Iron Responsive Element (IRE)-mediated responses to iron dyshomeostasis in Alzheimer's disease

Author Details:

Nhi Hin*[†a,b], Morgan Newman[b], Stephen Pederson[a], Michael Lardelli*[b]

a. Bioinformatics Hub, School of Biological Sciences, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
b. Alzheimer's Disease Genetics Laboratory, School of Biological Sciences, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

* corresponding authors

[†] Current Address: South Australian Genomics Centre, South Australian Health and Medical Institute, North Terrace, Adelaide, SA 5000, Australia

Author Full Names and Contact Details:

Nhi Hin – **Email:** nhi.hin@sahmri.com **Phone:** +61 8 8128 4621

Morgan Newman – **Email:** morgan.newman@adelaide.edu.au

Stephen Pederson – **Email:** stephen.pederson@adelaide.edu.au

Michael Lardelli – **Email:** michael.lardelli@adelaide.edu.au **Phone:** +61 8 8313 3212

Abstract

Background: Iron trafficking and accumulation has been associated with Alzheimer's disease (AD) pathogenesis. However, the role of iron dyshomeostasis in early disease stages is uncertain. Currently, gene expression changes indicative of iron dyshomeostasis are not well characterized, making it difficult to explore these in existing datasets.

Objective: We identified sets of genes predicted to contain Iron Responsive Elements (IREs) and used these to explore iron dyshomeostasis-associated gene expression responses in cultured cell, human, mouse, and zebrafish datasets.

Results: IRE gene sets were sufficiently sensitive to distinguish not only between iron overload and deficiency in cultured cells, but also between AD and other pathological brain conditions in different brain regions. Notably, we see changes in IRE transcript abundance as amongst the earliest observable changes in zebrafish familial AD (fAD)-like brains, preceding other AD-typical pathologies such as inflammatory changes. Unexpectedly, while some IREs in the 3' untranslated regions of transcripts show significantly increased stability under iron deficiency in line with current assumptions, many such transcripts instead show decreased stability, indicating that this is not a generalizable paradigm.

Conclusion: Our results reveal IRE gene expression changes as early markers of the pathogenic process in fAD and are consistent with iron dyshomeostasis as an important driver of this disease. IRE gene expression changes can also distinguish AD from other neuropathologies. Our work demonstrates how differences in the stability of IRE-containing transcripts can be used to explore and compare iron dyshomeostasis-associated gene expression responses in different species, tissues, and conditions.

Keywords

Iron responsive element, iron homeostasis, Alzheimer's disease, gene expression, Computational Biology, Transcriptomics, familial Alzheimer's disease

Introduction

The pathological processes underlying Alzheimer's Disease (AD) commence decades before symptoms become evident [1,2]. Since the early days of AD research, iron trafficking and accumulation has been observed to be altered in AD [3–7]. However, it is still unclear whether iron dyshomeostasis represents a late symptom or an early pathological driver of AD. Iron homeostasis is closely linked to many critical biological processes including cellular respiration, hypoxia, and immune responses, all of which are disrupted in AD. Consistent with this, discoveries over the past decade have shown that disruptions to iron homeostasis can drive feedback loops that worsen AD pathology [8,9]. However, evidence for iron dyshomeostasis as an early driver of the disease is only just emerging. While iron accumulation associated with amyloid plaque formation has been observed in various transgenic mouse models of AD [10–13], the age at which iron dyshomeostasis first occurs is uncertain [14].

Currently, gene expression patterns representing responses to iron dyshomeostasis are not well-characterized. Cellular responses to iron dyshomeostasis are complex and involve several systems and layers of regulation. The stability of the transcription factor HIF1α, (a component of HIF1, a master regulator of responses to hypoxia) is regulated by an iron-dependent mechanism so that transcriptional responses to iron deficiency can resemble hypoxia responses [15]. However, cellular iron homeostasis is also regulated at the post-transcriptional level by the IRP/IRE system [16–18]. In this system, altered levels of available

83 ferrous iron (Fe^{2+}) cause Iron Responsive Proteins (IRP1 or IRP2) to change conformation
84 or stability respectively [19,20]. This alters their ability to bind *cis*-regulatory Iron Responsive
85 Element (IRE) stem-loop motifs in the 3' or 5' untranslated regions (UTRs) of genes
86 encoding products related to iron metabolism. Only a few IRE-containing genes have been
87 characterized in detail, including transferrin receptor 1 (*TFR1*; 3' UTR IRE), divalent metal
88 transporter 1 (*DMT1* or *SLC11A2*; 3' UTR IRE), ferritin heavy chain and ferritin light chain (
89 *FTH1* and *FTL*; both 5' UTR IRE), and ferroportin (*SLC40A1*; 5' UTR IRE) [18]. In general,
90 these previously-characterized IRE-containing genes suggest that IRPs binding to 3' IREs
91 tend to stabilize transcripts to increase protein translation, while binding to 5' IREs
92 suppresses translation [17]. It is important to note that while the ancestral ferritin IRE is
93 present even in lower metazoans (including sponges), IRE elements in other genes have
94 likely more recently arisen through convergent evolution [21]. Perhaps because of this, some
95 heterogeneity in regulatory mechanisms and effects for IRE genes has been reported. For
96 example, *SLC11A2* is stabilized by IRP1 binding to a greater extent than by IRP2 binding
97 (as its 3' IRE has a greater affinity for IRP1) [22]. Similarly, while *CDC42BPA* and *TFRC*
98 both possess 3' IREs, *CDC42BPA* is stabilized to a greater extent than *TFRC* under iron
99 deficiency [23]. Currently however, global gene expression changes mediated by genes in
100 the IRP/IRE system have not been well-defined, and the overall expression patterns of IRE-
101 containing genes have not been explored in the context of AD. In addition, it is unclear how
102 expression of these genes might differ between AD and other neurodegenerative diseases,
103 or how AD risk factors such as aging and hypoxia might contribute.
104
105 In this study, we utilized the SIREs (Searching for Iron Responsive Elements) tool [24] to
106 predict and identify sets of IRE-containing genes in human, mouse, and zebrafish. We then
107 applied these gene sets to explore overall IRP/IRE-mediated iron dyshomeostasis-
108 associated responses in datasets involving: (1) a cultured cell line subjected to iron overload

109 and deficiency treatments, (2) a cohort of AD patients, healthy controls, and two other
110 pathological conditions affecting the brain, (3) 5XFAD mice used to model the amyloid and
111 tau pathology seen in AD, and (4) a zebrafish knock-in model possessing a familial AD
112 (fAD)-like mutation.
113
114 Our IRE gene sets displayed significant enrichment in AD, the 5XFAD mouse model, and an
115 fAD-like zebrafish model, demonstrating for the first time the early and extensive
116 involvement of IRP/IRE-mediated iron dyshomeostasis-associated gene expression
117 responses in the context of AD. IRE gene sets were sufficiently sensitive to distinguish not
118 only between iron overload and deficiency in a cultured cell line dataset, but also between
119 AD and other pathological conditions affecting the brain (pathological aging and progressive
120 supranuclear palsy), implying that the dysregulation of IRE-containing genes in AD may
121 differ from other conditions. Changes in IRE gene expression were already evident in young
122 adult brains of animal models of AD (3-month-old 5XFAD mice, 6-month-old fAD-like
123 zebrafish). Overall, our observations do not support the current assumption that IRP binding
124 to 3' IREs generally stabilizes transcripts as we observed both increases and decreases in
125 abundance of transcripts with either 3' or 5' IREs under most conditions.
126
127

128 # Results

129 ## Defining sets of genes containing Iron Responsive Elements (IREs)

130 We utilized the SIREs (Searching for Iron Responsive Elements) tool, to define species-
131 specific IRE gene sets by searching for IRE and IRE-like motifs in the 3' and 5' untranslated
132 regions (UTRs) of the reference transcriptomes of human (hg38), mouse (mm10), and
133 zebrafish (z11). *SIREs* assigns a quality-score to all detected IREs, with "high-quality"

134 scores corresponding to canonical IREs and “medium-quality” or “low-quality” scores
135 reflecting deviations from the canonical IRE (alternative nucleotide composition in the apical
136 loop, one bulge at the 3’ or one mismatch in the upper stem) that would still produce an IRE-
137 like motif with potential to be functional [24–28]. **Figure 1** summarizes the IRP/IRE
138 interaction effects on transcripts and gives examples of canonical and non-canonical IREs.
139
140 We defined four gene sets for each species as follows: ***HQ 3’ IREs*** (high-quality predicted 3’
141 IRE genes), ***HQ 5’ IREs*** (high-quality predicted 5’ IRE genes)*,* ***all 3’ IREs*** (including all low,
142 medium, and high-quality predicted 3’ IRE genes), and ***all 5’ IREs*** (including all low,
143 medium, and high-quality predicted 5’ IRE genes). The size of these gene sets for each
144 species and the overlap present is shown in **Figure 2** and the gene sets are provided in
145 **Supplementary Table 1**. Overall, searching through human UTRs uncovered the largest
146 number of predicted IRE genes, followed by mouse, and then zebrafish. Overlap between
147 homologous genes from the IRE gene sets of different species was generally poor. The
148 largest sets of genes shared between species were consistently found between human and
149 mouse, which is reflective of their closer evolutionary divergence.
150
151 While not many high-quality IRE genes were identified across all three species, the few we
152 identified are consistent with known and well-characterized IREs in the literature [29,30]. For
153 example, the single shared ***HQ 3’ IRE*** gene between the three species was *TFRC*
154 (transferrin receptor), while the shared ***HQ 5’ IRE*** genes between the three species included
155 *FTH1* (ferritin heavy chain 1) and *ALAS2* (5'-aminolevulinate synthase 2). We note that the
156 scoring of genes for the ***HQ 3’ IRE*** and ***HQ 5’ IRE*** sets is considered stringent (i.e. only
157 canonical IREs) and that other human genes with well-characterized IREs experimentally
158 shown to bind IRPs are recovered in the ***All 3’ IRE*** and ***All 5’ IRE*** sets across mouse and
159 zebrafish as well. These genes include *CDC14A* (3’ IRE; [31]), *CDC42BPA* (3’ IRE; [23,32],

160 *SLC40A1* (5' IRE; [33]), and *FTL* (5' IRE; [34]). In addition, the 3' IRE of *SLC11A2* (also
161 known as *DMT1*) [22] which is only found in mammals was also recovered in the mouse ***All***
162 ***3' IREs*** set (but not for zebrafish). The expression of these previously characterized IRE
163 genes across the human, mouse, and zebrafish datasets analyzed in this current work is
164 described in **Supplementary Text 1** and discussed further in the analyses we later describe.
165

166 IRE gene sets are over-represented within up-regulated AD genes, but
167 overall not well-represented in existing gene sets

168 We explored the biological relevance of the predicted IRE gene sets described above by
169 testing whether genes within them were over-represented in existing gene sets from the
170 Molecular Signatures Database (MSigDB; available at https://www.gsea-
171 msigdb.org/gsea/msigdb). MSigDB represents one of the largest publicly-available collection
172 of gene sets collated from existing studies and other biological databases [35]. We limited
173 our analysis to gene sets from the following collections: Hallmark (non-redundant sets of
174 ~200 genes each representing various biological activities), C2 (gene sets from databases
175 including KEGG and Reactome and published studies), C3 (gene sets containing genes with
176 motif elements), and C5 (gene sets based on gene ontology terms) (see **Methods**). We
177 performed over-representation analysis for the predicted IRE gene sets (***all 3' IREs*** and ***all***
178 ***5' IREs***) separately for each species (human, mouse, zebrafish), and used Wilkinson's
179 meta-analytic approach to determine gene sets that were significantly over-represented
180 across the three species' IRE gene sets. Our results indicated that 1,148 of 10,427 tested
181 gene sets displayed over-representation of IRE gene sets across all three species
182 (Bonferroni-adjusted Wilkinson's p-value < 0.05) (**Supplementary Table 2**). Remarkably,
183 the "Blalock Alzheimer's Disease Up" gene set from the C2 collection was significantly over-
184 represented in both the sets of ***all 3' IREs*** (Bonferroni-adjusted Wilkinson's p-value = 3.9e-
185 14) and ***all 5' IREs*** (Bonferroni-adjusted Wilkinson's p-value = 2.1e-26) in the meta-analysis,

186 and was also the gene set with the most significant over-representation of the human ***all 3'***
187 ***IREs*** set (Bonferroni-adjusted Fisher's exact test *p*-value = 7.8e-58) (**Supplementary Table**
188 **2**). This supports that disturbance of iron homeostasis particularly distinguishes Alzheimer's
189 disease from other disease conditions and pathways represented within the C2 collection. In
190 addition, the top 15 MSigDB gene sets showing the most significant over-representation
191 generally differed between species. However, in all cases, a large proportion of IRE genes
192 from the predicted IRE gene sets were not contained within any of these top-ranked MSigDB
193 gene sets or within the Heme Metabolism gene set belonging to the Hallmark collection
194 (**Figure 3**). This demonstrates that the predicted IRE genes we defined are not fully
195 captured by existing gene sets and so may be uniquely useful for investigating gene
196 expression changes during the IRE-IRP response to iron dyshomeostasis.
197

198 ## Involvement of transcription factor regulation within IRE gene sets

199 To be confident that IRE gene sets accurately capture information about the IRP/IRE-
200 mediated responses, we needed to investigate possible co-ordinate regulation by other
201 factors. As a starting point, we examined whether known binding motifs for transcription
202 factors were significantly over-represented in the promoter regions of genes within each IRE
203 gene set (***all 3' IREs***, ***all 5' IREs***, ***HQ 3' IREs***, and ***HQ 5' IREs***) for each species. We
204 detected significant over-representation of several transcription factors including the Klf14
205 motif in the zebrafish ***all 3' IREs*** set (FDR-adjusted *p*-value = 0.049), the E2F motif in the
206 human ***all 5' IREs*** set (FDR-adjusted *p*-value = 0.049) and the Bach1 (FDR-adjusted *p*-
207 value = 0.012), Nrf2 (FDR-adjusted *p*-value = 0.013), and NF-E2 (FDR-adjusted *p*-value =
208 0.019) motifs in the zebrafish ***HQ 5' IREs*** set (**Supplementary Table 3**). This suggests that
209 the following results should be interpreted with the knowledge that expression of subsets of
210 genes in some of the IRE gene sets may be influenced by factors other than iron ions.
211

212 ## Gene set enrichment testing approach

213 Our predicted IRE gene sets can be used in any standard gene set enrichment testing
214 approach to detect potential changes in iron homeostasis between conditions. Our workflow,
215 which we later apply on human, mouse, and zebrafish datasets, is shown in **Figure 4**. Due
216 to variability in the results produced by different gene set enrichment testing methods, we
217 were inspired by the EGSEA framework [36] to combine the results from different methods.
218 Based on an initial analysis using EGSEA, we chose *fry/mroast* [37,38], *camera* [39], and
219 *fgsea* [40,41] as the representative methods to use. (See **Supplementary Figure 2** for a
220 principal component analysis of results from different gene set enrichment analysis
221 approaches.) A summary of the different characteristics of the three methods is shown in
222 **Table 1**. The raw enrichment *p*-values from these approaches can be combined to obtain an
223 overall enrichment *p*-value for each gene set. In accordance with EGSEA default
224 parameters, we used Wilkinson's method to combine raw *p*-values, and then applied
225 adjustment for multiple testing on all combined *p*-values. Along with performing this gene set
226 enrichment testing on our IRE gene sets, we also recommend using the same method to
227 test for enrichment for the MSigDB Hallmark gene sets as the diverse biological activities
228 they represent help to provide context for interpreting the IRE enrichment results. To explore
229 further the results of IRE gene set enrichment analysis, we use UpSet plots to display the
230 overlap between sets of "leading-edge" genes in the ***all 3' IREs*** and ***all 5' IREs*** gene sets.
231 The leading-edge genes can be interpreted as the core (often biologically important) genes
232 of a gene set that account for the significant enrichment as calculated by GSEA [41].
233

234 ## Differences in IRE gene set enrichment during iron deficiency and iron
235 ## overload in a cultured cell line

236 We first tested how our enrichment approach illuminated the effects of iron overload and
237 deficiency in a cultured cell line microarray dataset from Caco-2 cells (GEO accession:

238 GSE3573). As only one cell type contributed to this dataset, interpretation of the IRE
239 enrichment results is simplified by not having to consider the differing iron requirements of
240 different cell types. This allowed us to focus on whether the iron dyshomeostasis treatments
241 (iron overload and iron deficiency) could be detected and distinguished in terms of their
242 IRP/IRE system-driven transcript abundance response. Because the dataset was from a
243 microarray experiment, we performed only the differential gene expression and gene set
244 enrichment testing portions of our workflow. The results of Principal Component Analysis
245 and differential gene expression analysis are provided in **Supplementary Figure 3**. We note
246 that this particular dataset was chosen despite not being derived from a central nervous
247 system cell type as maintaining iron homeostasis is fundamental to all cell types (regardless
248 of their tissue of origin). In addition, this particular dataset included iron deficiency and iron
249 overload treatments in the same batch and with sufficient sample number, and no other
250 cultured cell line gene expression dataset was publicly available at the time that fulfilled
251 these criteria.
252
253 In general, we found iron deficiency and iron overload treatments resulted in different gene
254 expression responses. In terms of classic IRE genes such as *TFRC, FTL,* and *FTH1*, iron
255 deficiency is associated with increased expression of *TFRC* (consistent with increased
256 transcript stabilization of this 3' IRE gene) and decreased expression of the 5' IRE genes
257 *FTL* and *FTH1,* while these changes are in the opposite direction during iron overload (see
258 **Supplementary Text 1** and **Supplementary Figure 1A**). Differential gene expression
259 across all detected genes in the dataset additionally revealed that iron deficiency was
260 associated with 96 differentially expressed genes of which 10 possessed predicted IREs
261 while iron overload was associated with 212 differentially expressed genes (FDR-adjusted *p*-
262 value from *limma* < 0.05) of which 33 possessed predicted IREs (**Supplementary Figure**
263 **3B**). There were 17 differentially expressed genes in common between iron deficiency and

264 iron overload treatments, and all moved in opposite directions according to the treatment
265 (i.e. increased abundance under iron deficiency and decreased abundance under iron
266 overload, or vice versa). These clear differences between iron deficiency and iron overload
267 were reflected in gene set enrichment analyses using the MSigDB Hallmark gene sets,
268 where the gene sets involved and the proportions of gene transcripts with increased or
269 decreased abundance differed (**Figure 5A**). As expected, IRE gene sets also showed
270 significant enrichment under iron deficiency and overload conditions (**Figure 5B**) (Bonferroni
271 adjusted p-value < 0.05).
272
273 We expected that iron deficiency would result in increased expression of genes with 3' IREs
274 under the IRP/IRE paradigm similar to the *TFRC* gene as seen in **Supplementary Text 1**.
275 However, both the 3' and 5' IRE gene sets displayed mixed patterns of increased and
276 decreased expression under both the iron deficiency and iron overload treatments (**Figure**
277 **5B**). This indicates it would be difficult to distinguish between these conditions based purely
278 on overall increases or decreases in the expression of IRE gene sets. Despite this, we see
279 that the iron deficiency and iron overload treatments can be distinguished by their "leading-
280 edge" genes (those genes contributing most to the enrichment signal for the predicted IRE
281 gene sets (***all 3' IREs*** and ***all 5' IREs***) (**Figure 5C**). This supports that gene set enrichment
282 using predicted IRE gene sets may be sufficiently sensitive to detect whether iron
283 dyshomeostasis is present and potentially more informative than individual IRE genes in
284 distinguishing between different IRP-IRE system-mediated gene expression responses in
285 iron deficiency and iron overload treatments in this cell line.
286

287 A distinct iron homeostasis response in human AD patients compared to

288 other neuropathologies

289 Given that IRE gene sets could distinguish between iron overload and deficiency in a
290 cultured cell line, we next tested our gene sets on a more complex dataset including
291 cerebellum and temporal cortex tissue samples from post-mortem human brains. The brains
292 originated from either healthy controls or patients afflicted with one of three conditions:
293 Alzheimer's disease (AD); pathological aging (PA), a condition involving amyloid pathology
294 but no significant dementia symptoms; or progressive supranuclear palsy (PSP), a
295 tauopathy without amyloid pathology [42]. An important characteristic of the dataset is that
296 both cerebellum and temporal cortex tissue samples were available from each patient. A
297 summary of the 236 patients whose data we included in this analysis is shown in **Table 2**
298 and the results of differential gene expression analysis and IRE gene set enrichment
299 analysis are shown in **Table 3** and **Figure 6A**. The expression of classic IRE genes
300 including *TFRC*, *FTL,* and *FTH1* in this dataset is shown in **Supplementary Text 1** and
301 **Supplementary Figure 1B-C**. In our analyses, we focus mainly on comparing conditions
302 within each tissue rather than between tissues. This is because we found significant
303 differences in the AD vs. control comparison in the temporal cortex compared to the
304 cerebellum (see **Supplementary Figure 4**).

305

306 Overall, our IRE enrichment analyses indicate significant enrichment of IRE gene sets in all
307 pathological conditions (AD, PA or PSP) compared to healthy controls within the cerebellum
308 and temporal cortex (**Figure 6A**). In all pathological conditions, 3' and 5' IRE gene
309 transcripts show overall mixed patterns of abundance (e.g. increased and decreased).
310 Further examination of the leading-edge genes from these IRE gene sets gives more insight
311 regarding potential differences and similarities between AD, PA, and PSP (**Figure 6B**).
312 Overall, AD, PSP, and PA appear to involve distinct yet partially overlapping IRE gene

313 expression responses in a tissue-specific manner. Within the temporal cortex, there are 435
314 3' IRE leading-edge genes and 178 5' IRE leading-edge genes exclusively present in the
315 "AD vs. control" comparison. These are greater numbers than for the other two conditions.
316 Interestingly, AD and PA share only relatively small numbers of leading-edge genes despite
317 the fact that many regard PA as a prodrome of AD [43]. Interestingly, in the cerebellum,
318 while AD and PSP share many 3' and 5' IRE leading-edge genes, PA is associated with a
319 large number of unique leading-edge 3' and 5' IRE genes, further emphasizing its difference
320 from AD and also PSP. These observations are supported when considering expression of
321 previously characterized IRE genes individually (see **Supplementary Text 1** and
322 **Supplementary Figure 1B-C**). The results in this Supplementary Text suggest that in
323 particular, some 3' IRE genes (e.g. *TFRC*, *CAV3*, *CDC14A*, *CDC42BPA, SERTAD2,*
324 *TRAM1*) show increased expression in AD (relative to controls) whereas PA and PSP do
325 not. Overall, our results suggest that IRP-IRE system-mediated gene expression changes
326 may be sufficiently sensitive to discern and identify potential differences in the way iron
327 homeostasis is regulated in AD compared to PA and PSP.
328

329 Age-dependent disruption of IRE-driven iron homeostasis in the 5XFAD 330 mouse model

331 Alzheimer's disease is thought to develop over decades [44–46]. However, detailed
332 molecular studies of AD brains can only be based on post-mortem tissues. To reveal the
333 early molecular changes that initiate the progression to AD we must study the brains of
334 animal disease models. Given that the IRE gene sets appear to work well in human
335 datasets, we then tested our mouse IRE gene sets on an RNA-seq dataset derived from
336 brain cortex tissue of the 5XFAD transgenic mouse AD model (GEO: GSE140286). The
337 5XFAD mouse is one of the most common systems used to model the amyloid beta
338 histopathology of AD brains and brain gene expression patterns in this model were

339 previously shown in a concordance analysis to resemble those of human AD brains more
340 than four other mouse models (3xTg, CK-p25, J20, and Tg2576) [47]. 5XFAD mice possess
341 two transgenes that include a total of five different mutations, each of which individually
342 causes fAD in humans. In this dataset, the mice analyzed were either 3, 6 or 12 months of
343 age. We acknowledge that the small sample size of this particular dataset means that results
344 should be interpreted tentatively (see Discussion).
345
346 Using gene set enrichment testing methods as before, we observed significant enrichment of
347 the ***all 3' IREs*** and ***all 5' IREs*** gene sets in several comparisons. These included 5XFAD vs.
348 wild type comparisons and wild type aging comparisons (Bonferroni-adjusted enrichment *p*-
349 value < 0.05) (**Figure 7**). Notably, even the youngest age group of 5XFAD mutant mice (3
350 months) displayed significant enrichment of genes containing 3' or 5' IREs compared to age-
351 matched wild types. This is consistent with an enrichment of immune-related Hallmark gene
352 sets that we observed in this age group (see **Figure 7B**) and with a previous transcriptome
353 analysis suggesting immune activation in 5XFAD mice as early as 2-4 months [48]. UpSet
354 plots of overlapping leading-edge genes suggest that the IRE responses due to aging and
355 due to the 5XFAD transgenes may involve partially overlapping gene expression changes
356 (**Figure 8**). In addition, the UpSet plots reveal subsets containing large numbers of IRE-
357 containing genes which uniquely contribute to enrichment of IRE gene sets in only one
358 comparison (e.g. 256 3' IRE genes only contribute to enrichment in the "*5xFAD vs WT at 6*
359 *months*" comparison). Notably, there are 126 shared 3' IRE genes contributing to enrichment
360 in the "*5xFAD vs WT at 6 months*" and "*5xFAD vs WT at 3 months*" comparisons, but no
361 genes shared between these comparisons and the "*5xFAD vs WT at 12 months*"
362 comparison. These 126 genes were found to show over-representation of Gene Ontology
363 terms including "Hsp70 protein binding" and "nuclear import signal receptor" (FDR-adj. *p*-
364 value < 0.05) (**Supplementary Table 4**), suggesting that subsets of leading-edge genes

365 may have biological relevance, and that these may give clues into the biologically distinct
366 age-dependent iron dyshomeostasis responses caused by the 5XFAD transgenes. Although
367 beyond the scope of our current analysis, further investigation into these subsets of genes
368 may further help to define differences in iron dyshomeostasis responses under different
369 biological conditions of interest.
370

371 Similarities in IRE-driven iron homeostasis responses during hypoxia
372 and a familial AD-like mutation in a zebrafish model

373 Concerns have been raised over the relevance of transgenic mouse models in modelling the
374 early stages of AD (reviewed by [49]). Meanwhile, knock-in models of familial AD (fAD) (that
375 possess a single, heterozygous mutation in an endogenous gene like the human version of
376 the disease) only exhibit subtle pathological changes with little to no visible amyloid or tau
377 pathology present [50]. We make an important distinction between these fAD-like models
378 that resemble the genetic state of human fAD as closely as possible and other knock-in
379 models that invoke genetic states which have never been observed in human fAD (e.g.
380 models possessing homozygous mutations that have only been observed in a heterozygous
381 state in humans and/or several fAD-causative mutations simultaneously as described by
382 Saito et al. [51] or Serneels et al. [52]). Despite the comparatively more subtle pathology
383 seen in fAD-like knock-in models, because they more closely mimic the genetic state of fAD,
384 it would be expected that gene expression analysis of knock-in models may reveal
385 informative early molecular changes that are relevant to early fAD states in human brains.
386 We had access to whole-brain RNA-seq data from a knock-in zebrafish model of fAD
387 possessing a single fAD-like mutation in its endogenous *psen1* gene (*psen1*^{Q96_K97del}/+)
388 (henceforth referred to as the “fAD-mutation-like” model in this paper). Previous analysis of a
389 subset of this dataset involving young adults (6-month-old brains) revealed gene expression
390 changes related to energy metabolism and mitochondrial function, especially ATP synthesis

391 and other ATP-dependent functions including lysosomal acidification [53]. (In contrast, the 3-
392 month-old young adult 5XFAD mouse brain dataset is dominated by immune/inflammation
393 responses, **Figure 7B**). Considering the critical role of iron homeostasis in energy
394 metabolism, we decided to revisit this zebrafish dataset. We performed IRE gene set
395 enrichment on the entire dataset in order to include exploration of the effects of aging and
396 acute hypoxia (two important risk factors for sporadic late onset AD) and to analyze how
397 these effects interact with the fAD-like *psen1*^{Q96_K97del}/+ mutant genotype. The experimental
398 design and the results of the differential gene expression analyses and gene set enrichment
399 tests are summarized in **Figure 9**.
400
401 We first turned our attention to the comparison between young adult (6-month-old)
402 *psen1*^{Q96_K97del}/+ mutant zebrafish and their wild-type siblings. At this age, gene expression
403 changes in the mutant fish likely represent early stresses driving the development of fAD in
404 humans. Gene set enrichment tests in this comparison identify alteration of energy
405 metabolism-related Hallmark gene sets (e.g. OXIDATIVE PHOSPHORYLATION,
406 GLYCOLYSIS) (**Figure 9B**) which is consistent with a previous analysis of gene ontology
407 terms with this dataset [53]. In addition, we see enrichment of other gene sets including
408 FATTY ACID METABOLISM, PI3K AKT MTOR SIGNALLING, MTORC1 SIGNALLING, and
409 HEME METABOLISM. All of these gene sets which show enrichment in 6-month-old
410 *psen1*^{Q96_K97del}/+ mutants relative to wild-type siblings are also enriched in 24-month-old
411 *psen1*^{Q96_K97del}/+ mutants relative to their wild-type siblings (**Figure 9B**). This supports that
412 the biological activities represented in these gene sets are amongst the earliest altered in
413 this particular fAD mutation model.
414
415 In Lumsden et al. [14] we predicted that fAD mutations in the major locus *PSEN1* would
416 cause an early cellular deficiency of ferrous iron due to the observation of insufficient

417 acidification of the endolysosomal pathway in in-vitro *PSEN1* mutation studies [54,55]. In
418 Newman et al. [53] we saw that GO analysis of 6 month old *psen1*^{Q96_K97del}/+ zebrafish brain
419 supported that lysosomal acidification was affected. Therefore, we decided to apply our IRE
420 enrichment analysis to test for evidence of iron dyshomeostasis in these fish. The
421 enrichment of the ***all 3' IREs*** set in 6-month-old *psen1*^{Q96_K97del}/+ zebrafish brains supports
422 that iron dyshomeostasis is an important stress in the early stages of fAD (**Figure 10A**).
423
424 While almost all pairwise comparisons in the zebrafish dataset show significant enrichment
425 of at least one IRE gene set (**Figure 10A**), the expression of IRE genes appears to differ in
426 terms of the proportions of IRE-containing transcripts which show increased versus
427 decreased abundance (**Figure 10B**). In addition, the Principal Component Analysis plot of
428 expression of the ***all 3' IREs*** set over all samples shown in **Figure 10C** suggests that
429 different conditions appear to have distinct expression patterns of these genes. Across the
430 first principal component, different age groups (6- and 24-month-old brains) differ in their
431 expression of predicted 3' IRE genes, while the second principal component appears to
432 separate *psen1*^{Q96_K97del}/+ mutants from their wild-type siblings. This separation between
433 *psen1*^{Q96_K97del}/+ mutants and wild type siblings is even more pronounced when both are
434 exposed to hypoxia.
435
436 To gain more insight into the similarities and differences between the IRE responses in the
437 *psen1*^{Q96_K97del}/+ vs. wild type comparison, we plotted UpSet plots of overlapping leading-
438 edge genes (**Figure 10D**). These plots suggest that the IRE responses during hypoxia,
439 aging, and due to this fAD-like mutation are mostly distinct from each other with unique
440 leading-edge genes. However, similarities in the IRE response between different conditions
441 are suggested by the existence of some shared leading-edge genes. For example, for the
442 set of ***all 3' IREs***, the "*6-month-old psen1*^{Q96_K97del}/+ *vs. wild type*" comparison shares 19

443 leading-edge genes with the "*24-month-old psen1*^{Q96_K97del}*/+ vs. wild type*" comparison. As
444 an initial exploration of the biological relevance of these genes, we submitted them to the
445 STRINGR tool. This indicated that the proteins products of these genes were significantly
446 associated with each other (e.g. in terms of text-mining from PubMed articles, experimentally
447 determined interactions, and co-expression). These proteins were significantly over-
448 represented in the sets "MAPK pathway", "AP-1 transcription factor", "Jun-like transcription
449 factor", and "signaling by TGF-beta family members" (FDR-adjusted over-representation p-
450 value < 0.05; **Supplementary Figure 5**). The AP-1 and MAPK pathways have previously
451 been shown to be stimulated by iron depletion [56,57]. Therefore, mechanistically, it is
452 possible that the IRE-IRP response to iron dyshomeostasis in the fAD-like zebrafish mutant
453 might involve these pathways, although this requires confirmation *in vivo*.
454
455 The fact that aging, hypoxia, and a fAD-like mutation all cause changes in the abundance of
456 IRE-containing transcripts in zebrafish raised concerns regarding the specificity of such
457 changes. Therefore, as a negative control, we examined changes in IRE transcript
458 abundance in a brain transcriptome dataset derived from zebrafish heterozygous for a *psen1*
459 mutation now thought not to be fAD-like, *psen1*^{K97fs} [58]. In this dataset, *psen1*^{K97fs}/+
460 zebrafish are compared to their wild type siblings at 6 and 24 months. We tested this dataset
461 for enrichment using our zebrafish 3' and 5' IRE gene sets but found no significant
462 enrichment of any of our predicted IRE gene sets in *psen1*^{K97fs}/+ vs. wild type comparisons
463 at any age (**Table 4**; **Supplementary Figure 6**). Reassuringly, we still observed significant
464 enrichment of both 3' and 5' IRE gene sets during wild type aging (*24-month-old wild types*
465 *vs. 6-month-old* wild types), consistent with the equivalent comparison in the
466 *psen1*^{Q96_K97del}/+ dataset. These results support that IRE-containing transcript abundance
467 changes are sufficiently sensitive to reflect differences in iron homeostasis between different
468 mutation models.

469

470 Simultaneous stabilization of some 3' IRE transcripts and destabilization

471 of others

472 In the cultured cell line dataset analyzed above, we noticed that even a straightforward iron
473 deficiency treatment resulted in the simultaneous increase and decrease in expression of 3'
474 IRE-containing genes. These findings are difficult to reconcile with the current paradigm that
475 stabilization of 3' IRE-containing genes occurs under iron deficiency and suggest that the
476 current model of the IRP/IRE system may be incomplete or insufficient for describing the
477 regulation of non-canonical IREs. Given that many predicted 3' IRE genes with non-
478 canonical IREs (e.g. in the ***all 3' IREs*** set) displayed enrichment and different gene
479 expression patterns in the fAD-like zebrafish dataset, we decided to explore further the
480 stability changes of these genes by comparing the expression of spliced and unspliced
481 transcripts for each gene (**Supplementary Text 2**). We found that transcripts of some
482 predicted 3' IRE genes were significantly increased in stability while others were significantly
483 decreased in stability (**Supplementary Figure 7**; **Supplementary Table 5**).

484 # Discussion

485 In this study we identified comprehensive sets of genes predicted to contain IREs in human,
486 mouse, and zebrafish. Surprisingly, IRE gene sets were generally not well-represented in
487 existing gene sets, and because of this, we believe they represent a novel resource for
488 researchers interested in exploring iron dyshomeostasis gene expression responses in
489 detail. Notably, our finding that IRE genes were most significantly over-represented in an
490 existing gene set previously shown to be up-regulated in AD brains (the "Blalock Alzheimer's
491 Disease Up" gene set from MSigDB) supports the unique importance of IRE genes in AD.
492 Furthermore, IRE gene sets displayed significant enrichment in postmortem brains from a
493 human AD cohort, 5XFAD mouse model, a zebrafish model of a fAD-like mutation. Taken

494 together, our results demonstrate for the first time the involvement of a coordinated IRP/IRE-
495 mediated gene expression response associated with AD.

496

497 Overall, IRE gene sets were sensitive enough to capture several interesting phenomena.
498 The relatively tightly controlled conditions in the mouse and zebrafish model datasets
499 revealed a strong age-dependent effect on the transcript abundances of these predicted
500 IRE-containing genes. Notably, 3' IRE gene expression changes were amongst the earliest
501 changes observable in the zebrafish fAD-like mutation model alongside changes in energy
502 metabolism. These 3' IRE gene expression changes preceded other signals of pathological
503 change in the transcriptome (such as altered expression of inflammatory response
504 pathways) commonly associated with AD. Our findings raise the question of how early IRE
505 gene expression changes and iron dyshomeostasis more broadly may start to occur in fAD
506 brains. In a recent analysis of transcriptome data from pools of 7-day-old larvae
507 heterozygous for the fAD-like mutation *psen1*^{Q96_K97del} (compared to wild type half-siblings),
508 we have observed over-representation of the GO term "iron ion transport" without significant
509 changes in 3' IRE or 5' IRE transcript abundance [59]. This supports that the 3' IRE
510 transcript abundance signal in 6-month-old brains of this mutant line reflects changes in
511 ferrous iron abundance that develop with age in response to an underlying defect in cellular
512 iron regulation.

513

514 In relation to our findings associated with the 5XFAD mouse model, we acknowledge that
515 the low sample number is a limitation in this analysis and constrains our interpretation of
516 results from this particular model. Unfortunately, to our knowledge, no other publicly-
517 available 5XFAD mouse dataset was available containing (1) both young and aged adults in
518 the same batch (where young is defined as ~2-4 months old mice as this is when the earliest
519 signs of pathology are observed for this model [60,61]), (2) 5XFAD and WT genotype mice

520 in the same batch, (3) information about biological sex for each sample (so that sex-specific
521 effects could be accounted for in the statistical model for differential gene expression), and
522 (4) separation of groups by biological factors (e.g. age, genotype) rather than unidentifiable
523 batch effects when inspected with Principal Component Analysis. Our original rationale for
524 analyzing the 5XFAD mouse model came from results from an independent concordance
525 analysis from various mouse models and human AD [47]. In that study, it was found that
526 5XFAD mice brain transcriptomes showed greatest (although still fairly poor) concordance to
527 human AD compared to other mouse models including the 3xTg model. However, we also
528 acknowledge the limitations associated with only analyzing the 5XFAD mouse model here,
529 and note that it would be useful to analyze other mouse models of AD in the future to
530 compare the involvement of IRE genes across different models. Likewise, we anticipate that
531 further analysis of fAD-like mutation models in general will help to extend our knowledge of
532 the early molecular changes contributing to the development of AD.
533

534 Notably, both classic IRE genes (e.g. *TFRC*, *FTL*, and *FTH1*) or IRE gene sets were
535 sufficiently sensitive to distinguish not only between iron overload and deficiency in a
536 cultured cell line dataset, but also between AD and other neuropathologies (i.e. PA and
537 PSP). This suggests that the dysregulation of IRE-containing genes and iron homeostasis in
538 general in AD may differ from other neuropathological conditions. The differences in both
539 classic IRE genes and IRE gene set expression between AD and PA are of particular
540 interest. Considering that presence of amyloid pathology is required to diagnose individuals
541 with AD [62,63] and sufficient for preclinical AD [2], it will be interesting to explore further the
542 biological relevance and impact of these IRE gene expression differences using other
543 techniques in future work to gauge whether PA is a prodrome of, or a separate pathological
544 condition to, AD.

545

546 One of our original intentions for IRE gene set analysis of the human AD and animal model
547 datasets was to gain insight on whether these brains were in a state of iron deficiency or
548 overload at the gene expression-level. Unfortunately, it became clear that interpretation of
549 IRE gene expression changes was more complex than expected, both because of the use of
550 bulk brain tissue samples for the human and animal model datasets (which obscured
551 differences between cell type and tissue) and because the regulation of IRE-containing
552 genes we observed did not follow the expected paradigm. To address the former issue, in
553 future work we would like to expand the use of IRE gene sets to single cell RNA-seq
554 (scRNA-seq) datasets so that the IRE gene expression patterns of different cell types can be
555 deconvoluted. Recently, a large-scale scRNA-seq study of AD brains indicated highly cell-
556 type specific changes related to inflammation during early stages of disease progression
557 within the prefrontal cortex [64], highlighting the importance of studying iron homeostasis
558 responses at the single-cell level in the context of AD. The failure of 3' IRE genes to follow
559 the expected expression paradigm is a limitation that we did not foresee when beginning this
560 work. Findings indicating that many predicted 3' IRE genes were able to show both
561 increased and decreased expression under iron deficiency in a cultured cell line (suggesting
562 they are not stabilized as would be predicted by the IRE paradigm) and the decreased
563 stability of many 3' IRE transcripts in the fAD-mutation-like zebrafish model are interesting
564 and challenge the simplistic paradigm of stabilization of 3' IRE-containing transcripts under
565 ferrous iron deficiency. Previous research holds that under iron deficiency, 3' IRE-containing
566 genes such *TFRC* are stabilized and hence increased in expression [20] in order to
567 increase ferrous iron availability. However, our findings reveal that this same principle may
568 not apply to other 3' IRE-containing genes, including those with canonical IREs (e.g.
569 *SLC11A2* in the cultured cell line), and genes with IRE-like elements deviating from the
570 canonical IRE sequence. Many of these IRE-like elements are likely to have some
571 functionality [26–28], and their expression changes in our analyses indicate that they are

572 associated with AD and able to distinguish AD from other neuropathological conditions (i.e.
573 PSP and PA). Detailed characterization of these genes and their biological significance is
574 beyond the scope of this current work. We expect our findings to provide a starting point for
575 further research involving IRE genes and their regulation beyond the classic paradigm.

576

577 It is important to acknowledge that at least some part of the complexity seen with IRE gene
578 expression may be due to other activities known to influence expression of IRE genes,
579 including oxidative stress [65–67]), oxidative phosphorylation [68], and cell cycle control
580 [69,70]. In addition, several transcription factor motifs potentially associated with other
581 biological pathways were also found in our analysis to be over-represented in IRE gene sets.
582 These biological activities may be expected to show some alteration in AD, other
583 neurodegenerative diseases, and during normal aging, which is likely to explain at least
584 some of the IRE gene set enrichment observed in the datasets analyzed, and potentially
585 some of the expression changes that are otherwise difficult to explain (e.g. decreased
586 expression of 3' IRE genes). Changes in expression patterns in predicted IRE genes which
587 we observed in these datasets must be carefully interpreted as not only reflecting iron
588 dyshomeostasis responses, but also co-occurring changes in other linked processes. The
589 overlap between gene expression responses to iron dyshomeostasis and other processes
590 was particularly evident in our analysis of the Caco-2 cultured cell line dataset where even
591 straightforward iron overload and iron deficiency treatments resulted in significant
592 enrichment in gene sets related to the reactive oxygen species pathway and hypoxia. We
593 expect that future work deconvoluting IRE expression patterns at the cellular level will be
594 required to determine to what extent IRE expression changes are influenced by cell / tissue
595 type compared to other linked biological activities.

596

597 Most previous work has assumed that accumulation of iron in the brain with age (a
598 phenomenon observed broadly across animal phyla [71,72]) is indicative of cellular iron
599 overload driving oxidative stress [73]. Indeed, iron accumulation and oxidative stress
600 apparently contribute to the insolubility of amyloid plaques and neurofibrillary tangles in AD
601 brains [74]. However, a recent publication by Yambire et al. [75] showed that disturbed
602 lysosomal function leading to increased lysosomal pH causes a deficiency of functional
603 ferrous iron (Fe^{2+}) while non-functional ferric iron (Fe^{3+}) accumulated in lysosomes. The
604 deleterious effects of this on the brains of mice (defective mitochondrial biogenesis and
605 function and stimulation of inflammatory responses) could be alleviated by increasing the
606 levels of iron in their diet. The observations of Nixon and colleagues that acidification of the
607 endolysosomal pathway is affected both by fAD mutations in *PSEN1* [54] and excessive
608 dosage of the amyloid beta A4 precursor protein (*Aβ PP*) gene [76], together with our
609 observations from our fAD-mutation-like *psen1*^{Q96_K97del}/+ mutant zebrafish both in this study
610 and in ref. [53] (where gene expression changes associated with lysosome and mitochondria
611 for this model were first described), support the possibility that AD brains may suffer a
612 ferrous iron deficiency in a background of ferric iron overload. This is discussed further
613 below.

614

615 Although the evidence from human AD brains is more tentative at this stage, our findings of
616 increased expression of *TFRC* expression and other well-characterized 3' IRE genes in the
617 temporal cortex tissue of human AD brains is consistent with a cellular ferrous iron
618 deficiency, although more research is required to confirm this at the single-cell level.
619 Intriguingly, the greatest genetic risk factor for late onset AD, the ε4 allele of the gene *APOE*,
620 appears to increase lysosomal pH [77] but ε4's increased risk of AD is alleviated in
621 individuals who possess the *HFE* 282Y allele that predisposes to the iron overload disease
622 hemochromatosis [78]. Given that many other risk loci for sporadic late onset AD also affect

623 endolysosomal pathway function (reviewed in [79]) it is reasonable to suggest that disturbed
624 iron homeostasis may afflict brains with this disease.

625

626 The proposal that fAD mutations drive a ferrous iron deficiency that contributes to AD
627 pathology is not necessarily in disagreement with the experimental data recently reported in
628 Tsatsanis et al. [80] regarding an association between amyloidogenic processing of AβPP
629 and impaired export of iron from cells. In that study, the authors extended investigation of
630 the idea that the protein product of the fAD locus *AβPP* stabilizes the iron exporting protein
631 ferroportin, providing a connection between *AβPP* and iron homeostasis in AD. The idea that
632 AβPP interacts with ferroportin has been the subject of some contention [81,82]. Tsatsanis et
633 al. observed that inhibition of the amyloidogenic cleavage of AβPP by the enzyme β-
634 secretase in mouse N2a neuroblastoma cells is associated with stabilization of ferroportin on
635 the cell surface and a reduced, intracellular labile iron pool. However, these authors did not
636 consider the evidence that the ferroxidase hephaestin, (well established as complexing with,
637 and stabilizing, ferroportin to export iron [83–85]), is likely also cleaved by β-secretase (see
638 Figure 7 in [86]). It is probable that reduced β-secretase activity increases the hephaestin
639 available to stabilize ferroportin.

640

641 We have previously noted that all the fAD mutations of AβPP (including gene duplications)
642 are expected to have the singular quality of increasing the abundance of the amyloidogenic
643 β-CTF / C99 fragment of AβPP (see Figure 2 in Lumsden et al. [14]). Recently Jiang et al.
644 [76] showed that increased β-CTF reduces lysosomal acidification. According to the results
645 of Yambire et al. cited above, this would be expected to reduce cellular ferrous iron levels
646 (while causing accumulation of ferric iron in lysosomes). This idea is consistent with the
647 existence of an IRE in the 5' UTR of AβPP transcripts that inhibits AβPP translation (and

648 hence β-CTF production) when cytosolic ferrous iron levels are low [87]. Thus, AβPP may
649 influence cellular ferrous iron levels by modulating lysosomal pH to regulate importation of
650 ferrous iron into the cytosol, rather than by modulating ferroportin stability to regulate iron
651 export from the cytosol.
652

653 Overall, our results demonstrate for the first time the involvement of a coordinated IRP/IRE-
654 mediated gene expression response in the context of AD. By searching entire human,
655 mouse, and zebrafish transcriptomes for all genes containing potential IREs, we formed
656 comprehensive gene sets applicable to the analysis of any human, mouse, or zebrafish
657 gene expression dataset. For the first time, our work suggests the existence of gene
658 expression changes associated with the IRP/IRE system in the young, pre-pathology brains
659 of *PSEN1* fAD mutation carriers, and is consistent with the idea of iron dyshomeostasis as
660 an early contributor to the disease process. It also reinforces the critical importance of IRE
661 genes in general in the development of the late onset, sporadic form of AD. More broadly,
662 our approach highlights how changes in the stability and abundance of IRE-containing
663 transcripts can be used to give insight into iron dyshomeostasis-associated gene expression
664 responses in different species, tissues, and conditions.

665

Methods

Defining IRE gene sets for human, mouse, and zebrafish

We extracted all 3' and 5' UTR sequences from the human, mouse and zebrafish genome assemblies using the Bioconductor packages BSgenome.Hsapiens.UCSC.hg38, BSgenome.Mmusculus.UCSC.mm10 and BSgenome.Drerio.UCSC.danRer11, and gene definitions from the Ensembl 94 release. Each set of UTR sequences was then submitted input into the SIREs web server (v.2.0 [24]). The SIREs algorithm assigns quality scores to predicted IREs taking into account whether the sequence is canonical, whether it contains any mismatches and bulges, and the free energy of the secondary structure. Canonical sequences are tagged by SIREs as being high-quality, while IRE-like sequences are tagged as low or medium quality. Given that High-quality IRE predictions miss the majority of true IREs, this enables a more comprehensive sampling of IRE motifs. For human, mouse, and zebrafish, we separately defined the following four gene sets: ***HQ 3' IREs*** and ***HQ 5' IREs*** (representing genes with high-quality predicted IREs), along with ***all 3' IREs*** and ***all 5' IREs***, which included genes containing any predicted IRE in the respective UTR. Comparisons between gene sets were performed using the UpSetR package (v.1.4.0 [88]) with gene ID mappings between species obtained by BioMart [89].

Over-representation of IRE gene sets in existing MSigDB gene sets

We downloaded the following gene set collections from MSigDB (v.6.0 [35]): Hallmark, C2 (gene sets from online pathway databases and biomedical literature, including KEGG and the REACTOME databases), C3 (motif gene sets based on regulatory targets), and C5 (gene sets defined by Gene Ontology terms). We excluded the following collections from analysis: C1 (specific to human chromosomes and cytogenetic bands, while our analysis involves different species), C4 (computationally-defined cancer-focused gene sets), C6

691 (oncogenic signatures) and C7 (immunologic signatures). C4, C6, and C7 were not included
692 as the level of detail in the gene sets in these specific collections is more domain-specific
693 rather than broad-level. We used Fisher's exact test to determine whether any IRE gene set
694 was significantly over-represented in each MSigDB gene set. Gene sets were defined as
695 having significant enrichment for IRE gene sets if the FDR-adjusted *p*-value from Fisher's
696 exact test was below 0.05. UpSet plots were produced using the UpSetR package (v.1.4.0
697 [88]) while network representations were produced in Gephi (v.0.9.3 [90]). To produce
698 network visualizations, we first exported node and edge tables from R. The nodes table
699 contained the following gene sets: top 15 gene sets (ranked by Fisher's exact test *p*-value),
700 Hallmark Heme Metabolism gene set, ***all 3' IREs***, ***all 5' IREs***, and all genes contained within
701 these gene sets. The edges table contained gene – gene set edges which indicated the
702 gene set(s) that genes belonged to. To create the network plots in Gephi, we used "Force
703 Atlas 2" as the initial layout algorithm, followed by the "Yifan Hu" [91] layout algorithm to
704 improve visual separation between distinct groupings of genes.
705

706 Over-representation analysis of transcription factor motifs in IRE gene
707 promoters

708 Defining promoter regions as being 1500 bp upstream and 200 bp downstream of the
709 transcription start site for each gene, we used the findMotifs.pl script from HOMER (v.4.11)
710 [92,93] to search for known transcription factor binding site (TFBS) motifs in the promoters of
711 each IRE gene set. The HOMER Motif database contains 363 vertebrate transcription factor
712 binding motifs based on analysis of high-quality public ChIP-seq datasets
713 (http://homer.ucsd.edu/homer/motif/HomerMotifDB/homerResults.html). We considered
714 TFBS motifs as being significantly enriched in a gene set if the FDR-adjusted *p*-value (q-
715 value provided in HOMER output) was less than 0.05.
716

717 ## Gene set enrichment testing

718 We performed all gene set enrichment tests in R v3.6.1 [94] using *fry* [37,38], *camera* [39],
719 and *fgsea* [40,41]. For *fry*, and *camera,* we used model fits obtained using *limma* [95,96],
720 whilst for *fgsea*, a ranked list was obtained using *moderated t*-statistics taken from *limma*. All
721 genes were used in gene set enrichment tests (i.e. not just DE genes). We combined the
722 raw *p*-values from *fry, camera*, and *fgsea* using Wilkinson's method [97] with default
723 parameters, followed by FDR-adjustment. When performing gene set enrichment testing on
724 the MSigDB Hallmark gene sets, we applied FDR-adjustment to combined *p*-values and
725 defined significant enrichment as gene sets having an adjusted *p*-value < 0.05. When
726 performing gene set enrichment on the four IRE gene sets (***all 3' IREs***, ***all 5' IREs***, ***HQ 3'***
727 ***IREs***, ***HQ 5' IREs***), we applied Bonferroni-adjustment to combined *p*-values to further
728 protect against Type I errors and defined significant enrichment as gene sets having an
729 adjusted *p*-value < 0.05. Depending on the species in the dataset being analyzed, we used
730 the respective IRE gene sets defined for human, mouse, or zebrafish.
731

732 ## Analysis of the Caco-2 cultured cell line dataset

733 We downloaded processed microarray data from GEO (accession number: GSE3573). This
734 study investigated gene expression responses to iron treatments, including iron deficiency
735 (cells in iron-free medium vs. cells cultivated in ferric ammonium nitrate, and iron overload
736 (cells cultivated in DMEM-FBS medium with hemin vs. cells in DMEM-FBS medium) [98].
737 We performed differential gene expression analysis using the "lmFit" and eBayes" functions
738 in *limma* [95]. Genes were defined as differentially expressed when their FDR-adjusted *p*-
739 value < 0.05.
740

741 ## Analysis of the Mayo Clinic RNA-seq dataset

742 No experiments of human subjects were performed in this study as all human data was
743 accessed from public databases and was anonymous. We downloaded processed CPM
744 count data from Synapse (https://www.synapse.org/#!Synapse:syn5550404). We matched
745 cerebellum and temporal cortex samples by their patient ID, and only retained genes which
746 were present across all samples and patients for which there were both cerebellum and
747 temporal cortex samples (n=236 patients with measurements for cerebellum and temporal
748 cortex, 472 samples in total). We performed analysis using *limma* [95,96] and determined
749 differentially expressed genes between conditions. In addition, we used the
750 "duplicateCorrelation" function in *limma*, setting the "block" parameter to the patient ID.
751 Genes were considered differentially expressed if their FDR-adjusted p-value < 0.05.
752

753 ## Zebrafish husbandry and animal ethics

754 All zebrafish work was conducted under the auspices of the Animal Ethics Committee
755 (permit numbers S-2017-089 and S-2017-073) and the Institutional Biosafety Committee of
756 the University of Adelaide. Tübingen strain zebrafish were maintained in a recirculated water
757 system.
758

759 ## fAD-like *psen1*$^{Q96_K97del/+}$ zebrafish

760 The isolation of the *psen1*^{Q96_K97del} mutation has previously been described [53]. Mutations
761 were only analyzed in the heterozygous state in this study.
762

763 ## Acute hypoxia treatment of zebrafish

764 *psen1*$^{Q96_K97del/+}$ mutants and their wild-type siblings were treated in low oxygen levels by
765 placing zebrafish in oxygen-depleted water for 3 hours (oxygen concentration of 6.6 ± 0.2
766 mg/L in normoxia and 0.6 ± 0.2 mg/L in hypoxia.
767

768 ## Whole brain removal from adult zebrafish

769 After normoxia or hypoxia treatment adult fish were euthanized by sudden immersion in an
770 ice water slurry for at least ~30 seconds before decapitation and removal of the entire brain
771 for immediate RNA or protein extraction. All fish brains were removed during late
772 morning/noon to minimize any influence of circadian rhythms.
773

774 ## RNA extraction from whole brain

775 Total RNA was isolated from heterozygous mutant and WT siblings using the *mir*Vana
776 miRNA isolation kit (Thermo Fisher). RNA isolation was performed according to the
777 manufacturer's protocol. First a brain was lysed in a denaturing lysis solution. The lysate
778 was then extracted once with acid-phenol:chloroform leaving a semi-pure RNA sample. The
779 sample was then purified further over a glass-fiber filter to yield total RNA. Total RNA was
780 DNase treated using the DNA-free™ Kit from Ambion, Life Technologies according to the
781 manufacturer's instructions. Total RNA was then sent to the Genomics Facility at the South
782 Australian Health and Medical Research Institute (Adelaide, Australia) to assess RNA quality
783 and for subsequent RNA sequencing *(using poly-A enriched RNA-seq technology, and*
784 *estimated gene expression from the resulting single-end 75 bp reads using the reference*
785 *GRCz11 zebrafish assembly transcriptome)*

786

787 ## Pre-processing of the fAD-like *psen1*$^{Q96_K97del/+}$ zebrafish dataset:

788 RNA-seq libraries contained single-end 75bp Illumina NextSeq reads. We performed quality
789 trimming with *AdapterRemoval* using default parameters, followed by quality assessment
790 with FastQC and ngsReports. Trimmed reads were pseudo-aligned to the reference
791 zebrafish transcriptome using *Kallisto* (v.0.45) [99] and transcript descriptions from Ensembl
792 release 94. The "catchKallisto" function from *edgeR* [100] was used to import and
793 summarise counts from transcript-level to gene-level, with all subsequent analyses
794 performed at the gene-level.

795

796 ## Differential gene expression analysis for fAD-like zebrafish dataset

797 For differential gene expression analysis, we retained all genes with expression of at least 1
798 cpm in 4 or more samples, and used voomWithQualityWeights to downweight lower quality
799 samples [96]. Contrasts were defined to include all relevant pairwise comparisons between
800 conditions, and genes were considered as differentially expressed using an FDR-adjusted *p*-
801 value < 0.05. The *limma* model used to create the design matrix for this dataset included
802 biological sex as a variable (~0 + biologicalSex + group, where biologicalSex is either male
803 or female and group refers to the combination of age, hypoxia treatment, and genotype of
804 the samples) to minimize any sex-specific effects in the gene expression analysis. More
805 details regarding the modest effects of biological sex in this dataset are provided in
806 **Supplementary Text 3**.

807

808 Estimation of spliced and unspliced gene expression in fAD-like zebrafish dataset
809 For spliced transcripts in Ensembl release 94, we additionally defined unspliced genes
810 including intronic regions. Unspliced transcripts were appended to the end of the reference

811 transcriptome and used to build a new Kallisto [99] index. Estimated counts for spliced
812 transcripts and unspliced genes were imported into R using the “catchKallisto” function from
813 *edgeR* [100].
814

815 Gene set enrichment tests for non-fAD-like (K97fs/+) zebrafish dataset

816 Please refer to ref. [53] for details on RNA-seq data processing and analysis of the non-fAD-
817 like dataset. In the current work, we used the gene expression counts matrix with *limma*. The
818 voom, design, and contrasts objects produced as part of the *limma* analysis were used for
819 gene set enrichment analysis with the zebrafish IRE gene sets we defined as well as the
820 MSigDB Hallmark gene sets. Significantly enriched Hallmark gene sets had FDR-adjusted *p*-
821 value < 0.05 while IRE gene sets were considered significantly enriched if the Bonferroni-
822 adjusted *p*-value < 0.05.
823

824 Gene set enrichment tests for 7-day-old Q96_K97del/+ dataset

825 Please refer to ref. [59] (GEO accession: GSE148631) for details on RNA-seq data
826 processing and analysis. In the current work, we used the gene expression counts matrix
827 with *limma* to perform gene set enrichment analysis. IRE gene sets we defined as well as
828 the MSigDB Hallmark gene sets. Significantly enriched Hallmark gene sets had FDR-
829 adjusted *p*-value < 0.05 while IRE gene sets were considered significantly enriched if the
830 Bonferroni-adjusted *p*-value < 0.05.
831

832 Differential transcript stability analysis

833 The estimated spliced and unspliced transcript count estimates from *kallisto* [99] were
834 imported into R using the catchKallisto function from *edgeR* [100]. We used *limma* [95] to
835 determine the logFC of spliced transcripts and unspliced transcripts for each comparison. To

836 test for whether there was a significant difference in the logFC of the spliced and unspliced
837 transcripts, we used Welch's *t*-test with the *s2.prior* values from *limma* as the variances of
838 the spliced and unspliced transcripts. We defined the null (no stabilization of transcript) and
839 alternate (stabilization of transcript) hypotheses for each gene as follows, where s and u
840 refer to the spliced and unspliced versions of a particular gene:
841

842 $$H_0\colon\ \mathrm{logFC_s} = \mathrm{logFC_u}$$

843 $$H_a\colon\ \mathrm{logFC_s} \neq \mathrm{logFC_u}$$

844

845 We defined genes with FDR-adjusted *p*-values < 0.05 as having differential stability.
846

847 Checks that observed gene expression differences are not artifacts of
848 differences in proportions of cell types

849 The Mayo Clinic RNA-seq study, 5XFAD mice, and fAD-like zebrafish datasets are bulk
850 RNA-seq datasets. To confirm that any gene expression changes were likely due to altered
851 transcriptional programs rather than changes in cell type proportions, we compared
852 expression of marker genes for four common neural cell types (astrocytes, neurons,
853 oligodendrocytes, microglia) in conditions within each dataset. The marker genes for
854 astrocytes, neurons, and oligodendrocytes were obtained from MSigDB gene sets from [101]
855 which were based on studies in mice. The marker genes for microglia were derived from
856 [102] which was based on studies in human and mouse. All gene IDs were converted to
857 human, mouse, or zebrafish Ensembl IDs using BioMart [89] for each dataset. Please refer
858 to **Supplementary Text 4** and **Supplementary Figures 8-10** for details.
859
860
861
862

Declarations

Acknowledgements and Funding

The authors thank the Carthew Family and Prof. David Adelson for their encouragement and helpful discussions. This work was financially supported by grants GNT1061006 and GNT1126422 from the National Health and Medical Research Council of Australia (NHMRC) and by funds from the Carthew Family Charity Trust. Lead author NH was supported by an Australian Government Research Training Program Scholarship and by the South Australian Genomics Centre after graduation. Co-author MN was supported by funds from the grants listed above. Joint senior co-authors SP and ML were/are academic employees of the University of Adelaide respectively. The authors also thank the University of Adelaide's Phoenix HPC team for providing supercomputing resources for this work.

Conflict of Interest / Disclosure Statement

The authors have no conflict of interest to report.

Ethics approval and consent to participate

No experiments of human subjects were performed in this study as all human data was accessed from public databases and was anonymous. All zebrafish work was conducted under the auspices of the Animal Ethics Committee (permit numbers S-2017-089 and S-2017-073) and the Institutional Biosafety Committee of the University of Adelaide.

Consent for publication

All authors give consent for publication of the manuscript.

887 ## Datasets and Data Articles

888 The fAD-like zebrafish dataset supporting the conclusions of this article is available in the
889 GEO repository with accession number GSE149149.

890

891 ## Authors' contributions

892 NH performed the bioinformatics analysis and drafted the manuscript. MN generated the
893 mutant zebrafish, isolated brain RNA and protein, drafted the description of this and edited
894 the manuscript. SP directly supervised NH, guided the analysis, and edited the manuscript.
895 ML conceived the project, supervised all aspects of it, and edited the manuscript. All authors
896 read and approved the manuscript.

897

898

899

900

901 # References

902

903 [1] Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB,
904 Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine
905 T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM,
906 Sperling R, Elliott C, Masliah E, Ryan L, Silverberg N (2018) NIA-AA Research
907 Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's*
908 *Dement* **14**, 535–562.

909 [2] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo
910 T, Jack CR, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E,
911 Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster M V.,
912 Phelps CH (2011) Toward defining the preclinical stages of Alzheimer's
913 disease: Recommendations from the National Institute on Aging-Alzheimer's
914 Association workgroups on diagnostic guidelines for Alzheimer's disease.
915 *Alzheimer's Dement* **7**, 280–292.

916 [3] Bartzokis G, Sultzer D, Mintz J, Holt LE, Marx P, Kelly Phelan C, Marder SR
917 (1994) In vivo evaluation of brain iron in Alzheimer's disease and normal
918 subjects using MRI. *Biol Psychiatry* **35**, 480–487.

919 [4] Bartzokis G, Tishler TA (2000) MRI evaluation of basal ganglia ferritin iron and
920 neurotoxicity in Alzheimer's and Huntingon's disease. *Cell Mol Biol (Noisy-le-*
921 *grand)* **46**, 821–833.

922 [5] Ding B, Chen KM, Ling HW, Sun F, Li X, Wan T, Chai WM, Zhang H, Zhan Y,
923 Guan YJ (2009) Correlation of iron in the hippocampus with MMSE in patients
924 with alzheimer's disease. *J Magn Reson Imaging* **29**, 793–798.

925 [6] Langkammer C, Ropele S, Pirpamer L, Fazekas F, Schmidt R (2014) MRI for
926 iron mapping in Alzheimer's disease. *Neurodegener Dis* **13**, 189–191.
927 [7] Van Duijn S, Bulk M, Van Duinen SG, Nabuurs RJA, Van Buchem MA, Van
928 Der Weerd L, Natté R (2017) Cortical Iron Reflects Severity of Alzheimer's
929 Disease. *J Alzheimer's Dis* **60**, 1533–1545.
930 [8] Mills E, Dong XP, Wang F, Xu H (2010) Mechanisms of brain iron transport:
931 Insight into neurodegeneration and CNS disorders. *Future Med Chem* **2**, 51–
932 64.
933 [9] Yamamoto A, Shin R-W, Hasegawa K, Naiki H, Sato H, Yoshimasu F,
934 Kitamoto T (2004) Iron (III) induces aggregation of hyperphosphorylated τ and
935 its reduction to iron (II) reverses the aggregation: implications in the formation
936 of neurofibrillary tangles of Alzheimer's disease. *J Neurochem* **82**, 1137–1147.
937 [10] Telling ND, Everett J, Collingwood JF, Dobson J, van der Laan G, Gallagher
938 JJ, Wang J, Hitchcock AP (2017) Iron Biochemistry is Correlated with Amyloid
939 Plaque Morphology in an Established Mouse Model of Alzheimer's Disease.
940 *Cell Chem Biol* **24**, 1205-1215.e3.
941 [11] Leskovjan AC, Kretlow A, Lanzirotti A, Barrea R, Vogt S, Miller LM (2011)
942 Increased brain iron coincides with early plaque formation in a mouse model of
943 Alzheimer's disease. *Neuroimage* **55**, 32–38.
944 [12] Gong NJ, Dibb R, Bulk M, van der Weerd L, Liu C (2019) Imaging beta
945 amyloid aggregation and iron accumulation in Alzheimer's disease using
946 quantitative susceptibility mapping MRI. *Neuroimage* **191**, 176–185.
947 [13] El Tannir El Tayara N, Delatour B, Le Cudennec C, Guégan M, Volk A,
948 Dhenain M (2006) Age-related evolution of amyloid burden, iron load, and MR

949 relaxation times in a transgenic mouse model of Alzheimer's disease.
950 *Neurobiol Dis* **22**, 199–208.
951 [14] Lumsden AL, Rogers JT, Majd S, Newman M, Sutherland GT, Verdile G,
952 Lardelli M (2018) Dysregulation of neuronal iron homeostasis as an alternative
953 unifying effect of mutations causing familial alzheimer's disease. *Front*
954 *Neurosci* **12**, 533.
955 [15] Lane DJR, Merlot AM, Huang MLH, Bae DH, Jansson PJ, Sahni S, Kalinowski
956 DS, Richardson DR (2015) Cellular iron uptake, trafficking and metabolism:
957 Key molecules and mechanisms and their roles in disease. *Biochim Biophys*
958 *Acta - Mol Cell Res* **1853**, 1130–1144.
959 [16] Gray NK, Hentze MW (1994) Iron regulatory protein prevents binding of the
960 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *EMBO J*
961 **13**, 3882–3891.
962 [17] Hentze MW, Muckenthaler MU, Andrews NC (2004) Balancing acts: Molecular
963 control of mammalian iron metabolism. *Cell* **117**, 285–297.
964 [18] Muckenthaler MU, Galy B, Hentze MW (2008) Systemic Iron Homeostasis and
965 the Iron-Responsive Element/Iron-Regulatory Protein (IRE/IRP) Regulatory
966 Network. *Annu Rev Nutr* **28**, 197–213.
967 [19] Basilion JP, Rouault TA, Massinople CM, Klausner RD, Burgesst WH (1994)
968 The iron-responsive element-binding protein: Localization of the RNA-binding
969 site to the aconitase active-site cleft (UV cross-linking/base hydrolysis).
970 *Biochemistry* **91**, 574–578.
971 [20] Pantopoulos K (2004) Iron metabolism and the IRE/IRP regulatory system: An
972 update. *Ann N Y Acad Sci* **1012**, 1–13.

973 [21] Piccinelli P, Samuelsson T (2007) Evolution of the iron-responsive element.
974 *RNA* **13**, 952–966.
975 [22] Gunshin H, Allerson CR, Polycarpou-Schwarz M, Rofts A, Rogers JT, Kishi F,
976 Hentze MW, Rouault TA, Andrews NC, Hediger MA (2001) Iron-dependent
977 regulation of the divalent metal ion transporter. *FEBS Lett* **509**, 309–316.
978 [23] Cmejla R, Petrak J, Cmejlova J (2006) A novel iron responsive element in the
979 3′UTR of human MRCKα. *Biochem Biophys Res Commun* **341**, 158–166.
980 [24] Campillos M, Cases I, Hentze MW, Sanchez M (2010) SIREs: searching for
981 iron-responsive elements. *Nucleic Acids Res* **38**, W360-367.
982 [25] Goforth JB, Anderson SA, Nizzi CP, Eisenstein RS (2010) Multiple
983 determinants within iron-responsive elements dictate iron regulatory protein
984 binding and regulatory hierarchy. *RNA* **16**, 154–169.
985 [26] Henderson BR, Menotti E, Kühn LC (1996) Iron regulatory proteins 1 and 2
986 bind distinct sets of RNA target sequences. *J Biol Chem* **271**, 4900–4908.
987 [27] Henderson BR, Menotti E, Bonnard C, Kuhn LC (1994) Optimal sequence and
988 structure of iron-responsive elements. Selection of RNA stem-loops with high
989 affinity for iron regulatory factor. *J Biol Chem* **269**, 17481–17489.
990 [28] Butt J, Kim HY, Basilion JP, Cohen S, Iwai K, Philpott CC, Altschul S, Klausner
991 RD, Rouault TA (1996) Differences in the RNA binding sites of iron regulatory
992 proteins and potential target diversity. *Proc Natl Acad Sci U S A* **93**, 4345–
993 4349.
994 [29] Surinya KH, Cox TC, May BK (1998) Identification and characterization of a
995 conserved erythroid-specific enhancer located in intron 8 of the human 5-
996 aminolevulinate synthase 2 gene. *J Biol Chem* **273**, 16798–16809.

997 [30] Theil EC (1990) Regulation of ferritin and transferrin receptor mRNAs. *J Biol*
998 *Chem* **265**, 4771–4774.
999 [31] Sanchez M, Galy B, Dandekar T, Bengert P, Vainshtein Y, Stolte J,
1000 Muckenthaler MU, Hentze MW (2006) Iron regulation and the cell cycle:
1001 Identification of an iron-responsive element in the 3′-untranslated region of
1002 human cell division cycle 14A mRNA by a refined microarray-based screening
1003 strategy. *J Biol Chem* **281**, 22865–22874.
1004 [32] Cmejla R, Ptackova P, Petrak J, Savvulidi F, Cerny J, Sebesta O, Vyoral D
1005 (2010) Human MRCKα is regulated by cellular iron levels and interferes with
1006 transferrin iron uptake. *Biochem Biophys Res Commun* **395**, 163–167.
1007 [33] Lymboussaki A, Pignatti E, Montosi G, Garuti C, Haile DJ, Pietrangelo A
1008 (2003) The role of the iron responsive element in the control of
1009 ferroportin1/IREG1/MTP1 gene expression. *J Hepatol* **39**, 710–715.
1010 [34] Aziz N, Munro HN (1987) Iron regulates ferritin mRNA translation through a
1011 segment of its 5' untranslated region. *Proc Natl Acad Sci U S A* **84**, 8478–
1012 8482.
1013 [35] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P,
1014 Mesirov JP, Bateman A (2011) Databases and ontologies Molecular
1015 signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740.
1016 [36] Alhamdoosh M, Law CW, Tian L, Sheridan JM, Ng M, Ritchie ME (2017) Easy
1017 and efficient ensemble gene set testing with EGSEA. *F1000Research* **6**, 2010.
1018 [37] Wu D, Lim E, Vaillant F, Asselin-Labat M-L, Visvader JE, Smyth GK (2010)
1019 ROAST: rotation gene set tests for complex microarray experiments.
1020 *Bioinformatics* **26**, 2176–2182.

1021 [38] Giner G, Smyth GK, Giner G, Smyth GK (2016) FRY: a fast approximation
1022 to ROAST gene set test with mean aggregated set statistics .
1023 *F1000Research* **5**,.
1024 [39] Wu D, Smyth GK (2012) Camera: a competitive gene set test accounting for
1025 inter-gene correlation. *Nucleic Acids Res* **40**, e133.
1026 [40] Sergushichev AA (2016) An algorithm for fast preranked gene set enrichment
1027 analysis using cumulative statistic calculation. *bioRxiv* 060012.
1028 [41] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA,
1029 Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set
1030 enrichment analysis: A knowledge-based approach for interpreting genome-
1031 wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550.
1032 [42] Allen M, Carrasquillo MM, Funk C, Heavner BD, Zou F, Younkin CS, Burgess
1033 JD, Chai HS, Crook J, Eddy JA, Li H, Logsdon B, Peters MA, Dang KK, Wang
1034 X, Serie D, Wang C, Nguyen T, Lincoln S, Malphrus K, Bisceglio G, Li M,
1035 Golde TE, Mangravite LM, Asmann Y, Price ND, Petersen RC, Graff-Radford
1036 NR, Dickson DW, Younkin SG, Ertekin-Taner N (2016) Human whole genome
1037 genotype and transcriptome data for Alzheimer's and other neurodegenerative
1038 diseases. *Sci Data* **3**, 1–10.
1039 [43] Ossenkoppele R, Van Berckel BN, Prins ND (2011) Amyloid imaging in
1040 prodromal Alzheimer's disease. *Alzheimer's Res Ther* **3**, 26.
1041 [44] Protas HD, Chen K, Langbaum JBS, Fleisher AS, Alexander GE, Lee W,
1042 Bandy D, De Leon MJ, Mosconi L, Buckley S, Truran-Sacrey D, Schuff N,
1043 Weiner MW, Caselli RJ, Reiman EM (2013) Posterior cingulate glucose
1044 metabolism, hippocampal glucose metabolism, and hippocampal volume in

1045 cognitively normal, late-middle-aged persons at 3 levels of genetic risk for
1046 alzheimer disease. *JAMA Neurol* **70**, 320–325.
1047 [45] Caldwell CC, Yao J, Brinton RD (2015) Targeting the Prodromal Stage of
1048 Alzheimer's Disease: Bioenergetic and Mitochondrial Opportunities.
1049 *Neurotherapeutics* **12**, 66–80.
1050 [46] Vermunt L, Sikkes SAM, van den Hout A, Handels R, Bos I, van der Flier WM,
1051 Kern S, Ousset PJ, Maruff P, Skoog I, Verhey FRJ, Freund-Levi Y, Tsolaki M,
1052 Wallin ÅK, Olde Rikkert M, Soininen H, Spiru L, Zetterberg H, Blennow K,
1053 Scheltens P, Muniz-Terrera G, Visser PJ, Vellas B, Reynish E, Ousset PJ,
1054 Andrieu S, Burns A, Pasquier F, Frisoni G, Salmon E, Michel JP, Zekry DS,
1055 Boada M, Dartigues JF, Olde-Rikkert MGM, Rigaud AS, Winblad B, Malick A,
1056 Sinclair A, Frölich L, Scheltens P, Ribera C, Touchon J, Robert P, Salva A,
1057 Waldemar G, Bullock R, Tsolaki M, Rodriguez G, Spiru L, Jones RW, Stiens
1058 G, Stoppe G, Eriksdotter Jönhagen M, Cherubini A, Lage PM, Gomez-Isla T,
1059 Camus V, Agüera-Morales E, Lopez F, Savy S, Cantet C, Coley N (2019)
1060 Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease
1061 in relation to age, sex, and APOE genotype. *Alzheimer's Dement* **15**, 888–898.
1062 [47] Hargis KE, Blalock EM (2017) Transcriptional signatures of brain aging and
1063 Alzheimer's disease: What are our rodent models telling us? *Behav Brain Res*
1064 **322**, 311–328.
1065 [48] Bundy JL, Vied C, Badger C, Nowakowski RS (2019) Sex-biased hippocampal
1066 pathology in the 5XFAD mouse model of Alzheimer's disease: A multi-omic
1067 analysis. *J Comp Neurol* **527**, 462–475.
1068 [49] Drummond E, Wisniewski T (2017) Alzheimer's disease: experimental models

1069 and reality. *Acta Neuropathol* **133**, 155–175.

1070 [50] Xia D, Watanabe H, Wu B, Lee SH, Li Y, Tsvetkov E, Bolshakov VY, Shen J,
1071 Kelleher RJ (2015) Presenilin-1 knockin mice reveal loss-of-function
1072 mechanism for familial alzheimer's disease. *Neuron* **85**, 967–981.

1073 [51] Saito T, Matsuba Y, Mihira N, Takano J, Nilsson P, Itohara S, Iwata N, Saido
1074 TC (2014) Single App knock-in mouse models of Alzheimer's disease. *Nat*
1075 *Neurosci* **17**, 661–663.

1076 [52] Serneels L, T'Syen D, Perez-Benito L, Theys T, Holt MG, De Strooper B
1077 (2020) Modeling the β-secretase cleavage site and humanizing amyloid-beta
1078 precursor protein in rat and mouse to study Alzheimer's disease. *Mol*
1079 *Neurodegener* **15**, 60.

1080 [53] Newman M, Hin N, Pederson S, Lardelli M (2019) Brain transcriptome analysis
1081 of a familial Alzheimer's disease-like mutation in the zebrafish presenilin 1
1082 gene implies effects on energy production. *Mol Brain* **12**, 43.

1083 [54] Lee JH, Yu WH, Kumar A, Lee S, Mohan PS, Peterhoff CM, Wolfe DM,
1084 Martinez-Vicente M, Massey AC, Sovak G, Uchiyama Y, Westaway D, Cuervo
1085 AM, Nixon RA (2010) Lysosomal proteolysis and autophagy require presenilin
1086 1 and are disrupted by Alzheimer-related PS1 mutations. *Cell* **141**, 1146–1158.

1087 [55] Lee JH, McBrayer MK, Wolfe DM, Haslett LJ, Kumar A, Sato Y, Lie PPY,
1088 Mohan P, Coffey EE, Kompella U, Mitchell CH, Lloyd-Evans E, Nixon RA
1089 (2015) Presenilin 1 Maintains Lysosomal Ca2+ Homeostasis via TRPML1 by
1090 Regulating vATPase-Mediated Lysosome Acidification. *Cell Rep* **12**, 1430–
1091 1444.

1092 [56] Yu Y, Richardson DR (2011) Cellular iron depletion stimulates the JNK and

1093 p38 MAPK signaling transduction pathways, dissociation of ASK1-thioredoxin,
1094 and activation of ASK1. *J Biol Chem* **286**, 15413–15427.
1095 [57] Riera H, Afonso V, Collin P, Lomri A (2015) A Central Role for JNK/AP-1
1096 Pathway in the Pro-Oxidant Effect of Pyrrolidine Dithiocarbamate through
1097 Superoxide Dismutase 1 Gene Repression and Reactive Oxygen Species
1098 Generation in Hematopoietic Human Cancer Cell Line U937. *PLoS One* **10**,
1099 e0127571.
1100 [58] Hin N, Newman M, Kaslin J, Douek AM, Lumsden A, Nik SHM, Dong Y, Zhou
1101 XF, Manucat-Tan NB, Ludington A, Adelson DL, Pederson S, Lardelli M (2020)
1102 Accelerated brain aging towards transcriptional inversion in a zebrafish model
1103 of the K115fs mutation of human PSEN2. *PLoS One* **15**, e0227258.
1104 [59] Dong Y, Newman M, Pederson S, Hin N, Lardelli M (2020) Transcriptome
1105 analyses of 7-day-old zebrafish larvae possessing a familial Alzheimer's
1106 disease-like mutation in psen1 indicate effects on oxidative phosphorylation,
1107 mcm functions, and iron homeostasis. *bioRxiv* 2020.05.03.075424.
1108 [60] Oakley H, Cole SL, Logan S, Maus E, Shao P, Craft J, Guillozet-Bongaarts A,
1109 Ohno M, Disterhoft J, Van Eldik L, Berry R, Vassar R (2006) Intraneuronal
1110 beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic
1111 mice with five familial Alzheimer's disease mutations: potential factors in
1112 amyloid plaque formation. *J Neurosci* **26**, 10129–10140.
1113 [61] Landel V, Baranger K, Virard I, Loriod B, Khrestchatisky M, Rivera S, Benech
1114 P, Féron F (2014) Temporal gene profiling of the 5XFAD transgenic mouse
1115 model highlights the importance of microglial activation in Alzheimer's disease.
1116 *Mol Neurodegener* **9**,.

1117 [62] Deture MA, Dickson DW (2019) The neuropathological diagnosis of
1118 Alzheimer's disease. *Mol Neurodegener* **14**, 1–18.
1119 [63] Jack CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC,
1120 Thies B, Phelps CH (2011) Introduction to the recommendations from the
1121 National Institute on Aging-Alzheimer's Association workgroups on diagnostic
1122 guidelines for Alzheimer's disease. *Alzheimer's Dement* **7**, 257–262.
1123 [64] Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ,
1124 Menon M, He L, Abdurrob F, Jiang X, Martorell AJ, Ransohoff RM, Hafler BP,
1125 Bennett DA, Kellis M, Tsai LH (2019) Single-cell transcriptomic analysis of
1126 Alzheimer's disease. *Nature* **570**, 332–337.
1127 [65] Martins EAL, Robalinho RL, Meneghini R (1995) Oxidative stress induces
1128 activation of a cytosolic protein responsible for control of iron uptake. *Arch*
1129 *Biochem Biophys* **316**, 128–134.
1130 [66] Pantopoulos K, Hentze MW (1995) Rapid responses to oxidative stress
1131 mediated by iron regulatory protein. *EMBO J* **14**, 2917–2924.
1132 [67] Yang L, Wang D, Wang X ting, Lu Y peng, Zhu L (2018) The roles of hypoxia-
1133 inducible Factor-1 and iron regulatory protein 1 in iron uptake induced by acute
1134 hypoxia. *Biochem Biophys Res Commun* **507**, 128–135.
1135 [68] Popovic Z, Templeton DM (2007) Inhibition of an iron-responsive element/iron
1136 regulatory protein-1 complex by ATP binding and hydrolysis. *FEBS J* **274**,
1137 3108–3119.
1138 [69] Yu Y, Kovacevic Z, Richardson DR (2007) Tuning cell cycle regulation with an
1139 iron key. *Cell Cycle* **6**, 1982–1994.
1140 [70] Fu D, Richardson DR (2007) Iron chelation and regulation of the cell cycle: 2

1141 Mechanisms of posttranscriptional regulation of the universal cyclin-dependent
1142 kinase inhibitor p21CIP1/WAF1 by iron depletion. *Blood* **110**, 752–761.
1143 [71] Zecca L, Youdim MBH, Riederer P, Connor JR, Crichton RR, Stankowski JN,
1144 Dawson VL, Dawson TM, Kawamoto M, Horibe T, Kohno M, Kawakami K,
1145 Duce JA, Tsatsanis A, Cater MA, James SA, Robb E, Wikhe K, Leong SL,
1146 Perez K, Johanssen T, Mark A, Cho H-H, Galatis D, Moir RD, Masters CL,
1147 McLean C, Tanzi RE, Cappai R, Barnham KJ, Ciccotosto D, Rogers JT, Bush
1148 AI, Greenough M a, Cho H-H, Galatis D, Moir RD, Masters CL, McLean C,
1149 Tanzi RE, Cappai R, Barnham KJ, Ciccotosto GD, Rogers JT, Bush AI, De
1150 Domenico I, McVey Ward D, Kaplan J (2011) Iron-export ferroxidase activity of
1151 β-amyloid precursor protein is inhibited by zinc in Alzheimer's disease. *Cell*
1152 **142**, 359.
1153 [72] Jacomin A-C, Geraki K, Brooks J, Tjendana-Tjhin V, Collingwood JF, Nezis IP
1154 (2019) Impact of Autophagy and Aging on Iron Load and Ferritin in Drosophila
1155 Brain. *Front Cell Dev Biol* **7**, 142.
1156 [73] Ayton S, Lei P, Bush AI (2013) Metallostasis in Alzheimer's disease. *Free*
1157 *Radic Biol Med* **62**, 76–89.
1158 [74] Smith MA, Harris PLR, Sayre LM, Perry G (1997) Iron accumulation in
1159 Alzheimer disease is a source of redox-generated free radicals. *Proc Natl*
1160 *Acad Sci U S A* **94**, 9866–9868.
1161 [75] Yambire KF, Rostosky C, Watanabe T, Pacheu-Grau D, Torres-Odio S,
1162 Sanchez-Guerrero A, Senderovich O, Meyron-Holtz EG, Milosevic I, Frahm J,
1163 Phillip West A, Raimundo N (2019) Impaired lysosomal acidification triggers
1164 iron deficiency and inflammation in vivo. *Elife* **8**,.

1165 [76] Jiang Y, Sato Y, Im E, Berg M, Bordi M, Darji S, Kumar A, Mohan PS,
1166 Bandyopadhyay U, Diaz A, Cuervo AM, Nixon RA (2019) Lysosomal
1167 dysfunction in down syndrome is app-dependent and mediated by APP-βCTF
1168 (c99). *J Neurosci* **39**, 5255–5268.
1169 [77] Prasad H, Rao R (2018) Amyloid clearance defect in ApoE4 astrocytes is
1170 reversed by epigenetic correction of endosomal pH. *Proc Natl Acad Sci U S A*
1171 **115**, E6640–E6649.
1172 [78] Tisato V, Zuliani G, Vigliano M, Longo G, Franchini E, Secchiero P, Zauli G,
1173 Paraboschi EM, Singh AV, Serino ML, Ortolani B, Zurlo A, Bosi C, Greco A,
1174 Seripa D, Asselta R, Gemmati D, Vikram Singh A, Serino ML, Ortolani B, Zurlo
1175 A, Bosi C, Greco A, Seripa D, Asselta R, Gemmati D (2018) Gene-gene
1176 interactions among coding genes of iron-homeostasis proteins and APOE-
1177 alleles in cognitive impairment diseases. *PLoS One* **13**, e0193867.
1178 [79] Van Acker ZP, Bretou M, Annaert W (2019) Endo-lysosomal dysregulations
1179 and late-onset Alzheimer's disease: Impact of genetic risk factors. *Mol*
1180 *Neurodegener* **14**, 1–20.
1181 [80] Tsatsanis A, Wong BX, Gunn AP, Ayton S, Bush AI, Devos D, Duce JA (2020)
1182 Amyloidogenic processing of Alzheimer's disease β-amyloid precursor protein
1183 induces cellular iron retention. *Mol Psychiatry* 1–9.
1184 [81] Dlouhy AC, Bailey DK, Steimle BL, Parker H V., Kosman DJ (2019)
1185 Fluorescence resonance energy transfer links membrane ferroportin,
1186 hephaestin but not ferroportin, amyloid precursor protein complex with iron
1187 efflux. *J Biol Chem* **294**, 4202–4214.
1188 [82] Lahiri DK, Maloney B, Wang R (2019) APPealing for a role in cellular iron

1189 efflux. *J Biol Chem* **294**, 9365.

1190 [83] Yeh K, Yeh M, Glass J (2011) Interactions between ferroportin and hephaestin
1191 in rat enterocytes are reduced after iron ingestion. *Gastroenterology* **141**, 292-
1192 299.e1.

1193 [84] Yeh K, Yeh M, Mims L, Glass J (2009) Iron feeding induces ferroportin 1 and
1194 hephaestin migration and interaction in rat duodenal epithelium. *Am J Physiol*
1195 *Liver Physiol* **296**, G55–G65.

1196 [85] Han O, Kim E-Y (2007) Colocalization of ferroportin-1 with hephaestin on the
1197 basolateral membrane of human intestinal absorptive cells. *J Cell Biochem*
1198 **101**, 1000–1010.

1199 [86] Pigoni M, Wanngren J, Kuhn PH, Munro KM, Gunnersen JM, Takeshima H,
1200 Feederle R, Voytyuk I, De Strooper B, Levasseur MD, Hrupka BJ, Müller SA,
1201 Lichtenthaler SF (2016) Seizure protein 6 and its homolog seizure 6-like
1202 protein are physiological substrates of BACE1 in neurons. *Mol Neurodegener*
1203 **11**, 1–18.

1204 [87] Rogers JT, Randall JD, Cahill CM, Eder PS, Huang X, Gunshin H, Leiter L,
1205 McPhee J, Sarang SS, Utsuki T, Greig NH, Lahiri DK, Tanzi RE, Bush AI,
1206 Giordano T, Gullans SR (2002) An iron-responsive element type II in the 5′-
1207 untranslated region of the Alzheimer's amyloid precursor protein transcript. *J*
1208 *Biol Chem* **277**, 45518–45528.

1209 [88] Conway JR, Lex A, Gehlenborg N UpSetR: an R package for the visualization
1210 of intersecting sets and their properties. *Bioinformatics*.

1211 [89] Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G,
1212 Kasprzyk A (2009) BioMart - Biological queries made easy. *BMC Genomics*

1213 **10**, 22.

1214 [90] Bastian M, Heymann S, Jacomy M (2009) Gephi: An Open Source Software
1215 for Exploring and Manipulating Networks Visualization and Exploration of
1216 Large Graphs. *Third Int AAAI Conf Weblogs Soc Media*.

1217 [91] Hu Y (2011) *Algorithms for Visualizing Large Networks*.

1218 [92] Benner C HOMER (Hypergeometric Optimization of Motif EnRichment).

1219 [93] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre
1220 C, Singh H, Glass CK (2010) Simple Combinations of Lineage-Determining
1221 Transcription Factors Prime cis-Regulatory Elements Required for
1222 Macrophage and B Cell Identities. *Mol Cell* **38**, 576–589.

1223 [94] Team RC (2018) R: A language and environment for statistical computing.

1224 [95] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) Limma
1225 powers differential expression analyses for RNA-sequencing and microarray
1226 studies. *Nucleic Acids Res* **43**, e47.

1227 [96] Law CW, Chen Y, Shi W, Smyth GK (2014) Voom: Precision weights unlock
1228 linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**,.

1229 [97] Wilkinson B (1951) A statistical consideration in psychological research.
1230 *Psychol Bull* **48**, 156–158.

1231 [98] Chicault C, Toutain B, Monnier A, Aubry M, Fergelot P, Le Treut A, Galibert
1232 MD, Mosser J (2006) Iron-related transcriptomic variations in CaCo-2 cells, an
1233 in vitro model of intestinal absorptive cells. *Physiol Genomics* **26**, 55–67.

1234 [99] Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic
1235 RNA-seq quantification. *Nat Biotechnol* **34**, 525–527.

1236 [100] Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: A Bioconductor

1237 package for differential expression analysis of digital gene expression data.
1238 *Bioinformatics* **26**, 139–140.
1239 [101] Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF,
1240 Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC,
1241 Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL,
1242 Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong HW, Dougherty JG, Duncan
1243 BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR,
1244 Fliss TP, Frensley C, Gates SN, Glattfelder KJ, Halverson KR, Hart MR,
1245 Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawal R, Kidney
1246 JM, Knapik RH, Kuan CL, Lake JH, Laramee AR, Larsen KD, Lau C, Lemon
1247 TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud
1248 MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE,
1249 Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA,
1250 Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova N V., Sivisay T,
1251 Slaughterbeck CR, Smith SC, Smith KA, Smith BI, Sodt AJ, Stewart NN,
1252 Stumpf KR, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson
1253 CL, Varnam LR, Visel A, Whitlock RM, Wohnoutka PE, Wolkey CK, Wong VY,
1254 Wood M, Yaylaoglu MB, Young RC, Youngstrom BL, Yuan XF, Zhang B,
1255 Zwingman TA, Jones AR (2007) Genome-wide atlas of gene expression in the
1256 adult mouse brain. *Nature* **445**, 168–176.
1257 [102] Bonham LW, Sirkis DW, Yokoyama JS (2019) The transcriptional landscape of
1258 microglial genes in aging and neurodegenerative disease. *Front Immunol* **10**,
1259 1170.
1260 [103] 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.

[104] Giner, Smyth G, K G (2016) FRY: a fast approximation to ROAST gene set test with mean aggregated set statistics. *F1000Research*.

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

[106] 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. *Proc Natl Acad Sci U S A* **102**, 15545–15550.

Tables

Table 1. Summary of gene set testing approaches used in our analyses.

Method	Hypothesis	Focused / battery approach	Accounts for inter-gene correlation	R implementation
Fry / Roast	Self-contained null hypothesis (genes within a set have no association with experimental condition)	Focused (each gene set is tested on its own terms, and multiple testing adjustment is applied afterwards)	Yes	fry() and mroast() functions in the *limma* Bioconductor package [103,104]
Camera	Competitive null hypothesis (genes within a set do not have a stronger association with experimental condition compared to a random gene set)	Focused (each gene set is tested on its own terms, and multiple testing adjustment is applied afterwards)	Yes	camera() function in the *limma* Bioconductor package [105]
fgsea (fast implementation of GSEA in R)	Self-contained null hypothesis (genes within a set have no association with experimental condition)	Battery (gene sets are pitted against each other to determine which ones are more significantly enriched)	No	fgsea() function in the *fgsea* Bioconductor package [40,106]

1280 **Table 2. Samples analyzed from Mayo Clinic RNA-seq study.**

1281

Group	Total number	% Female	Mean age ± s.d (years)	Diagnosis
Alzheimer's disease (AD)	134	56.7	82.6 ± 7.3	Braak Stage ≥ 4, diagnosis according to NINCDS-ADRDA criteria
Control	130	46.2	82.9 ± 8.3	Braak Stage ≤ 3, No or sparse CERAD neuritic and cortical plaque density, no diagnoses for any neurodegenerative disease
Pathological Aging (PA)	44	54.5	84.5 ± 4.3	Braak Stage ≤ 3, CERAD neuritic and cortical plaque density of 2 or more, no diagnoses for any neurodegenerative disease or mild cognitive impairment
Progressive supranuclear palsy (PSP)	164	40.2	73.8 ± 6.5	Braak Stage ≤ 3, no or sparse CERAD neuritic and cortical plaque density

1282

1283

1284

1285 **Table 3. Differential gene expression and IRE gene set enrichment results from Mayo**

1286 **Clinic RNA-seq study.** Asterisks (*) indicate significant enrichment (Bonferroni-adjusted

1287 Wilkinson's *p*-value from *fry*, *camera*, and *fgsea* < 0.05).

	Comparison	**No. of DE genes (FDR p-value < 0.05 and abs(logFC) > 0.5)**	***All 3' IREs*** enrichment *p*-value	***All 5' IREs*** enrichment *p*-value	***HQ 3' IREs*** enrichment *p*-value	***HQ 5' IREs*** enrichment *p*-value
Effect of Alzheimer's disease	AD vs. control (cerebellum)	Down: 1,422 Up: 1,308	0.0109*	0.128	0.770	0.770
	AD vs. control (temporal cortex)	Down: 1,650, Up: 1,929	9.03E-15*	2.44E-14*	4.50E-15*	8.95E-08*
Effect of pathological aging (amyloid pathology, no dementia symptoms)	PA vs. control (cerebellum)	Down: 254, Up: 463	0.000116*	0.00139*	0.0201*	0.770
	PA vs. control (temporal cortex)	Down: 466, Up: 512	0.000521*	0.0461*	0.000065*	0.0000682*
Effect of progressive supranuclear palsy (PSP)	PSP vs. control (cerebellum)	Down: 2,550, Up: 1,080	1.81E-22*	3.45E-22*	1.43E-12*	0.0399*
	PSP vs. control (temporal cortex)	Down: 1,271, Up: 622	3.58E-11*	5.17E-13*	1.39E-06*	0.0399*
Differences between Alzheimer's disease and pathological aging	AD vs. PA (cerebellum)	Down: 1,728, Up: 1,006	0.000116*	0.00139*	0.0201*	0.770
	AD vs. PA (temporal cortex)	Down: 2,633, Up: 2,174	0.000521*	0.0461*	0.000065*	0.0000682*
Differences between Alzheimer's disease and progressive supranuclear palsy	AD vs. PSP (cerebellum)	Down: 108, Up: 659	0.0019*	0.00184*	0.158	0.158
	AD vs. PSP (temporal cortex)	Down: 2,052, Up: 2,354	9.12E-22*	2.10E-22*	5.35E-16*	0.040*
Differences between cerebellum and temporal cortex tissue	Cerebellum vs. temporal cortex in controls	Down: 10,925, Up: 9,302	0.00E+00*	0.00E+00*	2.82E-269*	1.60E-81*
	Cerebellum vs. temporal cortex in AD	Down: 11,966, Up: 9,950	0.00E+00*	0.00E+00*	7.71E-302*	9.35E-92*
	Cerebellum vs. temporal cortex in PA	Down: 10,200, Up: 8,737	1.87E-153*	5.88E-160*	1.22E-89*	1.31E-88*
	Cerebellum vs. temporal cortex in PSP	Down: 11,552, Up: 9,576	0.00E+00*	0.00E+00*	0.00E+00*	4.74E-104*

1288 **Table 4. Enrichment of Iron Responsive Element (IRE) gene sets in fAD-like zebrafish**
1289 **dataset.** Raw *p*-values from *fry*, *camera* and *fgsea* were combined with Wilkinson's method,
1290 with combined *p*-values then Bonferroni-adjusted for multiple testing. The same process was
1291 repeated for the K97fs/+ dataset, which involves an independent family of fish (†).

	Comparison	***All 3' IREs*** enrichment *p*-value	***All 5' IREs*** enrichment *p*-value	***HQ 3' IREs*** enrichment *p*-value	***HQ 5' IREs*** enrichment *p*-value
Effect of *psen1* mutation	6-month-old Q96_K97del/+ vs. 6-month-old +/+ all under normoxia	0.0109	0.128	0.770	0.770
	6-month-old Q96_K97del/+ vs. 6-month-old +/+ all under hypoxia	9.03E-15	2.44E-14	4.50E-15	8.95E-08
	24-month-old Q96_K97del/+ vs. 24-month-old +/+ all under normoxia	0.000116	0.00139	0.0201	0.770
	24-month-old Q96_K97del/+ vs. 24-month-old +/+ all under hypoxia	0.000521	0.0461	0.000065	0.0000682
	6-month-old K97fs/+ vs. 6-month-old +/+ all under normoxia †	0.199	0.289	0.0197	0.289
	24-month-old K97fs/+ vs. 24-month-old +/+ all under normoxia †	0.116	0.115	0.00000118	0.979
Effect of hypoxia	6-month-old +/+ under hypoxia vs. 6-month-old +/+ under normoxia	1.10E-18	1.10E-18	8.19E-13	1.27E-08
	24-month-old +/+ under hypoxia vs. 24-month-old +/+ under normoxia	3.67E-15	7.98E-11	2.94E-07	3.31E-05
	6-month-old Q96_K97del/+ under hypoxia vs. 6-month-old Q96_K97del/+ under normoxia	1.13E-24	9.88E-22	2.28E-13	2.86e- 3
	24-month-old Q96_K97del/+ under hypoxia vs. 24-month-old Q96_K97del/+ under normoxia	0.000521	0.015	0.128	0.770
Effect of aging	24-month-old +/+ vs. 6-month-old +/+ all under normoxia	6.80E-35	4.82E-27	3.41E-18	1.26E-21
	24-month-old +/+ vs. 6-month-old +/+ all under hypoxia	1.55E-35	6.84E-31	2.36E-20	7.40E-23
	24-month-old Q96_K97del/+ vs. 6-month-old Q96_K97del/+ all under normoxia	1.85E-27	1.03E-21	8.09E-17	1.07E-17
	24-month-old Q96_K97del/+ vs. 6-month-old Q96_K97del/+ all under hypoxia	3.77E-30	1.59E-23	2.11E-19	2.46E-16
	24-month-old +/+ vs. 6-month-old +/+ all under normoxia (K97fs family) †	1.69E-10	3.96E-08	5.95E-10	4.59E-06
	24-month-old K97fs/+ vs. 6-month-old K97fs/+ all under normoxia †	0.199	0.289	0.0197	0.289

1292
1293

1294 # Figure Legends

1295 **Figure 1. Iron Responsive Element (IRE) background. A. Model of altered transcript**
1296 **stability due to the IRP/IRE system under iron dyshomeostasis.** When cellular iron
1297 levels (through FeS) are low, Iron Responsive Proteins (IRP1 or IRP2) bind IREs in the 5' or
1298 3' untranslated regions (UTRs) of genes involved in iron homeostasis. In general, genes with
1299 IREs in their 3' UTRs will be stabilised and hence increased in expression while genes with
1300 IREs in their 5' UTR will have repressed translation from IRP binding. However, when
1301 cellular iron levels are high, IRP1 undergoes a conformation change and is now unable to
1302 bind IREs. IRP1 now acts as a cytosolic aconitase (ACO1). IRP2 is degraded in high iron
1303 conditions and is also unable to bind IREs. Without IRPs binding to IREs, genes with 5' IREs
1304 will have increased translation, while those with 3' IREs will have decreased mRNA stability,
1305 causing their expression to be decreased. **B. Consensus IRE secondary structure and**
1306 **examples of high-quality and IRE-like motifs predicted by SIREs.** IRE-like motifs with
1307 non-canonical structure are able to be detected by SIREs if they have up to one mismatch
1308 pair in the upper stem (e.g. *Hao1*) or 1 unpaired bulge nucleotide on the 3' strand of the
1309 upper stem (e.g. *EPAS1*). For more details on the prediction of non-canonical IRE motifs,
1310 please refer to Figure 1 of Campillos et al. [24].

1311

1312 **Figure 2. Overlap between predicted IRE gene sets for human, mouse, and zebrafish**.
1313 The number of genes in the gene set for each species is shown at the bottom-left bars of
1314 each UpSet plot, while genes with shared homologs across species are indicated in the
1315 main plot region. IRE genes in mouse and zebrafish gene sets were excluded from this plot
1316 if they did not have a human homolog.

1317

1318 **Figure 3. MSigDB gene sets showing over-representation of predicted 3' and 5' IRE**
1319 **genes in human, mouse, and zebrafish**. The top 15 MSigDB gene sets ranked by Fisher's

1320 exact test *p*-value (testing for over-representation of the ***all 3' IREs*** and/or ***all 5' IREs*** sets)
1321 are shown for each species. In the network plots, the top 15 MSigDB gene sets are shown
1322 as large nodes, with genes represented as small nodes. Edges connecting genes to gene
1323 sets indicate the gene set(s) that a gene belongs to. Overall, the ***all 3' IREs*** and ***all 5' IREs***
1324 gene sets have a large proportion of genes which are not included in any of the top ranked
1325 MSigDB gene sets for each species.
1326
1327 **Figure 4. IRE-containing gene expression analysis workflow.** The section including
1328 identification of candidate 3' IRE genes stabilized under iron deficiency was only applied to
1329 the fAD-like zebrafish dataset due to unavailable raw RNA-seq reads for the other datasets
1330 needed to identify expression of unspliced genes.
1331
1332 **Figure 5. Analysis of Caco-2 cultured cell line microarray dataset. A. Gene set**
1333 **enrichment testing results for human predicted IRE gene sets for the iron overload**
1334 **and iron deficiency treatments.** Dots indicate if a gene set was considered significantly
1335 enriched (Bonferroni-adjusted *p*-value < 0.05) in the iron overload (on left) or iron deficiency
1336 (on right) treatment. **B. Gene set enrichment testing results for MSigDB Hallmark gene**
1337 **sets in the iron overload and iron deficiency treatments.** Dots indicate if a gene set was
1338 considered significantly enriched (FDR-adjusted *p*-value < 0.05) in the iron overload (on left)
1339 or iron deficiency (on right) treatments. **C. UpSet plots showing overlap between iron**
1340 **overload and iron deficiency treatments in GSEA leading-edge genes for the *All 3'***
1341 ***IREs* and *All 5' IREs* gene sets**. The bars to the lower-left indicate the number of the
1342 leading-edge genes for iron overload and iron deficiency treatments, while the bars in the
1343 main plot region indicate the number of leading-edge genes which are unique or shared
1344 between the treatments.
1345

1346 **Figure 6. Analysis of Mayo Clinic RNA-seq dataset with human IRE gene sets. A.**
1347 **Enrichment analysis results of IRE gene sets in comparisons of AD (Alzheimer's**
1348 **Disease), PA (Pathological Ageing) and PSP (Progressive Supranuclear Palsy) vs.**
1349 **control.** The proportions of IRE genes within each IRE gene set with increased, decreased,
1350 or unchanged expression are indicated as coloured bars. The dots indicate that the IRE
1351 gene set was significantly enriched (Bonferroni-adjusted enrichment *p*-value < 0.05). **B.**
1352 **UpSet plots showing overlap between leading-edge IRE genes in comparisons of AD**
1353 **(Alzheimer's Disease), PA (Pathological Aging), and PSP (Progressive Supranuclear**
1354 **Palsy) vs. control.** The bars to the lower-left region of each UpSet plot indicate the number
1355 of leading-edge IRE genes for each comparison, while bars in the main plot region indicate
1356 the number of leading-edge IRE genes which are unique or shared between comparisons.
1357
1358 **Figure 7. Analysis of 5XFAD mouse dataset. A. Experimental design and results of**
1359 **IRE gene set enrichment analysis.** The gene sets ***all 3' IREs*** and ***all 5' IREs*** derived from
1360 searching for IRE sequences in the UTRs of genes in the reference mouse genome mm10
1361 are represented here as "3' IRE" and "5' IRE" respectively. Asterisks (*) indicate that the
1362 gene set was significantly enriched in a particular comparison (Bonferroni-adjusted *p*-value <
1363 0.05). **B. Proportions of genes in IRE and MSigDB Hallmark gene sets which are**
1364 **increased (t > 2) or decreased (t < -2) in expression in all "5XFAD vs. wild type"**
1365 **comparisons**. A dot next to a bar indicates that the gene set was significantly enriched
1366 (FDR-adjusted *p*-value < 0.05).
1367
1368 **Figure 8. UpSet plots showing the overlap in GSEA leading-edge genes between all**
1369 **comparisons in the 5XFAD mouse datasets for the gene sets all 3' IREs and all 5'**
1370 **IREs**. Numbers of genes for each intersection are shown above intersection bars.
1371

1372 **Figure 9. Differential gene expression and gene set enrichment analysis in the fAD-**
1373 **like zebrafish dataset. A. Results of differential gene expression analysis. Genes**
1374 **which were significantly increased or decreased in expression are indicated in boxes**.
1375 These differentially expressed genes have FDR-adjusted *p*-value < 0.05. **B. Gene set**
1376 **enrichment with MSigDB Hallmark gene sets.** The comparisons between Q96_ K97del/+
1377 fAD-like mutants and their wild-type siblings are shown for the 6-month-old (young adult)
1378 and 24-month-old (infertile adult) age groups. Dots indicate gene sets which are significantly
1379 enriched (FDR-adjusted *p*-value < 0.05).
1380
1381 **Figure 10. Iron Responsive Element (IRE)-containing gene expression in the fAD-like**
1382 **zebrafish dataset. A. Results of gene set enrichment testing using predicted IRE gene**
1383 **sets.** We represent the gene sets all 3' IREs and all 5' IREs derived from searching for IRE
1384 and IRE-like sequences from z11 reference zebrafish gene UTRs as "3' IRE" and "5' IRE" in
1385 the panel. **B. Proportions of predicted IRE genes which are increased (t > 2) or**
1386 **decreased (t < -2) in expression for each pairwise comparison in the dataset. C.**
1387 **Principal component analysis of all genes in the sets all 3' IREs and all 5' IREs for all**
1388 **samples.** Circles on the 3' IRE plot show that different conditions generally have distinct
1389 expression of genes in the ***all 3' IREs set*** but not in the ***all 5' IREs*** set. **D. UpSet plot**
1390 **showing overlap in leading-edge genes for the *all 3' IREs* gene set for select**
1391 **comparisons.**

1392

1393

Supplementary Information

Supplementary Text

- **Supplementary Text 1.** Expression of previously characterized 3' and 5' IRE genes in the gene expression datasets analyzed in this paper
- **Supplementary Text 2.** Identifying predicted 3' IRE genes with altered stability.
- **Supplementary Text 3.** Effects of biological sex on gene expression in the fAD-mutation-like zebrafish dataset.
- **Supplementary Text 4.** Neural cell type proportions in the gene expression datasets analyzed in this paper.

Supplementary Tables

- **Supplementary Table 1.** Iron Responsive Element (IRE) gene sets defined from human, mouse, and zebrafish transcriptomes. The untranslated regions (UTR) of all known genes were searched for IRE and IRE-like motifs. Four gene sets are included for each species: ***all 3' IREs*** (all genes with predicted IRE in their 3' UTR), ***all 5' IREs*** (all genes with predicted IRE in their 5' UTR), ***HQ 3' IREs*** (genes with high-quality predicted IRE in 3' UTR), and ***HQ 5' IREs*** (genes with high-quality predicted IRE in 5' UTR). Genes are provided as ENSEMBL gene identifiers.
- **Supplementary Table 2.** Molecular Signature Database (MSigDB) gene sets with significant over-representation of IRE gene sets in human, mouse, and zebrafish. Gene sets were defined to have significant over-representation of IRE gene sets if the FDR-adjusted Fisher's exact test *p*-value was < 0.05.
- **Supplementary Table 3**. Promoter motif over-representation analysis results for all IRE gene sets in human, mouse, and zebrafish. Promoter regions were defined as being 1500 bp upstream and 500 bp downstream of genes. See Methods for details. We define significant over-representation as FDR-adjusted enrichment *p*-value < 0.05.
- **Supplementary Table 4.** Gene ontology (GO) analysis results for leading edge genes shared between the "5XFAD vs. WT (3-months-old)" and "5XFAD vs. WT (6-months-old)" comparisons. Significant over-representation of GO terms was determined as GO terms having an FDR-adjusted *p*-value < 0.05.

1427 • **Supplementary Table 5.** Altered stability of transcripts in the fAD-mutation-like
1428 zebrafish dataset. See Methods and Supplementary Text 2 for details on how
1429 stability was defined and calculated. Genes which contain a 3' IRE (belonging to one
1430 of the 3' IRE gene sets) or 5' IRE (belonging to one of the 5' IRE gene sets) are
1431 indicated with the "has3ire" or "has5ire" columns.
1432

1433 ## Supplementary Figures

1434 • **Supplementary Figure 1.** Expression of previously characterized IRE genes in
1435 datasets analyzed. **A.** Caco-2 cell line dataset. **B.** Mayo Clinic RNA-seq (Human)
1436 dataset – Temporal Cortex tissue. **C.** Mayo Clinic RNA-seq (Human) dataset –
1437 Cerebellum tissue. **D.** 5XFAD mouse dataset. **E.** fAD-mutation-like zebrafish dataset.
1438 • **Supplementary Figure 2.** Principal Component Analysis of results from different
1439 gene set testing methods.
1440 • **Supplementary Figure 3.** Analysis of cultured Caco-2 cell line dataset. **A.** Principal
1441 Component Analysis plot of gene expression in the cultured Caco-2 cell line dataset.
1442 **B.** Volcano plots indicating differential gene expression due to iron overload (DMEM-
1443 FBS Medium + Hemin vs. DMEM-FBS Medium) and iron deficiency (Iron-free
1444 medium vs. Iron-free medium + FAC) treatments in the Caco-2 cell line dataset.
1445 • **Supplementary Figure 4.** Gene set enrichment testing results for the AD vs. control
1446 comparison in cerebellum and temporal cortex.
1447 • **Supplementary Figure 5.** STRINGR protein-protein interaction network plot
1448 between the 19 shared leading-edge genes between the *psen1*$^{Q96_K97del/+}$ and
1449 *psen1*$^{+/+}$ comparisons at 6 and 24-months-old.
1450 • **Supplementary Figure 6.** Principal Component Analysis plots showing lack of clear
1451 association between IRE gene expression and the *psen1*$^{K97fs/+}$ mutant genotype.
1452 • **Supplementary Figure 7.** Predicted 3' IRE containing transcripts with significant
1453 differences in stability between conditions in the fAD-like zebrafish dataset.
1454 • **Supplementary Figure 8.** Age-dependent expression of neural marker genes
1455 (microglia, astrocyte, neuron, oligodendrocyte) in the Mayo Clinic RNA-seq dataset.
1456 • **Supplementary Figure 9.** Age-dependent expression of neural marker genes
1457 (microglia, astrocyte, neuron, oligodendrocyte) in 5XFAD mouse dataset.
1458 • **Supplementary Figure 10.** Age- and hypoxia- dependent expression of neural
1459 marker genes (microglia, astrocyte, neuron, oligodendrocyte) in fAD-like zebrafish
1460 dataset.

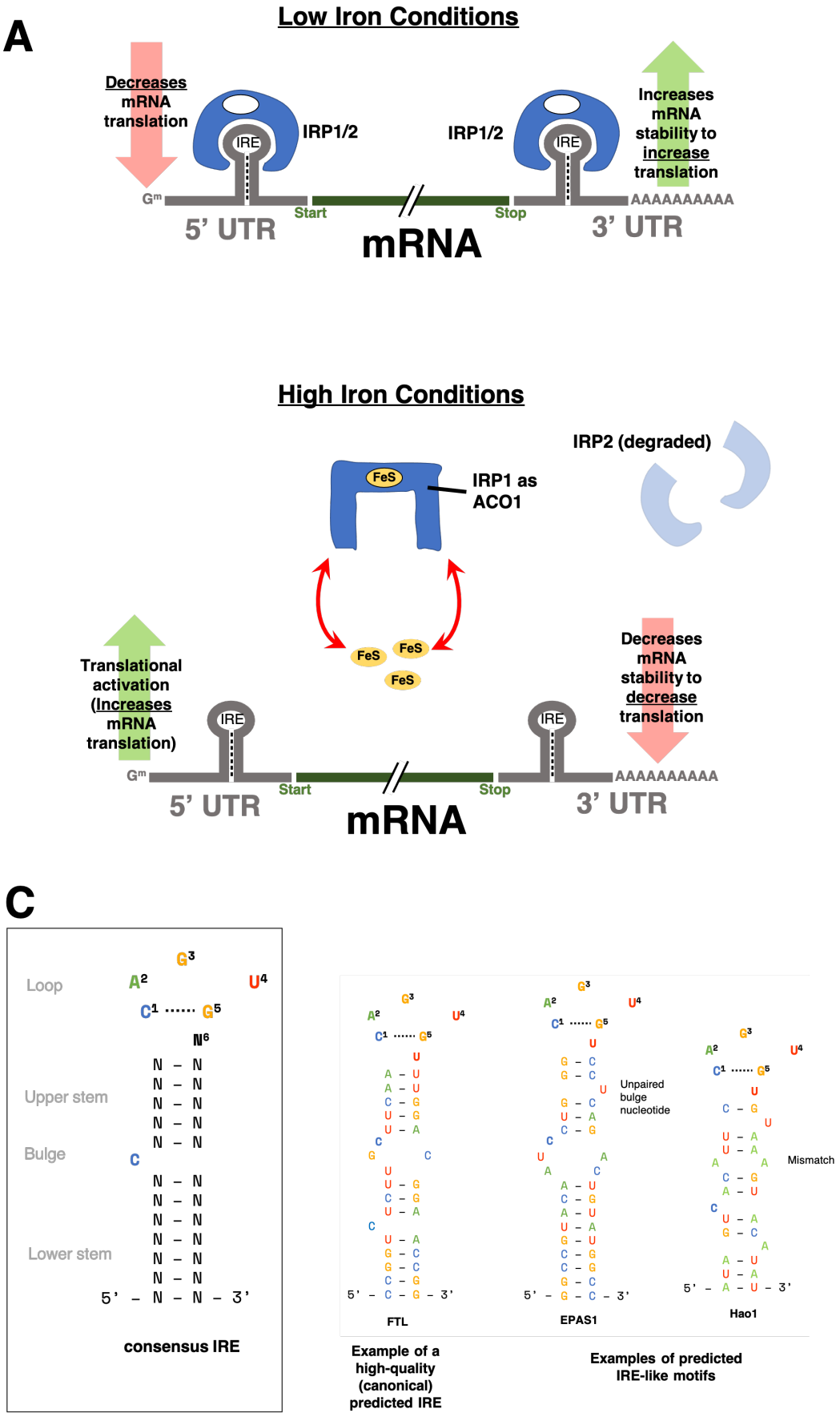

Figure 1. Iron Responsive Element (IRE) background. A. Model of altered transcript stability due to the IRP/IRE system under iron dyshomeostasis. When cellular iron levels (through FeS) are low, Iron Responsive Proteins (IRP1 or IRP2) bind IREs in the 5' or 3' untranslated regions (UTRs) of genes involved in iron homeostasis. In general, genes with IREs in their 3' UTRs will be stabilised and hence increased in expression while genes with IREs in their 5' UTR will have repressed translation from IRP binding. However, when cellular iron levels are high, IRP1 undergoes a conformation change and is now unable to bind IREs. IRP1 now acts as a cytosolic aconitase (ACO1). IRP2 is degraded in high iron conditions and is also unable to bind IREs. Without IRPs binding to IREs, genes with 5' IREs will have increased translation, while those with 3' IREs will have decreased mRNA stability, causing their expression to be decreased. **B. Consensus IRE secondary structure and examples of high-quality and IRE-like motifs predicted by SIREs.** IRE-like motifs with non-canonical structure are able to be detected by SIREs if they have up to one mismatch pair in the upper stem (e.g. Hao1) or 1 unpaired bulge nucleotide on the 3' strand of the upper stem (e.g. EPAS1). For more details on the prediction of non-canonical IRE motifs, please refer to Figure 1 of Campillos et al. [24].

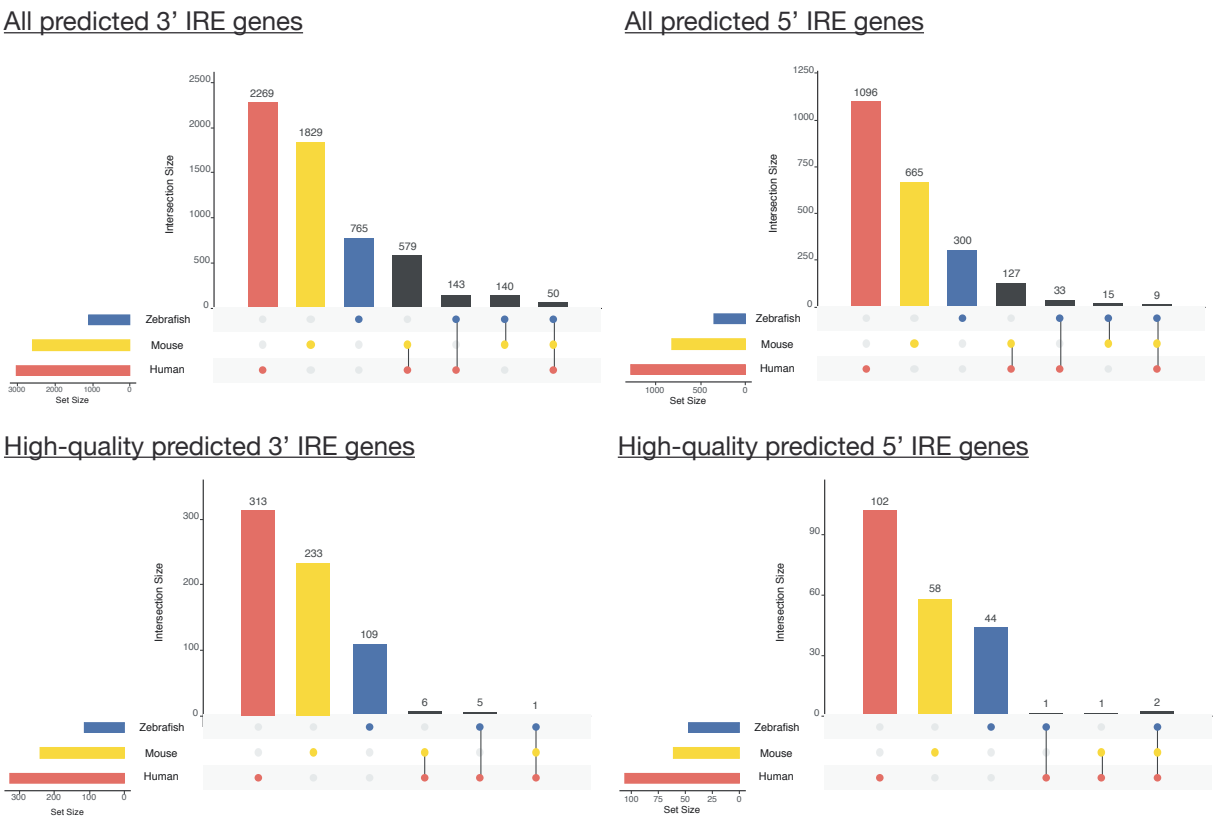

Figure 2. Overlap between predicted IRE gene sets for human, mouse, and zebrafish. The number of genes in the gene set for each species is shown at the bottom-left bars of each UpSet plot, while genes with shared homologs across species are indicated in the main plot region. IRE genes in mouse and zebrafish gene sets were excluded from this plot if they did not have a human homolog.

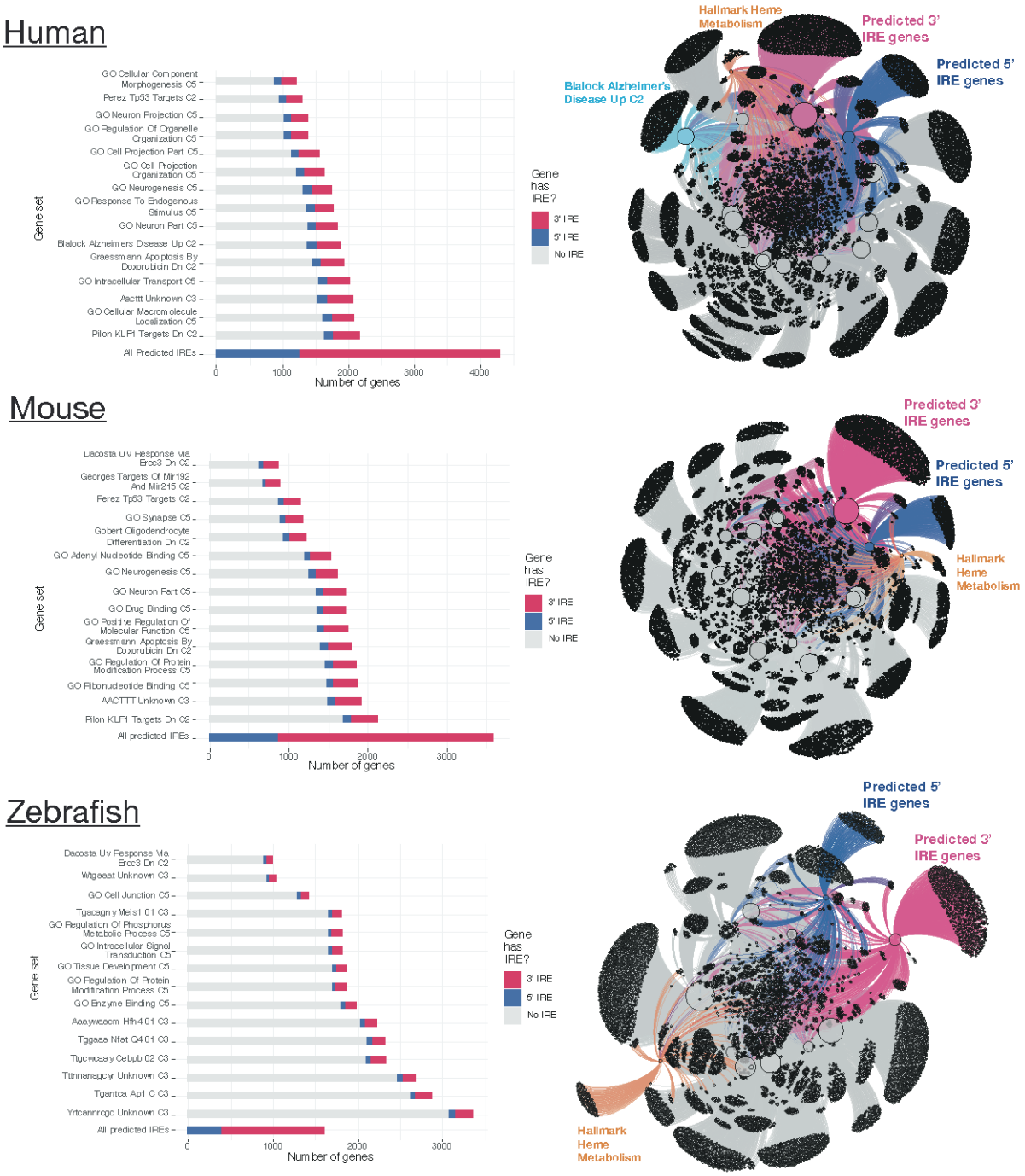

Figure 3. MSigDB gene sets showing over-representation of predicted 3' and 5' IRE genes in human, mouse, and zebrafish. The top 15 MSigDB gene sets ranked by Fisher's exact test *p*-value (testing for over-representation of the ***all 3' IREs*** and/or ***all 5' IREs*** sets) are shown for each species. In the network plots, the top 15 MSigDB gene sets are shown as large nodes, with genes represented as small nodes. Edges connecting genes to gene sets in-dicate the gene set(s) that a gene belongs to. Overall, the ***all 3' IREs*** and ***all 5' IREs*** gene sets have a large proportion of genes which are not included in any of the top ranked MSigDB gene sets for each species.

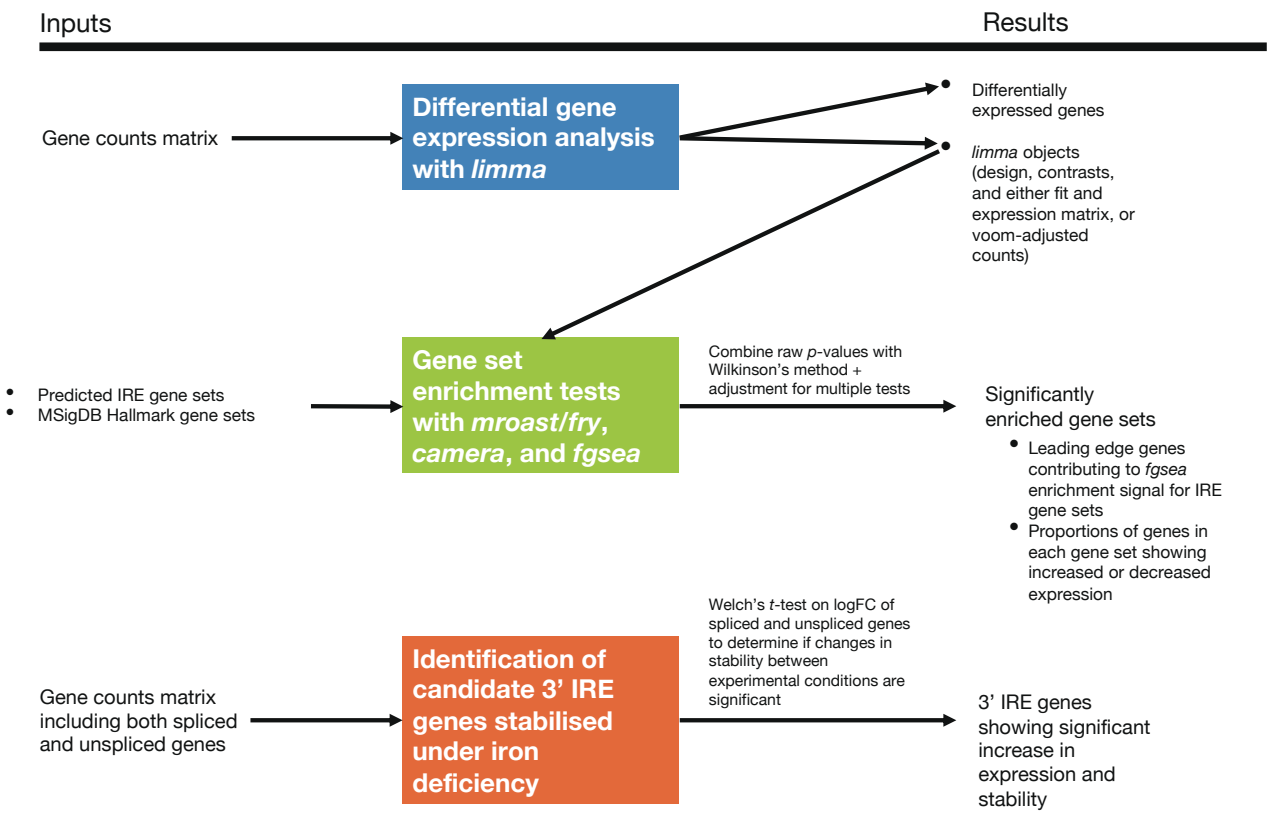

Figure 4. IRE-containing gene expression analysis workflow. The section including identification of can-didate 3' IRE genes stabilised under iron deficiency was only applied to the fAD-like zebrafish dataset due to unavailable raw RNA-seq reads for the other datasets needed to identify expression of unspliced genes.

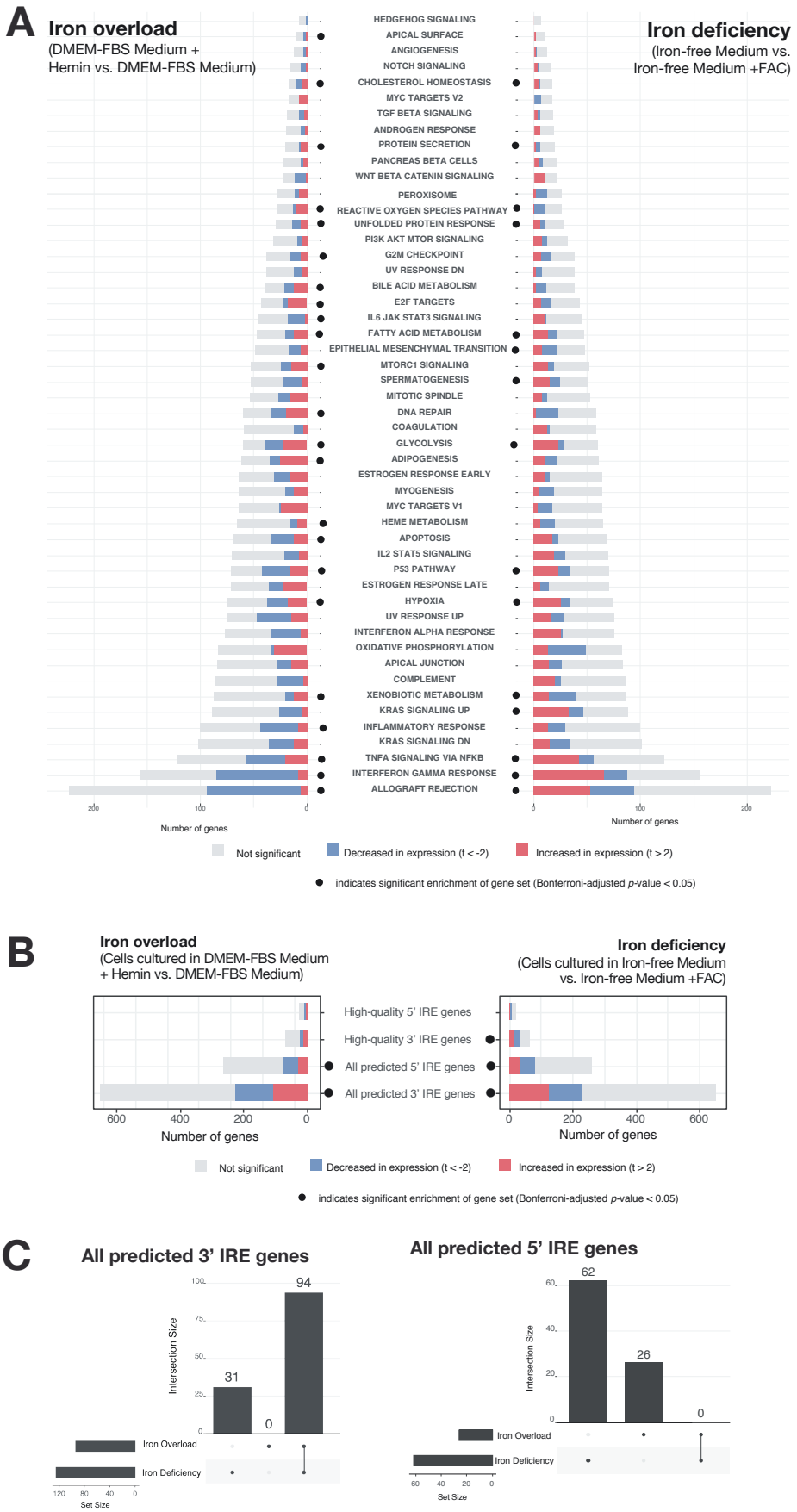

Figure 5. Analysis of Caco-2 cultured cell line microarray dataset. A. Gene set enrichment testing results for human predicted IRE gene sets for the iron overload and iron deficiency treatments. Dots indicate if a gene set was considered significantly enriched (Bonferroni-adjusted *p*-value < 0.05) in the iron overload (on left) or iron deficiency (on right) treatment. **B. Gene set enrichment testing results for MSigDB Hallmark gene sets in the iron overload and iron deficiency treatments.** Dots indicate if a gene set was considered significantly enriched (FDR-adjusted p-value < 0.05) in the iron overload (on left) or iron deficiency (on right) treatments. C. **UpSet plots showing overlap between iron overload and iron deficiency treatments in GSEA leading-edge genes for the "All predicted 3' IRE genes" and "All predicted 5' IRE genes" gene sets.** The bars to the lower-left indicate the number of the leading-edge genes for iron overload and iron deficiency treatments, while the bars in the main plot region indicate the number of leading-edge genes which are unique or shared between the treatments.

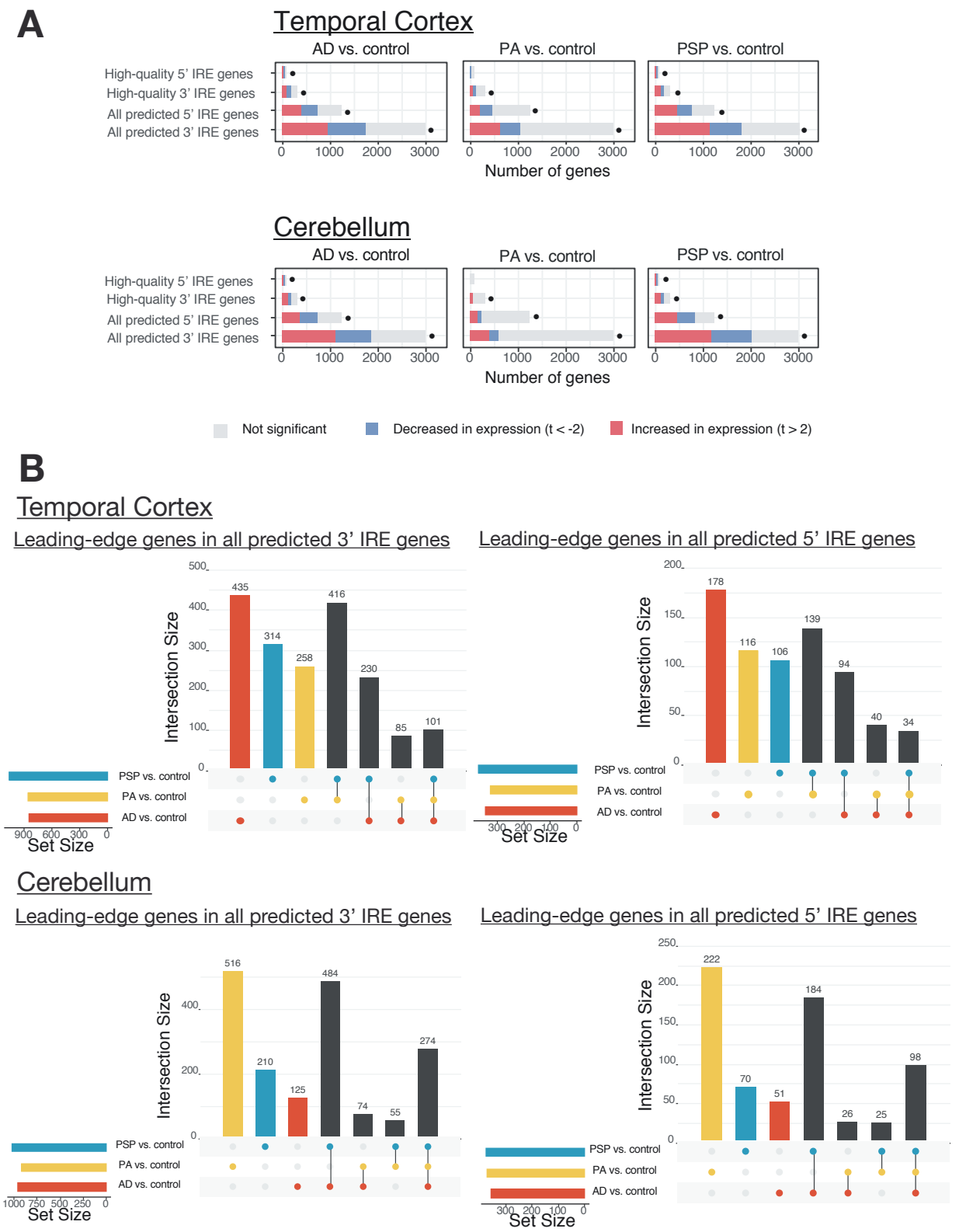

Figure 6. Analysis of Mayo Clinic RNA-seq dataset with human IRE gene sets. A. Enrichment analysis results of IRE gene sets in comparisons of AD (Alzheimer's Disease), PA (Pathological Aging), and PSP (Progressive Supranuclear Palsy) vs. control. The proportions of IRE genes within each IRE gene set with increased, decreased, or unchanged expression are indicated as coloured bars. The dots indicate that the IRE gene set was significantly enriched (Bonferroni-adjusted enrichment p-value < 0.05). **B. UpSet plots showing overlap between leading-edge IRE genes in comparisons of AD (Alzheimer's Disease), PA (Pathological Aging), and PSP (Progressive Supranuclear Palsy) vs. control.** The bars to the lower-left region of each UpSet plot indicate the number of leading-edge IRE genes for each comparison, while bars in the main plot region indicate the number of leading-edge IRE genes which are unique or shared between comparisons.

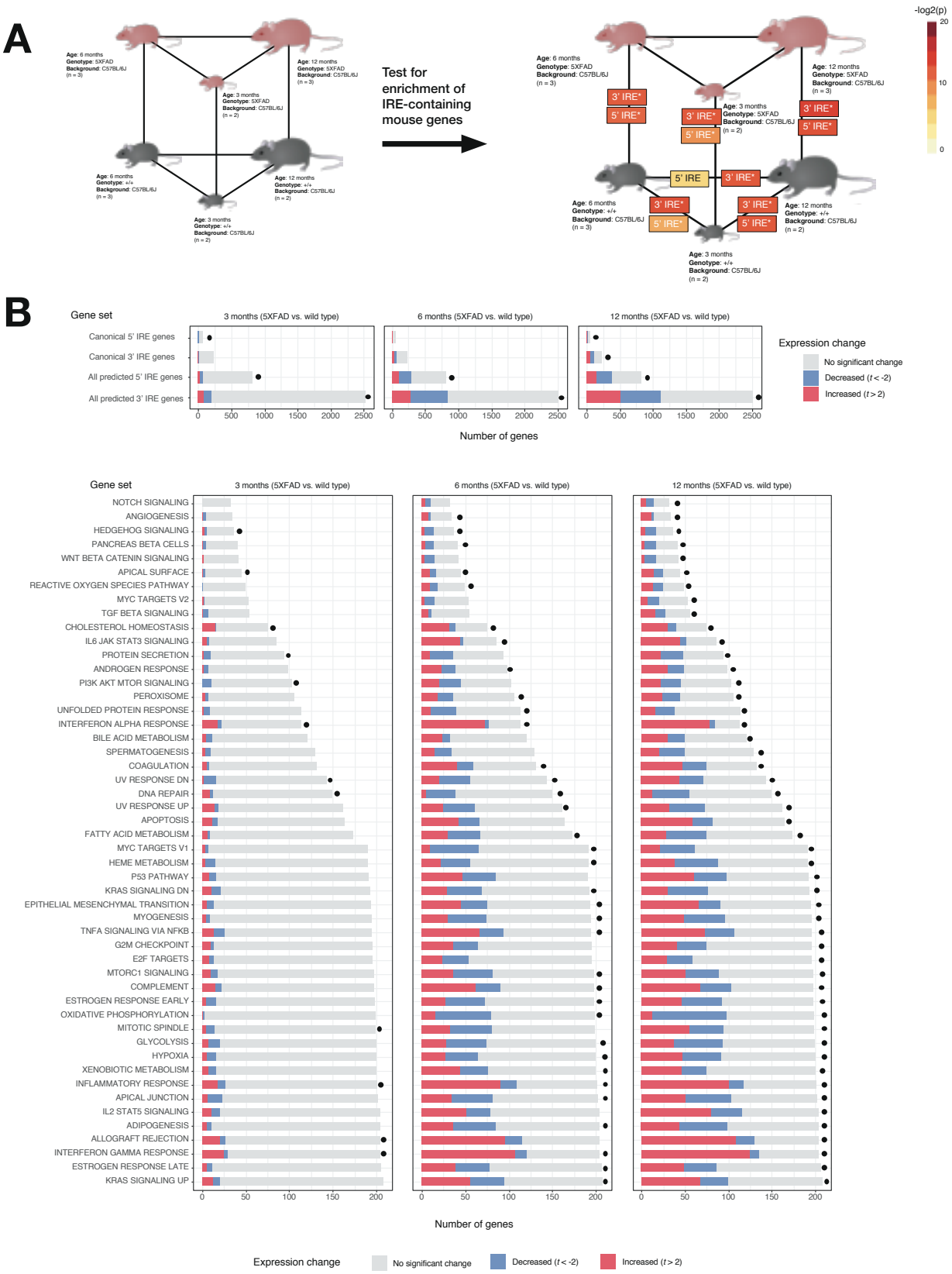

Figure 7. Analysis of 5XFAD mouse dataset. A. Experimental design and results of IRE gene set enrichment analysis. The gene sets ***all 3' IREs*** and all ***5' IREs*** derived from searching for IRE sequences in the UTRs of genes in the reference mouse genome mm10 are represented here as "3' IRE" and "5' IRE" respectively. Asterisks (*) indicate that the gene set was significantly enriched in a particular comparison (Bonferroni-adjusted *p*-value < 0.05). **B. Proportions of genes in IRE and MSigDB Hallmark gene sets which are increased (t > 2) or decreased (t < -2) in expression in all "5XFAD vs. wild type" comparisons.** A dot next to a bar indicates that the gene set was significantly enriched (FDR-adjusted p-value < 0.05).

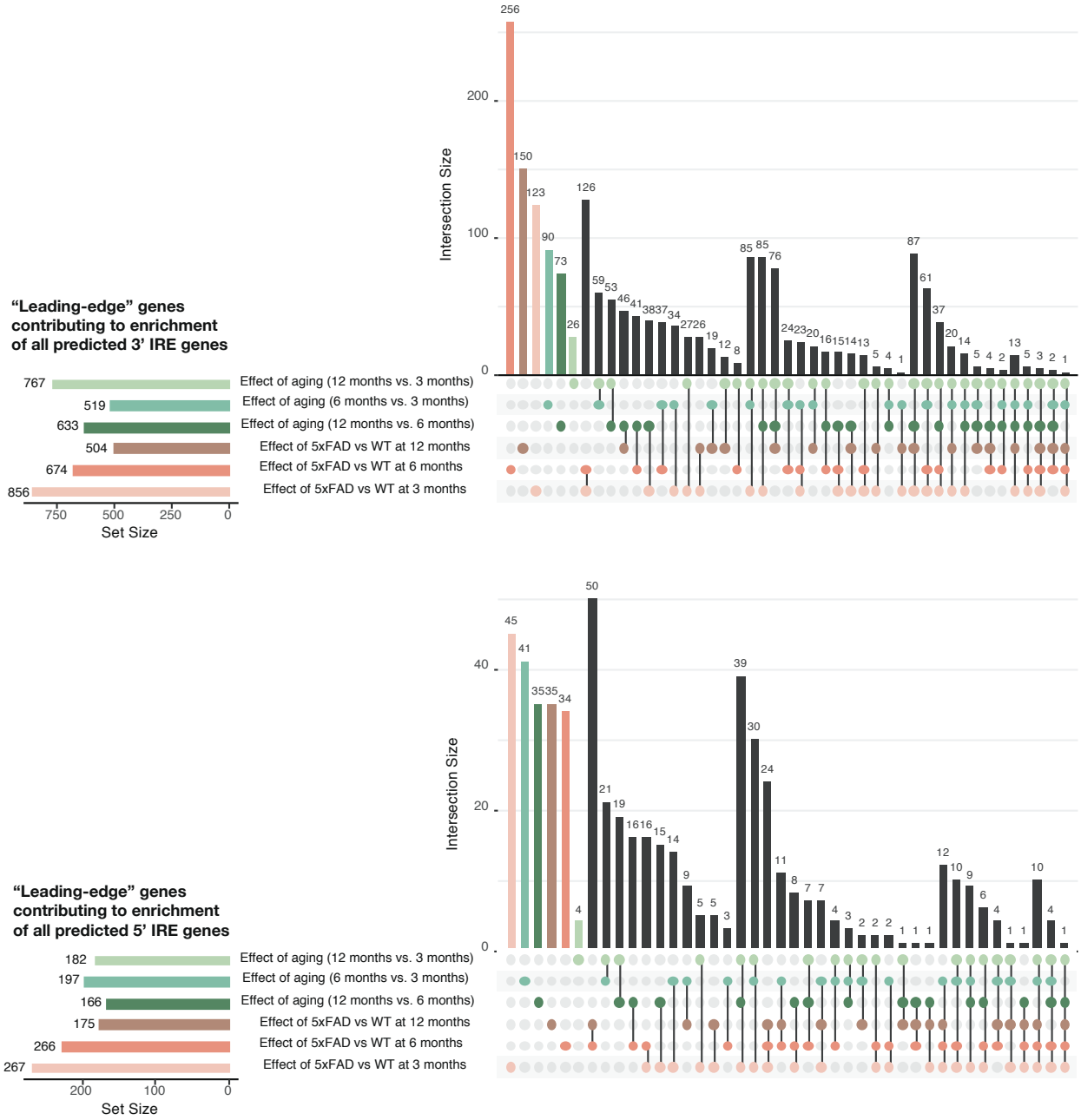

Figure 8. UpSet plots showing the overlap in GSEA leading-edge genes between all comparisons in the 5XFAD mouse datasets for the gene sets *all 3' IREs* and *all 5' IREs*. Numbers of genes for each intersection are shown above intersection bars.

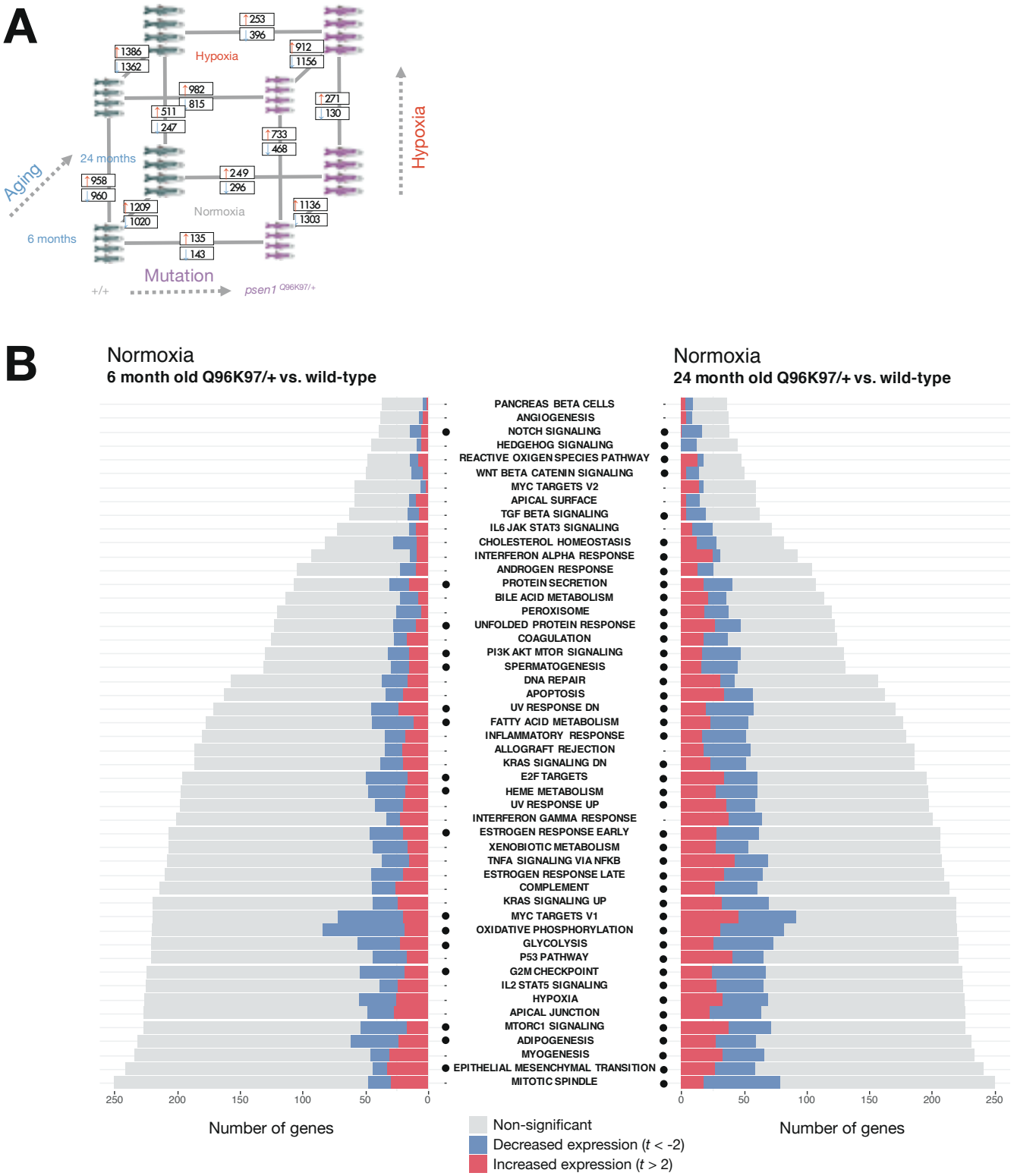

Figure 9. Differential gene expression and gene set enrichment analysis in the fAD-like zebrafish dataset. A. Results of differential gene expression analysis. Genes which were significantly increased or decreased in expression are indicated in boxes. These differentially expressed genes have FDR-adjusted p-value < 0.05. **B. Gene set enrichment with MSigDB Hallmark gene sets. The comparisons between Q96_K97del/+ fAD-like mutants and their wild-type siblings are shown for the 6-month-old (young adult) and 24-month-old (infertile adult) age groups.** Dots indicate gene sets which are significantly enriched (FDR-adjusted p-value < 0.05).

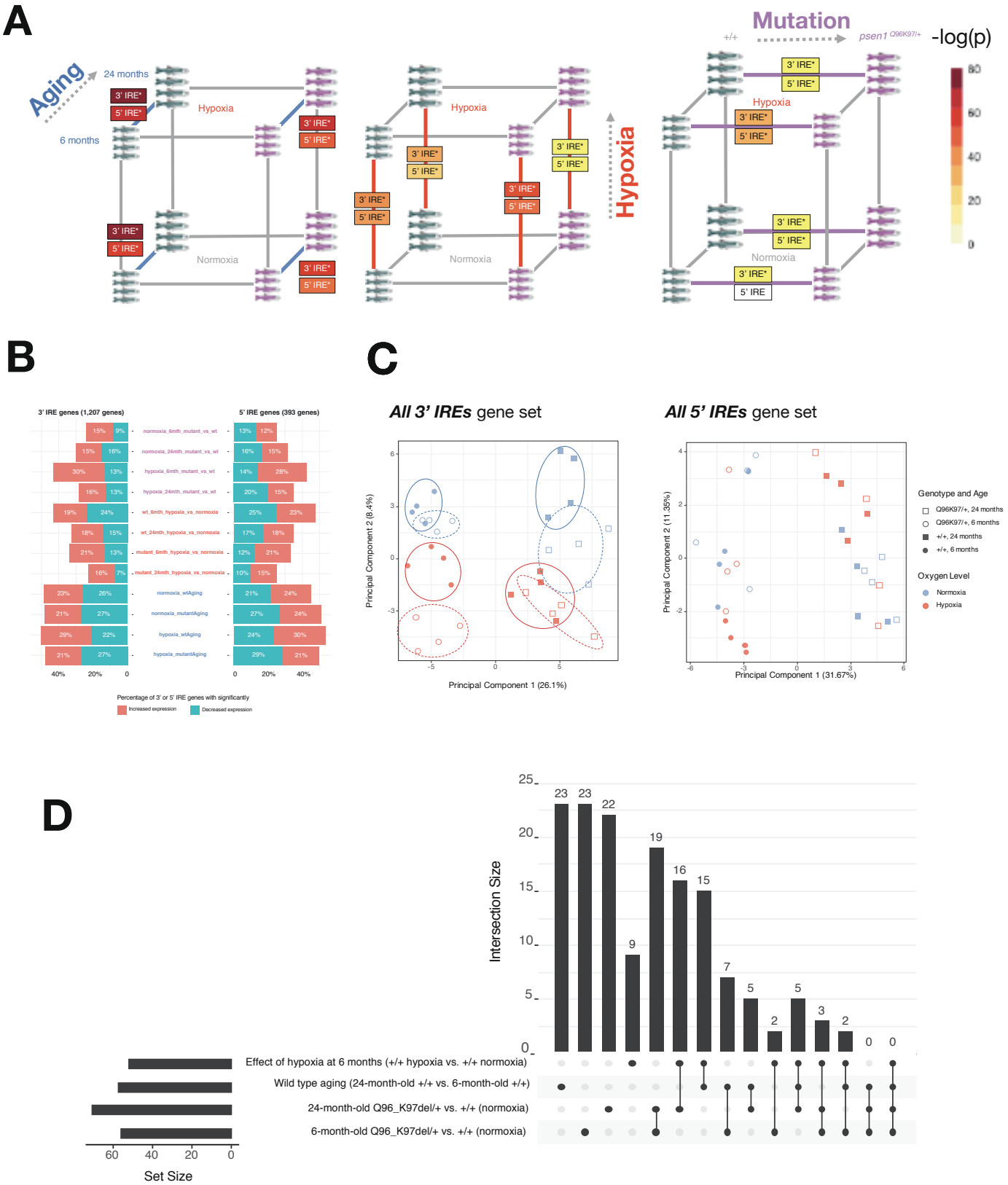

Figure 10. Iron Responsive Element (IRE)-containing gene expression in the fAD-like zebrafish dataset. A. Results of gene set enrichment testing using predicted IRE gene sets. We represent the gene sets ***all 3' IREs*** and ***all 5' IREs*** derived from searching for IRE and IRE-like sequences from z11 reference zebrafish gene UTRs as "3' IRE" and "5' IRE" in the panel. **B. Proportions of predicted IRE genes which are increased (t > 2) or decreased (t < -2) in expression for each pairwise comparison in the dataset. C. Principal component analysis of all genes in the sets *all 3' IREs* and *all 5' IREs* for all samples.** Circles on the 3' IRE plot show that different conditions generally have distinct expression of genes in the ***all 3' IREs*** set but not in the ***all 5' IREs*** set. **D. UpSet plot showing overlap in leading-edge genes for the "all predicted 3' IRE genes" for select comparisons.**