ALTERNATIVE TRANSCRIPTION AND SPLICING OF DUPLICATED mitf GENES IN THE FISH MEDAKA

ZHU FENG
(B.Sc., Nanjing 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.

Zhu Feng
ACKNOWLEDGEMENTS

Work presented in this thesis was accomplished at the Department of Biological Sciences, National University of Singapore from August 2011 to August 2015. I am very happy to have this opportunity to acknowledge everyone who made this work possible.

This thesis would not have been possible without the support and supervision of my PhD advisor, Prof. Hong Yunhan. I really appreciate his support and help to broaden my vision and push towards excellence.

I would also like to express my thanks to the thesis advisory committee members, Prof. Chan Woon Khiong and Prof. Yue Genhua for their very useful advice and comments.

I would like to thank all my labmates, Dr. Li Mingyou, Dr. Guan Guijun, Dr. Yuan Yongming, Dr. Lin Fan, Dr. Wang Tiansu, Chen Jianbin, Zhang Xi and Liu Qizhi. We had a good atmosphere in the lab and we made a good team. This made my candidature more fun.

Furthermore, I would like to appreciate the President’s Graduate Fellowship for releasing me from any financial problems.

Finally, I would like to thank my wonderful family for their unconditional love and support without which I could not have pursued my dreams.

Singapore, 12th Aug 2015
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .............................................................................................. iv
SUMMARY ...................................................................................................................... viii
LIST OF TABLES ........................................................................................................... xi
LIST OF FIGURES ......................................................................................................... xii
LIST OF ABBREVIATIONS ............................................................................................ xiii

Chapter 1: Introduction ................................................................................................. 1
1.1 Alternative transcription and splicing .................................................................. 2
1.2 Mitf as an example of alternative transcription and splicing ............................. 6
1.3 Whole genome duplication (WGD) in fish and alternative splicing ............... 11
1.4 Medaka model to study alternative mitf transcription and splicing .............. 15
1.5 RNA-Seq for revealing alternative transcription and splicing ....................... 16
1.6 Significance and objective .................................................................................. 17

Chapter 2: Materials and Methods ........................................................................... 19
2.1 Fish and chemicals ............................................................................................... 20
  2.1.1 Fish ................................................................................................................ 20
  2.1.2 Chemicals ....................................................................................................... 20
2.2 Molecular cloning work ...................................................................................... 20
  2.2.1 Agarose gel electrophoresis ........................................................................ 20
  2.2.2 RNA extraction ............................................................................................ 20
  2.2.3 Purification of DNA from agarose gel or enzyme reaction solution ......... 21
  2.2.4 cDNA library synthesis ............................................................................... 21
  2.2.5 Rapid amplification of cDNA ends (RACE) ............................................... 22
  2.2.6 RT-PCR ......................................................................................................... 22
  2.2.7 Real-time RT-PCR ....................................................................................... 23
  2.2.8 droplet digital PCR (ddPCR) ....................................................................... 24
  2.2.9 Cloning .......................................................................................................... 24
  2.2.10 Automated sequencing analysis ................................................................. 25
  2.2.11 Plasmids ...................................................................................................... 25
2.3 in situ hybridization ............................................................................................. 27
2.4 Cell culture and transfection ......................................................28
2.5 Luciferase reporter assay ..........................................................29
2.6 Microinjection and western blot .................................................30
2.7 Microscopy ................................................................................31
2.8 Statistical analysis .....................................................................31
2.9 Transcriptome sequencing ..........................................................31
2.10 Bioinformatics analysis of RNA-Seq ...........................................32
  2.10.1 Quality assessment and filter of raw reads ..............................32
  2.10.2 Alignment/Mapping reads to genome ....................................32
  2.10.3 Quality assessment of alignments .........................................33
  2.10.4 Assembly of transcripts .......................................................33
  2.10.5 Generation of transcriptome assembly ....................................33
  2.10.6 Counting the transcripts .......................................................33
  2.10.7 Annotation of transcripts .....................................................34

Chapter 3: Results ............................................................................35
3.1 Differential evolution of duplicated fish mitf genes .......................36
  3.1.1 Medaka Mitf1 and Mitf2 correspond to mammalian MitfA and MitfM .................................................................36
  3.1.2 Medaka mitf1 and mitf2 differ in sequence divergence rate .......41
  3.1.3 Medaka mitf1 and mitf2 genes are the products of WGD ..........43
  3.1.4 Medaka mitf1 and mitf2 differ in the expression level ...............44
  3.1.5 Medaka mitf1 and mitf2 differ in embryonic melanocytic expression ..............................................................................46
  3.1.6 Medaka Mitf2 possesses a reduced melanogenic activity .........47
  3.1.7 Medaka Mitf2 possesses a reduced melanocyte-inducing activity ..................................................................................50
3.2 Complimentary approaches for investigation of AT and AS ..........50
  3.2.1 Rapid amplification of cDNA Ends (RACEs) ............................50
  3.2.2 RNA-Seq .............................................................................52
3.3 Alternative transcription of duplicated fish mitf genes .................58
  3.3.1 Medaka mitf1 produces multiple AT isoforms .........................58
  3.3.2 Expression of mitf1 and mitf2 AT variants .................................61
  3.3.3 Activities of Mitf isoforms on target genes’ promoters ................62
  3.3.4 Induction of melanocytes and tyr RNA expression in ES cells ..................................................................................66
  3.3.5 Activities of Mitf isoforms on the expression of cell lineage markers ..............................................................................67
  3.3.6 Expression and stability of Mitf protein isoforms .....................69
3.4 Alternative splicing of duplicated fish mitf genes .........................73

Chapter 4: Discussion ......................................................................77
4.1 Gene duplication and differential evolution ........................................78
4.2 Differential evolution of mitf genes .................................................81
4.3 AT and the role of mitf genes .......................................................85
  4.3.1 Conserved mechanism of AT of mitf .................................85
  4.3.2 Role of mitf due to differential expression or differential activity ...............................................................86
4.4 AS and the role of mitf genes .....................................................89
4.5 Conclusions .............................................................................91

References .................................................................................93

Appendices ................................................................................103
SUMMARY

Alternative transcription (AT) and alternative splicing (AS) represent a hallmark of eukaryotic evolution. This thesis work focused on AT and AS by using microphthalmia-associated transcription factor (mitf) as a model gene. Mitf is a basic helix–loop–helix leucine zipper transcription factor that is conserved in the animal kingdom. It acts as the master regulator of normal pigment cells and also of cancerous pigment cells. Mitf mutations cause Waardenburg syndrome type 2A in human, which has defects in skin and eye colour due to melanocyte reduction. Mouse Mitf mutations can cause defects in neural-crest-derived melanocytes, manifested by a lack of pigment in the coat and inner ear, the latter leading to congenital deafness. A few of the mutations also induce osteoclasts defects as well as mast cell defects. In mammals, a single mitf gene produces multiple RNA variants by differential usage of alternative promoters and alternative first exons. In human, 9 alternative promoters/first exons are utilized to generate various MITF isoforms with unique N-termini including MITF-A, MITF-B, MITF-H and MITF-M. In fish, whole genome duplication (WGD) also provides more transcripts or protein isoforms. In this work, medaka (Oryzias latipes) is used as a
lower vertebrate model. Medaka has two mitf genes, mitf1 and mitf2. This thesis has identified mitf1 and mitf2 differential evolution after WGD by examining their sequence divergence, differential expression and differential activity. To further investigate into AT and AS of mitf genes, RACE (Rapid Amplification of cDNA Ends) and RNA-Seq were employed. We have identified 5' and 3' transcript variants of mitf1 via AT and AS, respectively. The results of RT-PCR prove differential expression of these mitf AT variants in different organs and stages of development. Isoforms predicted from these mitf1 AT variants differed dramatically in activating target gene expression. Intriguingly, these mitf1 isoforms possess considerable differences in inducing the expression of multiple cell lineage marker genes. These results demonstrate that AT is a conserved mechanism controlling mitf expression in vertebrates ranging from human to fish, and that different mitf AT isoforms possesses different activities. Interestingly, mitf2 has a single promoter and thus lacks AT but is able to produce 2 AS variants. We have also proven the differential expression of AS variants in both mitf1 and mitf2. Thus, the duplicated medaka mitf genes show a considerable difference in the retention and loss of AT and AS during their evolution. This finding sheds new lights on differential evolution
after gene duplication through AT and AS for increasing genetic complexity.
LIST OF TABLES

Table 1. Percentage Mitf sequence identities between fishes and tetrapods ................................................................. 42
Table 2. Summary statistics of the RNA sequencing. ......................... 53
LIST OF FIGURES

Figure 1.1. Alternative transcription initiation ...........................................3
Figure 1.2. Alternative splicing events ......................................................4
Figure 1.3. Organization of human MITF gene ...........................................10
Figure 1.4. Whole Genome Duplication (WGD) in fish .............................12
Figure 1.5. Evolutionary fate of single gene duplications ............................13
Figure 3.1. Nucleotide and deduced amino acid sequence of medaka mitf1 and mitf2 ..........................................................................................37
Figure 3.2. Mitf protein sequence alignment .............................................38
Figure 3.3. Phylogenetic tree of mitf genes ...............................................40
Figure 3.4. Fish mitf1 and mitf2 are homologs of the human Mitf by chromosome synteny ..........................................................43
Figure 3.5. RNA expression of mitf .............................................................45
Figure 3.6. Medaka Mitf2 has a reduced activity ......................................49
Figure 3.7. RNA-Seq discovers AT and AS variants ................................54
Figure 3.8. Medaka mitf transcripts revealed by RNA-Seq ......................55
Figure 3.9. Relative expression of mitf genes in the RNA-Seq experiments ..........................................................................................57
Figure 3.10. Structure and RNA expression of medaka Mitf1 and Mitf2 ..........................................................................................59
Figure 3.11. Alternative first exons .............................................................60
Figure 3.12. ddPCR analysis of mitf1 AT variants ....................................62
Figure 3.13. Relative activities of isoforms of Mitf1 and Mitf2 .................64
Figure 3.14. Induction of melanocyte differentiation and tyr RNA expression by Mitf proteins ..............................................................65
Figure 3.15. Mitf isoforms have different activities on lineage gene expression ..........................................................................................68
Figure 3.16. Expression and stability of Cherry-tagged Mitf isoforms ..........................................................................................71
Figure 3.17. Expression and stability of mCherry-tagged Mitf isoforms ..........................................................................................72
Figure 3.18. Alternative splicing of mitf genes in medaka .......................74
Figure 3.19. Differential expression of mitf AS variants ...........................75
Figure 3.20. Validation of mitf AS variants .............................................75
Figure 4.1. Different models of duplicate gene evolution .........................79
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Amino acid(s)</td>
</tr>
<tr>
<td>AFE</td>
<td>alternative first exon</td>
</tr>
<tr>
<td>AS</td>
<td>alternative splicing</td>
</tr>
<tr>
<td>AT</td>
<td>alternative transcription</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide(s)</td>
</tr>
<tr>
<td>cDNA</td>
<td>copied DNA complementary to RNA</td>
</tr>
<tr>
<td>CDS</td>
<td>coding sequence</td>
</tr>
<tr>
<td>DAPI</td>
<td>4',6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DIG</td>
<td>digoxigenin</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>ddPCR</td>
<td>droplet digital PCR</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>FISH</td>
<td>fluorescence in situ hybridization</td>
</tr>
<tr>
<td>h</td>
<td>hour(s)</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>IF</td>
<td>immunofluorescence</td>
</tr>
<tr>
<td>ISH</td>
<td>in situ hybridization</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>LB</td>
<td>lysogeny broth</td>
</tr>
<tr>
<td>M, mM</td>
<td>molar, millimolar</td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
</tr>
<tr>
<td>mitf, Mitf</td>
<td>microphthalmia-associated transcription factor</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>NTP</td>
<td>ribonucleotide triphosphate</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PGC</td>
<td>primordial germ cell</td>
</tr>
<tr>
<td>PFA</td>
<td>paraformaldehyde</td>
</tr>
<tr>
<td>PBST</td>
<td>phosphate buffered saline with 0.1% Tween-20</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>POD</td>
<td>peroxidase</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>SE</td>
<td>standard error</td>
</tr>
<tr>
<td>SS</td>
<td>splice site</td>
</tr>
<tr>
<td>SSC</td>
<td>sodium chloride trisodium citrate solution</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>TSA</td>
<td>tyramide signal amplification</td>
</tr>
<tr>
<td>WGD</td>
<td>whole genome duplication</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction
1.1 Alternative transcription and splicing

In recent years, with the application of genome wide analysis, the traditional approach of “one gene→one mRNA→one protein” has become an over-simplification. Wide-spread applications of high-throughput experiments, such as exon arrays and NGS (Next Generation Sequencing), have made us realized the prevalence of alternative transcription and splicing in almost all multi-exon genes (Pal et al., 2011; Sultan et al., 2008). In fact, 90% of the protein-coding genes use alternative transcriptional and alternative splicing events (Pan et al., 2008). For mammals, this number can reach 95% for alternative splicing event. In *Drosophila melanogaster*, 60% of genes undergo alternative splicing (Graveley et al., 2011).

Broadly speaking, alternative transcription and splicing consists of four categories, which are alternative transcription initiation, alternative splicing, alternative polyadenylation and alternative translation initiation (de Klerk and Hoen, 2015). For alternative transcription initiation, it could be divided into alternative first exons (AFEs) and alternative 5’UTR (A5UTR). As shown in Figure 1.1, cell- and tissue-specific AT transcripts are regulated by the binding of transcription factors (TFs) to promoters and enhancer regions. In some cases, whereas enhancers are located within the coding region, others are located in intergenic regions or within intronic regions of other genes (Hill and Lettice, 2013).
Figure 1.1. Alternative transcription initiation.

Alternative transcription initiation can be divided into (A). Alternative first exons and (B) Alternative 5’UTR. (C) Cell- and tissue-specific AT transcripts are regulated by the binding of transcription factors (TFs) to promoters and enhancer regions. (D) Long-range transcriptional control mediated by enhancers. Adopted from (de Klerk and Hoen, 2015).
Figure 1.2. Alternative splicing events. Five major AS events are illustrated, including (A) Exon skipping (ES), (B) Alternative 3’ splice site selection (A3SS), (C) Alternative 5’ splice site selection (A5SS), (D) Intron retention (IR) and (E) Mutually exclusive exons (MXE). Constitutive exons are shown in blue and alternatively spliced regions in purple. Introns are represented by solid lines, and dashed lines indicate splicing options. Adopted from (Keren et al., 2010)
As shown in Figure 1.2, alternative splicing (AS) events can be classified into four main subgroups (Keren et al., 2010). The first type is exon skipping (ES), in which an exon is skipped together with its flanking introns so that it does not appear in the final transcript (Figure 1.2A). Exon skipping can account for nearly 40% of higher eukaryotes AS events (Alekseyenko et al., 2007). The second and third types are alternative 3’ splice site (A3SS) and 5’ splice site (A5SS) (Figure 1.2B and C). These types of AS events occur when two or more splice sites are recognized at one end of an exon. A3SS and A5SS constitutes of 18.4% and 7.9% of all AS events in higher eukaryotes, respectively. The fourth AS event is intron retention (Figure 1.2D), in which the mature mRNA transcript includes an intron. This is the rarest AS event in vertebrates and invertebrates (Alekseyenko et al., 2007), but is the most common AS in plants (Kim et al., 2008). The last category includes mutually exclusive exons (Figure 1.2E), which is rare and more complex (Black, 2003).

Alternative splicing is regulated by splicing silencers which would interfere the recognition of strong and weak splice sites (SS) by small nuclear ribonucleoprotein (snRNP) splicing factor (Yu et al., 2008). Usually snRNPs recognize both strong and weak SS but splicing occurs only at the strong SS. However, with the existence of a splicing silencer sequence (SSS) located downstream of the strong SS, only the weak SS is used for splicing.

Alternative splicing can also be regulated by RNA secondary structures. When the short-range or long-range RNA secondary
structure masks the strong SS, a weaker SS can be recognized (Pervouchine et al., 2012).

The event of alternative promoter usage and/or alternative splicing results in a plethora of transcript variants corresponding to a single gene, which greatly evolves the generation of transcriptome and proteome complexity. The resultant transcript variants may remain as non-coding transcripts or be translated to different proteins that diverge in structure, activity or even functions. These differed proteins are called isoforms. For example, SRA1 produces both non-coding RNA and protein isoforms of steroid receptor RNA activator (Ulveling et al., 2011), and multiple BDNF transcript variants differing in the first exonic sequence translate the same BDNF protein (Pruunsild et al., 2007). The difference between AT and AS is that in AT, isoforms are produced by several pre-mRNAs, whereas in AS, isoforms all arise from a single pre-mRNA but in different splicing patterns. In this thesis work, we mainly focus on one specific AT event, usage of alternative promoters, which produces AFEs. AS events are also being studied.

1.2 Mitf as an example of alternative transcription and splicing

One of the best examples for alternative transcription perhaps is mitf, a highly conserved gene essential for pigment cells in diverse animal species.

The microphthalmia-associated transcription factor (mitf) encodes a basic helix–loop–helix leucine zipper transcription factor whose major role in vertebrates is to act as the master regulator of melanin-bearing
pigment cells (Hodgkinson et al., 1993; Widlund and Fisher, 2003). For instance, Human \textit{MITF} regulates melanocytes at specification, proliferation, survival, migration, differentiation and malignant transformation. It regulates retinal pigment epithelium (RPE) cells at the level of developmental specification, proliferation and function in retinal physiology. Mutation of \textit{MITF} causes Waardenburg syndrome (WS) type IIA which has melanocytic deficiencies in the eye, forelock and inner ear (Tassabehji et al., 1994). Common to nearly all \textit{Mitf} mutations in the mouse are defects in neural-crest-derived melanocytes, manifested by a lack of pigment in the coat and inner ear (Hallsson et al., 2000; Steingrimsson et al., 1994; Tsujimura et al., 1996). Most mutations also produce defects in retinal pigment epithelial cells, which in turn results in abnormal eye development. A few of the mutations also induce osteoclast defects, which can lead to severe osteoporosis, as well as mast cell defects (Roundy et al., 2003).

Genomic sequencing of \textit{mitf} in vertebrates has revealed a complex intron-exon structure with several widely spaced 5’ exons, which are expressed from different promoters, giving rise to protein products with different N-termini (Amae et al., 1998). In mammals, a single \textit{mitf} gene produces multiple RNA variants by differential usage of alternative promoters and alternative first exons (Bharti et al., 2008; Hallsson et al., 2000; Hershey and Fisher, 2005; Steingrimsson et al., 2004). In addition to alternative usage of 5’ exons, splice forms have been described that would lead to different Mitf proteins through the alternative uses of exons 2 through 9 (Takemoto et al., 2002; Udono et al., 2000).
In human, the *MITF* generates 4 ncRNAs and 13 protein-coding variants (http://asia.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000187098), and a total of 9 alternative promoters/first exons are known to generate various MITF isoforms with unique N-termini (Hou and Pavan, 2008). As shown in Figure 1.3, nine different promoters present in a 200kb region, assigning AFEs for each isoform. To better understand the how AFES affect protein isoforms, the structure of MITF protein isoforms are shown. Generally speaking, AFES will determine the 5’ specificity of isoforms. However, in some cases, alternative first exon usage can lead to same isoform production. For MITF-J, MITF-E and MITF-D, as they do not have ATG in their first exons, they share a common translated protein. Despite that, all isoforms share the same exons 2 to exon 9, where the functional domains reside (Udono et al., 2000), including the activation domain (AD) and the b-HLH-Zip involved in homo- or hetero-dimerization with the related TFs: TFE3, TFEB, and TFEC (Hallsson et al., 2007; Sato et al., 1997).

Some Conserved domain (CD) presents in some exons to serve as DNA-binding site or has regulatory functions (Hallsson et al., 2007). Furthermore, Mitf is also modified at some specific sites in the post-translational process, particularly phosphorylation by mitogen-activated protein kinase (MAPK) and glycogen synthase kinase-3b (GSK3b) (Hemesath et al., 1998; Takeda et al., 2000). These phosphorylation enhances the transcriptional activity of MITF. These kinases reside within various important homeostatic signalling pathways and might therefore
modulate Mitf transcriptional activity in response to specific environmental cues. MITF can be subjected to USP13-mediated deubiquitylation, thus preventing MITF from proteasomal degradation (Hartman and Czyz, 2015). In addition, MITF can be a target for other modifications, including SUMOylation (Murakami and Arnheiter, 2005) and caspase-mediated cleavage (Larribere et al., 2005).
Figure 1.3. Organization of human MITF gene.
(A). Genomic organization of human MITF gene. Nine different promoters/first exons of MITF are located throughout a region over 200kb while exons 2 to 9 are common. (B) Protein structure alignment of human MITF. The isoform-specific first exons (pink) are spliced to exon 1B1b (red) and then to the common exons 2–9 (grey), except for MITF-M. 7 isoforms are produced. Functional domains (AD, transactivation domain and b-HLH-Zip, basic/helix-loop-helix/leucine zipper) and some conserved domains (CD, conserved domain) are marked in yellow. (C). Structure of MITF-M, showing function domains and regulatory sites including phosphorylation, deubiquitylation, SUMOylation and Caspase-mediated cleavage. Adopted from (Hartman and Czyz, 2015)
Notably, human MITF isoforms show differential expression patterns. For example, MITF-A is expressed in a wide variety of organs including the eye and cartilage, and MITF-M is expressed preferentially in the neural crest and its pigment cell derivatives, where it acts as the master regulator of melanocytes, because its forced expression is sufficient to induce melanocyte formation in differentiated cells of mammals and embryonic stem (ES) cells of the fish medaka (Bejar et al., 2003).

1.3 Whole genome duplication (WGD) in fish and alternative splicing

In contrast to alternative transcription, it has been shown that fish has undergone a specific evolution and as a result, two Mitf proteins corresponding to the mouse MitfA and MitfM were generated (Altschmied et al., 2002; Braasch et al., 2009a). Gene duplication is a major force of evolution and plays a critical role in increasing complexity, and two rounds of whole genome duplication (WGD) have been proposed to occur during vertebrate evolution (Ohno et al., 1968). As shown in Figure 1.4, through the ancient WGD in the fish lineage ~350 million years ago (Yamanoue et al., 2006), many fish species have two or even more distinct genes that are homologous/paralogous to a single gene in mammals and birds.
Figure 1.4. Whole Genome Duplication (WGD) in fish. The teleost lineage splits from basal ray-finned fishes and started to diverge after a WGD event that took place 320–350 mya. For some fish species such as salmon and common carp, they have even more rounds of WGD. Adopted from (Glasauer and Neuhauss, 2014).
Figure 1.5. Evolutionary fate of single gene duplications. Single gene duplication most often results in a nonfunctional duplicate gene copy (A, nonfunctionalization). (B) In rare instances, the functional duplicate gene copy and the ancestral gene diverge in function; neofunctionalization means that one of the two genes retains the original function, while the other evolves a new, often beneficial function. Subfunctionalization implies that both the original and the duplicate genes mutate and evolve to full complementary functions already present in the original gene. Duplication via retrotransposition represents a particular case of sub- or neofunctionalization. (C). If there is no functional divergence, genetic robustness will be increased. Adopted from (Conrad and Antonarakis, 2007)
Duplicated genes may adopt three major modes of evolutionary divergence: loss by deleterious mutations (nonfunctionalization) (Figure 1.5A), acquisition of a new adaptive function (neofunctionalization) or division of an ancestral function (subfunctionalization) (Figure 1.5B). Neofunctionalization means that one of the two genes retains the original function, while the other evolves a new, often beneficial function. Subfunctionalization implies that both the original and the duplicate genes mutate and evolve to full complementary functions already present in the original gene. Duplication via retrotransposition represents a particular case of sub- or neofunctionalization. If there is no functional divergence, genetic robustness will be increased.

From the above introduction we know that both alternative splicing and gene duplication contribute to the proteomic functional diversity. Studies on the relation between AS and gene duplication across species (Abascal et al., 2015; Su et al., 2006) all support a function-sharing model that some proteomic components can be performed either by alternatively spliced genes or duplicate genes.

In fish, mitf has been identified in ~40 species, and the presence of two mitf genes has been documented in several model fish species including fugu, pufferfish, Xiphophorus (Altschmied et al., 2002) and zebrafish (Lister et al., 2001; Lister et al., 1999). The two zebrafish mitf genes differ in the first exon just as mammal isoforms of mitf. Nacre mutation of mitfa affects the development of pigment cells on the skin but not in the eye, leading to absence of melanocyte lineage cells. The eye pigmentation is controlled by mitfb, which is mainly expressed in,
and required for, RPE development. Interestingly, ectopic expression of
mitfb can rescue the pigment phenotype of mitfa mutant (Lister et al.,
2001; Lister et al., 1999). Therefore, in zebrafish, the two mitf genes
have undergone subfunctionalization to provide complementary
expression and functions in pigmentation (Altschmied et al., 2002; Lister
et al., 2001). The zebrafish studies have also provided the first evidence
that differential expression but not protein structure of the mitf isoforms
largely determines the ultimate mitf role in pigmentation.

Medaka model to study alternative mitf transcription and splicing.

In this thesis work, Medaka (Oryzias latipes) serves as a good model
to study pigment cell biology, function and evolution of mitf genes in fish.
The small laboratory fish is one excellent lower vertebrate model for
analyzing development (Wittbrodt et al., 2002) and in particular,
pigmentation biology. Compared to mammals, medaka has four types of
pigmented cells (melanophore, leucophore, iridophore and xanthophore)
instead of only one (melanocyte). There are more than 40 inbred lines
and mutant strains showing diverse pigmentation patterns and
abnormalities (Kelsh et al., 2004). In medaka, several pigmentation-
related genes have been identified, including tyrosinase (Koga and Hori,
1997), Oca2 (Fukamachi et al., 2001), somatolactin (Fukamachi et al.,
2004) and vps11 (Yu et al., 2006). Furthermore, cell lines of medaka ES
cell (Hong et al., 1996) and male germ stem cell (Hong et al., 2004) have
been established for studying pigment cell development in vitro.

Medaka has two mitf genes, mitf1 and mitf2, which encode mitf1 and
mitf2 that correspond to the mouse MitfA and MitfM, respectively, on the
basis of gene structure, chromosome synteny and phylogenetic sequence comparisons.

Above all, medaka is a perfect tool to study alternative mitf transcription and differential activity.

Previous studies in our lab have identified two medaka mitf genes, mitf1 and mitf2. Mitf is a transcriptional activator of medaka germline genes (Zhao et al., 2012) and both mitf1 and mitf2 RNAs are a component of the Balbiani’s body (Li et al., 2013a), which is a transient structure in the developing oocyte and thought of as being a stage-specific equivalent of germ plasm (Xu et al., 2010).

1.4 RNA-Seq for revealing alternative transcription and splicing

In the past ten years, with the development of technologies in sequencing broadens our vision on alternative transcription and alternative splicing, especially in a genome-wide manner. Compared to tilling array and EST sequencing, RNA-Seq has the invaluable advantage of high resolution and throughput and also has low required amount of RNA and low cost.

RNA-Seq has the power to give us the number of AT and AS transcripts, the structure of each transcript (i.e. how each transcript is formed by AT and AS process) and also the abundance of transcripts all in the same process. To a deeper extent, even the molecular mechanisms that form the transcripts can be elucidated. A thorough comparison of RNA-Seq platforms is reviewed in (van Dijk et al., 2014). In my study, I choose the most widely used and cost-effective platform:
HiSeq from illumina. In this way, my raw data would be pair end short reads.

The power of RNA-Seq to reveal AT and AS depends on a crucial step: transcriptome reconstruction. This step is to define a precise map of all transcripts that are expressed in the sample sequenced. It requires the assembly of read alignments. This is a difficult computational task due to the divergence of gene expression profile, existence of premature mRNA contamination and determination of short read alignment if many isoforms exist. One strategy to do this reconstruction is called Reference Annotation Based Transcript (RABT), which firstly provides a reference genome to mapping all the reads to it and then assembles overlapping reads into transcripts (Trapnell et al., 2012). This method was widely used and examined.

Researchers are now utilizing the power of RNA-Seq to investigate into more specific biological questions including the quantification of alternative splicing events in tissues (Cui et al., 2012; Wang et al., 2008), populations (Blekhman et al., 2010) and disease (Gamazon and Stranger, 2014; Wilhelm et al., 2011).

1.5 Significance and objective

*mitf* is a pleiotropic factor, because its mutations cause a variety of abnormality including pigmentation-related diseases, melanomas and cartilage defects. In mammals, *mitf* represents a family of isoforms generated by alternative promoter choice and alternative splicing. In fish, whole genome duplication also provides more transcripts or protein
isoforms. It is unknown whether alternative transcription is a mechanism for *mitf* expression that is unique to mammals or common to vertebrates, as no report on alternative transcription has been described for *mitf* in lower vertebrates. Furthermore, the pigmentary cell lineage comprises solely of melanocytes in mammals but 4 distinct pigment cells in fish, namely melanophores, iridophores, xanthophores and leucophores (Kelsh et al., 2004), making more potential target genes of *mitf*. More importantly, it is also unknown whether distinct Mitf roles are due to differential expression and/or varying activities of different isoforms.

The objective of this thesis is to answer three biological questions. The first aim is to reveal the differential evolution of fish duplicated *mitf* genes after WGD. The second aim is to test that AT and AS of *mitf* also exists in fish. The third is to explore whether that differentially expressed mitf isoforms have overt differences in activity and specificity of inducing target genes. One or more dominant isoforms will reveal their role in particular organs or stages of development, leading to discoveries of new target genes or upstream regulating genes and connecting *mitf* to new signalling pathways. To achieve these goals, molecular cloning, transcriptome analysis, gene expression profile, reporter assay and melanocyte induction are applied.
Chapter 2: Materials and Methods
2.1 Fish and chemicals

2.1.1 Fish

Work with fish followed the guidelines on the Care and Use of Animals for Scientific Purposes of the National Advisory Committee for Laboratory Animal Research in Singapore (Protocol Number: 067/12). Medaka strains HdrR and HB32C were maintained under an artificial photoperiod of 14-h/10-h light/darkness at 26°C as described (Hong et al., 2011).

2.1.2 Chemicals

In this study, chemicals were purchased from Sigma (MO, USA) unless otherwise indicated. Deionized water was collected from PURELAB Prima system (ELGA Lab). Ultrapure water was collected from Milli-Q Water Purification Systems (Merck Millipore).

2.2 Molecular cloning work

2.2.1 Agarose gel electrophoresis

For agarose gel electrophoresis, 1.0% or 1.5% agarose gels were casted with 1×GelRed DNA stain (Biotium). For each well, 2-5 μl sample with Loading Buffer (Takara) was loaded. The gels were run at 5-8V/cm in 1×TAE buffer (0.04 M Tris-base, 0.02 M acetic acid, 0.01 M EDTA, pH 8.0) with Horizontal Electrophoresis Systems (Bio-Rad).

2.2.2 RNA extraction

Total RNA isolated from adult fish organs or developing embryos was extracted by using the Trizol reagent (Invitrogen) following the
manufacturer’s instruction. 100 mg sample with the addition of 1 ml TRIzol was homogenized in 1.5 ml EP tube with a homogenizer on ice (Sigma). For the final step, RNA samples were dissolved in 20 μl RNase-free water. RNA quality were checked by using agarose gel electrophoresis and concentration were assessed on a WPA Biowave II UV/Visible Spectrophotometer from Biochrom (Cambridge, UK). For samples for next-generation sequencing, RNA quality and concentration were assessed using Agilent 2100 Bioanalyzer (Agilent Technologies) and CYTATION 3 Cell Imaging Multi-Mode Reader (Bio Tek). After that, ~15 μg of total RNA dissolved in 40 μl water was sent to BGI TECH SOLUTIONS (HONGKONG) CO., LIMITED (Hong Kong, China).

2.2.3 Purification of DNA from agarose gel or enzyme reaction solution

For purification of DNA from agarose gel, DNA bands of interest were identified and cut under UV light and weighed in an EP tube. For purification of DNA from enzyme digestion or other reaction, there was no such prior step. DNA ready for purification was then applied to MN NucleoSpin Gel and PCR Clean-up Kit (MACHEREY-NAGEL) following manufacturer’s instructions.

2.2.4 cDNA library synthesis

For RT-PCR validation, the cDNA libraries were synthesized from 2 μg of RNA extracted using Maxima First Strand cDNA Synthesis Kit (Life technologies) following the manufacturer’s instruction. Synthesis of SMART cDNA libraries was primed with oligo (dT)8-12 by using the
RACE cDNA Amplification Kit according to the manufacturers’ instructions (Clontech).

2.2.5 Rapid amplification of cDNA ends (RACE)

For 5’-RACE, the SMART library 5’-adapter primer (5’-AACGCAGAGTACGCAGGGG-3’) was used in combination with a gene-specific primer for a first round PCR, the latter being m1r5 (5’-CAGCTTGCGGATGTAATCTAC-3’) for mitf1 and m2r5 (5’-TGGTAACCTTCCGCTTCGT-3’) for mitf2. For 3’-RACE, the primer for mitf1 was 5’-CCTCAACTCCAACTGCGAGAAAGAGATGG-3’ and primer for mitf2 was 5’-ATCAAGGACTAACGCTGTCCAACCTGCTA-3’ for mitf2. The PCR reaction was conducted on Mastercycler from Eppendorf (Hamburg, Germany) in a volume of 50 µl, containing 5 ng of a SMART cDNA reaction, 200 µM of four dNTPs, 0.2 µM of each primer, and 1 µl Advantage 2 Polymerase Mix (Clontech). A second round of nest PCR was similarly run for 30 cycles by using 1 µl of the first PCR reaction and the 5’-adaptor primer together with m1r4 (5’-TGCCCTTTCCAGGCATGT-3’) for mitf1 and m2r4 (5’-GGCAAGGCTCTCAAACCTCGC-3’) for mitf2. PCR products were cloned into pGEM-T (Promega) and sequenced. The primers were designed with DNAMan software from Lynnon Biosoft (San Ramon).

2.2.6 RT-PCR

PCR analysis was done as described (Hong et al., 2004b; Yi et al., 2009). Briefly, PCR was run in a 25-µl volume containing 10 ng of cDNA reaction for 30 cycles (β-actin), 36 (mitf1), 38 (mitf1 RNA variants), 40 (mitf2) or 35 cycles (other genes) of 10 s at 94°C, 15 s at 58°C and 60 s
at 72°C. Primers used are listed in Appendices. Primers m1f1 plus m1r1 and m2f1 plus m2r1 (See Appendices) were used to amplify mitf1 and mitf2 RNAs, respectively. The four mitf1 variants were individually amplified by using the common 3’ primer 5’-ACAGCCCTGCTCGTTTTCCATCCAT and variant-specific primer for mitf1a (M1af; 5’-ATGCAGTCCGAATCCGGAATA-3’), mitf1b (M1bf; 5’-ATGACATCAGCAGCCATCCTGCTACGC-3’), mitf1h (M1hf; 5’-ATGGAGGGCCCATAGGAGTGCAAGTG-3’) and mitf1m (M1mf; 5’-ATGCTTGAAGATGCTCGGATAC-3’). The PCR products were separated on 1.0% agarose gels and documented with a bioimaging system (Synoptics).

2.2.7 Real-time RT-PCR

Real-time PCR was performed with SYBR® Green qPCR SuperMix (Invitrogen) on the IQ5 real-time PCR detection system (BioRad). The primers for detecting were tyr1514F (5’-TCACCATGCTTTCCATTGACAGC-3’) and tyr1621R for the medaka tyra, mitf1-QF (5’-TCTCCTGCCCATCATGCACA-3’) and mitf1-QR (5’-AGCGCTTACCTCAGGCTTCA-3’) for the medaka mitf1, mitf2-QF (5’-CAACTTGATTGAACGAAGGCG-3’) and mitf2-QR (5’-TGTTCCAGCGCATATCTGGAT-3’) for the medaka mitf2, actinqs (5’-AGAAGAGCTATGAGCTGCCTG-3’) and actinqr (5’-AACATCACACTTCATGATGCTG-3’) for the medaka β-actin as an internal control. The relative fold change of tyr quantities was calculated with 2(-delta delta Ct) method (Livak and Schmittgen, 2001).
2.2.8 Droplet digital PCR (ddPCR)

Quantification of RNA transcripts was performed with QX200 automated ddPCR system (Bio-rad) strictly following the manufacturer’s instructions. Briefly, cDNA templates were first mixed with 2× QX200 ddPCR EvaGreen Supermix to a total volume of 20 μl. The mixture was transferred to the 96-well plate (Eppendorf) and subjected to oil droplets generation with the automated generator (Bio-Rad). The oil droplets were further amplified with the following set-up: 95 °C for 10 min (1 cycle), 94 °C for 30 s and 60 °C for 1 min (40 cycles), 98 °C for 10 min (1 cycle) and 4 °C until droplet reading. Droplets were read by QX200 Droplet Reader (Bio-Rad) and analyzed with the QuantaSoft software (Bio-Rad) that determines concentration of target cDNA as copies per microliter (copies/μL) from the fraction of positive droplets using Poisson statistics (Beliakova-Bethell et al., 2014). The copy number of each mRNA was further normalized with β-actin.

2.2.9 Cloning

PCR products or other DNA segments of interest were applied to TA-cloning method for cloning. In our existing PCR system (described above), there was no need for a step of 3’ A overhang addition prior to TA-cloning. pGEM-T Easy vectors (Promega) were used to perform TA-cloning. After ligation, DNA products were transformed into HIT-DH5alpha high competent cells (BioAspect) according to manufacturer’s guide or self-prepared competent cells originated from strain TOP10F (Invitrogen) following protocol from Chapter 3 in Molecular Cloning: A Laboratory Manual (Sambrook and Russell, 2001).
2.2.10 Automated sequencing analysis

DNA Sequences were determined by using the BigDye Terminator v3.1 Cycle Sequencing Kit on ABI PrismR 3100 Genetic Analyzer (Applied Biosystems). Sequence alignment was run on the Vector NTI Suite 11 (Invitrogen).

A phylogenetic tree was constructed by using the DNAMAN software (version 4.15, Lynnon Biosoft) from a matrix of pairwise genetic distances according to the neighbor-joining (NJ) and minimum evolution (ME) methods, and 1,000 trials of bootstrap analyses were used to provide confidence estimates for tree topologies. Genomic organization and chromosomal locations were investigated by comparing the cDNA and corresponding genomic sequence.

2.2.11 Plasmids

Five luciferase (Luc) reporter plasmids were used in this study. pGL3-basic is promoterless and contains the firefly luc (Promega). pTALuc was derived by inserting the TATA-containing minimal promoter of the Herpes Simplex virus thymidine kinase upstream of luc in pGL3-basic (Promega). pDAZLuc was described (Zhao et al., 2012a). To construct pDAZLuc, the 4.2-kb promoter sequences of the medaka dazl were PCR-amplified from genomic DNA and inserted into MluI and XhoI sites that are upstream of the luc in pGL3-Basic, respectively. The PCR primers used were 5’-acgcgTCATCACCAGTGAGGTCC-3’ plus 5’-ctcgaGAATTCTTTCTAAATGATTAAAACCAAGAATTGG3’ (restriction sites are underlined) for DAZ. To construct pTYRLuc, the 3.3-kb promoter of a medaka tyrosinase gene (tyra; accession number
AB010101) was obtained by PCR from genomic DNA with primer 5'-TTACGCGTAGTGATGATGCACCTCTGC-3' and 5'-TCTCGAGATTCACCACATCTGTCC-3' (restriction sites are underlined) and inserted into MluI and XhoI sites that are upstream of the luc in pGL3-Basic. pRL-CMV expresses the Renilla Luc and was used as internal reference to calibrate transfection efficiency (Promega).

Briefly, pXmitf is a derivative of pCMVtkLUC+, in which the luc coding sequence was replaced by a fusion of a synthetic DNA fragment encoding the myc-tag and the *Xiphophorus* mitf-M coding sequence (Zhao et al., 2012a). Seven expression plasmids were used as transfection control or effectors. Plasmids pCVpf and pOmitf2 were described, which express the fusion between the puromycin acetyltransferase gene and GFP (Zhao et al., 2012a) and the medaka mitf2 (Li et al., 2013b). Plasmid pCVpf was constructed by inserting in frame the puromycin acetyltransferase gene to *gfp* in pEGFP-N1 (Clontech). pOmitf2 was constructed by inserting the medaka *mitf2* cDNAs in frame to the mCherry sequence in pCS2+. pOmitf1a, pOmitf1b, pOmitf1h, pOmitf1m and pOmitf1N were constructed by inserting the medaka *mitf1* cDNAs for isoforms A, B, H and M in frame to the mCherry sequence in pCS2+, respectively. The *mitf1* cDNA variants (*mitf1a, mitf1b, mitf1h, mitf1m and mitf1n*) were PCR-amplified by using forward primers M1aF, M1bF, M1hF, M1mF and M1nF in combination with a common reverse primer M1R (5'-ACAGCCCTGCTCGTTTTTCATCA-3'). The cDNAs for the five Mitf1 variants and Mitf2 were inserted between BamHI or EcoRI and ClaI site of pCS2cherry, resulting in pOmitf1a:ch,
pOmitf1b:ch, pOmitf1h:ch, pOmitf1m:ch, pOmitf1N:ch and pOmitf2:ch, which express a fusion between each of Mitf proteins and mCherry. Correct cloning was confirmed by test digestion and/or sequencing. Plasmid DNA was prepared by using the Plasmid Midprep kit (Qiagen).

2.3 in situ hybridization

Antisense RNA probes were synthesized from linearized pGEMtyr, pGEMmi1 and pGEMmi2 by using the DIG RNA labeling kit (Roche) as described (Li et al., 2011b). Embryos at stage 29 were subjected to in situ hybridization (ISH) with DIG-labeled probe followed by chromogenic staining with NBT/BCIP as described (Liu et al., 2009). Sections were subjected to two-color fluorescence in situ hybridization (Hershey and Fisher) by using the tyramide signal amplification (TSA™ Plus Fluorescence Systems according to the manufacturer’s instruction (NEL756, PerkinElmer Life Science) as described (Li et al., 2011b). Briefly, after hybridization and blocking as described above, the samples were incubated with the POD-conjugated anti-FITC-antibody (Sigma) at a 1:2000 dilution for 2 hours at room temperature to detect the FITC-labeled probes. Following six PBST washes the samples were incubated for 30 minutes in the TSA-Fluorecein at a 1:100 dilution in the TSA Amplification Buffer (0.004% H₂O₂ in 0.1 M borate buffer, pH8). The samples were then subjected to detection of the DIG-labeled probe: They were treated for 1 hour in 1% H₂O₂ and incubated for two hours with the horseradish peroxidase (POD)-conjugated anti-DIG antibody (Sigma) at a 1:2000 dilution, followed by a 30-min incubation in TSA-Cy3 for 30 min. The samples were finally stained for nuclei by using DAPI
and embedded in the Gold Antifade reagent (Invitrogen) for microscopy. In this FISH procedure, two differently labeled antisense RNA probes are co-incubated with a sample in the hybridization step and subjected to sequential color development into green and red fluorescence (Li et al., 2011a).

2.4 Cell culture and transfection

Plasmid DNA was prepared by using the Qiagen Midi- and Maxi-preps kit (Qiagen). The medaka ES cell line MES1 was maintained on gelatin-coated tissue culture plasticware in medium ESM4 containing 15% fetal bovine serum (Hong et al., 1996) and transfected as described (Hong et al., 2004a). In this cell line it has previously been shown that pXmitf transfection is sufficient to direct melanocyte differentiation (Bejar et al., 2003). Cells in 24-, 12- and 6-well plates were transfected with 0.5, 1 and 3 μg of plasmid DNA by using the DNAfectinTM 2200 reagent (Applied Biological Materials Inc.) according to the manufacturers’ instruction. Each of Mitf-expression plasmids was cotransfected with pCVpf at a 1:4 ratio. Control transfection received only pCVpf. Appearance of black pigmented melanocytes was monitored from day 3 post transfection (dpt) onwards. In some experiments, cell transfection was followed by drug selection to enrich for transgenic cells. For this, cells following cotransfection with pCVpf were grown in the presence of puromycin (final 1 μg/ml) for 2 days before harvest for RNA isolation and RT-PCR analysis.
2.5 Luciferase reporter assay

For reporter assays, cells were transfected in 96-well plates (Zhao et al., 2012b). Briefly, ~30,000 cells were seeded in 96-well plates (Biostar) and transfected with 102 ng of plasmid DNA. The following day, cells were at ~70% of confluence and transfected. For cotransfection, plasmid DNA for each well of 96-well plate contains 2 ng of pRL-CMV as the internal reference, 50 ng of a firefly luciferase reporter (pGL3-basic, pTAluc, pDAZluc or pTYRluc), 25 ng of Mitf-expressing plasmids or pCVpf. For control transfection, each well received 2 ng of pRL-CMV, 50 ng of a firefly luciferase reporter plasmid and 50 ng of pCVpf. Transfection efficiency was monitored by GFP expression from pCVpf in a duplicate 96-well plate to be at ~30%. At day 3 or 4 dpt, cells were measured for luciferase activity by using the Dual-GloTM Luciferase assay kit (Promega) and measured on GlomaxTM 96 Microplate Luminometer (Promega). Briefly, each well was replaced by 20 µl of fresh medium containing only 9% FBS. The plate was equilibrated at room temperature (RT) for 10 min. For measuring firefly luciferase activity, 20 µl of Dual-Glo™ Luciferase reagent was added to each well and mixed. After incubation at RT for 20 min, the firefly luminescence was measured in the GlomaxTM 96 Microplate Luminometer (Promega). Subsequently, 20 µl of Dual-Glo™ Stop & Glo® Reagent (1:100 dilution of substrate to buffer) was added to each well, mixed, incubated at RT for 10 min and measured similarly for the Renilla luciferase activity. The ratio between firefly luciferase activity and Renilla luciferase activity was used for comparisons. The reading from pGL3-basic was considered as
background, and that from pTALuc treated as unit activity. Relative luciferase activity from pTYRLuc or DAZLuc alone or in combination with Mitf-expressing plasmids was normalized to that of pCVpf.

2.6 Microinjection and western blot

Medaka embryos were injected at the 1-cell stage as described (Li et al., 2012). They were injected with each of plasmids for mCherry-tagged Mitf isoforms plus pCVpf at a 1:1 ratio (total 25 pg/nl) and analyzed at day 2 post fertilization (the mid-gastrula stage). Positive embryos were used for protein isolation. Western blot was done as described (Xu et al., 2005). Briefly, embryos were put in DMSO and dechorionated to remove yolk with very sharp forceps. When the yolk has run out, the embryos were transferred to a new tube and incubated in RIPA buffer (Pierce) for protein isolation. About 30 μg of total protein from the embryos extract was subjected to 10% SDS-PAGE Bio-Rad mini Protein electrophoresis system. The separated proteins were transferred to a PVDF membrane by using the Transblot mini-protein transfer system. The membrane was first incubated with anti-mouse mCherry primary antibody (ab125096, Abcam) at a dilution 1:2000 in 5% non-fat milk in TBST (100 Mm Tris-Cl, pH 7.5, 100Mm NaCl, 0.1% Tween 20) for 1 hour at room temperature. After three washes for 10 min in TBST, the membrane was incubated for 1 hour at room temperature with diluted (1:10000) anti-mouse HRP secondary antibody (A9044, Sigma). After another three washes of TBST, image was made by using ECL Plus Western Blotting Substrate (Pierce). The membrane was then stripped in 7 M Guanidine hydrochloride (Sigma), after two washes in TBST, the membrane was
subjected to anti-mouse GFP primary antibody (1:5000, ab1218, Abcam) and anti-mouse HRP secondary antibody (1:10000, A9044, Sigma) for a secondary round of analysis.

2.7 Microscopy

Microscopy was done as described (Xu et al., 2009). Briefly, Cells were observed and photographed on Zeiss Axiovert2 invert microscope and Axiovert200 upright microscope equipped with a Zeiss AxioCam MRc digital camera and AxioVision 4 software, and embryos and fry were visualized using a Leica MZFLIII stereo microscope (Leica) equipped with a Fluo III UV-light system and photographed by using a Nikon E4500 digital camera (Nikon Corp).

2.8 Statistical analysis

GRAPHPAD PRISM (version 6, La Jolla, CA, USA) software was used. To compare means, we used the unpaired T-test analysis. Statistical difference was considered significant and very significant if it had a two-sided P-value of <0.05 and <0.01.

2.9 Transcriptome sequencing

Transcriptome sequencing was carried out on Hiseq 2000 platform from illumina. Pools of RNA were prepared, containing RNA extracted from eye, testes and ovary from adult fish, stage 10 embryos and young fries of 7dpf. Checked RNA (for integrity, purity and concentration as described before) was sent to the service provider BGI TECH SOLUTIONS (HONGKONG) CO., LIMITED (Hong Kong, China) for library preparation and the subsequent sequencing. During the QC steps,
ABI StepOnePlus Real-Time PCR System (ABI) were used in quantification and qualification of the sample library. 2x100 bp paired-end reads were generated in a single lane from an Illumina HiSeq2000 following the service provider’s protocol.

2.10 Bioinformatics analysis of RNA-Seq

After sequencing, 2x100 bp paired-end reads were generated, which were called raw reads. To illustrate the biological meaning of these sequences, the sequenced reads were first mapped to the medaka genome in order to get the region that were aligned, showing the origin of these sequences. Alignments would then be assembled to generate transcripts of genes. After that, transcripts of different sample libraries would be merged into a final transcriptome assembly. Further application of the transcriptome assembly includes annotation of the transcripts, counting of the reads so that the expression level of a particular transcript can be measured. All applications were applied on a 64-bit Ubuntu operating system (version 14.04, Canonical Ltd)

2.10.1 Quality assessment and filter of raw reads

Quality assessment of raw reads was performed by FastQC software ([http://www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)) in its command line version, simply running the fastqc command with default options.

2.10.2 Alignment/Mapping reads to genome

For the alignment of reads to the genome, Tophat2 (version 2.0.13, [http://ccb.jhu.edu/software/tophat/](http://ccb.jhu.edu/software/tophat/)) was chosen as the reads mapper.
Medaka genome sequence (fasta format) and gene annotation (gtf format) was downloaded from Ensembl ftp site (version 77) as reference. To reduce the mapping of remaining rRNA, tRNA, non-coding RNAs were masked in the annotation file. The maximum intron length was set to 300000 for medaka.

2.10.3 Quality assessment of alignments

For quality assessment of alignments, Qualimap (http://qualimap.bioinfo.cipf.es/) was used in its command line version using the alignment files sorted and masked gene annotation file.

2.10.4 Assembly of transcripts

To assemble transcripts, Cufflinks (version 2.2.0, http://cole-trapnell-lab.github.io/cufflinks/cufflinks/) was chosen as the genome-guided transcriptome assembler. Transcript assembly was carried out using masked gene annotation file to guide RABT assembly.

2.10.5 Generation of transcriptome assembly

To assemble transcriptome generated from each sample library, Cuffmerge (http://cole-trapnell-lab.github.io/cufflinks/cuffmerge/) was used to perform the merging of these assemblies into a master transcriptome for downstream differential expression analysis of the new transcripts that had been assembled.

2.10.6 Counting the transcripts

To measure the expression level of transcripts, Htseq-count was used to count the transcripts that come from the final transcriptome assembly.
2.10.7 Annotation of transcripts

For annotation of transcripts in the final transcriptome assembly, BLAST+ suite from NCBI (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST) was used to blast the merged library against the NCBI nucleotide collection (Nt) database using BLASTN and NCBI non-redundant (Nr) and Swiss-Prot protein sequence database using BLASTX with default settings following instructions from the local blast website on NCBI.
Chapter 3: Results
3.1 Differential evolution of duplicated fish mitf genes

3.1.1 Medaka Mitf1 and Mitf2 correspond to mammalian MitfA and MitfM

The sequenced medaka genome has two annotated mitf genes called mitf1 and mitf2 (http://asia.ensembl.org/Oryzias_latipes/Gene/Summary?db=core;g=ENSORLG00000013461 and http://asia.ensembl.org/Oryzias_latipes/Gene/Summary?db=core;g=ENSORLG00000003123). We followed this annotated gene nomenclature. The sequence appears to be partial for mitf1 and complete for mitf2. A sequence comparison led to a putative full length of open reading frame (ORF) for mitf1. Cloning of PCR products using primers in arrow in cDNA library of embryos and sequencing led to the discovery of the full length cDNA sequences for mitf1 and mitf2, which predict proteins Mitf1 and Mitf2 of 512 and 406 amino acid residues, respectively (Figure 3.1). On a sequence alignment, both medaka Mitf proteins are highly similar to the known Mitf homologues from other fish species and human (Figure 3.2).

A single gene in birds and mammals produces multiple Mitf isoforms with distinct N-terminal sequences. Two major isoforms are MitfA and MitfM, which in human have 118 and 11 amino acids in the N-terminus, respectively. Together with a comparison in protein domain structure especially the first exon (Figure 3.2), the medaka Mitf1 corresponds to the human isoform MitfA, and the Mitf2 is equivalent to the human MitfM. Therefore, the two major fish Mitf proteins encoded by two distinct genes.
resemble the two major Mitf isoforms encoded by a single gene in higher vertebrates.

**Figure 3.1 Nucleotide and deduced amino acid sequence of medaka mitf1 and mitf2.**

(A) *mitf1*. (B) *mitf2*. Start codon and stop codon are shown in bold. Broken and solid arrows depict sequences of primers for PCR analysis and cDNA cloning, respectively. Sequences are deposited in GenBank under accession numbers KC249979 (*mitf1*) and KC249980 (*mitf2*).
Figure 3.2. Mitf protein sequence alignment. Common species names, Mitf proteins or isoforms and percentage sequence similarity values compared to the medaka Mitf1 and Mitf2 are seen. The basic helix-loop-helix leucine zipper sequences are indicated. Also indicated are activation domain 1 (AD1) and 2 (AD2), MAPK and GSK3β. For accession numbers see Figure 3.3B.
So far the fish *mitf* genes and their protein products have controversially been named in the literature. For example, fish genes whose protein products are similar to the mammalian MitfM isoform have been called *mitfa* in zebrafish, fugu and tetraodon or *mitfm* in *Xiphophorus*, whereas those whose protein products are similar to the mammalian MitfA isoform have been referred to as *mitf1* or *mitfb* (Altschmied et al., 2002; Delfgaauw et al., 2003; Lister et al., 2001). All the five sequenced genomes of fish species, namely zebrafish, medaka, fugu, tetradon and stickleback, contain two separate *mitf* genes. The gene pairs are annotated as *mitf1* and *mitf2* in both medaka and stickleback, which were followed in this study. On the basis of sequence alignment (see below), fish Mitf1 is more similar to the mammalian MitfA, and fish Mitf2 is equivalent of the mammalian MitfM, respectively. Hence, *mitfa* or *mitfm* genes previously described in other fish species are equivalents of the medaka *mitf2* gene, and previous *mitfb* genes correspond to the medaka *mitf1*. In this regard, the annotated stickleback Mitf1 and Mitf2 appears to be equivalents of medaka Mitf2 and Mitf1, respectively (Figure 3.3A). Designation of fish *mitf* genes as *mitf1* and *mitf2* will allows to similarly nomenclature fish Mitf isoforms, a salient feature of the *mitf* gene as has been demonstrated in mammals (Hou and Pavan, 2008;
Figure 3.3 Phylogenetic tree of Mitf proteins. (A) Phylogenetic tree by the NJ algorithm. (B) Phylogenetic tree by minimum evolution. The Drosophila Mitf serves as the out-group. Bootstrap values are given. Accession numbers follow organisms and original names of Mitf proteins.
By using an invertebrate Mitf protein from *Drosophila* (Hallsson et al., 2004) as the out-group, phylogenetic trees were constructed on the basis of amino acid sequences. In the phylogenetic tree constructed by using the neighbor-joining algorithm on the DNAMAN package, all vertebrate Mitf proteins fall into two major clades. One clade consists of fish mitf2 proteins only, and the other comprises fish Mitf1 proteins and all mammalian Mitf proteins, as represented by the human MitfA and MitfM isoforms (Figure 3.3A). However, fish Mitf2 but not Mitf1 proteins may clustered with tetrapod counterparts on phylogenetic trees constructed by minimum evolution methods (Figure 3.3B). The inter-species mitf1 sequence similarity is generally higher than intra-species comparisons between mitf protein pairs in the fish species, suggesting that fish mitf duplicates may have shared an ancient WGD event in the common ancestor.

**3.1.2 Medaka mitf1 and mitf2 differ in sequence divergence rate**

A total of ~40 fish mitf sequences available in public databases ([http://www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov); [http://www.ensembl.org](http://www.ensembl.org)). This allowed us to compare the sequence divergence of Mitf duplicates in various fish species by using tetrapod Mitf proteins as a reference (Table 1). Seven Mitf pairs of representative species fall into two categories. The first include *Xiphophorus*, fugu, two cichlid species and zebrafish and exhibits an apparent difference of 6% in sequence divergence between Mitf1 and Mitf2: On average, the 5 Mitf1 proteins are 73.5% identical to the tetrapod Mitf proteins, whereas this value decreases to 67.6% for the 5 Mitf2 proteins (Table 1). Medaka also belongs to this category but
represents an extreme. Notably, the medaka Mitf1 is 72.2% similar to the tetrapod reference. However, this value for the medaka Mitf2 sharply declines to 60.2%, leading to a difference of up to 12% between the medaka Mitf1 and Mitf2 (Table 1). The second category is represented by stickleback, in which Mitf1 and Mitf2 display 67.1~69.7% and 67.6~72% identity values to the tetrapod reference, respectively, and thus a similar rate of sequence divergence. Taken together, medaka Mitf1 exhibits a slower, and perhaps basal, evolutionary rate of change, whereas Mitf2 shows a significant acceleration in divergence rate.

<table>
<thead>
<tr>
<th>Fish Mitf (aa)</th>
<th>Tetrapod Mitf (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>New name</strong></td>
<td><strong>Species</strong></td>
</tr>
<tr>
<td>Mitf1</td>
<td>Xiphophorus</td>
</tr>
<tr>
<td></td>
<td>Fugu</td>
</tr>
<tr>
<td></td>
<td><em>Maylandia callaisos</em></td>
</tr>
<tr>
<td></td>
<td><em>Maylandia zebra</em></td>
</tr>
<tr>
<td></td>
<td>Zebrasfish</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
</tr>
<tr>
<td>Medaka</td>
<td>Mitf1 (523)</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
</tr>
<tr>
<td>Mitf2</td>
<td>Xiphophorus</td>
</tr>
<tr>
<td></td>
<td>Fugu</td>
</tr>
<tr>
<td></td>
<td><em>Maylandia callaisos</em></td>
</tr>
<tr>
<td></td>
<td><em>Maylandia zebra</em></td>
</tr>
<tr>
<td></td>
<td>Zebrasfish</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
</tr>
</tbody>
</table>

| Medaka | Mitf2 (466) |                  | 59.5       | 63.0       | 58.9       | 60.9       | 59.5       | 59.1       |
| **Average** |            |                  | 60.2       |            |            |            |            |            |
| stickleback | Mitf1 (404) |                  | 68.8       | 72.0       | 66.8       | 69.0       | 67.9       | 67.6       |

**Table 1. Percentage Mitf sequence identities between fishes and tetrapods.** Protein sequence accession number see Figure 3.3. Number in parenthesis shows the number of AAs in CDS.
3.1.3 Medaka \textit{mitf1} and \textit{mitf2} genes are the products of WGD

Comparisons of chromosome locations revealed clear syntenic relationships between the \textit{mitf}-bearing regions of human chromosome 3 and those of fish chromosome pairs, chromosomes 5 and 7 in medaka, and chromosomes 6 and 23 in zebrafish, respectively (Figure 3.4). These data strongly suggest that the presence of \textit{mitf} gene pairs in diverse fish species is the consequence of WGD, which has been dated \(\sim 450\) million years ago (Amores et al., 1998; Meyer and Schartl, 1999; Postlethwait et al., 2000). The chromosome syntenic relationship also demonstrates that the medaka \textit{mitf1} and \textit{mitf2} are homologs of the human \textit{Mitf} gene.

**Figure 3.4.** Fish \textit{mitf1} and \textit{mitf2} are homologs of the human Mitf by chromosome synteny. Medaka and zebrafish are compared to human. Data are compiled from [http://www.ensembl.org/index.html](http://www.ensembl.org/index.html). Chr, chromosome; numerals in parentheses, chromosomal positions.
3.1.4 Medaka *mitf1* and *mitf2* differ in the expression level

In mammals, a single *mitf* gene shows differential expression from alternative promoters and exons, and thereby producing multiple isoforms (Bharti et al., 2008; Hallsson et al., 2000; Hershey and Fisher, 2005; Steingrimsson et al., 2004). We performed RT-PCR analysis to examine *mitf* expression (Figure 3.5A). In adult tissues, the transcripts of both genes were detected in all samples examined, including the ectodermal skin, brain and eye, mesodermal muscle, heart, kidney and spleen, endodermal liver and gut, and male and female gonads (testis and ovary). There was no apparent difference between *mitf1* and *mitf2* in a ubiquitous adult expression pattern. However, the level of *mitf2* expression appeared to be substantially lower than *mitf1*. After 36 cycles of PCR, *mitf1* was easily detectable (Figure 3.5A), whereas *mitf2* was barely detectable (data not shown). *mitf2* became detectable only after 40 cycles (Figure 3.5A). In developing embryos, the transcripts of *mitf1* and *mitf2* were detected as early as 2-cell stage (Figure 3.5A), suggesting their maternal inheritance. The expression of *mitf1* and *mitf2* occurs throughout embryogenesis. Notably, the RNA is clearly more abundant for *mitf1* than for *mitf2* also in developing embryos. Real-time RT-PCR analyses (Figure 3.5B) revealed that the *mitf1* RNA was more abundant than *mitf2* by 3 and 8 fold in the adult skin and eye, and 3 and 2 fold at embryonic stages 16 and 25, respectively. However, at the 4-cell stage, the *mitf2* RNA was more abundant than *mitf1* by 5 fold. Taken together, *mitf1* and *mitf2* show different levels of RNA expression in embryogenesis and adult tissues.
Figure 3.5. RNA expression of mitf1 and mitf2 in medaka. (A) RT-PCR analysis in adult tissues and embryos. The number of PCR cycles is given in parenthesis for each gene. (B) Real-time PCR analysis. β-actin was used as the control for data calibration. Data are means ± s.d (error bars) of five determinations. **, p ≤ 0.01. (C-E) Chromogenic ISH with antisense riboprobes on whole mount embryos at stage 29. Melanocytes (asterisks) in the body, yolk sac (sc) and RPE are positive for tyr and mitf1 but not mitf2. (F and G) Cross-section of the adult posterior trunk after two-color fluorescence ISH with mitf1 (F) and mitf2 (G) antisense riboprobes. ms, muscle; nt, neural tube; sc, spinal cord; sk, skin; vt, vertebra. Scale bars, 100 µm.
3.1.5 Medaka *mitf1* and *mitf2* differ in embryonic melanocytic expression

To study the expression patterns of the medaka *mitf1* and *mitf2* in more detail, we performed in situ hybridization (ISH) in developing embryos at stage 29 (3 days post fertilization; dpf) (Xu et al., 2009), when pigment cells from the neural crest are clearly visible on the body and yolk sac, and pigmented RPE is evident in the eye (Iwamatsu, 2004). To prevent black pigmentation in melanocytes, we treated embryos with phenylthiourea and used them for ISH. We made use of the *tyr* transcript as a melanocyte marker, which encodes tyrosinase, the key enzyme for melanin synthesis in maturing/mature melanocytes (Koga and Hori, 1997). In stage-29 embryos, *tyr*-positive cells were found as individual cells on the embryonic body, yolk sac and in RPE (Figure 3.5C). We found that *mitf1* exhibited a similar expression pattern (Figure 3.5D). However, *mitf2* expression was barely detectable at this stage (Figure 3.5E). Taken together, the medaka *mitf1* resembles the mouse *mitfm* in predominant expression in the pigment lineage, whereas the medaka *mitf2* did degenerate its expression in embryonic melanocytes. This suggests that medaka *mitf1* but not *mitf2* might be a major player for embryonic melanocytes at this critical stage.

We then performed fluorescence ISH to examine the adult expression on cross-sections of the posterior trunk. Both genes exhibited a ubiquitous expression in a wide variety of tissues, including the skin, muscle, neural tube and vertebra (Figure 3.5F and G). We noticed that the *mitf2* signal was relatively stronger than *mitf1* in the spinal cord and
muscle. Hence, \textit{mitf1} and \textit{mitf2} show a generally overlapping expression pattern in adult tissues.

3.1.6 Medaka Mitf2 possesses a reduced melanogenic activity

Mitf controls pigmentation through two distinct aspects: One is to regulate melanocyte formation and the other is to activate the expression of enzymes (e.g. tyrosinase) involved in melanin synthesis (melanogenesis) in maturing/mature melanocytes. The observation that the medaka \textit{mitt2} exhibits an accelerated sequence divergence rate and a lower level of expression provoked us to examine whether this gene has maintained activity. To perform transfection experiments in medaka ES cell line MES1, which is capable of directed melanocyte differentiation following transfection by pXmitf that expresses the \textit{Xiphophorus mitfm} (Bejar et al., 2003), we constructed several expression vectors (Figure 3.6A). pTYRlac contains the 3.3-kb medaka tyrosinase gene a promoter (TYR) ahead of the luciferase (Luc), and pOmitf2 constitutively expresses the medaka \textit{mitf2} cDNA. In undifferentiated MES1 cells, \textit{mitf} expression is not detectable by RT-PCR analysis (Yi et al., 2009), but can be activated by forced XmitfM expression (Bejar et al., 2003). Transfection with pTYRlac alone or together with pCVpf as a control led to a low reporter activity, which was 6 times that of transfection with pTAluc that contains a TATA-box minimal promoter. When pOmitf2 was used for cotransfection with pTYRlac, a 6.5-fold increase in luciferase activity was observed (Figure 3.6B), demonstrating that the medaka Mitf2 maintains melanogenic activity. However, when pXmitf was used for cotransfection, an increase
by up to 178 fold was obtained (Figure 3.6B). Therefore, the melanogenic activity of the medaka Mitf2 is reduced by 27 fold compared to the *Xiphophorus* Mitf2. We then examined the induction of endogenous *tyr* expression by ectopic *mitf* expression in MES1 cells. Upon transfection with pOmitf2 and pXmitf, the *tyr* RNA expression was increased by 38 and 355 fold, respectively (Figure 3.6C). Compared to the *Xiphophorus* Mitf2, the medaka Mitf2 displays a 9.3-fold reduction in its activity to induce the endogenous *tyr* expression. Taken together, the medaka *mitf2* retains melanogenic activity that has considerably been degenerated compared to the *Xmitf*. 
Figure 3.6. Medaka Mitf2 has a reduced activity. MES1 cells were used for cotransfection and analyses. (A) Vectors. (B) Luciferase reporter assay. Each of reporter plasmids pGL3luc, pTAluc and pTYRluc in combination with pRLluc was cotransfected with effector plasmids pOmitf2 or pXmitf into MES1 cells, and Luc activity was determined at day 3 dpt. pRLluc was used as the internal control for calibration of different samples and transfection efficiencies. pGL3luc was used as the negative control, whose activity was considered as background. pTAluc was used as the standard for normalization, whose activity was defined as unit activity. Fold activity was obtained by compared to control cotransfection with pCVpf. Data are means ± s.d of three experiments with triplicates each. (C) Real-time RT-PCR analysis of induced endogenous tyr expression. Data are means ± s.d of five determinations each. (D) Melanocyte formation. Data are means ± s.d of triplicate determinations in 12-well plates. P values are given in (B-D). (E) Micrographs showing the number of melanocytes relative to the number of GFP-positive transgenic cells.
3.1.7 Medaka Mitf2 possesses a reduced melanocyte-inducing activity

We then addressed whether the medaka Mitf2 possessed melanocyte-inducing activity, the original function of the prototype mouse MitfM. To this end, we took the advantage of an in vitro system for melanocyte induction again by using the medaka ES cell line MES1, in which Xmitf expression is sufficient to direct melanocyte differentiation (Bejar et al., 2003). When pOmitf2 was used for transfection, 25 melanocytes were formed per well of a 12-well plate. This number increased to 1298 for pXmitf transfection (Figure 3.6D). This difference by 52 fold is not ascribed to variations in transfection efficiency, because many GFP-positive transgenic cells following pOmitf2 transfection did not become melanocytes, whereas the majority of GFP-positive cells following pXmitf transfection did (Figure 3.6E). Convincingly, the medaka Mitf2 possesses a reduced activity for melanocyte induction in vitro compared to the *Xiphophorus* counterpart.

In summary, the duplicated mitf genes in medaka adopt two distinct directions of evolution: While mitf1 displays conservation, mitf2 shows significant degeneration in protein sequence, expression and activity.

3.2 Complimentary approaches for investigation of AT and AS

3.2.1 Rapid amplification of cDNA Ends (RACEs)

To identify possible transcript variants, 5’-rapid amplification of cDNA end (5’-RACE) and 3’-rapid amplification of cDNA end (3’-RACE) was performed on both medaka mitf1 and mitf2.
For 5’-RACEs of *mitf1*, a total of 385 RACE clones from the cDNA libraries of adult eye (n = 101), testes (n = 48), ovary (n = 70), morula embryos (n = 50), blastula embryos (n = 41), and larvae (n = 65) were sequenced, leading to the identification of 5 major types of *mitf1* variants. For 5’-RACEs of *mitf2*, a total of 221 RACE clones from the cDNA libraries of morula embryos (n = 27), blastula embryos (n = 65), larvae (n = 65) and adult eyes (n = 64) were sequenced. The *mitf2* produced 3 types of cDNA, one of which encodes Mitf2M, an equivalent of the mammalian MitfM. The other two were alternatively spliced. One skipped exon 5b, which is unique in fish but absent in mammals and the other skipped exon 8. They are named *mitf2Δ5b* and *mitf2Δ8*. This proves the existence of AS for *mitf2*. For 3’-RACEs of *mitf1*, a total of 62 clones from the cDNA library of blastula embryos (n = 23) and larvae (n = 40) were sequenced, but only 2 types of cDNA were obtained. One was the original full length, one showed the skipping of exon 5b. This proves the existence of AS for *mitf1*. The results of RACE can be seen as a series of Bernoulli trials when we focus on the detection of a specific RNA variant. In this case, there is a formula to describe the binomial proportion confidence interval which is $p \pm z_{1-0.5\alpha}\sqrt{p(1-p)/n}$ where $p$ is estimated proportion of successes, $z_{1-0.5\alpha}$ is the 1-0.5$\alpha$ percentile of a standard normal distribution, $\alpha$ is the error percentile and $n$ is the sample size (Brown et al., 2001). At a 95% confidence level, $\alpha$ is 0.05 and $z_{1-0.5\alpha}$ is 1.96. At 90% confidence level, $z$ is 1.64. In this way, a sample size of 40 is sufficient to identify an unknown *mitf* RNA variant that is presumed to be 10% of all variants at 95% confidence level.
Increasing total number of clones to 150 can even be able to detect a rare variant whose abundance is only 2% at 90% confidence level. The total RACE clone numbers fulfilled this last stringent requirement but for individual cDNA library, the clone number may not be sufficient for a stringent conclusion. This required the other approach to investigate AT and AS for *mitf* genes: RNA-Seq.

### 3.2.2 RNA-Seq

The other approach to identify possible transcript variants was RNA-Seq. Similar to RACEs, cDNA libraries of adult eye, testes, ovary, blastula embryos and larvae was constructed and sequenced in the NGS platform. As shown in Table 2, for each sample, approximately 47 million 2x100 bp paired-end reads were generated. A total of 235,395,146 reads were obtained. To illustrate the biological meaning of the raw reads, the sequenced reads were mapped to the medaka genome in order to get the region that were aligned, showing the origin of these sequences. Alignments would then be assembled to generate transcripts of genes. After that, transcripts of different sample libraries would be merged into a final transcriptome assembly. A total of 94,515 contigs were assembled from these reads with N50 of 3,673 bp. This final transcriptome assembly was used for annotation of the transcripts and 29,535 unigenes were obtained after filtering. Annotation of these unigenes was carried out using BLASTX searches against the NCBI non-redundant (Nr) protein sequence database. 23,123 (78.28%) of unigenes shared significant similarities with the matched protein sequences with the threshold of $E \leq 1e^{-15}$. This number is larger than
gene number in public database from Ensembl (19,699 coding genes), published transcriptome analysis in other studies (Lai et al., 2015) and databases in the PhyloFish project (Accession: PRJNA255889) suggesting the good quality of the assembly.

<table>
<thead>
<tr>
<th>Category</th>
<th>Eye</th>
<th>Fry</th>
<th>Stage 10</th>
<th>Testis</th>
<th>Ovary</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>total number of reads</td>
<td>47,163,902</td>
<td>47,106,882</td>
<td>46,995,810</td>
<td>47,210,266</td>
<td>46,918,286</td>
<td>235,395,146</td>
</tr>
<tr>
<td>overall length (Gb)</td>
<td>4.716</td>
<td>4.711</td>
<td>4.700</td>
<td>4.721</td>
<td>4.692</td>
<td>23.54</td>
</tr>
<tr>
<td>total number of contigs</td>
<td>68503</td>
<td>62278</td>
<td>49296</td>
<td>68,444</td>
<td>52,814</td>
<td>94,515</td>
</tr>
<tr>
<td>total length (bp)</td>
<td>127,459,048</td>
<td>117,312,055</td>
<td>89,301,686</td>
<td>117,913,851</td>
<td>91,901,838</td>
<td>220,388,675</td>
</tr>
<tr>
<td>N50</td>
<td>2,774</td>
<td>2,699</td>
<td>2,539</td>
<td>2,492</td>
<td>2,492</td>
<td>3,673</td>
</tr>
<tr>
<td>total number of unigenes</td>
<td>28,820</td>
<td>28,523</td>
<td>27,194</td>
<td>28,765</td>
<td>28,396</td>
<td>29,539</td>
</tr>
</tbody>
</table>

Table 2. Summary statistics of the RNA sequencing.

Unlike RACE, the problem of test number was easily overcome due to large number of reads generated by the powerful sequencing platform. With the help of RNA-Seq, situations of exon skipping and alternative promoter usage can be directly seen by exon distributions and alignment of reads in the iGV viewer, which discovers AT and AS variants easily (Figure 3.7). For example, a skipping of exon 5b was observed in both mitf1 and mitf2 and an alternative promoter usage was found in mitf1 while it was not annotated in the known annotation file. In the final transcriptome assembly, a total of 11 variants of mitf duplicated genes were found, 8 for mitf1 and 3 for mitf2. As shown in Figure 3.8, mitf1 has both AT and AS while mitf2 only has AS. The relatively large ‘test’ numbers seemed to explain why RNA-Seq revealed more AS variants of mitf1 than RACEs. Detailed description of these AT and AS will be discussed separately in the following chapters.
Figure 3.7 RNA-Seq discovers AT and AS variants.
With the help of RNA-Seq, situations of (A) Exon skipping and (B) Alternative promoter usage can be directly seen by exon distributions and alignment of reads in the iGV viewer, which discovers AT and AS variants easily. Junctions of the reads are marked as blue lines linking the exons (grey boxes).
Figure 3.8. Medaka *mitf* transcripts revealed by RNA-Seq. By applying RNA-Seq, in total, 11 variants of *mitf* genes were found. *mitf1* has both AT and AS (A) while *mitf2* only has AS (B).
Since counts of the reads can be measured for a particular transcript, the expression level of a particular transcript or a gene can be assessed. For example, Morula embryos and ovary had a similar expression level of *mitf* genes, for both *mitf1* and *mitf2* (Figure 3.9). *mitf2* was expressed dramatically lower than *mitf1* in each organs or development stages examined. This trend is consistent with the results in Figure 3.5 A and B.

For *mitf1* expression, it was lower in early embryonic stages than in the adult organs. For *mitf2*, it was the opposite. However, RNA-Seq has the inevitable problem of false positive and false negative results which cannot be purely eliminated. For example, if we look for a closer examination of a specific *mitf* AT and AS variant, FPKM value was sometimes too small (lower than 0.1) and SE could be large (data not shown), which reduced the reliability of this digital expression and increased the necessity of further validation of AT and AS variants by PCR and cloning. Actually, this problem was to some extent already solved in the former RACEs experiments and only a small fraction of variants validation was needed.

In summary, the combination of RACE and RNA-Seq help to reveal the *mitf* AT and AS variants to a largest extent.
Figure 3.9. Relative expression of *mitf* genes in the RNA-Seq experiments. (A) *mitf1* (B) *mitf2*. Expression level was marked in FPKM.
3.3 Alternative transcription of duplicated fish *mitf* genes

3.3.1 Medaka *mitf1* produces multiple AT isoforms

As already shown in the above chapters, after applying 5′-RACE and RNA-Seq, 5 AT variants were found. According to the predicted protein sequences (Figure 3.10A), the 4 out of these 5 *mitf1* variants were found to encode Mitf1A, Mitf1B, Mitf1H and Mitf1M, equivalents of their mammalian isoforms MitfA, MitfB, MitfH and MitfM, which in human have 118, 83, 20 and 11 amino acid residues in the N-terminus, respectively. The remaining *mitf1* variant had a first exon of 81 bp, but the first exon did not include transcription start site. This first exon has no equivalent in mammals and other vertebrates (Figure 3.11A). Thus, this variant was named *mitf1n* due to the discovery of the novel first exon/promoter. For *mitf2*, 5′RACE and RNA-Seq revealed that the *mitf2* utilize the same first exon. The complete cDNA encodes Mitf2M, an equivalent of the mammalian MitfM. Therefore, the *mitf1*, but not *mitf2*, has evolved/retained the ability to utilize alternative promoter/first exons to produce isoforms as in mammals.

A comparison between cDNAs and genomic sequences revealed a conserved gene structure between both fish *mitf* genes and human gene, with the 4 alternative promoters being similarly positioned in fish and human, with 1n promoter residing between 1b and 1m promoter (Figure 3.10B). A closer examination identified conserved positioning and organization of alternative exon 1a, 1h, 1b and 1m of gene *mitf1* as well as the *mitf2* first exon 1m between fish and human (Figure 3.11B-
F). Taken together, the *mitf1*, but not *mitf2*, shows a highly conserved mechanism of alternative transcription.

**Figure 3.10. Structure and RNA expression of medaka Mitf1 and Mitf2.** (A) N-terminal sequences of medaka Mitf proteins and isoforms. Amino acid sequences of alternative exon 1 are highlighted in color. (B) Genomic organization highlighting alternative exon 1. Numbering of exons (boxes) and sizes of introns (thin lines) are given. Fish genes have lost the human exon 3 but acquired exon 5b. Arrowheads define primers used for RT-PCR analysis. (C and D) RT-PCR analysis in adult organs (C) and developing embryos (D). The number of cycles is shown in parenthesis.
Figure 3.11. Alternative first exons. (A) First exons are highlighted in the genomic organization and the length is labeled. They are named after the similarity of human MITF gene. Black box indicates ORF. (B-F) Exons are shown in capital letter together with single-letter amino acid sequences, introns in small-case letter. (B) Exon 1a. (C) Exon 1h. (D) Exon 1b. (E) Exon 1n. This novel exon has no equivalent in mammals and other vertebrates and does no encode amino acids. (F) Exon 1m. The exon is shown together with its junction intronic sequence and encoded 11 amino acids.
3.3.2 Expression of *mitf1* and *mitf2* AT variants

After identification of *mitf1* AT variants, we were interested in the expression patterns of them. As illustrated in Figure 1B, several combinations of primers were used to detect all the 5 *mitf1* AT variants (common primers spanning exon 2 to exon 5) and specific variants (a specific primer unique to one of the alternative first exons plus a primer in exon 5) by RT-PCR analyses. By using the common primers, the *mitf1* and *mitf2* RNAs were found to be present in all adult organs examined and throughout embryogenesis (Figure 3.10C and D). PCR with specific primers revealed differential expression of *mitf1* AT variants. Among the adult organs examined, *mitf1a* and *mitf1b* exhibited wide distribution and were most abundant in the liver, gut and testes. However, *mitf1a* had a relatively higher expression in the eye while *mitf1b* was expressed higher in the heart. *mitf1h* was limited to the eye, heart, liver and testes, whereas *mitf1m* was easily detectable in the skin and eye (Figure 3.10C). For *mitf1n* was found to be expressed in skin, brain, eye and gonads. Furthermore, these primers were further applied to droplet digital PCR (ddPCR) to obtain a more accurate and quantitative result. Differential RNA abundance in adult organs became more evident upon ddPCR analyses (Figure 3.12). This validated the RT-PCR results. Differential expression was seen also seen in medaka embryogenesis. For instance, *mitf1b* and *mitf2* RNAs are present already at the 4-cell stage and persists throughout subsequent stages of development until hatching, indicating that they are maternally supplied. The *mitf1* RNA variants *mitf1a*, *mitf1h* and *mitf1m* are not detectable until the late
gastrulation onwards (Figure 3.10D), suggesting that they are not maternally provided. Taken together, AT variants encoding the 5 Mitf1 protein isoforms show differential expression in adult and developing embryos.

**Figure 3.12. ddPCR analysis of mitf1 AT variants.** Each variant was colored according to Figure 3.10B. Values are copy numbers per µl normalized to β-actin.

### 3.3.3 Activities of Mitf isoforms on target genes’ promoters

To determine the relative activities of Mitf isoforms as effectors for reporter assays, a total of 6 constructs were made, which express Mitf2M, Mitf1A, Mitf1B, Mitf1H, Mitf1M, and Mitf1N, with the last being the expression of mitf1n which does no encode amino acids in its first exon (Figure 3.13A). The primary function of Mitf is to activate or upregulate the expression of genes involved in pigment cell differentiation/survival and melanin synthesis. One of the best studied Mitf target genes is tyr (Li et al., 2013b), which encodes tyrosinase essential for catalyzing the melanin synthesis from tyrosine. Recently, Mitf has been found to be a
transcriptional activator of germline genes dazl, dnd and vasa (Zhao et al., 2012). Therefore, TYR and DAZ, the promoters of medaka tyr and dazl genes, were linked to luc, resulting in pTYRluc and pDAZluc, were used as two separate reporter systems (Figure 3.13B). For reporter assays, either pTYRluc or pDAZluc was co-transfected with one of the effector plasmids into MES1 cells, and luciferase activity was determined at 3~4 days post transfection (dpt).

Upon cotransfection with pTYRluc, different Mitf proteins and isoforms exhibited considerably varying activities (Figure 3.13C). Specifically, Mtif1B displayed the strongest activity, Mitf1M and Mitf2M exhibited moderate activity, whereas Mitf1A and Mitf1H had little enhancing activity. Interestingly, Mitf1N was comparable to Mitf1M in activity, demonstrating that deletion of the N-terminus does not affect the Mitf1M’s activity on pTYRluc. Upon cotransfection with pDAZluc, Mitf1A, Mitf1B, MitfH and Mitf2M enhanced the activity by 2.5~3.8 fold, whereas Mitf1M and Mitf1N were inefficient in enhancing the activity (Figure 3.13C). Taken together, different Mitf1 isoforms and Mitf2M show dramatic differences in transcriptional activity on the TYR and DAZ promoters, and different N-terminal sequences largely determine the target-specific activity. One obvious exception to this is the difference in activity between Mitf1M and Mitf2M on the DAZ promoter, which may be due to considerable amino acid sequence divergence (61% identity) between the two proteins (Table 1) (Li et al., 2013a) (Li et al., 2013b).
**Figure 3.13. Relative activities of isoforms of Mitf1 and Mitf2.** (A) Effector and selection vectors. CV, the human cytomegalovirus immediate early enhancer/promoter; Omitf1a, b, h, m and n, cDNAs encoding the medaka isoforms Mitf1A, B, H, M and N; Omitf2, cDNA encoding the medaka Mitf2M; pf, cDNA encoding the fusion PF between puromycin acetyltransferase and green fluorescent protein. (B) Reporter vectors. DAZ, medaka dafl promoter; TA, minimal promoter containing a TATA-box; TYR, medaka tyrosinase promoter; luc, firefly luciferase; Rluc, renilla luciferase as the internal control; pGL3-basic, promoterless vector as the negative control. (C) Reporter assays by cotransfection in the medaka ES cell line MES1. The reading from pGL3-basic was defined as the background. The reading from pTAuc transfection was used for normalization. The reading from cotransfection with pCVpf as the control was defined as unit activity for comparison. Relative activities are means ± s.d. (bars above columns) from three independent experiments of transfection in triplicate each. ***, very significant difference at \( p \leq 0.01 \).
Figure 3.14. Induction of melanocyte differentiation and tyr RNA expression by Mitf proteins. (A) Melanocytes induced by Mitf1B, Mitf1M and Mitf2M. Melanocytes are visible as relatively large and flattened cells (asterisks) that are dendritic and rich in melanin granules (black). GFP expression from co-transfected pCVpf served a control for similar transfection efficiencies between experiments. (B) Different activities on melanocyte induction. (C) Effect on tyr RNA expression by real time RT-PCR 4 days after gene transfection. Data are means ± s.d (bars above columns) from three independent experiments in triplicate each. Scale bars, 10 µm.
3.3.4 Induction of melanocytes and tyr RNA expression in ES cells

Previously, we and others have shown that mammalian and Xiphophorus MitfM is sufficient to induce melanocyte formation in mouse fibroblasts (Tachibana et al., 1996), avian neuroretina cells (Planque et al., 1999), medaka ES cells (Bejar et al., 2003) and male germ stem cells (Thoma et al., 2011). The medaka Mitf2M is also capable of melanocyte induction in medaka ES cells (Figure 3.6). This time, further determination of whether Mitf isoforms other than MitfM also possessed this ability was carried out. For this, each of the 6 Mitf-expressing plasmids was introduced together with pCVpf into MES1 cells, and the cells at 7 dpt were observed for melanocyte formation by microscopy (Figure 3.14A and B) and for tyr RNA expression induction by real-time PCR (Figure 3.14C). It was found that not only Mitf1M and Mitf2M but also Mitf1B were capable of melanocyte induction, whereas Mitf1A, Mitf1H and Mitf1N were not (Figure 3.14A and B). Quantitatively, Mitf1M was the strongest of the three Mitf proteins, which induced 22 melanocytes per well, Mitf2M was moderate and produced 10 melanocytes per well, whereas Mitf1B was the weakest and produced only 6 melanocytes per well (Figure 3.14B). We compared the relative activities of the 3 melanocyte-inducing Mitf proteins on inducing tyr RNA expression. Mitf11M and Mitf2M produced 39- and 38-fold induction, whereas Mitf1B exhibited 64-fold induction. Collectively, Mitf isoforms show dramatic differences in the ability and efficiency in the induction of melanocyte differentiation and endogenous tyr RNA expression, and the N-terminus of the isoform Mitf1M is essential for melanocyte-inducing
activity. Furthermore, these data also provide first evidence that the melanocyte-inducing activity of a Mitf protein is associated with but not identical to the melanogenic gene-regulating activity.

3.3.5 Activities of Mitf isoforms on the expression of cell lineage markers

Differential expression and activities provoked us to determine the effect of Mitf1 isoforms as well as Mitf2M and XmitfM on inducing the expression of marker genes of diverse cell lineages (Figure 3.15). The pigmentary lineage comprises solely melanocytes in mammals but 4 distinct pigment cells, namely melanophores, iridophores, xanthophores and leucophores (Kelsh et al., 2004). Upon forced expression, XmitfM, Mitf2M and Mitf1 isoforms - including Mitf1N but except Mitf1H - were capable of inducing to varying extent the RNA expression of tyr and xanthine dehydrogenase (xdh), the later being a putative marker of xanthophores (Ben et al., 2003). Only Mitf1M remarkably activated tyrp1, a gene involved in melanogenesis, whereas Mitf1N exhibited a reduced activity. All the Mitf proteins was able to induce cart1, nkx2.5 and p53, the markers for cartilages (Zhao et al., 1993), cardiomyocytes (Lyons et al., 1995) and general ES cell differentiation (Bejar et al., 2003), but none of them activated pax3a, pax6a and hnf3b, the medaka homologs of markers for neural crest, neuroectoderm and endoderm (Hong et al., 2012; Li et al., 2009). Notably, Mitf1H exhibited the highest activity to preferentially induce actn2 and myhl1, markers for cardiomyocytes and muscles. Interestingly, Mitf1A, Mitf1B, Mitf1H and XmitfM were able to upregulate germline genes dazl and vasa, whereas Mitf1M, Mitf1N and
Mitf2M were not. Taken together, Mitf isoforms show clear differences in activity, specificity and efficiency for regulating gene expression in the pigmentary lineage and others, and the N-terminus plays an essential role in determining the activity and specificity of Mitf isoforms.

Figure 3.15. Mitf isoforms have different activities on lineage gene expression. The ES cell line MES1 in 12-well plates was transfected with pCVpf alone (control) or together with vectors expressing different Mitf proteins and medaka Mitf1 isoforms, grown in the presence of 1 µg/ml puromycin to enrich for transgenic cells and analyzed at day 4 post transfection for gene expression by RT-PCR. RT-PCR was run in a 20-µl volume containing 40 ng of cDNA for 40 cycles and the reaction was separated on 1% agarose gels. Germ layers or cell types and genes are shown to the left, sizes of PCR products to the right. Differentially expressed genes are shown in bold.
3.3.6 Expression and stability of Mitf protein isoforms

The fact that Mitf1 isoforms and Mitf2M upon forced expression exhibit apparent differences in regulating the expression of reporter genes and endogenous genes provoked us to examine whether this were due to differential expression and/or activity of protein isoforms. To this end, we made a series of six constructs that express the 5 Mitf1 isoforms and Mitf2M as a fusion with mCherry and examined the level of protein expression in vitro and in vivo. In vitro, the medaka ES cell line MES1 was used again for transfection with each of the constructs and analyzed by fluorescent microscopy at 2 dpt. As illustrated in Figure 3.16A, the 6 constructs produced a similar efficiency of cells that expressed mCherry. This suggests that the constructs are not different from each other in the level of tagged protein expression, and that the Mitf1 isoforms and Mitf2M appear to have similar stability in medaka ES cells. In vivo, medaka embryos at the 1-cell stage were microinjected with each of the constructs and analyzed at 2 dpf when they reached the gastrula stage (Figure 3.16B). Fluorescent microscopy revealed that the 6 constructs gave rise to a similarly high level of expression of Mitf1 isoforms and Mitf2M in fusion with mCherry (Figure 3.16C), as compared to GFP expression from coinjected plasmid pCVpf as the internal control (Figure 3.16D). Furthermore, microinjected embryos were pooled at the gastrula stage and subjected to Western blot analysis by using monoclonal antibodies against mCherry and GFP. This revealed a similar level of protein amount for each of the Mitf1 isoforms and Mitf2 when calibrated with the GFP signal from coinjected pCVpf (Figure 3.17).
Therefore, both in vivo during embryogenesis and in vitro in cell culture, Mitf1 isoforms and Mitf2M do not differ from each other in expression and/or stability, and differences in activity are largely due to different N-terminal sequences encoded by alternative exon 1 as well as sequence divergence between Mitf1M and Mitf2M.

In summary, one of the duplicated medaka mitf genes, mitf1 is capable of producing multiple RNA variants by alternative transcription, and Mitf isoforms show differential expression and significantly varying activities and efficiencies in regulating the expression of genes important for pigment cells, germ cells and other cell lineages.
**Figure 3.16. Expression and stability of Cherry-tagged Mitf isoforms.** (A) Micrographs showing expression of Mitf isoforms *in vitro*. MES1 cells were transfected with each of plasmids for mCherry-tagged Mitf isoforms plus pCVpf at a 1:4 ratio and were photographed at day 2 post transfection. (B-D) Micrographs showing expression of Mitf isoforms *in vivo*. Embryos at the 1-cell stage were injected with each of plasmids for mCherry-tagged Mitf isoforms plus pCVpf at a 1:1 ratio (total 25 pg/ml) and analyzed at day 2 post fertilization (the mid-gastrula stage). (B) Bright field. (C) mCherry. (D) GFP. Scale bars, 50 µm in (A) and 500 µm in (B-D).
Figure 3.17. Expression and stability of mCherry-tagged Mitf isoforms. (A) Embryos at the 1-cell stage were injected with each of plasmids for mCherry-tagged Mitf isoforms plus pCVpf at a 1:1 ratio (total 25 pg/ml) and analyzed at day 2 post fertilization (the mid-gastrula stage) to perform western blot with anti-mCherry and GFP antibody respectively. (B) Mitf isoform expression was quantified by mCherry level after normalization to GFP level. Bars indicate fold expression of mCherry vs GFP. The relative expression of mCherry in Mitf1A is arbitrarily defined as unit expression (black).
3.4 Alternative splicing of duplicated fish *mitf* genes

Both RNA-Seq and RACEs results proved the existence of AS in duplicate *mitf* genes. RNA-Seq revealed that the *mitf1* produced 3 types of AS variants (Figure 3.18A), one skipped the exon 5b in *mitf1b* and the other 2 had an alternative 3'-splice site in exon 6 which lead to a deletion of a Glutamic acid at position 288 (Mitf1a as standard) for *mitf1a* and *mitf1m*. In this way, the AS variants of *mitf1* were named *mitf1bΔ5b* and *mitf1aΔE288* and *mitf1mΔE288*. Since RACEs failed to discover the *mitf1ΔE288* and there is also possibilities of false discovery in RNA-Seq, further validation was carried out. The first step was to validate the existence of these 2 types of AS. Specific primers that could recognize the skipping of exon 5b and deletion of E288 were designed. In ddPCR, the *mitf1* AS variants were successfully found in adult organs but compared to *mitf1* AT variants, the expression level was quite low (Figure 3.19A). The second step was to further validate the specific AS variant. According to the above result, skin, eye, testes and ovary was chosen for RT-PCR. Primers that can distinguish specific variants were used. As shown in figure 3.20, *mitf1bΔ5b* is clearly seen in the eye but absent in the skin, testis and ovary.
Figure 3.18. Alternative splicing of mitf genes in medaka. The length of exon 5-9 are labeled. Black box indicates ORF. (A) *mitf1*. Three forms of alternative splicing are found. One skipped the exon 5b and the other two had an alternative splice acceptor in exon 6 which led to a deletion of a Glutamic acid. (B) *mitf2*. Two forms of AS were found. One skipped the exon 5b and the other skipped exon 8.
Figure 3.19. Differential expression of *mitf* AS variants.
Relative expression of *mitf*1 (A) and *mitf*2 (B) AS variants are shown. Values are copy numbers per µl normalized to β-actin.

Figure 3.20 Validation of *mitf* AS variants. The RNA of variant mitf1bΔ5b is clearly seen in the eye but absent in the skin, testis and ovary.
Consistent with RACEs results, RNA-Seq revealed that the *mitf2* produced 3 types of cDNA, one of which encodes Mitf2M, an equivalent of the mammalian MitfM(Figure 3.18B). The other two were alternatively spliced. One skipped exon 5b (similar to *mitf1*) the other skipped exon 8. They are named *mitf2Δ5b* and *mitf2Δ8*. To verify the expression patterns of *mitf2* variants to a more complete extent, ddPCR was carried out (Figure 3.19B). Primers that can distinguish *mitf2* AS variants were used. The results showed differential expression of these variants in different organs. *mitf2m* exhibited wide distribution while *mitf2Δ5b* showed expression limited in skin, brain, eye and ovary, and *mitf2Δ8* was uniquely expressed in the gonads.
Chapter 4: Discussion
4.1 Gene duplication and differential evolution

The role of gene duplication in evolution is perhaps best illustrated by the growth hormone (GH)-prolactin (PRL) family (Forsyth and Wallis, 2002). This family consists of GH, PRL, somatolactin (SL) and placental lactogen (PL), which has arisen from a single ancestor by gene duplication followed by extensive functional divergence. Although GH retains the original role in growth control in all vertebrate species, the other members have distinctive functions between fish and mammals: PRL is important for the mammary gland in mammals but for osmoregulation in fish (Sakamoto and McCormick, 2006), SL is fish-specific and selectively regulates proliferation and morphogenesis of pigment cells (Fukamachi et al., 2004), and PL is specific to mammals and essential for placental development. They also exhibit differences in expression pattern: GH, PRL and SL retain the pituitary expression, whereas PL expression occurs in the placenta.
Figure 4.1. Different models of duplicate gene evolution. A WGD event results in the formation of two identical duplicates of every gene. Duplicate genes can undergo different models of evolution that leads to functional divergence. In DDC model, duplicated genes undergo partial and complementary degenerative mutations to perform complementary functions that jointly match the function of a single ancestral gene. This would lead to subfunctionalization. In DCD model, one of the duplicated genes retain the original expression and function, and other undergoes degeneration in expression and/or sequence to acquire an emerging new function or total loss of its original function. In this way, nonfunctionalization or neofunctionalization would take place.
Fish models provide an ideal system to study the models and mechanisms of duplicated genes’ evolution in vertebrates. After split from the tetrapod lineage, fish has undergone a third WGD event in the common ancestor of all modern-day teleosts (Amores et al., 1998). Interestingly, the resultant genome-wide gene duplicates usually coexist, leading to more genes in fish (Meyer and Schartl, 1999). As shown in Figure 4.1, the duplication-degeneration-complementation (DDC) model was originally proposed in zebrafish to explain the preservation of duplicated genes: Duplicated genes undergo partial and complementary retention and degeneration to perform complementary functions that jointly match the function of a single ancestral gene (Force et al., 1999). This DDC model emphasizes that complementary mutations in the regulatory sequences of gene duplicates lead to the retention rather than the loss of gene duplicates following WGD (Force et al., 1999).

Neofunctionalization has recently been reported for the evolution of an antifreeze protein under escape from adaptive conflict in an antarctic zoarcid fish (Deng et al., 2010). In the case of neofunctionalization and nonfunctionalization, one of the duplicated genes must retain the original expression and function, and other undergoes degeneration in expression and/or sequence to acquire an emerging new function or total loss of function. Therefore, the duplication-conservation-degeneration (DCD) may represent a transitory status from gene redundancy to nonfunctionalization or neofunctionalization. The DCD model is elusive to study, because it is often difficult to determine whether a gene under degeneration is indeed a degenerating gene: its remnant expression
and/or activity could obscure observation and experimentation. Consequently, compelling evidence for the DCD model has been absent, and the mechanism underlying this model has remained poorly understood.

4.2 Differential evolution of *mitf* genes

We have recently identified Mitf as a transcriptional activator of medaka germline genes *in vitro* (Zhao et al., 2012), and both *mitf1* and *mitf2* mRNAs as the component of the Balbiani body (Li et al., 2013a), a structure resembling germ plasm that exist transient in growing oocytes (Xu et al., 2010). In chapter 3.1, we have identified the medaka *mitf1* and *mitf2* gene duplicate as the homologs of the mouse *mitf*. Strikingly, the medaka *mitf1* retains expression in the embryonic pigmentary lineage including both melanocytes and RPE, in which the medaka *mitf2* lacks detectable expression. We are suggesting that medaka fits the DCD model for the evolution of duplicated *mitf* genes: In the developing embryo, *mitf1* has retained the original melanocytic expression and perform the ancestral *mitf* function, whereas *mitf2* has undergone degeneration to such a degree that its melanocytic expression becomes hardly detectable at the critical stages. In contrast, zebrafish utilizes the DDC model for the evolution of its duplicated *mitf* genes, where both genes have undergone partial retention and degeneration to jointly display complementary expression and thus execute complementary subfunctions of a single ancestral *mitf* in both melanocytes and RPE. Whether *mitf2* degeneration is towards neofunctionalization or nonfunctionalization remains open.
Concurrent retention and loss of expression in the embryonic pigmentary lineage is a key feature of the DCD model. Since various cell lineages are formed during embryogenesis, loss of expression at critical stages of embryogenesis may exclude an essential role for the degenerating mitf2 in the pigment lineage establishment. Notably, the mitf2 (also named mitfa) promoter can drive transgene expression in adult pigment cells (Schartl et al., 2010).

Besides the loss of expression in the embryonic pigmentary lineage, the medaka mitf2 also shows an expression level that is much lower than the conserved mitf1 in both embryos and adult tissues. We show that reduced mitf2 expression is not compensated by increased melanocyte-inducing and melanogenic activity, but instead, accompanied by a remarkably degenerated melanogenic activity. These facts together with a higher divergence rate of the Mitf2 sequence strongly suggest that mitf2 has already been released from the constraint for pigment development and is in a transitory status towards neofunctionalization or nonfunctionalization. The mouse mitf is also expressed in several non-pigmentary lineages in the form of various variants. A substantially lower expression level would imply that mitf2 has degenerated its role to some extent also in non-pigmentary lineages. It deserves to note that in the case of medaka mitf gene pair, the DCD model operates at multiple aspects, including sequence divergence rate, expression and activity of the protein product. It is unknown whether these aspects act independently or collectively.
Expression of the *Xiphophorus mitf* gene pair has been examined by RT-PCR analysis in adult tissues: *mitfb* shows ubiquitous expression, whereas *mitfm* is expressed only in the eye, skin and melanomas (Altschmied et al., 2002). This suggests that the *Xiphophorus mitf* duplicate also adopt the DCD model: *mitfm* has retained its pigmentary expression, whereas *mitfb* has acquired non-pigmentary expression in the adulthood. It is unknown whether the *mitf* genes also use the DCD model during *Xiphophorus* embryogenesis.

To compare the relative protein activity, we have chosen the medaka Mitf2 and its *Xiphophorus* counterpart Xmitf, the latter being the only fish Mitf that has been studied both for expression in vivo (Delfgaauw et al., 2003) and activity in vitro (Bejar et al., 2003). This comparison has revealed a great reduction in the activity of the medaka *mitf2* protein product, consistent with its accelerated protein sequence divergence indicative of degeneration. Our finding that the wide adult expression of medaka *mitf2* accompanies the retention of melanocyte-inducing activity by its protein product in combination with the fact that Mitf is a melanocyte master regulator raises an interesting question as to why many *mitf2*-expressing cells do not become melanocytes. A low level of expression and a dramatically reduced activity for melanocyte induction may account for an insufficient activity for melanocyte induction in non-pigment cells. This notion is supported by the fact that the expression of highly active *Xmitf* is limited to the pigment cell-containing tissues in *Xiphophorus* (Altschmied et al., 2002). It deserves to note that the
medaka $mitf_2$ has been compared for melanogenic activity and melanocyte induction with the $Xmitf$ in this study.

Our observation that both $mitf_1$ and $mitf_2$ are maternal factors in medaka is intriguing, because in all other vertebrate species examined so far, the $mitf$ gene does not show maternal inheritance and its expression does not commence until late stages of embryogenesis when the pigment cell lineage starts to develop. For example, zygotic $mitf$ expression occurs at the 16-somite stage in zebrafish (Lister et al., 2001; Lister et al., 1999), stage 21/22 in *Xenopus* (Kumasaka et al., 2004), the 5-day-old embryo in chicken (Mochii et al., 1998) and the 9.5-day-old embryo in mouse (Nakayama et al., 1998). Interestingly, $mitf$ maternal inheritance has been described in diverse invertebrate species such as *Drosophila* (Hallsson et al., 2004) and ascidians (Yajima et al., 2003).

Our data illustrate that the $mitf$ gene duplicate may adopt different ways of evolutionary divergence in different species. This species-specific difference has recently reported also to $tyrp1$ encoding tyrosinase-related protein 1 (Typr1). In medaka embryos, $tyrp1a$ expression occurs in the RPE and melanophores while $tyrp1b$ is expressed only in melanophores. In zebrafish embryos, the duplicated $tyrp1$ genes exhibit overlapping expression in the RPE and melanophores, and knockdown of both $tyrp1$ genes leads to pigmentary defects (Braasch et al., 2009b). In this case, medaka adopts the DCD model, and zebrafish uses the DDC model as usual. It will be interesting to determine whether non-pigmentary genes also exhibit such a species-specific difference in evolutionary divergence.
Approximately 30% of zebrafish genes are duplicates (Amores et al., 1998; Postlethwait et al., 1998), their current co-existence may well be explained by the DDC model (Force et al., 1999). Evolution of many other gene duplicates must follow a different way, so that one copy can no longer be identified as a member of the gene pair. This copy may have undergone neofunctionalization or nonfunctionalization. During the past process of neofunctionalization or nonfunctionalization, the DCD model must have operated. Therefore, the DCD model used by the medaka mitf gene may represent an intermediate step for gene loss or neofunctionalization in the past, and towards gene loss or neofunctionalization in the future. During early stages of evolution, the DCD model might be insufficiently distinguishable to the DDC model because of remnant expression and/or function.

4.3 AT and the role of mitf genes

4.3.1 Conserved mechanism of AT of mitf

Mitf is best known for its AT and AS. The human Mitf gene, for example, has 9 alternative promoters and first exons that are separately spliced to exons 2~9 (Hou and Pavan, 2008). A comparison of sequenced genomes has revealed that the conservation of these 9 promoters and exons is limited to mammalian mitf genes (chimpanzee, mouse and rat), with the chicken mitf having only 3 of them (Hallsson et al., 2007). In fish, two mitf genes coexist which are the products of an ancient WGD event, and their protein products are equivalent to 2 of the mammalian isoforms. This has led to a notion that fish has evolved a genomic approach rather than alternative transcription to produce Mitf
proteins corresponding to mammalian Mitf isoforms (Altschmied et al., 2002; Braasch et al., 2009a). In this study (Chapter 3.3), we reveal that medaka mitf1 has at least 5 alternative promoters for generating 5 isoforms, providing first evidence that alternative transcription has also evolved in a fish mitf gene. The Drosophila mitf has two alternative promoters, which - in combination with alternative splicing – can produce up to 6 protein isoforms (http://asia.ensembl.org/Drosophila_melanogaster/Gene/Summary?g=FBgn0263112&db=core ). We conclude that alternative transcription is a conserved mechanism of mitf expression in both invertebrates and vertebrates.

4.3.2 Role of mitf due to differential expression or differential activity

Mitf is a pleiotropic factor in mammals, coincident with wide expression of its transcripts in many types of cells. Differential expression of various Mitf isoforms has best been documented for ubiquitous MitfA and melanocyte-specific MitfM in mouse and human (Hou and Pavan, 2008). In this study, we reveal that the 5 Mitf1 isoforms possess considerably differential expression in adult organs and developing embryos. These results suggest that differential expression and its regulatory mechanism are conserved between fish and mammals.

It has long been argued that the N-terminus may determine the activity of a particular Mitf isoform. It has been reported that the mast cell-specific isoform MitfMC functions differently from the melanocyte-specific isoform MitfM in its ability to activate cell type-specific target genes, because MitfMC functions only on a mast cell target promoter but
fails to activate, and even inhibit, a melanocyte target promoter despite binding to its E-box element (Takemoto et al., 2002). Data obtained in the present study corroborates and extends this report by showing that isoforms Mitf1A, Mtf1B, MitfH and Mitf1M possess different activities on activating target genes specific either to melanocytes or germ cells and on inducing melanocyte differentiation in ES cells. Most convincingly, we reveal that deletion of the N-terminus retains the ability to activate melanogenic expression but loses the capabilities of inducing melanocyte differentiation and activating germline genes’ expression, demonstrating the importance of the N-terminus in determining the activity and efficiency of major Mitf isoforms in different cell lineages.

Our observations that major Mitf1 isoforms show clear differences in expression, activity and specificity offer new insights into the Mitf functions and phenotypes. Of the 5 Mitf1 isoforms analyzed in this study, for example, only Mitf1M is efficient in inducing melanogenesis target genes’ expression and melanocyte differentiation, whereas the remainder has no or little activity. This explains why the mitf transcripts are present in a wide variety of cells, and only a subset of which are or will become melanocytes. For example, we have recently reported that Mitf is a transcriptional activator of germline genes and that the gonadal expression of both medaka mitf genes is specific to germ cells as evidenced by in situ hybridization using riboprobes capable of detecting all RNA variants (Zhao et al., 2012). In this study, we reveal that the mitf1 transcripts in the medaka gonads of both sexes are essentially variants for Mitf1A and Mitf1B, the former being incapable of, and latter being
inefficient, in melanocyte induction in the ES cell differentiation model. It is not surprising that medaka germ cells retain the germ cell identity while express a relatively high level of Mitf expression. This situation also applies similarly to cells in the brain, heart, spleen and liver where the Mitf1M-encoding variant is hardly detectable.

Two organs of mesodermal origin do express the Mitf1M-encoding variant at a level comparable to that in the pigmentary organs (skin and eye). These are the kidney and muscle, which express also variants for Mitf1A and Mitf1B. This allows for two explanations. One is that the level of Mitf1M is insufficient for trans-differentiating muscles and kidney cells into melanocytes. The other is an inhibitory effect of Mitf1A and/or Mitf1B on the melanocyte-inducing activity of Mitf1M. In support of the second possibility is the report that the mast cell-specific isoforms inhibits melanocytic gene expression (Takemoto et al., 2002).

Unlike the mammalian pigmentary lineage comprising solely melanocytes, the medaka pigmentary lineage consists also of iridophores, leucophores and xanthophores besides melanophore (Kelsh et al., 2004). In this study, we observe that Mitf2M and all Mitf1 isoforms but not Mitf1H are capable of activating the expression of xanthine dehydrogenase, a marker for xanthophores. This suggests a possible role for Mitf also in the development of this particular subtype of pigment cells. For zebrafish, the nacre mutation would also lead to reduction of xanthophores (Lister et al., 1999).

Melanocyte development involves melanocytogenesis and melanogenesis. The former refers to melanocytic formation, proliferation
and survival, and latter is a process of melanin synthesis for black pigmentation. Mitf controls both melanocytic and melanogenic events. In this study, we reveal, for the first time to our knowledge, that melanocytic and melanogenic gene-regulating activities of a Mitf protein are separable. For example, Mitf1M, Mitf2M and Mitf1N are comparable in activating the melanogenic tyrosinase promoter but dramatically different in melanocyte induction.

4.4 AS and the role of mitf genes

Recent studies based on RNA-Seq have revealed the prevalence of alternative splicing across many species (Pan et al., 2008; Sultan et al., 2008). This raises a question about the biological meaning of AS. For example, how many AS variants have functions? In the scale of mRNA, many low-abundant AS variants are not conserved, as they are suggested to be a stochastic noise during transcription (Melamud and Moul, 2009). However, in other studies, people found that when an AS event is conserved, its change of mRNA length is much more likely that it can be directly divided by 3 (Xing and Lee, 2005). In this way, the addition or deletion of this part of mRNA will lead to the change of some amino acids while the translation of downstream proteins still remains unaffected. This ‘removable’ part is regarded as a ‘modular’ protein-coding unit. However, the actual roles and mechanism of these modular exons in gene regulation and evolution still needs further demonstration.

A study of AS in human MITF showed that one MitfM AS isoform that are 6 residues shorter than the normal one has differences in DNA binding efficiency (Bismuth et al., 2005). Another AS isoform of MITF-M,
called MITF-MDel, was recently reported to be expressed only in melanocytes and melanoma cells (Wang et al., 2010). This isoform that is spliced at exon 2 and 6 was predicted to be a prognostic biomarker for melanoma. A more recent studies also identified an AS variant in melanoma cells with an 75 bp in-frame deletion in exon 2B for MITF-M (Simmons et al., 2014). In vivo study in the mice suggest an important role of exon 2 in melanoma progression and linking between AS and post-translational modifications (Debbache et al., 2012). All this AS studies are consistent with the concept of ‘modular’ protein-coding unit and are suggesting the role of some particular exons in gene regulation.

In our study, the exon 5b of both mitf1 and mitf2 also fulfilled the situation that the lengths of the skipped exon were 96 and 84, which both can be divided by 3. However, after comparing the genomic organization of human and fish mitf genes, an interesting finding can be drawn that the exon 5b was only obtained in fish mitf genes. In this way, there was no conservation of this exon between human and fish. Another finding was that there was a conservation between medaka mitf genes themselves. These finding suggest a differential evolution of fish and human mitf genes after the WGD event and the exon 5b could be a modular protein-coding unit if we limit this concept only in medaka mitf. If AS evidence of mitf in other fish species could be provided, this phenomena could even likely to be conserved in fish species. For skipping of mitf2 exon 8, it did not fulfil this situation as the length of exon 8 was 148. This means the downstream translation was affected due to non-synonymous mutations for mitf2Δ8 variant.
Apart from exon skipping, other AS events are also being studied. Researchers found that an insertion of long interspersed element-1 in the Mitf gene is associated with altered neurobehavior of the black-eyed white *Mitf*<sup>mi-bw</sup> mouse (Takeda et al., 2014)

### 4.5 Conclusions

This thesis has identified *mitf1* and *mitf2* differential evolution after WGD by examining divergent sequence, differential expression and differential activity. To further investigate into AT and AS of *mitf* genes, RACE (Rapid Amplification of cDNA Ends) and RNA-Seq were employed. We have identified 5 and 3 transcript variants of *mitf1* via AT and AS, respectively. The results of RT-PCR prove differential expression of these *mitf* AT variants in different organs and stages of development. Isoforms predicted from these *mitf1* AT variants differed dramatically in activating target gene expression. Intriguingly, these Mitf1 isoforms possess considerable differences in inducing the expression of multiple cell lineage marker genes. These results demonstrate that AT is a conserved mechanism controlling *mitf* expression in vertebrates ranging from human to fish, and that different Mitf AT isoforms possesses different activities. Interestingly, *mitf2* has a single promoter and thus lacks AT but is able to produce 2 AS variants. We have also proven the differential expression of AS variants in both *mitf1* and *mitf2*. Thus, the duplicated medaka *mitf* genes show a considerable difference in the retention and loss of AT and AS during their evolution. This finding sheds new lights on differential evolution
after gene duplication through AT and AS for increasing genetic complexity.
References


Lai, K.P., Li, J.W., Wang, S.Y., Chiu, J.M., Tse, A., Lau, K., Lok, S., Au,


tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev 9, 1654-1666.


years of next-generation sequencing technology. Trends Genet 30, 418-426.


### Table S1. PCR primers

<table>
<thead>
<tr>
<th>Category</th>
<th>Gene (accession/annotation)</th>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>PCR product (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>mitf gene</strong></td>
<td></td>
<td>m1f1</td>
<td>GTAGATTACATCCGCAAGCTG</td>
<td>501</td>
</tr>
<tr>
<td></td>
<td>(ENSRGLG000000013461)</td>
<td>m1r1</td>
<td>ACAGCCCTGCTCGTTTTCATCCAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>m2f1</td>
<td>AGCTCCGTGACATCTGGCATATT</td>
<td>201</td>
</tr>
<tr>
<td></td>
<td>(ENSRGLG000000003123)</td>
<td>m2r1</td>
<td>GTTGAACCTTGCCCTCGTT</td>
<td></td>
</tr>
<tr>
<td><strong>Neural crest</strong></td>
<td></td>
<td>Sox10-1</td>
<td>AAGCAGACGACGATACACAG</td>
<td>441</td>
</tr>
<tr>
<td></td>
<td>(AJ410868)</td>
<td>Sox10-2</td>
<td>CTGCTCTGTGCTGCTGATG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pax3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ENSRPLT00000011327)</td>
<td>pax3ae2f</td>
<td>TCCCCATGGTGCTGCTGTC</td>
<td>247</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pax3ae4r</td>
<td>ACTCCGCTACATCGAGCTG</td>
<td></td>
</tr>
<tr>
<td><strong>Melanophore</strong></td>
<td></td>
<td>tyra</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ABD010101)</td>
<td>tyrb2f</td>
<td>CGCTGACGAGTAGTACACAT</td>
<td>740</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tyrb3f</td>
<td>TACCGGTATTGCGCCAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trp1b</td>
<td>GATCGTGAGATGATGATG</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>(AF072305)</td>
<td>Trp1c</td>
<td>GCCCGAAGACGTCGAGTA</td>
<td></td>
</tr>
<tr>
<td><strong>Iridophore</strong></td>
<td></td>
<td>prnp4a</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ENSRPLT00000003413)</td>
<td>prnp4af</td>
<td>AAGGACCACATCAACTCC</td>
<td>278</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prnp4ar</td>
<td>ACTCCGCTACATCGAGCTG</td>
<td></td>
</tr>
<tr>
<td><strong>Xanophore</strong></td>
<td></td>
<td>xanthine dehydrogenase, xdh</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ENSRPLT00000001340)</td>
<td>xdhf</td>
<td>TCAGAGTGAGATTGACTG</td>
<td>294</td>
</tr>
<tr>
<td></td>
<td></td>
<td>xxhr</td>
<td>CTTGGAGCAGGAGATGCTT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>csf1r, receptor tyrosine kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ENSRPLT00000006111)</td>
<td>csf1rf</td>
<td>GATCGTGAGATGATGACTG</td>
<td>355</td>
</tr>
<tr>
<td></td>
<td></td>
<td>csf1rr</td>
<td>GTCACAAGACTGTCACAGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>csf1rb</td>
<td>CCTATTGGTGGCTGCTCCCAA</td>
<td>574</td>
</tr>
<tr>
<td></td>
<td></td>
<td>csf1rb</td>
<td>TCCGGAGACGAGGAATCTCC</td>
<td></td>
</tr>
<tr>
<td><strong>Schwan cell</strong></td>
<td></td>
<td>erb3a</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(AF072310)</td>
<td>Erb3f</td>
<td>TTGGGGACAAAGACGACTG</td>
<td>302</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Erb3r</td>
<td>CAAAATGAGTGCCGACCTG</td>
<td></td>
</tr>
<tr>
<td><strong>Cartilage</strong></td>
<td></td>
<td>cart1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ENSRPLT00000003418)</td>
<td>cart1f</td>
<td>CATAGGAAGACGCTGAGG</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cart1r</td>
<td>AGTTGTTCAAGAGACCTG</td>
<td></td>
</tr>
<tr>
<td><strong>Heart</strong></td>
<td></td>
<td>nkh2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(NM_001104912)</td>
<td>nkh2.5f</td>
<td>CGTACATCCTATGTTGGGG</td>
<td>557</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nkh2.5r</td>
<td>TGGTTGAGTGGAGTGGCTG</td>
<td></td>
</tr>
<tr>
<td><strong>Muscle</strong></td>
<td></td>
<td>myosin chain</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ENSRPLG000000018530)</td>
<td>myhf1f</td>
<td>CAAGAACAGCTCGAGAAAC</td>
<td>398</td>
</tr>
<tr>
<td></td>
<td></td>
<td>myh1r</td>
<td>GGACATATGACGCTTACTGACG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>actn2-1</td>
<td>GATTAATGATCTGGTTGAGTTCC</td>
<td>288</td>
</tr>
<tr>
<td></td>
<td>(ENSRPLG00000005187)</td>
<td>actn2-2</td>
<td>CATAGGGCCGAGATTGAATGC</td>
<td></td>
</tr>
<tr>
<td><strong>Tumor suppressor</strong></td>
<td></td>
<td>p53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(OLUS7306)</td>
<td>p53ex6f</td>
<td>CTGGAGAGACGAGGCGAC</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p53cde5R</td>
<td>ATTGCTCTACGCTTCCCA</td>
<td></td>
</tr>
<tr>
<td>Cell Type</td>
<td>Gene</td>
<td>Forward Primer</td>
<td>Reverse Primer</td>
<td>Position</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------</td>
</tr>
<tr>
<td>Germ cell</td>
<td>DazlF</td>
<td>AATACCTCATCATCCCTCACCTC</td>
<td>DazlR</td>
<td>328</td>
</tr>
<tr>
<td></td>
<td>VasaF</td>
<td>CGTCGCCACGTGGTAAGCT</td>
<td>VasaR</td>
<td>486</td>
</tr>
<tr>
<td>Mast cell</td>
<td>TrypF</td>
<td>GCTCCATCGACACAGAAC</td>
<td>TrypR</td>
<td>362</td>
</tr>
<tr>
<td>Neural ectoderm</td>
<td>pax6F</td>
<td>GCGAAGGCCCTATCATTTATC</td>
<td>pax6R</td>
<td>397</td>
</tr>
<tr>
<td></td>
<td>nx3F</td>
<td>CGAACCACCTTCCACACATAC</td>
<td>nx3R</td>
<td>396</td>
</tr>
<tr>
<td>Endoderm</td>
<td>sox17F</td>
<td>ATGAGTAGTCCCGATGCGGG</td>
<td>sox17R</td>
<td>450</td>
</tr>
<tr>
<td></td>
<td>hnf3BF</td>
<td>ATGAGTGCTTGGAGCATTA</td>
<td>hnf3BR</td>
<td>800</td>
</tr>
<tr>
<td>House keeping</td>
<td>β-actinF</td>
<td>TTTCACAGCGCTCCATGTAC</td>
<td>β-actinR</td>
<td>651</td>
</tr>
<tr>
<td>Gene</td>
<td>Primer</td>
<td>Sequence</td>
<td>Size (bp)</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mitf1a</td>
<td>mitf1aQF</td>
<td>TTCAGGACGAGCCTAAAACG</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mitf1aQR</td>
<td>CTTGAAGCTGCTCCCCCTCAC</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>mitf1b</td>
<td>mitf1bQF</td>
<td>GTCAGGTGCCCCGAGCAGTTC</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mitf1bQR</td>
<td>TTTCCCCAGGTATCGTCTCAC</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>mitf1h</td>
<td>mitf1hQF</td>
<td>CTGTCTCCCAGACCAACACA</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mitf1hQR</td>
<td>CTTGAAGCTGCTCCCCCTCAC</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>mitf1m</td>
<td>mitf1mQF</td>
<td>GCTTGAGATGCTCGGATACC</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mitf1mQR</td>
<td>TTTCCCCAGGTATCGTCTCAC</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>mitf1n</td>
<td>mitf1nQF</td>
<td>TTTGACGGTAGATTTCCAACAG</td>
<td>177</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mitf1nQR</td>
<td>TTTCCCCAGGTATCGTCTCAC</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>mitf1</td>
<td>mitf1QF</td>
<td>TCTGATGGACCCAGGACTTC</td>
<td>176</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mitf1QR</td>
<td>CTTGAAGCTGCTCCCCCTCAC</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>mitf2</td>
<td>mitf2QF</td>
<td>GCTTTGCTCACCCCTCAGTTC</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mitf2QR</td>
<td>GGGAGCTGTTGTTGATTGTTT</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>mitf1Δe288</td>
<td>Mitf1de288F</td>
<td>GCTTCCCGTGTTGAAGCTGAG</td>
<td>176</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mitf1dR</td>
<td>CTTGAGATGTTGCTTTTGTTT</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>mitf1Δ5b</td>
<td>Mitf1d5bF</td>
<td>GAATACTCAAGAGCTGAGGTAAGAGC</td>
<td>177</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mitf1dR</td>
<td>CTTGAGATGTTGCTTTTGTTT</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td>mitf2m</td>
<td>Mitf2mF</td>
<td>AACAGCTCTGCAGCATGAG</td>
<td>242</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mitf2mR</td>
<td>CCTGTTCCAGCGCATATCT</td>
<td>242</td>
<td></td>
</tr>
<tr>
<td>mitf2Δ5b</td>
<td>Mitf2d5bF</td>
<td>ATTAACAGAGACTGAGGTTAGAGCTTTT</td>
<td>243</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mitf2mR</td>
<td>CCTGTTCCAGCGCATATCT</td>
<td>242</td>
<td></td>
</tr>
<tr>
<td>mitf2Δ8</td>
<td>Mitf2d8R</td>
<td>AACAGCTCTGCAGCATGAG</td>
<td>235</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mitf2d8R</td>
<td>GCCATGTTCCAGCGCATATCT</td>
<td>242</td>
<td></td>
</tr>
<tr>
<td>β-actin</td>
<td>actinQF</td>
<td>AGAAGAGCTATGAGCCTGCTG</td>
<td>151</td>
<td></td>
</tr>
<tr>
<td></td>
<td>actinQR</td>
<td>AACATCACACTTCATGAGCTG</td>
<td>151</td>
<td></td>
</tr>
</tbody>
</table>