GENE EDITING OF *dnd* BY TALENS IN MEDAKA EMBRYOS AND ITS ROLE IN PGC FORMATION

WANG TIANSU

(B.Sc., Sun Yat-Sen University)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2015

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Wang Tiansu
21 Jan 2015

Acknowledgement

I would like to express my great appreciation to Associate Professor Hong Yunhan, my supervisor, for his patient guidance, enthusiastic encouragement and useful critiques of this research work.

I would also like to thank Professor Wang Shu, Dr He Yuehui and Assistant Professor Wu Min for being my Qualifying Examination committee members; Dr Gen Hua YUE and Associate Professor Chan Woon Khiong for being my Thesis Advisory Committee members. They have given me great critiques and suggestions that benefit my research.

Special thanks should be given to Dr Li Zhendong, Dr Li Mingyou, Dr Yuan Yongming and Dr Yan Yan for their mentorship on my experimental techniques and guidance on my research. I would like to extend my thanks to my labmates, Chen Jianbin, Zhang Xi, Zhu Feng and Liu Qizhi, for their help on my experiments, the discussion on my research projects and supports during hard times.

I would also like to extend my thanks to Ms Deng Jiaorong and Mr Zeng Qinghua for their caring on the fish, and to Ms Veronica Wong and Ms Foong Choy Mei for their help on administrative issues.

Finally, I wish to thank my family for their support and encouragement throughout my study. I greatly thank my beloved wife Huang Yuhan, my mom and my mother-in-law for taking good care of my son for the previous year, allowing me to focus on my thesis work.

Table of contents

SUMMARY

Targeted gene editing (GE) is powerful for generating genetic alterations in animal genomes. Programmable endonucleases such as zinc finger nucleases, transcription activator-like effector nucleases (TALENs) and CRISPR/Cas allow for GE directly in animal embryos and germ line transmission. In this thesis, procedures and parameters of TALEN-mediated GE of a germ line-specific gene *dead end* (*dnd*) in the fish medaka were examined. Embryos at the 1-cell stage were injected with TALEN mRNAs and examined for the survival rate and GE efficiency. Medaka embryos can tolerate a high dosage of TALEN-mRNA injection and exhibit a steadily increasing GE efficiency with the increasing of mRNA dosages until peaking at 100 ng/μl. This dosage produced ~24% efficiency for somatic GE. Several *dnd* GE families with various targeted alleles were established. These data demonstrate that TALEN is proficient for somatic and germ line GE in medaka embryos.

dnd was first identified in zebrafish as a germline-specific gene that encodes an RNA-binding protein essential for primordial germ cell (PGC) development. *dnd* exists in many other vertebrates and is essential for gametogenesis in mouse. In this thesis, the role of *dnd* in medaka germ cell development was analyzed by TALEN-mediated GE. A frame-shift mutant allele containing a deletion of 10 bp in exon 2 of *dnd* was chosen as a null mutant for fish family production and phenotypic analyses. Embryos heterozygous for this null mutant allele were able to grow into adult and fertile female and male. Crossing between heterozygotes led to the

segregation of 3 genotypes (*dnd*$^{+/+}$, *dnd*$^{+/-}$ & *dnd*$^{-/-}$) in a Mendelian manner at embryonic, larval and adult stages. A considerable difference in the number of PGCs was observed depending on the *dnd* genotype. At day 2 post fertilization, embryos of both wild type and mutant embryos reach the somitogenesis stage (6 somites). At this stage, a wild type embryo from a control mating between a normal male and a normal female has 41 PGCs on average, whereas embryos from a cross between heterozygous animals have fewer PGCs. Specifically, *dnd*$^{+/+}$ and *dnd*$^{+/-}$ embryos averagely have 28 and 27 PGCs, whereas this number decreases significantly to 13 in *dnd*$^{-/-}$ embryos. Wild type and mutant embryos did not show a difference in the position of PGCs, as they were aligned dorsobilaterally to the embryonic axis irrespective of *dnd* genotype. These results demonstrate an essential role of *dnd* in PGC formation.

Interestingly, PGCs in *dnd*$^{-/-}$ embryos disappeared at day 5 post fertilization when PGCs were clearly visible by transgenic GFP expression and RNA in situ hybridization in the developing gonad of *dnd*$^{+/+}$ and *dnd*$^{+/-}$ embryos, demonstrating an essential role of *dnd* in postmigratory PGC survival. These embryos completely free of PGCs are able to develop into adults that possess a severely underdeveloped gonad and show germ cell-less sterility. These results demonstrate that medaka *dnd* is dispensable for somatic development but indispensable for germline development such as PGC formation and survival in particular.

List of Tables

List of Figures

List of Abbreviations

AP	Alkaline phosphatatse
APS	Amomonium persulfate
BCIP	5-bromo, 4-chloro, 3-indolylphosphate
bp	Base pairs
BSA	Bovine serum albumin
cDNA	DNA complementary to RNA
CMV	Cytomegalovirus
CRISPR	Clustered regularly interspaced short palindromic repeats
Cas	CRISPR associated nuclease
dpf	days post fertilization
DEPC	Diethyl pyrocarbonate
DIG	Digoxygenin
DNA	Deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
eGFP	enhanced green fluorescent protein
ERM	Embryo rearing medium
FBS	Fetal bovine serum
GE	Gene editing
GT	Gene targeting
HD	RVD consists of Histidine and Aspartic acid
HMA	Heteroduplex mobility assay
HRP	Horseradish peroxidase
ISH	in situ hybridization
kb	kilobase
LB	Luria-bertani medium
LiCl	Lithium Chloride
MO	Morpholino oligo
mRNA	messenger RNA
NG	RVD consists of Asparagine and Glycine
NI	RVD consists of Asparagine and Isoleucine
NK	RVD consists of Asparagine and Lysine
NS	RVD consists of Asparagine and Serine
N*	RVD consists of only Asparagine
NBT	nitro blue tetrazolium chloride

Ol	Oryzias latipes
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PGC	Primodial germ cell
PFA	Paraformaldehyde
POD	Peroxidase enzyme
PTK	Proteinase K
PBT	PBS with 0.1% Tween-20
RNA	Ribonucleic acid
RFP	Red fluorescent protein, DsRed
RT-PCR	Reverse Trancriptase Polymerase Chain Reaction
RVD	repeat variable di-residue
SDS	Sodium dodecylsulfate
SSC	Sodium chloride tri-sodium citrate solution
TAE	tris-acetate EDTA
TALE	Transcription activator-like effector
TALEN	TALE nuclese
TBS	Tris buffered saline
TBST	Tris buffered saline with 0.1% Tween-20
TEMED	N,N,N’,N’-tetramethylethylene diamine
WISH	Whole-mount in situ hybridization
ZFN	Zinc finger nuclease

CHAPTER 1. INTRODUCTION

1.1. Genome editing

1.1.1. Overview: From gene targeting to genome editing

Gene targeting has been the gold standard of examining a certain gene's function. By completely removing a gene from the genome, gene targeting reveals what will happen to the organism without this gene. This way, the function of the gene can be convincingly demonstrated. Gene targeting utilizes homologous recombination to locate a specific gene out of the enormous genome by homologous sequence pairing and manipulates that gene in the way defined by design of the homologous sequence: knock-in, knock-out or both (replacement). Capecchi and his colleagues generated the first gene-targeted cell line in 1986 by correcting a previously introduced mutated *neo*^{r} gene (Thomas et al., 1986). Then they applied this technology on generating gene targeted mouse embryonic stem (ES) cells. In 1987, *hprt* gene was knocked out using a selection method that utilized the *hprt* gene's own function, which was using 6-thioguanine to kill the cells with untargeted *hprt* (Thomas and Capecchi, 1987). The frequency of homologous recombination was around 10^{-4} when using a targeting vector containing 14 kb of homologous sequence to target the *hprt* gene (Deng and Capecchi, 1992). Considering the ES cell line used was of male karyotype, it was hemizygous for *hprt* gene. Thus, the projected frequency of

homologous recombination for other genes that have two copies in the genome is even lower at 10^{-8}. Therefore, positive and negative selection was introduced to enrich gene-targeted cells (Mansour et al., 1988). In detail, G418 serves as the positive selection that kills cells without the recombination of *neo*^{r}-containing targeting vector. Gancyclovir serves as the negative selection that kills cells with the integration of HSV-*tk* gene, which is located outside the homologous recombination cassette on the targeting vector.

Subsequently, they injected gene-targeted embryonic stem (ES) cells into developing mouse embryos to form chimera mice. Some of the gene-targeted ES cells were able to contribute to the germline of the embryos (Thompson et al., 1989). So these mice were able to produce the first batch of gene-targeted mice among their offspring (Chisaka et al., 1992). However, gene targeting by homologous recombination was not achievable for other animals for a long time. The reason is simple: lack of germline-competitive ES cells. Gene targeting has to be performed in cell lines to allow for positive and negative selection of gene-targeted clones. In the same time, the ES cells must be germline-competitive to allow for the production of gene-targeted offspring. Rat has been the only other organism to have germline-competitive ES cells and thus gene targeting since mouse. Qi-long Ying and his colleagues generated p53 knockout rats by injecting gene targeting rat ES cells into rat embryos in 2010, which was almost two decades since the generation of knockout mice (Tong et al., 2010).

Although germline competitive ES cell is not an option for most organisms, a new kind of tools emerged to bypass this prerequisite and enables gene targeting in almost all organisms that reproduce with embryos. They are called programmable endonucleases. These nucleases can be engineered to specifically target almost any genomic locus and create double-strand DNA breaks (DSB). Before the discovery of programmable endonucleases, I-SceI has been used to prove that DSB greatly promotes homologous recombination based gene targeting, whose inherent frequency is very low, in drosophila, yeast and mouse and human ES cells (Choulika et al., 1995; Cohen-Tannoudji et al., 1998; Donoho et al., 1998; Gloor et al., 1991; Smih et al., 1995). This means the programmable endonucleases can enhance the efficiency of gene targeting to any locus of interest by creating sequence-specific DSBs. Because the gene targeting efficiency has been promoted so greatly, the positive-negative selection procedure is no longer necessary. Therefore, programmable endonucleases together with desired homologous recombination template can be microinjected into embryos to produce gene-targeted chimeras, which are comparable to those generated by injecting gene-targeted germline-competitive ES cells. Other than homologous recombination, programmable endonucleases can trigger another DNA repair mechanism when no homologous repairing template is present, which is called non-homologous end joining (NHEJ). This mechanism joins the broken DNA ends in an imperfect way, leaving a few bps' error. The error may disrupt the open reading frame of the target gene and produce knockouts of organisms.

With the invention of programmable nucleases, manipulation of chromosomal genes has been expanded a lot more than what homologous recombination based gene targeting was capable. Therefore, new names, such as gene editing, genome editing and genome engineering, have been given by the research community to research projects that use programmable endonucleases to manipulate genome in a defined manner. Detailed new approaches of genome engineering will be introduced after the introduction of various programmable endonucleases.

1.1.2. Zinc finger nucleases

Up until now, three kinds of programmable endonucleases emerged and prosper: zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and RNA-guided engineered nucleases (RGENs) derived from the bacterial clustered regularly interspaced short palindromic repeat (CRISPR)/Cas (CRISPR-associated) system.

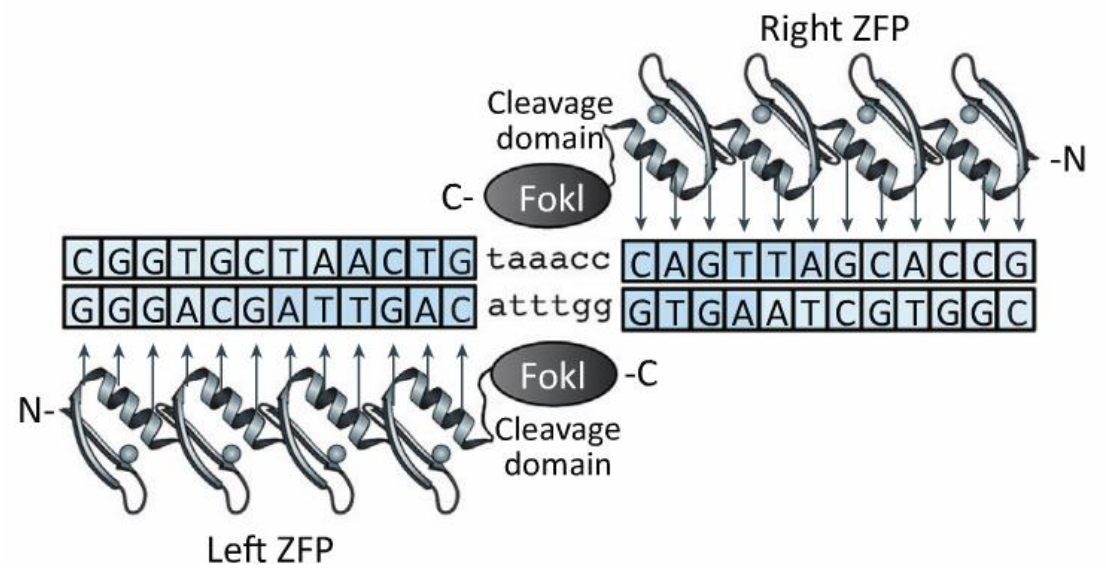


Figure 1.1. Cartoon of a ZFN dimer bound to DNA. Adapted from (Gaj et al., 2013).

Zinc finger proteins are among the most abundant proteins in eukaryotic genomes. They consist of zinc fingers that are the most common DNA recognition motif in all metazoan. Zinc fingers' structures are much diversified so that they can recognize equally diversified DNA sequences (Laity et al., 2001). Each zinc finger motif was found to bind to 3 nucleobases of DNA (Pavletich and Pabo, 1991). In 2003, Bibikova and her colleagues first engineered a pair of ZFNs that targets gene *yellow* in drosophila, and reached a gene targeting efficiency of 2% (Bibikova et al., 2003). Usually, three to four arrayed zinc finger motifs are fused together and tagged with a C-terminal FokI nuclease monomer to become a ZFN, reaching a sequence-recognition specificity of 9 to 12 bps (Fig. 1.1). By design, a pair of ZFNs will reciprocally bind to each side of the target site, spaced by several bps. This is to let the FokI nuclease monomers to form a dimer in between the zinc finger motifs, since the FokI nuclease only cut the DNA to generate a DSB after dimerization. In total, two ZFNs bind to 18 to 24 bps of sequence, which is, in principal, sufficient to

pick out a single target site even in a complex genome. However, in practical, the design and construction of zinc finger nucleases was not that easy. Simply fusing of the fingers that recognize target sequence doesn't always work. Several combinations need to be tested on yeast system and the best can be selected (Carroll et al., 2006; Mani et al., 2005). Accumulative evidences showed that zinc fingers not only bind to three nucleobases, but also interact with one more nucleobase, which is opposite to the nucleobase that is adjacent to the 3'- end of the three nucleobases recognized by the zinc finger(Isalan et al., 1997; Isalan et al., 1998). This has been the cause of incompatible zinc fingers in ZFNs. In 2011, Sander and his colleagues reported a method called context-dependent assembly (CoDA) to solve this problem (Sander et al., 2011). It was realized that design and construction of ZFNs can be much easier by referring to previously proven zinc finger arrays. Thanks to the development of ZFN community, there has been enough data to be referred to. The sequence of two zinc fingers is called a 'context'. If a context is already proven by a published ZFN, then the context can be used to combine with other contexts to construct new ZFNs. For example, if two ZFN arrays, one consisting zinc fingers A-B-C and the other consisting of zinc fingers D-A-E, was reported, then the contexts 'A-B' and 'D-A' are considered solid. A new array consisting of zinc fingers D-A-B can be constructed without testing the efficiency, because contexts A-B and D-A have been confirmed to be working.

1.1.3. Transcription activator-like effectors nucleases

Transcription activator-like effectors (TALEs) are DNA binding proteins found in plant bacteria *Xanthomonas*. The bacteria utilize a Hrp-type III secretion system to inject TALEs into the cytoplasm of plant cell. TALEs will be transport into plant cell's nucleus due to the nuclear localization signal near its C-terminus. They naturally bind to promoter regions of plant genes after localize to nucleus, just like transcriptional activators. The cellular environment will be altered to weaken the defense of plant cell, fit for bacteria proliferation, spreading and etc. (Sugio et al., 2007; Yang et al., 2006). The reason that TALEs can target various plant gene promoters is that it can alter its modular repeats to target different sequences.

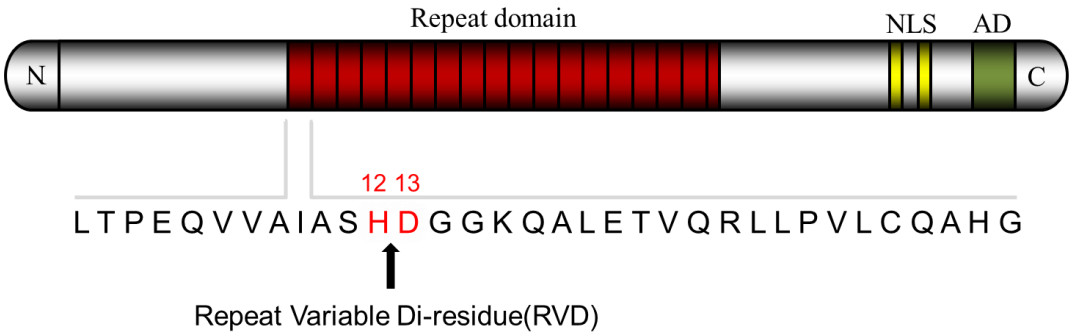
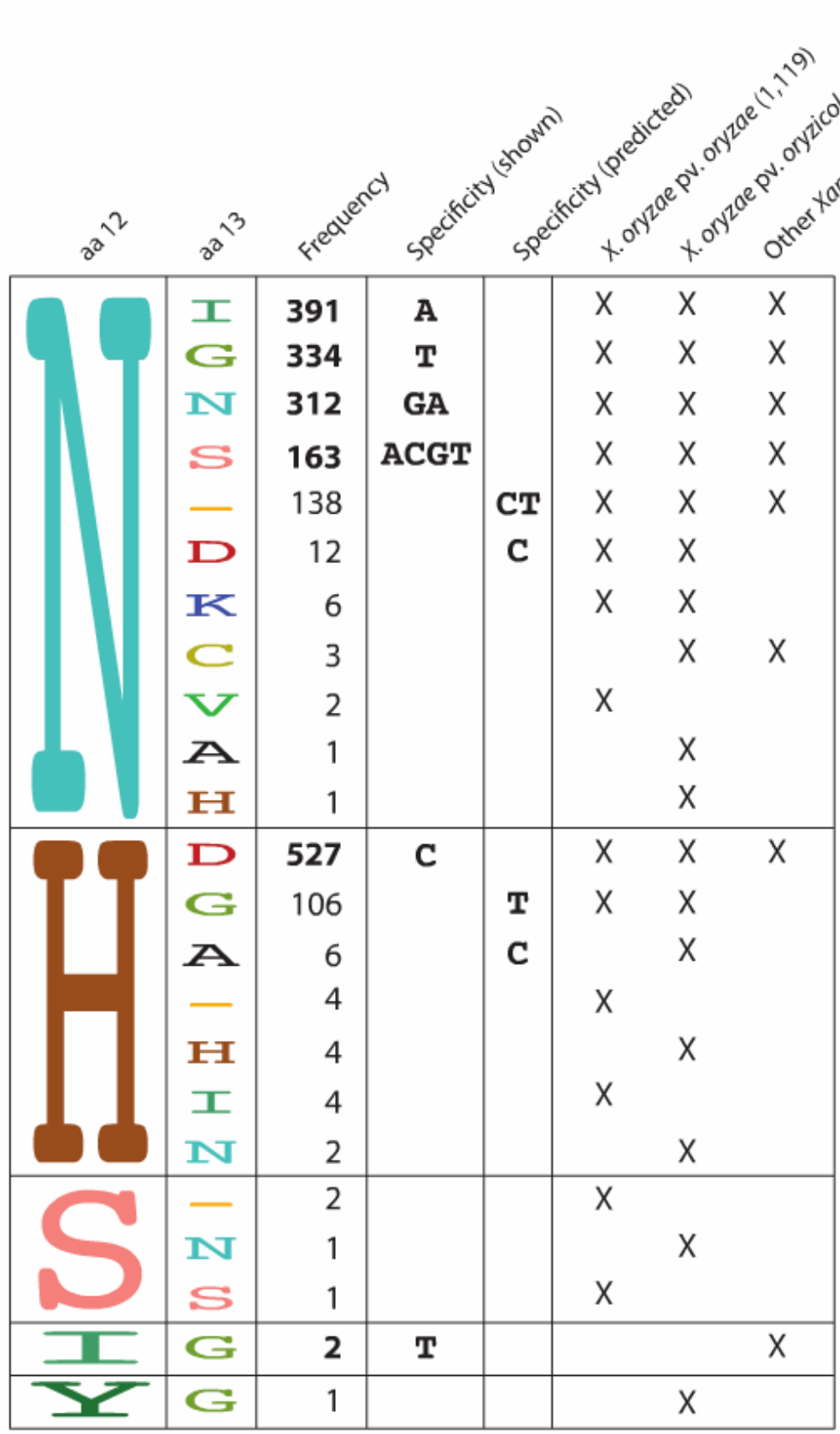


Figure 1.2. Cartoon of a native TALE. Adapted from (Boch et al., 2009b).

The TALE has a big family. Typical TALEs have a 283-290 aa N-terminus, a 274-297 aa C-terminus and the DNA binding domain of varying length in the middle (Boch and Bonas, 2010) (Fig. 1.2). It has been found out that the end of N-terminus is responsible for translocation by Hrp-type III secretion system (Szurek et al., 2002a). It has also been found out that the nuclear localization signal (NLS) locates

near the C-terminus (Yang and Gabriel, 1995), while the transcription activation domain resides at the end of C-terminus (Zhu et al., 1998).

The domain consisting tandem repeats in the middle of the TALEs are responsible for DNA recognizing (Herbers et al., 1992), turning out to be a novel type of modular DNA binding domain. It was found out that each one of the repeats is responsible for one base pair recognition, while the recognition sequence of TALEs always starts with a T, probably determined by the residues to the right of their DNA binding domain (Boch et al., 2009b; Moscou and Bogdanove, 2009). The tandem repeats, except for the last one, consist of various numbers of amino acids ranging from 30 to 42 aa, while the predominant majority of them consist of 34 amino acids (Boch and Bonas, 2010). The last repeat in the DNA binding domain only consists of 19 or 20 aa, seemingly a truncated half repeat. Within 34 aa repeats, the repeat units from across different family members are highly conserved, except for position 4, 12, 13 and 32 amino acids.



aa 12	aa 13	Frequency	Specificity (shown)	Specificity (predicted)	*X. oryzae* pv. *oryzae* (1,119)	*X. oryzae* pv. *oryzicola* (475)	Other *Xanthomonas* pvs. (429)
N	I	**391**	**A**		X	X	X
	G	**334**	**T**		X	X	X
	N	**312**	**GA**		X	X	X
	S	**163**	**ACGT**		X	X	X
	—	138		**CT**	X	X	X
	D	12		**C**	X	X	
	K	6			X	X	
	C	3				X	X
	V	2			X		
	A	1				X	
	H	1				X	
H	D	**527**	**C**		X	X	X
	G	106		**T**	X	X	
	A	6		**C**		X	
	—	4			X		
	H	4				X	
	I	4			X		
	N	2				X	
S	—	2			X		
	N	1				X	
	S	1			X		
I	G	**2**	**T**				X
Y	G	1				X	

Figure 1.3. RVD types in *Xanthomonas*. Adapted from (Boch and Bonas, 2010).

For most of the repeats, including the half ones, the 12th and 13th amino acids are hypervariable and together determine the specificity of the DNA recognition (Fig. 1.2) (Boch et al., 2009a; Moscou and Bogdanove, 2009). This pair of amino acids is called repeat variable di-residue (RVD). There are various compositions of RVDs found in the TALE family (Fig. 1.3). These RVDs have been allocated target nucleotides based on the target sequences of the TALEs where the RVDs were found. However, because of the abundance of RVDs varies, the deduction of RVD to nucleotide relationship is not always reliable because of sample size. Only the predominantly abundant RVDs have proven themselves in the tests of nucleotide-binding specificity (Streubel et al., 2012). There are 4 abundant RVDs that have experimentally proven specificity for binding DNA: NI, corresponding to A; NG, corresponding to T; NN, corresponding to both G and A, with a slight preference towards G; HD, corresponding to C (Boch et al., 2009a; Moscou and Bogdanove, 2009). NS is also an experimentally proven RVD, but it doesn't have a preference for recognizing nucleotides. Other abundant RVDs N* (losing the 13th aa) and HG only have the predicted binding preference but haven't been experimentally proven. Taken together, tandem repeats in the DNA binding domain of TALEs have a range of RVDs that will cover binding specificity for all 4 types of DNA nucleotides.

Since the repeats of TALE binding domain are highly modular and the RVD to DNA code has been deciphered, TALEs can also be engineered to be TALE nucleases (TALENs) similar to the way that ZFNs are engineered (Fig. 1.4). The early reported TALENs comprise full length of TALE's N- and C-terminus except

for the transcription activation domain for activating plant gene's transcription, which is not necessary for the DNA binding property of TALE (Cermak et al., 2010). A FokI nuclease is used to replace the activation domain of TALE to form a TALEN, a pair of which will reciprocally pair at the C-terminus to form a FokI dimer upon binding to target DNA sequences. The spacer sequence between the two TALEN target sequences will be cut by the FokI dimier. The N-terminus serves as a linker between the DNA binding domain and the FokI nuclease. This linker is very long as

Figure 1.4. Cartoon of a TALEN dimer bound to DNA.

compared to ZFN's, leading to a longer spacer between two TALEN targeting sequence (Cermak et al., 2010). Later on, researchers have tested truncated versions of TALEs in the constructions of TALENs, deleting various lengths of the N-terminus and C-terminus in TALEs, since most of the amino acids there are not relevant to their DNA binding ability. It has been shown that TALENs with truncated versions of N-terminus consisting of the first 153 or 147 residues to the right of the DNA binding domain gave rise to targeting efficiencies that were not significantly different from the full-length ones (Miller et al., 2011). On the other hand, it has been shown that TALENs with truncated versions of C-terminus consisting of 17, 28, 47 or 63 residues to the left of the DNA binding domain possessed similar targeting

efficiencies to the full-length ones when the spacer was 15 bp (Zhang et al., 2011). After these studies, new designer-TALENs are 200 residues shorter than the full-length ones, which at least reduce 20% of the size, making designer TALENs more compact.

Comparing to ZFN, each repeat of a TALEN in the DNA binding domain Has a one-to-one correlation to the nucleotides in the target binding sequence, thus TALENs are easier to design. Moreover, newly synthesized TALENs have a very high successful rate. In a report of high-throughput synthesis of TALENs, out of 96 pairs of TALENs targeting 96 different genes, 84 of them showed efficient NHEJ-mediated mutagenesis at their intended target sites, which is 88% (Reyon et al., 2012). However, ZFNs have a more compact protein structure, making ZFNs usually only one-third to half of TALENs in regard of the length of coding sequences or amino acids.

1.1.4. RNA-guided engineered nucleases

CRISPR/Cas system is found among the adaptive immune systems in bacteria. In bacteria, it recognizes and stores foreign DNA sequences in between clustered regularly interspaced short palindromic repeats and transcribes them as CRISPR RNAs (crRNAs) to lead the CRISPR associated nucleases (Cas) to target the foreign DNA upon next infection of the same pathogen (Barrangou et al., 2007; Brouns et al., 2008; Horvath and Barrangou, 2010). Researchers have utilized this mechanism and

make Cas a new kind of programmable nuclease (Jinek et al., 2012b). They also invented a hybrid of the two essential non-coding RNAs in the CRISPR/Cas system (Fig. 1.5). Specifically, minimal regions of crRNA and trans-activating crRNA (tracrRNA) were fused by a linker loop, making a single chimeric guide RNA (gRNA) that is able to guide the Cas. Engineered gRNA can be used to direct Cas to any target sequences that end with a protospacer adjacent motif (PAM), which is NGG for *Streptococcus pyogenes*. The DNA recognition in this system is RNA to DNA, different from the previous protein to DNA recognitions. This property makes the designing and engineering of gRNA much easier than engineering ZFN and TALEN.

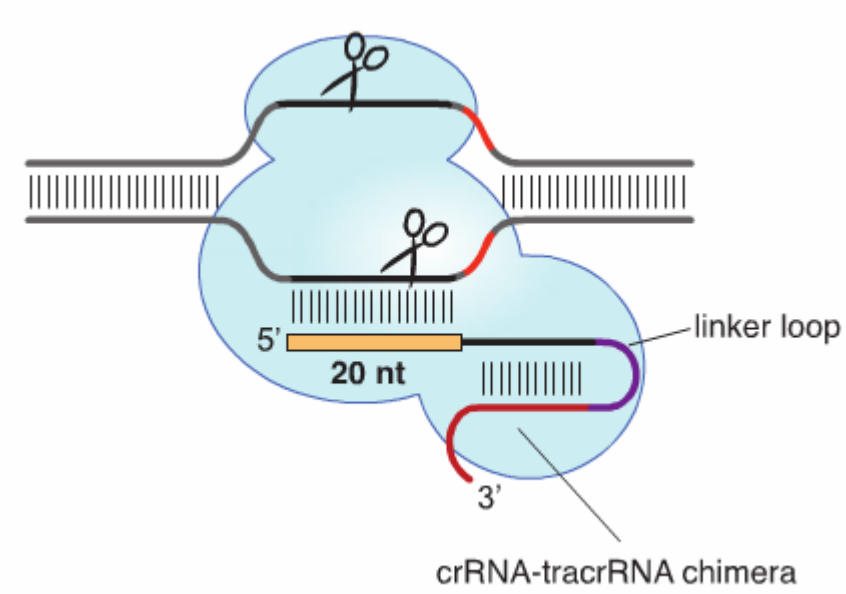


Figure 1.5. Cartoon of a Cas9 programmed by single chimeric RNA. Adapted from (Jinek et al., 2012a).

It has been reported that this system has a very high efficiency, enabling multiplex genome engineering in the same time (Cong et al., 2013), but may also possess a higher off-targeting probability (Fu et al., 2013; Hsu et al., 2013). The

higher off-targeting rate is partially due to the fact that, within the target sequence, only a seed sequence of 10 to 12 bp from PAM strongly restricts the specificity of CRISPR/Cas cleavage. Obviously, 10 to 12 bp is not enough for achieving satisfactory specificity in most vertebrate genomes. The off-targeting rate can be reduced by using two CRISPR/Cas nickase to reciprocally target two nearby sequences like ZFNs and TALENs, increasing the seed sequence to 20 to 24 bp (Scott et al., 2013). However, the two target site cannot be separated for too far away for the binding affinity between the two strands of DNA increases with the increase of nucleotides between the two nicks, making the design of gRNA to match PAM harder.

1.1.5. Genome engineering approaches

Empowered by programmable endonucleases, many new ways of manipulation of the genome has been developed, making genome engineering much more versatile. These ways of manipulation of genome can be generally allocated to two categories by the DNA repair mechanism that they utilize. One is non-homologous recombination end joining (NHEJ), the other is homologous dependent repairing (HDR).

NHEJ based genome engineering includes gene disruption, large fragment deletion and large fragment inversion. NHEJ joins broken DNA ends in an imperfect way, which will leave errors ranging from several to a few bps (Moore and Haber,

1996). If a programmable endonucleases-generated error is created at the beginning of the target gene's open reading frame, there will be a chance that the open reading frame is disrupted by a frame-shift mutation. Given that genetic codons consist of nucleotides in triplets, the chance of gene disruption should be 2 out of 3 times. Other than inducing frame-shift mutation, programmable endonucleases can also be used to delete large fragment of several to dozens kb from genome, by creating two DSBs flanking the large fragment that is to be deleted (Wang et al., 2013; Xiao et al., 2013). Another outcome of this creation of two DSBs is large-fragment inversion. Instead of joining the two ends of cleaved genome, there is a chance that the fragment is ligated back to the genome but in an inverted manner (Xiao et al., 2013).

HDR based genome engineering can be categorized according to the kind of donor DNA template used: double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA). Similar to traditional gene targeting, dsDNA donor triggers homologous recombination and swap genetic information into or out of the target site. However, for the programmable endonucleases-aided genome engineering, donors containing around 3 to 5 kb of homology are able to produce a much higher efficiency of gene targeting (Yang et al., 2014). This method is of most fidelity in genome engineering. Conditional knockout/in can be achieved by adding *loxP* sites to both sides of the targeting cassette in the donor. Short ssDNA is also reported to be able to direct HDR of the target site of programmable endonucleases (Bedell et al., 2012). Target sequence of dozens of bp can be manipulated using this method. An ssDNA with 60 bp homology arm on each side of the cassette, which is also several dozens of bp

long, can induce HDR upon DSB. By this method, *loxP* site, tags and restriction enzyme sites can be introduced to the target site with a high fidelity.

A newly emerged way of genome engineering utilizes microhomology-mediated end joining (MMEJ) to perform targeted knock-in into DSB site with the direction in control. Principally, 100 bp of homology sequence, with 50 bp homology sequence on each side, is enough to trigger the MMEJ pathway to repair the DSB in accordance to the 750 bp-cassette flanked by the 50 bp homology sequence (Orlando et al., 2010). The achieved integration efficiency is 10%, with half of the deletion sites inserted with the microhomology-cassette. In another report that uses TALENs and CRISPR instead of ZFN, only a microhomology of 5 to 25 bp in each side of the 1600 bp-targeting cassette achieved a 10% knock-in efficiency, which was not including the incomplete inserts (Nakade et al., 2014).

In conclusion, programmable endonucleases are powerful tools for modern life sciences and are rapidly expanding their application area. They are still continuing to grow and improve, enabling various ways of genome engineering and making great contributions to discoveries and inventions.

1.2. Medaka germline development and germ plasm

1.2.1. Medaka primordial germ cell specification and migration

Germ cells are the only cell type to be able to transmit genetic information through generations. The germline development of animals starts from primordial germ cell (PGC) formation, which is a set of cells that are set aside early in the embryonic development. There are two types PGC formation modes: preformation and epigenesis (Extavour and Akam, 2003). For example, in drosophila, PGC specification is categorized as preformation (Fig. 1.6, left). Maternally synthesized germline determinant (blue) containing proteins and mRNAs is deposited to the posterior pole of the embryo and dictates the formation of PGCs (blue in red circles) in the posterior pole later. These germline determinants are called germ plasm. Generally, germ plasm distribution dictates PGC specification in animals using preformation mode. In mouse, PGC specification is categorized as epigenesis (Fig. 1.6, right). There is no maternally deposited germ plasm observed in mouse. Instead, after the segregation of embryonic and extraembryonic tissues, a subpopulation from the pluripotent epiblast cells expresses some genes that are competent for germline commitment (striped circles) and are able to differentiate into PGCs (red circles) after receiving inductive signals from the extraembryonic ectoderm (blue) and endoderm (yellow).

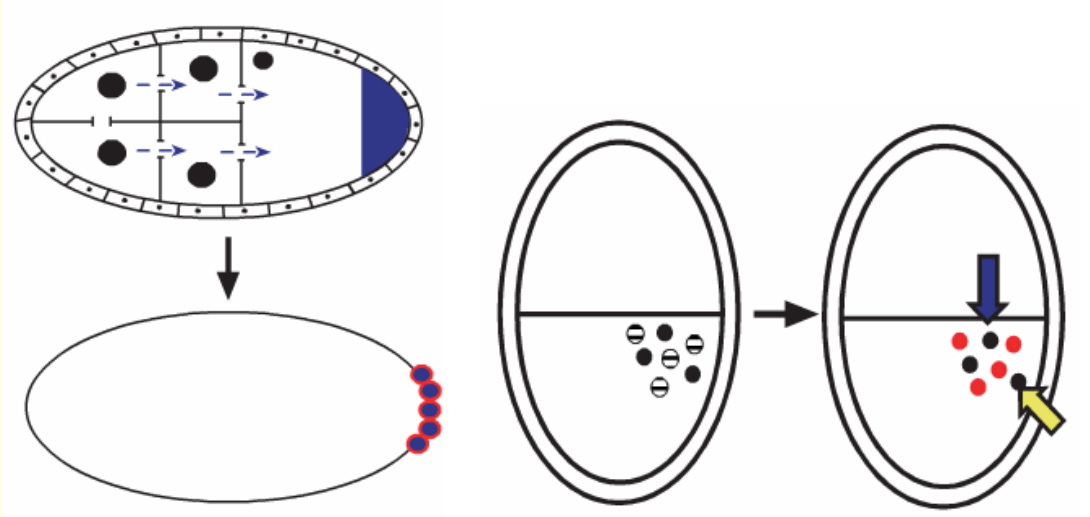


Figure 1.6. PGC specification of drosophila and mouse. Adopted from (Extavour and Akam, 2003).

Medaka germline specification is categorized as preformation. Therefore, there is maternally deposited germ plasm in medaka embryos. Cells who inherit germ plasm in the animal pole of early blastula (stage 10) become PGCs (Fig. 1.7, top left) (Kinoshita et al., 2009). These PGCs will remain there until gastrulation starts, which means the beginning of PGC migration. PGC migration is part of medaka germline development. PGCs will migrate into the gonadal primordium and develop with somatic gonadal cells to form gonad. The PGC migration is majorly controlled by a simple ligand-receptor interaction (Kurokawa et al., 2006). The ligand is a chemokine expressed in the lateral plate mesoderm, called SDF-1a. The receptor is CXCR4 from PGC side. During PGC migration, the SDF-1a's lateral plate mesoderm expression is restricted to the region with gonadal somatic precursors. The SDF-1a is the guiding signal sent from these precursors to tell where the PGCs should migrate towards. During gastrulation, PGCs actively migrate from the animal

pole to the dorsal half of the marginal zone within hours, where SDF-1a expression is concentrated (Fig. 1.7, stage 15). During the late gastrula stage to early neurula stage (Fig. 1.7, stage 16-17), the lateral plate mesoderm moves medially, driving PGCs toward the embryonic body by convergent extension. Then, PGCs align along the SDF-1a-expressing lateral plate mesoderm. The SDF-1a expression domain becomes restricted to the posterior part of the lateral plate mesoderm as somitogenesis proceeds (Fig. 1.7, stages 19-26), directing the PGCs to the posterior direction along the body axis.

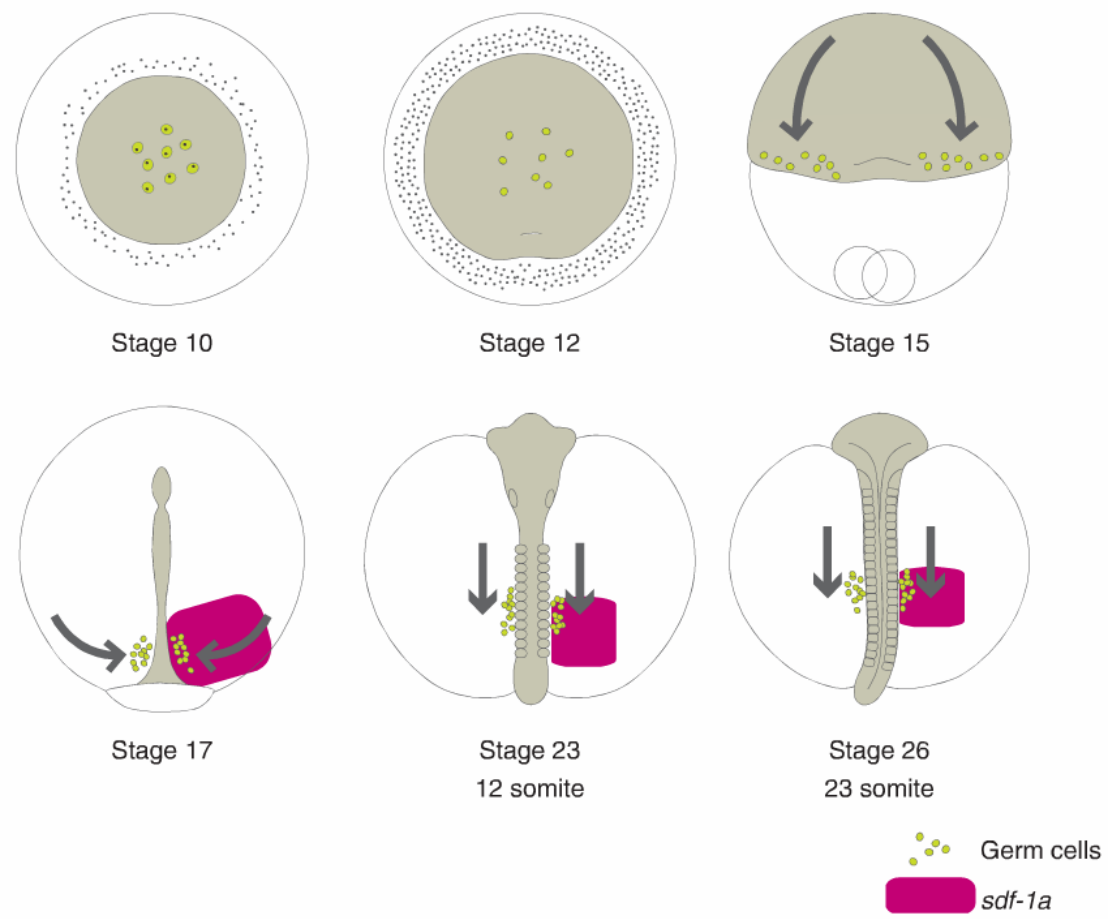


Figure 1.7. Medaka PGC migration (stages 10-26). Adopted from (Kinoshita et al., 2009).

PGCs arrive at the level of the 11th to 13th somites in the lateral plate mesoderm at around stage 26. This is the place where gonadal somatic precursors reside ((Nakamura et al., 2006)). They are spatially organized into two populations, which occupy the anteromedial and posterolateral parts of this region. Then the anterior population of gonadal somatic cells moves towards the middle of the body axis onto the dorsal side of the hindgut, which is the prospective gonadal region (Fig. 1.8, stage 27-30). In the same time, the posterior population goes to the lateral sides of the hindgut in the following of anterior population. After this, the two populations

express distinct marker genes: the anterior one expresses *sox9b*, and the posterior one expresses *ftz-f1*. In the same time, the PGCs also reach the region of posterior population and go on to the anterior population region, which is on the dorsal side of hindgut. The anterior sox9b-expressing population then envelops the PGCs during stages 31-33 (Fig. 1.8, top right). Now the gonadal primordium formation completes, and will separate into two lobes during stage 33 to 35 (Fig. 1.8, bottom). Upon this point, the medakas PGCs complete their migration.

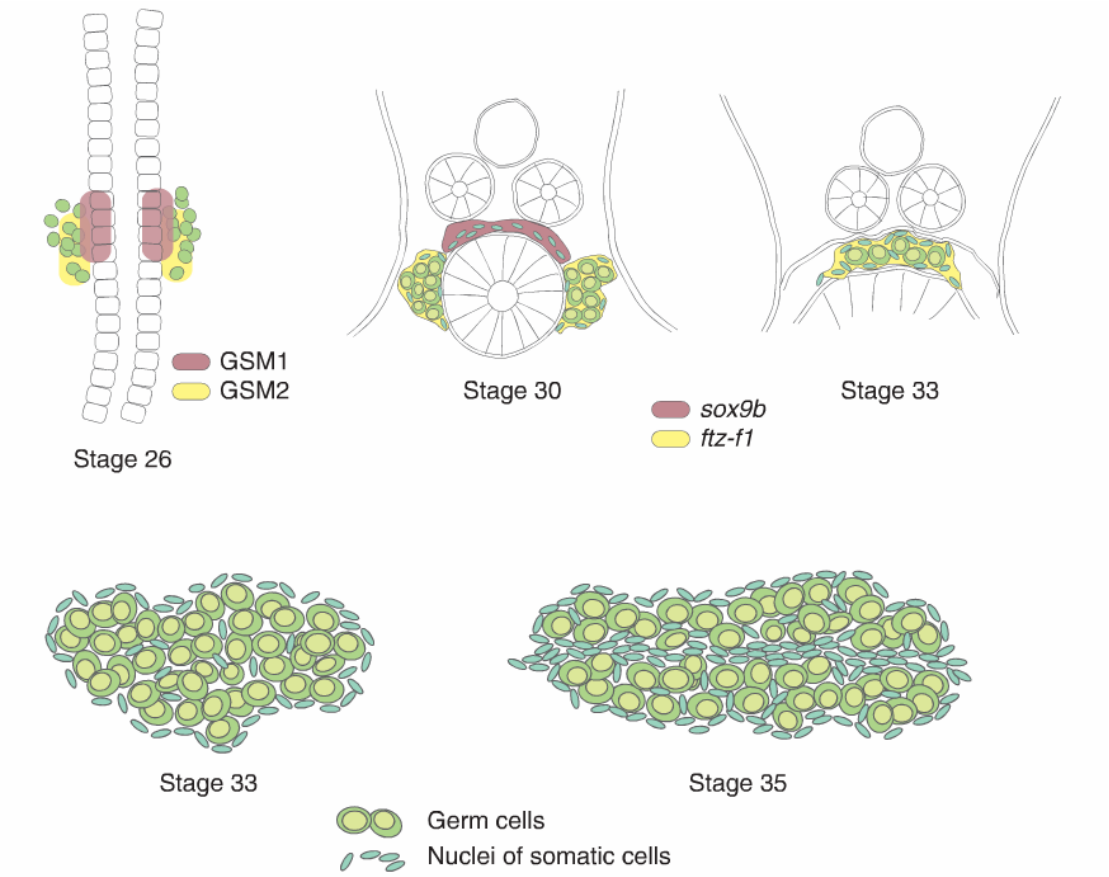


Figure 1.8. Medaka PGC migration (stages 26-35). Adopted from (Kinoshita et al., 2009).

1.2.2. Medaka *dead end* and germ plasm

Gene *dead end* (*dnd*) was first identified as a maternal gene that encodes RNA and protein as components of germ plasm in zebrafish (Weidinger et al., 2003). It was found that ablation of dnd induces sterility in zebrafish (Weidinger et al., 2003) and high frequency of testicular germ-cell tumors in mouse (Youngren et al., 2005). Further investigation into the mechanism of the phenomena showed that Dnd protects mRNAs from degradation by blocking the microRNA binding sites at the 3'-UTR of the target genes (Fig. 1.9) (Ketting, 2007). By dnd protein: messenger

RNAs of Germ plasm genes *nanos1* and *tdrd7* were protected from miR-430 mediated degradation in zebrafish PGCs; Oncogene miR-221 were prevented from degrading mRNA of p27 in human germline cells; and miR372 were prevented from degrading mRNA of LATS2 in human Ter1 cells (Kedde et al., 2007).

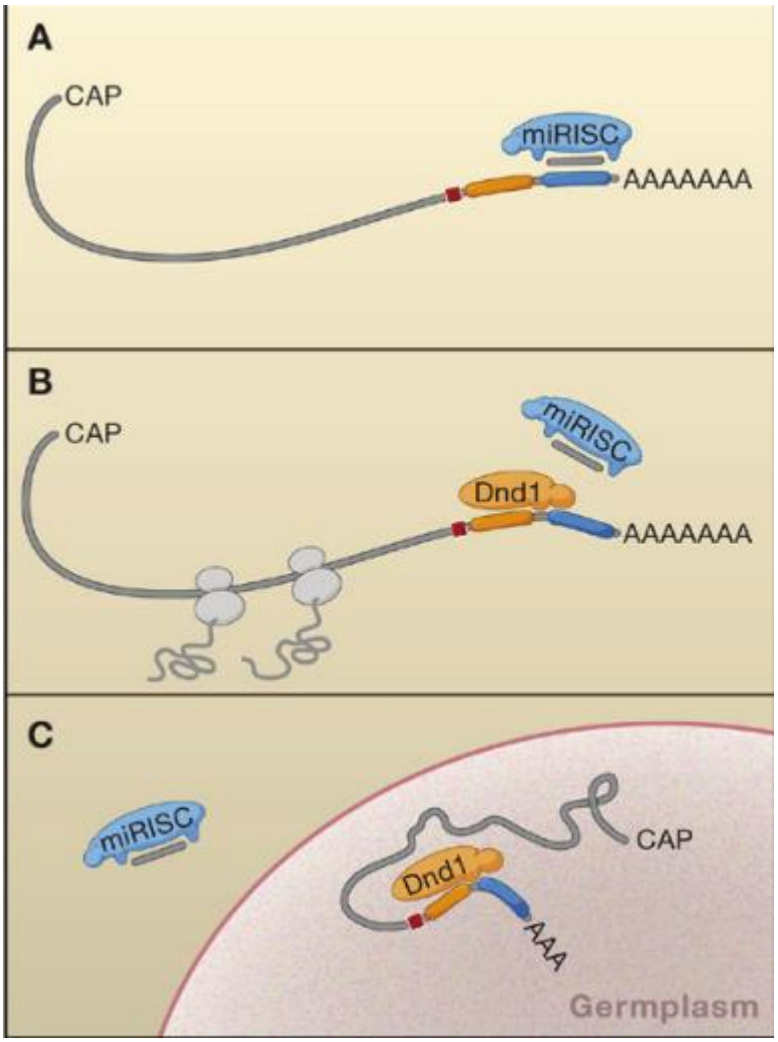


Figure 1.9. Dnd prevents microRNA mediated mRNA degradation. Adapted from (Ketting, 2007).

Medaka *dnd* is a maternal germline-specific gene, its expression is restricted in the PGCs and gonads (Liu et al., 2009). Observing the subcellular expression of Dnd by using Dnd-mCherry fusion protein showed that Dnd protein form granules in cytoplasm, co-localizing with Vasa. This result indicates that Dnd is a component of germ plasm, too. Amino acid alignment between species shows that the RNA recognition motif (RRM) is highly conserved across species, while the whole protein shares a 44% homology with zebrafish Dnd (Fig. 1.10).

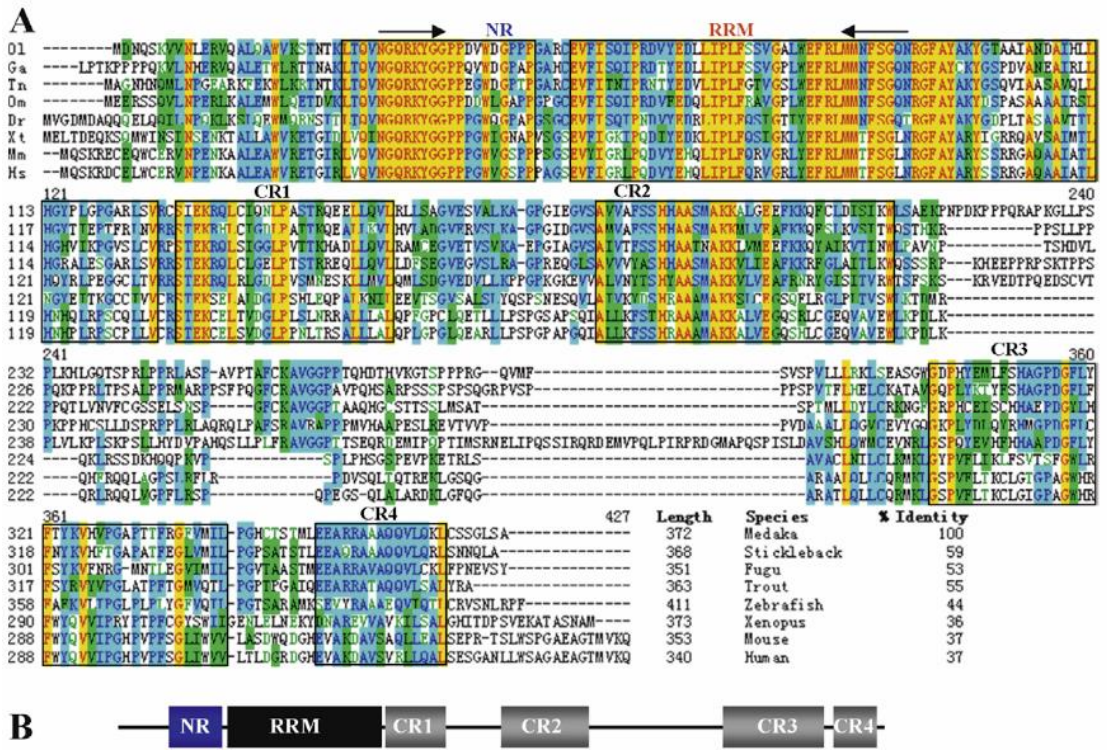


Figure 1.10. Dnd alignment and structure. Adapted from (Liu et al., 2009).

1.3. Objectives of the thesis

The main aim of this research project was to:

(1) Explore and validate the procedures and parameters for TALEN-mediated genome editing on the chromosomal locus *dnd* in medaka embryos via mRNA injection;

(2) Explore and characterize the phenotype of *dnd* deficiency in medaka embryos and fish.

CHAPTER 2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Fish

Medaka fish (*Oryzias Latipes*) are maintained in fresh water with a temperature range of 24-30 ℃ and a light/dark cycle of 14/10 hr. Strains HdrR, HdrR-II, Orange and AF (Auto-fluorescence free) were used for microinjection, crossing, in situ hybridization and photographing. All the experimental procedures are followed the Guidelines on the Care and Use of Animals for Scientific Purposes of the National Advisory Committee for Laboratory Animal Research in Singapore. Embryos are staged according to (Iwamatsu, 2004).

2.2. Reagents

2.2.1. Reagent list

The complete list of commercialized reagents, listed in a sequence according to where in the method section they are first mentioned, is detailed in Appendice I.

2.2.2. Reagent setups

(1) Tris-EDTA (TE buffer): 10 mM pH 8.0 Tris and 1 mM EDTA.

(2) TAE: 40 mM Tris-HCl, pH 8.0, 20 mM acetic acid and 1 mM EDTA.

(3) TBE: 89mM Tris-HCl, pH 8.0, 89 mM Boric acid and 2 mM EDTA.

(4) TNES-6U: 10mM Tris-HCl, pH 7.5; 125mM NaCl; 10mM EDTA; 1% SDS; 6M Urea.

(5) Direct PCR lysis buffer: 50 mM Tris-HC1 (pH 8.5), 1 mM EDTA, 0.5% SDS, 200 μg/ml proteinase K.

(6) 10 mg/ml Proteinase K: 0.1 g Proteinase was dissolved in 10 ml 60% Glycerol and store at -25 ℃.

(7) Preparation of Torula Yeast RNA for hybridization solution: One gram of torula RNA powder was added to a 50 ml Falcon tube. Ten ml of phenol and 15 ml of TE were added to each tube. The tube was mixed under very hard vortex until powder dissolves. After all the torula RNA powder in tubes was dissolved, the tubes were centrifuged in table top centrifuge for 10 min. Top phase was removed to a clean 50 ml Falcon tube followed by one phenol/chloroform extraction and one chloroform extraction. One ml of 3M NaOAc and 30 ml of ethanol were added to the supernatent. The mixture was let stand for 5 min and spin 20 min at full speed to pellet the RNA. The supernatant was decanted and an RNAse-free 80% ethanol rinse was performed to remove salts. The tube was centrifuged for 10 min , the supernatant was decanted and the pellet was air-dried. Then the pellet was dissolved in 15 ml RNase-free water. The concentration was measured on the spectrometer. The volume was increased to reach a 20 mg/ml final concentration. The torula RNA

was then aliquot into 15 ml falcon tubes and store at –20 ℃. (Jing, 2012).

(8) Hybridization buffer: Hybridization buffer used for whole mount in situ hybridization contains 50% formamide, 5xSSC (pH 7.0), 500 μg/ml torula yeast RNA, 50 μg/ml heparin, 0.1% Tween 20, and 9mM citric acid for adjusting the pH to 6.0-6.5.

(9) CCMB80: 10 mM KOAc pH 7.0 (10 ml of a 1M stock/L), 80 mM CaCl2.2H2O (11.8 g/L), 20 mM MnCl2.4H2O (4.0 g/L), 10 mM MgCl2.6H2O (2.0 g/L), 10% glycerol (100 ml/L), adjust pH DOWN to 6.4 with 0.1N HCl if necessary,

(10) Embryo rearing medium: To make 1L: NaCl 1.0 g, KCl 0.03 g, $CaCl_2.2H_2O$ 0.04 g, $MgSO_4.7H_2O$: 0.163 g, Methylene blue 0.01 % 10 ml, distilled Water to 1 L.

(11) Yamamoto's Ringer's solution: 0.75%, NaCl 0.02%, KCl 0.02%, $CaCl_2$ 0.002%, $NaHCO_3$ (Adjust to pH 7.3 with NaHCO3).

2.2.3. Equipments

Electrophoresis systems, including horizontal system and the vertical system, were from Bio-Rad.

Thermocyclers were Mastercycler® pro from Eppendorf.

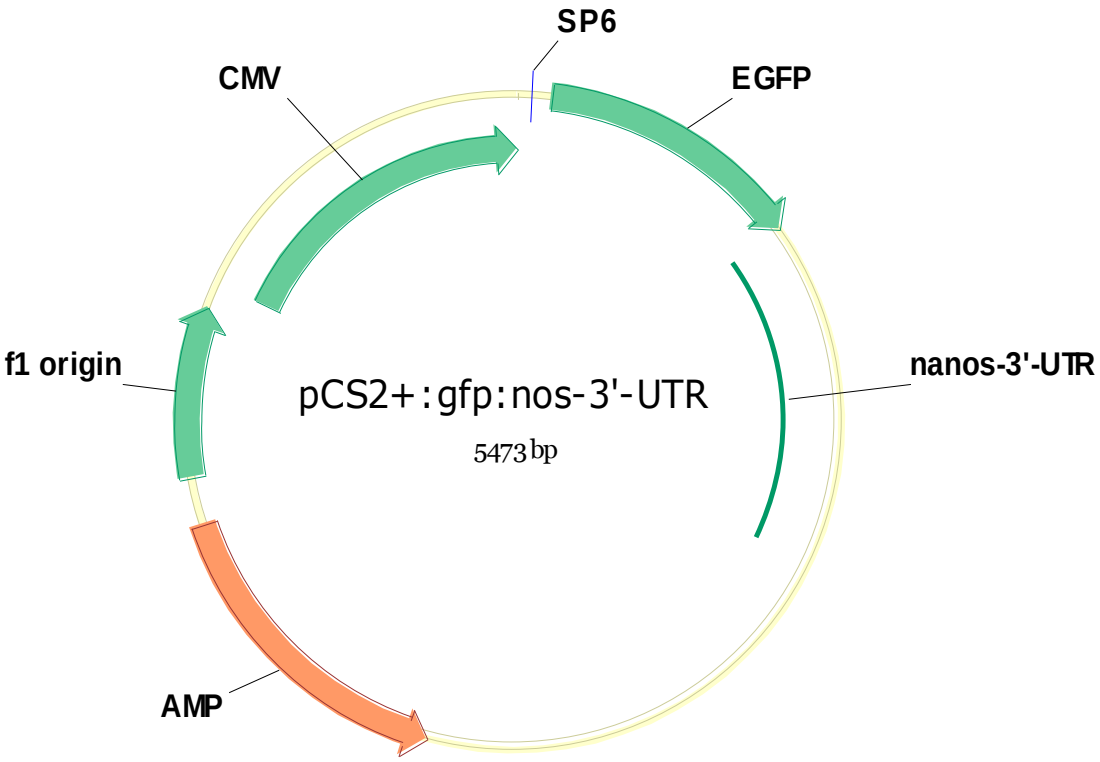
MZFIII stereomicroscope from Leica was used to observe the embryos and fish. MZ125 stereomicroscope together with a micromanipulator from Leca was used to

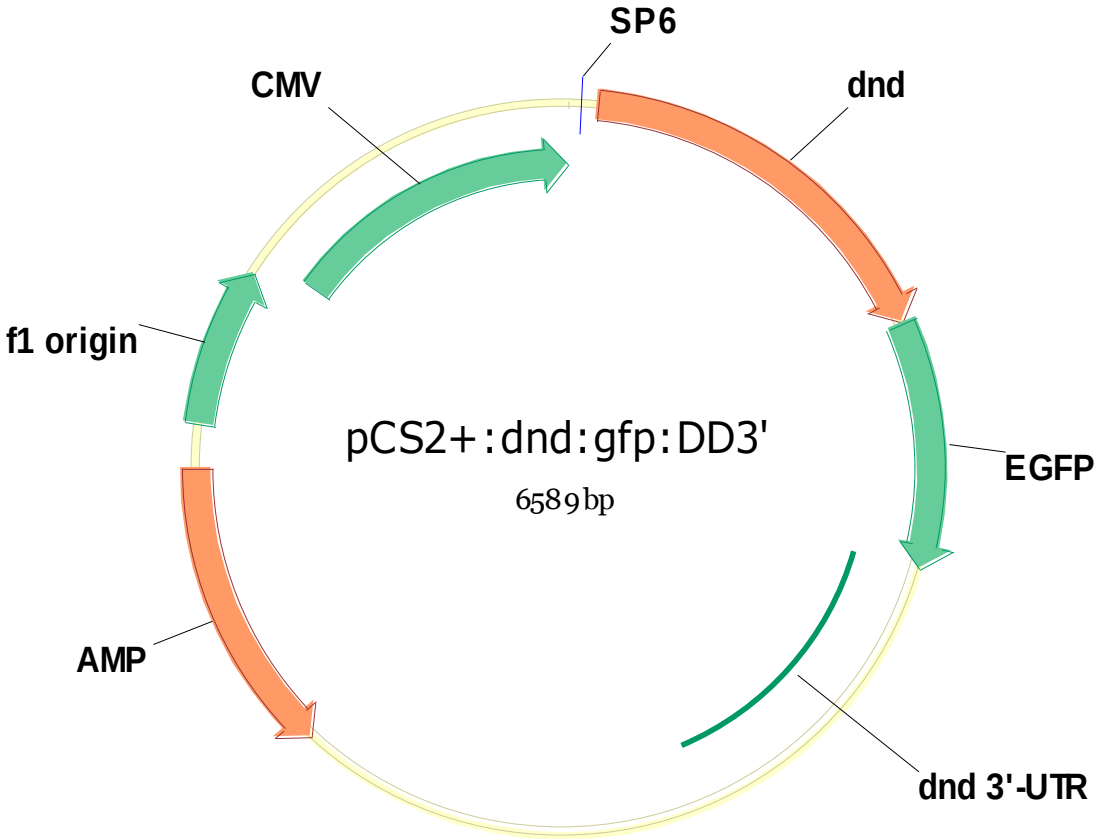
perform microinjection.

Centrifuges for 1.5 or 2 ml tubes were from Eppendorf and Thermo Scientific. Centrifuges for 15 or 50 ml tubes were from Eppendorf and Sigma.

2.2.4. Plasmids

pTNdnd1 and pTNdnd2 are the TALEN vectors constructed by using method and plasmid arrays published as (Cermak et al., 2011). pCS2:gfp:*nos*-3'-UTR was constructed by Prof. Li Mingyou, my senior and an alumni of our lab. pCS2:dnd:gfp:dnd-3'-UTR was constructed by changing the mCherry of pCS2:dnd:mCherry:dnd-3'-UTR to gfp. pCS2:dnd:mCherry:dnd-3'-UTR was also constructed by Prof. Li Mingyou.


SP6
CMV
EGFP
f1 origin
nanos-3'-UTR
pCS2+:gfp:nos-3'-UTR
5473 bp
AMP


SP6
CMV
dnd
f1 origin
pCS2+:dnd:gfp:DD3'
6589 bp
EGFP
AMP
dnd 3'-UTR

2.2. Methods

2.2.1. Genomic DNA extraction

Genomic DNAs of organs, tissues or embryos were extracted by lysis buffer TNES-6U. For each 25 mg of sample, 180 μl of lysis buffer plus 20 μl of 10 mg/ml proteinase K were used. The samples were homogenized in the lysis buffer using a plastic pestle and incubated in 55℃ for 3 hr or overnight. During the incubation, the tubes should be subjected to vortexes every half an hour. After digestion, the mixture was subjected to two phenol-chloroform extractions to remove the non-nucleic acid components and two chloroform extractions to remove the residue phenol. Isopropanol equals to 0.7 volume of the extracted solution and 0.1 volume of 5M NaCl were then added and mixed. The mixture was let stand for 15 min in room temperature and centrifuged for 15 min under 18000 rcf. The supernatant was then decanted and 1 ml of 70% ethanol was used to wash the pellet. The tube was centrifuged again to decant the 70% ethanol. This wash step was repeated once. The pellet was let dry after removing as much 70% ethanol as possible. Appropriate amount of TE buffer or water was used to resuspend the genomic DNA. The quality and quantity of the genomic DNA were checked and determined by electrophoresis with DNA ladder.

2.2.2. Chemically prepared competent cell

Hanahan & Inoue's methods of preparing chemically prepared competent cells

were adopted and combined (Hanahan et al., 1991; Inoue et al., 1990). Inoculation of exponentially growing E.coli: Streak stock bacteria culture on an antibiotic-free LB plate and pick up single colonies to inoculate at 37℃ overnight (16hrs). Pick single colony to inoculate in 2 to 5ml LB at 37℃ overnight. In the next morning, inoculate the overnight culture in a 1 to 100 ratio to a new flask of LB. The culture volume should not exceed one fifth of the max volume. This inoculation should be performed in low temperature, such as 18℃. The low temperature is critical to the transformation efficiency. The inoculation duration depends on the temperature used. Normally for 18 ℃, 40 hr is enough, the time increase with the decrease of temperature. The OD of the inoculation should be checked every hour during the last 5 hours of inoculation. An OD between 0.6 is acceptable. Re-suspension of the bacteria using transformation buffer CCMB80: Centrifuge the culture at 2000 g and 4C for 10mins. (Before this step, if the bacteria haven't been put on ice more than 20mins, wait until then.) Discard the supernatant and re-suspend the bacteria using CCMB80. For every 50ml culture, use 16ml CCMB80. Keep it on ice for 20mins. Centrifuge the culture at 2000 g and 4C for 10mins. Discard the supernatant and re-suspend the bacteria using CCMB80. For every 50ml culture, use 2~4ml CCMB80 (depending on O.D., 0.3 is to 2ml as 0.6 is to 4ml). Keep it on ice for 20mins. Aliquot the suspension to 50 or 100ul aliquots and keep them in -80C fridge.

2.2.3. TA cloning

PCR products were gel-extracted using gel-extraction kits and subjected to TA

cloning with a 3 to 1 insert-to-backbone-ratio, using TA cloning kit from Thermo or Promega. After 15 min to 1 hr of incubation, 2.5 μl the ligation product was transformed into 100 μl competent cells.

2.2.4. Plasmid DNA preparation

Mini-prep of plasmid DNA was prepared as following description. *E. coli* transformed with plasmid DNA was inoculated overnight in 37℃ shaker with 240 rpm of shaking. The bacteria were pelleted by 5 min of 6000 rcf centrifugation. After decanting the supernatant, plasmid extraction buffer I was used to resuspend the bacteria. Plasmid extraction buffer II was used to lyse the bacteria for 1 to 3 mins. Plasmid extraction buffer III was then added to neutralize the mixture and precipitate cell debris, proteins and genomic DNAs. After mixing, put the tubes on ice for 5 mins and centrifuge them for 15 min under 18000 rcf. Transfer the supernatant to a new 1.5 ml tube and add isopropanol equals to 0.7 volume of the solution were then added and mixed. The mixture was let stand for 15 min in room temperature and centrifuged for 15 min under 18000 rcf. The supernatant was then decanted and 1 ml of 70% ethanol was used to wash the pellet. The tube was centrifuged again to decant the 70% ethanol. This wash step was repeated once. The pellet was let dry after removing as much 70% ethanol as possible. Appropriate amount of TE buffer or water was used to resuspend the plasmid DNA. The quality and quantity of the plasmid DNA were checked and determined by electrophoresis with DNA ladder.

Mini- or midi-prep were also prepared using kits from Qiagen as described in the kits' manuals.

2.2.5. Restriction enzyme digestion

Some general guidelines for restriction enzyme digestion are as follows. For one hour digestion, amounts of used restriction enzyme were 5 units for every μg of Plasmid DNA to be digested. For overnight digestion, amounts of used restriction enzyme were 1.25 units for every μg of Plasmid DNA to be digested. Compatibility of double digestion should be checked in the enzyme's company's webpage. If the compatibility is not satisfactory or non-existent, double digestion can be performed instead. The DNA should be analyzed on a gel to confirm the complete digestion after incubation with first enzyme. The DNA should then be precipitated and resuspend with the second enzyme's reaction mix. For the second enzyme, because DNA is linear, only 1 unit is needed for 1 hr digestion, or 0.25 unit for overnight digestion. The incubation temperature of digestion should follow the manual of the specific enzyme.

2.2.6. Whole RNA extraction

Tissues or organs of interest up to 100 mg were submerged into 100 μl of trizol and homogenized. Additional trizol of 900 μl was added and the tube was let stand on ice for 5 mins. 200 μl of chloroform was then added to the tube, which was shaken vigorously for 15 seconds. The mixture was centrifuged for 15 min under

15000 rcf to separate the organic and aqueous phase. The upper aqueous phase was transferred to a new tube and extracted again with chloroform in order to remove residue phenol. Again transfer the upper aqueous phase to a new tube and add 500 µl of isopropanol, mix and stand for 10 min. Centrifuge the mixture for 10 min under 12000 rcf, decant the supernatant and wash the pellet with 1 ml of 75 % ethanol. Centrifuge the tube again for 5 min under 12000 rcf and discard the supernatant. After the pellet was air-dried, resuspend it with DEPC-treated water and mix it by passing the solution up and down several times through a pipette tip. Heat the solution on a 55℃ heat block for 10 min. The RNA is ready for downstream applications.

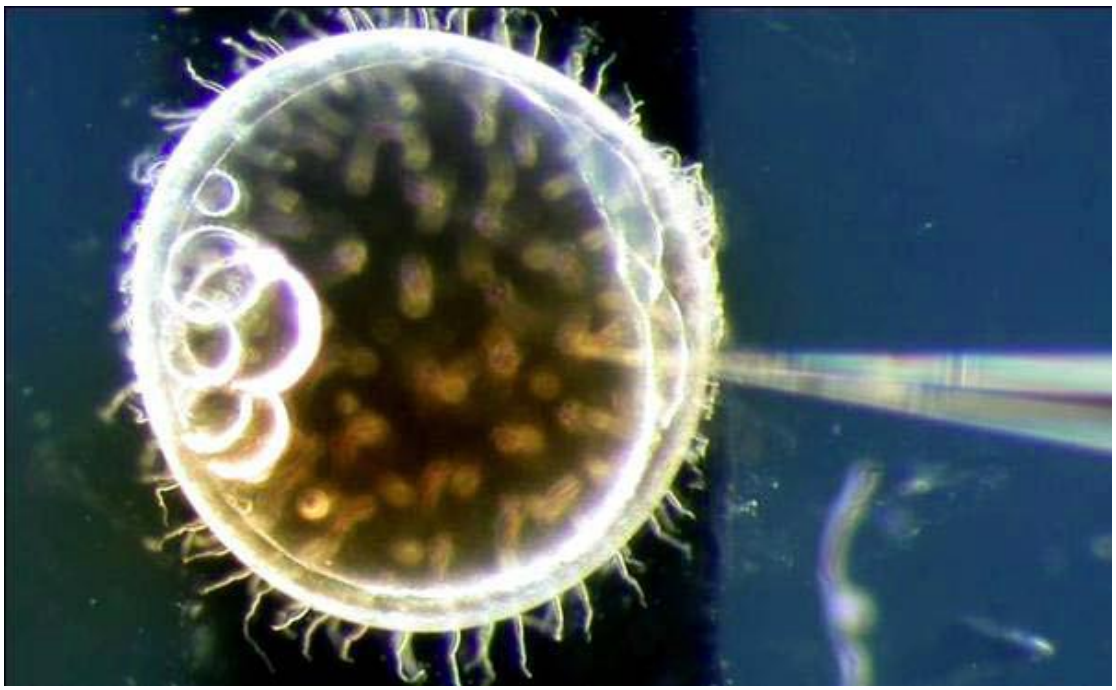
2.2.7. Synthesis of first strand cDNA

First strand cDNA was synthesized by using Maxima First Strand cDNA Synthesis Kit for RT-qPCR from Thermo Scientific. Briefly, 0.1 to 1 µg total RNA was mixed with 4 µL of 5X reaction mix, 2 µL maxima enzyme mix and water to 20 µL by passing the solution up and down through pipette tip. The reaction was performed on a thermo-cycler with the program: 25℃ for 10 min, 50℃ for 15 min and 85℃ for 5 min. if the GC content and the secondary structure of RNAs are high, increase the second temperature to 65℃.

2.2.8. Microinjection

One- or two-cell stage microinjection was performed as follows. To obtain

one-cell stage embryos, female and male fish were separated the day before injection. In the morning on the injection day, the male and female fish were put together and allowed for mating for 15 min before the embryos were collected. The collected one-cell stage embryos were dissociated by gently twisted between two dissecting needles and kept on ice. While waiting for the mating, centrifuge the to-be-injected solutions for 5 min under 15000 rcf. To load the centrifuged solutions to the microinjection needle: put some mineral oil on a piece of freshly opened parafilm; put 2 to 4 μl to-be-injected solution on the mineral oil; heat the needle briefly on an alcohol lamp and immediately placing the back of the needle on the solution; let the solution and the oil flow into the needle due to contracting of the heated air inside the needle. Afterwards, insert the loaded needle into the Leica micromanipulator and break the tip of the needle under microscope carefully. During the centrifugation, pour a 1.5% Ringer's-TAE agarose gel into a 10 cm petri dish and put the microinjection mold on the gel. After obtaining dissociated one-cell stage embryos put them into the lanes of the gel and align their cells to the top right direction. In vitro transcribed mRNAs were then injected into one-cell stage embryos with various concentrations. The injection volume was approximately 1 nl.



2.2.9. Polymerase chain reaction (PCR)

PCR for the purpose of screening using genomic DNA or cDNA as templates was performed by using Taq based polymerases, such as KAPA 2G, Takara's Ex-Taq and Thermo's DreamTaq. For each one of the thermocylcler programs, different temperatures and durations were used as the manuals suggest. Generally, a program is set as follows: 1 to 3 min of 95℃ denaturation; 25 to 35 cycles of 15 to 30 sec of 95℃ denaturation, 15 to 30 sec of 50 to 60 ℃ annealing and 1 to 1 min of 72℃ amplification; 5 min of 72℃ final elongation.

PCR for cloning was performed by using Pfu or Phusion polymerases for the low error rate of amplification. The thermocycler programs are similar to the previous one but the elongation needs more time because of the property of the enzymes. Moreover, since cloning primers are normally longer, the annealing

temperatures often go higher than 68℃, which needs to be coupled with another two cycle program. Specifically, it is: 1 to 3 min of 98℃ denaturation; 20 to 30 cycles of 15 sec 98℃ denaturation and 30 sec to several min of 68 to 73℃ annealing/amplification; 72℃ final elongation.

2.2.10. Direct PCR from lysed embryos or caudal fin clips

For screening of progeny fish of mutant families, direct PCR was used for convenience, due to the large amount of samples. It is a very simple and efficient method adopted from a cold spring harbor protocol (Goldenberger et al., 1995). Briefly, small amount of samples (one embryo or a little piece of caudal fin clip in this case) were lysed by 50 μl of Direct PCR lysis buffer for 2 hr. After briefly mix the lysate by gentle vortexing, they are ready for PCR. An appropriate amount of Tween 20 need to be added into the PCR system, depending on the amount of SDS introduced into the system by the sample. The correlations between Tween 20 and SDS are listed below (final concentration in PCR system): 2% Tween 20 for 0.05% SDS; 5% Tween 20 for 0.2% SDS.

2.2.11. Agarose gel electrophoresis of nucleic acid samples

0.8 to 1.5% agarose gel in 1xTAE was casted depending on the size nucleic acid samples that are to separate. The gels were run under 70V for 30 to 45 min also depending on the size of the sample to be analyzed.

2.2.12. Poly-acrylamide gel electrophoresis (PAGE) of PCR products

10% TBE-PAGE gel was used to reveal the heteroduplex PCR products of mutant alleles. They were casted using 10% AB (29:1), 0.005V 10% APS and 0.001V TEMED in 1xTBE. The electrophoresis was performed in 1xTBE under 100V for 2 hr. Afterwards, the gels were stained using 3xGelred in 1xTBE for 10min prior being photographed using gel documentation system.

2.2.13. Detecting disrupted alleles of *dnd* in PCR products using capillary electrophoresis

To prepare for capillary electrophoresis, 0.5 µl of PCR product, 0.5 µl of Genescan-500-ROX marker and 9 µl of HiDi-formamide were mixed together and subjected to ABI3130 or ABI3730 genetic analyzers. The results were analyzed in the Genemapper software.

2.2.14. Sequencing of DNA samples using BigDye terminator

Plasmid DNA or PCR products were used as the templates of BigDye terminator sequencing, using 100 to 200 ng or 10 to 25 ng of them, respectively. A 1/8 BigDye reaction was performed in 10 µl system for sequencing. Thermocycler program was set as: 1min 96℃ denaturation; 25 cycles of 10 sec 96℃ denaturation, 5 sec 50 to 55℃ annealing; 2 min 60℃ elongation. The 10µl sequencing reaction product was precipitated by adding 1 µl 3M NaAc, 2.5 µl EDTA and 3V of ethanol. The pellete

was washed for once by 70% ethanol and let dry. The dried pellete was resuspended by HiDi-formamide for sequencing on ABI3130 or ABI3730 genetic analyzer.

2.2.15. In vitro synthesis of RNA

Messenger and probe RNA was synthesized using mMESSAGE mMACHINE transcription kits. To prepare template for RNA synthesis, 10 μg plasmid DNA was linearized using restriction enzyme. The linearization was confirmed by agarose electrophoresis. The digestion was cleaned up by phenol-chloroform extraction, precipitated by 2.5 volumes ethanol and 0.1 volume 4M ammonium sulphate (from the kit), washed by 75% ethanol and resuspended by 30 μl of nuclease-free water. After the concentration of linearized plasmid DNA was determined, the mRNA was synthesized according the protocol from the kit manual. For microinjection, the synthesized RNA was resuspended with nuclease-free water or DEPC-treated Ringer's solution with 0.1% phenol red. For in situ hybridization probes, the synthesized RNA was resuspended with 50% formamide in DEPC-treated water.

2.2.16. Whole-mount In situ hybridization

2.2.16.1. Preparation of embryos

The embryos were fixed at 4℃ over one to two nights and washed in PBS/PBT twice for 5 min each time. Then the embryos were dechorionated and equilibrated in methanol for 5 mins. The methanol was changed and the embryos were stored in -20℃

overnight, embryos can be stored this way for several months. The embryos should be kept in -20℃ for at least 2 hours before proceeding to next step. The fixed embryos were rehydrated by soaking for 5 min each in: 75% methanol + 25% PBT; 50% methanol + 50% PBT; 25% methanol + 75% PBT. Young embryos (before 1 somite) seem to be rather fragile during rehydration. Then the embryos were washed for 5 min for 4 times in 100% PBT. The embryos were incubated in proteinase K (10 μg/ml PBT) at room temperature. The incubation time depends on the fixation and may vary for different batches of the proteinase K. The incubation time ideally should be titrated. In this thesis, 10 min, 15 min, 20 min, 25 min and 30 min were used for stage 21, 2-dpf, 3-dpf, 4-dpf and 5-dpf embryos, respectively. Wash with PBT twice, each for 5 mins. The embryos were refixed in 4% paraformaldehyde in PBS for 20 min at room temperature and washed 5 times, each for 5 min with PBT.

2.2.16.2. Hybridyzation

Embryos were pre-hybridized in hybridization buffer for at least 1 hour (preferably for 4-5 hours) at 70℃ in 50% formamide. The pre-hybridization buffer was discarded and hybridization buffer containing probe was added. Fifty ng of digoxigenin-labeled probe in 200 μl hybridization solution was used to hybridize the embryos. The tube containing the embryos and hybridization buffer was incubated overnight at 70℃ in a 1.5 ml microfuge tube in a water bath with gentle shaking. Probes are stable in hybridization buffer and can be reused several times. Washes were all done at the hybridization temperature with preheated solutions: 10 min 75%

hyb + 25% 2x SSC; 10 min 50% hyb + 50% 2x SSC; 10 min 25% hyb + 75% 2x SSC; 10 min 100% 2x SSC. Embryos were then washed for 30 min twice with 0.2x SSC at 70 ℃. Then the following washes were performed at room temperature: 5 min 75% 0.2x SSC + 25% PBT; 5 min 50% 0.2x SSC + 50% PBT; 5 min 25% 0.2x SSC + 75% PBT; 5 min 100% PBT.

2.2.16.3. Detection

The embryos from previous step were put into blocking buffer (2% goat serum, 2mg/ml BSA in PBT) for at least one and a half hours (ideally 3 to 4 hours) at room temperature. After blocking, the embryos were put into blocking buffer containing anti-dig-AP (1:5000) overnight at 4 ℃ with gentle shaking. The next day, the embryos were washed 6 times in PBT for 15 min at room temperature. Before two times of 5 min NTMT wash, a 5 min TBST wash was performed to avoid formation of magnesium phosphate precipitation. The washed embryos were stained in proper amount of diluted NBT/BCIP (200 µl stock solution in 10 ml NTMT) at room temperature or 4 ℃ in dark. The staining was stopped by one quick wash and two 15 min wash with stop solution when desired signal was reached. After stopping the development of staining, the embryos were fixed again with 4% PFA for 20 min and washed 3 times for 5 min each time to remove the residual PFA. The embryos were put into 3% methylcellulose for imaging. For long term storage, the embryos can be kept in 4% PFA in fridge.

2.2.17. Bioinformatic analysis

Analysis of variance (ANOVA) was performed by using Graphad PRISM software.

CHAPTER 3: RESULTS

3.1. Gene editing of *dnd* by TALENs in medaka embryos

The invention and utilization of programmable endonucleases has enabled efficient targeted genome editing in various organisms. In this thesis, the feasibility of transcription activator-like effector nucleases (TALENs) in generating gene-edited medaka fish via microinjecting TALEN mRNAs into 1-cell stage embryos was examined. Key parameters resulted from various dosages of TALEN mRNAs, such as survival rates, GE efficiencies in F_0 embryos and adults and germ-line transmission rates, were analyzed.

3.1.1. Experimental design for GE target locus and TALENs

For a pair of TALENs to successfully induce GE, the most essential rule is to let the targeting sequence to be within the exon and be as close to the starting codon as possible. Therefore, with frame-shift mutations abolishing whatever amino acids encoded after the targeting sequence, the majority of the protein is altered. Gene *dnd* has a transcript of 1119 nt and encodes a protein of 372 amino acids. The most important domain of Dnd is the RNA recognition motif (RRM), which ranges from amino acid 53 to 125 and its coding sequence ranges from the 159th base pair to the 375th. Because the role of RRM is critical for Dnd's biological function as an RNA-binding protein, the choosing of target sequence also takes this into consideration.

A sequence spanning 49 base pairs at the beginning of *dnd* exon 3 was chosen as the target sequence of TALENs. This sequence ranges from the 127th to the 175th base, so that the end of it has a 19-bp overlapping with the coding sequence of Dnd RRM. A frame-shift mutation created here abolishes 77% of the subsequent amino acids. Moreover, there may be frame-shift mutations that alter the overlapping sequence, part of RRM. These mutations can eliminate the possibility of gene function recovery by alternative splicing, when the functional motif itself is anticipated to be disrupted.

Two vectors, namely pTN1dnd and pTN2dnd, were constructed to express two TALENs targeting the target site discussed above (Fig. 3.1). They are shown for coding nucleotide sequence and the encoded protein sequences in Appendice II. This target sequence is present only once in the medaka genome by a search using BLAST (http://www.ensembl.org/Oryzias_latipes/Info/Index). Two primers were designed for PCR genotyping (Fig. 3.1, bottom), which flank a 196-bp fragment spanning the target site (Fig. 3.3A).

Overall, the target sequence of TALENs for dnd GE has been chosen to span from the beginning of *dnd* exon 3 to the beginning of RRM coding sequence, ensuring the ablation of Dnd protein's function.

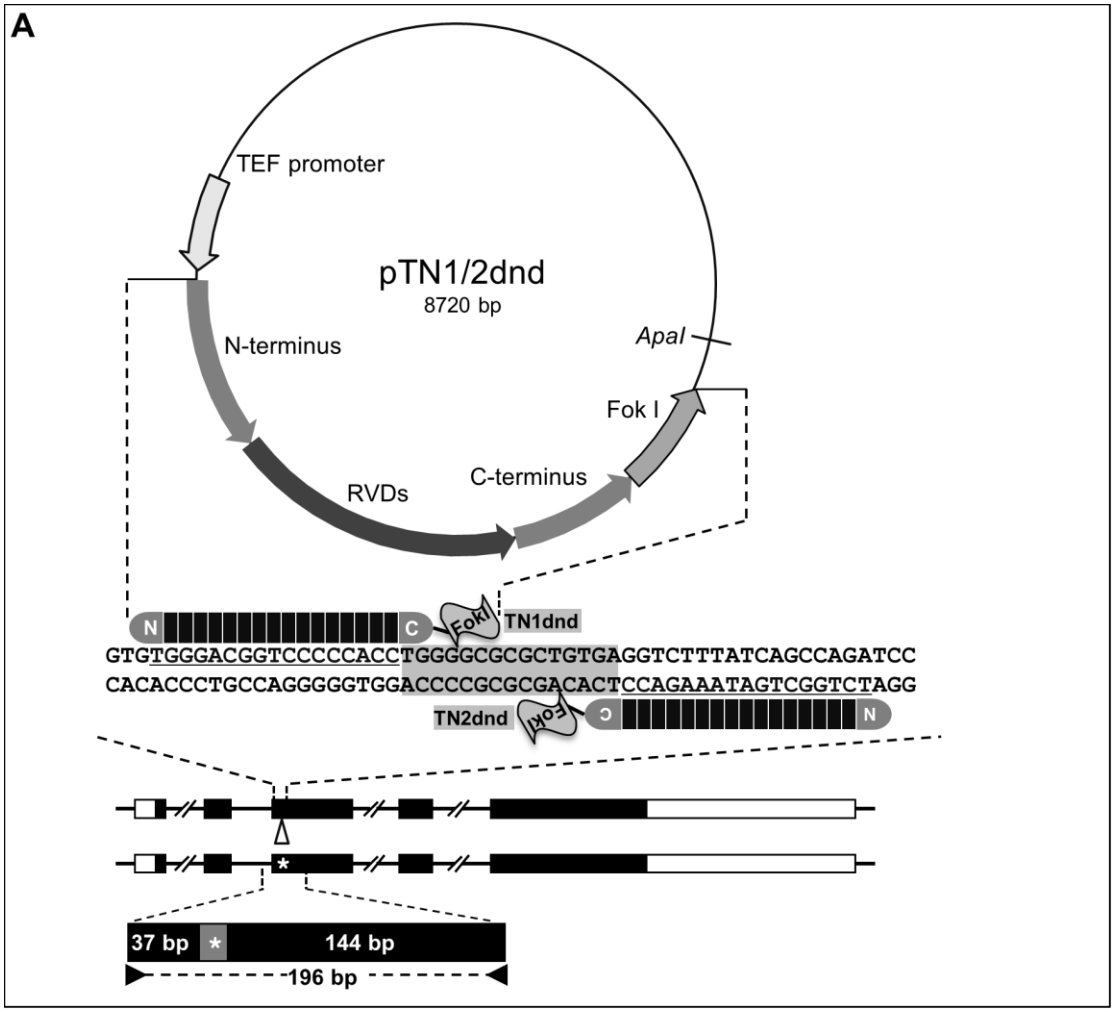


Figure 3.1. Medaka *dnd*'s protein & gene structure and its TALENs' design. Shown is the map of pTN1dnd and pTN2dnd (top), the TALEN target site (middle; open triangle) and targeted alleles (bottom). The target site consists of two recognition sequences (underlined) flanking a spacer. Each of TALENs has 16 RVDs (black rectangles between N- and C-terminus) and – together with a T at the 5′ start – recognizes a total of 17 bp. The target site contains two 17-bp recognition sequences and a 15-bp spacer and is located within exon 3 of gene dnd. TALENs generate a double-strand DNA break (asterisk) within the target site to initiate DNA repair and thus yield GE alleles with subtle additions and deletions within or near the target site. Black arrowheads below the PCR product depict primers to amplify a 196-bp PCR product flanking the target site.

3.1.2. Dose-dependent survival of TALEN-injected embryos

Programmable endonucleases are expected to be lethal to cells to some extent, because of the off-targeting effect (Szczepek et al., 2007). Therefore, the toxicity of TALENs needs to be characterized before going to test the GE efficiency. The effects of various dosages on embryo survival were examined. The mRNA concentrations of TALENs injected initiated from 25 ng/µl for each of the TALENs. The following concentrations were all doubled up until an extreme one of 400 ng/µl killed 99.5% of injected embryos. For the lower concentrations, the survival rates at the hatching stage were 93%, 82%, 83% and 71% in embryos injected with 25, 50, 100 and 200 ng/µl each of TALEN mRNAs, respectively (Table 1). DdH_2O-injected embryos served as control and 97% of them survived until hatching stage. In general, the survival rates were negatively correlated with the injection concentrations. However, they are all acceptable for experiment except for 400 ng/µl.

Table 3.1. Dose-dependent survival rates of TALEN-injected embryos

[TALEN mRNA] (ng/μl each)	Embryos injected, n	Survivor, n (%) [2)]		
		Gastrula	Hatching	Fry
0	30	30 (100)	30 (100)	29 (97)
25	44	44 (100)	43 (98)	41 (93)
50	66	62 (94)	56 (85)	54 (82)
100	95	89 (94)	81 (85)	79 (83)
200	111	103 (93)	91 (82)	79 (71)
400	64	46 (72)	7 (11)	3 (0.5)

1) The relationship between survival rate and various doses of TALEN mRNA was shown. TALEN mRNAs were injected at 25, 50, 100, 200 and 400 ng/μl. Water injection served as a negative control.

2) Percentage values were derived by comparison to the number of embryos injected.

3.1.3. Dose-dependent efficiency

In order to produce a higher number of targeted embryos, not only a high survival rate but also a high GE efficiency is required. The dosage that provides the most surviving embryos doesn't necessarily provide the most targeted surviving embryos. Therefore, the effects of different dosages on GE efficiency were tested on embryos. As expected, in PCR-heteroduplex mobility assay (PCR-HMA), no heteroduplex bands were revealed in PCR products of control embryos' DNA other than the homoduplex band, showing a pure wild type (WT) allele composition of the DNA. On the other hand, both a bright homoduplex band and a varying number of additional dimmer heteroduplex bands were detected in PCR products of the

embryos injected with TALEN mRNAs, indicating the existence of various altered alleles and the WT allele in those embryos. Both the number and the intensity of heteroduplex bands increased with the dosage of injected TALEN mRNAs (Fig. 3.2A). This trend was more evident after the enrichment of heteroduplex bands by another round of PCR-HMA assay using the extracted heteroduplex bands as template, for the rare mutant alleles was enriched (Fig. 3.2B).

A densitometry-based semi-quantitative analysis using the Gel-Pro software was performed to examine the effect of TALEN mRNA dosage on the relative abundance of GE alleles. The relative intensity of heteroduplex bands was found to be 10.6, 21.3 and 32.4% in embryos injected with 25, 50 and 100 ng/µl of TALEN mRNAs, respectively (Fig. 3.2C and Fig. 3.3). This revealed a correlation between GE efficiency and injected TALEN mRNA dosages. In parallel, the number of heteroduplex variants was found to be 2, 3, and 8 in embryos injected with 25, 50 and 100 ng/µl of TALEN mRNAs, respectively (Fig. 3.2C and Fig. 3.4D), again revealing a correlation between the number of GE events and injected TALEN mRNA dosage. An exception to both correlations was seen in embryos injected with an extremely high TALEN mRNA dosage (≥200 ng/µl), where a decrease rather than an increase in intensity and number of heteroduplex bands was recorded (Fig. 3.2C). These results indicate that 100 ng/µl each of TALEN mRNAs is the most effective concentration in regards of getting most targeted alleles in developing embryos, when both survival rate and GE efficiency are taken into consideration.

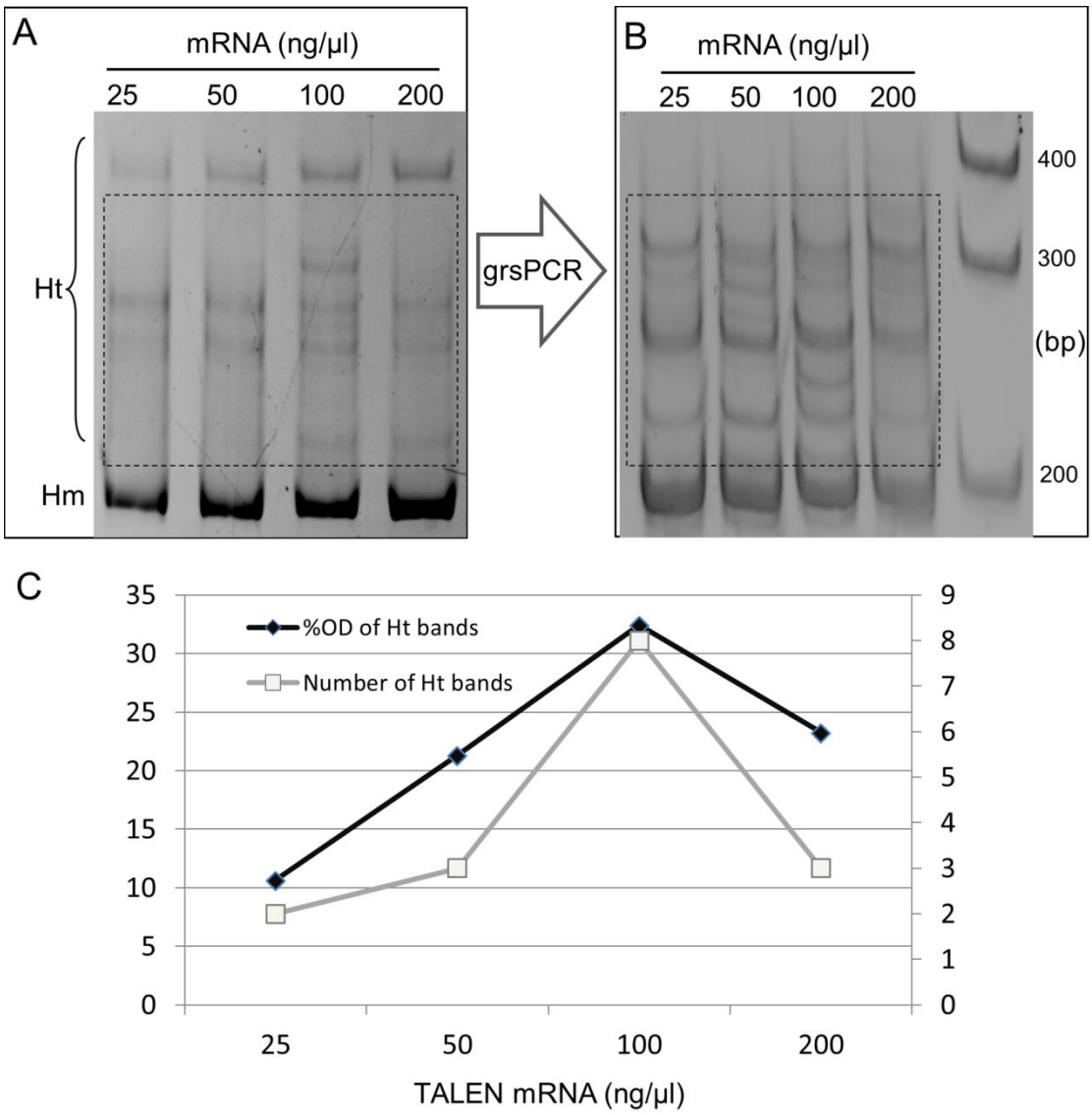
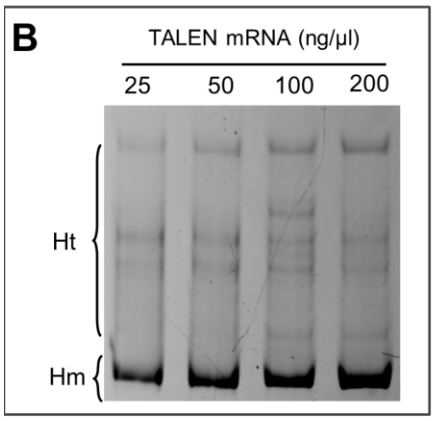


Figure 3.2. Dose-dependent dnd GE efficiency. Embryos were injected at the 1-cell stage with indicated doses of TALEN mRNAs. Genomic DNA was isolated from 25 pooled fry and PCR products were analyzed by PAGE. (A) GE alleles in heteroduplex bands. (B) GE alleles in heteroduplex bands following enrichment by gel recovery and successive PCR (grsPCR). DNA size markers are shown to the right. Heteroduplex (Ht) bands migrate slower and are above homoduplex (Hm) bands. (C) Dose-dependent abundance of GE alleles.

A

```
  1 TTTACATTTCGTGTTCAGTGTGGGACGGTCCCCCACCTGGGGCGCGCTGTGAGGTCTTTATCAGCCAGAT
 71 CCCACGGGATGTCTACGAGGACCTGCTCATCCCCCTCTTCAGCTCGGTGGGGGCACTTTGGGAATTCCGG
141 CTCATGATGAACTTCAGCGGTCAGAACCGGGGCTTCGCTTACGCCAAATACGGCAC
```



C

Ht band	TALEN mRNA (ng/μl)			
	25	50	100	200
OD%	10.647	21.314	32.432	23.243

D

Ht band	TALEN mRNA (ng/μl)			
	25	50	100	200
number	2	3	8	3

Figure 3.3. Analysis of TALEN-mediated dnd GE. (A) Primers used for PCR-based genotyping. Forward and reverse primers are underlined. Spacer is in bold. (B) Quantification of heteroduplex PCR products. (C) Optic density (OD) of heteroduplex PCR products. (D) Numbers of heteroduplex PCR products. Hm, homoduplex; Ht, heteroduplex.

3.1.4. GE efficiency in embryos and adults

According to the above trials of concentrations, 100 ng/μl has been chosen as the working dosage for the following experiments. The GE efficiencies mentioned above are observed in the grouping of embryos, representing overall GE efficiencies for respective dosages. Subsequently, GE efficiency in individual embryos was characterized. Embryos injected with 100 ng/μl of TALEN mRNAs were individually sampled at 7 dpf for genotyping by PCR-HMA. There were 11 out of 20 examined embryos showed obvious heteroduplex bands (Fig. 3.4A), demonstrating the generation of GE alleles-containing embryos at a 55% efficiency.

After the GE efficiency of working dosage was characterized in embryonic stage,

it also needed to be examined in adult stage. A total of 400 medaka embryos at the 1-cell stage were injected with 100 ng/μl of TALEN mRNAs. There were 122 of these injected embryos brought up into adulthood. In order to keep the fish alive to establish GE family, autopsy examination was performed to screen for adult fish that carry GE alleles. Fin clips were sampled and subjected to genotyping by PCR-HMA. It was found that 29 of the F0 fish exhibited heteroduplex bands in the assay, indicating they were putative founder animals (Fig. 3.3B). Various heteroduplex bands formed between GE and WT alleles showed retarded mobility compared to the homoduplex bands. Besides GD-containing heteroduplex bands, a faint GD-Homoduplex band may also be visible in certain F_0 fish such as fish #14 and #19. Taken together, injection with 100 ng/μl of TALEN mRNAs is capable of generating GD-alleles-carrying adults at a high efficiency of 23.8%.

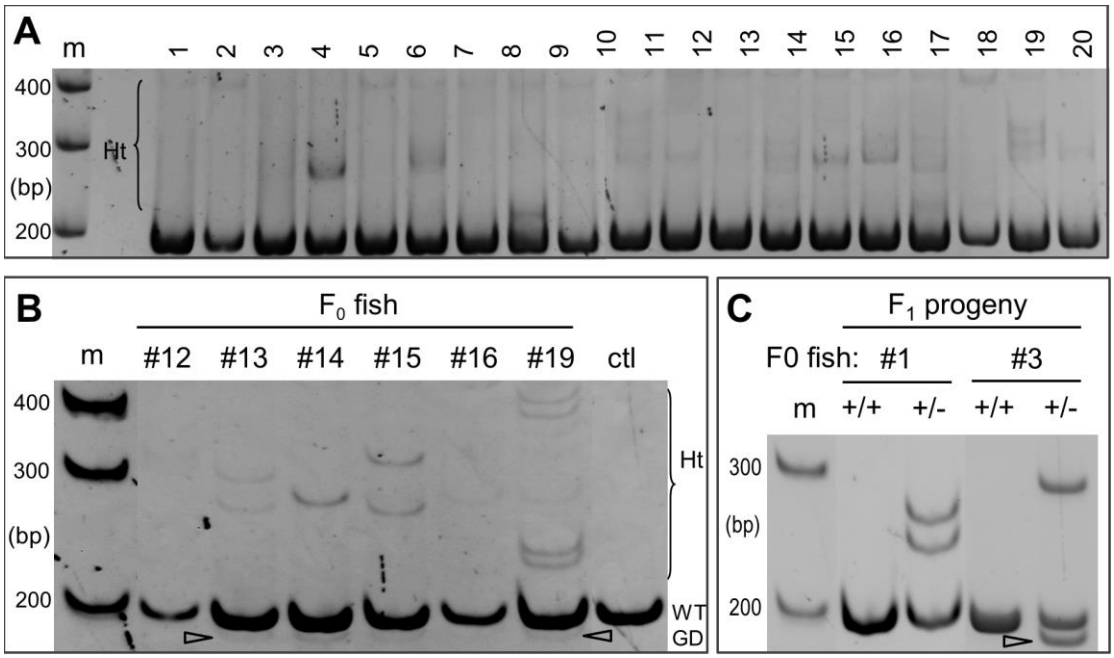


Figure 3.4. Genotyping of GE alleles. Embryos injected with 100 ng/µl mRNAs were analyzed for heteroduplex formation by using PAGE. (A) Individual embryos at 7 dpf. (B) Representative F0 adult fish from injected embryos. ctl, non-injected control fish. (C) F1 progeny adults from founder fish #1 and #3. Heteroduplex (Ht) bands containing GE alleles, and homoduplex (Hm) WT allele (WT) and GE alleles (GD) are indicated. Arrowheads depict Hm of GE alleles in (B and C). m, DNA size markers shown to the left.

3.1.5. Germline transmission

F_0 fish #1, #3 #19 and #58 derived from TALEN-injected embryos grew to adulthood. They were crossed with WT fish of opposite sexes, and the resultant progeny were examined for germline transmission. All of them were confirmed to be founders, as they transmitted GE alleles to their F_1 progenies (Table 2). Specifically, founder fish #1 was a female and transmitted a GE allele to 9 out of 84 progenies, demonstrating a ~11% germline transmission rate. Founder fish #3 transmitted a GE allele to 2 out of 28 F_1 progenies, thus demonstrating a ~7% germline transmission rate. Founder fish #19 has transmitted a GE allele to 2 out of 18 F_1 progenies, producing a ~11% germline transmission rate. Founder fish #58 transmitted a GE allele to 2 out of 15 F_1 progeny, giving a ~13% rate of germline transmission. These data suggest that TALEN is able to produce a high GE efficiency in both the soma and germline.

Importantly, the relative abundance or intensity of GE PCR products (i.e. heteroduplex band and homoduplex band of the GE allele) was comparable to that of the WT PCR product (homoduplex of WT allele) in the progeny of both founder fish #1 and #3 (Fig. 4C). This observation is in accordance with germline transmission of GE alleles and production of heterozygous F_1 animals. Taken together, TALENs are capable of generating GE in germline.

Table 3.2. Germline transmission of GE alleles

Founder	Mutant alleles [1)]		Frequency
	Fin	Germline	
#1	A4D7	A3	10.7% (9/84)
#3	N/D [2)]	D10	7.7% (2/26)
#7	D5	N/D	N/D (0/28)
#19	D4/A14D8	A1D11	11% (2/18)
#58	A4D15	D8	13% (2/15)

[1)] A, addition; D, Deletion.

[2)] N/D, Not Detected.

3.1.6. Variable targeted alleles

Sequencing of cloned PCR products from gel-recovered heteroduplex bands validated the GE events. The F_1 fish fins were sampled and subjected to PCR-HMA. GE alleles with various alterations were identified, which fall into three major categories, namely addition (alleles with added nucleotide), deletion (deleted nucleotide), and compound alteration (a combination of addition and deletion). Up to 10 different GE alleles were found in 6 F_0 fish (Fig. 3.5). In the fin clip samples, founder fish #13, #15 and #19 have deletion GE alleles missing 14, 11 and 4 nt, respectively. Founder fish #19 and #58 have compound GE alleles containing both addition and deletion.

The GD-positive progeny of founder fish #1 produced two heteroduplex bands

besides a major homoduplex band of the WT allele (Fig. 3.4C), indicating the presence of one or two altered alleles, since one loop formed by mismatching PCR products may generate one or two heteroduplex bands depending on DNA configuration (Chen et al., 2012). Sequencing of cloned PCR products revealed the presence of one and same allele, namely A4D7, in all of the 9 progenies. In A4D7, a region of 7 nt within the TALEN target sequence is replaced by 4 nt, leading to a net loss of 3 nt (Fig. 3.5). Notably, this germline-transmitted allele is different from the somatic GE allele A3 in the fin of the founder (Fig. 3.5), indicating that they were products of two independent GE events. PCR-HMA of the progeny of founder fish #3 produced one heteroduplex band and one homoduplex band of the GE allele besides a homoduplex band of the WT allele. Sequencing of the PCR product revealed that the GE allele in the founder fish #1 progenies had a deletion of 10 nt (D10). What deserves being noted is that autopsy of these progenies' founder #3 possesses only the WT allele, showing no detectable GE event in the fin (Fig. 3.5), which indicates that allele D10 was generated in the PGCs after the separation of germline from somatic lineages that developed to the fin.

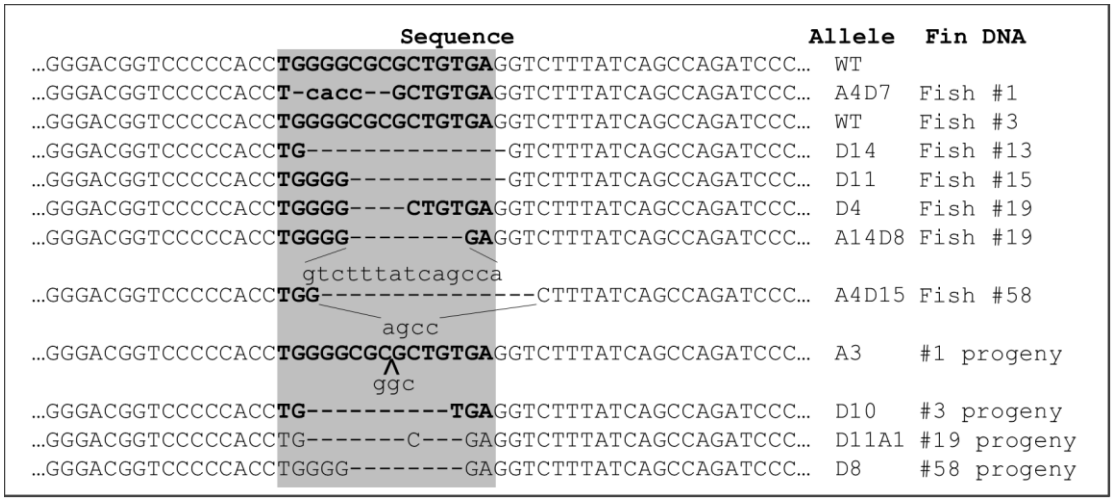


Figure 3.5. Sequence analysis of GE alleles. DNA was extracted from the fin of adult fish from TALEN-injected embryos or F1 progeny fish. Sequences shown here are for the WT and GE alleles within the target site with the spacer being highlighted. Small letter cases indicate introduced sequence substitutions. A, addition; D, deletion. Numerals after A or D indicate the numbers of nt added or deleted. Notably, fish #19 had two different GE alleles in the fin, fish #1 had a GE allele in the fin. All of #1, #19 and #58 transmitted a different allele to the F1 progeny. Fish #3 transmitted a GE allele, namely D10, to the F1 progeny without showing a detectable GE allele in the fin.

3.1.7. Family establishment

Fish derived from TALEN-injected embryos often developed into fertile female and male adults. Families of dnd GE fish should be established to facilitate the phenotypic analysis. However, F_1 fish often produce progenies that carry different GE alleles other than the alleles found in fin clip autopsies. Therefore, F_1 fish was chosen to be the pool of family founders and a closer inspection on the sequences of altered *dnd* alleles was needed to predict how those alleles affect the protein.

In the progenies of founder #1, the germline-transmitted allele A3 has 3 more nt in the mRNA of *dnd*, which was confirmed by cloning and sequencing of *dnd* transcript using the cDNA. Thus the mutation would not affect the open reading

frame of the gene, so it will still translate into a full length Dnd isoform, except that it contains one more amino acid (Fig. 3.6, second). The additional amino acid is not within the conserved box of RRM, showing a less possibility for it to interrupt Dnd's biological function. Therefore, the F1 fish that carry GE allele A3 were chosen to establish *dnd* GE family dgd1 as a non-frame-shift mutant.

In the progenies of founder #3, the germline-transmitted allele D10 has 10 fewer nt in the mRNA of *dnd* and thus represents a frame-shift mutation, which will affect the translation of Dnd. This was also confirmed by cloning and sequencing of *dnd* transcript using the cDNA of this family. The predicted protein product of the D10 allele is only 80 amino acids long and only has the first 48 amino acids correct (Fig. 3.6, third). This means the RRM is completely removed from the protein. Therefore, the F_1 fish that carry GE allele D10 were chosen to establish *dnd* GE family dgd2 as a *dnd*-null line for further phenotypic analysis.

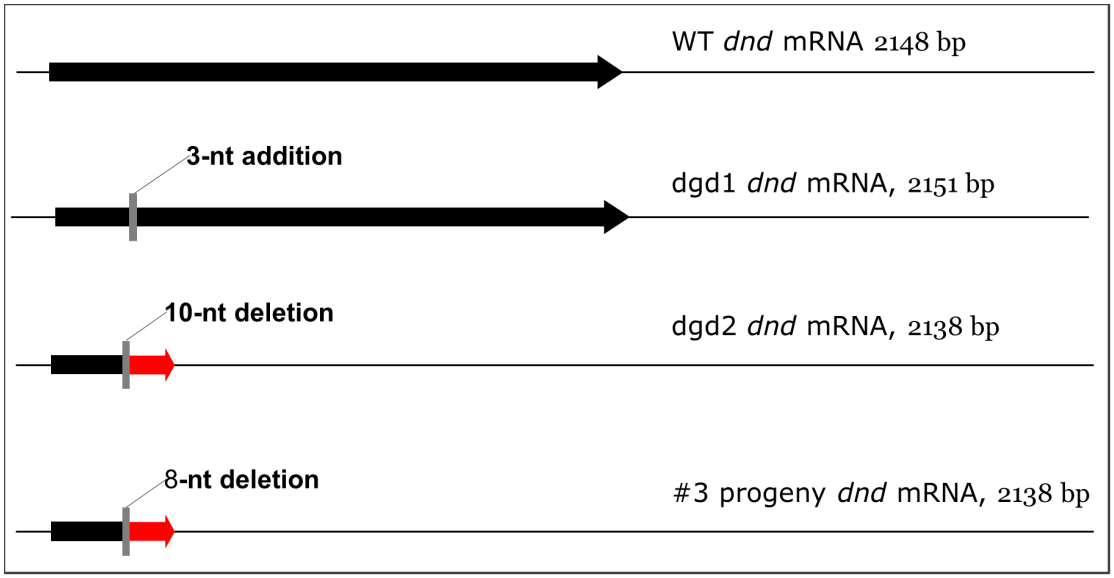


Figure 3.6. Schematic mRNA and protein products of GE alleles. Shown are the predicted mRNA (line) and protein (thick arrow) products of wild type and altered alleles. The grey bars indicate the positions of the alterations. The genuine Dnd protein is depicted in black within the protein arrow and the nonsense part of the protein products are depicted in red. The #1 founder carries and passes the non-frame-shift GE allele A3, producing a Dnd with one additional amino acid. The #3 and #58 founders carry and pass the GE alleles with 8- and 10-bp deletion, leading to a frame-shift mutation and disrupted Dnd sequence after the deletion site.

In conclusion, TALENs has been proved to be proficient for inducing germline and soma GE events at a high efficiency during medaka embryonic development by direct introduction in to the embryos via microinjection. These GE events include various targeted alleles that have different alterations, which lead to point mutation, in-frame mutation or frame-shift mutations. These GE alleles have been proved to be inheritable to progenies of the founders and sometimes the inherited alleles are different from the somatic sample of the founders due to chimeric state of the founders. The progeny fish heterozygous for either maternally or paternally inherited A3 or D10 alleles were found to have a normal somatic development and growth to their control siblings. These fish also have the ability to produce offspring. This result indicate that the gene *dnd* is haplosufficient for both somatic and germline development.

3.2. Phenotypic analysis of *dnd* knockouts

In order to observe the phenotype of loss of both alleles of *dnd*, F_1 fish of both sexes from the same family were paired together to produce F_2 fish. The embryos produced from these mating were collected and observed carefully each and every day during the embryonic development. For these F_2 embryos, it was anticipated that one fourth of them are homozygous for the *dnd* mutant allele that their founders carry, which means they are *dnd* knockout fish (Fig. 3.7A). Thus, if there were any phenotypes, they were probably to be found only in one fourth of the embryos produced. However, under white light stereomicroscope, there were no obvious phenotypes found. This confirmed the hypothesis that loss of *dnd* will not affect somatic development based on the fact that *dnd* is a germline specific gene as mentioned. The phenotype should be related to the development of germline cells and organs. Therefore, further investigation into the phenotype in germline, such as primordial germ cells, embryonic gonad and adult gonad, was needed.

3.2.1. *dnd*-deficient adults are sterile and show secondary sexual characteristics

Embryos collected from $dnd^{+/-}$ parents from dgd1 and dgd2 families were raised to adulthood. The fish hatched from these embryos were screened for homozygous GE fish after they matured (Fig. 3.7 and Fig. 3.8).

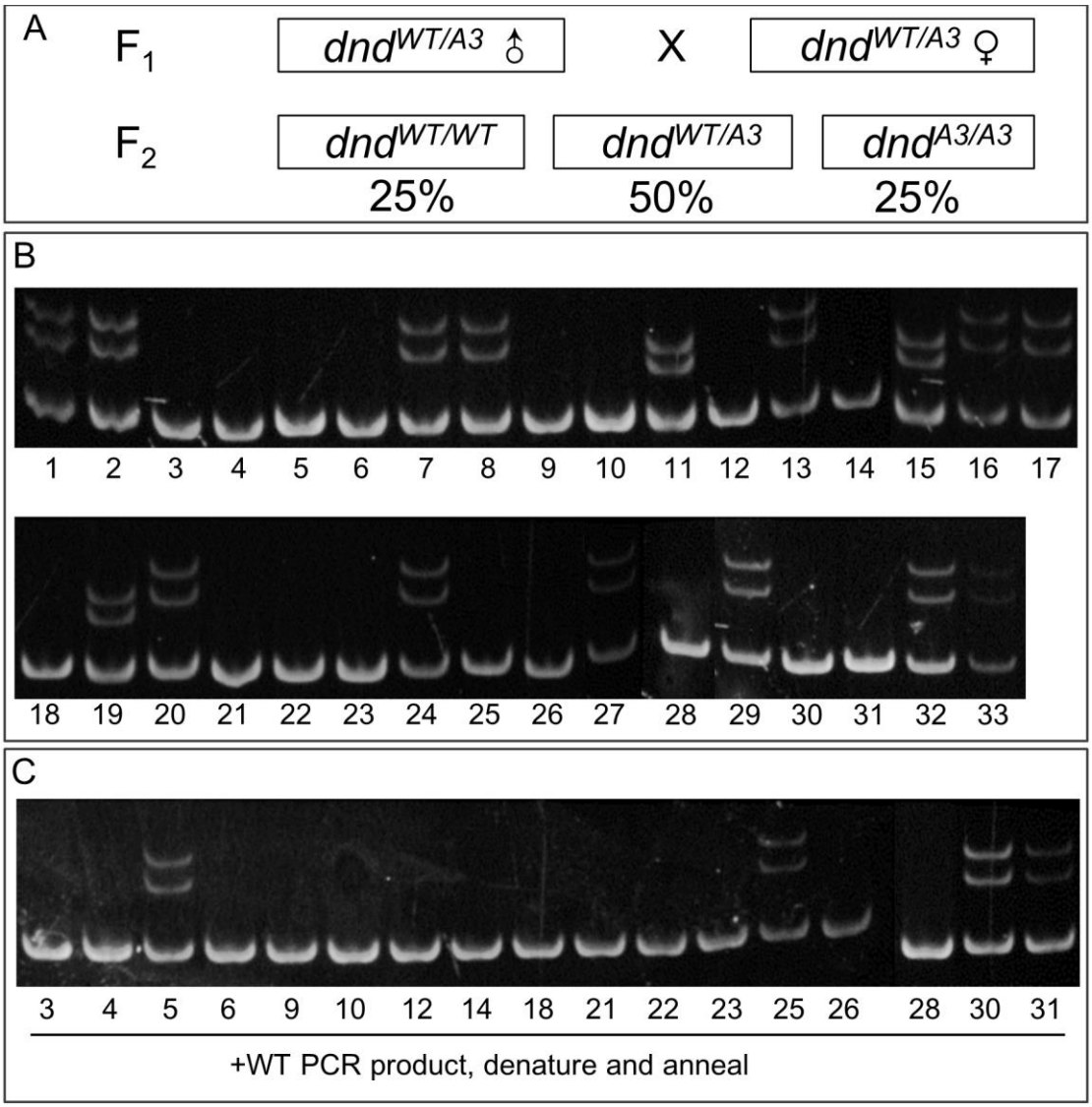


Figure 3.7. Progeny screening of family dgd1. F_2 fish of dgd1 family was screened after raised to adulthood. (A) Fish were produced by crossing two $dnd^{WT/A3}$ fish. The projected Mendel segregation is shown. (B) The F_2 fish genomic DNA were sampled from the end of caudal fin, and subjected to PCR. The PCR products were

subjected to PAGE to reveal the heteroduplex bands lagging behind the wild type bands. The samples exhibited heteroduplex bands are heteroduplex for the A3 allele. (C) The F_2 fish screening aims for homoduplex fish. The samples showed only homoduplex bands in the first PAGE were mixed with PCR products of wild type genomic DNA, denatured and re-annealed, thus the samples homogeneous for A3 allele will exhibit heteroduplex bands now. As shown, samples number 5, 25, 30 and 31 showed heteroduplex bands and were homozygous for A3 alleles.

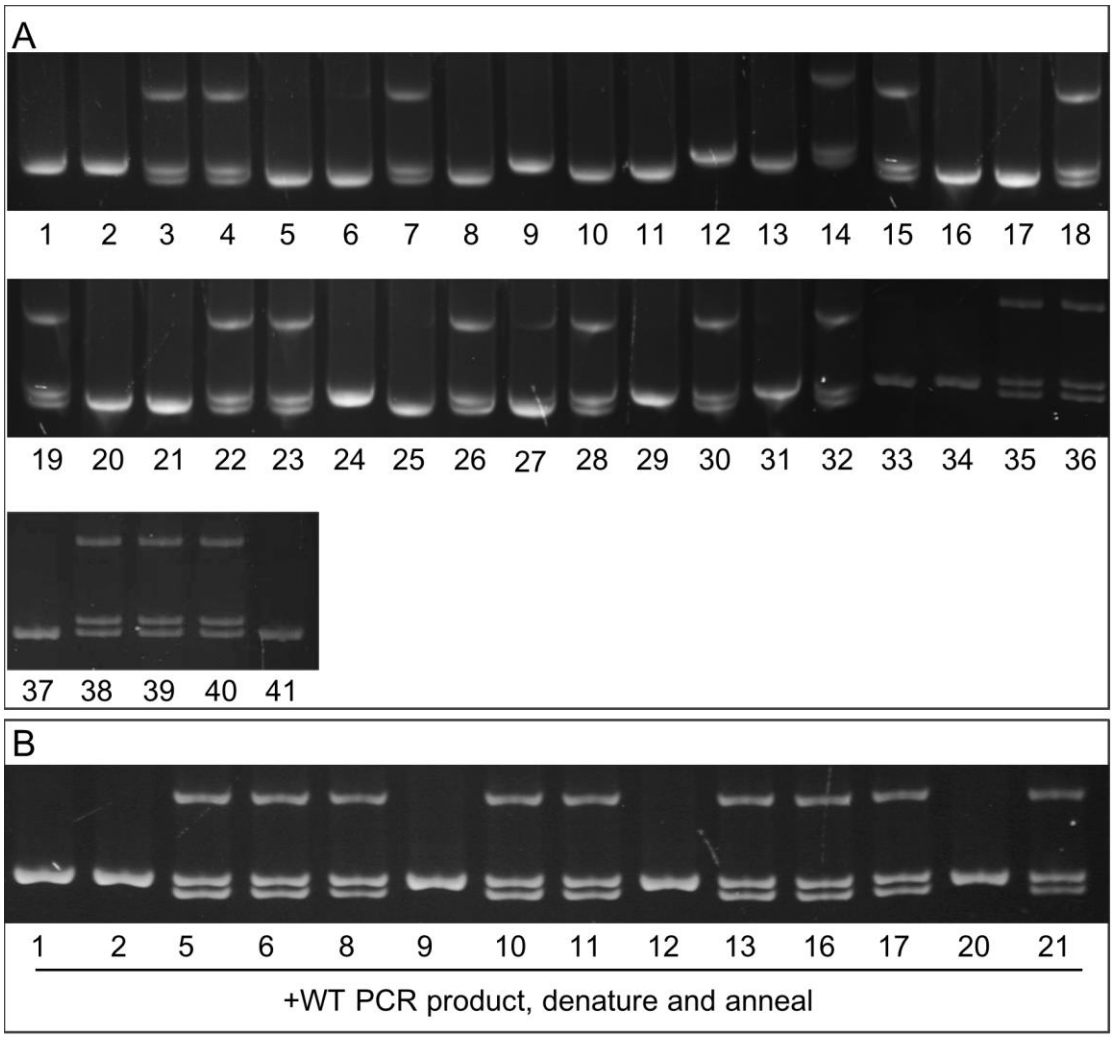


Figure 3.8. Progeny screening of family dgd2. Similar to dgd1 family, the F2 fish homozygous for D10 alleles were also screened out using two rounds of PCR-HMA. (A) The PCR-HMA of dgd2 F_2 fish. (B) The second PCR-HMA using PCR products of wild type genomic DNA to form heteroduplex bands with the samples homozygous for D10 alleles.

X/Y genotype was also analyzed by using genomic PCR of *dmrt1y* using the genomic DNA from caudal fin autopsy (data not shown). All the homozygous knockouts looked the same as their wild type counterparts, thus had the secondary sexual characteristics corresponding with their genomic sex (Fig. 3.9). Specifically, the fifth dorsal fin ray of X/Y genotypes wre further separated than that of the X/X ones (Fig. 3.9B); and the last anal fin ray of X/Y genotypes were also further separated as compared to that of the X/X ones (Fig. 3.9C).

The homozygous knockouts were not able to spawn after paired with each other. Wild type fish that were spawning or making females spawn were used to pair with the homozygous fish of opposite sex. Still, there was no spawning of eggs. In the fertility test, the fish were isolated in the afternoon and paired in the next morning. The results were observed 30 min later by checking if there were any embryos attached to the belly of the female. Because human presence would interfere with the mating of the fish, usually scaring them to a corner, the mating process wasn't observed during the fertility test. Subsequently, a digital camera was used to record the courtship and mating of the fish. The record showed that after the male kept performing courting towards the female, which is chasing to the tail in this case, they mated and the female spawned (Fig. 3.10B, D). However, if the female was replaced with a *dnd*-deficient one, the male lost the desire to chase after the female (Fig. 3.10C). On the other hand, if the male was replaced with a *dnd*-deficient one, it doesn't perform courting towards the female (Fig. 3.10A).

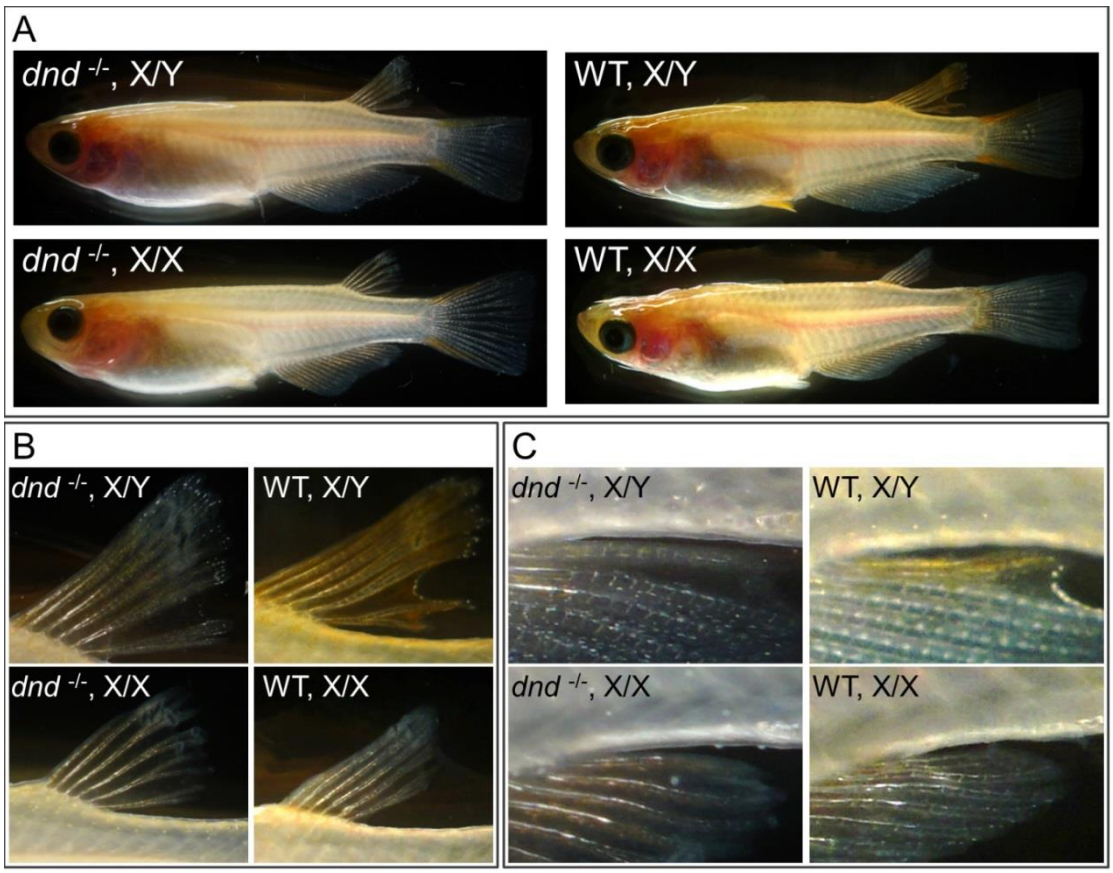


Figure 3.9. Sexuality of *dnd*-deficient fish. (A) Shown is sexually dimorphic color patterns in whole fish. (B) Dorsal fin in large magnification, highlighting sexual dimorphism. As compared to female fish, in both the dnd deficient and wild type fish, the dorsal fin of male fish showed widely separated 4th and 5th fin rays. (C) Anal fin in large magnification, highlighting sexual dimorphism. As compared to female fish, in both the dnd deficient and wild type fish, the anal fin of male fish showed one fin ray, which is nearest to the fish body, widely separated with the others.

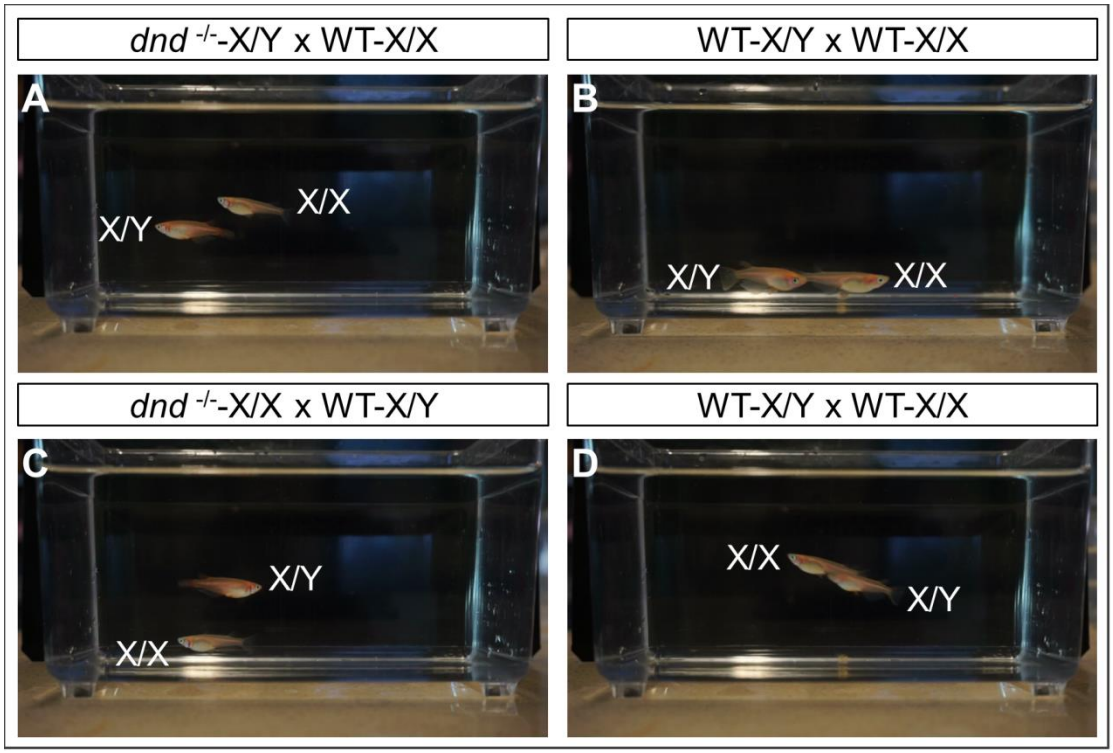


Figure 3.10. Courtship behavior of dnd deficient fish. Mating videos were recorded for the purpose of observing the courtships. Shown are the screenshots of the videos. As shown, the *dnd*$^{-/-}$X/Y showed no interest in WT-X/X (A), while the WT-X/Y kept chasing after WT-X/X (B). Similarly, the WT-X/Y didn't show any interest in the *dnd*$^{-/-}$X/X fish (C), in contrast to the wild type fish pair (D).

In conclusion, the *dnd*-deficient fish of both sexes were not able to give rise to progeny and male fish couldn't perform courtship to female while female fish couldn't attract courtship from male.

3.2.2. Loss of *dnd* leads to loss of germ cells during embryonic development

Subsequently, the PGCs were checked during embryonic development, since knockdown of *dnd* leads to PGC death during embryogenesis in zebrafish, which could also be true in medaka and have led to the infertility. Specifically, *gfp* mRNA with *nanos* 3'-UTR (*gfp*:*nos*-3'-UTR) was used to label the PGCs via 1-cell-stage

microinjection (Fig. 3.11). Because this labeling method utilizes the targeted clearance of mRNAs with *nanos* 3'-UTR in somatic cell lineage by microRNAs (Ketting, 2007), the mRNA injected will not be degraded in the embryonic stem cells before the separation of soma and germline, and survive only in the PGCs after the separation. This way, the injected mRNAs will be specifically expressed in PGCs, labeling them with GFP. However, the GFP translated during early embryonic development will be inherited by the somatic lineage and last for a while, for the clearance mechanism only interacts with the mRNAs. This leads to the formation of a GFP background and makes the *nanos* 3'UTR of no use in restricting the GFP signal in PGCs in early stages of embryos. As the background GFP degrades and the *gfp* mRNAs in PGCs continue to produce germline-specific GFP, the observation of PGCs is more and more eminent.

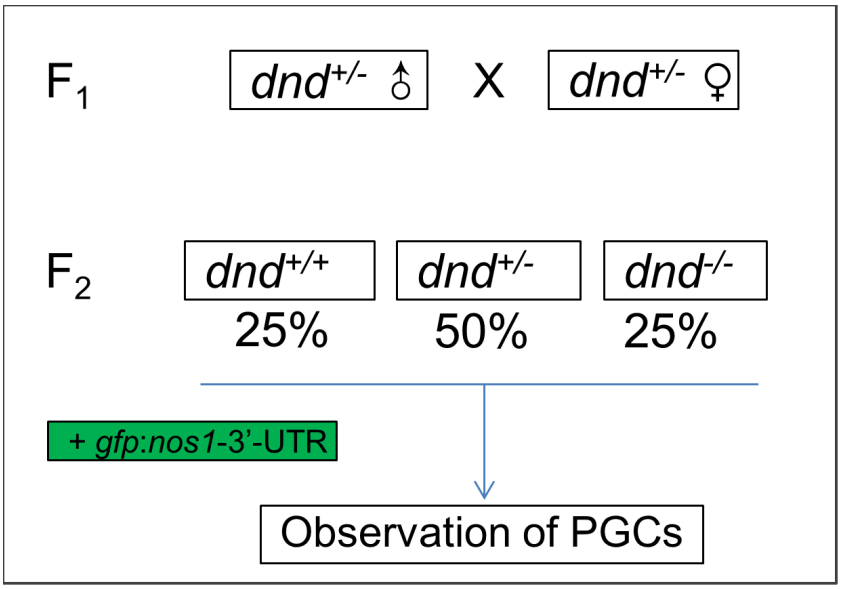


Figure 3.11. Flowchart of labeling PGCs with *gfp* mRNA with nanos 3'-UTR. F_2 fish heterozygous for GE alleles were crossed with each other. The produced embryos were injected with *gfp*:*nos1*-3'-UTR mRNAs and observed for PGCs. Note that 25% of the produced embryos are projected to be $dnd^{-/-}$.

After the F_2 embryos of dgd1 and dgd2 family were injected with the *gfp* mRNA as described above, they were observed under fluorescence stereomicroscope daily. The PGCs in the embryos of dgd1 family were completely normal through the observation period from 2- to 10-dpf as compared to wild type controls (data not shown). However, the PGCs in the embryos of dgd2 family only seemed normal for the first 2 to 3 days. On the third to fifth days post fertilization, the GFP labeled PGCs of some embryos were found to start disappearing (Fig. 3.12). The disappearance of GFP signal indicated that the PGCs either simply died out or changed their lineage to soma, which no longer sustain the protection for *gfp*:*nos*-3'-UTR. All of the PGCs in those embryos that have phenotype would disappear latest after 7 days post fertilization (data not shown).

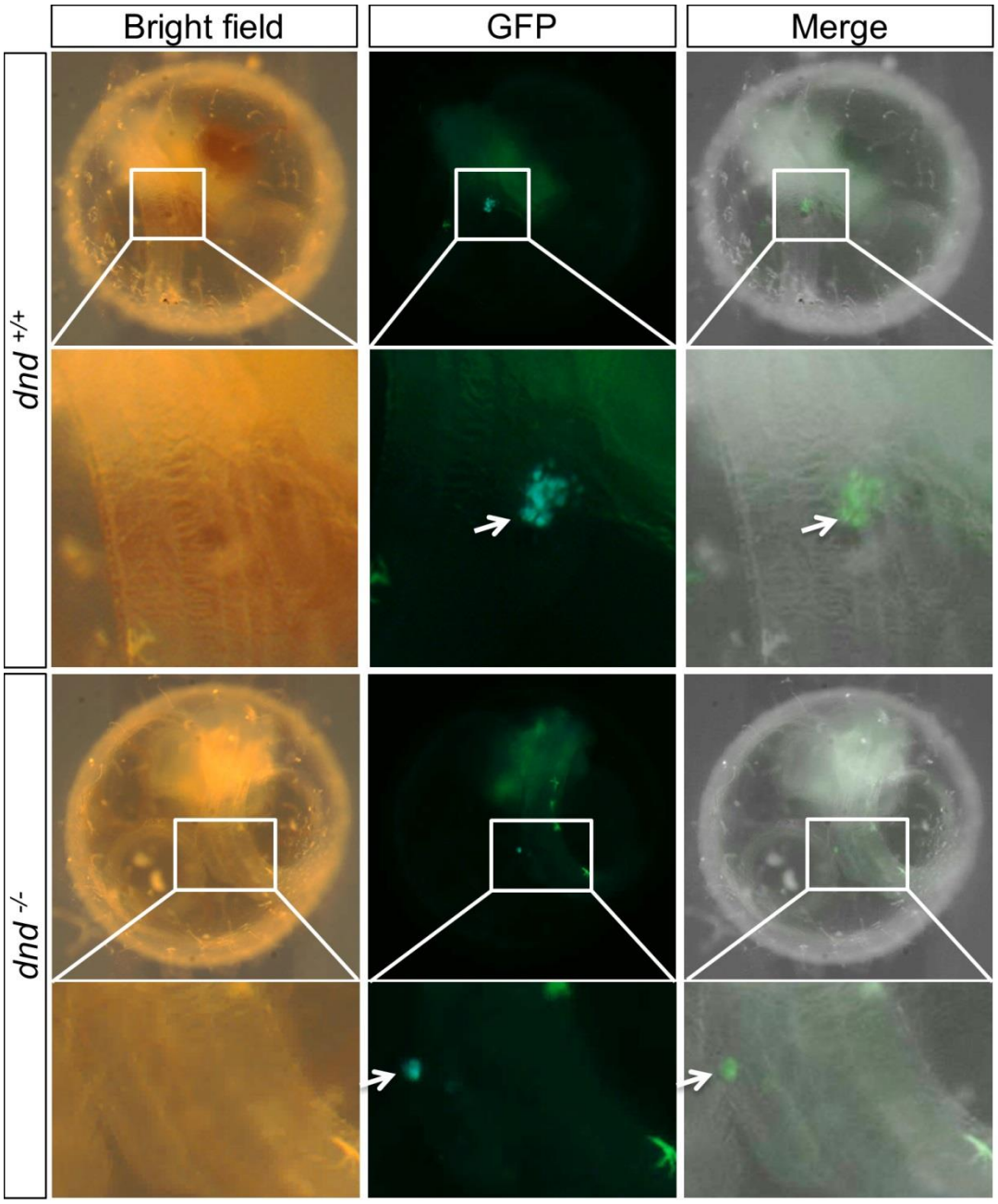


Figure 3.12. Observing PGCs labeled with *gfp* mRNA with nanos 3'-UTR. The phenotype was obviously revealed upon 5th day of embryonic development. The top 6 figures show the wild type embryo with the PGCs in the gonad. The bottom 6 figures show the phenotype of *dnd* $^{-/-}$ embryos with severely decreased PGC number in the gonad.

The start point of this observation is stage 21 (1 day 10 hr, or 34 hr post fertilization), when there are 6 somites in the embryo during the second day of development. The reason is that only from this point the PGCs are clearly visible,

because they align dorso-bilaterally along the axis of the embryo instead of being among other cells, thus the somatic cells with inherited background GFP do not interfere with the PGCs. With a closer observation of the F_2 embryos from dgd2 family, it was found that among the embryos who showed disappearance or severely decrease in number of PGCs upon 5-dpf (Fig. 3.13D and D', Fig. 14B, B', D and D'), some of the embryos showed impaired PGCs with dimmer GPF signal as early as 2-dpf (Fig. 3.13C and C'). In some other embryos, impaired PGC number was observed together with several PGCs with retarded migration (Fig. 3.14A and A'). However, some embryos showed no obvious phenotype at 2-dpf (Fig. 3.14C and C') even they later showed disappearance of PGCs at 5-dpf (Fig. 3.14D and D'). As the embryos develop, the PGCs continued to migrate to the gonad but many of them gradually disappeared in this process. From these observations, it can be concluded that some of the embryos produced by parents heterozygous for dnd D10 showed a phenotype of decreasing PGC number as early as 2-dpf, which finally led to the disappearance of all the PGCs upon 5-dpf.

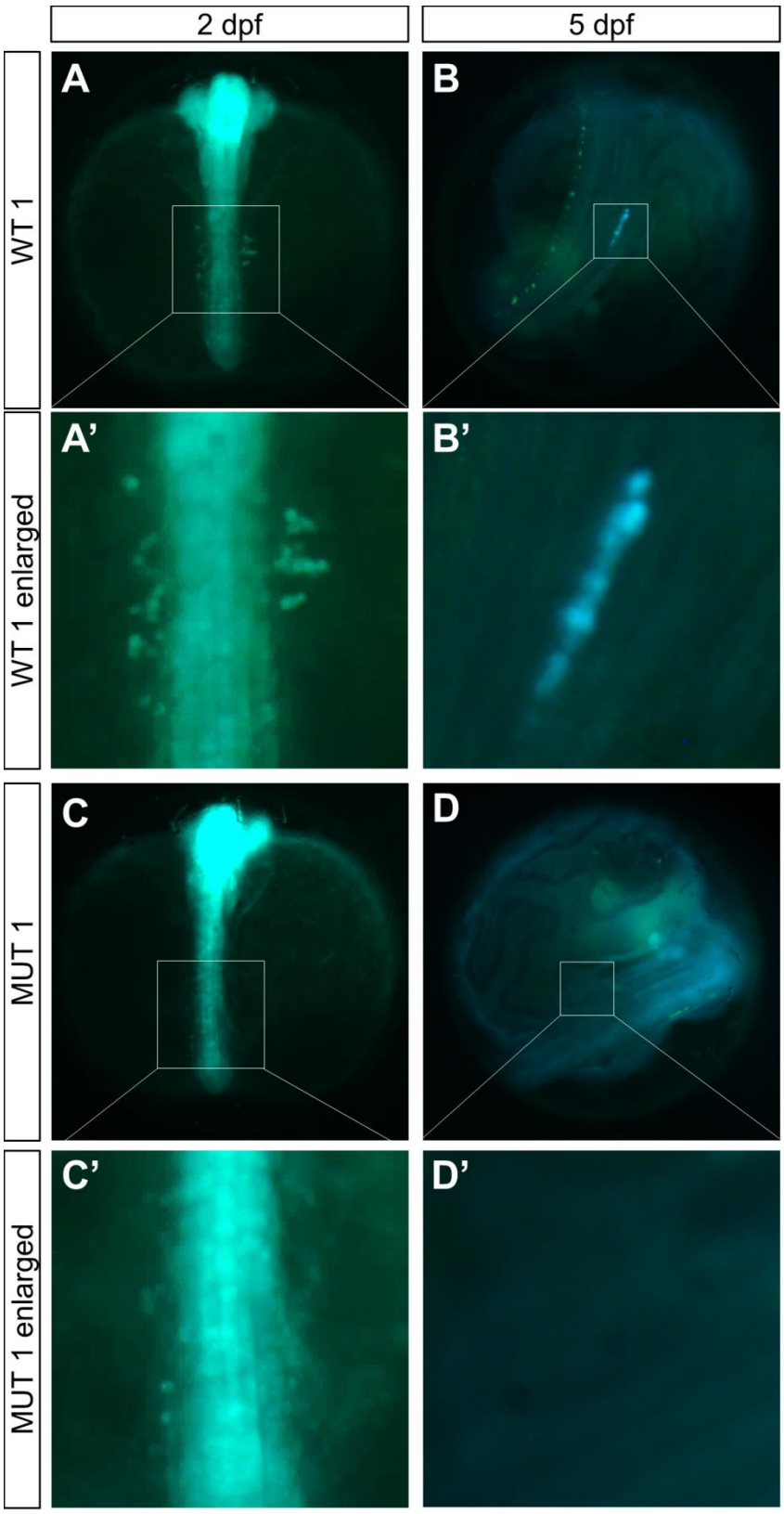


Figure 3.13. Further investigation of the phenotype. Some embryos exhibited decreased PGC number phenotype as early as 2 dpf, the PGCs disappeared upon 5-dpf. (A) A wild type 2-dpf embryo. (A') Enlarged view of the PGCs. (B) The 5-dpf photo of the same wild type embryo. (B') Enlarged view of the PGCs. (C) A 2-dpf embryo that shows phenotype. (C') Enlarged view of the decreased number of PGCs. (D) The 5-dpf photo of the same embryo that shows phenotype. (D') Enlarged view of the gonad position, no PGCs can be observed.

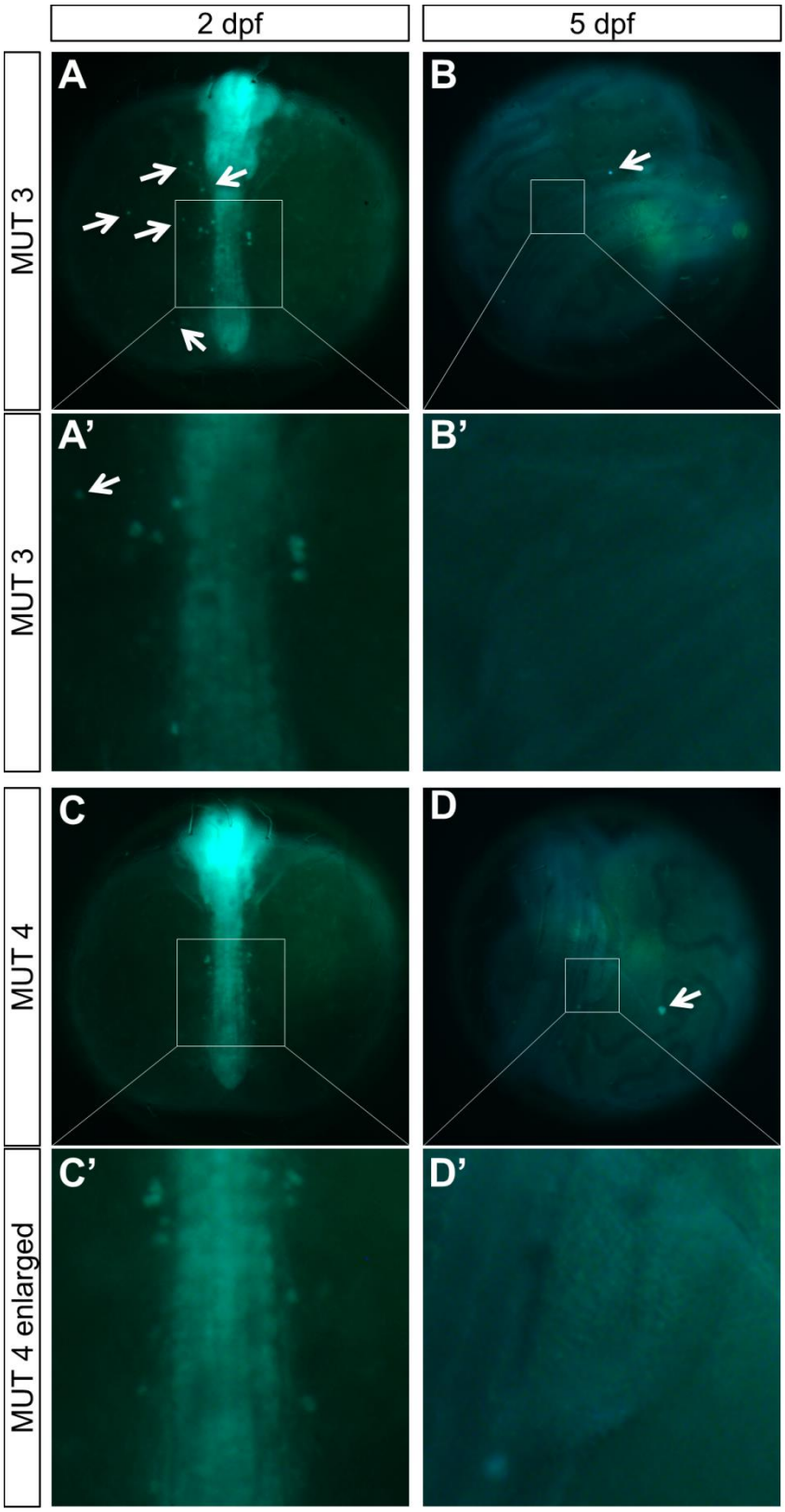


Figure 3.14. Phenotype affecting migration was also observed in some embryos. Some of the 2-dpf embryos exhibited PGCs with impaired migration. (A) A 2-dpf embryo with 5 PGCs (arrows) that are not in the normal position. (A') Enlarged view of the PGCs, note the number of them is also decreased. (B) The 5-dpf photo of the same embryo, arrowed is the PGC still with GFP signal that have not migrated properly. (B') Enlarged view of the gonad, no PGCs. (C) A 2-dpf embryo that shows no obvious phenotype. (C') Enlarged view of the PGCs. (D) The 5-dpf photo of the same embryo that shows phenotype. Note a cell with GFP signal is found (arrow), could be a PGC with impaired migration. (D') Enlarged view of the gonad position, only one PGC at the left bottom corner can be observed.

At this point, it can only be proposed that the embryos with a loss of PGC phenotype are homozygous for the *dnd*-deficient mutant allele D10. In order to confirm that the phenotypes are truly caused by the loss of functional Dnd, genotyping of the embryos with phenotypes were performed. In the PCR-HMA, all the amplicons from embryos with phenotype turned out to be a unified single band, indicating a homologous composition of the alleles (Fig. 3.15A). Subsequently, PCR products from wild type alleles were added to the ones from samples. A denaturing and annealing procedure was used to create heteroduplex of sample PCR products and wild type ones. The re-annealed PCR products were then used to perform another HMA. The second HMA revealed that the *dnd* alleles in the samples were different from the wild type, and they exhibited the same heteroduplex pattern as the D10 alleles (Fig. 3.15B). Sequencing results of the samples' PCR products confirmed they were D10 alleles. These results showed that all the embryos with impaired PGCs were proved to be of $dnd^{-/-}$ genotype.

Overall, some of the embryos produced by the mating of F_1 fish, which were heterozygous for *dnd* D10 mutant allele, showed impaired PGC numbers as early as stage 21, and these embryos' PGCs continued to disappear during the embryonic development. All of the PGCs in these embryos disappeared within 7 days post fertilization. These embryos were then confirmed to be homozygous for the *dnd* D10 mutation. In order to further characterize the phenotype of impaired PGC number at stage 21, additional investigation with the ability to give a higher resolution of the PGCs should be carried out, which will be discussed in section 3.2.5.

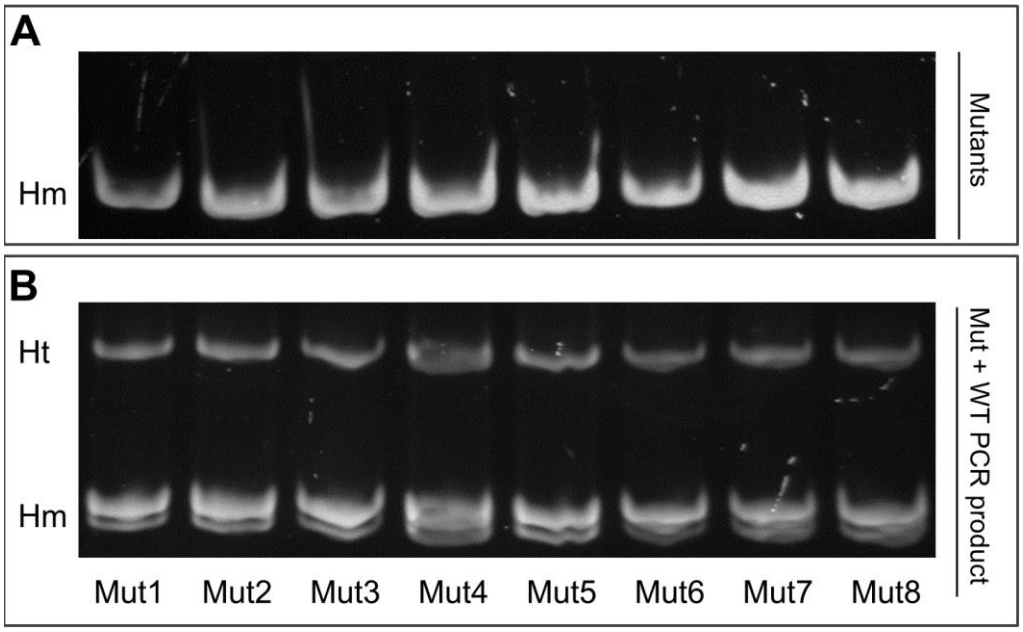


Figure 3.15. Embryos showing phenotypes were proven as *dnd*$^{-/-}$. Eight of the embryos with phenotypes were subjected to genomic DNA extraction and genotyping. Their genotypes were confirmed to be *dnd*$^{-/-}$, indicating the phenotypes were caused by homozygous loss of *dnd*. (A) Shown is the PCR-HMA by PAGE result, indicating the genomic DNA only consists of one kind of *dnd* allele. (B) Shown is the second HMA result of the eight embryos after mixing with wild type *dnd* PCR product and re-annealing. The heteroduplex bands have the same pattern with the dgd2 F_1 fish, proving the existence of D10 allele.

3.2.3. *dnd*-deficient adults have severely underdeveloped gonads

Provided that the germ cells of *dnd*-deficient fish had all disappeared during embryonic development, and the adults were all sterile, it was very likely that their gonads were affected. Therefore, those fish were dissected after genotyping, fertility test and photograph taking. When the trunks of the fish were opened from the ventral side, no ovaries or testes were revealed. It took very close and careful observations for one to find a very thin and tube-like organ within the area where normally ovaries and testes would locate. These organs were found to link to the gonadopore

when I was trying to isolate them by pulling them away from the fish. The found organs were likely to be the gonads without germ cells involved in their development and looked the same regardless of the genomic sex (Fig. 3. 16A and B).

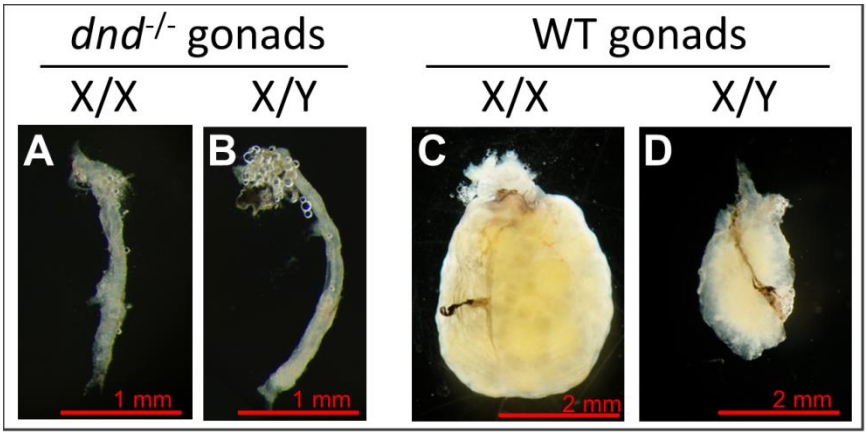


Figure 3.16. Primary sexuality of *dnd*-deficient fish. Morphology of gonads from fish of different genotypes. Note that the scale bars are different in size for underdeveloped gonads and the wild type ones. Thus underdeveloped gonads of *dnd* deficient fish are much smaller and thinner than the wild type ones. They are roughly half the length of the wild type ones.

3.2.4. Gene expression pattern of *dnd*-deficient gonads

Markers of medaka germline were examined in the *dnd*-deficient gonads. Specifically, the expressions of *dazl* and *vasa*, which are expressed in germ line stem cells of both sex, and *dmy*, which is expressed in male germ line stem cells, were examined by using RT-PCR. However, no germline markers were detected in the *dnd*-deficient gonads (Fig. 3.17). Subsequently, *sox9b*, a marker for gonadal somatic cells, was also examined. However, *sox9b* was not detectable, either. The results of routine RT-PCR indicated that the quality and amount of total RNAs extracted from the tiny gonads by Trizol method were not good enough. So I decided to use RNeasy

micro kit and digital droplet-PCR (dd-PCR) for higher recover rate of total RNA and more precise quantification.

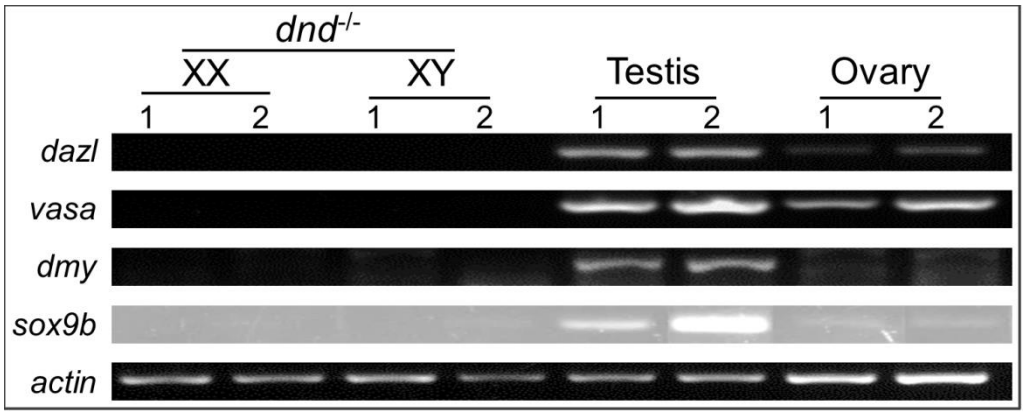


Figure 3.17. RT-PCR results of the gonads of *dnd* deficient fish. As shown, the gonads of *dnd* deficient fish lost the germline markers completely, which include *dazl*, *vasa*, *dmy*. Sox9b, a gonadal somatic cell marker, wasn't detected in the *dnd* deficient fish.

Indeed, the combination of RNeasy micro kit and dd-PCR gave rise to better results. Still, no germline stem cell markers were detectable in those gonads (data not shown), while some somatic gonadal markers were detectable (Figure 3.18). For example, ovary somatic marker aromatase and *foxl2* were detectable. However, while the aromatase's mRNA levels of the underdeveloped gonads were much lower than that of the wildtype ones, the *foxl2*'s mRNA levels of the underdeveloped gonads were much higher than that of the wildtype ones, but the X/X *dnd*-deficient gonads still showed higher levels of mRNA than the X/Y ones. For male somatic gonadal markers, dmrt1 showed similar transcription level in both sexes of *dnd*-deficient gonads, which were both one fifth of the wildtype male one. High transcription level of gsdf in wildtype testis was not seen in either genotypes of *dnd*-deficient gonads. The mRNA levels of *gsdf* in dnd-deficient gonads were similar

to the wildtype counterparts. However, the data is not repeated and consolidated for now. Further repeats will be done to consolidate the data and give a clearer view of the gene expression pattern in the *dnd*-deficient gonads.

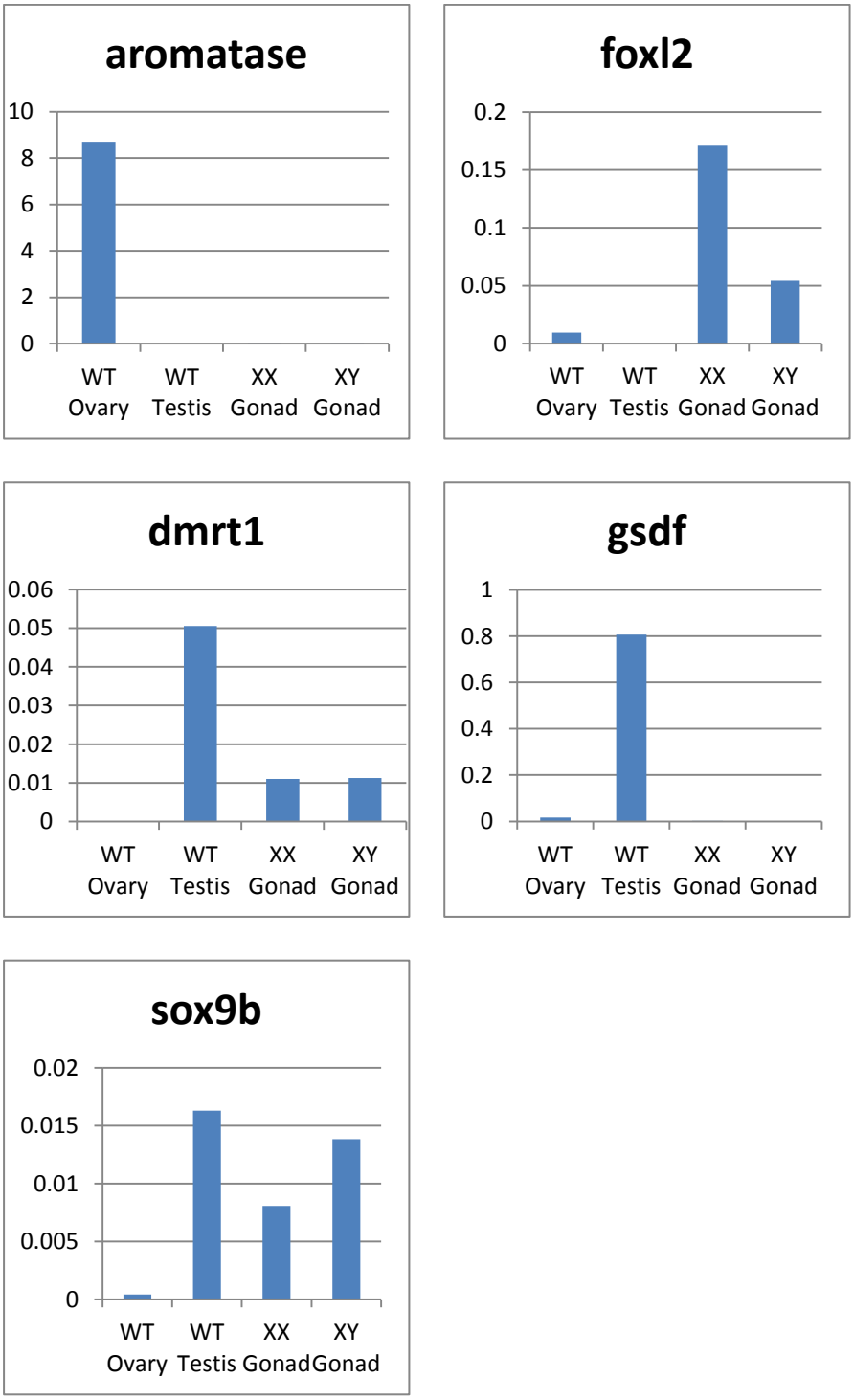


Figure 3.18. dd-PCR results of the gonads of *dnd* deficient fish. As shown, the gonads of *dnd* deficient fish lost the germline markers completely, which include *dazl*, *vasa*, *dmy*. Sox9b, a gonadal somatic cell marker, wasn't detected in the *dnd* deficient fish.

3.2.5. Loss of germ cells phenotype correlates with *dnd* quantitatively

The *gfp*:*nos*-3'-UTR labeling method comes with an inherent defect, which is the back ground created during early embryonic stages before the separation of germline and soma, as mentioned before. This background greatly reduced the signal to noise ratio, thus reducing the resolution of the image. However, the image quality needed to be improved to enable more detailed analysis of the phenotype caused by loss of *dnd* in the embryos. Therefore, whole-mount in situ hybridization of *vasa*, also a germline-specific gene, was used to label the PGCs in embryos instead of GFP. Subsequently, time-lapsed whole-mount in situ hybridization with the probe for *vasa* (*vasa*-WISH) was used to dissect the process of disappearance of PGCs (Fig. 3.18). Time points were selected as 34-hpf (embryonic stage 21), 2-dpf, 3-dpf and 4-dpf.

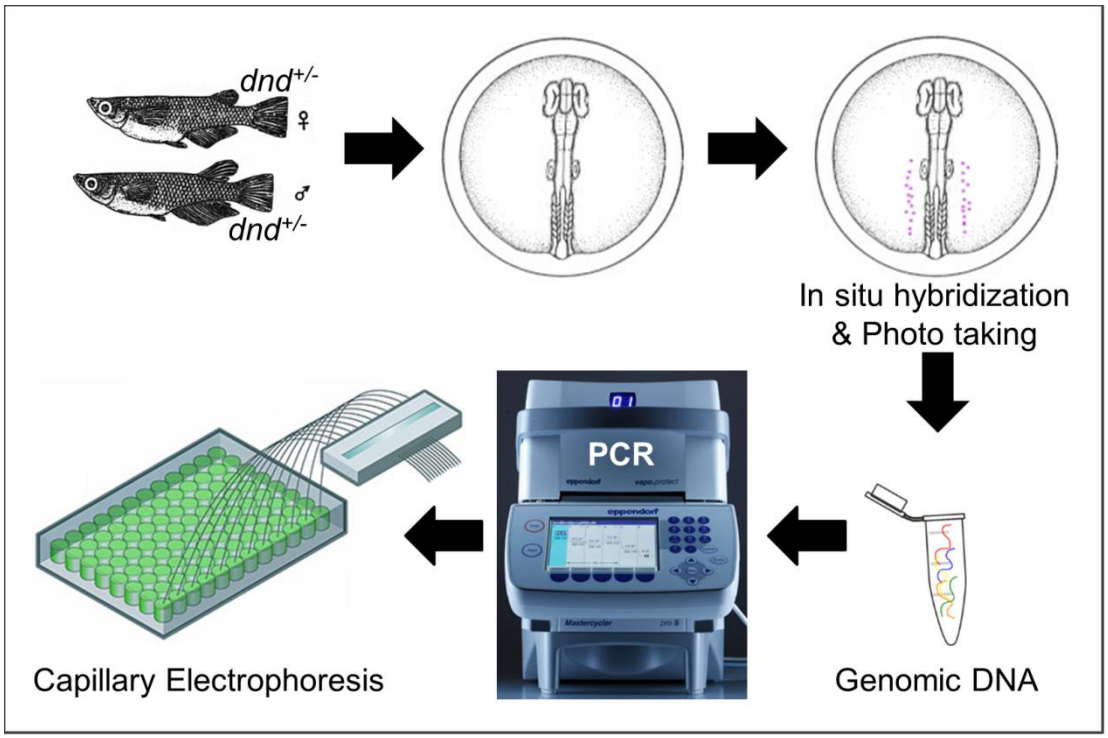


Figure 3.19. Flowchart of examination of the correlation between PGC number

and *dnd* genotype. This flowchart shows how the experiment was performed for combined whole mount in situ hybridization and genotyping using capillary electrophoresis.

In contrast to the *gfp*:*nos*-3'-UTR-labeled fluorescent images of the stage 21 embryos, the images of the in situ hybridization results had a very high resolution so that the single PGCs were distinguishable from each other. Because the PGCs align bilaterally along the axis of the embryos at stage 21 and generally disperse in one layer within this area, the PGC number can be counted (Fig. 3.19A to D). The *in situ* hybridization was performed on a total of 49 embryos spawned by $dnd^{+/-}$ parents and 7 embryos spawned by wild type parents as a control. In the assay, the wild type-produced embryos showed an average PGC number of 41 (Fig. 3.19E). On the other hand, the $dnd^{+/-}$-produced embryos showed an average PGC number of 25. Among these $dnd^{+/-}$-produced embryos, some had few PGCs while others had a number of PGCs, which is comparable to that of wild type-produced embryos. This finding provoked the analysis of the genotypes of these embryos by lysing the embryos that had gone through in situ hybridization and extracting their genomic DNAs for PCR and subsequent genotyping method. Because PCR-capillary electrophoresis (PCR-CE) was more convenient than PCR-HMA but can only be used in F_1 fish when the mutant allele of *dnd* was an investigated one, this method was chosen instead of PCR-HMA for the genotyping of those embryos. In PCR-CE, the $dnd^{+/+}$ embryos showed a single peak at 196 bp, representing the wild type PCR product; the $dnd^{-/-}$ embryos showed a single peak at 186 bp, representing the D10 allele's PCR product; the $dnd^{+/-}$ embryos showed both PCR products due to the

heterozygosity of the sample (Fig. 3.19F).

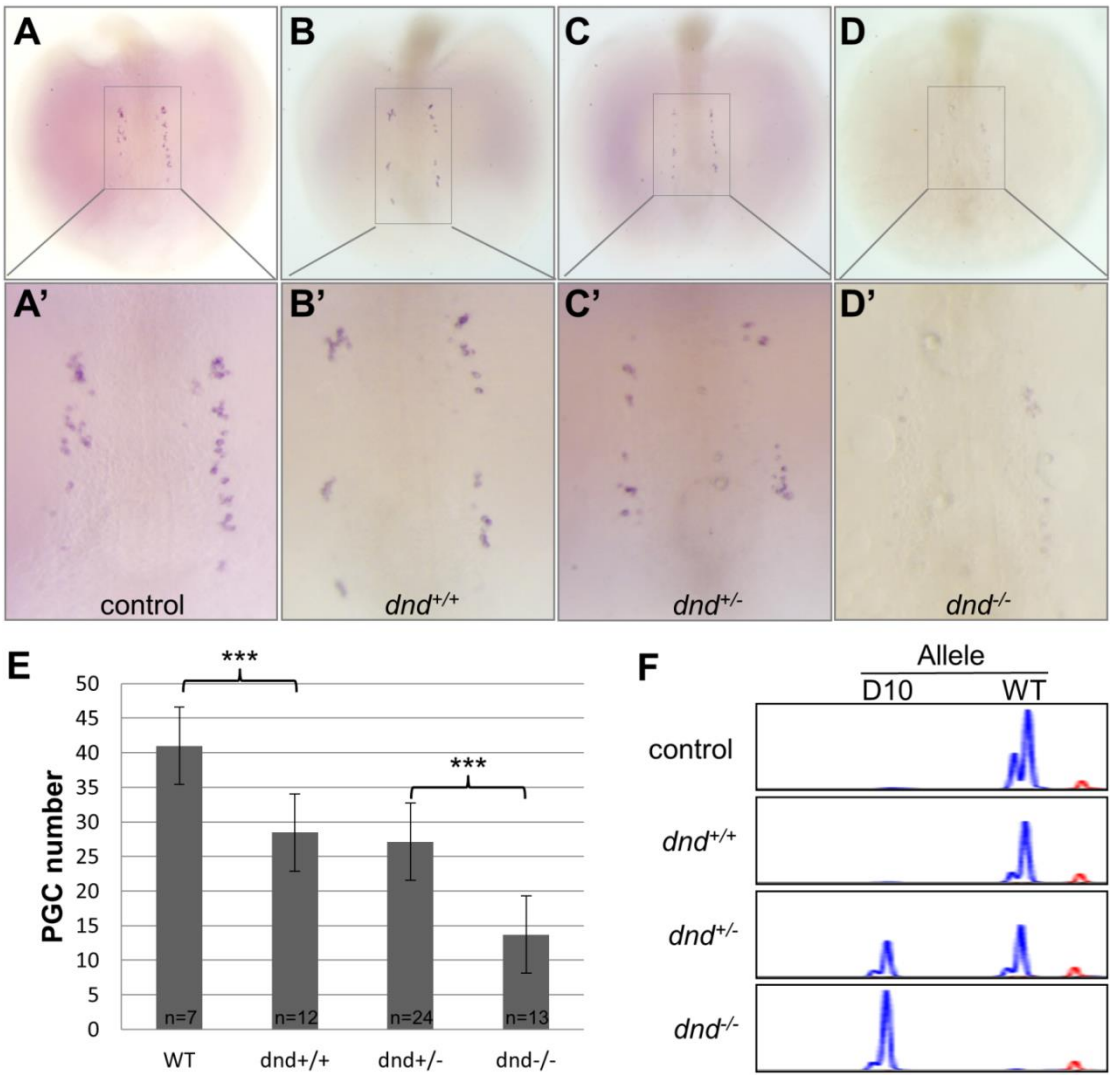


Figure 3.20. *dnd* dosage and PGC number in medaka embryos. Embryos from crosses between fish herterozygous for a mutant dnd allele (D10) were collected at stage 21 (6 somites), analyzed with WISH by using an antisense vasa riboprobe for counting the number of PGCs and genotyped. (A-D) Representative embryos of different genotypes, showing that somatic development is normal until stage 21. (A'-D') Large magnifications of areas boxed in A-D, highlighting the number and positions of PGCs. (E) Statistic analysis of PGC numbers. Numerals within columns refer to the numbers of embryos analyzed. (F) Representative result of genotyping of the embryos.

It turned out that the result of PGC number in the stage 21 embryos correlated with the genotypes of the embryos (Fig. 3.19E). Interestingly, as compared with the

wild type average PGC number of 41, the *dnd*$^{+/+}$ embryos from *dnd*$^{+/-}$ parents showed an average PGC number of 28, which is about one-fourth less than their counter parts from wild type parents. Although the genotypes of these two groups of embryos were the same, the PGC numbers diverged. The only difference between them was the parents' genotypes. For the *dnd*$^{+/-}$ embryos from *dnd*$^{+/-}$ parents, they showed an average PGC number of 27, which was close to their *dnd*$^{+/+}$ siblings, despite *dnd*$^{+/-}$ embryos had only one copy of functional genomic *dnd*. However, for the *dnd*$^{-/-}$ embryos, the PGC number was averagely 13, which was much lower than the previous ones.

The difference of PGC number in stage 21 between wildtype-born embryos and dnd heterozygotes-born ones indicates that maternal effect may have been playing a role here. Therefore, 30 embryos from WT parents and another 30 from dgd2$^{+/-}$ parents were smashed and the total RNA was extracted and subjected to dd-PCR using *dnd* and *actin* primers. The quantity of *dnd* mRNA of each group indeed differed and the WT group showed roughly twice the amount of dgd2$^{+/-}$ group. This data shows that less *dnd* mRNA will lead to less PGC formation during early embryonic development.

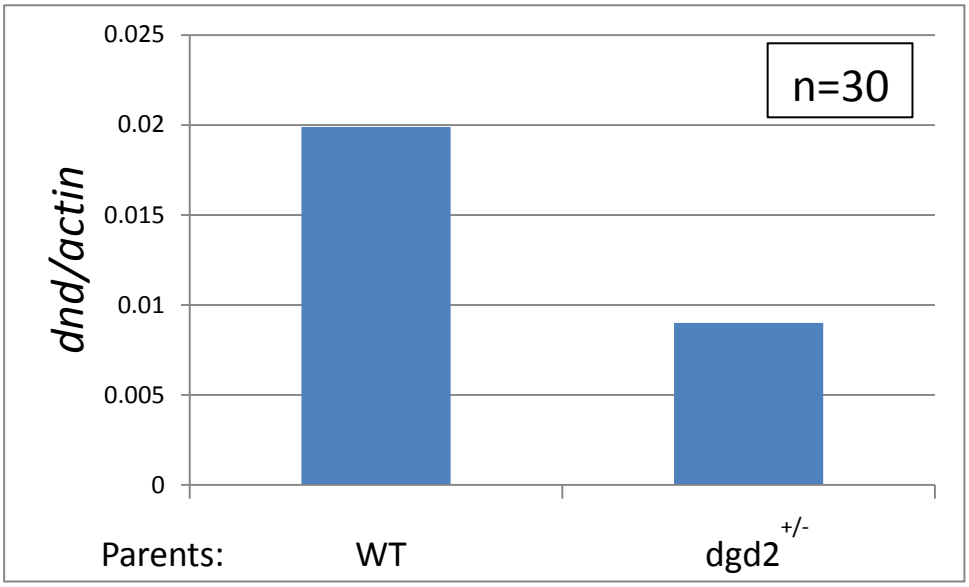


Figure 3.21. Quantification of maternal *dnd* mRNA in WT-born and dgd2$^{+/-}$-born embryos. 30 embryos of each group were smashed and the total RNA was extracted. Dd-PCR results of *dnd* normalized to actin are shown above.

Collectively, the PGC number of stage 21 embryos correlated with two parameters: the genotypes of the parents and their own genotypes. Generally, the embryos from wild type parents showed more PGCs than those from *dnd*$^{+/-}$ parents, which is caused by the different amount of maternal *dnd* mRNA. Specifically, within those embryos from *dnd*$^{+/-}$ parents, *dnd*$^{+/+}$ and *dnd*$^{+/-}$ embryos showed more PGCs than *dnd*$^{-/-}$ ones, which is possibly caused by the different level of zygotic *dnd* mRNA synthesis.

3.2.6. *Vasa*-WISH reveals earlier onset of disappearance of PGCs

Vasa-WISH was also performed on the 2-, 3- and 4-dpf embryos as described. Surprisingly, unlike results obtained by GFP-labeling method, *vasa*-WISH results showed that, instead of some, all of 2-dpf *dnd*-deficient embryos showed severe

decrease of PGC number (Fig. 3.20A and A'), as compared to wild type embryos (Fig. 3.20D and D'). 3-dpf embryos also showed similar phenotype (Fig. 3.20B and B'), as compared to wild type embryos (Fig 3.20E and E'). In 3-dpf embryos, the decrease of PGC number is more prominent, since the remaining number of PGCs in the dnd-deficient embryos is further decreased, while number of PGCs in the wild type embryos is increasing in line with embryonic development. In the *vasa*-WISH results of 4-dpf embryos, complete ablation of PGCs was observed (Fig. 3.20C and C'), as compared to wild type embryos (Fig. 3.20D and D'). This is earlier than the data from GFP-PGC labeling method.

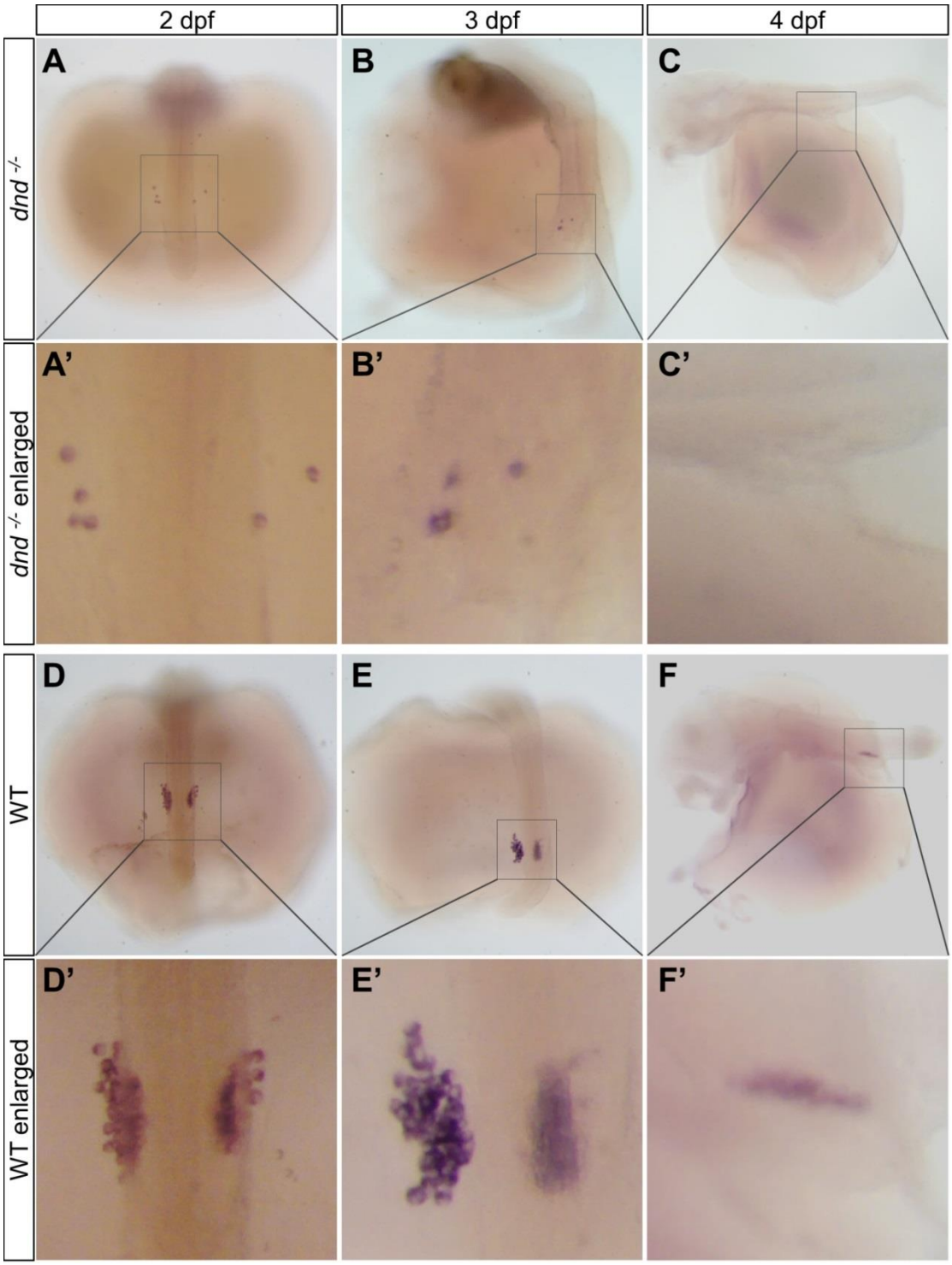


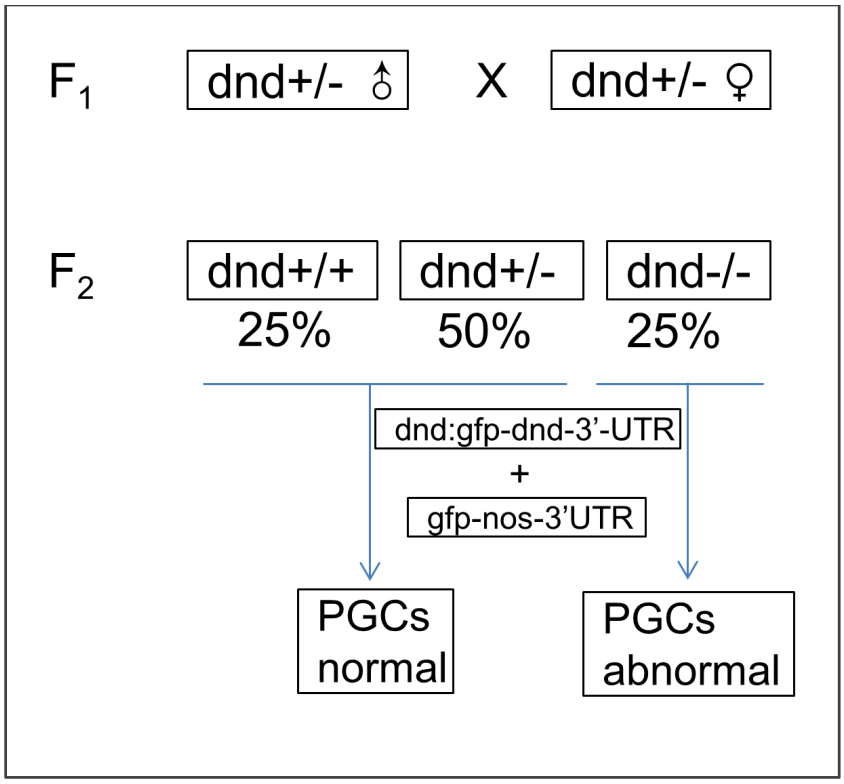
Figure 3.22. In situ hybridization of 2 to 4 days *dnd*$^{-/-}$ embryos. As shown, 2-dpf embryo with *dnd*$^{-/-}$ genotype showed severely decreased PGC number (A, A'), as compared to WT one (D, D'). 3-dpf *dnd*$^{-/-}$ embryos also showed the disappearance of most PGCs (B, B'), but not the wild type ones (E, E'). 4-dpf *dnd*$^{-/-}$ embryo showed complete disappearance of all PGCs (C, C'), as compared to wild type embryo (F, F').

3.2.7. *dnd* mRNA not only rescued the phenotype but also induced gain of function

In order to confirm that the phenotypes observed above were genuinely caused by the loss of Dnd, rescue experiment was performed (Fig. 3.21, top). If the observed phenotypes could be rescued by externally introduced Dnd, it could be concluded that they were caused by loss of Dnd. Embryos produced by $dnd^{+/-}$ fish were chosen as the subjects to be rescued. In vitro transcribed mRNA encoding for a fusion protein of Dnd and GFP were used to rescue the embryos. In the same time, the PGCs were visible because of fused GFP for the convenience of observation. This mRNA (*dnd*:*gfp*:*dnd*-3'-UTR) included a 3'-UTR which was derived from gene *dnd* itself, ensuring genuine germline-expression of Dnd.

One-cell stage embryos from the heterozygous parents were injected with *dnd*:*gfp*:*dnd*-3'-UTR. However, the ones injected with an mRNA concentration of 100 showed a very low survival rate of 6% only upon gastrulation (Fig. 3.21, bottom). The reason for initially using 100 ng/µl was that the same concentration used for *gfp*:*nos*-3'-UTR didn't cause any embryo deaths. The concentration was halved to 50 ng/µl and the survival rate for the microinjection experiment went up to 39%. It was expected that only one fourth of the embryos were $dnd^{-/-}$ ones – targets for rescue (Fig. 21, top). Therefore, a 39% survival rate means that only 10% of the experimented embryos would be surviving $dnd^{-/-}$ ones, making this concentration unsatisfactory. The concentration was then halved again to 25 ng/µl. With this

concentration, the survival rate increased to almost 90%, which was satisfactory.



dnd (ng/ul)	100	50	25
Survival	4/67 (5.97%)	21/54 (38.89%)	62/69 (89.86%)

Figure 3.23. Flowchart of rescuing loss of dnd phenotype and the survival rates of injecting *dnd*:*gfp*-*dnd*-3'-UTR. Top shows the flowchart of rescuing experiment. Bottom shows the survival rates increase to a satisfied level after the dosage has been decreased to 25 ng/μl.

A total of 50 embryos produced by $dnd^{+/-}$ parents were injected with 25 ng/μl *dnd*:*gfp*:*dnd*-3'-UTR and 100 ng/μl *gfp*:*nos*-3'-UTR. Another 50 embryos produced by $dnd^{+/-}$ parents were injected with 100 ng/μl *gfp*:*nos*-3'-UTR only, serving as control. These embryos were observed daily under fluorescent microscopy. The same

as previous experiments, some of the embryos from control group started to lose PGCs from day two post fertilization (Fig. 3.22B and B'), while others remain normal (Fig. 3.22A and A'). However, none of the embryos from experiment group showed losing of PGCs. In contrast, they exhibited increased PGC numbers (Fig. 3.22C and C'). When the embryos were observed 5 days post fertilization and injection, PGCs had already migrated into the gonads in most embryos of the control group (Fig. 3.22D and D'), but 12 of them exhibited loss of all germ cells (Fig. 3.22E and E'). In the same time, all of the experiment group's embryos exhibited excessive amounts of germ cells in the gonad (Fig. 3.22F and F'). Subsequently, all the embryos were genotyped using PCR-HMA. The results showed that the 12 germ cell-less embryos from the control group were all of *dnd*$^{-/-}$ genotypes (data not shown) and there were 13 *dnd*$^{-/-}$ embryos out of the 50 embryos with excessive PGCs from the experiment group.

Collectively, externally introduced *dnd* mRNA not only successfully prevented the loss of germ cells phenotype during embryonic development, but also enhanced the amounts of germ cells in the embryos. This result indicated that this phenotype is truly induced by loss of *dnd* in embryos.

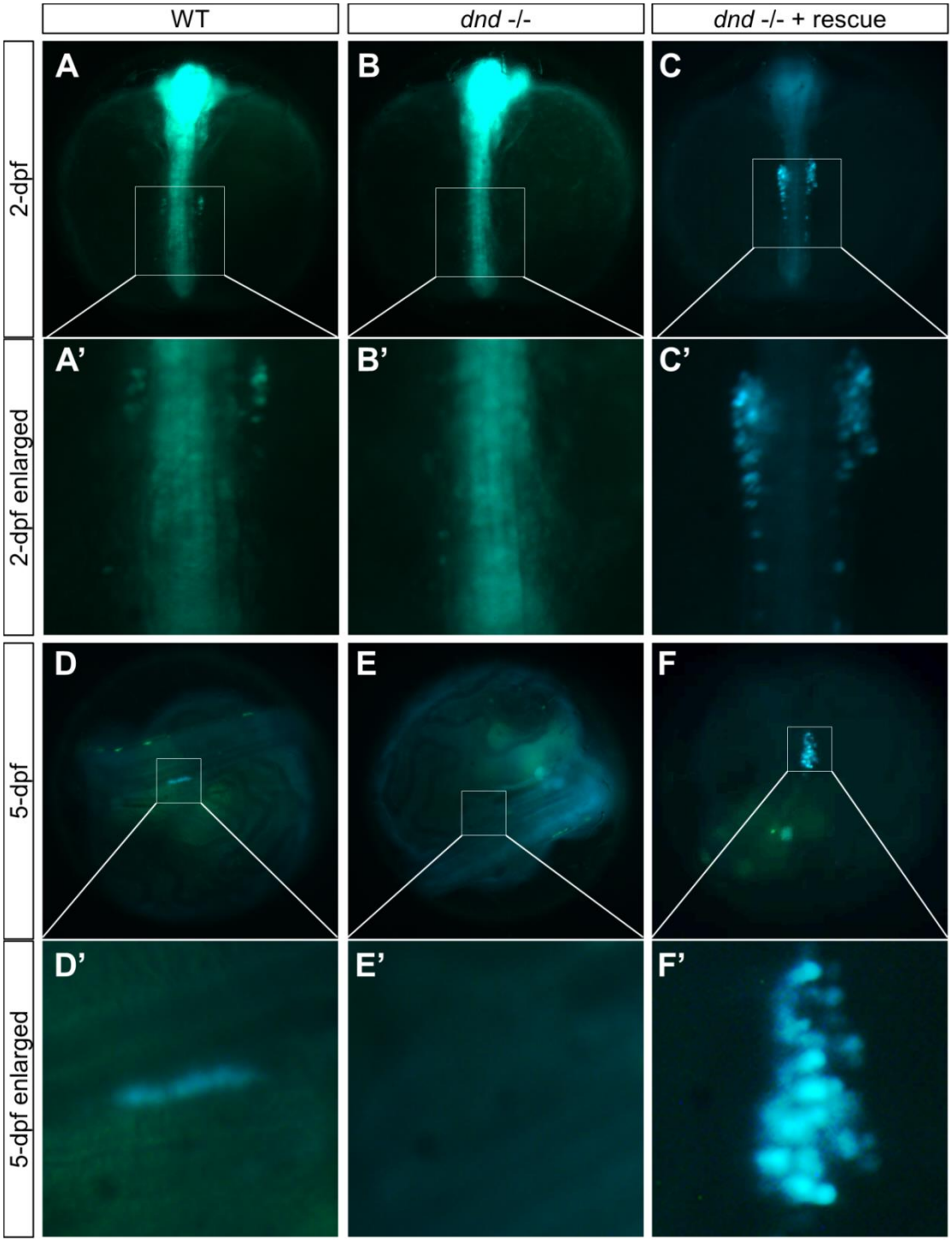


Figure 3.24. Rescue the loss of PGC phenotype in early stage embryos with *dnd* mRNA. The left and middle columns show gfp-nos-3'UTR-labeled wild type embryos (A, A', D, D') and embryos with phenotype (B, B', E, E'). Embryos rescued are shown at the right column (C, C', F, F'). Interestingly, all of the embryos rescued showed increased number of PGCs, regardless of the phenotypes. No rescued embryos showed disappearance of PGCs before hatching.

In sum, the efficiency of TALEN-mediated gene editing of dnd was explored and determined and the phenotypic results of loss of dnd were characterized. In the discussion section, different approaches of programmable nuclease-mediated genome engineering will be discussed as compared to this thesis. Moreover, the *dnd* deficient phenotype will also be discussed and compared with other similar approaches.

CHAPTER 4 DISCUSSION

4.1. Summary of results

In this thesis, I have: **(1)** established procedures and parameters for TALEN-mediated gene editing on the chromosomal locus *dnd* in medaka embryos via mRNA injection; **(2)** explored and characterized the phenotype of *dnd* deficiency in medaka embryos and fish; **(3)** provided evidence for a quantitative correlation between *dnd* and PGC number, indicating an important role of *dnd* in orchestrating PGC formation.

It is shown that the survival rate and GE efficiency are dependent on the dose of injected TALEN mRNAs. Specifically, TALENs at an optimal dose of 100 μg/μl is able to produce a high survival rate of up to 85% and a GE efficiency of 55% at the hatching stage. Both lower and higher dosages produce lower GE efficiency. The GD-positive embryos developed normally into fertile female and male adults capable of transmitting GE alleles into 7.7 to 13% progeny of the F1 generation. Therefore, the TALEN technology is proficient for GE in the soma and germline of medaka embryos. These data provide valuable information for TALEN-mediated GE in medaka as well as other fish species.

It is shown that the embryos without genomic *dnd*, but still with maternal *dnd* mRNA, initially have PGCs and lose them during embryonic development. The PGCs in these embryos start to decrease in number since stage 21 and disappear by the 5^{th} day post fertilization. The timings of disappearance of PGCs observed by GFP

labeling method and *vasa*-ISH method differs, with the prior lagging about one day. The adult dnd deficient fish show infertility and no courtship behavior. The gonads of these fish are severely underdeveloped, which are only half of the length of wild type ones and extremely thin. These gonads have lost genetic markers for germline and gonadal somatic cells. These data provide a platform for further studies on the mechanism of *dnd's* role in the germline development and adult germline maintenance.

4.2. Factors that affect TALEN-mediated GE efficiency

In medaka, direct GE by using the ZFN approach has been reported on the transgene *gfp* (Ansai et al., 2013) and the chromosomal gene *gsdf* (Zhang et al., 2014). In both studies, ZFNs produce a wide variety of GE alleles with subtle sequence alterations including additions, deletions and compound changes. In this thesis, TALENs targeting on the chromosomal locus *dnd* can also produce similar diversity of GE alleles. These results suggest that the TALEN and ZFN approach share a common ability to generate a wide variety of allelic alterations in medaka.

In medaka, *gsdf*-targeting ZFNs under optimized conditions have produced a 58% GE efficiency and a 48% survival rate (Zhang et al., 2014). In this study, *dnd*-targeting TALENs at an optimized dosage have given rise to a 55% GE efficiency and an 85% survival rate. More importantly, our results in this study demonstrate that all the four fish from TALEN-injected are capable of transmitting a GE allele to 7.7~13% of F_1 progeny, suggesting a high efficiency in generating germline transmitters. These twin studies appear to point to the advantage of TALENs over ZFNs in terms of GE efficiency, survival and ease with which that TALENs are engineered. This notion is supported by a recent report that the TALEN approach is highly efficient for *dj-1* GE (Ansai et al., 2013).

The GE efficiency observed in this thesis peaks at 32% and the germline transmission rate ranges from 7.7% to 13%. However, a nearly 100% GE efficiency and a germline transmission rate of 43.8%~100% has been obtained on *dj-1 GE* by using TALENs at 50~300 ng/μl (Ansai et al., 2013). Multiple factors may account

for such differences in efficiency. For example, it is known that the efficiency of TALENs is depending on genomic context, with different target genes producing varying efficiencies (Cade et al., 2012). In addition, other differences in the structure of TALENs such as the lengths of N- and C-terminus may also influence the efficiency. In the following sub-sections, factors affecting the efficiency of TALENs will be discussed.

4.2.1. RVDs

In the previous sections, it has already been introduced that HD, NI, NN and NG are commonly used for designing and constructing new TALENs for targeting nucleotides C, A, G and T, respectively. In fact, it has been reported that NK, also an abundant RVD, possesses a more specific binding affinity to G than NN (Miller et al., 2011). In the initial stage of this thesis, a pair of synthesized TALENs, choosing NK for the recognition of G, was first used to target gene *dnd*. The target sequence of that pair of TALENs is CATGATGAACTTCA and TGGCGTAAGCGAAG. However, that pair of TALENs didn't induce detectable GE in medaka embryos after many trials. This indicated that NK may not be a genuine good RVD for G recognition. Thus, I chose NN for G recognition and constructed the TALEN pair reported in the results section, finally getting satisfactory GE efficiency.

Later, a report re-examining both the specificity and the binding efficiency of the potential RVDs was published on Nature Biotechnology (Streubel et al., 2012), explaining why NK was not suitable for my first pair of TALENs. They first used

poly A-, C-, G- or T-boxes of 15 bp long to examine each RVDs' binding efficiency by constructing TALEs consisting of 15 NI, HD, NN, NG, NS, N* or HG and testing all these TALEs on the four poly nucleotide boxes. It was shown that only HD targeting C and NN targeting G showed a strong reporter gene expression (Fig. 4.1a). Other pairing of poly RVDs and poly nucleotide boxes only resulted in very weak expression of the reporter gene. This result indicates that only the binding of HD to C and NN to G is strong enough to trigger stable TALE binding to its target sequence. It was also shown that, in the case without NN or HD, 3 NK or 3 NH were not enough for the targeting of TALEs (Fig. 4.1c). Furthermore, 6 NH but not 6 NK were enough for the targeting of TALEs in this scenario (Fig. 4.1d). This result indicates that NH is an intermediate RVD for G and NK is a very weak one. Finally, specificity test showed that NN in TALEs that already consisting 3 HD not only exhibit strong targeting efficiency to G, but also to A (Fig. 4.1e). On the other hand, when NH or NK were used instead, the targeting efficiency is very specific to G only.

Collectively, for further TALEN RVD designs, 3 to 4 strong RVD (NN or HD) should be included to ensure the binding efficiency and NH should be used for G for the binding specificity if the first rule has already been met. In the case of lack of binding efficiency, NN should be used.

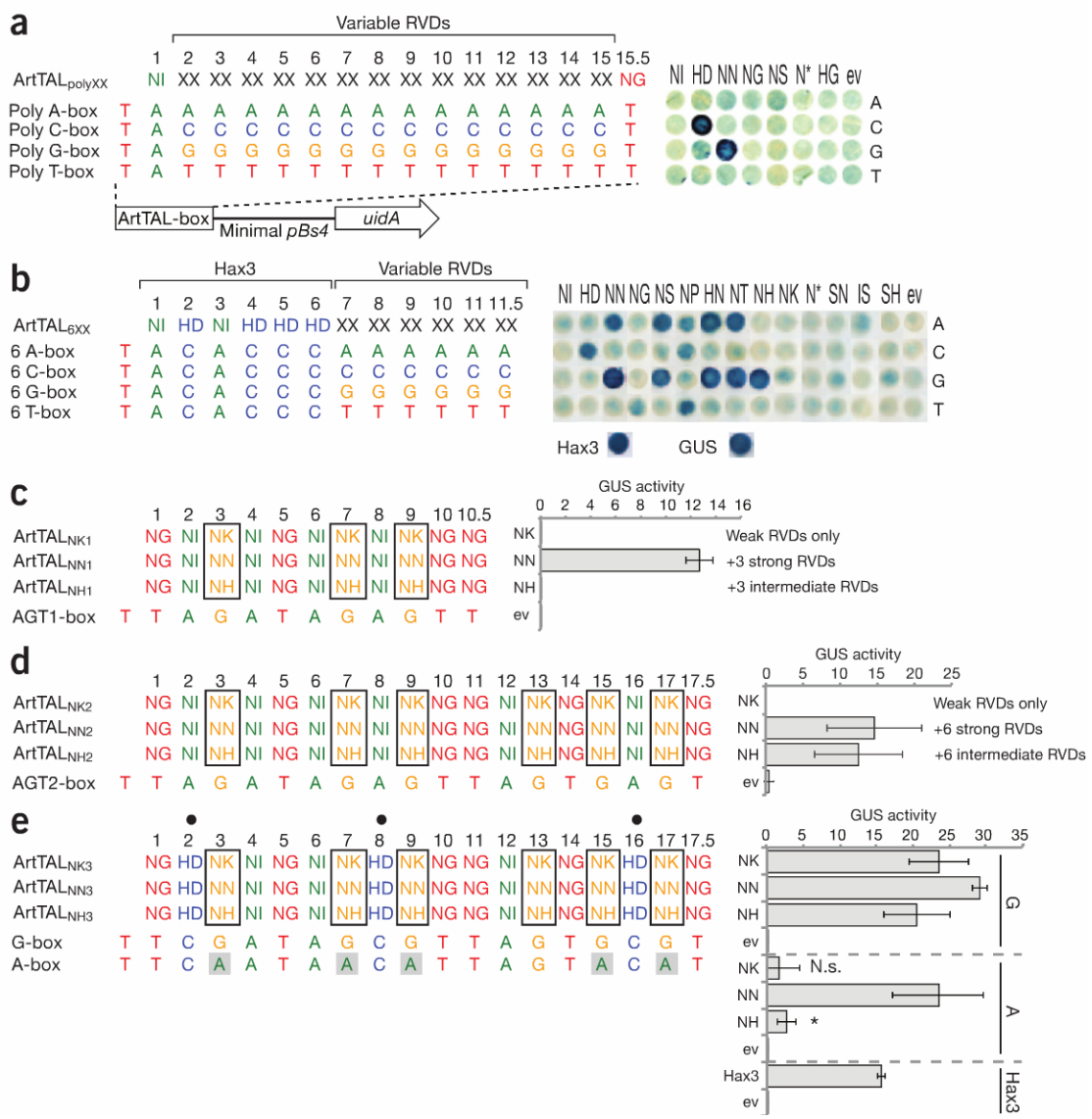


Figure 4.1. TALE RVD specificities and efficiencies. Adapted from (Streubel et al., 2012)

4.2.2. C- and N- terminus of TALE architecture

In this thesis, the TALENs consist of full length of C- and N-terminus, because the truncation experiment of TALENs was not published when this project initiated. However, longer TALENs without truncation may have contributed to the lower GE efficiency. The first TALEN architecture contains a 288 aa of N-terminus and a 231

aa C-terminus (Cermak et al., 2010). The first 136 aa of the N-terminus was found to help the TALEs to enter plant cells (Szurek et al., 2002b). Therefore, a TALE with only 153 aa N-terminus can be created, namely Δ153. A smaller TALEN architecture increases the targeting efficiency per ng/µl, since I didn't use molar concentration for GE efficiency characterization. For C-terminus, the improvement has a bigger impact on the GE efficiency, because the C-terminus is responsible for linking DNA binding domain to the FokI nuclease. Although I mentioned that different truncations in C-terminus didn't affect GE efficiency according to relevant reports, they actually improved the GE efficiency if the spacer is also adjusted according to the length of C-terminus (Christian et al., 2012). In a very aggressive version of TALEN architecture, only 18 aa were left for the C-terminus, namely NΔ153/C+18. This version exhibited 4 times more efficient than the full-length version targeting 14-bp spacer (Fig. 4.2). NΔ153/C+18's efficiency targeting 14-bp spacer is still 2 fold higher than the full-length version targeting its optimal spacer, which is 27 bp (Fig. 4.2). In this thesis, full length TALEN architecture was used, however, the spacer was 15 bp, leaving a lower than normal but still satisfactory GE efficiency. This explains why the various targeted alleles lack addition types, since addition types of GE alleles will induce higher cleavage rate due to elongated spacer, making this kind of alleles be re-cut and eliminated.

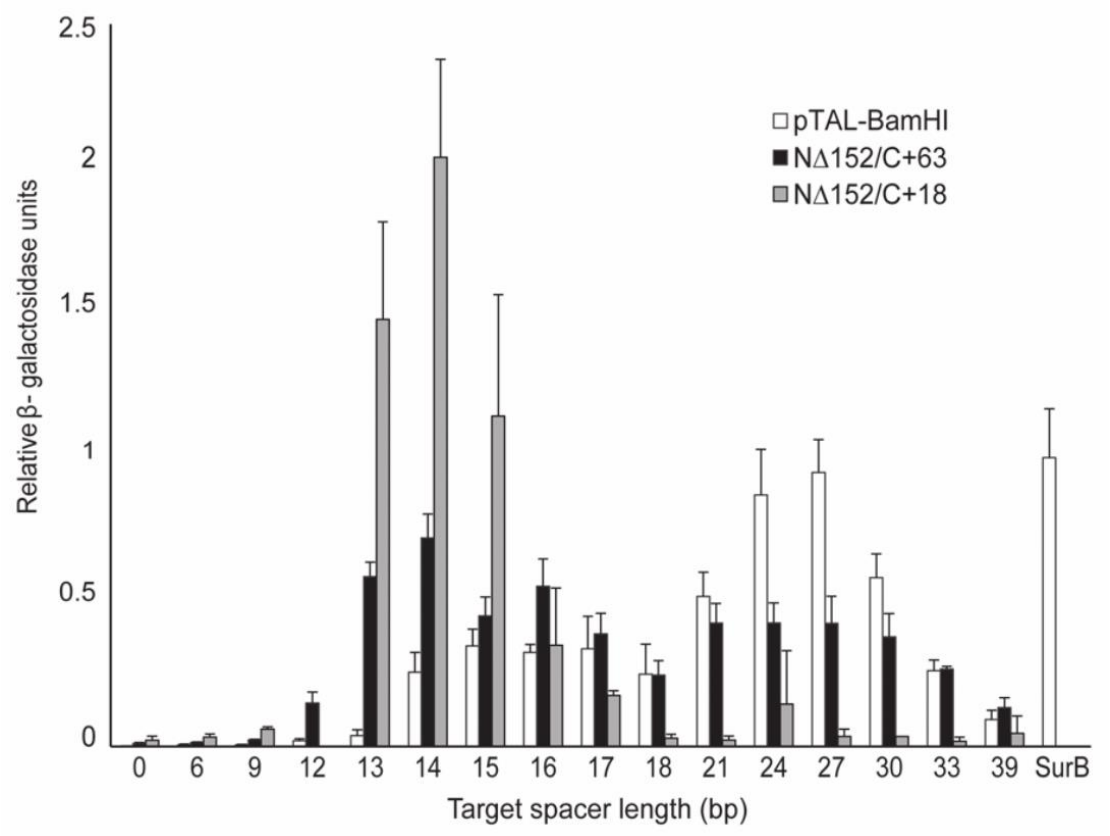


Figure 4.2. TALEN backbone architectures and spacer length optima. Adapted from (Christian et al., 2012).

4.2.3. FokI nucleases

The innate toxicity of ZFNs and TALENs comes from the off-targeting effect. The off-targeting effect comes from the homodimer formation of FokI nucleases in high concentration context (Alwin et al., 2005; Porteus and Baltimore, 2003). In detail, high concentration of ZFNs or TALENs means high concentration of FokI nuclease monomers. These monomers tend to dimerize at genomic locus with only one of them binding to the DNA when they are too concentrated, leading to off-target cleavages across genome. Some of these off-target cleavages induce genome instability, which may finally lead to toxicity to the cells. Therefore,

Heterodimer FokI nucleases were computationally designed and tested in cells (Szczepek et al., 2007). Asymmetric FokI heterodimers (DD/RR) and FokI homodimers with lower binding potential that only cleave after both of the two FokI nucleases bind to the DNA (IV/AI) were created and proven to be effective, reducing the toxicity by 3 fold (Fig. 4.3). By combining the asymmetric design and the lower binding potential design, RV/DA and KV/EA heterodimers were created and proved to be not only less toxic but also highly efficient (Mussolino et al., 2011; Szczepek et al., 2007).

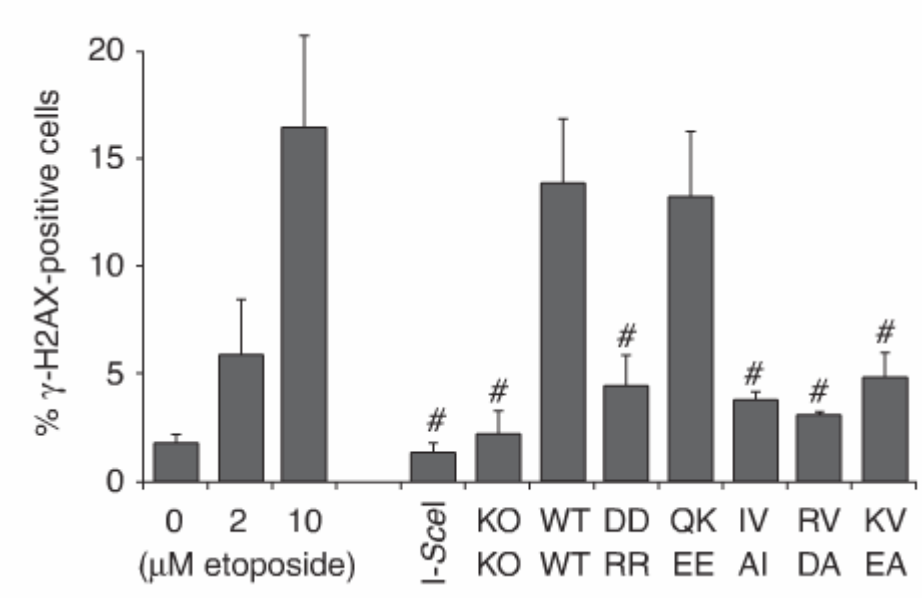


Figure 4.3. Percentage of highly g-H2AX-positive cells. The percentage represents the toxicity. Adapted from (Szczepek et al., 2007).

In summary, if all these parameters are taken into consideration, TALEN-mediated GE can be much more efficient with a very low to not detectable toxicity.

4.3. Dnd deficient medaka embryos and adults

In this thesis, medaka *dnd* deficient phenotypes were analyzed. Generally, loss of Dnd leads to retarded migration or direct death for PGCs, which is in accordance to the phenotype of zebrafish *dnd* knockdown (Weidinger et al., 2003). Dnds in both fish models are cell-autonomously essential for the PGC migration and survival. Being a germline-specific gene was part of the reason for choosing *dnd* as the TALENs' target. The disruption of this gene is neither likely to affect the somatic development nor to induce embryonic lethality. The TALENs' toxicity thus can be analyzed thoroughly. Indeed, the genotypes of all screened F_2 animals revealed that *dnd*$^{+/-}$ fish produced offspring in a Mendelian segregation manner (Table 4.1 and table 4.2).

Table 4.1. Mendelian-segregation inheritance of dnd alleles in F_2 embryos

	dnd$^{+/-}$ x dnd$^{+/-}$ embryos st 21		
	dnd$^{-/-}$	dnd$^{+/-}$	dnd$^{+/+}$
dgd2	13	24	12
%	26.5	49	24.5

Table 4.2. Mendelian-segregation inheritance of dnd alleles in F_2 fish

	$dnd^{+/-}$ x $dnd^{+/-}$ adults		
	$dnd^{-/-}$	$dnd^{+/-}$	$dnd^{+/+}$
dgd1 (%)	4 (12)	16 (49)	13 (40)
dgd2 (%)	13 (32)	19 (46)	9 (22)
dgd2 gene mapper (%)	15 (23)	39 (61)	10 (17)
Total	32 (23)	74 (54)	32 (23)

Germline transmitters in this thesis were highly mosaic in terms of cells that carry a disrupted *dnd* allele, which can be indicated by the germline transmission rate of 7.7 to 13%. Since double-allelic disruption of *dnd* is expected to cause PGC death in F_0 fish, this germline transmission rate represents only the percentage of cells with single-allelic disruption of *dnd* in the PGCs of the founders. Thus, F_0 founder fish are mosaic for cells with or without a single GE allele. Since all the germ cells containing a single GE allele are expectedly heterozygous before the completion of meiosis, they should be capable of normal development during embryonic and post-hatching stages. However, such a disrupted allele will stand alone in post-meiosis cells. The fact that female, and more importantly, male founders are able to transmit a mutant *dnd* allele to F_1 generation strongly suggests that genomic *dnd* is dispensable for germ cell development during the post-meiotic haploid phase.

Interestingly, in observing the disappearance of PGCs in $dnd^{-/-}$ embryos, *vasa*-WISH data showed the disappearance happens between 3 to 4 dpf, however,

the GFP labeling data showed the disappearance happens between 4 to 5 dpf. In the earlier stages (stage21 and 2-dpf), a bigger portion of the *vasa*-WISH data showed embryos with only few PGCs, which is nearly considered as disappearance. For GFP labeled embryos, only a very small portion showed this phenomenon. A possible explanation to this phenomenon is that the timing of loss of GFP signal lag behind the timing of losing PGC lineage trait. Even a cell dies, the GFP takes time to degrade. However, for *vasa* mRNA, it is expected to be degraded soon after the cell die or lose PGC lineage trait.

4.3.1. PGC number correlates with Dnd dosage

Dnd and its mRNA are maternally deposited into the embryos (Liu et al., 2009), so the $dnd^{-/-}$ embryos produced by heterozygous female are expected to retain Dnd until the maternally deposition runs out. According to the results section, dnd deficient embryos retained PGCs until 5-dpf, indicating that the decreasing of Dnd dosage may have caused the death of PGCs. On the other hand, since the female parent of the $dnd^{-/-}$ embryos is heterozygous for a functional *dnd* allele, it is expected that only about half of the normal amount of Dnd can be deposited into the oocytes before meiosis. The difference between WT female-produced embryos and $dnd^{+/-}$ female-produced embryos in PGC numbers at stage 21 indicates a correlation between the number of PGCs and the dosage of maternally deposited Dnd. Moreover, introduction of external Dnd at one-cell stage by microinjection not only rescue the PGC number decrease but also greatly enhance the number of PGCs, supporting the

Dnd dosage-PGC number relationship in the other way. However, further experiments are needed to prove the coincidence to be a cause and effect relationship: Examine the maternal Dnd dosage difference and the Dnd dosage in the PGCs during embryonic development.

4.3.2. Secondary sex characteristics of adult *dnd* deficient fish

According to the observation of adult fish, their secondary sex characteristics were not different from their wild type counterparts. Since the phenotype of *dnd* deficiency should have been caused by loss of PGCs, the results should be compared with the other study that investigated germ cell-loss phenotype, which was conducted by Hiromi et al. (Ohta et al., 2007). In that study, they used morpholino targeting CXCR4, which is essential for medaka PGC migration, to prevent the PGCs from reaching the gonadal primordium, generating germ cell-free fish at an efficiency of 32% of injected embryos. However, according to their observation, all the germ cell-free fish, including the X/X ones, exhibited male secondary sex characteristics. Moreover, PGC-ablated zebrafish also exhibited an all-male phenotype (Slanchev et al., 2005). It is unclear why the germ cell-free *dnd* deficient fish in this thesis retain secondary sex characteristics in accordance to their sex genotype. There is one important difference between this study and the two referred: this study uses knockout system to ablate PGCs, which leads to 100% of PGC ablation in $dnd^{-/-}$ fish; the other two use morpholino-mediated knock down system to ablate PGCs, which leads to an incomplete suppression of the target gene and 32%

of PGC ablation in embryos. The medaka gonads in germ cell-free fish observed by Hiromi ei al. were thicker and longer than the ones observed in this study. It is possible that knockdown approach was not able to completely remove the PGCs, which would leave a very small number of germ cells. According to their observation, some X/X embryos originally categorized as germ cell-free developed into fertile male fish. They explained this by stating that maybe small amount of PGCs escaped observation and led to the male development in an X/X context. This explanation would support my hypothesis, that a very small number of germ cells also escaped their observation and led to male development in X/X context, however, the amount of germ cells are even lower and thus the fish was not fertile.

Figure 4.4. Gonads of wild type and CXCR4 knockdown fish. Adapted from (Ohta et al., 2007)

CHAPTER 5. CONCLUSION

In this thesis, TALENs have been found proficient for direct genome editing in medaka embryos by one-cell stage microinjection, providing an easy one-step method for generation of medaka knockout fish for genetic analyzes. Results of this study showed TALENs generated a genome editing efficiency of 23% in F_1 adults and a germline transmission rate of 7.7 to 13%. Various aspects of the TALENs themselves are still able to be improved by new approaches in the development of TALEN backbones with higher efficiency and lower toxicity, which have been discussed in chapter 4. If the improvements are implemented, the TALENs are expected to be an even more convenient tool for genome editing, because lower toxicity means larger amount of TALENs can be applied, which means even higher efficiency.

On the other hand, this thesis demonstrates the phenotypes of *dnd* deficient fish in a knockout context. This thesis provided evidences that indicate that PGC number correlates with cellular Dnd dosage, since the PGC number decreases with the running out of maternal Dnd in *dnd*$^{-/-}$ embryos and the PGC number increases in the Dnd injected embryos. Additionally, the PGC number also correlates with the quantity of maternally deposited *dnd* mRNAs. This thesis also showed that the adult *dnd*-deficient fish exhibit a loss of germ cell phenotype. However, this phenotype is not the same as the phenotype observed in morpholino oligo-mediated germ cell deficient fish. The secondary sex characteristic is not reversed in X/X *dnd*-deficient fish and the *dnd*-deficient gonads in this thesis are thinner and shorter than those

observed in the morpholino oligo-mediated germ cell deficient gonads. Further studies are required to investigate the true reason of the difference in the germ cell-loss phenotypes by clearer characterizing the gene expression pattern in the *dnd*-deficient fish.

REFERENCES

Alwin, S., Gere, M.B., Guhl, E., Effertz, K., and Barbas, C.F. (2005). Custom zinc-finger nucleases for use in human cells. Molecular ….

Ansai, S., Sakuma, T., Yamamoto, T., and Ariga, H. (2013). Efficient targeted mutagenesis in medaka using custom-designed transcription activator-like effector nucleases. ….

Barrangou, R., Fremaux, C., Deveau, H., and Richards, M. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science.

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

Bibikova, M., Beumer, K., Trautman, J.K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. Science *300*, 764.

Boch, J., and Bonas, U. (2010). Xanthomonas AvrBs3 family-type III effectors: discovery and function. Annual review of phytopathology *48*, 419-436.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., and Hahn, S. (2009a). Breaking the code of DNA binding specificity of TAL-type III effectors. Science.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009b). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Brouns, S.J.J., Jore, M.M., Lundgren, M., and Westra, E.R. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science.

Cade, L., Reyon, D., Hwang, W.Y., and Tsai, S.Q. (2012). Highly efficient generation of heritable

zebrafish gene mutations using homo-and heterodimeric TALENs. Nucleic acids ….

Carroll, D., Morton, J.J., Beumer, K.J., and Segal, D.J. (2006). Design, construction and in vitro testing of zinc finger nucleases. Nature protocols *1*, 1329-1341.

Cermak, T., Doyle, E.L., Christian, M., and Wang, L. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic acids ….

Cermak, T., Doyle, E.L., Schmidt, C., and Zhang, F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Targeting DNA double-strand breaks with TAL effector nucleases.

Chen, J., Zhang, X., Wang, T., Li, Z., Guan, G., and Hong, Y. (2012). Efficient detection, quantification and enrichment of subtle allelic alterations. DNA research : an international journal for rapid publication of reports on genes and genomes *19*, 423-433.

Chisaka, O., Musci, T.S., and Capecchi, M.R. (1992). Developmental defects of the ear, cranial nerves and hindbrain resulting from targeted disruption of the mouse homeobox gene Hox-1.6. Nature *355*, 516-520.

Choulika, A., Perrin, A., Dujon, B., and Nicolas, J.F. (1995). Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. Molecular and cellular biology *15*, 1968-1973.

Christian, M.L., Demorest, Z.L., Starker, C.G., and Osborn, M.J. (2012). Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. PLoS One.

Cohen-Tannoudji, M., Robine, S., Choulika, A., Pinto, D., El Marjou, F., Babinet, C., Louvard, D., and Jaisser, F. (1998). I-SceI-induced gene replacement at a natural locus in embryonic stem cells. Molecular and cellular biology *18*, 1444-1448.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (New York, NY) *339*, 819-823.

Deng, C., and Capecchi, M.R. (1992). Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. Molecular and cellular biology *12*, 3365-3371.

Donoho, G., Jasin, M., and Berg, P. (1998). Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. Molecular and cellular biology *18*, 4070-4078.

Extavour, C.G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. Development (Cambridge, England) *130*, 5869-5884.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., and Reyon, D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature ….

Gaj, T., Gersbach, C.A., and Iii, B.C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in biotechnology.

Gloor, G.B., Nassif, N.A., Johnson-Schlitz, D.M., Preston, C.R., and Engels, W.R. (1991). Targeted gene replacement in Drosophila via P element-induced gap repair. Science *253*, 1110-1117.

Goldenberger, D., Perschil, I., and Ritzler, M. (1995). A simple" universal" DNA extraction procedure using SDS and proteinase K is compatible with direct PCR amplification. Genome ….

Hanahan, D., Jessee, J., and Bloom, F.R. (1991). Plasmid transformation of< i> Escherichia coli</i> and other bacteria. Methods in enzymology.

Herbers, K., Conrads-Strauch, J., and Bonas, U. (1992). Race-specificity of plant resistance to

bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. Nature *356*, 172-174.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science.

Hsu, P.D., Scott, D.A., Weinstein, J.A., and Ran, F.A. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature ….

Inoue, H., Nojima, H., and Okayama, H. (1990). High efficiency transformation of< i> Escherichia coli</i> with plasmids. Gene.

Isalan, M., Choo, Y., and Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. Proceedings of the National Academy of Sciences of the United States of America *94*, 5617-5621.

Isalan, M., Klug, A., and Choo, Y. (1998). Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. Biochemistry *37*, 12026-12033.

Iwamatsu, T. (2004). Stages of normal development in the medaka< i> Oryzias latipes</i>. Mechanisms of development.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., and Doudna, J.A. (2012a). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., and Doudna, J.A. (2012b). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science.

Jing, L., Burns,Caroline (2012). Preparation of Torula Yeast RNA for Hyb Solutions Bio-protocol *2*, e180.

Kedde, M., Strasser, M.J., Boldajipour, B., Oude Vrielink, J.A., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P.M., van Duijse, J., Ørom, U.A., *et al.* (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell *131*, 1273-1286.

Ketting, R.F. (2007). A dead end for microRNAs. Cell *131*, 1226-1227.

Kinoshita, M., Murata, K., Naruse, K., and Tanaka, M. (2009). Medaka: biology, management, and experimental protocols. Medaka: biology, management, and experimental protocols.

Kurokawa, H., Aoki, Y., Nakamura, S., Ebe, Y., Kobayashi, D., and Tanaka, M. (2006). Time-lapse analysis reveals different modes of primordial germ cell migration in the medaka Oryzias latipes. Development, growth & differentiation *48*, 209-221.

Laity, J.H., Lee, B.M., and Wright, P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. Current opinion in structural biology *11*, 39-46.

Liu, L., Hong, N., Xu, H., Li, M., Yan, Y., Purwanti, Y., Yi, M., Li, Z., Wang, L., and Hong, Y. (2009). Medaka dead end encodes a cytoplasmic protein and identifies embryonic and adult germ cells. Gene expression patterns : GEP *9*, 541-548.

Mani, M., Kandavelou, K., Dy, F.J., Durai, S., and Chandrasegaran, S. (2005). Design, engineering, and characterization of zinc finger nucleases. Biochemical and biophysical research communications *335*, 447-457.

Mansour, S.L., Thomas, K.R., and Capecchi, M.R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature *336*, 348-352.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., and Wang, J. (2011). A TALE nuclease architecture for efficient genome editing. Nature ….

Moore, J.K., and Haber, J.E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Molecular and cellular biology.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science.

Mussolino, C., Morbitzer, R., and Lütge, F. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic acids ….

Nakade, S., Tsubota, T., Sakane, Y., and Kume, S. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. Nature ….

Nakamura, S., Kobayashi, D., Aoki, Y., Yokoi, H., Ebe, Y., Wittbrodt, J., and Tanaka, M. (2006). Identification and lineage tracing of two populations of somatic gonadal precursors in medaka embryos. Developmental biology *295*, 678-688.

Ohta, K., Baba, T., and Morohashi, K. (2007). Germ cells are essential for sexual dimorphism in the medaka gonad. … National Academy ….

Orlando, S.J., Santiago, Y., and DeKelver, R.C. (2010). Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. Nucleic acids ….

Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science *252*, 809-817.

Porteus, M.H., and Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. Science *300*, 763.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D., and Joung, K.J. (2012). FLASH assembly of TALENs for high-throughput genome editing. Nature biotechnology *30*, 460-465.

Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y., *et al.* (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nature methods *8*, 67-69.

Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell.

Slanchev, K., Stebler, J., de la Cueva-Méndez, G., and Raz, E. (2005). Development without germ cells: the role of the germ line in zebrafish sex differentiation. Proceedings of the National Academy of Sciences of the United States of America *102*, 4074-4079.

Smih, F., Rouet, P., Romanienko, P.J., and Jasin, M. (1995). Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. Nucleic acids research *23*, 5012-5019.

Streubel, J., Blücher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. Nature biotechnology.

Sugio, A., Yang, B., Zhu, T., and White, F.F. (2007). Two type III effector genes of Xanthomonas oryzae pv. oryzae control the induction of the host genes OsTFIIAgamma1 and OsTFX1 during bacterial blight of rice. Proceedings of the National Academy of Sciences of the United States of America *104*, 10720-10725.

Szczepek, M., Brondani, V., Büchel, J., and Serrano, L. (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nature ….

Szurek, B., Rossier, O., Hause, G., and Bonas, U. (2002a). Type III-dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell. Molecular microbiology *46*, 13-23.

Szurek, B., Rossier, O., Hause, G., and Bonas, U. (2002b). Type III dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell. Molecular microbiology *46*, 13-23.

Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.

Thomas, K.R., Folger, K.R., and Capecchi, M.R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. Cell *44*, 419-428.

Thompson, S., Clarke, A.R., Pow, A.M., and Hooper, M.L. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell.

Tong, C., Li, P., Wu, N.L., Yan, Y., and Ying, Q.L. (2010). Production of p53 gene knockout rats by homologous recombination in embryonic stem cells. Nature *467*, 211-213.

Wang, Y., Li, Z., Xu, J., Zeng, B., Ling, L., You, L., and Chen, Y. (2013). The CRISPR/Cas System mediates efficient genome engineering in Bombyx mori. Cell research.

Weidinger, G., Stebler, J., Slanchev, K., and Dumstrei, K. (2003). < i> dead end</i>, a Novel Vertebrate Germ Plasm Component, Is Required for Zebrafish Primordial Germ Cell Migration and Survival. Current biology.

Xiao, A., Wang, Z., Hu, Y., Wu, Y., Luo, Z., and Yang, Z. (2013). Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. Nucleic acids ….

Yang, B., Sugio, A., and White, F.F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. Proceedings of the National Academy of Sciences of the United States of America *103*, 10503-10508.

Yang, H., Wang, H., and Jaenisch, R. (2014). Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. Nature protocols *9*, 1956-1968.

Yang, Y., and Gabriel, D.W. (1995). Xanthomonas avirulence/pathogenicity gene family encodes functional plant nuclear targeting signals. Molecular plant-microbe interactions : MPMI *8*, 627-631.

Youngren, K.K., Coveney, D., Peng, X., Bhattacharya, C., Schmidt, L.S., Nickerson, M.L., Lamb, B.T., Deng, J.M., Behringer, R.R., Capel, B., *et al.* (2005). The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. Nature *435*, 360-364.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., and Church, G.M. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nature ….

Zhang, X., Guan, G., Chen, J., Naruse, K., and Hong, Y. (2014). Parameters and efficiency of direct gene disruption by Zinc Finger Nucleases in Medaka Embryos. Marine Biotechnology.

Zhu, W., Yang, B., Chittoor, J.M., Johnson, L.B., and White, F.F. (1998). AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. Molecular plant-microbe interactions : MPMI *11*, 824-832.

Appendice I. Reagent list

Acylamide solution, 30%, A:B=29:1, Biorad, CAT#: 161-0156

Agar, Sigma.

Agarose, 1st base, molecular biology grade

Ampicillin, Sigma

Big-Dye Terminator cycle sequencing kit, Applied Biosystems, CAT#: 4336917

BSA, Sigma, A9418-100G

$CaCl_2 \bullet 2H_2O$, Merck

Chloroform, Merck, CAT#: 1.02445.2500

Diethylpyrocarbonate (DEPC)

Dimethyl sulfoxide, Sigma, CAT#: D2650

EDTA, Sigma

Ethanol, Merck, CAT#: 1.00983.2500

Ex Taq and HS Ex Taq polymerase, TaKaRa, CAT#: DRR100E

Formamide, Molecular biology grade, Calbiochem, CAT#: 344206

GeneRuler 1-kb DNA ladder, Fermentas, CAT#: SM0311

Genescan400HD, ABI

Glacial acetic acid, Merck, CAT#: 100056

Goat serum, Sigma, CAT#: G9023

Glycerol, 99.5%, Sigma

Heparin, Sigma

Hi-Di-formamide, Life technologies, CAT#: 4311320

Hydrogen peroxide, 35% by weight, Riedel-de Haen, CAT#: 18304

Isopropanol Sinopharm Chemical, CAT#: 19130020

Kanamycin, Sigma

KCl, Sigma, CAT#: P5405

KH2PO4, Merck, CAT#: 529568

LB powder mix, 1st Base

Liquid nitrogen, Cryoexpress Singapore

Methanol, QReC, CAT#: M2097-4-2501

$MgCl_2$, Sigma

Mineral oil, Sigma, CAT#:M8410-500ML

MMLV reverse transcriptase, Thermo, CAT#: EP0352

MN gel recovery and PCR purification kit NucleoSpin Extract II, Macherey-Nagel, CAT#: 740609.250

MN plasmid midi kit Nucleobond Xtra midi plus, Macherey-Nagel, CAT#: 740412.2

Na_2HPO_4, Merck, CAT#: 106566

NaCl, Sigma

$NaHCO_3$ Sigma, CAT#: S5761

NaOH, Sigma

NBT/BCIP substrate solution, Roche, CAT#: 11681451001

Paraformaldehyde PFA, Sigma, CAT#: P6148

PCR primers are synthesized by Sigma, 1st Base or IDT.

pGEM-T-easy vector system, Promega, CAT#: A1360

Phenol red, Sigma, CAT#: P3532

Phenol solution, Sigma, CAT#: P4557

Phenol:Chloroform:IAA 25:24:1, Ambion, CAT#: AM9732

Proteinase K, Invitrogen, CAT#: 25530-015

Restriction enzymes are purchased from Promega, New England BioLabs and Fermentas

RNase A, Sigma, CAT#: R6513

RO and Milli-Q water: From Millipore pure water system, Milli-Q water is autoclaved for molecular work

10% SDS solution, Sigma, CAT#: L4522

Secondary Antibody anti-DIG-AP fab fragments, Roche, CAT#: 11207733910

T4 DNA ligase, NEB, CAT#: M0202, Invitrogen, CAT#: 15224-025, Promega, CAT#: M180A

Tris-HCl, Sigma

Trisodium citrate Sigma, CAT#: S4641

Trizol reagent Invitrogen, CAT#: 15596-026

Tween 20, Sigma, CAT#: P2287

tRNA transfer ribonucleic acid, from wheat germ, Sigma, CAT#: R7876

X-gal and IPTG: 20 mg/mL of X-gal is prepared in dimethyl formamide

Appendice II. Coding sequences of pTN1dnd and pTN2dnd

Coding sequence of pTN1dnd:

1 ATGGCTTCCTCCCCTCCAAAGAAAAAGAGAAAGGTTAGTTGGAAGGACGCAAGTGGTTGGTCTAGAATGCATGCGGATCCCATTCGTCCGCGCAGGCCAA

101 GTCCTGCCCGCGAGCTTCTGCCCGGACCCCAACCGGATAGGGTTCAGCCGACTGCAGATCGTGGGGTGTCTGCGCCTGCTGGCAGCCCTCTGGATGGCTT

201 GCCCGCTCGGCGGACGGTGTCCCGGACCCGGCTGCCATCTCCCCCTGCGCCCTCACCTGCGTTCTCGGCGGGCAGCTTCAGCGATCTGCTCCGTCCGTTC

301 GATCCGTCGCTTCTTGATACATCGCTTCTTGATTCGATGCCTGCCGTCGGCACGCCGCATACAGCGGCTGCCCCAGCAGAGTGGGATGAGGCGCAATCGG

401 CTCTGCGTGCAGCCGATGACCCGCCACCCACCGTGCGTGTCGCTGTCACTGCCGCGCGGCCGCCGCGCGCCAAGCCGGCCCCGCGACGGCGTGCTGCGCA

501 ACCCTCCGACGCTTCGCCGGCCGCGCAGGTGGATCTACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGATCAAACCGAAGGTGCGTTCGACAGTG

601 GCGCAGCACCACGAGGCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGCGCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCACGT

701 ATCAGCACATAATCACGGCGTTGCCAGAGGCGACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGGCGCACGCGCCCTGGAGGCCTTGCTCAC

801 GGATGCGGGGGAGTTGAGAGGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGATTGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT

901 GCATCGCGCAATGCACTGACGGGTGCCCCCCTGAACCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGG

1001 TGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAAC

1101 GGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAA

1201 ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCG

1301 AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCT

1401 CGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCG

1501 CTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAG

1601 CGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCA

1701 AGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAG

1801 CAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCA

1901 AGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGG

2001 CAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGC

2101 GGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATG

2201 GCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAT

2301 TGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCAC

2401 GATGGCGGCAAGCAAGCGCTCGAAACGGTGCATGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCC

2501 AGCCACGATGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAGCCGGCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCACCTCGTCGCCT

2601 TGGCCTGCCTCGGCGGACGTCCTGCCATGGATGCAGTGAAAAAGGGATTGCCGCACGCGCCGGAATTGATCAGAAGAGTCAATCGCCGTATTGGCGAACG

2701 CACGTCCCATCGCGTTGCCGACTACGCGCAAGTGGTTCGCGTGCTGGAGTTTTTCCAGTGCCACTCCCACCCAGCGTACGCATTTGATGAGGCCATGACG

2801 CAGTTCGGGATGAGCAGGAACGGGTTGGTACAGCTCTTTCGCAGAGTGGGCGTCACCGAACTCGAAGCCCGCGGTGGAACGCTCCCCCCAGCCTCGCAGC

2901 GTTGGGACCGTATCCTCCAGGCATCAGGGATGAAAAGGGCCAAACCGTCCCCTACTTCAGCTCAAACACCGGATCAGGCGTCTTTGCATGCATTCGCCGA

3001 TTCGCTGGAGCGTGACCTTGATGCGCCCAGCCCAATGCACGAGGGAGATCAGACGCGGGCAAGCAGCCGTAAACGGTCCCGATCGGATCGTGCTGTCACC

3101 GGCCCCTCCGCACAGCAGGCTGTCGAGGTGCGCGTTCCCGAACAGCGCGATGCGCTGCATTTGCCCCTCAGCTGGAGGGTAAAACGCCCGCGTACCAGGA

3201 TCTGGGGCGGCCTCCCGGATCCGATATCTAGATCCCAGCTAGTGAAATCTGAATTGGAAGAGAAGAAATCTGAACTTAGACATAAATTGAAATATGTGCC

3301 ACATGAATATATTGAATTGATTGAAATCGCAAGAAATTCAACTCAGGATAGAATCCTTGAAATGAAGGTGATGGAGTTCTTTATGAAGGTTTATGGTTAT

3401 CGTGGTAAACATTTGGGTGGATCAAGGAAACCAGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTGATCGTTGATACTAAGGCATATT

3501 CAGGAGGTTATAATCTTCCAATTGGTCAAGCAGATGAAATGCAAAGATATGTCGAAGAGAATCAAACAAGAAACAAGCATATCAACCCTAATGAATGGTG

3601 GAAAGTCTATCCATCTTCAGTAACAGAATTTAAGTTCTTGTTTGTGAGTGGTCATTTCAAAGGAAACTACAAAGCTCAGCTTACAAGATTGAATCATATC

3701 ACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTGATTGGTGGAGAAATGATTAAAGCTGGTACATTGACACTTGAGGAAGTGAGAAGGAAAT

3801 TTAATAACGGTGAGATAAACTTTTAA

Coding sequence of pTN2dnd:

```
1    ATGGCTTCCTCCCCTCCAAAGAAAAAGAGAAAGGTTAGTTGGAAGGACGCAAGTGGTTGGTCTAGAATGCATGCGGATCCCATTCGTCCGCGCAGGCCAA
101  GTCCTGCCCGCGAGCTTCTGCCCGGACCCCAACCGGATAGGGTTCAGCCGACTGCAGATCGTGGGGTGTCTGCGCCTGCTGGCAGCCCTCTGGATGGCTT
201  GCCCGCTCGGCGGACGGTGTCCCGGACCCGGCTGCCATCTCCCCCTGCGCCCTCACCTGCGTTCTCGGCGGGCAGCTTCAGCGATCTGCTCCGTCCGTTC
301  GATCCGTCGCTTCTTGATACATCGCTTCTTGATTCGATGCCTGCCGTCGGCACGCCGCATACAGCGGCTGCCCCAGCAGAGTGGGATGAGGCGCAATCGG
401  CTCTGCGTGCAGCCGATGACCCGCCACCCACCGTGCGTGTCGCTGTCACTGCCGCGCGGCCGCCGCGCGCCAAGCCGGCCCCGCGACGGCGTGCTGCGCA
501  ACCCTCCGACGCTTCGCCGGCCGCGCAGGTGGATCTACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGATCAAACCGAAGGTGCGTTCGACAGTG
601  GCGCAGCACCACGAGGCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGCGCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCACGT
701  ATCAGCACATAATCACGGCGTTGCCAGAGGCGACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGGCGCACGCGCCCTGGAGGCCTTGCTCAC
801  GGATGCGGGGGAGTTGAGAGGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGATTGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT
901  GCATCGCGCAATGCACTGACGGGTGCCCCCCTGAACCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGG
1001 TGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAAC
1101 GGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAA
1201 ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCG
1301 AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCT
1401 CGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCG
1501 CTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAG
1601 CGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCA
1701 AGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAG
1801 CAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCA
1901 AGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGG
2001 CAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGC
2101 GGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATG
2201 GCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAT
2301 TGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCAC
2401 GATGGCGGCAAGCAAGCGCTCGAAACGGTGCATGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCC
2501 AGCCACGATGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAGCCGGCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCACCTCGTCGCCT
2601 TGGCCTGCCTCGGCGGACGTCCTGCCATGGATGCAGTGAAAAAGGGATTGCCGCACGCGCCGGAATTGATCAGAAGAGTCAATCGCCGTATTGGCGAACG
2701 CACGTCCCATCGCGTTGCCGACTACGCGCAAGTGGTTCGCGTGCTGGAGTTTTTCCAGTGCCACTCCCACCCAGCGTACGCATTTGATGAGGCCATGACG
2801 CAGTTCGGGATGAGCAGGAACGGGTTGGTACAGCTCTTTCGCAGAGTGGGCGTCACCGAACTCGAAGCCCGCGGTGGAACGCTCCCCCCAGCCTCGCAGC
2901 GTTGGGACCGTATCCTCCAGGCATCAGGGATGAAAAGGGCCAAACCGTCCCCTACTTCAGCTCAAACACCGGATCAGGCGTCTTTGCATGCATTCGCCGA
3001 TTCGCTGGAGCGTGACCTTGATGCGCCCAGCCCAATGCACGAGGGAGATCAGACGCGGGCAAGCAGCCGTAAACGGTCCCGATCGGATCGTGCTGTCACC
3101 GGCCCCTCCGCACAGCAGGCTGTCGAGGTGCGCGTTCCCGAACAGCGCGATGCGCTGCATTTGCCCCTCAGCTGGAGGGTAAAACGCCCGCGTACCAGGA
3201 TCTGGGGCGGCCTCCCGGATCCGATATCTAGATCCCAGCTAGTGAAATCTGAATTGGAAGAGAAGAAATCTGAACTTAGACATAAATTGAAATATGTGCC
3301 ACATGAATATATTGAATTGATTGAAATCGCAAGAAATTCAACTCAGGATAGAATCCTTGAAATGAAGGTGATGGAGTTCTTTATGAAGGTTTATGGTTAT
3401 CGTGGTAAACATTTGGGTGGATCAAGGAAACCAGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACGGTGTGATCGTTGATACTAAGGCATATT
3501 CAGGAGGTTATAATCTTCCAATTGGTCAAGCAGATGAAATGCAAAGATATGTCGAAGAGAATCAAACAAGAAACAAGCATATCAACCCTAATGAATGGTG
3601 GAAAGTCTATCCATCTTCAGTAACAGAATTTAAGTTCTTGTTTGTGAGTGGTCATTTCAAAGGAAACTACAAAGCTCAGCTTACAAGATTGAATCATATC
3701 ACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTGATTGGTGGAGAAATGATTAAAGCTGGTACATTGACACTTGAGGAAGTGAGAAGGAAAT
3801 TTAATAACGGTGAGATAAACTTTTAA
```

Protein sequence of pTN1dnd encoded TALEN

```
1    MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPSPARELLPGPQPDRVQPTADRGVSAPAGSPLDGLPAR
71   RTVSRTRLPSPPAPSPAFSAGSFSDLLRPFDPSLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALRAADD
141  PPPTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGH
211  GFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSGARALEALLTDAGELRGPPLQLD
281  TGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPD
351  QVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQV
421  VAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA
491  IASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIA
561  SNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASH
631  DGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDG
701  GKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK
771  QALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA
841  LESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVADYAQV
911  VRVLEFFQCHSHPAYAFDEAMTQFGMSRNGLVQLFRRVGVTELEARGGTLPPASQRWDRILQASGMKRAK
981  PSPTSAQTPDQASLHAFADSLERDLDAPSPMHEGDQTRASSRKRSRSDRAVTGPSAQQAVEVRVPEQRDA
1051 LHLPLSWRVKRPRTRIWGGLPDPISRSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEM
1121 KVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
1191 KHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLE
1261 EVRRKFNNGEINF
```

Protein sequence of pTN1dnd encoded TALEN

```
1    MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPSPARELLPGPQPDRVQPTADRGVSAPAGSPLDGLPAR
71   RTVSRTRLPSPPAPSPAFSAGSFSDLLRPFDPSLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALRAADD
141  PPPTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGH
211  GFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHEDIVGVGKQWSGARALEALLTDAGELRGPPLQLD
281  TGQLVKIAKRGGVTAMEAVHASRNALTGAPLNLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPD
351  QVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQV
421  VAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA
491  IASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIA
561  SNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASN
631  IGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIG
701  GKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK
771  QALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA
841  LESIVAQLSRPDPALAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNRRIGERTSHRVADYAQV
911  VRVLEFFQCHSHPAYAFDEAMTQFGMSRNGLVQLFRRVGVTELEARGGTLPPASQRWDRILQASGMKRAK
981  PSPTSAQTPDQASLHAFADSLERDLDAPSPMHEGDQTRASSRKRSRSDRAVTGPSAQQAVEVRVPEQRDA
1051 LHLPLSWRVKRPRTRIWGGLPDPISRSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEM
1121 KVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
1191 KHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLE
1261 EVRRKFNNGEINF
```