (A-D) The co-expression profile of selected differential expressed genes for four types of mouse in olfactory bulb. The transparency of links indicates the strength of connectivity, genes that are more strongly connected have darker color. Genes are clustered to different groups based on the cluster of WGCNA and those genes are plotted together.

**Discussions and Conclusions**

Based on the Circos plots shown above, we found some gene pairs that significantly standout in the co-expression network, and crucial genes were identified. It can be found that in hippocampus, *Pkm* and *Thy1* were crucial in the co-expression network of different mouse types, genes *Cox6c* and *Ubc* also had a high frequency and strength of connectivity. Among all these genes, only *Thy1* was identified to be upregulated in all brain region, others tended to show a trend of downregulated, which makes sense because *Thy1* was believed to be the promoter of expression for AD transgenes (Navarro et al., 2020). *Pkm* is short for protein kinase muscle, it encodes a pyruvate kinase which catalyzes the generation of ATP and pyruvate by transferring of the phosphoryl group from phosphoenolpyruvate to ADP. Also, Cox6c functions in the electron transport chain catalyzing electron transfer from reduced cytochrome c to oxygen. The downregulation of both genes can directly cause the deficiency of ATP generation; thus, neuron functions in hippocampus are likely to be impacted due to the deviation of energy production (Mosconi el al., 2008).

While in olfactory bulb, *Nr4a1* and *Ubc* were prevalently co-expressed in different mouse types and the gene *Fos* was highly co-expressed in both types of AD transgenic mouse, all genes were found to be downregulated in olfactory bulb (Navarro et al., 2020). *Ubc* encodes the protein that acting as the substrate for polyubiquitin reactions which related to degradation of toxic proteins like tau and amyloid . The decreasing of *Ubc* expression may promote the accumulation of tau and amyloid misfolding and inflammation, which can contribute to the pathological changes in olfactory bulb. *Nr4a1* and *Fos* are classified as the immediate-early genes that regulates intracellular signaling and synaptic plasticity, thus may also play crucial rules in the dysfunction of neural cells in olfactory bulb (Patel et al., 2019).

Based on these finding, we can readily conclude that hippocampus and olfactory bulb have different co-expression pattern for selected differentially expressed genes. The simultaneous downregulation of genes in hippocampus seems more related to energy deficiency in neural cells, while this pattern may relate more with protein misfolding and toxicity in olfactory bulb in AD mouse. While the overall pathological effect of AD can be a combination of all those factors, these results reveal potential spatially difference for pathological change in AD mouse brain.

Though the co-expression pattern for the selected genes could be identified directly from the Circos plots, some quantitative analysis is still available for further network summary and comparison. From the adjacent matrix constructed by WGCNA, it is always possible to identify the eigengenes of each network which may enable us to relate modules to each other and to clinical traits with SNPs. In fact, there is research used module eigengenes for different brain regions identified different regional vulnerability to AD (Wang el al. 2016). Also, the co-expression patterns can be found from different subregion of hippocampus or from different cell types. This is a promising direction due to the combination of emerging *in situ* sequencing technologies and network analysis methods such as WCGNA, and can greatly contribute to our understanding for spatial distribution of gene expression in AD.

**Reference**

1. Chen, W. T., Lu, A., Craessaerts, K., Pavie, B., Sala Frigerio, C., Corthout, N., Qian, X., Laláková, J., Kühnemund, M., Voytyuk, I., Wolfs, L., Mancuso, R., Salta, E., Balusu, S., Snellinx, A., Munck, S., Jurek, A., Fernandez Navarro, J., Saido, T. C., Huitinga, I., … De Strooper, B. (2020). Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell*, *182*(4), 976–991.e19. <https://doi.org/10.1016/j.cell.2020.06.038>
2. Hou, Y., Lautrup, S., Cordonnier, S., Wang, Y., Croteau, D. L., Zavala, E., Zhang, Y., Moritoh, K., O'Connell, J. F., Baptiste, B. A., Stevnsner, T. V., Mattson, M. P., & Bohr, V. A. (2018). NAD+ supplementation normalizes key Alzheimer's features and DNA damage responses in a new AD mouse model with introduced DNA repair deficiency. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(8), E1876–E1885. <https://doi.org/10.1073/pnas.1718819115>
3. Lane, C. A., Hardy, J., & Schott, J. M. (2018). Alzheimer's disease. *European journal of neurology*, *25*(1), 59–70. <https://doi.org/10.1111/ene.13439>
4. Mosconi, L., Pupi, A., & De Leon, M. J. (2008). Brain glucose hypometabolism and oxidative stress in preclinical Alzheimer's disease. *Annals of the New York Academy of Sciences*, *1147*, 180–195. <https://doi.org/10.1196/annals.1427.007>
5. Murphy C. (2019). Olfactory and other sensory impairments in Alzheimer disease. *Nature reviews. Neurology*, *15*(1), 11–24. <https://doi.org/10.1038/s41582-018-0097-5>
6. Navarro, J. F., Croteau, D. L., Jurek, A., Andrusivova, Z., Yang, B., Wang, Y., Ogedegbe, B., Riaz, T., Støen, M., Desler, C., Rasmussen, L. J., Tønjum, T., Galas, M. C., Lundeberg, J., & Bohr, V. A. (2020). Spatial Transcriptomics Reveals Genes Associated with Dysregulated Mitochondrial Functions and Stress Signaling in Alzheimer Disease. *iScience*, *23*(10), 101556. <https://doi.org/10.1016/j.isci.2020.101556>
7. Patel, H., Dobson, R., & Newhouse, S. J. (2019). A Meta-Analysis of Alzheimer's Disease Brain Transcriptomic Data. *Journal of Alzheimer's disease : JAD*, *68*(4), 1635–1656. <https://doi.org/10.3233/JAD-181085>
8. 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 medicine*, *8*(1), 104. <https://doi.org/10.1186/s13073-016-0355-3>
9. Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, *4*, Article17. <https://doi.org/10.2202/1544-6115.1128>