DEFINGING THE RELATIONSHIP BETWEEN PERICENTROMERIC HYPOMETHYLATION AND DISEASE PHENOTYPES: INSIGHTS FROM ZEBRAFISH MODELS OF IMMUNODEFICIENCY, CENTROMERE INSTABILITY, AND FACIAL ANOMALIES (ICF) SYNDROME

by

ALYSHA N. HIGGS

(Under the Direction of Mary Goll)

ABSTRACT

5-methylcytosine (5mC) is an epigenetic modification associated with repression of transcription. 5mC (DNA methylation) is highly enriched at repetitive sequences including pericentromeric satellite repeats that surround chromosome centromeres. Loss of DNA methylation at these repeats is seen in aging, cancer, and in the rare autosomal recessive disease Immunodeficiency, Centromere Instability, and Facial Anomalies (ICF) syndrome. To date the mechanisms that regulate DNA methylation at pericentromeric repeats remain unclear and we have a limited knowledge of how a loss of DNA methylation at pericentromeric satellite repeats contributes to disease pathology. Homozygous loss-of-function mutations in cell division cycle associated 7 (CDCA7) and zinc finger and BTB domain containing 24 (ZBTB24) have been identified in patients with ICF syndrome. However, the functions of these genes in the regulation of DNA methylation remain unclear. To address this deficiency, our laboratory has developed animal models of ICF syndrome by deleting the zebrafish orthologs of each gene. Homozygous deletion of either *zbtb24* or *cdca7a* results in a specific loss of DNA methylation at pericentromeric repeats similar to that observed in patients with ICF syndrome. However, while *zbtb24*$^{\Delta/\Delta}$ animals recapitulate most phenotypic hallmarks of ICF syndrome, *cdca7a*$^{\Delta/\Delta}$ mutants

only exhibit a subset of these phenotypes. This finding suggests not all ICF phenotypes are linked to pericentromeric hypomethylation. Mechanistically, we find that *zbtb24*$^{\Delta/\Delta}$ and *cdca7a*$^{\Delta/\Delta}$ mutants both show increasingly severe DNA methylation loss at pericentromeres in progressively older animals, suggesting these genes are involved in the long-term maintenance of DNA methylation at pericentromeric satellite repeats.

INDEX WORDS: DNA methylation, epigenetics, ICF syndrome, pericentromere, chromatin, zebrafish, repeats

DEFINING THE RELATIONSHIP BETWEEN PERICENTROMERIC HYPOMETHYLATION AND DISEASE PHENOTYPES: INSIGHTS FROM ZEBRAFISH MODELS OF IMMUNODEFICIENCY, CENTROMERE INSTABILITY, AND FACIAL ANOMALIES (ICF) SYNDROME

by

ALYSHA N. HIGGS

B.S., High Point University, 2017

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

Athens, Georgia

2022

DEFINING THE RELATIONSHIP BETWEEN PERICENTROMERIC HYPOMETHYLATION AND DISEASE PHENOTYPES: INSIGHTS FROM ZEBRAFISH MODELS OF IMMUNODEFICIENCY, CENTROMERE INSTABILITY, AND FACIAL ANOMALIES (ICF) SYNDROME

by

ALYSHA N. HIGGS

Major Professor: Mary Goll

Committee: Robert Schmitz
Zachary Lewis
Shelley Hooks
Michael White

Electronic Version Approved:

Ron Walcott
Vice President for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2022

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my sincere gratitude to my advisor, Dr. Mary Goll. I am so appreciative to her for giving me the opportunity to work on an exciting thesis project. During the duration of my time in her lab she was committed to training me to be a competent and confident scientist. She helped in any way she could to ensure that I had a safe and understanding environment in the midst of COVID which I will be forever thankful. I have received invaluable feedback that I will use throughout my scientific career.

Next, I would like to thank my friends and colleagues at the Goll lab for their feedback and support which have been helpful in my thesis research. I would like to acknowledge our previous postdoc Bagde Akdogan-Ozedilek and current graduate students Audrey Calvird and Katie Duval for endless protocol guidance. I would also like to thank Danielle Obiri and Anvith Reddy for being great mentees and bring fresh enthusiasm to the research. Lastly, I would like to thank past lab technicians Alexandria Scozzari and Chauncey Ragin for taking care of the lab and fish.

A special thank you to Fuel Hot Yoga in Athens, Georgia for contributing to my wellbeing both physical and mental.

Thank you to my parents and family for the support and encouragement to continue on my path to becoming a scientist.

Thank you, David, my best friend and confidante. Thank you for always being here on this journey to complete my education.

Thank you to Rachel for always making me laugh and giving me motivation to continue on this journey.

LIST OF FIGURES

Table of Contents

Chapter 1: INTRODUCTION

DNA METHYLATION

Methylated bases in eukaryotic genomes

Two types of DNA methylation have been detected in eukaryotic genomes, N6-methyladenine (6mA) and 5-methylcytosine (5mC). 5-methylcytosine (5mC) is a reversible epigenetic mark which consists of a methyl group added to the fifth position of the cytosine ring of DNA (Goll and Bestor 2005). 5mC is found across vertebrates, flowering plants, some fungal species, invertebrates, protist taxa and some bacterial species. However, there is little to no 5mC present in *C. elegans* and *D. melanogaster* (Goll and Bestor 2005). In vertebrates, 5mC is located across the genome and is found predominately in the symmetrical cytosine-guanine (CpG) dinucleotide context (Suzuki and Bird 2008).

While historically 5mC was thought to be the only modified base present in eukaryotic genomes, more recent research has uncovered the presence of N6-methyladenine (6mA) in low abundance in eukaryotes (O'Brown and Greer 2016; Luo et al. 2015; Luo and He 2017). It is found in unicellular eukaryotes, Drosophila, *C. elegans*, and lower fungi (Zhang et al. 2015; Greer et al. 2015; Mondo et al. 2017). In Drosophila, 6mA is mainly located in gene bodies of transposons and is shown to be correlated with the activation of those transposons (Zhang et al. 2015). In *C. elegans,* 6mA has been linked to transgenerational inheritance and has been shown to increase transgenerationally upon the decrease of the histone 3 lysine 4 dimethylation (H3K4me2) epigenetic mark (Greer et al. 2015). 6mA has recently been discovered to be detectable in vertebrates such has mice, zebrafish, and humans (Koziol et al. 2016; Liu et al.

2016). In zebrafish, 6mA is present at up to 0.2% of total adenines during embryogenesis. During this process, the 6mA epigenetic mark is located in repetitive regions of the genome. This decreases with the progression of development (Liu et al. 2016). In humans it has been found that 6mA is located at approximately 0.051% of all adenines within the genome. 6mA has been found in human cells to be located at the coding region of genes and to mark actively transcribed genes (Xiao et al. 2018). However, 6mA is an epigenetic mark that remains to be well characterized and is highly debated to whether 6mA seen in eukaryotes is just a form of contamination.

This thesis focuses on the more abundant and well characterized 5mC mark. For simplicity, throughout this thesis, the term DNA methylation will refer exclusively to 5mC.

Enzymes involved in DNA methylation

DNA methylation is established and maintained by de novo and maintenance DNA methyltransferases. DNA methyltransferases (DNMTs) transfer a methyl group from S-adenosyl-L-methionine to the fifth position of the cytosine ring (Goll and Bestor 2005). DNMT1 performs maintenance DNA methylation which adds a methyl group to hemimethylated strands of DNA. DNMT1 is able to copy methylation states by localizing to replication forks during mitosis and adding methyl groups to the cytosine ring onto the newly formed DNA strand (Li, Bestor, and Jaenisch 1992). For DNMT1 to localize to the replication forks, it needs the cofactor ubiquitin-like with PHD and ring finger domains 1 (UHRF1) (Bostick et al. 2007). DNMT1 localizes to the replication forks during S phase of the cell cycle. DNMT1 then copies the methylation state to new synthesized strands of DNA (Leonhardt et al. 1992). *Dnmt1* mutant mice are embryonic

lethal and exhibit a global loss of DNA methylation at approximately 90% reduction (Li, Bestor, and Jaenisch 1992).

DNMT3 family members perform de novo DNA methylation which adds a methyl group to unmethylated strands of DNA (Okano et al. 1999). In mammals, there is a global loss of DNA methylation during embryogenesis and DNA methylation is then reestablished by the de novo DNA methyltransferases DNMT3A and DNMT3B. *Dnmt3a* and *Dnmt3b* are partially redundant, and combined loss leads to hypomethylation of major satellite repeats and imprinted genes. Individually, a loss of *Dnmt3a* does not alter global DNA methylation. On the other hand, mice with a loss of *Dnmt3b* exhibit hypomethylation at minor satellite repeats (Okano et al. 1999). DNMT3A and DNMT3B have also been implicated to help DNMT1 perform maintenance DNA methylation at germline genes and some repetitive elements (Chen et al. 2003; Liang et al. 2002). Rodent genomes encode for a third DNMT3 ortholog, DNMT3C, which is needed in order to establish DNA methylation at the promoters of young transposable elements in the male germ line (Barau et al. 2016).

DNMT3L is another homolog of DNA methyltransferases that is specific to mammals. Intriguingly, DNMT3L lacks methyltransferase activity and instead works with DNMT3A to help establish genomic imprinting in the germline (Bourc'his et al. 2001; Chedin, Lieber, and Hsieh 2002).

Methods of detecting DNA methylation

Several methods have been developed for assessing DNA methylation. Bisulfite sequencing is able to determine differentially methylated regions at specific genes. Bisulfite sequencing works by using sodium bisulfite to turn cytosine into uracil through deamination. The

converted residues are then read as thymine after PCR. However, DNA methylation is resistant to this bisulfite conversion thus, methylated cytosines will be read as cytosine. After sequencing DNA, methylation status can be determined through comparison to an untreated sample (Kurdyukov and Bullock 2016).

Another way to detect DNA methylation changes is via restriction enzymes. Methylation specific restriction endonucleases are a fast and easy way to determine DNA methylation status. This analysis is done by selective DNA cleavage by restriction enzymes, which are unable to cut at a restriction site where there is DNA methylation present. The resulting fragments from the digestions are then analyzed on a gel, Southern Blot, or by quantitative PCR. The downside to this mechanism is that you can only use a restriction enzyme for specific restriction sites, and you can only assess one or a few sites at a time (Sestakova, Salek, and Remesova 2019).

DNA methylation across the genome

Animals in which DNA methylation is present exhibit an underrepresented frequency of CpG dinucleotides. For example, in mammals there is only 20-25% of the expected CpGs. In the mammalian genome 70-80% of these CpGs are methylated (Ehrlich et al. 1982). Most of the CpGs in the mammalian genome are found in regions like the intergenic and intronic regions of DNA, mainly within repeat sequences and transposable elements.

Transposable elements and pericentromeric satellite repeats are repressed by DNA methylation and this is conserved across many species. Repetitive DNA elements make up 50% of the human genome and contain approximately 52% of all CpGs in the human genome (Lander et al. 2001). Approximately 90% of DNA methylation is found at repetitive elements (Beisel and Paro 2011). In humans, repetitive elements are arranged as interspersed repeats or tandem

repeats. Interspersed repeats include transposable elements such as DNA transposons, LINE (Long Interspersed Elements), LTR (Long Terminal Repeats), and SINE (Short Interspersed Elements). On the other hand, tandem repeats include satellite and simple repeats that are found near the centromeres and telomeres (Lopez-Flores and Garrido-Ramos 2012).

Transposable elements are heavily methylated to prevent the invasion of these elements into new regions of the genome. Improper silencing of transposable elements leads to an increase in genomic instability because unmethylated transposable elements can cause chromosomal rearrangements and gene disruption. While the known function of DNA methylation at transposable elements is to prevent the selfish elements from transposing, the significance of DNA methylation at pericentromeric repeats is unclear.

Gene bodies are heavily methylated even though they are often CpG poor. Different to the role of DNA methylation in repression, gene body methylation is linked with transcribed genes (Jones 2012). Studies have implicated that gene body DNA methylation may have a role in regulating splicing (Shukla et al. 2011).

CpG islands are small CG-dense regions of mammalian genomes of about one kilobase and less than 10% of CpGs occur in these regions. CpG islands contain over 50% of the unmethylated CpGs. However, CpG islands only amount for approximately 1% of the genome and less than 10% of the total genome wide CpGs. CpG islands are found in the promoters and first exons of genes and are unmethylated (Deaton and Bird 2011). CpG islands are located in open chromatin and are enriched in acetylated histones H3 and H4. These epigenetic marks further mark CpG islands for activation of transcription (Tazi and Bird 1990). However, in unhealthy cells CpG islands are hypermethylated resulting in transcriptional silencing (Esteller 2002).

DNA methylation in vertebrate development

DNA methylation is necessary for proper development in vertebrates. It has been shown that reduced levels of DNMT1 in mouse embryos is recessive lethal and that homozygous embryos fail to develop beyond day 9 and die before day 11 (Li, Bestor, and Jaenisch 1992). Along with this, it has been shown that *Dnmt3a* mutant mice develop to term, but most become runted and die at about four-weeks old. On the other hand, *Dnmt3b* mutant mice are not viable. *Dnmt3b* mutant mice develop normally before E9.5 but then develop growth impairments and rostral neural tube defects (Okano et al. 1999).

DNA methylation is involved in regulating gene expression, X-chromosome inactivation and genomic imprinting. A global loss of DNA methylation has been linked to deregulation of gene expression and an increase in DNA damage and genome instability (Smith and Meissner 2013). X-chromosome inactivation involves genes that are monoallelically expressed. Thus, one gene needs to be silenced. Genes on the inactive X-chromosome are silenced due to DNA methylation located on CpG-rich promoters (da Rocha and Gendrel 2019).

Genomic imprinting is caused by epigenetic marks that result in monoallelic expression that is tied to a parent of origin. DNA methylation plays a big part in silencing imprinted genes. For example, *Dnmt1* mutant embryos are unable to sustain imprinted gene expression (Li, Beard, and Jaenisch 1993). In addition, DNMT3A and DNMT3L are needed for the establishment of DNA methylation at imprinting control regions of the mouse germline. *Dnmt3a* and *Dnmt3l* mutants have a similar phenotype in which the differentially methylated regions normally methylated on the maternal allele were unmethylated (Kaneda et al. 2004; Bourc'his et al. 2001). Without proper silencing, disorders such as Beckwith-Wiedemann Syndrome and Angelman syndrome can occur (Jin and Liu 2018). Beckwith-Wiedemann syndrome is a human genomic

imprinting disorder caused by epigenetic changes on a cluster of imprinted genes on chromosome 11p15.5-11p15.4 region. This region contains two independent domains with their own imprinting control region. Alterations in these regions leads to this overgrowth syndrome. Beckwith-Wiedemann is typically diagnosed neonatally or in childhood. This syndrome presents with macroglossia, abdominal wall defects, lateralized overgrowth, enlarged abdominal organs, and an increase risk for developing embryonal tumors (Wang et al. 2019). Another imprinting disorder, Angelman syndrome, is caused by lack of maternal contribution on chromosome region 15q11.2-q13. Patients with Angelman syndrome present with delayed growth, speech impairment, movement or balance disorder, and behavioral uniqueness (Williams, Driscoll, and Dagli 2010).

Changes in DNA methylation has also been seen in many cancers including breast, colon, liver, and bone. The onset of malignancies has been associated with genome-wide changes in DNA methylation. Some DNA methylation changes associated with cancer include tumor suppressor genes becoming hypermethylated and oncogenes and repetitive elements becoming hypomethylated (Jones and Laird 1999). Hypermethylation can prevent transcription of tumor suppressor genes and this, in combination with genetic mutations, can promote oncogenesis (Jones and Laird 1999). Hypomethylation is normally seen in cancer at repeat elements such as satellites and retrotransposons. Hypomethylation at repeat elements is seen in a variety of cancers and can lead to increased genomic instability and promote chromosomal rearrangements (Jin, Li, and Robertson 2011).

PERICENTROMERES

Pericentromeric repeats

Pericentromeric satellite repeats are repeats that flank centromeres. They are typically AT-rich, non-coding repeat sequences present in thousands of copies. Satellite repeats are not conserved at a sequence level. However, the organization of repeats at the centromere and pericentromere are conserved amongst eukaryotic cells (Garrido-Ramos 2017). Satellite repeats are important for maintaining heterochromatin, chromosome segregation during cell division and genome structure within the nucleus (Allshire and Madhani 2018; Plohl, Mestrovic, and Mravinac 2014; Jagannathan, Cummings, and Yamashita 2018).

The centromeric and pericentromeric regions consist of several thousand copies of alpha satellite DNA in humans. Alpha satellites have a monomer length of approximately 170bp. This monomer is repeated to form a higher repeat which is then further repeated in the centromeric and pericentromeric regions. The alpha-satellite repeats are found on all chromosomes. The alpha-satellites are important for kinetochore assembly, which aids in proper cell division (Tyler-Smith and Brown 1987). In humans, there are three types of classical satellite repeats found in the pericentromeric region. These classical satellites account for approximately 4-5% of the human genome. Classical satellite type I are short AT-rich sequences found at the pericentromere of most chromosomes. Classical satellites type II and type III are made up of a 5bp GGAAT repeat unit and are located in the pericentromere of chromosomes 1, 9, and 16 (Vourc'h and Biamonti 2011).

Mice have two types of repetitive DNA sequences at their centromeres. Major satellite repeats are made up of a 234bp monomer located at the pericentromere. The minor satellites are made up of a 120bp monomer (Joseph, Mitchell, and Miller 1989). In zebrafish, type I satellite

repeats (sat1) are made up of a 186bp monomer that is AT-rich and makes up 8% of the zebrafish genome (Ekker, Fritz, and Westerfield 1992). In zebrafish sat1 is located at the pericentromeric and centromeric regions of all chromosomes (Phillips and Reed 2000). Based upon location and sequence organization sat1 repeats in zebrafish most closely relate to alpha satellites in humans.

Pericentromeric chromatin

The repeats at the pericentromeres are enriched in canonical markers of heterochromatin. This includes the repressive marks histone H3 lysine 9 trimethyl (H3K9me3) and DNA methylation (Dejardin 2015). The pericentromeres in vertebrates are heavily methylated. However, the importance of DNA methylation at the pericentromere is unknown.

Even though the function of DNA methyltransferases is clear, the mechanisms by which DNA methylation is targeted to the satellite repeats by DNA methyltransferases is unclear. DNMT3B is predicted to be involved in de novo DNA methylation at satellite repeats. Minor satellite repeats are demethylated in *Dnmt3b*$^{-/-}$ mutant mouse embryos. However, demethylation of the minor satellites does not occur in *Dnmt3a*$^{-/-}$ mouse embryos (Okano et al. 1999). Also, in mouse, embryonic stem (ES) cells that are deficient in *Dnmt3a* and *Dnmt3b* progressively lose DNA methylation at repetitive elements after prolonged cell culture (Chen et al. 2003). It has been suggested that the risk of errors in maintenance DNA methylation at highly repetitive regions allows for de novo DNA methyltransferases to ensure faithful inheritance of DNA methylation patterns. Although, the mechanism of this role is unclear (Liang et al. 2002).

Some data suggests that H3K9me3 is involved in directing DNA methylation to satellite repeats. DNA methylation at satellite repeats is impaired in mouse ES cells deleted for the H3K9 methyltransferases *Suv39h1* and *Suv39h2*, however the effect is mild (Lehnertz et al. 2003).

Pericentromeric methylation loss in cancer

In several cancers, pericentromeric satellite repeats appear to be susceptible to hypomethylation (Enukashvily et al. 2007; Fanelli et al. 2008; Nakagawa et al. 2005; Narayan et al. 1998; Qu et al. 1999; Suzuki, Fujii, and Ayusawa 2002; Tsuda et al. 2002). It is hypothesized that a loss of DNA methylation at pericentromeric repeats may predispose cells to chromosomal rearrangements. However, a loss of DNA methylation at the pericentromere alone does not appear to be enough to cause chromosomal rearrangements in tumors (Ehrlich et al. 2003; Tsien et al. 2002). Additional studies are needed to determine the link between DNA hypomethylation and cancer progression. This has been a challenging feat due to a lack of models with a specific loss of DNA methylation at the pericentromere. In addition to promoting chromosomal rearrangements, a loss of DNA methylation can lead to aberrant satellite transcription from pericentromeres which is also seen in some cancers such as pancreas, lung, and kidney (Ugarkovic et al. 2022). Overexpression of satellite transcripts can increase genome instability. For example, increased satellite transcripts can bind breast cancer gene 1 (BRCA1) and hinder the stability of the replication fork and induce DNA damage and genome instability to promote the formation of breast cancer (Zhu et al. 2018).

IMMUNODEFICIENCY, CENTROMERE INSTABILITY AND FACIAL ANAMOLIES SYNDROME

ICF syndrome phenotypes

Immunodeficiency, Centromere instability and Facial anomalies (ICF) syndrome is a rare autosomal recessive disorder characterized by loss of DNA methylation at pericentromeric sequences. At present, there are less than 100 patients diagnosed with ICF syndrome worldwide (van den Boogaard et al. 2017).

Although global methylation is relatively intact in ICF patients, patients exhibit hypomethylation of classical satellites 2 and 3 DNA on chromosome 1, 9, and 16. In addition to a loss of DNA methylation at satellites 2 and 3, a subset of ICF syndrome patients exhibit a loss of DNA methylation at alpha satellites. Loss of DNA methylation is associated with severe chromosomal abnormalities in mitogen stimulated lymphoid cells, including whole arm deletions and multiradial chromosomes involving chromosomes 1, 9, and 16 (Ehrlich, Jackson, and Weemaes 2006).

Patients with ICF syndrome exhibit facial anomalies which include a flattened nasal bridge, widely spaced eyes, enlarged tongue, and low set ears. ICF syndrome also leads to growth retardation, gastrointestinal defects, and an overall failure to thrive. Some patients have also been reported to have variable differences in neurological defects. Most patients with ICF syndrome exhibit a decrease in immunoglobulins in the presence of normal levels of lymphoid cells, resulting in immunodeficiency. Due to immunodeficiency, patients have recurrent infections that often lead to death at a young age. The cause of the immunodeficiency in ICF syndrome patients is unclear. To treat the recurrent infections, patients are given intravenous infusions of immunoglobulins.

Genetic and molecular basis of ICF syndrome

Using homozygosity mapping and whole exome sequencing, four subtypes of ICF syndrome have been identified. The four types of ICF syndrome are caused by mutations in DNA methyltransferase 3B (DNMT3B), zinc-finger and BTB domain containing 24 (ZBTB24), cell division cycle associated 7 (CDCA7), and helicase, lymphoid specific (HELLS) respectively (de Greef et al. 2011; Thijssen et al. 2015; Xu et al. 1999). Several additional patients exist that do not appear to have causative mutations in any of these genes.

ICF syndrome type 1 makes up approximately half of the reported ICF syndrome cases, and harbor predicted hypomorphic mutations in DNMT3B. Most of the described DNMT3B mutations cause an amino acid substitution within the c-terminal catalytic domain (Xu et al. 1999). Patients with ICF syndrome type I exhibit a loss of DNA methylation at sat2 and sat3, but do not exhibit DNA hypomethylation of alpha-satellites.

ICF syndrome type 2 is caused by homozygous predicted loss-of-function mutations in ZBTB24. Such mutations are found in approximately 30% of ICF syndrome patients (de Greef et al. 2011). ICF syndrome type 2 patients exhibit hypomethylation of sat2, sat3, and alpha satellites. The molecular function of ZBTB24 is unclear. However, it is thought to function as a transcription factor, and has been shown to be required for transcription of CDCA7 (Wu et al. 2016). Other studies have suggested that ZBTB24 can bind to the pericentromere in the absence of DNA methylation (Nitta et al. 2013) and that ZBTB24 can work with DNMT3B to control DNA methylation at intergenic sites (Thompson et al. 2018). In B cells, knockdown of ZBTB24 results in reduced proliferation (Liang et al. 2016; Unoki et al. 2019).

ICF syndrome type 3 is caused by homozygous predicted loss-of-function mutations in CDCA7. This form of ICF syndrome is relatively rare, making up only 4 % of total cases.

CDCA7 is a zinc finger protein of unknown function. Some data suggests it may be a transcription factor under the control of c-myc (Gill et al. 2013). Other research has suggested that CDCA7 forms a complex with KU70 and KU80 factors involved in non-homologous end joining (Unoki et al. 2019).

ICF syndrome type 4 is caused by homozygous predicted loss-of-function mutations in HELLS. HELLS is a well-known chromatin remodeler that is involved in de novo DNA methylation through its interaction with DNMT3B (Myant and Stancheva 2008). *Hells* homozygous null mutant mice are perinatal lethal and exhibit a global loss of DNA methylation (Tao et al. 2011).

It has been suggested that the four known causative genes for ICF syndrome may converge in one pathway. It is thought that ZBTB24 acts as a transcription factor for CDCA7 (Wu et al. 2016). CDCA7 and HELLS then form a complex allowing HELLS to remodel nucleosomes opening DNA to be methylated by DNMT3B (Jenness et al. 2018). However, this model does not clarify how DNA methylation is being targeted to the pericentromere. This model also fails to clarify why only ICF types 2-4 exhibit hypomethylation of alpha satellites.

Thus far, attempts to generate mouse models of ICF syndrome have been unsuccessful. Mice with ICF-like hypomorphic mutations in *Dnmt3b* exhibit small size and facial anomalies but die within 24 hours of birth. They also exhibit a T-cell defect but fail to recapitulate other ICF like phenotypes (Ueda et al. 2006). Deletion of *Zbtb24* in mice leads to embryonic lethality; however, the DNA methylation status of these mutants is not known (Wu et al. 2016). Animal models with mutations in *Cdca7* have not yet been reported. Perinatal lethality was observed following deletion of the mouse *Hells* ortholog. Along with this, *Hells* mutant mice exhibit a

global loss of DNA methylation of around 50%. DNA hypomethylation was seen at both repeat sequences and genes (Tao et al. 2011).

ZEBRAFISH AS A MODEL FOR THE STUDY OF PERICENTROMERIC METHYLATION AND ICF SYNDROME

Overview of the zebrafish model

Zebrafish are an important model organism that can recapitulate human diseases ranging from Duchenne muscular dystrophy to melanoma (Ceol et al. 2011; Berger and Currie 2012). In addition, zebrafish are low-cost, small, and have external development. With the use of zebrafish as a model organism, you have access to hundreds of synchronous embryos. Along with this comes the availability of large-scale genome mutagenesis and gene mapping, transgenics, and protein overexpression or knockdown.

DNA methylation in zebrafish

Danio rerio (Zebrafish) is a model organism in which DNA methylation is present. Zebrafish are a great model organism to study epigenetic regulation because approximately 70% of mammalian genes have a zebrafish ortholog and proteins involved in DNA methylation are highly conserved (Howe et al. 2013). The DNMT machinery is conserved in zebrafish. Zebrafish have one ortholog of DNMT1 and six orthologs of DNMT3 (Goll and Halpern 2011). Mutation of *dnmt1* in zebrafish leads to a global loss of DNA methylation. Along with this, morpholino-based knockdown of *dnmt1* results in a global loss of DNA methylation (Rai et al. 2006; Anderson et al. 2009). *uhrf1* mutant zebrafish also display a global loss of DNA methylation (Feng et al. 2010). Mutation of *dnmt3aa* and *dnmt3ab* in zebrafish results in no change in global

DNA methylation. However, these mutants do express different behavioral alterations such as predator avoidance (Lai et al. 2020). To date there have been no attempts to mutate all six zebrafish Dnmt3 orthologs.

Approximately 80% of CpG dinucleotides in the zebrafish genome are methylated. Zebrafish have a depletion of DNA methylation at non-methylated islands located at gene promoters and have heavily methylated repetitive elements and gene bodies similar to that of mammals (Feng et al. 2010). Even though zebrafish have similar functions for DNA methylation to that of mammals, zebrafish are unique and do not have imprinting or X-chromosome inactivation (Corley-Smith, Lim, and Brandhorst 1996). Zebrafish do not exhibit sex chromosomes in the lab. In zebrafish, the establishment of DNA methylation during early development is unique when compared to mammals. Both parental genomes for mammals undergo a genome-wide erasure and re-establishment of global methylation patterns (Smith and Meissner 2013). However, zebrafish do not undergo the same genome-wide erasure. The maternal methylome of zebrafish undergoes some remodeling to match that of the paternal genome that is stably inherited (Jiang et al. 2013; Potok et al. 2013).

Zebrafish as a model of ICF syndrome

Previously, our lab described the first viable model of ICF syndrome (Rajshekar et al. 2018). Rajshekar et al. describes that mutation of *zbtb24* in zebrafish leads to DNA hypomethylation. The hypomethylation seen in these mutants is located at the pericentromere. DNA methylation elsewhere across the genome is relatively intact. This ICF syndrome model is the first to recapitulate key phenotypic hallmarks of ICF syndrome. A mutation of *zbtb24* leads to a progressive loss of DNA methylation at the pericentromere, facial anomalies characterized

by elongation of the snout, reduced growth and life span, and a decrease in immunoglobulins in the presence of normal lymphoid cell counts (Rajshekar et al. 2018).

Dissertation overview

The role of pericentromeric DNA methylation remains unclear. The zebrafish model for ICF syndrome offers an avenue to research pericentromeric hypomethylation and its relation to disease pathology. My dissertation work aimed to establish a second viable model for ICF syndrome along with determining how pericentromeric hypomethylation relates to disease pathology in ICF syndrome. To this end, my thesis focuses on mutation of the zebrafish ortholog of *cdca7a*, which is thought to function downstream of Zbtb24 in the regulation of pericentromeric methylation. In chapter 2, I compare general developmental phenotypes between *cdca7a* and *zbtb24* homozygous deletion mutants, uncovering unexpected differences between these two models. In chapter 3, I complement this phenotypic analysis by defining the DNA methylation loss in *cdca7a* homozygous mutants, demonstrating a progressive loss of 5mC in *cdca7a* mutants which initiates earlier than that observed in *zbtb24* mutants. In chapter 4, I identify altered transcriptional signatures associated with the onset of methylation loss in *cdca7a* mutants. Due to immunodeficiency being a common cause of lethality in ICF patients, in chapter 5, I perform additional characterization of the immune system in *cdca7a* mutants.

Chapter 2: CHARACTERIZING AN ANIMAL MODEL OF ICF SYNDROME

PREFACE

There are four known causative genes for ICF syndrome. ICF syndrome patients harbor homozygous hypomorphic mutations in DNMT3B or are homozygous for predicted loss of function mutations in either ZBTB24, CDCA7 or HELLS (de Greef et al. 2011; Thijssen et al. 2015; Xu et al. 1999). Thus far, attempts to make a mouse model of ICF syndrome has been unsuccessful. Mice with ICF-like mutations in *Dnmt3b* exhibit smaller body size and facial abnormalities. However, these mice die within 24 hours of birth (Ueda et al. 2006). Deletion of *Zbtb24* in mice leads to embryonic lethality, and detailed characterization of these mutants has not been reported (Wu et al. 2016). Homozygous mutation of *Hells* causes perinatal lethality and a global loss of DNA methylation (Tao et al. 2011). There have been no current publications on an attempt to generate an ICF mice model by deleting *Cdca7*.

Our lab has focused on establishing zebrafish models of ICF syndrome to gain a better understanding of the consequences of a loss of DNA methylation specifically at the pericentromere. A previous publication by our laboratory reported that homozygous deletion of *zbtb24* caused growth retardation beginning at two weeks post fertilization. This size reduction persisted into adulthood. Mutation of *zbtb24* also leads to a decrease in survival. Fewer than ten percent of *zbtb24* mutants survived past eight months post fertilization. Similar to ICF syndrome patients, *zbtb24* mutants have facial abnormalities characterized by elongated snouts, and show immunodeficiency characterized by immunoglobulin deficiency. Along with these phenotypes, the *zbtb24* mutants are sterile (Rajshekar et al. 2018).

In this chapter, I focus on Cdca7. In zebrafish, and in humans, there are two *cdca7* genes. We chose to focus on *cdca7a* as it is most similar to the human CDCA7 ortholog implicated in ICF syndrome. The initial deletion mutation in *cdca7a* was created using Clustered Regularly Interspaced Short Palindromic Repeats Cas9 by former graduate student Ben Tischler. I recovered these mutations, developed the described genotyping protocols and characterized the mutants.

Analysis in this chapter includes both zygotic only and maternal zygotic *cdca7a* mutants. Zygotic mutants are mutants born to heterozygous parents. A heterozygous female deposits mRNA that the zygote uses until its own genome gets transcribed at the maternal to zygotic transition. The zygote relies on the maternal RNAs to translate needed proteins before the maternal to zygotic transition. *cdca7a* mRNA is maternally deposited (White et al. 2017) (Fig. 1). Thus, *cdca7a* mRNA is present in the zygote and the initially loaded *cdca7a* mRNA gets translated into protein until the zygotic genome turns on and the mutant *cdca7a* mRNA is transcribed. Maternal zygotic mutants born to a homozygous mutant female do not have any maternally deposited *cdca7a* mRNA. Thus, maternal zygotic mutants do not ever translate wild-type Cdca7a.

Given data suggesting Zbtb24 acts primarily to regulate *cdca7*a transcription, our initial hypothesis was that we would see similar phenotypes in *zbtb24* and *cdca7a* homozygous mutant zebrafish. However, I find that this is not the case.

RESULTS

Zebrafish CDCA7 and CDCA7L orthologs

The human genome encodes for two related genes: Cell Division Cycle-Associated 7 (CDCA7) which is mutated in ICF syndrome, and Cell Division Cycle-Associated 7 Like (CDCA7L). The human CDCA7 is located on chromosome 2 and has nine exons. The translated protein is 371 amino acids. The CDCA7 protein has two main domains. One is a domain known to interact with myc (Goto et al. 2006). The other is a four-CXXC zinc finger domain. The human CDCA7L is located on chromosome 7 and has ten exons. The translated protein is 454 amino acids. The CDCA7L protein has two domains. The domains include a myc interacting domain and a four-CXXC zinc finger domain (Fig. 2).

Similarly, the zebrafish genome encodes for two related genes: cell division cycle-associated 7a (*cdca7a*) and cell division cycle-associated 7b (*cdca7b*). *cdca7a* is located on chromosome 9 and has ten exons. The translated protein is 394 amino acids. The Cdca7a protein has a four-CXXC zinc finger domain similar to the human CDCA7 protein. *cdca7b* is located on chromosome 19 and has ten exons. The translated protein is 428 amino acids. The Cdca7b protein also has a four-CXXC zinc finger domain (Fig. 2).

The zebrafish *cdca7a* has a greater amino acid identity to the human ortholog CDCA7 than *cdca7b* (60.37% vs 44.63%) (Fig. 3). While neither *cdca7b* nor *cdca7a* have particularly high amino acid identity to the human ortholog CDCA7L (44.99% vs 42.69%). According to a phylogenetic tree that I generated, *cdca7a* is more closely related to CDCA7 than *cdca7b*. Thus, we chose to move forward with deleting *cdca7a* in zebrafish (Fig. 4).

Description of the zebrafish cdca7a deletion allele

Deletion of *cdca7a* was achieved by CRISPR/Cas9 mediated mutagenesis. Guide RNAs designed by former graduate student Ben Tischler were designed to exons 2 and 8. This mutant allele resulted in a 6 kb deletion which removed exons 3-7. This deletion encompasses amino acids encoding the zinc finger domain. I identified deletion mutants by screening candidate animals by PCR using primers spanning the deletion. Generation of this deletion was confirmed via Sanger sequencing. As expected, there were no transcripts detected in the region of the deletion by qRT-PCR. However, I was able to detect transcripts using primers designed to a region that was 5' to the deletion. These findings suggest that *cdca7a* mRNA is not subject to nonsense mediated decay. The translational start site is deleted in the *cdca7a* mutants. The largest open reading frame remaining encodes for a protein which is approximately 72 amino acids long and contains no identifiable motifs (Fig. 5).

To generate a second allele of *cdca7a*, I designed CRISPR RNAs to delete the transcriptional start site along with the translational start site (Fig. 6). I injected the crRNAs along with Cas9 protein to generate new mutant alleles, which will be recovered and analyzed by others in the laboratory.

Distinctions between phenotypes associated with zygotic loss of cdca7a and zbtb24

Survival

Previous work from our laboratory found that zygotic *zbtb24* homozygous mutant larva appeared grossly normal until 2-3 weeks post fertilization, at which point they began to show reduced growth (Rajshekar et al. 2018). Decreased survival was noted beginning at four months post fertilization with few animals surviving beyond eight months post fertilization.

To assess viability in *cdca7a* homozygous deletion mutants, I first intercrossed the *cdca7a* heterozygous adults. Homozygous *cdca7a* mutants born to heterozygous parents were recovered in normal Mendelian ratios (25%). I found that *cdca7a* homozygous mutants grew to adulthood with no obvious abnormalities (Fig. 7).

Growth

As adults, zygotic *zbtb24* homozygous mutants exhibited approximately a 66% reduction in body weight and exhibited a reduction in length compared to wild-type siblings. Reduced body length was apparent as early as three weeks post fertilization. To address whether similar phenotypes were observed in *cdca7a* mutants, I first compared length between homozygous mutant adults and wild-type siblings from heterozygous parents raised under identical conditions. Standard length from snout to fin was measured in static images of wild-type and homozygous *cdca7a* mutant juveniles at 3.5 weeks post fertilization and in adults at eight months post fertilization as described by Parichy et al. 2009. Measurements were taken for four wild-type and four homozygous mutant animals at 3.5 weeks post fertilization and for six wild-type and six homozygous mutant individuals at eight months of age. GraphPad PRISM software was used to perform a student unpaired 2-tailed t-test, which revealed no significant difference in length between wild-type and *cdca7a* homozygous mutant animals. Next, I compared weight in *cdca7a* mutants and wild-type. Weight was taken for five wild-type and five *cdca7a* homozygous mutants. GraphPad PRISM software was used to perform a t-test, which revealed no significant difference in weight between wild-type and homozygous mutant animals (Fig. 7).

Facial Anomalies

Another feature often observed in ICF patients is facial anomalies, characterized by a flattened nasal bridge, widely spaced eyes, enlarged tongue, small jaw, and low set ears. Previous work from our laboratory documented that *zbtb24* mutant zebrafish also exhibit facial abnormalities including an elongated snout. To see if this phenotype was recapitulated in the *cdca7a* mutants, I measured the facial structure of the fish. To normalize this measurement, I calculated the ratio of the distance from the snout to the anterior of the eye relative to the distance between the snout and the posterior of the eye as previously described by Rajshekar et. al 2018. Comparison of this ratio in six wild-type and six mutant zebrafish revealed no significant difference between *cdca7a* mutants and *cdca7a* wild-type siblings (Fig. 7). This finding suggests that *cdca7a* mutants to do not recapitulate facial abnormalities reported in *zbtb24* mutant zebrafish and many ICF syndrome patients.

Fertility

A previous publication from our laboratory showed that *zbtb24* mutants were sterile (Rajshekar et al. 2018). In contrast, I found that intercrosses between *cdca7a* homozygous mutant fish were generally successful in generating fertile progeny. Although I did not conduct rigorous breeding experiments, casual observations suggested that *cdca7a* homozygous mutant fish were able to lay eggs in amounts similar to their wild-type siblings, and eggs showed high levels of fertility that were not obviously different than observed for wild-type crosses. This allowed me to assess for the same morphological phenotypes in embryos without maternally loaded *cdca7a* mRNA.

Phenotypic characterization of maternal zygotic cdca7a mutants

Survival

To assess viability in *cdca7a* maternal zygotic homozygous deletion mutants, I first intercrossed the *cdca7a* homozygous mutant adults. Homozygous parents produced clutches of fertilized embryos of comparable size to wild-type (~200/ cross) and nearly all maternal zygotic progeny grew to adulthood with no obvious abnormalities (Fig. 8). Greater than 90% of *cdca7a* maternal zygotic mutant adults were viable at eight months post fertilization.

Growth

To assess growth phenotypes in the *cdca7a* maternal zygotic mutants, I first intercrossed the *cdca7a* mutants. Weight was taken at eight months post fertilization for five wild-type and five maternal zygotic mutants. Standard length was taken at both 3.5 weeks post fertilization and at eight months post fertilization as described above. The mutants exhibited similar body weight and length compared to wild-type raised under identical conditions (Fig. 8).

Facial Anomalies

To see if the facial anomalies phenotype is recapitulated in the *cdca7a* maternal zygotic mutants, I measured the facial structure of six fish per genotype using similar methods to those described above. Again, I observed no difference in the facial structure between *cdca7a* maternal zygotic mutants and *cdca7a* wild-type (Fig. 8). This finding suggests that *cdca7a* maternal zygotic mutants do not recapitulate the facial anomalies that have been reported in ICF syndrome patients.

Fertility

Intercrosses between maternal zygotic mutant fish were successful. Although not rigorously tested, *cdca7a* maternal zygotic mutant fish were able to lay embryos in amounts similar to that of wild-type.

DISCUSSION

This deletion in *cdca7a* represents the first animal model of *cdca7a* deletion. The maternal zygotic mutants give us an avenue to better understand the molecular phenotypes associated with early loss of *cdca7a*. Intriguingly, my findings suggest that, at least in zebrafish, *cdca7a* mutants do not recapitulate the failure to thrive and facial anomalies that have been reported in ICF syndrome patients.

Given that previous studies have shown that ZBTB24 regulates CDCA7 expression, the differences in phenotypes between *zbtb24* and *cdca7a* mutants was somewhat unexpected. One possibility is that the second homolog of CDCA7 might compensate for some functions of *cdca7a*. In zebrafish, it has been shown that if the promoter and transcription start site are not deleted it could lead to mutant mRNA decay which leads to transcriptional compensation of a closely related gene (El-Brolosy et al. 2019). Thus, *cdca7b* might be performing transcriptional compensation for *cdca7a*. However, I do not find evidence of nonsense mediated decay in the *cdca7a* deletion mutants, and zebrafish Cdca7b appears fairly diverged from Cdca7a, especially in the N terminus.

Another intriguing possibility is that given Zbtb24's predicted roles as a transcription factor, it may have roles in growth and development that are independent of its role in regulating

cdca7a and DNA methylation. Zbtb24 could be a transcription factor that regulates other genes that leads to the morphological phenotypes seen in the *zbtb24* mutants.

ICF syndrome in patients exists on a spectrum. Some patients exhibit the morphological phenotypes that have been reported such as growth retardation and facial anomalies. However, this is not seen in all ICF syndrome patients. Intriguingly, *CDCA7* mutations are much rarer than *ZBTB24* mutations in ICF patients (~4% vs ~30%), raising the possibility that additional undiagnosed patients exist with *CDCA7* mutations. These patients may have been overlooked for clinical diagnosis due to less obvious developmental abnormalities.

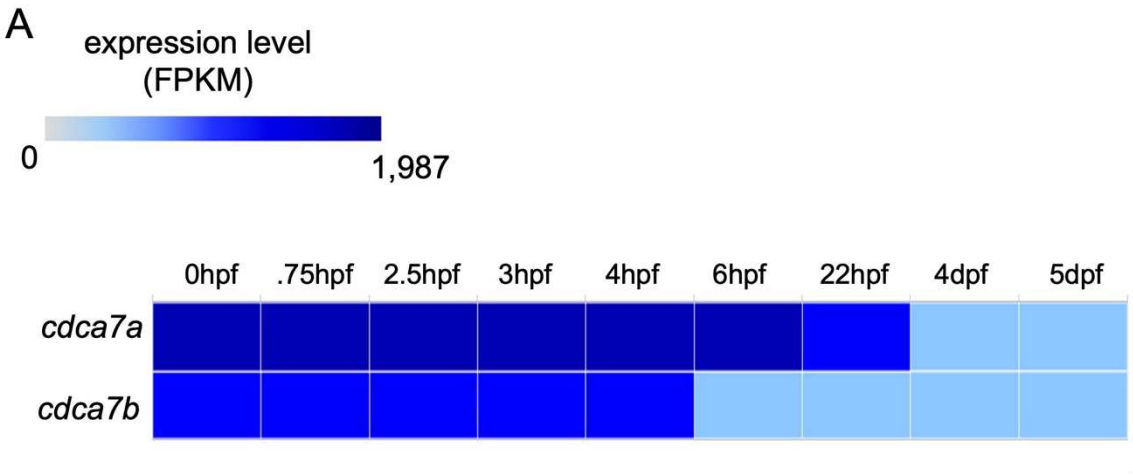


Figure 1: *cdca7a* and *cdca7b* are maternally deposited
(A) Expression levels of *cdca7a* and *cdca7b* measured in FPKM. Values are from expression atlas which obtained values from RJ White and Colleagues expression study of zebrafish mRNA levels across developmental time points (White et al. 2017).

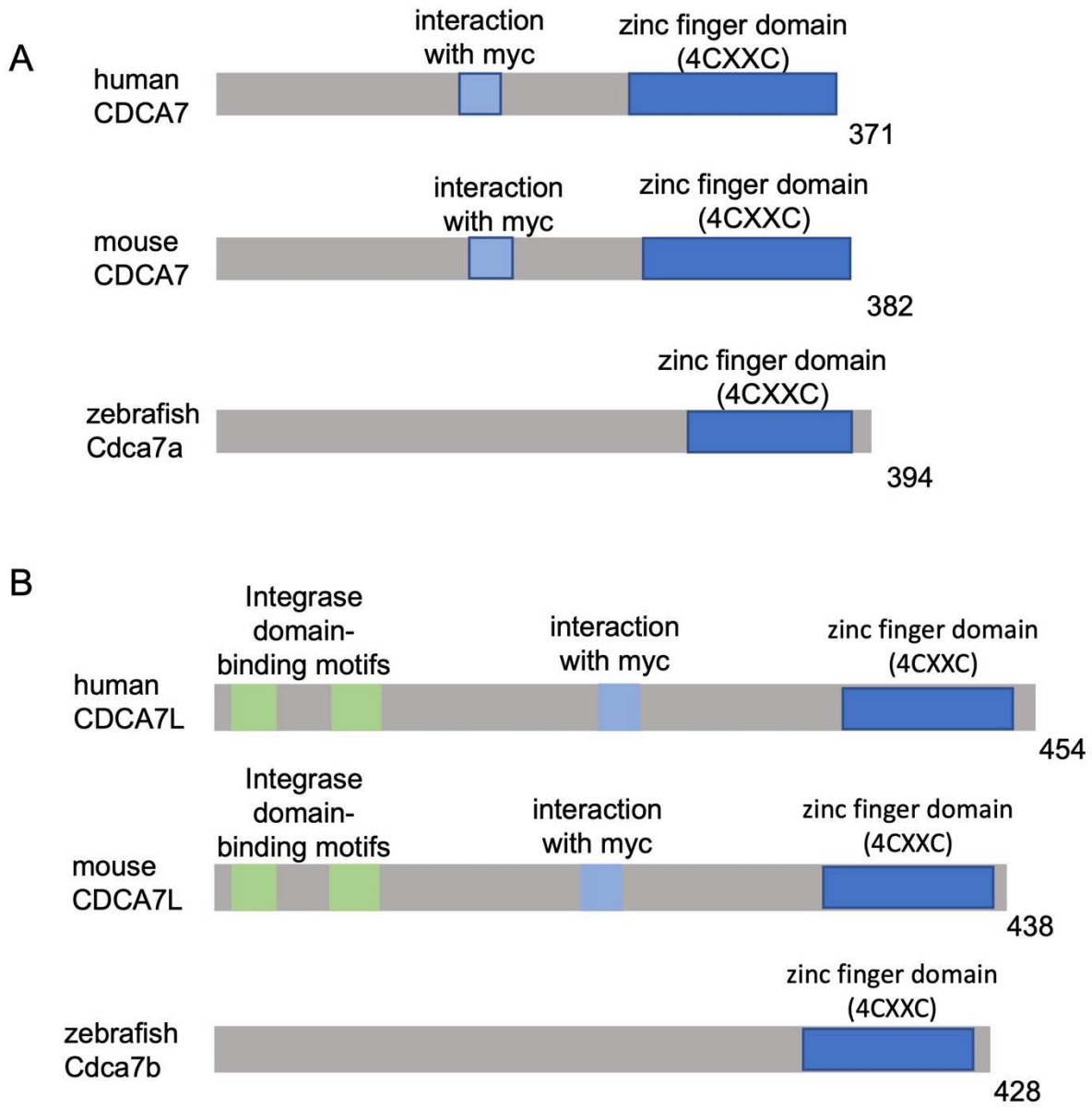


Figure 2: Zebrafish Cdca7a is conserved
(A) Schematic of human, mouse, and zebrafish Cdca7 proteins. Predicted domains highlighted and marked. (B) Schematic of human and mouse CDCA7L proteins compared to zebrafish Cdca7b protein. Predicted domains highlighted and marked.

A

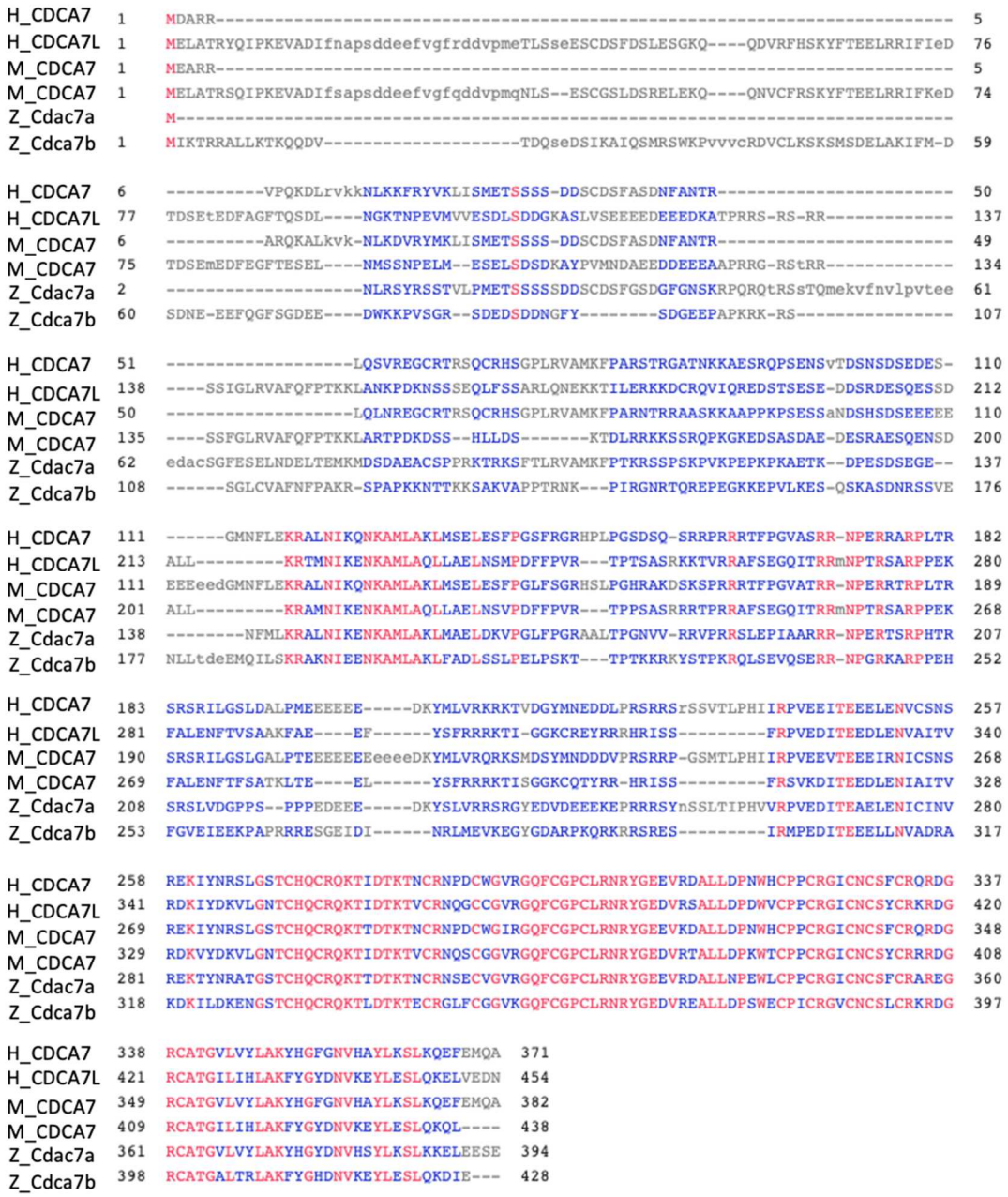


Figure 3: Protein sequence alignment of CDCA7
(A) Sequence alignment of human (H) CDCA7, human CDCA7L, mouse (M) CDCA7, mouse CDCA7L, zebrafish (Z) Cdca7a, and zebrafish Cdca7b. Alignment was generated using COBALT (Papadopoulos and Agarwala 2007).

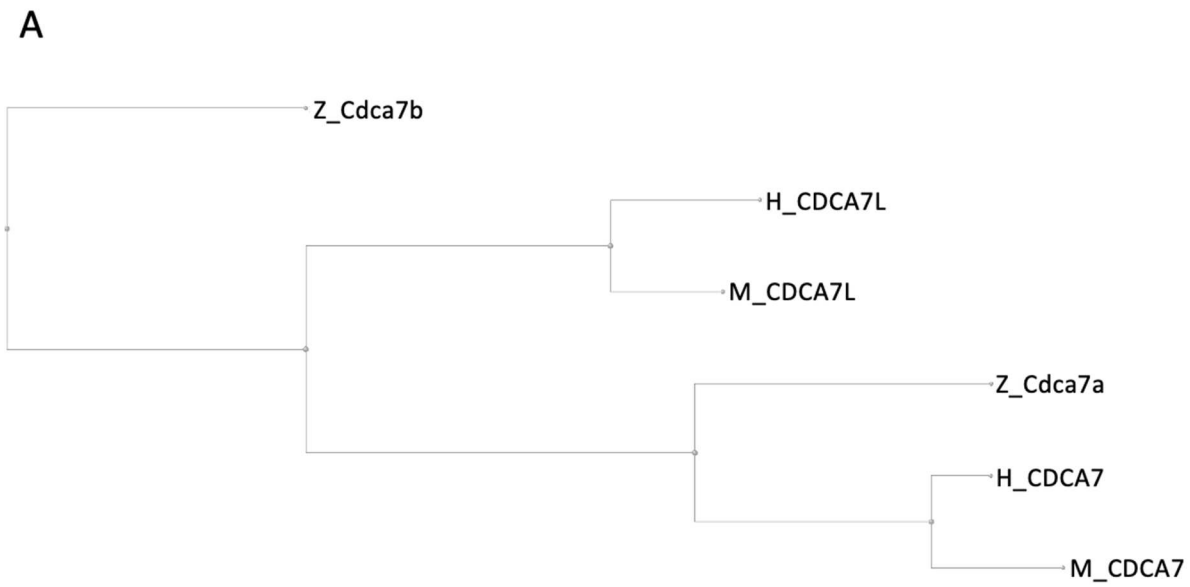
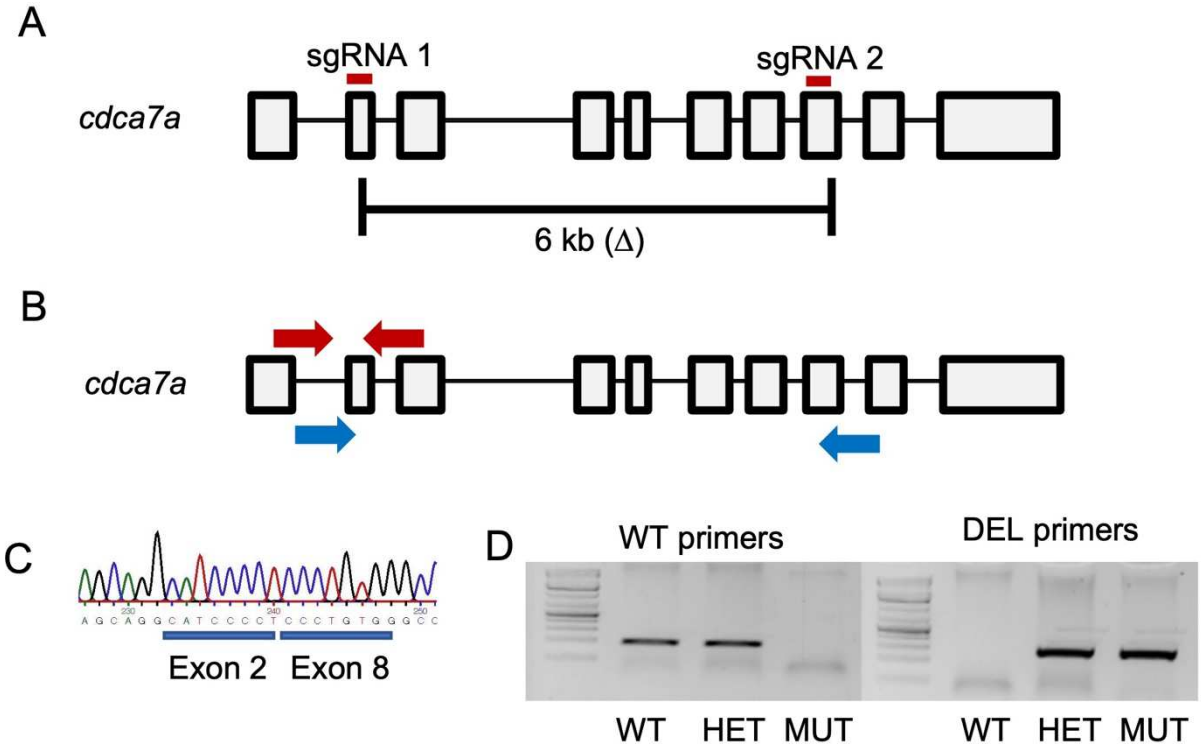


Figure 4: Zebrafish Cdca7a is more like human and mouse CDCA7
(A) Phylogenetic tree of human (H) CDCA7, human CDCA7L, mouse (M) CDCA7, mouse CDCA7L, zebrafish (Z) Cdca7a, and zebrafish Cdca7b. Phylogenetic tree was generated using COBALT (Papadopoulos and Agarwala 2007).


A
sgRNA 1
sgRNA 2
cdca7a
6 kb (Δ)
B
cdca7a
C
Exon 2
Exon 8
D
WT primers
DEL primers
WT
HET
MUT
WT
HET
MUT

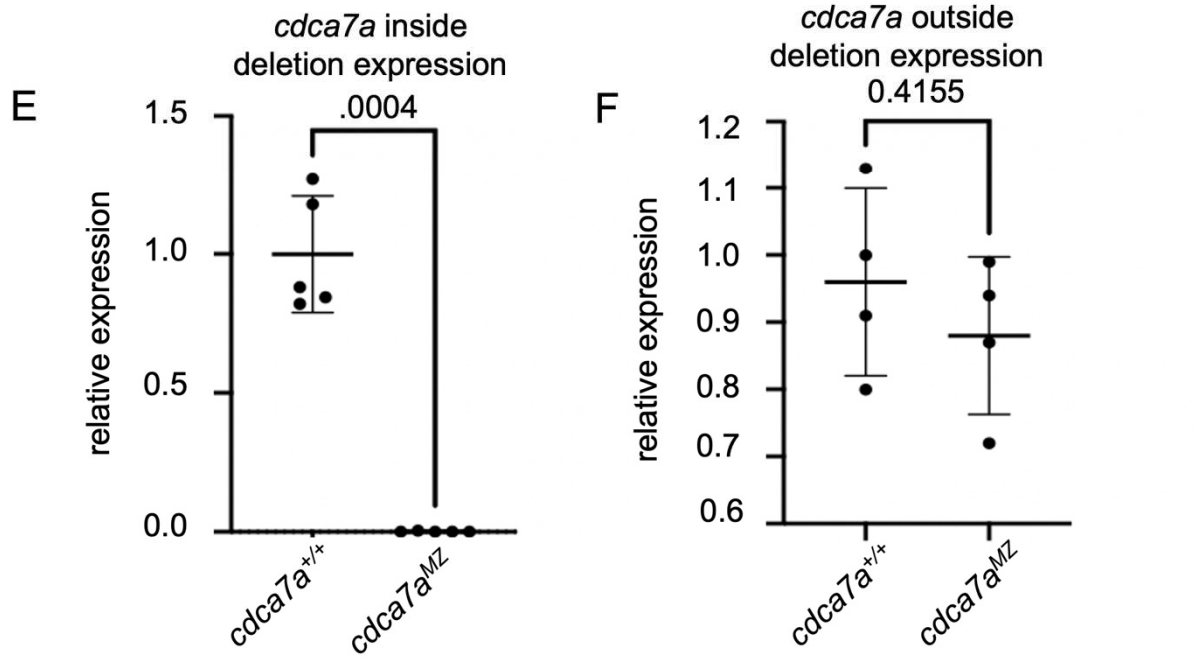

E
cdca7a inside
deletion expression
.0004
relative expression
1.5
1.0
0.5
0.0
cdca7a+/+
cdca7aMZ
F
cdca7a outside
deletion expression
0.4155
relative expression
1.2
1.1
1.0
0.9
0.8
0.7
0.6
cdca7a+/+
cdca7aMZ

Figure 5: Generating a deletion in the zebrafish ortholog of *cdca7a*
(A) Schematic of zebrafish *cdca7a* with the location of sgRNA target sequences indicated in red. Brackets represent the region deleted. (B) Schematic of zebrafish *cdca7a* gene showing the location of genotyping primers. Red arrows indicate wild-type primers. Blue arrows represent deletion primers. (C) Sequence trace confirming the generation of a deletion in *cdca7a*. (D) Primers from panel B used to amplify product in specified lanes. (E) qRT-PCR analysis of *cdca7a* mRNA with primers inside of the deletion in $cdca7a^{+/+}$ and $cdca7a^{-/-}$ zebrafish at 6 weeks post fertilization (wpf) (n=5 for each group). (F) qRT-PCR analysis of *cdca7a* mRNA with primers outside of the deletion in $cdca7a^{+/+}$ and $cdca7a^{-/-}$ zebrafish at 6 wpf (n=4 for each group).

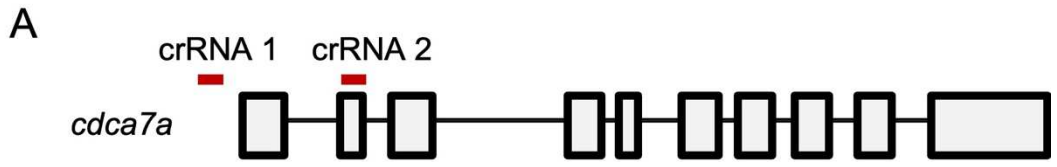


Figure 6: Generating a new mutation allele in the zebrafish ortholog of *cdca7a*

(A) Schematic of zebrafish *cdca7a* with the location of crRNA target sequences indicated in red.

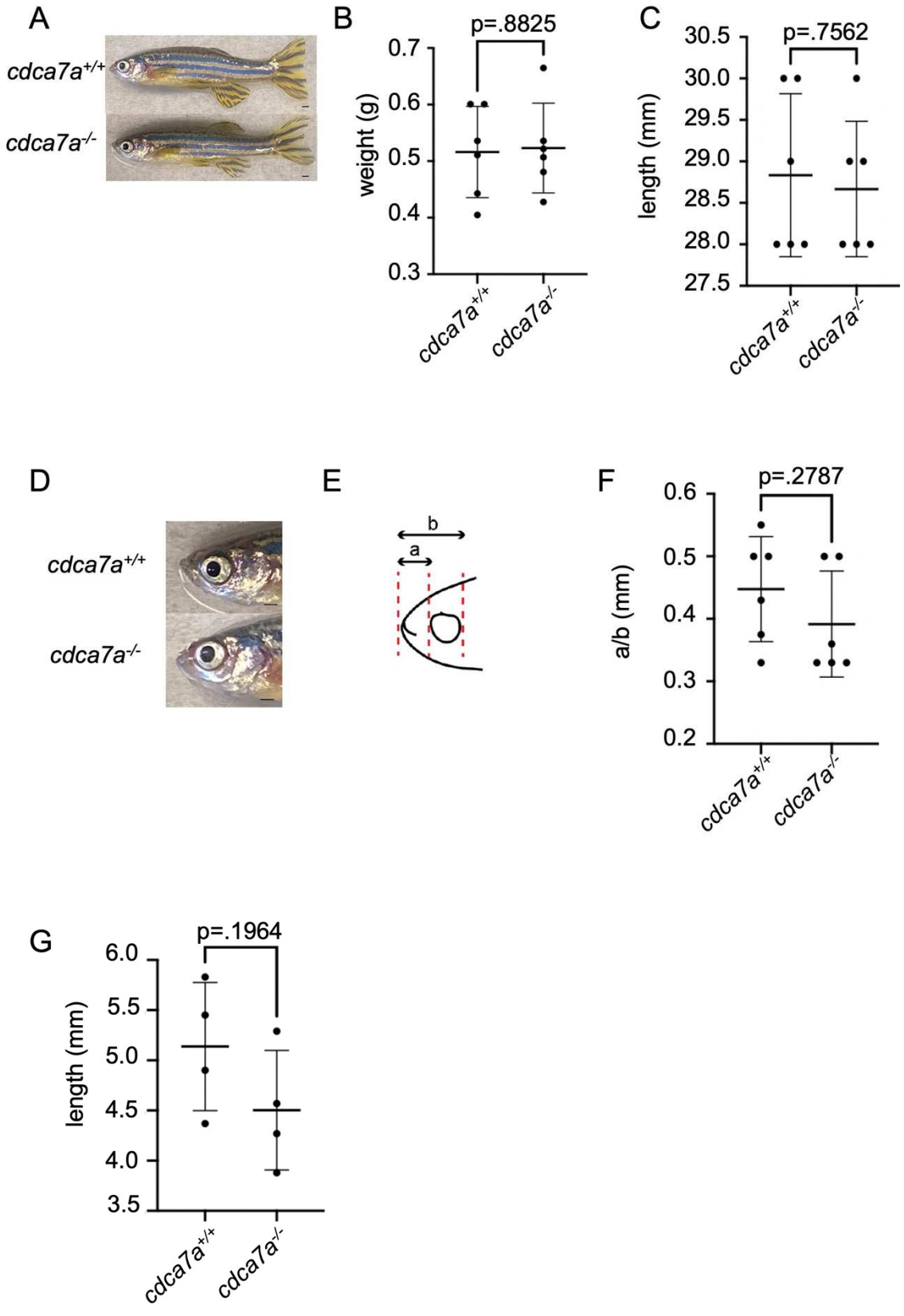

A
cdca7a+/+
cdca7a-/-
B
p=.8825
weight (g)
0.7
0.6
0.5
0.4
0.3
cdca7a+/+
cdca7a-/-
C
p=.7562
length (mm)
30.5
30.0
29.5
29.0
28.5
28.0
27.5
cdca7a+/+
cdca7a-/-
D
cdca7a+/+
cdca7a-/-
E
b
a
F
p=.2787
a/b (mm)
0.6
0.5
0.4
0.3
0.2
cdca7a+/+
cdca7a-/-
G
p=.1964
length (mm)
6.0
5.5
5.0
4.5
4.0
3.5
cdca7a+/+
cdca7a-/-

Figure 7: Growth defects in *cdca7a*$^{-/-}$
(A) Representative images of *cdca7a* wild-type and mutant zebrafish at 8 months post fertilization (mpf). Scale bar: 1mm. (B) Weight of *cdca7a* wild-type and mutant zebrafish at egiht mpf (n=5 for each group). (C) Standard length measurements for *cdca7a* wild-type and mutant zebrafish at 8 mpf (n=6 for each group). (D) Representative images of facial structure of *cdca7a* wild-type and mutant zebrafish at 8 mpf. Scale bar: 1mm. (E) Schematic of facial structure measurement. (F) Quantification of facial structure in *cdca7a* wild-type and mutant zebrafish at 8mpf (n=6 for each group). (G) Standard length measurements for *cdca7a* wild-type and mutant zebrafish at 3.5 wpf (n=4 for each group).

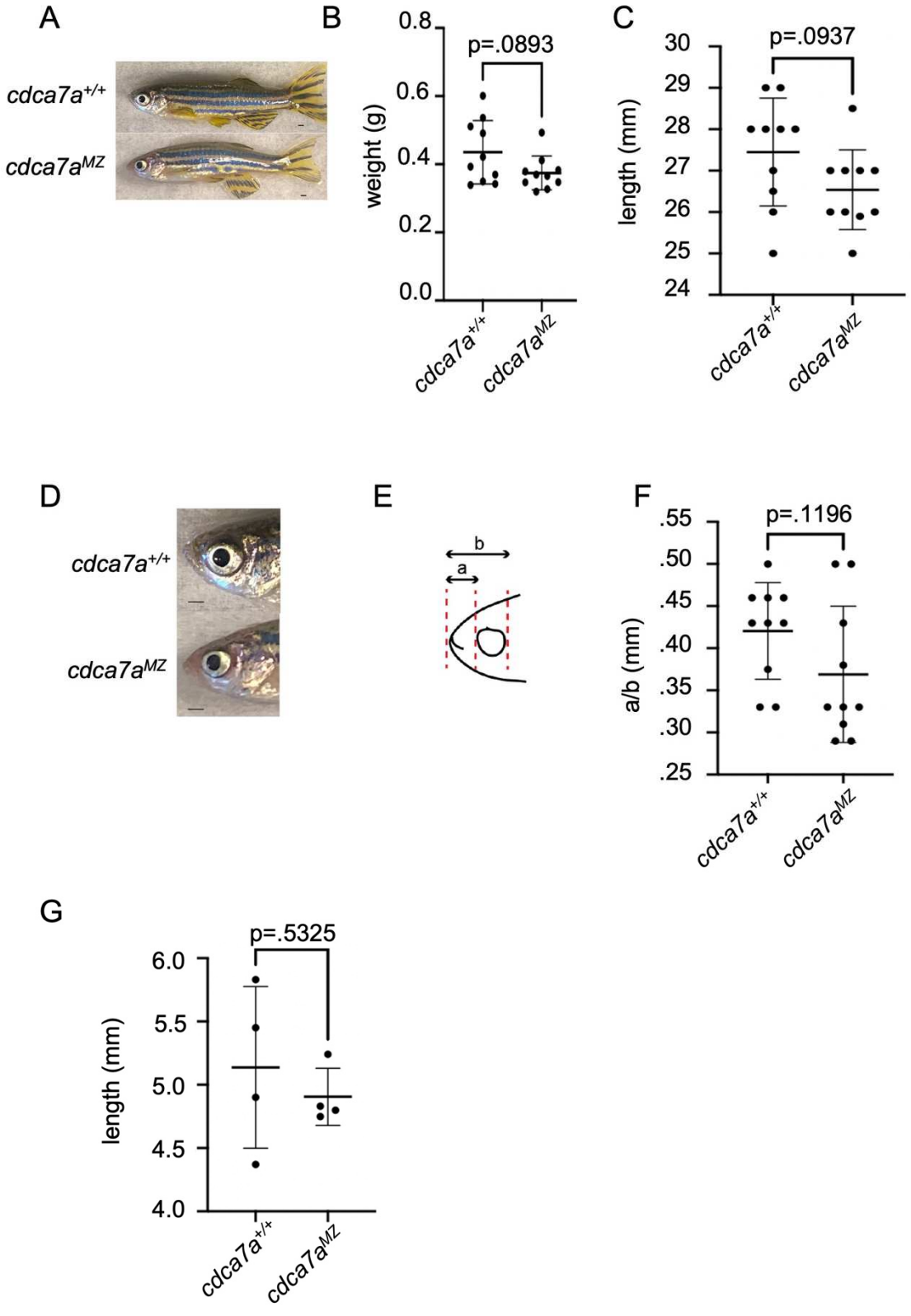

A
cdca7a+/+
cdca7aMZ
B
weight (g)
0.8
0.6
0.4
0.2
0.0
p=.0893
cdca7a+/+
cdca7aMZ
C
length (mm)
30
29
28
27
26
25
24
p=.0937
cdca7a+/+
cdca7aMZ
D
cdca7a+/+
cdca7aMZ
E
b
a
F
a/b (mm)
.55
.50
.45
.40
.35
.30
.25
p=.1196
cdca7a+/+
cdca7aMZ
G
length (mm)
6.0
5.5
5.0
4.5
4.0
p=.5325
cdca7a+/+
cdca7aMZ

Figure 8: Growth defects in *cdca7a*^{MZ}
(A) Representative images of *cdca7a* wild-type and maternal zygotic mutant zebrafish at eight mpf. Scale bar: 1mm. (B) Weight of *cdca7a* wild-type and maternal zygotic mutant zebrafish at eight mpf (n=5 for each group). (C) Standard length measurements for *cdca7a* wild-type and maternal zygotic mutant zebrafish at 8 mpf (n=6 for each group). (D) Representative images of facial structure of *cdca7a* wild-type and maternal zygotic mutant zebrafish at 8 mpf. Scale bar: 1mm. (E) Schematic of facial structure measurements. (F) Quantification of facial structure in *cdca7a* wild-type and maternal zygotic mutant zebrafish at 8mpf (n=6 for each group). (G) Standard length measurements for *cdca7a* wild-type and maternal zygotic mutant zebrafish at 3.5 wpf (n=4 for each group).

Chapter 3: EPIGENETIC ANALYSIS OF ANIMAL MODELS OF ICF SYNDROME

PREFACE

DNA methylation involves the addition of a methyl group to the fifth position of the cytosine ring. It is a highly conserved epigenetic mark that is necessary for vertebrate viability. The establishment of DNA methylation is performed by DNMT3B and DNMT3A. DNA methylation is maintained by DNMT1. DNA methylation is found across the genome and is heavily enriched at repetitive elements such as pericentromeric repeats and transposable elements. The sequences of pericentromeric repeats are not conserved across species. However, species with DNA methylation all have heavily methylated pericentromeric repeats. It is not known how DNA methylation is established at pericentromeric repeats or how a specific loss of DNA methylation at pericentromeric sequences relates to disease pathology.

A key molecular hallmark of ICF syndrome is a specific loss of DNA methylation at the pericentromere. Patients with ICF syndrome type 1 - 4 exhibit a loss of DNA methylation at pericentromeric satellite repeats 2 and 3, while patients with ICF syndrome types 2 - 4 also have hypomethylation of alpha satellite repeats.

In zebrafish, it has been shown through Southern Blots that zygotic loss of *zbtb24* leads to a progressive loss of DNA methylation of pericentromeric repeats starting at two weeks post fertilization. This loss of DNA methylation coincided with the onset of the ICF like morphological phenotypes (Rajshekar et al. 2018). We hypothesize that maternally deposited *zbtb24* supports wild-type methylation levels during early development in these mutants.

In this chapter I aim to discover if *cdca7a* mutants exhibit a loss of DNA methylation at pericentromeric repeats. I use two orthogonal methods to analyze DNA methylation at the pericentromere. One method for analyzing DNA methylation at the pericentromere is through the use of qualitative Southern Blots. The second method I used to access the level of DNA methylation at the pericentromere is through the quantitative bisulfite-sequencing.

RESULTS

Comparable methylation loss in cdca7a and zbtb24 homozygous mutant adults

One explanation for the discrepancies in phenotypes between *zbtb24* and *cdca7a* zygotic mutants described in chapter 2 is that there could be differing levels of hypomethylation at sat1 in the two mutant lines. To test this hypothesis, I first compared DNA methylation levels at sat1 repeats in *cdca7a* and *zbtb24* homozygous zygotic mutant adults at eight months post fertilization, using a Southern Blot strategy. I used the methylation-sensitive restriction enzyme HpyCH4IV, which has a restriction site in the zebrafish satellite 1 repeat (sat1). This restriction enzyme is resistant to digestion when DNA methylation is present (Fig. 9). After running sequences on an agarose gel and transferring to a membrane, detection of digestion at zebrafish sat1 sequences was performed by using a biotinylated probe that was complementary to the sat1 sequence. Consistent with previous observations demonstrating that wild-type sat1 sequence is heavily methylated, we observed that genomic DNA isolated from wild-type animals was highly resistant to digestion with HpyCH4IV. In contrast, I found that genomic DNA from both mutants showed extensive hypomethylation (Fig. 10). From this observation, I conclude that similar to *zbtb24*, *cdca7a* is required for DNA methylation at sat1 pericentromeric sequences in adult zebrafish.

Next, I compared methylation levels in juvenile mutants. DNA was collected from wild-type, *cdca7a* maternal zygotic, *cdca7a* zygotic and *zbtb24* zygotic mutant animals at one month post fertilization, and analyzed by Southern Blot as described above. At this stage, comparable methylation loss was observed in *zbtb24* and *cdca7a* zygotic mutants, while much more substantial hypomethylation was observed in the *cdca7a* maternal zygotic mutants (Fig. 10). This suggests that similar to *zbtb24*, maternal *cdca7a* is sufficient to support initial methylation, and that this methylation may be progressively lost in zygotic mutants once the maternal protein is depleted.

cdca7a homozygous mutants exhibit an early loss of DNA methylation

Our previous research has shown that *zbtb24* zygotic mutants initially loose methylation around two weeks post fertilization (Rajshekar et al. 2018). In contrast, when we examined DNA methylation in *cdca7a* zygotic mutants as early as 2 - 3 days post fertilization, we observed substantial loss of DNA methylation (Fig. 10). This finding suggests, that *cdca7a* zygotic mutants begin to lose DNA methylation at an earlier time point than *zbtb24* zygotic mutants.

Finally, I examined *cdca7a* maternal zygotic mutants to determine if animals with extensive hypomethylation from early in development could be obtained. Comparison of samples from wild-type and maternal zygotic mutants at six hours post fertilization revealed significant digestion via Southern Blot in *cdca7a* maternal zygotic mutants when compared to wild-type (Fig. 10). This finding suggests, that *cdca7a* maternal zygotic mutants exhibit a loss of DNA methylation at the pericentromere early in development.

New method for analyzing DNA methylation at the pericentromere

Analysis of satellite repeats in zebrafish is typically assessed by performing a Southern Blot. However, this analysis is somewhat qualitative. To gain a better assessment of DNA methylation loss at sat1, we aimed to develop a method to quantitatively measure DNA methylation at sat1. I used MethylC sequencing to quantitatively measure DNA methylation at the highly repetitive region of sat1 (Urich et al. 2015). This method relies on sodium bisulfite conversion of DNA which causes the deamination of unmethylated cytosines to uracil. Only methylated cytosines remain after the bisulfite conversion (Fig. 11). This approach is able to provide nucleotide resolution of DNA methylation in the genome. Typically, whole genome bisulfite sequencing analysis is expensive as extensive sequencing coverage is required for analysis. In contrast, I used a strategy where I assessed sat1 repeats in bulk, allowing for a bulk assessment of methylation using significantly lower sequencing depths and at a much lower cost. This work was performed in collaboration with Christina Ethridge from the laboratory of Dr. Bob Schmitz, who assisted with the conversion and preparation of sequencing libraries.

Following bisulfite conversion and sequencing, I performed the computational analysis. I aligned the reads to sat1 and took the ratio of methylated reads to the total unmethylated and methylated reads to determine the percent methylation at sat1. I found that the use of MethylC-seq confirmed the results of the Southern Blot. I found that at three months post fertilization, both *cdca7a* mutants and maternal zygotic mutants showed hypomethylation at sat1. There was a reduction of 26% DNA methylation between *cdcda7a* zygotic and wild-type, while there was a 31% reduction of DNA methylation between *cdca7a* maternal zygotic mutants and wild-type. Through the use of GraphPad PRISM software, I found that at three months post fertilization there is no statistical difference in the DNA methylation status between *cdca7a* zygotic mutants

and maternal zygotic mutants (Fig. 12). I was able to confirm that at eight months post fertilization there is a similar loss of DNA methylation between *cdca7a* and *zbtb24* mutants (Fig. 12). My findings give a new way to analyze DNA methylation at highly repetitive regions of the genome in zebrafish.

DISSCUSSION

These findings suggest that despite the fact that phenotypes in *cdca7a* zygotic and maternal zygotic mutants are less severe than those observed in *zbtb24* homozygous mutants, *cdca7a* zygotic mutants and maternal zygotic mutants exhibit comparable hypomethylation of their pericentromeres in adulthood and begin to lose DNA methylation much earlier than *zbtb24* mutants. The *cdca7a* mutants represent the first animal model in which clear pericentromeric methylation is observed without obvious morphological phenotypes.

Our lab has previously generated a viable model for ICF syndrome. Mutation of *zbtb24* in zebrafish leads to a loss of pericentromeric DNA methylation at two weeks post fertilization (Rajshekar et al. 2018). This loss of DNA methylation coincided with the onset of the morphological phenotypes. One may assume that the loss of DNA methylation at the pericentromere is what caused the morphological phenotypes seen in the *zbtb24* mutants. However, *cdca7a* mutants begin to lose DNA methylation earlier than the *zbtb24* mutants. *cdca7a* mutants start to lose DNA methylation at the pericentromere as early as two days post fertilization, while *zbtb24* mutants do not start to lose DNA methylation until two weeks post fertilization. The early loss of DNA methylation seen in *cdca7a* mutants does not lead to ICF-like morphological phenotypes. Plus, *zbtb24* and *cdca7a* mutants exhibit a similar loss of DNA methylation at the pericentromere in adulthood. Thus, my findings suggest that a loss of DNA

methylation at the pericentromere does not lead to the ICF like morphological phenotypes. This suggests that the morphological phenotypes seen in *zbtb24* mutants are most likely due to Zbtb24 acting as a transcription factor for other genes that have a causative role in the morphological phenotypes present in the mutants.

Patients with ICF syndrome type 3 account for less than 10% of all ICF syndrome cases. Patients with a mutation in CDCA7 are rare and I hypothesize that they are underdiagnosed due to the mild phenotypes, but these people will still exhibit the molecular hallmark of ICF syndrome.

A

GATCCAGCCATAAAATGCATCATTCTTTTTTGTTTTAGAC
AACATTTCATGCACTGTTAAACATGTTAAAGCAAGTTGC
AAGTGAAAATCTATGTCTCTGACTGAGTTTGCATTACTGT
GATTTGACCTCTCTGCTGGCTGAGATAAGCTCATTTTCA
ACGTCCAATTCAGAATGTAATAAAACCTGATC

B

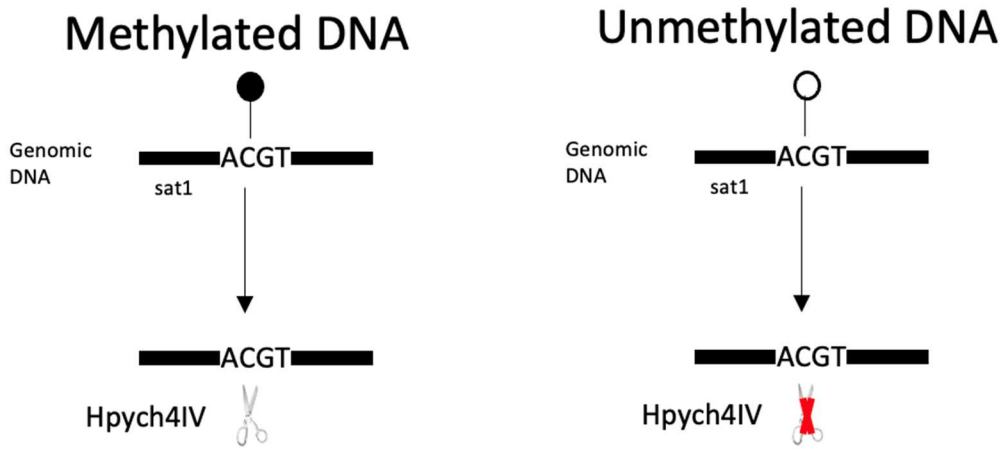


Figure 9: Zebrafish satellite sequence and restriction digestion
(A) Schematic of zebrafish satellite 1 sequence. Highlighted in yellow is the Hypch4IV restriction enzyme digestion site. (B) Schematic of restriction enzyme digestion completed before Southern Blot.

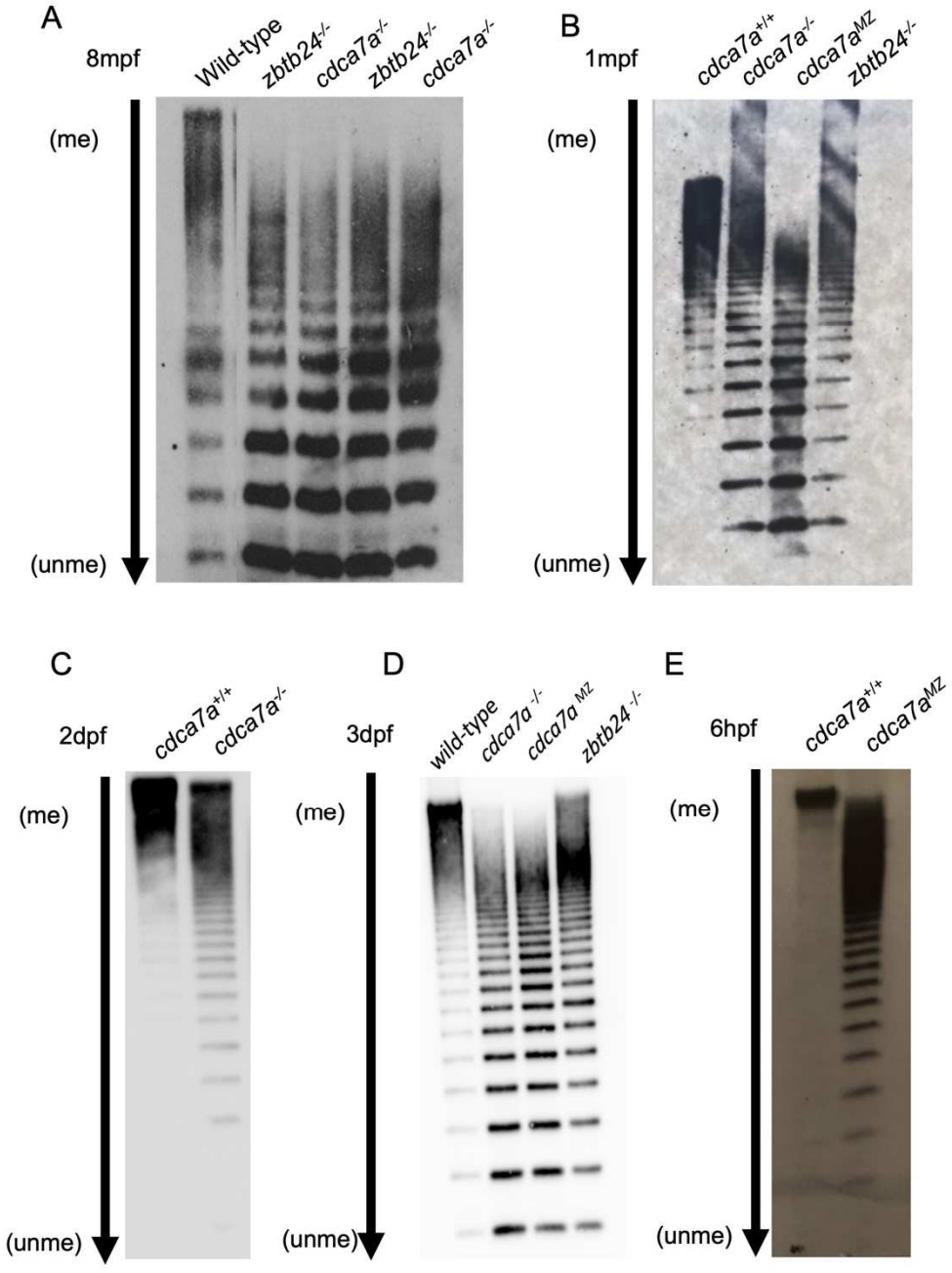

A
8mpf
Wild-type
zbtb24-/-
cdca7a-/-
zbtb24-/-
cdca7a-/-
(me)
(unme)
B
1mpf
cdca7a+/+
cdca7a-/-
cdca7aMZ
zbtb24-/-
(me)
(unme)
C
2dpf
cdca7a+/+
cdca7a-/-
(me)
(unme)
D
3dpf
wild-type
cdca7a -/-
cdca7a MZ
zbtb24 -/-
(me)
(unme)
E
6hpf
cdca7a+/+
cdca7aMZ
(me)
(unme)

Figure 10: DNA methylation loss in *cdca7a*$^{-/-}$
(A) Southern Blot of genomic DNA from 8 mpf wild-type, *cdca7a* mutant, and *zbtb24* mutant zebrafish digested with a methylation sensitive enzyme and probed for zebrafish sat1. Each lane represents one .5ug from one adult zebrafish fin. (B) Southern Blot of genomic DNA from 1 mpf wild-type, *cdca7a* mutant, *cdca7a* maternal zygotic, and *zbtb24* mutant zebrafish digested with HpyCH4IV and probed for zebrafish sat1. Each lane represents genomic DNA isolated from one adult zebrafish. (C) Southern Blot of genomic DNA from 2 dpf *cdca7a* wild-type and mutant zebrafish digested with HpyCH4IV and probed for zebrafish sat1. Each lane represents 1ug of DNA isolated from a pool of fifty embryos. (D) Southern Blot of genomic DNA from three days post fertilization wild-type, *cdca7a* maternal zygotic, *cdca7a* zygotic, and *zbtb24* zygotic mutants digested with HpyCH4IV and probed for zebrafish sat1. Each lane represents 0.5 ug of DNA isolated from a pool of ten larva. (E) Southern Blot of genomic DNA from six hours post fertilization *cdca7a* wild-type and maternal zygotic mutant zebrafish digested with a methylation sensitive enzyme HpyCH4IV and probed for zebrafish sat1. Each lane represents pooled samples of fifty embryos with 1ug of genomic DNA.

A

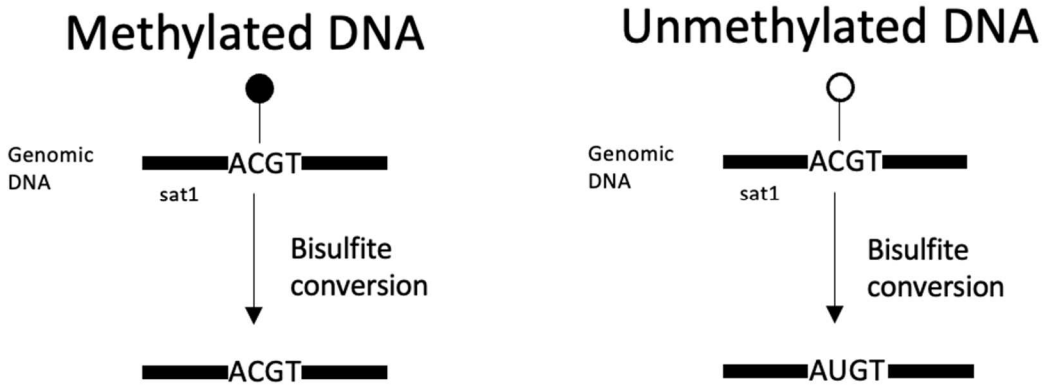


Figure 11: Bisulfite sequencing conversion
(A) Schematic of bisulfite conversion before PCR amplification and subsequent sequencing.

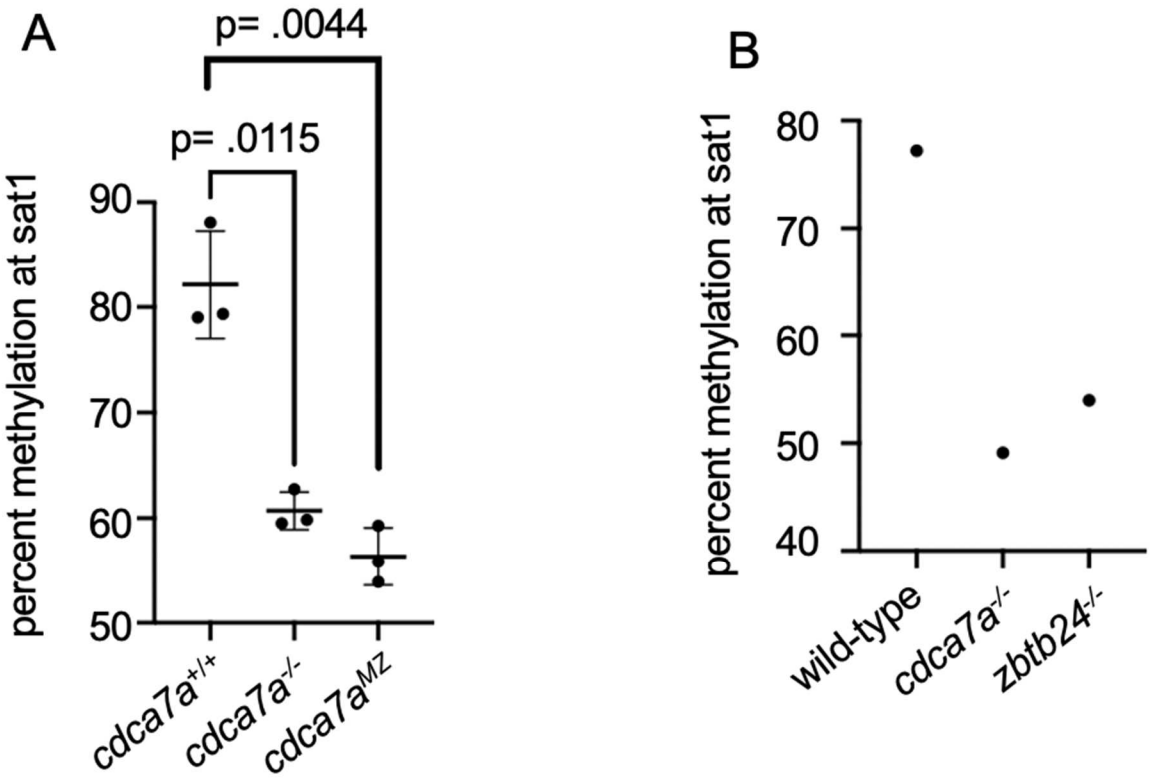


Figure 12: WGBS-seq analysis on *cdca7a*$^{-/-}$ and *zbtb24*$^{-/-}$
(A) Percent methylation at sat1 from three months post fertilization *cdca7a* wild-type, *cdca7a* mutant, and *cdca7a* maternal zygotic (n=3 for each group). (B) Percent methylation at sat1 from eight months post fertilization wild-type, *cdca7a* mutant and *zbtb24* mutant (n=1 for each group).

CHAPTER 4: EARLY CONSEQUENCES OF PERICENTROMERIC DNA HYPOMETHYLATION

PREFACE

Previous analysis of RNA transcription was performed in *zbtb24* homozygous mutants at two weeks post fertilization, when these mutants begin to show a loss of DNA methylation (Rajshekar et al. 2018). Here, I have described another animal model of ICF syndrome in which the phenotypic hallmark of pericentromeric DNA hypomethylation is present. This loss of pericentromeric DNA methylation occurs earlier in development, which gives us the avenue to assess the possible consequences of a loss of DNA methylation at early stages of development. A loss of *cdca7a* in homozygous mutants occurs as early as two days post fertilization at a time in which *cdca7a* maternal zygotic mutants also exhibit a loss of DNA methylation at the pericentromere. In this chapter I perform transcriptome analysis to identify some of the early consequences of a loss of DNA methylation at the pericentromere.

RESULTS

Data validation and normalization

I analyzed three homozygous mutant genotypes, *cdca7a* maternal zygotic, *cdca7a* zygotic and *zbtb24* zygotic mutants, compared to wild-type controls at three days post fertilization. To assess the early consequences of DNA pericentromeric hypomethylation, I chose to perform mRNA sequencing at three days post fertilization a time in which I had two viable mutants (*cdca7a* zygotic and maternal zygotic) that exhibited a loss of pericentromeric DNA methylation and one mutant (*zbtb24* zygotic) that was still heavily methylated at the pericentromere.

For my *cdca7a* maternal zygotic analysis, I compared three mutant samples to three wild-type controls. Samples generated an average of 19.2 million single end reads and produced an average of 92.57% reads that uniquely mapped to the genome. To ensure that my two conditions segregated away from each other I performed a principal component analysis (PCA) in R. Upon completion of the PCA I found that one sample from the *cdca7a* maternal zygotic mutants did not cluster together with the other *cdca7a* maternal zygotic mutants (Fig. 13). Thus, I removed this outlier from the analysis. This left two *cdca7a* maternal zygotic samples compared to three wild-type samples.

For my *cdca7a* zygotic analysis, I compared three mutant samples to three wild-type controls. Samples generated an average of 14.5 million single end reads and produced an average of 92.37% uniquely mapped reads. Upon completion of the PCA, I found that one sample from the *cdca7a* zygotic mutants did not cluster together with the other homozygous mutants (Fig. 14). Thus, I removed this outlier from the analysis. This left two *cdca7a* zygotic samples compared to three wild-type samples.

For my *zbtb24* zygotic analysis, I compared three mutant samples to three wild-type controls. Samples generated an average of 18 million single end reads and produced an average of 92.24% reads that uniquely mapped to the genome. Upon completion of the PCA I found that all mutants and wild-type samples clustered appropriately (Fig. 15).

Identification of differentially expressed genes

For broad visualization of the differentially expressed genes, I generated volcano plots for each of my three comparisons. For the comparison between *cdca7a* maternal zygotic mutants and wild-type there was 106 genes with a log2fold change of greater than 1.5 and an adjusted p-value of less than 0.05. On the other hand, there was 148 genes with a log2fold change of less

than -1.5 with an adjusted p-value of less than 0.05 (Fig. 16). For the comparison between *cdca7a* zygotic mutants and wild-type there was 103 genes with a log2fold change of greater than 1.5 with a significant adjusted p-value. While there were 80 genes with a log2fold change of less than -1.5 with a significant adjusted p-value (Fig. 17). The comparison between *zbtb24* zygotic mutants and wild-type generated 665 genes with a log2fold change of greater than 1.5 with an adjusted p-value of less than 0.05. Although, there was 166 genes with a log2fold change of less than -1.5 with an adjusted p-value of less than 0.05 (Fig. 18).

Intersections of increasing dysregulated genes in three mutants

At three days post fertilization I compared the dysregulated genes that were increasing between the three mutants using Venny (Oliveros 2015) . I found that there were 28 genes that were increased that were shared between only *cdca7a* maternal zygotic and zygotic mutants. Of these shared genes four were found to be involved in apoptosis according to an online resource, gProfiler. These genes were *si:dkey-239j18.3, zgc:174855, zgc:174153,* and *si:dkey-28g8.5.* There were 57 unique genes increased in *cdca7a* maternal zygotic and 43 unique genes increased in *cdca7a* zygotic mutants. Between *cdca7a* maternal zygotic mutants and *zbtb24* mutants there were 14 increasing shared genes. *cdca7a* zygotic and *zbtb24* zygotic mutants shared 25 genes that were increased. There were seven genes that were upregulated between all three genotypes (Fig. 19 and Table 3).

Intersections of decreasing dysregulated genes in three mutants

I compared decreasing genes that were dysregulated between the three mutants. There were six dysregulated genes between all three mutants. One of these genes, *gck*, according to

gProfiler is involved in neomycin, kanamycin, and gentamicin biosynthesis. Between *cdca7a* maternal zygotic and zygotic mutants there were 12 genes the same. There were 10 genes shared between *cdca7a* maternal zygotic and *zbtb24* zygotic mutants. One of these genes, *cepet1a,* is involved in phosphonate and phosphinate metabolism according to gProfiler. There were also 10 genes in common between *cdca7a* and *zbtb24* zygotic mutants (Fig. 19 and Table 4).

Identification of disrupted pathways using KEGG

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a pathway database that can be used to conduct genomic pathway analyses in many different species. Upon performing KEGG analysis in R I found that both *cdca7a* maternal zygotic and zygotic mutants both expressed dysregulated genes involved in apoptosis (Fig. 20 and Fig. 21). While the dysregulated apoptotic genes were not seen in *zbtb24* mutants (Fig. 22). Upon further examination, I found that apoptotic genes were increased in the *cdca7a* maternal zygotic and zygotic mutants, but not in the *zbtb24* mutants (Fig. 23 and Fig. 24). This could suggest that a loss of pericentromeric DNA hypomethylation could lead to an increase in apoptosis.

In addition to the apoptotic genes found dysregulated through KEGG analysis, I also found that there was dysregulation of the Fanconi anemia pathway, homologous recombination, and mismatch repair pathways in *cdca7a* maternal zygotic mutants. This could be a sign that there are changes in DNA damage signatures as early as three days post fertilization in *cdca7a* maternal zygotic mutants. In addition, I also saw that there was dysregulation of DNA replication and mismatch repair pathways in *zbtb24* zygotic mutants. I hypothesize that Zbtb24 could regulate transcription of some genes involved in these pathways.

DISSCUSSION

My findings suggest that *cdca7a* maternal zygotic and zygotic mutants exhibit an increase in apoptosis at three days post fertilization from whole larva. However, this is not seen in *zbtb24* mutants. The genes that I have shown to be increased in *cdca7a* maternal zygotic and zygotic mutants are *zgc:174855, zgc:174153, si:dkey-23j18.3,* and *si:dkey-26g8.5*. *zgc:174855, zgc:174153* and *si:dkey-26g8.5* are located in a cluster of genes on chromosome 12. It is thought that these genes could be tandem duplicates of *ctsl. ctsl* is a cytosine protease that is involved in the degradation of many proteins. Knockdown of CTSL in PC-12 cells lead to a decrease in apoptosis (Shen et al. 2019). Although CTSL is involved in apoptosis it is also thought to be involved in antigen processing and MHC class II immune responses (Gomes et al. 2020). Further research will be needed to determine if these genes actually lead to an increase in apoptosis.

Further research needs to be done to due to some limitations of my data set. There could be additional consequences of pericentromeric DNA methylation not seen in this data set due to batch effects. In this data set I used wild-type strain to compare my mutants to. A better control for this experiment would be to compare mutants to their wild-type siblings. In addition to this there could be tissue specific consequences that I do not see due to this being whole larva. Lastly, I could have picked a different time in which to analyze the consequences of a loss of DNA methylation at the pericentromere. There could be some effects that I am not seeing due to the three days post fertilization time point.

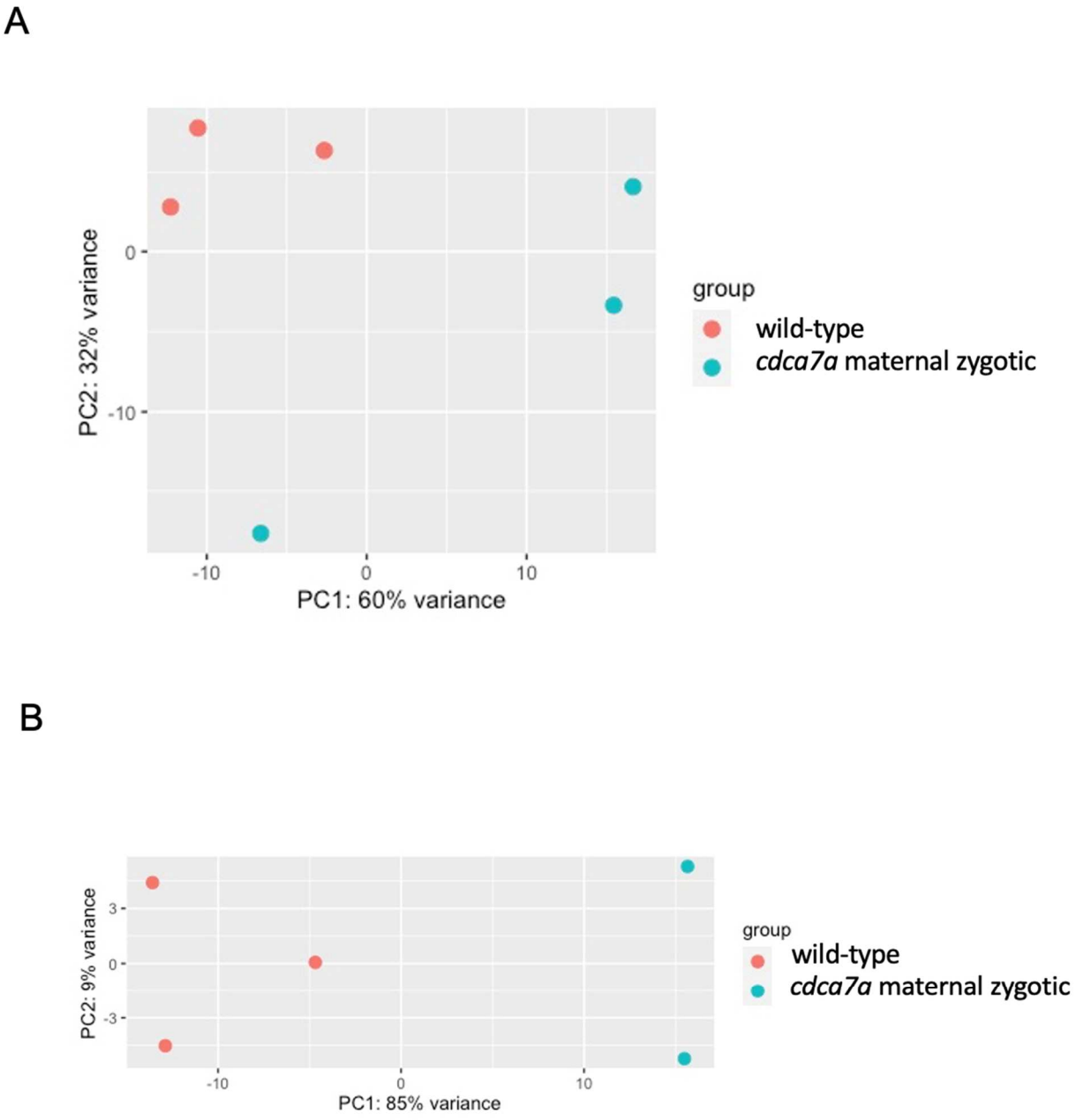
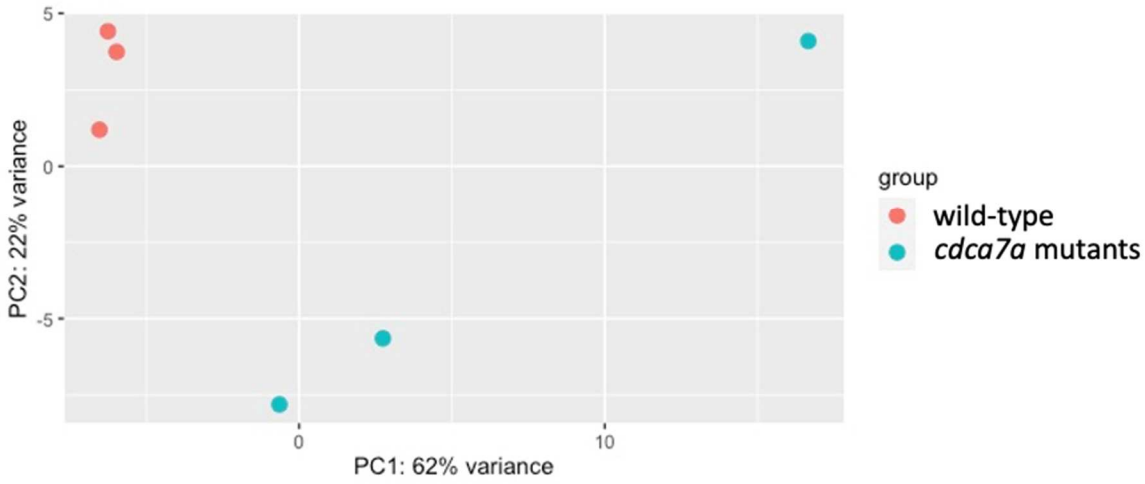


Figure 13: *cdca7a*^{MZ} 3dpf RNA sequencing clustering
(A) Principal components analysis of *cdca7a* maternal zygotic and wild-type. Plotted on the x-axis is PC1 with 60% of the variance and on the y-axis is PC2 with 32% of the variance. (B) Principal components analysis of *cdca7a* maternal zygotic and wild-type with the outlier *cdca7a* maternal zygotic removed. Plotted on the x-axis is PC1 with 85% of the variance and on the y-axis is PC2 with 9% of the variance.

A


5
0
-5
PC2: 22% variance
0
10
PC1: 62% variance
group
wild-type
cdca7a mutants

B

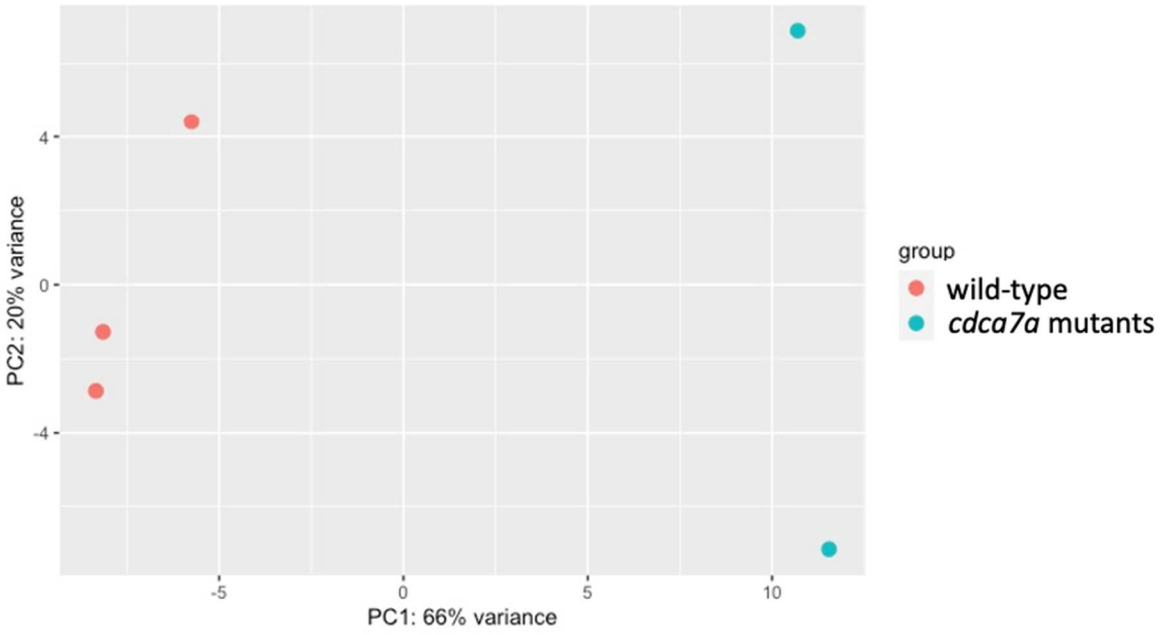

4
0
-4
PC2: 20% variance
-5
0
5
10
PC1: 66% variance
group
wild-type
cdca7a mutants

Figure 14: *cdca7a*$^{-/-}$ three days post fertilization (dpf) RNA sequencing clustering
(A) Principal components analysis of *cdca7a* zygotic and wild-type. Plotted on the x-axis is PC1 with 62% of the variance and on the y-axis is PC2 with 22% of the variance. (B) Principal components analysis of *cdca7a* zygotic and wild-type with the outlier *cdca7a* zygotic mutant removed. Plotted on the x-axis is PC1 with 66% of the variance and on the y-axis is PC2 with 20% of the variance.

A

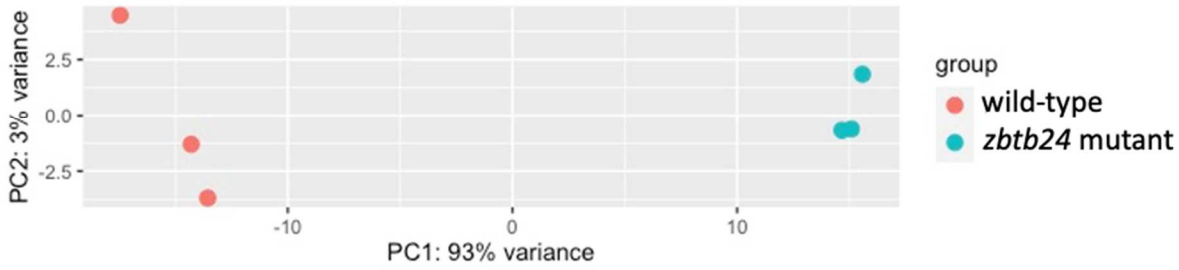


Figure 15: *zbtb24*$^{-/-}$ 3dpf RNA sequencing clustering
(A) Principal components analysis of *zbtb24* zygotic and wild-type. Plotted on the x-axis is PC1 with 93% of the variance and on the y-axis is PC2 with 3% of the variance.

A

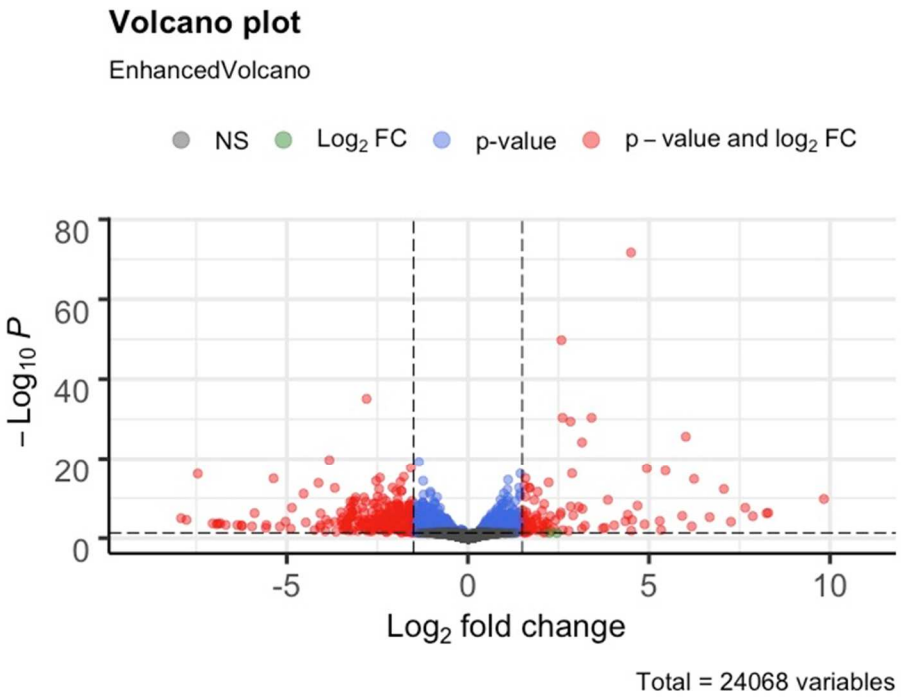


Figure 16: Differentially expressed genes in *cdca7a*^{MZ}
(A) Volcano plot of differentially expressed genes. Highlighted in gray are genes that with no significant difference with a p-value > 0.05. Highlighted in green are genes with a log2fold of either > 1.5 or < -1.5 and a p-value of > 0.05. Highlighted in blue are genes with a log2fold change between -1.5 and 1.5. They also have a p-value < 0.05. Highlighted in red are genes with a log2fold change of either > 1.5 or < -1.5. However, these genes also have a p-value < 0.05.

A

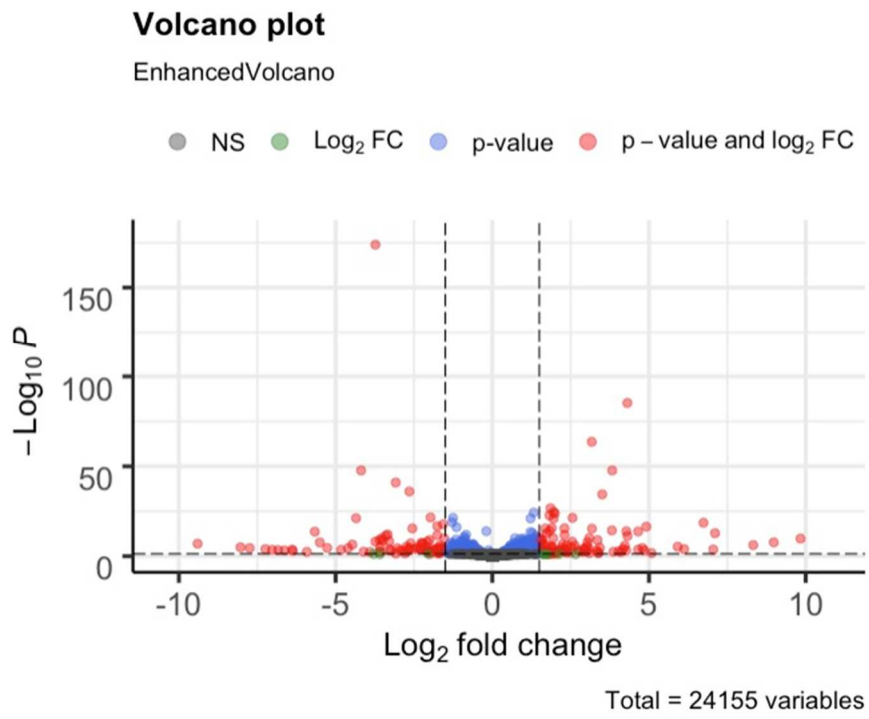


Figure 17: Differentially expressed genes in *cdca7a*$^{-/-}$ zygotic mutants
(A) Volcano plot of differentially expressed genes. Highlighted in gray are genes that with no significant difference with a p-value > 0.05. Highlighted in green are genes with a log2fold of either > 1.5 or < -1.5 and a p-value of > 0.05. Highlighted in blue are genes with a log2fold change between -1.5 and 1.5. They also have a p-value < 0.05. Highlighted in red are genes with a log2fold change of either > 1.5 or < -1.5. However, these genes also have a p-value < 0.05.

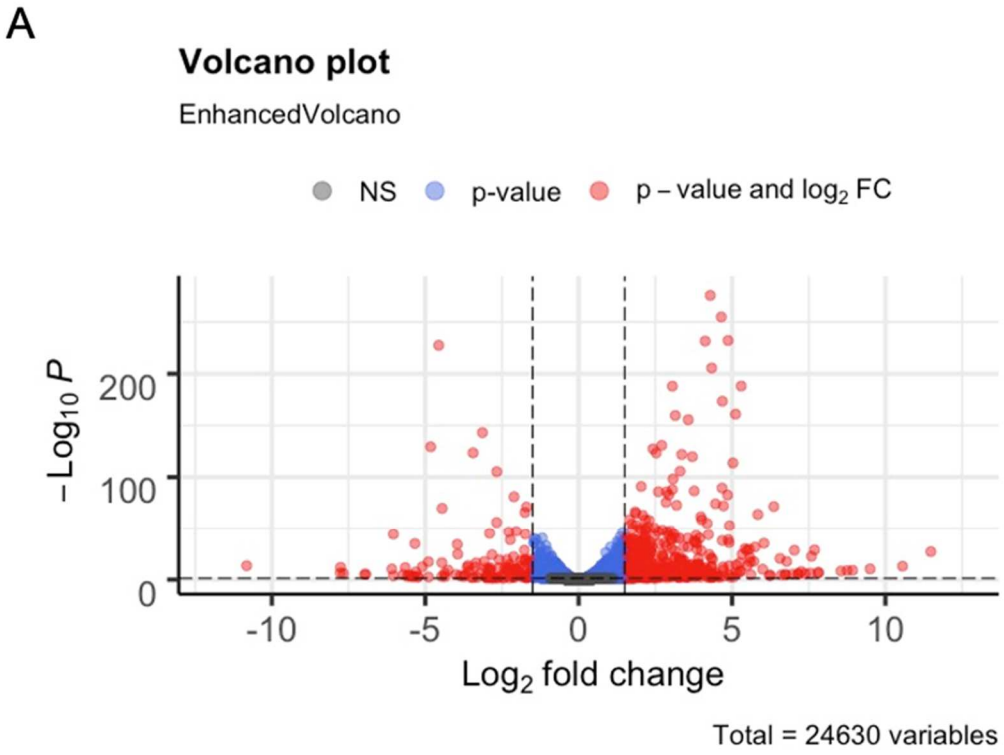
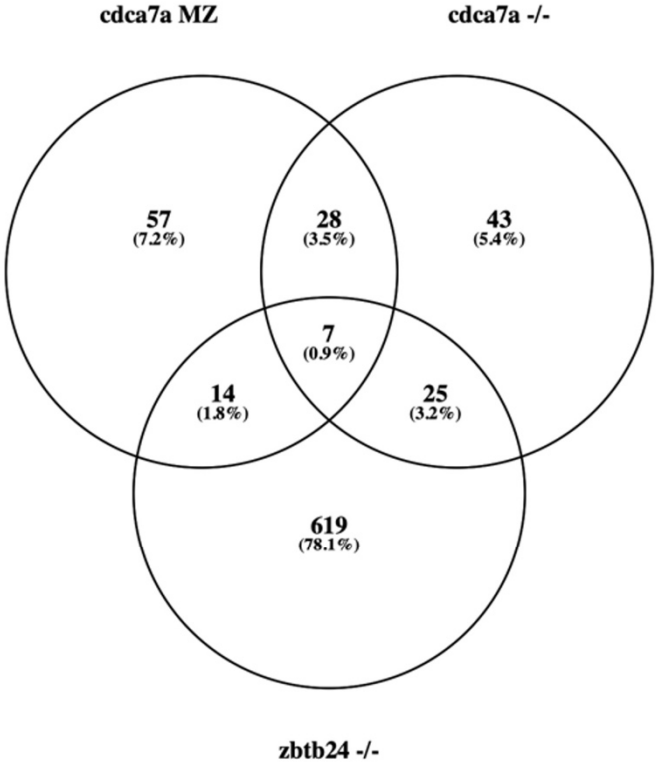


Figure 18: Differentially expressed genes in *zbtb24*$^{-/-}$
(A) Volcano plot of differentially expressed genes. Highlighted in gray are genes that with no significant difference with a p-value > 0.05. Highlighted in green are genes with a log2fold of either > 1.5 or < -1.5 and a p-value of > 0.05. Highlighted in blue are genes with a log2fold change between -1.5 and 1.5. They also have a p-value < 0.05. Highlighted in red are genes with a log2fold change of either > 1.5 or < -1.5. However, these genes also have a p-value < 0.05.

A


cdca7a MZ
cdca7a -/-
57
(7.2%)
28
(3.5%)
43
(5.4%)
7
(0.9%)
14
(1.8%)
25
(3.2%)
619
(78.1%)
zbtb24 -/-

B

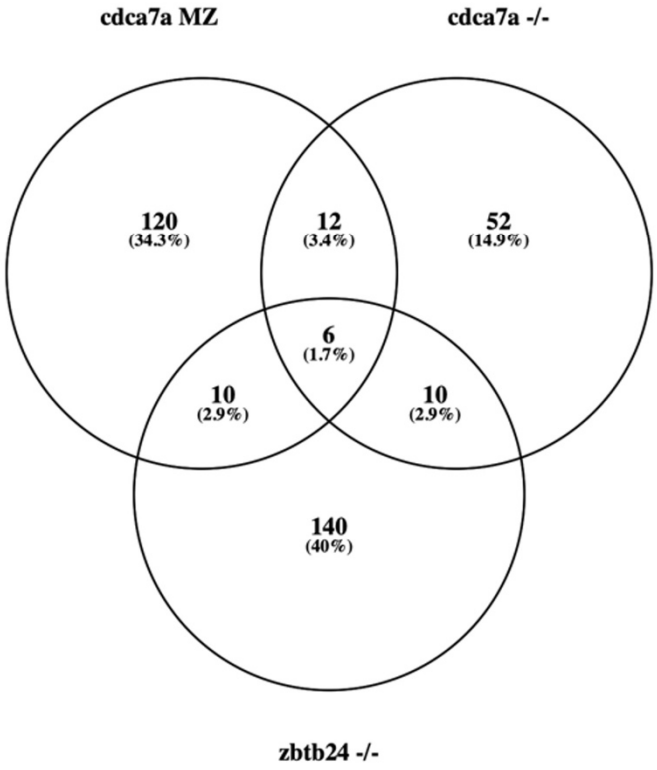

cdca7a MZ
cdca7a -/-
120
(34.3%)
12
(3.4%)
52
(14.9%)
6
(1.7%)
10
(2.9%)
10
(2.9%)
140
(40%)
zbtb24 -/-

Figure 19: Intersection of dysregulated genes
(A) Venn diagram showing the intersection of genes that are increased among *cdca7a* maternal zygotic, *cdca7a* zygotic, and *zbtb24* zygotic mutants. (B) Venn diagram showing the intersection of genes that are decreased among *cdca7a* maternal zygotic, *cdca7a* zygotic, and *zbtb24* zygotic mutants. Both Venn diagrams were made using Venny 2.1 (Oliveros 2015).

Shared increasing genes between all three mutants	**Shared increasing genes between *cdca7a* maternal zygotic and zygotic mutants**	**Shared increasing genes between *cdca7a* maternal zygotic and *zbtb24* mutants**	**Shared increasing genes between *cdca7a* zygotic and *zbtb24* mutants**
zgc:174862	*znf1179*	*tcap*	*agt*
ftr52p	*si:dkey-19a16.7*	*CCDC134*	*si:ch211-147m6.1*
tec	*zgc:171497*	*CR318588.3*	*si:dkey-239b22.1*
si:ch211-103n10.5	*hbbe1.2*	*BX465228.2*	*si:ch211-284o19.8*
si:ch211-207n23.2	*si:dkey-239j18.3*	*mhc1uma*	*mpeg1.2*
UBLCP1	*fgfr1bl*	*BX465228.1*	*slc2a6*
ccl34a.4	*comtd1*	*si:ch211-266g18.9*	*acod1*
	si:ch211-197g15.6	*CABZ01075611.1*	*zgc:153932*
	si:rp71-36a1.3	*CABZ01068366.1*	*ctsl.1*
	AL928685.2	*socs3a*	*si:dkey-102g19.3*
	BX005417.1	*si:dkey-29p23.1*	*si:ch73-181m17.1*
	sccpdha.2	*micall1b.2*	*ccl19a.2*
	si:dkey-15h8.17	*sfxn5a*	*si:ch211-209j10.6*
	cyp2aa11	*mhc1zfa*	*SLC16A6*
	ctslb		*hamp*
	cd28l		*mhc1uba*
	CR388157.1		*cxcl18a.1*
	he1.2		*asb15b*
	duox		*aqp10a*
	zgc:174855		*plekhs1*
	irg1l		*pcdh1g11*
	zgc:174275		*FO904966.1*
	he1.1		*si:ch211-194e1.7*
	BX322603.1		*si:ch211-226h7.3*
	zgc:174153		*BX322618.1*
	tgm1l4		
	si:ch211-153b23.3		
	si:dkey-26g8.5		

Table 3: List of genes from intersection that are increased among the mutants

Shared decreasing genes between all three mutants	**Shared decreasing genes between *cdca7a* maternal zygotic and zygotic mutants**	**Shared decreasing genes between *cdca7a* maternal zygotic and *zbtb24* mutants**	**Shared decreasing genes between *cdca7a* zygotic and *zbtb24* mutants**
zgc:162509 *gck* *CABZ01063602.1* *ptk2ba* *hce2l1* *rfesd*	*si:ch211-243p7.3* *kbtbd7* *zgc:194221* *CABZ01044746.1* *n4bp3* *AL928650.3* *msh4* *CU972512.1* *FO904898.4* *npm2a* *acot15* *cdh29*	*CCDC39* *si:dkey-9c18.3* *CR847936.6* *adgre9* *CR788249.1* *selenol* *si:dkey-178j11.5* *BX927374.1* *CABZ01070527.1* *cept1a*	*slc16a6a* *cry1a* *per2* *apoa4a* *si:ch211-236p5.3* *CABZ01065076.1* *hbae5* *CABZ01019904.1* *si:dkey-147f3.4* *CABZ01038699.1*

Table 4: List of genes from intersection that are decreased among the mutants

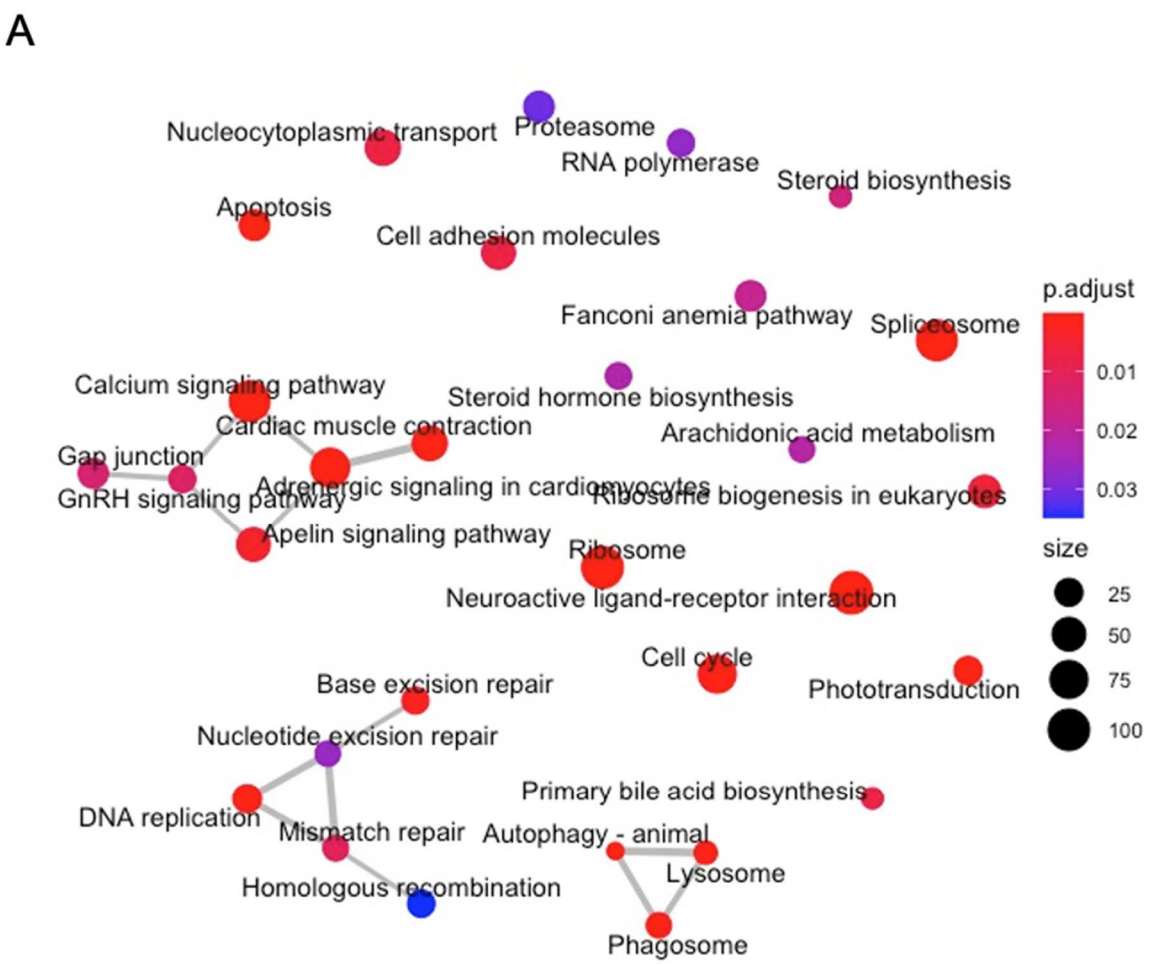


Figure 20: KEGG analysis in *cdca7a*^{MZ}

(A) Connectivity map showing the clustering from KEGG analysis. The color signifies the adjusted p-value. The size of the dot relates to how many genes are in the term.

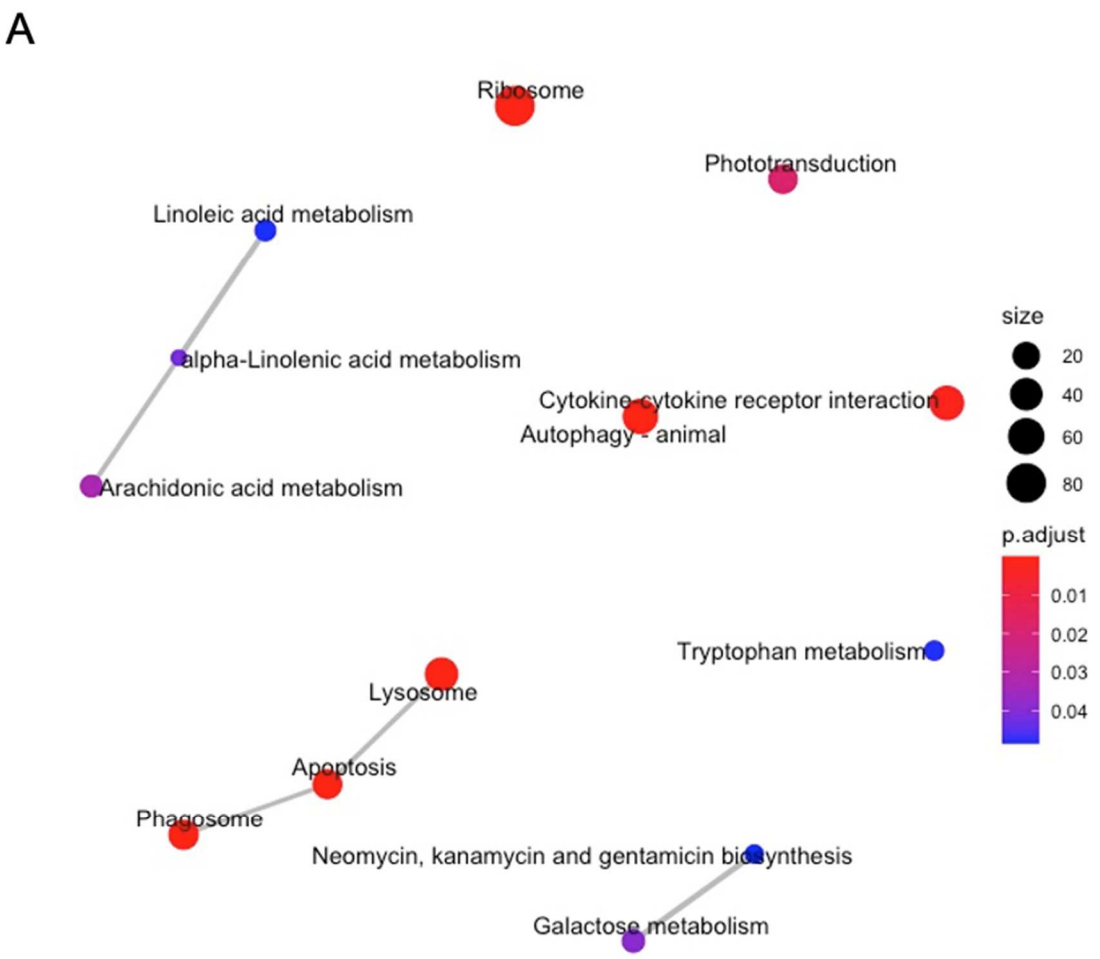


Figure 21: KEGG analysis in *cdca7a*$^{-/-}$

(A) Connectivity map showing the clustering from KEGG analysis. The color signifies the adjusted p-value. The size of the dot relates to how many genes are in the term.

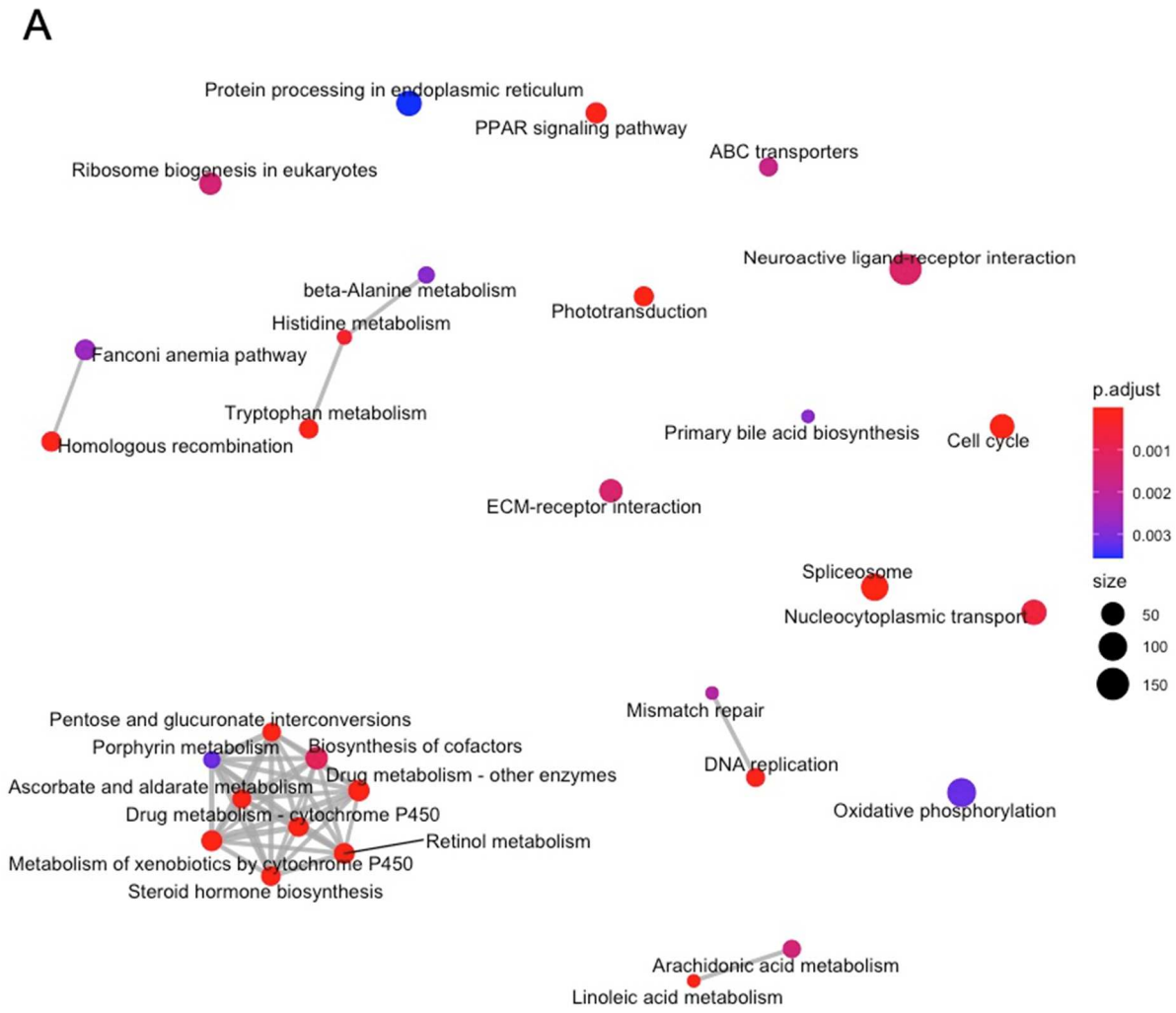


Figure 22: KEGG analysis in *zbtb24*$^{-/-}$

(A) Connectivity map showing the clustering from KEGG analysis. The color signifies the adjusted p-value. The size of the dot relates to how many genes are in the term.

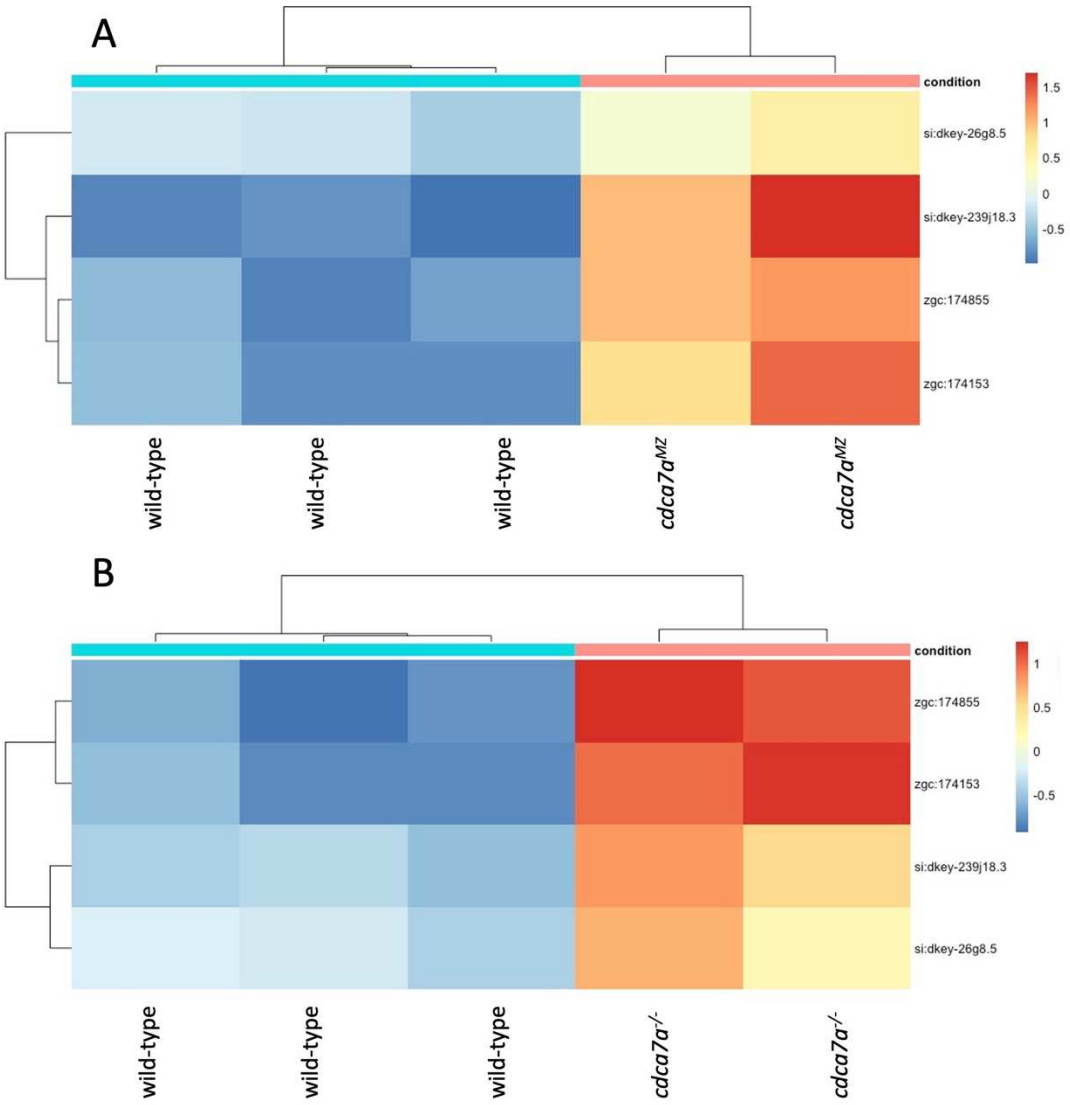


Figure 23: Dysregulation of apoptotic genes in *cdca7a* mutants
(A) Heatmap of dysregulated apoptotic genes in *cdca7a* maternal zygotic mutants. (B) Heatmap of dysregulated apoptotic genes in *cdca7* zygotic mutants.

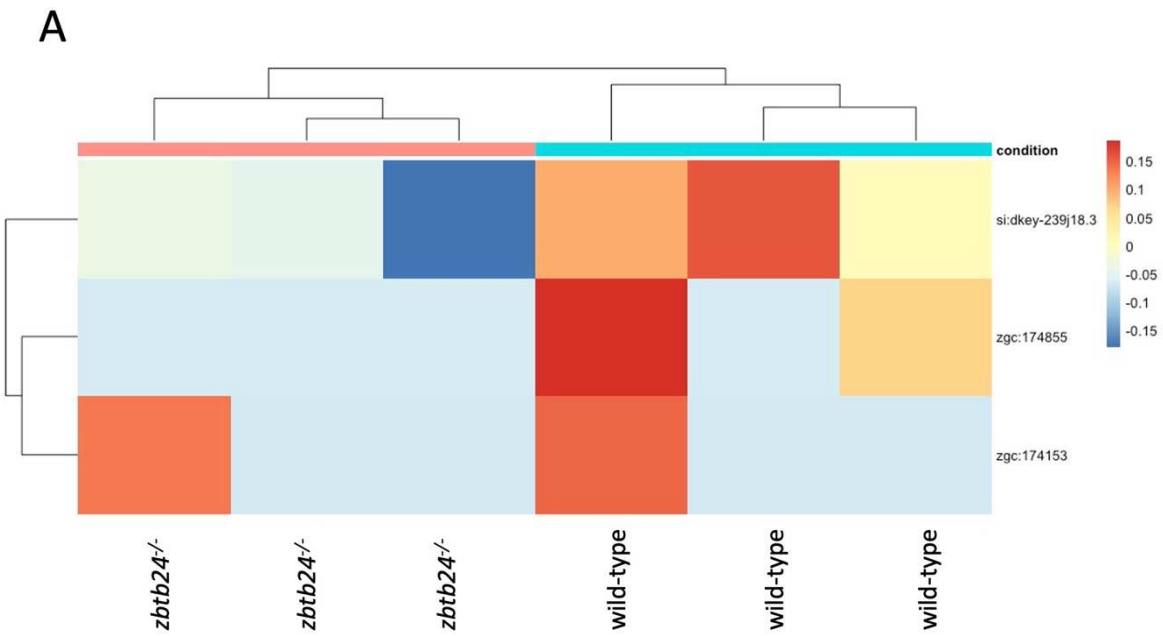


Figure 24: Dysregulation of apoptotic genes in *zbtb24*$^{-/-}$
(A) Heatmap of dysregulated apoptotic genes in *zbtb24* zygotic mutants.

Chapter 5: CHARACERIZING THE IMMUNOLOGICAL PHENOTYPES IN ICF SYNDROME MUTANT ZEBRAFISH

PREFACE

Patients with ICF syndrome are characterized by having severe immunodeficiency. They can exhibit a decrease in immunoglobulins despite having normal levels of lymphoid cells. Although the type of immunoglobulin deficiency is variable. Immunodeficiency in patients can vary from a mild decrease in immune response to agammaglobulinemia. Some patients even have low levels of T cells and/or B cells. Due to the immunodeficiency, patients have recurrent infections that often lead to death at a young age. Typically, patients with ICF syndrome have severe respiratory infections and over half of ICF syndrome patients have recurrent gastrointestinal infections. There have also been reports that patients also have pericarditis, ear infections, septicemia and oral Candida infections (Ehrlich, Jackson, and Weemaes 2006). The cause of the immunoglobulin deficiency in ICF syndrome patients is unknown. To treat the recurrent infections patients are given intravenous infusions of immunoglobulins.

In zebrafish B cell development does not occur until six weeks post fertilization. At six weeks post fertilization in *zbtb24* homozygous mutant zebrafish our lab was shown to have a decrease in *IgM*, *IgD*, and *IgZ* expression. This decrease in immunoglobulins was accompanied by normal lymphoid cell numbers (Rajshekar et al. 2018). Collectively, this data shows that *zbtb24* mutants exhibit ICF-like immunological phenotypes.

In this chapter I aim to discover if *cdca7a* homozygous mutants exhibit a decrease in immunoglobulins. I use two orthogonal methods to analyze immunoglobulin levels. One method

for analyzing immunoglobulins is through the use of qPCR. The second method I used to assess the level of immunoglobulins is through RNA sequencing.

RESULTS

cdca7a mutants do not exhibit a decrease in immunoglobulins via qPCR

To examine the immunoglobulin levels in *cdca7a* homozygous maternal zygotic and zygotic mutants, I first performed quantitative PCR, examining levels of transcripts for the immunoglobulins *IgM*, *IgD* and the zebrafish specific *IgZ* at 6 weeks post fertilization. I found that *cdca7a* maternal zygotic mutants exhibited clearly detectable levels of immunoglobulin transcripts that were comparable to levels detected in wild-type siblings (Fig. 25). Similar immunoglobulin levels were also noted when I compared *cdca7a* zygotic mutants to age matched wild-type controls (Fig. 26).

cdca7a mutants do not exhibit a decrease in immunoglobulins via RNA sequencing

To further explore potential phenotypes related to immunodeficiency in *cdca7a* mutant zebrafish, I performed RNA sequencing on kidney marrow dissected from *cdca7a* zygotic mutant adults. I chose kidney marrow for this analysis because in zebrafish, this is the site of definitive hematopoiesis. In this way, the zebrafish kidney marrow is similar to human bone marrow. At four months post fertilization, I surgically removed the kidney marrow from adult wild-type and *cdca7a* mutant siblings and extracted total RNA. Library preparation was then performed from a protocol adapted from the Illumina TruSeq mRNA Stranded Library Kit. Upon completion of sequencing, I preformed the computational analysis. I obtained on average 24.7 million single end reads and approximately 72% of those reads mapped to the zebrafish genome.

After I aligned the reads to the zebrafish genome and obtained the counts file, I made sure that my two conditions segregated from each other (Fig. 27). In this data set, I found that 74 genes had a log2fold change of less than -1.5 with a significant adjusted p-value. There were 501 genes that had a log2fold change of over 1.5 with a significant adjusted p-value (Fig. 28). After ensuring proper clustering, I found that the RNA sequencing confirmed the results obtained from qPCR. I found no statistical difference in the immunoglobulin genes between wild-type and *cdca7a* zygotic mutants.

To further assess potential immune defects, I next assessed transcription of other genes that are known to be important for immune function. While most genes exhibited expression levels that were comparable to wild-type, *cdca7a* zygotic mutants did show significant decreases in several genes important for immune response including *cish*, *prf1.8*, *nkl.1*, and *ccl34a.4* (Fig. 29). This decrease in genes related to the immune response could potentially be similar to the decreases in immune response genes seen in some ICF syndrome patients. Further analysis will need to be done to determine whether this decrease in immune response genes would lead to a lessened response to infections.

DISSCUSSION

Unlike what is seen in *zbtb24* mutants, my findings suggest that *cdca7a* maternal zygotic and zygotic mutants do not exhibit a decrease in expression of immunoglobulin genes. However, *cdca7a* zygotic mutants do have a decrease in several genes related to the immune response.

Cytokine Inducible SH2 Containing Protein (CISH) has been shown to be involved in regulating immune cell development and function. *Cish* homozygous mutant mice exhibit increased sensitivity to IL-15 in NK cells and increases IL-4 and IL-2 signaling in $CD4^+$ T cells

(Delconte et al. 2016; Putz et al. 2017). It was found that people with *CISH* variants are more susceptible to infectious diseases (Khor et al. 2010; Sun et al. 2014).

prf1.8 is the zebrafish ortholog of PRF1. PRF1has been shown to be an effector molecule for both T cell and NK cell mediated cytolysis. Also, mutations in PRF1 in people has been show to lead to Hemophagocytic Lymphohistiocytosis (HLH). HLH is an autosomal recessive disorder caused by mutations in several genes involved in NK and T cell granule mediated cytotoxic function (Bordbar et al. 2017).

Both *nkl.1* and *ccl34a.4* do not have an associated human ortholog. Both of these genes are predicted to be involved in the immune response in zebrafish.

ICF syndrome immunological defects exist on a spectrum. Although most patients exhibit a decrease in immunoglobulins not all patients exhibit this phenotype. Some ICF syndrome patients only exhibit a decrease in immune response leading to them getting recurrent infections. This zebrafish model represents a new ICF syndrome model in which the decrease in immunoglobulin genes is not present. This new model seems to recapitulate a mild case of ICF syndrome in which the morphological phenotypes are also not seen.

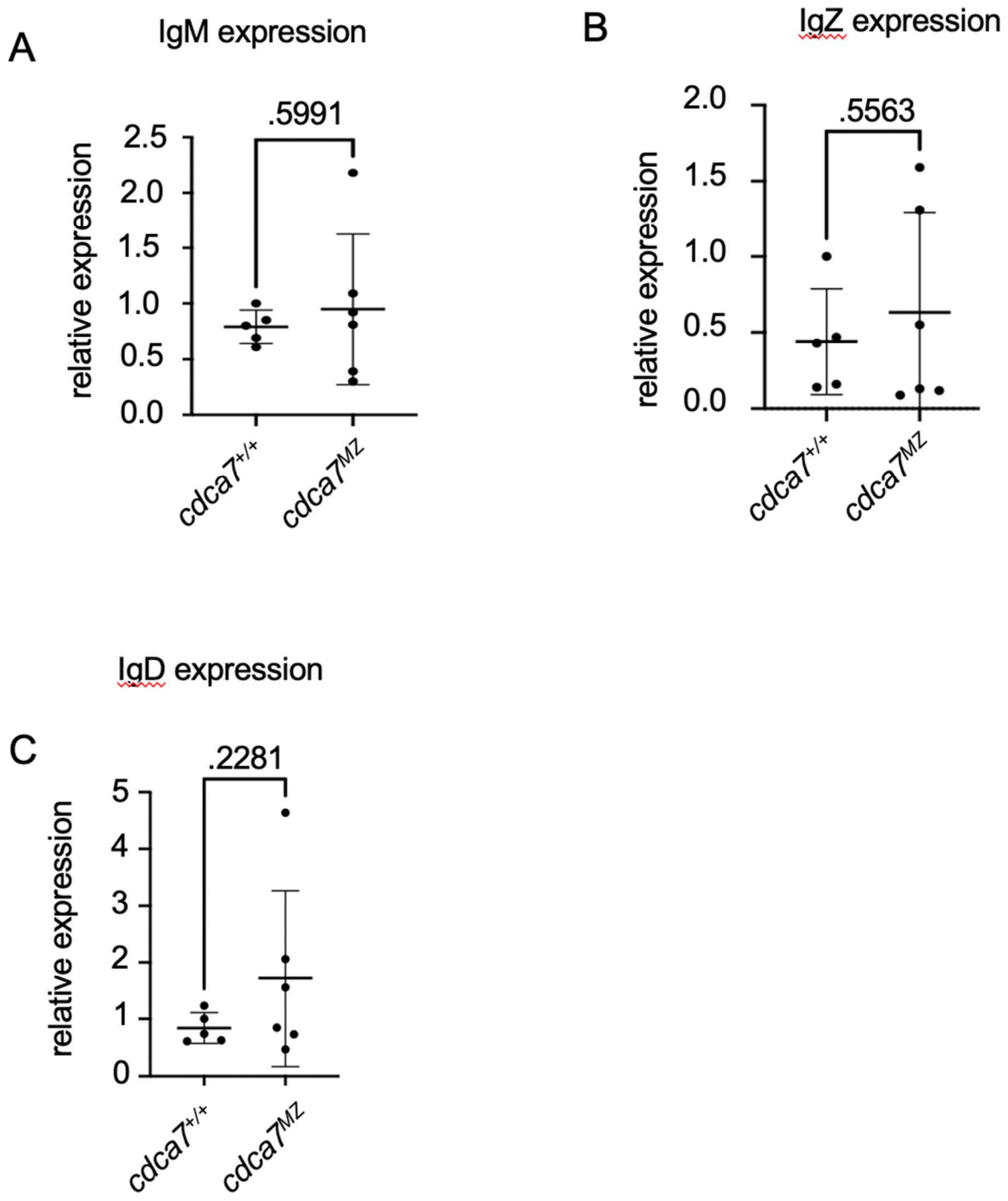


Figure 25: Immunological characterization in *cdca7a*^{MZ}
(A) Relative expression of *IgM* at six weeks post fertilization in *cdca7a* wild-type and maternal zygotic mutant zebrafish (n=5 for each group). (B) Relative expression of *IgZ* at six weeks post fertilization in *cdca7a* wild-type and maternal zygotic mutant zebrafish (n=5 for each group). (C) Relative expression of *IgD* at six weeks post fertilization in *cdca7a* wild-type and maternal zygotic mutant zebrafish (n=5 for each group).

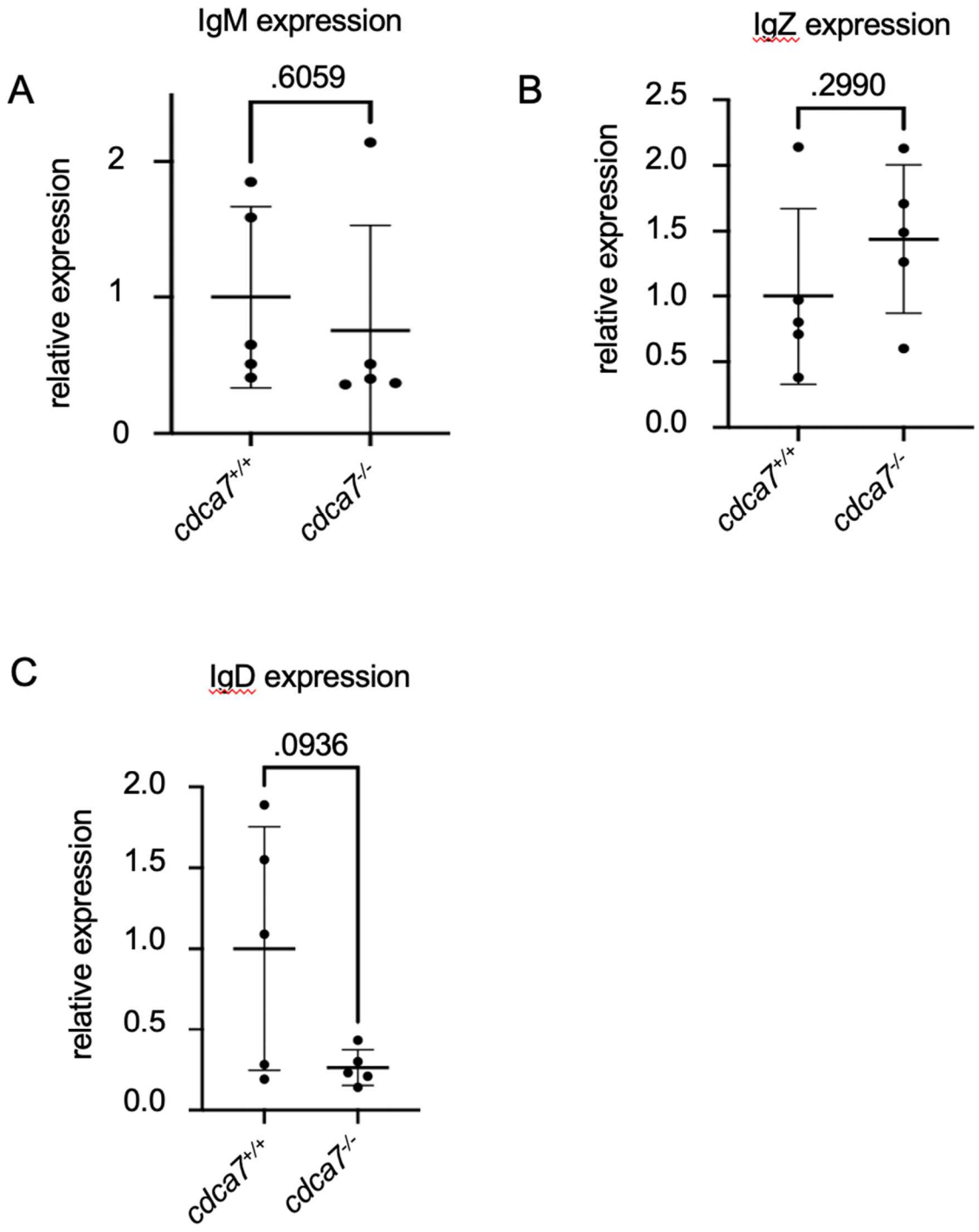


Figure 26: Immunological characterization in *cdca7a*$^{-/-}$
(A) Relative expression of *IgM* at six weeks post fertilization in *cdca7a* wild-type and mutant zebrafish (n=5 for each group). (B) Relative expression of *IgZ* at six weeks post fertilization in *cdca7a* wild-type and mutant zebrafish (n=5 for each group). (C) Relative expression of *IgD* at six weeks post fertilization in *cdca7a* wild-type and mutant zebrafish (n=5 for each group).

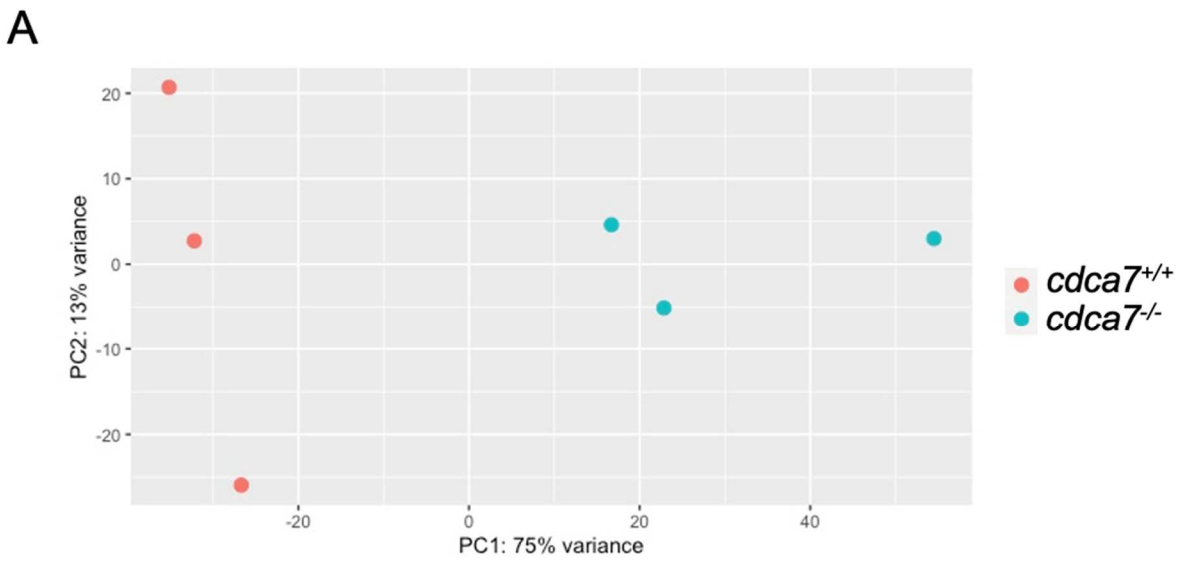


Figure 27: *cdca7a*$^{-/-}$ kidney RNA sequencing clustering
(A) Principal components analysis of *cdca7a* wild-type and zygotic mutants. Plotted on the x-axis is PC1 with 75% of the variance and on the y-axis is PC2 with 13% of the variance.

A

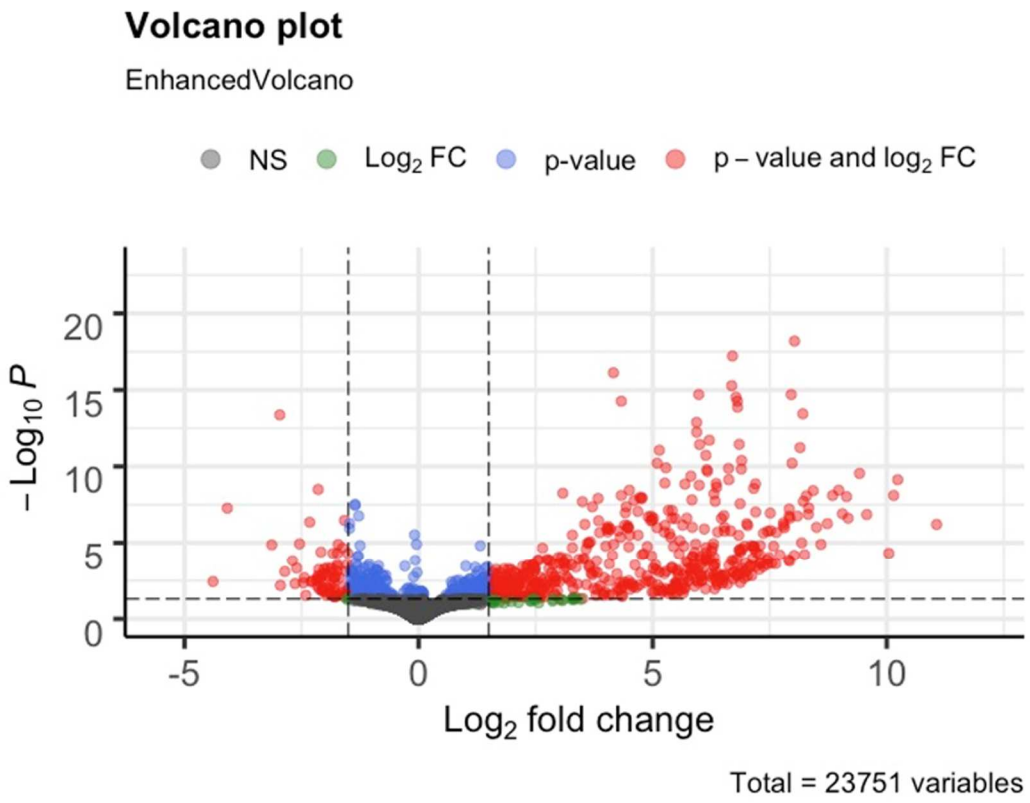


Figure 28: Differentially expressed genes in zebrafish kidney marrow
(A) Volcano plot of differentially expressed genes. Highlighted in gray are genes that with no significant difference with a p-value > 0.05. Highlighted in green are genes with a log2fold of either > 1.5 or < -1.5 and a p-value of > 0.05. Highlighted in blue are genes with a log2fold change between -1.5 and 1.5. They also have a p-value < 0.05. Highlighted in red are genes with a log2fold change of either > 1.5 or < -1.5. However, these genes also have a p-value < 0.05.

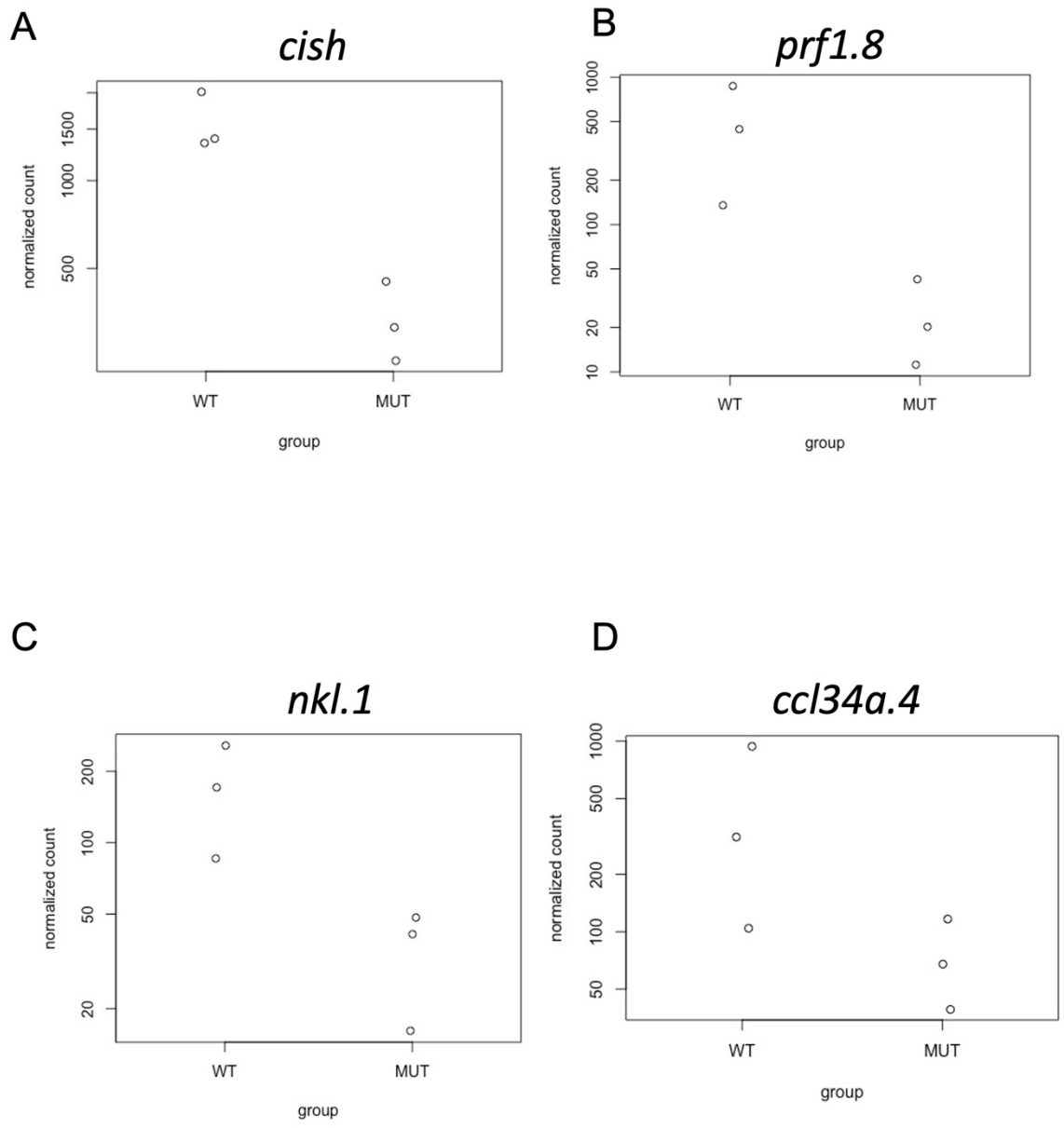


Figure 29: Downregulated immune genes in *cdca7a*$^{-/-}$
(A) Schematic of *cish* expression from normalized count data obtained in R using DESeq2. (B) Schematic of *prf1.8* expression from normalized count data obtained in R using DESeq2. (C) Schematic of *nkl.1* expression from normalized count data obtained in R using DESeq2. (D) Schematic of *ccl34a.4* expression from normalized count data obtained in R using DESeq2.

Chapter 6: SUMMARY AND PERSPECTIVES

My research has focused on developing a second zebrafish model for ICF syndrome. I have found that deletion of *cdca7a* does not recapitulate some of the ICF like phenotypes, which were previously noted in a zebrafish *zbtb24* deletion model of ICF syndrome. I have shown that *cdca7a* mutants appear morphologically the same as wild-type counterparts. This was an unexpected finding considering deletion of *zbtb24* leads to a smaller stature, elongated snout, and a reduced lifespan, and Zbtb24 has been previously reported to act upstream of *cdca7a* (Rajshekar et al. 2018). Despite the lack of obvious morphological phenotypes, I found that *cdca7a* maternal zygotic and zygotic mutants have more severe pericentromeric DNA hypomethylation compared to *zbtb24* mutants at early stages of development. I found that *cdca7a* maternal zygotic mutants lose pericentromeric DNA methylation as early as six hours post fertilization. In addition, I showed that *cdca7a* zygotic mutants lose DNA methylation at the pericentromere well before *zbtb24* mutants. These findings suggest that pericentromeric DNA hypomethylation does not lead to the ICF like phenotypes observed in *zbtb24* mutants, and that Zbtb24 may have additional functions beyond regulation of pericentromeric methylation.

During the course of methylation analysis, I also developed a new way of analyzing DNA methylation at the pericentromere that is quantitative and could possibly be used for other repeat analysis in conjunction with Southern Blots. This quantitative approach to DNA methylation analysis allows for an extensive look into repetitive regions of the genome that are difficult to

map to. In addition, this analysis makes the way for a more robust analysis of multiple samples at one time.

These findings also show that *cdca7* is important for DNA methylation at the pericentromere. Increased genomic digestion of DNA shows that as the fish age there is a progressive loss of DNA methylation at the pericentromere. This suggests that *cdca7* is involved in the maintenance of DNA methylation at the pericentromere. This can be explored further through WGBS sequencing to get a better understanding of the percent of methylation loss as the fish age.

I sought after to perform RNA sequencing to gain a better understanding of the early consequences of DNA pericentromeric hypomethylation. In doing so I found that *cdca7a* maternal zygotic and zygotic mutants exhibit an increase in apoptotic genes at 3 days post fertilization. This increase in apoptotic genes at 3 days post fertilization is not seen in *zbtb24* mutants at three days post fertilization. This finding suggests that a decrease in DNA methylation at the pericentromere may lead to an increase in apoptosis as an early consequence. This could be followed up with by using TUNEL staining to show an increase in apoptosis.

In addition to characterizing the morphological phenotypes, I sought after to explore the causality behind the immunodeficiency seen in ICF syndrome patients. To do this I first examined whether there was a decrease in immunoglobulins in the *cdca7a* maternal zygotic and zygotic mutants. I found that there were no statistical differences between the mutants and wild-type counterparts. Next, I performed RNA sequencing on kidney marrow from adult zebrafish. I confirmed that there were no changes in immunoglobulins. However, there was a decrease in several genes associated with the immune system. This is an avenue for future exploration to

determine if these genes play a role in the immunodeficiency seen in some ICF syndrome patients and if this decrease could lead to recurrent infections.

Materials and methods

Zebrafish husbandry

The care of the zebrafish was in accordance with animal care and use guidelines with approval by the Institutional Animal Care and Use Committees at the University of Georgia. Zebrafish embryos were raised under standard conditions at 28°C and transferred to animal facility located at the University of Georgia URAR Life Sciences Facility at 5 days post fertilization. Wild-type lines are of the AB background.

DNA extraction

Tissue for adult DNA extraction was obtained from adult zebrafish caudal fin clips. DNA extraction was performed via one of two ways: (1) with a quick prep protocol using 50mM NaOH and TrisHCl (pH 8.0) or (2) using DNA lysis buffer and phenol chloroform. DNA extraction for larval zebrafish was obtained via DNA lysis buffer and phenol choloroform.

CRISPR mutagenesis

CHOPCHOP was used to determine target selection for CRISPR/Cas9 mediated mutagenesis. sgRNA templates were generated by using the oligo-based approach (Kelley et al. 2016). Invitro transcription of the sgRNAs was performed using T7 RNA polymerase (Promega) as per manufacturer protocol. Cas9 was invitro transcribed from the pT3TS-nls-zCas9-nls plasmid using the T3 mMessage mMachine Kit (Ambion). Approximately, 200-400ng of sgRNA and 500ng of Cas9 mRNA was 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.

Analysis of DNA methylation at sat1

DNA methylation was analyzed using Southern Blots and bisulfite sequencing. For Southern Blot analysis, 1 microgram of total genomic DNA was digested with methylation sensitive restriction enzyme (Hypch4iv from neb) fractionated by electrophoresis through a 0.9% agarose gel. DNA was then transferred to a nylon membrane. Probes were PCR amplified using primers designed to Sat1. Probes are biotinylated and labeled following the North2South Biotin Random Prime DNA labeling Kit (ThermoFisher). Hybridization signals were detected following manufacturer protocol from North2South Chemiluminescent Hybridization and Detection Kit (ThermoFisher).

For bisulfite sequencing analysis 1 microgram of total genomic DNA is used in library preparation. MethylC-seq library preparation is performed (Urich et al. 2015). The bisulfite conversion is performed using EZ DNA Methylation – Lightning Kit according to manufacturer protocol (Zymo). Upon sequencing completion, the zebrafish reference sequence was downloaded from UCSC genome browser, and I used RepeatMasker (v4.0.7-foss-2016b) to mask the genome for the zebrafish sat1 repeat. I then added the sat1 repeat as an extra chromosome to the genome. Reads were trimmed with Trim_Galore (v0.4.5-foss-2016b) to remove Illumina adapters with default parameters. Quality checks for the reads were completed using FastQC and MultiQC. Bismark (v0.22.1-foss-2016b) was used to align reads with one mismatch allowed. Bismark was also used to extract methylation analysis from the aligned reads.

RNA expression analysis

For qRT-PCR, total RNA was extracted using Trizol and isopropanol was used to precipitate out the RNA. Total RNA was cleaned with RNA clean and concentrator kit (Zymo) according to

manufacturer protocol. cDNA was generated by following GoScript Reverse Transcriptase Kit (Promega) according to manufacturer protocol. Real Time PCR was performed using QuantStudio 3 (ThermoFisher). Analysis was performed using the $2^{-\Delta\Delta Ct}$ method, with relative expression levels normalized to β-actin.

Transcriptome sequencing

After quantification of total RNA, up to 1.3ug of total RNA underwent polyA selection and TruSeq library preparation according to instructions provided by Illumina (TruSeq Standard mRNA LT Kit) with 13 cycles of PCR. Using the TruSeq SBS v4 Kit (Illumina) samples were run on a HiSeq 2500 High Output in a 50bp single end run. An average of 15.7 million reads was generated per sample.

Genome alignment and read mapping and expression analysis

The Linux based GACRC computing cluster at the University of Georgia was used for bioinformatic analysis. along with the R programming environment. The reference sequence and genome annotation was obtained from ENSEMBL (GRCz11). FASTQC and MultiQC was used for quality checks. Trim_Galore (v0.6.5-GCCcore-8.3.0-Java-11-Python-3.7.4) was used to trim RNA sequencing adapters using default parameters. Alignment of the reads to the genome was performed by STAR (v2.7.3a-GCC-8.3.0) with the option to output one output alignment for multimappers. FeatureCounts (Subread v2.0.1-GCC-8.3.0) was used to generate a counts text file from the STAR output using a provided GTF file downloaded from ENSEMBL. DESeq2 in the R programming environment was used to perform the differential expression analysis. Pipeline was adapted from Love et. al, 2014.

Phylogenetic tree

The phylogenetic tree was completed using Geneious Prime 2022.1.1 default parameters were used when supplied with protein sequences from UCSC genome browser.

Statistical Analysis

The student unpaired 2-tailed t-test, Welch's t test, or Wald's t test was performed using GraphPad PRISM software depending on sample number, data distributions and type. Significant P values are described to be at 0.05. Statistical analysis was carried out using the R programming environment or GraphPad Prism.

References

Allshire, R. C., and H. D. Madhani. 2018. "Ten principles of heterochromatin formation and function." *Nat Rev Mol Cell Biol* 19 (4):229-244. doi: 10.1038/nrm.2017.119.

Anderson, R. M., J. A. Bosch, M. G. Goll, D. Hesselson, P. D. Dong, D. Shin, N. C. Chi, C. H. Shin, A. Schlegel, M. Halpern, and D. Y. Stainier. 2009. "Loss of Dnmt1 catalytic activity reveals multiple roles for DNA methylation during pancreas development and regeneration." *Dev Biol* 334 (1):213-23. doi: 10.1016/j.ydbio.2009.07.017.

Barau, J., A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Herault, F. Guillou, and D. Bourc'his. 2016. "The DNA methyltransferase DNMT3C protects male germ cells from transposon activity." *Science* 354 (6314):909-912. doi: 10.1126/science.aah5143.

Beisel, C., and R. Paro. 2011. "Silencing chromatin: comparing modes and mechanisms." *Nat Rev Genet* 12 (2):123-35. doi: 10.1038/nrg2932.

Berger, J., and P. D. Currie. 2012. "Zebrafish models flex their muscles to shed light on muscular dystrophies." *Dis Model Mech* 5 (6):726-32. doi: 10.1242/dmm.010082.

Bordbar, M. R., F. Modarresi, M. A. Farazi Fard, H. Dastsooz, N. Shakib Azad, and M. A. Faghihi. 2017. "A case report of novel mutation in PRF1 gene, which causes familial autosomal recessive hemophagocytic lymphohistiocytosis." *BMC Med Genet* 18 (1):49. doi: 10.1186/s12881-017-0404-9.

Bostick, M., J. K. Kim, P. O. Esteve, A. Clark, S. Pradhan, and S. E. Jacobsen. 2007. "UHRF1 plays a role in maintaining DNA methylation in mammalian cells." *Science* 317 (5845):1760-4. doi: 10.1126/science.1147939.

Bourc'his, D., G. L. Xu, C. S. Lin, B. Bollman, and T. H. Bestor. 2001. "Dnmt3L and the establishment of maternal genomic imprints." *Science* 294 (5551):2536-9. doi: 10.1126/science.1065848.

Ceol, C. J., Y. Houvras, J. Jane-Valbuena, S. Bilodeau, D. A. Orlando, V. Battisti, L. Fritsch, W. M. Lin, T. J. Hollmann, F. Ferre, C. Bourque, C. J. Burke, L. Turner, A. Uong, L. A. Johnson, R. Beroukhim, C. H. Mermel, M. Loda, S. Ait-Si-Ali, L. A. Garraway, R. A. Young, and L. I. Zon. 2011. "The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset." *Nature* 471 (7339):513-7. doi: 10.1038/nature09806.

Chedin, F., M. R. Lieber, and C. L. Hsieh. 2002. "The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a." *Proc Natl Acad Sci U S A* 99 (26):16916-21. doi: 10.1073/pnas.262443999.

Chen, T., Y. Ueda, J. E. Dodge, Z. Wang, and E. Li. 2003. "Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b." *Mol Cell Biol* 23 (16):5594-605. doi: 10.1128/MCB.23.16.5594-5605.2003.

Corley-Smith, G. E., C. J. Lim, and B. P. Brandhorst. 1996. "Production of androgenetic zebrafish (Danio rerio)." *Genetics* 142 (4):1265-76. doi: 10.1093/genetics/142.4.1265.

da Rocha, S. T., and A. V. Gendrel. 2019. "The influence of DNA methylation on monoallelic expression." *Essays Biochem* 63 (6):663-676. doi: 10.1042/EBC20190034.

de Capoa, A., F. R. Febbo, F. Giovannelli, A. Niveleau, G. Zardo, S. Marenzi, and P. Caiafa. 1999. "Reduced levels of poly(ADP-ribosyl)ation result in chromatin compaction and hypermethylation as shown by cell-by-cell computer-assisted quantitative analysis." *FASEB J* 13 (1):89-93. doi: 10.1096/fasebj.13.1.89.

de Greef, J. C., J. Wang, J. Balog, J. T. den Dunnen, R. R. Frants, K. R. Straasheijm, C. Aytekin, M. van der Burg, L. Duprez, A. Ferster, A. R. Gennery, G. Gimelli, I. Reisli, C. Schuetz, A. Schulz, Dfcm Smeets, Y. Sznajer, C. Wijmenga, M. C. van Eggermond, M.

M. van Ostaijen-Ten Dam, A. C. Lankester, M. J. D. van Tol, P. J. van den Elsen, C. M. Weemaes, and S. M. van der Maarel. 2011. "Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2." *Am J Hum Genet* 88 (6):796-804. doi: 10.1016/j.ajhg.2011.04.018.

Deaton, A. M., and A. Bird. 2011. "CpG islands and the regulation of transcription." *Genes Dev* 25 (10):1010-22. doi: 10.1101/gad.2037511.

Dejardin, J. 2015. "Switching between Epigenetic States at Pericentromeric Heterochromatin." *Trends Genet* 31 (11):661-672. doi: 10.1016/j.tig.2015.09.003.

Delconte, R. B., T. B. Kolesnik, L. F. Dagley, J. Rautela, W. Shi, E. M. Putz, K. Stannard, J. G. Zhang, C. Teh, M. Firth, T. Ushiki, C. E. Andoniou, M. A. Degli-Esposti, P. P. Sharp, C. E. Sanvitale, G. Infusini, N. P. Liau, E. M. Linossi, C. J. Burns, S. Carotta, D. H. Gray, C. Seillet, D. S. Hutchinson, G. T. Belz, A. I. Webb, W. S. Alexander, S. S. Li, A. N. Bullock, J. J. Babon, M. J. Smyth, S. E. Nicholson, and N. D. Huntington. 2016. "CIS is a potent checkpoint in NK cell-mediated tumor immunity." *Nat Immunol* 17 (7):816-24. doi: 10.1038/ni.3470.

Ehrlich, M., M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, and C. Gehrke. 1982. "Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells." *Nucleic Acids Res* 10 (8):2709-21. doi: 10.1093/nar/10.8.2709.

Ehrlich, M., N. E. Hopkins, G. Jiang, J. S. Dome, M. C. Yu, C. B. Woods, G. E. Tomlinson, M. Chintagumpala, M. Champagne, L. Dillerg, D. M. Parham, and J. Sawyer. 2003. "Satellite DNA hypomethylation in karyotyped Wilms tumors." *Cancer Genet Cytogenet* 141 (2):97-105. doi: 10.1016/s0165-4608(02)00668-4.

Ehrlich, M., K. Jackson, and C. Weemaes. 2006. "Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF)." *Orphanet J Rare Dis* 1:2. doi: 10.1186/1750-1172-1-2.

Ekker, M., A. Fritz, and M. Westerfield. 1992. "Identification of two families of satellite-like repetitive DNA sequences from the zebrafish (Brachydanio rerio)." *Genomics* 13 (4):1169-73. doi: 10.1016/0888-7543(92)90033-o.

El-Brolosy, M. A., Z. Kontarakis, A. Rossi, C. Kuenne, S. Gunther, N. Fukuda, K. Kikhi, G. L. M. Boezio, C. M. Takacs, S. L. Lai, R. Fukuda, C. Gerri, A. J. Giraldez, and D. Y. R. Stainier. 2019. "Genetic compensation triggered by mutant mRNA degradation." *Nature* 568 (7751):193-197. doi: 10.1038/s41586-019-1064-z.

Enukashvily, N. I., R. Donev, I. S. Waisertreiger, and O. I. Podgornaya. 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 (1):42-54. doi: 10.1159/000106440.

Esteller, M. 2002. "CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future." *Oncogene* 21 (35):5427-40. doi: 10.1038/sj.onc.1205600.

Fanelli, M., S. Caprodossi, L. Ricci-Vitiani, A. Porcellini, F. Tomassoni-Ardori, S. Amatori, F. Andreoni, M. Magnani, R. De Maria, A. Santoni, S. Minucci, and P. G. Pelicci. 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 (3):358-65. doi: 10.1038/sj.onc.1210642.

Feng, S., S. J. Cokus, X. Zhang, P. Y. Chen, M. Bostick, M. G. Goll, J. Hetzel, J. Jain, S. H. Strauss, M. E. Halpern, C. Ukomadu, K. C. Sadler, S. Pradhan, M. Pellegrini, and S. E. Jacobsen. 2010. "Conservation and divergence of methylation patterning in plants and animals." *Proc Natl Acad Sci U S A* 107 (19):8689-94. doi: 10.1073/pnas.1002720107.

Garrido-Ramos, M. A. 2017. "Satellite DNA: An Evolving Topic." *Genes (Basel)* 8 (9). doi: 10.3390/genes8090230.

Gill, R. M., T. V. Gabor, A. L. Couzens, and M. P. Scheid. 2013. "The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation." *Mol Cell Biol* 33 (3):498-513. doi: 10.1128/MCB.00276-12.

Goll, M. G., and T. H. Bestor. 2005. "Eukaryotic cytosine methyltransferases." *Annu Rev Biochem* 74:481-514. doi: 10.1146/annurev.biochem.74.010904.153721.

Goll, M. G., and M. E. Halpern. 2011. "DNA methylation in zebrafish." *Prog Mol Biol Transl Sci* 101:193-218. doi: 10.1016/B978-0-12-387685-0.00005-6.

Gomes, C. P., D. E. Fernandes, F. Casimiro, G. F. da Mata, M. T. Passos, P. Varela, G. Mastroianni-Kirsztajn, and J. B. Pesquero. 2020. "Cathepsin L in COVID-19: From Pharmacological Evidences to Genetics." *Front Cell Infect Microbiol* 10:589505. doi: 10.3389/fcimb.2020.589505.

Goto, Y., R. Hayashi, T. Muramatsu, H. Ogawa, I. Eguchi, Y. Oshida, K. Ohtani, and K. Yoshida. 2006. "JPO1/CDCA7, a novel transcription factor E2F1-induced protein, possesses intrinsic transcriptional regulator activity." *Biochim Biophys Acta* 1759 (1-2):60-8. doi: 10.1016/j.bbaexp.2006.02.004.

Greer, E. L., M. A. Blanco, L. Gu, E. Sendinc, J. Liu, D. Aristizabal-Corrales, C. H. Hsu, L. Aravind, C. He, and Y. Shi. 2015. "DNA Methylation on N6-Adenine in C. elegans." *Cell* 161 (4):868-78. doi: 10.1016/j.cell.2015.04.005.

Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M. Caccamo, C. Churcher, C. Scott, J. C. Barrett, R. Koch, G. J. Rauch, S. White, W. Chow, B. Kilian, L. T. Quintais, J. A. Guerra-Assuncao, Y. Zhou, Y. Gu, J. Yen, J. H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S. F. Maguire, G. K. Laird, D. Lloyd, E. Kenyon, S. Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R. Clark, C. Riddle, D. Elliot, G. Threadgold, G. Harden, D. Ware, S. Begum,

B. Mortimore, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Lloyd, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K. Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthravadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C. Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A. Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S. Martin, P. Wray, A. Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. Cooper, A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley, J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A. Berger, C. M. Dooley, Z. Ersan-Urun, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kirn, J. Konantz, M. Konantz, M. Oberlander, S. Rudolph-Geiger, M. Teucke, C. Lanz, G. Raddatz, K. Osoegawa, B. Zhu, A. Rapp, S. Widaa, C. Langford, F. Yang, S. C. Schuster, N. P. Carter, J. Harrow, Z. Ning, J. Herrero, S. M. Searle, A. Enright, R. Geisler, R. H. Plasterk, C. Lee, M. Westerfield, P. J. de Jong, L. I. Zon, J. H. Postlethwait, C. Nusslein-Volhard, T. J. Hubbard, H. Roest Crollius, J. Rogers, and D. L. Stemple. 2013. "The zebrafish reference genome sequence and its relationship to the human genome." *Nature* 496 (7446):498-503. doi: 10.1038/nature12111.

Jagannathan, M., R. Cummings, and Y. M. Yamashita. 2018. "A conserved function for pericentromeric satellite DNA." *Elife* 7. doi: 10.7554/eLife.34122.

Jenness, C., S. Giunta, M. M. Muller, H. Kimura, T. W. Muir, and H. Funabiki. 2018. "HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome." *Proc Natl Acad Sci U S A* 115 (5):E876-E885. doi: 10.1073/pnas.1717509115.

Jiang, L., J. Zhang, J. J. Wang, L. Wang, L. Zhang, G. Li, X. Yang, X. Ma, X. Sun, J. Cai, J. Zhang, X. Huang, M. Yu, X. Wang, F. Liu, C. I. Wu, C. He, B. Zhang, W. Ci, and J. Liu. 2013. "Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos." *Cell* 153 (4):773-84. doi: 10.1016/j.cell.2013.04.041.

Jin, B., Y. Li, and K. D. Robertson. 2011. "DNA methylation: superior or subordinate in the epigenetic hierarchy?" *Genes Cancer* 2 (6):607-17. doi: 10.1177/1947601910393957.

Jin, Z., and Y. Liu. 2018. "DNA methylation in human diseases." *Genes Dis* 5 (1):1-8. doi: 10.1016/j.gendis.2018.01.002.

Jones, P. A. 2012. "Functions of DNA methylation: islands, start sites, gene bodies and beyond." *Nat Rev Genet* 13 (7):484-92. doi: 10.1038/nrg3230.

Jones, P. A., and P. W. Laird. 1999. "Cancer epigenetics comes of age." *Nat Genet* 21 (2):163-7. doi: 10.1038/5947.

Joseph, A., A. R. Mitchell, and O. J. Miller. 1989. "The organization of the mouse satellite DNA at centromeres." *Exp Cell Res* 183 (2):494-500. doi: 10.1016/0014-4827(89)90408-4.

Kaneda, M., M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, and H. Sasaki. 2004. "Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting." *Nature* 429 (6994):900-3. doi: 10.1038/nature02633.

Kelley, M. L., Z. Strezoska, K. He, A. Vermeulen, and Av Smith. 2016. "Versatility of chemically synthesized guide RNAs for CRISPR-Cas9 genome editing." *J Biotechnol* 233:74-83. doi: 10.1016/j.jbiotec.2016.06.011.

Khor, C. C., F. O. Vannberg, S. J. Chapman, H. Guo, S. H. Wong, A. J. Walley, D. Vukcevic, A. Rautanen, T. C. Mills, K. C. Chang, K. M. Kam, A. C. Crampin, B. Ngwira, C. C. Leung, C. M. Tam, C. Y. Chan, J. J. Sung, W. W. Yew, K. Y. Toh, S. K. Tay, D. Kwiatkowski, C. Lienhardt, T. T. Hien, N. P. Day, N. Peshu, K. Marsh, K. Maitland, J. A. Scott, T. N. Williams, J. A. Berkley, S. Floyd, N. L. Tang, P. E. Fine, D. L. Goh, and A. V. Hill. 2010.

"CISH and susceptibility to infectious diseases." *N Engl J Med* 362 (22):2092-101. doi: 10.1056/NEJMoa0905606.

Koziol, M. J., C. R. Bradshaw, G. E. Allen, A. S. H. Costa, C. Frezza, and J. B. Gurdon. 2016. "Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications." *Nat Struct Mol Biol* 23 (1):24-30. doi: 10.1038/nsmb.3145.

Kurdyukov, S., and M. Bullock. 2016. "DNA Methylation Analysis: Choosing the Right Method." *Biology (Basel)* 5 (1). doi: 10.3390/biology5010003.

Lai, Y. H., G. Audira, S. T. Liang, P. Siregar, M. E. Suryanto, H. C. Lin, O. Villalobos, O. B. Villaflores, E. Hao, K. H. Lim, and C. D. Hsiao. 2020. "Duplicated dnmt3aa and dnmt3ab DNA Methyltransferase Genes Play Essential and Non-Overlapped Functions on Modulating Behavioral Control in Zebrafish." *Genes (Basel)* 11 (11). doi: 10.3390/genes11111322.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S.

Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki, and Consortium International Human Genome Sequencing. 2001. "Initial sequencing and analysis of the human genome." *Nature* 409 (6822):860-921. doi: 10.1038/35057062.

Lehnertz, B., Y. Ueda, A. A. Derijck, U. Braunschweig, L. Perez-Burgos, S. Kubicek, T. Chen, E. Li, T. Jenuwein, and A. H. Peters. 2003. "Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin." *Curr Biol* 13 (14):1192-200. doi: 10.1016/s0960-9822(03)00432-9.

Leonhardt, H., A. W. Page, H. U. Weier, and T. H. Bestor. 1992. "A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei." *Cell* 71 (5):865-73. doi: 10.1016/0092-8674(92)90561-p.

Li, E., C. Beard, and R. Jaenisch. 1993. "Role for DNA methylation in genomic imprinting." *Nature* 366 (6453):362-5. doi: 10.1038/366362a0.

Li, E., T. H. Bestor, and R. Jaenisch. 1992. "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." *Cell* 69 (6):915-26. doi: 10.1016/0092-8674(92)90611-f.

Liang, G., M. F. Chan, Y. Tomigahara, Y. C. Tsai, F. A. Gonzales, E. Li, P. W. Laird, and P. A. Jones. 2002. "Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements." *Mol Cell Biol* 22 (2):480-91. doi: 10.1128/MCB.22.2.480-491.2002.

Liang, J., R. Yan, G. Chen, J. Feng, W. W. Wu, W. Ren, C. Zhu, Y. Zhao, X. M. Gao, and J. Wang. 2016. "Downregulation of ZBTB24 hampers the G0/1- to S-phase cell-cycle transition via upregulating the expression of IRF-4 in human B cells." *Genes Immun* 17 (5):276-82. doi: 10.1038/gene.2016.18.

Liu, J., Y. Zhu, G. Z. Luo, X. Wang, Y. Yue, X. Wang, X. Zong, K. Chen, H. Yin, Y. Fu, D. Han, Y. Wang, D. Chen, and C. He. 2016. "Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig." *Nat Commun* 7:13052. doi: 10.1038/ncomms13052.

Lopez-Flores, I., and M. A. Garrido-Ramos. 2012. "The repetitive DNA content of eukaryotic genomes." *Genome Dyn* 7:1-28. doi: 10.1159/000337118.

Luo, G. Z., M. A. Blanco, E. L. Greer, C. He, and Y. Shi. 2015. "DNA N(6)-methyladenine: a new epigenetic mark in eukaryotes?" *Nat Rev Mol Cell Biol* 16 (12):705-10. doi: 10.1038/nrm4076.

Luo, G. Z., and C. He. 2017. "DNA N(6)-methyladenine in metazoans: functional epigenetic mark or bystander?" *Nat Struct Mol Biol* 24 (6):503-506. doi: 10.1038/nsmb.3412.

McCurley, A. T., and G. V. Callard. 2008. "Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment." *BMC Mol Biol* 9:102. doi: 10.1186/1471-2199-9-102.

Mondo, S. J., R. O. Dannebaum, R. C. Kuo, K. B. Louie, A. J. Bewick, K. LaButti, S. Haridas, A. Kuo, A. Salamov, S. R. Ahrendt, R. Lau, B. P. Bowen, A. Lipzen, W. Sullivan, B. B. Andreopoulos, A. Clum, E. Lindquist, C. Daum, T. R. Northen, G. Kunde-Ramamoorthy, R. J. Schmitz, A. Gryganskyi, D. Culley, J. Magnuson, T. Y. James, M. A. O'Malley, J. E. Stajich, J. W. Spatafora, A. Visel, and I. V. Grigoriev. 2017. "Widespread adenine N6-methylation of active genes in fungi." *Nat Genet* 49 (6):964-968. doi: 10.1038/ng.3859.

Myant, K., and I. Stancheva. 2008. "LSH cooperates with DNA methyltransferases to repress transcription." *Mol Cell Biol* 28 (1):215-26. doi: 10.1128/MCB.01073-07.

Nakagawa, T., Y. Kanai, S. Ushijima, T. Kitamura, T. Kakizoe, and S. Hirohashi. 2005. "DNA hypermethylation on multiple CpG islands associated with increased DNA methyltransferase DNMT1 protein expression during multistage urothelial carcinogenesis." *J Urol* 173 (5):1767-71. doi: 10.1097/01.ju.0000154632.11824.4d.

Narayan, A., W. Ji, X. Y. Zhang, A. Marrogi, J. R. Graff, S. B. Baylin, and M. Ehrlich. 1998. "Hypomethylation of pericentromeric DNA in breast adenocarcinomas." *Int J Cancer* 77 (6):833-8. doi: 10.1002/(sici)1097-0215(19980911)77:6<833::aid-ijc6>3.0.co;2-v.

Nitta, H., M. Unoki, K. Ichiyanagi, T. Kosho, T. Shigemura, H. Takahashi, G. Velasco, C. Francastel, C. Picard, T. Kubota, and H. Sasaki. 2013. "Three novel ZBTB24 mutations

identified in Japanese and Cape Verdean type 2 ICF syndrome patients." *J Hum Genet* 58 (7):455-60. doi: 10.1038/jhg.2013.56.

O'Brown, Z. K., and E. L. Greer. 2016. "N6-Methyladenine: A Conserved and Dynamic DNA Mark." *Adv Exp Med Biol* 945:213-246. doi: 10.1007/978-3-319-43624-1_10.

Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." *Cell* 99 (3):247-57. doi: 10.1016/s0092-8674(00)81656-6.

Page, D. M., V. Wittamer, J. Y. Bertrand, K. L. Lewis, D. N. Pratt, N. Delgado, S. E. Schale, C. McGue, B. H. Jacobsen, A. Doty, Y. Pao, H. Yang, N. C. Chi, B. G. Magor, and D. Traver. 2013. "An evolutionarily conserved program of B-cell development and activation in zebrafish." *Blood* 122 (8):e1-11. doi: 10.1182/blood-2012-12-471029.

Papadopoulos, J. S., and R. Agarwala. 2007. "COBALT: constraint-based alignment tool for multiple protein sequences." *Bioinformatics* 23 (9):1073-9. doi: 10.1093/bioinformatics/btm076.

Phillips, R. B., and K. M. Reed. 2000. "Localization of repetitive DNAs to zebrafish (Danio rerio) chromosomes by fluorescence in situ hybridization (FISH)." *Chromosome Res* 8 (1):27-35. doi: 10.1023/a:1009271017998.

Plohl, M., N. Mestrovic, and B. Mravinac. 2014. "Centromere identity from the DNA point of view." *Chromosoma* 123 (4):313-25. doi: 10.1007/s00412-014-0462-0.

Potok, M. E., D. A. Nix, T. J. Parnell, and B. R. Cairns. 2013. "Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern." *Cell* 153 (4):759-72. doi: 10.1016/j.cell.2013.04.030.

Putz, E. M., C. Guillerey, K. Kos, K. Stannard, K. Miles, R. B. Delconte, K. Takeda, S. E. Nicholson, N. D. Huntington, and M. J. Smyth. 2017. "Targeting cytokine signaling checkpoint CIS activates NK cells to protect from tumor initiation and metastasis." *Oncoimmunology* 6 (2):e1267892. doi: 10.1080/2162402X.2016.1267892.

Qu, G. Z., P. E. Grundy, A. Narayan, and M. Ehrlich. 1999. "Frequent hypomethylation in Wilms tumors of pericentromeric DNA in chromosomes 1 and 16." *Cancer Genet Cytogenet* 109 (1):34-9. doi: 10.1016/s0165-4608(98)00143-5.

Rai, K., L. D. Nadauld, S. Chidester, E. J. Manos, S. R. James, A. R. Karpf, B. R. Cairns, and D. A. Jones. 2006. "Zebra fish Dnmt1 and Suv39h1 regulate organ-specific terminal differentiation during development." *Mol Cell Biol* 26 (19):7077-85. doi: 10.1128/MCB.00312-06.

Rajshekar, S., J. Yao, P. K. Arnold, S. G. Payne, Y. Zhang, T. V. Bowman, R. J. Schmitz, J. R. Edwards, and M. Goll. 2018. "Pericentromeric hypomethylation elicits an interferon response in an animal model of ICF syndrome." *Elife* 7. doi: 10.7554/eLife.39658.

Sestakova, S., C. Salek, and H. Remesova. 2019. "DNA Methylation Validation Methods: a Coherent Review with Practical Comparison." *Biol Proced Online* 21:19. doi: 10.1186/s12575-019-0107-z.

Shen, X., Y. F. Zhao, S. Q. Xu, L. Wang, H. M. Cao, Y. Cao, Y. Zhu, Y. Wang, and Z. Q. Liang. 2019. "Cathepsin L induced PC-12 cell apoptosis via activation of B-Myb and regulation of cell cycle proteins." *Acta Pharmacol Sin* 40 (11):1394-1403. doi: 10.1038/s41401-019-0286-9.

Shukla, S., E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg, and S. Oberdoerffer. 2011. "CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing." *Nature* 479 (7371):74-9. doi: 10.1038/nature10442.

Smith, Z. D., and A. Meissner. 2013. "DNA methylation: roles in mammalian development." *Nat Rev Genet* 14 (3):204-20. doi: 10.1038/nrg3354.

Sun, L., Y. Q. Jin, C. Shen, H. Qi, P. Chu, Q. Q. Yin, J. Q. Li, J. L. Tian, W. W. Jiao, J. Xiao, and A. D. Shen. 2014. "Genetic contribution of CISH promoter polymorphisms to susceptibility to tuberculosis in Chinese children." *PLoS One* 9 (3):e92020. doi: 10.1371/journal.pone.0092020.

Suzuki, M. M., and A. Bird. 2008. "DNA methylation landscapes: provocative insights from epigenomics." *Nat Rev Genet* 9 (6):465-76. doi: 10.1038/nrg2341.

Suzuki, T., M. Fujii, and D. Ayusawa. 2002. "Demethylation of classical satellite 2 and 3 DNA with chromosomal instability in senescent human fibroblasts." *Exp Gerontol* 37 (8-9):1005-14. doi: 10.1016/s0531-5565(02)00061-x.

Tao, Y., S. Xi, J. Shan, A. Maunakea, A. Che, V. Briones, E. Y. Lee, T. Geiman, J. Huang, R. Stephens, R. M. Leighty, K. Zhao, and K. Muegge. 2011. "Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences." *Proc Natl Acad Sci U S A* 108 (14):5626-31. doi: 10.1073/pnas.1017000108.

Tazi, J., and A. Bird. 1990. "Alternative chromatin structure at CpG islands." *Cell* 60 (6):909-20. doi: 10.1016/0092-8674(90)90339-g.

Thijssen, P. E., Y. Ito, G. Grillo, J. Wang, G. Velasco, H. Nitta, M. Unoki, M. Yoshihara, M. Suyama, Y. Sun, R. J. Lemmers, J. C. de Greef, A. Gennery, P. Picco, B. Kloeckener-Gruissem, T. Gungor, I. Reisli, C. Picard, K. Kebaili, B. Roquelaure, T. Iwai, I. Kondo, T. Kubota, M. M. van Ostaijen-Ten Dam, M. J. van Tol, C. Weemaes, C. Francastel, S. M. van der Maarel, and H. Sasaki. 2015. "Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome." *Nat Commun* 6:7870. doi: 10.1038/ncomms8870.

Thompson, J. J., R. Kaur, C. P. Sosa, J. H. Lee, K. Kashiwagi, D. Zhou, and K. D. Robertson. 2018. "ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation." *Nucleic Acids Res* 46 (19):10034-10051. doi: 10.1093/nar/gky682.

Tsien, F., E. S. Fiala, B. Youn, T. I. Long, P. W. Laird, K. Weissbecker, and M. Ehrlich. 2002. "Prolonged culture of normal chorionic villus cells yields ICF syndrome-like chromatin decondensation and rearrangements." *Cytogenet Genome Res* 98 (1):13-21. doi: 10.1159/000068543.

Tsuda, H., T. Takarabe, Y. Kanai, T. Fukutomi, and S. Hirohashi. 2002. "Correlation of DNA hypomethylation at pericentromeric heterochromatin regions of chromosomes 16 and 1 with histological features and chromosomal abnormalities of human breast carcinomas." *Am J Pathol* 161 (3):859-66. doi: 10.1016/S0002-9440(10)64246-0.

Tyler-Smith, C., and W. R. Brown. 1987. "Structure of the major block of alphoid satellite DNA on the human Y chromosome." *J Mol Biol* 195 (3):457-70. doi: 10.1016/0022-2836(87)90175-6.

Ueda, Y., M. Okano, C. Williams, T. Chen, K. Georgopoulos, and E. Li. 2006. "Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome." *Development* 133 (6):1183-92. doi: 10.1242/dev.02293.

Ugarkovic, D., A. Sermek, S. Ljubic, and I. Feliciello. 2022. "Satellite DNAs in Health and Disease." *Genes (Basel)* 13 (7). doi: 10.3390/genes13071154.

Unoki, M., H. Funabiki, G. Velasco, C. Francastel, and H. Sasaki. 2019. "CDCA7 and HELLS mutations undermine nonhomologous end joining in centromeric instability syndrome." *J Clin Invest* 129 (1):78-92. doi: 10.1172/JCI99751.

Urich, M. A., J. R. Nery, R. Lister, R. J. Schmitz, and J. R. Ecker. 2015. "MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing." *Nat Protoc* 10 (3):475-83. doi: 10.1038/nprot.2014.114.

van den Boogaard, M. L., P. E. Thijssen, C. Aytekin, F. Licciardi, A. A. Kiykim, L. Spossito, Vash Dalm, G. J. Driessen, R. Kersseboom, F. de Vries, M. M. van Ostaijen-Ten Dam, A. Ikinciogullari, F. Dogu, M. Oleastro, E. Bailardo, L. Daxinger, E. Nain, S. Baris, M. J. D. van Tol, C. Weemaes, and S. M. van der Maarel. 2017. "Expanding the mutation spectrum in ICF syndrome: Evidence for a gender bias in ICF2." *Clin Genet* 92 (4):380-387. doi: 10.1111/cge.12979.

Vourc'h, C., and G. Biamonti. 2011. "Transcription of Satellite DNAs in Mammals." *Prog Mol Subcell Biol* 51:95-118. doi: 10.1007/978-3-642-16502-3_5.

Wang, K. H., J. Kupa, K. A. Duffy, and J. M. Kalish. 2019. "Diagnosis and Management of Beckwith-Wiedemann Syndrome." *Front Pediatr* 7:562. doi: 10.3389/fped.2019.00562.

White, R. J., J. E. Collins, I. M. Sealy, N. Wali, C. M. Dooley, Z. Digby, D. L. Stemple, D. N. Murphy, K. Billis, T. Hourlier, A. Fullgrabe, M. P. Davis, A. J. Enright, and E. M. Busch-Nentwich. 2017. "A high-resolution mRNA expression time course of embryonic development in zebrafish." *Elife* 6. doi: 10.7554/eLife.30860.

Williams, C. A., D. J. Driscoll, and A. I. Dagli. 2010. "Clinical and genetic aspects of Angelman syndrome." *Genet Med* 12 (7):385-95. doi: 10.1097/GIM.0b013e3181def138.

Wu, H., P. E. Thijssen, E. de Klerk, K. K. Vonk, J. Wang, B. den Hamer, C. Aytekin, S. M. van der Maarel, and L. Daxinger. 2016. "Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals." *Hum Mol Genet* 25 (18):4041-4051. doi: 10.1093/hmg/ddw243.

Xiao, C. L., S. Zhu, M. He, Chen, Q. Zhang, Y. Chen, G. Yu, J. Liu, S. Q. Xie, F. Luo, Z. Liang, D. P. Wang, X. C. Bo, X. F. Gu, K. Wang, and G. R. Yan. 2018. "N(6)-Methyladenine DNA Modification in the Human Genome." *Mol Cell* 71 (2):306-318 e7. doi: 10.1016/j.molcel.2018.06.015.

Xu, G. L., T. H. Bestor, D. Bourc'his, C. L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russo, and E. Viegas-Pequignot. 1999. "Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene." *Nature* 402 (6758):187-91. doi: 10.1038/46052.

Zardo, G., and P. Caiafa. 1998. "The unmethylated state of CpG islands in mouse fibroblasts depends on the poly(ADP-ribosyl)ation process." *J Biol Chem* 273 (26):16517-20. doi: 10.1074/jbc.273.26.16517.

Zardo, G., M. D'Erme, A. Reale, R. Strom, M. Perilli, and P. Caiafa. 1997. "Does poly(ADP-ribosyl)ation regulate the DNA methylation pattern?" *Biochemistry* 36 (26):7937-43. doi: 10.1021/bi970241s.

Zhang, G., H. Huang, D. Liu, Y. Cheng, X. Liu, W. Zhang, R. Yin, D. Zhang, P. Zhang, J. Liu, C. Li, B. Liu, Y. Luo, Y. Zhu, N. Zhang, S. He, C. He, H. Wang, and D. Chen. 2015. "N6-methyladenine DNA modification in Drosophila." *Cell* 161 (4):893-906. doi: 10.1016/j.cell.2015.04.018.

Zhu, Q., N. Hoong, A. Aslanian, T. Hara, C. Benner, S. Heinz, K. H. Miga, E. Ke, S. Verma, J. Soroczynski, J. R. Yates, 3rd, T. Hunter, and I. M. Verma. 2018. "Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer." *Mol Cell* 70 (5):842-853 e7. doi: 10.1016/j.molcel.2018.04.023.

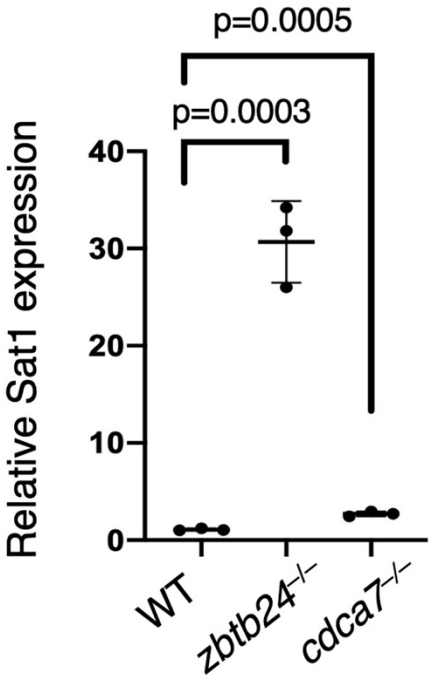
Zimmerman, A. M., F. M. Moustafa, K. E. Romanowski, and L. A. Steiner. 2011. "Zebrafish immunoglobulin IgD: unusual exon usage and quantitative expression profiles with IgM and IgZ/T heavy chain isotypes." *Mol Immunol* 48 (15-16):2220-3. doi: 10.1016/j.molimm.2011.06.441.

APPENDIX

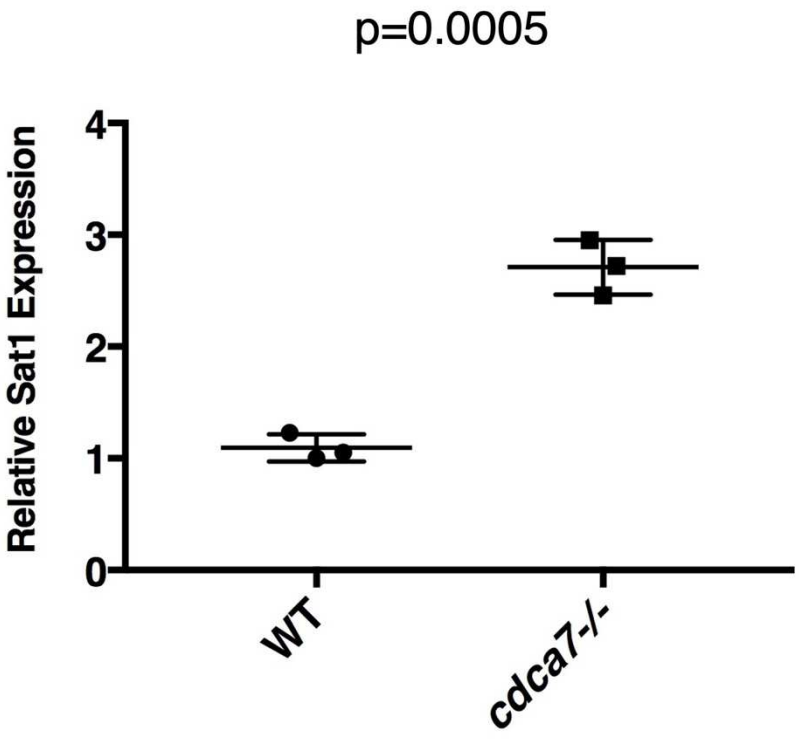
APPENDIX 1: RELATIVE SATELLITE 1 EXPRESSION IN ICF MODELS

Our laboratory has previously shown that a loss of pericentromeric DNA methylation leads to an upregulation of satellite transcripts in *zbtb24* zygotic mutants (Rajshekar et al. 2018). I set out to determine if that same satellite transcript upregulation was seen in *cdca7a* zygotic mutants. I performed qPCR on the satellite transcripts and determined that *cdca7a* zygotic mutants also exhibited an upregulation of satellite transcripts. However, this upregulation in satellite transcripts was not as severe as that seen in *zbtb24* mutants. This could be due to *zbtb24* being a transcription factor and having independent roles in DNA methylation and transcription.

A


p=0.0005
p=0.0003
Relative Sat1 expression
40
30
20
10
0
WT
zbtb24-/-
cdca7-/-

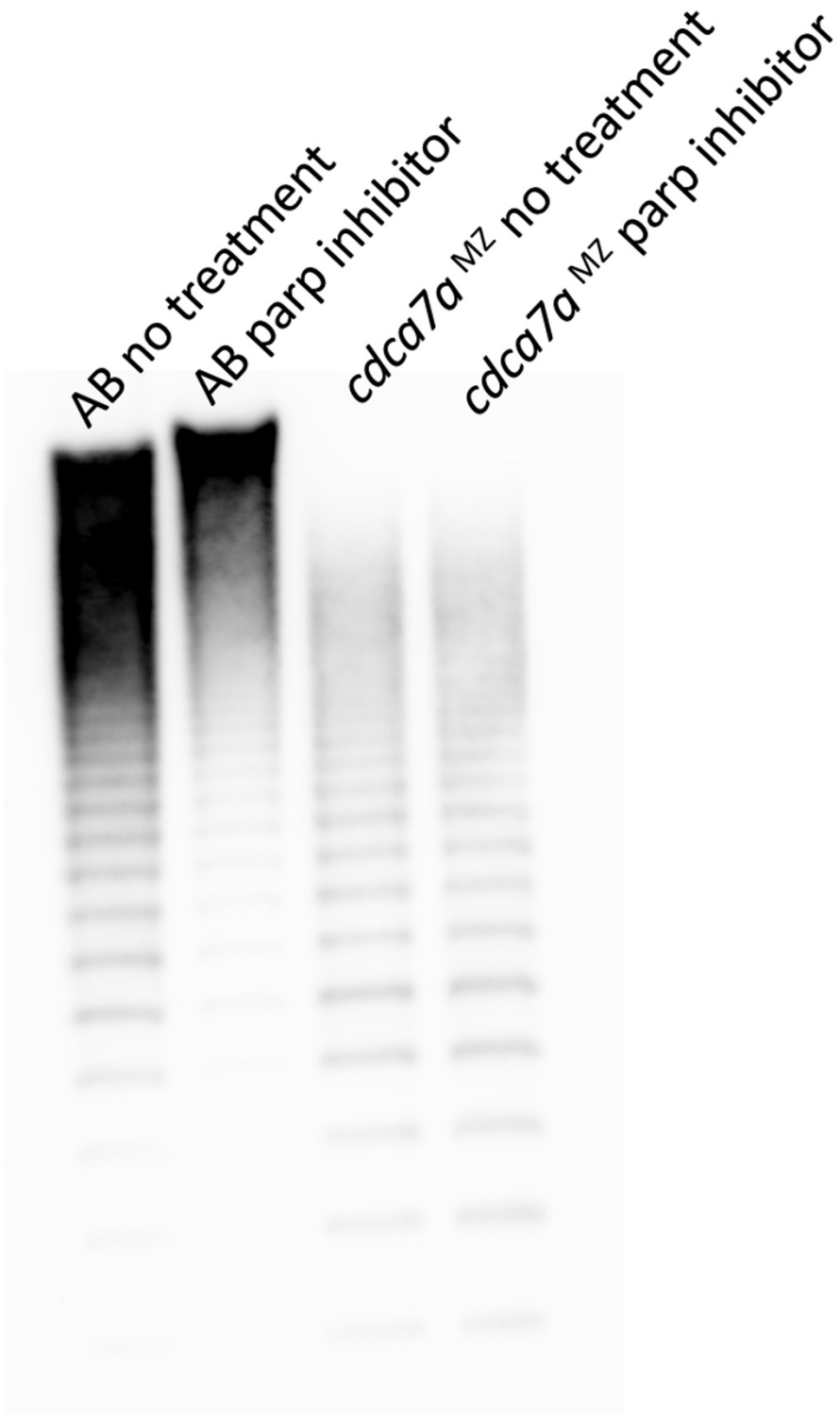
B


p=0.0005
Relative Sat1 Expression
4
3
2
1
0
WT
cdca7-/-

APPENDIX 2: PARP INHIBITION 1DPF SOUTHERN BLOT

Through a protein array it was found that CDCA7 binds to PARP2. PARP2 is a protein involved in catalyzing poly(ADP-ribosyl)ation (PARylation). PARylation is involved in many cellular functions including DNA methylation through its control of Dnmt1. It has been shown that inhibition of PARylation leads in an increase in DNA methylation in both genomic DNA and CpG island regions (de Capoa et al. 1999; Zardo et al. 1997; Zardo and Caiafa 1998). Thus, I set out to determine whether PARylation was involved in DNA methylation at the pericentromere. I instructed an undergraduate student, Erik Schouten, in performing a parp inhibitor experiment. He added 5uM of Niraparib to inhibit PARylation in wild-type and *cdca7a* maternal zygotic mutants. For the no treatment controls he used 1% DMSO. Embryos were treated with DMSO or Niraparib at six hours post fertilization and collected at twenty-four hours post fertilization. Upon parp inhibitor treatment I performed that twenty-four hour post fertilization Southern Blot. This Southern Blot shows that upon treatment of parp inhibitor wild-type samples become less resistant to digested. Indicating that there is an increase in DNA methylation at the pericentromere upon inhibition of PARylation. However, this was not seen in the *cdca7a* maternal zygotic mutants upon inhibition of PARylation. In order to make any conclusions this experiment would need to be repeated. However, I hypothesize that Cdca7a binds to PARylated PARP2 instead of DNA, but once PARylation is inhibited then it is able to bind to DNA leading to an increase in DNA methylation at the pericentromere.

A


AB no treatment
AB parp inhibitor
cdca7a MZ no treatment
cdca7a MZ parp inhibitor

APPENDIX 3: LIST OF PRIMERS

qPCR

q_IgZ_F	AAAGCAACGATACCAAAGTG (Page et al. 2013)
q_IgZ_R	AACAGCTTGCAAGACAATTC (Page et al., 2013)
q_IgD_F	GACACATTAGCCCATCAGCA (Page et al., 2013)
q_IgD_R	CTGGAGAGCAGCAAAAGGAT (Page et al., 2013)
q_IgM_F	GAAGCCTCCAATTCTGTTGG (Zimmerman et al. 2011)
q_IgM_R	CCGGGCTAAACACATGAAG (Zimmerman et al., 2011)
q_bactin1_F	CGAGCAGGAGATGGGAACC (McCurley and Callard 2008)
q_bactin1_R	CAACGGAAACGCTCATTGC (McCurley and Callard, 2008)
q_Sat1_F	GTCTCTGACTGAGTTTGCATTAC (Rajshekar et al., 2018)
q_Sat1_R	ACATTCTGAATTGGACGTTGA (Rajshekar et al., 2018)

gRNAs and crRNAs

*cdca7a*_gRNA2	GAAATTAATACGACTCACTATAGGTTCATCCTGGTGGACGGAGTTTTAGAGCTAGAAAT
*cdca7a*_gRNA5	GAAATTAATACGACTCACTATAGGCGTCCGCGGTCAATTCTGGTTTTAGAGCTAGAAAT
*cdca7a*_crRNA1	TTCCAGGATTGAACTTGAAT
*cdca7a*_crRNA2	GAGCTCAACAGTTTTGCCTA

Genotyping primers

Name of allele	Mutation in	Genotyping primers	Reference
mk22	*zbtb24*	F(P1):AGTCCTCGCTCTGCACTCAG R(P3):CTCTTGGCGGTGAAACACTT	Rajshekar et. al 2018
	cdca7a	F:TGGTTCCATTTTGTAAAACATGACC R: GCGTCACGGACTTCTTCTCC	This publication