CLARIFYING THE RELATIONSHIP BETWEEN PERICENTROMERIC HYPOMETHYLATION, DNA DAMAGE, AND MALIGNANCY: INSIGHTS FROM ZEBRAFISH MODELS OF IMMUNODEFICIENCY, CENTROMERE INSTABILITY, AND FACIAL ANOMALIES (ICF) SYNDROME

by

ANVITH P. REDDY

(Under the Direction of Mary Goll)

ABSTRACT

Hypomethylation of pericentromeric sequences is commonly observed in cancer, and it has been speculated that this hypomethylation may contribute to genome instability and aneuploidy. At present, the relationship between pericentromeric methylation and cancer is only correlative, and a causative role is yet to be established. Zebrafish models of ICF syndrome present an opportunity to dissect the relationship between pericentromeric hypomethylation, genome instability, and cancer. I aimed to test whether transcriptional signatures associated with DNA damage are associated with pericentromeric methylation loss and if ICF-like mutations lead to increased DNA damage susceptibility. Further, I examine if a pericentromeric hypomethylation is associated with decreased viability in a sensitized genetic background**.** We suggest that pericentromeric hypomethylation leads to an upregulation of DNA-damage associated transcripts. Further, we suggest that ICF- mutant zebrafish exhibit greater levels of DNA damage and DNA damage susceptibility as evidenced by γH2AX and RNA:DNA hybrids. Lastly, we suggest a loss of *cdca7* acts to increase death in *cdca7*$^{-/-}$ *p53*$^{-/-}$ zebrafish.

INDEX WORDS: DNA methylation, Epigenetics, DNA damage, Cancer, ICF syndrome

CLARIFYING THE RELATIONSHIP BETWEEN PERICENTROMERIC HYPOMETHYLATION, DNA DAMAGE, AND MALIGNANCY: INSIGHTS FROM ZEBRAFISH MODELS OF IMMUNODEFICIENCY, CENTROMERE INSTABILITY, AND FACIAL ANOMALIES (ICF) SYNDROME

By

ANVITH P. REDDY

B.S., The University of Georgia, 2021

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

MASTER OF SCIENCE

Athens, Georgia
2022

CLARIFYING THE RELATIONSHIP BETWEEN PERICENTROMERIC HYPOMETHYLATION, DNA DAMAGE, AND MALIGNANCY: INSIGHTS FROM ZEBRAFISH MODELS OF IMMUNODEFICIENCY, CENTROMERE INSTABILITY, AND FACIAL ANOMALIES (ICF) SYNDROME

By

ANVITH P. REDDY

Major Professor: Mary Goll

Committee: Robert Schmitz
Zachary Lewis

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2022

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge Dr. Mary Goll for giving me the opportunity to work on such an exciting project. I am extremely grateful for her patience and commitment to developing my scientific and critical thinking skills throughout my time in her lab. I have learned invaluable skills that I will continue to use in my future scientific career. Further, I would like to acknowledge Alysha Higgs for her invaluable support and guidance. I would also like to extend thanks to the other members of the Goll Lab for their scientific and compassionate support throughout the duration of my project. I would also like to thank Dr. Zachary Lewis and Dr. Bob Schmitz for agreeing to be on my graduate committee. Finally, I am deeply thankful for all my friends and family who have supported my endeavors throughout my time at the University of Georgia.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

5mC	5 methyl-cytosine
CDCA7	Cell division cycle associated 7
CpG	Cytosine-guanine dinucleotide
DNMT	DNA Methyltransferase
dpf	Days post fertilization
HELLS	Helicase, lymphoid-specific
hpf	Hours post fertilization
ICF	Immunodeficiency, Centromere Instability, and Facial Anomalies
LINE	Long Interspersed Nuclear Elements
LTR	Long Terminal Repeats
mpf	Months post fertilization
MZ	maternal zygotic
NHEJ	Non-Homologous End Joining
Sat1	Satellite-1
SINE	Short Interspersed Nuclear Elements
ssDNA	Single-Stranded DNA
UHRF1	Ubiquitin-like with PHD and ring finger domains 1
ZBTB24	Zinc-finger and BTB domain containing 24

CHAPTER 1: INTRODUCTION

DNA methylation and 5-methylcytosine

Addition of a methyl group to the 5' position of the cytosine ring (5-methylcytosine, 5mC) is common in many eukaryotic, fungal, protist, and bacterial genomes (Goll and Bestor, 2005). Although other types of DNA methylation such as 5-hydroxymethylcytosine and 6-methyladenine exist, 5-methylcytosine is by far the most common DNA modification in eukaryotic genomes and is often referred to simply as DNA methylation. The distribution of 5mC varies among different species, however, vertebrate genomes are predominantly methylated at CpG dinucleotides with approximately 70-80% of these sequences methylated. Most methylation at CpG dinucleotides occurs in intergenic and intronic regions of the genome including transposons and other highly repetitive sequences, such as pericentromeric satellite repeats and telomeres (Bird et al., 1985; Jin et al., 2011).

DNA methyltransferase enzymes catalyze 5-methylcytosine (5mC)

5-methylcytosine is deposited by the covalent transfer of a methyl group from the co-factor S-adenosyl-L-methionine to the 5' carbon of the cytosine ring of DNA (Moore et al., 2013). This event is catalyzed by DNA methyltransferase enzymes (DNMTs). Vertebrate genomes encode for two types of DNA methyltransferases. Methyltransferases of the Dnmt3 family are primarily responsible for the *de novo* establishment of new methyl marks on cytosine residues. Methyltransferases of the Dnmt1 family typically maintain 5-methylcytosine through replication after establishment by Dnmt3 enzymes (Goll and Bestor, 2005).

Dnmt1 mediated maintenance methylation is facilitated by interactions with the Ubiquitin-like with PHD and ring finger domains 1 protein, Uhrf1. During the S phase of the cell cycle, Uhrf1 helps localize Dnmt1 to the replication fork where Dnmt1 is able to recognize hemi-methylated CpG dinucleotides and "copy" methylation onto the symmetrical CpG site on the newly synthesized DNA strand (Li et al., 1992). Through this mechanism, methylation states can be faithfully propagated through semi-conservative rounds of replication without outside stimulus as long as the Dnmt1 and Uhrf1 enzymes are functional.

All vertebrate genomes examined to date only encode for a single Dnmt1 ortholog (Alvarez-Ponce et al, 2018). However, the number of Dnmt3 de novo methyltransferases varies significantly between genomes. For example, the human genome encodes for two Dnmt3 family members (*DNMT3A* and *DNMT3B*), mice encode for three de novo methyltransferases (*Dnmt3A*, *Dnmt3B*, and *Dnmt3C*) and zebrafish encode for six Dnmt3 orthologs (Xi et al., 1999; Molaro et al., 2020; Campos et al., 2012). The mechanisms that target DNMT3 de novo methyltransferases to different genomic sequences to establish DNA methylation are not fully clear.

DNA methyltransferases are essential for normal vertebrate development

Studies in mice:

To date, most of what we know about the consequences of DNA methylation loss in vertebrate species comes from studies in mice. Studies of mice harboring null mutations suggest Dnmt3a and Dnmt3b share partially redundant functions in promoting *de novo* methylation. Mice with a complete loss of Dnmt3a activity are viable to birth but become runted and die 4 weeks after birth with a marked loss of germ cells in males (Okano et al., 1999). *Dnmt3a* mutant mice have a relatively intact methylome with intact methylation at C-type retroviral elements, major satellite DNA, and minor satellite repeats. *Dnmt3b* knockout mice are not viable to term, die before

E15.5, and exhibit developmental defects including growth impairment and rostral neural tube defects. These mice also have methylation defects including modest hypomethylation at C-type retroviral elements and substantial hypomethylation at minor satellite repeats. *Dnmt3a/Dnmt3b* double mutants exhibit a smaller size and abnormal morphology; these mice die before E11.5. As *Dnmt3a/Dnmt3b* double mutant mice lack somites and do not undergo embryonic turning, it is likely that their morphogenesis arrests shortly after gastrulation. *Dnmt3a/Dnmt3b* double mutants exhibit substantial hypomethylation at C-type retroviral elements, slight hypomethylation at major satellite repeats, and *Dnmt3b* mutant-comparable hypomethylation at minor satellite repeats (Okano et al., 1999).

Dnmt1 mutant mice present with extensive reductions in genome-wide methylation and are embryonic lethal (Li et al., 1992; Okano et al., 1999). *Dnmt1* mutant hypomethylation phenotypes are more severe across the entire genome when compared to *Dnmt3a*, *Dnmt3b*, and *Dnmt3a/Dnmt3b* mutants. Like *Dnmt1* mutants, loss-of-function *Uhrf1* mutations also result in genome-wide hypomethylation and embryonic lethality in mice (Sharif et al., 2007; Yamashita et al., 2018).

Studies in Zebrafish:

In Zebrafish, *dnmt3bb.2* is the only *de novo* methyltransferase for which phenotypes have been reported. *Dnmt3bb.2* morpholino-injected embryos die by 96 hours post fertilization with phenotypic abnormalities including small brains, defective pharyngeal arch formation, and abnormal retinal neural epithelial differentiation (Rai et al., 2010).

Homozygous loss of function for *dnmt1* in zebrafish is associated with decreased levels of DNA methylation; this is consistent with morpholino-knockdown studies of Dnmt1 which describe a 40% reduction in DNA methylation (Tittle et al., 2010; Anderson et al., 2009; Rai et

al., 2006). Other *dnmt1* mutant phenotypes include a loss of expression of terminal differentiation markers in the intestine, exocrine pancreas, and retina. Subsequently, mutant fish have a smaller exocrine pancreas as well as liver, eye, and jaw defects by 4 days post fertilization. Lastly, *dnmt1* mutant zebrafish also exhibit molecular phenotypes including increased levels of p53-mediated apoptosis (Anderson et al., 2009). Like *dnmt1* mutants, hypomorphic mutations in zebrafish *uhrf1* result in a roughly 50% decrease in global DNA methylation (Sadler et al., 2007; Feng et al., 2010).

Known functions of DNA methylation

DNA methylation is implicated in (1) regulating gene expression at some gene promoters, (2) silencing of some repetitive elements, and (3) promoting genomic stability.

DNA methylation at gene promoters and regulatory regions, has been shown to repress transcription. Although a large portion of CpG sites within the genome are methylated, CpG dense regions in promoters termed "CpG islands" generally remain unmethylated (Rakyan et al., 2004; Meissner et al., 2008). Further, methylation at upstream regulatory regions with lower CpG content has also been associated with repression of gene expression in a variety of tissues and cell types. In addition to upstream regulatory regions, methylation of gene bodies has also been noted. The role of gene body methylation remains unclear, however contrary to canonical roles of DNA methylation in mediated repression, some studies suggest DNA methylation in gene bodies might lead to an increase in gene expression (Ball et al., 2009).

DNA methylation also plays a role in the silencing of repetitive elements and is implicated in maintaining genomic stability, particularly at repeat sequences. Roughly 50% of the mammalian genome is made up of transposable and retroviral elements (Walsh et al., 1998; Ayarpadikannan and Kim, 2014). These regions are usually highly repetitive and heavily methylated. Repetitive

DNA contains roughly 52% of all CpG dinucleotides and 90% of 5-methylcytosine in humans (Lander et al., 2001; Beisel and Paro, 2011).

Methylation at repeat sequences

In humans, repetitive elements fall into two categories: interspersed repeats and tandem repeats. Interspersed repeats are composed of SINEs (Short Interspersed Elements), LINEs (Long Interspersed Elements), LTRs (Long Terminal Repeats), and DNA transposons (López-Flores et al., 2012. Within these intergenic regions, normal levels of DNA methylation act to repress transcription and translocation. These repetitive elements are potentially harmful and could lead to unwanted parasitic integration, gene disruption, and DNA damage if spuriously expressed (Kuster et al., 1997; López-Flores and Garrido-Ramos, 2012).

Tandem repeats include satellite repeats (100 kb to over 1 Mb), minisatellites (1 kb to 20 k), and microsatellites (less than 150 bp). Alpha-satellite repeats are typically associated with the centromere, classical satellite repeats are associated with the pericentromeres, minisatellites are associated with the telomeres, and microsatellites are found scattered throughout the genome (Garrido-Ramos, 2017). Satellite repeats are implicated in sustaining heterochromatin architecture, facilitating kinetochore assembly and separation of chromosomes during the cell cycle, and chromocenter formation (Allshire and Madhani, 2018; Plohl et al., 2014; Jagannathan et al., 2018). These satellite repeats are often highly methylated, which may contribute to the establishment of centromeric architecture as well as the binding of centromeric proteins such as CENP-B and CENP-C (Scelfo and Fachinetti, 2019). There is still, however, much work that needs to be done to further elucidate the significance of DNA methylation abundance at pericentromeric repeats.

DNA methylation in cancer

Genome-wide changes in the epigenetic landscape are a hallmark of cancer and are associated with the onset of malignancy. Changes in DNA methylation patterns observed in cancer include: (1) gain of methylation at specific CpG islands and regulatory regions associated with tumor-suppressor genes and (2) decreases in methylation at specific oncogenes and repetitive elements (Jones and Laird, 1999; Smet and Loriot, 2010; Sheaffer et al., 2016). These changes often occur early in cancer development and may contribute to cancer initiation and progression.

Hypermethylation may impede the docking of transcriptional regulators and allow for the formation of chromatin through interactions with various other epigenetic modifications. Silencing through methylation can act in tandem with genetic mutations to promote oncogenesis (Jones and Laird, 1999). Advances in identifying site-specific hypermethylation have led to better detection of early cancer states. For example, hypermethylation of promoter regions in *APC*, *RASSF1A*, and *TP53* genes are considered as common epigenetic markers for early detection and development of cancer (Moosavi and Ardekani, 2016).

Hypomethylation in cancer is often observed at normally heavily methylated repeat elements, including satellites such as SAT2, retrotransposons such as LINEs, and centromeric repeats (Jin et al., 2011; Sharma et al, 2019). Hypomethylation at repeat sequences has been seen in a variety of different types of cancers including epithelial carcinoma, colorectal cancer, breast cancer, and glioblastoma (Suzuki et al., 2002; Fanelli et al., 2008; Narayan et al., 1998; Tsuda et al., 2002). This loss of methylation might lead to increased genomic instability and promote chromosomal rearrangements. However, the at present the link between hypomethylation at repeat sequences and cancer is mainly correlative, and a causative link has yet to be clearly established.

Immunodeficiency, Centromere instability, and Facial anomalies (ICF) syndrome

Immunodeficiency, Centromere instability, and Facial anomalies (ICF) syndrome is a rare autosomal recessive disease characterized by the selective loss of DNA methylation at pericentromeres. To date, there are 118 documented cases of ICF syndrome spanning 94 families (Kiaee et al., 2020).

Pericentromeric methylation loss in ICF syndrome

At the molecular level, ICF syndrome is characterized by a selective loss of methylation at the pericentromeric satellite repeats that flank chromosome centromeres. All individuals afflicted with ICF syndrome exhibit loss of methylation at classical satellite 2 and 3 sequences which are found predominately on chromosomes 1, 9, and 16 (Jeanpierre et al., 1993). In addition, individuals with ICF syndrome subtypes 2, 3, and 4 display hypomethylation of pericentromeric alpha-satellite repeats (Miniou et al., 1997). Few methylation abnormalities are seen outside of the pericentromeres; when seen, these abnormalities are localized to sub-telomeric repeats and CpG islands contained within intergenic regions of the genome (Simo-Riudalbas et al., 2015).

Centromeric instability and chromosomal abnormalities in ICF syndrome

Pericentromeric hypomethylation in ICF syndrome is associated with centromeric region instability, particularly affecting chromosomes 1, 9 and 16, at regions of classical satellite DNA. When induced to undergo high levels of replication, mitogen-stimulated cultured lymphoblastoid cell lines display dramatic chromosomal rearrangements, including the formation of multi-radial chromosomes (Ehrlich et al., 2018). Other common chromosomal abnormalities seen in ICF syndrome include chromosome de-condensation, chromatid/chromosomal breaks, somatic pairing, translocations between homologous and nonhomologous chromosomes, and the abnormal

formation of micronuclei (Haas, 1990; Maraschio et al., 1992; Smeets et al., 1994; Stacey et al., 1995; Weemaes et al., 2013).

Immunodeficiency in ICF syndrome

Individuals with ICF syndrome exhibit Immunodeficiency characterized by hypogammaglobulinemia. Although ICF individuals present with normal levels of lymphoid cells, they exhibit low levels of immunoglobulins, notably IgA, IgM, and IgE (Tiepolo et al., 1979; Gimelli et al., 1993). Low levels of circulating immunoglobulins are associated with recurrent and prolonged infections, most commonly of the respiratory system, skin, and digestive system. In addition, roughly half of ICF individuals present with T-lymphocytopenia (low T-cell counts) and an altered $CD4^+/CD8^+$ ratio (Smeets et al., 1994; Pezzolo et al., 2001). Individuals often succumb to death due to infections in late childhood or early adulthood. Clinically, intravenous supplementation of immunoglobulins is a common treatment. Further, allogeneic hematopoietic stem cell transplantation in ICF infants has also shown promise in ameliorating immunodeficiency and improving life expectancy (Gössling et al., 2017).

Additional phenotypes associated with ICF syndrome

Facial anomalies including hypertelorism, a flat nasal bridge, epicanthal folds, protrusion of the tongue, micrognathia, and low set ears are often reported in ICF individuals (Haas, 1990; Fasth et al., 1990; Weemaes et al., 2013). Individuals also often present with growth retardation, gastrointestinal defects, and an overall failure to thrive. Intellectual impairment and neurological defects have been described in 63% of documented cases (Kiaee et al., 2020).

Genetic basis of ICF Syndrome

Through homozygosity mapping and whole-exome sequencing, the genes responsible for the four subtypes of ICF syndrome (ICF type 1-4) have been identified (Fig. 1). These include (1) DNA Methyltransferase 3B (DNMT3B, ICF type-1), (2) Zinc-finger and BTB domain containing 24 (ZBTB24, ICF type-2), (3) Cell division cycle associated 7 (CDCA7, ICF type-3), and (4) Helicase, lymphoid specific (HELLS, ICF type-4). There is a small subcategory of ICF-X individuals that do not fall under these four categories (Xu et al., 1999; de Greef et al., 2011; Thijssen et al., 2015).

(1) DNA Methyltransferase 3B (DNMT3B, ICF-1): 60% of individuals (71 cases) diagnosed with ICF are homozygous for mutations in the gene encoding for the DNA methyltransferase DNMT3B (Weemaes et al., 2013). ICF mutations in DNMT3B are rarely nonsense. Instead, most are missense mutations that cluster in the catalytic domain of DNMT3B, suggesting the strong likelihood that they are hypomorphic alleles which impair methyltransferase activity (Wijimenga et al., 2000).

(2) Zinc-finger and BTB/POZ domain-containing 24 (ZBTB24, ICF-2): 30% of individuals (35 cases) of ICF syndrome result from homozygous predicted loss-of-function mutations in ZBTB24 (Weemaes et al., 2013). ZBTB24 is thought to primarily function as a transcription factor that binds DNA via its zinc finger domains and is highly expressed in the lymphoid lineage (de Greef et al., 2011; Siggs and Beutler, 2012).

(3) Cell division cycle associated 7 (CDCA7, ICF-3): 4% of individuals (5 cases) have mutations in CDCA7 (Thijssen et al., 2015; Kiaee et al., 2020). All mutations in CDCA7 are in the conserved Zinc Finger Domain (4 CXXC-motifs). CDCA7 is a known c-Myc responsive gene and has been shown to interact with Ku70 and Ku80, proteins involved in the non-homologous

end joining (NHEJ) repair pathway (Gill et al., 2013; Unoki et al., 2018). In cell culture, knockdown of CDCA7 has been shown to cause CpG hypomethylation at centromeric repeats as well as increased susceptibility to DNA damage (Thijssen et al., 2015; Unoki et al., 2019).

(4) Helicase, lymphoid-specific (HELLS, ICF-4): 6% of individuals (seven cases) have mutations in (Kiaee et al., 2020). Apart from one missense mutation, all mutations in HELLS are homozygous or compound heterozygous nonsense mutations (Thijssen et al., 2015). Knockdown of HELLS in cell culture has been shown to result in CpG hypomethylation at centromeric repeats. (Thijssen et al., 2015). Further, HELLS has been shown to interact with CDCA7 and remodel nucleosomes. It has been shown that the CDCA7/HELLS complex is required for the facilitation of non-homologous end joining (NHEJ). It has also been implicated in accumulation of DNMT1/UHRF1 methylation maintenance machinery on newly synthesized DNA and the resolution of R-loops composed of DNA:RNA hybrids and ssDNA (Unoki et al., 2019, Unoki et al., 2020).

All four known ICF genes are thought to mechanistically interact in a singular pathway (Fig. 2). Most upstream, ZBTB24 is thought to promote the expression of CDCA7 (Wu et al., 2016). Current understanding of CDCA7 function is that it forms a complex with HELLS (Jenness et al., 2018). This nucleosome remodeling complex interacts with DNMT3B to promote pericentromeric DNA methylation. (Wu et al., 2016; Jenness et al., 2018). It has also been suggested by Thompson et. al that ZBTB24 interacts with DNMT3B to influence DNA methylation. However other studies have failed to detect such interactions (Thompson et. al, 2018; Hardikar et al., 2020). Although this proposed pathway aligns with methylation defects seen in ICF-2, 3, and 4 individuals, several questions remain unanswered, including how CDCA7 recruits

HELLS to centromeric satellite repeats, how DNMT3B is localized to the pericentromeres, and why alpha-satellite methylation differs among ICF subtypes.

Animal models of ICF syndrome

To date, attempts to generate mouse models of ICF have only been partially successful. Although mice carrying ICF-like point mutations in *Dnmt3b* exhibit some characteristics of ICF syndrome including facial anomalies and T-cell defects, they fail to exhibit other key aspects of ICF syndrome including immunoglobulin deficiency (Ueda et al., 2006). Mice null for *Zbtb24* exhibit embryonic lethality; methylation changes in these mutants have not yet been reported (Wu et al., 2016). Mice mutant for the *Hells* orthologue, *Lsh*, exhibit perinatal lethality and a roughly 50% reduction in 5mC. Methylation loss was observed at both repeat sequences and some single copy genes (Tao et al., 2011). Mice mutants for *Cdca7* have not yet been reported.

Zebrafish are an emerging model organism to study development and epigenomic changes. Rajshekar et al., describe a viable model of pericentromeric hypomethylation through a 9kb deletion in *zbtb24* removing exons 2 through 5 (Fig. 3) (Rajshekar et al., 2018). Unlike all ICF mice models thus far reported, Zebrafish *zbtb24* mutants exhibit key phenotypic hallmarks of the disease. These mutants exhibit hypomethylation (Fig. 4) and de-repression of Sat1 pericentromeric satellite repeats; methylation loss in these mutants is progressive over the lifetime of the organism. Like ICF syndrome, *zbtb24* mutants also exhibit facial anomalies characterized by elongation of the snout. Furthermore, these mutants exhibit low levels of immunoglobulins (*IgM*, *IgD*, and *IgZ*) but normal levels of lymphoid cells at 6 weeks post fertilization. Consistent with growth retardation and the small stature observed in ICF syndrome, *zbtb24* mutants are quantifiably smaller than wild-type siblings at 3-4 weeks post fertilization. These mutants exhibit reduced lifespans and fail to intercross or outcross.

Our lab has also generated viable *cdca7* homozygous mutants using CRISPR-Cas9 mutagenesis (Fig. 3) (Higgs, A., *Unpublished Data*). These mutants are homozygous for a 6kb deletion of exons 3 through 8. These mutants exhibit a progressive loss of methylation at the pericentromeres. *Cdca7* mutants experience an earlier loss of methylation than *zbtb24* mutants, and *cdca7* maternal zygotic progeny exhibit methylation loss as early as 6 hours post fertilization (Fig. 4). *cdca7* mutants are fertile and yield maternal zygotic progeny viable to adulthood. Immunoglobulin deficiency evidenced by *IgM* expression is also observed in *cdca7* mutants, however, this phenotype is stronger in *zbtb24* mutants. Unlike *zbtb24* mutants, *cdca7* mutants do not exhibit any facial anomalies or reduction in body size.

These zebrafish models of ICF syndrome open opportunities to study pericentromeric hypomethylation, its relation to cancer, and its mechanistic underpinnings in the context of a developing vertebrate animal system. My thesis work uses our viable zebrafish models of ICF syndrome to clarify the relationship between ICF mutations, pericentromeric methylation loss, and DNA damage.

CHAPTER 1: FIGURES

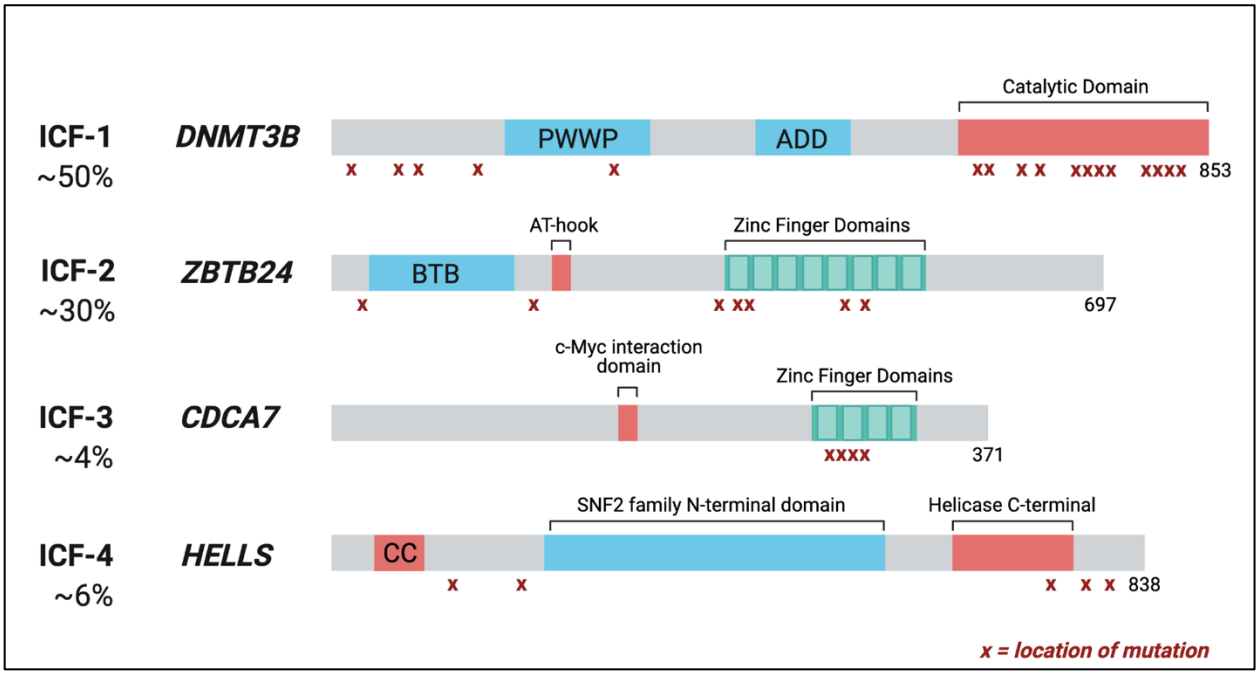


Figure 1: Genetic basis of ICF syndrome and zebrafish mutant alleles. Four genes have been implicated in ICF syndrome. *DNMT3B* (ICF-1), *ZBTB24* (ICF-2), *CDCA7* (ICF-3), and *HELLS* (ICF-4). (Adapted from Kiaee et al., 2020; BioRender)

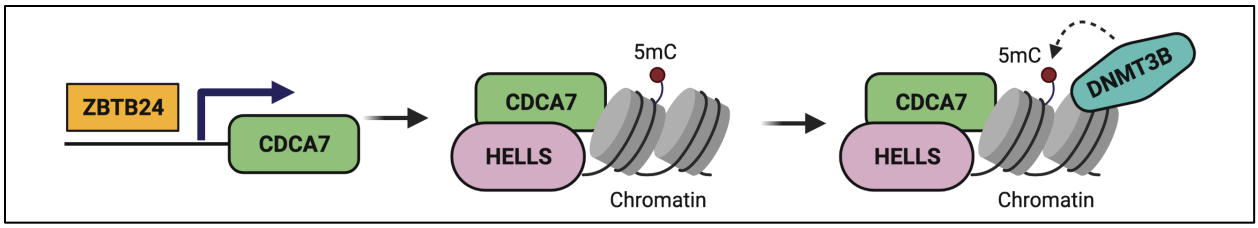


Figure 2: Molecular basis of ICF syndrome and zebrafish mutant alleles. Predicted model for the interaction of all four ICF genes. Zbtb24 stimulates transcription of Cdca7. Cdca7 and Hells form a complex and reposition nucleosomes, giving Dnmt3b access to DNA for methylation. (Adapted from Jenness et al., 2018; BioRender)

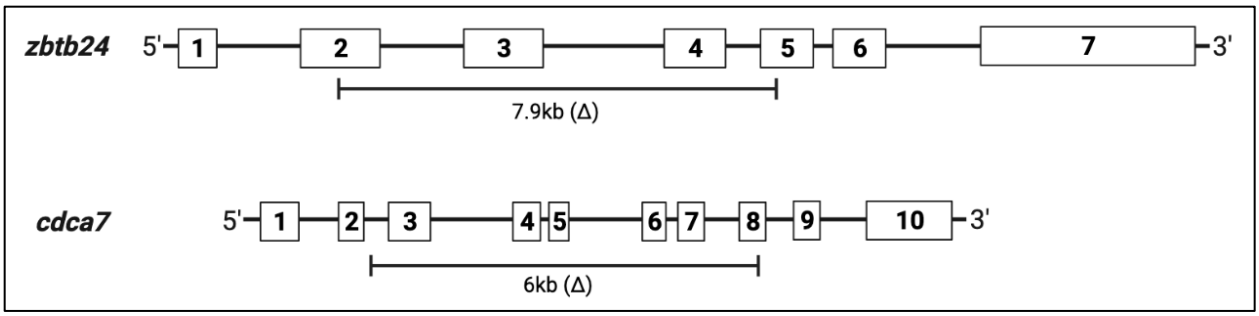


Figure 3: Deletion mutations in zebrafish models of ICF syndrome. 7.9kb and 6kb deletion mutants in zebrafish *zbtb24* and *cdca7* respectively.

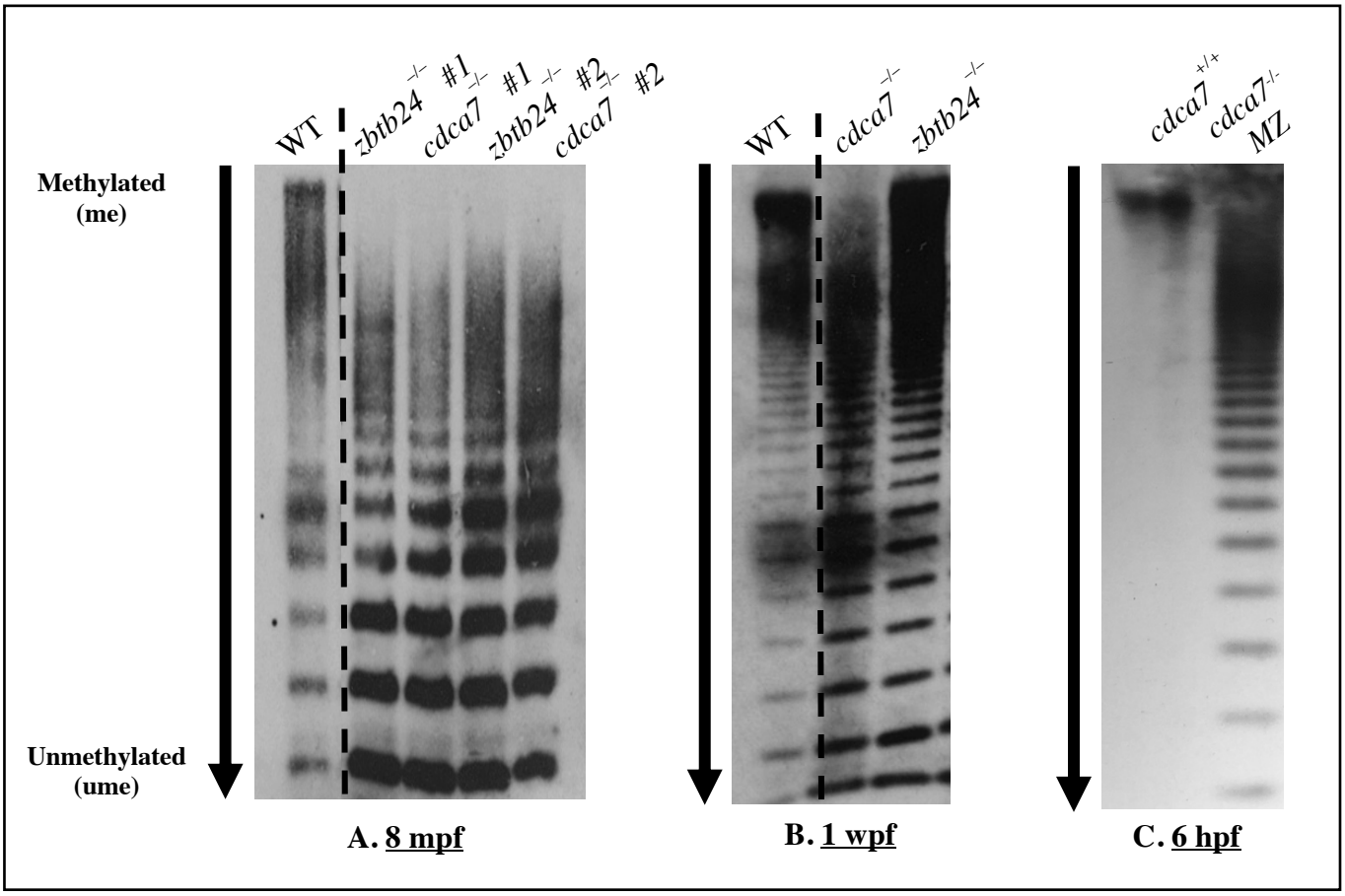


Figure 4: *cdca7*$^{-/-}$ zebrafish exhibit earlier methylation loss than *zbtb24*$^{-/-}$ *mutants.* Methylation sensitive southern blots were performed by A. Higgs (*unpublished data*). Genomic DNA was digested with the methylation sensitive enzyme HpyCH4IV, and a biotin labeled probe was used to visualize satellite repeats. Hypomethylation is evidenced by increased digestion. (A) *zbtb24*$^{-/-}$ and *cdca7*$^{-/-}$ mutants exhibit a comparable loss of methylation at 8 mpf. (B) *cdca7*$^{-/-}$ mutants exhibit a loss of methylation at earlier developmental timepoints than *zbtb24*$^{-/-}$ mutants. (C) *cdca7*$^{-/-}$ maternal zygotic mutants lose methylation as early as 6 hpf.

CHAPTER 2: UPREGULATION OF DNA DAMAGE RESPONSE GENES IN ZEBRAFISH MODELS OF ICF SYNDROME

PREFACE

In order to better understand gene expression changes associated with pericentromeric hypomethylation, our laboratory performed RNA-seq on wildtype zebrafish larvae and larvae that were homozygous mutant for either *zbtb24* or *cdca7* at 2 weeks post fertilization. This stage was chosen because it is the stage at which pericentromeric hypomethylation is first observed in *zbtb24* homozygous mutants (Rajshekar et al., 2018). At this time point, we also find extensive pericentromeric hypomethylation in *cdca7* homozygous mutants (Fig. 4).

My thesis work involved reanalysis of previously reported RNA-seq data comparing expression in wildtype and *zbtb24*$^{-/-}$ zebrafish (Rajshekar et al., 2018). I extended the original analysis to include a more in-depth gene-set and pathway analysis than previously performed. Simultaneously, I performed primary analysis of *cdca7*$^{-/-}$ RNA-seq data generated in our laboratory. I independently extended this analysis by performing gene-ontology, gene-set enrichment, and pathways analyses. Finally, I independently confirmed select RNA-seq results by quantitative reverse transcriptase PCR (qRT-PCR).

RESULTS

Data validation and Normalization

For our *zbtb24* analysis, we compared three mutant samples to two wildtype controls. Samples generated an average of 35.4 million total paired end reads and produced an average of 94.77% uniquely mapped reads, for a total of 33.6 million successfully assigned alignments. For

cdca7, we compared two wildtype and three mutant samples. Single-end RNA-seq generated an average of 15.7 million reads per-sample. With an average of 90.94% uniquely mapped reads, there was a total average of 12.3 million successfully assigned alignments per sample.

Initially, I performed quality control tests to assess the quality of RNA-seq data. I normalized the datasets for sequencing depth and RNA composition. Before normalization (Fig. 5A and Fig. 6A), variance in the mean log counts distribution prevents accurate downstream comparisons between samples. After normalization (Fig. 5B and Fig. 6B), all samples have the same mean log counts value, indicating proper quality control.

I stabilized the data to account for genes with low counts using a blind, parametric rlog transformation. Before stabilization for count variance (Fig. 5A and Fig. 6A), log normalized counts show a fanning out of datapoints in the lower left quadrant. Stabilization uses the variance model for count data to shrink together log-transformed counts for genes with very low counts. Stabilization values (rlog) are very close to the original values (log2) for genes with medium and high counts. After stabilization for count variance (Fig. 5B and Fig. 6B), log normalized counts display a tapering of the bottom left quadrant and shrinkage towards the line, indicating proper variance stabilization. These quality controls ensure downstream analyses are not skewed.

Mutant zebrafish are expected to have a distinct expression profile when compared to wildtype controls. To assess whether or not these two populations (mutant vs. wildtype) segregated from one another, unsupervised clustering was performed (Fig. 5C-D and Fig. 6C-D). As replicate wildtype samples segregate from replicate mutant samples, we can conclude that mutant samples had a distinct expression profile when compared to wildtype samples for both the *cdca7* and *zbtb24* datasets.

Identification of differentially expressed genes

zbtb24$^{-/-}$ *vs. wild-type*

Volcano plots display differentially expressed genes in context of the entire dataset (Fig. 7A). This initial analysis allows for a broad visualization of differentially expressed genes and shows that there are there are a greater number of upregulated genes than downregulated genes which pass the threshold for adjusted p-value and false discovery rate (FDR). The top 50 misregulated genes are visualized by a heatmap (Fig. 7B). Upon further investigation, a few sets of related genes were seen. This includes a cluster of immune-related genes (including *irf1b*, *ifi45*, and *anxa1a*), a cluster of metabolic-related genes (including *alas2*, *bbox1*, and *cahz*), and a cluster of misregulated heat shock proteins (*hsp70l*, *hsp70.1*, *hsp70.2*, and *hsp70.3*). This data directly compares with the previous analysis by Rajshekar et al., 2018.

cdca7$^{-/-}$ *vs. wild-type*

Compared to the *zbtb24* dataset, there are a much higher number of differentially expressed genes in the *cdca7* analysis (Fig. 8A). However, there is still a larger proportion of upregulated genes to downregulated genes that pass the threshold for adjusted p-value and false discovery rate (FDR). The top 50 misregulated genes are visualized by a heatmap (Fig. 8B). Clusters of misregulated genes in this heatmap include metabolic-related genes (including *ucp1*, *soat2*, *tdh*, and *hmgcs1*), DNA replication/repair related genes (including *pcna*, *rrm1*, *tyms*, *top2a*, *lig1*, and *pole*), and cell-structure genes (including *tuba8l*, *tuba8l3*, *tuba8l4*, *kifc*1, and *kif11*).

Enrichment analysis for Gene Ontology (GO) terms

The genome is much more complex than genes in isolation, and, to dig deeper into the dataset and look for patterns, the next step was to perform an enrichment analysis for Gene

Ontology (GO) terms. The Gene Ontology (GO) Consortium is based in the United States and is funded by the NHGRI. The consortium has been in existence for almost 20 years and aims to is systematically define natural biological processes, molecular functions, and cellular components. Gene enrichment categories and terms are based on *in silico* research and confirmed by laboratory evidence. Gene-Ontology analyses were performed in the R programming environment using clusterProfiler version 3.0.4 (Yu et al., 2012). To perform enrichment analysis, I first input the normalized gene expression data. This was followed by gene-wise correlation and differential expression analysis. If a certain GO-term is enriched there will be a very low probability under the null hypothesis that the GO-term of interest follows the expected distribution.

zbtb24$^{-/-}$ vs. wild-type

GO analysis brought up many over-represented GO-terms. Of particular interest were chromatin, the nucleosome, the DNA-packaging complex, and protein-DNA interactions (Fig. 9A). In addition, many of these over-represented GO-terms had common genes between them (Fig. 9B-C). A cluster of epigenetic related GO-terms indicates that genes in these cluster are dysregulated in *zbtb24* mutant zebrafish as indicated by a connectivity map (Fig. 9C). When taking a look at common genes that underlie these multiple GO-terms, we see downregulation of the following genes: *npas4a*, *klf2a*, *egr2a*, and *esrrd, hsp70l, hsp70.2, hsp70.3, grik1b, rx2, pimr71* and an upregulation of the following genes: *cgas, ccl39.4, fbxl13, irf1b, spic, lgals9l4, prr11, plac8.2,* and *psmb8a*.

cdca7$^{-/-}$ vs. wild-type

GO analysis of *cdca7* data revealed many GO-terms involving the cell cycle including cell cycle process, miotic cell cycle, DNA replication. We also see an enrichment of chromosomal GO-

terms including nuclear chromosome segregation, chromosome segregation, sister chromatid segregation, chromosome condensation, and nuclear division (Fig. 10A-B). When looking at commonalities between GO-terms, upregulated genes between terms include: *tpx2, ska3, cepe, cdk2, cenpk, ncaph, tacc3*, and *bub1* among others. Downregulated genes were much less common as only three downregulated genes were shared among the GO-terms analyzed, *msh4, spo11,* and *nek11*. A full connectivity map of the top 11 enriched GO-terms is visualized in Fig. 10C.

Identification of disrupted gene-sets using GSEA

Gene-Set Enrichment Analysis (GSEA) is an alternative method for identifying disrupted gene-sets. This method creates an enrichment score based on a pre-defined statistic (e.g., fold change); it uses this to rank and compare pathway genes. GSEA is able to provide and analysis of both activated and suppressed gene-sets whereas GO-ontology can only provide analyses on activated/enriched gene-sets. GSEA analyses were performed in the R programming environment using clusterProfiler version 3.0.4 (Yu et al., 2012).

zbtb24$^{-/-}$ vs. wild-type

GSEA analysis surfaced a few terms of interest, notably, condensed chromosome /centromeric region, immune response, immune-effector process, and DNA replication which were all activated (Fig. 11A). There was also a subset of enriched cell cycle related terms: mitotic cell process, nuclear division, cell cycle process, and cell cycle. These gene-sets are of interest as they relate to ICF syndrome and might explain some ICF phenotypes as well as the mechanisms of changes seen in ICF individuals. Many of the suppressed pathways relate to signaling, sensory perception, and channel/ion activity (e.g., synaptic signaling, ion channel complex, voltage-gated ion channel activity, and visual perception). GSEA plots for immune response, chromosome-

centromeric region, and cell-cell signaling were generated to visualize Running Enrichment Score and the ranked dataset (Fig. 11B-D).

cdca7$^{-/-}$ *vs. wild-type*

In the *cdca7* dataset there were similarities to *zbtb24* dataset including an activation of the chromosome/centromeric region as well as the upregulation of the cell cycle and cell cycle process. However, there was a marked increase in terms related to other chromosomal terms including mitotic sister chromatid segregation, chromosomal region, and chromosome organization. Notably, there was also an increase DNA damage and repair terms. These include DNA repair, cellular response to DNA damage stimulus, double-strand break repair, and DNA replication Similar to the *zbtb24* dataset, suppressed pathways included synaptic signaling, cation channel activity, and sensory perception (Fig. 12A). GSEA plots for DNA repair, chromosome organization, and double strand break repair were generated to visualize Running Enrichment Score and the ranked dataset (Fig. 12B-D).

Identification of disrupted pathways using KEGG

Kyoto Encyclopedia of Genes and Genomes (KEGG) is another database to conduct genomic and pathway analyses. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a consortium based in Japan. KEGG is most useful for its robust curation of pathways in human and other many other species. KEGG has systemically defined both normal/healthy and disease-related pathways. Unlike GSEA or GO-analyses which are just based on related genes, KEGG analyses provide insight into gene interactions and pathway diagrams. GSEA analyses were performed in the R programming environment using clusterProfiler version 3.0.4 (Yu et al., 2012).

zbtb24$^{-/-}$ *vs. wild-type*

My KEGG pathway analysis (Fig. 13) highlights numerous upregulated genes relevant to ICF syndrome and ICF phenotypes. These include Homologous Recombination, Mismatch Repair, Fanconi Anemia, Immune Network for IgA production, Cytokine receptor interaction, Nucleotide excision repair, p53-signaling, and the cell cycle. A large over-representation of DNA damage associated pathways indicates that a loss of pericentromeric hypomethylation might lead to an increase in DNA damage.

An advantage to using the KEGG method of analysis is the ability to create and render complex pathway diagrams displaying dysregulation (Fig. 15-18). I created a pathway analysis with all relevant proteins for homologous recombination, providing a method to analyze where ICF genotypes can dysregulate a pathway. From this pathway analysis of Homologous Recombination (Fig. 12), one notable cluster of genes is the *rad51*, *rad52*, and associated *rad51* paralogues which play a role in double-strand break repair.

cdca7$^{-/-}$ *vs. wild-type*

The KEGG analysis of the *cdca7* dataset revealed similarly misregulated pathways to *zbtb24*, once again including DNA damage associated pathways such as Homologous Recombination, Fanconi Anemia, Mismatch repair, and p53-signaling. Further, the non-homologous end joining pathway was also enriched in *cdca7* where it was not in *zbtb24*. These DNA damage response pathways were more severely misregulated in *cdca7* than *zbtb24* as evidenced by greater enrichment values.

KEGG Pathway analysis also revealed more severe mis-regulation of *rad51*, *rad52*, and *rad51*-associated paralogues in the *cdca7* dataset (Fig. 16). Furthermore, pathway analysis of Non-homologous end joining (Fig. 17) revealed an associated upregulation of DNA repair subunits

Ku70 and Ku80 as well as a downregulation of the double-strand break repair protein TdT. Pathway analysis of p53-signaling (Fig. 18) revealed a downregulation of cell cycle associated proteins including cyclins as well as a downregulation of the ATM and ATR kinases.

qRT-PCR reveals increased expression of DNA damage response genes is specific to later stages of development

To clarify if increased expression of DNA-damage response genes was also detected at other earlier developmental timepoints, I performed quantitative reverse-transcriptase PCR (qRT-PCR) to assess expression of *rad51* and *rad52* genes at 1-, 2-, 5- and 14-days post fertilization. As a stronger mis-regulation phenotype was seen in the *cdca7* mutants than *zbtb24* mutants (Table 1), I sought to first assay the expression of DNA-damage associated markers in *cdca7* mutants. After primer design and validation (Appendix 2), qRT-PCR was performed comparing *cdca7* maternal zygotic mutants to age-matched wild-type controls. No time points had a significant fold change for *rad51* expression. On the other hand, *rad52*, *rad51c*, and *rad51ap1* data showed the significant upregulation specifically at 2 weeks post fertilization. *rad51b* was significant at every time point analyzed except 5 dpf (Fig. 19).

DISCUSSION

RNA-seq differential expression analysis reveals a misregulation of ICF syndrome related gene-sets and pathways including condensed chromosome, centromeric region, immune effector process, chromosomal region, chromosome organization, and intestinal network for IgA production. Furthermore, there is an upregulation of DNA damage associated genes, gene-sets, and pathways in both *cdca7* and *zbtb24* mutants. Misregulated gene-sets and pathways related to

DNA damage include DNA repair, homologous recombination, Fanconi anemia, p53-signaling pathway, mismatch repair, and non-homologous end-joining.

Upon closer inspection of specific genes in these gene-sets, a particular group of double-strand break repair associated genes (*rad51, rad52, rad51b, rad51c, rad51d,* and *rad51ap1*) are upregulated in RNA-seq analyses of both *cdca7* and *zbtb24* mutants. These misexpression phenotypes are stronger in *cdca7* mutants than *zbtb24* mutants. At first, these results were unexpected as Zbtb24 is required for *cdca7* expression, therefore we would expect that gene expression changes in *cdca7* mutants would also be observed in *zbtb24* mutants. We have identified two potential explanations for this discrepancy.

First, we find that there is greater loss of methylation in *cdca7* mutants compared to *zbtb24* mutants at 2 weeks post fertilization (the same timepoint as both RNA-seq experiments described in this paper). This is due to the fact that *cdca7* mutants exhibit an earlier loss of methylation compared to *zbtb24* mutants (Fig. 4B). Thus, at 2 wpf, *cdca7* mutants have experienced hypomethylation for a longer period of time, potentially leading to a stronger mis-regulation of DNA-damage response genes.

Alternatively, while functional Cdca7 is completely abolished in *cdca7* mutant animals, in *zbtb24* mutants, we still detect low levels of residual *cdca7* expression. It is possible that residual *cdca7* is capable of providing some of the required activity.

qRT-PCR shows a mis-regulation of *rad52*, *rad51c*, and *rad51ap1* at 2 weeks post fertilization and no significance at 1 dpf, 2 dpf, and 5 dpf. *rad51b* is significant at every time point analyzed except 5 dpf. These analyses were performed in *cdca7* maternal zygotic mutants which lose methylation as early as 6 hpf (Fig. 4C). Therefore, it is intriguing that changes in DNA damage associated gene expression are not observed in earlier timepoints where methylation loss is already

present. We hypothesize that this might be caused by an “accumulation” of DNA damage rather than a sudden increase in DNA damage signatures.

Several previous reports suggest that the CDCA7/HELLS chromatin remodeling complex may facilitate non-homologous end-joining (NHEJ) mediated repair. Defective NHEJ during immunoglobulin class-switch recombination has also been observed in *Zbtb24* mutants (Helfricht et al., 2020). One intriguing possibility is that diminished non-homologous end-joining may force DNA repair to occur more extensively through homologous recombination-based repair pathways. Given that homologous recombination is often involved branch migration and DNA synthesis outside of replication, DNA-repair mediated through this pathway may lead to loss of DNA methylation at repaired sites. This could be one potential cause of the hypomethylation observed in ICF syndrome.

CHAPTER 2: FIGURES

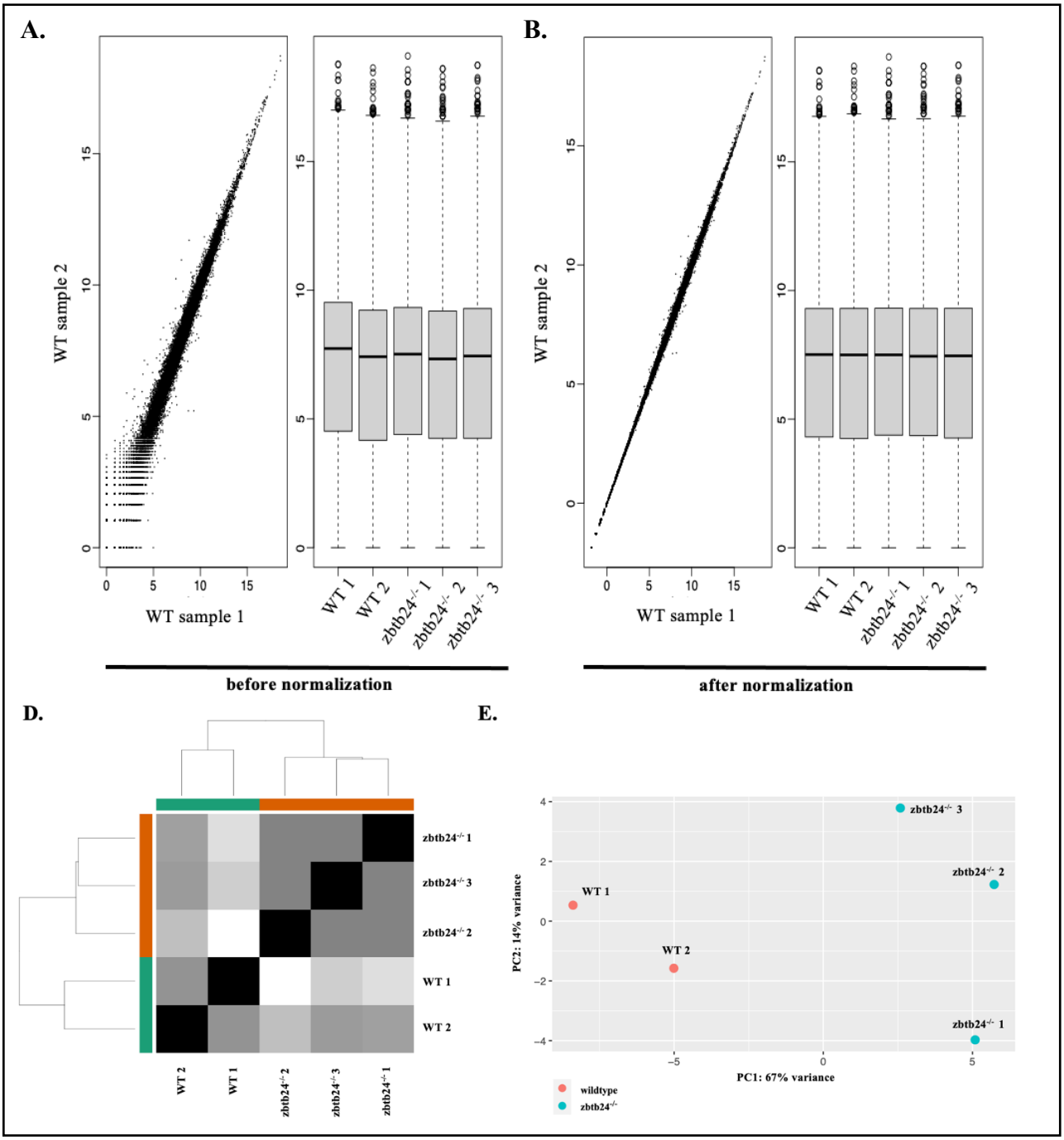


Figure 5: *zbtb24*$^{-/-}$ RNA-seq normalization, stabilization, and clustering. (A) Scatterplot of log normalized counts for both wild-type controls and boxplot of all five samples before normalization and variance stabilization. (B) Scatterplot of log normalized counts for both wild-type controls and boxplot of all five samples after normalization and variance stabilization. (C) Heatmap of all five RNA-seq samples with an overlayed cluster dendrogram tree. (D) Principal Components Analysis of WT and *zbtb24*$^{-/-}$ samples. Plotted on the x-axis is PC1 holding 67% of the variance and on the y-axis is PC2 holding 14% of the variance.

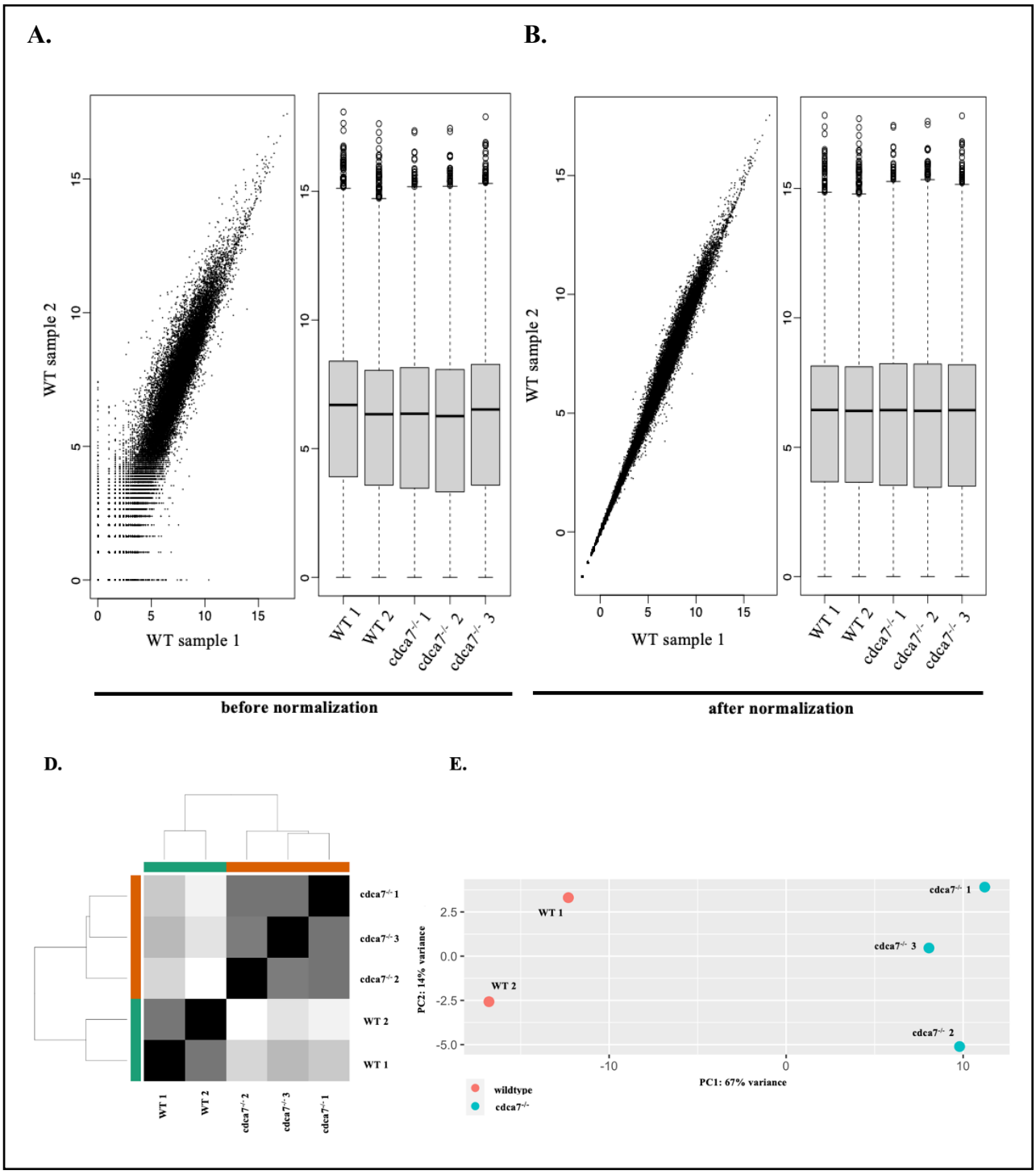


Figure 6: *cdca7*$^{-/-}$ RNA-seq validation, normalization, and clustering. (A) Scatterplot of log normalized counts for both wild-type controls and boxplot of all five samples before normalization and variance stabilization. (B) Scatterplot of log normalized counts for both wild-type controls and boxplot of all five samples after normalization and variance stabilization. (C) Heatmap of all five RNA-seq samples with an overlayed cluster dendrogram tree. (D) Principal Components Analysis of WT and *cdca7*$^{-/-}$ samples. Plotted on the x-axis is PC1 holding 67% of the variance and on the y-axis is PC2 holding 14% of the variance.

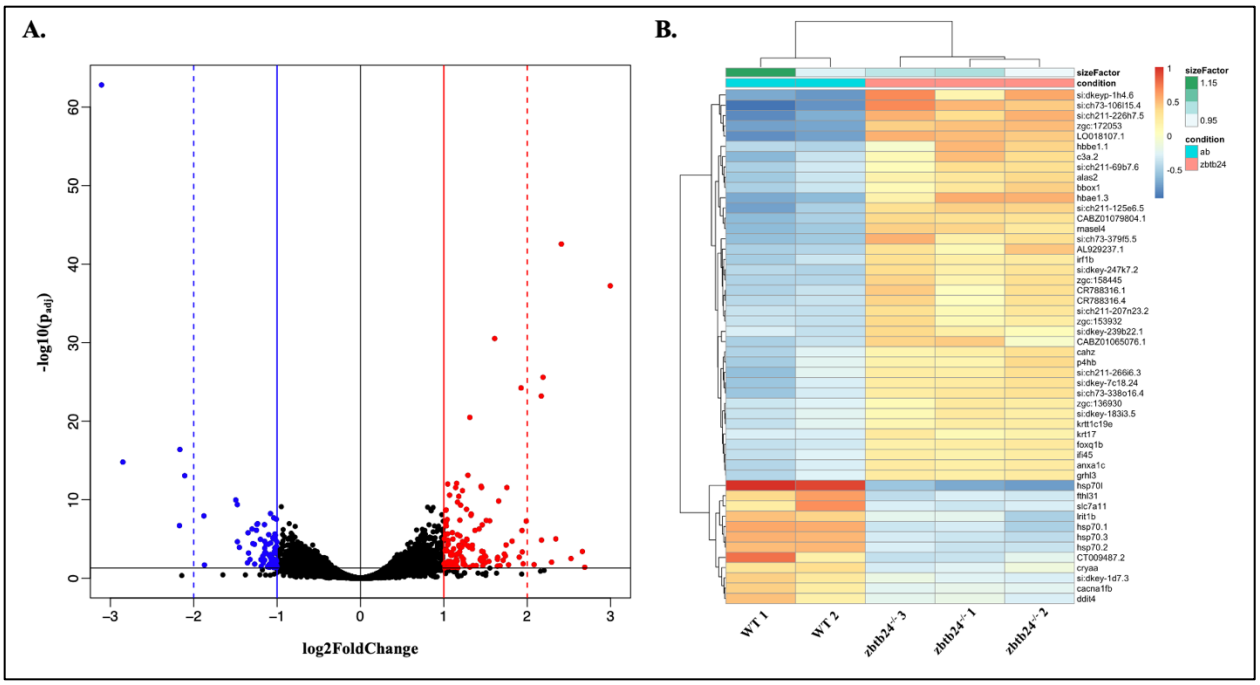


Figure 7: Top differentially expressed genes in *zbtb24*$^{-/-}$ mutants. (A) Volcano plot of differentially expressed genes. Highlighted in blue are under-expressed genes with a log2fc<-1 and p-value<0.05. Highlighted in red are over-expressed genes with a log2fc>1 and p-value<0.05. Solid color bars delineate genes with a log2fold change >1 and <-1. Dotted color bars delineate genes with a log2fold change >2 and <-2. (B) Heatmap representing the top fifty differentially expressed genes between wildtype samples (first two columns) and mutant samples (last three columns). Shown are Z-score normalized gene expression values. Size factor is represented in green. The heatmap is scaled from +1 (red) to -1 (blue).

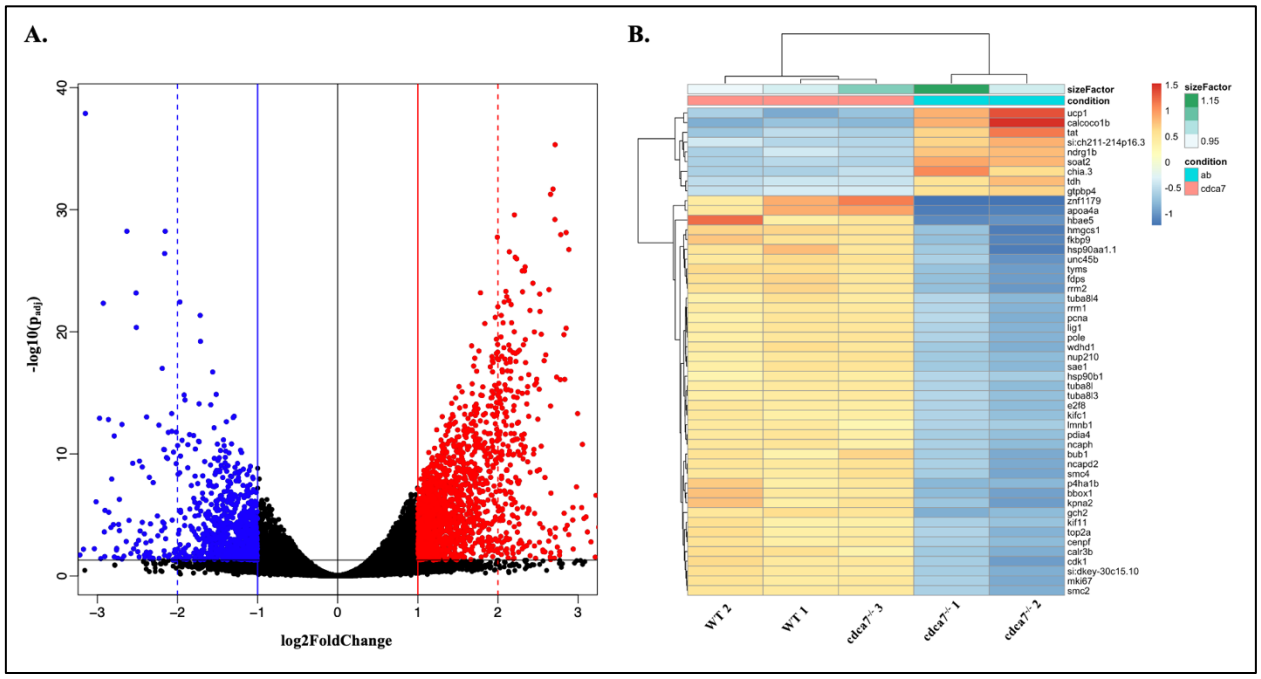


Figure 8: Top differentially expressed genes in *cdca7*$^{-/-}$ mutants. (A) Volcano plot of differentially expressed genes. Highlighted in blue are under-expressed genes with a log2fc<-1 and p-value<0.05. Highlighted in red are over-expressed genes with a log2fc>1 and p-value<0.05. Solid color bars delineate genes with a log2fold change >1 and <-1. Dotted color bars delineate genes with a log2fold change >2 and <-2. (B) Heatmap representing the top fifty differentially expressed genes between wildtype samples (first two columns) and mutant samples (last three columns). Shown are Z-score normalized gene expression values. Size factor is represented in green. The heatmap is scaled from +1 (red) to -1 (blue).

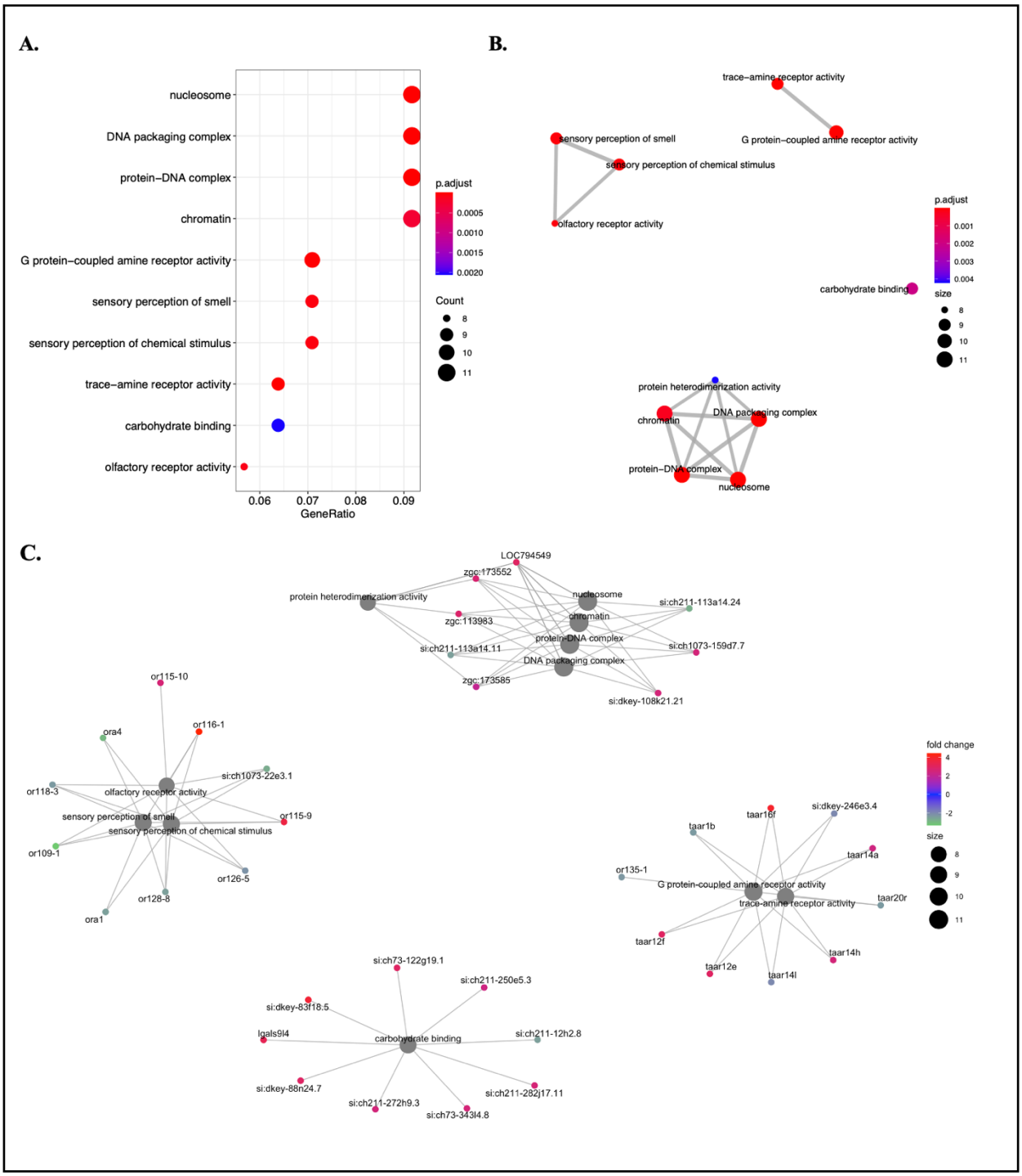


Figure 9: Gene-Ontology (GO) enrichment analysis in *zbtb24*$^{-/-}$ mutants. (A) Dot plot displaying the top 10 enriched GO-terms from an ontology analysis. Color displays adjusted p-value, and the size of the dot is the number of genes contained within that term (B) Connectivity map displaying the clustering of GO-terms from panel A. Color displays adjusted p-value, and the size of the dot is the number of genes contained within that term. (C) Connectivity map with gene web displaying the intersections between GO-terms. Color displays fold change, and the size of the dot is the number of genes contained within that term.

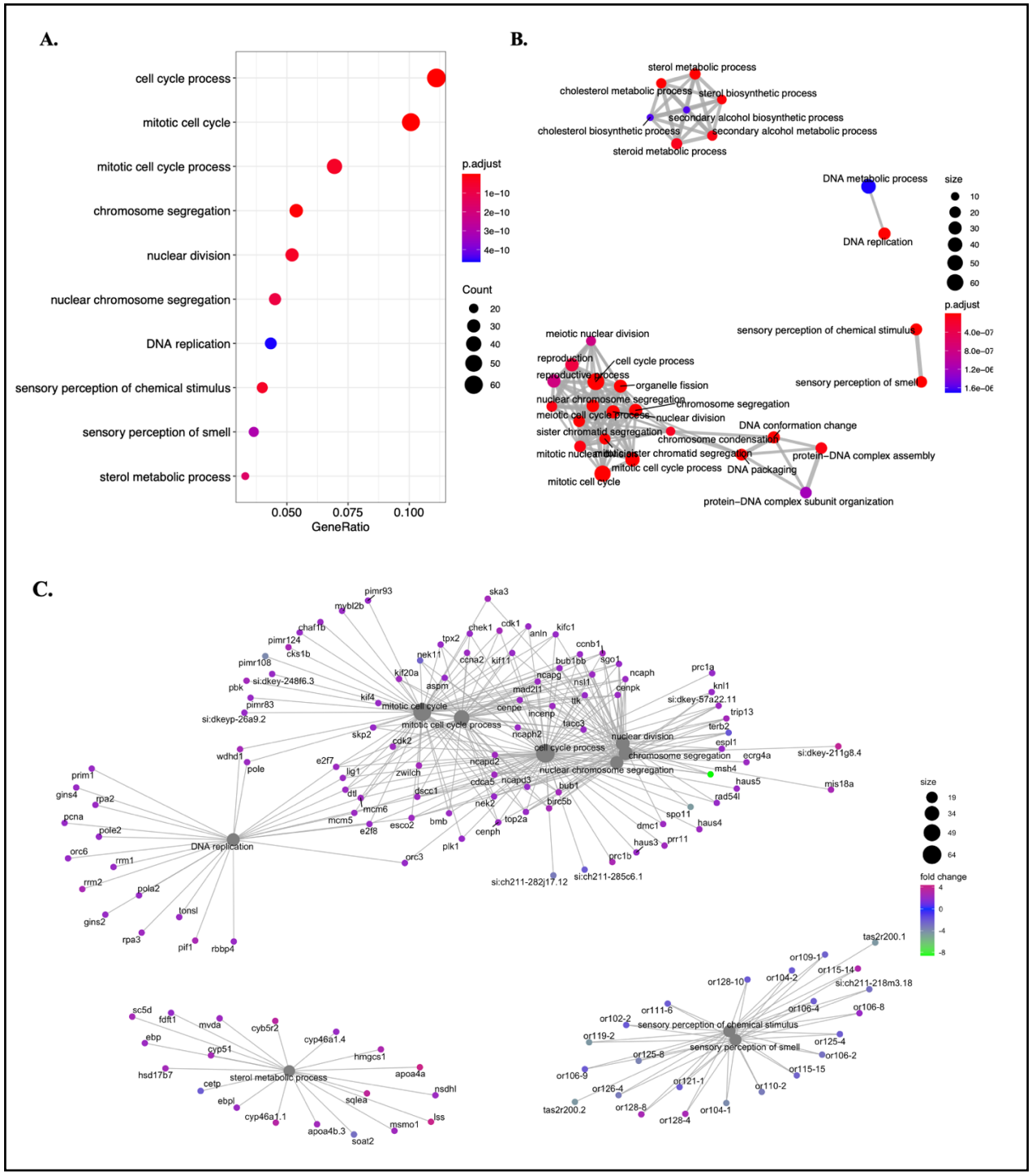


Figure 10: Gene-Ontology (GO) enrichment analysis in *cdca7*$^{-/-}$ mutants. (A) Dot plot displaying the top 10 enriched GO-terms from an ontology analysis. Color displays adjusted p-value, and the size of the dot is the number of genes contained within that term (B) Connectivity map displaying the clustering of GO-terms from panel A. Color displays adjusted p-value, and the size of the dot is the number of genes contained within that term. (C) Connectivity map with gene web displaying the intersections between GO-terms. Color displays fold change, and the size of the dot is the number of genes contained within that term.

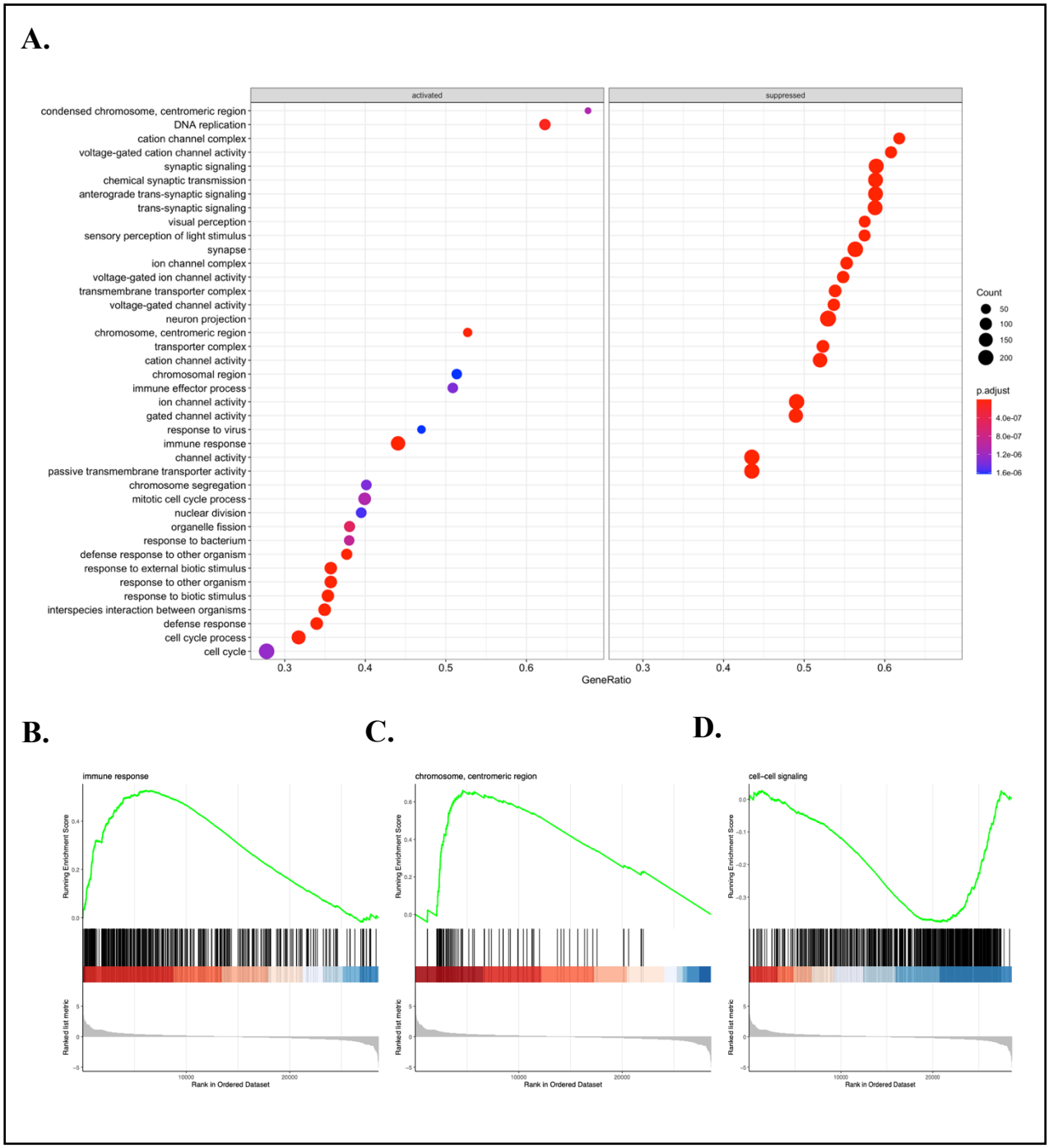


Figure 11: GSEA and plots of ICF related gene-sets in *zbtb24*$^{-/-}$ mutants. (A) Dot plot of the top 50 activated and suppressed gene-sets from a GSEA analysis. Color displays adjusted p-value, and the size of each dot is the number of genes contained within that term. (B-D) GSEA of immune response, chromosome/centromeric region, and cell-cell signaling, respectively.

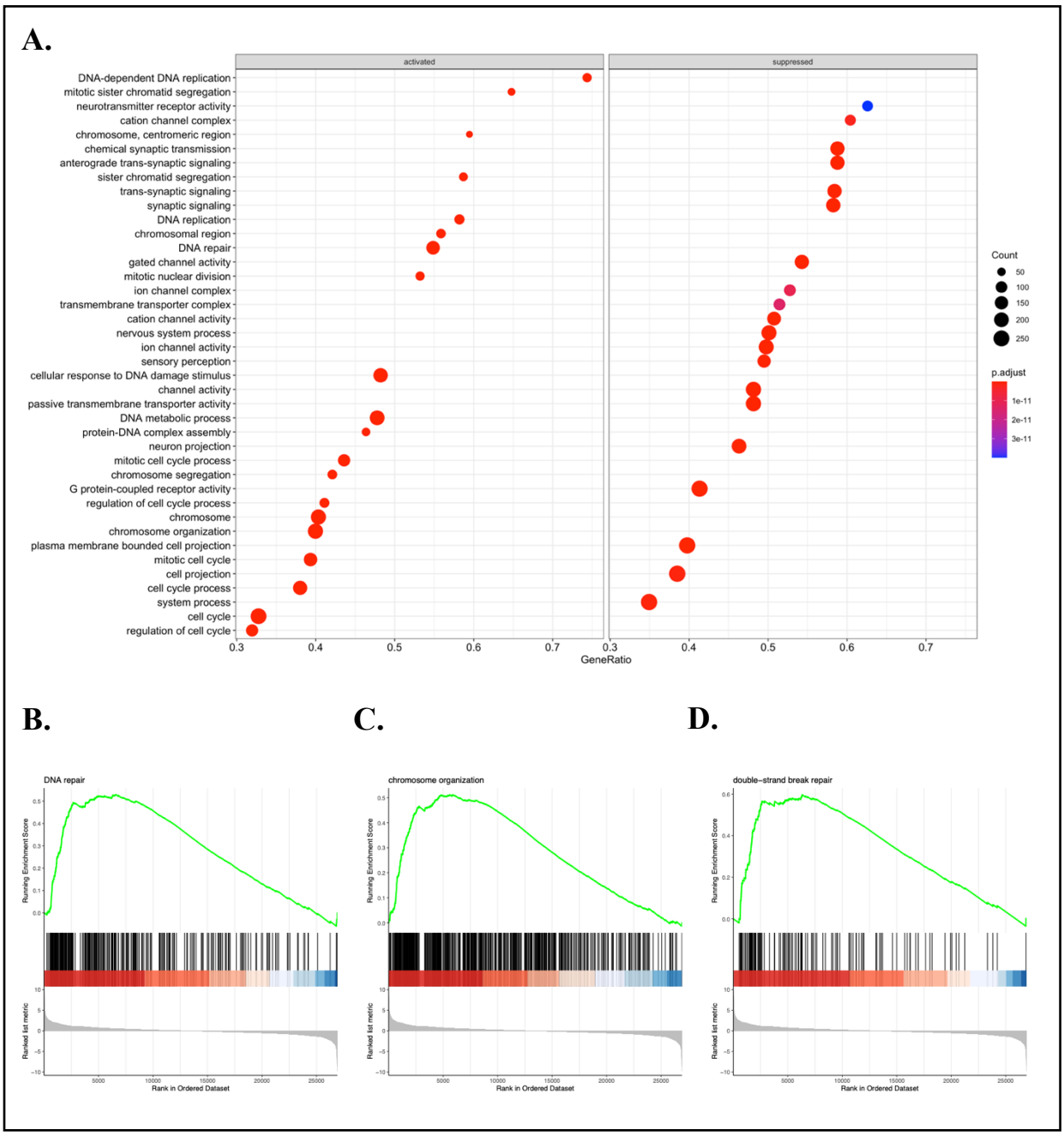


Figure 12: GSEA and plots of ICF related gene-sets in *cdca7*$^{-/-}$ mutants. (A) Dot plot of the top 50 activated and suppressed gene-sets from a GSEA analysis. Color displays adjusted p-value, and the size of each dot is the number of genes contained within that term. (B-D) GSEA of DNA repair, chromosome organization, and double-strand break repair, respectively.

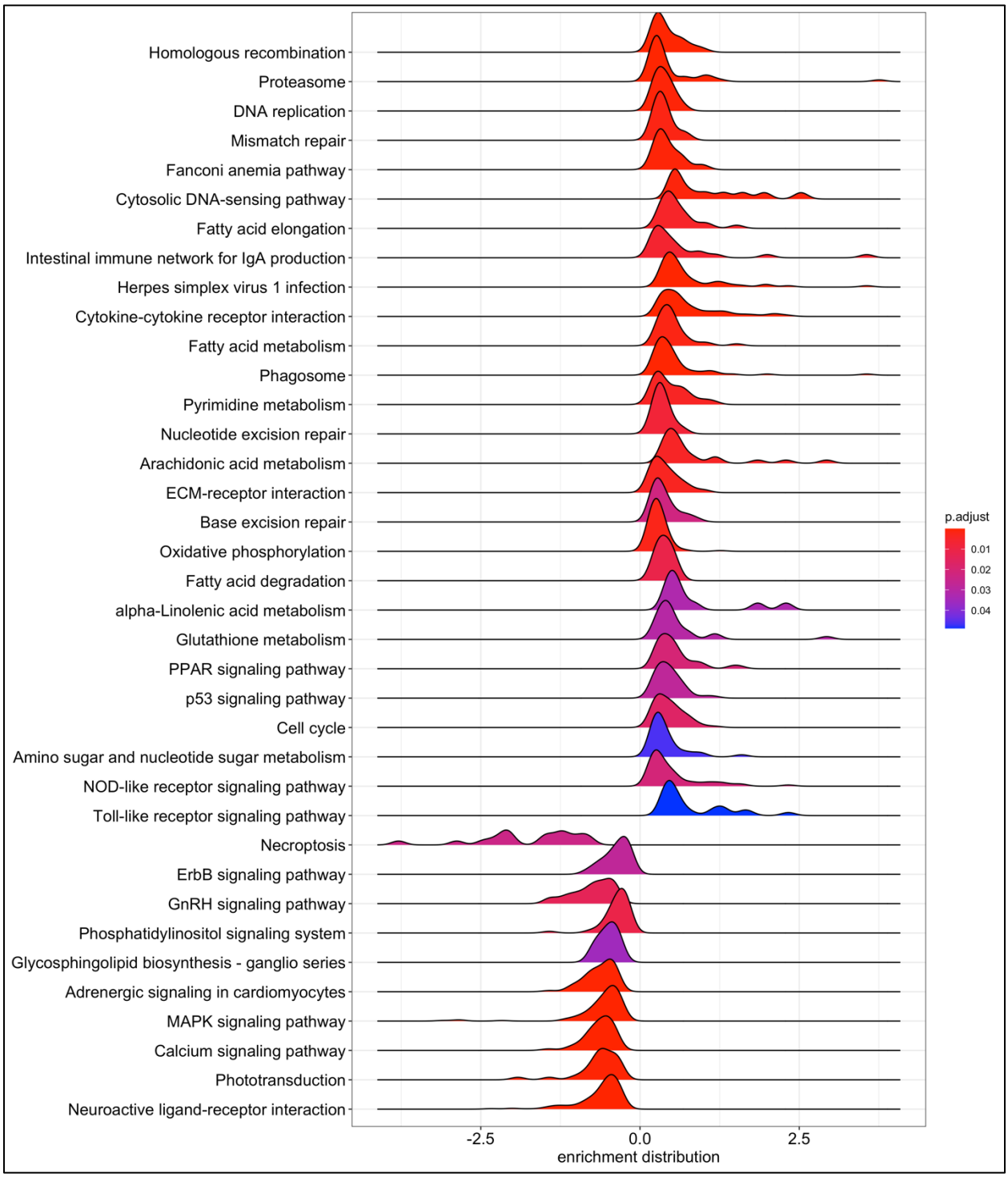


Figure 13: KEGG analysis of top dysregulated pathways in *zbtb24*$^{-/-}$ mutants. Ridge plot representing the top 50 upregulated and suppressed pathways in a KEGG pathway analysis. Enrichment scores are plotted on the y-axis. Color displays adjusted p-value.

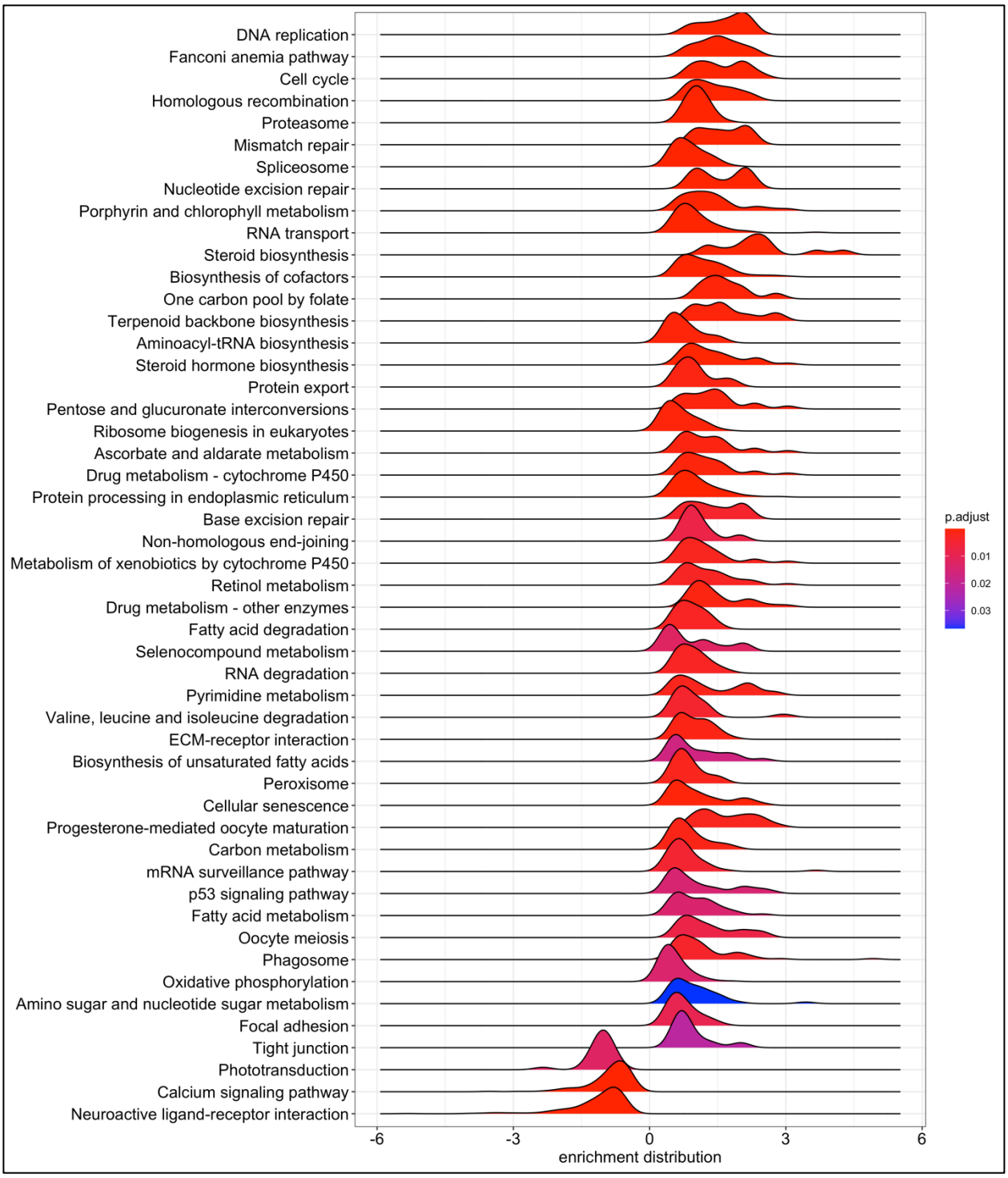


Figure 14: KEGG analysis of top dysregulated pathways in *cdca7*$^{-/-}$ mutants. Ridge plot representing the top 50 upregulated and suppressed pathways in a KEGG pathway analysis. Enrichment scores are plotted on the y-axis. Color displays adjusted p-value.

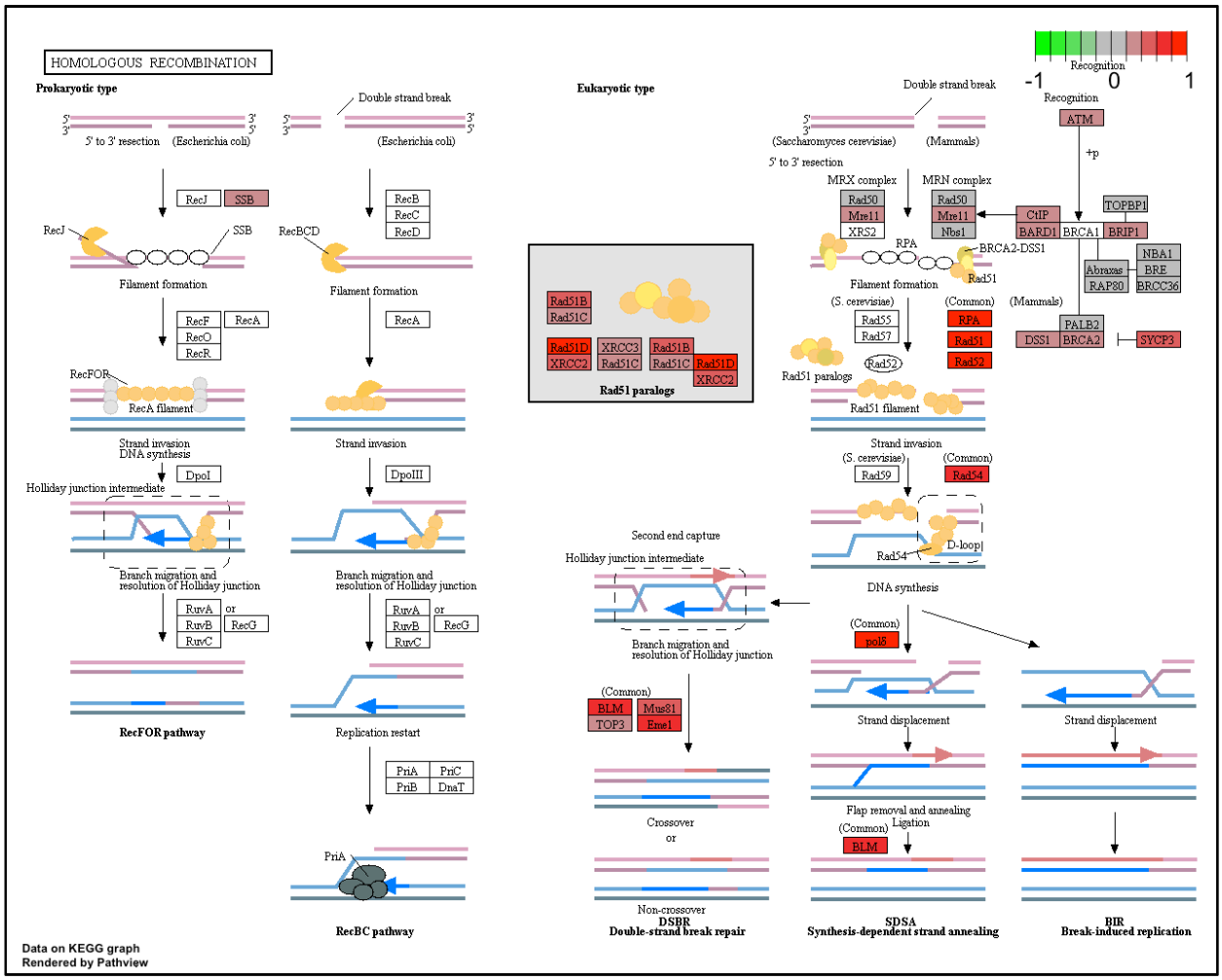


Figure 15: Misregulation of genes in the homologous recombination pathway in *zbtb24*$^{-/-}$ mutants. KEGG Illustrative pathway of Homologous Recombination overlayed with differentially expressed genes. Red represents over-expressed genes; green represents under-expressed genes. Rendered using Pathview in R.

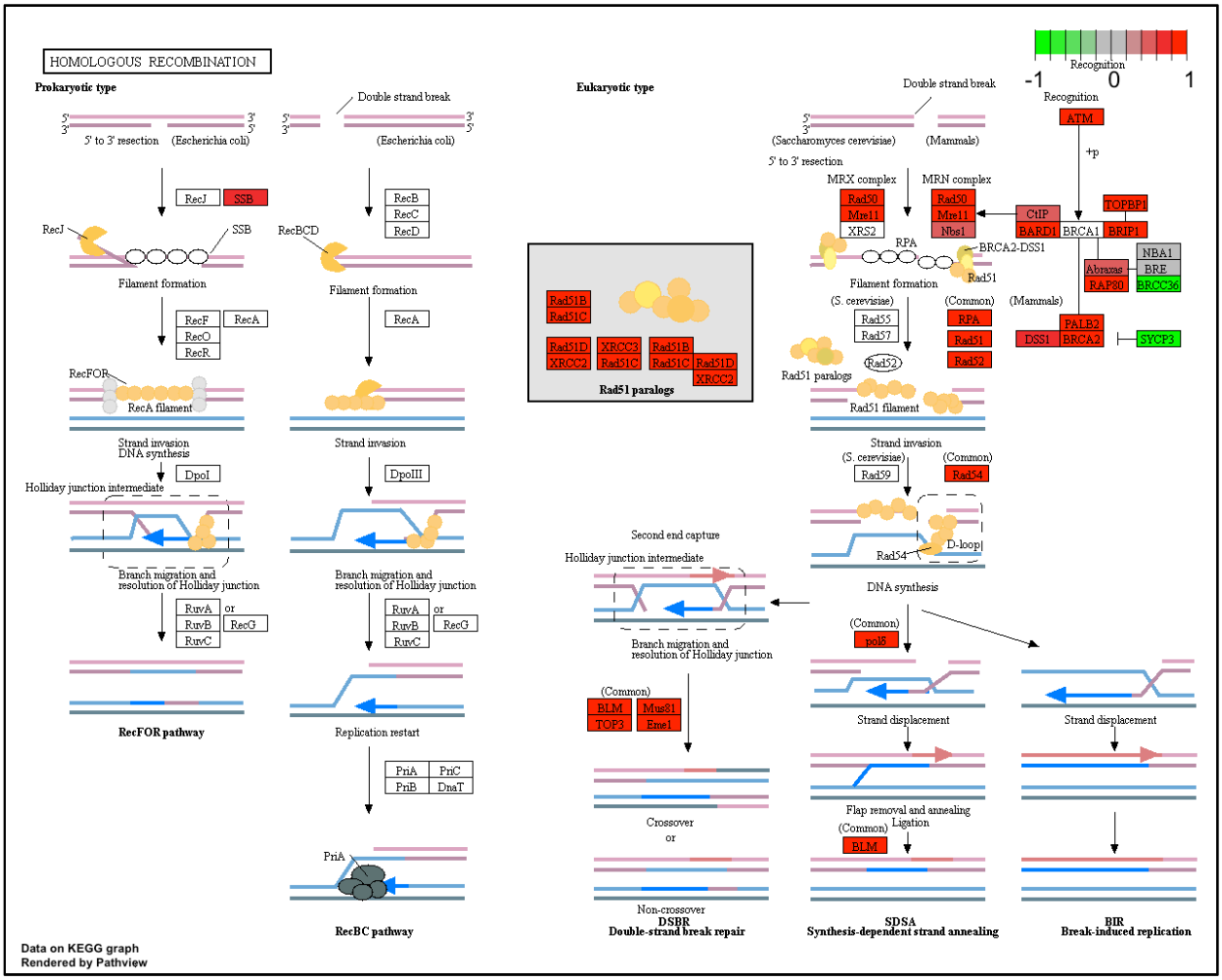


Figure 16: Misregulation of genes in the homologous recombination pathway in *cdca7*$^{-/-}$ mutants. KEGG Illustrative pathway of Homologous Recombination overlayed with differentially expressed genes. Red represents over-expressed genes; green represents under-expressed genes. Rendered using Pathview in R.

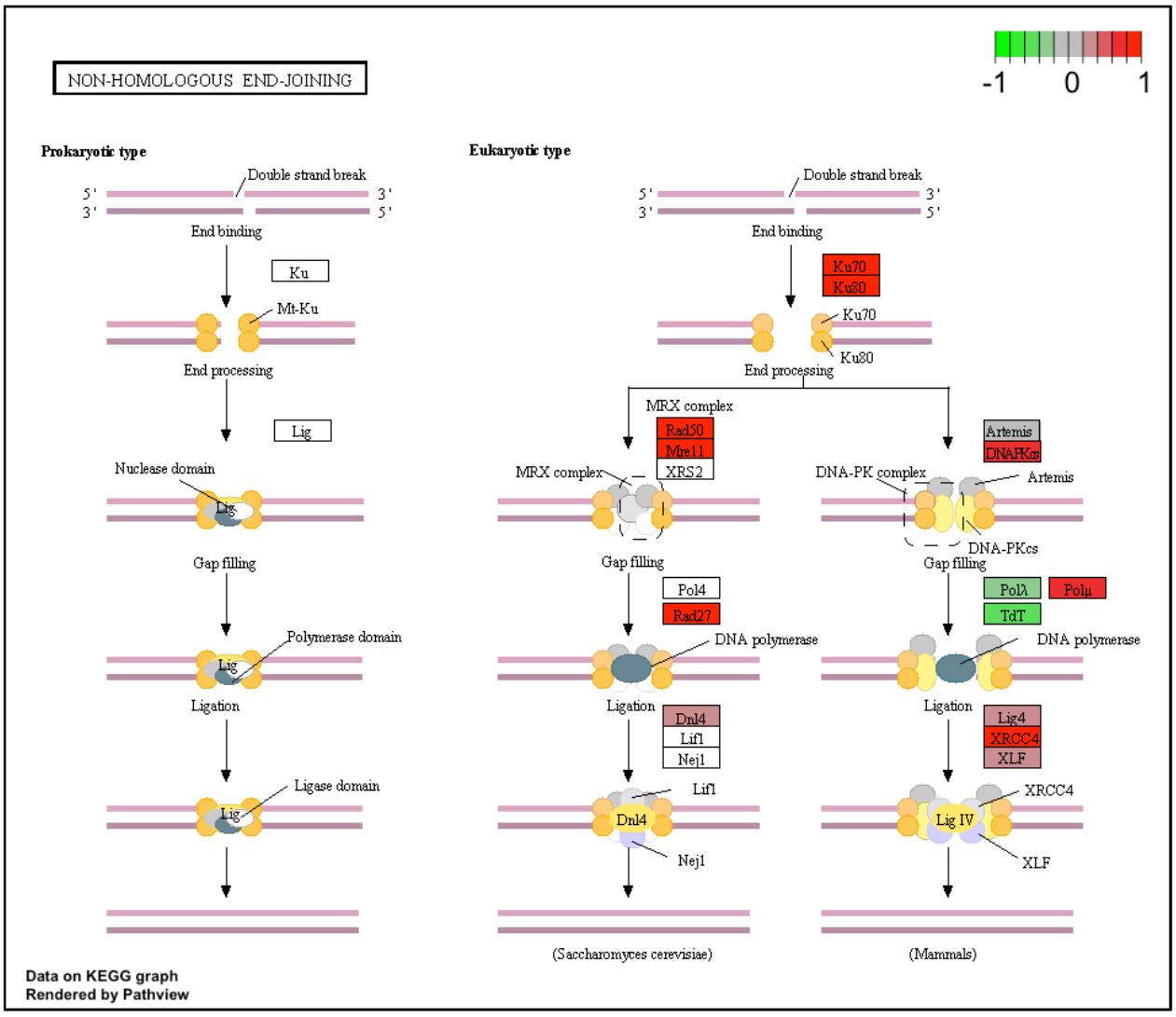


Figure 17: Misregulation of genes in the non-homologous end-joining pathway in *cdca7*$^{-/-}$ mutants. KEGG Illustrative pathway of Non-Homologous End-Joining overlayed with differentially expressed genes. Red represents over-expressed genes; green represents under-expressed genes. Rendered using Pathview in R.

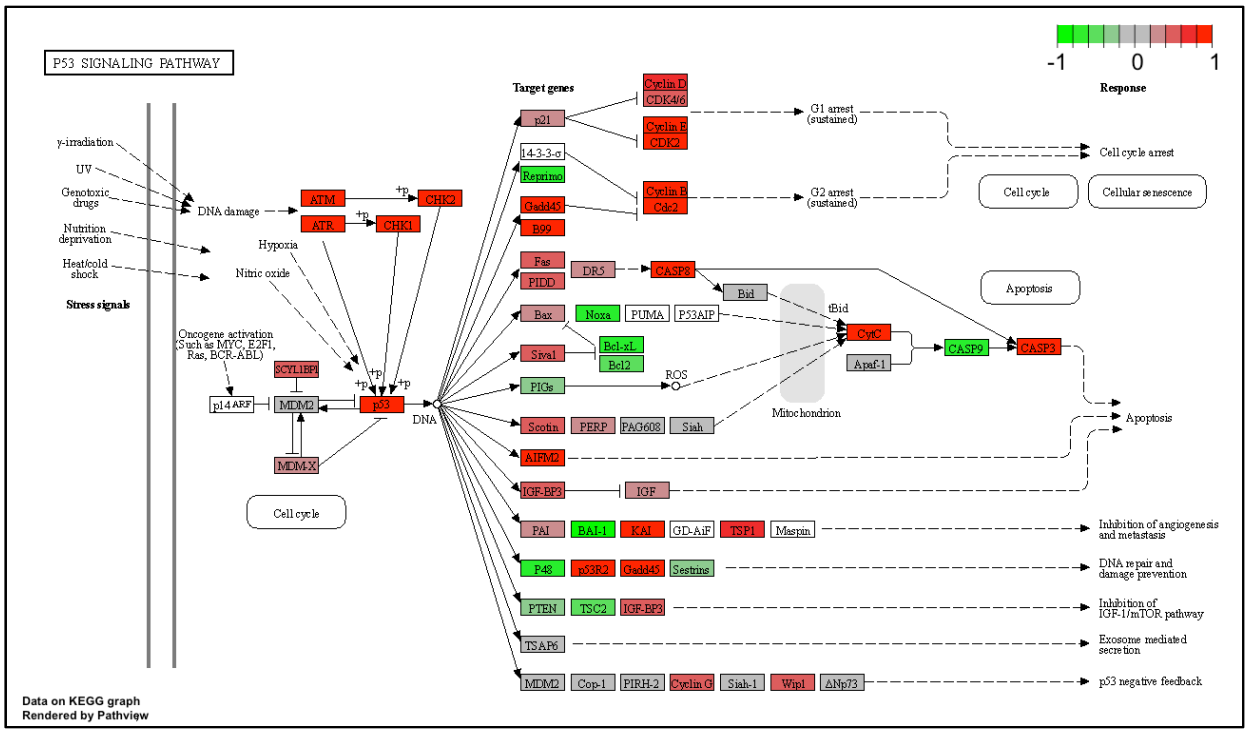


Figure 18: Misregulation of genes in the p53-signaling pathway in *cdca7*$^{-/-}$ mutants. KEGG Illustrative pathway of p53-signaling overlayed with differentially expressed genes. Red represents over-expressed genes; green represents under-expressed genes. Rendered using Pathview in R.

Table 1: Upregulation of DNA damage associated genes in both *zbtb24*$^{-/-}$ and *cdca7*$^{-/-}$ mutants. Differential expression analysis of Rad51, Rad52, and Rad51ap1, Rad51b, Rad51c, and Rad51d in (A) *zbtb24*$^{-/-}$ mutants and (B) *cdca7*$^{-/-}$ mutants. Log2FoldChange is highlighted in yellow. P-value statistics are highlighted in green. Blue indicated normalized counts for wildtype samples, and pink indicated normalized counts for mutant samples.

A.

Gene	log2FoldChange	lfcSE	stat	pvalue	padj	WT 1	WT 2	zbtb24-/- 1	zbtb24-/- 2	zbtb24-/- 3
rad51	**0.990287748**	0.20796875	4.76171415	1.92E-06	0.00016471	134.325457	196.683954	327.276168	359.411001	295.595343
rad52	**0.834796786**	0.27337272	3.05369457	0.00226042	0.02764258	57.3236025	78.2506054	113.706222	129.907591	117.428287
rad51ap1	**0.864739953**	0.24216576	3.57085964	0.00035581	0.00792181	88.9799203	108.916383	162.15496	214.347525	162.982364
rad51b	**0.454178787**	0.38366663	1.18378497	0.23649819	0.49856138	34.2230463	33.8380997	43.5049892	51.9630363	44.541764
rad51c	**0.302572324**	0.30816102	0.98186436	0.32616668	0.5886773	58.1791787	75.0782836	88.9874778	92.0178767	64.7880203
rad51d	**0.879021883**	0.22487321	3.9089667	9.27E-05	0.00295152	108.658172	112.088705	178.963705	233.833663	196.388687

B.

Gene	log2FoldChange	lfcSE	stat	pvalue	padj	WT 1	WT 2	cdca7-/- 1	cdca7-/- 2	cdca7-/- 3
rad51	**1.237547624**	0.23295629	-5.3123598	1.08E-07	2.38E-06	181.676167	154.912929	475.773257	340.703239	375.858896
rad52	**1.803544809**	0.40391468	-4.4651628	8.00E-06	0.00010422	19.167669	23.0275976	70.6802842	77.1819634	72.1724065
rad51ap1	**3.053972889**	0.41549295	-7.3502399	1.98E-13	1.50E-11	15.0007844	13.6072168	114.987328	152.158728	90.9184861
rad51b	**2.264826855**	0.3620337	-6.2558454	3.95E-10	1.48E-08	25.0013073	20.9341796	107.602821	105.84955	119.037606
rad51c	**1.049317281**	0.52913067	-1.9830967	0.04735663	0.13323365	10.8338998	24.0743066	32.7028181	38.5909817	35.6175513
rad51d	**1.306719703**	0.4112989	-3.1770561	0.00148778	0.00884399	31.6683226	19.8874706	60.1309881	67.2585681	65.6112787

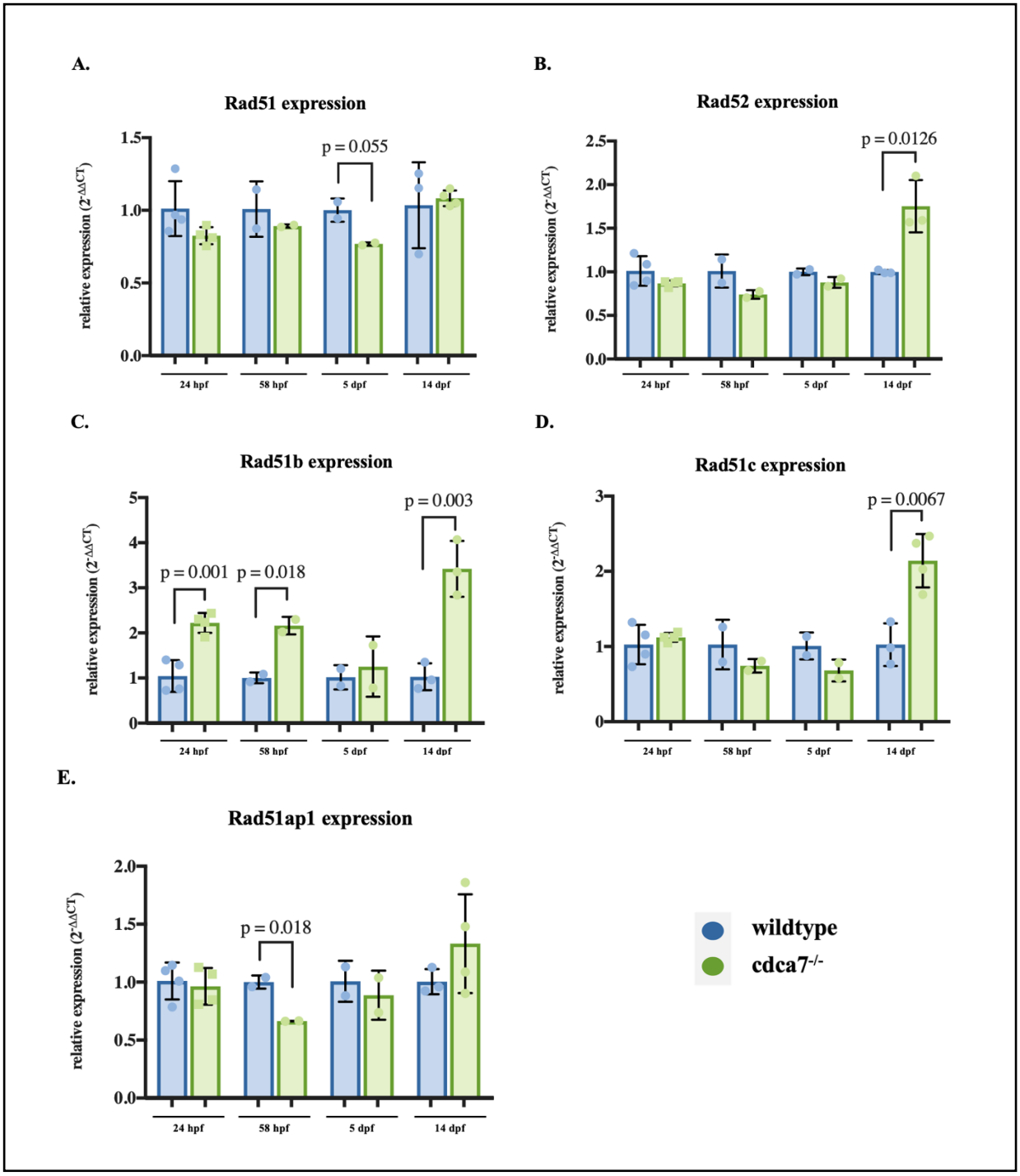


Figure 19: qRT-PCR assay of DNA damage associated gene expression in *cdca7*$^{-/-}$ mutants. qRT-PCR data of Rad51, Rad52, and Rad51ap1, Rad51b, and Rad51c expression in *cdca7*$^{-/-}$ mutants compared to WT controls. qRT-PCR was performed at 24 hpf, 58 hpf, 5 dpf, and 14 dpf. Relative expression is quantified by $-2^{\Delta\Delta CT}$. Blue data points are wild-type and green data points are *cdca7*$^{-/-}$ mutants.

CHAPTER 3: INCREASED DNA DAMAGE SUSCEPTIBILITY IN CDCA7$^{-/-}$ MUTANT ZEBRAFISH

PREFACE

Previous literature has associated global hypomethylation with DNA damage and genomic instability (Kim et al, 2004; Palii et al, 2008). These genomic changes may be responsible for promoting the onset of malignancy. However, there is a lack of experimental data addressing if DNA damage can result from a specific loss of methylation at the pericentromeres *in vivo*. To address this, our lab leveraged our viable zebrafish models of ICF syndrome with a mutation in *cdca7*. We hypothesized that a loss of pericentromeric methylation through a mutation in *cdca7* will lead to increased DNA damage. This could potentially arise directly from a loss of DNA methylation at pericentromeric sequences or alternatively through functions of Cdca7 in promoting DNA repair through pathways such as nonhomologous end joining. In addition, we were interested in testing if *cdca7* mutant zebrafish were more susceptible to DNA damaging agents.

EXPIREMENTAL DESIGN

Hydroxyurea Assay for Phenotypic Abnormalities

Wildtype and *cdca7*$^{-/-}$ MZ mutant zebrafish were exposed to Hydroxyurea dissolved into their fish water. Hydroxyurea is an antimetabolite which acts as a ribonucleotide reductase inhibitor, effectively reducing the available pool of deoxyribonucleotides available for DNA synthesis. This would cause the cell to enter S-phase, replicative stress, and thus undergo DNA damage. Further, DNA modifications after hydroxyurea exposure are primarily through base oxidation and secondarily by depurination (Sakano et al., 2001). Hydroxyurea has been previously in humans used as a chemotherapeutic agent in the treatment of cancers such as chronic myeloid

leukemia and advanced squamous cell carcinoma (Madaan et al., 2012). In zebrafish, hydroxyurea has been shown to inhibit the cell cycle by interfering with DNA replication forks and potentially activating apoptotic programs (Verduzco et al., 2011). Because of our initial hypothesis that pericentromeric methylation loss and DNA damage may be linked to replicative stress, I chose hydroxyurea as my default DNA damaging agent for the rest of the work described in this thesis.

Before we could conduct full-scale experiments with hydroxyurea, we had to first identify working doses with adequate phenotype-to-survival ratios. This work was originally conducted by Danielle Obiri who treated wild-type animals with hydroxyurea concentrations between 25 mM and 200 mM. The fish were exposed to hydroxyurea at 24 hpf and observed for phenotypic abnormalities and developmental issues over the course of 144 hours. As expected, wildtype fish exposed to higher levels of hydroxyurea experienced more severe phenotypes including cardiac edema, bent body, failure to inflate the swim bladder, pigmentation defects, late de-chorionation, and death (Figure 21). Fish exposed to 100 mM and 200 mM hydroxyurea levels experienced significant levels of death. However, fish exposed to 50 mM hydroxyurea survived through the course of analysis and displayed the range of observable phenotypes. Therefore, we set our upper threshold for all future analyses at 50 mM.

Immunofluorescence for markers of DNA damage

Additionally, were interested in assessing changes at the molecular level as a result of the cdca7 mutation and in response to DNA damaging agents. In particular, we were interested in two markers of DNA damage: (1) γH2AX and (2) RNA:DNA hybrids.

(1) Although there are many forms of DNA damage, one of the most severe is the formation of double-strand breaks. Double-strand breaks are followed by the phosphorylation of the histone variant H2AX. H2AX is a variant of the H2A histone octamer in nucleosomes. As a result of DNA

damage, this histone variant is typically phosphorylated by kinases including ataxia telangiectasia mutated (ATM) and ATM-Rad3-related (ATR) as part of the PI3K signaling pathway (Kuo and Yang, 2008). Once phosphorylated, the histone variant gamma-H2AX plays a role in recruiting and localizing DNA repair proteins to the site of double-strand breaks. We chose γH2AX as the first proxy to measure DNA damage in our immunofluorescence experiments as there is a commercially available and widely tested antibody (Genetex, GTX127342).

(2) The second structure we assessed was RNA:DNA hybrids (R-loops). R-loops consist of a three-stranded nucleic acid structure containing an RNA:DNA hybrid and a displaced ssDNA strand. The exact roles of RNA:DNA hybrids currently remain uncertain. RNA:DNA hybrids are not typically present across the genome and have been suggested to play a role in a range of human diseases (Nadel et al., 2015). RNA:DNA hybrids may form when a nascent transcript inappropriately rehybridizes with its DNA behind the transcription complex. When R-loops are formed, there is subsequently a higher chance to form double-stranded breaks due to stalled replication forks. On the contrary, DNA damage, in particular double-strand break formation, itself may cause the formation of RNA:DNA hybrids (Bader and Bushell, 2020). It has been suggested that these RNA:DNA hybrids play a role in marking sites of DNA damage or that they act as templates for DNA repair pathways (Lu et al., 2018).

It has been suggested that RNA:DNA hybrids are enriched at loci with decreased DNA methylation, which would be of particular interest to the pericentromeric methylation phenotype in ICF syndrome. Further, there has been literature suggested that Cdca7 provides a protective role in preventing R-loop formation at pericentromeric repeats (Unoki et al., 2020). To test if RNA:DNA hybrids are formed *in vivo*, we conducted immunofluorescence in our *cdca7* mutant and wildtype zebrafish. For each of these immunofluorescence experiments, *cdca7* MZ and AB

fish were pooled from multiple separate clutches. Fish were treated with Propylthiouracil (PTU), a commonly used pigment inhibitor in zebrafish, at 6 hpf. Fish were subsequently treated with Hydroxyurea at 12 hours post fertilization (Figure 20).

RESULTS

Hydroxyurea Assay for DNA damage susceptibility

To assess DNA damage sensitivity, we treated dishes of embryos pooled from crosses between wild-type adults or *cdca7*$^{-/-}$ maternal zygotic mutants with hydroxyurea at four conditions: no treatment, 10mM, 25mM, and 50mM. Fish were exposed to hydroxyurea at 12 hpf and monitored at 24 hpf, 48 hpf, 72 hpf, and 96 hpf for evidence of phenotypic and developmental bnormalities (Figure 20).

The first round of hydroxyurea experiments was conducted by a previous undergraduate in the laboratory, Danielle Obiri. She observed that nearly all of fish of both genotypes across all treatments appeared normal at 24 hpf and 48 hpf (Figure 22 and Figure 24). However, by 72 hpf she found increased levels of death, body bend, and cardiac edema in hydroxyurea treated *cdca7* mutants when compared to treated wild-type controls (~20% and 100% respectively) (Figure 25 and Figure 26). These results were significant as assayed by a two-way ANOVA ($p<0.0001$) (Figure 22).

Interestingly, when this experiment was repeated by myself and another undergraduate in the lab, Eric Schouten, we obtained conflicting results (Figure 23). I observed that *cdca7* mutant fish showed no distinct changes in development when compared to WT controls at each treatment level. There were no significant differences between *cdca7*$^{-/-}$ and wildtype controls by statistical analysis at the timepoints where significant results were previously obtained. We hypothesize that this may be a result of tank-to-tank differences in sensitivity when embryos were collected.

Alternatively, my embryos were pooled together before hydroxyurea treatment whereas the previous experiment was conduct using individual zebrafish clutches. These experiments need to be repeated with a higher number of embryos to obtain more conclusive results.

Immunofluorescence for markers of DNA damage

Although *cdca7* maternal zygotic zebrafish may or may not show a higher prevalence of phenotypic abnormalities when compared to wildtype counterparts, we wanted to test if they showed any phenotypic changes at the molecular level. First, we aimed to assess when DNA damage can be detected in the developing zebrafish organism. To examine this, I performed immunofluorescence staining for γH2AX in cdca7 and WT zebrafish at 24 hpf, 48 hpf, and 72 hpf. We saw increased levels of γH2AX foci in cdca7 fish compared to wildtype controls at both 48hpf and 72 hpf, however, we did not observe the formation of γH2AX at no treatment conditions in either *cdca7* or wildtype fish. We did observe foci formation at 50 mM conditions across both genotypes. Because we saw the clearest phenotypes both at the developmental level (Figure 25) and molecularly in our 72 hpf samples, all further analyses were conducted at 72 hpf.

Next, we were interested in assessing if there are increased levels of DNA damage in *cdca7* mutants compared to WT fish at baseline (no treatment). I examined *cdca7*$^{-/-}$ and wildtype zebrafish at 72 hpf for formation of γH2AX foci. In our wildtype zebrafish, I observed few sparsely localized γH2AX foci mostly in the head of the organism. In our *cdca7* mutant zebrafish, I observed a marked increase in these foci in the head of the organism (Figure 27). Quantification of these results confirms an increase in double-strand break formation in *cdca7*$^{-/-}$ fish compared to WT fish at 72 hpf (Figure 28). Quantification was performed on lower magnification pictures taken with a widefield fluorescent microscope.

Further, we were interested in examining if there was increased DNA damage susceptibility in *cdca7* mutants after exposure to DNA damaging agents. To perform this experiment, we used Hydroxyurea dissolved in fish water at four concentrations (no treatment, 10 mM, 25 mM, and 50 mM). We saw an increase in γH2AX staining in *cdca7* maternal zygotic fish compared to their respective wildtype controls at each concentration assayed (Figure 27). We observed a significant increase in double-strand break formation in *cdca7*$^{-/-}$ fish compared to WT fish after treatment with 10 mM hydroxyurea (Figure 29). Quantification was performed on lower magnification pictures taken with a widefield fluorescent microscope. This quantification of these results was performed on fish treated with 10 mM hydroxyurea by calculating area-normalized mean fluorescent intensity in ImageJ.

Next, we assessed RNA:DNA hybrid formation in the same treatment conditions above. RNA:DNA hybrids were not observed in either genotype at baseline (no treatment). Interestingly, RNA:DNA hybrids were only observed after high exposure to DNA damaging agents. These foci were localized to the myocardium and yolk sack extension of the organism (Figure 30). *Cdca7*$^{-/-}$ organisms had notably more staining for RNA:DNA hybrids when compared to wildtype zebrafish.

DISCUSSION AND FUTURE DIRECTIONS

Taken together, my results suggest that *cdca7* mutant zebrafish undergo higher levels of DNA damage and experience higher sensitivity to DNA damaging agents than wildtype counterparts. According to the results from my hydroxyurea sensitivity screen, *Cdca7*$^{-/-}$ fish do not experience developmental changes or phenotypic abnormalities compared to wildtype fish when reared in normal fish water or at higher treatment levels. However, these fish do exhibit a significant increase in γH2AX staining, indicating that a loss of Cdca7 function or loss of

pericentromeric DNA methylation may lead to DNA damage at the molecular level. These changes were localized to more highly replicative tissues (e.g., eyes and vertebral column) and might result from increased replicative stress in these regions.

I suggest that *cdca7* mutant zebrafish undergo a greater amount of DNA damage or have a greater difficulty repairing that damage than their wildtype counterparts. The results of this experiment could point to two potential hypotheses. First, a lack of pericentromeric hypomethylation could lead to increased DNA damage susceptibility and a greater formation of double strand breaks. Secondarily, Cdca7 has been proposed to promote proper non-homologous end-joining. Therefore, if Cdca7 function is lost, cells would have a greater difficulty repairing DNA damage. In addition, both of these phenomena could act in concert to influence greater levels of DNA damage in cdca7 mutant zebrafish before and after treatment with hydroxyurea.

Previous literature has suggested that RNA:DNA hybrids can lead to DNA damage in ICF cell lines (Unoki et al., 2020). The relationship between RNA:DNA hybrids and DNA damage was suggested to be causal, where an increase in RNA:DNA hybrids resulting from ICF mutations leads to increased DNA damage. However, we did not observe the formation of RNA:DNA hybrids until the induction of DNA damage by hydroxyurea. This finding suggests that RNA:DNA hybrids are formed as a consequence of DNA damage instead of causing DNA damage themselves, contrary to previous literature. Further, we suggest that increased DNA damage levels in *cdca7*$^{-/-}$ mutants may lead to the increase in RNA:DNA hybrid formation we observe through immunofluorescence staining.

Further directions include determining why DNA damage shows up in specific tissues. We predict DNA damage is localized to more highly replicating tissues, however, further experiments are needed to confirm this. Further, it would be interesting to test if certain genomic sites are more

susceptible to DNA damage. We predict that DNA damage would be localized to pericentromeric repeats, where we see a loss of methylation in our *cdca7* mutants. We could perform ChIP-qPCR or DNA Fluorescence In Situ Hybridization (FISH) to test whether or not γH2AX co-localizes with Sat1 repeats. Since γH2AX is a marker of DNA damage, an increased number of foci at Sat1 repeats, would suggest that these repeat sequences are more susceptible to DNA damage than other chromosomal regions.

In addition, we anecdotally observe that RNA:DNA hybrids largely do not localize to the same tissues as γH2AX in our *cdca7* mutants. DNA damage in the form of γH2AX was largely prevalent in the eyes of the organism, however, RNA:DNA hybrids were mainly localized to the myocardium. We have not yet tested if these two marks co-localize through simultaneous co-staining with both antibodies. Further, we have not yet tested if both of these marks are observed at the same genomic sites (e.g. pericentromeric repeats).

Additionally, it would be intriguing to assess different DNA damaging reagents which cause distinct forms of damage than Hydroxyurea. Using agents such as doxorubicin, a topoisomerase II inhibitor in DNA replication, we could independently confirm our findings. Moving forward, it will be important to further clarify which pathways of DNA damage and repair are affected by pericentromeric methylation loss.

CHAPTER 3: FIGURES

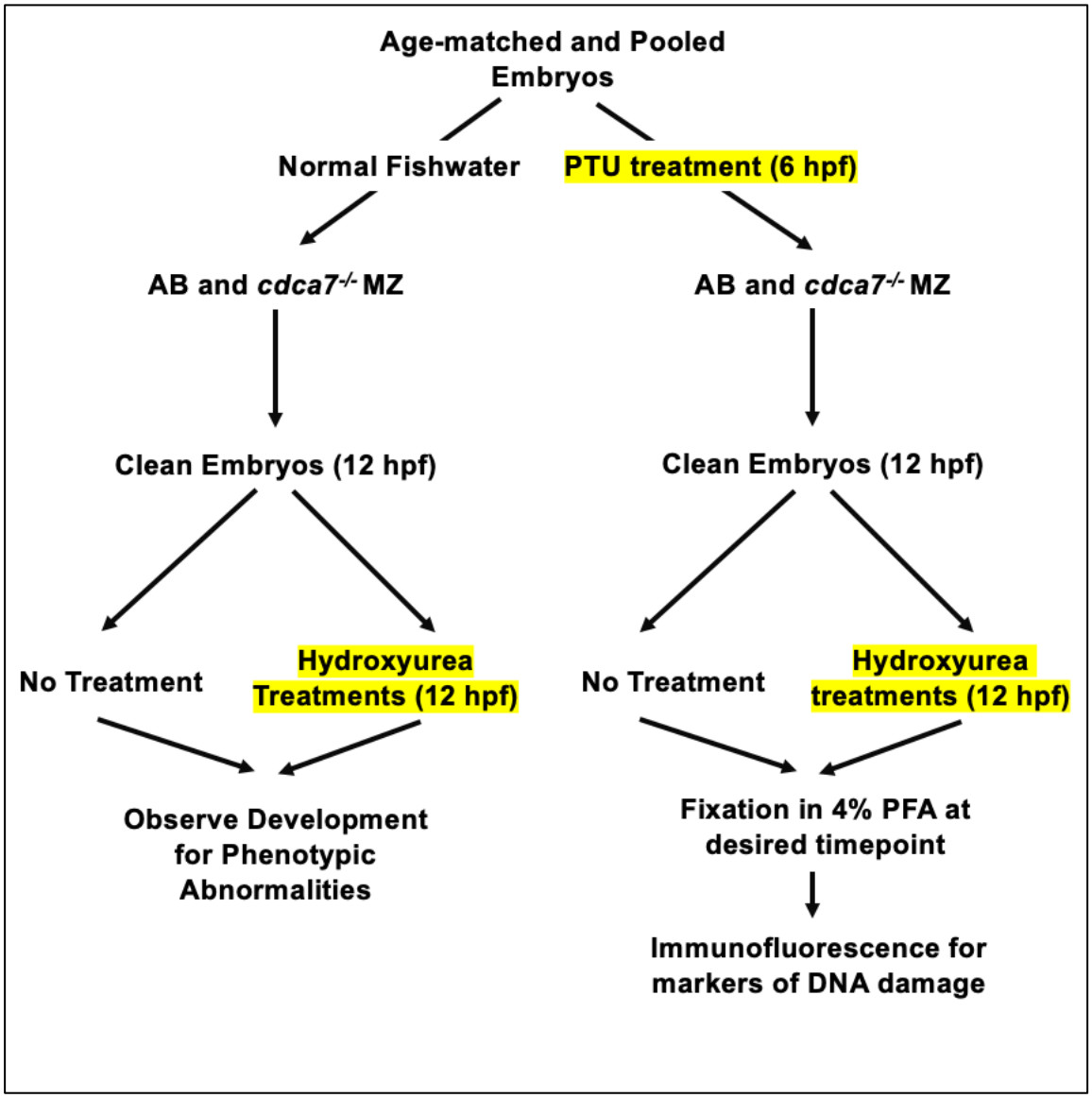


Figure 20: Experimental design of Hydroxyurea and Immunofluorescence Screens for developmental abnormalities and DNA damage.

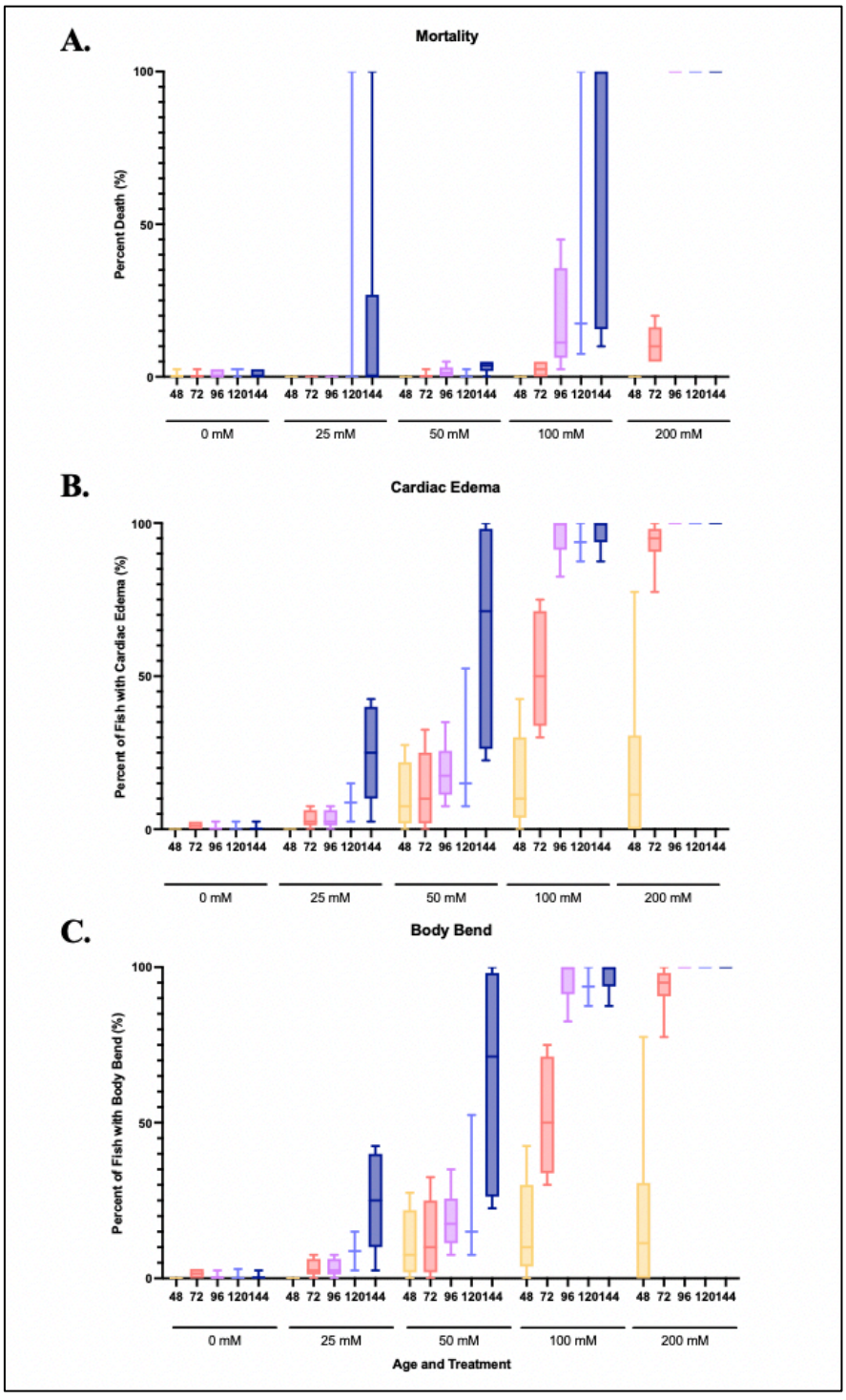


Figure 21: Higher levels of (A) Mortality, (B) Cardiac Edema, and (C) Body Bend observed at higher Hydroxyurea concentrations. Zebrafish were observed across development (48 hpf through 144 hpf) at five treatments (0 mM, 25 mM, 50 mM, 100 mM, and 200 mM). This data was collected by Danielle Obiri and plotted by Anvith Reddy in Prism.

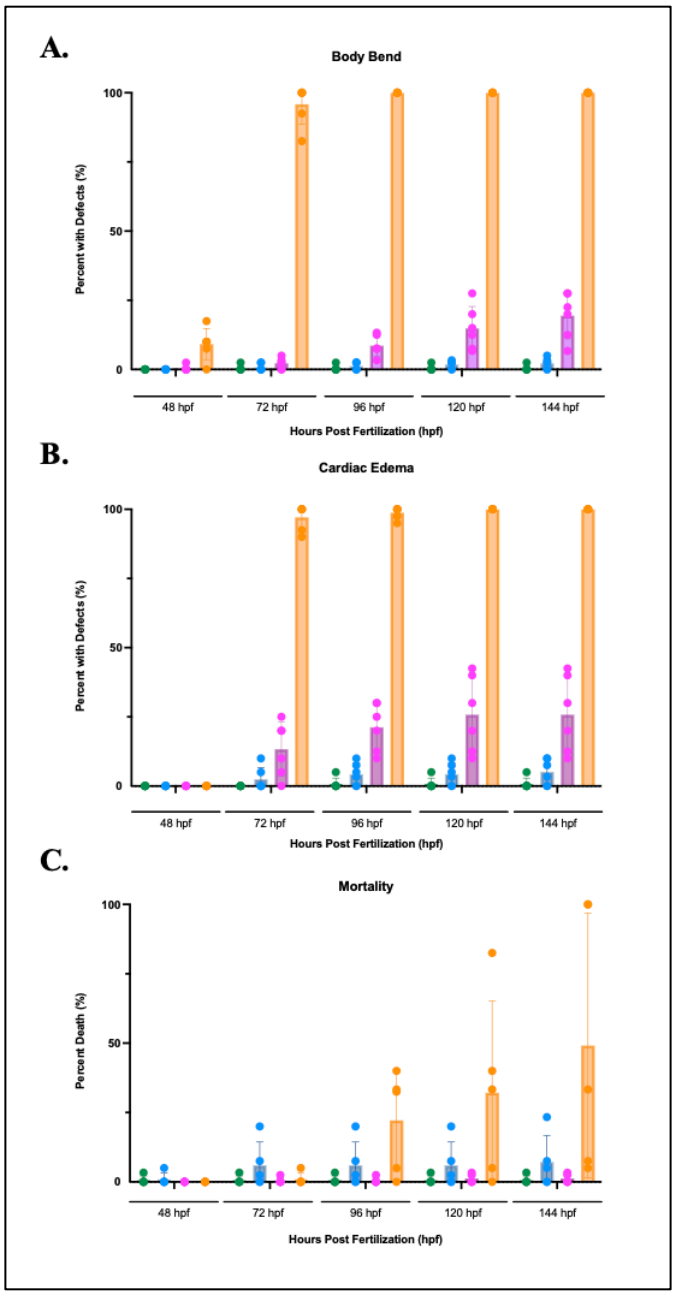


Figure 22: Graphs displaying the distribution of cardiac edema (A) and body bend (B) between wild-type and *cdca7*$^{-/-}$ MZ zebrafish at two treatment levels. Green data points correspond to wildtype untreated fish, blue corresponds to *cdca7*$^{-/-}$ MZ untreated fish, pink corresponds to wildtype fish treated with 50 mM hydroxyurea, and orange corresponds to *cdca7*$^{-/-}$ MZ fish treated with 50 mM hydroxyurea. Levels of all three phenotypes assayed are consistent across treatments at 48 hpf. However, *cdca7*$^{-/-}$ MZ fish (50 mM) display a significant increase in Cardiac Edema and Body Bend at 72 hpf, 96 hpf, 120 hpf, and 144 hpf when compared to wildtype controls at the same treatment. Mortality was significantly higher in *cdca7*$^{-/-}$ MZ fish (50 mM) at 96 hpf, 120 hpf, and 144 hpf. Results were significant as assessed by a two-way ANOVA, <0.0001. This data was collected by Danielle Obiri and plotted by Anvith Reddy in Prism.

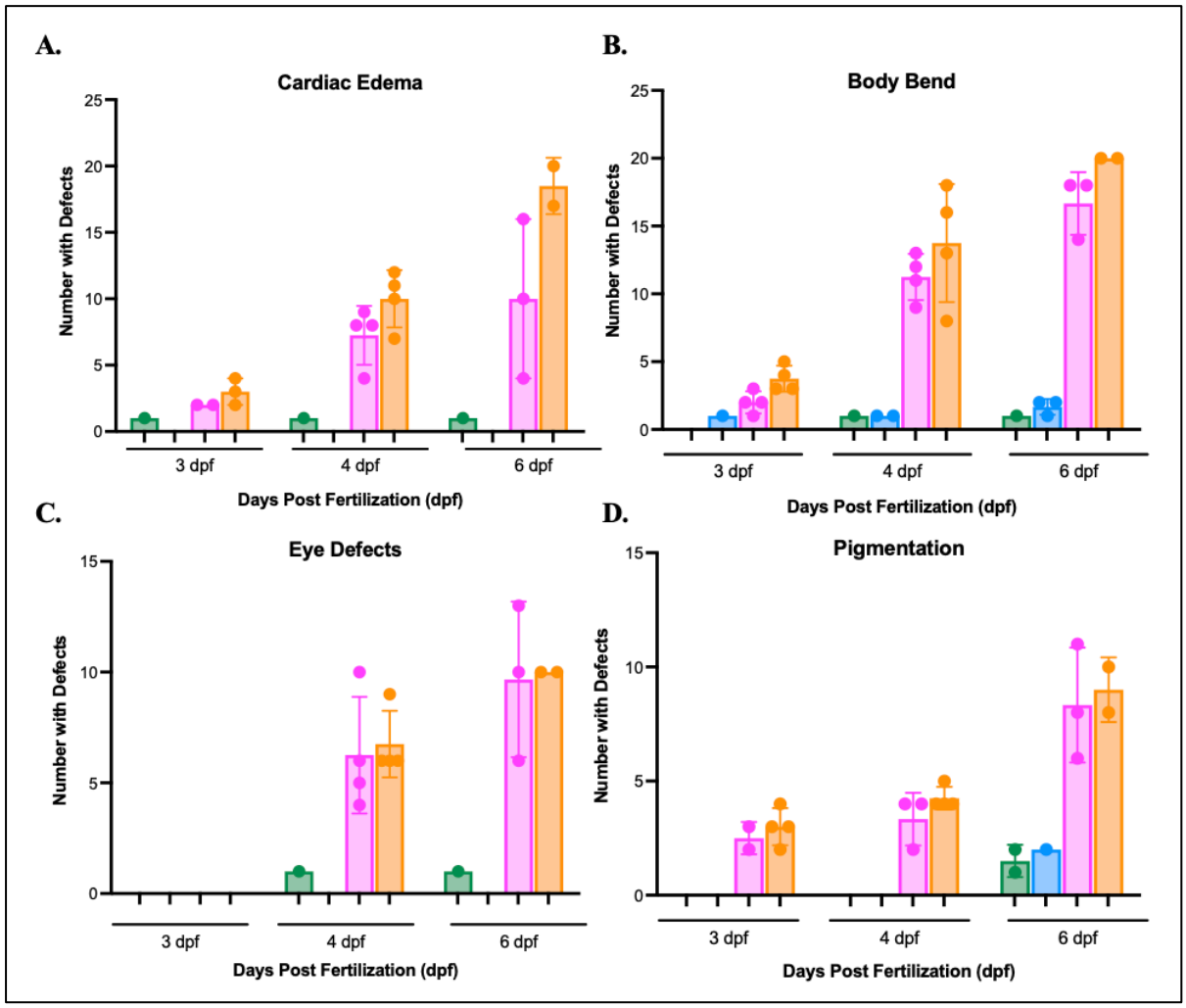


Figure 23: Plots displaying the distribution of (A) cardiac edema, (B) body bend, (C) eye defects, and (D) between wild-type and *cdca7*$^{-/-}$ MZ zebrafish at two treatment levels. There are no significant differences in phenotypic severity between *cdca7*$^{-/-}$ and wildtype zebrafish at the timepoints displayed as examined by pairwise t-tests (ns).

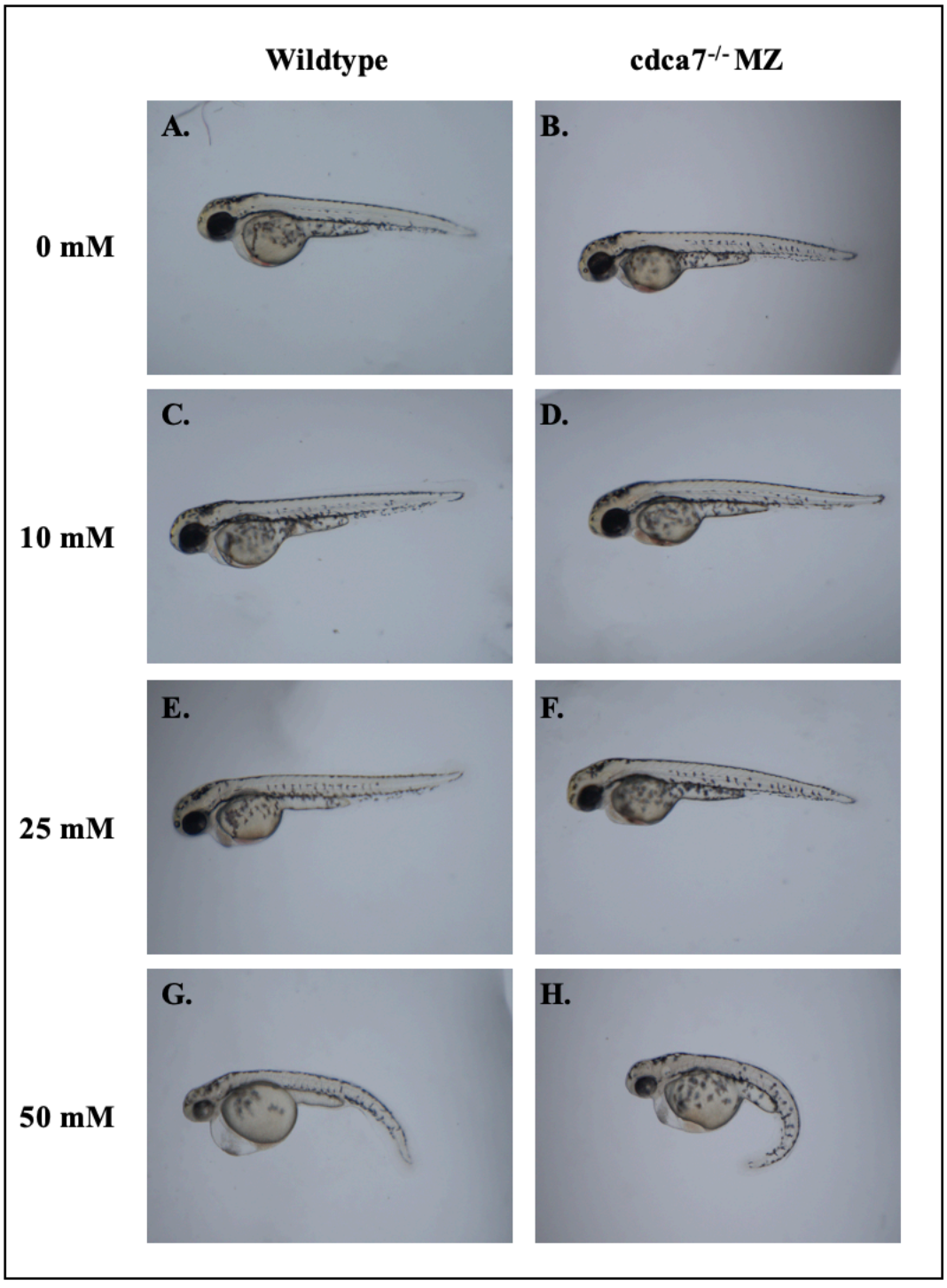


Figure 24: Bright field, lateral images of AB (wildtype) and *cdca7*$^{-/-}$ MZ zebrafish at four treatment conditions. All fish are pictured at 48 hpf. Genotypes are displayed vertically and increasing concentrations of hydroxyurea are displayed horizontally. Images were captured by Eric Schouten.

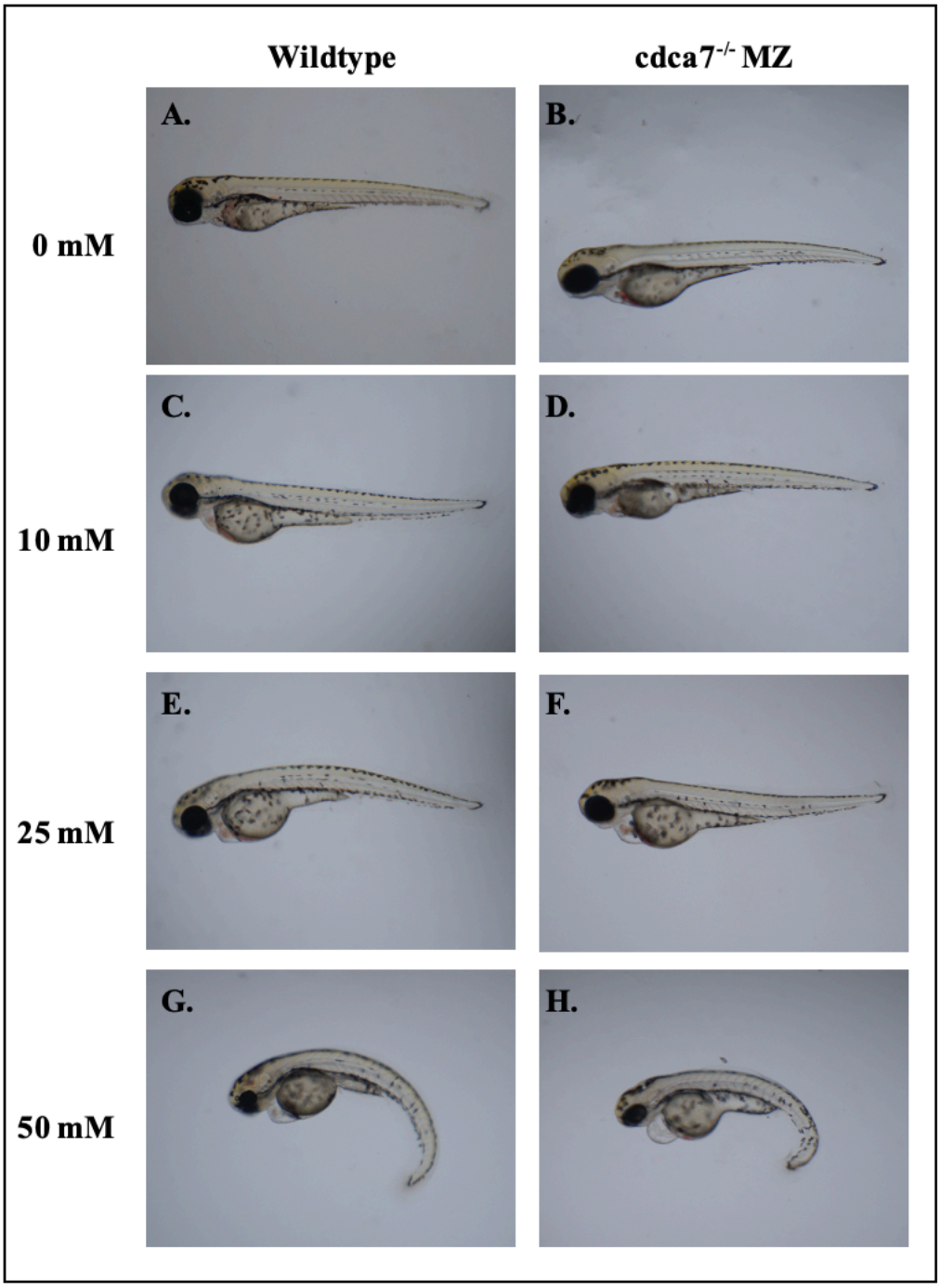


Figure 25: Bright field, lateral images of AB (wildtype) and *cdca7*$^{-/-}$ MZ zebrafish at four treatment conditions. All fish are pictured at 72 hpf. Genotypes are displayed vertically and increasing concentrations of hydroxyurea are displayed horizontally. Images were captured by Eric Schouten.

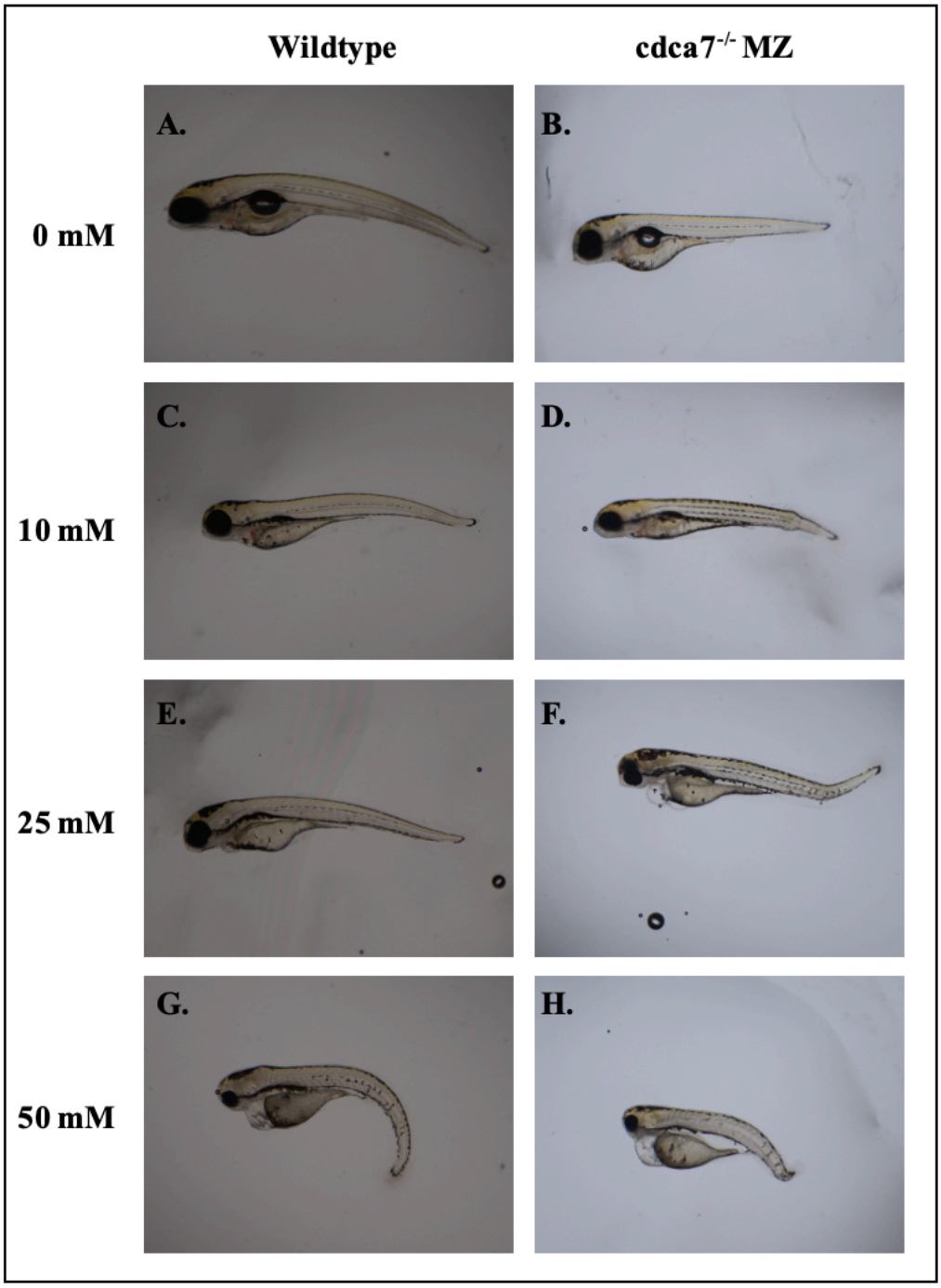


Figure 26: Bright field, lateral images of AB (wildtype) and *cdca7*$^{-/-}$ MZ zebrafish at four treatment conditions. All fish are pictured at 96 hpf. Genotypes are displayed vertically and increasing concentrations of hydroxyurea are displayed horizontally. Images were captured by Eric Schouten.

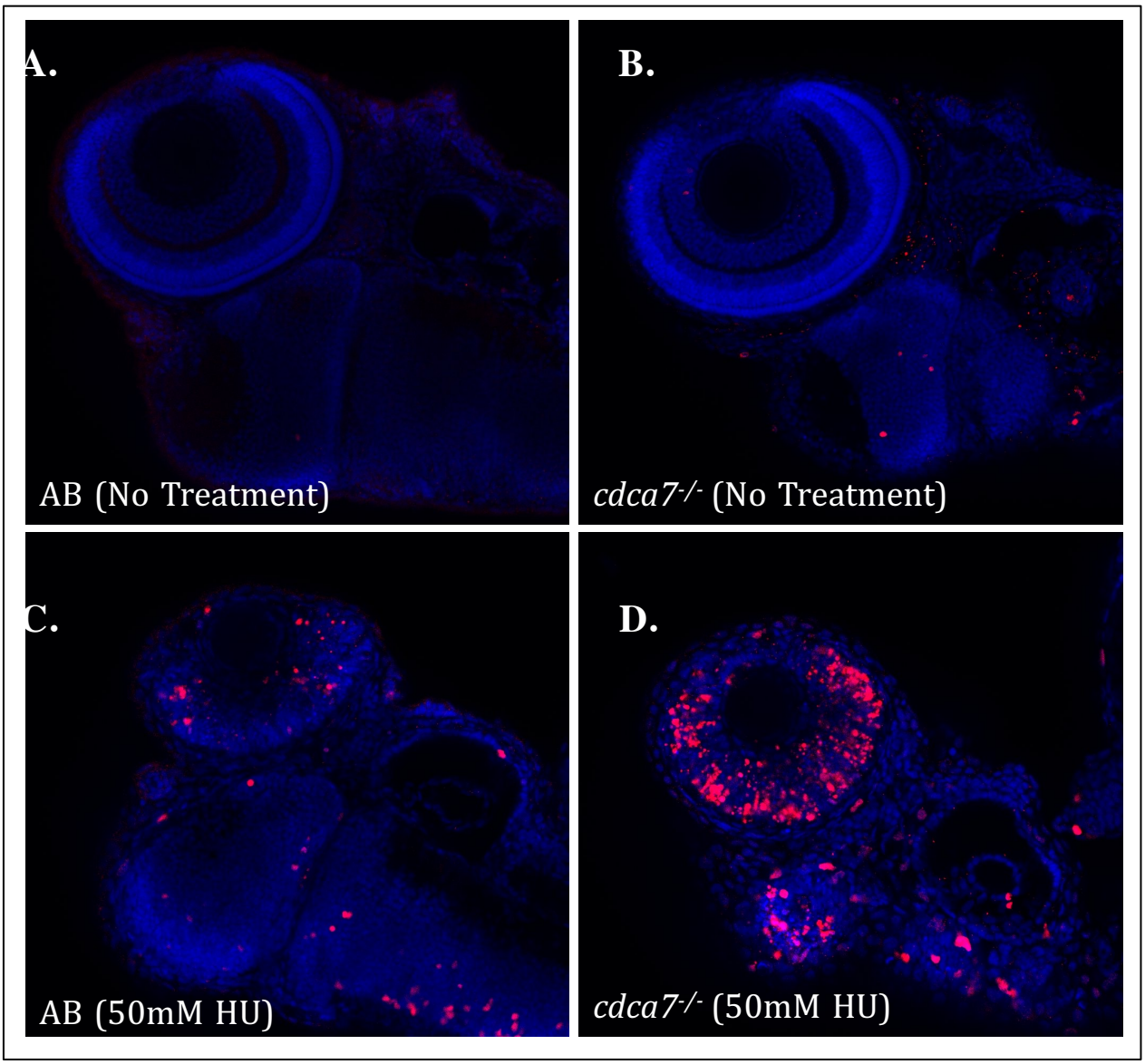
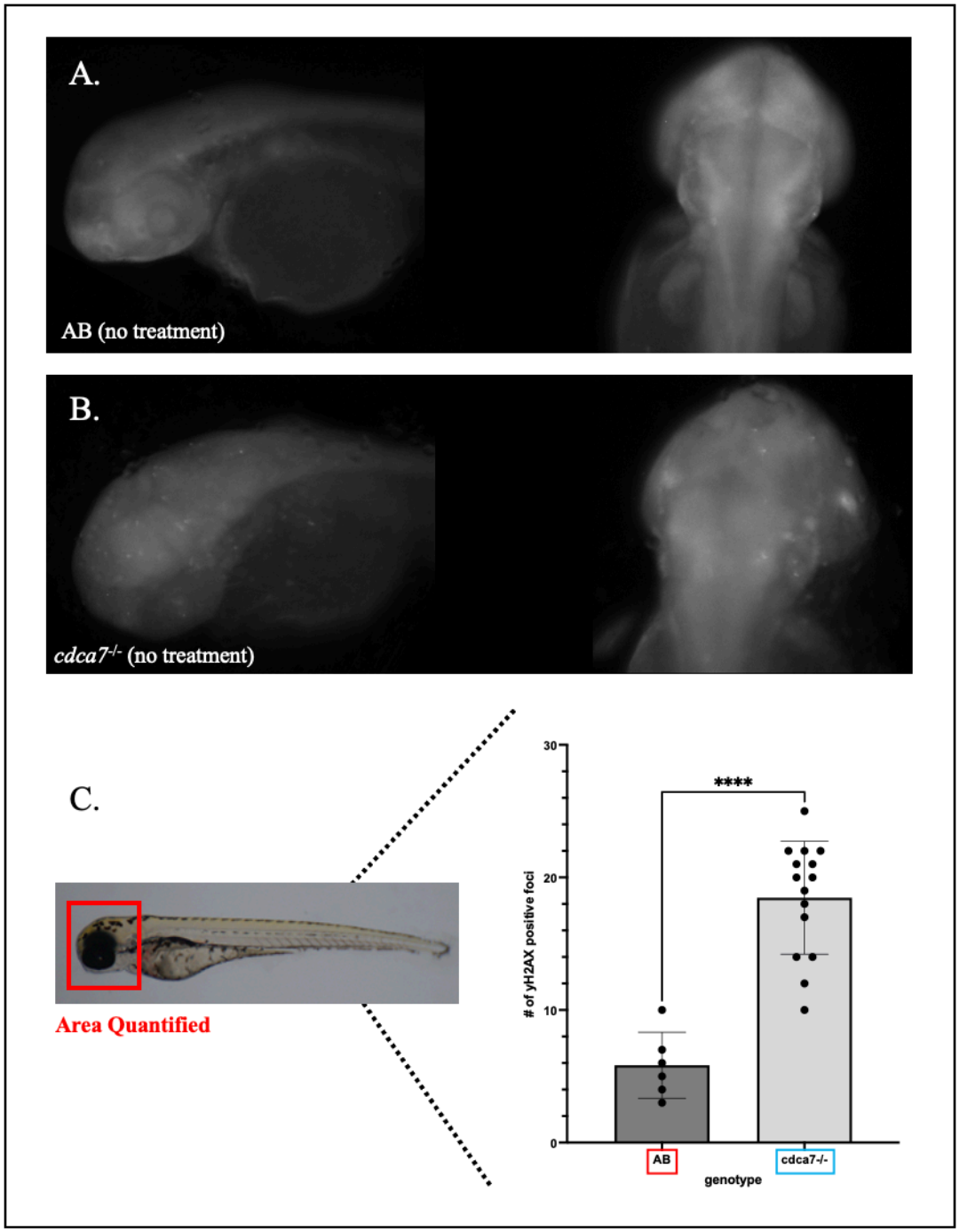


Figure 27: Higher levels of γH2AX staining in *cdca7*$^{-/-}$ mutant zebrafish compared to AB (wildtype) at no treatment and 50 mM Hydroxyurea levels. Fish are staged at 72 hpf and imaged on a confocal microscope with the retina in focus. Wildtype fish are displayed in panels A and C. *Cdca7*$^{-/-}$ MZ mutant fish are displayed in panels B and D. Fish were reared in normal fish water in panels A and B. Fish were reared in 50 mM Hydroxyurea in panels C and D.



C

Figure 28: Quantification of γH2AX staining in untreated *cdca7*$^{-/-}$ MZ and WT zebrafish at 72 hpf. (A and B) Representative images of AB (WT) and *cdca7*$^{-/-}$ embryos, respectively, stained for γH2AX. (C) Quantification of foci in the head of the organism (AB, n=6; *cdca7*$^{-/-}$ n=14). Results were significant by Student's t-test ($p<0.0001$).

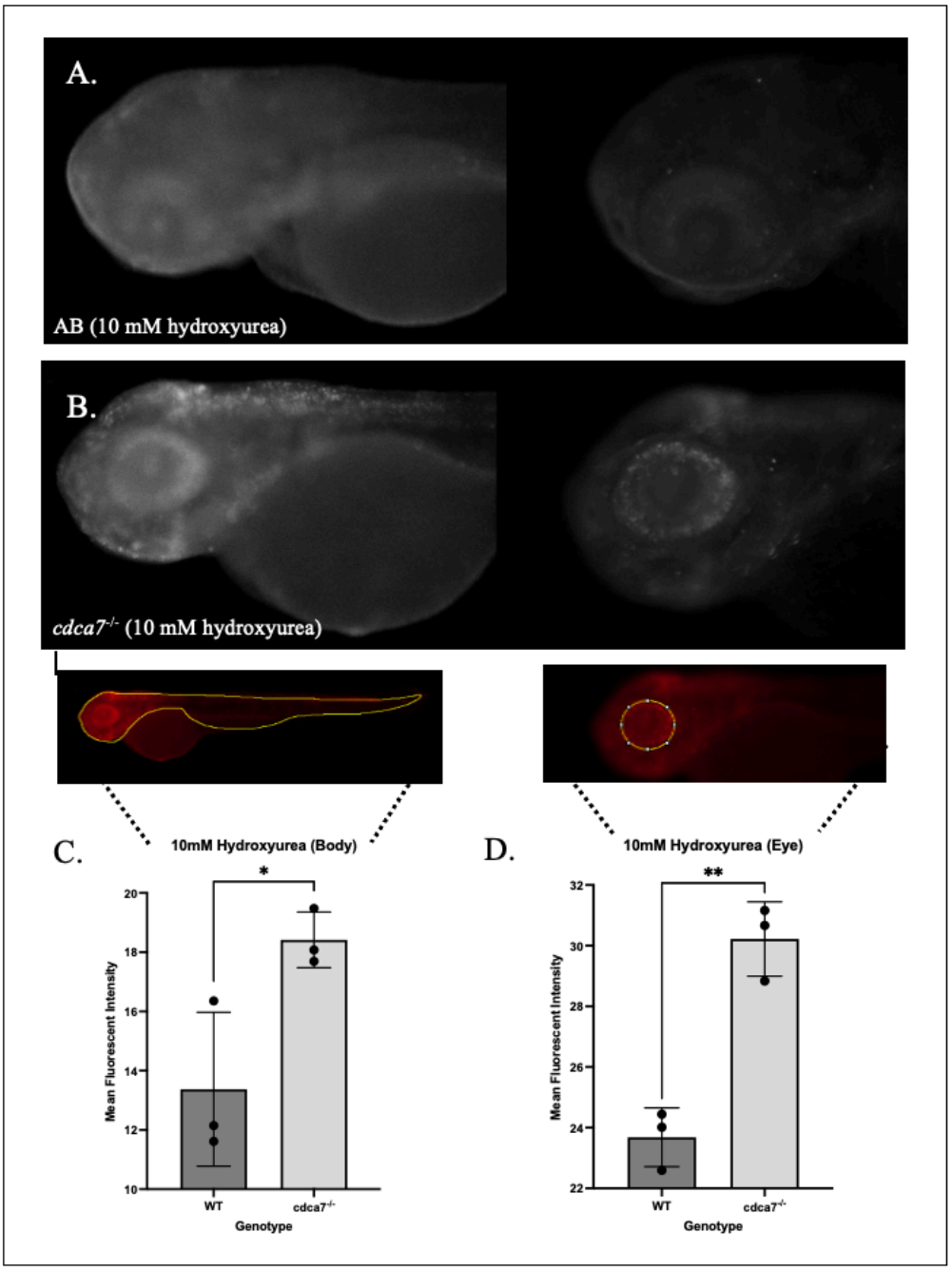


Figure 29: Quantification of γH2AX staining in *cdca7*$^{-/-}$ MZ and WT zebrafish treated with 10 mM hydroxyurea. (A and B) Representative images of AB and *cdca7*$^{-/-}$ embryos, respectively, stained for γH2AX. Mean Fluorescent Intensity, adjusted for area, was assessed for three fish per genotype and treatment level. Student's t-test was performed between the genotype groups. (C) ImageJ analysis of the zebrafish body, excluding the yolk sack. Results were significant ($p<0.05$) (D) ImageJ analysis of the zebrafish eye. Results were significant ($p<0.01$).

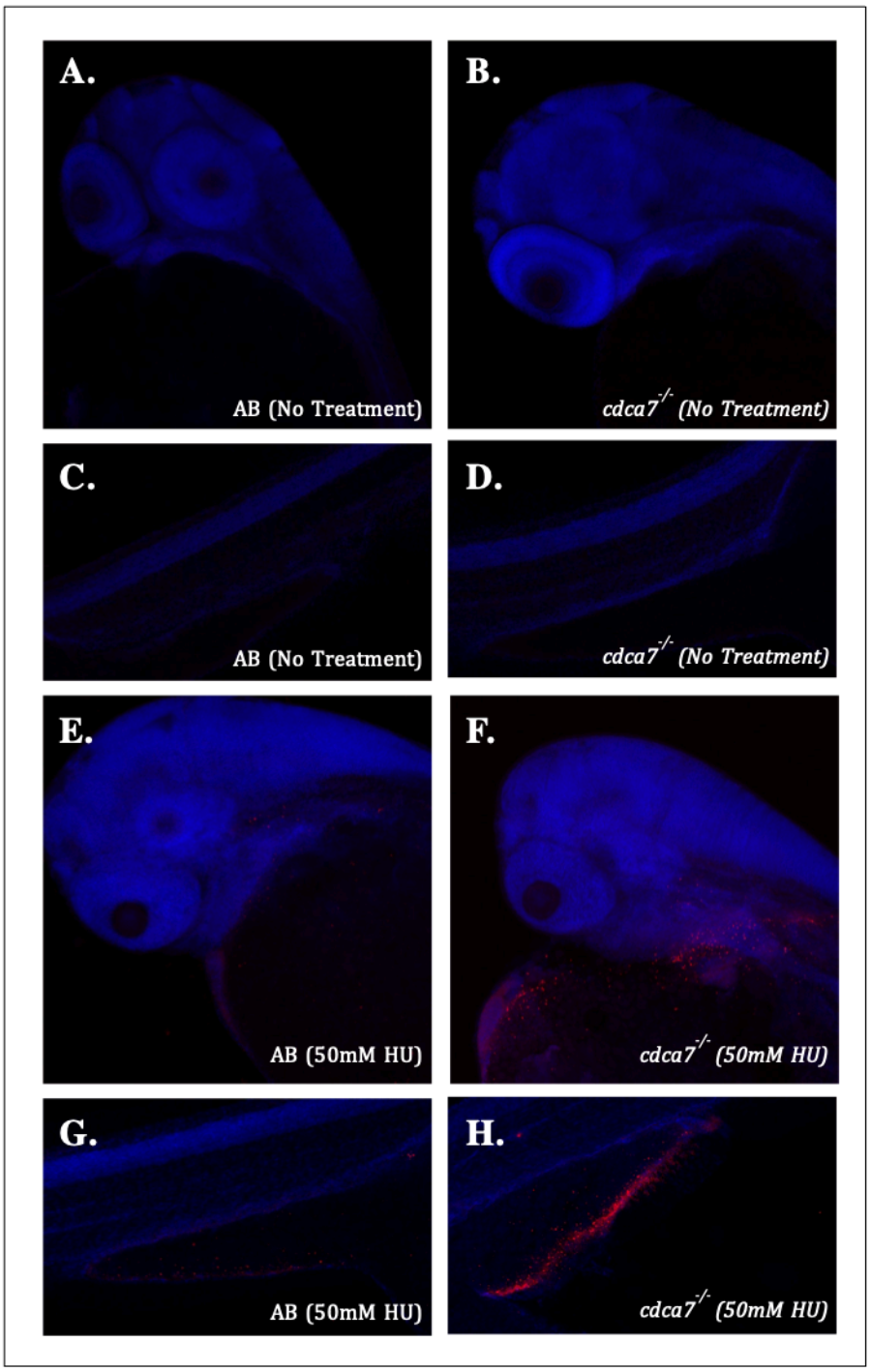


Figure 30: Higher staining for RNA:DNA hybrids in *cdca7*$^{-/-}$ MZ mutants after treatment with 50 mM hydroxyurea. Fish are staged at 72 hpf. Panels A-D are untreated fish and panels E-H are zebrafish reared in 50 mM Hydroxyurea. Panels A, C, E, and G represent AB (wildtype) fish and panels B, D, F, and H represent *cdca7*$^{-/-}$ MZ mutant fish. Panels A, B, E, and F display the head of the zebrafish with most staining observed in the myocardium. Panels C, D, G, and H depict the yolk sack extension.

CHAPTER 4: CLARIFYING THE LINK BETWEEN PERICENTROMERIC METHYLATION LOSS AND MALIGNANCY

PREFACE

Loss of DNA methylation at pericentromeric repeats has been observed in many cancers including Breast adenocarcinomas (Tsuda et al., 2002), Wilms Tumors (Qu et al., 1999b), Ovarian Epithelial Tumors (Qu et al., 1999a), Glioblastoma (Fanelli et al., 2008), and Urothelial Carcinomas (Nakagawa et al., 2005). However, a causative link between pericentromeric hypomethylation and malignancy has yet to be established. Our *in vivo* model of ICF syndrome allows us to probe whether or not hypomethylation acts to increase malignancy or rather if hypomethylation is a passive byproduct of malignancy.

Previously in our lab, antidotally, we have not observed increased malignancy in *cdca7* or *zbtb24* homozygous mutant adults when this mutation is present on an otherwise wildtype background. This finding suggests that on its own mutation of either gene is not sufficient to drive malignancy. In this chapter, I set up experimental cohorts to test whether mutation of *cdca7* promotes malignancy when introduced onto a cancer susceptible genetic background.

To this end, I crossed the homozygous *cdca7* mutation into a *p53* mutant background. *p53* acts as a tumor suppressor by mediating response to DNA damage and regulating the cell cycle/apoptosis. Loss of *p53* function, is a common feature in the majority of human cancers. The p53 background we are using contains an A to T transversion causing a methionine to lysine change at residue 214 (Berghmans et al., 2005). This mutation abolishes normal p53 function. Adult $p53^{-/-}$ zebrafish are viable, and up to 30% of homozygous adults get malignant peripheral nerve sheath tumors (MPNSTs) in the second year of life.

EXPIREMENTAL APPROACH AND SURVIVAL DATA

To determine if a homozygous mutation in *cdca7* results in increased tumor susceptibility, I set out to compare cohorts of zebrafish that were wildtype, homozygous for *cdca7*$^{-/-}$ only, homozygous *p53*$^{-/-}$ only, or homozygous mutant for both *cdca7*$^{-/-}$ and p53$^{-/-}$. The analysis began with ~200 embryos of each genotype pooled from 3-6 crosses each. At 1.5 mpf, fish were split and redistributed to ~30 fish per tank. Each genotype is being reared in separate tanks and counted every four weeks. In addition, fish are inspected weekly for tumors, as evidenced by protruding or discolored tissue around the eyes and body. At the time of this thesis, zebrafish from this experiment are 15 months old.

Thus far, I have observed tumor formation in both homozygous *p53*$^{-/-}$ and homozygous *cdca7*$^{-/-}$ p53$^{-/-}$ background. Tumors are generally formed in either the body of the animal or in the ocular region. There has been one observance of a melanoma of the tail in a *cdca7*$^{-/-}$ *p53*$^{-/-}$ mutant fish. Survival curves were generated with counts starting at four months post fertilization. This date was chosen as it was when our fish passed the larval death phase and were classified as adults. At this stage, the fish were re-balanced to 22 fish per tank across each genotype. Survival analysis was conducted using Log rank (Mantel-Cox) statistical analysis. At 15 months post fertilization and the time of this thesis, we see a significant decreased in survival of the *cdca7*$^{-/-}$ *p53*$^{-/-}$ when compared to both single homozygous genotypes (Figure 31).

DISCUSSION

Currently, we hypothesize that an increase in death in our double mutant (*cdca7*$^{-/-}$ *p53*$^{-/-}$) cohorts might implicate the loss of methylation as an important factor in cancer progression. We aim to replicate this experiment to confirm these findings.

At present, analysis has been restricted to visual assessment of fish, which is likely to overlook many types of malignancy, especially at early stages. Potential future studies include fixing, sectioning, and cytological analysis of tumors from different genotypes as well as H&E staining of surviving animals at 6 months to identify fish lacking obvious surface level tumors but that might have early signs of malignant transformation. Further, karyotyping of tumor cells can be carried out to identify any chromosomal abnormalities.

Our sensitized zebrafish exhibited a greater amount of death after the introduction of the *cdca7* mutation (**chapter 4**), and our *cdca7* mutant zebrafish exhibit greater DNA damage and DNA damage susceptibility (**chapter 3**). Although this data is preliminary, taken together, this may suggest that a loss of pericentromeric hypomethylation leads to an increase in DNA damage and subsequently tumor formation/decreased viability.

CHAPTER 4: FIGURE

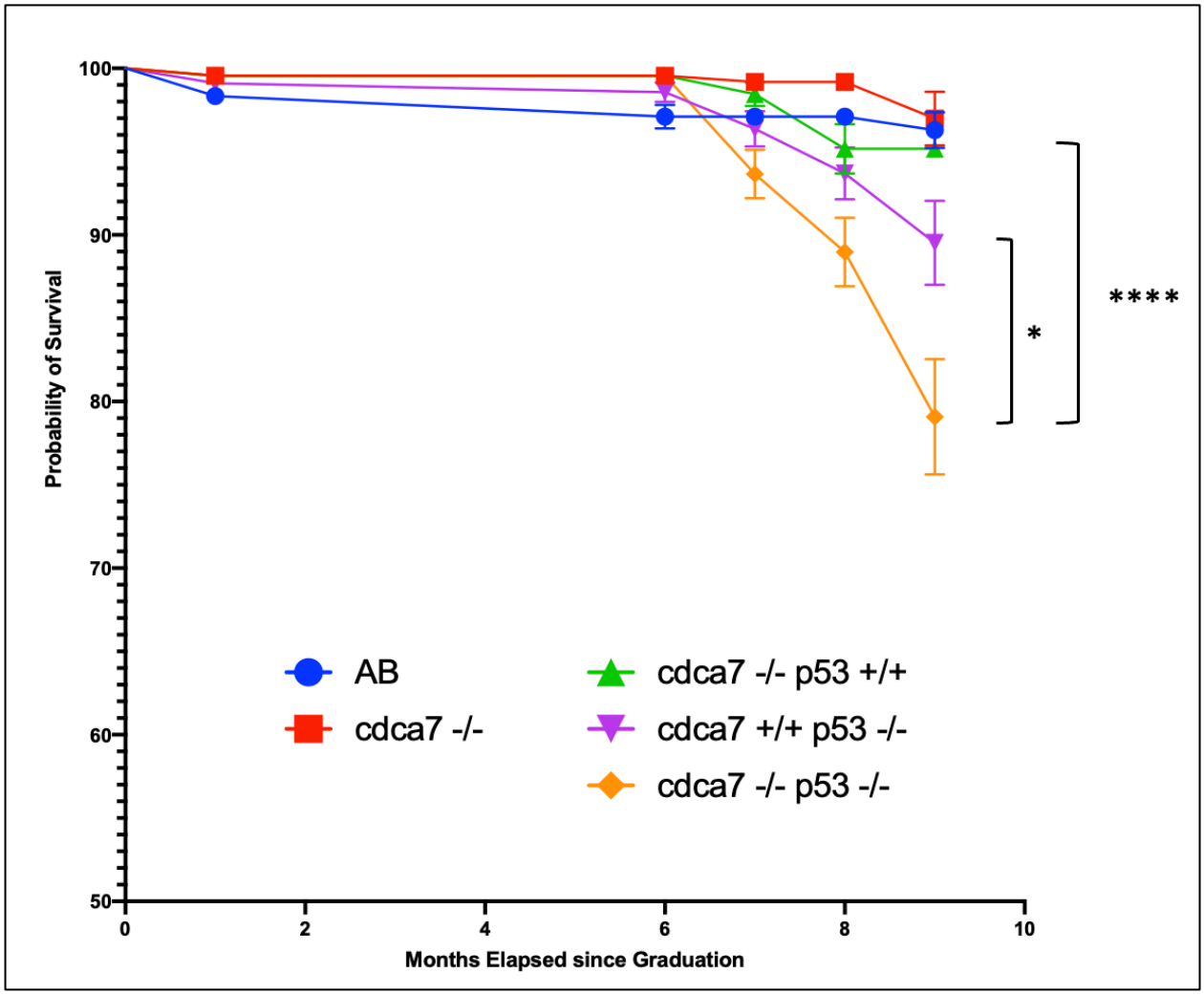


Figure 31: Decreased survival probability in *cdca7/p53* double mutants when compared to both single mutant cohorts. Line color indicates genotype. P-values were generated using Log rank (Mantel-Cox) survival tests.

CHAPTER 5: SUMMARY AND PERSPECTIVES

Using viable zebrafish models of ICF syndrome which carry mutations in *zbtb24* and *cdca7* (Rajshekar et al., 2018; A. Higgs, *Unpublished data*), our lab sought to clarify the relationship between pericentromeric methylation loss, DNA damage and malignancy. In this thesis, I first describe analysis of two RNA-seq datasets comparing *cdca7* and *zbtb24* mutants to wildtype controls. Using gene-set enrichment and KEGG pathway analyses, I found that increased DNA-damage associated signatures in both *cdca7* and *zbtb24* mutants with a stronger misexpression in *cdca7.* Using qRT-PCR, I sought to verify misexpression of a subset of these DNA-damage associated genes (*rad51*, *rad52*, and *rad51* paralogues). Moving forward, we hope to continue verifying these changes through qRT-PCR as well as investigate mis-regulation of other DNA-damage associated markers.

Next, I tested if *cdca7* mutant zebrafish exhibit an increased sensitivity to DNA damaging agents as seen through developmental abnormalities. These results were inconclusive. Further, I assayed the accumulation of two DNA damage-associated signatures, γH2AX and DNA:RNA hybrids, *in vivo*. I saw that *cdca7* mutant zebrafish exhibited greater signs of DNA damage at baseline as well as after exposure to hydroxyurea, an S-phase specific DNA damaging agent. γH2AX formation was observed at baseline, however, we detected RNA:DNA hybrids only after the introduction of hydroxyurea.

In this thesis, I also describe an experiment to assess if pericentromeric hypomethylation is associated with decreased viability and/or increased tumor susceptibility in a sensitized genetic background. At present, we observe a statistically significant difference between our *cdca7* mutants in a *p53* mutant background when compared to the *p53* mutant cohort or *cdca7* mutant

cohort alone. However, this experiment will need to be repeated across more tanks with a more robust analysis of death and tumor incidence. Our sensitized zebrafish exhibited a greater amount of death after the introduction of the *cdca7* mutation (**chapter 4**), and our *cdca7* mutant zebrafish exhibit greater DNA damage and DNA damage susceptibility (**chapter 3**). Although this data is preliminary, taken together, this may suggest that a loss of pericentromeric hypomethylation leads to an increase in DNA damage and subsequently tumor formation/decreased viability.

MATERIALS AND METHODS

Zebrafish husbandry

Zebrafish husbandry and care were conducted in full accordance with animal care and use guidelines with approval by the Institutional Animal Care and Use Committees at the University of Georgia. Zebrafish were raised and in compliance with relevant protocols and ethical regulations at the University of Georgia URAR Life Sciences Facility. Wild-type lines were of the AB background. Fertilized eggs were obtained by breeding single pairs unless otherwise stated. Embryos were reared in system water at 28°C and transferred to the animal facility 5 days post fertilization.

DNA-prep and genotyping

Tissue for DNA-prep was obtained from adult zebrafish caudal fin-clips. DNA was isolated with a quick-prep protocol using 50mM NaOH and TrisHCl (pH 8.0). Standard PCR conditions were used (appendix 1).

RNA extraction and cDNA synthesis

Total RNA was extracted from zebrafish at 24 hours (10 pooled embryos), 58 hours (10 pooled embryos), 5 days (5 pooled fish), 7 days (5 pooled fish), 14 days (5 pooled fish) post-fertilization. Total RNA was extracted using TRIzol Reagent (Ambion by Life Technology) and precipitated with Isopropanol. RNA was subsequently treated with DNAse using the Turbo DNA-*free* Kit (Invitrogen by Thermo Fisher Scientific). cDNA was synthesized from total RNA using GoScript™ Reverse Transcription System (Promega).

qRT-PCR

Quantitative real-time PCR was performed using Power SYBR Green PCR Master Mix (REF 4367659, Thermo Fisher) and the Applied Biosystems QuantStudio 3 Thermal Cycler System (96 well, 0.2ml block). Analysis was performed using the $2^{-\Delta\Delta Ct}$ method, with relative mRNA levels of all transcripts normalized to β-actin. Amplification conditions were 30 seconds at 95°C and then 40 cycles each consisting of 5 seconds at 95°C and 30 seconds at 60°C.

qRT-PCR primer design and verification

Primer design was carried out using the PubMed gene database and the NCBI Primer BLAST tool. Primers were designed to ensure the amplicon size was between 70 and 200 bp. Minimum and maximum melt temp were set to 60°C and 63°C respectively. Primers were designed to span an exon-exon junction to avoid genomic DNA contamination. 40-60% GC content was targeted. Primer verification (appendix 2) was carried out by cDNA serial dilution to 1, 1:2, 1:4, 1:8, 1:16, and 1:32.

RNA-sequencing

ZBTB24$^{-/-}$*:* 500 ng of total RNA was used for polyA enrichment and TruSeq library preparation after RiboGreen quantification and quality control by Agilent BioAnalyzer. The TruSeq Stranded mRNA LT Kit (Illumina) was used with 8 cycles of PCR. Using the TruSeq SBS v4 Kit (Illumina), barcoded samples were run on a HiSeq 2500 High Output in a 50bp/50bp paired end run. An average of 35.4 million reads per paired end per sample were generated. GC content averaged 46% across all samples. *CDCA7*$^{-/-}$*:* Up to 1300 ng of total RNA was used for polyA enrichment and TruSeq library preparation after Qubit

quantification and quality control by Agilent BioAnalyzer. The TruSeq Stranded mRNA LT Kit (Illumina) was used with 13 cycles of PCR. Using the TruSeq SBS v4 Kit (Illumina), barcoded samples were run on a HiSeq 2500 High Output in a 50bp single end run. Fragment analysis, bead cleanup, and sequencing was performed at the Georgia Genomics and Bioinformatics Core (GGBC). An average of 15.7 million reads per sample were generated. GC content averaged 52% across all samples

Genome alignment and read mapping

Bioinformatic analyses were performed using the Linux based GACRC computing cluster at the University of Georgia as well as the R programming environment. The reference sequence and genome annotation of *Danio rerio* used were sourced from ENSEMBL (GRCz11.100 genome assembly). Trim_Galore (v0.6.5-GCCcore-8.3.0-Java-11-Python-3.7.4) was used to trim RNA-seq adaptors. STAR (v2.7.3a-GCC-8.3.0) was used to align the reads supplied with a reference General Features File (GFF) to the *Danio rerio* GRCz11.100 reference genome assembly. To generate counts for the samples, featureCounts (Subread v2.0.1-GCC-8.3.0) was used. Quality checks were performed with FastQC and MultiQC. The scripts used on the GACRC computing cluster were adapted from the Goll Lab (University of Georgia) pipeline.

Expression analysis

Differential expression analysis was performed in R using the DESeq2 pipeline described by Love et. al, 2014. The DESeq2 pipeline estimates gene-wise dispersions and shrinks them to generate accurate counts. The dispersion parameters are set based on a log-normal prior distribution centered on the dataset's mean normalized counts. Gene-wise estimates flagged as outliers are not shrunk towards the fitted values. The normalization pipeline

used was adapted from the Goll Lab (University of Georgia) and Turner Lab (University of Virginia). Normalization was performed by dividing counts by sample-specific size factors determined through the median ratio of gene counts relative to geometric mean per gene. This method considers both sequencing depth and RNA composition.

GO, GSEA, and KEGG analyses

GO: For Gene-Ontology Analysis the Benjamini & Hochberg p-value adjustment method was used with a p-value cutoff of 0.01 and a q-value cutoff of 0.05. The keytype used was ENSEMBL and the ontology method used was "ALL". EnrichGO (clusterProfiler) (Yu et al., 2012) was used to generate the GO-object. *GSEA:* For GSEA analysis, a minimum gene-set size (GSS) of 3 and a maximum GSS of 800 were used. A p-value cutoff of 0.05 was used with an eps argument of 0 and no p-value adjustment. The keytype used was ENSEMBL and the ontology method used was "ALL". gseGO (clusterProfiler) (Yu et al., 2012) was used to generate the GSEA-object. *KEGG:* All KEGG arguments except the keytype (ncbi-geneid) were the same as the GSEA analysis. gseKEGG (clusterProfiler) (Yu et al., 2012) was used to generate the KEGG-object.

Fixation and Immunofluorescence

Whole zebrafish at the desired stage of analysis were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) overnight at 4°C. Fish were washed with 100% methanol and stored in 100% MeOH at -20 degrees Celcius. Fish were rehydrated in increasing percentages of MeOH/PBS. Fish were blocked in 10% BSA/1% Normal Goat Serum in PBST (PBS/0.1% Tween-20). All antibodies were diluted in 10% BSA/1% Normal Goat Serum/0.1% PBST. anti-y-H2AX (Genetex, GTX127342) was diluted 1:400 and anti-RNA-DNA-hybrid was diluted 1:50. Alexa-596 goat anti- rabbit and

Alexa-488 goat anti-rabbit secondary antibodies (abcam ab150088) were diluted 1:1000 and 1:500 respectively. Low magnification images were acquired using an upright Olympus MVX10 MacroView fluorescence microscope and an inverted Zeiss AxioObserver Z1 microscope. High magnification images were acquired using the upright Zeiss LSM 880 Confocal Microscope system (AXIO Imager Z2) and inverted Zeiss LSM 710 confocal microscope system (AXIO Observer Z1) at the University of Georgia Biomedical Microscopy core.

ImageJ Analysis

Analysis of immunofluorescent zebrafish was conducted using pictures taken under a widefield fluorescent microscope using the ImageJ software. For each genotype and treatment condition, three larvae were randomly chosen, and lateral images of the zebrafish embryos were taken. All images were taken at the magnification and exposure time. In the ImageJ software, a section of the zebrafish (either the eye or the body, excluding the yolk sack) was selected and assessed for mean fluorescent intensity. Each measurement was normalized to both area and background intensity.

Statistics

Student's *t* test, Welch's *t* test, or Wald's t-test were performed for statistical analysis based on sample numbers, data distributions, and data type. Bonferroni correction and the Benjamini & Hochberg adjustment were applied for multiple comparisons. Significant *P* values were described to be at the 5% level unless otherwise indicated. All statistical tests were two-sided, unless otherwise indicated. Statistical analysis was carried out using the R programming environment version 3.4.2 (R Foundation for Statistical

Computing, Vienna, Austria, 2020) or GraphPad Prism version 7.000 (GraphPad Software, La Jolla, CA, USA, 2020).

REFERENCES

Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* 19, 229-244.

Alvarez-Ponce, D., Torres-Sánchez, M., Feyertag, F., Kulkarni, A., & Nappi, T. (2018). Molecular evolution of DNMT1 in vertebrates: Duplications in marsupials followed by positive selection. *PloS one*, 13(4), e0195162.

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

Ayarpadikannan, S., & Kim, H. S. (2014). The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. *Genomics & informatics*, 12(3), 98–104.

Bader, A. S., Hawley, B. R., Wilczynska, A., & Bushell, M. (2020). The roles of RNA in DNA double-strand break repair. British journal of cancer, 122(5), 613–623.

Ball, M. P., Li, J. B., Gao, Y., Lee, J. H., LeProust, E. M., Park, I. H., Xie, B., Daley, G. Q., & Church, G. M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature biotechnology*, 27(4), 361–368.

Beisel, C., & Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. Nature reviews. *Genetics*, 12(2), 123–135.

Berghmans, S., Murphey, R.D., Wienholds, E., Neuberg, D., Kutok, J.L., Fletcher, C.D., Morris, J.P., Liu, T.X., Schulte-Merker, S., Kanki, J.P., et al. (2005). tp53 mutant zebrafish develop malignant peripheral nerve sheath tumors. *Proc Natl Acad Sci U S A* 102, 407-412.

Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40(1), 91–99.

Campos, C., Valente, L. M., & Fernandes, J. M. (2012). Molecular evolution of zebrafish dnmt3 genes and thermal plasticity of their expression during embryonic development. *Gene*, 500(1), 93–100.

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

Ehrlich, M., Sanchez, C., Shao, C., Nishiyama, R., Kehrl, J., Kuick, R., Kubota, T., & Hanash, S. M. (2008). ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. *Autoimmunity*, 41(4), 253–271.

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

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

Fasth A, Forestier E, Holmberg E, Holmgren G, Nordenson I, Söderström T, Wahlström J (1990) Fragility of the centromeric region of chromosome 1 associated with combined immune deficiency in siblings. *Acta Paediatr Scand* 79:605–612

Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A* 107, 8689-8694.

Garrido-Ramos, M.A. (2017). Satellite DNA: An Evolving Topic. *Genes (Basel)* 8.

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

Gimelli, G., Varone, P., Pezzolo, A., Lerone, M., & Pistoia, V. (1993). ICF syndrome with variable expression in sibs. *Journal of medical genetics*, 30(5), 429–432.

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

Gössling, K. L., Schipp, C., Fischer, U., Babor, F., Koch, G., Schuster, F. R., Dietzel-Dahmen, J., Wieczorek, D., Borkhardt, A., Meisel, R., & Kuhlen, M. (2017). Hematopoietic Stem Cell Transplantation in an Infant with Immunodeficiency, Centromeric Instability, and Facial Anomaly Syndrome. *Frontiers in immunology*, 8, 773.

Haas O. A. (1990). Centromeric heterochromatin instability of chromosomes 1, 9, and 16 in variable immunodeficiency syndrome--a virus-induced phenomenon? *Human genetics*, 85(2), 244–246.

Hardikar, S., Ying, Z., Zeng, Y., Zhao, H., Liu, B., Veland, N., McBride, K., Cheng, X., & Chen, T. (2020). The ZBTB24-CDCA7 axis regulates HELLS enrichment at centromeric satellite repeats to facilitate DNA methylation. *Protein & cell*, 11(3), 214–218.

Helfricht, A., Thijssen, P. E., Rother, M. B., Shah, R. G., Du, L., Takada, S., Rogier, M., Moritz, J., IJspeert, H., Stoepker, C., van Ostaijen-Ten Dam, M. M., Heyer, V., Luijsterburg, M. S., de Groot, A., Jak, R., Grootaers, G., Wang, J., Rao, P., Vertegaal, A., van Tol, M., … van Attikum, H. (2020). Loss of ZBTB24 impairs nonhomologous end-joining and class-switch recombination in patients with ICF syndrome. *The Journal of experimental medicine*, 217(11), e20191688.

Jagannathan, M., Cummings, R., and Yamashita, Y.M. (2018). A conserved function for pericentromeric satellite DNA. *Elife* 7.

Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A., & Viegas-Pequignot, E. (1993). An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Human molecular genetics*, 2(6), 731–735.

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

Jin, B., Li, Y., & Robertson, K. D. (2011). DNA methylation: superior or subordinate in the epigenetic hierarchy?. *Genes & cancer*, 2(6), 607–617.

Jones, P. A., & Laird, P. W. (1999). Cancer epigenetics comes of age. *Nature genetics*, 21(2), 163–167.

Kiaee, F., Zaki-Dizaji, M., Hafezi, N., Almasi-Hashiani, A., Hamedifar, H., Sabzevari, A., Shirkani, A., Zian, Z., Jadidi-Niaragh, F., Aghamahdi, F., Goudarzvand, M., Yazdani, R., Abolhassani, H., Aghamohammadi, A., & Azizi, G. (2020). Clinical, Immunologic, and Molecular Spectrum of Patients with Immunodeficiency, Centromeric instability, and Facial anomalies (ICF) syndrome: a Systematic Review. *Endocrine, metabolic & immune disorders drug targets*, Advance online publication.

Kim, M., B.N. Trinh, T.I. Long, S. Oghamian, and P.W. Laird, Dnmt1 deficiency leads to enhanced microsatellite instability in mouse embryonic stem cells. *Nucleic Acids Res*, 2004. 32(19): p. 5742-9.

Kuo, L. J., & Yang, L. X. (2008). Gamma-H2AX - a novel biomarker for DNA double-strand breaks. In vivo (Athens, Greece), 22(3), 305–309.

Kuster, J.E., Guarnieri, M.H., Ault, J.G., Flaherty, L., and Swiatek, P.J. (1997). IAP insertion in the murine LamB3 gene results in junctional epidermolysis bullosa. *Mamm Genome* 8, 673-681.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915-926.

López-Flores, I., & Garrido-Ramos, M. A. (2012). The repetitive DNA content of eukaryotic genomes. Genome dynamics, 7, 1–28. https://doi.org/10.1159/000337118

Lu, W. T., Hawley, B. R., Skalka, G. L., Baldock, R. A., Smith, E. M., Bader, A. S., Malewicz, M., Watts, F. Z., Wilczynska, A., & Bushell, M. (2018). Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair. Nature communications, 9(1), 532.

Madaan, K., Kaushik, D., & Verma, T. (2012). Hydroxyurea: a key player in cancer chemotherapy. Expert review of anticancer therapy, 12(1), 19–29.

Mah, LJ., El-Osta, A. & Karagiannis, T. γH2AX: a sensitive molecular marker of DNA damage and repair. Leukemia 24, 679–686 (2010).

Maraschio, P., Cortinovis, M., Dainotti, E., Tupler, R., & Tiepolo, L. (1992). Interphase cytogenetics of the ICF syndrome. *Annals of human genetics*, 56(3), 273–278.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205), 766–770.

Miniou, P., Jeanpierre, M., Bourc'his, D., Coutinho Barbosa, A. C., Blanquet, V., & Viegas-Péquignot, E. (1997). alpha-satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. *Human genetics*, 99(6), 738–745.

Molaro, A., Malik, H. S., & Bourc'his, D. (2020). Dynamic Evolution of De Novo DNA Methyltransferases in Rodent and Primate Genomes. *Molecular biology and evolution*, 37(7), 1882–1892.

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, 38(1), 23–38.

Moosavi, A., & Motevalizadeh Ardekani, A. (2016). Role of Epigenetics in Biology and Human Diseases. *Iranian biomedical journal*, 20(5), 246–258.

Nadel, J., Athanasiadou, R., Lemetre, C. et al. RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. Epigenetics & Chromatin 8, 46 (2015).

Nakagawa, T., Kanai, Y., Ushijima, S., Kitamura, T., Kakizoe, T., and Hirohashi, S. (2005). DNA hypomethylation on pericentromeric satellite regions significantly correlates with loss of heterozygosity on chromosome 9 in urothelial carcinomas. *J Urol* 173, 243-246.

Narayan, A., Ji, W., Zhang, X.Y., Marrogi, A., Graff, J.R., Baylin, S.B., and Ehrlich, M. (1998). Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *Int J Cancer* 77, 833-838.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247-257.

Palii, S.S., B.O. Van Emburgh, U.T. Sankpal, K.D. Brown, and K.D. Robertson, DNA methylation inhibitor 5-Aza-2'-deoxycytidine induces reversible genome-wide DNA damage that is distinctly influenced by DNA methyltransferases 1 and 3B. *Mol Cell Biol*, 2008. 28(2): p. 752-71.

Pezzolo, A., Prigione, I., Facchetti, P., Castellano, E., Viale, M., Gimelli, G., & Pistoia, V. (2001). T-cell apoptosis in ICF syndrome. *The Journal of allergy and clinical immunology*, 108(2), 310–312.

Plohl, M., Mestrovic, N., and Mravinac, B. (2014). Centromere identity from the DNA point of view. *Chromosoma* 123, 313-325.

Qu, G., Dubeau, L., Narayan, A., Yu, M.C., and Ehrlich, M. (1999a). Satellite DNA hypomethylation vs. overall genomic hypomethylation in ovarian epithelial tumors of different malignant potential. Mutat Res 423, 91-101.

Qu, G.-z., Grundy, P.E., Narayan, A., and Ehrlich, M. (1999b). Frequent Hypomethylation in Wilms Tumors of Pericentromeric DNA in Chromosomes 1 and 16. *Cancer Genetics and Cytogenetics* 109, 34-39.

Rai, K., Jafri, I.F., Chidester, S., James, S.R., Karpf, A.R., Cairns, B.R., and Jones, D.A. (2010). Dnmt3 and G9a cooperate for tissue-specific development in zebrafish. *J Biol Chem* 285, 4110-4121.

Rai, K., Nadauld, L.D., Chidester, S., Manos, E.J., James, S.R., Karpf, A.R., Cairns, B.R., and Jones, D.A. (2006). Zebra fish Dnmt1 and Suv39h1 regulate organ-specific terminal differentiation during development. *Mol Cell Biol* 26, 7077-7085.

Rajshekar, S., Yao, J., Arnold, P. K., Payne, S. G., Zhang, Y., Bowman, T. V., Schmitz, R. J., Edwards, J. R., & Goll, M. (2018). Pericentromeric hypomethylation elicits an interferon response in an animal model of ICF syndrome. *eLife*, 7, e39658.

Rakyan, V. K., Hildmann, T., Novik, K. L., Lewin, J., Tost, J., Cox, A. V., Andrews, T. D., Howe, K. L., Otto, T., Olek, A., Fischer, J., Gut, I. G., Berlin, K., & Beck, S. (2004). DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS biology*, 2(12), e405.

Robertson, K. D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F. A., & Jones, P. A. (1999). The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic acids research*, 27(11), 2291–2298.

Sadler, K. C., Krahn, K. N., Gaur, N. A., & Ukomadu, C. (2007). Liver growth in the embryo and during liver regeneration in zebrafish requires the cell cycle regulator, uhrf1. *Proceedings of the National Academy of Sciences of the United States of America*, 104(5), 1570–1575.

Sakano, K., Oikawa, S., Hasegawa, K., & Kawanishi, S. (2001). Hydroxyurea induces site-specific DNA damage via formation of hydrogen peroxide and nitric oxide. Japanese journal of cancer research : Gann, 92(11), 1166–1174.

Scelfo, A., & Fachinetti, D. (2019). Keeping the Centromere under Control: A Promising Role for DNA Methylation. *Cells*, 8(8), 912.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908-912.

Sharma, A., Jamil, M. A., Nuesgen, N., Dauksa, A., Gulbinas, A., Schulz, W. A., Oldenburg, J., & El-Maarri, O. (2019). Detailed methylation map of LINE-1 5'-promoter region reveals hypomethylated CpG hotspots associated with tumor tissue specificity. *Molecular genetics & genomic medicine*, *7*(5), e601.

Siggs, O.M., and Beutler, B. (2012). The BTB-ZF transcription factors. *Cell Cycle* 11, 3358-3369.

Simo-Riudalbas, L., Diaz-Lagares, A., Gatto, S., Gagliardi, M., Crujeiras, A. B., Matarazzo, M. R., Esteller, M., & Sandoval, J. (2015). Genome-Wide DNA Methylation Analysis Identifies Novel Hypomethylated Non-Pericentromeric Genes with Potential Clinical Implications in ICF Syndrome. *PloS one*, 10(7), e0132517.

Smeets, D. F., Moog, U., Weemaes, C. M., Vaes-Peeters, G., Merkx, G. F., Niehof, J. P., & Hamers, G. (1994). ICF syndrome: a new case and review of the literature. *Human genetics*, 94(3), 240–246.

Stacey, M., Bennett, M. S., & Hulten, M. (1995). FISH analysis on spontaneously arising micronuclei in the ICF syndrome. *Journal of medical genetics*, 32(7), 502–508.

Suzuki, T., Fujii, M., and Ayusawa, D. (2002). Demethylation of classical satellite 2 and 3 DNA with chromosomal instability in senescent human fibroblasts. *Exp Gerontol* 37, 1005-1014.

Tao, Y., Xi, S., Shan, J., Maunakea, A., Che, A., Briones, V., Lee, E.Y., Geiman, T., Huang, J., Stephens, R., et al. (2011). Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. *Proc Natl Acad Sci U S A* 108, 5626-5631.

Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., et al. (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nat Commun* 6, 7870.

Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., et al. (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nat Commun* 6, 7870.

Thompson, J.J., Kaur, R., Sosa, C.P., Lee, J.H., Kashiwagi, K., Zhou, D., and Robertson, K.D. (2018). ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. *Nucleic acids research*, 46(19), 10034–10051

Tiepolo, L., Maraschio, P., Gimelli, G., Cuoco, C., Gargani, G. F., & Romano, C. (1979). Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. *Human genetics*, 51(2), 127–137.

Tittle, R. K., Sze, R., Ng, A., Nuckels, R. J., Swartz, M. E., Anderson, R. M., Bosch, J., Stainier, D. Y., Eberhart, J. K., & Gross, J. M. (2011). Uhrf1 and Dnmt1 are required for development and maintenance of the zebrafish lens. *Developmental biology*, 350(1), 50–63.

Tsuda, H., Takarabe, T., Kanai, Y., Fukutomi, T., and Hirohashi, S. (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, 859-866.

Tsuda, H., Takarabe, T., Kanai, Y., Fukutomi, T., and Hirohashi, S. (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, 859-866.

Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K., and Li, E. (2006). Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development* 133, 1183-1192.

Unoki, M., Funabiki, H., Velasco, G., Francastel, C., & Sasaki, H. (2019). CDCA7 and HELLS mutations undermine nonhomologous end joining in centromeric instability syndrome. *The Journal of clinical investigation*, *129*(1), 78–92.

Unoki, M., Funabiki, H., Velasco, G., Francastel, C., & Sasaki, H. (2019). CDCA7 and HELLS mutations undermine nonhomologous end joining in centromeric instability syndrome. *The Journal of clinical investigation*, 129(1), 78–92.

Unoki, M., Sharif, J., Saito, Y., Velasco, G., Francastel, C., Koseki, H., & Sasaki, H. (2020). CDCA7 and HELLS suppress DNA:RNA hybrid-associated DNA damage at pericentromeric repeats. *Scientific reports*, 10(1), 17865.

Verduzco, D., & Amatruda, J. F. (2011). Analysis of cell proliferation, senescence, and cell death in zebrafish embryos. Methods in cell biology, 101, 19–38.

Walsh, C. P., Chaillet, J. R., & Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature genetics*, 20(2), 116–117.

Weemaes, C. M., van Tol, M. J., Wang, J., van Ostaijen-ten Dam, M. M., van Eggermond, M. C., Thijssen, P. E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E., Ferster, A., Furthner, D., Gimelli, G., Gennery, A., Kloeckener-Gruissem, B., Meyn, S., Powell, C., Reisli, I., Schuetz, C., Schulz, A., … van der Maarel, S. M. (2013). Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. *European journal of human genetics* : EJHG, 21(11), 1219–1225.

Wijmenga, C., Hansen, R. S., Gimelli, G., Björck, E. J., Davies, E. G., Valentine, D., Belohradsky, B. H., van Dongen, J. J., Smeets, D. F., van den Heuvel, L. P., Luyten, J. A., Strengman, E., Weemaes, C., & Pearson, P. L. (2000). Genetic variation in ICF syndrome: evidence for genetic heterogeneity. *Human mutation*, 16(6), 509–517.

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

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 402, 187-191.

Yamashita, M., Inoue, K., Saeki, N., Ideta-Otsuka, M., Yanagihara, Y., Sawada, Y., Sakakibara, I., Lee, J., Ichikawa, K., Kamei, Y., Iimura, T., Igarashi, K., Takada, Y., & Imai, Y. (2018). Uhrf1 is indispensable for normal limb growth by regulating chondrocyte

differentiation through specific gene expression. *Development (Cambridge, England)*, 145(1)

G Yu, LG Wang, Y Han, QY He (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, 16(5): 284-287.

APPENDIX

Appendix 1: Mutant Alleles and Genotyping Primers

Name	Mutation	Mutation Type	Genotyping Primer Sequences	Annealing Temp (°C)
mk22	*zbtb24*	~8kb large deletion	F(P1): AGTCCTCGCTCTGCACTCAG WT-R(P3): TCTCGTCCACCAACACGAC	55°C
			F(P1): AGTCCTCGCTCTGCACTCAG Mut-R(P2): CTCTTGGCGGTGAAACACTT	
cdca7	*cdca7*	~6kb large deletion	WT-F: CGTTTCTGTCTTTGGGTGG WT-R: TTGCCGAATCCATCAGAGCC	53°C
			Mut-F: TGGTTCCATTTTGTAAAACATGACC Mut-R: GCGTCACGGACTTCTTCTCC	
m214k	*p53*	A to T transversion at the 3' end	F(P1): GATAGCCTAGTGCGAGCACACTCTT WT-R(P2): AGCTGCATGGGGGGGAT	55°C
			F(P1): GATAGCCTAGTGCGAGCACACTCTT Mut-R(P3): AGCTGCATGGGGGGGAA	

Appendix 2: qRT-PCR primers and primer verification

Name	Primer Sequence
b-actin	F: CGAGCAGGAGATGGGAACC
	R: CAACGGAAACGCTCATTGC
Rad51_set1	F: TGTGGCTGAACGGTATGGTC
	R: GGACTCGGTCATCATAGCGG
Rad51_set2	F: CCCGCCTAGAGCAAAGTGG

	R: CTTGTCAGCTTTGGCTTCGC
Rad52_set1	F: AATACCAAAGCAGCCTCCTCC
	R: CTGCTCTATGGGGATGTTGGC
Rad52_set2	F: CAGCCAACATCCCCATAGAGC
	R: CGTGGTATCAGATGCTGACCTGG
Rad51ap1_set1	F: TTTGCCTGTGTGAAACCTCCT
	R: AGCATCCAGAGACACCCTCTT
Rad51ap1_set2	F: GCTAAAAGCACAGCTCCCAGT
	R: GCAGGTTTGGAGATCGGTGG
Rad51b_set1	F: TCTGTGGCCTCTGTTGTAAGGA
	R: TCTGGTTAGTTAAAACCACCGGG
Rad51b_set2	F: ATTCTGTGGCCTCTGTTGTAAGG
	R: CTGGTTAGTTAAAACCACCGGGAT
Rad51c_set1	F: TGGGGTCTTAATCTGCATTTGAGG
	R: TCTTGAGAAATACCTGCCTCTTTGC
Rad51c_set2	F: CAGTCTGGGGTCTTAATCTGCAT
	R: CTTGAGAAATACCTGCCTCTTTGC
Rad51d_set1	F: GTGGCGGGCGCTAATATGAA
	R: CCACCATGACTCCCAGATTGC
Rad51d_set2	F: GGCTTTCTAGCAATCTGGGAGTC
	R: CAAAAGCACACGCCGGAC

Appendix 3: RNA-sequencing read mapping.

cdca7$^{-/-}$ maternal zygotic dataset

sample	# of input reads	uniquely mapped reads (%)	successfully assigned alignments (#)
AB rep1	18,822,751	91.44%	14,845,979
AB rep2	14,977,766	88.26%	11,256,973
cdca7$^{-/-}$ MZ rep1	14,598,396	91.58%	11,346,421
cdca7$^{-/-}$ MZ rep2	13,920,704	92.29%	11,066,310
cdca7$^{-/-}$ MZ rep3	16,259,190	91.11%	13,123,889

zbtb24$^{-/-}$ dataset

sample	# of input reads	uniquely mapped reads (%)	successfully assigned alignments (#)
AB rep1			
first of pair	39,619,488	94.27%	37,350,921
second of pair	39,619,488	94.27%	37,350,921
AB rep2			
first of pair	32,686,731	95.00%	31,053,682
second of pair	32,686,731	95.00%	31,053,682
zbtb24$^{-/-}$ rep1			
first of pair	36,237,006	94.76%	34,336,606
second of pair	36,237,006	94.76%	34,336,606
zbtb24$^{-/-}$ rep2			
first of pair	33,089,362	95.07%	31,459,602
second of pair	33,089,362	95.07%	31,459,602
zbtb24$^{-/-}$ rep3			
first of pair	35,553,797	94.77%	33,693,789
second of pair	35,553,797	94.77%	33,693,789