RESEARCH ARTICLE

mRNA processing in mutant zebrafish lines generated by chemical and CRISPR-mediated mutagenesis produces unexpected transcripts that escape nonsense-mediated decay

Jennifer L. Anderson[1], Timothy S. Mulligan[1], Meng-Chieh Shen[1], Hui Wang[2,3], Catherine M. Scahill[4], Frederick J. Tan[1], Shao J. Du[2], Elisabeth M. Busch-Nentwich[4,5], Steven A. Farber[1]*

1 Carnegie Institution for Science, Department of Embryology, Baltimore, Maryland, United States of America, 2 University of Maryland School of Medicine, Department of Biochemistry and Molecular Biology, Baltimore, Maryland, United States of America, 3 College of Animal Science and Technology, Shandong Agricultural University, Tai'an, China, 4 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, United Kingdom, 5 Department of Medicine, University of Cambridge, Cambridge, United Kingdom

* farber@carnegiescience.edu

Abstract

As model organism-based research shifts from forward to reverse genetics approaches, largely due to the ease of genome editing technology, a low frequency of abnormal phenotypes is being observed in lines with mutations predicted to lead to deleterious effects on the encoded protein. In zebrafish, this low frequency is in part explained by compensation by genes of redundant or similar function, often resulting from the additional round of teleost-specific whole genome duplication within vertebrates. Here we offer additional explanations for the low frequency of mutant phenotypes. We analyzed mRNA processing in seven zebrafish lines with mutations expected to disrupt gene function, generated by CRISPR/Cas9 or ENU mutagenesis methods. Five of the seven lines showed evidence of altered mRNA processing: one through a skipped exon that did not lead to a frame shift, one through nonsense-associated splicing that did not lead to a frame shift, and three through the use of cryptic splice sites. These results highlight the need for a methodical analysis of the mRNA produced in mutant lines before making conclusions or embarking on studies that assume loss of function as a result of a given genomic change. Furthermore, recognition of the types of adaptations that can occur may inform the strategies of mutant generation.

Author summary

The recent rise of reverse genetic, gene targeting methods has allowed researchers to readily generate mutations in any gene of interest with relative ease. Should these mutations have the predicted effect on the mRNA and encoded protein, we would expect many

more abnormal phenotypes than are typically being seen in reverse genetic screens. Here we set out to explore some of the reasons for this discrepancy by studying seven separate mutations in zebrafish. We present evidence that thorough cDNA sequence analysis is a key step in assessing the likelihood that a given mutation will produce hypomorphic or null alleles. This study reveals that mRNA processing in the mutant background often produces transcripts that escape nonsense-mediated decay, thereby potentially preserving gene function. By understanding the ways that cells avoid the deleterious consequences of mutations, researchers can better design reverse genetic strategies to increase the likelihood of gene disruption.

Introduction

The recent increased use of reverse genetic approaches has been largely driven by the ease, affordability of construction, and implementation of the CRISPR/Cas9 and TALEN systems. Recent communications recount numerous cases of generated mutations in genes of interest lacking an expected effect on phenotypes [1, 2]. The shift from antisense-based knockdown (morpholinos, RNAi) to mutant generation (gene targeting/TILLING methods) resulted in discrepancies in phenotypes, leading researchers to question the specificity and mechanisms of anti-sense technologies and also the methods by which mutants are generated [3]. A screen for essential genes performed in a human cultured cell line found little correlation between genes identified with short hairpin RNA (shRNA) silencing and CRISPR/Cas9 methods [4]. While genome editing methods, such as the CRISPR/Cas9 and TALEN systems, have proven to be an efficient and effective way to reduce or eliminate gene function, a frequent lack of a mutant phenotype is observed, often explained by genetic compensation. This is a process wherein related genes or pathway members are differentially regulated in the mutants to compensate for the targeted loss of a specific gene [3].

In addition to genetic compensation, other mechanisms to recover the function of genes harboring homozygous mutations involve variations in processing of mRNA. For example, variations in essential splice sites (ESS) in humans often lead to loss of function resulting in disease [5, 6]; however, there are several well described ways that function may be recovered [7–9]. In canonical pre-mRNA splicing, joining exons for a functional product requires the presence of a 5' splice donor sequence (intronic GU), a branchpoint adenosine, the polypyrimidine tract, and a splice acceptor sequence (intronic AG). Base variations in the ESSs lead to one of four outcomes, in order of frequency: 1) exon skipping, 2) activation of cryptic splice sites, 3) activation of cryptic start sites producing a pseudo-exon within the intron or 4) intron inclusion, in the case of short or terminal introns [10]. Mutations in ESSs that lead to skipped exons may result in transcripts that escape nonsense-mediated decay (NMD), the surveillance system that reduces errors in gene expression, if the exon skip does not lead to a frame shift and premature translation termination codon (PTC) [11]. Cryptic splice sites are present throughout the genome both by chance and through evolution of introns [12] and their activation and use by splicing machinery is typical when exon definitions (such as the natural splice sequences) have been altered [13, 14]. Depending on the location of the cryptic splice site used, and the impact on the sequence and frame, functional transcripts may still be generated.

Nonsense-associated alternative splicing, in which a PTC-containing exon is skipped, may also restore the reading frame of a mutated gene [8]. Again, if the exon skip does not lead to a frame shift and new PTC, and the skipped exon does not contain essential motifs, transcripts may be generated that escape nonsense-mediate decay. Location of the PTC also determines

whether the nascent transcripts will be subject to NMD [15]; however, even though these transcripts escape the surveillance system that detects PTCs, these transcripts may either be functional or aberrant. Translation of the transcripts may result in wildtype or deleterious function.

Recently, discussions on how to produce successful knockout models have been renewed [16–18]. To better inform the generation of future mutations, this report analyzes the genetic consequences of several chemical- and CRISPR-induced zebrafish mutant lines in depth. Zebrafish are amenable to current genome editing methods [19, 20] and are a well-established vertebrate model routinely used to assign functions to genes through the use of classical genetic approaches [21]. To begin to investigate the type and frequency of adaptations that may lead to unexpected splicing in the context of mutation, we carried out studies of mutant lines that included measuring transcript levels and analyzing mRNA splicing (cDNA sequence) and genomic sequence for the presence and use of cryptic splice or start sites. Of the seven mutant lines presented in this study, we show five examples that result in altered mRNA processing. Our findings emphasize the need to analyze putative mutant lines at the level of the mRNA sequence and not assume that a mutation will have the predicted effect on mRNA and/or loss of function.

Results

To begin to investigate how a functional product could be made from a gene containing a putative loss-of-function mutation, we randomly selected seven mutant zebrafish lines generated through chemical- or CRISPR-mediated mutagenesis to study (Table 1). Six zebrafish lines carrying mutations in genes involved in lipid metabolism were obtained through the Zebrafish Mutation Project (ZMP, Wellcome Trust Sanger Institute) ($abca1a^{sa9624}$, $abca1b^{sa18382}$, $cd36^{sa9388}$, $creb3l3a^{sa18218}$, $pla2g12b^{sa659}$, and $slc27a2a^{sa30701}$). Lines were generated with point mutations throughout the genome using classical ENU mutagenesis [22], followed by association of the induced mutations with protein-coding genes using whole exome sequencing methods [1]. We selected five lines that had mutations in essential splice sites and one line with a nonsense mutation ($creb3l3a^{sa18218}$), with the aim to investigate the genome's ability to compensate for induced mutations through the generation of novel alternative transcripts. In addition to the 6 ZMP lines with ENU-induced base changes, a line with a CRISPR/Cas9-generated deletion was included. cDNA sequence and transcript levels of pooled wildtype and homozygous mutant larvae were analyzed.

Two of the five ZMP lines with ESS mutations lose the predicted exon

To determine whether each ZMP line containing an ESS mutation results in the predicted skipping of an exon, adult heterozygote mutant zebrafish were incrossed and their offspring

Table 1. Mutations in this study.

Gene mutated	Ensembl ID	Allele	Nature of mutation	Method of mutation
abca1b	ENSDARG00000079009	sa18382	point mutation in ESS	ENU
slc27a2a	ENSDARG00000036237	sa30701	point mutation in ESS	ENU
abca1a	ENSDARG00000074635	sa9624	point mutation in ESS	ENU
cd36	ENSDARG00000032639	sa9388	point mutation in ESS	ENU
pla2g12b	ENSDARG00000015662	sa659	point mutation in ESS	ENU
creb3l3a	ENSDARG00000056226	sa18218	nonsense mutation	ENU
smyd1a	ENSDARG00000009280	mb4	7-bp deletion in exon	CRISPR/Cas9

https://doi.org/10.1371/journal.pgen.1007105.t001

were pooled or individually processed into total RNA and cDNA. PCR amplification of this cDNA template revealed amplicons of the expected size from each ZMP line, while *abca1b*[sa18382] (ATP-binding cassette transporter, sub-family A, member 1B) and *slc27a2a*[sa30701] (solute carrier family 27, member 2a; protein is fatty acid transport protein, member 2a) also had a shorter amplicon (116 and 210 bp shorter, respectively; S1 Fig) that matches the predicted length of amplicons of these cDNAs after the omission of the affected exon. The *pla2g12b*[sa659] (phospholipase A2, group XIIB) mutant allele has a mutation in the essential splice acceptor site preceding its final (fourth) exon and could not be investigated for the loss of that exon using these methods.

To confirm that exons are skipped in *abca1b*[sa18382] and *slc27a2a*[sa30701] and determine why the mutations in *abca1a*[sa9624] (ATP-binding cassette transporter, sub-family A, member 1A) and *cd36*[sa9388] (cluster of differentiation 36, aka fatty acid translocase) did not appear to lead to the predicted skipping of exons (S2 Fig), individual 6-dpf larvae underwent genotyping, gDNA and cDNA sequence analysis, and qPCR studies.

An ESS mutation in *abca1b* leads to a skipped exon and early termination signal

abca1b[sa18382] has a point mutation in the essential splice acceptor site of intron 33–34 (g.64427G>T) (Fig 1). To determine whether the point mutation results in skipping of the subsequent exon (e34) and use of the (next) essential splice acceptor site of intron 34–35, we performed PCR amplification using primers targeted to flanking exons of cDNA (synthesized from individual, genotyped larvae; 713-bp amplicon), followed by Sanger sequencing. As expected, cDNA sequencing
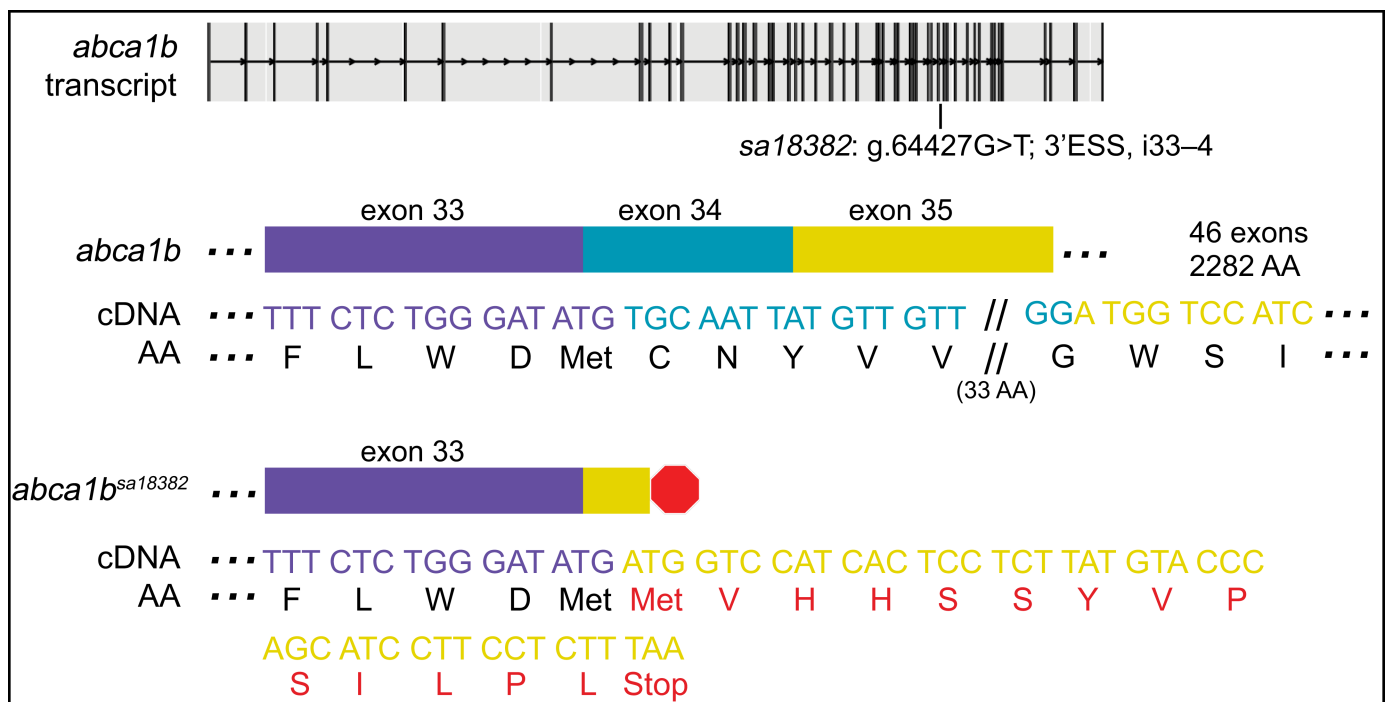
Fig 1. An ENU-induced G>T mutation in the 3' essential splice site (ESS) of i33–34 of *abca1b* leads to a skipped exon and an early termination signal. Analysis of cDNA sequence and agarose gel electrophoresis (S1 Fig) indicates that loss of exon 34 (116 bases) causes a frame shift, followed by a short ORF (13 AA) and an early termination codon (shown in red). qPCR studies revealed transcript levels down 3.5-fold in 6-dpf *abca1b*[sa18382/sa18382] zebrafish. See S1 Table for sequence spanning mutation and predicted outcome.

https://doi.org/10.1371/journal.pgen.1007105.g001

confirmed that exon 34 (116 bases) is skipped and leads to a frame shift in *abca1b*[sa18382/+] and *abca1b*[sa18382/sa18382] larvae. Following the frame shift, the mutant cDNA encodes a 13 AA open reading frame (ORF) and an early termination signal that would direct the loss of exons 34–46. qPCR studies reveal transcript levels are down 3.5-fold in 6-dpf *abca1b*[sa18382/sa18382] zebrafish (ANOVA with Tukey's test, p = 0.049).

An ESS mutation in *slc27a2a* leads to an expected skipped exon but not a frame shift

slc27a2a[sa30701] has a point mutation in the essential splice donor site of intron 2–3 (g.3431G>A) (Fig 2). cDNA sequencing confirms omission of exon 2 in *slc27a2a*[sa30701/+] and *slc27a2a*[sa30701/sa30701] larvae. No frame shift is observed since exon 2 is 210 bases long (encoding 70 AA). By qPCR, transcript levels in 6-dpf *slc27a2a*[sa30701/sa30701] zebrafish did not differ from those of their wild-type siblings (ANOVA with Tukey's test; p-value greater than threshold of 0.05).

An ESS mutation in *abca1a* leads to use of a nearby cryptic splice acceptor site and loss of a single AA

abca1a[sa9624] has a point mutation in the 3' ESS of intron 29–30 (g.48320G>A) (Fig 3). Analysis of cDNA sequence from individual genotyped larvae revealed the loss of three bases, "TAG", at the start of exon 30 in heterozygous and homozygous mutants. To look for cryptic splice sites, a flanking region of gDNA was PCR amplified and sequenced. A cryptic splice acceptor site ("AG") was found 2 and 3 bases downstream of the mutated wildtype splice acceptor site, in exon 30. Use of this cryptic splice acceptor site splices out the first three bases of exon 3 ("TAG") and the protein this message encodes would lack one Serine (and remain in frame with the wildtype product). Transcript levels in 6-dpf *abca1a*[sa9624/sa9624] zebrafish did not differ from their wildtype siblings (ANOVA with Tukey's test; p-value>0.05).
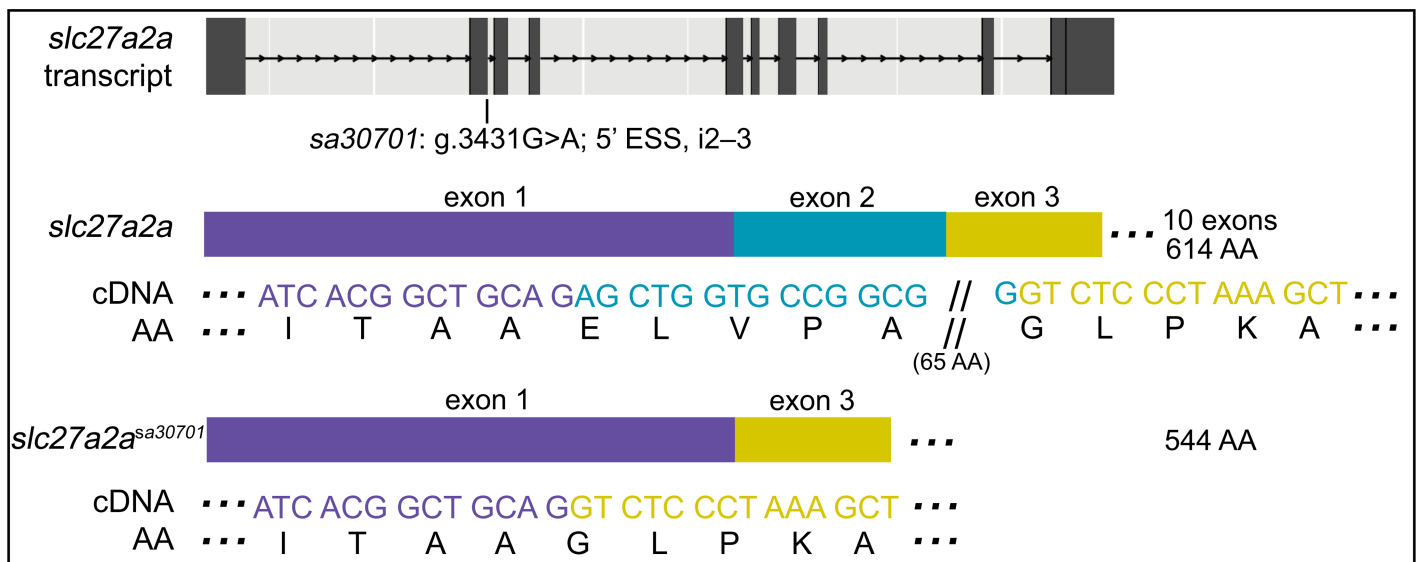
Fig 2. An ENU-induced G>A mutation in the 5' essential splice site (ESS) of i2–3 of *slc27a2a* leads to a skipped exon. Analysis of cDNA sequence and agarose gel electrophoresis (S1 Fig) reveals a loss of exon 2 (70 AA) with no frame shift. By qPCR, transcript levels of 6-dpf *slc27a2a*[sa30701/sa30701] zebrafish were not found to be different than their siblings. See S1 Table for sequence spanning mutation and predicted outcome.

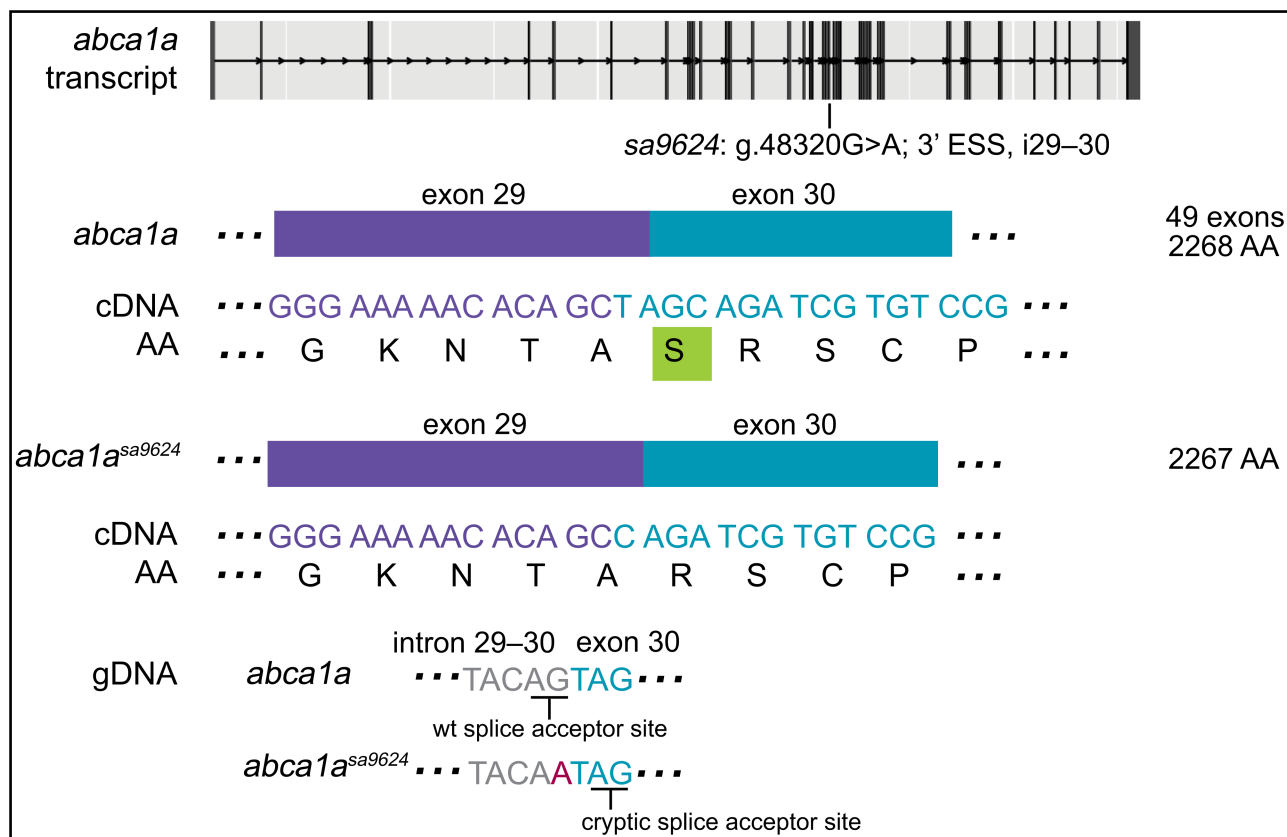https://doi.org/10.1371/journal.pgen.1007105.g002

Fig 3. An ENU-induced G>A mutation in the 3' essential splice site (ESS) of i29–30 of *abca1a* leads to use of a nearby cryptic splice site and loss of a single AA. The base change causes a missed splice acceptor and sequence immediately following the mutated base to be used as a cryptic splice site, as confirmed by analysis of cDNA sequence. A single Serine is lost (boxed in green) and the product remains in frame. By qPCR, transcript levels of 6-dpf *abca1a$^{sa9624/sa9624}$* zebrafish were not found to be different than their siblings. See S2 Fig for agarose gel electrophoresis of amplified cDNA and S1 Table for sequence spanning mutation and predicted outcome.

https://doi.org/10.1371/journal.pgen.1007105.g003

An ESS mutation in *cd36* leads to use of a cryptic splice donor site, frame shift, PTC, but not NMD

cd36^{sa9388} has a point mutation in the 5' ESS (splice donor site) of intron 10–11 (g.11242G>A) (Fig 4). cDNA sequencing of individual, genotyped larvae reveals incorporation of extra bases "ATAT" in between the sequence for exon 10 and exon 11, which leads to a frame shift in the mutant allele. After the frame shift, 18 AA and a PTC follow, predicting the loss of exon 12 (154 AA). The PTC position sits at the last exon-exon junction and thus transcripts are predicted to escape NMD. Transcript levels of 6-dpf *cd36$^{sa9388/sa9388}$* larvae did not differ significantly from their wildtype siblings (ANOVA with Tukey's test; p-value>0.05).

To look for the use of a cryptic splice donor site, a flanking region of gDNA isolated from individual larvae was amplified and sequenced. The wildtype sequence at the 5' end of intron 10–11 includes the splice donor "GT". However, in the mutant allele, the first base is mutated to an "A", resulting in the loss of the splice donor site. The mutated intronic sequence begins "ATATGT...", which provides a cryptic splice donor site ("GT") 3 and 4 bases downstream of the mutated wildtype splice donor site (Fig 4).
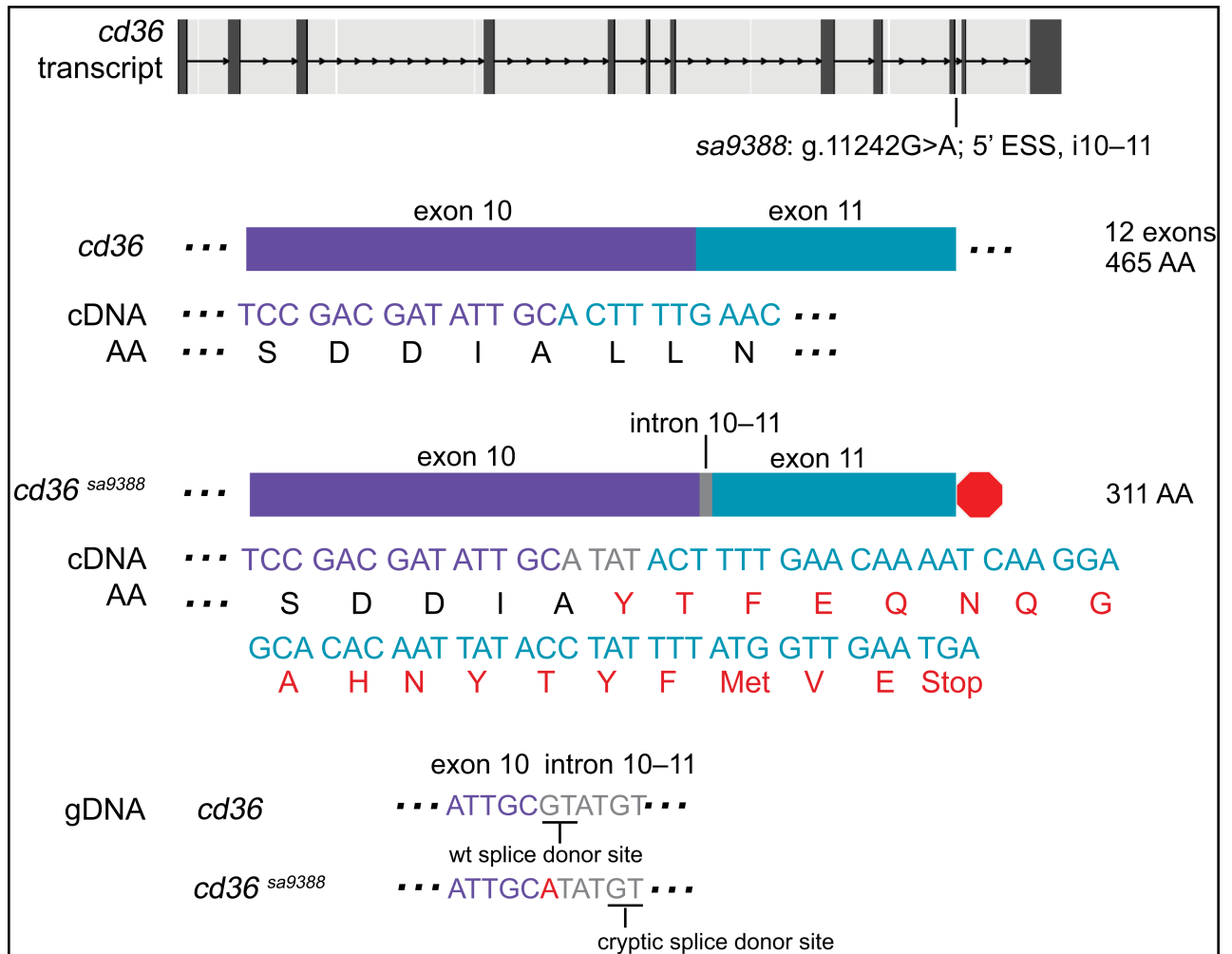
Fig 4. An ENU-induced G>A mutation in the 5' essential splice site (ESS) of i10–11 of *cd36* leads to use of a cryptic splice site and a frame shift. The base change causes loss of a splice donor and use of a cryptic splice site 3 and 4 bases downstream, as confirmed by analysis of cDNA sequence. The intronic sequence "ATAT" preceding the cryptic splice site is thus incorporated, leading to a frame shift. After the frame shift, 18 AA follow before a stop codon (shown in red) directs early termination and loss of exon 12 (154 AA). By qPCR, transcript levels of 6-dpf *cd36*^{sa9388/sa9388} zebrafish were not found to be different than their siblings. See S2 Fig for agarose gel electrophoresis of amplified cDNA and S1 Table for sequence spanning mutation and predicted outcome.

https://doi.org/10.1371/journal.pgen.1007105.g004

An ESS mutation in *pla2g12b* is associated with lowered transcript counts and a mutant phenotype

pla2g12b^{sa659} has a point mutation in the splice acceptor site of intron 3–4 (g.10194A>T) and is predicted to skip the last (4 of 4) exon; thus, exons flanking the mutation could not be PCR amplified to confirm the loss of exon 4 in mutants. Attempts to amplify an alternative transcript with retention of either the final exon 4 or the intron 3–4 did not succeed when using cDNA synthesized from homozygous mutant larvae as the template. During phenotypic screening and genotyping of 5-dpf larvae from heterozygous incrosses, a total of 29 *pla2g12b*^{sa659/sa659} larvae exhibited a darkened yolk phenotype while 52 *pla2g12b*^{+/+} siblings did not (2 experiments; S3 Fig). Correspondingly, RNA expression profiling demonstrates a 3-fold decrease in mutants/siblings (adj. p = 3.68 x 10^{−8}) [23] (S2 Table).
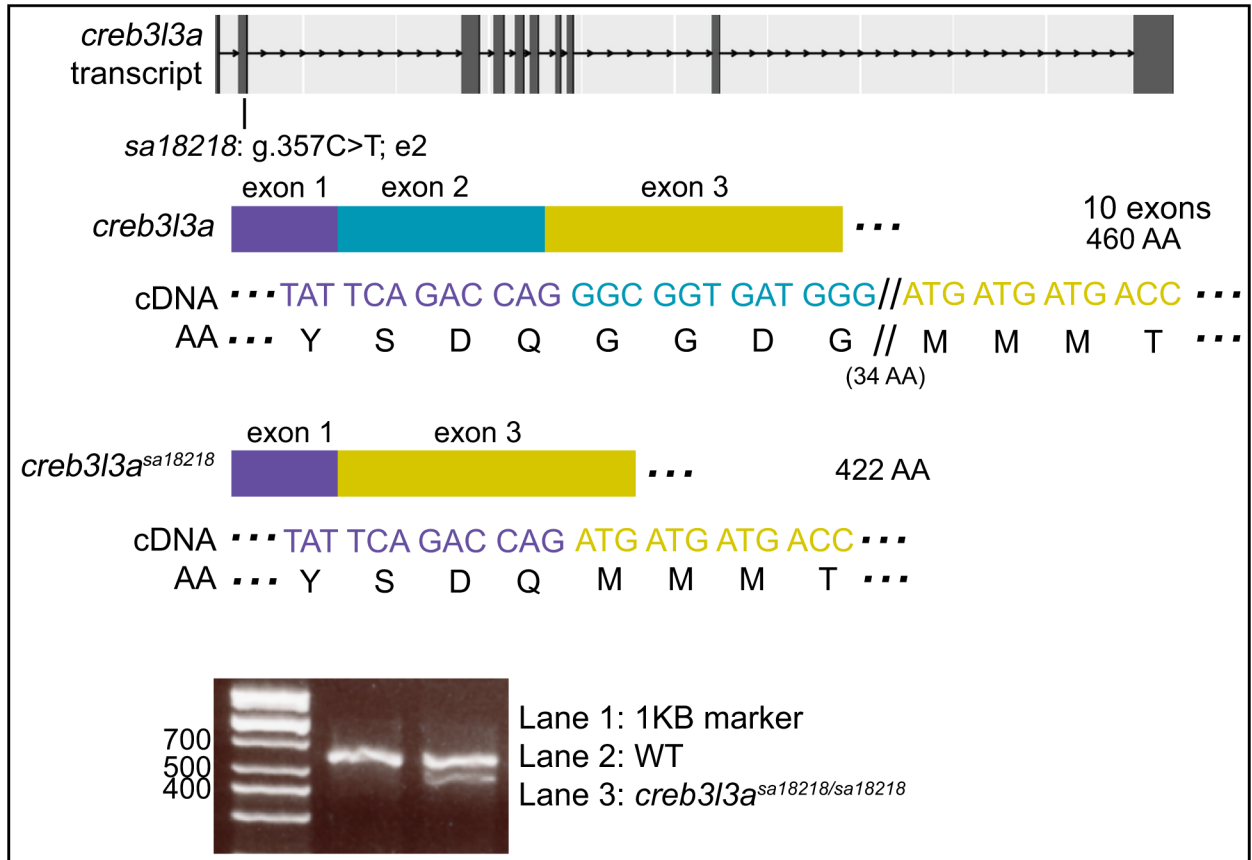
Fig 5. An ENU-induced C>T nonsense mutation in exon 2 of *creb3l3a* leads to a skipped exon. Analysis of cDNA sequence reveals a loss of exon 2 (38 AA) with no frame shift. Agarose gel electrophoresis of amplified cDNA (from pooled intestines) revealed an additional band in the homozygous mutants matching the expected size of a product with a skipped exon 2. By qPCR, transcript levels of 6-dpf *creb3l3a*$^{sa18218/sa18218}$ zebrafish were not found to be different than their siblings. See S1 Table for sequence spanning mutation and predicted outcome. Band sizes (number of bases) for the ladder is indicated.

A nonsense mutation in *creb3l3a* leads to an unexpected skipped exon but no frame shift

creb3l3a^{sa18218} has a nonsense mutation in exon 2 of 10 (g.357C>T), which changes codon <u>C</u>AA to <u>T</u>AA, a PTC (Fig 5). PCR amplification of cDNA (wildtype and homozygous pooled larval intestines) followed by gel electrophoresis revealed two bands in homozygous mutants but only the expected wildtype band in the wildtype siblings (Fig 5). cDNA sequencing of the bands showed alternative transcripts with the unexpected omission of exon 2 in homozygous mutant but not in wildtype larvae. Splicing out exon 2 (114bp encoding 38 AA) does not lead to a frame shift. The nonsense mutation was found to occur in a predicted exonic splice enhancer (ESE) sequence using the web-based prediction tool, ESEFinder [24]. Mutation of an ESE, an important aspect of exon definition, could explain a reduction of transcripts that include exon 2 in mutant cDNA. By qPCR, transcript levels of 6-dpf *creb3l3a*$^{sa18218/sa18218}$ dissected intestines did not differ significantly from their wildtype siblings (Wilcoxon-Mann-Whitney test; p>0.05).

Analysis of the frequency of zebrafish exons that are divisible by 3

It has been shown that conserved alternative exons have a high percentage of preservation of reading frame [11] and 41% of all human exons are symmetrical (divisible by 3) [25]. To

determine whether zebrafish exons 2–10 were symmetrical at a higher frequency than expected by random chance, we summarized the remainder (0, 1, and 2) for all coding genes in Ensembl GRCz10 (Release 90; August 2017) across each exon. Our analysis of zebrafish coding exons 2–10 revealed between a 5.1% and 7.2% increase over chance (33.33%) in exons divisible by 3 (Chi-squared test, df = 2, p-value < 2.2e-16); S4 Fig).

A CRISPR-induced deletion in *smyd1a* correlates with the use of upstream cryptic splice sites

To confirm that these adaptive phenomena are not specific to ENU-mutagenized lines, we analyzed a 7-bp deletion in exon 3 of *smyd1a* (g.6948_6955del; SET and MYND domain containing 1A) which was generated using CRISPR/Cas9 targeting methods. The 7-bp deletion leads to a predicted frame shift and PTC (48/485 AA produced) (Fig 6). To look for evidence
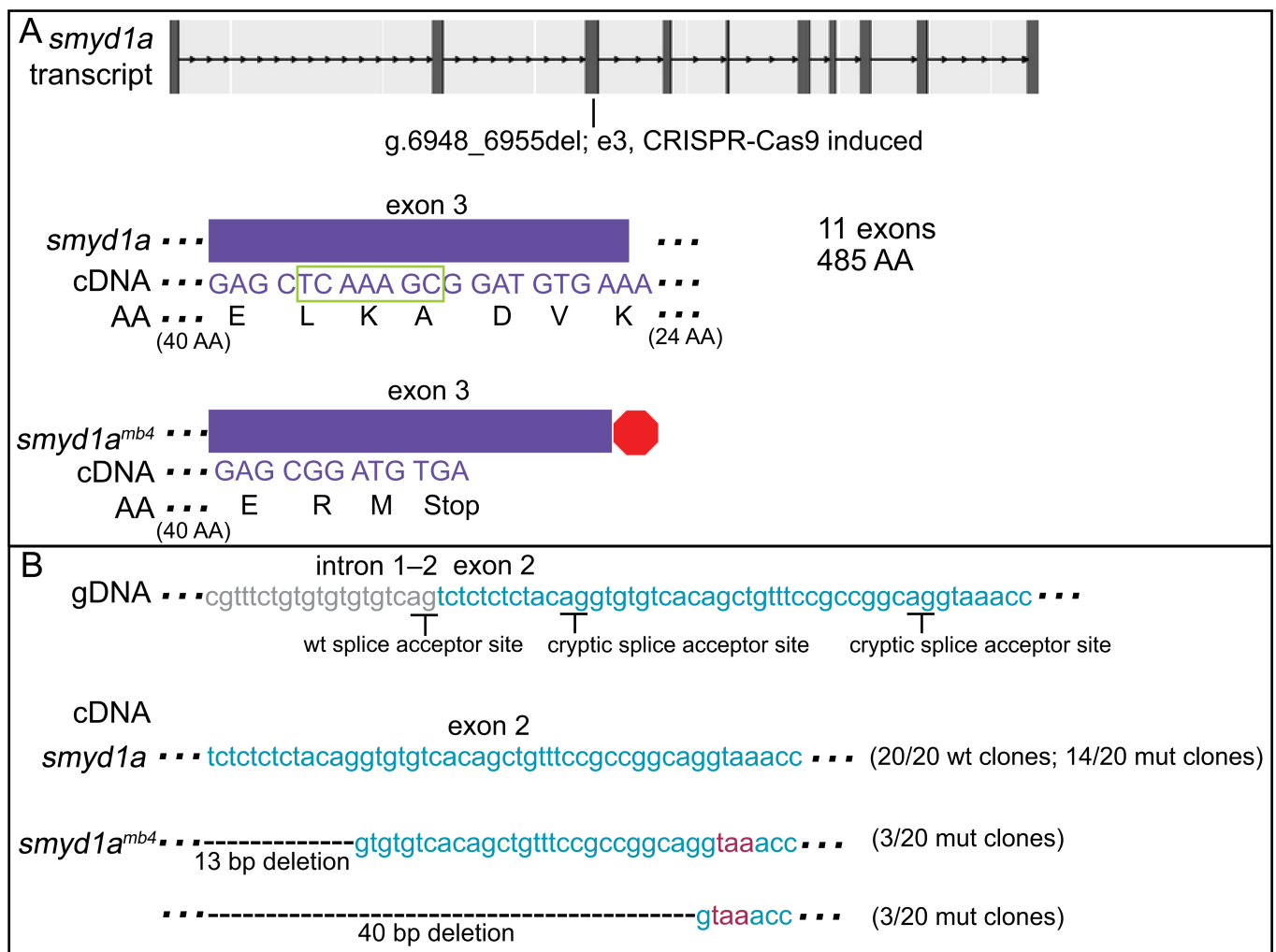
Fig 6. A CRISPR-induced deletion (7 bp) in exon 3 of *smyd1a* corresponds to use of cryptic splice sites in exon 2 in mutant clones. A. A 7-bp deletion in exon 3 (boxed in green) is predicted to lead to a frame shift mutation and early termination. Of 20 mutant clones sequenced, all 20 had the expected 7-bp deletion. **B.** In addition, 3 revealed use of one cryptic splice site and another 3 revealed use of a second cryptic splice site in exon 2, leading to a 13- and 40-bp deletion, respectively. Both deletions result in a frame shift and premature termination codon, shown in red text. 20 randomly selected wt clones did not show alternative splicing. By qPCR, transcript levels of 1- and 2-dpf *smyd1a^mb4/mb4* zebrafish were down thirteen-fold compared to wildtype siblings. See S1 Table for sequence spanning mutation and predicted outcome.

https://doi.org/10.1371/journal.pgen.1007105.g006

of novel alternative splicing, *smyd1a* cDNA was sequenced from wildtype and mutant embryos by cloning full-length PCR products. As expected, all 20 wildtype cDNA clones had the *smyd1a* wildtype sequence and all 20 clones from the homozygous mutant embryos contained the 7-bp deletion in exon 3. However, 6 of the 20 cDNA clones from mutant embryos exhibited alternative splicing at exon 2 (Fig 6). Three clones had an alternative splice event using a cryptic splice acceptor site ("AG") in exon 2, located 13-bp downstream of the wildtype splice acceptor site, leading to a 13-bp deletion at the 5' end of the exon 2. Similarly, sequence data from another three clones show the use of a cryptic splice acceptor site ("AG") 40 bp downstream of the wildtype splicing site, resulting in a 40-bp deletion at the 5' region of exon 2. Both deletions are predicted to lead to a frame shift and premature translation termination. qPCR studies revealed transcript levels of 1- and 2-dpf *smyd1a*$^{mb4/mb4}$ zebrafish were down 13-fold compared to wildtype siblings (Wilcoxon-Mann-Whitney test, p = 0.000077).

Discussion

In this report, we analyzed the compensatory mechanisms that function through permissive mRNA processing in the context of ENU- and CRISPR-induced mutations (Table 2). Recently, Popp et al. reviewed how the process of exon-junction-complex-mediated NMD influences the success of creating loss-of-function mutations with CRISPR/Cas9 [17]. Most notable is their earlier finding that NMD cannot occur if a PTC is within 50–55 nucleotides (nt) of the last exon-exon junction [15]. In our study, we found one example of this phenomenon. For the *cd36*^{sa9388} allele, the resultant PTC is within 1 nt of the last exon junction (e11–e12) and as predicted, we observed wildtype transcript levels in the homozygous mutants.

Others have proposed identifying potential cryptic start sites before the construction of any CRISPR or TALEN vectors, after finding wildtype expression levels in *in vitro* mouse NIH3T3 cell lines harboring frame-shift mutations in Gli3 [16]. Loss of function from mutations near the translation initiation site may be recovered by utilizing nearby downstream alternative translation initiation sites [26]. The mutations in our lines were closer to the middle or 3' end of genes. We did identify use of cryptic splice sites in the mutant allele in three of seven lines (*abca1a*^{sa9624}, *cd36*^{sa9388}, *smyd1a*^{mb4}), underlining the importance of identifying potential cryptic splice sites prior to basing studies on presumed lack of gene function.

We also described an example of nonsense-associated splicing (*creb3l3a*^{sa18218}), wherein a PTC-containing exon is spliced out and *creb3l3a*$^{sa18218/sa18218}$ larvae have wildtype transcript

Table 2. Summary of outcomes from this study.

Gene mutated	Ensembl ID	Allele	Nature of mutation	Method of mutation	Outcome on cDNA sequence	Outcome on mutant transcript levels
abca1b	ENSDARG00000079009	*sa18382*	point mutation in ESS	ENU	skipped exon (116 bp), frame shift, PTC	3.5-fold down
slc27a2a	ENSDARG00000036237	*sa30701*	point mutation in ESS	ENU	skipped exon (210 bp), frame maintained	WT levels
abca1a	ENSDARG00000074635	*sa9624*	point mutation in ESS	ENU	downstream cryptic splice site used, loss of single Serine, frame maintained	WT levels
cd36	ENSDARG00000032639	*sa9388*	point mutation in ESS	ENU	downstream cryptic splice site used, frame shift, PTC	WT levels
pla2g12b	ENSDARG00000015662	*sa659*	point mutation in ESS	ENU	no evidence found for retention of exon or preceding intron	3-fold down
creb3l3a	ENSDARG00000056226	*sa18218*	nonsense mutation	ENU	skipped affected exon (114 bp), frame maintained	WT levels
smyd1a	ENSDARG00000009280	*mb4*	7-bp deletion in exon	CRISPR/Cas9	(deletion) frame shift, PTC; two cryptic splice sites used in mutant clones, frame shift, PTC	13-fold down

levels. The mechanisms underlying this process are still being explored: in many cases, mutations in conserved splice elements (such as exonic splice enhancers; ESE) have been shown to cause nonsense-associated splicing [27–31]. Prykhozhij et al. also recently illustrated the need for careful mutation analysis, beyond the level of gDNA sequence. They found only one of three mutant zebrafish lines resulted in the predicted frameshift [18, 32]. Of the remaining two lines, one displayed an exon skip, possibly due to a mutation in an ESE, and the other used an alternative start site. Moreover, there have been two recent publications documenting numerous cases of exon skipping in response to CRISPR/Cas9-mediated mutations [33, 34]. Our analysis of whether zebrafish coding exons 2–10 are divisible by 3 greater than 33.33% of the time revealed a 5.1–7.2% increase over expected. Taken together, these data suggest exon skipping in response to mutation is more common than generally thought and support our suggestion that when possible, researchers target exons that are not divisible by 3.

Since ESS mutations often lead to human disease [6], *in vivo* models are critical to our understanding. However, we found that skipping an exon may still lead to a viable product: if the exon is divisible by 3 and thus its omission does not lead to a frame shift and PTC, transcript levels were not subject to NMD (*creb3l3a^sa18218^*, *slc27a2a^sa30701^*). In both lines found to skip an exon in our study, sequence alignment with their human ortholog revealed no known essential motifs in the skipped exons (S5 Fig). While the skipped exon 2 in zebrafish contains part of the ATP/AMP binding motif responsible for fatty acid activation (through acyl-CoA synthetase activity), data from functional studies suggest that it functions efficiently in long-chain fatty acid transport through the FATP/VLACS motif [35]. Of the two human splice isoforms, FATP2a and FATP2b, the latter lacks the ATP/AMP binding motif but has the FATP/VLACS motif. Expressing FATP2b in yeast and mammalian cultured cells revealed that it functions in long chain fatty acid transport.

Examination of intron-spanning reads from available temporal expression data revealed no evidence of the alternative transcripts we identified in this study in wildtype larvae [36], suggesting that they did not result from wildtype alternative splicing events. These data are not consistent with a low-abundance mRNA variant that is normally expressed in the WT background emerging to partially compensate for the loss of the major WT mRNA variant in the mutant background.

In this study, we report that five of seven analyzed zebrafish lines with induced mutations show evidence of compensation through altered mRNA processing and contribute to the growing data of how to produce successful knockout models. Our data support a hypothesis that there may be a surveying mechanism that could detect mutations and adapt mRNA alternative splicing to cope with potential loss of function. Our findings are consistent with an analysis of 418 nonsense gene variants in the human population that catalog very similar adaptations and suggests "that permissive RNA processing and translation in human cells facilitates the accumulation of otherwise deleterious genetic variation in the human population" [37]. Analysis of cDNA sequence in mutant alleles may allow for prediction of compensation, simply by scanning for proximal cryptic splice and initiation sites that might be used for alternative transcripts. Moreover, it is entirely possible that splice-blocking morpholinos could engage some of the same compensatory mechanisms described in this study. This hypothesis can be tested by future studies in which cDNA from morphants are subjected to sequencing and supports our contention that researchers always sequence the cDNA in mutants and morphants.

Employing multiple "guide" RNAs in the CRISPR/CAS9 system can result in large intron-spanning deletions in or the elimination of targeted genes. While this approach has been used to generate loss-of-function alleles, it can lead to the deletion of the genomic regions needed for post-transcriptional regulation of gene expression or transcriptional regulation of other

genes. It is estimated that 30–80% of human coding genes are post-transcriptionally regulated, at least in part, by microRNAs (miRNAs) [38]; so far, 2,619 miRNAs and 324,219 miRNA-target interactions have been annotated in human (miRTarBase) [39] and approximately 40% of miRNA genes are located within the introns of protein-coding genes [40]. Rather than creating large deletions or removing an entire gene, other approaches, such as those used to generate nonsense mutations or small deletions, may work better to generate loss-of-function alleles that retain these regulatory regions. As we have shown, alternative transcripts may escape nonsense-mediated decay so 1) analyze the DNA sequence for nearby cryptic splice sites, especially those in frame to the natural/altered cryptic splice site, 2) check whether a nonsense mutation is in a predicted splicing enhancer sequence using available web tools, and 3) in the case of expected exon skip, analyze the exonic sequence for essential motifs and whether the exon length is divisible by 3. Since shorter introns that precede expected affected exons may be retained (instead of exon skip), intron length is also a factor to consider when generating mutants. Performing these steps near the start of a project can inform the nature and location of mutations that would most likely result in a loss-of-function mutant with a phenotype of interest.

Materials and methods

Ethics statement

All procedures using zebrafish were approved by the Carnegie Institution Animal Care and Use Committee (Protocol# 139) or the Institutional Animal Care and Use Committee of the University of Maryland (Permit Number: 0610009).

Zebrafish husbandry

All lines were raised and crossed according to zebrafish husbandry guidelines [41].

Genotyping carriers (ZMP lines)

Heterozygotes for each mutation were identified through a fin-clip based gDNA isolation (REDExtract-N-Amp Tissue PCR kit; Sigma-Aldrich), PCR amplification of a 400–600 bp region around the mutation using designed primer sets (MacVector, Primer 3), and Sanger sequencing using a nested sequencing primer. (Primer sets and conditions are in S3 Table.)

For the *creb3l3a*^{sa18218} line, an NaOH-based DNA extraction method was used to extract gDNA from fin tissues. Genotyping primers were designed using dCAPS finder 2.0 with one mismatch (http://helix.wustl.edu/dcaps/dcaps.html; [42]). The primer introduces EcoRV restriction sites in the mutant amplicons but not in the WT amplicons.

The genomic location of each mutation is based on Ensembl genome assembly GRCz10 and calculated according to the guidelines of the Human Genome Variation Society.

To look for evidence of a skipped exon in the lines with mutations in essential splice sites (ZMP lines)

For each line with a mutation in an ESS, larvae were collected from incrosses of identified heterozygotes and 10–20 6-dpf larvae were pooled for generating RNA samples (using above protocol). RNA samples served as template to generate cDNA (iScript cDNA Synthesis Kit, Bio-Rad). cDNA samples were PCR amplified to provide amplicon sizes of 400–700 bp) and the products were separated and sized using gel electrophoresis. For lines that showed evidence of a skipped exon, individual larvae were genotyped and treated similarly to above to correlate amplicon size with genotype.

To generate cDNA after genotyping individual larvae for qPCR analysis (ZMP lines)

Adults that were found to carry the ESS mutation were incrossed and the progeny collected. Individual 6-dpf larvae underwent a Trizol-based RNA prep adapted from Macedo and Ferreira (2014) [43] to include an additional chloroform extraction. To genotype individual samples, residual gDNA in the unpurified RNA samples was PCR amplified and sent for Sanger sequencing. After genotypes were determined (SnapGene Viewer to view peak trace files), RNA samples were DNAase I-treated and purified (RNA Clean and Concentrator, Zymo Research), served as templates for cDNA synthesis (iScript cDNA Synthesis Kit, Bio-Rad), and ultimately used in qPCR studies to analyze transcript levels.

To check transcript levels using Real-time PCR

qPCR methods included SYBR Green-based methods (Sigma-Aldrich, *abca1a*, *creb3l3a*, *smyd1a*) and Taqman gene expression assays (ThermoFisher Scientific; *cd36*, *slc27a2a*, and *abca1b*). *ef1α* (for *smyd1a*) or 18s rRNA (for all others) levels were used as reference genes. Primer and assay information is shown in S2 Table.

cDNA samples for individual larvae, along with "No RT" controls and "No transcript" controls were run on the CFX96 Touch Real-Time PCR Detection System (Bio-Rad) or on the 7500 Fast-Time PCR System (AB Applied Biosystem). Three technical replicates were run for each sample and a minimum of three biological replicates were used for each genotype for each line. Data was analyzed through calculation of Delta Ct values (18s rRNA as internal control) and either one-way analysis of variance (ANOVA) with the Tukey post hoc test or the Wilcoxon-Mann-Whitney test [44].

To perform transcript counts

Transcript counting data for the *pla2g12b*^{sa659} mutant line was obtained from the Sanger Zebrafish Mutation Project [1] and processed as described [36]. Differential expression analysis was performed using DESeq2 (2010) [45]. The data are deposited at ENA under accession number ERP004581 (samples ERS401972-ERS401991).

To analyze cDNA sequencing (ZMP lines)

cDNA for individual larvae was generated as described above and sequencing was obtained using Sanger sequencing methods. Peak trace files were analyzed manually using SnapGene Viewer (GSL Biotech, LLC) or MacVector. To assist in determining the two alleles of interest (wildtype and potential mutant) for each line, Poly Peak Parser [46] and alignment of wildtype and mutant alleles in MacVector (Align to Reference) were used.

Generation of the *smyd1a*^{mb4} mutant

The *smyd1a* allele containing a 7-bp deletion was generated using the CRISPR/Cas9 targeting method (Cai and Du, in preparation). The target site (5'-GGACCTGAAGGAGCTCAAA-3') was located in exon 3 of the *smyd1a* gene. Genotyping was carried out by using gDNA extracted from the caudal fin as a template for PCR followed by SacI digestion of the resulting amplicons. The 7-bp deletion abolished the SacI site, allowing resolution of bands by agarose gel electrophoresis.

DNA and mRNA sequence analysis of the *smyd1a* ^{mb4} mutant

Homozygous *smyd1a* mutants were identified by PCR and SacI digestion, and confirmed by sequencing the PCR product. To compare the *smyd1a* mRNA sequences from wild type (WT) and mutant embryos, total RNA was isolated from a pool of 50 WT and homozygous *smyd1a* mutant embryos at 48 hpf. cDNAs were generated using the RevertAid First Strand cDNA Synthesis Kit (ThermoFisher, K1621). The full length *smyd1a* cDNA was amplified from the WT and mutant template using Phusion High-Fidelity DNA Polymerase (NEB, M0530S). The amplicons were A-tailed using Taq DNA polymerase (Promega, M8295) and subsequently cloned into pGEM-T easy (Promega, A1360).

Summing symmetry in zebrafish exons

Length of all feature-tagged exons (1–10) in Ensembl genome assembly GRCz10 (Release 90; August 2017) were divided by 3 and remainder (0, 1, or 2) noted. Number of exons analyzed were as follows: for exon 1, 62895; 2, 57137; 3, 48568; 4, 41336; 5, 35312; 6, 30144; 7, 25617; 8, 21969; 9, 19057; 10, 16738. For each exon number, percentage of remainders 0, 1, and 2 were calculated. Chi-squared test for given probabilities was performed using the R Project for Statistical Computing v3.3.1.

Supporting information

S1 Fig. Two lines with ESS mutations have a shorter amplicon, supporting an exon skip. *slc27a2a*^{sa30701} (solute carrier family 27, member 2a) and *abca1b*^{sa18382} (ATP-binding cassette transporter, sub-family A, member 1B) had a shorter amplicon (210 and 116 bp shorter, respectively) that matches the predicted length of amplicons of these cDNAs (pooled from heterozygous incrosses) after the omission of the affected exon. A. Red arrows indicate shorter amplicons. "//" indicates where the gel image was digitally split in order to add ladder sizes. Band sizes of the ladders are show in bases.
B. Primers used and sizes anticipated are listed. C. Primer locations are indicated by arrows. The exons that are expected to be skipped as a result of ESS mutations are shown as dotted lines.
(EPS)

S2 Fig. Three lines with ESS mutations do not reveal shorter amplicons expected with an exon skip. *abca1a*^{sa9624} (ATP-binding cassette, sub-family A, member 1A), *cd36*^{sa30701} (cluster of differentiation cd), and *pla2g12b*^{sa659} (phospholipase A2 Group XIIB) yield PCR products (using pooled larvae from heterozygous incrosses) that match the predicted length of amplicons of the wildtype cDNAs and do not reveal a shorter amplicon expected with the omission of the affected exon (A). Modifying PCR conditions to look for evidence of a retained intron (i3–4) in *pla2g12b*^{sa659} did not yield a product in homozygous mutants. Band sizes of the ladders are shown in bases.
B. Primers used and sizes anticipated are listed. C. Primer locations are indicated by arrows. The exons that are expected to be skipped as a result of ESS mutations are shown as dotted lines. For *pla2g12b*^{sa659}, the splice acceptor sequence preceding exon 4 is mutated (and there is no downstream natural splice acceptor sequence).
(PDF)

S3 Fig. 6-dpf *pla2g12b*^{sa659/sa659} larvae (bottom) display a darkened yolk sac compared to wildtype siblings (top).
(EPS)

S4 Fig. Exons 2–10 of zebrafish coding genes are divisible by 3 (symmetrical) more frequently than one-third of the time. Length of exons 1–10 of all coding genes in Ensembl genome assembly GRCz10 were divided by 3 and remainder (0, 1, or 2) noted. For each exon, percentage of remainders 0, 1, and 2 were calculated and displayed.
(EPS)

S5 Fig. Sequence alignments of *slc27a2* and *creb3l3a*. The shaded yellow region indicates the skipped exon (2). For *slc27a2*, the shaded green region indicates the ATP/AMP motif and the shaded blue region indicates the FATP/VLACS motif [47]. For *creb3l3a*, the shaded green region indicates the basic region and the shaded blue region indicates the leucine zipper domain (identified in [48]).
(PDF)

S1 Table. Mutations analyzed in this study and their predicted outcome.
(TIF)

S2 Table. RNA expression profiling reflect a 3-fold decrease in *pla2g12b*$^{sa659/sa659}$ larvae compared to wildtype siblings (adj. p = 3.68 x 10^{-8}).
(XLSX)

S3 Table. List of primers and methods used in this study.
(TIFF)

Author Contributions

Conceptualization: Jennifer L. Anderson, Timothy S. Mulligan, Steven A. Farber.

Formal analysis: Jennifer L. Anderson, Timothy S. Mulligan, Meng-Chieh Shen, Hui Wang, Catherine M. Scahill, Frederick J. Tan, Shao J. Du, Elisabeth M. Busch-Nentwich, Steven A. Farber.

Investigation: Jennifer L. Anderson, Timothy S. Mulligan, Meng-Chieh Shen, Hui Wang, Catherine M. Scahill, Frederick J. Tan, Elisabeth M. Busch-Nentwich.

Methodology: Jennifer L. Anderson, Timothy S. Mulligan, Steven A. Farber.

Supervision: Shao J. Du, Elisabeth M. Busch-Nentwich, Steven A. Farber.

Visualization: Jennifer L. Anderson.

Writing – original draft: Jennifer L. Anderson, Timothy S. Mulligan, Steven A. Farber.

Writing – review & editing: Jennifer L. Anderson, Timothy S. Mulligan, Meng-Chieh Shen, Catherine M. Scahill, Shao J. Du, Elisabeth M. Busch-Nentwich, Steven A. Farber.

References

1. Kettleborough RN, Busch-Nentwich EM, Harvey SA, Dooley CM, de Bruijn E, van Eeden F, et al. A systematic genome-wide analysis of zebrafish protein-coding gene function. Nature. 2013; 496 (7446):494–7. https://doi.org/10.1038/nature11992 PMID: 23594742; PubMed Central PMCID: PMCPMC3743023.

2. Kok FO, Shin M, Ni CW, Gupta A, Grosse AS, van Impel A, et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. Dev Cell. 2015; 32 (1):97–108. https://doi.org/10.1016/j.devcel.2014.11.018 PMID: 25533206; PubMed Central PMCID: PMCPMC4487878.

3. Rossi A, Kontarakis Z, Gerri C, Nolte H, Holper S, Kruger M, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature. 2015; 524(7564):230–3. https://doi.org/10.1038/nature14580 PMID: 26168398.

4. Morgens DW, Deans RM, Li A, Bassik MC. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. Nat Biotechnol. 2016; 34(6):634–6. https://doi.org/10.1038/nbt.3567 PMID: 27159373; PubMed Central PMCID: PMCPMC4900911.

5. Caminsky N, Mucaki EJ, Rogan PK. Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. F1000Res. 2014; 3:282. https://doi.org/10.12688/f1000research.5654.1 PMID: 25717368; PubMed Central PMCID: PMCPMC4329672.

6. Lewandowska MA. The missing puzzle piece: splicing mutations. Int J Clin Exp Pathol. 2013; 6 (12):2675–82. PubMed PMID: 24294354; PubMed Central PMCID: PMCPMC3843248.

7. Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. J Med Genet. 2005; 42(10):737–48. https://doi.org/10.1136/jmg.2004.029538 PMID: 16199547; PubMed Central PMCID: PMCPMC1735933.

8. Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. Cell. 1999; 96(3):307–10. PubMed PMID: 10025395.

9. Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. Nat Rev Genet. 2016; 17(7):407–21. https://doi.org/10.1038/nrg.2016.46 PMID: 27240813; PubMed Central PMCID: PMCPMC5154377.

10. Nakai K, Sakamoto H. Construction of a novel database containing aberrant splicing mutations of mammalian genes. Gene. 1994; 141(2):171–7. PubMed PMID: 8163185.

11. Magen A, Ast G. The importance of being divisible by three in alternative splicing. Nucleic Acids Res. 2005; 33(17):5574–82. Epub 2005/09/30. https://doi.org/10.1093/nar/gki858 PMID: 16192573; PubMed Central PMCID: PMCPMC1236976.

12. Kapustin Y, Chan E, Sarkar R, Wong F, Vorechovsky I, Winston RM, et al. Cryptic splice sites and split genes. Nucleic Acids Res. 2011; 39(14):5837–44. https://doi.org/10.1093/nar/gkr203 PMID: 21470962; PubMed Central PMCID: PMCPMC3152350.

13. Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet. 1992; 90(1–2):41–54. PubMed PMID: 1427786.

14. Treisman R, Orkin SH, Maniatis T. Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. Nature. 1983; 302(5909):591–6. PubMed PMID: 6188062.

15. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when non-sense affects RNA abundance. Trends Biochem Sci. 1998; 23(6):198–9. PubMed PMID: 9644970.

16. Makino S, Fukumura R, Gondo Y. Illegitimate translation causes unexpected gene expression from on-target out-of-frame alleles created by CRISPR-Cas9. Sci Rep. 2016; 6:39608. https://doi.org/10.1038/srep39608 PMID: 28000783; PubMed Central PMCID: PMCPMC5175197.

17. Popp MW, Maquat LE. Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. Cell. 2016; 165(6):1319–22. https://doi.org/10.1016/j.cell.2016.05.053 PMID: 27259145; PubMed Central PMCID: PMCPMC4924582.

18. Prykhozhij SV, Rajan V, Berman JN. A Guide to Computational Tools and Design Strategies for Genome Editing Experiments in Zebrafish Using CRISPR/Cas9. Zebrafish. 2016; 13(1):70–3. https://doi.org/10.1089/zeb.2015.1158 PMID: 26683213.

19. Gonzales AP, Yeh JR. Cas9-based genome editing in zebrafish. Methods Enzymol. 2014; 546:377–413. https://doi.org/10.1016/B978-0-12-801185-0.00018-0 PMID: 25398350.

20. Irion U, Krauss J, Nusslein-Volhard C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. Development. 2014; 141(24):4827–30. https://doi.org/10.1242/dev.115584 PMID: 25411213; PubMed Central PMCID: PMCPMC4299274.

21. Lieschke GJ, Currie PD. Animal models of human disease: zebrafish swim into view. Nat Rev Genet. 2007; 8(5):353–67. https://doi.org/10.1038/nrg2091 PMID: 17440532.

22. Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, Kane DA, et al. The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. Development. 1996; 123:1–36. PMID: 9007226

23. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11 (10):R106. https://doi.org/10.1186/gb-2010-11-10-r106 PMID: 20979621; PubMed Central PMCID: PMCPMC3218662.

24. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res. 2003; 31(13):3568–71. PubMed PMID: 12824367; PubMed Central PMCID: PMCPMC169022.

25. Schad E, Kalmar L, Tompa P. Exon-phase symmetry and intrinsic structural disorder promote modular evolution in the human genome. Nucleic Acids Res. 2013; 41(8):4409–22. Epub 2013/03/06. https://doi.org/10.1093/nar/gkt110 PMID: 23460204; PubMed Central PMCID: PMCPMC3632108.

26. Kochetov AV. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. Bioessays. 2008; 30(7):683–91. https://doi.org/10.1002/bies.20771 PMID: 18536038.

27. Buchner DA, Trudeau M, Meisler MH. SCNM1, a putative RNA splicing factor that modifies disease severity in mice. Science. 2003; 301(5635):967–9. https://doi.org/10.1126/science.1086187 PMID: 12920299.

28. Caputi M, Kendzior RJ Jr., Beemon KL. A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer. Genes Dev. 2002; 16(14):1754–9. https://doi.org/10.1101/gad.997502 PMID: 12130535; PubMed Central PMCID: PMCPMC186389.

29. Liu HX, Cartegni L, Zhang MQ, Krainer AR. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat Genet. 2001; 27(1):55–8. https://doi.org/10.1038/83762 PMID: 11137998.

30. Pagani F, Buratti E, Stuani C, Baralle FE. Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. J Biol Chem. 2003; 278(29):26580–8. https://doi.org/10.1074/jbc.M212813200 PMID: 12732620.

31. Valentine CR. The association of nonsense codons with exon skipping. Mutat Res. 1998; 411(2):87–117. PubMed PMID: 9806422.

32. Prykhozhij SV, Steele SL, Razaghi B, Berman JN. A rapid and effective method for screening, sequencing and reporter verification of engineered frameshift mutations in zebrafish. Dis Model Mech. 2017; 10(6):811–22. https://doi.org/10.1242/dmm.026765 PMID: 28280001.

33. Mou H, Smith JL, Peng L, Yin H, Moore J, Zhang XO, et al. CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. Genome Biol. 2017; 18(1):108. Epub 2017/06/16. https://doi.org/10.1186/s13059-017-1237-8 PMID: 28615073; PubMed Central PMCID: PMCPMC5470253.

34. Lalonde S, Stone OA, Lessard S, Lavertu A, Desjardins J, Beaudoin M, et al. Frameshift indels introduced by genome editing can lead to in-frame exon skipping. PLoS One. 2017; 12(6):e0178700. Epub 2017/06/02. https://doi.org/10.1371/journal.pone.0178700 PMID: 28570605; PubMed Central PMCID: PMCPMC5453576.

35. Melton EM, Cerny RL, Watkins PA, DiRusso CC, Black PN. Human fatty acid transport protein 2a/very long chain acyl-CoA synthetase 1 (FATP2a/Acsvl1) has a preference in mediating the channeling of exogenous n-3 fatty acids into phosphatidylinositol. J Biol Chem. 2011; 286(35):30670–9. https://doi.org/10.1074/jbc.M111.226316 PMID: 21768100; PubMed Central PMCID: PMC3162428.

36. White RJ, Collins JE, Sealy IM, Wali N, Dooley CM, Digby Z, et al. A high-resolution mRNA expression time course of embryonic development in zebrafish. bioRxiv. 2017.

37. Jagannathan S, Bradley RK. Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. Genome Res. 2016; 26(12):1639–50. Epub 2016/09/21. https://doi.org/10.1101/gr.205070.116 PMID: 27646533; PubMed Central PMCID: PMCPMC5131816.

38. Lu J, Clark AG. Impact of microRNA regulation on variation in human gene expression. Genome Res. 2012; 22(7):1243–54. https://doi.org/10.1101/gr.132514.111 PMID: 22456605; PubMed Central PMCID: PMCPMC3396366.

39. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. 2016; 44(D1):D239–47. https://doi.org/10.1093/nar/gkv1258 PMID: 26590260; PubMed Central PMCID: PMCPMC4702890.

40. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. Genome Res. 2004; 14(10A):1902–10. https://doi.org/10.1101/gr.2722704 PMID: 15364901; PubMed Central PMCID: PMCPMC524413.

41. Westerfield M. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed. Eugene OR: University of Oregon Press; 2000.

42. Neff MM, Turk E, Kalishman M. Web-based primer design for single nucleotide polymorphism analysis. Trends Genet. 2002; 18(12):613–5. PMID: 12446140

43. Macedo NJ, Ferreira TL. Maximizing Total RNA Yield from TRIzol Reagent Protocol: A Feasibility Study. ASEE Zone I Conference. 2014;Student Papers Proceedings Archive

44. Marx A, Backes C, Meese E, Lenhof HP, Keller A. EDISON-WMW: Exact Dynamic Programing Solution of the Wilcoxon-Mann-Whitney Test. Genomics Proteomics Bioinformatics. 2016; 14(1):55–61. https://doi.org/10.1016/j.gpb.2015.11.004 PMID: 26829645; PubMed Central PMCID: PMCPMC4792850.

45. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15(12):550. Epub 2014/12/18. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281; PubMed Central PMCID: PMCPMC4302049.

46. Hill JT, Demarest BL, Bisgrove BW, Su YC, Smith M, Yost HJ. Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. Dev Dyn. 2014; 243(12):1632–6. https://doi.org/10.1002/dvdy.24183 PMID: 25160973; PubMed Central PMCID: PMCPMC4525701.

47. Melton EM, Cerny RL, Watkins PA, DiRusso CC, Black PN. Human fatty acid transport protein 2a/very long chain acyl-CoA synthetase 1 (FATP2a/Acsvl1) has a preference in mediating the channeling of exogenous n-3 fatty acids into phosphatidylinositol. J Biol Chem. 2011; 286(35):30670–9. Epub 2011/07/20. https://doi.org/10.1074/jbc.M111.226316 PMID: 21768100; PubMed Central PMCID: PMCPMC3162428.

48. Omori Y, Imai J, Watanabe M, Komatsu T, Suzuki Y, Kataoka K, et al. CREB-H: a novel mammalian transcription factor belonging to the CREB/ATF family and functioning via the box-B element with a liver-specific expression. Nucleic Acids Res. 2001; 29(10):2154–62. Epub 2001/05/23. PubMed PMID: 11353085; PubMed Central PMCID: PMCPMC55463.