**Still figuring out a good title**

**Abstract (very rough so far)**

Alzheimer’s disease (AD) is a complex and heterogenous disease that still evades any effective pharmaceutical intervention. Animal models containing familial Alzheimer’s disease (fAD) mutations have often been used to understand pathogenesis of AD in humans, including the more common sporadic version of AD (sAD). The 5XFAD mouse model remains one of the most popular models used to study AD despite its genetic state being very different to familial Alzheimer’s disease. In this study, we compare 5XFAD mice brain transcriptomes to fAD-like zebrafish brain transcriptomes for the first time and find numerous differences in the young adult brains that suggest that 5XFAD brains may not be effective in modelling the early stages of AD, particularly aspects of AD related to energy metabolism. In addition, we also adapt a robust sparse PCA approach to perform cross-species analysis between aged brains of these animal models and human fAD and sAD brains. We find evidence that fAD-like zebrafish show similar gene expression changes to human fAD brains, unlike 5XFAD mice. We also find a large amount of heterogeneity in sAD brains compared to fAD brains, but a clear subset of sAD brains that resemble fAD brains. Taken together, our results support careful consideration of gene expression findings from animal models when applied to translational aspects of human AD, and emphasises the need to better characterise and understand the heterogeneity of sAD at the molecular level.

**Introduction**

Paragraph 1 – Introduction to Alzheimer’s Disease and the problem statement, which is that we still don’t understand early stages of AD, and that no one has really evaluated how effective animal models are in modelling these earliest stages.

Paragraph 2 – Studies that have looked at concordance between animal models (in particular mouse models) and AD (e.g. the Eric Blalock study) in aged brains found little similarity. Our lab has developed fAD-like knock-in models which resemble the genetic state of fAD more closely.

Paragraph 3 – Comparing gene expression between species is difficult. Matrix factorisation and dimension reduction approaches to reduce noise in the data appear to be effective for this.

Paragraph 4 – The main aim of this paper is to compare the young adult brain transcriptomes of 5XFAD mice and fAD-like zebrafish; and then compare the aged brain transcriptomes of both of these models to human fAD and sAD.

**Results**

1. Young adult fAD-like zebrafish brains (6 months) differ significantly from young adult 5XFAD mouse brains (3 months) **(Figure 1A)**
2. When comparing aged fAD-like zebrafish brains (24 months) and aged 5XFAD mouse brains (11-12 months), many biological activities show changes. This supports that dysregulation of these processes is likely to accumulate with age. The proportions of up and down-regulated genes differ, in addition to the actual genes involved, indicating that there are differences in the pathological state of aged fAD-like zebrafish and 5XFAD mouse brains, likely because of the differences at younger ages. **(Figure 1B)**
3. We find that using AES-PCA (adaptive elastic net sparse PCA) is a robust technique to compare and summarise gene set activity across different gene expression datasets, including across different species.
4. Using AES-PCA to compare fAD-like zebrafish to human fAD and 5XFAD mouse to human fAD. **(Figure 2)**
5. Using AES-PCA to compare human fAD and human sAD reveals that sAD is very heterogenous. There are many clusters representing different subsets of patients. fAD patients are similar to each other and appear to resemble a subset of the sAD patients. **(Figure 3)**

**Figures**

**AA screenshot of a cell phone

Description automatically generated**

**B A screenshot of a cell phone

Description automatically generated**

**Figure 1. Comparison of gene set enrichment results in fAD-like zebrafish whole brains and 5XFAD mouse cortex. A. Young brains from 6-month-old zebrafish and 3-month-old mice). B. Aged brains from 24-month-old zebrafish and 11-12-month-old mice.**

A screenshot of a cell phone

Description automatically generated

**Figure 2. Preservation of brain gene expression changes between AD-like models and human early-onset fAD**. Gene expression changes shown here are PC1 values calculated using AES-PCA and standardised through scaling by dataset-specific mean and standard deviation. Significant preservation was determined through linear mixed model (FDR-adjusted *p*-value < 0.05). Zebrafish values derived from whole-brain gene expression, mouse values derived from cortex gene expression, human values derived from posterior cingulate gene expression.

A close up of a map

Description automatically generated

**Figure 3. Principal Component Analysis of PC1 values of Hallmark gene sets from familial Alzheimer’s disease (fAD) and sporadic Alzheimer’s disease (sAD) patients**. Clusters 1-5 represent subsets of patients in the sAD dataset.

**Results / Discussion**

* **Discussion Point 1.** Brains of young 5XFAD mice model an entirely different pathological state compared to young fAD-like zebrafish. While mice brains show differences that are more inflammatory / immune response related, zebrafish show differences that are more metabolism related. There are some similarities (e.g. fatty acid metabolism, cholesterol homeostasis).
* **Discussion Point 2.** However, aged brains across both zebrafish and mouse show dysregulation of many more gene sets (although not exactly the same). This would indicate that more pathology accumulates with age, although the earliest start to the “cascade” could differ in these models. This has implications for studies using 5XFAD mice to model the early stages of AD.
* **Discussion Point 3.** Aged fAD-like zebrafish are more similar to human post-mortem fAD brains than 5XFAD mice. This would support that aged fAD-like zebrafish are able to model aspects of late fAD. This finding also makes sense as the genetic basis of 5XFAD is not similar genetically to fAD in humans.
* **Discussion Point 4.** fAD and sAD are overall similar in terms of gene expression in humans. However, sAD is significantly more heterogeneous and there is considerable variation in the pathology seen in both AD and control brains in the sAD dataset. In contrast, fAD is quite homogenous and consistent.
* **Limitations**
  + Different expression patterns in different cell types and tissues. Although there can be broad similarities across different tissue types we can’t necessarily make precise conclusions