Pericentromeric hypomethylation elicits an interferon response in an animal model of ICF syndrome

Authors

Srivarsha Rajshekar[1,2,5], Jun Yao[2], Paige K. Arnold[3], Sara G. Payne[4], Yinwen Zhang[5], Teresa V. Bowman[4], Robert J. Schmitz[6], John R. Edwards[7], Mary G. Goll[2,6*]

[1]Program in Biochemistry and Structural Biology, Cell and Developmental Biology, and Molecular Biology, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY, 10065, USA

[2]Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

[3]Louis V. Gerstner, Jr., Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

[4]Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, Bronx, NY, 10461, USA

[5]Institute of Bioinformatics, University of Georgia, Athens, GA, 30602, USA

[6]Department of Genetics, University of Georgia, Athens, GA, 30602, USA

[7]Center for Pharmacogenomics, Department of Medicine, Washington University in St. Louis School of Medicine, St. Louis, MO, 63110, USA

*Current address

Corresponding Author: Mary.Goll@uga.edu

Abstract

Pericentromeric satellite repeats are enriched in 5-methylcytosine (5mC). Loss of 5mC at these sequences is common in cancer and is a hallmark of Immunodeficiency, Centromere and Facial abnormalities (ICF) syndrome. While the general importance of 5mC is well-established, the specific functions of 5mC at pericentromeres are less clear. To address this deficiency, we generated a viable animal model of pericentromeric hypomethylation through mutation of the ICF-gene *ZBTB24*. Deletion of zebrafish *zbtb24* caused a progressive loss of 5mC at pericentromeres and ICF-like phenotypes. Hypomethylation of these repeats triggered derepression of pericentromeric transcripts and activation of an interferon-based innate immune response. Injection of pericentromeric RNA is sufficient to elicit this response in wild-type embryos, and mutation of the MDA5-MAVS dsRNA-sensing machinery blocks the response in mutants. These findings identify activation of the innate immune system as an early consequence of pericentromeric hypomethylation, implicating derepression of pericentromeric transcripts as a trigger of autoimmunity.

Introduction

In vertebrate genomes, the majority of cytosine residues within CpG dinucleotides are methylated at the 5 position of the cytosine ring (5-methylcytosine, 5mC) (Suzuki and Bird, 2008). 5mC is established by the *de novo* DNA methyltransferases of the Dnmt3 family, and is propagated by the maintenance DNA methyltransferase, Dnmt1 (Goll and Bestor, 2005). In mice, frogs and zebrafish, mutation or morpholino-mediated depletion of *Dnmt1* results in extensive genome-wide methylation loss and embryonic lethality (Anderson et al., 2009; Lei et al., 1996; Rai et al., 2006; Stancheva and Meehan, 2000). In these species, global methylation deficiencies are linked to a variety of adverse outcomes including deregulation of gene expression, derepression of transposons, elevated levels of DNA damage and increased genome instability during mitosis (Smith and Meissner, 2013). Recent studies have further linked global hypomethylation to activation of antiviral signaling pathways in zebrafish mutated for *dnmt1* and in cancer cells treated with the DNA methyltransferase inhibitor 5-azacytidine (Chernyavskaya et al., 2017; Chiappinelli et al., 2015; Roulois et al., 2015). While these studies reinforce the general importance of DNA methylation in vertebrate development and tissue homeostasis, the extensive genome-wide loss of methylation in these models makes it difficult to assign significance to methylation deficiencies at any particular subclass of sequence.

The pericentromeric satellite sequences that juxtapose chromosome centromeres represent an essential structural component of chromosomes and a significant source of 5mC in vertebrate genomes. These highly repetitive sequences appear particularly susceptible to methylation loss in cancer and senescent cells, although the consequences of this hypomethylation are not well understood (Enukashvily et al., 2007; Fanelli et al., 2008; Nakagawa et al., 2005; Narayan et al., 1998; Qu et al., 1999; Suzuki et al., 2002; Tsuda et al., 2002). In contrast to global hypomethylation, loss of 5mC at pericentromeric repeats is compatible with human development. Individuals with the rare, autosomal

recessive disorder Immunodeficiency, Centromere and Facial anomalies (ICF) syndrome show extensive hypomethylation of pericentromeric repeats, while methylation across the rest of the genome is relatively intact (Tuck-Muller et al., 2000; Weisenberger et al., 2005). Affected individuals usually die in late childhood or early adulthood, and exhibit variable symptoms including immunoglobulin deficiency, facial dysmorphism, growth retardation and a generalized failure to thrive (Ehrlich et al., 2008). Chromosome anomalies including whole-arm deletions and multiradial chromosomes have also been reported in mitogen-stimulated lymphocytes from ICF-patients. However, similar chromosome anomalies are not observed in primary tissues from affected individuals (Ehrlich, 2003).

Homozygosity mapping and whole-exome sequencing have separately implicated four genes in ICF syndrome: DNA Methyltransferase 3B (*DNMT3B,* ICF type-1), Zinc-finger and BTB domain containing 24 (*ZBTB24,* ICF type-2), Cell division cycle associated 7 (*CDCA7,* ICF type-3) and Helicase, lymphoid-specific (*HELLS,* ICF type-4) (de Greef et al., 2011; Thijssen et al., 2015; Xu et al., 1999). Most of the described mutations in *DNMT3B* cause amino acid substitutions within the C-terminal catalytic domain, suggesting they may be hypomorphic. In contrast, the majority of mutations in *ZBTB24*, *CDCA7* and *HELLS* are predicted to cause loss of function. Mechanistically, ZBTB24, CDCA7 and HELLS are thought to converge in a singular pathway that facilitates DNMT3B access to pericentromeric DNA (Jenness et al., 2018; Wu et al., 2016).

To date, most studies of pericentromeric 5mC loss have been performed using transformed B-cell lines derived from ICF patients carrying mutations in *DNMT3B* (Ehrlich et al., 2008). Attempts to generate viable mouse models of pericentromeric hypomethylation through mutation of ICF genes have had limited success. Mice harboring ICF-like mutations in *Dnmt3b* exhibit some characteristics of ICF syndrome including small size and facial anomalies. However, most mice die within 24 hours of birth (Ueda et al., 2006). Global methylation profiles were not assessed in these mutants; but significant

hypomethylation was reported at both pericentromeric repeats and retroviral sequences. Similar perinatal lethality was observed following deletion of the mouse *HELLS* orthologue. In this case, mutations were accompanied by roughly 50% reductions in 5mC, and methylation loss was noted at pericentromeres, retroviruses and some single copy sequences (Tao et al., 2011). Deletion of the mouse *Zbtb24* gene was reported to cause embryonic lethality; but methylation changes in these mutants have not been investigated (Wu et al., 2016).

Here, we describe a viable model of pericentromeric methylation loss, generated through mutation of the zebrafish *zbtb24* gene. Homozygous mutant adults exhibited key phenotypic hallmarks of ICF syndrome including hypomethylation of pericentromeric satellite repeats. Hypomethylation of these repeats was first detected in mutants at 2 weeks post fertilization (wpf) and became more severe as animals matured. This progressive methylation loss allowed us to investigate the primary consequences of pericentromeric hypomethylation in the context of a vertebrate animal. Using this model, we link derepression of transcripts from hypomethylated pericentromeres to activation of an interferon-based innate immune response, and we demonstrate that this response is mediated through the MDA5-MAVS dsRNA sensing machinery. Our findings provide insights into the earliest consequences of pericentromeric hypomethylation, demonstrating an unappreciated function for methylation of pericentromeric repeats in protecting against autoimmunity.

Results

Mutation of zebrafish *zbtb24* causes ICF syndrome-like phenotypes

The zebrafish genome encodes a single, well-conserved orthologue of ZBTB24, which we mutated using TAL effector nucleases (TALENs) (**Figure 1A** and **Figure 1-figure supplement 1 and 2**). The recovered 7.9 kb deletion allele (*zbtb24*^{mk22}; here after referred to as *zbtb24*$^{\Delta}$), eliminates coding sequence between exons 2 and 5 (**Figure 1B**). Animals that were homozygous for this deletion lacked detectable *zbtb24* transcripts, suggesting *zbtb24*$^{\Delta}$ is a null allele (**Figure 1-figure supplement 2D**). *Zbtb24*$^{\Delta/\Delta}$ embryos were born to heterozygous parents at the expected Mendelian ratios and had no obvious morphological abnormalities during the first two weeks of development (**Figure 1C**). Phenotypes that were reminiscent of ICF syndrome emerged as animals matured. Consistent with the small stature observed in ICF syndrome, by 3-4 weeks post fertilization (wpf), *zbtb24*$^{\Delta/\Delta}$ mutant zebrafish were smaller than wild-type siblings raised under identical conditions, and this size reduction persisted into adulthood (**Figure 1D-F**). As adults, *zbtb24*$^{\Delta/\Delta}$ mutants exhibited facial anomalies that were characterized by a quantifiable elongation of the snout (**Figure 1G-H**). We also noted evidence of hypogammagloblulinemia in the presence of normal lymphoid cell numbers, which is an immunological hallmark of ICF syndrome (**Figure 1I-J**). Significant death was noted among homozygous mutants at 4 months of age and fewer than 10% of *zbtb24*$^{\Delta/\Delta}$ animals survived beyond 8 months (**Figure 1K**). Similar phenotypes were observed in zebrafish that were homozygous for a second independently-isolated mutant allele of *zbtb24* (*zbtb24*^{mk19}) (**Figure 1-figure supplement 3**). Taken together, these findings identify *zbtb24* homozygous mutant zebrafish as a faithful animal model of ICF syndrome phenotypes.

Progressive methylation loss at pericentromeric repeats in *zbtb24* mutants

Pericentromeric satellite type-1 (Sat1) repeats are found on all zebrafish chromosomes and comprise 5-8% of the zebrafish genome (Phillips and Reed, 2000). As expected, we found that Sat1 sequences from wild-type adults were resistant to digestion with the methylation sensitive restriction enzyme HpyCH4IV, indicating that these pericentromeric repeats were heavily methylated. In contrast to wildtype, Sat1 sequences from *zbtb24*$^{\Delta/\Delta}$ and *zbtb24*$^{mk19/mk19}$ mutant adults were readily digested with HpyCH4IV, indicating extensive loss of methylation at these repeats (**Figure 2A-B** and **Figure 2-figure supplement 1A**)**.** Comparable Sat1 methylation deficiencies were observed when DNA was isolated from dissected adult brain, skin, muscle and fin, suggesting that these sequences were similarly hypomethylated in most adult tissues (**Figure 2-figure supplement 1B** and **Figure 2-figure supplement 2A**).

Somewhat unexpectedly, we found that pericentromeric methylation loss in *zbtb24*$^{\Delta/\Delta}$ mutants was progressive. While extensive hypomethylation of Sat1 sequences was detected in adults lacking *zbtb24*, similar hypomethylation was not observed in mutants at 1 wpf (**Figure 2C-D**). At 2 wpf, *zbtb24* mutants exhibited roughly 3-fold increases in HpyCH4IV digestion, and sensitivity to digestion became increasingly pronounced in older animals (**Figure 2C-D**). By 32 weeks, Sat1 sequences from *zbtb24* mutants exhibited a 23-fold increase in HpyCH4IV digestion compared to wildtype, suggesting a greater than 95% reduction in methylation of these repetitive sequence blocks.

Zbtb24 **mutants exhibit modest reductions in 5mC at non-pericentromeric sequences**

To clarify whether additional sequences were hypomethylated in *zbtb24* mutants, we performed Enhanced Reduced Representation Bisulfite Sequencing (ERRBS) using genomic DNA isolated from the fins of three *zbtb24*$^{\Delta/\Delta}$ mutant adults and three wild-type

siblings at 25 wpf (Garrett-Bakelman et al., 2015). At this stage, Sat1 sequences from isolated fins were 20-fold more sensitive to HypCH4IV in *zbtb24* mutants compared to controls, indicating extensive loss of DNA methylation at pericentromeric repeats (**Figure 2-figure supplement 2A-B**). We then used ERRBS data to interrogate the methylation status of 979,971 non-pericentromeric CpG sites across the genome in the same tissue samples. Our analysis revealed a strong correlation between genome wide 5mC levels in wild-type and *zbtb24*$^{\Delta/\Delta}$ mutant adults (Pearson's correlation value of 0.928), although overall methylation levels were reduced by ~10% in mutants (**Figure 2E** and **Figure 2-figure supplement 2C**). This 10% methylation reduction in mutants consisted primarily of small-magnitude changes in 5mC across the genome, with only 1.3% (13,205) of examined CpG dinucleotides exhibiting methylation differences of greater than 20%. Consistent with this finding, at a threshold of 20% change (p-value<0.01), only 55 differentially methylated regions (DMRs) were identified between wild-type and *zbtb24*$^{\Delta/\Delta}$ adults (**Supplementary File 4**). Methylation levels at endogenous retroviruses and other transposable elements were also examined by methylation sensitive restriction digest. All tested elements were similarly resistant to digestion in *zbtb24*$^{\Delta/\Delta}$ mutant adults and wild-type siblings, indicating that these sequences are comparably methylated in both genotypes (**Figure 2-figure supplement 3**). Collectively, these data reveal limited methylation changes at non-pericentromeric CpG sites across the genome.

Mutation of *zbtb24* causes activation of innate immune response genes

To gain insights into the early consequences of methylation loss in *zbtb24* mutants, we performed transcriptome analysis on RNA isolated from wild-type and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 2 wpf. At this stage, *zbtb24*$^{\Delta/\Delta}$ mutants remain morphologically indistinguishable from wildtype, but show clear hypomethylation of pericentromeric sequences. RNA-seq

identified 58 genes that were downregulated by more than 2-fold in *zbtb24*$^{\Delta/\Delta}$ larvae at 2 wpf, while 119 were upregulated by 2-fold or more (**Figure 3A**). No gene enrichment signature was observed among downregulated genes. However, roughly 30% of upregulated genes were associated with activation of the innate immune system. In particular, we noted that upregulated transcripts included those associated with interferon stimulated genes (ISGs) and inflammatory cytokines (**Figure 3B**). Consistent with these observations, Gene Set Enrichment Analysis (GSEA) identified significant enrichment of genes involved in viral response, a key function of innate immune pathways (**Figure 3C**). Upregulation of ISGs was also observed in *zbtb24*$^{\Delta/\Delta}$ and *zbtb24*$^{mk19/mk19}$ mutants by qRT-PCR at 3 wpf, whereas the same genes were expressed at wild-type levels at 1 wpf (**Figure 3D-E** and **Figure 3-figure supplement 1**). No immune-related genes (and only 1 gene differentially upregulated in the RNA-Seq) were found within 100kb of identified DMRs, suggesting that direct loss of methylation at these sequences was unlikely to cause the response (**Figure 2-figure supplement 2D** and **Supplementary File 4)**.

The innate immune response in *zbtb24* mutants is mediated by sensors of cytosolic RNA

The innate immune system represents an ancient defense system in which pathogen-associated molecular patterns (PAMPs) are recognized by pattern recognition receptors (PRRs). These PRRs induce signaling cascades that drive the production of interferons and other inflammatory cytokines with antiviral and immune modulating functions (Schneider et al., 2014). In addition to extracellular pathogens, PRRs also recognize PAMPs associated with cell-intrinsic stimuli including DNA damage, endogenous retroviral RNA and RNA-DNA hybrids (Chiappinelli et al., 2015; Hartlova et al., 2015; Mankan et al., 2014; Roulois et al., 2015).

To clarify the origin of the response in *zbtb24* mutants, we examined the major families of PRRs involved in innate immunity. These include the Toll-like receptors (TLRs), which have broad functions in detecting PAMPs, the RIG-I like receptors (RLRs), which are involved in the detection of cytosolic RNA and cGAMP synthase (cGAS), which functions as a cytosolic sensor of DNA and RNA/DNA hybrids (Crowl et al., 2017). Mutations in key mediator proteins required to propagate interferon signaling from each PRR family were introduced onto the *zbtb24* mutant background and we tested the effect on ISG expression. Mutations in the zebrafish orthologs of *mitochondrial antiviral-signaling protein* (*mavs)*, which is an intermediate in RLR signaling and *stimulator of interferon genes* (*sting),* which is involved in *cGAS* signaling were generated using CRISPR/Cas9 technology (**Figure 4-figure supplement 1A-B**).The mutant allele of *Myeloid differentiation primary response 88* (*myd88),* which is required for signaling through most TLRs, was previously described (van der Vaart et al., 2013).

As in prior experiments, significant increases of the ISGs, *signal transducer and activator of transcription 1b* (*stat1b)* and *interferon regulatory factor* (*irf7)* were observed in *zbtb24*$^{\Delta/\Delta}$ larvae at 3 wpf by qRT-PCR (**Figure 4A-C**). Introduction of *myd88* or *sting* mutations had little impact on expression of these ISGs, as similar transcript levels were detected in *zbtb24*$^{\Delta/\Delta}$ single mutant animals compared to *myd88*$^{hu3568/hu3568}$; *zbtb24*$^{\Delta/\Delta}$ or *sting*$^{mk30/mk30}$; *zbtb24*$^{\Delta/\Delta}$ double mutants (**Figure 4A-B**). Sustained ISG expression in these double mutants suggests limited roles for TLR and cGAS PRRs in mediating the interferon response in *zbtb24* mutants. In contrast to *myd88* and *sting*, mutation of *mavs* suppressed *stat1b* and *irf7* upregulation in *zbtb24*$^{\Delta/\Delta}$ mutant animals. Expression levels of *irf7* and *stat1b* were reduced 2- and 4-fold respectively in *mavs*$^{mk28/mk28}$; *zbtb24*$^{\Delta/\Delta}$ double mutants when compared to *zbtb24*$^{\Delta/\Delta}$ single mutant zebrafish, indicating a requirement for *mavs* in the upregulation of these ISGs (**Figure 4C)**. This finding implicates RLR signaling in the

activation of the innate immune system in *zbtb24* mutants and suggests a cytosolic RNA trigger for the response.

Pericentromeric RNA transcripts are sufficient to trigger the interferon response in *zbtb24* mutants

Given known roles for DNA methylation in transcriptional repression, we next tested whether loss of methylation at pericentromeric sequence resulted in increased levels of Sat1 transcripts that could trigger the RNA mediated interferon response. Consistent with this model, strong derepression of Sat1 RNA from hypomethylated pericentromeres was noted in *zbtb24* mutant adults (**Figure 5A** and **Figure 5-figure supplement 1A**), whereas transcripts for other dispersed repetitive elements remained unchanged between mutants and wildtype (**Figure 5-figure supplement 1B**). Increases in Sat1 transcripts correlated with levels of *irf7* expression in adult zebrafish (r=0.77), and upregulation of Sat1 transcripts coincided with the window of ISG induction during development (**Figure 5B-C**). Both sense and antisense transcripts were detected in mutants using TAG-aided sense/antisense transcript detection (TASA-TD) strand-specific PCR (Henke et al., 2015), suggesting the potential for derepressed Sat1 transcripts to form double stranded RNAs (**Figure 5D-E**).

To determine whether Sat1 transcripts were sufficient to activate an innate immune response, in vitro synthesized RNA corresponding to Sat1 sense and antisense sequence was injected into wild-type embryos at the 1-cell stage. Expression of the ISGs *stat1b*, *irf7*, *irf1b* and *mxa* was then assessed at 8 hours post fertilization. Injection of Sat1 RNA was sufficient to cause a 3 to 4-fold upregulation in expression of these ISGs, whereas injection of control transcripts encoding the dsRed fluorophore had no effect on expression of these genes (**Figure 5F**). These results functionally link the derepression of Sat1 transcripts to the activation of the innate immune response in *zbtb24* mutants.

The cytosolic dsRNA helicase MDA5 is required for the interferon response in *zbtb24* mutants

Finally, we sought to identify the specific PRR required for the interferon response in *zbtb24* mutants. The RLR family of PRRs consists of two RNA helicases that signal through Mavs: Melanoma Differentiation-Associated protein 5 (Mda5) and Retinoic acid-inducible gene I (Rig-I). Rig-I binds 5' triphosphorylated RNA molecules, whereas Mda5 recognizes long double-stranded RNAs in the cytosol (Crowl et al., 2017). Given that 5' triphosphorylation of RNAs is a typical viral signature that is unlikely to be present on endogenous RNA transcripts, we reasoned that Mda5 was a more likely candidate for the receptor. To test the requirement for *mda5*, we generated a 7 base-pair deletion in this gene that disrupted the DEAD box helicase domain (**Figure 4-figure supplement 1C**). This *mda5*^{mk29} allele was then introduced onto the *zbtb24* mutant background, and expression of the ISGs *stat1b* and *irf7* was examined at 3 wpf. Homozygous mutation of *mda5* was sufficient to restore *stat1b* and *irf7* expression to wild-type levels in *zbtb24*$^{\Delta/\Delta}$ mutant larvae, suggesting that Mda5 is the primary PPR required for the response (**Figure 6A**). This requirement was further validated by RNA-seq, which revealed that a broad panel of ISGs that showed elevated expression in *zbtb24* single mutants were no longer upregulated in *mda5*$^{mk29/mk29}$; *zbtb24*$^{\Delta/\Delta}$ double mutants (**Figure 6B-C).**

Taken together, these results support a model in which derepression of transcripts from hypomethylated pericentromeres triggers activation of the innate immune system through the Mda5/Mavs viral RNA recognition pathway (**Figure 6D**). These findings identify roles for pericentromeric RNA as a trigger of autoimmunity and reveal important functions for pericentromeric methylation in suppressing the generation of these immunostimulatory transcripts. Based on these results, we propose that induction of the innate immune system is one of the earliest *in vivo* consequences of pericentromeric methylation loss.

Discussion

In this study, we describe a viable animal model of ICF syndrome which recapitulates key phenotypic hallmarks of the disease including slow growth, facial anomalies, immunoglobulin deficiencies and reduced lifespan. Given that previous attempts to model ICF syndrome have resulted in perinatal or embryonic lethality (Geiman et al., 2001; Ueda et al., 2006; Wu et al., 2016), this zebrafish model provides an important new resource for understanding ICF disease etiology during juvenile and adult life stages. In particular, *zbtb24* mutant zebrafish will be useful for understanding phenotypes such as immunoglobulin deficiency, which have not been observed in mouse models and are difficult to study in cell culture systems.

Methylation analysis of *zbtb24* mutant zebrafish suggests that the general methylation landscape in these mutants is comparable with that in ICF syndrome. Recent methylome analysis of primary blood from ICF patients identified methylation changes of greater than 20% at roughly 3% of examined CpG dinucleotides. Significant changes in methylation of retroviruses and other dispersed repeats were not observed in these patients (Velasco et al., 2018). Consistent with these findings, our ERRBS analysis revealed methylation changes of greater than 20% at roughly 1.3% of assayed CpG dinucleotides and found methylation of dispersed repeats to be similar between wildtype and in *zbtb24* mutant zebrafish.

In the current study, methylation levels at pericentromeric Sat1 sequences could not be quantified by ERRBS, as this technique relies on Msp1 restriction digest to enrich for CpG containing sequences, and zebrafish Sat1 repeats are lacking in this restriction site. Instead we employed digestion with the methylation sensitive restriction enzyme HpyCH4IV to measure methylation at these repeats. This approach remains the most effective way to assess methylation at highly repetitive sequences. At 32 wpf, we observed increases in sensitivity to HpyCH4IV digestion that are consistent with up to 95%

reductions in methylation at Sat1 pericentromeric repeats in *zbtb24* mutants, while methylation of these sequences was similar to wildtype at 1 wpf. This progressive loss of 5mC implicates Zbtb24 in regulating the long-term maintenance of methylation at pericentromeric repeats. We note that the onset of ICF-like growth defects in *zbtb24* mutant zebrafish emerged in the weeks following Sat1 methylation loss. In at least one case of ICF syndrome type 2, growth reductions and immunodeficiency were also reported to develop with age, raising the possibility that similar progressive methylation loss may impact ICF etiology in humans (von Bernuth et al., 2014). It is also possible that Zbtb24 functions in both maintenance and establishment of pericentromeric methylation, but that requirements for establishment are masked by maternally deposited RNA in *zbtb24* mutant zebrafish lines. Unfortunately, *zbtb24* homozygous mutant zebrafish are sterile, preventing the generation of the maternal-zygotic mutants required to address this question.

In the current study, we take advantage of the progressive Sat1 methylation loss in *zbtb24* mutants to identify activation of interferon signaling as one of the earliest *in vivo* consequences of pericentromeric hypomethylation. This phenotype cannot be attributed to defects in adaptive immunity, as the zebrafish adaptive immune system is not functional until roughly 4 wpf (Trede et al., 2004). Induction of an interferon response has been reported in the context of global hypomethylation in cancer cell lines treated with the DNA methyltransferase inhibitor 5-azacytidine and in zebrafish mutated for the maintenance DNA methyltransferase machinery (Chernyavskaya et al., 2017; Chiappinelli et al., 2015; Roulois et al., 2015). In each of these cases induction of the interferon response was attributed to massive derepression of endogenous retroviral elements.

Our results are distinguished from these earlier studies in that we identify hypomethylation of pericentromeric sequences and subsequent derepression of associated satellite transcripts as a previously unappreciated trigger of innate immunity.

Immunostimulatory motifs have been noted in pericentromeric RNAs derived from mouse and humans, and transcripts derived from these repeats have been observed in *p53* null mouse fibroblasts following global methylation loss (Leonova et al., 2013; Tanne et al., 2015). However, while these studies suggest the potential for pericentromeric hypomethylation to drive an interferon response in diverse vertebrate species, experimental evidence in support of this model has been lacking. Here we demonstrate a causative link between derepression of pericentromeric RNAs and the interferon response, and identify a requirement for Mda5/Mavs in mediating the response. Our findings suggest that aberrant upregulation of both sense and antisense transcripts derived from pericentromeric repeats creates an abundance of double stranded RNAs within the cytosol, which mimic features of double stranded RNA viruses. This finding raises the possibility that this pathway may also recognize additional endogenous long dsRNAs that lack viral origin.

While mutation of *mda5/mavs* rescued the interferon response in *zbtb24* mutants, *mda5/mavs* mutation had little impact on other ICF phenotypes observed in *zbtb24* mutants. Therefore, we find it unlikely that the interferon response drives ICF etiology. Rather this response represents an additional consequence of pericentromeric hypomethylation. Hypomethylation of pericentromeric sequences is compatible with human viability and is observed in abnormal cell contexts including cancer and senescence. Massive increases in pericentromeric transcripts and upregulation of interferon genes have both been noted in cancer (Cheon et al., 2014; Ting et al., 2011). Our data raise the possibility that pericentromeric hypomethylation and subsequent derepression of associated RNAs represents an important but underappreciated trigger of autoimmunity in a variety of disease states.

Materials and Methods

Zebrafish husbandry

Zebrafish husbandry and care were conducted in full accordance with animal care and use guidelines with approval by the Institutional Animal Care and Use Committees at Memorial Sloan Kettering Cancer Center and the University of Georgia. Zebrafish were raised under standard conditions at 28° C. Wild-type lines were of the AB background. All mutant alleles are summarized in Supplementary File 1.

TALEN and CRISPR mutagenesis

TALEN sequences were selected using Targeter 2.0 software (Doyle et al., 2012). TAL repeat assembly was achieved using the Golden Gate assembly method, and assembled repeats were integrated into the GoldyTALEN scaffold (Bedell et al., 2012; Cermak et al., 2011). Assembled vectors served as templates for in vitro mRNA transcription using the T3 mMessage mMachine kit (Ambion) according to manufacturer's instructions. 50–100pg mRNA was injected into wild-type embryos at the one-cell stage. Injected embryos were raised to adulthood and F1 progeny were screened for germline transmission of mutations as previously described (Li et al., 2015). Primers used for detection of mutations and subsequent genotyping are included in Supplementary File 1.

Target selection for CRISPR/Cas9 mediated mutagenesis was performed using CHOPCHOP (Labun et al., 2016). sgRNA templates were generated either by cloning into pT7-gRNA as described by (Jao et al., 2013) or using the oligo-based approach described in (Gagnon et al., 2014) and (Burger et al., 2016). All template oligos are listed in Supplementary File 3. sgRNAs were in vitro transcribed from their respective templates using T7 RNA polymerase (Promega) as per manufacturer protocol. Cas9 RNA was in

vitro transcribed from the pT3TS-nls-zCas9-nls plasmid (Jao et al., 2013) using the T3 mMessage mMachine Kit (Ambion). For mutagenesis, 200-400ng of sgRNA and ~500ng of Cas9 mRNA were co-injected into wild-type embryos at the one-cell stage. Injected embryos were raised to adulthood, and F1 progeny were screened for germline transmission of mutations as previously described (Li et al., 2015). Primers used for detection of mutations and subsequent genotyping are included in Supplementary File 1.

Zebrafish imaging and length measurements

All bright field imaging of zebrafish larvae and adult was performed using Olympus MVX10 with CellSens Standard software. Standard-length was documented using ImageJ as defined in (Parichy et al., 2009). Photoshop (Adobe) adjustments to brightness and contrast were equally applied to all images of whole zebrafish in order to improve visualization.

FACS Analysis of Whole Kidney Marrow

Adult zebrafish at 6 months were sacrificed with a combination of tricaine (Sigma-Aldrich, CAS number 886-86-2) and rapid chilling. Whole kidneys were dissected using forceps and placed in 0.9× PBS/5% FCS. Manual disaggregation using a P1000 pipette resulted in single cell suspensions. Cells were filtered over a 40 μm nylon mesh filter, and resuspended in PBS/FCS to give a final concentration of 100,000 cells/μl. FACS sorting of single cells were analyzed for forward/side scatter profiles. FACS data were analyzed using FloJo software.

DNA Methylation Analysis

For Southern blot analysis, 1 μg of purified total genomic DNA was digested with the indicated methylation sensitive restriction enzyme, fractionated by electrophoresis through

a 0.9% agarose gel and transferred to nylon membrane. Probes were PCR amplified using primers in Supplementary File 2 and radiolabeled with ^{32}P-dCTP using Rediprime™ II Random Prime Labelling System (Amersham) according to manufacturer protocol. Hybridization signals were imaged and analyzed using a Typhoon phosphorimager (GE Life Sciences). Signal intensities were measured using ImageJ. Methylation changes at Sat1 was quantified as a ratio of the intensity of the unmethylated / methylated blot regions as indicated in the respective blot.

HypCH4IV was selected for Sat1 methylation analysis over the more traditional Msp1/HpaII isoschizomer pair because Sat1 sequences lack the CCGG sites that are recognized by these enzymes.

Enhanced Reduced Representation Bisulfite Sequencing (ERRBS)

50 ng of high quality genomic DNA was prepared from fin tissue from adult zebrafish at 24 wpf as previously described (Garrett-Bakelman et al., 2015). DNA was digested with Msp1. Bisulphite conversion rates (calculated using non-CpG methylation conversion rates) ranged from 99.6 to 99.7% for all samples (**Figure 2-figure supplement 2C**). Amplified libraries were sequenced on the Hiseq2000 platform for 50 cycles single end read runs. ERRBS data were filtered for sequence adapters, limited to the first 29 bp of the read (Boyle et al., 2012), and mapped to the zebrafish genome (danRer7) using BSmap (v 2.90) (Xi and Li, 2009). Methylation scores were calculated as the number of unconverted reads divided by the number of total reads at each CpG site. DMRs with at least a 0.2 change in methylation were determined using DSS (delta=0.2, p.threshold=0.01) (Park and Wu, 2016). Sat1 sequences are deficient in Msp1 sites, and are therefore not included in ERRBS data.

RNA Expression Analysis

For qRT-PCR, total RNA was isolated using Trizol (Invitrogen) and precipitated with isopropanol. RNA used for assaying expression of repeat sequences subsequently was treated with DNase using TURBO DNA-*free*™ Kit (Ambion) prior to analyses. RNA was converted to cDNA using GoScript™ Reverse Transcriptase Kit (Promega) and Real Time PCR was performed using an Applied Biosystems 7500 PCR Machine. Analysis was performed using the $2^{-\Delta\Delta Ct}$ method, with relative mRNA levels of all transcripts normalized to β-actin1. All primer sequences are listed in Supplementary File 2.

For Northern blot analysis, total RNA was extracted with using Trizol (Invitrogen). 2 μg of RNA was subjected to electrophoresis on 1% agarose gel and transferred to Amersham Hybond-N+ membrane (GE Healthcare). The membrane was probed with ^{32}P-dCTP radiolabeled Sat1 DNA probe at 42°C. Hybridization signals were imaged and analyzed using a Typhoon phosphorimager (GE Life Sciences).

TAG-aided sense/antisense transcript detection (TASA-TD) strand-specific PCR was performed as described by (Henke et al., 2015). Oligos used are listed in Supplementary File 3.

Transcriptome sequencing

After RiboGreen quantification and quality control by Agilent BioAnalyzer, 500 ng of total RNA underwent polyA selection and TruSeq library preparation according to instructions provided by Illumina (TruSeq Stranded mRNA LT Kit), with 8 cycles of PCR. Samples were barcoded and run on a HiSeq 2500 High Output in a 50bp/50bp paired end run, using the TruSeq SBS v4 Kit (Illumina). An average of 45.3 million paired reads was generated per sample. The percent of mRNA bases averaged 62.8%.

For single-mutant RNA-seq analysis presented in Fig 3, reads were mapped to the Zebrafish genome (danRer7) using the rnaStar aligner v2.5.0a (Dobin et al., 2013). We used the two pass mapping method outlined in (Engstrom et al., 2013). The first mapping pass used a list of known annotated junctions from Ensemble. Novel junctions found in the first pass were then added to the known junctions and a second mapping pass was done (on the second pass the RemoveNoncanoncial flag was used). Expression counts (counts per million, cpm) were computed from the mapped reads using HTSeq v0.5.3 (Anders et al., 2015) and Ensemble D.rerio v79 gene annotations. Normalization and differential expression was performed using DESeq (Anders and Huber, 2010).

For RNA-seq analysis presented in Fig 6, raw RNA-seq FASTQ reads were trimmed for adapters and preprocessed to remove low-quality reads using Trimmomatic v0.33 (arguments: LEADING:3 TRAILING:3 MINLEN:36) (Bolger et al., 2014) prior to mapping to the *Danio rerio* GRCz10 reference genome assembly. Reads were mapped using TopHat v2.1.1 (Kim et al., 2013) supplied with a reference General Features File (GFF) to the *Danio rerio*GRCz10 reference genome assembly, and with the following arguments: -i 10 -I 5000 --library-type fr-firststrand. Gene expression was estimated using Cuffquant (a tool from Cufflinks v2.2.1), with following arguments --library-type fr-firststrand. Expression level were normalized in FPKM units by Cuffnorm (a tool from Cufflinks v2.2.1), with following arguments --library-type fr-firststrand.

RNA synthesis and injections

Sat1 RNA and dsRed RNAs were *in vitro* transcribed using Riboprobe® in vitro transcription systems (Promega). Oligos to amplify the DNA template for *in vitro* transcription are included in Supplementary File 3. Sense and anti-sense transcripts were transcribed in vitro using the T3 and T7 RNA polymerases respectively. RNA was purified

illustra MicroSpin G-50 Columns (GE Healthcare) and 50 ng of sense and antisense RNA was co-injected into zebrafish embryos at the 1-cell stage.

Statistical Analysis

The Student unpaired 2-tailed t-test was used for statistical analysis unless specified otherwise. Statistical analysis was performed using GraphPad PRISM software.

Accession Number

All ERRBS and RNA-Seq data reported in this paper have been deposited in GEO under the accession GSE116360. The data is currently private and will be released at publication.

Acknowledgements

This research was supported by a grant from the National Institutes of Health (R01GM11009) to M.G.G. We thank the Goll laboratory for helpful discussions and critical reading of the manuscript. We acknowledge the use of the Integrated Genomics Operation Core of MSKCC, funded by the NCI Cancer Center Support Grant (CCSG, P30 CA08748), Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology. We also acknowledge the use of the Bioinformatics core of MSKCC for support with sequence analysis. ERRBS was performed in the Weill Cornell Medicine Epigenomics Core Facility.

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Anderson, R.M., Bosch, J.A., Goll, M.G., Hesselson, D., Dong, P.D., Shin, D., Chi, N.C., Shin, C.H., Schlegel, A., Halpern, M.*, et al.* (2009). Loss of Dnmt1 catalytic activity reveals multiple roles for DNA methylation during pancreas development and regeneration. Dev Biol *334*, 213-223.

Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., 2nd, Tan, W., Penheiter, S.G., Ma, A.C., Leung, A.Y.*, et al.* (2012). In vivo genome editing using a high-efficiency TALEN system. Nature *491*, 114-118.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Boyle, P., Clement, K., Gu, H., Smith, Z.D., Ziller, M., Fostel, J.L., Holmes, L., Meldrim, J., Kelley, F., Gnirke, A.*, et al.* (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. Genome Biol *13*, R92.

Burger, A., Lindsay, H., Felker, A., Hess, C., Anders, C., Chiavacci, E., Zaugg, J., Weber, L.M., Catena, R., Jinek, M.*, et al.* (2016). Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. Development *143*, 2025-2037.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res *39*, e82.

Cheon, H., Borden, E.C., and Stark, G.R. (2014). Interferons and their stimulated genes in the tumor microenvironment. Semin Oncol *41*, 156-173.

Chernyavskaya, Y., Mudbhary, R., Zhang, C., Tokarz, D., Jacob, V., Gopinath, S., Sun, X., Wang, S., Magnani, E., Madakashira, B.P.*, et al.* (2017). Loss of DNA methylation in zebrafish embryos activates retrotransposons to trigger antiviral signaling. Development *144*, 2925-2939.

Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A.*, et al.* (2015). Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. Cell *162*, 974-986.

Crowl, J.T., Gray, E.E., Pestal, K., Volkman, H.E., and Stetson, D.B. (2017). Intracellular Nucleic Acid Detection in Autoimmunity. Annu Rev Immunol *35*, 313-336.

de Greef, J.C., Wang, J., Balog, J., den Dunnen, J.T., Frants, R.R., Straasheijm, K.R., Aytekin, C., van der Burg, M., Duprez, L., Ferster, A.*, et al.* (2011). Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. Am J Hum Genet *88*, 796-804.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., Vandyk, J.K., and Bogdanove, A.J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res *40*, W117-122.

Ehrlich, M. (2003). The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clinical Immunology *109*, 17-28.

Ehrlich, M., Sanchez, C., Shao, C., Nishiyama, R., Kehrl, J., Kuick, R., Kubota, T., and Hanash, S.M. (2008). ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. Autoimmunity *41*, 253-271.

Engstrom, P.G., Steijger, T., Sipos, B., Grant, G.R., Kahles, A., Ratsch, G., Goldman, N., Hubbard, T.J., Harrow, J., Guigo, R.*, et al.* (2013). Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods *10*, 1185-1191.

Enukashvily, N.I., Donev, R., Waisertreiger, I.S., and Podgornaya, O.I. (2007). Human chromosome 1 satellite 3 DNA is decondensed, demethylated and transcribed in senescent cells and in A431 epithelial carcinoma cells. Cytogenet Genome Res *118*, 42-54.

Fanelli, M., Caprodossi, S., Ricci-Vitiani, L., Porcellini, A., Tomassoni-Ardori, F., Amatori, S., Andreoni, F., Magnani, M., De Maria, R., Santoni, A.*, et al.* (2008). Loss of pericentromeric DNA methylation pattern in human glioblastoma is associated with altered DNA methyltransferases expression and involves the stem cell compartment. Oncogene *27*, 358-365.

Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., and Schier, A.F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PLoS One *9*, e98186.

Garrett-Bakelman, F.E., Sheridan, C.K., Kacmarczyk, T.J., Ishii, J., Betel, D., Alonso, A., Mason, C.E., Figueroa, M.E., and Melnick, A.M. (2015). Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. J Vis Exp, e52246.

Geiman, T.M., Tessarollo, L., Anver, M.R., Kopp, J.B., Ward, J.M., and Muegge, K. (2001). Lsh, a SNF2 family member, is required for normal murine development. Biochim Biophys Acta *1526*, 211-220.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem *74*, 481-514.

Hartlova, A., Erttmann, S.F., Raffi, F.A., Schmalz, A.M., Resch, U., Anugula, S., Lienenklaus, S., Nilsson, L.M., Kroger, A., Nilsson, J.A.*, et al.* (2015). DNA damage primes the type I interferon system via the cytosolic DNA sensor STING to promote anti-microbial innate immunity. Immunity *42*, 332-343.

Henke, C., Strissel, P.L., Schubert, M.T., Mitchell, M., Stolt, C.C., Faschingbauer, F., Beckmann, M.W., and Strick, R. (2015). Selective expression of sense and antisense transcripts of the sushi-ichi-related retrotransposon--derived family during mouse placentogenesis. Retrovirology *12*, 9.

Jao, L.E., Wente, S.R., and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc Natl Acad Sci U S A *110*, 13904-13909.

Jenness, C., Giunta, S., Muller, M.M., Kimura, H., Muir, T.W., and Funabiki, H. (2018). HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. Proc Natl Acad Sci U S A *115*, E876-E885.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B., and Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Research *44*, W272-W276.

Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195-3205.

628 Leonova, K.I., Brodsky, L., Lipchick, B., Pal, M., Novototskaya, L., Chenchik, A.A., Sen,
629 G.C., Komarova, E.A., and Gudkov, A.V. (2013). p53 cooperates with DNA methylation
630 and a suicidal interferon response to maintain epigenetic silencing of repeats and
631 noncoding RNAs. Proc Natl Acad Sci U S A *110*, E89-98.
632 Li, C., Lan, Y., Schwartz-Orbach, L., Korol, E., Tahiliani, M., Evans, T., and Goll, M.G.
633 (2015). Overlapping Requirements for Tet2 and Tet3 in Normal Development and
634 Hematopoietic Stem Cell Emergence. Cell Rep *12*, 1133-1143.
635 Mankan, A.K., Schmidt, T., Chauhan, D., Goldeck, M., Honing, K., Gaidt, M., Kubarenko,
636 A.V., Andreeva, L., Hopfner, K.P., and Hornung, V. (2014). Cytosolic RNA:DNA hybrids
637 activate the cGAS-STING axis. EMBO J *33*, 2937-2946.
638 Nakagawa, T., Kanai, Y., Ushijima, S., Kitamura, T., Kakizoe, T., and Hirohashi, S. (2005).
639 DNA hypomethylation on pericentromeric satellite regions significantly correlates with loss
640 of heterozygosity on chromosome 9 in urothelial carcinomas. J Urol *173*, 243-246.
641 Narayan, A., Ji, W., Zhang, X.Y., Marrogi, A., Graff, J.R., Baylin, S.B., and Ehrlich, M.
642 (1998). Hypomethylation of pericentromeric DNA in breast adenocarcinomas. Int J Cancer
643 *77*, 833-838.
644 Parichy, D.M., Elizondo, M.R., Mills, M.G., Gordon, T.N., and Engeszer, R.E. (2009).
645 Normal table of postembryonic zebrafish development: staging by externally visible
646 anatomy of the living fish. Dev Dyn *238*, 2975-3015.
647 Park, Y., and Wu, H. (2016). Differential methylation analysis for BS-seq data under
648 general experimental design. Bioinformatics *32*, 1446-1453.
649 Phillips, R.B., and Reed, K.M. (2000). Localization of repetitive DNAs to zebrafish (Danio
650 rerio) chromosomes by fluorescence in situ hybridization (FISH). Chromosome Res *8*, 27-
651 35.
652 Qu, G.-z., Grundy, P.E., Narayan, A., and Ehrlich, M. (1999). Frequent Hypomethylation
653 in Wilms Tumors of Pericentromeric DNA in Chromosomes 1 and 16. Cancer Genetics
654 and Cytogenetics *109*, 34-39.
655 Rai, K., Nadauld, L.D., Chidester, S., Manos, E.J., James, S.R., Karpf, A.R., Cairns, B.R.,
656 and Jones, D.A. (2006). Zebra fish Dnmt1 and Suv39h1 regulate organ-specific terminal
657 differentiation during development. Mol Cell Biol *26*, 7077-7085.
658 Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen, S.Y., Han, H., Liang,
659 G., Jones, P.A., Pugh, T.J*., et al.* (2015). DNA-Demethylating Agents Target Colorectal
660 Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. Cell *162*, 961-973.
661 Schneider, W.M., Chevillotte, M.D., and Rice, C.M. (2014). Interferon-stimulated genes: a
662 complex web of host defenses. Annu Rev Immunol *32*, 513-545.
663 Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development.
664 Nat Rev Genet *14*, 204-220.
665 Stancheva, I., and Meehan, R.R. (2000). Transient depletion of xDnmt1 leads to
666 premature gene activation in Xenopus embryos. Genes Dev *14*, 313-327.
667 Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from
668 epigenomics. Nat Rev Genet *9*, 465-476.
669 Suzuki, T., Fujii, M., and Ayusawa, D. (2002). Demethylation of classical satellite 2 and 3
670 DNA with chromosomal instability in senescent human fibroblasts. Exp Gerontol *37*, 1005-
671 1014.
672 Tanne, A., Muniz, L.R., Puzio-Kuter, A., Leonova, K.I., Gudkov, A.V., Ting, D.T.,
673 Monasson, R., Cocco, S., Levine, A.J., Bhardwaj, N*., et al.* (2015). Distinguishing the
674 immunostimulatory properties of noncoding RNAs expressed in cancer cells. Proc Natl
675 Acad Sci U S A *112*, 15154-15159.
676 Tao, Y., Xi, S., Shan, J., Maunakea, A., Che, A., Briones, V., Lee, E.Y., Geiman, T.,
677 Huang, J., Stephens, R*., et al.* (2011). Lsh, chromatin remodeling family member,

678 modulates genome-wide cytosine methylation patterns at nonrepeat sequences. Proc Natl
679 Acad Sci U S A *108*, 5626-5631.
680 Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara,
681 M., Suyama, M., Sun, Y.*, et al.* (2015). Mutations in CDCA7 and HELLS cause
682 immunodeficiency-centromeric instability-facial anomalies syndrome. Nat Commun *6*,
683 7870.
684 Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino,
685 G., Deshpande, V., Iafrate, A.J., Letovsky, S.*, et al.* (2011). Aberrant overexpression of
686 satellite repeats in pancreatic and other epithelial cancers. Science *331*, 593-596.
687 Trede, N.S., David M. Langenau, David Traver, A. Thomas Look, and Zon, a.L.I. (2004).
688 The Use of Zebrafish to Understand Immunity. Immunity *20*, 367-379.
689 Tsuda, H., Takarabe, T., Kanai, Y., Fukutomi, T., and Hirohashi, S. (2002). Correlation of
690 DNA hypomethylation at pericentromeric heterochromatin regions of chromosomes 16
691 and 1 with histological features and chromosomal abnormalities of human breast
692 carcinomas. Am J Pathol *161*, 859-866.
693 Tuck-Muller, C.M., Narayan, A., Tsien, F., Smeets, D.F., Sawyer, J., Fiala, E.S., Sohn,
694 O.S., and Ehrlich, M. (2000). DNA hypomethylation and unusual chromosome instability
695 in cell lines from ICF syndrome patients. Cytogenet Cell Genet *89*, 121-128.
696 Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K., and Li, E. (2006). Roles
697 for Dnmt3b in mammalian development: a mouse model for the ICF syndrome.
698 Development *133*, 1183-1192.
699 van der Vaart, M., van Soest, J.J., Spaink, H.P., and Meijer, A.H. (2013). Functional
700 analysis of a zebrafish myd88 mutant identifies key transcriptional components of the
701 innate immune system. Dis Model Mech *6*, 841-854.
702 Velasco, G., Grillo, G., Touleimat, N., Ferry, L., Ivkovic, I., Ribierre, F., Deleuze, J.F.,
703 Chantalat, S., Picard, C., and Francastel, C. (2018). Comparative methylome analysis of
704 ICF patients identifies heterochromatin loci that require ZBTB24, CDCA7 and HELLS for
705 their methylated state. Hum Mol Genet.
706 von Bernuth, H., Ravindran, E., Du, H., Frohler, S., Strehl, K., Kramer, N., Issa-Jahns, L.,
707 Amulic, B., Ninnemann, O., Xiao, M.S.*, et al.* (2014). Combined immunodeficiency
708 develops with age in Immunodeficiency-centromeric instability-facial anomalies syndrome
709 2 (ICF2). Orphanet J Rare Dis *9*, 116.
710 Weisenberger, D.J., Campan, M., Long, T.I., Kim, M., Woods, C., Fiala, E., Ehrlich, M.,
711 and Laird, P.W. (2005). Analysis of repetitive element DNA methylation by MethyLight.
712 Nucleic Acids Res *33*, 6823-6836.
713 Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K., Wang, J., den Hamer, B., Aytekin, C., van
714 der Maarel, S.M., and Daxinger, L. (2016). Converging disease genes in ICF syndrome:
715 ZBTB24 controls expression of CDCA7 in mammals. Hum Mol Genet *25*, 4041-4051.
716 Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program.
717 BMC Bioinformatics *10*, 232.
718 Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M.,
719 Qu, X., Russo, J.J., and Viegas-Pequignot, E. (1999). Chromosome instability and
720 immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene.
721 Nature *402*, 187-191.
722
723

Figure Legends

Figure 1. Mutation of *zbtb24* causes ICF syndrome-like phenotypes in zebrafish.

(**A**) Schematic of human and zebrafish Zbtb24 proteins. The BTB/POZ domain is indicated in blue and C2H2-type zinc fingers in dark grey. (**B**) Schematic of zebrafish *zbtb24* gene. Location of TALEN target sequences are indicated in red (not to scale). Brackets indicate the region deleted by the $zbtb24^{mk22(\Delta)}$ allele. (**C**) Representative images of $zbtb24^{+/+}$ and $zbtb24^{\Delta/\Delta}$ zebrafish at 2 wpf. Scale bar: 1 mm. (**D**) Standard length measurements for $zbtb24^{+/+}$ and $zbtb24^{\Delta/\Delta}$ zebrafish at 1, 2, 3 and 4 wpf (n≥6 for each group). (**E**) Representative images of $zbtb24^{+/+}$ and $zbtb24^{\Delta/\Delta}$ zebrafish at 5 months. Scale bar: 5 mm. (**F**) Average weight of $zbtb24^{+/+}$ and $zbtb24^{\Delta/\Delta}$ zebrafish at 5 months (n=5 for each group). (**G**) Representative images of facial abnormalities in $zbtb24^{+/+}$ and $zbtb24^{\Delta/\Delta}$ adults at 6 months. Scale bar: 2 mm. (**H**) Schematic and quantification of facial abnormalities in $zbtb24^{\Delta/\Delta}$ zebrafish (n=5 for each group). (**I**) Abundance of *IgM*, *IgD* and *IgZ* transcripts in $zbtb24^{+/+}$ and $zbtb24^{\Delta/\Delta}$ zebrafish at 6 weeks post fertilization (n=5 for each group). (**J**) Quantification of lymphoid cell populations in total blood isolated from $zbtb24^{+/+}$ or $zbtb24^{\Delta/\Delta}$ kidney marrow from adults, measured by Forward/Side scatter flow cytometry (n=11 for each group). (**K**) Kaplan-Meier curve indicating survival among groups of $zbtb24^{+/+}$ and $zbtb24^{\Delta/\Delta}$ zebrafish (n=12 for each group). All error bars indicate standard error of the mean (SEM).

Figure 2. Mutation of *zbtb24* causes progressive methylation loss at pericentromeric satellite repeats.

(A) Southern blot of genomic DNA digested with 5mC-sensitive restriction enzyme HpyCH4IV and probed with zebrafish Sat1 sequence. Each lane represents DNA isolated from one adult individual of the indicated genotype. DNA from $dnmt1^{-/-}$ zebrafish larvae at 7 days post fertilization and their phenotypically wild-type siblings (WT) provides a positive control. **(B)** Quantification of methylation changes at Sat1 sequences in panel **A**. Error bars indicate SEM from the 3 biological replicates. **(C)** Southern blot of genomic DNA digested with 5mC-sensitive restriction enzyme HpyCH4IV and probed with zebrafish Sat1 sequence. Genomic DNA was isolated from $zbtb24^{+/+}$ and $zbtb24^{\Delta/\Delta}$ animals at 1, 2, 4, 16 and 32 wpf as indicated. **(D)** Quantification of methylation changes at Sat1 sequences in panel **C**. Data represent averages from two independent experiments. Error bars

represent the standard deviation (SD). **(E)** Heat map of CpG methylation in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ as assessed by ERRBS (Data reflects 3 biological replicates of each genotype).

Figure 3. Mutation of *zbtb24* leads to activation of innate immune response genes. (**A**) Volcano plot representation of differential gene expression in *zbtb24*$^{+/+}$ vs *zbtb24*$^{\Delta/\Delta}$ zebrafish at 2 wpf. Blue and red points mark genes with >2 fold downregulation or upregulation respectively. (**B**) RNA-seq heatmap showing innate immune genes upregulated in *zbtb24*$^{\Delta/\Delta}$ mutant compared to *zbtb24*$^{+/+}$ siblings. Shown are Z-score normalized gene expression values. (**C**) GSEA of a set of genes involved in Response to Virus in zebrafish comparing *zbtb24*$^{+/+}$ vs *zbtb24*$^{\Delta/\Delta}$. NES, normalized enrichment score; FDR, false discovery rate. (**D**) qRT-PCR demonstrating upregulated interferon and inflammatory response genes in *zbtb24*$^{\Delta/\Delta}$ mutants at 3 wpf. Expression levels are reported relative to *β-actin.* Error bars indicate SEM from at least 3 independent biological replicates with n=8 total animals for each replicate. (**E**) qRT-PCR analysis reveals similar expression of interferon genes in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ larvae at 1 wpf. Error bars represent SEM from at least 5 biological replicates.

Figure 4. Interferon response in *zbtb24* mutants is mediated by sensors of cytosolic RNA.

(**A**) Expression of interferon signaling genes *stat1b* and *irf7* in indicated genotypes at 3 wpf. n = 4 biological replicates. (**B**) Expression of the ISGs *stat1b* and *irf7* in indicated genotypes at 3 wpf. n ≥ 7 biological replicates. (**C**) Expression of interferon signaling genes *stat1b* and *irf7* in indicated genotypes at 3 wpf. n ≥ 5 biological replicates. All error bars indicate SEM.

Figure 5. Pericentromeric transcripts are sufficient to induce the innate immune response in *zbtb24* mutants.

(**A**) Northern blot analysis of Sat1 transcripts in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 6 wpf. Each lane represents a biological replicate. The lower panel represents the cropped ethidium-bromide stained gel as loading control. (**B**) qRT-PCR for Sat1 transcripts in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 1, 3 and 6 wpf. Error bars indicate SEM of at least 4 biological replicates in each group. (**C**) Correlation between the expression of Sat1 and *irf7* in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ at 6 weeks (n=15). (**D**) TASA-TD PCR amplified sense (s) and antisense (as) transcripts Sat1 (114 bp) and *β-actin* (125bp) from first strand *zbtb24*$^{\Delta/\Delta}$ cDNA. PCR primers: gene-specific (GS); TAG. The products from TASA-TD PCR were

run on the same gel, then cropped and presented. This panel is representative of two independent biological replicates. (**E**) Quantification of TASA-TD from panel **D**. Error bars indicate SD from 2 biological replicates. (**F**) Expression of interferon stimulated genes in wild-type embryos injected with Sat1 or dsRed RNA. 50pg of in vitro transcribed sense and antisense transcripts were injected into wild-type zebrafish embryos at the 1-cell stage. Total RNA was extracted at 8 hours post fertilization for qRT-PCR analysis. Error bars indicate SEM from at least 3 biological replicates with n=20 embryos for each biological replicate.

Figure 6. Mutation of cytosolic dsRNA receptor Mda5 mitigates the interferon response in *zbtb24*$^{\Delta/\Delta}$ zebrafish.

(**A**) Expression of interferon signaling genes *stat1b* and *irf7* in indicated genotypes at 3 wpf. n ≥ 7 biological replicates. (**B**) RNA-seq heatmap of interferon stimulated genes upregulated in *zbtb24*$^{\Delta/\Delta}$ zebrafish and rescued in *mda5*$^{mk29/mk29}$*; zbtb24*$^{\Delta/\Delta}$ zebrafish at 3 wpf. Shown are Z-score normalized gene expression values. (**C**) Expression of interferon signaling genes *stat1b* and *irf7* in indicated genotypes at 6 wpf. n=6 biological replicates. (**D**) Model for the activation of interferon response in *zbtb24* mutants. Loss of Zbtb24 function causes hypomethylation of pericentromeric Sat1 repeats, which leads to derepression of associated Sat1 transcripts. These pericentromeric transcripts are recognized by the dsRNA helicase Mda5 which signals through Mavs and Irf7 to upregulate ISGs. Autoregulatory feedback implicates *irf7* as both an ISG and a key downstream effector of Mda5/Mavs signaling.

Figure 1-figure supplement 1. Zbtb24 conservation in vertebrate species.

(**A**) Sequence alignment of mouse (Mm), human (Hs) and zebrafish (Dr) Zbtb24 showing conservation of BTB (blue) and Zinc finger (orange) domains. (**B**) Phylogenetic tree of Zbtb24 with closest members of Zbtb family of proteins, Zbtb14 and Zbtb16 mouse, rat, human, chick, and Tetradon. Alignments were performed on full-length amino acid sequences via ClustalW and DRAWTREE (http://mobyle.pasteur.fr/cgi-bin/portal.py).

Figure 1-figure supplement 2. TALEN design for introducing mutations at the endogenous *zbtb24* zebrafish gene.

(**A**) Schematic of TALEN sequences showing target sites for introducing mutations in *zbtb24*. P1, P2 and P3 indicate locations for genotyping primers. Sequence in blue indicate target site. Sequence in yellow indicates site of restriction enzyme digestion. (**B**)

Sequence trace confirming generation of large deletion, *zbtb24*$^{mk22(\Delta)}$. (**C**) Representative genotyping of *zbtb24*$^{mk22(\Delta)}$ allele. L: Ladder. P1, P2, P3 represent primers from panel A used for amplifying product in specified lane. (**D**) qRT-PCR analysis of *zbtb24* mRNA in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 2 wpf (n=6 for each group).

Figure 1-figure supplement 3. A second mutant allele of *zbtb24* recapitulates key features of ICF Syndrome.

(**A**) Schematic of *zbtb24*^{mk19} with premature stop codon at aa 82. (**B**) Representative genotyping of *zbtb24*^{mk19} allele. (**C**) Sequence trace confirming generation of small deletion allele, *zbtb24*^{mk19}. (**D**) Representative bright field image of *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish at 4 days and 2 months after fertilization. Scale bar: 1mm. (**E**) Quantification of body weight of *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish at 5 months (n ≥ 6 biological replicates). (**F**) Facial abnormalities in *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish. Scale bar: 3 mm. (**G**) qRT-PCR analysis of IgM, IgD and IgZ zebrafish immunoglobulins in *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish measured at 6 weeks post fertilization (n=5 biological replicates). (**H**) Quantification of lymphoid cell populations in total blood isolated from *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ adult kidney marrow, measured by Forward/Side scatter flow cytometry (n=16 biological replicates). All error bars indicate standard error of the mean (SEM).

Figure 2-figure supplement 1. *Zbtb24* mutation causes methylation loss at pericentromeric repeats.

(**A**) Southern blot of genomic DNA digested with 5mC-sensitive restriction enzyme HpyCH4IV and probed with zebrafish Sat1 sequence. Genomic DNA was isolated at 1 month from *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ animals. Each lane represents one biological replicate for the indicated genotype. (**B**) Southern blot of genomic DNA from different tissues in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ digested with 5mC-sensitive restriction enzyme HpyCH4IV and probed with zebrafish Sat1 sequence. Each lane represents pooled DNA samples of the indicated tissue from 3 zebrafish adults.

Figure 2-figure supplement 2. *Zbtb24* mutants exhibit modest reductions in 5mC at non-pericentromeric sequences

(A-B) Southern Blot and quantification of HpyCH4IV digestion at Sat1 sequences in fin tissues used for ERRBS. (**C**) Basic statistics of ERRBS analysis in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish. (**D**) Venn diagram summarizing the overlap between differentially upregulated

genes from RNA-Seq (right circle) and genes within 100kb of all DMRs in ERRBS tested for differential expression (left circle).

Figure 2-figure supplement 3. Methylation at interspersed repeats is unaffected in *zbtb24* mutants.

Southern blot analysis of DNA methylation at interspersed repeats (**A**) Short Interspersed Nuclear Element (SINE), DANA and SINE_HE1 (**B**) Long Interspersed Nuclear Element (LINE), L1-10 (**C**) DNA Transposon, Kolobok (**D**) Endogenous Retroviral (ERV) elements, ZFERV2, ERV1-3, ERV4 and Gypsy21 in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ 1-month old zebrafish. Each lane represents a biological replicate of the indicated genotype. Genomic DNA from *dnmt1*$^{-/-}$ embryos is used as a positive control. The methylation-sensitive enzyme used to digest genomic DNA is indicated below the respective blot.

Figure 3-figure supplement 1. Mutation in *zbtb24* leads to activation of innate immune response pathways.

(**A**) qRT-PCR validation of genes that are part of the innate immune response pathway that are also upregulated in *zbtb24*$^{mk19/mk19}$ zebrafish at 3 wpf. Error bars indicate SEM from 5 biological replicates.

Figure 4-figure supplement 1. Mutation of zebrafish orthologs of *mavs*, *sting*, and *mda5*.

(**A**) Schematic of mutation in *mavs* including position of targeted gRNA, sequence information of the mutation, and location of predicted STOP codon. (**B**) Schematic of mutation in *sting* including position of targeted gRNA, sequence information of the mutation, and location of predicted STOP codon. (**C**) Schematic of mutation in *mda5* including position of targeted gRNA, sequence information of the mutation, and location of predicted STOP codon. The site of deletion is indicated with a red arrow on the sequence trace.

Figure 5-figure supplement 1. Mutation in *zbtb24* upregulates Sat1 transcripts but not transposons.

(**A**) qRT-PCR of Sat1 RNA in *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish at 6 wpf. Error bars indicate SEM from at least 3 biological replicates. (**B**) qRT-PCR reveals similar levels of expression from transposable elements in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 6 wpf. Error bars indicate SEM from 4-8 biological replicates.

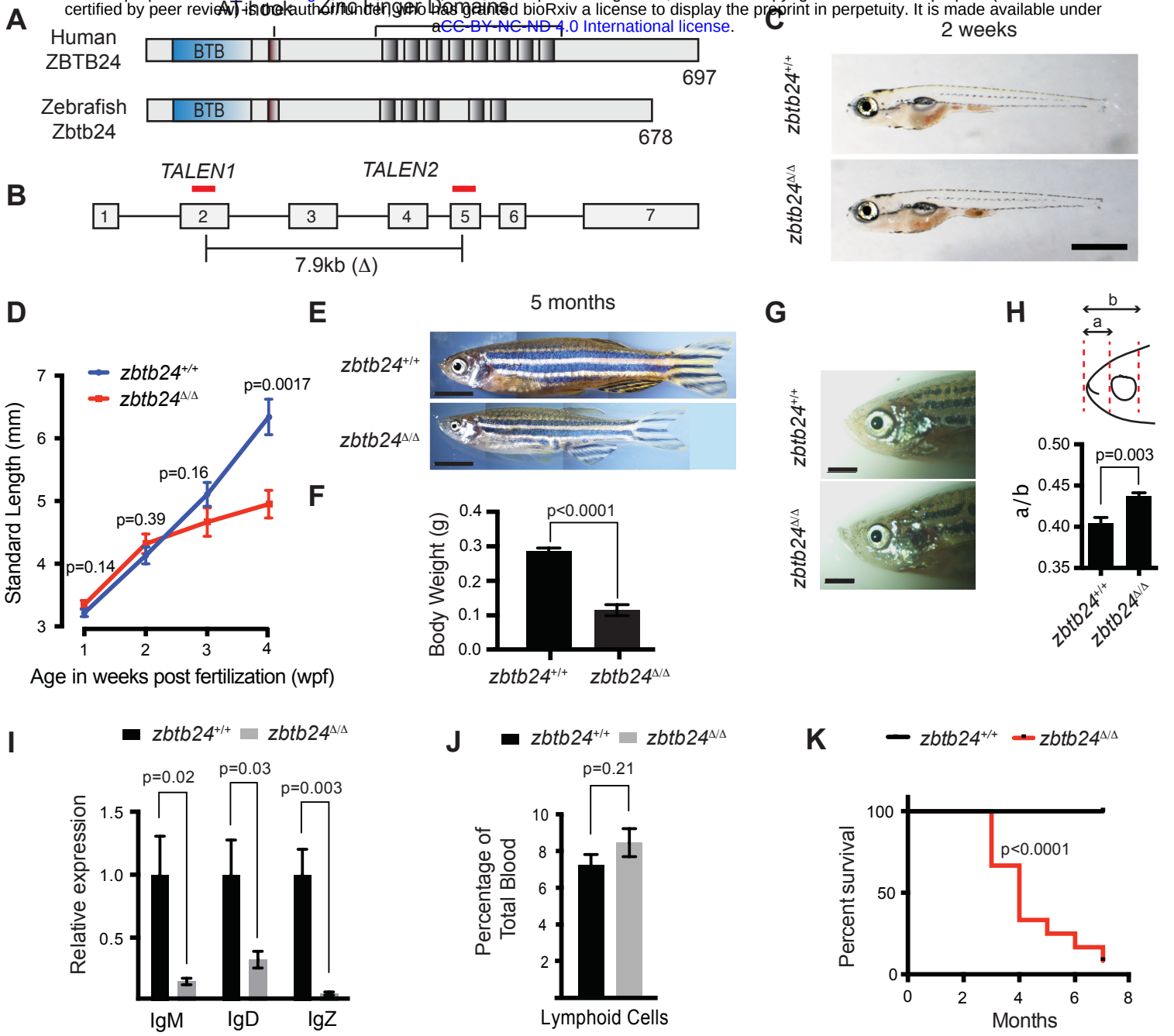


Figure 1. Mutation of *zbtb24* causes ICF syndrome-like phenotypes in zebrafish.
(**A**) Schematic of human and zebrafish Zbtb24 proteins. The BTB/POZ domain is indicated in blue and C2H2-type zinc fingers in dark grey. (**B**) Schematic of zebrafish *zbtb24* gene. Location of TALEN target sequences are indicated in red (not to scale). Brackets indicate the region deleted by the *zbtb24*$^{mk22(\Delta)}$ allele. (**C**) Representative images of *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 2 wpf. Scale bar: 1 mm. (**D**) Standard length measurements for *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 1, 2, 3 and 4 wpf (n≥6 for each group). (**E**) Representative images of *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 5 months. Scale bar: 5 mm. (**F**) Average weight of *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 5 months (n=5 for each group). (**G**) Representative images of facial abnormalities in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ adults at 6 months. Scale bar: 2 mm. (**H**) Schematic and quantification of facial abnormalities in *zbtb24*$^{\Delta/\Delta}$ zebrafish (n=5 for each group). (**I**) Abundance of *IgM*, *IgD* and *IgZ* transcripts in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 6 weeks post fertilization (n=5 for each group). (**J**) Quantification of lymphoid cell populations in total blood isolated from *zbtb24*$^{+/+}$ or *zbtb24*$^{\Delta/\Delta}$ kidney marrow from adults, measured by Forward/Side scatter flow cytometry (n=11 for each group). (**K**) Kaplan-Meier curve indicating survival among groups of *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish (n=12 for each group). All error bars indicate standard error of the mean (SEM).

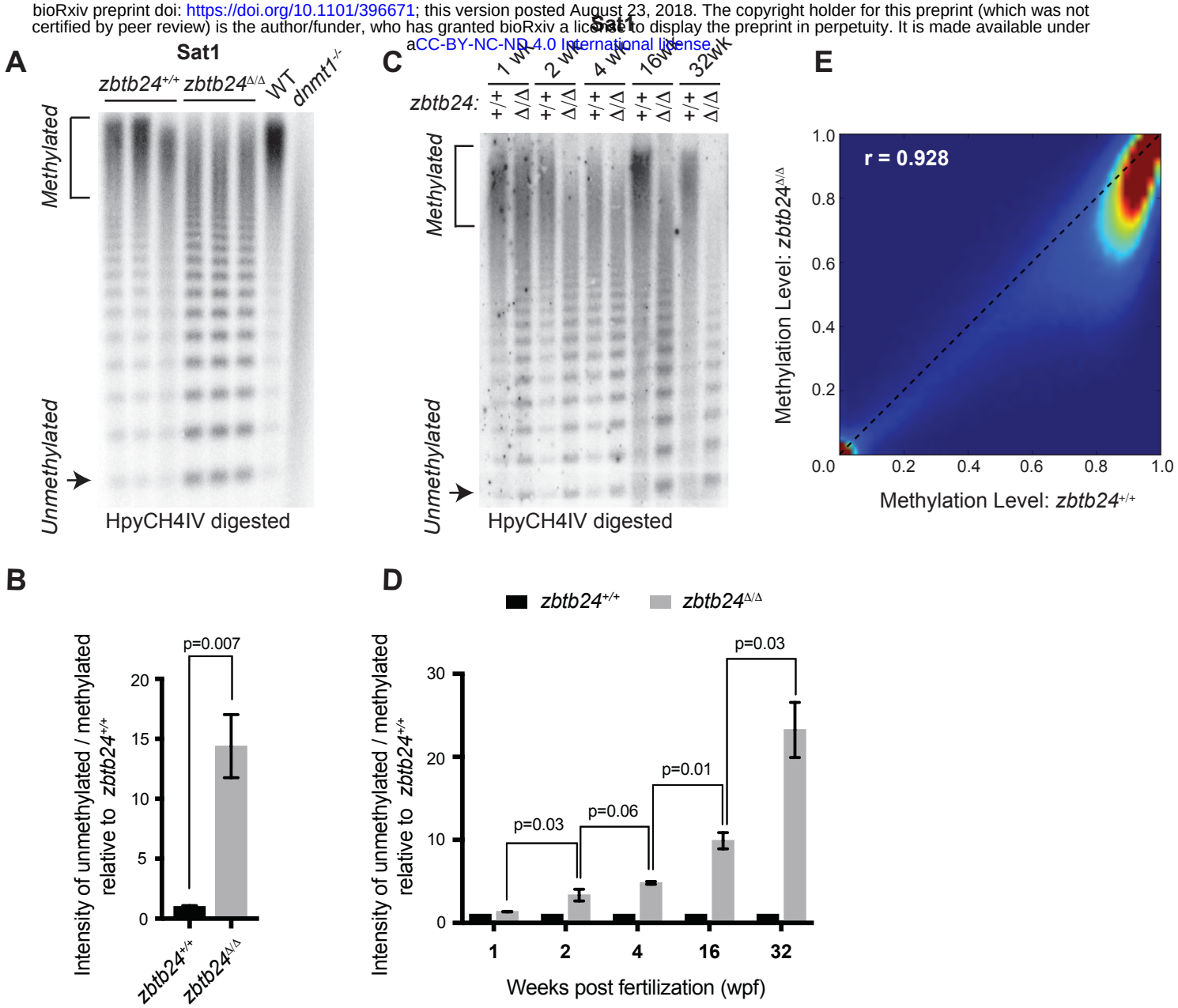


Figure 2. Mutation of *zbtb24* causes progressive methylation loss at pericentromeric satellite repeats.
(**A**) Southern blot of genomic DNA digested with 5mC-sensitive restriction enzyme HpyCH4IV and probed with zebrafish Sat1 sequence. Each lane represents DNA isolated from one adult individual of the indicated genotype. DNA from *dnmt1*$^{-/-}$ zebrafish larvae at 7 days post fertilization and their phenotypically wild-type siblings (WT) provides a positive control. (**B**) Quantification of methylation changes in panel A. Error bars indicate SEM from the 3 biological replicates. (**C**) Southern blot of genomic DNA digested with 5mC-sensitive restriction enzyme HpyCH4IV and probed with zebrafish Sat1 sequence. Genomic DNA was isolated from *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ animals at 1, 2, 4, 16 and 32 wpf as indicated. (**D**) Quantification of methylation changes at Sat1 sequences in panel C. Data represent averages from two independent experiments. Error bars represent the standard deviation (SD). (**E**) Heat map of CpG methylation in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ as assessed by ERRBS (Data reflects 3 biological replicates of each genotype).

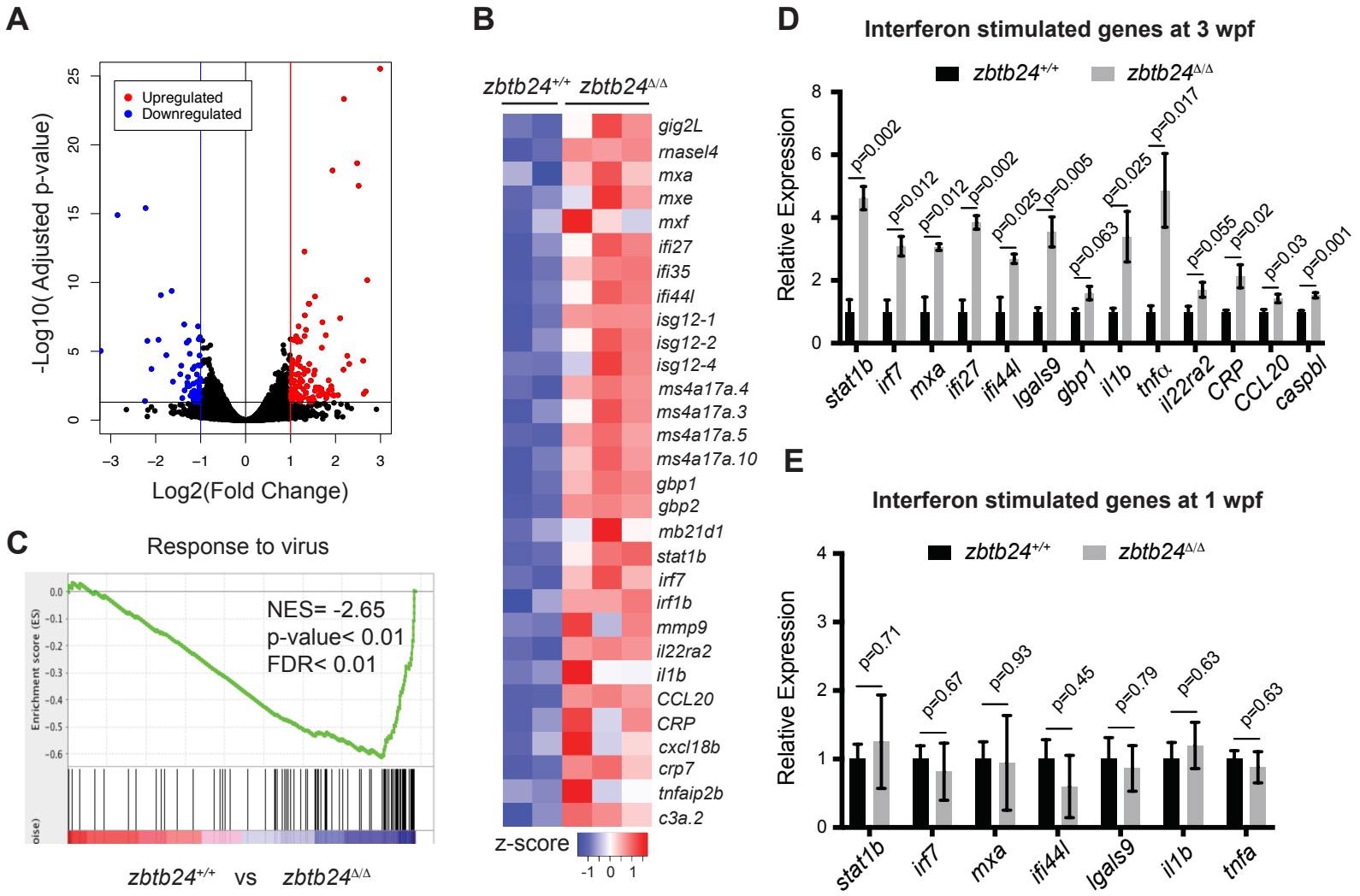


Figure 3. Mutation of *zbtb24* leads to activation of innate immune response genes.
(**A**) Volcano plot representation of differential gene expression in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 2 wpf. Blue and red points mark genes with >2 fold downregulation or upregulation respectively. (**B**) RNA-seq heatmap showing innate immune genes upregulated in *zbtb24*$^{\Delta/\Delta}$ mutant compared to *zbtb24*$^{+/+}$ siblings. Shown are Z-score normalized gene expression values. (**C**) GSEA of a set of genes involved in Response to Virus in zebrafish comparing *zbtb24*$^{+/+}$ vs *zbtb24*$^{\Delta/\Delta}$. NES, normalized enrichment score; FDR, false discovery rate. (**D**) qRT-PCR demonstrating upregulated interferon and inflammatory response genes in *zbtb24*$^{\Delta/\Delta}$ mutants at 3 wpf. Expression levels are reported relative to β-actin. Error bars indicate SEM from at least 3 independent biological replicates with n=8 total animals for each replicate. (**E**) qRT-PCR analysis reveals similar expression of interferon genes in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ larvae at 1 wpf. Error bars represent SEM from at least 5 biological replicates.

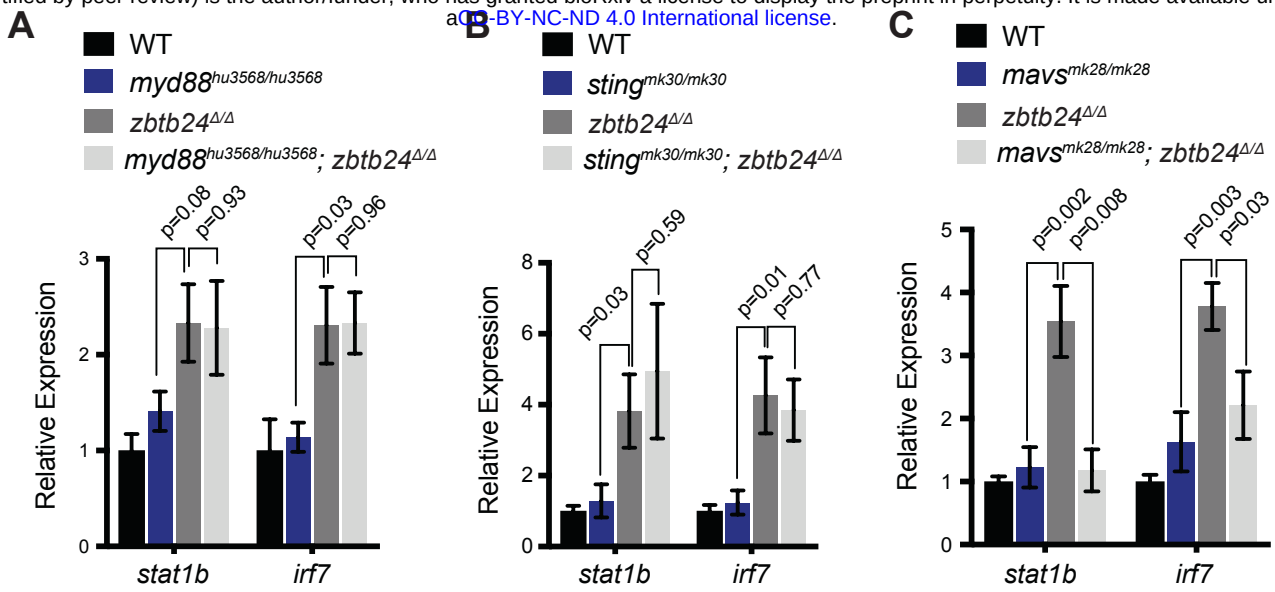


Figure 4. Interferon response in *zbtb24* mutants is mediated by sensors of cytosolic RNA.
(**A**) Expression of interferon signaling genes *stat1b* and *irf7* in indicated genotypes at 3 wpf. n = 4 biological replicates. (**B**) Expression of the ISGs *stat1b* and *irf7* in indicated genotypes at 3 wpf. n ≥ 7 biological replicates. (**C**) Expression of interferon signaling genes *stat1b* and *irf7* in indicated genotypes at 3wpf. n ≥ 5 biological replicates. All error bars indicate SEM.

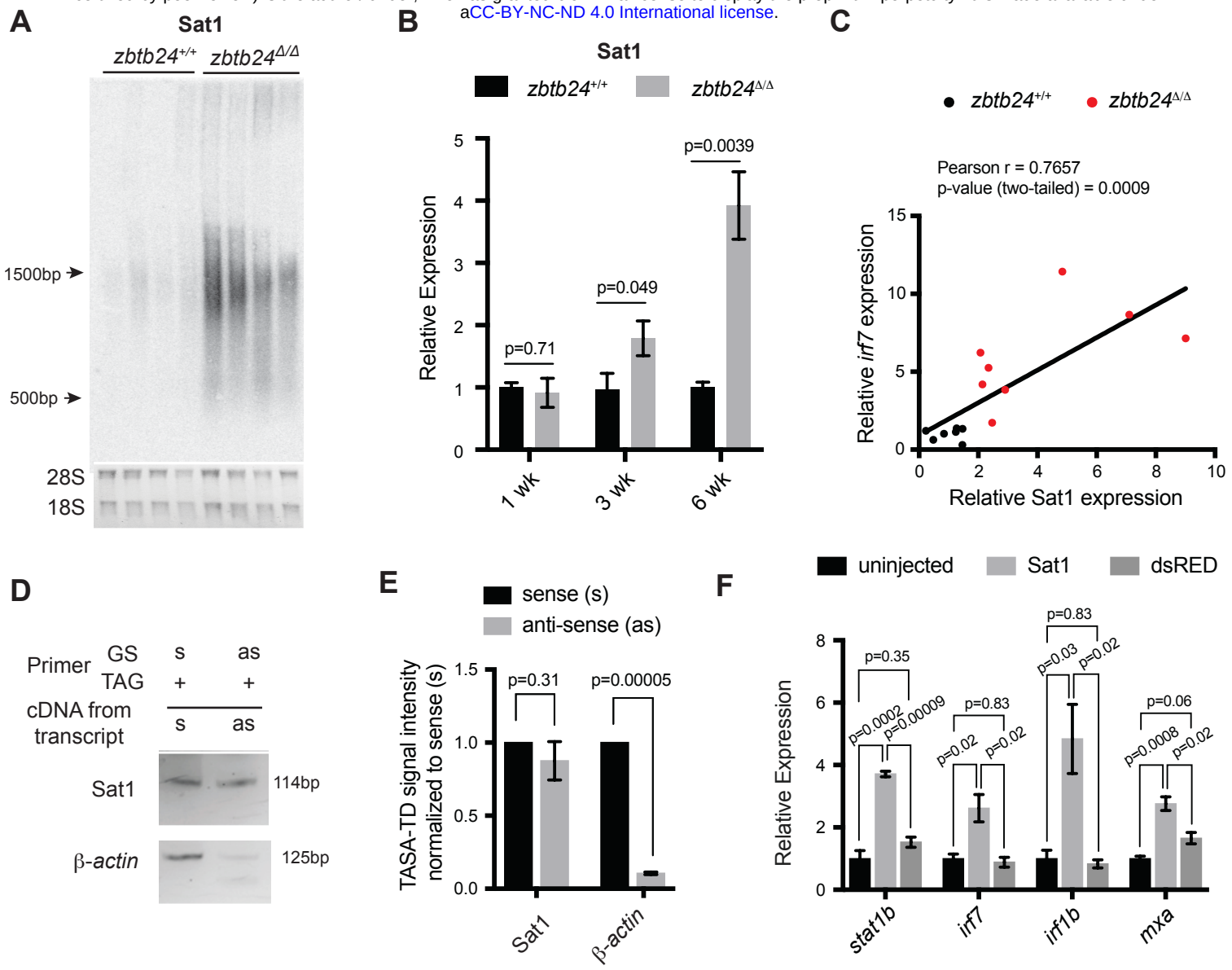


Figure 5. Pericentromeric transcripts are sufficient to induce the innate immune response in *zbtb24* mutants.
(**A**) Northern blot analysis of Sat1 transcripts in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 6 wpf. Each lane represents a biological replicate. The lower panel represents the cropped ethidium-bromide stained gel as loading control. (**B**) qRT-PCR for Sat1 transcripts in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 1, 3 and 6 wpf. Error bars indicate SEM of at least 4 biological replicates in each group. (**C**) Correlation between the expression of Sat1 and *irf7* in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ at 6 weeks (n=15). (**D**) TASA-TD PCR amplified sense (s) and antisense (as) transcripts Sat1 (114 bp) and β-*actin* (125bp) from first strand *zbtb24*$^{\Delta/\Delta}$ cDNA. PCR primers: gene-specific (GS); TAG. The products from TASA-TD PCR were run on the same gel, then cropped and presented. (**E**) Quantification of TASA-TD from panel **D**. Error bars indicate SD from 2 biological replicates. (**F**) Expression of interferon stimulated genes *stat1b*, *irf7, irf1b* and *mxa* in wildtype embryos injected with Sat1 or dsRed RNA. 50pg of *in vitro* transcribed sense and antisense transcripts were injected into wildtype zebrafish embryos at the 1-cell stage. Total RNA was extracted at 8 hours post fertilization for qRT-PCR analysis. Error bars indicate SEM from at least 3 biological replicates with n=20 embryos for each biological replicate.

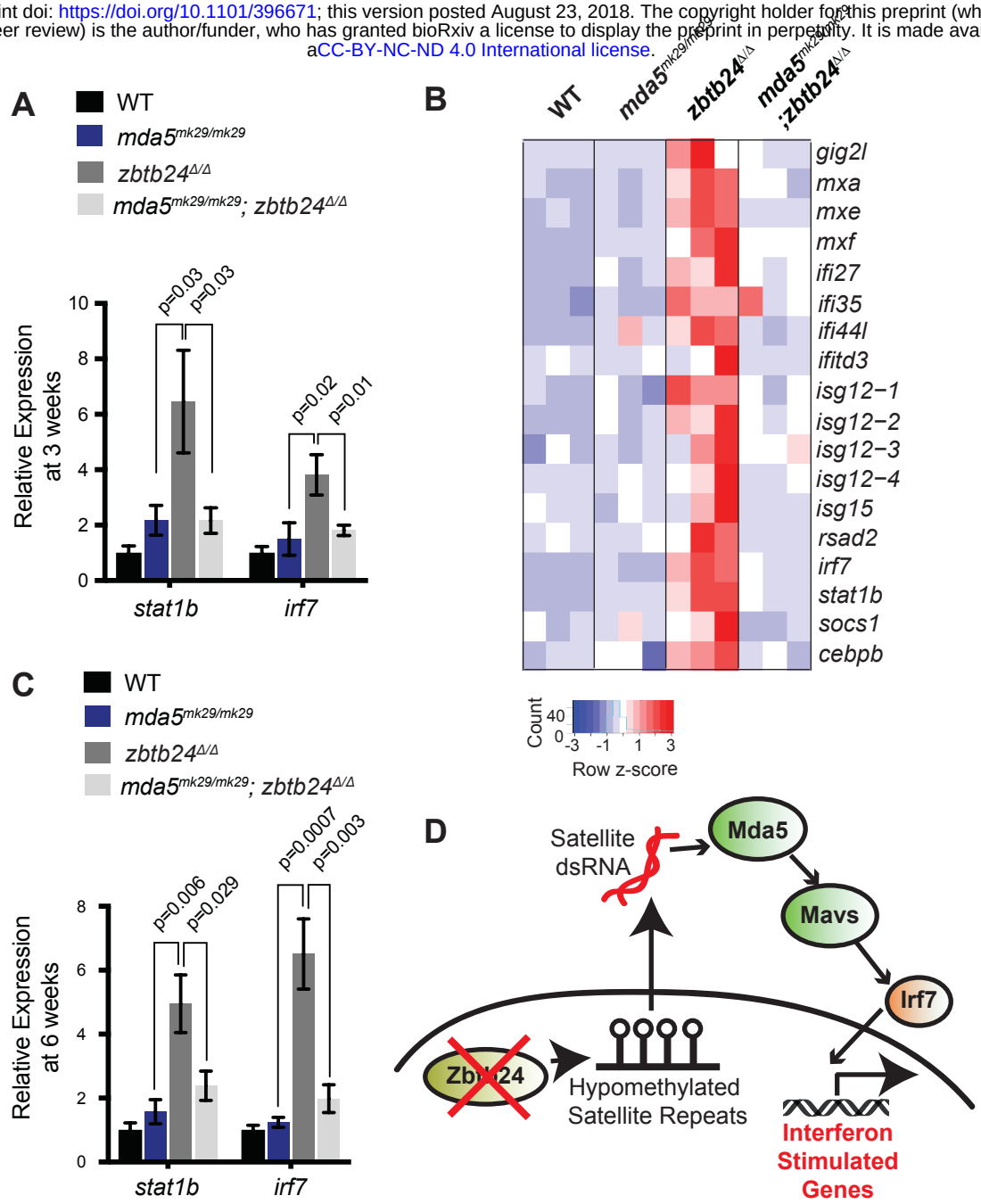
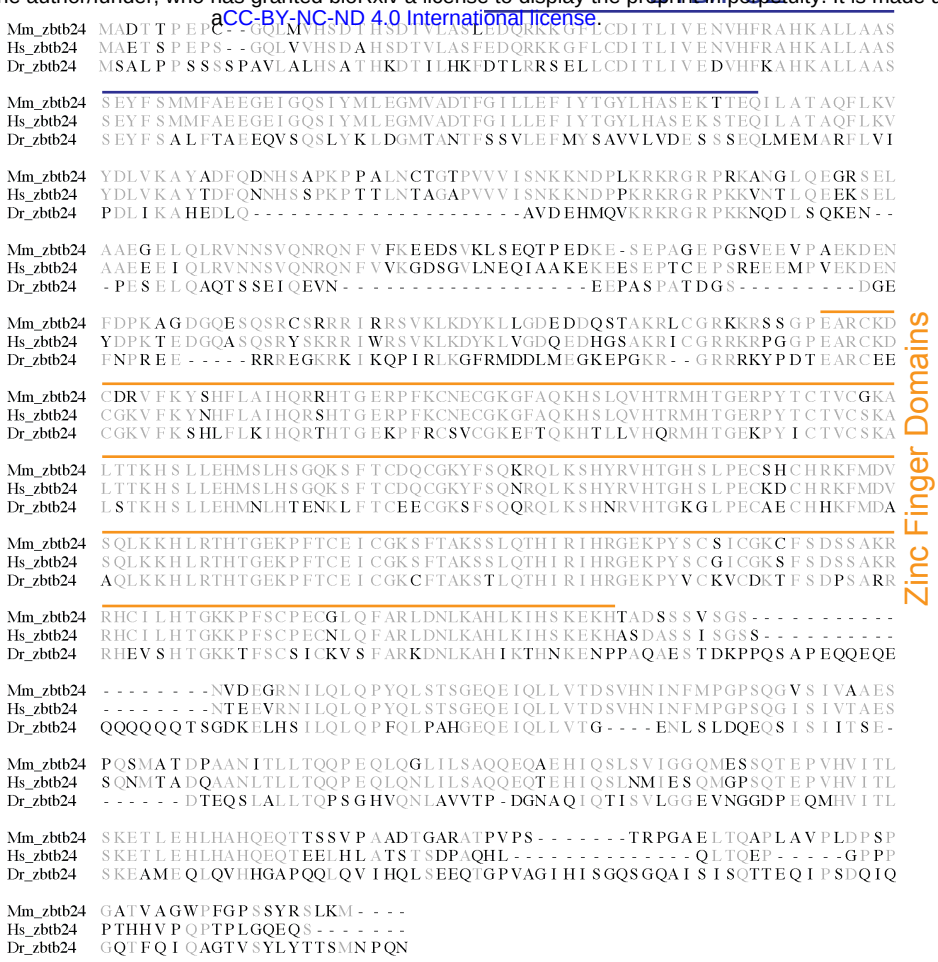


Figure 6. Mutation of cytosolic dsRNA receptor Mda5 mitigates the interferon response in *zbtb24*$^{\Delta/\Delta}$ zebrafish.
(**A**) Expression of interferon signaling genes *stat1b* and *irf7* in indicated genotypes at 3 wpf. n ≥ 7 biological replicates. (**B**) RNA-seq heatmap of interferon stimulated genes upregulated in *zbtb24*$^{\Delta/\Delta}$ zebrafish and rescued in *mda5*$^{mk29/mk29}$; *zbtb24*$^{\Delta/\Delta}$ zebrafish at 3 wpf. Shown are Z-score normalized gene expression values. (**C**) Expression of interferon signaling genes *stat1b* and *irf7* in indicated genotypes at 6 wpf. n=6 biological replicates. (**D**) Model for the activation of interferon response in *zbtb24* mutants. Loss of Zbtb24 function causes hypomethylation of pericentromeric Sat1 repeats, which leads to depression of associated Sat1 transcripts. These pericentromeric transcripts are recognized by the dsRNA helicase Mda5 which signals through Mavs and Irf7 to upregulate ISGs. Auto-regulatory feedback implicates irf7 as both an ISG and a key downstream effector of Mda5/Mavs signaling.

A



B

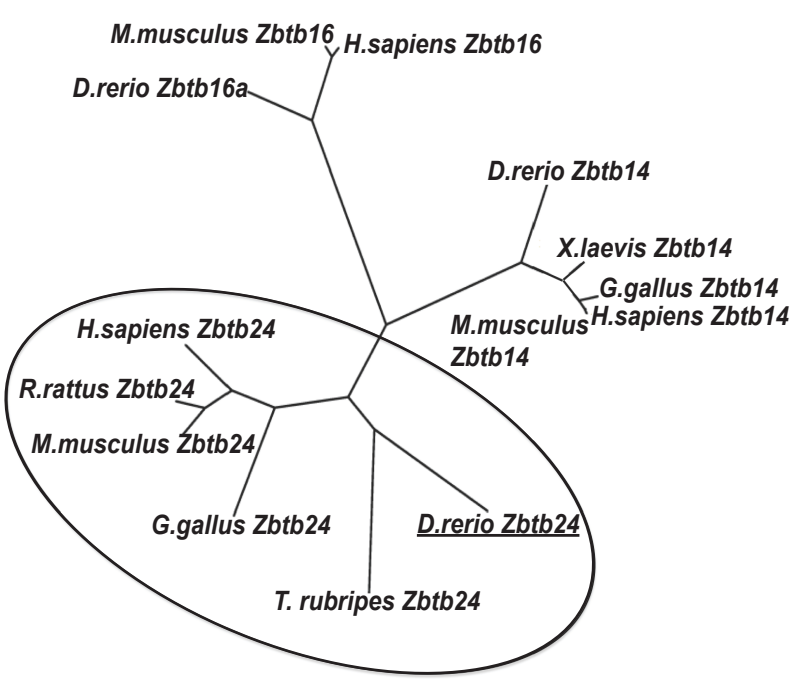


Figure 1-figure supplement 1. Zbtb24 conservation in vertebrate species.
(**A**) Sequence alignment of mouse (Mm), human (Hs) and zebrafish (Dr) Zbtb24 showing conservation of BTB (blue) and Zinc finger (orange) domains. (**B**) Phylogenetic tree of Zbtb24 with closest members of Zbtb family of proteins, Zbtb14 and Zbtb16 mouse, rat, human, chick, and Tetradon. Alignments were performed on full-length amino acid sequences via ClustalW and DRAWTREE (http://mobyle.pasteur.fr/cgi-bin/portal.py).

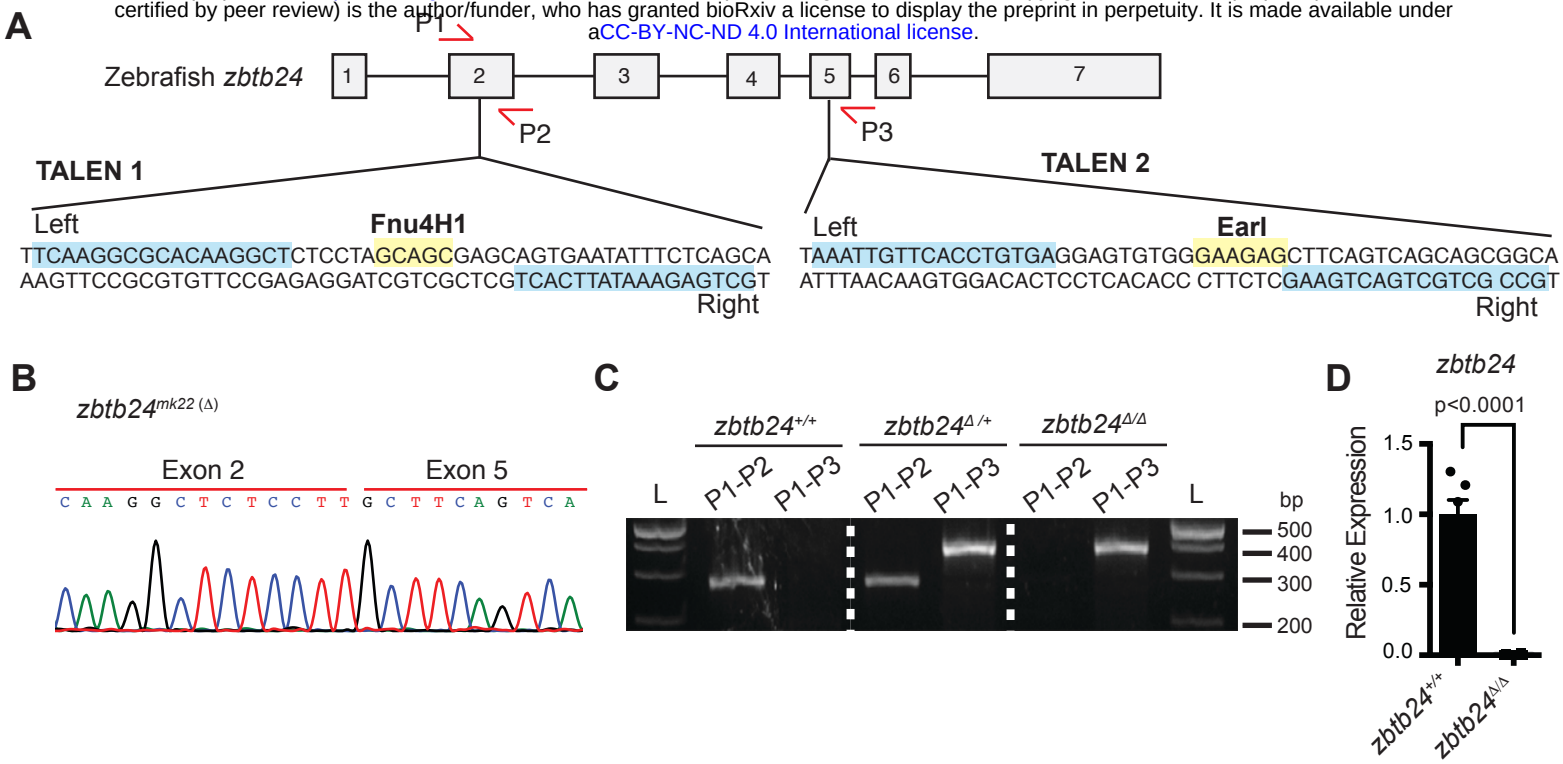


Figure 1-figure supplement 2. TALEN design for introducting mutations at the endogenous *zbtb24* zebrafish gene. (**A**) Schematic of TALEN sequences showing target sites for introducing mutations in *zbtb24*. P1, P2 and P3 indicate locations for genotyping primers. Sequence in blue indicate target site. Sequence in yellow indicates site of restriction enzyme digestion. (**B**) Sequence trace confirming generation of large deletion, *zbtb24*$^{mk22\ (\Delta)}$. (**C**) Representative genotyping of *zbtb24*$^{mk22\ (\Delta)}$ allele. L: Ladder. P1, P2, P3 represent primers from panel **A** used for amplifying product in specified lane. (**D**) qRT-PCR analysis of *zbtb24* mRNA in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 2 wpf (n=6 for each group).

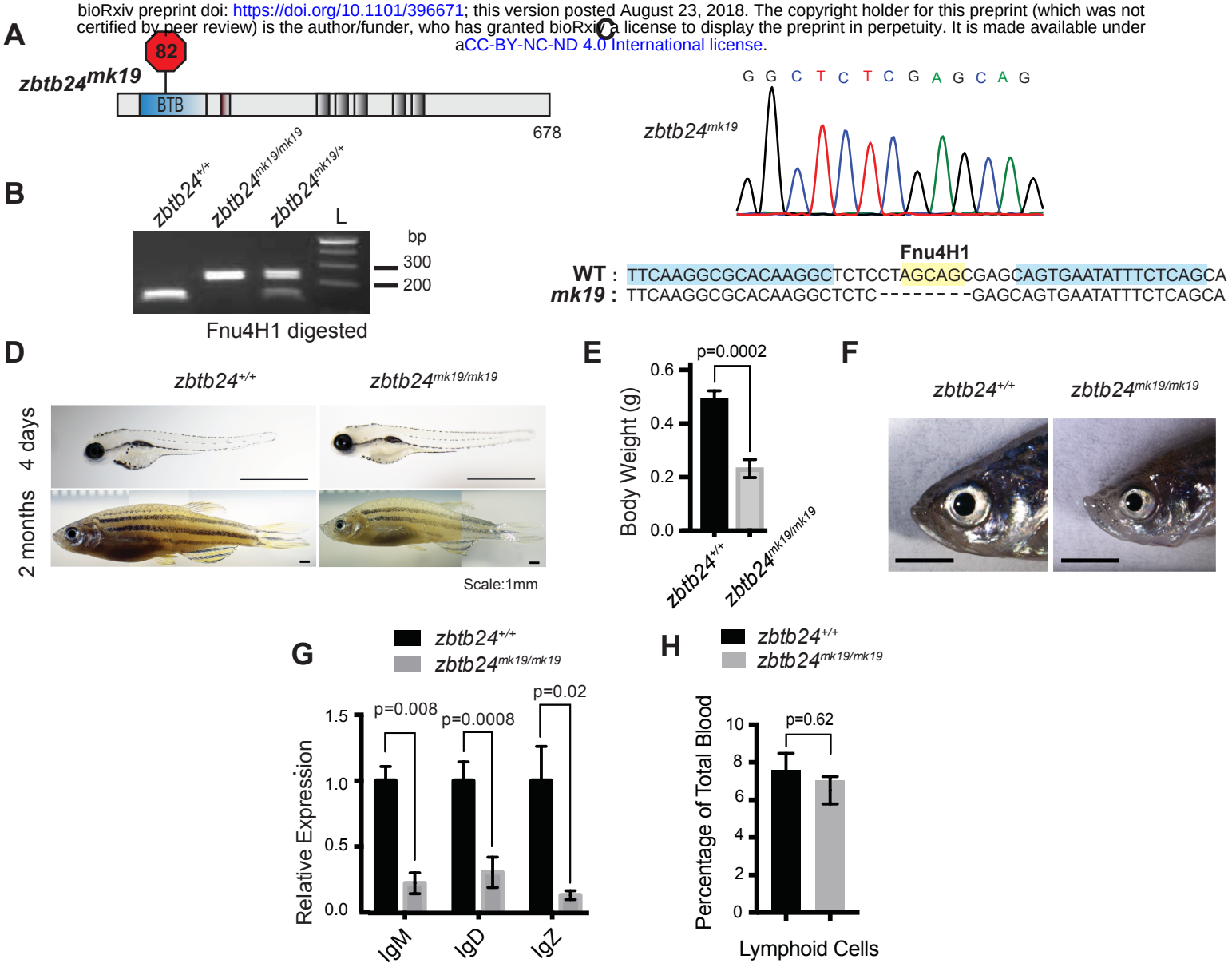


Figure 1-figure supplement 3. A second mutant allele of *zbtb24* recapitulates key features of ICF Syndrome.
(**A**) Schematic of *zbtb24*^{mk19} with premature stop codon at aa 82. (**B**) Representative genotyping of *zbtb24*^{mk19} allele. (**C**) Sequence trace confirming generation of small deletion allele, *zbtb24*^{mk19}. (**D**) Representative bright field image of *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish at 4 days and 2 months after fertilization. Scale bar: 1mm. (**E**) Quantification of body weight of *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish at 5 months (n ≥ 6 biological replicates). (**F**) Facial abnormalities in *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish. Scale bar: 3 mm. (**G**) qRT-PCR analysis of IgM, IgD and IgZ zebrafish immunoglobulins in *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish measured at 6 weeks post fertilization (n=5 biological replicates). (**H**) Quantification of lymphoid cell populations in total blood isolated from *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ adult kidney marrow, measured by Forward/Side scatter flow cytometry (n=16 biological replicates). All error bars indicate standard error of the mean (SEM).

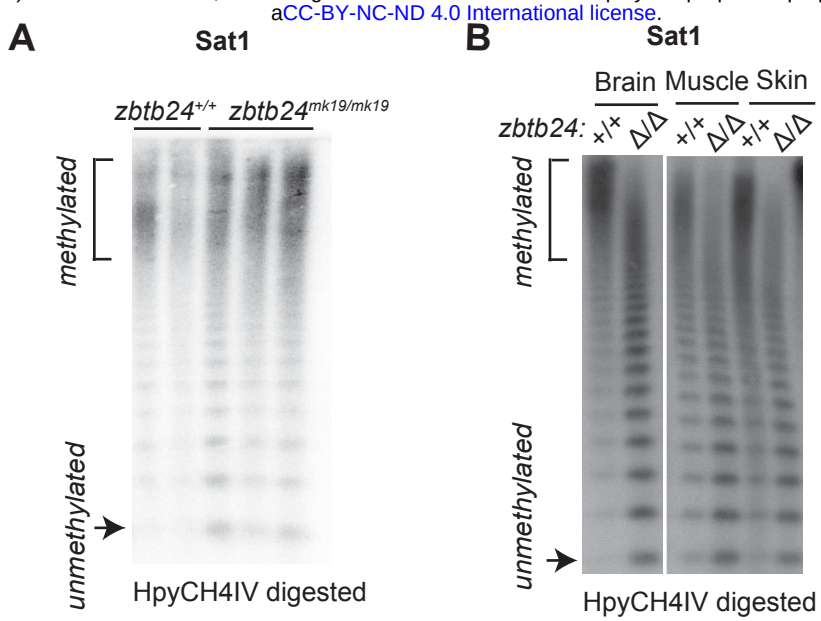
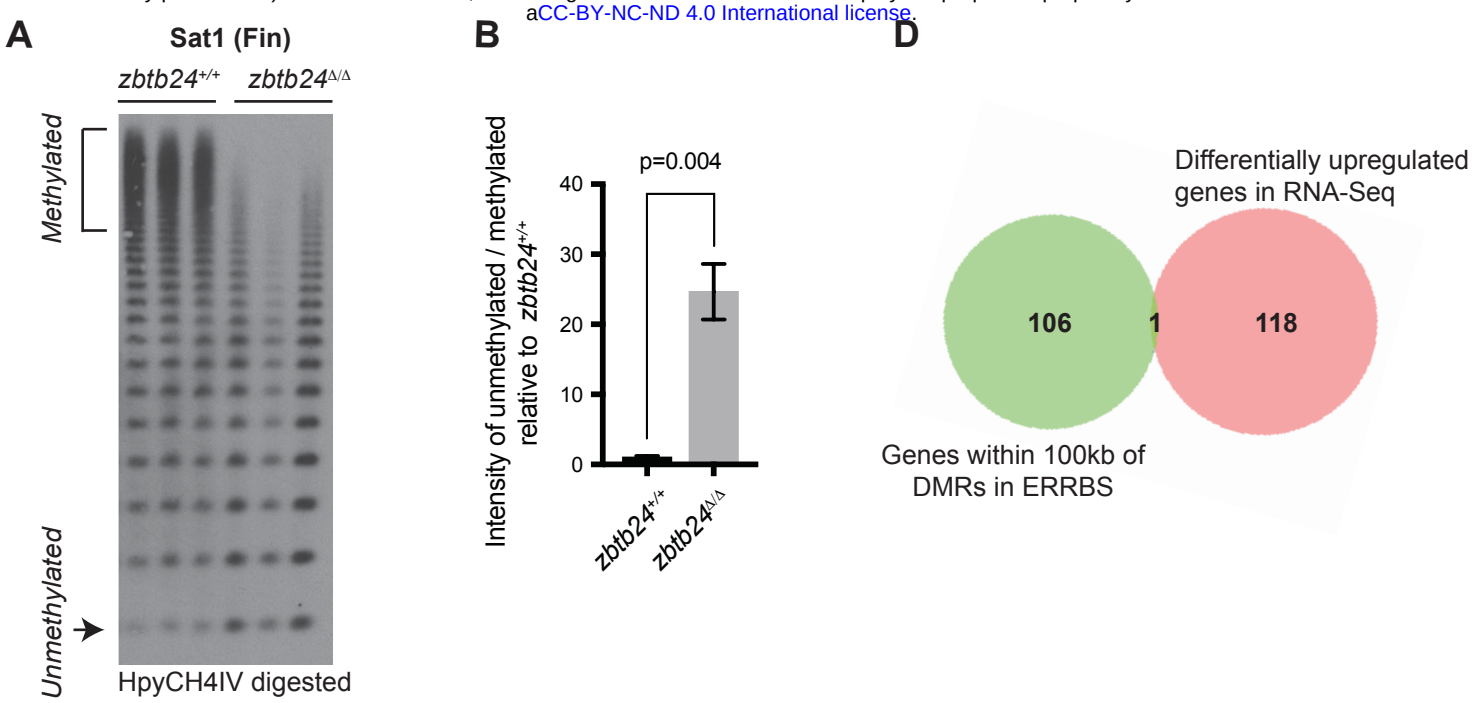


Figure 2-figure supplement 1. *Zbtb24* mutation causes methylation loss at pericentromeric repeats.
(**A**) Southern blot of genomic DNA digested with 5mC-sensitive restriction enzyme HpyCH4IV and probed with zebrafish Sat1 sequence. Genomic DNA was isolated at 1 month from *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ animals. Each lane represents a biological replicate for the indicated genotype. (**B**) Southern blot of genomic DNA from different tissues in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ digested with 5mC-sensitive restriction enzyme HpyCH4IV and probed with zebrafish Sat1 sequence. Each lane represents pooled DNA samples of the indicated tissue from 3 zebrafish adults.



C

	WT_1	WT_2	WT_3	Δ/Δ_1	Δ/Δ_2	Δ/Δ_3
Total Reads	91,715,863	82,987,666	87,720,686	78,979,041	86,934,288	78,008,892
Aligned Reads	20,187,443	19,433,013	20,277,484	18,296,141	16,303,335	23,170,337
Unique Reads	20,072,743	19,433,013	20,277,484	18,296,141	16,303,335	23,170,337
Non-Unigue Reads	16,471,055	16,176,498	17,096,320	15,779,058	13,085,223	18,297,942
CpGs	1,755,842	1,750,384	1,749,329	1,728,118	1,728,061	1,703,669
CpGs (cov>=5)	1,348,631	1,329,238	1,324,407	1,270,289	1,294,706	1,312,726
CpGs (cov>=10)	1,092,951	1,088,543	1,079,920	1,035,382	1,025,718	1,100,081
Average Sample Methylation	0.7627	0.7638	0.7625	0.6826	0.6746	0.6843
Average Coverage	29.20	28.28	29.81	27.18	23.75	34.49
Conversion Rates (non-CG)	0.997	0.996	0.997	0.997	0.997	0.997

Figure 2-figure supplement 2. *Zbtb24* mutants exhibit modest reductions in 5mC at non-pericentromeric sequences
(**A-B**) Southern Blot and quantification of HpyCH4IV digestion at Sat1 sequences in fin tissues used for ERRBS. (**C**) Basic statistics of ERRBS analysis in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish. (**D**) Venn diagram summarizing the overlap between differentially upregulated genes from RNA-Seq (right circle) and genes within 100kb of all DMRs in ERRBS tested for differential expression (left circle).

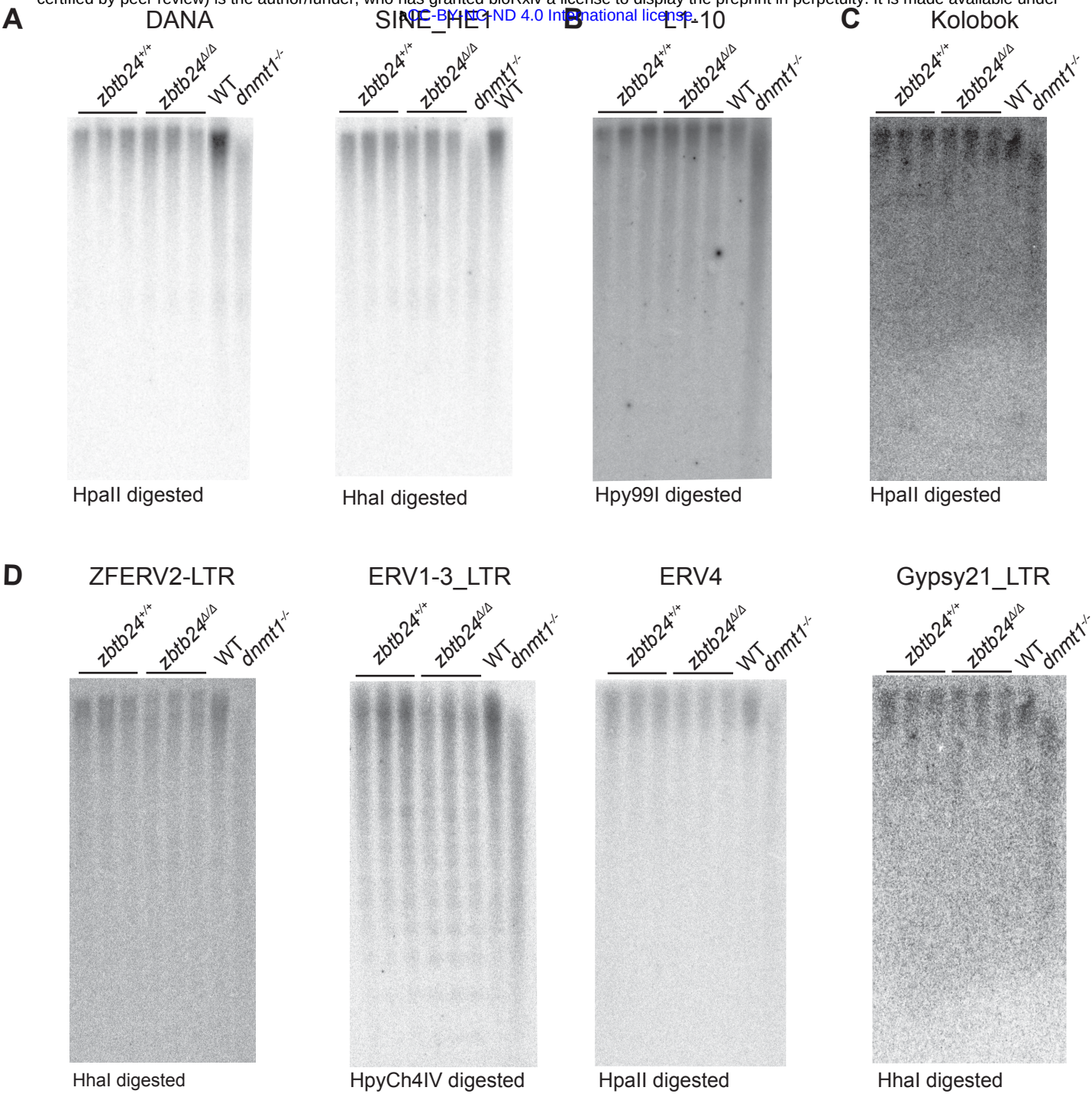


Figure 2-figure supplement 3. Methylation at interspersed repeats is unaffected in *zbtb24* mutants.
Southern blot analysis of DNA methylation at interspersed repeats (**A**) Short Interspersed Nuclear Element (SINE), DANA and SINE_HE1 (**B**) Long Interspersed Nuclear Element (LINE), L1-10 (**C**) DNA Transposon, Kolobok (**D**) Endogenous Retroviral (ERV) elements, ZFERV2, ERV1-3, ERV4 and Gypsy21 in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ 1-month old zebrafish. Each lane represents a biological replicate of the indicated genotype. Genomic DNA from *dnmt1*$^{-/-}$ embryos is used as a positive control. The methylation-sensitive enzyme used to digest genomic DNA is indicated below the respective blot.

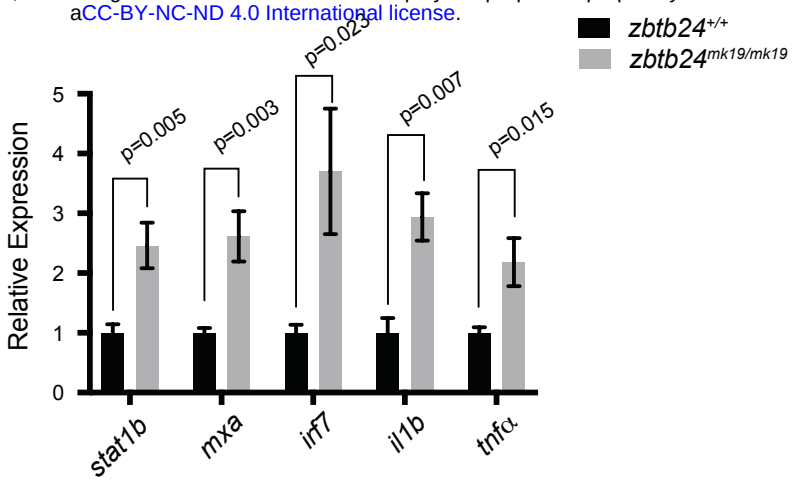


Figure 3-figure supplement 1. Mutation in *zbtb24* leads to activation of innate immune response pathways.
qRT-PCR validation of genes that are part of the innate immune response pathway that are also upregulated in zbtb24$^{mk19/mk19}$ zebrafish at 3 wpf. Error bars indicate SEM from 5 biological replicates.

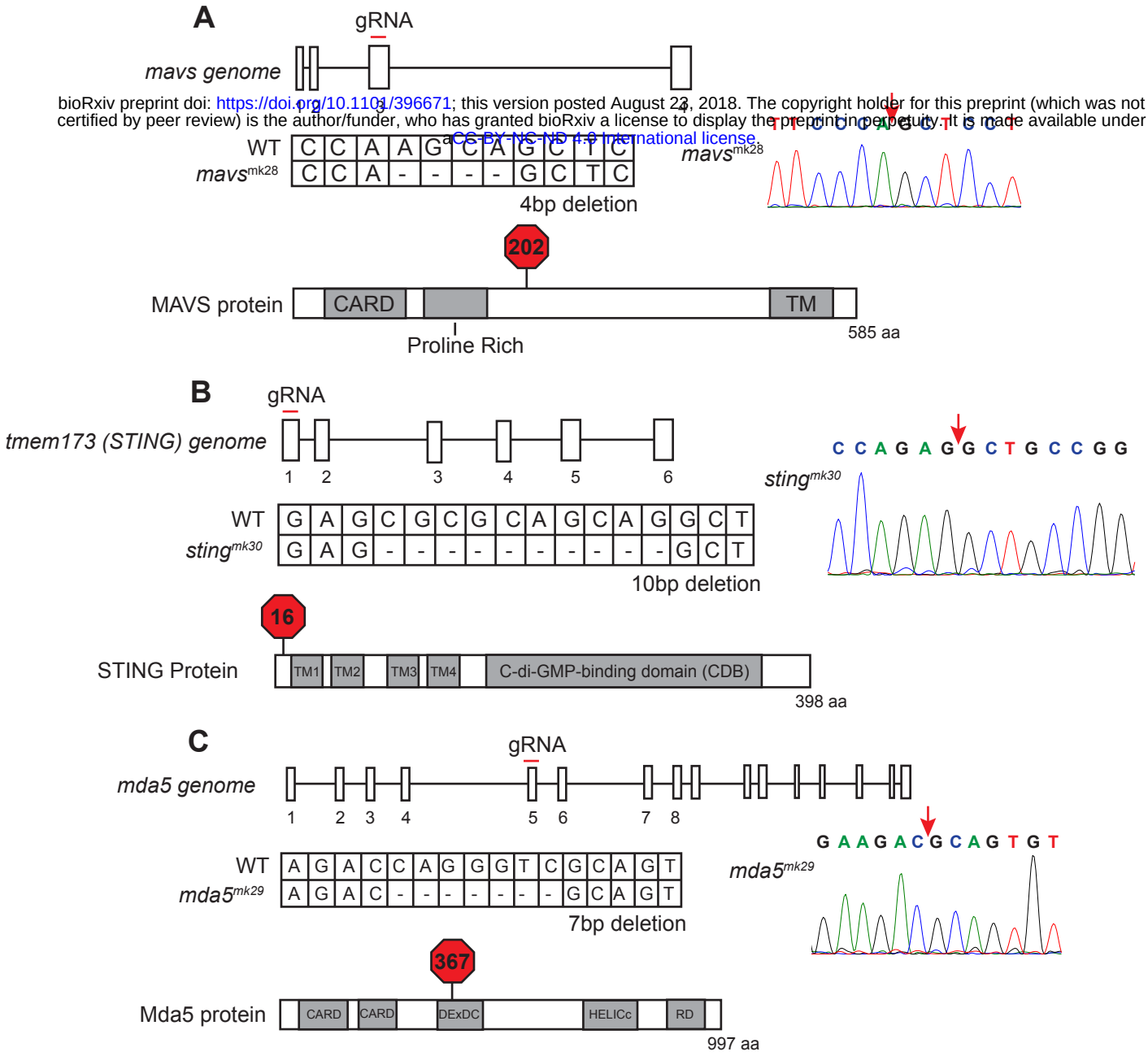


Figure 4-figure supplement 1. Mutation of zebrafish orthologs of *mavs*, *sting*, and *mda5*.
(**A**) Schematic of mutation in *mavs* including position of targeted gRNA, sequence information of the mutation, and location of predicted STOP codon. (**B**) Schematic of mutation in *sting* including position of targeted gRNA, sequence information of the mutation, and location of predicted STOP codon. (**C**) Schematic of mutation in *mda5* including position of targeted gRNA, sequence information of the mutation, and location of predicted STOP codon. The site of deletion is indicated with a red arrow on the sequence trace.

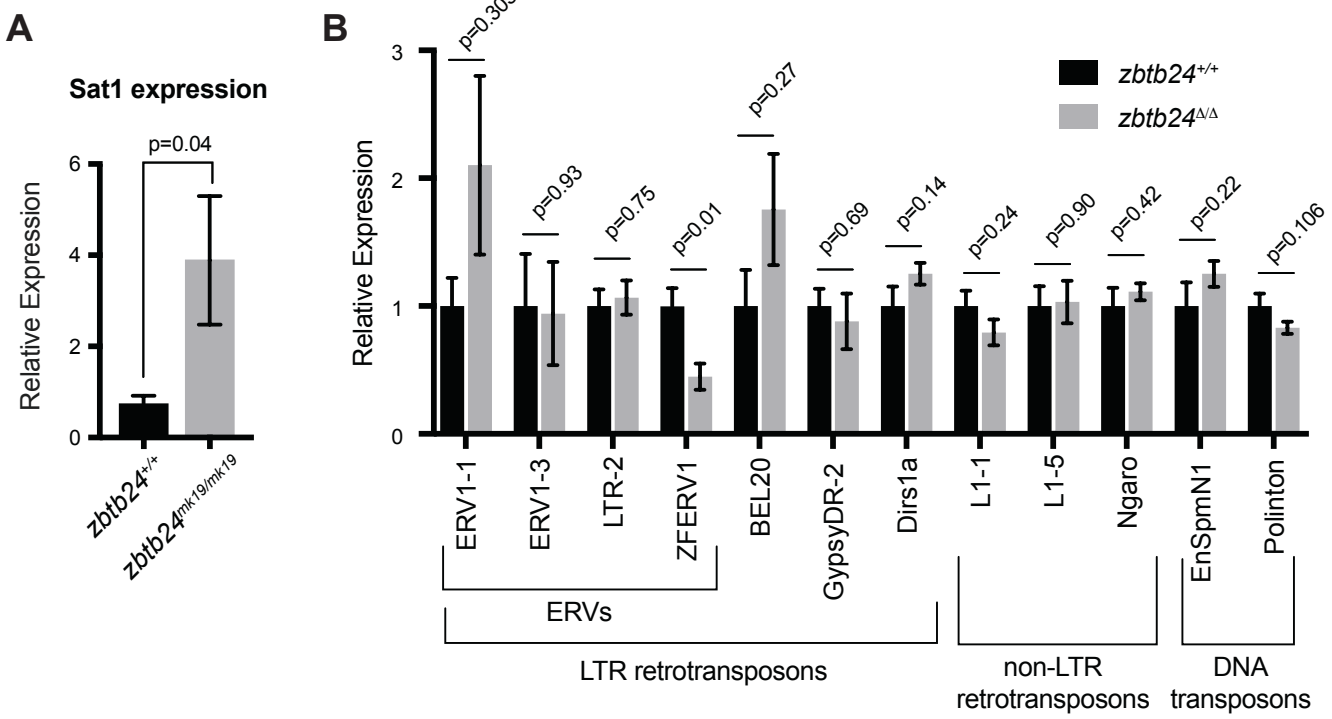


Figure 5-figure supplement 1. Mutation in *zbtb24* upregulates Sat1 transcripts but not transposons.
(**A**) qRT-PCR of Sat1 RNA in *zbtb24*$^{+/+}$ and *zbtb24*$^{mk19/mk19}$ zebrafish at 6 wpf. Error bars indicate SEM from at least 3 biological replicates. (**B**) qRT-PCR reveals similar levels of expression from transposable elements in *zbtb24*$^{+/+}$ and *zbtb24*$^{\Delta/\Delta}$ zebrafish at 6 wpf. Error bars indicate SEM from 4-8 biological replicates.