



Zebrafish Chromosome 14 Gene Differential Expression in the *fmr1*^{hu2787} Model of Fragile X Syndrome

OPEN ACCESS

Edited by:

Silvia De Rubeis,
Icahn School of Medicine at Mount
Sinai, United States

Reviewed by:

Andrea Rossi,
Leibniz-Institut für
Umweltmedizinische Forschung (IUF),
Germany
Julia Dallman,
University of Miami, United States

*Correspondence:

Michael Lardelli
michael.lardelli@adelaide.edu.au

†These authors have contributed
equally to this work and share first
authorship

‡These authors have contributed
equally to this work and share senior
authorship

Specialty section:

This article was submitted to
Neurogenomics,
a section of the journal
Frontiers in Genetics

Received: 03 November 2020

Accepted: 14 April 2021

Published: 31 May 2021

Citation:

Barthelson K, Baer L, Dong Y,
Hand M, Pujic Z, Newman M,
Goodhill GJ, Richards RI,
Pederson SM and Lardelli M (2021)
Zebrafish Chromosome 14 Gene
Differential Expression
in the *fmr1*^{hu2787} Model of Fragile X
Syndrome. *Front. Genet.* 12:625466.
doi: 10.3389/fgene.2021.625466

Karissa Barthelson^{1†}, Lachlan Baer^{1†}, Yang Dong^{1†}, Melanie Hand^{2†}, Zac Pujic^{3†},
Morgan Newman¹, Geoffrey J. Goodhill^{3,4}, Robert I. Richards¹, Stephen M. Pederson^{3†}
and Michael Lardelli^{1**}

¹ School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia, ² Bioinformatics Hub, University of Adelaide, Adelaide, SA, Australia, ³ Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia, ⁴ School of Mathematics and Physics, University of Queensland, Brisbane, QLD, Australia

Zebrafish represent a valuable model for investigating the molecular and cellular basis of Fragile X syndrome (FXS). Reduced expression of the zebrafish *FMR1* orthologous gene, *fmr1*, causes developmental and behavioural phenotypes related to FXS. Zebrafish homozygous for the hu2787 non-sense mutation allele of *fmr1* are widely used to model FXS, although FXS-relevant phenotypes seen from morpholino antisense oligonucleotide (morpholino) suppression of *fmr1* transcript translation were not observed when hu2787 was first described. The subsequent discovery of transcriptional adaptation (a form of genetic compensation), whereby mutations causing non-sense-mediated decay of transcripts can drive compensatory upregulation of homologous transcripts independent of protein feedback loops, suggested an explanation for the differences reported. We examined the whole-embryo transcriptome effects of homozygosity for *fmr1*^{hu2787} at 2 days post fertilisation. We observed statistically significant changes in expression of a number of gene transcripts, but none from genes showing sequence homology to *fmr1*. Enrichment testing of differentially expressed genes implied effects on lysosome function and glycosphingolipid biosynthesis. The majority of the differentially expressed genes are located, like *fmr1*, on Chromosome 14. Quantitative PCR tests did not support that this was artefactual due to changes in relative chromosome abundance. Enrichment testing of the “leading edge” differentially expressed genes from Chromosome 14 revealed that their co-location on this chromosome may be associated with roles in brain development and function. The differential expression of functionally related genes due to mutation of *fmr1*, and located on the same chromosome as *fmr1*, is consistent with R.A. Fisher’s assertion that the selective advantage of co-segregation of particular combinations of alleles of genes will favour, during evolution, chromosomal rearrangements that place them in linkage disequilibrium on the same chromosome. However, we cannot exclude that the

apparent differential expression of genes on Chromosome 14 genes was (if only in part), caused by differences between the expression of alleles of genes unrelated to the effects of the *fmr1*^{hu2787} mutation and made manifest due to the limited, but non-zero, allelic diversity between the genotypes compared.

Keywords: fragile X syndrome, zebrafish, transcriptional adaptation, transcriptome analysis, FMR1, chromosome evolution, linkage disequilibrium, homeostasis

BACKGROUND

Fragile X Syndrome (FXS) is an X-chromosome-linked form of human inherited intellectual disability affecting, on a pan-ethnic basis, approximately 1 in every 2,400 male fetuses (Owens et al., 2018) (As females are diploid for the X-chromosome, the proportion of female fetuses affected is far lower). FXS most commonly arises due to expansion of a CGG trinucleotide repeat sequence in the human *FMR1* gene that codes for an RNA-binding protein involved in multiple aspects of RNA metabolism (Salcedo-Arellano et al., 2020). The trinucleotide repeat expansion promotes methylation and repression of *FMR1* transcription (Pieretti et al., 1991). However, other forms of mutation that reduce gene expression can cause the syndrome (Quan et al., 1995; Hammond et al., 1997). The syndrome includes non-cognitive features such as craniofacial changes, macroorchidism, and abnormalities of connective tissue, blood, and pigmentation (reviewed by Salcedo-Arellano et al., 2020). The syndrome is complicated by the existence of subtle cognitive effects due to limited expansion of the trinucleotide repeat known as a “premutation” (Salcedo-Arellano et al., 2020). The syndrome also displays the phenomenon of anticipation whereby phenotype severity increases in successive generations due to expansion of the trinucleotide repeat beyond a premutation range (Sutherland et al., 1991).

Numerous models of FXS in animals as divergent as *Drosophila* and mice have been constructed to facilitate research where access to human brain material is restricted or not possible (reviewed in Dahlhaus, 2018). As expected, all these models have their limitations. For example, mouse models of trinucleotide repeat expansion in the *Fmr1* gene show reduced FMRP protein expression but do not show increased methylation or reduced mRNA levels (Brouwer et al., 2007). For this reason, research is dominated by null mutation models and, conveniently, the mouse *Fmr1* gene is located on the mouse X-chromosome. Mouse *Fmr1* gene knockout models show effects on cognition, postsynaptic dendritic spines, craniofacial changes, and macroorchidism (The Dutch-Belgian Fragile X Consortium, 1994; Kooy et al., 1996; Heulens et al., 2013;

Dahlhaus, 2018). Mouse and zebrafish models of *FMR1* mutation have revealed impaired processing of visual (Goel et al., 2018) and auditory (Constantin et al., 2020) information, respectively, and are systems in which testing of therapeutics offers hope of ameliorating FXS (Chen et al., 2011; Dinday and Baraban, 2015; Wiley et al., 2017; Goel et al., 2018).

In zebrafish, the *FMR1* orthologous gene, *fmr1*, is autosomal so that analysis of loss-of-function phenotypes requires examination of homozygous mutants or embryos injected with morpholino antisense oligonucleotides (morphants). We published the first analysis of *fmr1* morphant zebrafish embryos in 2006 (Tucker et al., 2006). We described changes in axonal branching that could be rescued by treatment with the mGluR antagonist 2-methyl-6-(phenylethynyl)pyridine (MPEP), similar to observations of mGluR effects in mouse and *Drosophila* analyses of *FMR1* activities. We also described changes in craniofacial structure in zebrafish *fmr1* morphants. Subsequently, den Broeder and colleagues described isolation of a null mutation in *fmr1*, allele hu2787, from a TILLING (targeted induced local lesions in genomes) mutation screen (den Broeder et al., 2009). They did not observe neurite branching or craniofacial defects (or any other *FMR1*-related phenotypes) and asserted their model to be superior to analysis of *fmr1* morphants, arguing that the morphant phenotypes were likely artefactual. Analysis of the *fmr1*^{hu2787} mutant now dominates FXS modelling in zebrafish and, fortunately, subsequent analyses of this mutant have described numerous syndrome-relevant phenotypes (Ng et al., 2013; Shamay-Ramot et al., 2015; Marsden et al., 2018; Constantin et al., 2020).

Kok et al. (2015) summarised the puzzling, frequent discordance between phenotypes caused by gene mutations compared to morphant phenotypes caused by reduction of gene expression due to injection of morpholinos. Rossi et al. (2015), then described the phenomenon of “genetic compensation” (now referred to as “transcriptional adaptation” Kontarakis and Stainier, 2020) as contributing to this discordance. As elaborated in a subsequent paper from that laboratory (El-Brolosy et al., 2019), non-sense-mediated decay (NMD) of transcripts with premature termination codons can (in a manner independent of protein feedback loops) increase the abundance of transcripts of genes with homologous sequences that, presumably, are partially functionally redundant and ameliorate the effects of the mutation. The discovery of this phenomenon raises questions regarding the definition of “null” mutant phenotypes and reveals that reducing gene expression using morpholinos may, in some cases, provide more focussed functional effects at the molecular level that are simpler to interpret than those caused by mutations inducing NMD.

Abbreviations: CPM, counts per million; DE, differentially expressed; FC, fold change; FDR, false discovery rate; *FMR1*, *FMRP translational regulator 1* (human gene); *Fmr1*, mouse orthologue of human *FMR1*; *fmr1*, zebrafish orthologue of human *FMR1*; FXS, Fragile X Syndrome; gDNA, genomic DNA; GATK Genome Analysis Toolkit; GEO, Gene Expression Omnibus; KEGG, Kyoto Encyclopaedia of Genes and Genomes; MD, mean-difference; mGluR, metabotropic glutamate receptor; *mitfa/nacre melanocyte inducing transcription factor a* gene; morpholino, morpholino antisense oligonucleotide; MPEP, 2-methyl-6-(phenylethynyl)pyridine; NMD, non-sense-mediated decay; PC, principal component; *psen1*, *presenilin 1* gene; *rpl13*, *Ribosomal Protein L13* gene; SNP single nucleotide polymorphism; TILLING, Targeting Induced Local Lesions in Genomes.

During the billions of years of biological evolution on Earth, there has been enormous selective pressure for the development of robustly stable cellular/physiological systems (homeostasis) that can survive and reproduce in the face of constant environmental and genetic change (mutation). Transcriptional adaptation is one example of a broadly acting homeostatic mechanism. Transcriptome analysis can reveal the genome-wide effects of mutations in particular genes. When a mutated gene normally functions in cellular systems controlling transcription or transcript stability, some of the consequent, genome-wide changes in transcript abundance or structure (“differential expression” of genes) may represent the direct, deleterious effects of the mutation on gene expression. However, the majority of gene expression changes detected in transcriptome analysis likely represent homeostatic responses to maintain the cellular/physiological functions in which the mutated gene normally acts.

We were curious to investigate the effects of *fmr1* mutation on gene expression and whether the homeostatic transcriptional adaptation phenomenon (detected as increased abundance of transcripts from genes containing homologous sequences) might contribute to the reported relatively milder phenotype of homozygosity for the hu2787 mutant allele of *fmr1* compared to that reported for morpholino-injected embryos. Therefore, we compared the transcriptomes of batches of entire 2 days post fertilisation (dpf) embryos homozygous for the hu2787 allele with those of batches of wild type embryos of the same age. This age was selected as a time when *fmr1* is known to be expressed (Thisse and Thisse, 2004; van 't Padje et al., 2005) and near the end of embryogenesis (Kimmel et al., 1995) but within the developmental interval examined in the den Broeder and Tucker et al. papers (Tucker et al., 2006; den Broeder et al., 2009). We observed statistically significant changes in transcript levels for *fmr1* and 20 other genes. However, significantly altered expression of transcripts from genes containing sequences with homology to the *fmr1* transcript were not observed. Curiously, 12 of the 21 genes observed to be differentially expressed, including *fmr1*, are located on Chromosome 14. Bioinformatic analysis of gene expression in the hu2787 homozygous embryos implied effects on lysosomal function and glycosphingolipid biosynthesis consistent with observations in mouse models of *Fmr1* mutation. This suggests that the colocalization of genes on Chromosome 14 affected by the *fmr1*^{hu2787} mutation may be associated with their coaction with *fmr1* in brain development and function. The differential expression of functionally related genes located on the same chromosome as *fmr1* is consistent with R.A. Fisher's assertion that the selective advantage of co-segregation of particular combinations of alleles of genes will favour evolutionary chromosomal rearrangements that place them in linkage disequilibrium on the same chromosome. However, due to the structure of the mating scheme used to produce the batches of homozygous mutant and wild type embryos examined, we cannot exclude that some of the apparent differential gene expression observed reflects the differential expression of different alleles of genes and is unrelated to any effects of the *fmr1*^{hu2787} mutation.

MATERIALS AND METHODS

Collection of Zebrafish Embryos

The *fmr1*^{hu2787} mutation was originally described by den Broeder et al. (2009). All the zebrafish embryos used in this work were descendants of a population of *fmr1*^{hu2787} heterozygous and homozygous zebrafish larvae imported from the laboratory of Howard Sirotkin. Heterozygous individuals selected from this population were crossed to produce homozygous, heterozygous and non-mutant (*fmr1*^{+/+}) progeny. From these progeny, pairs of homozygous and non-mutant fish were mated to produce the clutches of embryos used for transcriptome analysis. Each clutch was derived from a different pair of parents. At 2 dpf, embryos were dechorionated in E3 embryo medium and then chilled on ice for 15 min. The E3 medium was then removed and approximately 100 embryos were added to 5 volumes of RNAlater™ Stabilisation Solution (Invitrogen, Carlsbad, CA, United States) at 4°C, chilled on ice for 1 h, and then stored overnight at 4°C. The tissue was then stored at -20°C until purification of RNA or DNA.

The recessive w2 allele of the gene *mitfa* (*nacre*) (Lister et al., 1999) was subsequently discovered to have been present (in a heterozygous state) in some individuals of the original, imported population and was present in some of the clutches of embryos examined. None of the parents of the clutches used for transcriptome analysis were homozygous for the w2 allele. *Post hoc* analysis showed that the presence of this allele did not greatly affect the transcriptome analysis (see **Supplementary Data File 2**). Some of the *fmr1*^{hu2787/hu2787} parents used to produce embryos for qPCR analysis of relative chromosome copy number were also homozygous for the w2 allele of *mitfa*, but any *mitfa*^{w2/w2} embryos (i.e., lacking melanotic pigmentation) were removed from these clutches before analysis.

RNA Extraction From Larval Clutches

Total RNA was isolated from four *fmr1* homozygous mutant clutches (labelled “S2,” “S4,” “S5,” “S8”) and four wild type clutches (“A,” “D,” “G,” “L”) of 2 dpf embryos using the *mirVana* miRNA isolation kit (Thermo Fisher Scientific Inc., Waltham, MA United States). A sample size of $n = 4$ was chosen based on our experience with transcriptome analyses of zebrafish brains (Newman et al., 2019; Hin et al., 2020a,b). RNA isolation was performed according to the manufacturer's protocol. First, the larval clutch was lysed in a denaturing lysis solution. The lysate was then extracted once with acid-phenol:chloroform leaving a semi-pure RNA sample. The sample was then purified further over a glass-fibre filter to yield total RNA. DNases were removed from the RNA samples using a DNA-free™ kit (Ambion, Austin, TX, United States). To prevent degradation of samples prior to RNA sequencing, RNA was stabilised using RNAsable (Biomatrix, San Diego, CA, United States) as per the manufacturer's protocol. Briefly, RNA was resuspended in DEPC-treated water, applied directly into an RNAsable tube and dried with a vacuum concentrator. Total RNA was then delivered to Novogene (Hong Kong, China) to assess RNA quality and for subsequent RNA sequencing.

RNA-seq Analysis

Strand-specific, paired end (150 bp insert), polyA + library preparation and RNA sequencing (RNA-seq) was conducted by Novogene using the Illumina HiSeq 4000 platform (Illumina Inc., San Diego, CA, United States). Base-calling was performed using Illumina Casava (v1.8). We evaluated the quality of the supplied demultiplexed fastq files by fastQC (Andrews, 2010) and ngsReports (Ward et al., 2019). We then pre-processed the reads before alignment by removing adapters from any reads derived from RNA fragments less than 300 bp, removing bases from the ends of reads when the quality score fell below 20, and discarding reads less than 35 bp in length after trimming of adapters using AdapterRemoval v2.2.1 (Schubert et al., 2016). The remaining reads were then aligned to the zebrafish genome [Ensembl Release 94 (GRCz11)] using STAR v2.5.3a (Dobin et al., 2013). Aligned reads were counted using featureCounts from the Subread package (v1.5.2) (Liao et al., 2014) only if they were unique and mapped to strictly exonic regions.

Gene Differential Expression and Gene Set Enrichment Analyses

Analysis of count data was performed using R (Zhao and Tan, 2006). We only retained genes which had more than 1 count in 4 or more of the 8 RNA-seq libraries, leaving 18,280 genes for downstream analysis. We performed differential gene expression analysis using the generalised linear model capabilities and likelihood-ratio tests from the package edgeR (Robinson et al., 2010), after calculation of normalisation offsets using Conditional Quantile Normalisation (Hansen et al., 2012). We considered genes to be differentially expressed (DE) due to homozygosity for the hu2787 allele of *fmr1* if the false discovery rate (FDR)-adjusted *p*-value was less than 0.05. Complete results of this differential expression analysis are shown in **Supplementary Data File 1**. We tested for over-representation of pre-defined gene sets within the DE genes using goseq (Young et al., 2010), using a weighted average GC content as input for the probability weighting function (PWF). The gene sets we used were the KEGG (v7.1) and HALLMARK (v7.1) gene sets available from MSigDB (Liberzon et al., 2011). We downloaded the gene sets as a .gmt file with human Entrez gene identifiers and converted these to zebrafish Ensembl IDs using a mapping file downloaded from BioMart (Smedley et al., 2015). We also tested whether there was over-representation of genes on any of the 26 chromosomes (i.e. including the mitochondrial chromosome) in the DE gene lists with goseq. We considered a gene set to be over-represented in the DE genes if the Bonferroni-adjusted *p*-value was < 0.05.

To obtain a more complete view on the changes to gene expression due to *fmr1* genotype, we performed the self-contained, fast rotation gene set testing method fry (Wu et al., 2010) on the cqn-adjusted logCPM values.

Implementation of the RNAseq Short Variant Discovery (SNPs + Indels) Workflow

Variant calls were generated for each sample at sites across the genome where there was sufficient evidence for a non-reference nucleotide in at least one sample. Variants were determined

based on the Genome Analysis Toolkit (GATK) best practices workflow (Van der Auwera and O'Connor, 2020) for *RNAseq short variant discovery (SNPs + Indels)*. Briefly, trimmed reads were aligned to the GRCz11 (Ensembl release 94) genome using STAR v2.7.7a (Dobin et al., 2013) two-pass mode to achieve better alignments around novel splice junctions. PCR duplicates were marked using Picard (Broad Institute, (n.d.), see References) such that only independent observations were counted, and reads that spanned introns were split into multiple supplementary alignments with the GATK v4.2 (Van der Auwera and O'Connor, 2020) SplitNCigar tool to allow for downstream processing. Before base quality score recalibration, a set of known variants was generated by the bootstrapping method described in GATK's best practices. This involved an initial round of variant calling with the GATK v4.2 HaplotypeCaller tool on the unrecalibrated alignments. Variants with a minimum phred-scaled confidence threshold of 20 were assigned as the set of known variants for base recalibration, which was achieved with the GATK v4.2 BaseRecalibrator tool. A final round of HaplotypeCaller was run on the recalibrated alignments and the resulting variant calls were filtered based on GATK's recommended specific hard filters (phred-scaled *p*-value using Fisher's exact test for strand bias (FS) > 30, variant confidence/quality by depth (QD) > 2). Lastly, variants were selected for only single nucleotide polymorphisms using the GATK v4.2 SelectVariants tool. See also **Supplementary Data File 4**.

Analyses to Detect Transcriptional Adaptation

We obtained the list of genes with sequences similar to *fmr1* by performing a BLASTn search using the Ensembl BLAST server with the *fmr1* cDNA sequence as input (i.e., the mRNA sequence including untranslated regions). Firstly, we queried this sequence against the zebrafish (GRCz11, Ensembl version 101) reference transcriptome with parameters set for distant homology. Detected homologous sequences with an *E*-value < 1 were tested to determine whether they were upregulated due to *fmr1* genotype. To determine if *fmr1* shares homology with other features shown to influence transcriptional adaptation, such as promoters and introns (El-Brolosy et al., 2019), the *fmr1* cDNA sequence was similarly queried against the entire zebrafish genome (GRCz11, Ensembl version 101). This returned no additional information. An exploration of similarity thresholds was performed similar to that of El-Brolosy et al. (2019) (see Methods section "Sequence similarity and subsampling analyses"). *p*-values were computed by bootstrapping random subsamples. Thresholds were chosen such that each threshold included at least one additional unique alignment. No significant correlations between sequence similarity and upregulation of expression were found at any threshold for any of the three parameters analysed (alignment length, bit score, and *E*-value, see **Supplementary Data File 5**).

DNA Extraction for Relative Standard Curve Quantitative PCR

Three clutches of 2 dpf zebrafish embryos homozygous for *fmr1*^{hu2787} were collected. Each clutch contained approximately

100 embryos. Three batches of around 100 wild type zebrafish embryos were generated from three different tanks of wild type zebrafish families and collected at 2 dpf. Genomic DNA (gDNA) was extracted and purified using the Wizard® genomic DNA purification kit (Promega, Madison, Wisconsin, United States). Genomic DNA (gDNA) concentrations were estimated using a NanoDrop Microvolume Spectrophotometer (Thermo Fisher Scientific). Each 25 μ L qPCR reaction contained 40 ng of gDNA, 0.2 μ M of each PCR primer and Power SYBR green master mix PCR solution (Applied Biosystems, Thermo Fisher Scientific). The relative standard curve method was used for quantification to determine the amounts of *fmr1*, *psen1* and *rpl13* in wild type samples relative to *fmr1*^{hu2787} homozygous samples. The standard curve was generated by a serial dilution, having 40, 20, and 10 ng of wild type gDNA per reaction. The qPCR was performed on an ABI 7000 Sequence Detection System (Applied Biosystems) using a 96-well plate. The amplification consisted of a holding stage and a cycling stage. The holding stage was 50°C for 2 min and then 95°C for 10 min, and the cycling stage had 40 cycles of 95°C for 15 s and 60°C for 1 min. Three technical PCR replicates were performed on each DNA sample and their mean was used to represent the quantity. The quantities of the genes *fmr1* and *psen1* were calculated relative to *rpl13*. Raw quantitative PCR data are shown in **Supplementary Data File 1**. The PCR primers used and the relative chromosome ratios calculated are described in **Supplementary Data File 3**.

RESULTS AND DISCUSSION

Differential Gene Expression in *fmr1*^{hu2787}/*hu2787* vs. +/+ Embryos at 2 dpf

Transcriptome (RNA-seq) analysis was conducted on clutches of wild type and *fmr1*^{hu2787} homozygous mutant embryos ($n = 4$ clutches for each genotype) that had been allowed to develop for 2 dpf at ~28.5°C. Each clutch of embryos was the product of a unique pair mating (i.e., no clutches shared any parents).

Bioinformatic analysis is described in detail in **Supplementary Data File 2** and the results are summarised here. Principal Component Analysis (**Figure 1A**) did not suggest clearly discrete clustering of genotypes which was reflected in the paucity of differentially expressed genes detected (**Figures 1B,C** and see below).

Genes Differentially Expressed at 2 dpf Due to Homozygosity for *fmr1*^{hu2787}

Using a False Discovery Rate (FDR) of 0.05, only 21 genes were detected as differentially expressed (DE, see **Table 1**). Levels of *fmr1* transcripts in the mutant embryos were approximately half of that in the wild type embryos consistent with nonsense-mediated decay (NMD) due to the premature termination codon generated by the *fmr1*^{hu2787} mutation. While NMD has not been demonstrated formally for transcripts of this allele (by stabilisation through inhibition of translation) and neither do we do so here, the location of the stop codon more than

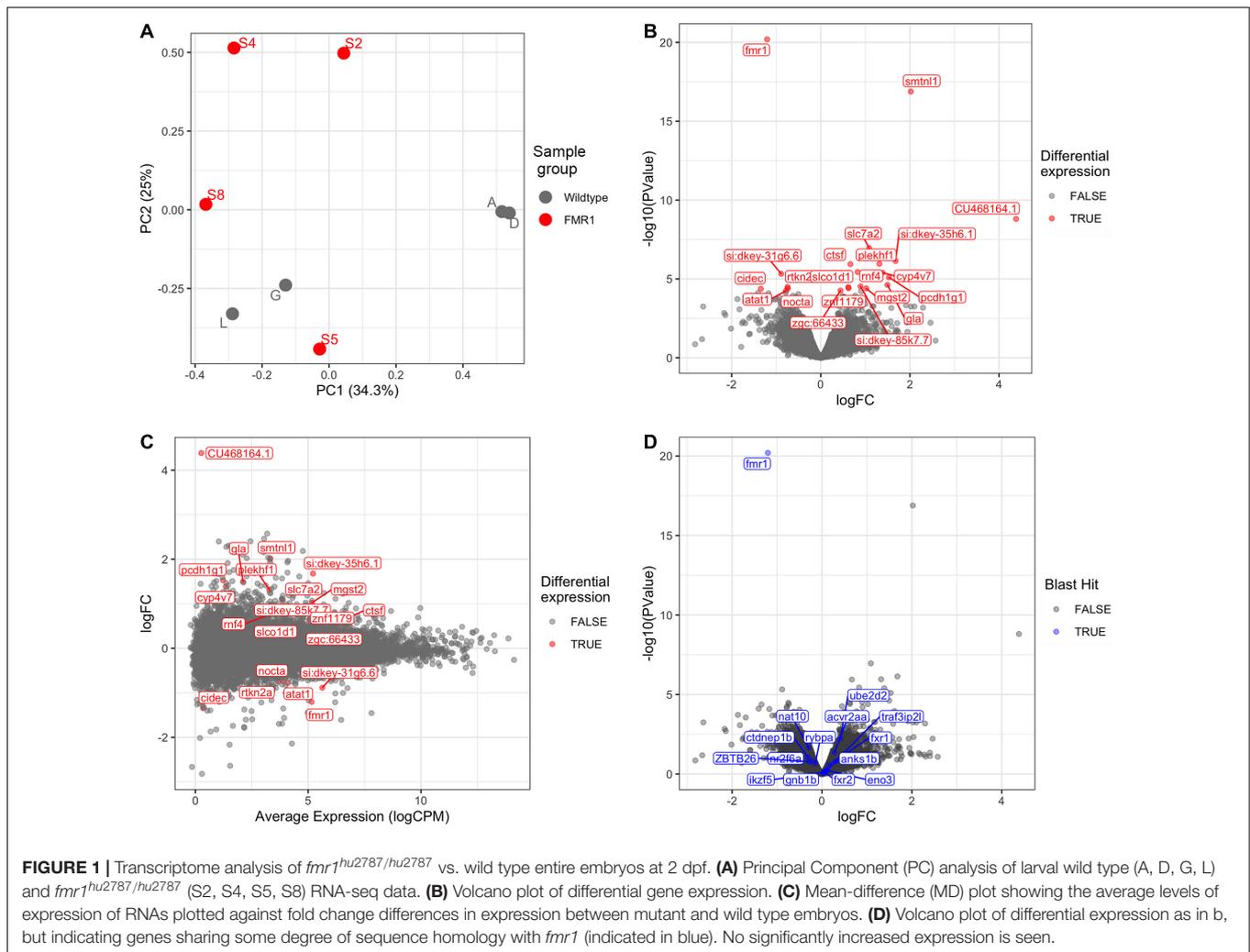
80 nucleotides upstream of the next exon/exon boundary is consistent with NMD activity (Nagy and Maquat, 1998).

Curiously, we observed that 12 of the 21 genes found to be DE at a false discovery rate (FDR) of less than 0.05 are located on zebrafish Chromosome 14. This was found to be a highly significant over-representation by goseq (Bonferroni adjusted p -value $< 2 \times 10^{-16}$, see **Figure 2A** and **Supplementary Data File 2**). This included genes both up- and down-regulated (**Table 1**) at loci widely dispersed on the chromosome and transcribed from either DNA strand (**Figure 2B**). As Chromosome 14 (DRE13 in Phillips et al., 2006) has a p arm/entire chromosome length ratio of 0.27 (Phillips et al., 2006) it is likely that these loci are distributed on both chromosome arms.

One possible explanation for many genes on one chromosome appearing differentially expressed would be if the *fmr1*^{hu2787} mutation was associated with meiotic or mitotic instability of Chromosome 14, although this would not be consistent with the simultaneous up- and down-regulation of Chromosome 14 genes observed. Nevertheless, to test this possibility we conducted quantitative PCR on genomic DNA extracted from an additional three groups each of 2 dpf *fmr1*^{hu2787}/*hu2787* and +/+ embryos. We used PCR primers amplifying genomic DNA sequences from the genes *fmr1* (Chromosome 14), *psen1* (Chromosome 17) and *rpl13* (Chromosome 7). No difference was observed in the amplification of *fmr1* sequence compared to *rpl13* and the amplification of *psen1* sequence relative to *rpl13* in clutches of either *fmr1*^{hu2787}/*hu2787* or +/+ embryos (**Table 2**) supporting that the stability of Chromosome 14 is normal (see also **Supplementary Data File 3**).

Is the Enrichment for DE Genes on Chromosome 14 an Allelic Expression Artefact?

Enrichments for DE genes on the same chromosome as a mutation have been observed previously. For example, in a study of neural crest development in zebrafish, Dooley et al. (2019) saw this for embryos homozygous for mutations in *sox10* (Chromosome 3) or *mitfa* (Chromosome 6). However, they were unable to distinguish whether this was due to reduced genetic variation around the homozygous loci (allowing genes with alleles having divergent levels of expression to be observed as differentially expressed) or due to real regulatory effects of the mutations. To find evidence possibly supporting one or the other of these alternatives in our analysis of *fmr1*^{hu2787}/*hu2787* embryos compared to wild type embryos, we examined allelic diversity within the homozygous mutant and wild type transcriptome data sets. We implemented the “RNAseq short variant discovery (SNPs + Indels)” workflow to identify and quantify exonic single nucleotide polymorphism (SNP) variants in the DE genes on Chromosome 14 and other chromosomes. As expected, this did identify the absence of allelic diversity at Chromosome 14 DE genes proximal to *fmr1* in the *fmr1*^{hu2787}/*hu2787* embryos while some of the Chromosome 14 DE genes more distal to *fmr1* showed greater allelic diversity (see **Supplementary Data File 4**). Greater, although still limited, allelic diversity was seen for all the



Chromosome 14 DE gene transcripts in the wild type embryos and for most of the non-Chromosome 14 DE gene transcripts in both *fmr1*^{hu2787/hu2787} and wild type embryos. Overall, the generally moderate levels of allelic diversity observed for the DE genes of our analysis do not allow us to exclude that many of the genes identified as DE on Chromosome 14 are displaying differences in allelic expression levels rather than responses to mutation of *fmr1*.

Interestingly, transcripts of the non-Chromosome 14 DE genes *cidec* and *plekhf1* showed variation between genotypes but no variation within genotypes supporting the possibility that their apparent differential expression may reflect allelic expression differences rather than responses to mutation of *fmr1* (see **Supplementary Data File 4**).

A function-based explanation for why mutation of *fmr1* might cause enriched differential expression of Chromosome 14 genes may lie in selective pressure during evolution for homeostatic robustness combined with low rates of meiotic recombination along zebrafish chromosomes. Indeed, meiotic recombination is suppressed in male zebrafish so that most chromosomes have recombination distance lengths of less than

50 centimorgans (Singer et al., 2002). This means that, in zebrafish male meiosis, no two genes on Chromosome 14 will show completely independent assortment of alleles. If a fitness advantage existed for co-segregation of particular pairs of alleles of functionally related genes, then chromosomal rearrangements leading to the co-location of those alleles within 50 centimorgans on the same chromosome might be under positive selection during evolution. This phenomenon was first suggested by RA Fisher in 1930 in his book, *The genetical theory of natural selection* (Fisher, 1930). In discussing pairs of alleles (“factors”) of functionally interacting genes in linkage disequilibrium he stated, “. . .the presence of pairs of factors in the same chromosome, the selective advantage of each of which reverses that of the other, will always tend to diminish recombination, and therefore to increase the intensity of linkage in the chromosomes of that species.”

Similar selective forces may have driven the accumulation of genes important for neural function during evolution of the human X-chromosome (for which meiotic recombination is limited in human males) (Skuse, 2005; Nguyen and Disteche, 2006) and within which the human *FMRI* gene resides. This idea is consistent with the observation by Salih and Adelson (2009)

TABLE 1 | Genes differentially expressed in 2 dpf homozygous *fmr1*^{hu2787} embryos.

Ensembl Gene ID	Location	Human chr.	Gene name	logFC	logCPM	P-value	FDR
ENSDARG00000037433	Chromosome 14: 20156477–20191281 (+)	Xq27.3	<i>fmr1</i>	−1.20	5.15	6.41E-21	1.17E-16
ENSDARG000000041257	Chromosome 14: 17108003–17121713 (−)	11q12.1	<i>smtnl1</i>	2.02	3.32	1.25E-17	1.14E-13
ENSDARG000000052609	Chromosome 14: 9056071–9066756 (+)		<i>CU468164.1</i>	4.35	0.26	2.16E-09	1.32E-05
ENSDARG000000037097	Chromosome 14: 30367602–30390145 (−)	8p22	<i>slc7a2</i>	1.09	3.87	1.11E-07	5.09E-04
ENSDARG00000104572	Chromosome 14: 13036560–13048355 (−)		<i>si:dkey-35h6.1</i>	1.68	5.22	7.39E-07	2.70E-03
ENSDARG00000102435	Chromosome 7: 45975537–45976956 (+)	19q12	<i>plekhf1</i>	1.32	3.28	1.09E-06	3.27E-03
ENSDARG000000063095	Chromosome 14: 7140997–7164646 (+)	11q13.2	<i>ctsf</i>	0.66	6.70	1.25E-06	3.27E-03
ENSDARG000000090993	Chromosome 14: 17566383–17576662 (−)	4p16.3	<i>mf4</i>	0.83	3.68	3.66E-06	8.14E-03
ENSDARG000000061585	Chromosome 14: 29975268–29990757 (+)		<i>cyp4v7</i>	1.39	1.35	4.01E-06	8.14E-03
ENSDARG00000101393	Chromosome 6: 18364048–18366339 (−)		<i>si:dkey-31g6.6</i>	−0.88	5.63	5.00E-06	9.13E-03
ENSDARG00000104007	Chromosome 10: 21650828–21820057 (+)		<i>pcdh1g1</i>	1.53	1.23	8.07E-06	1.34E-02
ENSDARG000000036155	Chromosome 14: 38932441–38946808 (−)	Xq22.1	<i>gla</i>	1.50	2.11	2.35E-05	3.58E-02
ENSDARG00000101135	Chromosome 7: 21859337–21874346 (+)		<i>si:dkey-85k7.7</i>	0.89	4.91	2.96E-05	3.84E-02
ENSDARG000000053558	Chromosome 17: 43610894–43623370 (−)	10q21.2	<i>rtkn2a</i>	−0.75	3.67	3.32E-05	3.84E-02
ENSDARG00000104108	Chromosome 4: 77389668–77432218 (−)		<i>slco1d1</i>	0.62	4.35	3.33E-05	3.84E-02
ENSDARG00000104613	Chromosome 22: 1796057–1800444 (+)		<i>znf1179</i>	0.62	5.05	3.87E-05	3.84E-02
ENSDARG000000059651	Chromosome 6: 21202639–21212367 (+)	3p25.3	<i>cidec</i>	−1.35	0.34	3.95E-05	3.84E-02
ENSDARG000000071345	Chromosome 14: 46216703–46220099 (+)	4q31.1	<i>mgst2</i>	1.02	5.17	3.97E-05	3.84E-02
ENSDARG000000077726	Chromosome 14: 46028003–46042136 (+)		<i>nocta</i>	−0.74	3.88	3.99E-05	3.84E-02
ENSDARG000000071697	Chromosome 14: 20918701–21001756 (+)		<i>zgc:66433</i>	0.44	5.32	5.36E-05	4.90E-02
ENSDARG00000004472	Chromosome 19: 27479563–27515651 (+)	6p21.33	<i>atat1</i>	−0.78	4.09	5.69E-05	4.95E-02

Gene positions on zebrafish chromosomes are taken from reference genome assembly GRCz11. Those on Chromosome 14 are highlighted in green. "(+)" denotes the nominal forward strand and "(−)" the reverse strand of chromosomes. Where human orthologous genes are known, their chromosomal locations from GRCh38 are given under "Human chr." Log₂ values are given for Fold Change (FC) and Counts Per Million (CPM).

that particular Gene Ontology (GO) terms, when mapped to genes spanning the bovine genome, appear clustered. The clustering of GOs was subsequently observed in the genomes of other species (Tiirikka et al., 2014). Also, numerous examples exist of genome rearrangements during evolution leading to clustering of functionally related genes (e.g., Qi et al., 2004; Wong and Wolfe, 2005; Hill et al., 2019). There is also evidence that chromosomal rearrangements (inversions) experience natural selection to maintain associations between functionally related alleles (Said et al., 2018).

If limited (but non-zero) allelic diversity in transcriptome analyses of genotypes such as we have performed here can generate significant levels of apparent differential gene expression unrelated to the genotypes of interest, then that is an important constraint on interpretation of transcriptome data. It is important that future work examine the extent to which limited allelic diversity can cause the appearance of such gene differential expression between genotypes.

No Evidence for Increased Expression of Gene Transcripts With Homology to *fmr1*

The phenomenon of transcriptional adaptation may contribute to explaining the frequent disparity between embryo development phenotypes caused by loss of gene function due to inhibition of transcript splicing/translation by morpholino antisense oligonucleotides and loss of gene function due to gene mutation (Kok et al., 2015). El-Brolosy et al. (2019) showed that a transcript subject to NMD could cause increased abundance

of transcripts from genes possessing sequences with homology to that transcript. To test whether transcriptional adaptation might contribute to the reported differences in developmental phenotypes observed between *fmr1*^{hu2787/hu2787} embryos (den Broeder et al., 2009) and embryos injected with morpholinos inhibiting *fmr1* transcript translation (Tucker et al., 2006), we searched the zebrafish genome and transcriptome (GRCz11, Ensembl version 101) for features with sequence homology to *fmr1* cDNA. These are listed in **Supplementary Data File 5**. None of these genes were significantly differentially expressed in our comparison of *fmr1*^{hu2787/hu2787} and +/+ larval clutches (**Figure 1D**).

Testing for the enrichment of groups of genes (gene sets) in transcriptome data sets can provide more sensitive detection of significant changes in expression than examining genes individually. Therefore, we defined as a set the genes identified with homology to *fmr1* and subjected this set to the gene set testing method fry (Wu et al., 2010) (from the limma package Ritchie et al., 2015) with a directional hypothesis (increased transcript abundance). This gave a *p*-value for enrichment of 0.6 which does not support that the expression of this set of genes is increased.

Our analyses have not been able to detect transcriptional adaptation due to homozygosity for the *fmr1*^{hu2787} allele. However, we cannot exclude that NMD of *fmr1*^{hu2787} allele transcripts might induce this phenomenon. It is possible that the restricted expression of *fmr1* in 2 dpf embryos limits the sensitivity of our analyses to detect differential expression of genes with homology to *fmr1* by diluting their relative

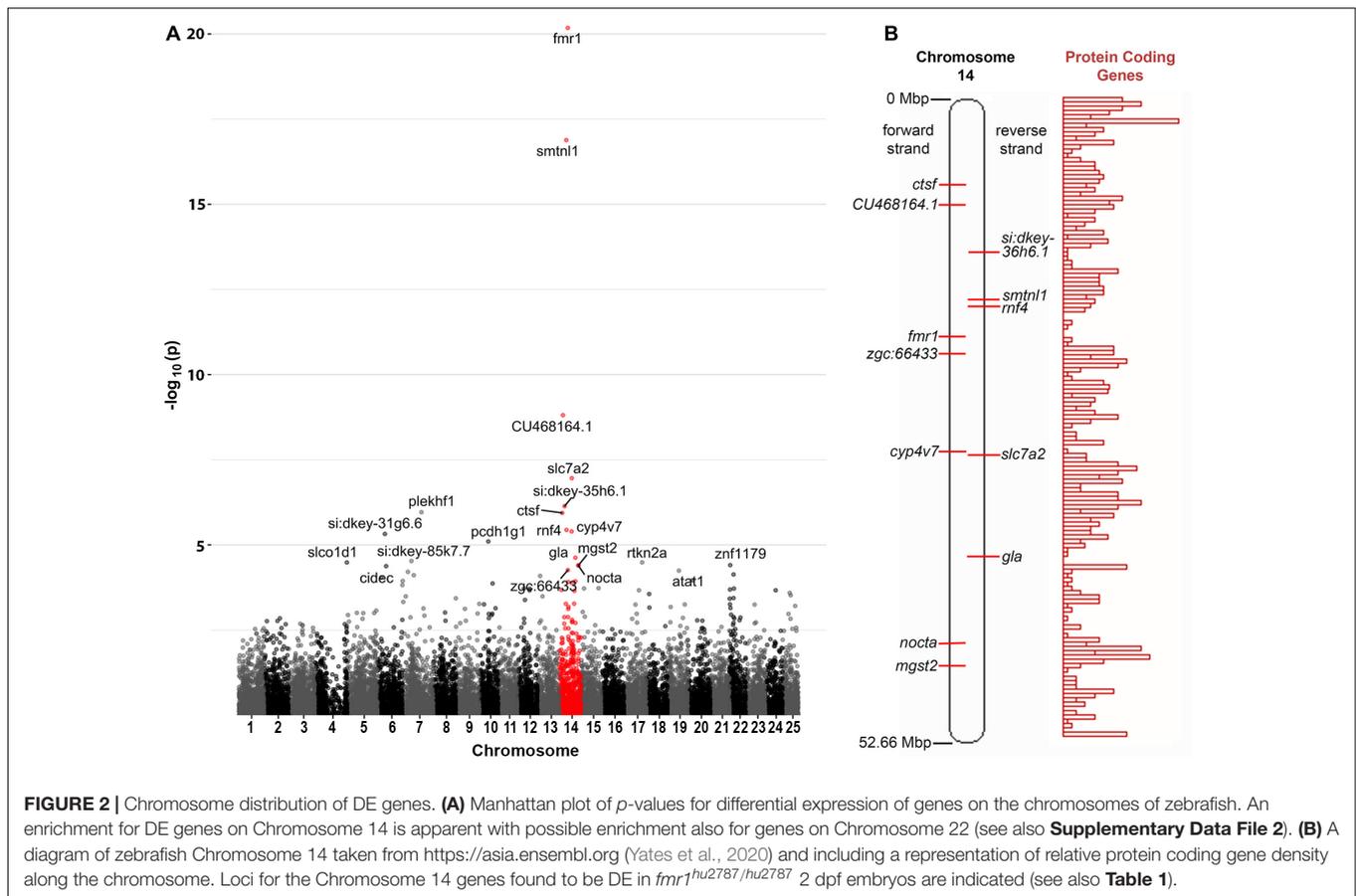


TABLE 2 | PCR primer pairs used to quantify relative genomic DNA copy number.

Gene	Chromosomal location (ENSEMBL)	PCR primer sequences	Mean ratio gene/ <i>rpl13</i>	
			+/+	<i>fmr1/fmr1</i>
<i>fmr1</i>	Chromosome 14: 20,156,477–20,191,281 forward strand.	5'-TCCAGGACCAGGAGGCTGTA-3' 5'-CCCACCTCTTGCCAATGACTTTTC-3'	0.99	0.98
<i>psen1</i>	Chromosome 17: 51,206,225–51,224,800 reverse strand.	5'-CTACACACAGAAGGACGGACAGC-3' 5'-CCATCCCTAAACTGCTCCTACT-3'	0.98	0.93
<i>rpl13</i>	Chromosome 7: 67,467,702–67,477,495 forward strand.	5'-TGGAATGGATGAATAGGTTTTTA-3' 5'-CTCTCTTCTGCCAGTCTTTATGA-3'	1	1

abundance with transcripts from cells not expressing *fmr1*. Alternatively, the transcriptional adaptation phenomenon may not function for the *fmr1* gene at this particular developmental stage, although this seems unlikely considering the supposed generality of the phenomenon.

Interestingly, unlike for the *fmr1*^{hu2787} allele, a recently published loss-of-function mutation in *fmr1* induced using the CRISPR-Cas9 system and that deletes the translation start codon does produce craniofacial abnormalities (Hu et al., 2020). However, these abnormalities appear more subtle than reported by Tucker et al. (2006) for morpholino inhibition of *fmr1* translation. Also, in work submitted for publication, careful re-examination of craniofacial morphology in larvae homozygous for the *fmr1*^{hu2787} mutation itself does, in fact,

reveal structural differences compared to wild type larvae (Zhu et al., 2021). Therefore, the differences between the phenotypic effects of the *fmr1*^{hu2787} mutation and injection of morpholinos blocking *fmr1* mRNA translation may not be as extensive as originally believed.

Prediction of Cellular Functions Affected by Homozygosity for *fmr1*^{hu2787}

To gain insight into the cellular functions that might be affected by the *fmr1*^{hu2787} mutation we performed a functional analysis within the set of DE genes using goseq (Young et al., 2010) with the HALLMARK and KEGG gene sets as defined in MSigDBv7.2 via the R package msigdb

(Dolgalev, 2020). The requisite probability weighting function (PWF), for the probability of a gene being considered DE, was estimated using the average GC content across all transcripts of a gene. We found that the KEGG gene sets for *lysosome* and *glycosphingolipid biosynthesis globo series* were significantly over-represented in the DE genes, although only one or two DE genes were driving this over-representation. The apparent effect of zebrafish *fmr1* mutation on lysosomal function is consistent with observations that loss of *Fmr1* gene function in mice affects lysosome function (Yan et al., 2018).

We also performed the gene set testing method fry using chromosome position, HALLMARK and KEGG gene sets to obtain a more complete view of the changes to gene expression due to *fmr1* genotype. No significant gene sets were observed. However, the genes on Chromosome 14 were those closest to being significantly altered as a group with an FDR adjusted *p*-value of 0.16.

We investigated further the possible functional significance of changes to expression of genes from Chromosome 14 by performing an over-representation analysis of those genes driving the enrichment of the set of Chromosome 14 genes (the “leading edge” genes from gene set enrichment analysis, GSEA, see **Supplementary Data File 2**). Relative to all the detected genes in the RNA-seq analysis, the leading edge Chromosome 14 genes were significantly over-represented in the KEGG gene sets *neuroactive ligand receptor interactions* and *RNA polymerase*. Interestingly, the *neuroactive ligand receptor interactions* gene set was previously observed to be enriched in the cerebellum of 8–10 week old male *Fmr1* knockout mice compared to wild type mice (Kong et al., 2014) and this is consistent with FXS as a developmental neuropathology. The enrichment of genes from this gene set among the leading edge genes expressed from zebrafish Chromosome 14 suggests that their co-location on this chromosome may reflect selective pressure during evolution for linkage disequilibrium between advantageous alleles of functionally related genes.

Limitations

Our ability to identify a number of differentially expressed genes, some at very small FDR corrected *p*-values despite moderate fold changes in expression, supports our confidence that we would have been able to detect transcriptional adaptation (increased abundance of transcripts from genes possessing homologous sequences) if it was induced by the *hu2787* allele of *fmr1* at 2 dpf. However, the small number of gene sets identified as statistically significantly enriched in *fmr1^{hu2787}* homozygous mutants may reflect the limited sample size used. It is also conceivable that transcriptional adaptation effects on transcript abundance are dependent on developmental stage and that we did not examine the transcriptome at an appropriate time to observe these effects. Future studies using larger sample sizes and examining a range of developmental stages would give improved definition of enriched gene sets (suggesting altered cellular functions) and would also likely identify additional

genes on Chromosome 14 with expression affected by mutation of *fmr1*.

CONCLUSION

Despite the apparent NMD of *fmr1^{hu2787}* transcripts, and the reported milder developmental phenotype of *fmr1^{hu2787}* homozygotes relative to individuals in which the function of this gene is suppressed using morpholinos, we did not see evidence for transcriptional adaptation by increased transcription of genes possessing sequences with homology to *fmr1*, at least at 2 dpf. This illustrates possible variability of the occurrence of the recently discovered transcriptional adaptation mechanism, and that more research is required to understand the factors modulating it. Our analysis has revealed the accumulation on zebrafish Chromosome 14 of genes for which expression is affected by mutation of *fmr1*. This may be the result of selection over evolutionary time for robust cellular homeostasis together with linkage between particular combinations of alleles of functionally related genes. Loss of *fmr1* gene function in the zebrafish *fmr1^{hu2787}* model produces similar effects on the brain transcriptome to loss of the orthologous gene in mice, and supports the translatability to humans of discoveries regarding *fmr1* function made using zebrafish.

DATA AVAILABILITY STATEMENT

All data described in this article are available within the article itself, or in the **Supplementary Data Files**. The RNA-seq data are available at the Gene Expression Omnibus (GEO) database under Accession Number GSE151443. The code used to perform this analysis is available at https://github.com/UofABioinformaticsHub/20190129_Lardelli_FMR1_RNASeq and at https://github.com/baerlachlan/210305_fmr1_snv.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Ethics Committee of the University of Queensland. Permit Number QBI/061/17.

AUTHOR CONTRIBUTIONS

ML and RIR conceived the study. GJG administered, and ZP performed, zebrafish breeding for embryo collection, and preservation. MN prepared RNA for, and administered, RNA-seq analysis. YD designed, performed, and analysed qPCR. SMP supervised all bioinformatics analysis. MH and LB identified genes with sequences possessing *fmr1* homology and performed initial DE gene analysis. LB performed further tests to detect transcriptional adaptation and the analysis of allelic diversity in differentially expressed genes. KB performed

all other bioinformatics analysis. All authors participated in drafting the manuscript.

FUNDING

This research was supported by the funds from the School of Biological Sciences of the University of Adelaide and by a grant from the Simons Foundation (SFARI 569051, GJG). KB and LB were supported by the Australian Government Research Training Program Scholarships. YD was supported by the Adelaide Graduate Research Scholarship from the University of Adelaide. No funding body played any part in the design of this study, or in the collection, analysis, or interpretation of data, or in the writing of the manuscript.

REFERENCES

- Andrews, S. (2010). FASTQC. A Quality Control Tool for High Throughput Sequence Data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Broad Institute, (n.d.) *Picard, A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF*. Broad Institute, GitHub repository. Available online at: <http://broadinstitute.github.io/picard/> (accessed March 10, 2021).
- Brouwer, J. R., Mientjes, E. J., Bakker, C. E., Nieuwenhuizen, I. M., Severijnen, L. A., Van Der Linde, H. C., et al. (2007). Elevated FMR1 mRNA levels and reduced protein expression in a mouse model with an unmethylated fragile X full mutation. *Exp. Cell Res.* 313, 244–253. doi: 10.1016/j.yexcr.2006.10.002
- Chen, S. Y., Zhu, Y. X., Xia, W. M., Xia, S. R., and Xu, X. Y. (2011). Automated analysis of zebrafish images for phenotypic changes in drug discovery. *J. Neurosci. Methods* 200, 229–236. doi: 10.1016/j.jneumeth.2011.06.015
- Constantin, L., Poulsen, R. E., Scholz, L. A., Favre-Bulle, I. A., Taylor, M. A., Sun, B., et al. (2020). Altered brain-wide auditory networks in a zebrafish model of fragile X syndrome. *BMC Biol.* 18:125. doi: 10.1186/s12915-020-00857-6
- Dahlhaus, R. (2018). Of men and mice: modeling the fragile X syndrome. *Front. Mol. Neurosci.* 11:41. doi: 10.3389/fnmol.2018.00041
- den Broeder, M. J., Van Der Linde, H., Brouwer, J. R., Oostra, B. A., Willemsen, R., and Ketting, R. F. (2009). Generation and characterization of FMR1 knockout zebrafish. *PLoS One* 4:e7910. doi: 10.1371/journal.pone.0007910
- Dinday, M. T., and Baraban, S. C. (2015). Large-Scale phenotype-based antiepileptic drug screening in a zebrafish model of dravet syndrome. *eNeuro*. 2. doi: 10.1523/ENEURO.0068-15.2015 [Epub ahead of print].
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Dolgalev, I. (2020). *msigdb: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format*.
- Dooley, C. M., Wali, N., Sealy, I. M., White, R. J., Stemple, D. L., Collins, J. E., et al. (2019). The gene regulatory basis of genetic compensation during neural crest induction. *PLoS Genet.* 15:e1008213. doi: 10.1371/journal.pgen.1008213
- El-Brolosy, M. A., Kontarakis, Z., Rossi, A., Kuenne, C., Gunther, S., Fukuda, N., et al. (2019). Genetic compensation triggered by mutant mRNA degradation. *Nature* 568, 193–197. doi: 10.1038/s41586-019-1064-z
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: The Clarendon press.
- Goel, A., Cantu, D. A., Guilfoyle, J., Chaudhari, G. R., Newadkar, A., Todisco, B., et al. (2018). Impaired perceptual learning in a mouse model of fragile X syndrome is mediated by parvalbumin neuron dysfunction and is reversible. *Nat. Neurosci.* 21, 1404–1411. doi: 10.1038/s41593-018-0231-0

ACKNOWLEDGMENTS

The line of *fmr1*^{hu2787} fish used in this work is kindly provided by Howard Sirotkin (The mutation is now available through the Zebrafish International Resource Center, ZIRC). We wish to thank Prof. David Adelson for referral to the work of RA Fisher.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.625466/full#supplementary-material>

- Hammond, L. S., Macias, M. M., Tarleton, J. C., and Shashidhar Pai, G. (1997). Fragile X syndrome and deletions in FMR1: new case and review of the literature. *Am. J. Med. Genet.* 72, 430–434. doi: 10.1002/(sici)1096-8628(19971112)72:4<430::aid-ajmg11>3.0.co;2-s
- Hansen, K. D., Irizarry, R. A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* 13, 204–216. doi: 10.1093/biostatistics/kxr054
- Heulens, I., Suttie, M., Postnov, A., De Clerck, N., Perrotta, C. S., Mattina, T., et al. (2013). Craniofacial characteristics of fragile X syndrome in mouse and man. *Eur. J. Hum. Genet.* 21, 816–823. doi: 10.1038/ejhg.2012.265
- Hill, J., Rastas, P., Hornett, E. A., Neethiraj, R., Clark, N., Morehouse, N., et al. (2019). Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Sci. Adv.* 5:eaa3648. doi: 10.1126/sciadv.aau3648
- Hin, N., Newman, M., Kaslin, J., Douek, A. M., Lumsden, A., Nik, S. H. M., et al. (2020a). Accelerated brain aging towards transcriptional inversion in a zebrafish model of the K115fs mutation of human PSEN2. *PLoS One* 15:e0227258. doi: 10.1371/journal.pone.0227258
- Hin, N., Newman, M., Pederson, S., and Lardelli, M. (2020b). Iron Responsive Element (IRE)-mediated responses to iron dyshomeostasis in Alzheimer's disease. *bioRxiv* [Preprint] doi: 10.1101/2020.05.01.071498
- Hu, J., Chen, L., Yin, J., Yin, H., Huang, Y., and Tian, J. (2020). Hyperactivity, memory defects, and craniofacial abnormalities in zebrafish FMR1 mutant larvae. *Behav. Genet.* 50, 152–160. doi: 10.1007/s10519-020-09995-7
- Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., and Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Dev. Dyn.* 203, 253–310. doi: 10.1002/aja.1002030302
- Kok, F. O., Shin, M., Ni, C. W., Gupta, A., Grosse, A. S., Van Impel, A., et al. (2015). Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev. Cell* 32, 97–108. doi: 10.1016/j.devcel.2014.11.018
- Kong, S. W., Sahin, M., Collins, C. D., Wertz, M. H., Campbell, M. G., Leech, J. D., et al. (2014). Divergent dysregulation of gene expression in murine models of fragile X syndrome and tuberous sclerosis. *Mol. Autism* 5:16. doi: 10.1186/2040-2392-5-16
- Kontarakis, Z., and Stainier, D. Y. R. (2020). Genetics in light of transcriptional adaptation. *Trends Genet.* 36, 926–935. doi: 10.1016/j.tig.2020.08.008
- Kooy, R. F., D'hooge, R., Reyniers, E., Bakker, C. E., Nagels, G., De Boulle, K., et al. (1996). Transgenic mouse model for the fragile X syndrome. *Am. J. Med. Genet.* 64, 241–245.
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656
- Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740. doi: 10.1093/bioinformatics/btr260

- Lister, J. A., Robertson, C. P., Lepage, T., Johnson, S. L., and Raible, D. W. (1999). *nacre* encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. *Development* 126, 3757–3767.
- Marsden, K. C., Jain, R. A., Wolman, M. A., Echeverry, F. A., Nelson, J. C., Hayler, K. E., et al. (2018). A *Cyfp2*-dependent excitatory interneuron pathway establishes the innate startle threshold. *Cell Rep.* 23, 878–887. doi: 10.1016/j.celrep.2018.03.095
- Nagy, E., and Maquat, L. E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* 23, 198–199. doi: 10.1016/s0968-0004(98)01208-0
- Newman, M., Hin, N., Pederson, S., and Lardelli, M. (2019). Brain transcriptome analysis of a familial Alzheimer's disease-like mutation in the zebrafish presenilin 1 gene implies effects on energy production. *Mol. Brain* 12:43.
- Ng, M. C., Yang, Y. L., and Lu, K. T. (2013). Behavioral and synaptic circuit features in a zebrafish model of fragile X syndrome. *PLoS One* 8:e51456. doi: 10.1371/journal.pone.0051456
- Nguyen, D. K., and Distèche, C. M. (2006). High expression of the mammalian X chromosome in brain. *Brain Res.* 1126, 46–49. doi: 10.1016/j.brainres.2006.08.053
- Owens, K. M., Dohany, L., Holland, C., Dare, J., Mann, T., Settler, C., et al. (2018). FMR1 premutation frequency in a large, ethnically diverse population referred for carrier testing. *Am. J. Med. Genet. A* 176, 1304–1308. doi: 10.1002/ajmg.a.38692
- Phillips, R. B., Amores, A., Morasch, M. R., Wilson, C., and Postlethwait, J. H. (2006). Assignment of zebrafish genetic linkage groups to chromosomes. *Cytogenet. Genome Res.* 114, 155–162. doi: 10.1159/000093332
- Pieretti, M., Zhang, F. P., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., et al. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 66, 817–822. doi: 10.1016/0092-8674(91)90125-i
- Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourn, A. (2004). A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8233–8238. doi: 10.1073/pnas.0401301101
- Quan, F., Zonana, J., Gunter, K., Peterson, K. L., Magenis, R. E., and Popovich, B. W. (1995). An atypical case of fragile X syndrome caused by a deletion that includes the FMR1 gene. *Am. J. Hum. Genet.* 56, 1042–1051.
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Holper, S., Kruger, M., et al. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* 524, 230–233. doi: 10.1038/nature14580
- Said, I., Byrne, A., Serrano, V., Cardeno, C., Vollmers, C., and Corbett-Detig, R. (2018). Linked genetic variation and not genome structure causes widespread differential expression associated with chromosomal inversions. *Proc. Natl. Acad. Sci. U.S.A.* 115, 5492–5497. doi: 10.1073/pnas.1721275115
- Salcedo-Arellano, M. J., Dufour, B., McLennan, Y., Martinez-Cerdeno, V., and Hagerman, R. (2020). Fragile X syndrome and associated disorders: clinical aspects and pathology. *Neurobiol. Dis.* 136:104740. doi: 10.1016/j.nbd.2020.104740
- Salih, H., and Adelson, D. L. (2009). QTL global meta-analysis: are trait determining genes clustered? *BMC Genomics* 10:184. doi: 10.1186/1471-2164-10-184
- Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9:88. doi: 10.1186/s13104-016-1900-2
- Shamay-Ramot, A., Khmermesh, K., Porath, H. T., Barak, M., Pinto, Y., Wachtel, C., et al. (2015). FMRP interacts with adar and regulates RNA editing, synaptic density and locomotor activity in zebrafish. *PLoS Genet.* 11:e1005702. doi: 10.1371/journal.pgen.1005702
- Singer, A., Perlman, H., Yan, Y., Walker, C., Corley-Smith, G., Brandhorst, B., et al. (2002). Sex-specific recombination rates in zebrafish (*Danio rerio*). *Genetics* 160, 649–657.
- Skuse, D. H. (2005). X-linked genes and mental functioning. *Hum. Mol. Genet.* 14 Spec No 1, R27–R32.
- Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., et al. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* 43, W589–W598.
- Sutherland, G. R., Haan, E. A., Kremer, E., Lynch, M., Pritchard, M., Yu, S., et al. (1991). Hereditary unstable DNA: a new explanation for some old genetic questions? *Lancet* 338, 289–292. doi: 10.1016/0140-6736(91)90426-p
- The Dutch-Belgian Fragile X Consortium, Bakker, C. E., Verheij, C., Willemsen, R., van der Helm, R., Oerlemans, F., et al. (1994). *Fmr1* knockout mice: a model to study fragile X mental retardation. *Cell* 78, 23–33.
- Thisse, B., and Thisse, C. (2004). Fast Release Clones: A High Throughput Expression Analysis, *ZFIN Direct Data Submission*. Available online at: www.zfin.org
- Tiirikka, T., Siermala, M., and Vihinen, M. (2014). Clustering of gene ontology terms in genomes. *Gene* 550, 155–164. doi: 10.1016/j.gene.2014.06.060
- Tucker, B., Richards, R. I., and Lardelli, M. (2006). Contribution of mGluR and FMR1 functional pathways to neurite morphogenesis, craniofacial development and fragile X syndrome. *Hum. Mol. Genet.* 15, 3446–3458. doi: 10.1093/hmg/ddl422
- Van der Auwera, G. A., and O'Connor, B. D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. Newton, MA: O'Reilly Media, Inc.
- van't Padje, S., Engels, B., Blondin, L., Severijnen, L. A., Verheijen, F., Oostra, B. A., et al. (2005). Characterisation of FMRP in zebrafish: evolutionary dynamics of the FMR1 gene. *Dev. Genes Evol.* 215, 198–206. doi: 10.1007/s00427-005-0466-0
- Ward, C. M., To, T.-H., and Pederson, S. M. (2019). ngsReports: a bioconductor package for managing FastQC reports and other NGS related log files. *Bioinformatics* 36, 2587–2588. doi: 10.1093/bioinformatics/btz937
- Wiley, D. S., Redfield, S. E., and Zon, L. I. (2017). Chemical screening in zebrafish for novel biological and therapeutic discovery. *Methods Cell Biol.* 138, 651–679. doi: 10.1016/bs.mcb.2016.10.004
- Wong, S., and Wolfe, K. H. (2005). Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat. Genet.* 37, 777–782. doi: 10.1038/ng1584
- Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M. L., Visvader, J. E., and Smyth, G. K. (2010). ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* 26, 2176–2182. doi: 10.1093/bioinformatics/btq401
- Yan, J., Porch, M. W., Court-Vazquez, B., Bennett, M. V. L., and Zukin, R. S. (2018). Activation of autophagy rescues synaptic and cognitive deficits in fragile X mice. *Proc. Natl. Acad. Sci. U.S.A.* 115, E9707–E9716.
- Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., et al. (2020). *Ensembl 2020*. *Nucleic Acids Res.* 48, D682–D688.
- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11:R14.
- Zhao, J. H., and Tan, Q. (2006). Integrated analysis of genetic data with R. *Hum. Genomics* 2, 258–265. doi: 10.1186/1479-7364-2-4-258
- Zhu, S., McCullough, M., Pujic, Z., Sibberas, J., Sun, B., Bucknall, B., et al. (2021). *FMR1* mutation interacts with sensory experience to alter the early development of behavior and sensory coding in zebrafish. *bioRxiv* [Preprint] doi: 10.1101/2021.03.16.435742

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Barthelson, Baer, Dong, Hand, Pujic, Newman, Goodhill, Richards, Pederson and Lardelli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.