Frameshift and frame-preserving mutations in zebrafish *presenilin 2* affect different cellular functions in young adult brains

Karissa **Barthelson**[1*], Stephen Martin **Pederson**[2], Morgan **Newman**[1] Haowei **Jiang**[3], and Michael **Lardelli**[1]

1. Alzheimer's Disease Genetics Laboratory, School of Biological Sciences, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

2. Bioinformatics Hub, School of Biological Sciences, University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

3. School of Pharmacy, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China

* Corresponding author:

Karissa Barthelson: karissa.barthelson@adelaide.edu.au

Abstract

Background: Mutations in *PRESENILIN 2* (*PSEN2*) cause early disease onset familial Alzheimer's disease (EOfAD) but their mode of action remains elusive. One consistent observation for all *PRESENILIN* gene mutations causing EOfAD is that a transcript is produced with a reading frame terminated by the normal stop codon – the "reading frame preservation rule". Mutations that do not obey this rule do not cause the disease. The reasons for this are debated.

Methods: A frameshift mutation (*psen2^{N140fs}*) and a reading frame-preserving mutation (*psen2^{T141_L142delinsMISLISV}*) were previously isolated during genome editing directed at the N140 codon of zebrafish *psen2* (equivalent to N141 of human *PSEN2*). We mated a pair of fish heterozygous for each mutation to generate a family of siblings including wild type and heterozygous mutant genotypes. Transcriptomes from young adult (6 months) brains of these genotypes were analysed. Bioinformatics techniques were used to predict cellular functions affected by heterozygosity for each mutation.

Results: The reading frame preserving mutation uniquely caused subtle, but statistically significant, changes to expression of genes involved in oxidative phosphorylation, long term potentiation and the cell cycle. The frameshift mutation uniquely affected genes involved in Notch and MAPK signalling, extracellular matrix receptor interactions and focal adhesion. Both mutations affected ribosomal protein gene expression but in opposite directions.

Conclusion: A frameshift and frame-preserving mutation at the same position in zebrafish *psen2* cause discrete effects. Changes in oxidative phosphorylation, long term potentiation and the cell cycle may promote EOfAD pathogenesis in humans.

Key words

Alzheimer's disease, RNA-seq, zebrafish, PSEN2, mitochondria.

Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder which develops silently over decades. The pathological hallmarks of AD include the presence of extracellular senile neuritic plaques consisting, primarily, of amyloid β (Aβ) peptides, intracellular neurofibrillary tangles (primarily consisting of hyperphosphorylated tau proteins), and progressive neuronal loss (reviewed in [1]). The majority of therapeutics for AD are aimed at reducing Aβ levels (reviewed in [2]). However, all compounds to date show little or no effect on cognitive symptoms (reviewed in [3, 4]). This likely reflects our ignorance of the pathogenic mechanism underlying AD. Also, once cognitive changes occur, damage to the brain is considerable and may be irreversible. A comprehensive understanding of the early molecular changes/stresses occurring many years before disease onset is required to develop preventative treatments to reduce the prevalence of AD.

The majority of AD cases arise sporadically with an age of onset after 65 years (late onset AD, LOAD). Genome-wide association studies (GWAS) have identified at least 20 genes associated with increased risk for LOAD, with the most influential being the *APOLIPOPROTEIN E* (*APOE*) locus [5, 6]. Rare, familial forms of AD also exist. Early onset familial AD (EOfAD) cases have an age of onset before 65 years, and

arise due to mutations in the *PRESENILIN* genes (*PSEN1* and *PSEN2*), and the genes *AMYLOID B A4 PRECURSOR PROTEIN* (*AβPP*) and *SORTILIN-RELATED RECEPTOR 1* (*SORL1*) (reviewed in [7-9]).

PSEN2 is the less commonly mutated *PSEN* gene in EOfAD. To date, only 13 pathogenic mutations have been described in *PSEN2* compared to over 185 pathogenic mutations in the homologous gene *PSEN1* [10]. The first *PSEN2* EOfAD mutation characterised was N141I, found in Volga German families and affecting the second transmembrane domain of the PSEN2 protein [11]. We previously attempted to introduce this mutation into zebrafish by genome editing. While we were not successful at introducing an exact equivalent of the human N141I mutation into zebrafish, we did isolate a reading frame-preserving mutation at this position: *psen2$^{T141\_L142delinsMISLISV}$* [12]. This mutation alters two codons and inserts five additional novel codons and was predicted to largely preserve the transmembrane structure of the resultant protein. We also isolated a frameshift mutation at the same position: *psen2^{N140fs}*. This is a deletion of 7 nucleotides resulting in a premature stop codon at novel codon position 142 [12]. Since all EOfAD mutations in the *PSENs* follow a "reading frame preservation rule" [13], *psen2$^{T141\_L142delinsMISLISV}$* models such a mutation in *PSEN2*. In contrast, the *psen2^{N140fs}* allele is predicted to express a truncated protein, is not EOfAD-like, and so can act as a negative control to investigate the differences between reading frame-preserving and -destructive mutations of *PSEN2*.

The overall goal of the Alzheimer's Disease Genetics Laboratory has been to identify the cellular processes affected in common by EOfAD mutations in different genes in young adult, heterozygous mutant brains (i.e. to establish an early EOfAD brain transcriptomic signature). We exploit zebrafish as a model organism for this work as

4

the EOfAD genes are conserved in zebrafish, and large families containing both EOfAD-like mutant fish and their wild type siblings can be generated from single mating events between single pairs of fish (reviewed in [14]). Analysing the transcriptomes of sibling fish raised in the same environment (aquaria within a single recirculated water system) reduces genetic and environmental sources of variation, (noise) and allows subtle effects of mutations to be identified. After introducing EOfAD-like mutations into the endogenous, zebrafish orthologues of EOfAD genes (i.e. knock-in models), we analyse their effects on brain transcriptomes in heterozygous fish to mimic the genetic state of such mutations in the human disease. We term this experimental strategy Between Sibling Transcriptome (BeST) analysis. We have previously performed BeST analyses for EOfAD-like mutations in *psen1* [15] and *sorl1* [16, 17] and for a complex (but possibly EOfAD-like) mutation in *psen2*: $psen2^{S4Ter}$ [18].

In the work described in this report, we performed a BeST analysis using a family of sibling fish generated by mating two fish with genotypes $psen2^{T141\_L142delinsMISLISV}$/+ and $psen2^{N140fs}$/+. The family included fish heterozygous for the mutations $psen2^{T141\_L142delinsMISLISV}$ (for simplicity, hereafter referred to as "EOfAD-like") and $psen2^{N140fs}$ (for simplicity, hereafter referred to as "FS") and their wild type siblings, all raised together in the same tank until 6 months of age. This strategy allowed direct comparisons of the brain transcriptomes from the two *psen2* mutant genotypes to their wild type siblings while reducing confounding effects from genetic and environmental variation. Despite this, considerable gene expression variation (noise) is still observed between individuals and the effects of the mutations are subtle so that few genes are identified initially as differentially expressed. However, when analysed at the pathway level, heterozygosity for each mutation shows distinct

5

transcriptomic effects. The sets of genes involved in oxidative phosphorylation, long-term potentiation, and the cell cycle are only altered significantly in the EOfAD-like brains, while the FS mutant brains shows apparent changes in Notch and MAPK signalling, focal adhesion, and ECM receptor interaction. Both mutation states affected genes encoding ribosomal subunits but the effect of each mutation was in opposite directions. Our results are consistent with the concept that only dominant, reading frame-preserving mutations of the *PRESENILIN* genes cause EOfAD and support that effects on oxidative phosphorylation may be a common signature of such mutations.

Methods

Zebrafish lines and ethics statement.

Generation of the *psen2^{T141_L142delinsMISLISV}* and *psen2^{N140fs}* lines of zebrafish used in this study is described in [12]. All experiments involving zebrafish were conducted under the auspices of the Animal Ethics Committee of the University of Adelaide, permit numbers S-2017-089 and S-2017-073.

Breeding strategy

Two fish heterozygous for these mutations were mated to generate a family of sibling fish with *psen2* genotypes EOfAD-like/+, FS/+, EOfAD-like/FS (transheterozygous) or wild type (**Figure 1A**). This family was raised together in the same tank until 6 months of age, at which time 24 fish were randomly selected and sacrificed in a loose ice slurry (to allow for n = 5 of each genotype in the RNA-seq analysis). The heads of these fish were removed and stored in 600 µL of RNAlater solution, while

6

their tails were removed for genomic DNA extraction and genotyping by polymerase chain reactions (PCRs).

PCR genotyping

Allele-specific genotyping PCRs were performed on genomic DNA extracted from fin biopsies as described in [17]. The allele-specific primer sequences can be found in **Table 1**. Primers were synthesised by Sigma Aldrich (St. Louis, MO, USA).

Table 1: Genotyping primer sequences.	
Primer Name	**Sequence (5' to 3')**
Wild type specific forward primer	TGAATTCGGTGCTCAACACTC
T141_L142delinsMISLISV specific forward primer	TGAATTCGGTGCTCAACATG
N140fs specific forward primer	TGCTGAATTCGGTGCTCTG
Common reverse primer	TCACCAAGGACCACTGATTCAGC

RNA-seq data generation and analysis

Preparation of RNA for RNA-seq and the subsequent analysis of the data was performed mostly as in [17]. Briefly, brains of n = 5 fish of wild type, EOfAD-like/+ and FS/+ genotypes (each group contained 3 females and 2 males) were carefully removed from the heads preserved RNAlater Stabilization Solution (Invitrogen, Carlsbad, CA, USA), followed by total RNA extraction using the mirVana™ miRNA Isolation Kit (Ambion, Austin, TX, USA) and DNase treatment using the DNA-free™ Kit (Ambion). Total RNA was then delivered to the Genomics service at the South

7

Australian Health and Medical Research Institute (SAHMRI, Adelaide, AUS) for polyA+, stranded library preparation and RNA sequencing using the Illumina Nextseq platform.

Processing of the demultiplexed fastq files provided by SAHMRI (75 bp single-end reads) was performed as in [17]. However, the pseudo-alignment using *kallisto* (v0.43.1) [19] was performed using a modified version of the zebrafish transcriptome (Ensembl release 94) which additionally included the two novel mutant *psen2* transcript sequences.

Transcript counts as estimated by *kallisto* were imported into *R* [20] using *tximport* [21], summing the transcript-level counts to gene-level counts. We also imported the transcript-level counts using the *catchKallisto* function from the package *edgeR* [22] to assess simply the allele-specific expression of the different *psen2* transcripts.

We removed genes considered as undetectable (less than one count per million in at least five of the samples) from downstream analysis, leaving library sizes ranging between 17 million and 27 million reads. We normalised for these differences in library sizes using the trimmed mean of M-values (TMM) method [23], followed by removal of one factor of unwanted variation using the *RUVg* method *RUVseq* [24] as described in [17].

Differential gene expression analysis was performed using a generalised linear model (GLM) and likelihood ratio test using *edgeR* [22], specifying a design matrix with an intercept as the wild type genotype, and the coefficients as the FS/+ and EOfAD-like/+ genotypes, and the W_1 coefficient from *RUVg*. Genes were considered differentially expressed (DE) in each specific comparison if the FDR-adjusted p-value was below 0.05.

Enrichment analysis was performed on the KEGG and GO gene sets from the Molecular Signatures Database (MSigDB) [25], with GO terms excluded if the shortest path back to the root node was < 3 steps. We also tested for possible iron dyshomeostasis in the RNA-seq data by performing an enrichment analysis on the iron-responsive element (IRE) gene sets described in [26]. To test for over-representation of these gene sets within the DE gene lists, we used *goseq* [27], using the average transcript length per gene to estimate sampling bias for DE genes. For enrichment testing on the entire list of detectable genes, we calculated the harmonic mean p-value [28] of raw p-values from *fry* [29], *camera* [30] and *fgsea* [31, 32]. Gene sets were considered significantly altered if the FDR-adjusted harmonic mean p-value was below 0.05. Visualisation of RNA-seq data analysis was performed using *ggplot2* [33], *pheatmap* [34] and *UpSetR* [35].

Reproducibility and data availability

The raw fastq files and the gene-level counts have been deposited in the GEO database with accession number GSE158233. All code to reproduce this analysis can be found at

https://github.com/UofABioinformaticsHub/20181113_MorganLardelli_mRNASeq.

Results

We first inspected the transcript-level counts to ensure that expression of the *psen2* alleles was as expected (**Additional File 1**). Consistent with observations in [12], expression of the EOfAD-like transcript was at a similar level to that of the wild type *psen2* transcript in the EOfAD-like/+ brains, while the decreased expression of the FS allele of *psen2* relative to the wild type allele was suggestive of nonsense-

mediated decay (NMD). We observed that one of the samples had been incorrectly

genotyped during sample preparation and was actually transheterozygous for the

EOfAD-like and FS alleles of *psen2*. Therefore, we omitted this sample from

subsequent analyses. We also examined the relationship between samples by

principal component analysis (PCA) on the gene-level counts. Samples did not

cluster by genotype in the PCA plot, supporting that heterozygosity for the *psen2*

mutations in this study does not give widespread effects on the brain transcriptome.

Additionally, samples did not cluster by sex, consistent with our previous

observations that sex does not show extensive effects in zebrafish brain

transcriptomes [16, 17, 36]. In the PCA plot, we noticed that one of the wild type

brain samples (12_WT_4) separated greatly from the others, appearing to be an

outlier (**Additional File 1**). Sample weights, as calculated using the

voomWithQualityWeights algorithm from the *limma* package [37] on the gene-level

counts, revealed that this wild type sample was highly down-weighted. Therefore, we

also omitted this sample from subsequent analyses.

Heterozygosity for an EOfAD-like or an FS mutation in *psen2* results in limited differential expression of genes

To identify which genes were dysregulated due to heterozygosity for the EOfAD-like

or FS alleles of *psen2,* we performed differential gene expression analysis using a

GLM and likelihood ratio test with *edgeR*. This revealed 32 differentially expressed

(DE) genes due to the FS mutation, and 5 DE genes due to the EOfAD-like mutation

relative to wild type (**Figure 1, Additional File 2**). Only two genes were seen to be

significantly upregulated in both comparisons relative to wild type: *AL929206.1* and

BX004838.1. However, these genes are currently not annotated and have no known

function. We tested for over-representation within the DE genes using *goseq* (using

the average transcript length per gene as the predictor variable) of gene ontology (GO) terms which, as the name suggests, use ontologies to annotate gene function; KEGG gene sets which can give insight into changes in various biological pathways and reactions; and our recently defined IRE gene sets [26] which can give insight to possible iron dyshomeostasis. However, due to the low numbers of DE genes in both comparisons, no significantly over-represented GO terms or gene sets were identified. (For the top 10 most significantly over-represented GO terms and gene sets in the DE genes, see **Additional File 3**).
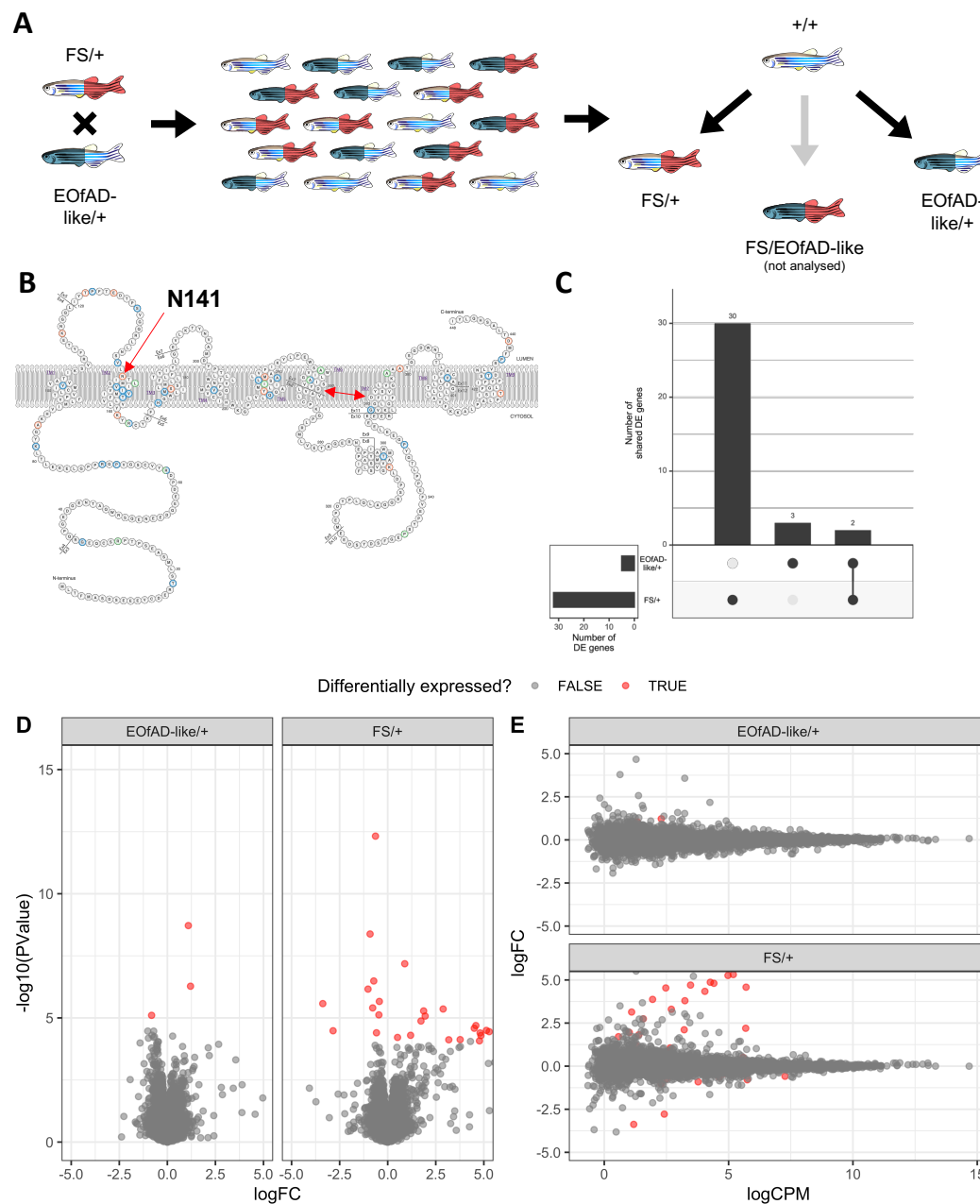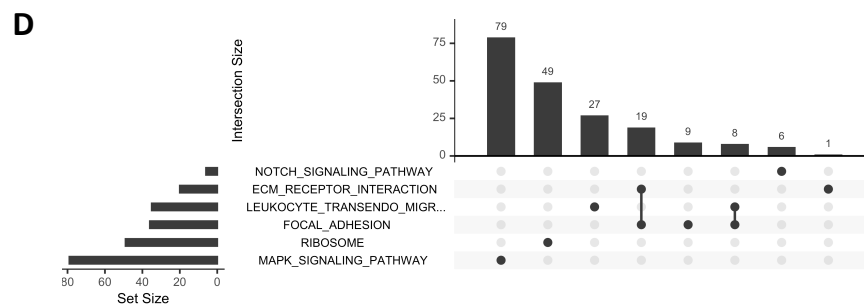
Figure 1:

A) Mating strategy and experimental design. **B)** Schematic of the human PSEN2 protein (adapted from https://www.alzforum.org/mutations/psen-2 with permission from FBRI LLC (Copyright © 1996–2020 FBRI LLC. All Rights Reserved. Version 2.7 – 2020)). Amino acid residues are color-coded to indicate whether substitutions have been observed to be pathogenic (orange), non-pathogenic (green) or of uncertain pathogenicity (blue). The site of the human EOfAD mutation N141I, in the second transmembrane domain (TMD) is indicated by the red single-headed arrow. The aspartate residues critical for γ-secretase catalysis are indicated by a red double-headed arrow. **C)** Upset plot indicating the number of differentially expressed (DE) genes in each comparison of the *psen2* mutant

genotypes to wild type noting that only 2 genes appear to be DE in common between both comparisons. **D)** Mean-difference (MD) plot and **E)** volcano plot of DE genes in each comparison. Note that the logFC axis limits in D) and E) are constrained to -5 and 5 for visualisation purposes.

Heterozygosity for an EOfAD-like or an FS mutation in *psen2* has distinct effects on biological processes in young adult zebrafish brains

We next performed enrichment analysis on the entire list of detectable genes in the RNA-seq experiment to obtain a more complete view on the changes to gene expression due to heterozygosity for the EOfAD-like or FS mutations in *psen2.* Due to the highly overlapping nature of GO terms (many GO terms include many of the same genes), we tested only the KEGG and IRE gene sets. We found statistical evidence for 8 KEGG gene sets in the EOfAD/+ brains and 6 KEGG gene sets in the FS/+ brains to be significantly altered as a group. Notably, no IRE gene sets were found to be significantly altered (**Figure 2A**). Genes in the KEGG_RIBOSOME gene set were significantly altered in both comparisons. However, the genes were mostly downregulated in EOfAD-like/+ brains and upregulated in FS/+ brains (**Figure 2B**). For additional visualisations of our enrichment analysis, see **Additional File 4**.

To determine whether the statistical significance of the enrichments of the gene sets was being driven by the expression of the same genes, we inspected the genes from the "leading edge" from *fgsea*. (These can be thought of as the core genes driving the enrichment of each gene set). The leading edge genes in each comparison were mostly independent of one another, except for the KEGG gene sets for oxidative phosphorylation, Parkinson's disease and Alzheimer's disease, which share 39 genes (**Figure 2C and D**).

Figure 2:

A) Heatmap showing the significant KEGG and IRE gene sets in *psen2* heterozygous mutant brains, clustered based on their Euclidean distance. The columns represent the gene sets which are significant in at least one comparison, and the rows are the two comparisons. The colour of each cell represents the significance value with more significant gene sets appearing lighter. The numbers within each cell are the FDR-adjusted harmonic mean p-values (HMP). ns, not significant. **B)** Heatmap indicating the logFC of genes in the KEGG_RIBOSOME gene set in EOfAD/+ and FS/+

brains. **C)** Upset plot showing the overlap of leading edge genes from the *fgsea* algorithm in gene sets found to be altered in EOfAD-like/+ brains, and **D)** FS/+ brains.

Discussion

Here, we aimed to identify similarities and differences in the effects on brain transcriptomes due to heterozygosity for an EOfAD-like mutation compared to a non-EOfAD-like mutation in *psen2*. To this end we used our BeST strategy, involving analysis of sibling fish raised together in an identical environment. BeST analysis reduces confounding effects in direct comparisons of two heterozygous *psen2* mutations relative to wild type, allowing detection of subtle transcriptome state differences.

We observed subtle, but highly statistically significant, changes to the transcriptomes of the *psen2* mutant fish relative to wild type. At the single gene level, relatively few genes were observed to be statistically significantly DE in each mutant genotype, and differentially enriched biological pathways were not detected in these DE gene lists (**Additional File 3**). As a comparison, heterozygosity for an EOfAD-like mutation in zebrafish *psen1,* (orthologous to the human gene most commonly mutated in EOfAD [38]), gives 251 DE genes in young adult zebrafish brains [15]. However, we were able to identify statistically significant changes in the *psen2* mutant brains by performing enrichment analysis on the entire list of detectable genes in the RNA-seq experiment using an ensemble approach. This involves combining raw p-values by calculation of a harmonic p-value – a strategy thought to be less restrictive than other methods of combining p-values. An additional benefit of this method is that it has been shown specifically to be robust to dependent p-values [28]. The subtlety of the *psen2* mutation effects is consistent with a generally later age of onset, and

variable penetrance of EOfAD mutations in *PSEN2* [39-41] relative to EOfAD mutations in *PSEN1* and *APP*.

Interestingly, the pathways which were affected by the two *psen2* mutations showed very little overlap. Gene sets which were significantly altered in EOfAD-like/+ brains were not significantly altered in the FS/+ brains and vice versa. Only one biological process appears to be significantly affected by both mutations: ribosome function (the KEGG_RIBOSOME gene set). However, in general, the direction of differential expression of genes in this gene set was observed to be opposite in the two mutants, implying different mechanisms of action of each mutation (**Figure 2B**). The low concordance between the brain transcriptome effects of these two mutations supports that presenilin EOfAD mutations do not act through a simple loss of function but, instead, through a gain of function mechanism connected with encoding an abnormal, but "full-length" protein [13]. Notably, the EOfAD-like mutation of *psen2* is predicted to affect the process of long term potentiation thought to be critical for memory formation. We also observed the dysregulation of genes involved in controlling the cell cycle. Genes driving the enrichment of this gene set (i.e. the leading edge genes) include the *minichromosome maintenance* (*mcm*) genes (see **Additional File 4**), which we observed to be significantly dysregulated due to heterozygosity for an EOfAD-like mutation in *psen1* in 7 day old zebrafish larvae [42]. Together, these results provide support for the "two-hit hypothesis" of Mark Smith and colleagues [43], which postulates that "cell cycle events" (CCEs) involving inappropriate attempts at cell division by neurons are critical for development of AD. The second "hit" of this hypothesis is oxidative stress, which is an anticipated outcome of disturbance of oxidative phosphorylation. Notably, the EOfAD-like mutation of *psen2* is predicted to affect oxidative phosphorylation, which is in

16

common with all the other EOfAD-like mutations of genes we have examined in zebrafish [16, 18, 26, 44].

One of the most characterised functions of presenilin proteins is their role as the catalytic core of gamma-secretase (γ-secretase) complexes. γ-secretase activity is responsible for cleavage of a wide range of type I transmembrane proteins including AβPP and NOTCH. We observed that genes involved in the Notch signalling pathway were only significantly altered in FS/+ genotype brains, indicating that γ-secretase activity may be altered by the FS mutation. Interestingly, inspection of the direction of changes to gene expression in this pathway's gene set suggest an increase in γ-secretase activity (**Additional File 4**), as *notch*, *delta* and, particularly, *hes* genes (that are Notch signalling targets) are all observed as upregulated. This seems counterintuitive, as the FS mutation results in a truncated protein lacking critical aspartate residues needed for γ-secretase activity (**Figure 1B**). A minor, (although statistically non-significant) increase in expression of *psen1* is observed in FS/+ brains (also in EOfAD-like/+ brains) (**Additional File 4**) raising the possibility that transcriptional adaptation (formerly known as "genetic compensation") might explain an apparent increase in γ-secretase activity [45]. Since presenilin holoproteins are known to form multimeric complexes, we have suggested previously that this may be involved in the regulation of the conversion of the holoprotein into its γ-secretase-active form by endoproteolysis [13]. N-terminal fragments of presenilin proteins are also known to multimerise [46] and the possibility exists that these may disrupt holoprotein multimer formation/stability. In our previous publication identifying the EOfAD-like and FS mutations of *psen2* analysed here, we saw that both mutations reduced adult melanotic surface pigmentation when homozygous [12]. This implied that both mutations reduce γ-secretase activity, at least in

melanosomes. These apparently conflicting observations support that analysis of brain biochemistry in these EOfAD-like and FS zebrafish *psen2* mutants may reveal important details of the relationship between presenilin holoproteins and their γ-secretase-active endoproteolysed forms including between the holoproteins encoded by *PSEN1* and *PSEN2*.

Conclusion

In conclusion, our results support that frameshift and reading frame-preserving mutations in the *presenilins* have distinct effects on the brain transcriptome, consistent with the reading frame preservation rule of presenilin EOfAD-causative mutations. The data presented here, along with our growing collection of BeST analyses, indicates that changes to mitochondrial and ribosomal functions are effects-in-common of heterozygosity for EOfAD-like mutations in different genes. These may be early cellular stresses which eventually lead to AD pathology, and warrant investigation for discovery of therapeutic targets.

Declarations
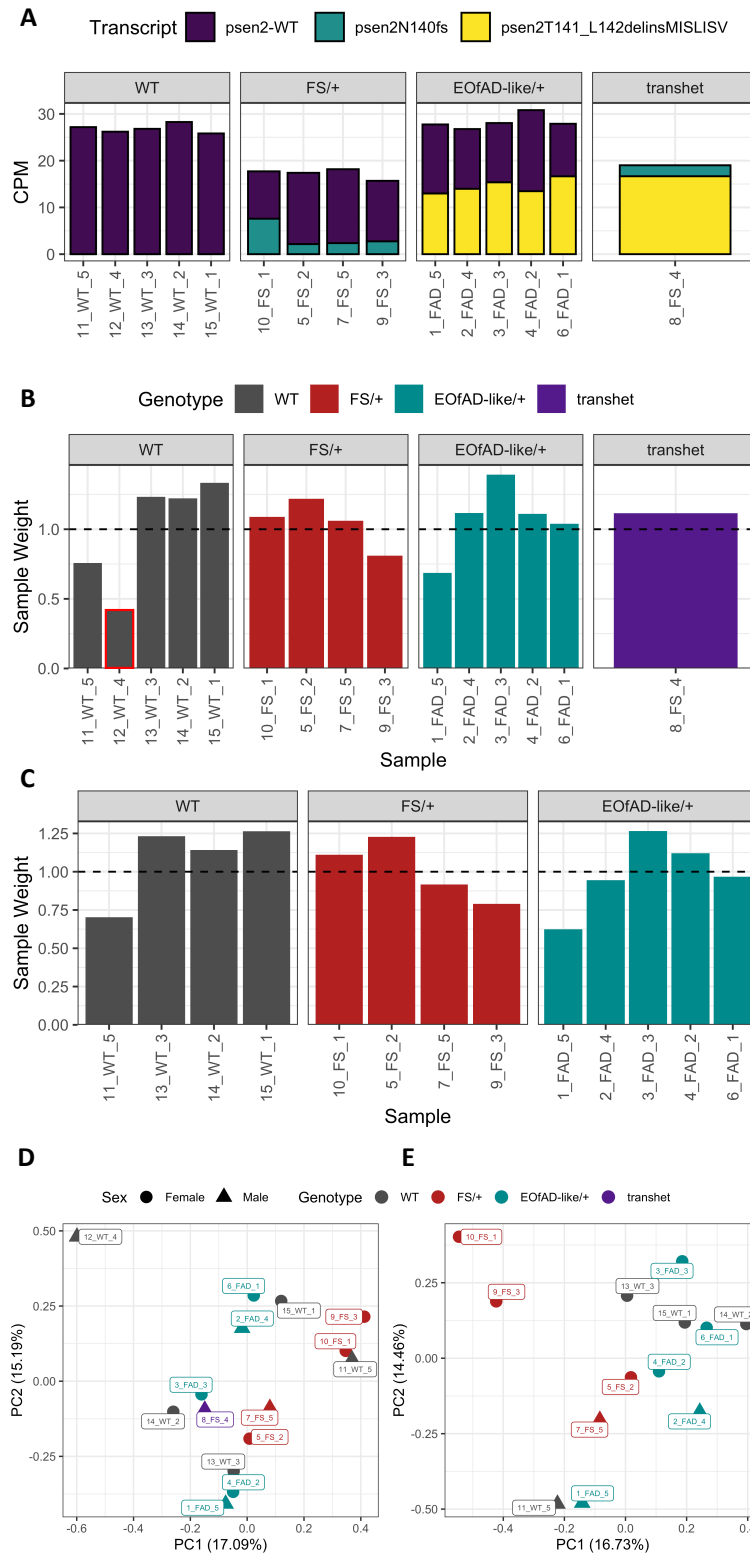
The authors have no conflict of interest to declare.

Acknowledgements

Additional Files

Additional File 1

A. Allele-specific expression (in counts per million, CPM) of *psen2* transcripts in young adult zebrafish brains. Reduced expression of the *psen2^N140fs* allele is observed in FS/+ brains, consistent with our previous observation that transcripts of this allele are subject to nonsense-mediated decay. Sample 8_FS_4 does not express the wild type allele of *psen2* and is actually a transheterozygous (transhet) sample. Therefore, it was omitted from the analysis. **B.** Sample weights as calculated by the *voomWithQualityWeights* algorithm on all samples sequenced. Sample 12_WT_4 is highly downweighted relative to all other samples and was omitted from subsequent analysis. **C.** Sample weights recalculated after exclusion of samples 8_FS_4 and 12_WT_4. **D.** Principal component 1 (PC1) against PC2 from a principal component analysis (PCA) on all samples of the experiment. Sample 12_WT_4 does not cluster with the other samples. **E.** PCA plot of the *RUVseq-*normalised counts after exclusion of samples 8_FS_4 and 12_WT_4.

Additional File 2: Full results of differential gene expression analysis.

Additional File 3: Over-representation analysis using goseq.

Table 1: The KEGG, GO and IRE gene sets approaching being over-represented in the DE genes list for the FS mutation. The 5 DE genes due to the EOfAD-like mutations are not found in any of the gene sets. Therefore, the results of the over-representation analysis are not shown.

Gene set	p-value	Number of DE genes in gene set	Number of genes in gene set	FDR adjusted p-value
GO_PHOSPHOTRANSFERASE _ACTIVITY_NITROGENOUS_G ROUP_AS_ACCEPTOR	1.94E-05	2	5	0.28952
GO_TRNA_3_END_PROCESSI NG	6.94E-05	2	9	0.517902

GO_NUCLEOSIDE_DIPHOSPHATE_METABOLIC_PROCESS	1.01E-03	3	142	1
KEGG_ALZHEIMERS_DISEASE	1.11E-03	3	146	1
GO_NCRNA_3_END_PROCESSING	1.39E-03	2	39	1
GO_PYRIMIDINE_NUCLEOBASE_TRANSPORT	2.85E-03	1	2	1
GO_UREA_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	2.85E-03	1	2	1
GO_TRNA_BINDING	3.58E-03	2	63	1
GO_GLYCEROL_CHANNEL_ACTIVITY	4.26E-03	1	3	1
GO_UREA_TRANSPORT	4.27E-03	1	3	1

Additional File 4: Additional RNA-seq visualisations

A. Pathview [47] visualisation of the logFC of genes in the KEGG_NOTCH_SIGNALLING_PATHWAY gene set in FS/+ brains. **B.** Expression of *psen1* in log counts per million (logCPM). **C.** Pat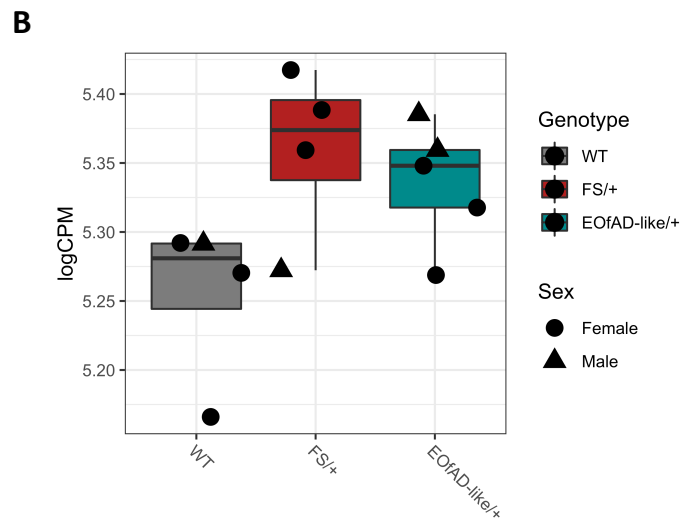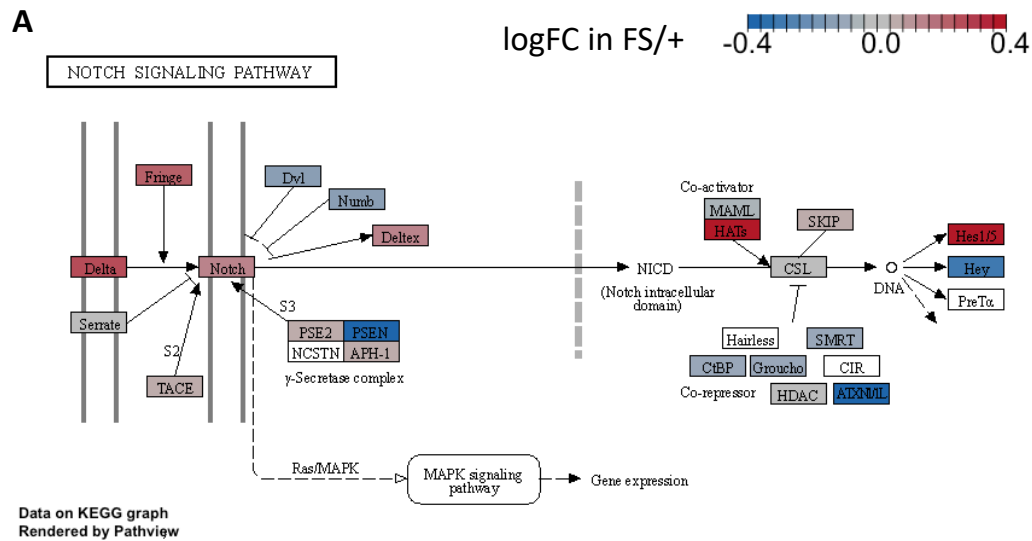hview [47] visualisation of the logFC of genes in the KEGG_OXIDATIVE_PHOSPHORYLATION gene set in EOfAD-like/+ brains. **D.** Pathview [47] visualisation of the logFC of genes in the KEGG_CELL_CYCLE gene set in EOfAD-like/+ brains. **E.** Heatmap of the logFC of genes in the KEGG_LONG_TERM_POTENTIATION gene set in EOfAD-like/+ brains.
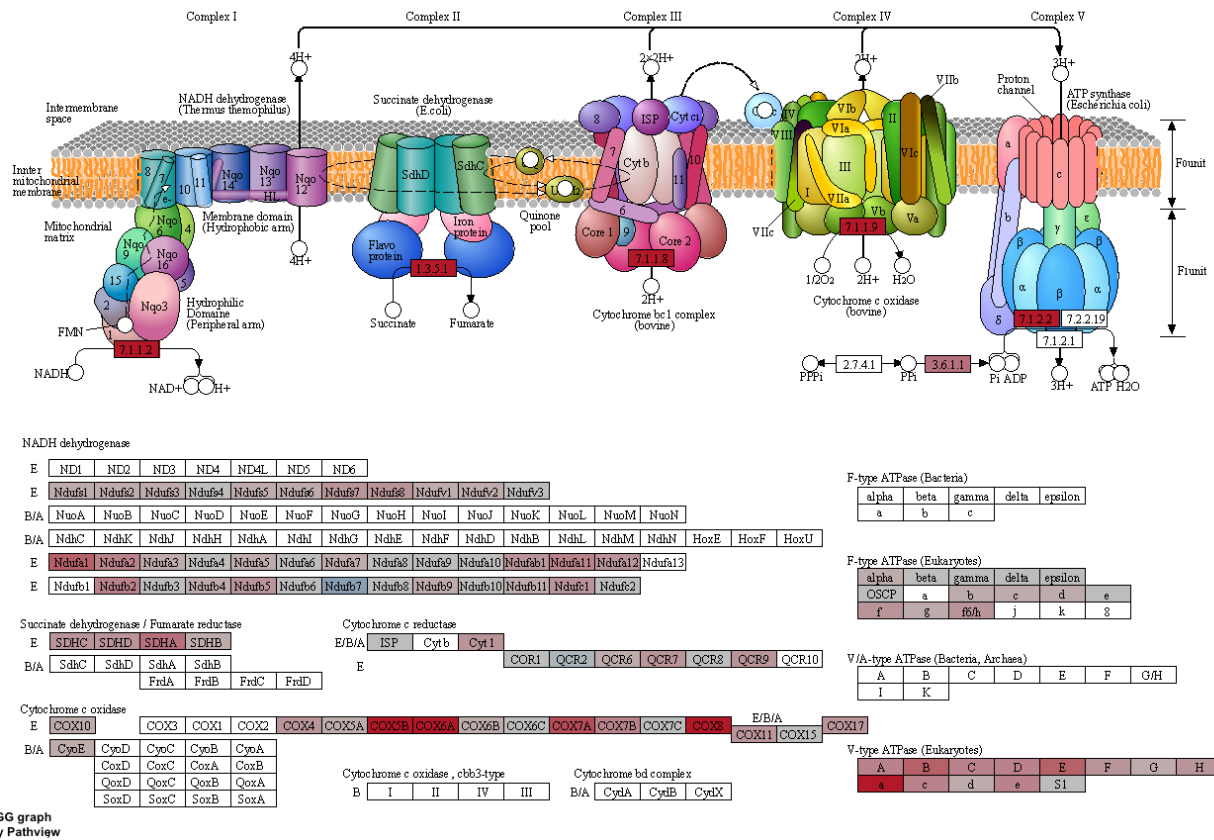
C

D

E

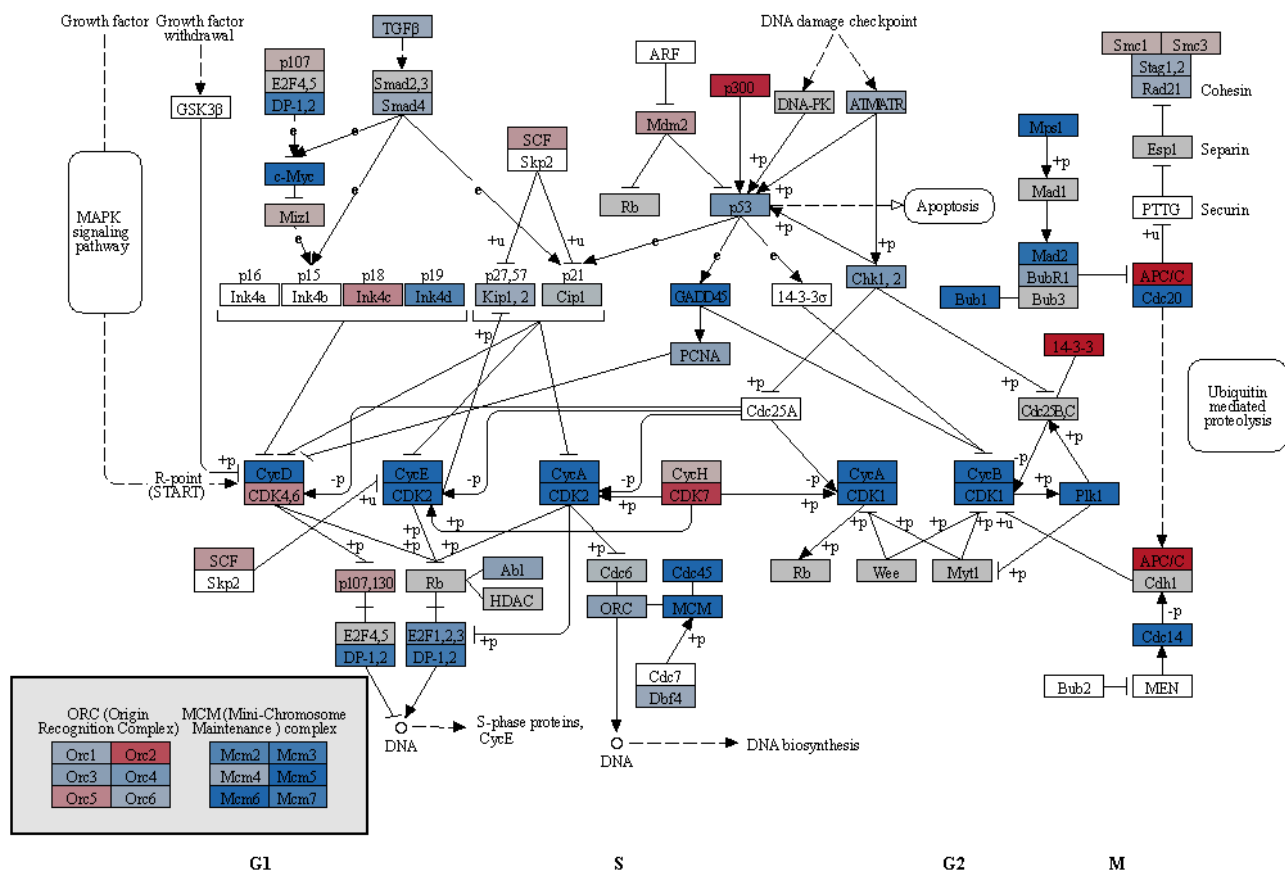KEGG_LONG_TERM_POTENTIATION

References

[1] Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL (2015) Alzheimer's disease. *Nature Reviews Disease Primers* **1**, 15056.

[2] Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K (2019) Alzheimer's disease drug development pipeline: 2019. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **5**, 272-293.

[3] Elmaleh DR, Farlow MR, Conti PS, Tompkins RG, Kundakovic L, Tanzi RE (2019) Developing Effective Alzheimer's Disease Therapies: Clinical Experience and Future Directions. *Journal of Alzheimer's Disease* **71**, 715-732.

[4] Modrego P, Lobo A (2019) A good marker does not mean a good target for clinical trials in Alzheimer's disease: the amyloid hypothesis questioned. *Neurodegenerative Disease Management* **9**, 119-121.

[5] Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, Van Der Lee SJ, Amlie-Wolf A, Bellenguez C, Frizatti A, Chouraki V, Martin ER, Sleegers K, Badarinarayan N, Jakobsdottir J, Hamilton-Nelson KL, Moreno-Grau S, Olaso R, Raybould R, Chen Y, Kuzma AB, Hiltunen M, Morgan T, Ahmad S, Vardarajan BN, Epelbaum J, Hoffmann P, Boada M, Beecham GW, Garnier J-G, Harold D, Fitzpatrick AL, Valladares O, Moutet M-L, Gerrish A, Smith AV, Qu L, Bacq D, Denning N, Jian X, Zhao Y, Del Zompo M, Fox NC, Choi S-H, Mateo I, Hughes JT, Adams HH, Malamon J, Sanchez-Garcia F, Patel Y, Brody JA, Dombroski BA, Naranjo MCD, Daniilidou M, Eiriksdottir G, Mukherjee S, Wallon D, Uphill J, Aspelund T, Cantwell LB, Garzia F, Galimberti D, Hofer E, Butkiewicz M, Fin B, Scarpini E, Sarnowski C, Bush WS, Meslage S, Kornhuber J, White CC, Song Y, Barber RC, Engelborghs S, Sordon S, Voijnovic D, Adams PM, Vandenberghe R, Mayhaus M, Cupples LA, Albert MS, De Deyn PP, Gu W, Himali JJ, Beekly D, Squassina A, Hartmann AM, Orellana A, Blacker D, Rodriguez-Rodriguez E, Lovestone S, Garcia ME, Doody RS, Munoz-Fernadez C, Sussams R, Lin H, Fairchild TJ, Benito YA, Holmes C, Karamujić-Čomić H, Frosch MP, Thonberg H, Maier W, Roshchupkin G, Ghetti B, Giedraitis V, Kawalia A, Li S, Huebinger RM, Kilander L, Moebus S, Hernández I, Kamboh MI, Brundin R, Turton J, Yang Q, Katz MJ, Concari L, Lord J, Beiser AS, Keene CD, Helisalmi S, Kloszewska I, Kukull WA, Koivisto AM, Lynch A, Tarraga L, Larson EB, Haapasalo A, Lawlor B, Mosley TH, Lipton RB, Solfrizzi V, Gill M, Longstreth WT, Montine TJ, Frisardi V, Diez-Fairen M, Rivadeneira F, Petersen RC, Deramecourt V, Alvarez I, Salani F, Ciaramella A, Boerwinkle E, Reiman

EM, Fievet N, Rotter JI, Reisch JS, Hanon O, Cupidi C, Andre Uitterlinden AG, Royall DR, Dufouil C, Maletta RG, De Rojas I, Sano M, Brice A, Cecchetti R, George-Hyslop PS, Ritchie K, Tsolaki M, Tsuang DW, Dubois B, Craig D, Wu C-K, Soininen H, Avramidou D, Albin RL, Fratiglioni L, Germanou A, Apostolova LG, Keller L, Koutroumani M, Arnold SE, Panza F, Gkatzima O, Asthana S, Hannequin D, Whitehead P, Atwood CS, Caffarra P, Hampel H, Quintela I, Carracedo Á, Lannfelt L, Rubinsztein DC, Barnes LL, Pasquier F, Frölich L, Barral S, McGuinness B, Beach TG, Johnston JA, Becker JT, Passmore P, Bigio EH, Schott JM, Bird TD, Warren JD, Boeve BF, Lupton MK, Bowen JD, Proitsi P, Boxer A, Powell JF, Burke JR, Kauwe JSK, Burns JM, Mancuso M, Buxbaum JD, Bonuccelli U, Cairns NJ, McQuillin A, Cao C, Livingston G, Carlson CS, Bass NJ, Carlsson CM, Hardy J, Carney RM, Bras J, Carrasquillo MM, Guerreiro R, Allen M, Chui HC, Fisher E, Masullo C, Crocco EA, Decarli C, Bisceglio G, Dick M, Ma L, Duara R, Graff-Radford NR, Evans DA, Hodges A, Faber KM, Scherer M, Fallon KB, Riemenschneider M, Fardo DW, Heun R, Farlow MR, Kölsch H, Ferris S, Leber M, Foroud TM, Heuser I, Galasko DR, Giegling I, Gearing M, Hüll M, Geschwind DH, Gilbert JR, Morris J, Green RC, Mayo K, Growdon JH, Feulner T, Hamilton RL, Harrell LE, Drichel D, Honig LS, Cushion TD, Huentelman MJ, Hollingworth P, Hulette CM, Hyman BT, Marshall R, Jarvik GP, Meggy A, Abner E, Menzies GE, Jin L-W, Leonenko G, Real LM, Jun GR, Baldwin CT, Grozeva D, Karydas A, Russo G, Kaye JA, Kim R, Jessen F, Kowall NW, Vellas B, Kramer JH, Vardy E, Laferla FM, Jöckel K-H, Lah JJ, Dichgans M, Leverenz JB, Mann D, Levey AI, Pickering-Brown S, Lieberman AP, Klopp N, Lunetta KL, Wichmann HE, Lyketsos CG, Morgan K, Marson DC, Brown K, Martiniuk F, Medway C, Mash DC, Nöthen MM, Masliah E, Hooper NM, McCormick WC, Daniele A, McCurry SM, Bayer A, McDavid AN, Gallacher J, McKee AC, Van Den Bussche H, Mesulam M, Brayne C, Miller BL, Riedel-Heller S, Miller CA, Miller JW, Al-Chalabi A, Morris JC, Shaw CE, Myers AJ, Wiltfang J, O'Bryant S, Olichney JM, Alvarez V, Parisi JE, Singleton AB, Paulson HL, Collinge J, Perry WR, Mead S, Peskind E, Cribbs DH, Rossor M, Pierce A, Ryan NS, Poon WW, Nacmias B, Potter H, Sorbi S, Quinn JF, Sacchinelli E, Raj A, Spalletta G, Raskind M, Caltagirone C, Bossù P, Orfei MD, Reisberg B, Clarke R, Reitz C, Smith AD, Ringman JM, Warden D, Roberson ED, Wilcock G, Rogaeva E, Bruni AC, Rosen HJ, Gallo M, Rosenberg RN, Ben-Shlomo Y, Sager MA, Mecocci P, Saykin AJ, Pastor P, Cuccaro ML, Vance JM, Schneider JA, Schneider LS, Slifer S, Seeley WW, Smith AG, Sonnen JA, Spina S, Stern RA, Swerdlow RH, Tang M, Tanzi RE, Trojanowski JQ, Troncoso JC, Van Deerlin VM, Van Eldik LJ, Vinters HV, Vonsattel JP, Weintraub S, Welsh-Bohmer KA,

Wilhelmsen KC, Williamson J, Wingo TS, Woltjer RL, Wright CB, Yu C-E, Yu L, Saba Y, Pilotto A, Bullido MJ, Peters O, Crane PK, Bennett D, Bosco P, Coto E, Boccardi V, De Jager PL, Lleo A, Warner N, Lopez OL, Ingelsson M, Deloukas P, Cruchaga C, Graff C, Gwilliam R, Fornage M, Goate AM, Sanchez-Juan P, Kehoe PG, Amin N, Ertekin-Taner N, Berr C, Debette S, Love S, Launer LJ, Younkin SG, Dartigues J-F, Corcoran C, Ikram MA, Dickson DW, Nicolas G, Campion D, Tschanz J, Schmidt H, Hakonarson H, Clarimon J, Munger R, Schmidt R, Farrer LA, Van Broeckhoven C, C. O'Donovan M, Destefano AL, Jones L, Haines JL, Deleuze J-F, Owen MJ, Gudnason V, Mayeux R, Escott-Price V, Psaty BM, Ramirez A, Wang L-S, Ruiz A, Van Duijn CM, Holmans PA, Seshadri S, Williams J, Amouyel P, Schellenberg GD, Lambert J-C, Pericak-Vance MA (2019) Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature Genetics* **51**, 414-430.

[6]   Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, DeStefano AL, Bis JC, Beecham GW, Grenier-Boley B, Russo G, Thornton-Wells TA, Jones N, Smith AV, Chouraki V, Thomas C, Ikram MA, Zelenika D, Vardarajan BN, Kamatani Y, Lin C-F, Gerrish A, Schmidt H, Kunkle B, Dunstan ML, Ruiz A, Bihoreau M-T, Choi S-H, Reitz C, Pasquier F, Hollingworth P, Ramirez A, Hanon O, Fitzpatrick AL, Buxbaum JD, Campion D, Crane PK, Baldwin C, Becker T, Gudnason V, Cruchaga C, Craig D, Amin N, Berr C, Lopez OL, De Jager PL, Deramecourt V, Johnston JA, Evans D, Lovestone S, Letenneur L, Morón FJ, Rubinsztein DC, Eiriksdottir G, Sleegers K, Goate AM, Fiévet N, Huentelman MJ, Gill M, Brown K, Kamboh MI, Keller L, Barberger-Gateau P, McGuinness B, Larson EB, Green R, Myers AJ, Dufouil C, Todd S, Wallon D, Love S, Rogaeva E, Gallacher J, St George-Hyslop P, Clarimon J, Lleo A, Bayer A, Tsuang DW, Yu L, Tsolaki M, Bossù P, Spalletta G, Proitsi P, Collinge J, Sorbi S, Sanchez-Garcia F, Fox NC, Hardy J, Naranjo MCD, Bosco P, Clarke R, Brayne C, Galimberti D, Mancuso M, Matthews F, European Alzheimer's Disease I, Genetic, Environmental Risk in Alzheimer's D, Alzheimer's Disease Genetic C, Cohorts for H, Aging Research in Genomic E, Moebus S, Mecocci P, Del Zompo M, Maier W, Hampel H, Pilotto A, Bullido M, Panza F, Caffarra P, Nacmias B, Gilbert JR, Mayhaus M, Lannfelt L, Hakonarson H, Pichler S, Carrasquillo MM, Ingelsson M, Beekly D, Alvarez V, Zou F, Valladares O, Younkin SG, Coto E, Hamilton-Nelson KL, Gu W, Razquin C, Pastor P, Mateo I, Owen MJ, Faber KM, Jonsson PV, Combarros O, O'Donovan MC, Cantwell LB, Soininen H, Blacker D, Mead S, Mosley Jr TH, Bennett DA, Harris TB, Fratiglioni L, Holmes C, de Bruijn RFAG, Passmore

P, Montine TJ, Bettens K, Rotter JI, Brice A, Morgan K, Foroud TM, Kukull WA, Hannequin D, Powell JF, Nalls MA, Ritchie K, Lunetta KL, Kauwe JSK, Boerwinkle E, Riemenschneider M, Boada M, Hiltunen M, Martin ER, Schmidt R, Rujescu D, Wang L-S, Dartigues J-F, Mayeux R, Tzourio C, Hofman A, Nöthen MM, Graff C, Psaty BM, Jones L, Haines JL, Holmans PA, Lathrop M, Pericak-Vance MA, Launer LJ, Farrer LA, van Duijn CM, Van Broeckhoven C, Moskvina V, Seshadri S, Williams J, Schellenberg GD, Amouyel P (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* **45**, 1452.

[7] Dourlen P, Kilinc D, Malmanche N, Chapuis J, Lambert J-C (2019) The new genetic landscape of Alzheimer's disease: from amyloid cascade to genetically driven synaptic failure hypothesis? *Acta Neuropathologica* **138**, 221-236.

[8] Neuner SM, Tcw J, Goate AM (2020) Genetic architecture of Alzheimer's disease. *Neurobiology of Disease* **143**, 104976.

[9] Barthelson K, Newman M, Lardelli M (2020) Sorting Out the Role of the Sortilin-Related Receptor 1 in Alzheimer's Disease. *J Alzheimers Dis Rep* **4**, 123-140.

[10] Mutations, https://www.alzforum.org/mutations, Accessed 1 November.

[11] Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K, et al. (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973-977.

[12] Jiang H, Newman M, Lardelli M (2018) The zebrafish orthologue of familial Alzheimer's disease gene PRESENILIN 2 is required for normal adult melanotic skin pigmentation. *PLOS ONE* **13**, e0206155.

[13] Jayne T, Newman M, Verdile G, Sutherland G, Munch G, Musgrave I, Moussavi Nik SH, Lardelli M (2016) Evidence for and Against a Pathogenic Role of Reduced gamma-Secretase Activity in Familial Alzheimer's Disease. *Journal of Alzheimer's Disease*.

[14] Newman M, Ebrahimie E, Lardelli M (2014) Using the zebrafish model for Alzheimer's disease research. *Frontiers in Genetics* **5**.

[15] Newman M, Hin N, Pederson S, Lardelli M (2019) Brain transcriptome analysis of a familial Alzheimer's disease-like mutation in the zebrafish presenilin 1 gene implies effects on energy production. *Molecular Brain* **12**.

[16]   Barthelson K, Pederson S, Newman M, Lardelli M (2020) Transcriptome analysis of a protein-truncating mutation in sortilin-related receptor 1 associated with early-onset familial Alzheimer's disease indicates effects on mitochondrial and ribosome function in young-adult zebrafish brains. *bioRxiv*, 2020.2009.2003.282277.

[17]   Barthelson K, Pederson SM, Newman M, Lardelli M (2020) Brain transcriptome analysis reveals subtle effects on mitochondrial function and iron homeostasis of mutations in the SORL1 gene implicated in early onset familial Alzheimer's disease. *Molecular Brain* **13**, 142.

[18]   Jiang H, Pederson SM, Newman M, Dong Y, Barthelson K, Lardelli M (2020) Transcriptome analysis indicates dominant effects on ribosome and mitochondrial function of a premature termination codon mutation in the zebrafish gene psen2. *PLOS ONE* **15**, e0232559.

[19]   Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525.

[20]   Team RC (2019) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.

[21]   Soneson C, Love M, Robinson M (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 1; peer review: 2 approved]. *F1000Research* **4**.

[22]   Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.

[23]   Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**, R25.

[24]   Risso D, Ngai J, Speed TP, Dudoit S (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology* **32**, 896-902.

[25]   Liberzon A (2014) A description of the Molecular Signatures Database (MSigDB) Web site. *Methods Mol Biol* **1150**, 153-160.

[26]   Hin N, Newman M, Pederson SM, Lardelli MM (2020) Iron Responsive Element (IRE)-mediated responses to iron dyshomeostasis in Alzheimer's disease. *bioRxiv*, 2020.2005.2001.071498.

[27]   Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**, R14.

[28]   Wilson DJ (2019) The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences* **116**, 1195.

[29]   Wu D, Lim E, Vaillant F, Asselin-Labat M-L, Visvader JE, Smyth GK (2010) ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* **26**, 2176-2182.

[30]   Wu D, Smyth GK (2012) Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic acids research* **40**, e133-e133.

[31]   Sergushichev AA (2016) An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*, 060012.

[32]   Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545.

[33]   Wickham H (2016) Springer-Verlag New York.

[34]   Kolde R (2019).

[35]   Conway JR, Lex A, Gehlenborg N (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938-2940.

[36]   Newman M, Nik HM, Sutherland GT, Hin N, Kim WS, Halliday GM, Jayadev S, Smith C, Laird AS, Lucas CW, Kittipassorn T, Peet DJ, Lardelli M (2020) Accelerated loss of hypoxia response in zebrafish with familial Alzheimer's disease-like mutation of presenilin 1. *Human Molecular Genetics* **29**, 2379-2394.

[37]   Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47-e47.

[38]   Cruts M, Theuns J, Van Broeckhoven C (2012) Locus-specific mutation databases for neurodegenerative brain diseases. *Human Mutation* **33**, 1340-1344.

[39]   Thordardottir S, Rodriguez-Vieitez E, Almkvist O, Ferreira D, Saint-Aubert L, Kinhult-Ståhlbom A, Thonberg H, Schöll M, Westman E, Wall A, Eriksdotter M, Zetterberg H, Blennow K, Nordberg A, Graff C (2018) Reduced penetrance of the PSEN1 H163Y autosomal dominant Alzheimer

mutation: a 22-year follow-up study. *Alzheimer's Research & Therapy* **10**, 45.

[40] Rossor M, Fox N, Beck J, Campbell T, Collinge J (1996) Incomplete penetrance of familial Alzheimer's disease in a pedigree with a novel presenilin-1 gene mutation. *The Lancet* **347**, 1560.

[41] Sherrington R, Froelich S, Sorbi S, Campion D, Chi H, Rogaeva EA, Levesque G, Rogaev EI, Lin C, Liang Y, Ikeda M, Mar L, Brice A, Agid Y, Percy ME, Clerget-Darpoux F, Piacentini S, Marcon G, Nacmias B, Amaducci L, Frebourg T, Lannfelt L, Rommens JM, St George-Hyslop PH (1996) Alzheimer's Disease Associated with Mutations in Presenilin 2 is Rare and Variably Penetrant. *Human Molecular Genetics* **5**, 985-988.

[42] Dong Y, Newman M, Pederson S, Hin N, Lardelli M (2020) Transcriptome analyses of 7-day-old zebrafish larvae possessing a familial Alzheimer's disease-like mutation in psen1 indicate effects on oxidative phosphorylation, mcm functions, and iron homeostasis. *bioRxiv*, 2020.2005.2003.075424.

[43] Zhu X, Lee H-g, Perry G, Smith MA (2007) Alzheimer disease, the two-hit hypothesis: An update. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1772**, 494-502.

[44] Barthelson K, Pederson SM, Newman M, Lardelli M (2020) Brain transcriptome analysis reveals subtle effects on mitochondrial function and iron homeostasis of mutations in the SORL1 gene implicated in early onset familial Alzheimer's disease. *bioRxiv*, 2020.2007.2017.207787.

[45] El-Brolosy MA, Kontarakis Z, Rossi A, Kuenne C, Günther S, Fukuda N, Kikhi K, Boezio GLM, Takacs CM, Lai S-L, Fukuda R, Gerri C, Giraldez AJ, Stainier DYR (2019) Genetic compensation triggered by mutant mRNA degradation. *Nature* **568**, 193-197.

[46] Cervantes S, Gonzàlez-Duarte R, Marfany G (2001) Homodimerization of presenilin N-terminal fragments is affected by mutations linked to Alzheimer's disease. *FEBS Lett* **505**, 81-86.

[47] Luo W, Pant G, Bhavnasi YK, Blanchard SG, Jr., Brouwer C (2017) Pathview Web: user friendly pathway visualization and data integration. *Nucleic Acids Research* **45**, W501-W508.