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

Author Details:

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

a. South Australian Genomics Centre, South Australian Health and Medical Research Institute, North Terrace, Adelaide, SA 5000, Australia
b. Alzheimer's Disease Genetics Laboratory, School of Biological Sciences, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
c. Dame Roma Mitchell Cancer Research Laboratories, Adelaide Medical School, Faculty of Health & Medical Sciences, University of Adelaide

* 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 is 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: To identify sets of genes predicted to contain Iron Responsive Elements (IREs) and use these to explore possible iron dyshomeostasis-associated gene expression responses in AD.

Methods: Comprehensive sets of genes containing predicted IRE or IRE-like motifs in their 3' or 5' untranslated regions (UTRs) were identified in human, mouse, and zebrafish reference transcriptomes. Further analyses focusing on these genes were applied to a range of cultured cell, human, mouse, and zebrafish gene expression datasets.

Results: IRE gene sets are sufficiently sensitive to distinguish not only between iron overload and deficiency in cultured cells, but also between AD and other pathological brain conditions. Notably, changes in IRE transcript abundance are 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 display 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. 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 across 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

83 component of HIF1, a master regulator of responses to hypoxia) is regulated by an iron-
84 dependent mechanism so that transcriptional responses to iron deficiency can resemble
85 hypoxia responses [15]. However, cellular iron homeostasis is also regulated at the post-
86 transcriptional level by the IRP/IRE system [16–18]. In this system, altered levels of available
87 ferrous iron (Fe^{2+}) cause Iron Regulatory Proteins (IRP1 or IRP2) to change conformation or
88 stability respectively [19,20]. This alters their ability to bind *cis*-regulatory Iron Responsive
89 Element (IRE) stem-loop motifs in the 3' or 5' untranslated regions (UTRs) of genes
90 encoding products related to iron metabolism. Only a few IRE-containing genes have been
91 characterized in detail, including transferrin receptor 1 (*TFR1*; 3' UTR IRE), divalent metal
92 transporter 1 (*DMT1* or *SLC11A2*; 3' UTR IRE), ferritin heavy chain and ferritin light chain (
93 *FTH1* and *FTL*; both 5' UTR IRE), and ferroportin (*SLC40A1*; 5' UTR IRE) [18]. In general,
94 these previously-characterized IRE-containing genes suggest that IRPs binding to 3' IREs
95 tend to stabilize transcripts to increase protein translation, while binding to 5' IREs
96 suppresses translation [17]. It is important to note that while the ancestral ferritin IRE is
97 present even in lower metazoans (including sponges), IRE elements in other genes have
98 likely arisen more recently through convergent evolution [21]. Perhaps because of this, the
99 properties of IRE genes can vary in different ways. For example, *SLC11A2* is stabilized by
100 IRP1 binding to a greater extent than by IRP2 binding (as its 3' IRE has a greater affinity for
101 IRP1) [22]. Similarly, while *CDC42BPA* and *TFRC* both possess 3' IREs, *CDC42BPA* is
102 stabilized to a greater extent than *TFRC* under iron deficiency [23]. There are also IRE-
103 containing genes that bind only IRP1 or IRP2 exclusively, at least in mice [24]. However,
104 global gene expression changes mediated by genes with IREs have not yet been well-
105 defined, and the overall expression patterns of IRE-containing genes have not been
106 explored in the context of AD. In addition, it is unclear how expression of these genes might
107 differ between AD and other neurodegenerative diseases, or how AD risk factors such as
108 aging and hypoxia might contribute.

In this study, we utilized the SIREs (Searching for Iron Responsive Elements) tool [25] to predict and identify sets of IRE-containing genes in human, mouse, and zebrafish. We then applied these gene sets to explore overall IRP/IRE-mediated iron dyshomeostasis-associated responses in datasets involving: (1) a cultured cell line subjected to iron overload and deficiency treatments, (2) a cohort of AD patients, healthy controls, and two other pathological conditions affecting the brain, (3) 5XFAD mice used to model the amyloid and tau pathology seen in AD, and (4) a zebrafish knock-in model possessing a familial AD (fAD)-like mutation.

Our IRE gene sets displayed significant enrichment in AD, the 5XFAD mouse model, and an fAD-like zebrafish model, demonstrating for the first time the early and extensive involvement of IRP/IRE-mediated iron dyshomeostasis-associated gene expression responses in the context of AD. IRE gene sets were sufficiently sensitive to distinguish not only between iron overload and deficiency in a cultured cell line dataset, but also between AD and other pathological conditions affecting the brain (pathological aging and progressive supranuclear palsy), implying that the dysregulation of IRE-containing genes in AD may differ from other conditions. Changes in IRE gene expression were already evident in young adult brains of animal models of AD (3-month-old 5XFAD mice, 6-month-old fAD-like zebrafish). Overall, our observations do not support the current assumption that IRP binding to 3' IREs generally stabilizes transcripts as, under most conditions, we observed both increases and decreases in the abundances of transcripts with either 3' or 5' IREs.

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. Note that potential IREs have also been identified within protein coding sequences [24,26], but were not included in our analyses under the initial assumption that the classic IRE paradigm was valid, as their possible effects on transcript stability are unknown. Each set of UTR sequences was then submitted into the SIREs web server (v.2.0 [25]). 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 [27]) with gene ID mappings between species obtained by BioMart [28].

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

We downloaded the following gene set collections from MSigDB (v.6.0 [29]): 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

159 (gene sets defined by Gene Ontology terms). We excluded the following collections from
160 analysis: C1 (specific to human chromosomes and cytogenetic bands, while our analysis
161 involves different species), C4 (computationally-defined cancer-focused gene sets), C6
162 (oncogenic signatures) and C7 (immunologic signatures). C4, C6, and C7 were not included
163 as the level of detail in the gene sets in these specific collections is more domain-specific
164 rather than broad-level. We used Fisher's exact test to determine whether any IRE gene set
165 was significantly over-represented in each MSigDB gene set. Gene sets were defined as
166 having significant enrichment for IRE gene sets if the FDR-adjusted *p*-value from Fisher's
167 exact test was below 0.05. UpSet plots were produced using the UpSetR package (v.1.4.0
168 [27]) while network representations were produced in Gephi (v.0.9.3 [30]). To produce
169 network visualizations, we first exported node and edge tables from R. The nodes table
170 contained the following gene sets: top 15 gene sets (ranked by Fisher's exact test *p*-value),
171 Hallmark Heme Metabolism gene set, ***all 3' IREs***, ***all 5' IREs***, and all genes contained within
172 these gene sets. The edges table contained gene – gene set edges which indicated the
173 gene set(s) that genes belonged to. To create the network plots in Gephi, we used "Force
174 Atlas 2" as the initial layout algorithm, followed by the "Yifan Hu" [31] layout algorithm to
175 improve visual separation between distinct groupings of genes.
176

177 Over-representation analysis of transcription factor motifs in IRE gene 178 promoters

179 Defining promoter regions as being 1500 bp upstream and 200 bp downstream of the
180 transcription start site for each gene, we used the findMotifs.pl script from HOMER (v.4.11)
181 [32,33] to search for known transcription factor binding site (TFBS) motifs in the promoters of
182 each IRE gene set. The HOMER Motif database contains 363 vertebrate transcription factor
183 binding motifs based on analysis of high-quality public ChIP-seq datasets
184 (http://homer.ucsd.edu/homer/motif/HomerMotifDB/homerResults.html). We considered

185 TFBS motifs as being significantly enriched in a gene set if the FDR-adjusted *p*-value (q-
186 value as provided in HOMER output) was less than 0.05.
187

188 ## Gene set enrichment testing

189 We performed all gene set enrichment tests in R v3.6.1 [34] using *fry* [35,36], *camera* [37],
190 and *fgsea* [38,39]. For *fry*, and *camera,* we used model fits obtained using *limma* [40,41],
191 whilst for *fgsea*, a ranked list was obtained using *moderated t*-statistics taken from *limma*. All
192 genes were used in gene set enrichment tests (i.e. not just DE genes). We combined the
193 raw *p*-values from *fry, camera*, and *fgsea* using Wilkinson's method [42] with default
194 parameters, followed by FDR-adjustment. When performing gene set enrichment testing on
195 the MSigDB Hallmark gene sets, we applied FDR-adjustment to combined *p*-values and
196 defined significant enrichment as gene sets having an adjusted *p*-value < 0.05. When
197 performing gene set enrichment on the four IRE gene sets (***all 3' IREs***, ***all 5' IREs***, ***HQ 3'***
198 ***IREs***, ***HQ 5' IREs***), we applied Bonferroni-adjustment to combined *p*-values to further
199 protect against Type I errors and defined significant enrichment as gene sets having an
200 adjusted *p*-value < 0.05. Depending on the species in the dataset being analyzed, we used
201 the respective IRE gene sets defined for human, mouse, or zebrafish.
202

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

204 We downloaded processed microarray data from GEO (accession number: GSE3573). This
205 study investigated gene expression responses to iron treatments, including iron deficiency
206 (cells in iron-free medium vs. cells cultivated in ferric ammonium nitrate, and iron overload
207 (cells cultivated in DMEM-FBS medium with hemin vs. cells in DMEM-FBS medium) [43].
208 We performed differential gene expression analysis using the "lmFit" and eBayes" functions
209 in *limma* [40]. Genes were defined as differentially expressed when their FDR-adjusted *p*-
210 value < 0.05.

211

212 Analysis of the Mayo Clinic RNA-seq dataset

213 No experiments of human subjects were performed in this study as all human data was
214 accessed from public databases and was anonymous. We downloaded processed CPM
215 count data from Synapse (https://www.synapse.org/#!Synapse:syn5550404). We matched
216 cerebellum and temporal cortex samples by their patient ID, and only retained genes which
217 were present across all samples and patients for which there were both cerebellum and
218 temporal cortex samples (n=236 patients with measurements for cerebellum and temporal
219 cortex, 472 samples in total). We performed analysis using *limma* [40,41] and determined
220 differentially expressed genes between conditions. In addition, we used the
221 "duplicateCorrelation" function in *limma*, setting the "block" parameter to the patient ID.
222 Genes were considered differentially expressed if their FDR-adjusted p-value < 0.05.

223

224 Zebrafish husbandry and animal ethics

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

229

230 fAD-like *psen1*$^{Q96\_K97del/+}$ zebrafish

231 The isolation of the *psen1*$^{Q96\_K97del}$ mutation has previously been described [44]. Zebrafish
232 mutations were only analyzed in the heterozygous state in this study.

233

234 Acute hypoxia treatment of zebrafish

235 *psen1*$^{Q96\_K97del/+}$ mutants and their wild-type siblings were treated in low oxygen levels by
236 placing zebrafish in oxygen-depleted water for 3 hours (oxygen concentration of 6.6 ± 0.2
237 mg/L in normoxia and 0.6 ± 0.2 mg/L in hypoxia.
238

239 Whole brain removal from adult zebrafish

240 After normoxia or hypoxia treatment adult fish were euthanized by sudden immersion in an
241 ice water slurry for at least ~30 seconds before decapitation and removal of the entire brain
242 for immediate RNA or protein extraction. All fish brains were removed during late
243 morning/noon to minimize any influence of circadian rhythms.
244

245 RNA extraction from whole brain

246 Total RNA was isolated from heterozygous mutant and WT siblings using the *mir*Vana
247 miRNA isolation kit (Thermo Fisher). RNA isolation was performed according to the
248 manufacturer's protocol. First a brain was lysed in a denaturing lysis solution. The lysate
249 was then extracted once with acid-phenol:chloroform leaving a semi-pure RNA sample. The
250 sample was then purified further over a glass-fiber filter to yield total RNA. Total RNA was
251 DNase treated using the DNA-free™ Kit from Ambion, Life Technologies according to the
252 manufacturer's instructions. Total RNA was then sent to the Genomics Facility at the South
253 Australian Health and Medical Research Institute (Adelaide, Australia) to assess RNA quality
254 and for subsequent RNA sequencing *(using poly-A enriched RNA-seq technology, and*
255 *estimated gene expression from the resulting single-end 75 bp reads using the reference*
256 *GRCz11 zebrafish assembly transcriptome)*
257

258 Pre-processing of the fAD-like *psen1*$^{Q96\_K97del/+}$ zebrafish dataset:

259 RNA-seq libraries contained single-end 75bp Illumina NextSeq reads. We performed quality
260 trimming with *AdapterRemoval* using default parameters. Trimmed reads were pseudo-

261 aligned to the reference zebrafish transcriptome using *Kallisto* (v.0.45) [45] and transcript
262 descriptions from Ensembl release 94. The “catchKallisto” function from *edgeR* [46] was
263 used to import and summarise counts from transcript-level to gene-level, with all subsequent
264 analyses performed at the gene-level.
265

266 Differential gene expression analysis for fAD-like zebrafish dataset

267 For differential gene expression analysis, we retained all genes with expression of at least 1
268 cpm in 4 or more samples, and used voomWithQualityWeights to downweight lower quality
269 samples [41]. Contrasts were defined to include all relevant pairwise comparisons between
270 conditions, and genes were considered as differentially expressed using an FDR-adjusted *p*-
271 value < 0.05. The *limma* model used to create the design matrix for this dataset included
272 biological sex as a variable (~0 + biologicalSex + group, where biologicalSex is either male
273 or female and group refers to the combination of age, hypoxia treatment, and genotype of
274 the samples) to minimize any sex-specific effects in the gene expression analysis. More
275 details regarding the modest effects of biological sex in this dataset are provided in
276 **Supplementary Text 1**.
277

278 Estimation of spliced and unspliced gene expression in fAD-like zebrafish dataset
279 For spliced transcripts in Ensembl release 94, we additionally defined unspliced genes
280 including intronic regions. Unspliced transcripts were appended to the end of the reference
281 transcriptome and used to build a new Kallisto [45] index. Estimated counts for spliced
282 transcripts and unspliced genes were imported into R using the “catchKallisto” function from
283 *edgeR* [46].
284

Gene set enrichment tests for other zebrafish datasets

Please refer to ref. [44] for details on RNA-seq data processing and analysis of the non-fAD-like dataset, and ref. [47] (GEO accession: GSE148631) for details on RNA-seq data processing and analysis of the 7-day-old Q96_K97del/+ dataset. In the current work, we used the gene expression counts matrix with *limma*. The voom, design, and contrasts objects produced as part of the *limma* analysis were used for gene set enrichment analysis with the zebrafish IRE gene sets we defined as well as the MSigDB Hallmark gene sets. Significantly enriched Hallmark gene sets had FDR-adjusted *p*-value < 0.05 while IRE gene sets were considered significantly enriched if the Bonferroni-adjusted *p*-value < 0.05.

Differential transcript stability analysis

The estimated spliced and unspliced transcript count estimates from *kallisto* [45] were imported into R using the catchKallisto function from *edgeR* [46]. We used *limma* [40] to determine the logFC of spliced transcripts and unspliced transcripts for each comparison. To test for whether there was a significant difference in the logFC of the spliced and unspliced transcripts, we used Welch's *t*-test with the *s2.prior* values from *limma* as the variances of the spliced and unspliced transcripts. We defined the null (no stabilization of transcript) and alternate (stabilization of transcript) hypotheses for each gene as follows, where s and u refer to the spliced and unspliced versions of a particular gene:

$$H_0: \ \mathrm{logFC_s} = \mathrm{logFC_u}$$

$$H_a: \ \mathrm{logFC_s} \neq \mathrm{logFC_u}$$

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

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

The Mayo Clinic RNA-seq study, 5XFAD mice, and fAD-like zebrafish datasets are bulk RNA-seq datasets. To confirm that any gene expression changes were likely due to altered transcriptional programs rather than changes in cell type proportions, we compared expression of marker genes for four common neural cell types (astrocytes, neurons, oligodendrocytes, microglia) in conditions within each dataset. The marker genes for astrocytes, neurons, and oligodendrocytes were obtained from MSigDB gene sets from [48] which were based on studies in mice. The marker genes for microglia were derived from [49] which was based on studies in human and mouse. All gene IDs were converted to human, mouse, or zebrafish Ensembl IDs using BioMart [28] for each dataset. Please refer to **Supplementary Text 2** for details.

Expression of previously validated IRE genes in cell line datasets

Gene symbols from Figure 2 and Supplementary Tables S1 and S2 of Sanchez et al. [50] were matched to corresponding Ensembl gene identifiers using the *biomaRt* R package [28], and filtered to include only genes present in ***HQ 3' IRE*** and ***HQ 5' IRE*** gene sets for human and/or mouse. Processed cell line microarray datasets were downloaded from GEO and included a human Caco-2 cell line dataset (GEO accession GSE3573), and mouse splenic B cell dataset (GEO accession GSE77306). The datasets were filtered for validated IRE genes and boxplots were used to visualize expression between groups. *t*-tests were used to test for differences in expression between groups, with p-values ≤ 0.05 considered statistically significant. The Sankey plot was produced using the *networkD3* R package [51].

Results

Defining sets of genes containing Iron Responsive Elements (IREs)

We utilized the SIREs (Searching for Iron Responsive Elements) tool, to define species-specific IRE gene sets by searching for IRE and IRE-like motifs in the 3' and 5' untranslated regions (UTRs) of the reference transcriptomes of human (hg38), mouse (mm10), and zebrafish (z11). *SIREs* assigns a quality-score to all detected IREs, with "high-quality" scores corresponding to canonical IREs and "medium-quality" or "low-quality" scores reflecting deviations from the canonical IRE (alternative nucleotide composition in the apical loop, one bulge at the 3' or one mismatch in the upper stem) that would still produce an IRE-like motif with potential to be functional [25,52–55]. **Figure 1** summarizes the IRP/IRE interaction effects on transcripts and gives examples of canonical and non-canonical IREs.

We defined four gene sets for each species as follows: ***HQ 3' IREs*** (high-quality predicted 3' IRE genes), ***HQ 5' IREs*** (high-quality predicted 5' IRE genes), ***all 3' IREs*** (including all low, medium, and high-quality predicted 3' IRE genes), and ***all 5' IREs*** (including all low, medium, and high-quality predicted 5' IRE genes). The size of these gene sets for each species and the overlap present is shown in **Figure 2** and the gene sets are provided in **Supplementary Table 1**. Overall, searching through human UTRs uncovered the largest number of predicted IRE genes, followed by mouse, and then zebrafish. Overlap between homologous genes from the IRE gene sets of different species was generally poor. The largest sets of genes shared between species were consistently found between human and mouse, which is reflective of their closer evolutionary divergence.

While not many high-quality IRE genes were identified across all three species, the few we identified are consistent with known and well-characterized IREs in the literature [56,57]. For example, the single ***HQ 3' IRE*** gene shared between the three species was *TFRC*

360 (transferrin receptor), while the ***HQ 5' IRE*** genes shared between the three species included
361 *FTH1* (ferritin heavy chain 1) and *ALAS2* (5'-aminolevulinate synthase 2). We note that the
362 scoring of genes for the ***HQ 3' IRE*** and ***HQ 5' IRE*** sets is considered stringent (i.e. only
363 canonical IREs). These two gene sets recover known IRE genes in human, including
364 *CDC14A* (3' IRE; [50]), *CDC42BPA* (3' IRE; [23,58], *SLC40A1* (5' IRE; [59]), and *FTL* (5'
365 IRE; [60]). These genes are also recovered in the ***All 3' IRE*** and ***All 5' IRE*** sets across
366 mouse and zebrafish too. In addition, the 3' IRE of *SLC11A2* (also known as *DMT1*) [22]
367 which is only found in mammals, was also recovered in the human ***HQ 5' IREs*** set and
368 mouse ***All 3' IREs*** set (but not for zebrafish). The expression of these previously
369 characterized IRE genes across the human, mouse, and zebrafish datasets analyzed in this
370 current work is described in **Supplementary Text 3** and discussed further in the analyses
371 we later describe.
372

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

375 We explored the biological relevance of the predicted IRE gene sets described above by
376 testing whether genes within them are over-represented in existing gene sets from the
377 Molecular Signatures Database (MSigDB; available at https://www.gsea-
378 msigdb.org/gsea/msigdb). MSigDB represents one of the largest publicly-available
379 collections of gene sets collated from existing studies and other biological databases [29].
380 We limited our analysis to gene sets from the following collections: Hallmark (non-redundant
381 sets of ~200 genes each representing various biological activities), C2 (gene sets from
382 databases including KEGG and Reactome and published studies), C3 (gene sets containing
383 genes with motif elements), and C5 (gene sets based on gene ontology terms) (see
384 **Methods**). We performed over-representation analysis for the predicted IRE gene sets (***all***
385 ***3' IREs*** and ***all 5' IREs***) separately for each species (human, mouse, zebrafish), and used

386 Wilkinson's meta-analytic approach to determine gene sets that were significantly over-
387 represented across the three species' IRE gene sets. Our results indicated that 1,148 of
388 10,427 tested gene sets displayed over-representation of IRE gene sets across all three
389 species (Bonferroni-adjusted Wilkinson's *p*-value < 0.05) (**Figure 3**; **Supplementary Table**
390 **2**). The bar plots in **Figure 3** show the 15 MSigDB gene sets with the most significant
391 enrichment for human, mouse, and zebrafish. Remarkably, the "Blalock Alzheimer's Disease
392 Up" gene set from the C2 collection was significantly over-represented in both the sets of ***all***
393 ***3' IREs*** (Bonferroni-adjusted Wilkinson's *p*-value = 3.9e-14) and ***all 5' IREs*** (Bonferroni-
394 adjusted Wilkinson's *p*-value = 2.1e-26) in the meta-analysis, and was also the gene set with
395 the most significant over-representation of the human ***all 3' IREs*** set (Bonferroni-adjusted
396 Fisher's exact test *p*-value = 7.8e-58) (**Supplementary Table 2**). This supports that
397 disturbance of iron homeostasis particularly distinguishes Alzheimer's disease from other
398 disease conditions and pathways represented within the C2 collection. In addition, the top 15
399 MSigDB gene sets showing the most significant over-representation generally differed
400 between species. However, in all cases, a large proportion of IRE genes from the predicted
401 IRE gene sets were not contained within any of these top-ranked MSigDB gene sets or
402 within the Heme Metabolism gene set belonging to the Hallmark collection (**Figure 3**). This
403 is visualized through the network plots of **Figure 3**, which show that the predicted IRE genes
404 we defined are not fully captured by existing gene sets. This suggests that IRE sets may be
405 uniquely useful for investigating gene expression changes during the IRE-IRP response to
406 iron dyshomeostasis.
407

408 Involvement of transcription factor regulation within IRE gene sets

409 To be confident that IRE gene sets accurately capture information about IRP/IRE-mediated
410 responses, we needed to investigate possible co-ordinate regulation by other factors. As a
411 starting point, we examined whether known binding motifs for transcription factors were

412 significantly over-represented in the promoter regions of genes within each IRE gene set (***all***
413 ***3' IREs***, ***all 5' IREs***, ***HQ 3' IREs***, and ***HQ 5' IREs***) for each species. We detected significant
414 over-representation of several transcription factors including the Klf14 motif in the zebrafish
415 ***all 3' IREs*** set (FDR-adjusted *p*-value = 0.049), the E2F motif in the human ***all 5' IREs*** set
416 (FDR-adjusted *p*-value = 0.049) and the Bach1 (FDR-adjusted *p*-value = 0.012), Nrf2 (FDR-
417 adjusted *p*-value = 0.013), and NF-E2 (FDR-adjusted *p*-value = 0.019) motifs in the
418 zebrafish ***HQ 5' IREs*** set (**Supplementary Table 3**). This suggests that the following results
419 should be interpreted with the knowledge that expression of subsets of genes in some of the
420 IRE gene sets may be influenced by factors other than iron ions.
421

422 Gene set enrichment testing approach

423 Our predicted IRE gene sets can be used in any standard gene set enrichment testing
424 approach to detect potential changes in iron homeostasis between conditions. Our workflow,
425 which we later apply to human, mouse, and zebrafish datasets, is shown in **Figure 4**. Due to
426 variability in the results produced by different gene set enrichment testing methods, we were
427 inspired by the EGSEA framework [61] to combine the results from different methods. Based
428 on an initial analysis using EGSEA, we chose *fry/mroast* [35,36], *camera* [37], and *fgsea*
429 [38,39] as the representative methods to use. (See **Supplementary Figure 2** for a principal
430 component analysis of results from different gene set enrichment analysis approaches.) A
431 summary of the different characteristics of the three methods is shown in **Table 1**. The raw
432 enrichment *p*-values from these approaches can be combined to obtain an overall
433 enrichment *p*-value for each gene set. In accordance with EGSEA default parameters, we
434 used Wilkinson's method to combine raw *p*-values, and then applied adjustment for multiple
435 testing on all combined *p*-values. Along with performing this gene set enrichment testing on
436 our IRE gene sets, we also recommend using the same method to test for enrichment for the
437 MSigDB Hallmark gene sets as the diverse biological activities they represent help to

438 provide context for interpreting the IRE enrichment results. To explore further the results of
439 IRE gene set enrichment analysis, we use UpSet plots to display the overlap between sets
440 of "leading-edge" genes in the ***all 3' IREs*** and ***all 5' IREs*** gene sets. The leading-edge
441 genes can be interpreted as the core (often biologically important) genes of a gene set that
442 account for the significant enrichment as calculated by GSEA [39].
443

444 Differences in IRE gene set enrichment during iron deficiency and iron
445 overload in a cultured cell line

446 We first tested how our enrichment approach illuminated the effects of iron overload and
447 deficiency in a cultured cell line microarray dataset from Caco-2 cells (GEO accession:
448 GSE3573). As only one cell type contributed to this dataset, interpretation of the IRE
449 enrichment results is simplified by not having to consider the differing iron requirements of
450 different cell types. This allowed us to focus on whether the iron dyshomeostasis treatments
451 (iron overload and iron deficiency) could be detected and distinguished in terms of their
452 IRP/IRE system-driven transcript abundance response. Because the dataset was from a
453 microarray experiment, we performed only the differential gene expression and gene set
454 enrichment testing portions of our workflow. The results of Principal Component Analysis
455 and differential gene expression analysis are provided in **Supplementary Figure 3**. We note
456 that this particular dataset was chosen despite not being derived from a central nervous
457 system cell type, as maintaining iron homeostasis is fundamental to all cell types (regardless
458 of their tissue of origin). In addition, this particular dataset included iron deficiency and iron
459 overload treatments in the same batch and with sufficient sample number, and no other
460 cultured cell line gene expression dataset was publicly available at the time that fulfilled
461 these criteria.
462

463 In general, we found iron deficiency and iron overload treatments resulted in different gene
464 expression responses. In terms of classic IRE genes such as *TFRC, FTL,* and *FTH1*, iron
465 deficiency is associated with increased expression of *TFRC* (consistent with increased
466 transcript stabilization of this 3' IRE gene) and decreased expression of the 5' IRE genes
467 *FTL* and *FTH1,* and these changes are in the opposite direction during iron overload (see
468 **Supplementary Text 3** and **Supplementary Figure 1A**). Differential gene expression
469 across all detected genes in the dataset additionally revealed that iron deficiency was
470 associated with 96 differentially expressed genes of which 10 possessed predicted IREs
471 while iron overload was associated with 212 differentially expressed genes (FDR-adjusted *p*-
472 value from *limma* < 0.05) of which 33 possessed predicted IREs (**Supplementary Figure**
473 **3B**). There were 17 differentially expressed genes in common between iron deficiency and
474 iron overload treatments, and all moved in opposite directions according to the treatment
475 (i.e. increased abundance under iron deficiency and decreased abundance under iron
476 overload, or vice versa). These clear differences between iron deficiency and iron overload
477 were reflected in gene set enrichment analyses using the MSigDB Hallmark gene sets,
478 where the gene sets involved and the proportions of gene transcripts with increased or
479 decreased abundance differed (**Figure 5A**). As expected, IRE gene sets also showed
480 significant enrichment under iron deficiency and overload conditions (**Figure 5B**) (Bonferroni
481 adjusted *p*-value < 0.05).
482
483 We expected that iron deficiency would result in increased expression of genes with 3' IREs
484 under the IRP/IRE paradigm similar to the *TFRC* gene as seen in **Supplementary Text 3**.
485 However, both the 3' and 5' IRE gene sets displayed mixed patterns of increased and
486 decreased expression under both the iron deficiency and iron overload treatments (**Figure**
487 **5B**). This indicates it would be difficult to distinguish between these conditions based purely
488 on overall increases or decreases in the expression of IRE gene sets. Despite this, we see

489 that the iron deficiency and iron overload treatments can be distinguished by their "leading-
490 edge" genes (those genes contributing most to the enrichment signal for the predicted IRE
491 gene sets (***all 3' IREs*** and ***all 5' IREs***) (**Figure 5C**). This supports that gene set enrichment
492 using predicted IRE gene sets may be sufficiently sensitive to detect whether iron
493 dyshomeostasis is present and potentially more informative than individual IRE genes in
494 distinguishing between different IRP-IRE system-mediated gene expression responses in
495 iron deficiency and iron overload treatments in this cell line.
496

497 A distinct iron homeostasis response in human AD patients compared to
498 other neuropathologies

499 Given that IRE gene sets could distinguish between iron overload and deficiency in a
500 cultured cell line, we next tested our gene sets on a more complex dataset including
501 cerebellum and temporal cortex tissue samples from post-mortem human brains. The brains
502 originated from either healthy controls or patients afflicted with one of three conditions:
503 Alzheimer's disease (AD); pathological aging (PA), a condition involving amyloid pathology
504 but no significant dementia symptoms; or progressive supranuclear palsy (PSP), a
505 tauopathy without amyloid pathology [62]. An important characteristic of the dataset is that
506 both cerebellum and temporal cortex tissue samples were available from each patient. A
507 summary of the 236 patients whose data we included in this analysis is shown in **Table 2**
508 and the results of differential gene expression analysis and IRE gene set enrichment
509 analysis are shown in **Table 3** and **Figure 6A**, and summarised in comparison to the mouse
510 and zebrafish datasets later analysed in **Table 5**. The expression of classic IRE genes
511 including *TFRC*, *FTL,* and *FTH1* in this dataset is shown in **Supplementary Text 3** and
512 **Supplementary Figure 1B-C**. In our analyses, we focus mainly on comparing conditions
513 within each tissue rather than between tissues. This is because we found significant

514 differences in the AD vs. control comparison in the temporal cortex compared to the
515 cerebellum (see **Supplementary Figure 4**).
516
517 Overall, our IRE enrichment analyses indicate significant enrichment of IRE gene sets in all
518 pathological conditions (AD, PA or PSP) compared to healthy controls within the cerebellum
519 and temporal cortex (**Figure 6A**). In all pathological conditions, 3' and 5' IRE gene
520 transcripts show overall mixed patterns of abundance (e.g. increased and decreased).
521 Further examination of the leading-edge genes from these IRE gene sets gives more insight
522 regarding potential differences and similarities between AD, PA, and PSP (**Figure 6B**).
523 Overall, AD, PSP, and PA appear to involve distinct yet partially overlapping IRE gene
524 expression responses in a tissue-specific manner. Within the temporal cortex, there are 435
525 3' IRE leading-edge genes and 178 5' IRE leading-edge genes exclusively present in the
526 "AD vs. control" comparison. These are greater numbers than for the other two conditions.
527 Interestingly, AD and PA share only relatively small numbers of leading-edge genes despite
528 the fact that many regard PA as a prodrome of AD [63]. Also interesting is that, in the
529 cerebellum, while AD and PSP share many 3' and 5' IRE leading-edge genes, PA is
530 associated with a large number of unique leading-edge 3' and 5' IRE genes, further
531 emphasizing its difference from AD and also PSP. These observations are supported when
532 considering expression of previously characterized IRE genes individually (see
533 **Supplementary Text 3** and **Supplementary Figure 1B-C**). The results in this
534 Supplementary Text suggest that, in particular, some 3' IRE genes (e.g. *TFRC*, *CAV3*,
535 *CDC14A*, *CDC42BPA, SERTAD2, TRAM1*) show increased expression in AD (relative to
536 controls) whereas PA and PSP do not. Overall, our results suggest that IRP-IRE system-
537 mediated gene expression changes may be sufficiently sensitive to discern and identify
538 potential differences in the way iron homeostasis is regulated in AD compared to PA and
539 PSP.

540

541 ## Age-dependent disruption of IRE-driven iron homeostasis in the 5XFAD
542 mouse model

543 Alzheimer's disease is thought to develop over decades [64–66]. However, detailed
544 molecular studies of AD brains can only be based on post-mortem tissues. To reveal the
545 early molecular changes that initiate the progression to AD we must study the brains of
546 animal disease models. Given that the IRE gene sets appear to work well in human
547 datasets, we then tested our mouse IRE gene sets on an RNA-seq dataset derived from
548 brain cortex tissue of the 5XFAD transgenic mouse AD model (GEO: GSE140286). The
549 5XFAD mouse is one of the most common systems used to model the amyloid beta
550 histopathology of AD brains and, in a concordance analysis, brain gene expression patterns
551 in this model were previously shown to resemble those of human AD brains to a greater
552 degree than for four other mouse models (3xTg, CK-p25, J20, and Tg2576) [67]. 5XFAD
553 mice possess two transgenes that include a total of five different mutations, each of which
554 individually causes fAD in humans. In this dataset, the mice analyzed were either 3, 6 or 12
555 months of age. We acknowledge that the small sample size of this particular dataset means
556 that results should be interpreted tentatively (see Discussion).

557

558 Using gene set enrichment testing methods as before, we observed significant enrichment of
559 the ***all 3' IREs*** and ***all 5' IREs*** gene sets in several comparisons. These included 5XFAD vs.
560 wild type comparisons and wild type aging comparisons (Bonferroni-adjusted enrichment *p*-
561 value < 0.05) (**Figure 7**). Notably, even the youngest age group of 5XFAD mutant mice (3
562 months) displayed significant enrichment of genes containing 3' or 5' IREs compared to age-
563 matched wild types. This is consistent with an enrichment of immune-related Hallmark gene
564 sets that we observed in this age group (see **Figure 7B**) and with a previous transcriptome
565 analysis suggesting immune activation in 5XFAD mice as early as 2-4 months [68]. UpSet

566 plots of overlapping leading-edge genes suggest that the IRE responses due to aging and
567 due to the 5XFAD transgenes may involve partially overlapping gene expression changes
568 (**Figure 8**). In addition, the UpSet plots reveal subsets containing large numbers of IRE-
569 containing genes which uniquely contribute to enrichment of IRE gene sets in only one
570 comparison (e.g. 256 3' IRE genes only contribute to enrichment in the "*5xFAD vs WT at 6*
571 *months*" comparison). Notably, there are 126 shared 3' IRE genes contributing to enrichment
572 in the "*5xFAD vs WT at 6 months*" and "*5xFAD vs WT at 3 months*" comparisons, but no
573 genes shared between these comparisons and the "*5xFAD vs WT at 12 months*"
574 comparison. These 126 genes were found to show over-representation of Gene Ontology
575 terms including "Hsp70 protein binding" and "nuclear import signal receptor" (FDR-adj. *p*-
576 value < 0.05) (**Supplementary Table 4**), suggesting that subsets of leading-edge genes
577 may have biological relevance, and that these may give insight into the biologically distinct
578 age-dependent iron dyshomeostasis responses caused by the 5XFAD transgenes. Although
579 beyond the scope of our current analysis, further investigation into these subsets of genes
580 may contribute to definition of differences in iron dyshomeostasis responses under different
581 biological conditions of interest.
582

583 Similarities in IRE-driven iron homeostasis responses during hypoxia
584 and from a familial AD-like mutation in a zebrafish model

585 Concerns have been raised over the relevance of transgenic mouse models in modelling the
586 early stages of AD (reviewed by [69]). Meanwhile, knock-in models of familial AD (fAD) (that
587 possess a single, heterozygous mutation in an endogenous gene like the human version of
588 the disease) only exhibit subtle pathological changes with little to no visible amyloid or tau
589 pathology present [70]. Despite the comparatively more subtle pathology seen in
590 heterozygous, single fAD-like mutation knock-in models, since they closely mimic the
591 genetic state of fAD, analysis of their brain gene expressionthem may reveal molecular

592 changes relevant to early fAD states in human brains. We had access to a whole-brain RNA-
593 seq dataset from a knock-in zebrafish model of fAD possessing a heterozygous, single fAD-
594 like mutation in its endogenous *psen1* gene (*psen1*$^{Q96\_K97del}$/+) (henceforth referred to as the
595 "fAD-mutation-like" model in this paper). Previous analysis of a subset of this dataset
596 involving young adult (6-month-old) brains revealed gene expression changes related to
597 energy metabolism and mitochondrial function, especially ATP synthesis and other ATP-
598 dependent functions including lysosomal acidification [44]. (In contrast, the 3-month-old
599 young adult 5XFAD mouse brain dataset is dominated by immune/inflammation responses,
600 **Figure 7B**). Considering the critical role of iron homeostasis in energy metabolism, we
601 decided to revisit this zebrafish dataset. We performed IRE gene set enrichment on the
602 entire dataset in order to include exploration of the effects of aging and acute hypoxia (two
603 important risk factors for sporadic late onset AD) and to analyze how these effects interact
604 with the fAD-like *psen1*$^{Q96\_K97del}$/+ mutant genotype. The experimental design and the results
605 of the differential gene expression analyses and gene set enrichment tests are summarized
606 in **Figure 9**.
607
608 We first turned our attention to the comparison between young adult (6-month-old)
609 *psen1*$^{Q96\_K97del}$/+ mutant zebrafish and their wild-type siblings. At this age, gene expression
610 changes in the mutant fish likely represent early stresses driving the development of fAD in
611 humans. Gene set enrichment tests in this comparison identify alteration of energy
612 metabolism-related Hallmark gene sets (e.g. OXIDATIVE PHOSPHORYLATION,
613 GLYCOLYSIS) (**Figure 9B**) which is consistent with a previous analysis of gene ontology
614 terms with this dataset [44]. In addition, we see enrichment of other gene sets including
615 FATTY ACID METABOLISM, PI3K AKT MTOR SIGNALLING, MTORC1 SIGNALLING, and
616 HEME METABOLISM. All of these gene sets showing enrichment in 6-month-old
617 *psen1*$^{Q96\_K97del}$/+ mutants relative to wild-type siblings are also enriched in 24-month-old

618 ("middle-aged") *psen1*$^{Q96\_K97del}$/+ mutants relative to their wild-type siblings (**Figure 9B**). This
619 supports that the biological activities represented in these gene sets are amongst the
620 earliest altered in this particular fAD mutation model.
621
622 In Lumsden et al. [14] we predicted that fAD mutations in the major locus *PSEN1* might
623 cause an early cellular deficiency of ferrous iron due to the observation of insufficient
624 acidification of the endolysosomal pathway in in-vitro *PSEN1* mutation studies [71,72]. In
625 Newman et al. [44] we saw that Gene Ontology enrichment analysis of 6-month-old
626 *psen1*$^{Q96\_K97del}$/+ zebrafish brains supported that lysosomal acidification was affected.
627 Therefore, we applied our IRE enrichment analysis to test for evidence of iron
628 dyshomeostasis in these fish. The enrichment of the ***all 3' IREs*** gene set in 6-month-old
629 *psen1*$^{Q96\_K97del}$/+ zebrafish brains supports that iron dyshomeostasis is an important stress in
630 the early stages of fAD (**Figure 10A**).
631
632 While almost all pairwise comparisons in the zebrafish dataset show significant enrichment
633 of at least one IRE gene set (**Figure 10A**), the expression of IRE genes appears to differ in
634 terms of the proportions of IRE-containing transcripts which show increased versus
635 decreased abundance (**Figure 10B**). In addition, the Principal Component Analysis plot of
636 expression of the ***all 3' IREs*** gene set over all samples shown in **Figure 10C** suggests that
637 different conditions appear to drive distinct patterns of expression of these genes. Across the
638 first principal component, different age groups (6- and 24-month-old brains) differ in their
639 expression of predicted 3' IRE genes, while the second principal component appears to
640 separate *psen1*$^{Q96\_K97del}$/+ mutants from their wild-type siblings. This separation between
641 *psen1*$^{Q96\_K97del}$/+ mutants and wild type siblings is even more pronounced when both are
642 exposed to hypoxia.
643

644 To gain more insight into the similarities and differences between the IRE responses in the
645 *psen1*$^{Q96\_K97del}$*/+* vs. wild type comparison, we plotted UpSet plots of overlapping leading-
646 edge genes (**Figure 10D**). These plots suggest that the IRE responses during hypoxia,
647 during aging, and due to this fAD-like mutation, are mostly distinct from each other with
648 unique leading-edge genes. However, similarities in the IRE responses between different
649 conditions are suggested by the existence of some shared leading-edge genes. For
650 example, for the set of ***all 3' IREs***, the "*6-month-old psen1*$^{Q96\_K97del}$*/+ vs. wild type*"
651 comparison shares 19 leading-edge genes with the "*24-month-old psen1*$^{Q96\_K97del}$*/+ vs. wild*
652 *type*" comparison. As an initial exploration of the biological relevance of these genes, we
653 submitted them to the STRINGR tool. This indicated that the proteins products of these
654 genes were significantly associated with each other (e.g. in terms of text-mining from
655 PubMed articles, experimentally determined interactions, and co-expression). These
656 proteins were significantly over-represented in the gene sets "MAPK pathway", "AP-1
657 transcription factor", "Jun-like transcription factor", and "signaling by TGF-beta family
658 members" (FDR-adjusted over-representation p-value < 0.05; **Supplementary Figure 5**).
659 The AP-1 and MAPK pathways have previously been shown to be stimulated by iron
660 depletion [73,74]. Therefore, mechanistically, it is possible that the IRE-IRP response to iron
661 dyshomeostasis in the fAD-like zebrafish mutant involves these pathways.
662
663 The fact that aging, hypoxia, and a fAD-like mutation all cause changes in the abundance of
664 IRE-containing transcripts in zebrafish raised concerns regarding the specificity of such
665 changes. Therefore, as a negative control, we examined changes in IRE transcript
666 abundance in a brain transcriptome dataset derived from zebrafish heterozygous for a *psen1*
667 mutation now thought not to be fAD-like, *psen1*^{K97fs} [75]. In this dataset, *psen1*^{K97fs}/+
668 zebrafish are compared to their wild type siblings at 6 and 24 months. We tested this dataset
669 for enrichment using our zebrafish 3' and 5' IRE gene sets but found no significant

670 enrichment of any of our predicted IRE gene sets in *psen1*^{K97fs}/+ vs. wild type comparisons
671 at any age (**Table 4**; **Supplementary Figure 6**). Reassuringly, we still observed significant
672 enrichment of both 3' and 5' IRE gene sets during wild type aging (*24-month-old* wild types
673 *vs. 6-month-old* wild types), consistent with the equivalent comparison in the
674 *psen1*$^{Q96\_K97del}$/+ dataset. These results support that IRE-containing transcript abundance
675 changes are sufficiently sensitive to reflect differences in iron homeostasis between different
676 mutation models.
677

678 Simultaneous stabilization of some 3' IRE transcripts and destabilization
679 of others

680 In the cultured cell line dataset analyzed above, we noticed that even a straightforward iron
681 deficiency treatment resulted in the simultaneous increase and decrease in expression of 3'
682 IRE-containing genes. These findings are difficult to reconcile with the current paradigm that
683 stabilization of 3' IRE-containing genes occurs under iron deficiency and suggest that the
684 current model of the IRP/IRE system may be incomplete or insufficient for describing the
685 regulation of a broader collection of IRE genes. Given that many predicted 3' IRE genes with
686 non-canonical IREs (e.g. in the ***all 3' IREs*** set) displayed enrichment and different gene
687 expression patterns in the different genotype/age/oxygenation groups of the fAD-like
688 zebrafish dataset, we decided to explore further the stability changes of these genes by
689 comparing the expression of spliced and unspliced transcripts for each gene
690 (**Supplementary Text 4**). We found that transcripts of some predicted 3' IRE genes were
691 significantly increased in stability while others were significantly decreased in stability
692 (**Supplementary Figure 7**; **Supplementary Table 5**).

693

694 Some validated IRE genes demonstrate expression changes

695 inconsistent with the classic IRE paradigm

696 The work described above discovered both increased and decreased stability of transcripts
697 predicted to possess IREs in fAD-like zebrafish brains relative to wild-type brains. However,
698 it did not address directly whether such changes in stability occur in transcripts confirmed as
699 targets of IRP binding and in systems subjected to disturbance of iron homeostasis. To
700 support that our sets of 5'- and 3'-IRE-containing genes include both transcripts that
701 increase in abundance and transcripts that decrease in abundance as iron levels change,
702 we examined several IRE genes previously demonstrated by Sanchez et al. [24] to produce
703 transcripts bound by IRPs (see below). We examined these transcripts in transcriptome
704 datasets produced from cell lines subjected to changes in iron levels (**Figure 11**).

705

706 In Sanchez et al. [24], a total of 258 mRNAs were identified that immunoprecipitated with
707 mouse IRP1 and/or IRP2 (44 immunoprecipitating with both IRP1 and IRP2, 113
708 immunoprecipitating exclusively with IRP1, and 101 immunoprecipitating exclusively with
709 IRP2). We filtered these mRNAs for those with high-quality IREs (as predicted using SIREs)
710 in the UTR of either the human or the mouse Ensembl gene model (**Figure 11A**). We found
711 30 mRNAs with high-quality IREs in human Ensembl gene models, and 50 mRNAs with
712 high-quality IREs in mouse Ensembl gene models. These were then separated into those
713 with IREs in their 3' UTRs and those with IREs in their 5' UTRs (see **Figure 11A** for more
714 details), since transcripts with 3' IREs are expected to show altered abundance under the
715 classic IRE paradigm while the abundance of transcripts containing 5' IREs is not expected
716 to change (see **Figure 1**). Lists of experimentally-validated IRE genes filtered using these
717 steps are available in **Supplementary Table 6**.

718

719 We then examined the expression of these experimentally-validated IRE-containing genes in
720 transcriptome datasets from human and mouse cells subjected to iron level-altering
721 treatments. We sought to observe whether transcript abundance differences were consistent
722 with those predicted under the classic IRE paradigm. The datasets included the human
723 Caco-2 cultured cell line dataset (previously described above), and a mouse splenic B cell
724 dataset [76] (GEO Accession: GSE77306).
725
726 The Caco-2 dataset includes both iron overload (DMEM-FBS Medium + Hemin vs. DMEM-
727 FBS Medium) and iron deficiency (Iron-free Medium vs. Iron-free Medium +FAC) treatments,
728 while the mouse B cell dataset includes only a treatment with deferoxamine (DFO) to create
729 iron deficiency (DFO-treated cells vs. control). We note that the DFO and control samples in
730 the mouse B cell dataset are not directly comparable within each group, due to the
731 unstimulated or stimulated anti-IgM or LPS treatments also used in this study. However, a
732 Principal Component Analysis plot of the samples shows that the DFO-treated samples
733 apparently separate from the control samples across Principal Component 2, indicating that
734 DFO-treatment is likely to be one of the main causes of differential gene expression in this
735 dataset (**Supplementary Figure 8**).
736
737 Overall, in the mouse B cells and human Caco-2 cells affected by changes in iron levels,
738 approximately half of the genes identified with validated IREs show significant differential
739 expression. Of those, similar proportions show increased or decreased transcript abundance
740 in either the 3' IRE or 5' IRE groups and, of those, there are numerous genes showing
741 changes in expression inconsistent with the IRE paradigm. In particular, in the Caco-2 cells,
742 three 3' IRE genes (*EIF4EB2, LAS1L,* and *ZNF207*) show decreased expression in cells
743 treated to create iron deficiency (**Figure 11B**; t-test p-values ≤ 0.05). In the mouse B cells,
744 DFO-treatment to cause iron deficiency increases the abundance of transcripts of four 3'

745 IRE-containing genes, including *Tfrc,* in accordance with the classic paradigm (**Figure 11C**;
746 *t*-test *p*-values ≤ 0.05). In contrast, transcripts of four such 3' IRE genes show decreased
747 abundance (*Bub3, Cxcl16*, *Gstm6*, *Zfp71-rs1*) (**Figure 11C**; *t*-test *p*-values ≤ 0.05).
748
749 Although these sample sizes are small, they clearly indicate that the presence of an IRE in
750 either the 5' or 3' UTR of a transcript cannot, by itself, be used to predict how the abundance
751 of the transcript will change in response to iron availability.
752

753 Discussion

754 In this study we identified comprehensive sets of genes predicted to contain IREs in
755 humans, mice, and zebrafish. The idea of analysing gene expression datasets using gene
756 sets is not novel, with many well-established techniques for gene set testing [35–39] that
757 offer the opportunity to analyze the behavior of large groups of genes to detect statistically
758 significant changes in state, even though these may be quite subtle for individual genes.
759 **Table 5** summarises representative enriched Hallmark gene sets (representing common
760 biological pathways/activities) across the human, mouse, and zebrafish datasets analysed in
761 this work. Surprisingly, we found that IRE gene sets are generally not well-represented in
762 existing gene sets. The existing gene set in which IRE genes are most highly represented is
763 one previously shown to be up-regulated in AD brains (the "Blalock Alzheimer's Disease Up"
764 gene set from MSigDB). This supports a unique importance of IRE genes in AD.
765 Consequently, the IRE gene sets identified here represent a novel resource for researchers
766 interested in exploring iron dyshomeostasis gene expression responses in detail in different
767 species, conditions, and treatments. In our results, we demonstrate that IRE gene sets are
768 sufficiently sensitive to reveal broad differences across different conditions (both in human
769 AD and animal models), while also offering the opportunity to study individual
770 genes/transcripts contributing to enrichment in more detail. IRE gene sets displayed

771 significant enrichment in postmortem brains from a human AD cohort, the 5XFAD mouse
772 model, and a zebrafish model of a fAD-like mutation. Taken together, our results
773 demonstrate, for the first time, the involvement of coordinated IRP/IRE-mediated gene
774 expression responses associated with AD.

775

776 Overall, IRE gene sets were sufficiently sensitive to reveal several interesting phenomena.
777 The relatively tightly controlled conditions in the mouse and zebrafish model datasets
778 revealed strongly age-dependent effects on the transcript abundances of these predicted
779 IRE-containing genes. Notably, 3' IRE gene expression changes were amongst the earliest
780 changes observable in the zebrafish heterozygous single fAD-like mutation model alongside
781 changes in energy metabolism. These 3' IRE gene expression changes preceded other
782 signals of pathological change in the transcriptome (such as altered expression of
783 inflammatory response pathways) commonly associated with AD. Our findings raise the
784 question of how early IRE gene expression changes, and iron dyshomeostasis more
785 broadly, may begin in fAD brains. In a recent analysis of transcriptome data from pools of 7-
786 day-old zebrafish larvae heterozygous for the fAD-like mutation *psen1*$^{Q96\_K97del}$ (compared to
787 wild type half-siblings), we observed over-representation of the GO term "iron ion transport"
788 without significant changes in 3' IRE or 5' IRE transcript abundance [47]. This supports that
789 the 3' IRE transcript abundance signal in 6-month-old brains of this mutant line reflects
790 changes in ferrous iron abundance that develop with age in response to an underlying defect
791 in cellular iron regulation.

792

793 In relation to our findings associated with the 5XFAD mouse model, we acknowledge that
794 low sample number is a limitation in this analysis and constrains interpretation of our results.
795 Unfortunately, to our knowledge, no other publicly-available 5XFAD mouse dataset was
796 available containing (1) both young and aged adults in the same batch (where young mice

797 are defined as ~2-4 months old as this is when the earliest signs of pathology are observed
798 for this model [77,78]), (2) 5XFAD and WT genotype mice in the same batch, (3) information
799 about biological sex for each sample (so that sex-specific effects could be accounted for in
800 the statistical model for differential gene expression), and (4) separation of groups by
801 biological factors (e.g. age, genotype) rather than by unidentifiable batch effects when
802 inspected through Principal Component Analysis. Our original rationale for analyzing the
803 5XFAD mouse model came from results from an independent concordance analysis of
804 various mouse models and human AD [67]. In that study, it was found that 5XFAD mouse
805 brain transcriptomes showed greatest (although still fairly poor) concordance to human AD
806 compared to other mouse models including the 3xTg model. However, we also acknowledge
807 the limitations associated with only analyzing the 5XFAD mouse model here. It would be
808 useful to analyze other mouse models of AD in the future to compare the involvement of IRE
809 genes across different models. Likewise, we anticipate that further analysis of fAD-like
810 mutation models in general will help to extend our knowledge of the early molecular changes
811 contributing to the development of AD.
812

813 Notably, both classic IRE genes (e.g. *TFRC*, *FTL*, and *FTH1*) and the IRE gene sets were
814 sufficiently sensitive to distinguish not only between iron overload and deficiency in a
815 cultured cell line dataset, but also between AD and other neuropathologies (i.e. PA and
816 PSP). This suggests that the dysregulation of IRE-containing genes, (and iron homeostasis
817 in general), in AD may differ from other neuropathological conditions. The differences in both
818 classic IRE genes and IRE gene set expression between AD and PA are of particular
819 interest. Considering that presence of amyloid pathology is required to diagnose individuals
820 with AD [79,80] and contributes to diagnosis of preclinical AD [1,2], it will be interesting to
821 explore further the biological relevance and impact of these IRE gene expression differences

822 using other techniques to gauge whether PA is a prodrome of, or a separate pathological
823 condition to, AD.

824

825 One of our original intentions for IRE gene set analysis of the human AD and animal model
826 datasets was to gain insight into whether these brains were in a state of iron deficiency or
827 overload using gene expression-level evidence. Unfortunately, it became clear that
828 interpretation of IRE gene expression changes was more complex than expected, both
829 because of the use of bulk brain tissue samples in the human and animal model datasets
830 (which obscures differences between cell type and tissue types) and because the regulation
831 of both 3' and 5' IRE-containing genes we observed did not follow the expected paradigm.
832 To address the former issue, future work should expand the use of IRE gene sets to single-
833 cell RNA-seq (scRNA-seq) datasets so that the IRE gene expression patterns of different
834 cell types can be deconvoluted. Recently, a large-scale scRNA-seq study of AD brains
835 indicated highly cell-type specific changes related to inflammation during early stages of
836 disease progression within the prefrontal cortex [81], highlighting the importance of studying
837 iron homeostasis responses at the single-cell level in the context of AD. The failure of 3' IRE
838 genes to follow the expected expression paradigm is a limitation we did not anticipate when
839 we began this work. In particular, our observations indicating that (1) the stability of some 3'
840 IRE transcripts (rather than their "expression"/"levels" as addressed in preceding analyses)
841 is significantly different under different conditions in zebrafish, and that (2) validated 3' IRE
842 genes show both increased and decreased expression under iron deficiency in two cell line
843 datasets (suggesting they are not stabilized as predicted by the IRE paradigm), challenge
844 the simplistic generalisation that 3' IREs stabilise transcripts under ferrous iron deficiency.

845

846 Previous research found that, under iron deficiency, the transcripts of 3' IRE-containing
847 genes such as *TFRC* are stabilized causing increased expression [20] to increase ferrous
848 iron availability. However, our findings reveal that this principle (the "classic IRE paradigm")
849 does not necessarily apply to other 3' IRE-containing genes, including those with canonical
850 IREs (e.g. *SLC11A2* in the Caco-2 cultured cell line), and genes with IRE-like elements
851 deviating from the canonical IRE sequence. Interestingly, Tybl et al. [82] found that the 3'
852 IRE in *SLC11A2* (*DMT1)* appears to influence transcript abundance in a development-
853 specific manner in mice, by promoting transcript expression during postnatal growth and
854 suppressing expression in adulthood. Together with our findings, this demonstrates that the
855 regulation of IRE genes is more complex than commonly appreciated.

856

857 Aside from 3' IRE genes, 5' IRE genes also showed unexpected behavior in the datasets we
858 analyzed. In general, the classic IRE paradigm holds that binding of IRPs to 5' IREs blocks
859 translation but is not expected to change transcript abundance. We initially sought to
860 investigate 5' IRE transcripts as a negative control group to which changes in the stability of
861 the 3' IRE transcripts could be compared. Unexpectedly, our results indicated that 5' IRE
862 gene sets are significantly enriched under various conditions. Upon reflection, this is not
863 unreasonable, considering that binding of any protein to a transcript must, inevitably,
864 changes the transcript's stability to some degree. However, the direction of that stability
865 change is difficult to predict as there is no generalizable cellular "rule" for the outcome of
866 protein binding. This is consistent with the reality that each gene represents a largely
867 independently evolving system constrained only by cells' requirements for its functionality.
868 For example, inactivation of RNA-binding proteins has been demonstrated previously to
869 cause both up- and down-regulation of different genes simultaneously [83]. In addition, the 5'
870 canonical IRE and surrounding acute box sequences of the H-ferritin (*FTH*) transcript have
871 been shown to stabilize it, which might appear contrary to the expected function of a 5' IRE

872 (of decreasing protein translation under iron deficient conditions) [84]. H-ferritin is one of the
873 better-characterised IRE genes, and it is possible that other, yet-to-be characterised
874 regulatory mechanisms involving 5' IREs could contribute to stability changes in other
875 transcripts. Aside from 3' IREs, other 3' UTR regulatory motifs are known that bind to
876 proteins to regulate multiple biological activities (e.g. AU-rich 3' UTR motifs as described in
877 [85]). Therefore, mechanistically, it is reasonable to assume that IRE-based regulation of
878 transcript stability may be co-opted for other biological activities beyond iron homeostasis
879 (and see below). Alternative splicing of IRE genes may be one mechanism for this, where a
880 non-IRE-containing isoform is expressed in certain tissues or conditions, as is the case for
881 mammalian *DMT-1* (*SLC11A2*) [22]. Differential transcript isoform inclusion of IREs is
882 beyond the scope of our current work, but will be important to explore in future. We expect
883 our findings to provide a starting point for further research involving IRE genes and their
884 complex regulation beyond the classic paradigm.

885

886 It is important to acknowledge that some of the complexity involved with IRE gene
887 expression, and the large number of genes identified with putative IREs, may reflect the
888 involvement of IREs in functions other than iron homeostasis. These functions include
889 regulation of oxidative stress [86–88]), oxidative phosphorylation [89], and cell cycle control
890 [90,91]. In 2019, it was proposed that the tricarboxylic acid cycle (TCA) may have evolved
891 around the reactivity of iron [92], and TCA intermediates themselves also perform important
892 cellular regulatory functions [93]. In addition, our analyses identified several transcription
893 factor motifs potentially associated with other biological pathways as over-represented in
894 IRE gene sets. Similarly, although beyond the scope of our current work, a regulatory role
895 for micro RNAs (miRNAs) in binding IRE structures is emerging [94–97]. This may be
896 particularly relevant to AD, as a miRNA has been described that can bind the 5' IRE-like
897 structure in AβPP transcripts to up-regulate AβPP its expression [98]. Co-regulation of IRE-

898 containing genes within other biological pathways may explain some of the IRE gene set
899 enrichment observed in the datasets we analyzed, and some of the unexpected expression
900 changes (e.g. decreased expression of 3' IRE genes in conditions where stabilization of their
901 transcripts might be expected, and increased expression of 5' IRE genes even when no
902 change in stability would be expected). Hence, the changes in expression of predicted IRE
903 genes we observed must be interpreted carefully as reflecting not only iron dyshomeostasis
904 responses, but also other co-occurring changes in linked processes. This was particularly
905 evident in our analysis of the Caco-2 cultured cell line dataset where even straightforward
906 iron overload and iron deficiency treatments resulted in significant enrichment of gene sets
907 related to reactive oxygen species and hypoxia pathways. We note that the relationship
908 between iron homeostasis and hypoxia would be interesting to explore in future work,
909 particularly their similarities at the the level of gene expression. Previous research has
910 demonstrated that desferrioxamine (that induces iron deficiency) can also induce HIF-1
911 activity, which activates erythropoietin (important for red blood cell formation to carry
912 oxygen) [99]. This supports that iron deficiency and hypoxia involve similar molecular
913 changes in cells (i.e. the view that iron deficiency generates a form of "pseudohypoxia"). We
914 expect that future work deconvoluting the expression of IRE genes at the cellular level will
915 be required to determine to what extent such expression changes are influenced by cell /
916 tissue type compared to other linked biological activities.

917

918 Most previous work has assumed that accumulation of iron in the brain with age (a
919 phenomenon observed broadly across animal phyla [100,101]) is indicative of cellular iron
920 overload driving oxidative stress [102]. Indeed, iron accumulation and oxidative stress
921 apparently contribute to the insolubility of amyloid plaques and neurofibrillary tangles in AD
922 brains [103]. However, a recent publication by Yambire et al. [104] showed disturbed
923 lysosomal function leading to increased lysosomal pH and causing a deficiency of functional

924 ferrous iron (Fe^{2+}) while non-functional ferric iron (Fe^{3+}) accumulated in lysosomes. The
925 deleterious effects of this on the brains of mice (defective mitochondrial biogenesis and
926 function and stimulation of inflammatory responses) could be alleviated by dietary iron
927 supplementation. The observations of Nixon and colleagues that acidification of the
928 endolysomal pathway is affected both by fAD mutations in *PSEN1* [71] and excessive
929 dosage of the amyloid beta A4 precursor protein (*A6PP*) gene [105], together with our
930 observations from our fAD-mutation-like *psen1*$^{Q96\_K97del}$/+ mutant zebrafish in both this study
931 and in ref. [44] (where gene expression changes associated with lysosome and mitochondria
932 for this model were first described), support the possibility that AD brains may suffer a
933 ferrous iron deficiency in a background of ferric iron overload. This is discussed further
934 below.

935

936 Although the evidence from human AD brains is more tentative at this stage, our findings of
937 increased expression of *TFRC* expression and other well-characterized 3' IRE genes in the
938 temporal cortex tissue of human AD brains is consistent with the existence of a cellular
939 ferrous iron deficiency, although more research is required to confirm this at the single-cell
940 level. Intriguingly, the greatest genetic risk factor for late onset AD, the ε4 allele of the gene
941 *APOE*, appears to increase lysosomal pH [106] but ε4's increased risk of AD is alleviated in
942 individuals who possess the *HFE* 282Y allele that predisposes to the iron overload disease
943 hemochromatosis [107]. Given that many other risk loci for sporadic late onset AD also
944 affect endolysosomal pathway function (reviewed in [108]), it is reasonable to suggest that
945 disturbed iron homeostasis may afflict brains with this disease.

946 Our work provides evidence consistent with Yambire et al. [104] and Lee et al. [71] to
947 suggest that fAD mutations may drive a ferrous iron deficiency that contributes to AD
948 pathology. However, iron export is a complex process, and it remains unclear whether iron

949 deficiency or overload is more relevant to the pathology seen in AD. Much of the existing
950 research has focused on the role of AβPP and its contributions to elevating iron levels within
951 cells and to increasing iron export to the cerebral interstitium in AD. For example,
952 Venkataramani et al. [109] demonstrated that decreased AβPP levels resulted in
953 stabilisation of ferroportin, resulting in increased iron export from neuroblastoma cells in two
954 mouse models. Later work by McCarthy et al. [110] proposed a mechanism for this
955 phenomenon, involving a motif in sAβPPa (cleavage product of AβPP) binding to ferroportin
956 and increasing iron export from cells through increasing the number of ferroportin molecules
957 in the plasma membrane but not their activity. A complicating factor is that ferroportin-
958 mediated iron export from cells requires both (1) stabilisation of ferroportin in the membrane,
959 and (2) ferroxidase activity (provided by either haephaestin or ceruloplasmin) [110]. The role
960 of haephestin as a ferroxidase in this process of iron export from cells has sometimes been
961 overlooked, despite that it is well-established as complexing with, and stabilizing, ferroportin
962 to export iron [111–113]). For example, recently Tsatsanis et al. [114] described the effects
963 on intracellular iron and iron export of changes in amyloidogenic cleavage (via β-secretase)
964 and non-amyloidogenic cleavage (via γ-secretase) of AβPP. It was observed that
965 amyloidogenic cleavage of AβPP led to destabilisation of ferroportin on the cell surface and
966 increased iron retention in cells. In addition, inhibiting β-secretase mitigated this effect,
967 increasing ferroportin stabilization and decreasing iron retention in cells. However, an
968 alternative explanation not considered by those authors is that haephestin may also be
969 cleaved by β-secretase (see Figure 7 in [115]), so it is possible that reduced β-secretase
970 activity could increase the haephestin available and also contribute to the effects on iron
971 export observed. Overall, the situation around iron export from cells is a complex area of
972 ongoing research involving many cellular factors, and it will be increasingly important in
973 future work to understand these mechanisms more completely in the context of AD.
974

975 We have previously noted that all the fAD mutations of AβPP (including gene duplications)
976 are expected to have the singular quality of increasing the abundance of the amyloidogenic
977 β-CTF / C99 fragment of AβPP (see Figure 2 in Lumsden et al. [14]). Recently Jiang et al.
978 [105] showed that increased β-CTF reduces lysosomal acidification. According to the results
979 of Yambire et al. cited above, this would be expected to reduce cellular ferrous iron levels
980 (while causing accumulation of ferric iron in lysosomes). This idea is consistent with the
981 existence of an IRE in the 5' UTR of AβPP transcripts that inhibits AβPP translation (and
982 hence β-CTF production) when cytosolic ferrous iron levels are low [116]. Thus, AβPP may
983 influence cellular ferrous iron levels by modulating lysosomal pH to regulate importation of
984 ferrous iron into the cytosol, rather than by modulating ferroportin stability to regulate iron
985 export from the cytosol.
986

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

999

1000 # Tables

1001

1002 **Table 1. Summary of gene set testing approaches used in our analyses.**

1003

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 [35,36]
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 [37]
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 [38,39]

1004

1005

1006

1007 **Table 2. Samples analyzed from Mayo Clinic RNA-seq study.**

1008

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

1009

1010

1011

1012 **Table 3. Differential gene expression and IRE gene set enrichment results from Mayo**

1013 **Clinic RNA-seq study.** Asterisks (*) indicate significant enrichment (Bonferroni-adjusted

1014 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*

1015

1016 **Table 4. Enrichment of Iron Responsive Element (IRE) gene sets in fAD-like zebrafish**
1017 **dataset.** Raw *p*-values from *fry*, *camera* and *fgsea* were combined with Wilkinson's method,
1018 with combined *p*-values then Bonferroni-adjusted for multiple testing. The same process was
1019 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

1020
1021

1022 **Table 5. Summary of enriched gene sets from the MSigDB Hallmark Gene Set**
1023 **Collection in Datasets Analyzed**. Gene sets shown here are representative enriched gene
1024 sets from the MSigDB Hallmark Collection (Bonferroni-adjusted Wilkinson's *p*-value from *fry*,
1025 *camera*, and *fgsea* < 0.05). For full lists of enriched gene sets, see relevant figures or
1026 Supplemental Tables). For gene sets which had a significant direction, an up (↑) or down (↓)
1027 arrow is used to indicate whether most of the genes in the gene set are increased or
1028 decreased in expression. No arrow indicates gene sets involving both increased and
1029 decreased expression from different subsets of genes.

Dataset	Comparison	Hallmark Gene Sets Enriched*
Caco-2 cell line (see **Fig. 5**)	Iron Overload	↓ Interferon Gamma Response ↑ Reactive Oxygen Species Pathway
	Iron Deficiency	↑ Interferon Gamma Response ↓ Reactive Oxygen Species Pathway
Human Mayo-Clinic RNA-seq Study (see Supplementary Table 7)	AD vs. control (Temporal Cortex)	↓ Oxidative Phosphorylation ↑ TNFA signalling via NFKB ↑ Interferon Alpha Response ↑ Inflammatory Response ↑ Interferon Gamma Response
	AD vs. control (Cerebellum)	↑ G2M Checkpoint ↑ Mitotic Spindle ↑ PI3K AKT MTOR Signalling ↑ Hypoxia
	PA vs. control (Temporal Cortex)	Estrogen Response Heme Metabolism Complement Adipogenesis Inflammation
	PA vs. control (Cerebellum)	Estrogen Response Glycolysis Hypoxia Unfolded Protein Response
	PSP vs. control (Temporal Cortex)	↓ Interferon Alpha Response ↑ Reactive Oxygen Species Signalling Pathway ↑ Glycolysis ↑ Hypoxia ↑ Estrogen Response
	PSP vs. control (Cerebellum)	↑ Myc Targets ↑ G2M Checkpoint ↑ Mitotic Spindle ↑ Heme Metabolism ↑ Hypoxia ↓ Fatty Acid Metabolism
Mouse 5XFAD model (see **Fig. 7**)	5XFAD vs. WT (3 months old)	↑ Interferon Gamma Response ↑ Interferon Alpha Response ↑ Cholesterol Homeostasis
	5XFAD vs. WT (6 months old)	Above also enriched + many other gene sets, see Fig. 7 for more details)
	5XFAD vs. WT (12 months old)	Above also enriched + many other gene sets, see Fig. 7 for more details)
Zebrafish *psen1*Q96K97del/+ model (see **Fig. 9**)	Q96K97del/+ vs. WT (6 months)	↓ Oxidative Phosphorylation ↓ Fatty Acid Metabolism MTORC1 signalling PI3K AKT MTOR signalling
	Q96K97del/+ vs. WT (24 months)	Above also enriched + many other gene sets, see Fig. 9 for more details)

1030

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 (SAGC) post-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.

Datasets and Data Articles

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

Authors' contributions

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

1069 # References

1070

1071 [1] Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB,
1072 Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine
1073 T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM,
1074 Sperling R, Elliott C, Masliah E, Ryan L, Silverberg N (2018) NIA-AA Research
1075 Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's*
1076 *Dement* **14**, 535–562.

1077 [2] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo
1078 T, Jack CR, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E,
1079 Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster M V.,
1080 Phelps CH (2011) Toward defining the preclinical stages of Alzheimer's
1081 disease: Recommendations from the National Institute on Aging-Alzheimer's
1082 Association workgroups on diagnostic guidelines for Alzheimer's disease.
1083 *Alzheimer's Dement* **7**, 280–292.

1084 [3] Bartzokis G, Sultzer D, Mintz J, Holt LE, Marx P, Kelly Phelan C, Marder SR
1085 (1994) In vivo evaluation of brain iron in Alzheimer's disease and normal
1086 subjects using MRI. *Biol Psychiatry* **35**, 480–487.

1087 [4] Bartzokis G, Tishler TA (2000) MRI evaluation of basal ganglia ferritin iron and
1088 neurotoxicity in Alzheimer's and Huntingon's disease. *Cell Mol Biol* **46**, 821–
1089 833.

1090 [5] Ding B, Chen KM, Ling HW, Sun F, Li X, Wan T, Chai WM, Zhang H, Zhan Y,
1091 Guan YJ (2009) Correlation of iron in the hippocampus with MMSE in patients
1092 with Alzheimer's disease. *J Magn Reson Imaging* **29**, 793–798.

1093 [6] Langkammer C, Ropele S, Pirpamer L, Fazekas F, Schmidt R (2014) MRI for
1094 iron mapping in Alzheimer's disease. *Neurodegener Dis* **13**, 189–191.
1095 [7] Van Duijn S, Bulk M, Van Duinen SG, Nabuurs RJA, Van Buchem MA, Van
1096 Der Weerd L, Natté R (2017) Cortical Iron Reflects Severity of Alzheimer's
1097 Disease. *J Alzheimer's Dis* **60**, 1533–1545.
1098 [8] Mills E, Dong XP, Wang F, Xu H (2010) Mechanisms of brain iron transport:
1099 Insight into neurodegeneration and CNS disorders. *Future Med Chem* **2**, 51–
1100 64.
1101 [9] Yamamoto A, Shin R-W, Hasegawa K, Naiki H, Sato H, Yoshimasu F,
1102 Kitamoto T (2004) Iron (III) induces aggregation of hyperphosphorylated τ and
1103 its reduction to iron (II) reverses the aggregation: implications in the formation
1104 of neurofibrillary tangles of Alzheimer's disease. *J Neurochem* **82**, 1137–1147.
1105 [10] Telling ND, Everett J, Collingwood JF, Dobson J, van der Laan G, Gallagher
1106 JJ, Wang J, Hitchcock AP (2017) Iron Biochemistry is Correlated with Amyloid
1107 Plaque Morphology in an Established Mouse Model of Alzheimer's Disease.
1108 *Cell Chem Biol* **24**, 1205-1215.e3.
1109 [11] Leskovjan AC, Kretlow A, Lanzirotti A, Barrea R, Vogt S, Miller LM (2011)
1110 Increased brain iron coincides with early plaque formation in a mouse model of
1111 Alzheimer's disease. *Neuroimage* **55**, 32–38.
1112 [12] Gong NJ, Dibb R, Bulk M, van der Weerd L, Liu C (2019) Imaging beta
1113 amyloid aggregation and iron accumulation in Alzheimer's disease using
1114 quantitative susceptibility mapping MRI. *Neuroimage* **191**, 176–185.
1115 [13] El Tannir El Tayara N, Delatour B, Le Cudennec C, Guégan M, Volk A,
1116 Dhenain M (2006) Age-related evolution of amyloid burden, iron load, and MR

1117 relaxation times in a transgenic mouse model of Alzheimer's disease.
1118 *Neurobiol Dis* **22**, 199–208.
1119 [14] Lumsden AL, Rogers JT, Majd S, Newman M, Sutherland GT, Verdile G,
1120 Lardelli M (2018) Dysregulation of neuronal iron homeostasis as an alternative
1121 unifying effect of mutations causing familial alzheimer's disease. *Front*
1122 *Neurosci* **12**, 533.
1123 [15] Lane DJR, Merlot AM, Huang MLH, Bae DH, Jansson PJ, Sahni S, Kalinowski
1124 DS, Richardson DR (2015) Cellular iron uptake, trafficking and metabolism:
1125 Key molecules and mechanisms and their roles in disease. *Biochim Biophys*
1126 *Acta - Mol Cell Res* **1853**, 1130–1144.
1127 [16] Gray NK, Hentze MW (1994) Iron regulatory protein prevents binding of the
1128 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *EMBO J*
1129 **13**, 3882–3891.
1130 [17] Hentze MW, Muckenthaler MU, Andrews NC (2004) Balancing acts: Molecular
1131 control of mammalian iron metabolism. *Cell* **117**, 285–297.
1132 [18] Muckenthaler MU, Galy B, Hentze MW (2008) Systemic Iron Homeostasis and
1133 the Iron-Responsive Element/Iron-Regulatory Protein (IRE/IRP) Regulatory
1134 Network. *Annu Rev Nutr* **28**, 197–213.
1135 [19] Basilion JP, Rouault TA, Massinople CM, Klausner RD, Burgesst WH (1994)
1136 The iron-responsive element-binding protein: Localization of the RNA-binding
1137 site to the aconitase active-site cleft (UV cross-linking/base hydrolysis).
1138 *Biochemistry* **91**, 574–578.
1139 [20] Pantopoulos K (2004) Iron Metabolism and the IRE/IRP Regulatory System:
1140 An Update. *Ann N Y Acad Sci* **1012**, 1–13.

1141 [21] Piccinelli P, Samuelsson T (2007) Evolution of the iron-responsive element.
1142 *RNA* **13**, 952–966.
1143 [22] Gunshin H, Allerson CR, Polycarpou-Schwarz M, Rofts A, Rogers JT, Kishi F,
1144 Hentze MW, Rouault TA, Andrews NC, Hediger MA (2001) Iron-dependent
1145 regulation of the divalent metal ion transporter. *FEBS Lett* **509**, 309–316.
1146 [23] Cmejla R, Petrak J, Cmejlova J (2006) A novel iron responsive element in the
1147 3′UTR of human MRCKα. *Biochem Biophys Res Commun* **341**, 158–166.
1148 [24] Sanchez M, Galy B, Schwanhaeusser B, Blake JJ, Bähr-Ivacevic T, Benes V,
1149 Selbach M, Muckenthaler MU, Hentze MW (2011) Iron regulatory protein-1
1150 and -2: transcriptome-wide definition of binding mRNAs and shaping of the
1151 cellular proteome by iron regulatory proteins. *e-Blood* **118**, e168–e179.
1152 [25] Campillos M, Cases I, Hentze MW, Sanchez M (2010) SIREs: searching for
1153 iron-responsive elements. *Nucleic Acids Res* **38**, W360-367.
1154 [26] Tanzi RE, Hyman BT (1991) Alzheimer's mutation. *Nat 1991 3506319* **350**,
1155 564–564.
1156 [27] Conway JR, Lex A, Gehlenborg N UpSetR: an R package for the visualization
1157 of intersecting sets and their properties. *Bioinformatics*.
1158 [28] Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G,
1159 Kasprzyk A (2009) BioMart - Biological queries made easy. *BMC Genomics*
1160 **10**, 22.
1161 [29] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P,
1162 Mesirov JP, Bateman A (2011) Databases and ontologies Molecular
1163 signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740.
1164 [30] Bastian M, Heymann S, Jacomy M (2009) Gephi: An Open Source Software

1165 for Exploring and Manipulating Networks Visualization and Exploration of
1166 Large Graphs. *Third Int AAAI Conf Weblogs Soc Media*.
1167 [31] Hu Y (2011) *Algorithms for Visualizing Large Networks*.
1168 [32] Benner C HOMER (Hypergeometric Optimization of Motif EnRichment).
1169 [33] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre
1170 C, Singh H, Glass CK (2010) Simple Combinations of Lineage-Determining
1171 Transcription Factors Prime cis-Regulatory Elements Required for
1172 Macrophage and B Cell Identities. *Mol Cell* **38**, 576–589.
1173 [34] Team RC (2018) R: A language and environment for statistical computing.
1174 [35] Giner G, Smyth GK (2016) FRY: a fast approximation to ROAST gene set test
1175 with mean aggregated set statistics. *F1000Research* **5**,.
1176 [36] Wu D, Lim E, Vaillant F, Asselin-Labat M-L, Visvader JE, Smyth GK (2010)
1177 ROAST: rotation gene set tests for complex microarray experiments.
1178 *Bioinformatics* **26**, 2176–2182.
1179 [37] Wu D, Smyth GK (2012) Camera: a competitive gene set test accounting for
1180 inter-gene correlation. **40**, e133.
1181 [38] Sergushichev AA (2016) An algorithm for fast preranked gene set enrichment
1182 analysis using cumulative statistic calculation. *bioRxiv* 060012.
1183 [39] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA,
1184 Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set
1185 enrichment analysis: a knowledge-based approach for interpreting genome-
1186 wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550.
1187 [40] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) Limma
1188 powers differential expression analyses for RNA-sequencing and microarray

1189 studies. *Nucleic Acids Res* **43**, e47.

1190 [41] Law CW, Chen Y, Shi W, Smyth GK (2014) Voom: Precision weights unlock
1191 linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**,.

1192 [42] Wilkinson B (1951) A statistical consideration in psychological research.
1193 *Psychol Bull* **48**, 156–158.

1194 [43] Chicault C, Toutain B, Monnier A, Aubry M, Fergelot P, Le Treut A, Galibert
1195 MD, Mosser J (2006) Iron-related transcriptomic variations in CaCo-2 cells, an
1196 in vitro model of intestinal absorptive cells. *Physiol Genomics* **26**, 55–67.

1197 [44] Newman M, Hin N, Pederson S, Lardelli M (2019) Brain transcriptome analysis
1198 of a familial Alzheimer's disease-like mutation in the zebrafish presenilin 1
1199 gene implies effects on energy production. *Mol Brain* **12**, 43.

1200 [45] Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic
1201 RNA-seq quantification. *Nat Biotechnol* **34**, 525–527.

1202 [46] Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: A Bioconductor
1203 package for differential expression analysis of digital gene expression data.
1204 *Bioinformatics* **26**, 139–140.

1205 [47] Dong Y, Newman M, Pederson S, Hin N, Lardelli M (2020) Transcriptome
1206 analyses of 7-day-old zebrafish larvae possessing a familial Alzheimer's
1207 disease-like mutation in psen1 indicate effects on oxidative phosphorylation,
1208 mcm functions, and iron homeostasis. *bioRxiv* 2020.05.03.075424.

1209 [48] Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF,
1210 Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC,
1211 Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL,
1212 Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong HW, Dougherty JG, Duncan

1213 BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR,
1214 Fliss TP, Frensley C, Gates SN, Glattfelder KJ, Halverson KR, Hart MR,
1215 Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawal R, Kidney
1216 JM, Knapik RH, Kuan CL, Lake JH, Laramee AR, Larsen KD, Lau C, Lemon
1217 TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud
1218 MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE,
1219 Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA,
1220 Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova N V., Sivisay T,
1221 Slaughterbeck CR, Smith SC, Smith KA, Smith BI, Sodt AJ, Stewart NN,
1222 Stumpf KR, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson
1223 CL, Varnam LR, Visel A, Whitlock RM, Wohnoutka PE, Wolkey CK, Wong VY,
1224 Wood M, Yaylaoglu MB, Young RC, Youngstrom BL, Yuan XF, Zhang B,
1225 Zwingman TA, Jones AR (2007) Genome-wide atlas of gene expression in the
1226 adult mouse brain. *Nature* **445**, 168–176.
1227 [49] Bonham LW, Sirkis DW, Yokoyama JS (2019) The transcriptional landscape of
1228 microglial genes in aging and neurodegenerative disease. *Front Immunol* **10**,
1229 1170.
1230 [50] Sanchez M, Galy B, Dandekar T, Bengert P, Vainshtein Y, Stolte J,
1231 Muckenthaler MU, Hentze MW (2006) Iron regulation and the cell cycle:
1232 Identification of an iron-responsive element in the 3′-untranslated region of
1233 human cell division cycle 14A mRNA by a refined microarray-based screening
1234 strategy. *J Biol Chem* **281**, 22865–22874.
1235 [51] Allaire J, Christopher G, Russell K, Yetman C (2017) networkD3: D3
1236 JavaScript Network Graphs from R.

1237 [52] Goforth JB, Anderson SA, Nizzi CP, Eisenstein RS (2010) Multiple
1238 determinants within iron-responsive elements dictate iron regulatory protein
1239 binding and regulatory hierarchy. *RNA* **16**, 154–169.
1240 [53] Henderson BR, Menotti E, Kühn LC (1996) Iron regulatory proteins 1 and 2
1241 bind distinct sets of RNA target sequences. *J Biol Chem* **271**, 4900–4908.
1242 [54] Henderson BR, Menotti E, Bonnard C, Kuhn LC (1994) Optimal sequence and
1243 structure of iron-responsive elements. Selection of RNA stem-loops with high
1244 affinity for iron regulatory factor. *J Biol Chem* **269**, 17481–17489.
1245 [55] Butt J, Kim HY, Basilion JP, Cohen S, Iwai K, Philpott CC, Altschul S, Klausner
1246 RD, Rouault TA (1996) Differences in the RNA binding sites of iron regulatory
1247 proteins and potential target diversity. *Proc Natl Acad Sci U S A* **93**, 4345–
1248 4349.
1249 [56] Surinya KH, Cox TC, May BK (1998) Identification and characterization of a
1250 conserved erythroid-specific enhancer located in intron 8 of the human 5-
1251 aminolevulinate synthase 2 gene. *J Biol Chem* **273**, 16798–16809.
1252 [57] Theil EC (1990) Regulation of ferritin and transferrin receptor mRNAs. *J Biol*
1253 *Chem* **265**, 4771–4774.
1254 [58] Cmejla R, Ptackova P, Petrak J, Savvulidi F, Cerny J, Sebesta O, Vyoral D
1255 (2010) Human MRCKα is regulated by cellular iron levels and interferes with
1256 transferrin iron uptake. *Biochem Biophys Res Commun* **395**, 163–167.
1257 [59] Lymboussaki A, Pignatti E, Montosi G, Garuti C, Haile DJ, Pietrangelo A
1258 (2003) The role of the iron responsive element in the control of
1259 ferroportin1/IREG1/MTP1 gene expression. *J Hepatol* **39**, 710–715.
1260 [60] Aziz N, Munro HN (1987) Iron regulates ferritin mRNA translation through a

1261 segment of its 5' untranslated region. *Proc Natl Acad Sci U S A* **84**, 8478–
1262 8482.
1263 [61] Alhamdoosh M, Law CW, Tian L, Sheridan JM, Ng M, Ritchie ME (2017) Easy
1264 and efficient ensemble gene set testing with EGSEA. *F1000Research* **6**, 2010.
1265 [62] Allen M, Carrasquillo MM, Funk C, Heavner BD, Zou F, Younkin CS, Burgess
1266 JD, Chai HS, Crook J, Eddy JA, Li H, Logsdon B, Peters MA, Dang KK, Wang
1267 X, Serie D, Wang C, Nguyen T, Lincoln S, Malphrus K, Bisceglio G, Li M,
1268 Golde TE, Mangravite LM, Asmann Y, Price ND, Petersen RC, Graff-Radford
1269 NR, Dickson DW, Younkin SG, Ertekin-Taner N (2016) Human whole genome
1270 genotype and transcriptome data for Alzheimer's and other neurodegenerative
1271 diseases. *Sci Data* **3**, 1–10.
1272 [63] Ossenkoppele R, Van Berckel BN, Prins ND (2011) Amyloid imaging in
1273 prodromal Alzheimer's disease. *Alzheimer's Res Ther* **3**, 26.
1274 [64] Protas HD, Chen K, Langbaum JBS, Fleisher AS, Alexander GE, Lee W,
1275 Bandy D, De Leon MJ, Mosconi L, Buckley S, Truran-Sacrey D, Schuff N,
1276 Weiner MW, Caselli RJ, Reiman EM (2013) Posterior cingulate glucose
1277 metabolism, hippocampal glucose metabolism, and hippocampal volume in
1278 cognitively normal, late-middle-aged persons at 3 levels of genetic risk for
1279 alzheimer disease. *JAMA Neurol* **70**, 320–325.
1280 [65] Caldwell CC, Yao J, Brinton RD (2015) Targeting the Prodromal Stage of
1281 Alzheimer's Disease: Bioenergetic and Mitochondrial Opportunities.
1282 *Neurotherapeutics* **12**, 66–80.
1283 [66] Vermunt L, Sikkes SAM, van den Hout A, Handels R, Bos I, van der Flier WM,
1284 Kern S, Ousset PJ, Maruff P, Skoog I, Verhey FRJ, Freund-Levi Y, Tsolaki M,

1285 Wallin ÅK, Olde Rikkert M, Soininen H, Spiru L, Zetterberg H, Blennow K,
1286 Scheltens P, Muniz-Terrera G, Visser PJ, Vellas B, Reynish E, Ousset PJ,
1287 Andrieu S, Burns A, Pasquier F, Frisoni G, Salmon E, Michel JP, Zekry DS,
1288 Boada M, Dartigues JF, Olde-Rikkert MGM, Rigaud AS, Winblad B, Malick A,
1289 Sinclair A, Frölich L, Scheltens P, Ribera C, Touchon J, Robert P, Salva A,
1290 Waldemar G, Bullock R, Tsolaki M, Rodriguez G, Spiru L, Jones RW, Stiens
1291 G, Stoppe G, Eriksdotter Jönhagen M, Cherubini A, Lage PM, Gomez-Isla T,
1292 Camus V, Agüera-Morales E, Lopez F, Savy S, Cantet C, Coley N (2019)
1293 Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease
1294 in relation to age, sex, and APOE genotype. *Alzheimer's Dement* **15**, 888–898.
1295 [67] Hargis KE, Blalock EM (2017) Transcriptional signatures of brain aging and
1296 Alzheimer's disease: What are our rodent models telling us? *Behav Brain Res*
1297 **322**, 311–328.
1298 [68] Bundy JL, Vied C, Badger C, Nowakowski RS (2019) Sex-biased hippocampal
1299 pathology in the 5XFAD mouse model of Alzheimer's disease: A multi-omic
1300 analysis. *J Comp Neurol* **527**, 462–475.
1301 [69] Drummond E, Wisniewski T (2017) Alzheimer's disease: experimental models
1302 and reality. *Acta Neuropathol* **133**, 155–175.
1303 [70] Xia D, Watanabe H, Wu B, Lee SH, Li Y, Tsvetkov E, Bolshakov VY, Shen J,
1304 Kelleher RJ (2015) Presenilin-1 knockin mice reveal loss-of-function
1305 mechanism for familial alzheimer's disease. *Neuron* **85**, 967–981.
1306 [71] Lee JH, Yu WH, Kumar A, Lee S, Mohan PS, Peterhoff CM, Wolfe DM,
1307 Martinez-Vicente M, Massey AC, Sovak G, Uchiyama Y, Westaway D, Cuervo
1308 AM, Nixon RA (2010) Lysosomal proteolysis and autophagy require presenilin

1309 1 and are disrupted by Alzheimer-related PS1 mutations. *Cell* **141**, 1146–1158.
1310 [72] Lee JH, McBrayer MK, Wolfe DM, Haslett LJ, Kumar A, Sato Y, Lie PPY,
1311 Mohan P, Coffey EE, Kompella U, Mitchell CH, Lloyd-Evans E, Nixon RA
1312 (2015) Presenilin 1 Maintains Lysosomal Ca2+ Homeostasis via TRPML1 by
1313 Regulating vATPase-Mediated Lysosome Acidification. *Cell Rep* **12**, 1430–
1314 1444.
1315 [73] Yu Y, Richardson DR (2011) Cellular iron depletion stimulates the JNK and
1316 p38 MAPK signaling transduction pathways, dissociation of ASK1-thioredoxin,
1317 and activation of ASK1. *J Biol Chem* **286**, 15413–15427.
1318 [74] Riera H, Afonso V, Collin P, Lomri A (2015) A Central Role for JNK/AP-1
1319 Pathway in the Pro-Oxidant Effect of Pyrrolidine Dithiocarbamate through
1320 Superoxide Dismutase 1 Gene Repression and Reactive Oxygen Species
1321 Generation in Hematopoietic Human Cancer Cell Line U937. *PLoS One* **10**,
1322 e0127571.
1323 [75] Hin N, Newman M, Kaslin J, Douek AM, Lumsden A, Nik SHM, Dong Y, Zhou
1324 XF, Manucat-Tan NB, Ludington A, Adelson DL, Pederson S, Lardelli M (2020)
1325 Accelerated brain aging towards transcriptional inversion in a zebrafish model
1326 of the K115fs mutation of human PSEN2. *PLoS One* **15**, e0227258.
1327 [76] Jiang Y, Li C, Wu Q, An P, Huang L, Wang J, Chen C, Chen X, Zhang F, Ma
1328 L, Liu S, He H, Xie S, Sun Y, Liu H, Zhan Y, Tao Y, Liu Z, Sun X, Hu Y, Wang
1329 Q, Ye D, Zhang J, Zou S, Wang Y, Wei G, Liu Y, Shi Y, Eugene Chin Y, Hao
1330 Y, Wang F, Zhang X (2019) Iron-dependent histone 3 lysine 9 demethylation
1331 controls B cell proliferation and humoral immune responses. *Nat Commun*
1332 *2019 101* **10**, 1–15.

1333 [77] Oakley H, Cole SL, Logan S, Maus E, Shao P, Craft J, Guillozet-Bongaarts A,
1334 Ohno M, Disterhoft J, Van Eldik L, Berry R, Vassar R (2006) Intraneuronal
1335 beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic
1336 mice with five familial Alzheimer's disease mutations: potential factors in
1337 amyloid plaque formation. *J Neurosci* **26**, 10129–10140.
1338 [78] Landel V, Baranger K, Virard I, Loriod B, Khrestchatisky M, Rivera S, Benech
1339 P, Féron F (2014) Temporal gene profiling of the 5XFAD transgenic mouse
1340 model highlights the importance of microglial activation in Alzheimer's disease.
1341 *Mol Neurodegener* **9**,.
1342 [79] Deture MA, Dickson DW (2019) The neuropathological diagnosis of
1343 Alzheimer's disease. *Mol Neurodegener* **14**, 1–18.
1344 [80] Jack CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC,
1345 Thies B, Phelps CH (2011) Introduction to the recommendations from the
1346 National Institute on Aging-Alzheimer's Association workgroups on diagnostic
1347 guidelines for Alzheimer's disease. **7**, 257–262.
1348 [81] Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ,
1349 Menon M, He L, Abdurrob F, Jiang X, Martorell AJ, Ransohoff RM, Hafler BP,
1350 Bennett DA, Kellis M, Tsai LH (2019) Single-cell transcriptomic analysis of
1351 Alzheimer's disease. *Nature* **570**, 332–337.
1352 [82] Tybl E, Gunshin H, Gupta S, Barrientos T, Bonadonna M, Celma NF, Palais G,
1353 Karim Z, Sanchez M, Andrews NC, Galy B, Celma Nos F, Palais G, Karim Z,
1354 Sanchez M, Andrews NC, Galy B (2020) Control of Systemic Iron Homeostasis
1355 by the 3' Iron-Responsive Element of Divalent Metal Transporter 1 in Mice.
1356 *HemaSphere* **4**, e459.

1357 [83] Hasan A, Cotobal C, Duncan CDS, Mata J (2014) Systematic Analysis of the
1358 Role of RNA-Binding Proteins in the Regulation of RNA Stability. *PLoS Genet*
1359 **10**, e1004684.
1360 [84] Thomson AM, Cahill CM, Cho HH, Kassachau KD, Epis MR, Bridges KR,
1361 Leedman PJ, Rogers JT (2005) The acute box cis-element in human heavy
1362 ferritin mRNA 5′- untranslated region is a unique translation enhancer that
1363 binds poly(C)-binding proteins. *J Biol Chem* **280**, 30032–30045.
1364 [85] Ostareck-Lederer A, Ostareck DH, Standart N, Thiele BJ (1994) Translation of
1365 15-lipoxygenase mRNA is inhibited by a protein that binds to a repeated
1366 sequence in the 3' untranslated region. *EMBO J* **13**, 1476.
1367 [86] Martins EAL, Robalinho RL, Meneghini R (1995) Oxidative stress induces
1368 activation of a cytosolic protein responsible for control of iron uptake. *Arch*
1369 *Biochem Biophys* **316**, 128–134.
1370 [87] Pantopoulos K, Hentze MW (1995) Rapid responses to oxidative stress
1371 mediated by iron regulatory protein. *EMBO J* **14**, 2917–2924.
1372 [88] Yang L, Wang D, Wang X ting, Lu Y peng, Zhu L (2018) The roles of hypoxia-
1373 inducible Factor-1 and iron regulatory protein 1 in iron uptake induced by acute
1374 hypoxia. *Biochem Biophys Res Commun* **507**, 128–135.
1375 [89] Popovic Z, Templeton DM (2007) Inhibition of an iron-responsive element/iron
1376 regulatory protein-1 complex by ATP binding and hydrolysis. *FEBS J* **274**,
1377 3108–3119.
1378 [90] Yu Y, Kovacevic Z, Richardson DR (2007) Tuning cell cycle regulation with an
1379 iron key. *Cell Cycle* **6**, 1982–1994.
1380 [91] Fu D, Richardson DR (2007) Iron chelation and regulation of the cell cycle: 2

1381 Mechanisms of posttranscriptional regulation of the universal cyclin-dependent
1382 kinase inhibitor p21CIP1/WAF1 by iron depletion. *Blood* **110**, 752–761.
1383 [92] Muchowska KB, Varma SJ, Moran J (2019) Synthesis and breakdown of
1384 universal metabolic precursors promoted by iron. *Nature* **569**, 104–107.
1385 [93] Martínez-Reyes I, Chandel NS (2020) Mitochondrial TCA cycle metabolites
1386 control physiology and disease. *Nat Commun* **11**, 1–11.
1387 [94] Andolfo I, Falco L De, Asci R, Russo R, Colucci S, Gorrese M, Zollo M,
1388 Iolascon A (2010) Regulation of divalent metal transporter 1 (DMT1) non-IRE
1389 isoform by the microRNA Let-7d in erythroid cells. *Haematologica* **95**, 1244.
1390 [95] Davis M, Clarke S (2013) Influence of microRNA on the Maintenance of
1391 Human Iron Metabolism. *Nutr 2013, Vol 5, Pages 2611-2628* **5**, 2611–2628.
1392 [96] Freiesleben S, Hecker M, Zettl UK, Fuellen G, Taher L (2016) Analysis of
1393 microRNA and Gene Expression Profiles in Multiple Sclerosis: Integrating
1394 Interaction Data to Uncover Regulatory Mechanisms. *Sci Rep* **6**, 34512.
1395 [97] Sangokoya C, Doss JF, Chi J-T (2013) Iron-Responsive miR-485-3p
1396 Regulates Cellular Iron Homeostasis by Targeting Ferroportin. *PLOS Genet* **9**,
1397 e1003408.
1398 [98] Long JM, Maloney B, Rogers JT, Lahiri DK (2018) Novel upregulation of
1399 amyloid-β precursor protein (APP) by microRNA-346 via targeting of APP
1400 mRNA 5′-untranslated region: Implications in Alzheimer's disease. *Mol*
1401 *Psychiatry 2018 243* **24**, 345–363.
1402 [99] Guo M, Song L-P, Jiang Y, Liu W, Yu Y, Chen G-Q (2006) Hypoxia-mimetic
1403 agents desferrioxamine and cobalt chloride induce leukemic cell apoptosis
1404 through different hypoxia-inducible factor-1α independent mechanisms.

1405 *Apoptosis 2006 111* **11**, 67–77.

1406 [100] Zecca L, Youdim MBH, Riederer P, Connor JR, Crichton RR, Stankowski JN,
1407 Dawson VL, Dawson TM, Kawamoto M, Horibe T, Kohno M, Kawakami K,
1408 Duce JA, Tsatsanis A, Cater MA, James SA, Robb E, Wikhe K, Leong SL,
1409 Perez K, Johanssen T, Mark A, Cho H-H, Galatis D, Moir RD, Masters CL,
1410 McLean C, Tanzi RE, Cappai R, Barnham KJ, Ciccotosto D, Rogers JT, Bush
1411 AI, Greenough M a, Cho H-H, Galatis D, Moir RD, Masters CL, McLean C,
1412 Tanzi RE, Cappai R, Barnham KJ, Ciccotosto GD, Rogers JT, Bush AI, De
1413 Domenico I, McVey Ward D, Kaplan J (2011) Iron-export ferroxidase activity of
1414 β-amyloid precursor protein is inhibited by zinc in Alzheimer's disease. *Cell*
1415 **142**, 359.

1416 [101] Jacomin A-C, Geraki K, Brooks J, Tjendana-Tjhin V, Collingwood JF, Nezis IP
1417 (2019) Impact of Autophagy and Aging on Iron Load and Ferritin in Drosophila
1418 Brain. *Front Cell Dev Biol* **7**, 142.

1419 [102] Ayton S, Lei P, Bush AI (2013) Metallostasis in Alzheimer's disease. *Free*
1420 *Radic Biol Med* **62**, 76–89.

1421 [103] Smith MA, Harris PLR, Sayre LM, Perry G (1997) Iron accumulation in
1422 Alzheimer disease is a source of redox-generated free radicals. *Proc Natl*
1423 *Acad Sci U S A* **94**, 9866–9868.

1424 [104] Yambire KF, Rostosky C, Watanabe T, Pacheu-Grau D, Torres-Odio S,
1425 Sanchez-Guerrero A, Senderovich O, Meyron-Holtz EG, Milosevic I, Frahm J,
1426 Phillip West A, Raimundo N (2019) Impaired lysosomal acidification triggers
1427 iron deficiency and inflammation in vivo. *Elife* **8**,.

1428 [105] Jiang Y, Sato Y, Im E, Berg M, Bordi M, Darji S, Kumar A, Mohan PS,

1429 Bandyopadhyay U, Diaz A, Cuervo AM, Nixon RA (2019) Lysosomal
1430 dysfunction in down syndrome is app-dependent and mediated by APP-βCTF
1431 (c99). *J Neurosci* **39**, 5255–5268.
1432 [106] Prasad H, Rao R (2018) Amyloid clearance defect in ApoE4 astrocytes is
1433 reversed by epigenetic correction of endosomal pH. *Proc Natl Acad Sci U S A*
1434 **115**, E6640–E6649.
1435 [107] Tisato V, Zuliani G, Vigliano M, Longo G, Franchini E, Secchiero P, Zauli G,
1436 Paraboschi EM, Singh AV, Serino ML, Ortolani B, Zurlo A, Bosi C, Greco A,
1437 Seripa D, Asselta R, Gemmati D, Vikram Singh A, Serino ML, Ortolani B, Zurlo
1438 A, Bosi C, Greco A, Seripa D, Asselta R, Gemmati D (2018) Gene-gene
1439 interactions among coding genes of iron-homeostasis proteins and APOE-
1440 alleles in cognitive impairment diseases. *PLoS One* **13**, e0193867.
1441 [108] Van Acker ZP, Bretou M, Annaert W (2019) Endo-lysosomal dysregulations
1442 and late-onset Alzheimer's disease: Impact of genetic risk factors. *Mol*
1443 *Neurodegener* **14**, 1–20.
1444 [109] Venkataramani V, Doeppner TR, Willkommen D, Cahill CM, Xin Y, Ye G, Liu
1445 Y, Southon A, Aron A, Au-Yeung HY, Huang X, Lahiri DK, Wang F, Bush AI,
1446 Wulf GG, Ströbel P, Michalke B, Rogers JT (2018) Manganese causes
1447 neurotoxic iron accumulation via translational repression of amyloid precursor
1448 protein and H-Ferritin. *J Neurochem* **147**, 831–848.
1449 [110] McCarthy RC, Park Y, Kosman DJ (2014) sAPP modulates iron efflux from
1450 brain microvascular endothelial cells by stabilizing the ferrous iron exporter
1451 ferroportin . *EMBO Rep* **15**, 809–815.
1452 [111] Yeh K, Yeh M, Glass J (2011) Interactions between ferroportin and hephaestin

1453 in rat enterocytes are reduced after iron ingestion. *Gastroenterology* **141**, 292-
1454 299.e1.
1455 [112] Yeh K, Yeh M, Mims L, Glass J (2009) Iron feeding induces ferroportin 1 and
1456 hephaestin migration and interaction in rat duodenal epithelium. *Am J Physiol*
1457 *Liver Physiol* **296**, G55–G65.
1458 [113] Han O, Kim E-Y (2007) Colocalization of ferroportin-1 with hephaestin on the
1459 basolateral membrane of human intestinal absorptive cells. *J Cell Biochem*
1460 **101**, 1000–1010.
1461 [114] Tsatsanis A, Wong BX, Gunn AP, Ayton S, Bush AI, Devos D, Duce JA (2020)
1462 Amyloidogenic processing of Alzheimer's disease β-amyloid precursor protein
1463 induces cellular iron retention. *Mol Psychiatry* 1–9.
1464 [115] Pigoni M, Wanngren J, Kuhn PH, Munro KM, Gunnersen JM, Takeshima H,
1465 Feederle R, Voytyuk I, De Strooper B, Levasseur MD, Hrupka BJ, Müller SA,
1466 Lichtenthaler SF (2016) Seizure protein 6 and its homolog seizure 6-like
1467 protein are physiological substrates of BACE1 in neurons. *Mol Neurodegener*
1468 **11**, 1–18.
1469 [116] Rogers JT, Randall JD, Cahill CM, Eder PS, Huang X, Gunshin H, Leiter L,
1470 McPhee J, Sarang SS, Utsuki T, Greig NH, Lahiri DK, Tanzi RE, Bush AI,
1471 Giordano T, Gullans SR (2002) An iron-responsive element type II in the 5′-
1472 untranslated region of the Alzheimer's amyloid precursor protein transcript. *J*
1473 *Biol Chem* **277**, 45518–45528.
1474
1475

1476 # Figure Legends

1477 **Figure 1. Iron Responsive Element (IRE) background. A. The classical paradigm of**
1478 **altered transcript stability due to the IRP/IRE system under iron dyshomeostasis.**
1479 When cellular iron levels (through FeS) are low, Iron Regulatory Proteins (IRP1 or IRP2)
1480 bind IREs in the 5' or 3' untranslated regions (UTRs) of genes involved in iron homeostasis.
1481 In general, genes with IREs in their 3' UTRs will be stabilised and hence increased in
1482 expression while genes with IREs in their 5' UTR will have repressed translation from IRP
1483 binding. However, when cellular iron levels are high, IRP1 undergoes a conformation
1484 change and is now unable to bind IREs. IRP1 now acts as a cytosolic aconitase (ACO1).
1485 IRP2 is degraded in high iron conditions and is also unable to bind IREs. Without IRPs
1486 binding to IREs, genes with 5' IREs will have increased translation, while those with 3' IREs
1487 will have decreased mRNA stability, causing their expression to be decreased. **B.**
1488 **Consensus IRE secondary structure and examples of high-quality and IRE-like motifs**
1489 **predicted by SIREs.** IRE-like motifs with non-canonical structure can be detected by SIREs
1490 if they have up to one mismatch pair in the upper stem (e.g. *Hao1*) or 1 unpaired bulge
1491 nucleotide on the 3' strand of the upper stem (e.g. *EPAS1*). For more details on the
1492 prediction of non-canonical IRE motifs, please refer to Figure 1 of Campillos et al. [25].
1493

1494 **Figure 2. Overlap between predicted IRE gene sets for human, mouse, and zebrafish**.
1495 The number of genes in the gene set for each species is shown at the bottom-left bars of
1496 each UpSet plot, while genes with shared homologs across species are indicated in the
1497 main plot region. IRE genes in mouse and zebrafish gene sets were excluded from this plot
1498 if they did not have a human homolog.
1499

1500 **Figure 3. MSigDB gene sets showing over-representation of predicted 3' and 5' IRE**
1501 **genes in human, mouse, and zebrafish**. The top 15 MSigDB gene sets ranked by Fisher's

1502 exact test *p*-value (testing for over-representation of the ***all 3' IREs*** and/or ***all 5' IREs*** sets)
1503 are shown for each species. In the network plots, the top 15 MSigDB gene sets are shown
1504 as large nodes, with genes represented as small nodes. Edges connecting genes to gene
1505 sets indicate the gene set(s) that a gene belongs to. Overall, the ***all 3' IREs*** and ***all 5' IREs***
1506 gene sets have a large proportion of genes which are not included in any of the top ranked
1507 MSigDB gene sets for each species.
1508
1509 **Figure 4. IRE-containing gene expression analysis workflow.** The section including
1510 identification of candidate 3' IRE genes stabilized under iron deficiency was only applied to
1511 the fAD-like zebrafish dataset due to unavailable raw RNA-seq reads for the other datasets
1512 needed to identify expression of unspliced genes.
1513
1514 **Figure 5. Analysis of Caco-2 cultured cell line microarray dataset. A. Gene set**
1515 **enrichment testing results for human predicted IRE gene sets for the iron overload**
1516 **and iron deficiency treatments.** Dots indicate if a gene set was considered significantly
1517 enriched (Bonferroni-adjusted *p*-value < 0.05) in the iron overload (on left) or iron deficiency
1518 (on right) treatment. **B. Gene set enrichment testing results for MSigDB Hallmark gene**
1519 **sets in the iron overload and iron deficiency treatments.** Dots indicate if a gene set was
1520 considered significantly enriched (FDR-adjusted *p*-value < 0.05) in the iron overload (on left)
1521 or iron deficiency (on right) treatments. **C. UpSet plots showing overlap between iron**
1522 **overload and iron deficiency treatments in GSEA leading-edge genes for the *All 3'***
1523 ***IREs* and *All 5' IREs* gene sets**. The bars to the lower-left indicate the number of the
1524 leading-edge genes for iron overload and iron deficiency treatments, while the bars in the
1525 main plot region indicate the number of leading-edge genes which are unique or shared
1526 between the treatments.
1527

1528 **Figure 6. Analysis of Mayo Clinic RNA-seq dataset with human IRE gene sets. A.**
1529 **Enrichment analysis results of IRE gene sets in comparisons of AD (Alzheimer's**
1530 **Disease), PA (Pathological Ageing) and PSP (Progressive Supranuclear Palsy) vs.**
1531 **control.** The proportions of IRE genes within each IRE gene set with increased, decreased,
1532 or unchanged expression are indicated as coloured bars. The dots indicate that the IRE
1533 gene set was significantly enriched (Bonferroni-adjusted enrichment *p*-value < 0.05). **B.**
1534 **UpSet plots showing overlap between leading-edge IRE genes in comparisons of AD**
1535 **(Alzheimer's Disease), PA (Pathological Aging), and PSP (Progressive Supranuclear**
1536 **Palsy) vs. control.** The bars to the lower-left region of each UpSet plot indicate the number
1537 of leading-edge IRE genes for each comparison, while bars in the main plot region indicate
1538 the number of leading-edge IRE genes which are unique or shared between comparisons.
1539
1540 **Figure 7. Analysis of 5XFAD mouse dataset. A. Experimental design and results of**
1541 **IRE gene set enrichment analysis.** The gene sets ***all 3' IREs*** and ***all 5' IREs*** derived from
1542 searching for IRE sequences in the UTRs of genes in the reference mouse genome mm10
1543 are represented here as "3' IRE" and "5' IRE" respectively. Asterisks (*) indicate that the
1544 gene set was significantly enriched in a particular comparison (Bonferroni-adjusted *p*-value <
1545 0.05). **B. Proportions of genes in IRE and MSigDB Hallmark gene sets which are**
1546 **increased (t > 2) or decreased (t < -2) in expression in all "5XFAD vs. wild type"**
1547 **comparisons**. A dot next to a bar indicates that the gene set was significantly enriched
1548 (FDR-adjusted *p*-value < 0.05).
1549
1550 **Figure 8. UpSet plots showing the overlap in GSEA leading-edge genes between all**
1551 **comparisons in the 5XFAD mouse datasets for the gene sets all 3' IREs and all 5'**
1552 **IREs**. Numbers of genes for each intersection are shown above intersection bars.
1553

1554 **Figure 9. Differential gene expression and gene set enrichment analysis in the fAD-**
1555 **like zebrafish dataset. A. Results of differential gene expression analysis. Genes**
1556 **which were significantly increased or decreased in expression are indicated in boxes**.
1557 These differentially expressed genes have FDR-adjusted *p*-value < 0.05. **B. Gene set**
1558 **enrichment with MSigDB Hallmark gene sets.** The comparisons between Q96_ K97del/+
1559 fAD-like mutants and their wild-type siblings are shown for the 6-month-old (young adult)
1560 and 24-month-old (infertile adult) age groups. Dots indicate gene sets which are significantly
1561 enriched (FDR-adjusted *p*-value < 0.05).
1562
1563 **Figure 10. Iron Responsive Element (IRE)-containing gene expression in the fAD-like**
1564 **zebrafish dataset. A. Results of gene set enrichment testing using predicted IRE gene**
1565 **sets.** We represent the gene sets all 3' IREs and all 5' IREs derived from searching for IRE
1566 and IRE-like sequences from z11 reference zebrafish gene UTRs as "3' IRE" and "5' IRE" in
1567 the panel. **B. Proportions of predicted IRE genes which are increased (t > 2) or**
1568 **decreased (t < -2) in expression for each pairwise comparison in the dataset. C.**
1569 **Principal component analysis of all genes in the sets all 3' IREs and all 5' IREs for all**
1570 **samples.** Circles on the 3' IRE plot show that different conditions generally have distinct
1571 expression of genes in the ***all 3' IREs set*** but not in the ***all 5' IREs*** set. **D. UpSet plot**
1572 **showing overlap in leading-edge genes for the *all 3' IREs* gene set for select**
1573 **comparisons.**
1574
1575 **Figure 11. Expression of validated IRE genes in two cell line datasets. A. Summary of**
1576 **workflow**. mRNAs found to bind mouse IRP1 and/or IRP2 via immunoprecipitation in
1577 Sanchez et al. [24] were filtered for those which contained a high-quality IRE in their 3' or 5'
1578 UTR in human and mouse Ensembl gene models. Expression of these validated IRE genes
1579 was examined in a mouse splenic B cell dataset and a human Caco-2 cell line dataset.

1580 Within each dataset, *t*-tests were used to test for significant expression differences of
1581 validated IRE genes between iron deficiency-treated cells and control cells. *t*-test *p*-values ≤
1582 0.05 were considered to indicate statistically significant differences. Statistically significant 3'
1583 IRE genes were categorized as either "consistent with the IRE paradigm" (increased
1584 expression in iron deficiency-treated cells) or "inconsistent with the IRE paradigm"
1585 (decreased expression in iron deficiency-treated cells). 5' IRE genes were not expected to
1586 show significant differences according to the IRE paradigm, but 5' IRE genes showing
1587 significant differences in either direction in iron deficiency-treated cells were also recorded
1588 and plotted. **B. Expression of validated IRE genes showing statistically significant**
1589 **differences in iron deficiency-treated human Caco-2 cells.** Iron deficiency-treated cells
1590 refers to the "Iron-free Medium" cells relative to "Iron-free Medium + FAC" cells. Numbers
1591 over boxplots are *p*-values from *t*-tests between groups. **C. Expression of validated IRE**
1592 **genes showing statistically significant differences in iron deficiency-treated mouse**
1593 **splenic B cells.** Iron deficiency-treated cells refers to the deferoxamine (DFO)-treated cells
1594 relative to control cells. *p*-values are from *t*-tests between DFO-treated and control cells.

1595

1596

Supplementary Information

Supplementary Text

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

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.

1630 • **Supplementary Table 5.** Altered stability of transcripts in the fAD-mutation-like
1631 zebrafish dataset. See Methods and Supplementary Text 4 for details on how
1632 stability was defined and calculated. Genes which contain a 3' IRE (belonging to one
1633 of the 3' IRE gene sets) or 5' IRE (belonging to one of the 5' IRE gene sets) are
1634 indicated with the "has3ire" or "has5ire" columns.
1635 • **Supplementary Table 6.** Validated IRE genes with high-quality IREs in their 3' or 5'
1636 UTRs in human and mouse, based on Ensembl gene models. The genes in this table
1637 are derived from information contained within Figure 2 and Supplementary Tables S1
1638 and S2 of Sanchez et al. [24], who used immunoprecipitation to determine mRNA
1639 transcripts binding to IRP1 and/or IRP2 in mouse.
1640 • **Supplementary Table 7.** Gene set test results for all comparisons in the Mayo Clinic
1641 RNA-seq dataset. Gene sets tested are those in the Hallmark Collection from the
1642 Molecular Signatures Database. Gene sets were considered significantly enriched if
1643 their Bonferroni-adjusted combined p-value from *fry*, *camera*, and *fgsea* was < 0.05.
1644

1645 ## Supplementary Figures

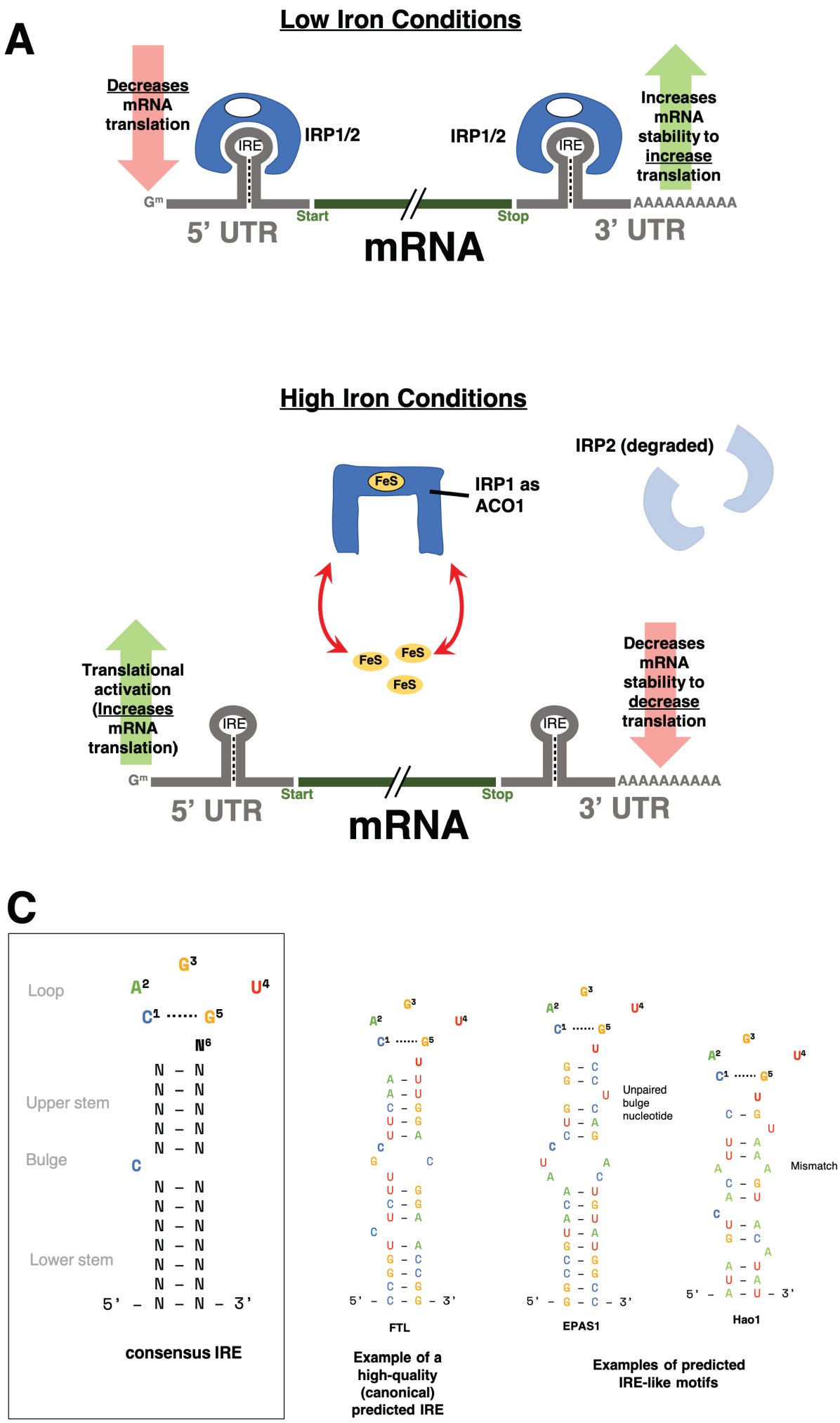
1646
1647 • **Supplementary Figure 1.** Expression of previously characterized IRE genes in
1648 datasets analyzed. **A.** Caco-2 cell line dataset. **B.** Mayo Clinic RNA-seq (Human)
1649 dataset – Temporal Cortex tissue. **C.** Mayo Clinic RNA-seq (Human) dataset –
1650 Cerebellum tissue. **D.** 5XFAD mouse dataset. **E.** fAD-mutation-like zebrafish dataset.
1651 • **Supplementary Figure 2.** Principal Component Analysis of results from different
1652 gene set testing methods.
1653 • **Supplementary Figure 3.** Analysis of cultured Caco-2 cell line dataset. **A.** Principal
1654 Component Analysis plot of gene expression in the cultured Caco-2 cell line dataset.
1655 **B.** Volcano plots indicating differential gene expression due to iron overload (DMEM-
1656 FBS Medium + Hemin vs. DMEM-FBS Medium) and iron deficiency (Iron-free
1657 medium vs. Iron-free medium + FAC) treatments in the Caco-2 cell line dataset.
1658 • **Supplementary Figure 4.** Gene set enrichment testing results for the AD vs. control
1659 comparison in cerebellum and temporal cortex.
1660 • **Supplementary Figure 5.** STRINGR protein-protein interaction network plot
1661 between the 19 shared leading-edge genes between the *psen1*$^{Q96\_K97del/+}$ and
1662 *psen1*$^{+/+}$ comparisons at 6 and 24-months-old.
1663 • **Supplementary Figure 6.** Principal Component Analysis plots showing lack of clear

1664 association between IRE gene expression and the *psen1*$^{K97fs/+}$ mutant genotype.

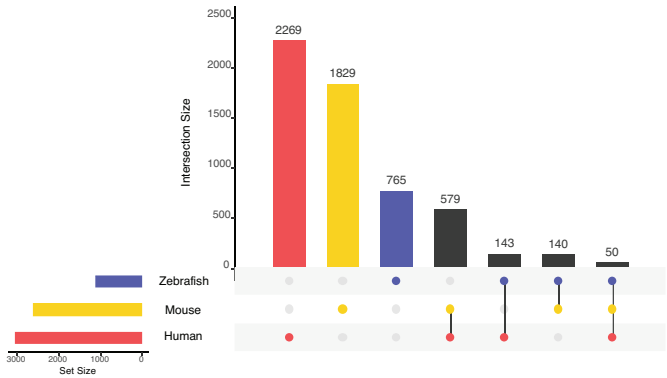
1665 • **Supplementary Figure 7.** Predicted 3' IRE-containing transcripts with significant
1666 differences in stability between conditions in the fAD-like zebrafish dataset.

1667 • **Supplementary Figure 8.** Principal Component Analysis (PCA) plot of splenic
1668 mouse B cell dataset (GEO Accession: GSE77306).

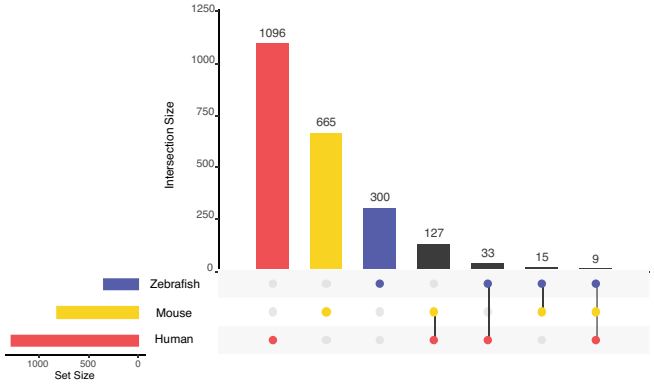
1669


A
Low Iron Conditions
Decreases mRNA translation
IRE
IRP1/2
IRP1/2
IRE
Increases mRNA stability to increase translation
Gm
Start
Stop
AAAAAAAAAA
5' UTR
mRNA
3' UTR
High Iron Conditions
IRP2 (degraded)
FeS
IRP1 as ACO1
FeS
FeS
FeS
Translational activation (Increases mRNA translation)
IRE
IRE
Decreases mRNA stability to decrease translation
Gm
Start
Stop
AAAAAAAAAA
5' UTR
mRNA
3' UTR
C
Loop
G3
A2
U4
C1 ..... G5
N6
Upper stem
N – N
N – N
N – N
N – N
N – N
Bulge
C
N – N
N – N
N – N
N – N
Lower stem
N – N
N – N
5' – N – N – 3'
consensus IRE
FTL
Example of a high-quality (canonical) predicted IRE
Unpaired bulge nucleotide
EPAS1
Mismatch
Hao1
Examples of predicted IRE-like motifs

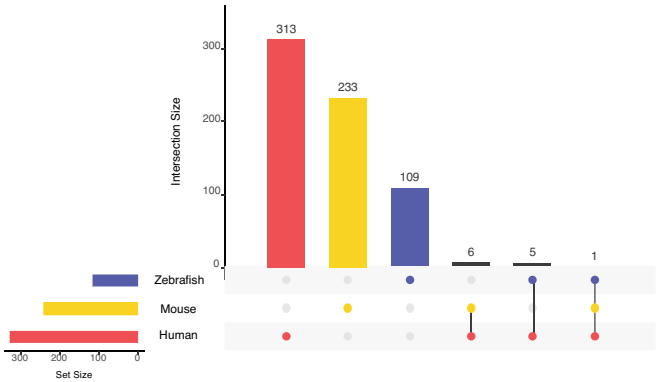
All predicted 3’ IRE genes



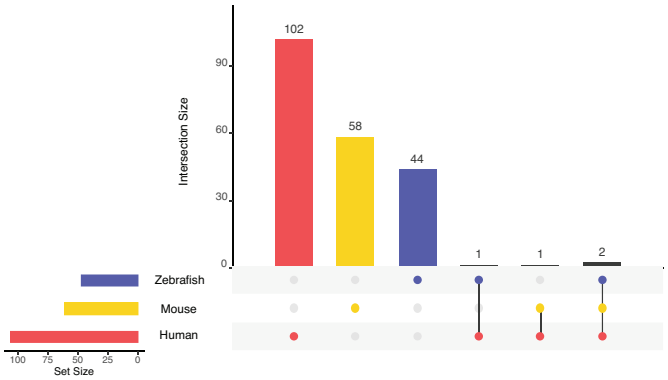
All predicted 5’ IRE genes



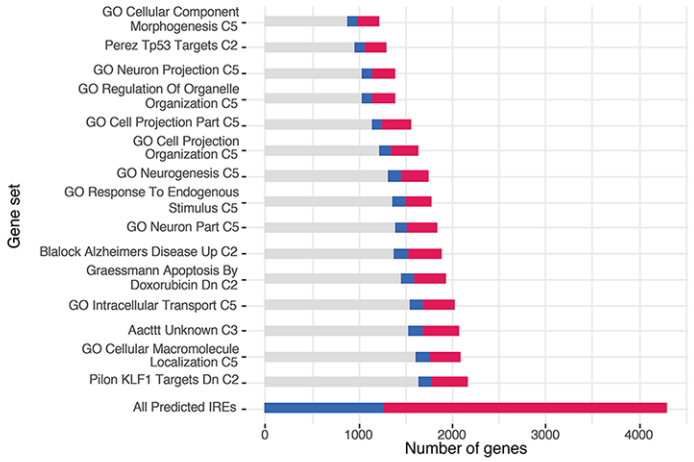
High-quality predicted 3’ IRE genes

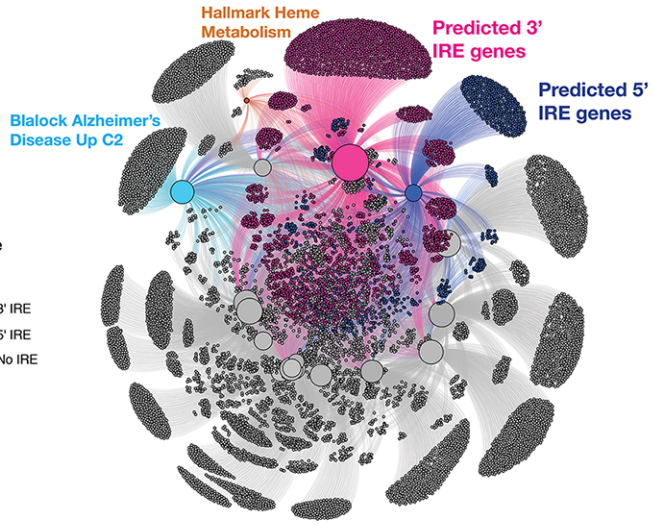


High-quality predicted 5’ IRE genes

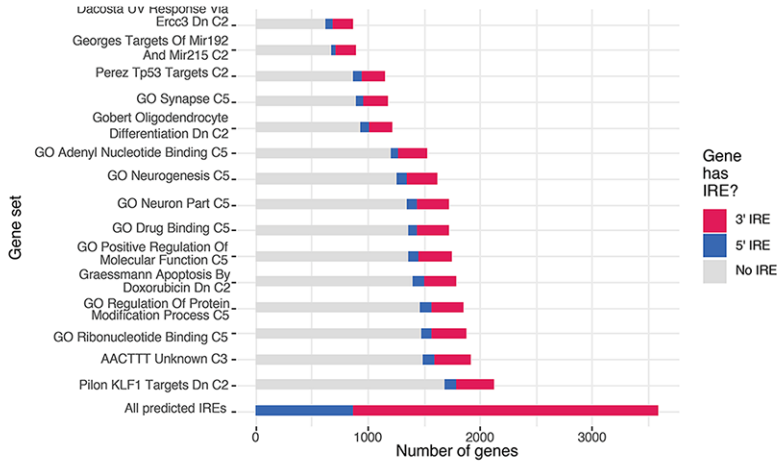


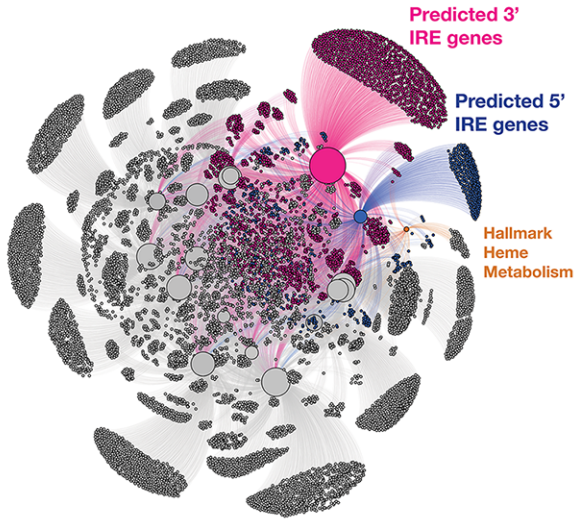
Human



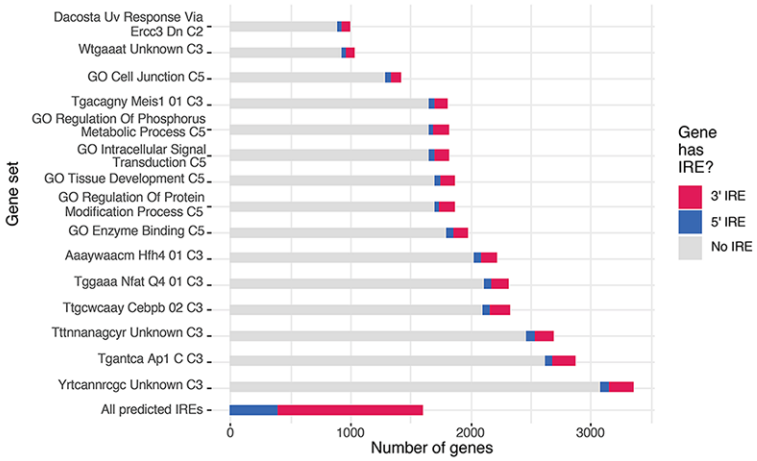


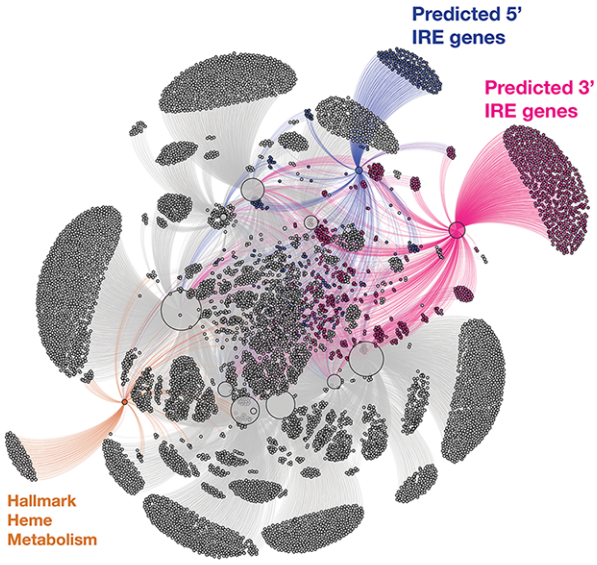
Mouse

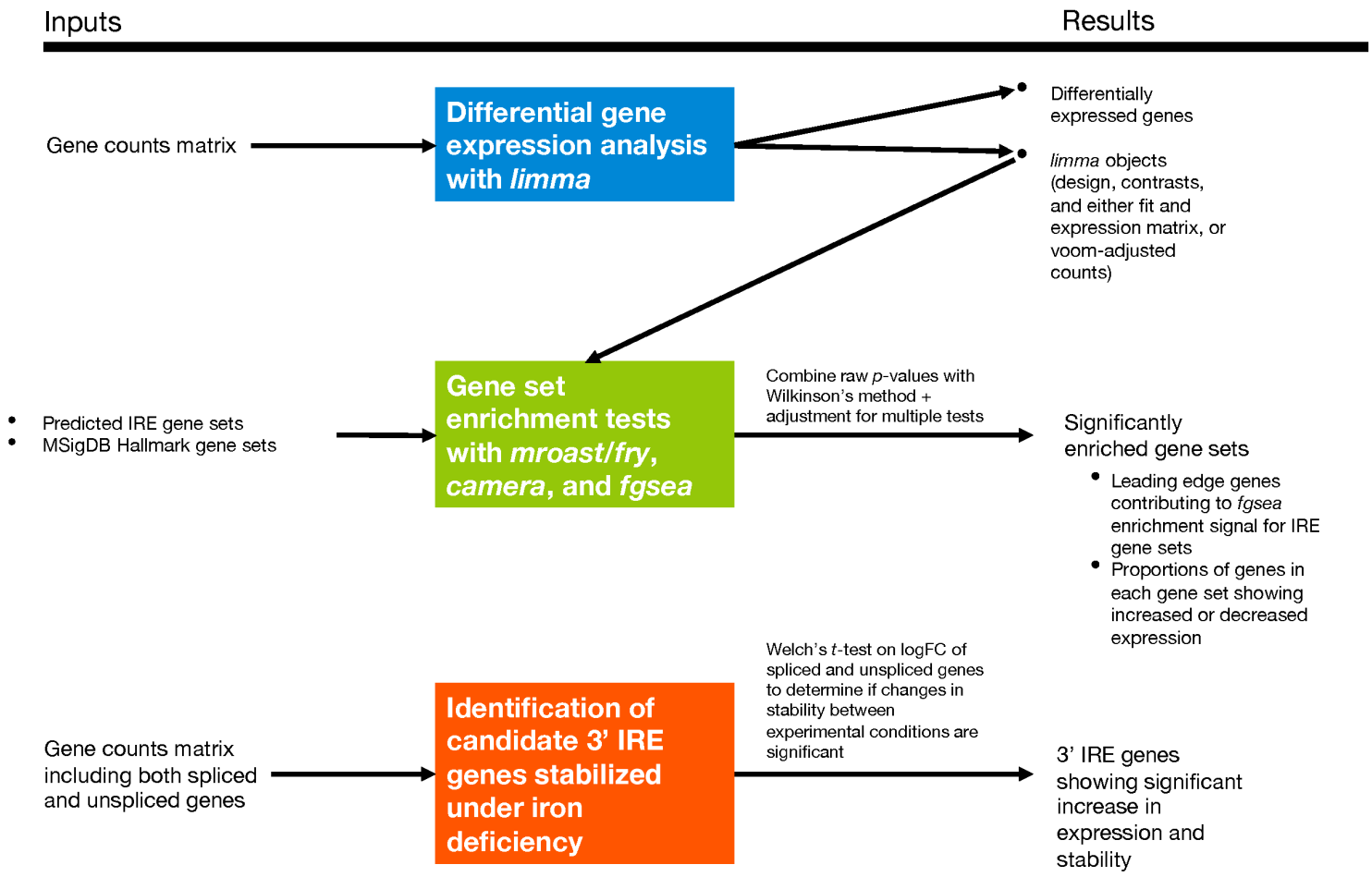


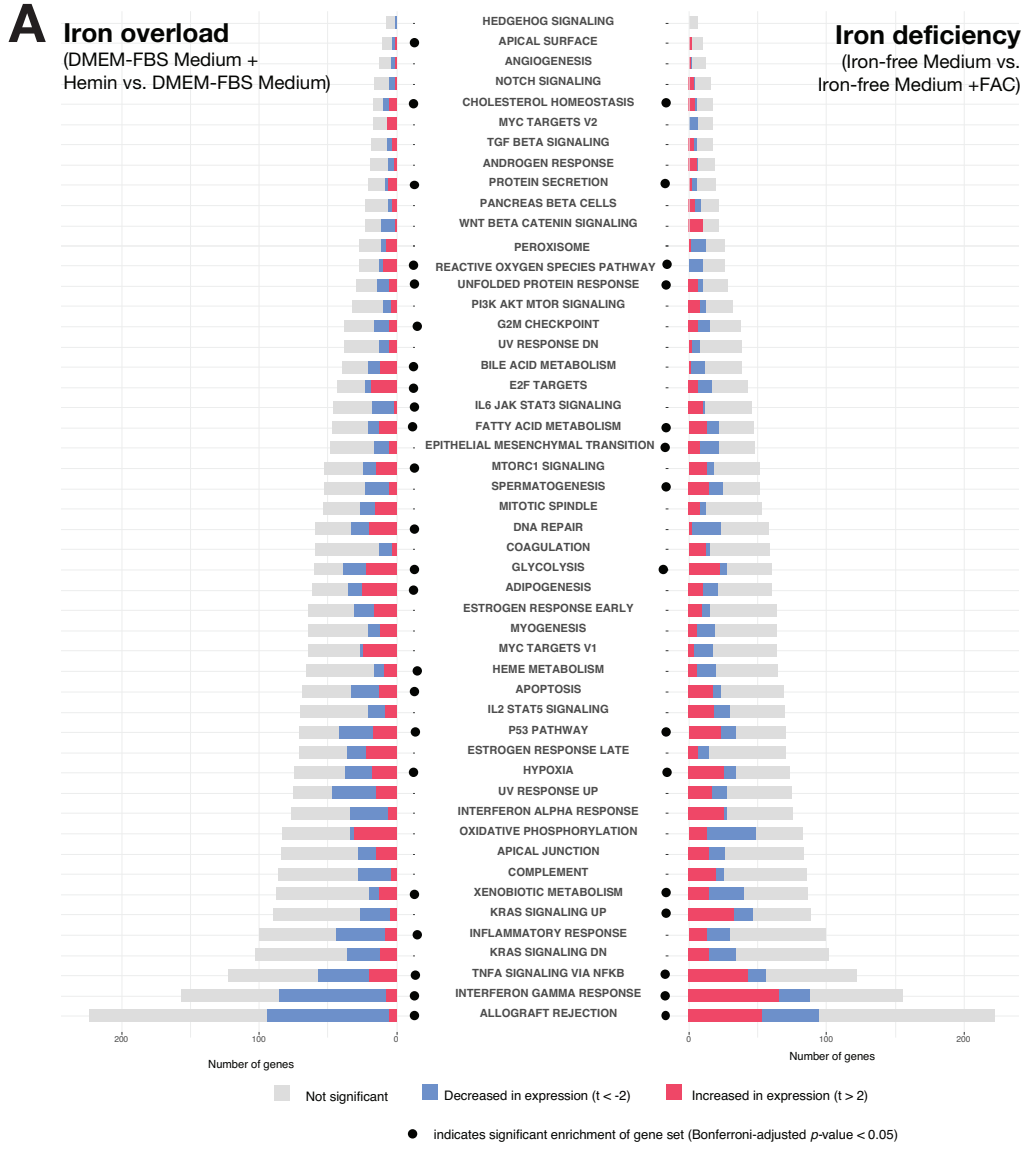


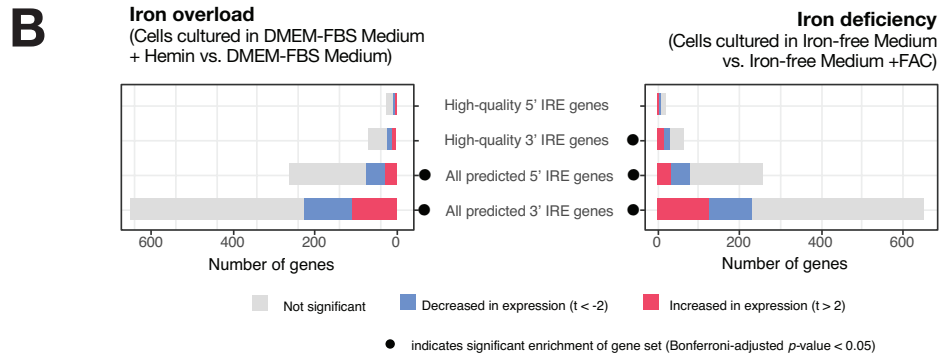
Zebrafish

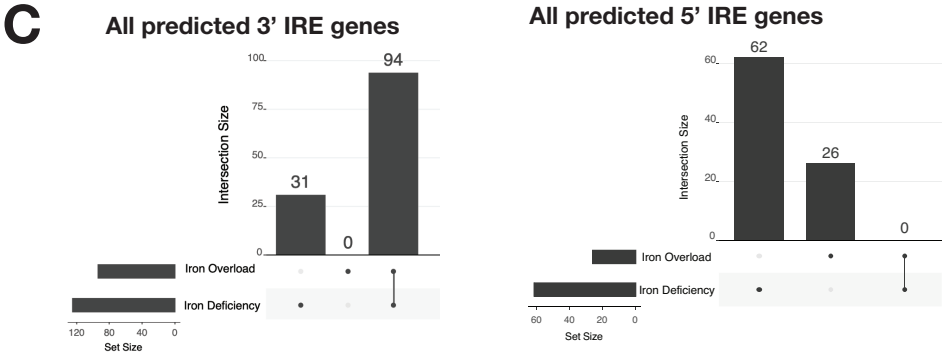





Inputs
Results
Gene counts matrix
Differential gene expression analysis with limma
Differentially expressed genes
limma objects (design, contrasts, and either fit and expression matrix, or voom-adjusted counts)
Predicted IRE gene sets
MSigDB Hallmark gene sets
Gene set enrichment tests with mroast/fry, camera, and fgsea
Combine raw p-values with Wilkinson's method + adjustment for multiple tests
Significantly enriched gene sets
Leading edge genes contributing to fgsea enrichment signal for IRE gene sets
Proportions of genes in each gene set showing increased or decreased expression
Gene counts matrix including both spliced and unspliced genes
Identification of candidate 3' IRE genes stabilized under iron deficiency
Welch's t-test on logFC of spliced and unspliced genes to determine if changes in stability between experimental conditions are significant
3' IRE genes showing significant increase in expression and stability

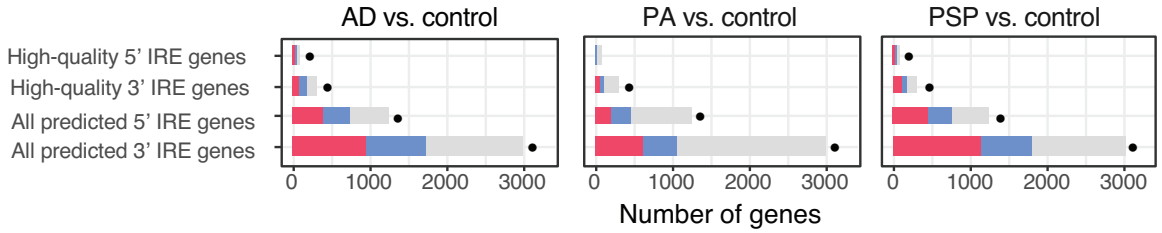

A
Iron overload
(DMEM-FBS Medium + Hemin vs. DMEM-FBS Medium)
Iron deficiency
(Iron-free Medium vs. Iron-free Medium +FAC)
HEDGEHOG SIGNALING
APICAL SURFACE
ANGIOGENESIS
NOTCH SIGNALING
CHOLESTEROL HOMEOSTASIS
MYC TARGETS V2
TGF BETA SIGNALING
ANDROGEN RESPONSE
PROTEIN SECRETION
PANCREAS BETA CELLS
WNT BETA CATENIN SIGNALING
PEROXISOME
REACTIVE OXYGEN SPECIES PATHWAY
UNFOLDED PROTEIN RESPONSE
PI3K AKT MTOR SIGNALING
G2M CHECKPOINT
UV RESPONSE DN
BILE ACID METABOLISM
E2F TARGETS
IL6 JAK STAT3 SIGNALING
FATTY ACID METABOLISM
EPITHELIAL MESENCHYMAL TRANSITION
MTORC1 SIGNALING
SPERMATOGENESIS
MITOTIC SPINDLE
DNA REPAIR
COAGULATION
GLYCOLYSIS
ADIPOGENESIS
ESTROGEN RESPONSE EARLY
MYOGENESIS
MYC TARGETS V1
HEME METABOLISM
APOPTOSIS
IL2 STAT5 SIGNALING
P53 PATHWAY
ESTROGEN RESPONSE LATE
HYPOXIA
UV RESPONSE UP
INTERFERON ALPHA RESPONSE
OXIDATIVE PHOSPHORYLATION
APICAL JUNCTION
COMPLEMENT
XENOBIOTIC METABOLISM
KRAS SIGNALING UP
INFLAMMATORY RESPONSE
KRAS SIGNALING DN
TNFA SIGNALING VIA NFKB
INTERFERON GAMMA RESPONSE
ALLOGRAFT REJECTION
200
100
0
Number of genes
0
100
200
Number of genes
Not significant
Decreased in expression (t < -2)
Increased in expression (t > 2)
indicates significant enrichment of gene set (Bonferroni-adjusted p-value < 0.05)


B
Iron overload
(Cells cultured in DMEM-FBS Medium + Hemin vs. DMEM-FBS Medium)
Iron deficiency
(Cells cultured in Iron-free Medium vs. Iron-free Medium +FAC)
High-quality 5' IRE genes
High-quality 3' IRE genes
All predicted 5' IRE genes
All predicted 3' IRE genes
600
400
200
0
Number of genes
0
200
400
600
Number of genes
Not significant
Decreased in expression (t < -2)
Increased in expression (t > 2)
indicates significant enrichment of gene set (Bonferroni-adjusted p-value < 0.05)

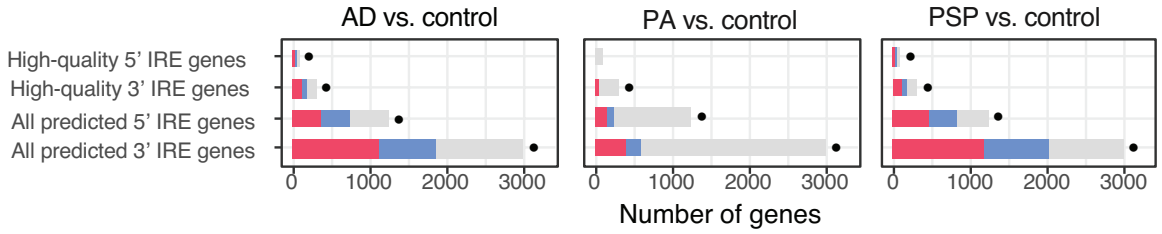

C
All predicted 3' IRE genes
Intersection Size
100
75
50
25
0
31
0
94
Iron Overload
Iron Deficiency
120
80
40
0
Set Size
All predicted 5' IRE genes
Intersection Size
60
40
20
0
62
26
0
Iron Overload
Iron Deficiency
60
40
20
0
Set Size

A

Temporal Cortex



Cerebellum

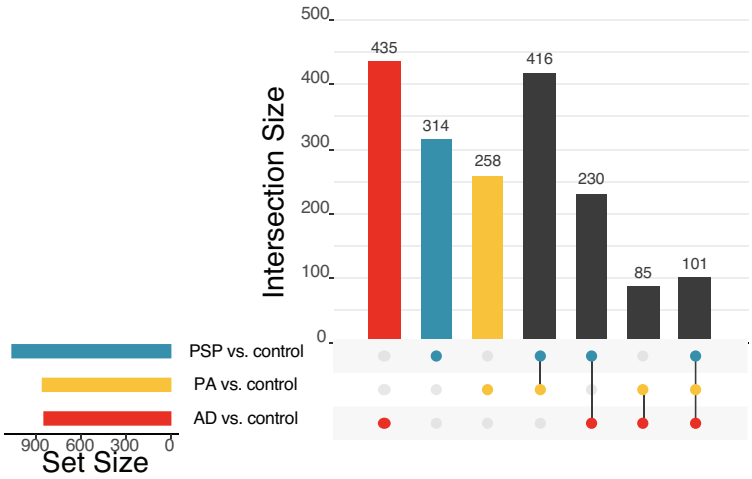


Not significant | Decreased in expression (t < -2) | Increased in expression (t > 2)

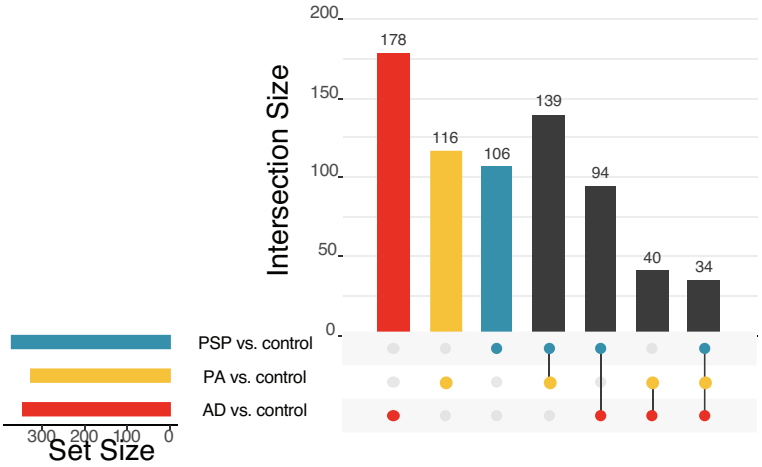
B

Temporal Cortex

Leading-edge genes in all predicted 3' IRE genes

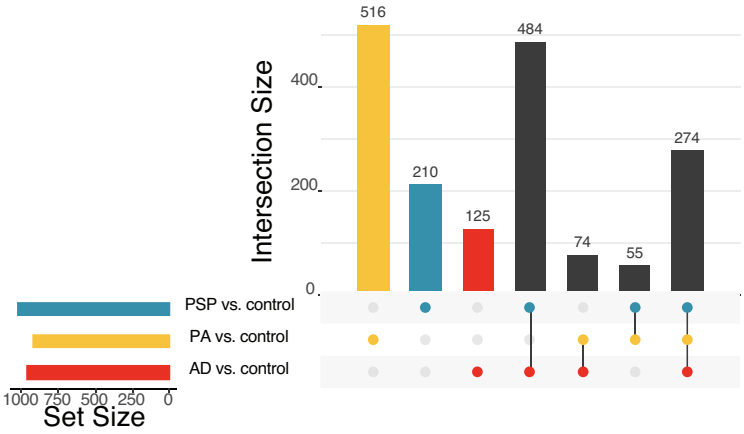


Leading-edge genes in all predicted 5' IRE genes

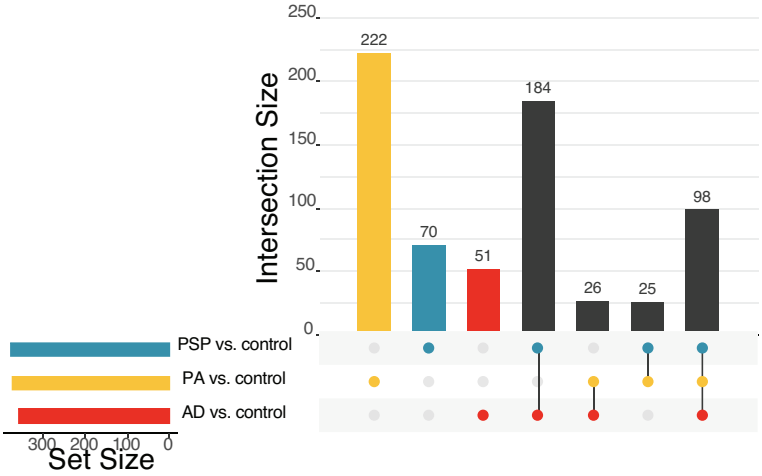


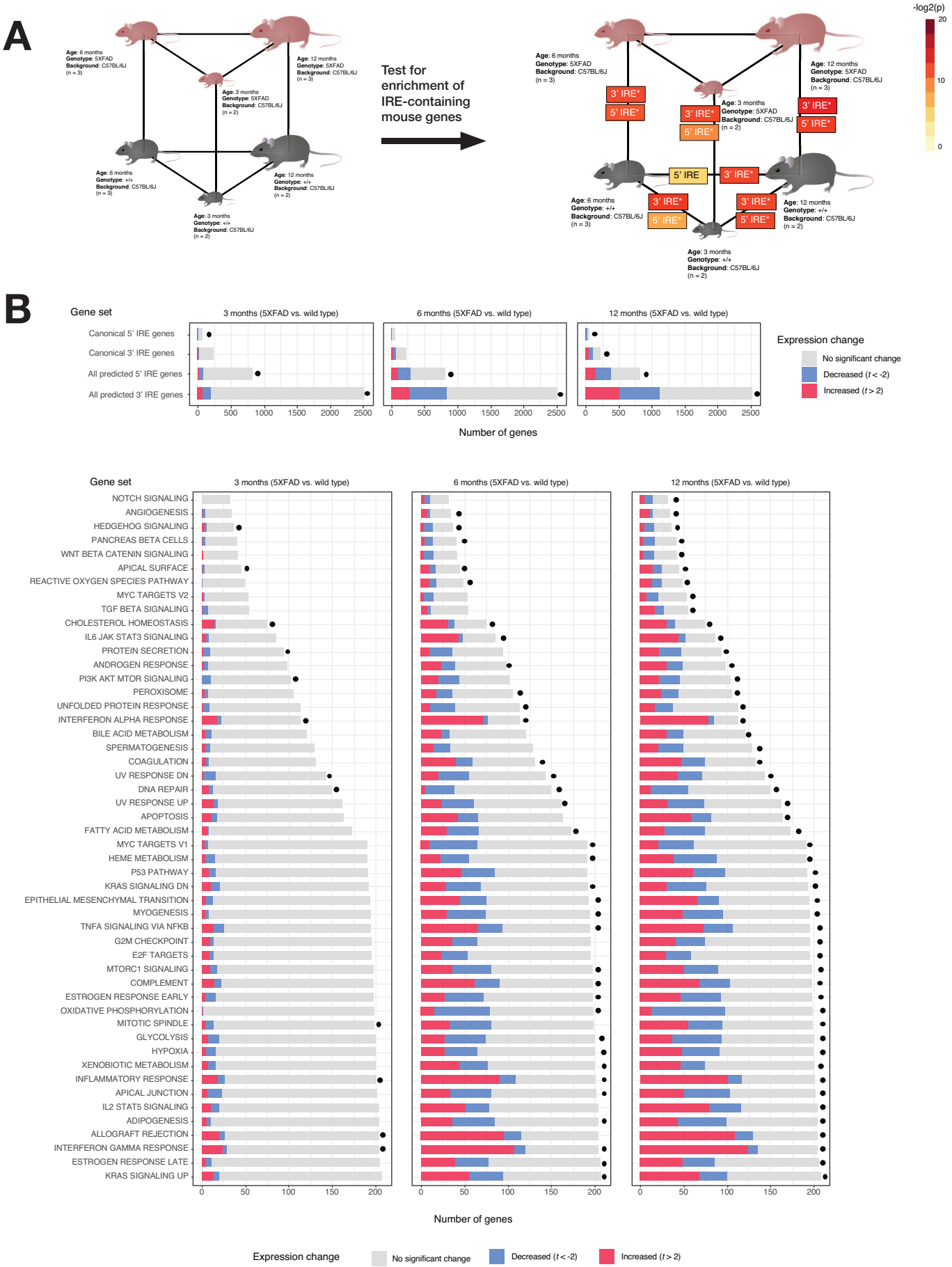
Cerebellum

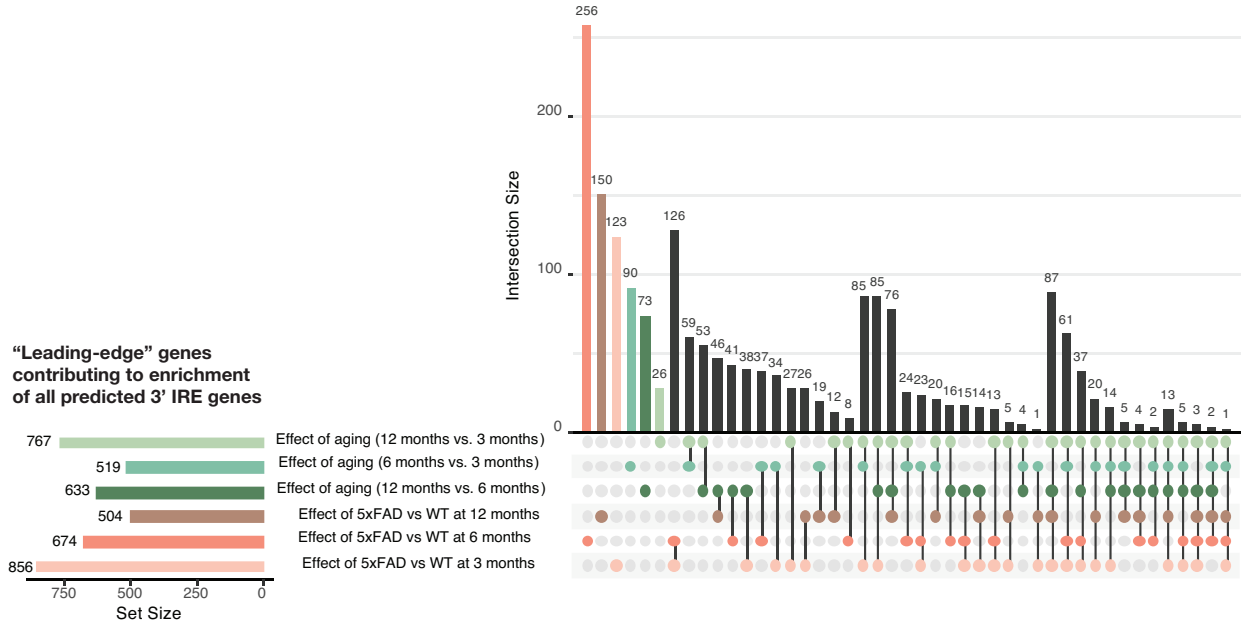
Leading-edge genes in all predicted 3' IRE genes

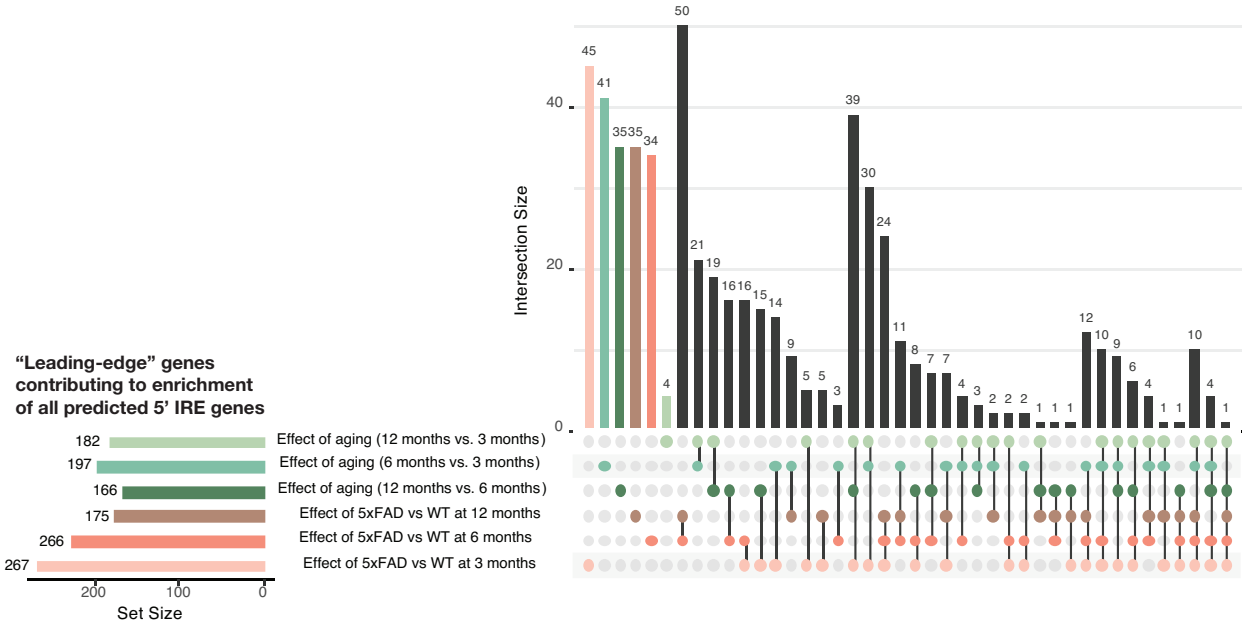


Leading-edge genes in all predicted 5' IRE genes

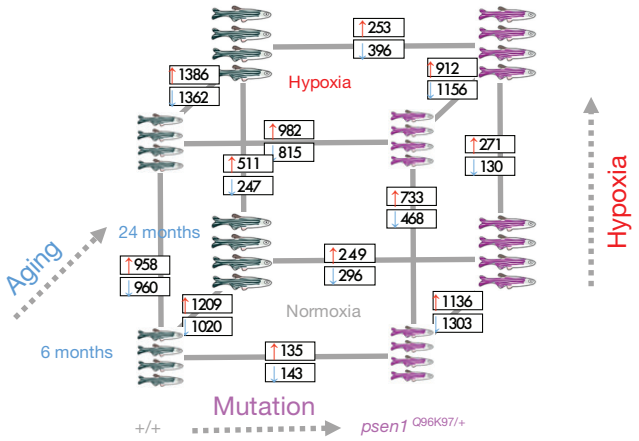



A
Age: 6 months
Genotype: 5XFAD
Background: C57BL/6J
(n = 3)
Age: 12 months
Genotype: 5XFAD
Background: C57BL/6J
(n = 3)
Age: 3 months
Genotype: 5XFAD
Background: C57BL/6J
(n = 2)
Age: 6 months
Genotype: +/+
Background: C57BL/6J
(n = 3)
Age: 12 months
Genotype: +/+
Background: C57BL/6J
(n = 2)
Age: 3 months
Genotype: +/+
Background: C57BL/6J
(n = 2)
Test for enrichment of IRE-containing mouse genes
-log2(p)
20
10
0
3' IRE*
5' IRE*
5' IRE
B
Gene set
3 months (5XFAD vs. wild type)
6 months (5XFAD vs. wild type)
12 months (5XFAD vs. wild type)
Canonical 5' IRE genes
Canonical 3' IRE genes
All predicted 5' IRE genes
All predicted 3' IRE genes
Expression change
No significant change
Decreased (t < -2)
Increased (t > 2)
Number of genes
NOTCH SIGNALING
ANGIOGENESIS
HEDGEHOG SIGNALING
PANCREAS BETA CELLS
WNT BETA CATENIN SIGNALING
APICAL SURFACE
REACTIVE OXYGEN SPECIES PATHWAY
MYC TARGETS V2
TGF BETA SIGNALING
CHOLESTEROL HOMEOSTASIS
IL6 JAK STAT3 SIGNALING
PROTEIN SECRETION
ANDROGEN RESPONSE
PI3K AKT MTOR SIGNALING
PEROXISOME
UNFOLDED PROTEIN RESPONSE
INTERFERON ALPHA RESPONSE
BILE ACID METABOLISM
SPERMATOGENESIS
COAGULATION
UV RESPONSE DN
DNA REPAIR
UV RESPONSE UP
APOPTOSIS
FATTY ACID METABOLISM
MYC TARGETS V1
HEME METABOLISM
P53 PATHWAY
KRAS SIGNALING DN
EPITHELIAL MESENCHYMAL TRANSITION
MYOGENESIS
TNFA SIGNALING VIA NFKB
G2M CHECKPOINT
E2F TARGETS
MTORC1 SIGNALING
COMPLEMENT
ESTROGEN RESPONSE EARLY
OXIDATIVE PHOSPHORYLATION
MITOTIC SPINDLE
GLYCOLYSIS
HYPOXIA
XENOBIOTIC METABOLISM
INFLAMMATORY RESPONSE
APICAL JUNCTION
IL2 STAT5 SIGNALING
ADIPOGENESIS
ALLOGRAFT REJECTION
INTERFERON GAMMA RESPONSE
ESTROGEN RESPONSE LATE
KRAS SIGNALING UP
Number of genes
Expression change
No significant change
Decreased (t < -2)
Increased (t > 2)

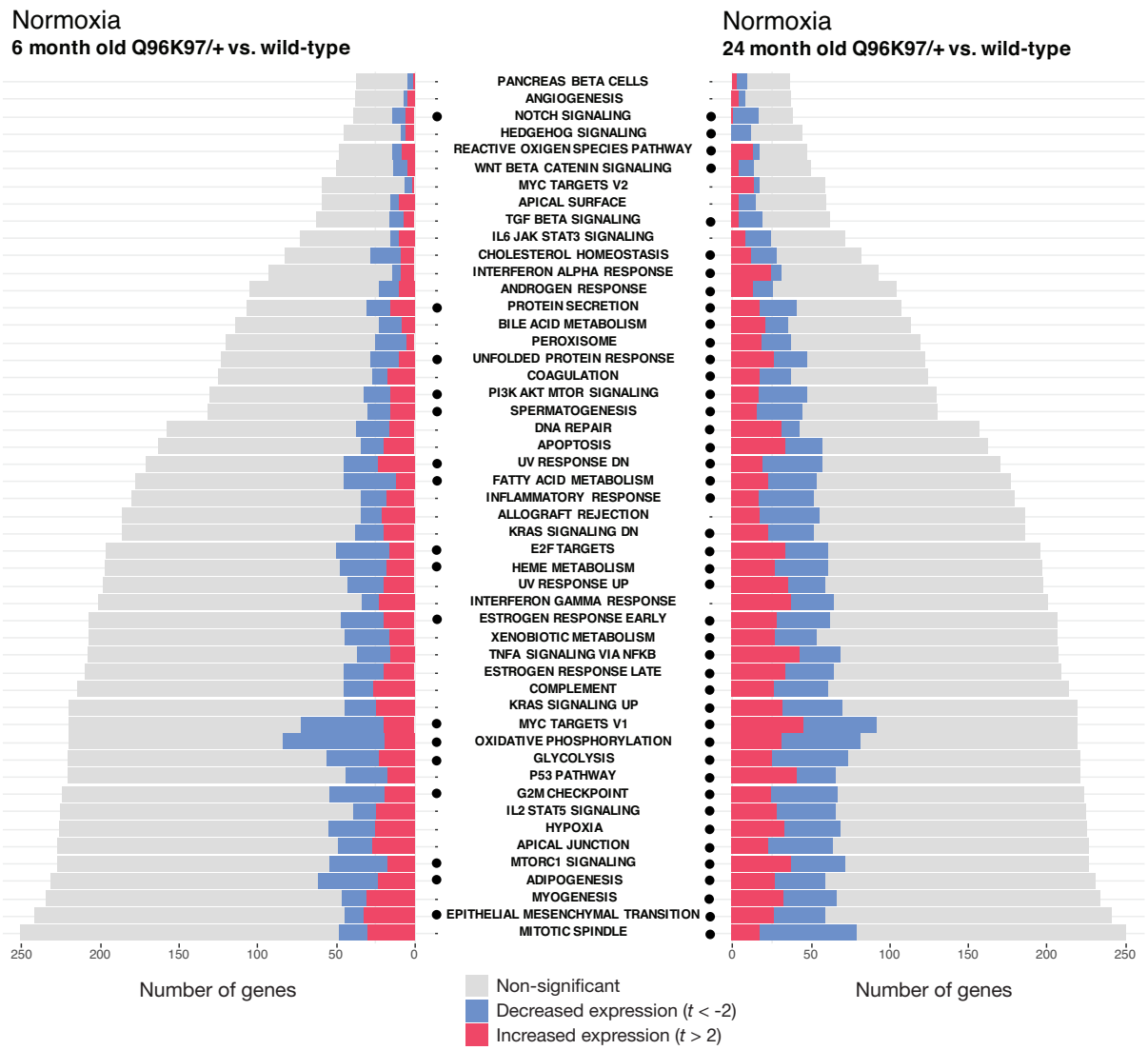

"Leading-edge" genes contributing to enrichment of all predicted 3' IRE genes
767
519
633
504
674
856
750
500
250
0
Set Size
Effect of aging (12 months vs. 3 months)
Effect of aging (6 months vs. 3 months)
Effect of aging (12 months vs. 6 months)
Effect of 5xFAD vs WT at 12 months
Effect of 5xFAD vs WT at 6 months
Effect of 5xFAD vs WT at 3 months
Intersection Size
0
100
200

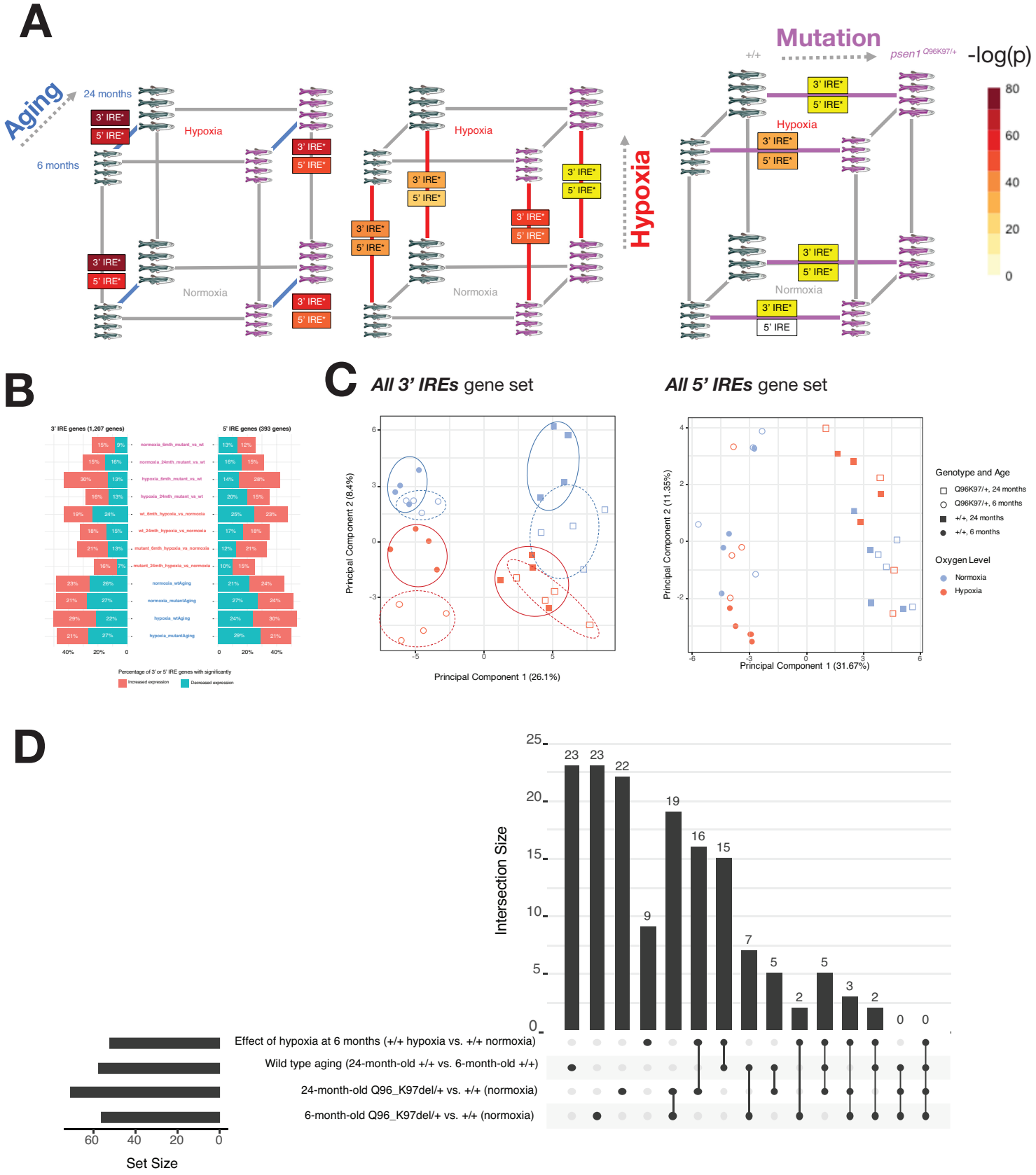

"Leading-edge" genes contributing to enrichment of all predicted 5' IRE genes
182
197
166
175
266
267
200
100
0
Set Size
Effect of aging (12 months vs. 3 months)
Effect of aging (6 months vs. 3 months)
Effect of aging (12 months vs. 6 months)
Effect of 5xFAD vs WT at 12 months
Effect of 5xFAD vs WT at 6 months
Effect of 5xFAD vs WT at 3 months
Intersection Size
0
20
40

A


253
396
1386
1362
912
1156
Hypoxia
982
815
511
247
271
130
733
468
Hypoxia
Aging
24 months
958
960
249
296
1209
1020
Normoxia
1136
1303
6 months
135
143
Mutation
+/+
psen1 Q96K97/+

B


Normoxia
6 month old Q96K97/+ vs. wild-type
Normoxia
24 month old Q96K97/+ vs. wild-type
PANCREAS BETA CELLS
ANGIOGENESIS
NOTCH SIGNALING
HEDGEHOG SIGNALING
REACTIVE OXIGEN SPECIES PATHWAY
WNT BETA CATENIN SIGNALING
MYC TARGETS V2
APICAL SURFACE
TGF BETA SIGNALING
IL6 JAK STAT3 SIGNALING
CHOLESTEROL HOMEOSTASIS
INTERFERON ALPHA RESPONSE
ANDROGEN RESPONSE
PROTEIN SECRETION
BILE ACID METABOLISM
PEROXISOME
UNFOLDED PROTEIN RESPONSE
COAGULATION
PI3K AKT MTOR SIGNALING
SPERMATOGENESIS
DNA REPAIR
APOPTOSIS
UV RESPONSE DN
FATTY ACID METABOLISM
INFLAMMATORY RESPONSE
ALLOGRAFT REJECTION
KRAS SIGNALING DN
E2F TARGETS
HEME METABOLISM
UV RESPONSE UP
INTERFERON GAMMA RESPONSE
ESTROGEN RESPONSE EARLY
XENOBIOTIC METABOLISM
TNFA SIGNALING VIA NFKB
ESTROGEN RESPONSE LATE
COMPLEMENT
KRAS SIGNALING UP
MYC TARGETS V1
OXIDATIVE PHOSPHORYLATION
GLYCOLYSIS
P53 PATHWAY
G2M CHECKPOINT
IL2 STAT5 SIGNALING
HYPOXIA
APICAL JUNCTION
MTORC1 SIGNALING
ADIPOGENESIS
MYOGENESIS
EPITHELIAL MESENCHYMAL TRANSITION
MITOTIC SPINDLE
250
200
150
100
50
0
Number of genes
Non-significant
Decreased expression (t < -2)
Increased expression (t > 2)


A
Aging
24 months
6 months
3' IRE*
5' IRE*
Hypoxia
Normoxia
Hypoxia
Mutation
+/+
psen1 Q96K97/+
-log(p)
80
60
40
20
0
3' IRE
5' IRE
B
3' IRE genes (1,207 genes)
5' IRE genes (393 genes)
Percentage of 3' or 5' IRE genes with significantly
Increased expression
Decreased expression
C
All 3' IREs gene set
All 5' IREs gene set
Principal Component 2 (8.4%)
Principal Component 1 (26.1%)
Principal Component 2 (11.35%)
Principal Component 1 (31.67%)
Genotype and Age
Q96K97/+, 24 months
Q96K97/+, 6 months
+/+, 24 months
+/+, 6 months
Oxygen Level
Normoxia
Hypoxia
D
Intersection Size
23
23
22
9
19
16
15
7
5
2
5
3
2
0
0
Effect of hypoxia at 6 months (+/+ hypoxia vs. +/+ normoxia)
Wild type aging (24-month-old +/+ vs. 6-month-old +/+)
24-month-old Q96_K97del/+ vs. +/+ (normoxia)
6-month-old Q96_K97del/+ vs. +/+ (normoxia)
Set Size