**Do animal models capture molecular signatures of Alzheimer’s disease? Adaptive elastic-net sparse PCA for robust cross-species analysis of complex gene expression data**

**Abstract**

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 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**

Alzheimer’s disease (AD) is a complex neurodegenerative disease that still lacks any disease-modifying pharmaceutical intervention. The structural and metabolic alterations seen in the brains of young adults with familial AD (fAD) or ApoE mutations suggest that AD may progress for decades before noticeable symptoms appear. However, while brain pathology evident in the late stages of AD have been extensively characterised, the earliest pathological changes in the brain predisposing individuals to AD are still unclear and widely debated.

Much of our knowledge regarding the progression of molecular events in AD has been derived from animal models of AD, particularly transgenic mice possessing several familial Alzheimer’s disease (fAD)-causing mutations (e.g. 5XFAD). These transgenic mice develop amyloid pathology and tau aggregates thought to be characteristic of AD in humans. Unfortunately, interventions that prevent or reduce pathology in these mouse models have consistently failed to slow disease progression when applied to human AD. Moreover, the gene expression patterns seen in aged brains of transgenic mouse models showed little concordance with each other or with human AD (Hargis et al. 2018). Taken together, these findings suggest that some aspects of pathology in these models may not be directly translatable to human AD at the molecular level. More recently, knock-in mouse models have been gaining traction as they have a more similar genetic background to human fAD. While they show subtle amyloid pathology, aspects of these models are considered to be more translatable to human AD. In our previous work, we demonstrated the utility of knock-in fAD-like zebrafish models by showing that the earliest detectable gene expression changes in young adult brain included altered energy metabolism and iron homeostasis, consistent with changes seen in human AD.

Gene expression patterns in any knock-in fAD model and 5XFAD brains have not previously been compared in detail. However, our previous research saw preliminary evidence that the earliest detectable changes in 5XFAD mouse brains differed substantially to fAD-like (*psen1*Q96\_K97del/+) zebrafish brains. In particular, the earliest detectable gene expression changes in 5XFAD mouse brains included inflammatory and immune-related pathways, with energy metabolism not showing significant dysfunction until old age. This finding suggested that the pathological states present in aged fAD-like and 5XFAD brains likely were caused by different factors and highlighted the need for better characterisation of early stages of AD at the molecular level.

In this current work, we expand upon these findings by adapting a more rigorous and robust approach involving adaptive elastic-net Principal Component Analysis (AES-PCA) to perform cross-species analysis between fAD-like zebrafish, 5XFAD mouse, and human fAD brain gene expression datasets. AES-PCA is an unsupervised method for dimension reduction and its use with existing gene set databases has been successful in biological datasets, although it has not yet been used in the context of cross-species comparisons. We extend these results through comparisons of gene expression patterns in fAD and sporadic AD (sAD) brains, which are thought to be considerably more heterogeneous due to different environmental and genetic factors contributing to disease. Our results give support to the idea that the earliest detectable pathological changes in the brain differ in fAD-like and 5XFAD brains, and that this eventually leads to the aged brains showing distinct pathological states characterised by different alterations in biological activities. Interestingly, gene expression patterns in the aged brains of both of these models appear to have complementary utility in modelling different pathological changes seen in post-mortem human fAD brains. In addition, we show that while gene expression patterns in human fAD brains appear consistent, sAD brains display a large amount of heterogeneity, with only a subset of sAD brains displaying gene expression changes closely resembling the fAD brains and animal models. 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.

**Results**

**Hallmark gene sets capture differences in gene expression between fAD-like and 5XFAD brains.**

AES-PCA uses gene sets to reduce dimensions of a gene expression dataset. We intended to use AES-PCA with the Hallmark gene set collection from the Molecular Signatures Database (MSigDB), which covers 50 diverse biological activities summarised from existing studies/gene sets. Each Hallmark gene set is represented by 200 representative genes. To confirm that the Hallmark gene set collection would be able to accurately capture gene expression differences and similarities between the fAD-like and 5XFAD brains, we initially performed gene set enrichment testing using the approach described in Hin et al. (2020). We found significant enrichment (FDR-adjusted Wilkinson’s *p* < 0.05) across both age groups in both the fAD-like zebrafish and 5XFAD mice compared to their wild-type siblings (**Figure 1A**). The gene sets which were significantly enriched in the young adult age group (zebrafish: 6 months; mice: 3 months) were largely non-overlapping. Gene sets uniquely enriched in 5XFAD brains included “interferon alpha response”, “interferon gamma response”, “complement”, and “inflammatory response” while those uniquely enriched in fAD-like zebrafish brains included “glycolysis”, “oxidative phosphorylation” and “MTORC1 signalling”. Gene sets that were enriched in both included “fatty acid metabolism”, “heme metabolism”, and “protein secretion”. The enriched gene sets in young adult 5XFAD and fAD-like brains are consistent with known information about these models (cite Morgan’s Q96K97 paper here as well as 5XFAD paper showing progression of symptoms – inflammation at young age). This gave us the confidence that the subset of genes represented in the Hallmark gene set collection would be sufficient to capture the relevant biological signals within the datasets.

In the older age group (zebrafish: 24 months; mouse: 11-12 months), we see a strong age-dependent effect of the fAD-like mutation and 5XFAD mutations, with the majority of gene sets now being significantly enriched compared to wild-type siblings (FDR-adjusted Wilkinson *p* < 0.05) (**Figure 1B**). This makes sense in the context of pathological changes accumulating with age, eventually resulting in extensive dysfunction of diverse biological activities compared to healthy wild-type brains.

**Using adaptive, elastic-net principal component analysis (AES-PCA) to summarise molecular signatures**

**Nhi to remake Figure 1**

Initially, we looked for differences in the gene expression changes in fAD-like and 5XFAD brains when compared to their wild-type siblings at two different time points – when they were young adults (zebrafish: 6 months old; mice: 3 months old), and when they were aged (zebrafish: 24 months old; mice: 11-12 months old).

1. **Gene expression changes in young adult fAD-like and 5XFAD brains differ substantially**

Initially we sought to confirm that

1. **Differences between aged fAD-like and 5XFAD brains are less pronounced**

Fgfg

1. **Using adaptive elastic-net sparse PCA to perform cross-species analysis and summarise molecular signatures**