universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's thesis

„Development of a model system to study nuclear OXPHOS gene regulation in the zebrafish“

verfasst von / submitted by

Magdalena Figlmüller, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie UG2002
Betreut von / Supervisor:	Assoz. Prof. Dr. Florian Raible

Statutory Declaration

I declare in lieu of an oath that I have written this master's thesis myself and that I have not used any sources or resources other than stated for its preparation. I further declare that I have indicated all direct and indirect quotations to the best of my knowledge. Should, against my best efforts, a copyright infringement become known, I ask to be notified.

This master's thesis has not been submitted elsewhere for examination purposes.

Vienna, January 15, 2022

Magdalena Figlmüller, BSc

Acknowledgements

I thank Prof. Florian Raible, my internal supervisor at the University of Vienna, for providing helpful feedback before and during the practical part of my thesis internship in Switzerland and in the process of writing this thesis.

A major thank you goes to Prof. Nadia Mercader-Huber for inviting me to join her groups' research efforts at the University of Bern for an extended master's thesis internship. She not only provided helpful and motivating weekly feedback, but also gave me the opportunity to join project meetings with collaborators, encouraged my participation in scientific discussions in- and outside of the lab, and generally made me feel welcome with her supportive, productive, generous, and positive approach to leading the group. I am deeply grateful to my external supervisor Dr. Indre Piragyte-Langa, who patiently shared her knowledge at the bench and significantly furthered my problem-solving skills with her style of supervision. Indre actively encouraged me to develop my own ideas and provided valuable, critical feedback along every step of the way. We frequently worked in parallel and shared and discussed results daily. I am very thankful that Indre and Xavi supported me throughout my stay in Switzerland also outside the lab. Their friendship and fun, caring and inclusive nature made the first year of the COVID-19 pandemic tolerable.

I thank Alex for patiently providing microscopy and Fiji-related support and Xavi and Anna for taking care of the animals. I appreciate how they, and everyone else in the group, including Myra, Inês, Benedetta, Eleonora, Andres, Carolina, Prateek, João, Andrej and Marius, welcomed me warmly, generously provided advice whenever needed and made my time in the lab a nice experience. Furthermore, I am grateful to our collaborators in the HFSP project which this thesis project was a part of, for welcoming me into their group and including me in discussions and meetings. I thank the HFSP project lead Prof. José Antonio Enríquez, whose motivation and confidence in the project were inspiring, and Dr. José Luis Cabrera Alarcón for providing the Ndufb7 variant models included in this thesis. In addition, I thank my brother Michael for his help with digitalizing my thesis illustrations.

On a more personal note, I thank Ali from the bottom of my heart for his continuous emotional support and for proving that long-distance relationships can work well, even in times of global lockdowns and travel restrictions. My gratitude extends to Pooja, the best lockdown-flatmate I can imagine, and I thank my family and close friends in Austria for their love and support.

Herzlichen Dank Euch allen!

Kurzfassung

Eukaryotische Organismen synthetisieren den Großteil der erforderlichen Energie durch oxidative Phosphorylierung (OXPHOS) mithilfe der bigenomisch codierten, mitochondrialen OXPHOS-Maschinerie. Im Normalfall verhindern selektive regulatorische Mechanismen, dass dem Organismus durch das Vorhandensein verschiedener mitochondrialer (mt) OXPHOS-Allele ein Nachteil entsteht. Ob natürliche Variabilität in nuklearen (n) OXPHOS-Genen ähnliche Konsequenzen nach sich zieht, ist jedoch ungewiss. Haben diploide Organismen Selektionsmechanismen entwickelt, um sicherzustellen, dass besser funktionierende Allele vom OXPHOS-System bevorzugt werden? Ziel dieser Masterarbeit war es, ein Tiermodell zu entwickeln, das die Untersuchung Allel-spezifischer nOXPHOS Genkontrolle im Zebrafisch ermöglicht.

RNA-Sequenzierung des heterozygoten Zebrafisch-Stammes WIK deutete auf eine ungleiche Expression der Allele des nOXPHOS Proteins Ndufb7 hin, was Ndufb7 zu einem vielversprechenden Testobjekt macht. Um die biologische Funktion des hochempfindlichen OXPHOS-Systems möglichst wenig zu beeinflussen, gleichzeitig aber die experimentellen Möglichkeiten des Tiermodells zu maximieren, wurden drei verschiedene Strategien getestet: (1) Farbcodierung der beiden *ndufb7*-Allele durch unterschiedliche Fluorophor-Tags, (2) Unterscheidung zwischen den beiden Allelen durch kurze Epitop-Tags und (3) Mutagenese von *ndufb7* (K84>W) um eine kürzlich identifizierte, heterozygote Patientin mit schwerem neuropathologischen Phänotyp zu modellieren.

Im Rahmen dieser Masterarbeit konnte ich zeigen, dass alle drei Strategien technisch durchführbar sind. Meine Ergebnisse legen jedoch nahe, dass Epitop-Tags und *in vivo*-Mutagenese einer Fluorophor-Markierung zur Untersuchung des nOXPHOS-Gens *ndufb7* überlegen sind. Darüber hinaus identifizierte ich bisher unbeschriebene, Zebrafisch-stammabhängige Unterschiede in der Überexpression von synthetischer mRNA. Zusammenfassend liefern die hier beschriebenen Ergebnisse wertvolle Einblicke in technische Aspekte der Tiermodellentwicklung, die über den Kontext der Untersuchung von nOXPHOS-Genregulation im Zebrafisch hinaus von Relevanz sind.

Abstract

Eukaryotic organisms produce most of the energy required for survival through the bigenomically encoded mitochondrial OXPHOS machinery. While the presence of several mitochondrial (mt)OXPHOS alleles could be detrimental to the organism, tight regulatory machinery is present to limit such outcome. On the other hand, the consequence of natural variability in nuclear (n)OXPHOS genes remains undetermined. Have diploid organisms evolved selection mechanisms to ensure the best fitting alleles are preferred in the OXPHOS system? The aim of this thesis project was to create tools to investigate allele-specific nOXPHOS gene control in the zebrafish model.

RNA sequencing revealed skewed allelic expression of the nOXPHOS subunit Ndufb7 in the heterozygous zebrafish WIK strain, thus making it a promising target. Aiming to minimize the risk of interference with the biological function of the highly sensitive OXPHOS system while maximizing experimental possibilities, I tested the feasibility of three different strategies: (1) color-coding the two *ndufb7* alleles by differential fluorophore-tagging, (2) differentiating between the two alleles by short epitope-tagging, and (3) mutagenizing *ndufb7* (K84>W) to model a recently identified human patient with severe neuropathological phenotype.

While all three strategies are technically feasible, my results suggest that short epitope-tagging and *in vivo* mutagenesis are superior to fluorophore-tagging when studying nOXPHOS gene expression. Furthermore, I identified previously undescribed strain-dependent differences in the overexpression of synthetic mRNA. In conclusion, the findings described here provide valuable insights into technical aspects of animal model development that may be relevant beyond the context of studying nOXPHOS gene regulation in the zebrafish.

Table of Contents

Table of Figures and Illustrations

List of Abbreviations

ADP / ATP	adenosine diphosphate / triphosphate
AB strain	pet shop A line x B line zebrafish strain
BN-PAGE	Blue Native Polyacrylamide Gel Electrophoresis
bp	base pairs
BSA	Bovine Serum Albumin
cDNA	complementary DNA
CDS	coding sequence
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats / Cas9 Nuclease
Cryaa	crystallin alpha a
DAPI	4′,6-diamidino-2-phenylindole
ddG	delta delta Gibbs free energy
dpf	days post fertilization
DSB	double strand break
eGFP	enhanced green fluorescent protein
F0 / F1 / F2	parental / first filial / second filial generation
FLAG	DYKDDDDK peptide sequence
gDNA	genomic DNA
gnomAD	genome aggregation database
gRNA	guide RNA
H28	wildtype Ndufb7 variant (amino acid 28: Histidine)
HA	Hemagglutinin
HDR	homology directed repair
het.	heterozygous

hom.	homozygous
hpf	hours post fertilization
IMS	intermembrane space
IQR	interquartile range
K84 variant	wildtype Ndufb7 variant (amino acid 84: Lysine)
KI	knock-in
L1	Linker peptide 1 (GGG)
L2	Linker peptide 2 $(GGGGS)_3$
L3	Linker peptide 3 $(EAAAK)_3$
mCerulean	monomeric Cerulean fluorophore
mCitrine	monomeric Citrine fluorophore
mtDNA	mitochondrial DNA
mEGFP	monomeric enhanced green fluorescent protein
mtOXPHOS subunit	mitochondrial encoded OXPHOS subunit
mRNA	messenger RNA
mRuby2	monomeric Ruby2 fluorophore
mTurquoise2	monomeric Turquoise2 fluorophore
N28	SNP Ndufb7 variant (amino acid 28: Asparagine)
NADH	nicotinamide adenine dinucleotide hydrogen
nDNA	nuclear DNA
N-module	NADH oxidation module of OXPHOS complex I
nOXPHOS subunit	nuclear encoded OXPHOS subunit
OXPHOS	oxidative phosphorylation
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction

P_D / P_P module	distal / proximal proton-pumping module of OXPHOS complex I
pLoF	predicted loss of function
Q-module	ubiquinone reduction module of OXPHOS complex I
R84 variant	humanized Ndufb7 variant (amino acid 84: Arginine)
r.m K84 variant	reverse mutagenized Ndufb7 variant (K84 wildtype)
RNA-seq	RNA sequencing
RNP	ribonucleoprotein
ROS	reactive oxygen species
RT	repair template
sgRNA	single guide RNA
SNP(s)	single nucleotide polymorphism(s)
ssODN	single stranded oligodinucleotide
Tol2	Tol2 transposase
T7E1	T7 Endonuclease I
TU strain	Tübingen zebrafish strain
Ubb	ubiquitin b
UTR	untranslated region
W84 variant	patient Ndufb7 variant (amino acid 84: Tryptophan)
WIK	Wild Indian Karyotype zebrafish strain
wt	wildtype

1 Introduction

1.1 The critical role of multiprotein complexes for oxidative phosphorylation

Oxidative phosphorylation (OXPHOS) is one of the processes by which cellular energy is produced in the form of ATP, the universal energy currency of all living organisms. It has been estimated that the mass of ATP turned over by an average person per day amounts to 50 - 65 kg (Rich, 2003; Alberts *et al.*, 2014). Cells employ several other pathways for ATP production besides OXPHOS, including glycolysis, fermentation, the Krebs cycle, and fatty acid oxidation. The complete breakdown of one glucose molecule, for example, yields a net gain of approximately 30 ATP molecules by gradual oxidation (Alberts *et al.*, 2014). Out of these, two ATPs are gained from glycolysis in the cytoplasm and a further two GTP / ATP molecules are generated through the Krebs cycle in mitochondria. The rest (approximately 90 %) of ATP molecules are generated further along the chain of interconnected pathways by mitochondrial oxidative phosphorylation (Lane and Martin, 2010; Alberts *et al.*, 2014). Mitochondria consist of an outer and inner membrane which, separated by an intermembrane space (IMS), enclose the mitochondrial matrix. The inner membrane is enlarged and heavily folded, forming invaginations known as cristae that accommodate the five enzymatic complexes, which collectively perform oxidative phosphorylation (Pfanner, Warscheid and Wiedemann, 2019).

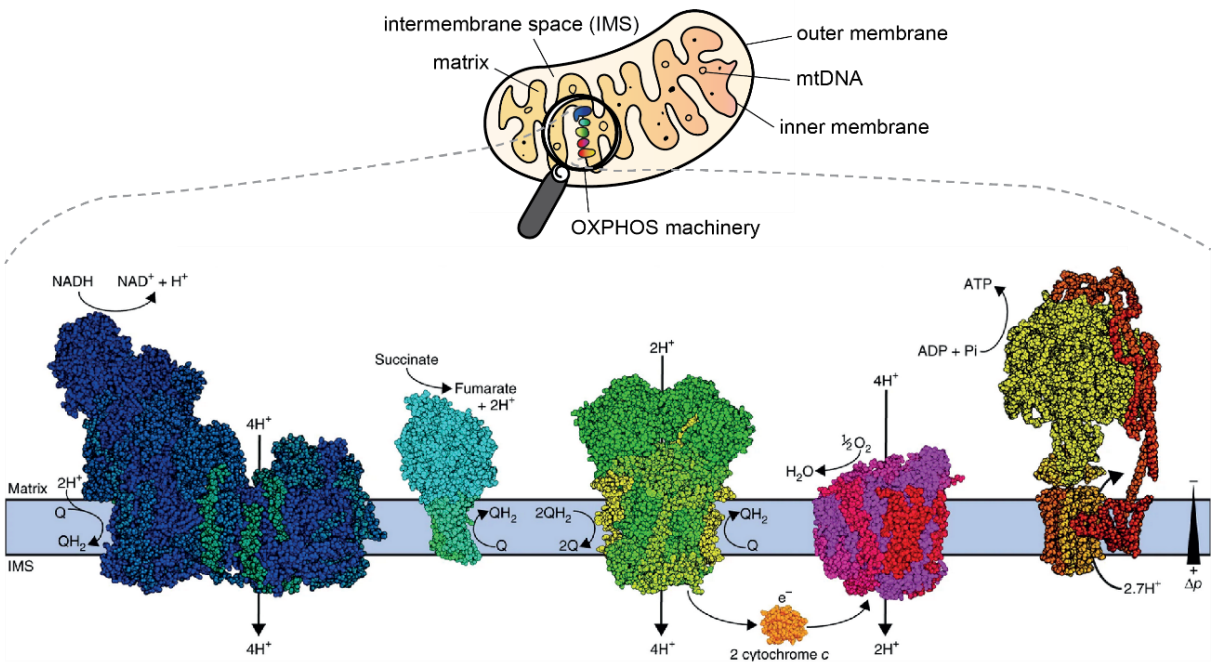
Essentially, energy is produced by electron flow from the mitochondrial matrix across three proton pumps (complexes I, III and IV) into the mitochondrial intermembrane space, thus establishing a proton gradient across the inner mitochondrial membrane (Mitchell, 1961). The potential energy stored in this electrochemical proton gradient is harvested by complex V, an ATP synthase which uses it to generate ATP from ADP and inorganic phosphate by rotary catalysis (Kühlbrandt, 2019). A fifth player, complex II, serves as a second independent entrance point of electrons into the respiratory chain. Two electron carriers, ubiquinone and cytochrome c, furthermore contribute by transferring electrons between the complexes I, II, III and IV. The electron transport chain is fueled by NADH or succinate which are the products of glycolysis, fatty acid oxidation and the citric acid cycle (Chaban, Boekema and Dudkina, 2014; Pfanner, Warscheid and Wiedemann, 2019).

As depicted in **Figure 1**, each OXPHOS complex consists of several individual subunits encoded in most cases by both the nuclear (nDNA) and the mitochondrial (mtDNA) genome:

- *Complex I* (NADH dehydrogenase) consists of 45 subunits, 7 mtDNA and 38 nDNA encoded, and is the largest enzyme of the electron transport chain with a total mass of about 980 kDa

(Carroll *et al.*, 2006). It binds NADH substrate, transfers two electrons to the carrier molecule ubiquinone and by doing so pumps four protons into the intermembrane space.

- *Complex II* (succinate dehydrogenase) comprises 4 subunits, all of which are encoded in the nucleus. The 124 kDa (Sun *et al.*, 2005) complex oxidizes succinate and transfers electrons to ubiquinone.
- *Complex III* (cytochrome c oxidoreductase) contains 11 subunits, only one of which is mitochondrially encoded. It is present in the mitochondrial membrane as an approximately 490 kDa dimer (Xia *et al.*, 1997). By oxidizing ubiquinone to ubiquinol, complex III pumps two protons across the inner mitochondrial membrane. The electrons from ubiquinol are then passed on to the carrier molecule cytochrome c.
- *Complex IV* (cytochrome c oxidase), approximately 200 kDa in size, is formed from 13-14 subunits, 3 of which are encoded by mtDNA (Wittig, Braun and Schägger, 2006; Zong *et al.*, 2018). It serves as electron acceptor from cytochrome c and uses the electrons to convert an oxygen molecule into two water molecules. At the same time, four protons are pumped into the mitochondrial intermembrane space.
- *Complex V* (ATP synthase) finally utilizes the protons pumped into the intermembrane space by complex I, III and IV to convert ADP into ATP. The 600 kDa complex consists of approximately 16 subunits: 2 mtDNA and 14 nDNA encoded (Stock *et al.*, 2000).



	Complex I	Complex II	Complex III	Complex IV	Complex V
Number of subunits:	ca. 45	4	11	13	16
mtDNA encoded:	7	0	1	3	2
nDNA encoded:	38	4	10	10	14

Figure 1. Five jointly encoded multiprotein complexes facilitate oxidative phosphorylation (OXPHOS) in mitochondria. Whereas all five OXPHOS complexes are located in the inner mitochondrial membrane (top), they consist of several subunits that are encoded either by mitochondrial (mt) or nuclear (n) DNA (bottom). Image adapted from *Letts and Sazanov, 2017.*

The formation of each of these complexes as well as the correct integration into the inner mitochondrial membrane happens in a series of carefully orchestrated steps, which involve a variety of assembly factors (Ghezzi and Zeviani, 2012; Formosa and Ryan, 2018; Formosa *et al.*, 2018; Signes and Fernandez-Vizarra, 2018). Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE) revealed that some OXPHOS complexes are not only present as monomers or dimers but also form supercomplexes with one another ($I + III_2$, $III_2 + IV_{1-2}$ or $I + III_2 + IV_{1-4}$) (Wittig, Braun and Schägger, 2006; Chaban, Boekema and Dudkina, 2014). Interestingly, complex II was only rarely reported to be involved in supercomplex formation (Acín-Pérez *et al.*, 2008; Guo *et al.*, 2017), probably due to its direct participation in the citric acid cycle (Dudkina *et al.*, 2010; Rutter, Winge and Schiffman, 2010).

1.2 Mitonuclear interactions, co-adaptation and incompatibilities

As outlined above, while most OXPHOS subunits are encoded by nuclear DNA, some of the "core" proton-translocating subunits are transcribed from the mitochondrial genome. To ensure functionality of the entire OXPHOS system, the products of the two genomes must therefore be able to properly interact (assemble) with each other. Hence, efficient respiration requires a suitable combination of mitochondrial and nuclear (mitonuclear) genotypes, which presumably facilitates co-adaptation between the two genomes (Bar-Yaacov, Blumberg and Mishmar, 2012; Hill *et al.*, 2019).

The mutation rate of mitochondrial DNA is much higher than that of nuclear DNA in many eukaryotes. This is due to several reasons, including: 1) reactive oxygen species (ROS) causing more damage to mtDNA than nDNA (Van Houten, Hunter and Meyer, 2016), 2) the absence of nucleotide excision repair, a robust nuclear repair system, in mitochondria (Kazak, Reyes and Holt, 2012), 3) the mtDNA polymerase Pol γ having very little translesion synthesis capacity compared to some nuclear DNA polymerases (Kasiviswanathan *et al.*, 2012), and 4) the rare occurrence of recombination in mitochondria (Brown, George and Wilson, 1979; Lynch, 1997; Neiman and Taylor, 2009; Havird and Sloan, 2016). Therefore, organisms have evolved several mechanisms to counteract the risk of damage induced by mtDNA variability, one of which is a drastic decrease in the number of mtDNA molecules during the genetic bottleneck in oogenesis (Burr, Pezet and Chinnery, 2018). Some reports suggest that mitochondria remain active within primordial germ cells and developing oocytes (Kasashima, Nagao and Endo, 2014; Hayashi *et al.*, 2017; Floros *et al.*, 2018), which possibly enables natural selection by screening oocytes based on their mitonuclear compatibility and metabolic integrity (Hill *et al.*, 2019). A second mechanism of selection is the elimination of paternal mitochondrial DNA during early stages of development: whereas nuclear OXPHOS genes are inherited biparentally, mitochondrial DNA is predominantly maternally inherited, resulting only in viable offspring if the combination of

paternal and maternal genetic material is compatible. The selection for mitonuclear compatibility is thought to continue throughout development after fertilization (Latorre-Pellicer *et al.*, 2016).

Mitonuclear incompatibilities impact overall fitness in a plethora of eukaryotic organisms ranging from yeast over animals such as nematodes, fruit flies, fish and mammals, to plants (reviewed in Lane, 2011; Nguyen *et al.*, 2020). Several studies suggest that allele specific selection mechanisms may be in place to increase overall fitness. For example, it was demonstrated in marine copepod that high fitness of F2 hybrids (characterized as increased developmental and ATP synthesis rate) is associated with biased selection of the maternal allele. This allelic bias was not observed in low-fitness hybrids, thus suggesting that selection of the nuclear allele matching the mitochondrial genome represents an advantage to overall fitness of the organism (Healy and Burton, 2020). Similarly, it was shown that the co-occurrence of missense SNPs in nuclear and mitochondrially encoded genes can negatively affect OXPHOS function in *Drosophila*. By using mitochondrial-nuclear *D. simulans-D. melanogaster* hybrid strains, Meiklejohn and colleagues observed disrupted mitochondrial protein synthesis, followed by decreased OXPHOS activity, caused by disturbed interaction between a nuclear encoded tRNA synthetase and the corresponding mitochondrially encoded mt-tRNA^{Tyr} when each of the two players harbors a particular SNP. Consequentially, the affected hybrids suffered from decreased fitness due to disturbed development, metamorphosis and a decrease in female fertility (Meiklejohn et al., 2013). These observations were confirmed by more recent studies focused on *Drosophila* oogenesis, embryonic survival (Zhang et al., 2017) and increased ROS production during ageing (Pichaud et al., 2019) as a consequence of said mitonuclear mismatch. Interestingly, mtDNA variation was found to affect the transcription of nuclear encoded mitochondrial genes differently depending on the diet, thus suggesting a retrograde signaling effect in transcription regulation that is at least partially influenced by the environment (Mossman et al., 2016).

Evidence for mitonuclear incompatibility was also found in several mouse models hosting mitochondria from distant genetic backgrounds, resulting in altered ROS levels, interferon immune response, cellular metabolism, and susceptibility to cardiac stress (Betancourt et al., 2014; Dunham-Snary and Ballinger, 2015; Fetterman et al., 2013; Latorre-Pellicer et al., 2016; Lopez Sanchez et al., 2020). Using conplastic mice, Latorre-Pellicer *et al* demonstrated that compatibility of nuclear and mitochondrial genomes affects OXPHOS function, and that segregation of mtDNA is driven by the nuclear genetic context (Latorre-Pellicer et al., 2019). Mitochondrial DNA heteroplasmy is sensed during oocyte maturation and embryo development in the mouse (Latorre-Pellicer et al., 2019). Similarly, a study on human mitochondrial DNA diversity found that selective forces act within the female germ line, under the influence of the nuclear genome (Wei et al., 2019). It is therefore plausible that allele-specific pre-selection of nOXPHOS subunits may occur in the female germ line.

Mitonuclear incompatibilities are sometimes not revealed until the F2 generation, depending on the dominance between different alleles, potentially because F1 offspring receive a haploid set of maternal nuclear genes that functions well with its mitochondrially encoded counterparts (Turelli and Orr, 2000; Burton, Ellison and Harrison, 2006; Raj *et al.*, 2010; Hill *et al.*, 2019). It is currently unknown whether this is, because the stringently pre-selected maternal gene variants simply mask the effects of less-functional nuclear variants provided by the male, or because there are additional mechanisms in place that ensure allele-specific utilization of the better suited nuclear encoded OXPHOS variants. While several studies reported changes in nuclear encoded OXPHOS genes that evolved complementary to sequence changes in their respective mitochondrially encoded interaction partners (Osada and Akashi, 2012; Barreto and Burton, 2013; Sloan *et al.*, 2014; Van Der Sluis *et al.*, 2015; Barreto *et al.*, 2018; Yan, Ye and Werren, 2018), this may be the exception.

1.3 Studying genetic variability of nuclear DNA

Depending on its maternal and paternal genetic background, a diploid organism may carry two different variants of one and the same nuclear gene (Wright, 2005). In a study of 2,504 individuals from 26 populations, the 1000 Genomes Project reported that a typical human genome differs from the reference genome at up to 5 million sites. According to the authors of the study, some of these differences can be explained by structural variants such as large deletions, insertions, or copy-number variants. The vast majority (>99.9 %) of variants, however, is attributed to short indels and single nucleotide polymorphisms (SNPs) (Auton *et al.*, 2015).

While some SNPs that are located in protein-coding regions of the genome remain silent, others induce changes in amino acid sequence, which in turn may alter the function of the corresponding gene product. In extreme cases (for example if a recessive, disease causing SNP is homozygous), this can cause detrimental effects on health, as has been reported in a variety of disease cases where nuclear encoded subunits of the OXPHOS system were affected (Bourgeron *et al.*, 1995; Birch-Machin *et al.*, 2000; Barel *et al.*, 2008; Haack *et al.*, 2012; Friederich *et al.*, 2017). It is virtually impossible to exert necessary controls over factors such as diet, lifestyle and environment in the study of diseases caused by such genetic variations in a human setting. Therefore, researchers have long utilized animal models to study genetic determinants underlying normal biological processes and disease (Balik-Meisner *et al.*, 2018).

1.4 The zebrafish: a small, partially transparent model organism with high genetic variability

The zebrafish (*Danio rerio*) is an attractive model system, in part due to rapid development and transparency of embryos, high fecundity rates and its remarkable regenerative capacities (Gemberling *et al.*, 2013; Marques, Lupi and Mercader, 2019). As vertebrate model with a large set of protein-coding genes that are >70 % orthologous to human genes, the zebrafish has proven useful for studying developmental processes and human diseases (Howe *et al.*, 2013). Notably, there is a large abundance of SNPs (5-15 million) segregating in zebrafish populations (Balik-Meisner *et al.*, 2018). The considerable genetic variance between different zebrafish strains (Guryev *et al.*, 2006) makes it a suitable model system for studying the expression of nOXPHOS variants. Further advantages include: 1) the ability to work with high numbers of offspring (50 to 300 eggs per adult female per week) throughout the year, 2) the possibility to use non-invasive (live-) imaging techniques to detect and follow the impact of a modified gene product in the transparent embryos, 3) the lower amount of space and funds required to maintain zebrafish compared to rodents, and 4) the external fertilization and rapid development which readily enable the examination and manipulation of zebrafish embryos from the moment of fertilization (Grunwald and Eisen, 2002; Burke, 2016).

2 Research Aims

As part of an HFSP-funded, collaborative research effort, this master's thesis project was designed to help generate a zebrafish model suitable for subsequent investigation of the main questions asked by the collaborators. The main goal of this thesis project was the development of a suitable zebrafish model with the help of genetic engineering tools. To put the demands on animal model design into context, this section first outlines the main questions asked by the research collaborators (overall HFSP project aims), before introducing the aims regarding model design that are specific to this thesis project (thesis goals) at the end of the chapter.

Eukaryotic organisms store most of the energy required for survival in the form of ATP. As described in **section 1.1**, the vast majority of ATP is produced by the mitochondrial OXPHOS machinery, which is encoded by both the nuclear and the mitochondrial genome. To ensure the functionality of this machinery, it is crucial that the biparentally inherited nOXPHOS gene products interact with the maternally inherited mtOXPHOS gene products to form functional multi-subunit complexes of the respiratory chain. Considering the above-described extent of natural genetic variability (**sections 1.3** and **1.4**), the occurrence of SNPs in nOXPHOS genes is not surprising. Introducing variations into the tightly regulated OXPHOS system, however, may affect the ability of individual subunits to interact adequately with one another and therefore poses a potential threat to the fitness of the organism (Stoldt *et al.*, 2018).

This master's thesis is embedded in a larger (HFSP funded) research project that tries to enhance our understanding of how nature deals with the inevitable genetic variability caused by biparental inheritance in diploid organisms. The HFSP project aims to address the following research questions:

Are nuclear encoded OXPHOS subunits regulated in an allele-specific manner in diploid organisms?

Have diploid organisms evolved a mechanism to ensure functionality of the OXPHOS system when different variants of nOXPHOS subunits are present? Or in other words: if the allelic variant inherited from one parent brings a functional benefit in combination with its mitochondrial and nuclear encoded interaction partners compared to the variant inherited from the other parent, is it preferentially utilized by the organism?

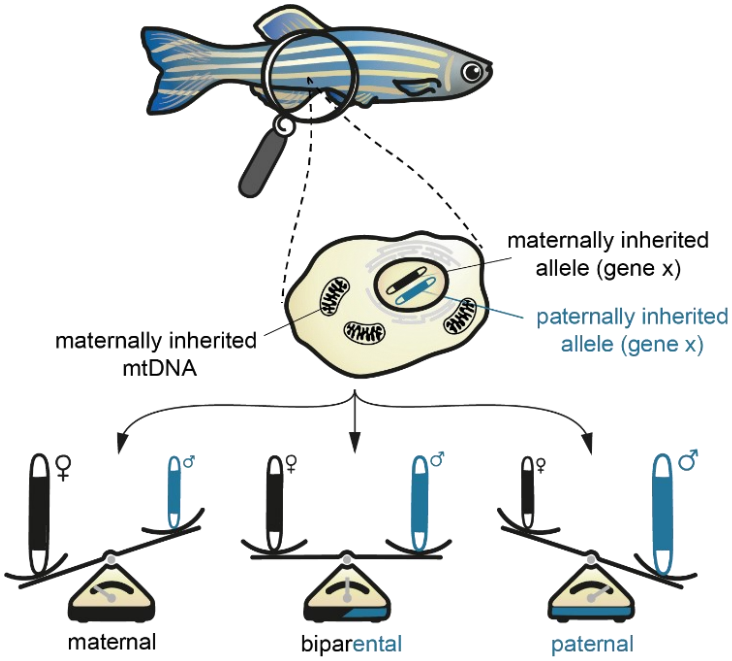


Figure 2. Possible allelic tendencies of nuclear gene expression. In diploid organisms, the utilization of nuclear gene products can in theory be balanced (biparental) or biased towards expression of the maternally (black) or the paternally (blue) inherited allele.

In diploid organisms such as the zebrafish, there are three possible allelic selection scenarios: in the absence of allele specific selection, gene expression from the maternally and paternally inherited allele (including utilization of the resulting proteins) is equally balanced. However, if a selection mechanism exists, gene expression and/or protein levels are skewed towards utilization of the maternal or the paternal allele, respectively (**Figure 2**). In the context of this thesis, the term “gene expression” refers to the “production of an observable phenotype by a gene — usually by directing the synthesis of a protein” as previously defined (Johnson, Lewis and Et Al., 2002). Thus, it encompasses all intermediate processes between transcription and the final processing steps required for the generation of a functional protein.

There are several levels at which allele-specific selection of one nOXPHOS gene variant over its allelic counterpart may take place within a eukaryotic cell:

1) At the DNA level: by means of chromatin or transcription control in the nucleus.
2) At the mRNA level: by control of mRNA processing, nuclear export, mRNA stability or translation.
3) At the protein level: by control through posttranslational processing, mitochondrial import, protein folding, OXPHOS complex assembly or protein degradation.

The levels of gene expression control at which SNP-dependent variant selection of a nOXPHOS subunit may happen in the process from transcription to OXPHOS complex I assembly are depicted in **Figure 3**.

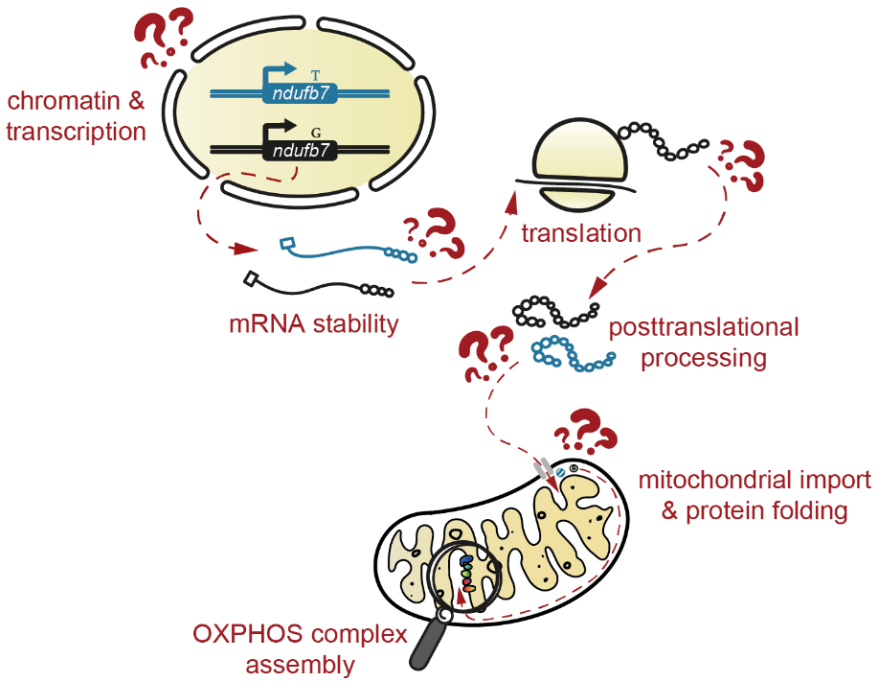


Figure 3. Potential levels of gene expression control of the nOXPHOS gene *ndufb7* up until integration into the OXPHOS machinery in mitochondria. The colors black and blue represent the two allelic variants of *ndufb7*, which differ from each other by a single nucleotide polymorphism (SNP: G-T).

Higher eukaryotic organisms frequently utilize intracellular control mechanisms differently depending on cell type, tissue type or organ, as has for example been demonstrated by tissue specific transcription profiling (Campbell and Marlow, 2013; Mella-Alvarado *et al.*, 2013; Lin *et al.*, 2014; Cokus *et al.*, 2019). Furthermore, according to a recent report by Lechuga-Vieco and colleagues, mtDNA haplotypes are selected in a cell type-specific manner. We, the below described group of researchers, therefore hypothesize that some SNPs in nuclear encoded OXPHOS genes may be selected for or against differentially depending on the cell type, as is the case for mtOXPHOS variants during mtDNA haplotype selection (Lechuga-Vieco *et al.*, 2020).

In an international collaborative effort between four research groups, the HFSP funded project introduced here aims to answer the research question by utilizing various model systems: human diploid and haploid cell lines, a mouse model (at whole organism and single cell level), structural computational modelling and a zebrafish model. Depending on the genetic background (SNP profile) of the different model organisms, each group studies a distinct set of nOXPHOS gene variants. The aim of this thesis project is to validate the feasibility of using zebrafish to address the research question by utilizing a naturally occurring SNP in Ndufb7, a nuclear encoded accessory subunit integral for functional assembly of OXPHOS complex I (Stroud *et al.*, 2016).

To enable studying allele-specific nOXPHOS gene expression, an established zebrafish model would ideally fulfill two main requirements. First, it should allow differentiation between the two alleles of a nOXPHOS gene at all three levels of gene expression control (DNA, mRNA and protein) to facilitate detection of potential allelic selection mechanisms. Second, to unveil

potential cell type or tissue-specific control mechanisms, the model should enable straightforward visualization of the two allelic variants by imaging or similar techniques in a cell type and/or tissue-specific manner. However, alteration of an OXPHOS subunit, for example by introduction of an artificial tag, may negatively influence its function or interaction with other subunits during assembly of the highly sensitive OXPHOS machinery. Thus, uncertainty whether a tagged Ndufb7 subunit would remain functional, together with potential disturbance of the OXPHOS system, are major concerns.

In this master's thesis, I set out to help with the development of such a model by contributing to three independent strategies to enable differentiation between the two alleles of the nOXPHOS gene *ndufb7* in zebrafish. Each of these strategies addresses different aspects of the above listed requirements and/or concerns. The main goal of my thesis project was to investigate and compare the feasibility of the following three strategies (depicted in **Figure 4**):

- **Strategy I:** Differential **fluorophore-tagging** of two naturally occurring *ndufb7* allelic variants using CRISPR/Cas9;
- **Strategy II:** Differential **short epitope-tagging** of two naturally occurring *ndufb7* variants;
- **Strategy III:** Differentiation of a wildtype (K84) *ndufb7* allele and a **disease variant (W84) *ndufb7* allele** by CRISPR/Cas9-mediated mutagenesis.

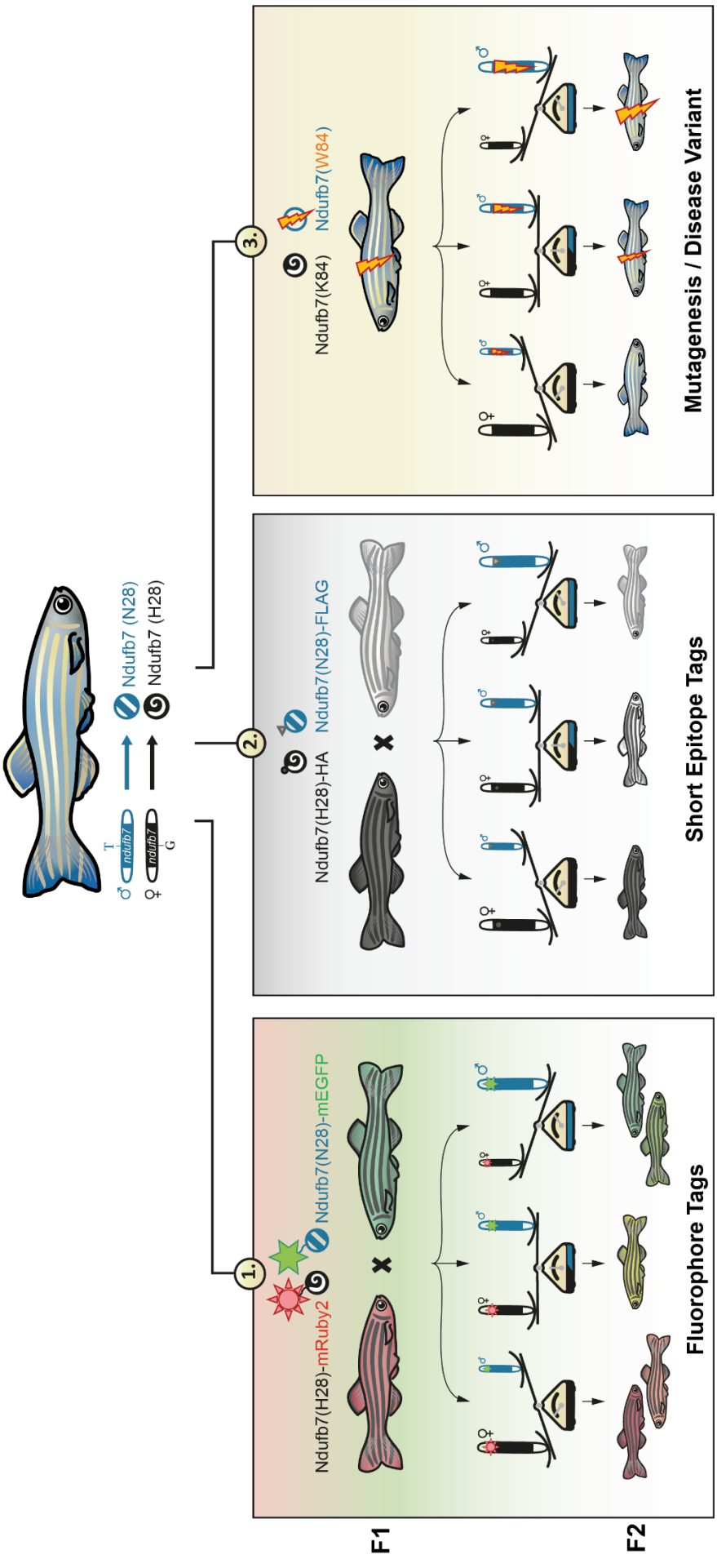


Figure 4. Three strategies devised for establishment of a zebrafish model to study allelic expression of the nOXPHOS subunit ndufb7. 1. Strategy I: differential tagging of the two naturally occurring ndufb7 alleles with the fluorophores mRuby2 and mEGFP. **2.** Strategy II: differential tagging of the two naturally occurring ndufb7 alleles with the short epitope tags HA and FLAG. **3.** Strategy III: differentiation of a wildtype (K84) allele and a disease variant (W84) allele. The colors blue and black represent the paternal and maternal allele or the disease and wildtype allele respectively. The two alleles differ by a SNP (T-G) which leads to a change in amino acid at position 28 (N: Asparagine, H: Histidine). The anticipated outcomes, depending on potential allelic preferences by the organism, are shown in the bottom row (F2 generation).

3 Materials and Methods

3.1 Zebrafish maintenance

The zebrafish (*D. rerio*) strains used in this project, AB (Streisinger *et al.*, 1981), WIK (Rauch, Granato and Haffter, 1997) and *Tg(actc1b:tomm20-ZsGreen),* a gift from Philipp Gut, were maintained at the animal facility of the University of Bern Institute of Anatomy at constant water conditions: 28°C, pH 7.5, 650-700 µS/cm, with a daily water exchange rate of 10%. Adult animals maintained for mating purposes were fed three times per day, once artemia (Ocean Nutrition) and two times dry food (Gemma Micron 300), and kept at a light:dark cycle of 14:10 hours respectively. Following approval by the “Amt für Landwirtschaft und Natur” from the Canton of Bern, Switzerland, all animal experiments were conducted with zebrafish embryos up to 5 days post fertilization (dpf) under the following licenses: BE64/18, BE11/17 and the license for genetic modifications G BE8/19.

3.2 Zebrafish injections and fluorescence imaging

Zebrafish embryos were injected (FemtoJet 4x Series, Eppendorf) at one-cell stage with the corresponding injection solution (**Table 1**). Prior to fluorophore knock-in injections, the repair template was digested with I-SceI (NEB, R0694S) in CutSmart Buffer (NEB, B7204S) at 37°C overnight where indicated. CRISPR RNAs (IDT, 100 µm Alt-R^{TM} CRISPR-Cas9 crRNA / tracrRNA) were combined into single guide RNAs in nuclease free Duplex Buffer (IDT, 11-05-01-12) by heating at 95°C for 5 min. Cas9 protein (IDT, 1081061) was diluted in Cas9 buffer (20 mM HEPES, 150 mM KCl, pH 7.5) and assembled with gRNAs at 37°C for 10 min prior to injection.

Injected embryos were maintained in E3 “blue” medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$, 10^{-5} % Methylene Blue) for 24 hours at 28°C. From days 2-5 embryos were kept in E3 medium without Methylene Blue. Fish were anesthetized with 0.02 mg/mL tricaine for fluorescence imaging via stereoscope (SMZ25, Nikon) at 10x magnification or euthanized with 0.04 mg/ml tricaine and collected for further analysis.

Table 1. Injection setup: injection solutions and amounts injected per embryo.

Experiment	Injection solution	Amount
transient overexpression	mRNA in nuclease free H_2O	0.45 - 1.3 ng
	pCS2_ndufb7-GGG-mEGFP plasmid in nuclease free H_2O	0.06 ng
stable overexpression	plasmid (25 ng/µL), *tol2* mRNA (200 ng/µL), KCl (200 mM) in nuclease free H_2O	1 nL
	control: plasmid (25 ng/µL), KCl (200 mM) in nuclease free H_2O	1 nL
fluorophore knock-in	*Cas9 mRNA mix:* repair template *I-SceI* digest (125 ng/µL), *Cas9* mRNA (370 ng/µL), CRISPR for gDNA (240 ng/µL), CRISPR for plasmid (240 ng/µL), KCl (200 mM) in Duplex Buffer	1 nL
	Cas9 protein mix: repair template *I-SceI* digest (58 - 85 ng/µL), Cas9 protein (600 ng/µL), CRISPR for gDNA (240 ng/µL), CRISPR for plasmid (240 ng/µL), KCl (200 mM) in Duplex Buffer	1 nL
	control: repair template *I-SceI* digest (58 - 85 ng/µL), KCl (200 mM) in Duplex buffer / nuclease free H_2O	1 nL
mutagenesis knock-in	*CRISPR/Cas9 mix:* Cas9 protein (600 ng/µL), CRISPR for gDNA (240 ng/µL), CRISPR for plasmid (240 ng/µL), KCl (200 mM) in Duplex Buffer	1 nL
	repair template mix: 2 µM *ndufb7*^{K84W} ssODN in nuclease free H_2O	1 nL

3.3 Genomic DNA extraction

Genomic DNA (gDNA) was extracted from 1-5 dpf embryos based on established protocols (Meeker *et al.*, 2007). In brief, when simultaneous RNA extraction was not required, embryos were euthanized, washed with PBS and lysed in 50 mM NaOH at 98°C for 20 min. 1M TRIS pH 7.5 was added to neutralize NaOH to 1:10 dilution. Following centrifugation, the supernatant containing gDNA was directly used as input for downstream applications (PCR). Samples were stored at -20°C.

When extracting RNA from the same sample, embryos were euthanized at 36 hpf, dechorionated with 1 mg/mL Pronase E (Serva, 33635.02) and washed with PBS. On ice, embryos were homogenized in PBS by pipetting and processed with TRI Reagent according to manufacturer's instructions (Invitrogen, 15596018). In brief: following chloroform addition, phase separation and extraction of RNA from the aqueous phase (see next section), gDNA was isolated from the residual interphase and lower phenol-chloroform phase by precipitation with 100 % EtOH. The precipitate was washed twice with 0.1 M sodium citrate in 10 % EtOH and once with 75 % EtOH. The resulting DNA pellet was resuspended in 8 mM NaOH, neutralized with 16 mM TRIS-HCl pH 7.5, centrifuged, and the final DNA-containing supernatant was stored at -20°C.

3.4 RNA extraction and cDNA synthesis

Total RNA was extracted from 36 hpf zebrafish embryos using TRI Reagent following the above described dechorionation, euthanasia, and tissue homogenization steps. The RNA-containing aqueous phase was isolated according to manufacturer's instructions (Invitrogen, 15596018) and purified with the RNA Clean & Concentrator kit (Zymo Research, R1015). cDNA was prepared using the Maxima First Strand cDNA Synthesis Kit with dsDNase (Thermo Scientific, K1672) according to manufacturer's instructions and the resulting cDNA was stored at -20°C.

3.5 Generation of plasmids and mRNA for transient overexpression

The coding sequence of *ndufb7*-201 was amplified from AB (*H28* variant) or WIK (*N28* variant) cDNA using Q5 HiFi DNA Polymerase (NEB, M0491) followed by purification via Zymoclean Gel DNA Recovery Kit (Zymo Research, D4008) according to manufacturer's instructions. In the same way, the coding sequence of the fluorophores was obtained from the following plasmids: pSems-Cox8a-mRuby2 and pSems-CIIIK_Clover (gifts from Karin Busch), mCitrine-N1 (Addgene, 54594 (Griesbeck *et al.*, 2001)), mTurquoise2-N1 (Addgene, 54843 (Goedhart *et al.*, 2012)), mEGFP (Addgene, 18696, a gift from Karel Svoboda) and eGFP (PCR product, a gift from A. Sanz).

The plasmids for generation of C-terminally tagged *ndfub7* mRNA were produced in the following steps:

1) amplification of *ndufb7* and fluorophore CDS from the respective templates,
2) attachment of a linker sequence (GGG, $(GGGGS)_3$ or $(EAAAK)_3$ in the case of mRuby2 and GGG in the case of all other fluorophores) to the C-terminus of $ndufb7^{AB}$ or addition of a C-terminal epitope tag (1x HA in the case of $ndufb7^{AB}$, 1x FLAG for $ndufb7^{WIK}$) and
3) generation of 25 bp overhangs via Q5 PCR,
4) restriction digest of the pCS2_nCas9n vector backbone (Addgene, 47929 (Jao, Wente and Chen, 2013)) with XbaI (Thermo Scientific, ER0682), NcoI (Thermo Scientific, FD0573) and EcoRI (Thermo Scientific, FD0274) followed by gel purification of the backbone band using the Zymoclean Gel DNA Recovery Kit (Zymo Research, D4008),
5) assembly of purified fragments with NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621),
6) transformation into competent DH5α cells, growth on Ampicillin LB-Agar plates, and
7) plasmid harvest using the NucleoSpin Plasmid Miniprep Kit (Macherey-Nagel, 740588) followed by sequence verification via Sanger sequencing.

All primers are listed in **Table 2**. The resulting mRNA template plasmids were linearized with NotI (Thermo Scientific, FD0594) or ApaI (Thermo Scientific, FD1414). Complete digestion was confirmed by agarose gel electrophoresis and linearized plasmids were purified using the DNA Clean & Concentrator™-5 kit (Zymo Research, D4008). Finally, mRNA was generated with the mMESSAGE mMACHINE SP6 Transcription Kit (Invitrogen, AM1340) according to manufacturer's instructions. Following TURBO DNase treatment, the generated mRNA was purified with the RNA Clean & Concentrator kit (Zymo Research, R1015). The concentration was measured with NanoDrop and the mRNA was aliquoted and stored at -80°C until injection.

3.6 Mutant *ndufb7* mRNA generation by plasmid mutagenesis

The newly generated pCS2_ndufb7^{H28}-(GGGGS)$_3$-mRuby2 mRNA plasmid was used as template for mutagenesis PCR with: 2 mM dNTPs, 50 ng template plasmid, 2-3 units Pfu-Polymerase (Promega, M7741), 10x Pfu-Buffer and the corresponding mutagenesis primers listed in **Table 2**. The PCR reaction was treated with 5 units DpnI (Thermo Scientific, ER1701) at 37°C for 1 hour to digest the template plasmid, grown in DH5α cells and harvested with the NucleoSpin Plasmid Miniprep Kit (Macherey-Nagel, 740588). Following sequence verification by Sanger sequencing, the mutagenized plasmids were linearized and used for mRNA generation as described in the previous paragraph.

3.7 Generation of stable overexpression plasmids

The coding sequences for *ndufb7^{H28}-1xHA* and *ndufb7^{N28}-1xFLAG* were amplified from the above described mRNA plasmids using Q5 HiFi DNA Polymerase (NEB, M0491). Similarly, the backbone sequence was obtained by PCR amplification from a pDestTol2_Ubi: Rox-Phi-Rox-loxP-tagBFPmyc-loxP-GFP-p2A-NTR_Cryaa:mCerulean vector (generated by A. Sanz, N. Mercader lab). A ubiquitin promoter element was obtained from the same plasmid backbone by restriction digest with NotI (Thermo Scientific, FD0594) and XhoI (Thermo Scientific, FD0695). All individual fragments were purified with the Zymoclean Gel DNA Recovery Kit (Zymo Research, D4008). The *ndufb7^{H28}-1xHA* and *ndufb7^{N28}-1xFLAG* coding sequences were each placed under the control of a ubiquitin promoter element and assembled with the pDestTol2_Cryaa:mCerulean backbone fragment using the NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621). The resulting plasmids were amplified and sequenced as described above. All primers are listed in **Table 2**.

3.8 *ndufb7* CRISPR-Cas9 ssODN, crRNA and plasmid repair template design

CRISPR sgRNA candidates were selected with the online tool CRISPOR and commercially synthetized by IDT. For K84W mutagenesis in exon 2 of *ndufb7* the following Alt-R® CRISPR-Cas9 crRNAs (2 nmol) were ordered: crRNA 134 (5'-AGTATTCCCAGTCGTGTTTC-3'), crRNA 147 (5'-GCGACCACCAGAAACACGAC-3'). Both mutagenesis crRNAs have no predicted exonic off-target binding sites (0-3 mismatches) and induce a cut in direct vicinity (0-3 bp) of the mutagenesis site. For the knock-in of the *ndufb7*^{K84W} mutation, the following 90 nucleotides long, + strand ssODN repair template was designed and ordered through the IDT website: 5'-TTACTCCTGATGTTGACAGTATTCCCAGTCGTGCCACTGGTGGTCGCAGGCCATGTAGTT TGGAAACATGTCTCTCCTGCACTTCATCAG -3'

Similarly, the following Alt-R® CRISPR-Cas9 crRNAs (2 nmol) were ordered for C-terminal tagging of *ndufb7*: crRNA 1 (5'-AGCTGAATGGAGACGAGGGA-3'), crRNA 2 (5'-GACTGAACTTGAGGAAGAAG-3'), crRNA 3 (5'-TCGAGCCTTCAGAACATCTG-3'). They were selected due to vicinity to the STOP codon (10-80 bp distance) and least predicted exonic off-target binding sites.

The repair templates for C-terminal fluorophore tagging of *ndufb7* were generated by molecular cloning. 5' and 3' homology arms flanking the *ndufb7* STOP codon, including addition of 25 bp overhangs, were amplified from WIK gDNA with Q5 HiFi DNA Polymerase (NEB, M0491), followed by purification via Zymoclean Gel DNA Recovery Kit (Zymo Research, D4008) according to manufacturer's instructions. The coding sequences of $(GGGGS)_3$-mEGFP and $(GGGGS)_3$-mRuby2 were amplified from the previously generated and sequenced mRNA plasmids. The pKHR4 backbone (Addgene, 74592 (Kazuyuki Hoshijima, Jurynec and Grunwald, 2016)) was linearized by restriction digest with NotI (Thermo Scientific, FD0594) and XhoI (Thermo Scientific, FD0695). The plasmid was assembled using NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621) in a vector:insert ratio of 1:2, followed by transformation into competent DH5α cells. Plasmids were harvested using the NucleoSpin Plasmid Miniprep Kit (Macherey-Nagel, 740588) and the sequence was verified by Sanger sequencing. All primers are listed in **Table 2**.

3.9 Determination of CRISPR-Cas9 targeting and *in vivo* mutagenesis efficacy via T7 Endonuclease 1 mutagenesis assay

Embryos injected with the CRISPR-Cas9 mixtures described in **Table 1** in the absence of a repair template were assessed for cutting efficiency by T7 Endonuclease 1 (T7E1) assay (NEB, M0302) at 1 dpf. First, the respective genomic regions flanking the CRISPR-Cas9 target sites in *ndufb7* exon 2 or near the STOP codon in *ndufb7* exon 3 were amplified from extracted gDNA using Q5 HiFi DNA Polymerase (NEB, M0491) according to manufacturer's instructions. Second, the PCR amplified fragments from gDNA of uninjected embryos were annealed with the PCR amplified fragments from the gDNA of CRISPR-Cas9 injected embryos as described in the kit instructions. The re-annealed fragments were treated with T7 Endonuclease 1 for 30' at 37°C and cutting efficiency was visualized via gel electrophoresis on a freshly prepared 2 % agarose gel. sgRNAs with a cutting efficiency of at least 80 % were chosen for genome editing experiments. All primer sequences are listed in **Table 2**.

Following the same procedure, single embryos injected with sgRNA 147 and ssODN were collected at 1 dpf and subjected to T7E1 assay, to test the knock-in success of the *ndufb7*^{K84W} mutation. The PCR fragments obtained from single embryos positive for T7E1 cutting were purified with the DNA Clean & Concentrator™-5 kit (Zymo Research, D4008) and sent for Sanger sequencing.

Table 2. List of all primer sequences used for plasmid generation, T7E1 assay and sequencing.

mRNA plasmid generation	
Target	**Primer sequence (5'-3')**
ndufb7-201_F	ATGGGCGCTCATCTTGTGC
ndufb7-201_R	TGCTGCGTTCGCCTCGAT
ndufb7-NcoI_F	cgatccaatgtcgacgccac*catgg*CTTCTCCAATGGGCGCTCATCTTGTG
ndufb7-ggg-mRuby2_R	cccttagacactcctccaccTGCTGCGTTCGCCTCGAT
ndufb7-ggggs3_R1	ctccaccactacctccaccaccTGCTGCGTTCGCCTCGAT
ggggs3-mRuby2_F	gaggtagcggtggaggaggtagtGTGTCTAAGGGCGAAGAGC
ndufb7-ggggs3_R2	acactacctcctccaccgctacctcCTCCACCACTACCTCCACC
ndufb7-eaaak3_R1	tgctgcctccttagctgctgcctcTGCTGCGTTCGCCTCGAT
eaaak3-mRuby2_F	gctaaggaggcagcagctaagGTGTCTAAGGGCGAAGAGC
ndufb7-eaaak3_R2	acaccttagctgctgcctccttagcTGCTGCCTCCTTAGCTGCTG
ndufb7-ggg-GFP_R	cccttgctcacTCCTCCACCTGCTGCGTTCGCCTCGAT
ggg-mRuby2_F	ggtggaggaGTGTCTAAGGGCGAAGAGCT
mRuby2_R	CTTGTACAGCTCGTCCATCCC
mRuby2-XbaI_R	taatacgactcactatagttctagaTTACTTGTACAGCTCGTC
ggg-GFP_F	ggtggaggaGTGAGCAAGGGCGAGGAG

GFP_R	CTTGTACAGCTCGTCCATGC
ndufb7-FLAG-stop_R1	ttacttgtcatcgtcgtccttgtaatcTGCTGCGTTCGCCTCGAT
FLAG-stop-XbaI_R2	gtaatacgactcactatagttctagaTTACTTGTCATCGTCGTCCT
ndufb7-HA-stop_R1	ttaagcgtaatctggtacgtcgtatgggtaTGCTGCGTTCGCCTCGAT
HA-stop-XbaI_R2	gtaatacgactcactatagttctagaTTAAGCGTAATCTGGTACGTCGT
Plasmid mutagenesis	
Target	**Primer sequence (5'-3')**
ndufb7(K84W) mut_F	TGCGACCACCAGtggCACGACTGGGAA
ndufb7(K84W) mut_R	TTCCCAGTCGTGccaCTGGTGGTCGCA
ndufb7(W84R) mut_F	TGCGACCACCAGagaCACGACTGGGAA
ndufb7(W84R) mut_R	TTCCCAGTCGTGtctCTGGTGGTCGCA
ndufb7(Asn28) SNP_F	CAATATGACCCGaATTTAGGCTTCG
ndufb7(Asn28) SNP_R	CGAAGCCTAAATtCGGGTCATATTG
ndufb7(84K, wt) rev mut_F	TGCGACCACCAGAAACACGACTGGGAA
ndufb7(84K, wt) rev mut_R	TTCCCAGTCGTGTTTCTGGTGGTCGCA
pTol2 transgenic plasmid generation	
Target	**Primer sequence (5'-3')**
kozak-ndufb7_F1	cgccaccatggcttctccaATG
Ubi-kozak-ndufb7_F2	tcttactttgaatttgtttacagGCGGCCGCcgccaccatggcttctccaATG
pDestTol2_F	tctagaactatagtgagtcgtattacgtagatccagacat
pDestTol2_R	gacaaattctagaactttgctggtctcgagggcccatctggcctgtgttt
Repair template plasmid generation	
Target	**Primer sequence (5'-3')**
5' hom-ndufb7_F1	AATCAAAGGGGGACTTATTTTTCTAT
5' hom-ndufb7_F2	cgggtagggataacagggtaatggcAATCAAAGGGGGACTTATTTTTCTAT
5' hom-ndufb7_mut_R	TGCTGCGTTCGCCTCGATtCTCTTtTTCCT
ndufb7-3' UTR hom_F1	TAAACATCCCTCCCTCGTCT
ndufb7-3' hom_R1	AGCAGTGCTGCCAATTGCTTT
ndufb7-3' UTR hom_mut_F2	TAAACATCCCTtCCTCGTaTCCATT
mRuby2-3' UTR hom_F3	tggtgggatggacgagctgtacaagTAAACATCCCTtCCTCGTaTCCATT
ndufb7-3' hom-Xho1_R2	ggataacagggtaatggcggctcgagAGCAGTGCTGCCAATTGCTTT
ggggs3-mEGFP_F	GAGGTAGCGGTGGAGGAGGTAGTgtgagcaagggcgaggag
mEGFP-ndufb7 3' UTR_F	tctcggcatggacgagctgtacaagTAAACATCCCTtCCTCGTaTCCATT
T7E1 assay, quality control and sequencing	
Target	**Primer sequence (5'-3')**
ndufb10_F	GGCCAGCGTTTATCCTTAGACA
ndufb10_R	TACAATTTGATGCACAGATGAGAGA
mRuby2_R2	GACCACCTTCATACGCATATTTTC
mRuby2_R3	GTCAAAGGCAAATGGCAGGG
mRuby2_F2	GTGTCTAAGGGCGAAGAGCT

mRuby2_F3	CCGCCTGGAAAGGTTAGAGG
ndufb7 exon 3_T7 F	TTACGTGATGCGGATGAAAGAGT
ndufb7_3' genomic_T7 R1	TGCTCTGCAACTCATGAACGTG
ndufb7_3' genomic_T7 R2	TTGGTCATACTGCCCAGCTCT
ndufb7_exon 2_F	AGATGAACCTGGCGATGTTGC
ndufb7_exon 1_F	ACCAGAACCCAACAAACCGTCT
ndufb7_3'genomic_R3	CGTGAGGCTCAGAAAGATGC
ndufb7_3'genomic_R4	AGCGTTCAGAAACCACGACT
ndufb7 exon 2 mut_T7 F	TGTCAGTAATGGTGGCCACGCAG
ndufb7 exon 2 mut_T7 R	GAGCATCTCTCACGCTGGCCATC

3.10 Immunostaining and confocal imaging

Embryos were collected at 1 dpf following treatment with 0.04 mg/mL tricaine. Immunostaining was performed as previously described (González-Rosa *et al.*, 2011). In brief, dechorionated embryos were deyolked with Ringer's solution (116 mM NaCl, 2.9 mM KCl, 5 mM HEPES, pH 7.2) containing protease inhibitor (Roche, 04693132001) and 1 mM EDTA. Deyolked embryos were fixed overnight with 2 % PFA in PBS at 4°C, followed by overnight permeabilization with PBS/Triton X-100 0.5 % at 4°C. Several washing steps were carried out using PBS/Tween 0.1 %. Non-specific binding was blocked by 3-hour (RT) or overnight (4°C) incubation in histoblock (5 % BSA, 5 % goat serum, 20 mM $MgCl_2$). Samples were incubated with the respective primary antibodies (see **Table 3**) in PBS/BSA 5 % overnight at 4°C.

Following secondary antibody incubation (see **Table 3**) for 4 hours at room temperature, washing and 5 minutes of DAPI staining, the embryos were embedded in low melting agarose (1%) and imaged using a Zeiss LSM 880 at 20x and 63x magnification.

Table 3. Primary and secondary antibodies used for immunostaining and confocal imaging.

Primary Antibodies		
Target	**Dilution**	**Antibody specifications**
anti-mCherry (mRuby2)	1:250	Rat IgG2a, monoclonal (Thermo Scientific, M11217)
anti-GFP	1:500	Chicken IgGY (Aves, GFP-1010)
anti-mtCO1	1:300	Mouse IgG2a, monoclonal (Invitrogen, 1D6E1A88)
anti-HA	1:150	Rabbit IgG, polyclonal (Invitrogen, SG77)
anti-FLAG	1:150	Mouse IgG1, monoclonal (Sigma-Aldrich, F3165)
Secondary Antibodies		
Target	**Dilution**	**Antibody specifications**
anti-Rat IgG	1:250	Goat anti-Rat IgG (H+L), Alexa Fluor 568 (Invitrogen, A-11077)

anti-Chicken IgY	1:1000	Goat anti-Chicken IgY (H+L), Alexa Fluor 488 (Invitrogen, A-11039)
anti-Mouse IgG2a	1:250	Goat anti-Mouse IgG2a, Alexa Fluor 488 (Invitrogen, A-21131)
anti-Mouse IgG2a	1:250	Goat anti-Mouse IgG2a, Alexa Fluor 568 (Invitrogen, A-21134)
anti-Mouse IgG2a	1:250	Goat anti-Mouse IgG2a, Alexa Fluor 647 (Invitrogen, A-21241)
Biotin-Sp anti-Mouse IgG	1:250	Biotin-SP (long spacer) AffiniPure $F(ab')_2$ Fragment Goat Anti-Mouse IgG (H+L), Jackson Immuno (115-066-062)
Streptavidin Cy5	1:300	Streptavidin, Cy5 conjugate (Invitrogen, SA1011)
anti-Rabbit IgG	1:250	Goat anti-Rabbit IgG (H+L), Alexa Fluor 568 (Invitrogen, A-11036)
anti-Mouse IgG1	1:250	Goat anti-Mouse IgG1, Alexa Fluor 488 (Invitrogen, A-21121)

3.11 Ndufb7 variant protein modeling (J. L. Cabrera)

At the time of writing, a solved protein structure of zebrafish Ndufb7 was not available through the Protein Data Bank (PDB). The structures described here are a courtesy of J. L. Cabrera, who modeled the Ndufb7 variants as follows:

The reference model Ndufb7^{H28} (AB strain, wildtype) was obtained by homology modeling / Rosetta comparative modeling (Song *et al.*, 2013), producing 1000 models using the following solved PDB structures from other species: 6g2j (Agip *et al.*, 2018) and 5xtd (Guo *et al.*, 2017). After clustering these 1000 models, the lowest energy model was selected from the most crowded cluster. The templates 6g2j and 5xtd correspond to mouse and human respiratory complex I, respectively. In both species the first amino acid, methionine, is removed, therefore it is also absent in the generated zebrafish Ndufb7 model.

The SNP variant Ndufb7^{N28} (WIK strain) and the two mutant variants Ndufb7$^{H28+K84W}$ and Ndufb7$^{N28+K84W}$ were obtained according to a previously published rosetta ddG monomer protocol (Kellogg, Leaver-Fay and Baker, 2011). Point mutation ddG calculation follows the expression: $\Delta\Delta G$ = mutant energy – wild type energy. The final SNP and mutant models were selected as lowest energy trajectories from unique clusters obtained in all cases. To determine the effect size of SNP and mutation, position specific mutability degrees of freedom were obtained by exchanging reference amino acids by all other 19 amino acids. Then, ddG values for SNP and mutation (H28N and K84W) were compared against median absolute values of ddG (|ddG|) from 19 possible amino acid changes for these positions.

4 Results

*Following target gene selection (**section 4.1**), I tested the feasibility of the three zebrafish model strategies in parallel. Thus, procedures related to construct design, mRNA generation and mRNA overexpression by microinjection largely overlapped for the three strategies. This is reflected by the recurring strain-dependent differences observed during mRNA overexpression experiments, particularly regarding linearization of the plasmids with either ApaI or NotI restriction enzyme (see **sections 4.2.2, 4.3.1 and 4.4.2**). Technical issues related to mRNA plasmid linearization are therefore briefly discussed in **section 4.4.2**. Unless indicated otherwise, the plasmids and results described in this section were generated by me. I performed experimental planning and target selection in close collaboration with Dr. Piragyte-Langa. When describing decisions or conclusions we made together, I therefore use "we" in some of the following subsections.*

4.1 Target gene selection

As outlined in the introduction, the OXPHOS machinery contains 13 mitochondrially encoded and approximately 90 nuclear encoded subunits, ~50% of which are part of OXPHOS complex I (Anderson *et al.*, 1981; Carroll *et al.*, 2006). To select a nOXPHOS target gene for investigating the potential of SNP-dependent allele-specific selection, Dr. Piragyte-Langa and I turned to RNA-seq data previously generated from two common zebrafish strains present in the laboratory: AB and WIK (unpublished data). The RNA-seq data was aligned to the TU strain reference genome GRCz11. Hence, gene variants matching the reference genome are hereafter referred to as "wildtype".

In total, 22 missense SNPs were detected that result in an amino acid change in nOXPHOS genes in either the AB or the WIK strain (see **Table 4**). Out of these 22 initial candidates, 17 SNPs lead to a change in amino acid properties (aromatic, acidic, basic, aliphatic, amidic, hydroxylic or sulfur-containing). We considered candidate SNPs inducing a change in amino acid properties more likely to interfere with protein function and therefore potentially capable of activating a selection mechanism in a heterozygous environment, in case such mechanisms exist. To avoid complications by mixing the genetic backgrounds of the two zebrafish strains, we finally considered only those nOXPHOS subunits present in heterozygosity (wildtype and SNP-variant allele) within one strain.

Interestingly, for two heterozygous nOXPHOS genes encoding subunits of complex I a slight allelic bias (>60 %) in transcript counts was detected: a 67 % bias towards expression of the wildtype allele in the case of *ndufb7* in WIKs and a 61 % bias towards expression of the wildtype

allele of *ndufb10* in comparison to one of its detected SNPs in ABs. These two SNPs are highlighted in **Table 4** in orange and yellow respectively.

Table 4. List of single nucleotide polymorphisms (SNPs) in nuclear OXPHOS genes by which the two working zebrafish strains AB and WIK differ from each other. SNP-induced changes are indicated in red. In the case of heterozygosity, the occurrence of each allelic variant is shown in percent. SNPs with a transcript imbalance >60 % are highlighted. hom: homozygous, het: heterozygous, -: SNP not present. *RNA-seq results excerpt is a courtesy of I. Piragyte-Langa.*

OXPHOS complex	nOXPHOS subunit	Chromosome	SNP name	Amino acid change	Change in amino acid properties?	AB strain	WIK strain
I	Ndufs1	6	rs512932187	D207 > E	both acidic	-	hom
	Ndufs7	11	rs512924184	F49 > L	aromatic > aliphatic	hom	-
	Ndufv1	19	rs513872267	S36 > T	both hydroxylic	-	het (T 46%, A 54%)
			rs504677439	Q42 > H	amidic > basic	-	het (A 44%, C 55%)
	Ndufb7	1	rs508855218	H28 > N	basic > amidic	-	het (G 67%, T 33%)
	Ndufb10	3	rs318240422	S54 > R	hydroxylic > basic	het (T 57%, G 43%)	-
			rs318240423	L137 > Q	aliphatic > amidic	het (A 61%, T 38%)	hom
	Ndufc2	15	rs511003198	N89 > D	amidic > acidic	het (A 51%, G 49%)	hom
			rs514012019	E92 > Q	acidic > amidic	het (G 52%, C 48%)	hom
			rs504809169	Q93 > L	amidic > aliphatic	het (A 51%, T 49%)	hom
II	Sdha	19	rs506272961	C18 > S	sulfur-containing > hydroxylic	het (A 55%, T 45%)	hom
			rs499758675	D545 > N	acidic > amidic	het (A 51%, T 49%)	-
III	Uqcrc1	6	rs507563988	Y104 > F	both aromatic	het (A 47%, T 53%)	-
	Uqcrq	14	rs501713507	D82 > E	both acidic	hom	-
IV	Cox10	12	rs512754587	S35 > L	hydroxylic > aliphatic	hom	-
	Cox4i1	18	rs40667660	T5 > K	hydroxylic > basic	hom	het (C 58%, A 42%)
	Cox5ab	18	rs513387084	A4 > S	aliphatic > hydroxylic	hom	het (G 44%, T 56%)
	Cox6c	6	rs505506371	E69 > Q	acidic > amidic	-	hom
	Cox7a2a	17	rs512510675	I11 > V	both aliphatic	het (T 51%, C 49%)	het (T 46%, C 54%)
	Cox15	13	rs508412800	A52 > T	aliphatic > hydroxylic	hom	-
V	Atp5f1c	4	rs40860200	M201 > I	sulfur-containing > aliphatic	het (C 52%, T 48%)	-
	Atp5f1e	6	rs512569373	G134 > S	aliphatic > hydroxylic	het (C 46%, T 54%)	-

While it is important to consider the limitations of the RNA-seq data analysis (potential artificial introduction of allelic bias due to positioning of the short reads in relation to individual SNPs), the following aspects make both Ndufb7 and Ndufb10 suitable targets, regardless of whether or not the detected allelic bias actually reflects the natural situation in the sequenced zebrafish strains:

First, both Ndufb7 and Ndufb10 are required for the assembly and functionality of OXPHOS complex I. Knock out of either subunit results in loss of the assembled complex I and certain compound heterozygous mutations in human NDUFB10 reportedly cause disease (Stroud *et al.*, 2016; Friederich *et al.*, 2017). Should there be mechanisms in place to select one allelic variant over the other in a heterozygous environment, we thus hypothesized that selecting the “best fitting” version of these subunits for integration into the assembled OXPHOS complex I may represent an evolutionary advantage.

Second, both subunits are located in the intermembrane space (IMS) in the distal arm of complex I. As such, they are neither in direct contact with the core catalytic subunits responsible for electron transport in the assembled complex, nor are they fully embedded in the inner mitochondrial membrane (Hunte, Zickermann and Brandt, 2010; Szklarczyk *et al.*, 2011; Wirth *et al.*, 2016). The location of Ndufb7 and Ndufb10 within complex I might therefore allow for attachment of a tag without compromising the overall assembly and function of complex I.

Third, both Ndufb7 and Ndufb10 are expressed across a wide variety of tissues and cell types in zebrafish and humans from early development on, according to the public Zebrafish Information Network database ZFIN (http://zfin.org/; Ruzicka *et al.*, 2019), the zfRegeneration database (Nieto-Arellano and Sánchez-Iranzo, 2019), and the Human Protein Atlas (http://www.proteinatlas.org; Uhlén *et al.*, 2015). For the scope of this thesis, we selected *ndufb7* as primary OXPHOS complex I target gene for parallel establishment of the below described animal models.

With its approximately 45 individual subunits, OXPHOS complex I is the largest of the five respiratory chain complexes. In its final form, complex I consists of three structurally and functionally distinct modules that form an L-shaped structure. One arm extends into the mitochondrial matrix (N- and Q-modules, responsible for NADH binding/oxidation and electron transfer) while the other arm is embedded in the inner mitochondrial membrane (P-module). The P-module, which harbors proton pumping activity, can be further divided into proximal (P_P) and distal (P_D-a, b) submodules. P_D-b, which is located at the distal end of the membrane arm, comprises the central mtOXPHOS protein MT-ND5, surrounded by the nOXPHOS accessory subunits NDUFAB1, NDUFB7, NDUFB3, NDUFB8, NDUFB9 and NDUFB2 (Guerrero-Castillo *et al.*, 2017). According to structural analysis by Zhu *et al*, NDUFB7 is located at the bottom of the P_D-b module within the intermembrane space (Zhu *et al.*, 2015; Wirth *et al.*, 2016). Interestingly, despite its location on the surface of complex I, deletion of NDUFB7 leads to loss of the entire P_D-b module in yeast, thereby causing a 50 % decrease in proton pumping activity (Dröse *et al.*, 2011). Similarly, a functional knockout study in HEK293T cells revealed that loss of NDUFB7 results in a drastic decrease in complex I levels (Stroud *et al.*, 2016).

The nOXPHOS gene *ndufb7* is encoded by three exons located on chromosome 1 in the zebrafish. In comparison, human *NDUFB7* and mouse *Ndufb7* are located on chromosome 19 and chromosome 8 respectively. Following transcription, nuclear export and translation in the cytoplasm, the initiator methionine is removed from the unfolded peptide chain and the N-terminal glycine is myristoylated (Walker *et al.*, 1992). Because Ndufb7 does not contain an N-terminal extension that could act as mitochondrial import signal, it is thought that the hydrophobic myristol group aids mitochondrial import via the Tomm40/Chchd4 import pathway (Udenwobele *et al.*, 2017; Ueda *et al.*, 2019). In the zebrafish, myristoylated Ndufb7 is predominantly detected during early development at 24 hpf when compared to 72 hpf and 120 hpf time points (Broncel *et al.*, 2015). The Ndufb7 peptide sequence contains a $(Cx_9C)_2$ domain. Once the peptide chain passes the outer mitochondrial membrane via Tomm40, it is folded by the chaperone Chchd4 through oxidation of the cysteines in the $(Cx_9C)_2$ domain, followed by formation of intramolecular disulfide bonds. The two disulfide bonds trap the folded protein in the intermembrane space, where it is finally assembled into OXPHOS complex I via the P_D-b module (Banci et al., 2009; Chacinska et al., 2004; Reinhardt et al., 2020; Terziyska et al., 2005).

Zebrafish Ndufb7 consists of a 120 amino acid-long polypeptide with a molecular weight of 14.48 kDa (Q6P6E5). In comparison, human NDUFB7 comprises 137 amino acids at a molecular weight of 16.40 kDa (P17568). The SNP detected by RNA-seq in the WIK strain induces an amino acid change at position 28 in the peptide chain, leading to conversion of the basic amino acid Histidine (H) to Asparagine (N), thus resulting in loss of the positive charge and ionizable imidazole ring (see model depicted in **Figure 21**). In comparison, this SNP is absent in ABs which are homozygous for the H28 wildtype variant.

4.2 Strategy I: Differential fluorophore-tagging of two naturally occurring *ndufb7* allelic variants using CRISPR/Cas9

In the first approach, I tested the feasibility of visualizing gene expression of the two *ndufb7* alleles by attachment of two differently colored fluorophore tags (one color per allelic variant). This method has several advantages: besides tag-specific sequence amplification at the DNA and mRNA level, fluorophores enable antibody-based differentiation between the two tagged allelic variants at the protein level *in vivo* and *in vitro*. Color-coding the two alleles allows for a direct readout of (tissue- or cell type-specific) allelic expression tendencies by imaging and it is a widely applicable method that enables real-time assessment of protein expression patterns throughout development by live imaging (Toseland, 2013). Thus, fluorophore-tagging fulfills the main requirements outlined in section 2.

However, there are major concerns associated with the use of fluorophores in the context of studying the OXPHOS machinery in its natural form: first, size and the potential of fluorophore multimerization are a valid concern. Commonly used fluorophores, such as variants of Green Fluorescent Protein (GFP), have a molecular weight of approximately 27 kDa (Zacharias *et al.*, 2002). In comparison, zebrafish Ndufb7 weighs just 14.5 kDa. Second, photo-excitation of fluorophores can induce phototoxic effects through generation of reactive oxygen species (ROS), which cause unwanted perturbations to biological systems (Dobrucki, Feret and Noatynska, 2007; Zheng *et al.*, 2014).

We did not find reports on whether these concerns play a role in the context of tagging the nOXPHOS subunit Ndufb7. With the goal in mind to establish a zebrafish line with two differentially fluorophore-tagged *ndufb7* alleles, I therefore investigated the potential of this strategy in two steps: 1) by testing the suitability of several peptide linkers and fluorophore candidates by transient overexpression of the tagged Ndufb7 wildtype variant via mRNA and/or plasmid injections, and 2) by testing the feasibility of CRISPR/Cas9-mediated knock-in at the genomic *ndufb7* locus.

4.2.1 Construct design: Red fluorophore and linker sequence testing

I first tested the red candidate fluorophore monomeric **mRuby2** (Lam *et al.*, 2012; Heppert *et al.*, 2016). This RFP variant had previously been used to tag *ndufb10* in zebrafish (data not shown) and other nOXPHOS subunits in human cell lines and yielded promising results (Rieger *et al.*, 2017). mRuby2 (26.5 kDa) exceeds Ndufb7 (14.5 kDa) nearly two-fold in molecular weight, hence I first tested three different linker sequences separating mRuby2 from Ndufb7 by transient mRNA overexpression:

1) 3 amino acid **short and flexible linker 1 (L1): GGG**
2) 15 amino acid **long and flexible linker 2 (L2): $(GGGGS)_3$**
3) 15 amino acid **long and rigid linker 3 (L3): $(EAAAK)_3$**

The three linker candidates were selected based on a review of commonly used linker sequences (Chen, Zaro and Shen, 2013): in brief, L1 and L2 provide flexibility and allow different degrees of mobility of the attached fluorophore, which may be important during OXPHOS complex assembly. The rigid linker L3 on the other hand maintains strict separation between Ndufb7 and the attached fluorophore due to the amino acid composition of the peptide sequence (Argos, 1990; Aurora *et al.*, 1997; George and Heringa, 2002; Biterge and Schneider, 2014). We reasoned that this separating characteristic and rigidity might prevent potential interference of the fluorophore with protein folding or function of Ndufb7.

To test these linker sequences in combination with mRuby2, I generated three mRNA plasmids (one per linker) containing SP6 promoter, Kozak sequence, the coding sequence for *ndufb7* and the C-terminally attached linker and fluorophore coding sequence, followed by stop codon and SV40 poly A tail. Following plasmid linearization with NotI and mRNA synthesis with SP6 RNA polymerase, I separately injected the three types of generated mRNA into zebrafish embryos at one-cell stage (up to 16-cell stage) and monitored the overall fluorescence of injected survivors by *in vivo* whole embryo fluorescence microscopy. This experimental approach is outlined in **Figure 5A**.

Interestingly, while some published zebrafish injection protocols recommend injection of less than 0.2 ng mRNA per embryo for overexpression of fluorophore-tagged proteins (Yuan and Sun, 2009), I did not observe fluorescence at similar mRNA concentrations (see **Table 5**). Instead, the observed percentage of fluorescent embryos at 1 dpf varied greatly depending on the strain as well as the mRNA preparation method directly preceding injection (thawing on ice, heating 95°C, heating + snap cooling or RNA clean-up).

As highlighted in **Table 5**, I achieved the highest percentage of fluorescent embryos when injecting between 0.9 and 1.3 ng of heated and snap cooled mRNA. At mRNA amounts >1 ng, I observed a maximum of 14 % (AB) and 75 % (WIK) of overall fluorescence for *ndufb7-**L1**-*

mRuby2 at 1 dpf. Injection of >1 ng of *ndufb7-**L2**-mRuby2* mRNA yielded a maximum of 33 % (AB) and 59 % (WIK) overall fluorescence the next day. In contrast, the number of fluorescent embryos observed following *ndufb7-**L3**-mRuby2* mRNA injection was lower, with maximum 6 % in ABs and 4 % in WIKs. Irrespective of the linker sequence, in most cases the fluorescent signal was lost by 2 dpf (data not shown).

Table 5. Percentage of fluorescent survivors per injected mRNA construct and zebrafish strain at 1 day post fertilization (dpf). mRNA constructs were linearized with NotI prior to mRNA synthesis. n.a. not assessed

mRNA construct	injected amount (ng)	fluorescent survivors at 1 dpf (%)		mRNA preparation preceding injection
		AB	WIK	
*ndufb7-**L1**-mRuby2*	0.15	0	n.a.	thawed on ice
	0.30	0	0	thawed on ice
	0.60	0	n.a.	thawed on ice
	0.90	n.a.	3-5	heated (95°C)
	1.09	n.a.	1-75	heated (95°C) + snap cooled
	1.24	8-14	20-70	heated (95°C) + snap cooled
*ndufb7-**L2**-mRuby2*	0.15	0	n.a.	thawed on ice
	0.30	0-16	15	RNA clean-up or heated (95°C)
	0.45	2-18	n.a.	thawed on ice or heated (95°C)
	0.60	10	n.a.	heated (95°C)
	0.90	13	n.a.	heated (95°C)
	1.05	5-33	36-59	heated (95°C) + snap cooled
	1.22 - 1.50	18-55	18	heated (95°C) + snap cooled
*ndufb7-**L3**-mRuby2*	0.30	n.a.	0	thawed on ice
	0.60	0	0	thawed on ice
	0.90	6	4	heated (95°C)
	1.34	2	4	heated (95°C) + snap cooled

Whole-embryo fluorescence microscopy showed that the amount of fluorescence within individual embryos varies greatly depending on the linker type. As shown in **Figure 5B**, I observed the most abundant fluorescent signal within individual immunostained embryos following *ndufb7-**L2**-mRuby2* mRNA overexpression (1.24 ng mRNA; n = 32), and overexpression of *ndufb7-**L1**-mRuby2* (1.22 ng mRNA; n = 14). mRNA encoding the long and rigid linker *ndufb7-**L3**-mRuby2* (1.34 ng mRNA; n = 9) yielded not only the lowest number of fluorescent embryos at 1 dpf, but also the lowest amount of fluorescent signal per embryo, which was in most cases limited to a few individual speckles.

To evaluate whether tagged Ndufb7 co-localizes with mitochondria, I collected fluorescent embryos at 1 dpf and co-stained with antibodies against mRuby2 (mCherry) and the mitochondrial marker mt-Co1 (Schmidt *et al.*, 2001). I confirmed mitochondrial co-localization for the three constructs by confocal imaging of embryonic tail muscle and skin regions (**Figures 6, 7** and **Suppl. Figure 2**). However, out of 9 fluorescent *ndufb7-**L3**-mRuby2* injected embryos, only two could be imaged. In the residual 7 embryos, the fluorescent signal was either out of reach or present in (likely dead) cells at the embryonic surface, as judged by the absence of DAPI signal. I validated the mt-Co1 antibody as mitochondrial marker by immunostaining of transgenic embryos expressing *tomm20-ZsGreen* under muscle specific α actin promoter *actc1b* prior to first use. Tomm20 is a translocase located in the outer mitochondrial membrane (Arribat

et al., 2019) and the endogenous ZsGreen signal overlaps nicely with immunostained mt-Co1 (see **Suppl. Figure 1**).

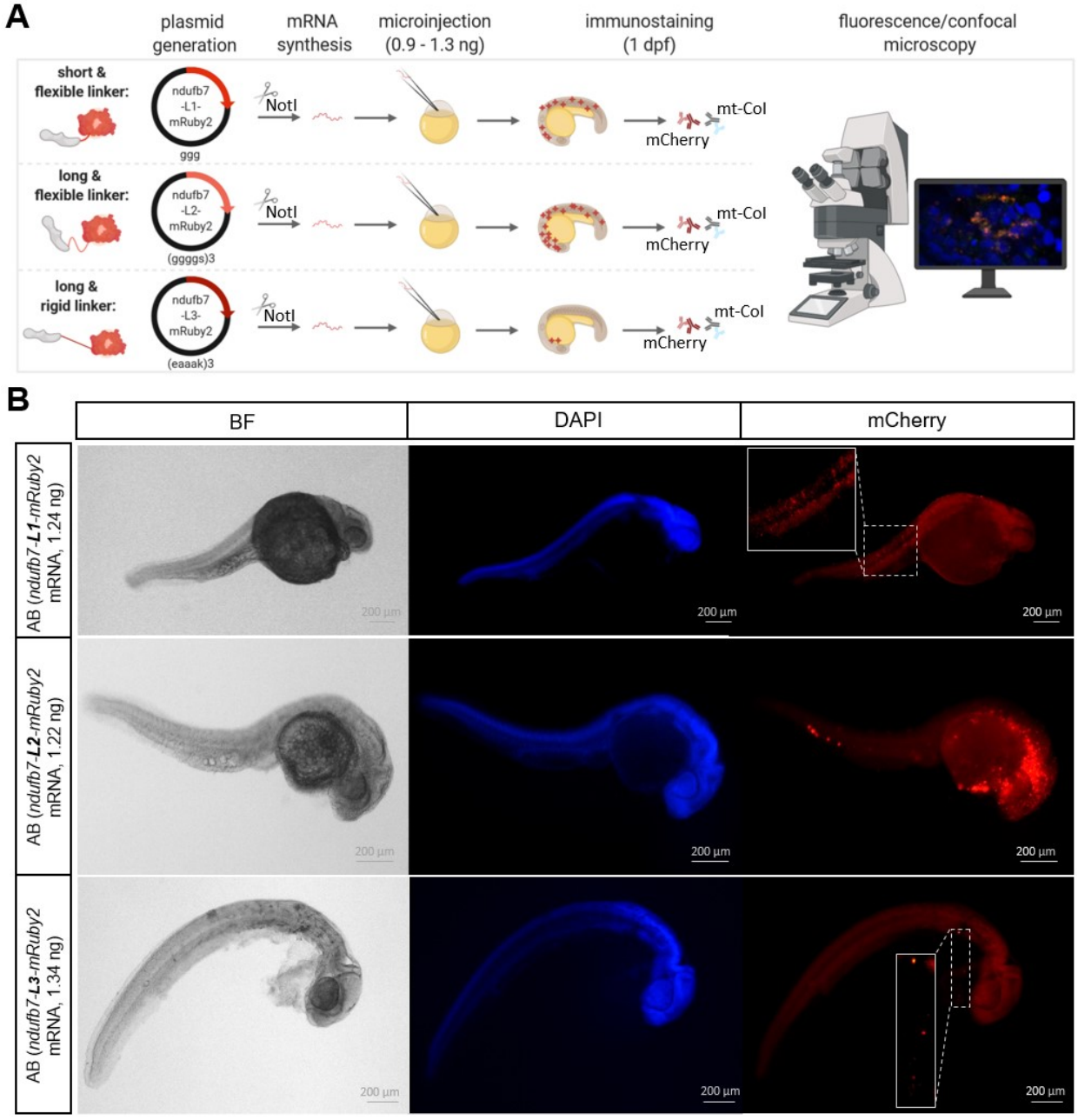


Figure 5. Red fluorophore and linker testing for C-terminal tagging of *ndufb7*. A schematic depiction of the experimental procedure generated with BioRender. Three constructs were tested containing the coding sequence for *ndufb7* C-terminally tagged with the red fluorophore *mRuby2*. Gene and fluorophore are separated by distinct linker sequences: a short and flexible linker 1 (L1, GGG), a long and flexible linker 2 (L2, $(GGGGS)_3$) and a long and rigid linker 3 (L3, $(EAAAK)_3$). Plasmids were linearized with NotI enzyme prior to mRNA synthesis, microinjection, immunostaining and imaging. **B** Fluorescence stereoscope images of partially deyolked AB strain embryos following mRNA overexpression and immunostaining at 1 dpf. BF: bright field, blue: DAPI (nuclei), red: mCherry (Ndufb7-Linker-mRuby2). Scale bar: 200 µm.

The depicted confocal imaging results suggest that some of the cells expressing mRuby2-tagged Ndufb7 undergo apoptosis (indicated by round apoptotic nuclear bodies visible in bright blue in Ndufb7-**L2/L3**-mRuby2 expressing embryos in **Figure 6, panel III and Figure 7, panel II**), while

other cells expressing the injected mRNA appear morphologically normal (observed for Ndufb7-**L1/L2/L3**-mRuby2).

We observed that the injected mRNA and its resulting protein product are initially distributed throughout the organism at 10-16 hpf (data not shown, see **Figure 9B** for an example). However, the fluorescent pattern became more concentrated (speckled) by 1 dpf, potentially indicating rapid protein degradation or accumulation and potential multimerization of mRuby2-tagged Ndufb7 within some cells of the injected embryos. The observed nuclear fragmentation of individual, brightly fluorescent cells may indicate that cell death is a direct consequence when the amount of the injected mRNA and/or its protein product become too high within an individual cell (Xu, Lai and Hua, 2019).

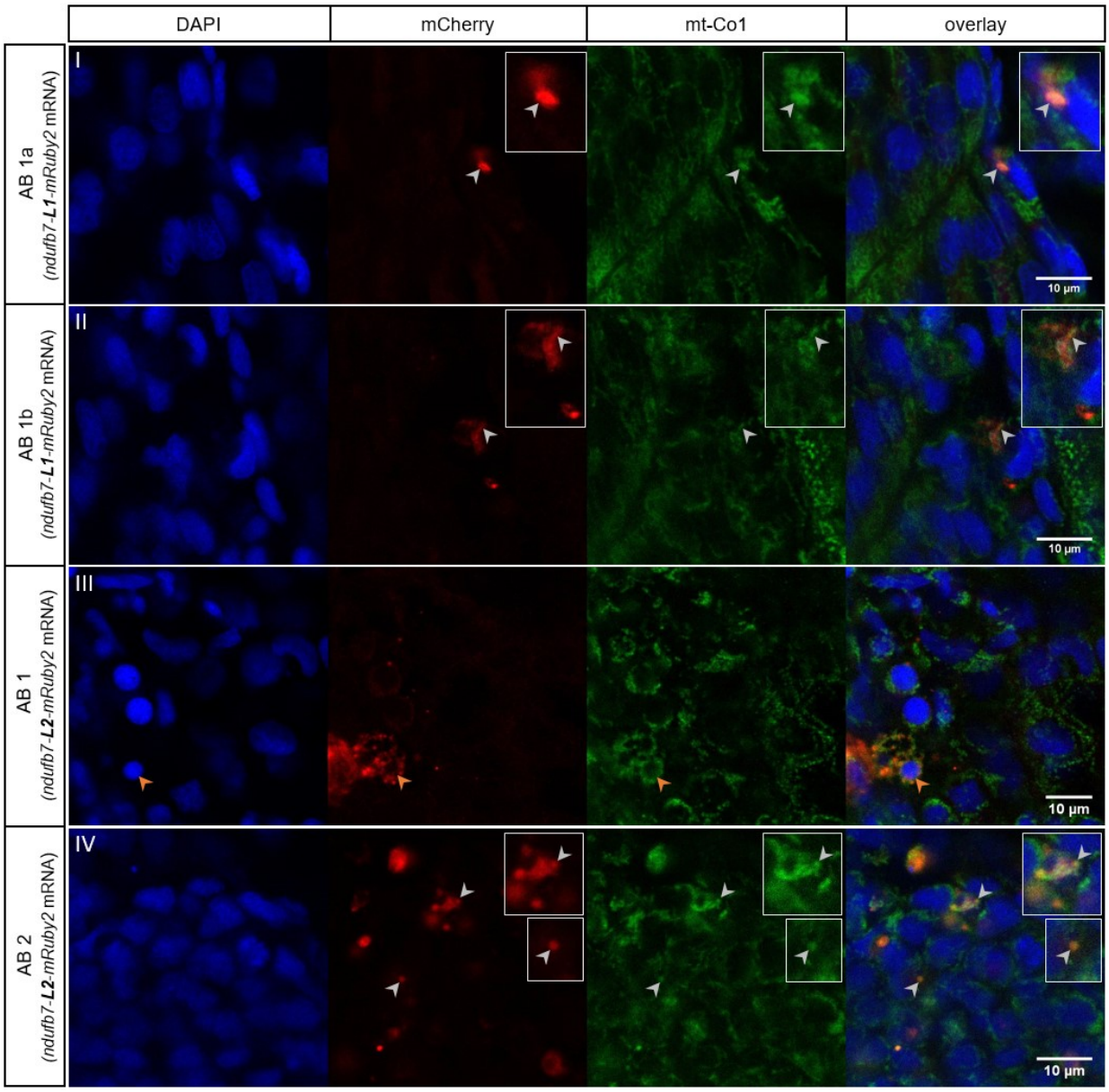


Figure 6. Immunostaining of tagged Ndufb7 following *ndufb7-L1/L2-mRuby2* mRNA overexpression in AB strain embryos at 1 dpf. Co-staining of mRuby2 with the mitochondrial marker mt-Co1 suggests mitochondrial localization of the tagged protein in the embryonic tail and at the tail surface. Grey arrows point to examples for overlapping signals which are also shown as zoom-ins in grey boxes. Orange arrows mark a dying cell (nuclear body).

Blue: DAPI (nuclei), red: anti-mCherry antibody (Ndufb7-L-mRuby2), green: anti-mt-Co1 antibody (mitochondria). Scale bar: 10 µm.

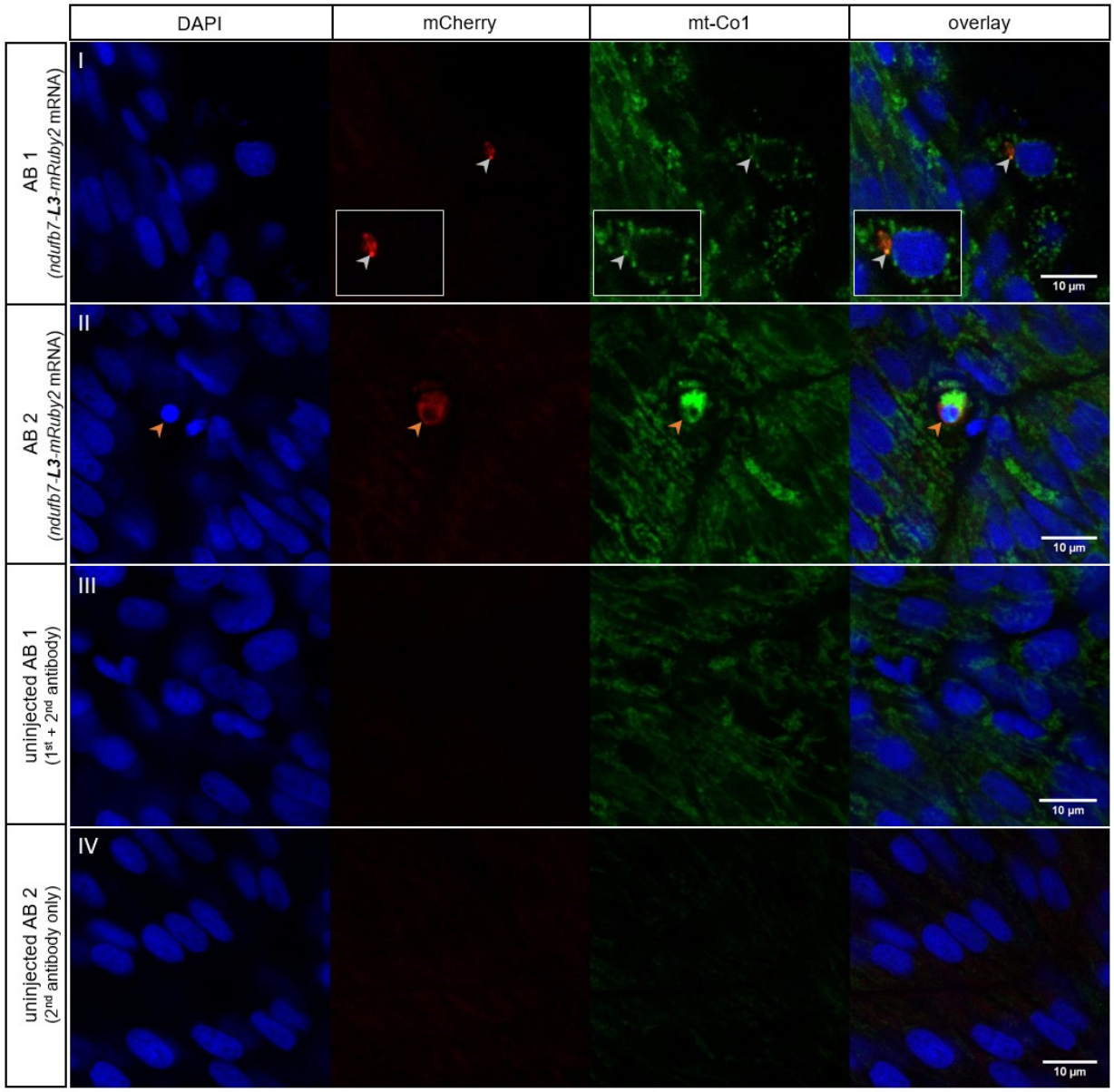


Figure 7. Immunostaining of tagged Ndufb7 following *ndufb7-L3-mRuby2* mRNA overexpression in AB strain embryos at 1 dpf. Co-staining of mRuby2 with the mitochondrial marker mt-Co1 suggests mitochondrial localization of the tagged protein at the tail surface. Grey arrows point to examples for overlapping signals, also shown as zoom-ins in grey boxes. Orange arrows in panel II mark Ndufb7-L3-mRuby2 expression in a dying cell (nuclear body). Row 3 and 4: negative control. Blue: DAPI (nuclei), red: anti-mCherry antibody (Ndufb7-L3-mRuby2), green: anti-mt-Co1 antibody (mitochondria). Scale bar: 10 µm.

In conclusion, all three tested mRuby2 mRNA constructs induce translation of the tagged protein, followed by localization to mitochondria. However, in the case of the L3 construct, which we found to be expressed at much lower levels than L1 and L2 constructs, more data would be required to definitely confirm this observation. While I did not observe pronounced cytotoxic effects associated with moderate mRuby2 expression (<10 % abnormal phenotype detected per mRNA type), on a cellular level, strong fluorescent signal coincided with the occurrence of apoptotic bodies in several samples, indicating cell death. This effect may be artificially induced by the overexpression of large quantities of injected mRNA (>1 ng/embryo) and may not be an

issue when expressed at endogenous levels. Thus, the flexible linkers L1 and L2 appear suitable for C-terminal tagging of Ndufb7 with the L2 construct yielding consistently higher amounts of signal within the fluorescent embryos compared to L1 and L3.

4.2.2 Construct design: Green fluorophore testing

While confirming the suitability of flexible linkers L1 and L2 in combination with the red fluorophore mRuby2 for C-terminal tagging of one of the allelic variants of Ndufb7, we simultaneously sought to determine a suitable green fluorescent protein to tag the second allele. Preliminary experiments for tagging of *ndufb10* suggested that, while no physical abnormalities were detected upon tagging with mRuby2, the zebrafish embryos reacted sensitively to the green fluorescent protein variant mClover: none of the injected embryos expressing the mClover fusion protein survived past 1 dpf, possibly due to phototoxic effects caused by the fluorophore (data not shown). For this reason, I tested a panel of five different GFP variants in combination with Ndufb7 and the flexible GGG linker L1 by transient mRNA overexpression: mClover, mTurquoise2, mCitrine, eGFP and mEGFP.

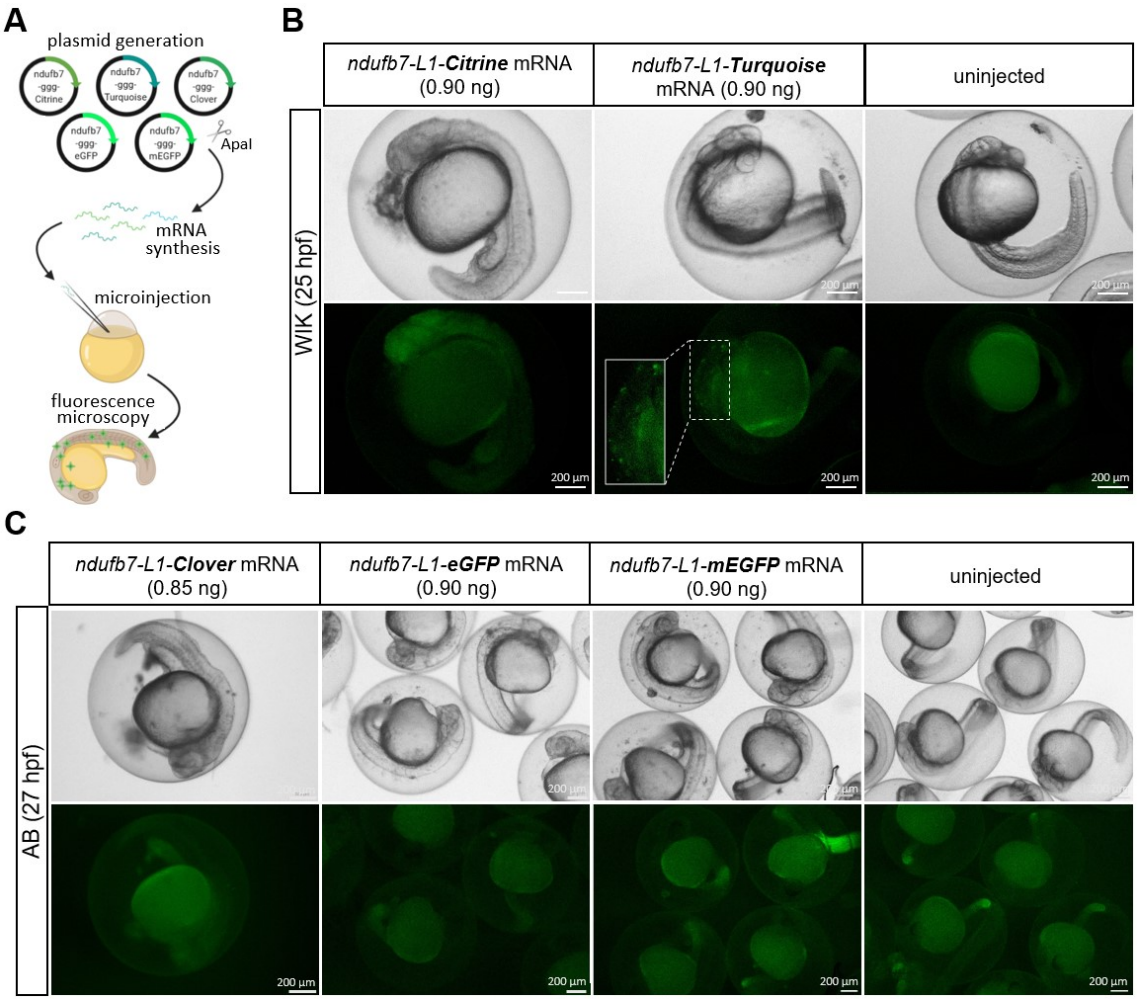


Figure 8. Green fluorophore testing by C-terminally tagged *ndufb7* mRNA overexpression in WIK and AB strain embryos. A schematic depiction of the experimental procedure generated with BioRender. Five GFP variants were tested by C-terminally tagging *ndufb7*: Citrine, Turquoise, Clover, eGFP and mEGFP. Plasmids were linearized

with ApaI enzyme prior to mRNA synthesis, microinjection and fluorescence microscopy. **B** Bright field and green channel fluorescence stereoscope images of 1 dpf WIK embryos injected with *ndufb7-L1-Citrine* or *-Turquoise* mRNA, compared to uninjected control. **C** Bright field and green channel fluorescence stereoscope images of 1 dpf AB embryos injected with *ndufb7-L1-Clover, -eGFP or -mEGFP mRNA* compared to uninjected control. Scale bar: 200 µm.

Using the same mRNA plasmid backbone (pCS2) as for the above described mRuby2 constructs, I generated five different plasmids containing SP6 promoter, Kozak sequence, coding sequence for *ndufb7-L1-**Clover/Citrine/Turquoise/eGFP** or **mEGFP*** followed by stop codon and SV40 poly A tail, which I linearized with ApaI prior to mRNA synthesis. I injected embryos at 1-cell stage (up to 16 cell-stage) with mRNA amounts ranging from 0.45 ng to 1.30 ng per embryo, followed by monitoring of fluorescent signal via whole-embryo live fluorescence microscopy at 11-16 hpf and at 1 dpf. The experimental approach is outlined in **Figure 8A**.

I observed fluorescent embryos approximately 11 hours after mRNA injection. The number of fluorescent embryos reached up to 43 % in ABs and up to 79 % in WIKs, thus confirming the functionality of the injected mRNA. However, this number dropped drastically overnight, with only up to 2 % of ABs (in the case of Clover; all other GFP variants: 0 %) and up to 26 % of WIKs remaining fluorescent at 1 dpf. Examples of fluorescent WIK embryos injected with mRNA are shown in **Figure 8B** for Citrine and Turquoise and **Figure 9B** for mEGFP. Injection details and the corresponding embryo counts are listed in **Table 6**.

Strikingly, AB embryos appeared more sensitive to the injected mRNA in general, as reflected by the percentage of embryos with abnormal phenotype at 1 dpf (including severe malformations such as absent or heavily deformed tail or head structures and less severe phenotypes such as shortened or kinked embryonic tails; see **Table 6**). The percentage of abnormal survivors at 1 dpf exceeded 15 % in 8 out of 10 injections into AB embryos (0.9 - 1.2 ng mRNA per embryo).

Table 6. Percentage of fluorescent survivors and abnormal phenotype per injected mRNA construct and zebrafish strain at 11 hpf compared to 1 dpf. mRNA constructs were linearized with ApaI prior to mRNA synthesis. The highlighted rows indicate mRNA injections corresponding to the embryos depicted in **Figure 8**. Red letters indicate whole plasmid injection. vs: versus, - : not recorded, n.a.: not assessed

mRNA construct	injected amount (ng)	fluorescent survivors at 11 hpf vs 1 dpf (%)		abnormal phenotype at 1 dpf (%) per strain	mRNA preparation preceding injection
		AB	WIK		
ndufb7-L1- ***Citrine***	0.45	n.a.	14 vs 2	-	heated (95°C) + snap cooled
	0.90	n.a.	10 vs 3	-	heated (95°C) + snap cooled
	1.00	few vs 0	n.a.	30 (AB)	heated (95°C) + snap cooled
	1.22	- vs 0	n.a.	ca. 95 (AB)	heated (95°C) + snap cooled
ndufb7-L1- ***Turquoise***	0.45	n.a.	17 vs 2	-	heated (95°C) + snap cooled
	0.90	n.a.	34 vs 3	-	heated (95°C) + snap cooled
	1.26	- vs 0	n.a.	ca. 89 (AB)	heated (95°C) + snap cooled
ndufb7-L1- ***Clover***	0.85	43 vs 2	n.a.	63 (AB)	thawed on ice
	1.00	many vs 1	n.a.	70 (AB)	heated (95°C) + snap cooled
ndufb7-L1- ***eGFP***	0.90	20 vs 0	n.a.	27 (AB)	heated (95°C) + snap cooled
ndufb7-L1- ***mEGFP***	0.90	15 vs 0	n.a.	2 (AB)	heated (95°C) + snap cooled
	1.20	0 vs 0	66 vs 4	8 (AB), 4 (WIK)	heated (95°C) + snap cooled
		7 vs 0	63 vs 8	18 (AB), 1 (WIK)	thawed on ice
	1.30	20 vs 0	51 vs 11	37 (AB), 3 (WIK)	thawed on ice
	> 1.30	n.a.	79 vs 26	37 (WIK)	heated (95°C) + snap cooled
	0.06	n.a.	82 vs 55	42 (WIK)	whole plasmid injection

One such abnormal, non-fluorescent AB embryo injected with *ndufb7-L1-Clover* mRNA is depicted in **Figure 8C**, together with non-fluorescent examples of other tested green fluorophores. In comparison, WIKs only presented with an abnormal phenotype >10 % when injected with more than 1.3 ng mRNA per embryo, or upon injection of plasmid DNA (see **Table 6, bottom line**). Uninjected AB and WIK control embryos generally appeared healthy and did not reflect the phenotypic abnormalities observed upon mRNA overexpression of GFP variants in the AB strain.

While all tested green fluorophores appeared problematic in the context of Ndufb7-tagging particularly in the AB strain, I observed fewer physical abnormalities with *ndufb7-L1-mEGFP* than with other GFP variants when injecting similar amounts of mRNA. At the same time, I achieved a higher score of fluorescent WIK embryos following *ndufb7-L1-mEGFP* mRNA injections compared to other GFP variants. mEGFP tops the list of most monomeric fluorescent proteins in a quantitative assessment of 63 different fluorophores (Cranfill *et al.*, 2016), making it an attractive candidate for tagging a nOXPHOS subunit.

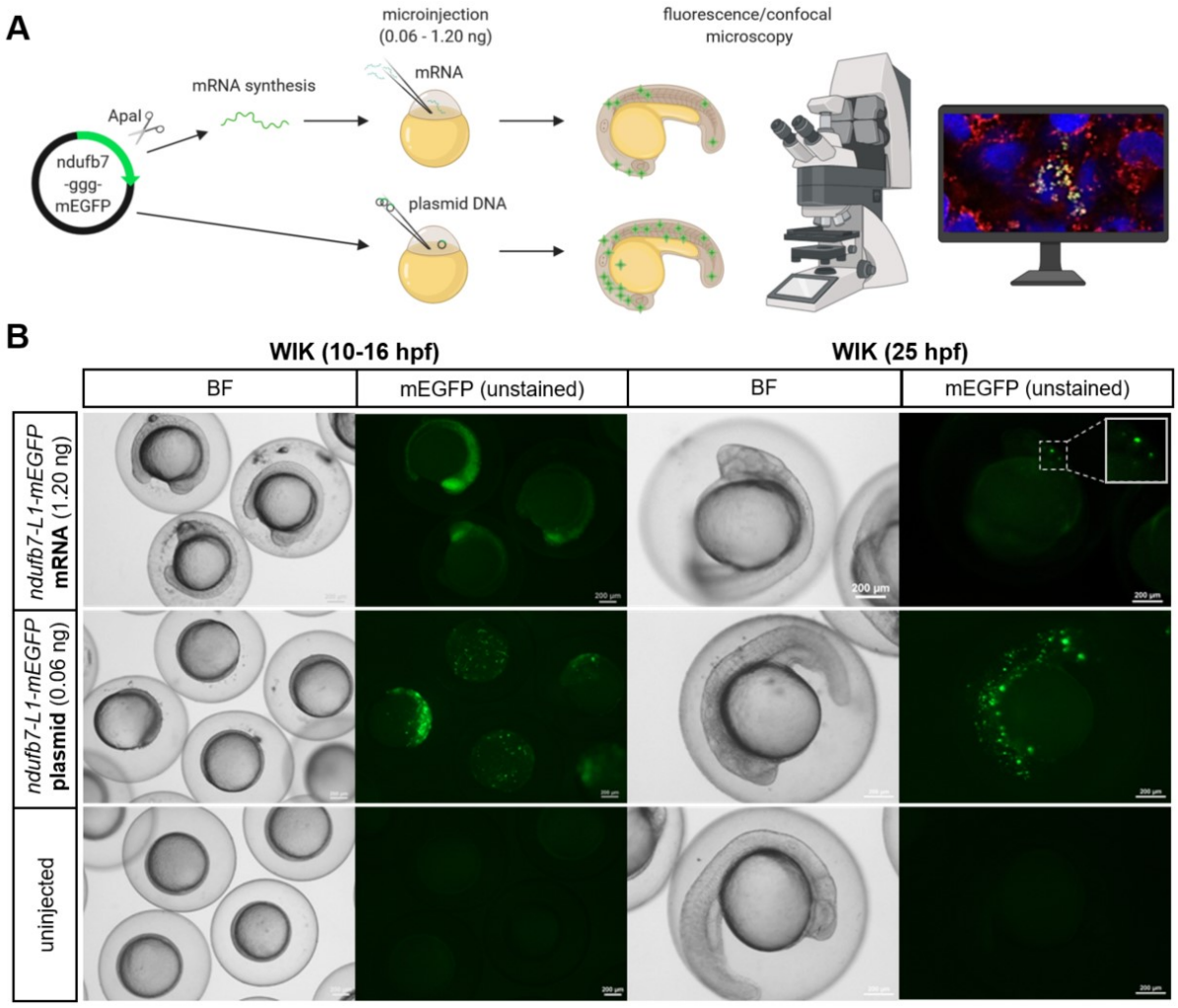


Figure 9. Comparison between *ndufb7-L1-mEGFP* mRNA and *ndufb7-L1-mEGFP* plasmid injections in WIK embryos. A the experimental scheme shows the differences in the procedure between mRNA and plasmid injections (created with BioRender). **B** comparison in fluorescence patterns following mRNA and plasmid injections at two major time points: 10 hpf (plasmid) or 16 hpf (mRNA) compared to 1 dpf. BF: bright field, scale bar: 200 µm.

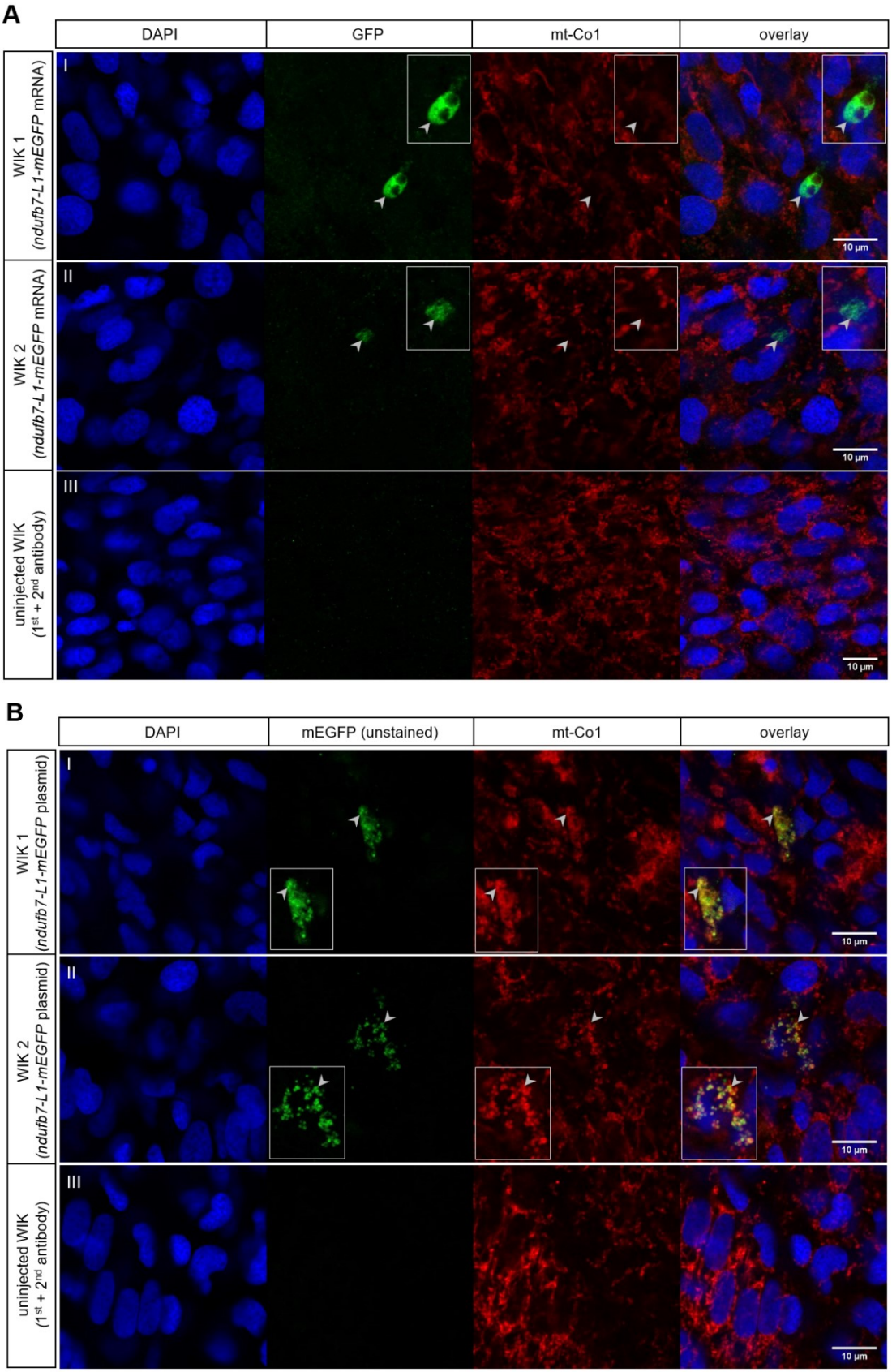


Figure 10. Confocal imaging of *ndufb7-L1-mEGFP* mRNA or plasmid injected WIK embryos at 1 dpf. A Co-staining of mEGFP (GFP antibody) with the mitochondrial marker mt-Co1 following *ndufb7-L1-mEGFP* mRNA injection suggests little to no mitochondrial co-localization. Grey arrows point to examples for non-overlapping signals for better comparison between channels. **B** Comparison of Ndufb7-L1-mEGFP expression following plasmid injection (no staining) with the mitochondrial marker mt-Co1 suggests mitochondrial localization of the tagged nOXPHOS subunit in the embryonic tail. Grey boxes: zoom-ins of regions marked with grey arrows. Blue: DAPI (nuclei), green: GFP antibody (A) or unstained mEGFP signal (B), red: anti-mtCo1 antibody (mitochondria). Scale bar: 10 µm.

To test whether overexpression from a double stranded DNA template (as opposed to less stable mRNA) results in expression of a more persistent fluorescent signal, I overexpressed *ndufb7-L1-mEGFP* by injecting the plasmid itself. The pCS2 backbone contains a CMV promoter element before the SP6 promoter, which is recognized in the zebrafish (Mella-Alvarado *et al.*, 2013). Following injection, I compared Ndufb7-L1-mEGFP expression from mRNA and plasmid by confocal microscopy. The two approaches are shown in **Figure 9** with examples of the resulting whole-embryo fluorescence patterns at 10-16 hpf and 1 dpf.

Judging by the number of abnormal survivors at 1 dpf (42 %), the injected plasmid amount of 0.06 ng/embryo was likely too high. Still, overexpression from the plasmid yielded: 1) a sufficient amount of physiologically normal fluorescent embryos to proceed with confocal imaging (**Figure 10**) and 2) a higher amount of fluorescence per embryo at 1 dpf when compared to embryos injected with *ndufb7-L1-mEGFP* mRNA. Thanks to these improvements, I was able to confirm mitochondrial localization of Ndufb7-L1-mEGFP in morphologically normal cells by mt-Co1 co-staining (**Figure 10B**, n = 37). In contrast, the sparse fluorescent signal observed following *ndufb7-L1-mEGFP* mRNA overexpression did not co-localize with mt-Co1 (**Figures 9B and 10A**, n = 26 faintly fluorescent embryos).

Considering that the issues associated with mRNA overexpression were largely resolved by switching to plasmid injection, we concluded that mEGFP is a suitable candidate for establishing a fluorophore-tagged *ndufb7* zebrafish model by CRISPR/Cas9-mediated knock-in.

4.2.3 CRISPR/Cas9-mediated knock-in of mRuby2 and mEGFP at the 3’ end of *ndufb7*

Having confirmed the suitability of two fluorophore candidates (mRuby2 and mEGFP) for differential C-terminal tagging of the two *ndufb7* alleles, I next tested the feasibility of CRISPR/Cas9 mediated fluorophore knock-in in the WIK strain, which is heterozygous for the two *ndufb7* alleles of interest. We chose the long and flexible linker 2 (L2) as connecting element, because it produced the highest fluorescence levels per embryo during mRNA overexpression experiments.

To test whether the genomic region surrounding the STOP codon of *ndufb7* is accessible to CRISPR/Cas9, I tested three different single guide RNA (sgRNA) candidates: sgRNA 1, which induces a cut 10 bp after the STOP signal, sgRNA 2 which induces a cut in the coding sequence of *ndufb7* 24 bp before the STOP signal and sgRNA 3, which targets the end of the 3’ UTR (91 bp after STOP). Upon injection of the individual sgRNAs complexed with Cas9 into 1-cell stage embryos, I isolated gDNA from ten single embryos at 1-5 dpf and amplified the target region with the primers indicated in **Figure 11B**. T7 Endonuclease 1 assay (T7 assay; Mashal et al., 1995) showed a cutting efficiency of 90 % for sgRNA 1 and sgRNA 2, and 100 % cutting efficiency for

sgRNA 3 (**Figure 11C**). With a cutting efficiency above the anticipated 80 %, all three sgRNA candidates were suitable for knock-in.

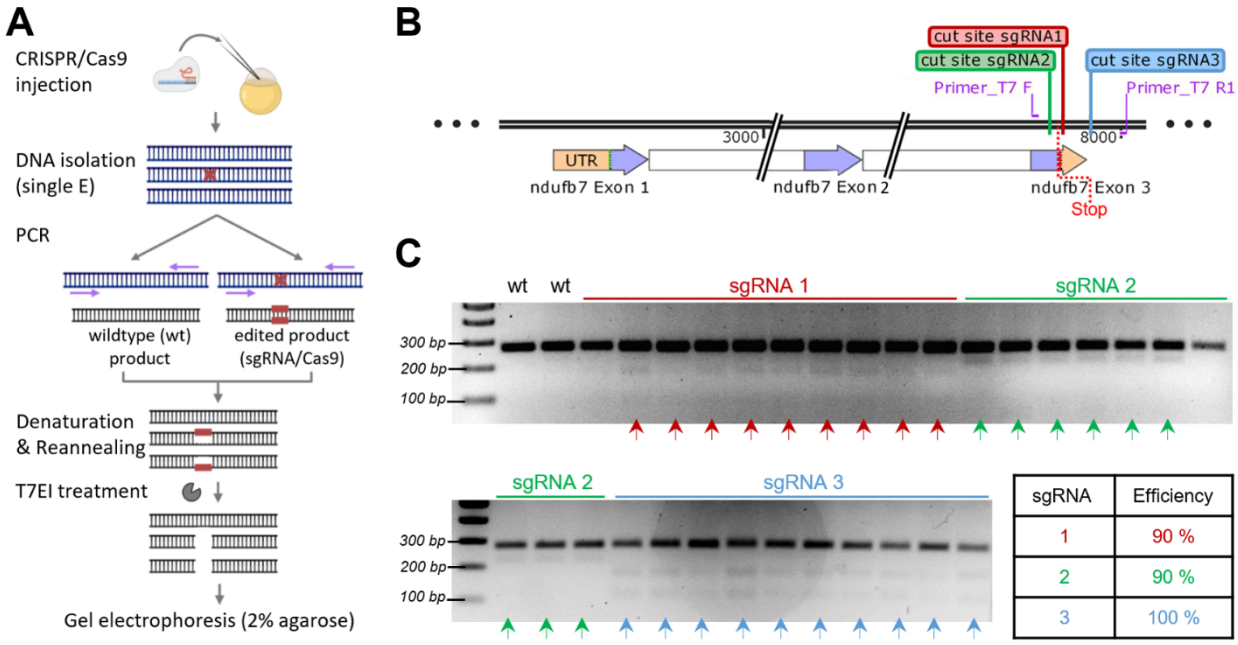


Figure 11. sgRNA efficiency test for CRISPR/Cas9-mediated C-terminal tagging of *ndufb7* by T7 Endonuclease 1 (T7E1) assay. A theoretic concept of the T7E1 assay; generated with BioRender. In brief: following injection of Cas9 complexed with a sgRNA at 1 cell-stage, genomic DNA (gDNA) is isolated from single embryos and the region of interest is amplified separately from wildtype and CRISPR/Cas9-modified gDNA. Wt and CRISPR/Cas9 edited PCR products are mixed, reannealed, treated with T7E1 and cutting efficiency is visualized by gel electrophoresis. **B** genomic locus of the zebrafish *ndufb7* gene, including cutting sites of the tested sgRNAs 1, 2 and 3 and PCR primer binding sites used for amplification of the target region. **C** T7E1 assay results for sgRNA 1 (90 % cutting efficiency, expected bands at 183 + 92 bp), sgRNA 2 (90 % cutting efficiency, expected band at 220 bp; 55 bp band not visible) and sgRNA 3 (100 % cutting efficiency, expected bands at 170 + 105 bp), unmodified wildtype band: 275 bp.

To achieve fluorophore knock-in at the 3' end of *ndufb7*, I generated repair templates containing the coding sequence of either L2-mRuby2 or L2-mEGFP, flanked by two long homology arms complementary to the genomic region surrounding the sgRNA target sites in the WIK strain (5' homology arm: 828 bp, 3' homology arm: 1017 bp). A schematic depiction of the L2-mRuby2 repair template is shown in **Figure 14B**. Each repair template was digested with I-SceI enzyme and injected into one-cell stage embryos in combination with CRISPR/Cas9 RNP or mRNA (sgRNA1 + sgRNA2).

I performed L2-mRuby2 and L2-mEGFP knock-in either by injection of one type of repair template per embryo or via double knock-in of the two fluorophores. The knock-in strategy is outlined in **Figure 12A**. In most cases, the resulting fluorescence resembled the speckled expression patterns of mRNA injections discussed previously at 1 dpf, except for one double KI embryo which expressed mEGFP seemingly throughout the entire body. Unfortunately, this particular embryo died shortly thereafter (**Figure 12B, row 1**). Examples for the speckled fluorophore expression observed at 1 dpf are shown in **Figure 12B, row 2** in the case of mEGFP + mRuby2 double KI and in **Figures 14A and 15A, B** for mRuby2 KI alone.

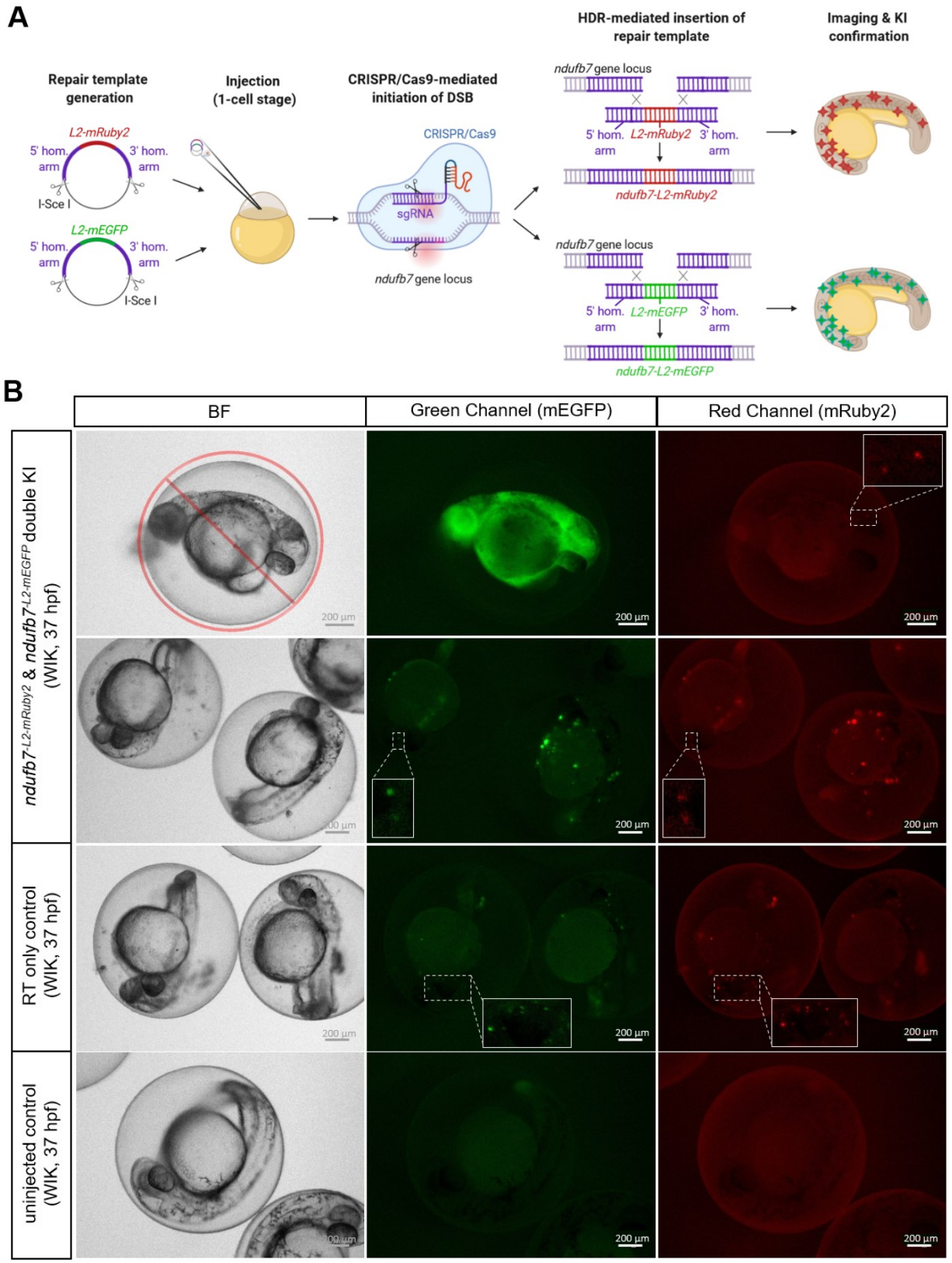


Figure 12. CRISPR/Cas9 + HDR-mediated knock-in of L2-mRuby2 and L2-mEGFP at the C-terminus of *ndufb7* in WIK embryos. A scheme of the experimental knock-in procedure, created with BioRender. DSB: double strand break, HDR: homology directed repair, hom. arm: homology arm **B** live imaging examples for fluorescence observed at **1 dpf** following double knock-in (KI) of *L2-mEGFP* and *L2-mRuby2* at the C-terminus of *ndufb7* exon 3 in WIK embryos, compared to repair template (RT) only injection and uninjected control. The zoom boxes highlight some of the regions with likely overlapping mRuby2 and mEGFP signals. The highly fluorescent embryo in row 1 died by 2 dpf. This figure is extended in **Figure 13**. BF: bright field, scale bar: 200 µm.

Interestingly, a large amount of the resulting fluorescent signal was observed in the skin cells at the embryo surface and covering the yolk. Surprisingly, all initially fluorescent embryos (1 dpf) lost fluorophore expression by 5 dpf (see **Figures 12B and 13** for direct comparison of time points 1 dpf and 3 dpf).

As shown in **Figure 12B**, control embryos injected with only the repair template (without CRISPR/Cas9) were expressing the fluorophore at similar levels compared to KI animals suggesting that there may be random integration of the repair template. I therefore proceeded to check for correct integration of the fluorophore both at the gDNA and mRNA level in embryos injected with the L2-mRuby2 repair template and sgRNA1+2/Cas9 injection solution.

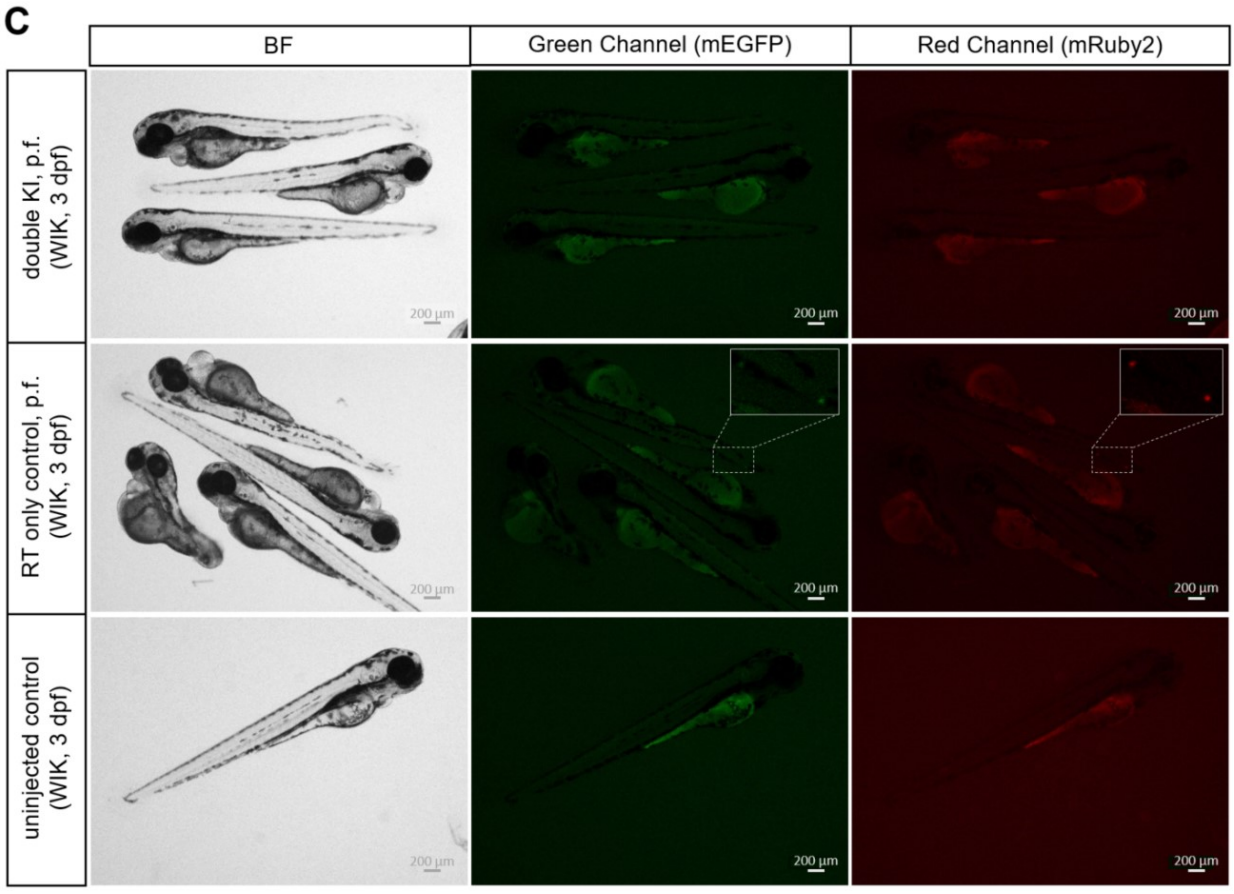


Figure 13. Loss of fluorescence following CRISPR/Cas9 + HDR-mediated double knock-in of *L2-mRuby2* and *L2-mEGFP* at the C-terminus of *ndufb7* in WIK embryos at 3 dpf (extension of Figure 12). The zoom boxes highlight some a region with overlapping mRuby2 and mEGFP signals. KI: knock-in, RT: repair template BF: bright field, scale bar: 200 µm, p.f.: previously fluorescent embryos.

From 266 injected embryos, 51 were fluorescent at 1 dpf (28 physiologically normal, kept until 5 dpf + 23 deformed embryos, collected/discarded). In total, I was able to generate sufficient product for Sanger sequencing by PCR amplification from 3 out of 12 previously fluorescent single embryos collected at 5 dpf. The three sequenced embryos, for one of which correct integration of L2-mRuby2 was confirmed as described below, had a slightly abnormal phenotype at the time point of collection: all three had a shorter tail than their physiologically normal siblings did.

After having generated enough product for sequence verification by Sanger sequencing in two consecutive rounds of gDNA PCR amplifications, I was finally able to confirm successful knock-in at the gDNA level in at least one previously fluorescent embryo collected at 5 dpf (see **Figure 14B** and **Suppl. Information 1** for alignment of the entire sequence of embryo 1). As observed for the double KI, the fluorescent signal had largely disappeared by the time of gDNA extraction at 5 dpf in all collected embryos. Regardless, I was able to confirm correct integration at the 3' end of *ndufb7* for one formerly fluorescent embryo (embryo 1 in **Figure 14B**) by amplification with a forward primer that binds inside the repair template (in *mRuby2*) in combination with a reverse primer that binds outside the repair template in the 3' downstream genomic sequence of the *ndufb7* gene. Thus, locus-specific knock in is indeed feasible. However, it should be noted that for the other checked embryos (such as embryo 4) I did not achieve PCR amplification and sequencing far outside the 3' homology arm of the repair template. The sequence surrounding the 5' homology arm in intron 2 of *ndufb7* was not assessed.

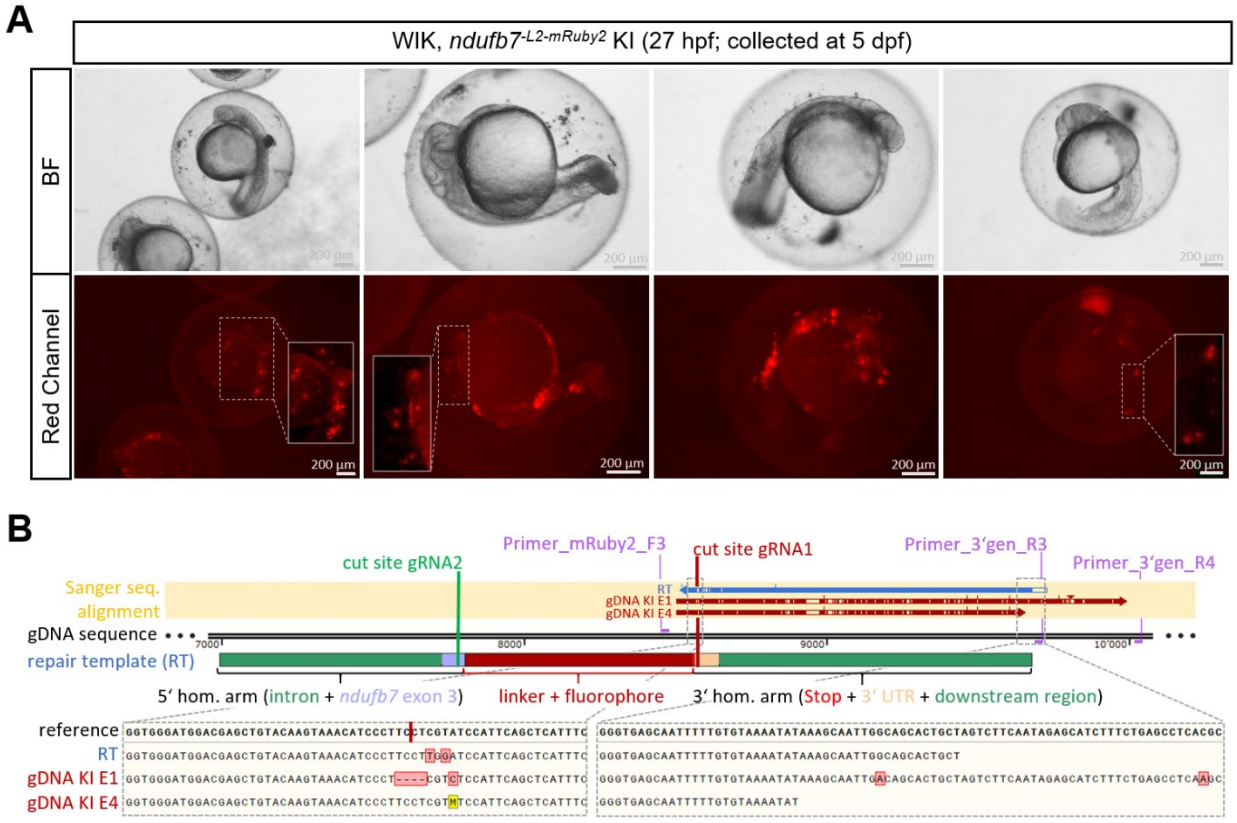


Figure 14. Confirmation of *L2-mRuby2* KI success at the C-terminus of *ndufb7* (gDNA level). A 1 dpf WIK embryos following CRISPR/Cas9 + HDR-mediated knock-in of *L2-mRuby2*. All embryos lost fluorescence by day 5 and were collected for confirmation of correct KI at the *ndufb7* genomic locus. BF: bright field, scale bar: 200 µm. **B** fluorophore integration confirmed by alignment of the PCR amplified and sequenced genomic region surrounding the STOP codon of the *ndufb7* gene (gDNA isolated from formerly fluorescent embryos at 5 dpf). Primer binding sites inside (F3) and outside (R3, R4) the repair template used for amplification and sgRNA cutting sites are shown. Alignments, shown as red and blue arrows and within the grey dotted boxes (performed in SnapGene) show that KI embryo 1 (E1) contains L2-mRuby2 at the correct location, whereas locus specificity remains unclear for KI embryo 4 (E4). Corresponding full alignments are provided in **Suppl. Information 1**. KI: knock-in; RT: repair template; UTR: untranslated region.

In both sequenced embryos (1 and 4) depicted in **Figure 14B**, the recombination event appears to have happened directly at the sgRNA 1 target site, because the mutations introduced into the

repair template at this locus were not present in the sequenced PCR amplicons. I observed several mismatches between repair template (reference) sequence and the sequence obtained from knock-in embryos downstream of the sgRNA target site in the 3' homology arm (see **Suppl. Information 1**), which further supports this theory.

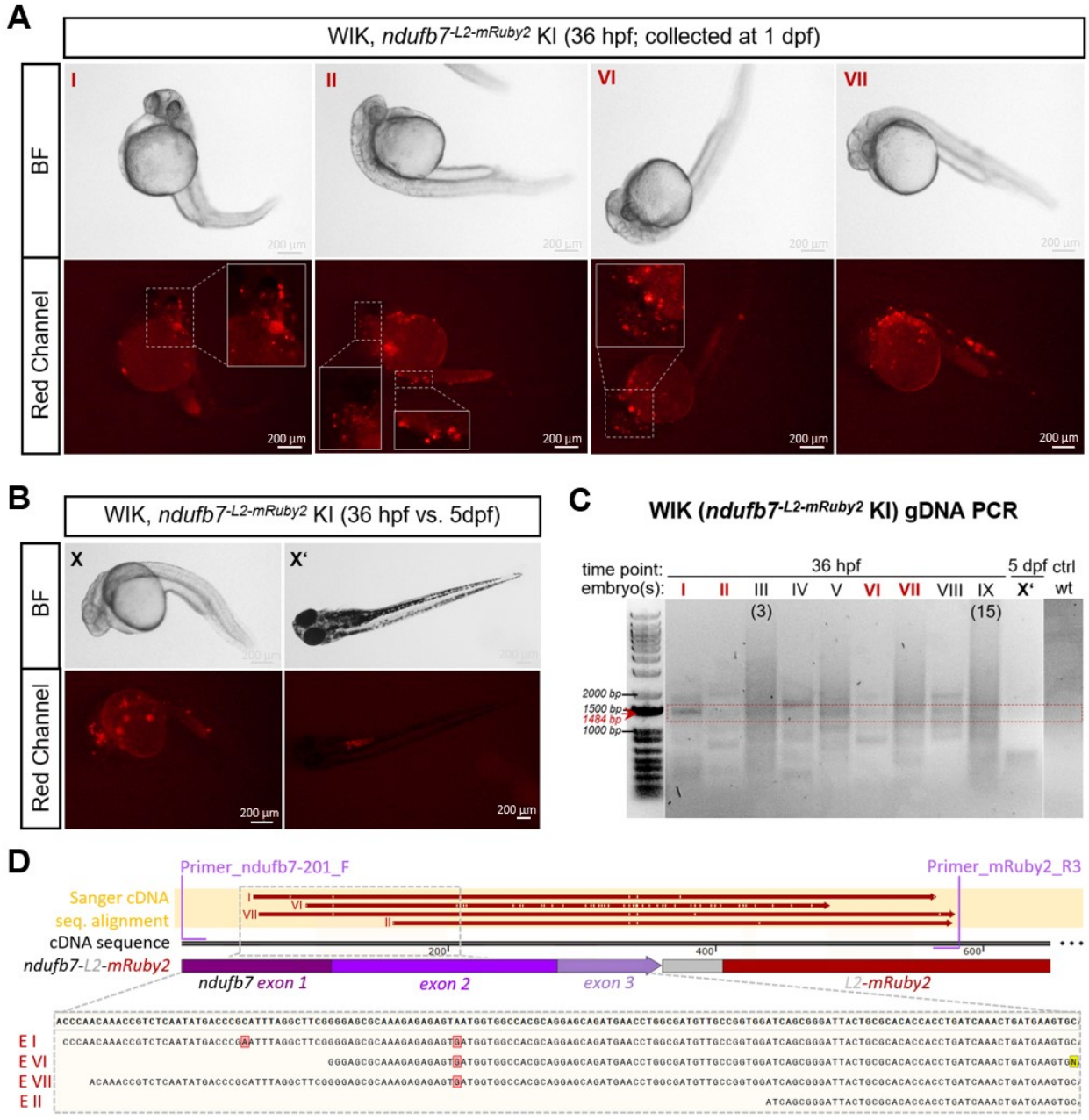


Figure 15. Proof for ndufb7-L2-mRuby2 in vivo transcription one day after fluorophore KI at the C-terminal of ndufb7 in WIK embryos (mRNA level KI confirmation). A embryos collected for simultaneous mRNA and gDNA extraction one day after knock-in of L2-mRuby2. BF: bright field, scale bar: 200 µm. **B** formerly fluorescent (1 dpf) embryo collected for gDNA extraction at 5 dpf after knock-in of L2-mRuby2. BF: bright field, scale bar: 200 µm. **C** gDNA level confirmation of L2-mRuby2 KI at the C-terminus of ndufb7 by PCR amplification and gel electrophoresis. The expected band size in embryos containing the fluorophore at the correct genomic location is highlighted in red. Bold lane labels correspond to the respective embryos depicted in **A, B** and **D**. **D** confirmation of *ndufb7-L2-mRuby2* expression following KI, by alignment of the PCR amplified and sequenced cDNA sequence spanning several exons of ndufb7, the linker sequence and a part of the fluorophore. mRNA used for cDNA generation was isolated from fluorescent embryos depicted in **A** at 1 dpf. Primer binding sites are shown in purple. Alignments, shown as red arrows and within the grey dotted box, were performed in SnapGene. Corresponding full alignments are provided in **Suppl. Information 2**. E: embryo, F: forward primer, R: reverse primer.

Having confirmed that correct integration is feasible, although not very efficient, I injected the same L2-mRuby2 KI mixture as above and extracted mRNA (simultaneously with gDNA) from single fluorescent embryos at 1 dpf. As shown in **Figure 15**, I was able to confirm expression of the correctly (in-frame) tagged *ndufb7* gene by amplification and sequencing of cDNA from 4 out of 8 single embryos collected at 1 dpf (embryos I, II, VI and VII). The alignment of the complete sequences is provided in **Suppl. Information 2**. While embryo I shows some phenotypical abnormalities, embryos II, VI and VII appear healthy (**Figure 15A**).

After having proven that fluorophore KI at mRNA and gDNA level is feasible, I reinjected several hundred WIK embryos for KI of either *L2-mRuby2* or *L2-mEGFP* individually or in combination. Promising fluorescent, physiologically normal embryos were selected at 1 dpf (mRuby2 KI: 34 embryos, mEGFP KI: 10 embryos, double KI: 51 embryos) and are currently being raised to adulthood for determination of KI success, germline transmission and to analyze whether fluorophore-tagging of Ndufb7 interferes with OXPHOS complex assembly in the F1 generation.

4.3 Strategy II: Differential short epitope-tagging of two naturally occurring *ndufb7* variants

As alternative tagging approach for the two different *ndufb7* alleles present in the WIK strain, we selected the short epitope tags HA and FLAG, both of which had previously been used to tag components of the OXPHOS machinery (Zhang *et al.*, 2020). HA and FLAG are short peptides consisting of 8-9 amino acids, with a molecular weight of 1.10 kDa and 1.01 kDa respectively. Due to their size, short epitope tags are less likely to interfere with protein function and/or complex assembly compared to fluorophores (Kimple, Brill and Pasker, 2013; Lotze et al., 2016). Besides addressing the size-issue of fluorophore-tags, this strategy further circumvents potential complications related to phototoxicity and fluorophore multimerization. While differentiation of allelic expression tendencies by live imaging is not possible with short peptide tags, tag-specific sequence amplification at the DNA and mRNA level and antibody-based differentiation between the two tagged allelic variants at the protein level (*in vivo* and *in vitro*) remain an option.

I followed two approaches to validate the suitability of single short epitope tags for differentiation between the two *ndufb7* alleles: 1) transient *ndufb7-HA* and *ndufb7-FLAG* mRNA overexpression and 2) generation of a transgenic line to stably overexpress *ndufb7-HA* and *ndufb7-FLAG* in combination with a selection marker. Both methods were combined with immunostaining and confocal imaging to confirm mitochondrial localization of the tagged OXPHOS subunit, and to confirm detectability of the single tags *in vivo*.

4.3.1 Transient overexpression of *ndufb7^{H28}-HA* and *ndufb7^{N28}-FLAG* by mRNA injection

In line with the fluorophore-tagging approach (strategy I), I simultaneously generated mRNA plasmids for both allelic variants of *ndufb7*, C-terminally tagged with a single HA or FLAG tag. I then linearized pCS2 plasmids containing the correct *ndufb7^{H28}-HA* or *ndufb7^{N28}-FLAG* coding sequence with ApaI, followed by SP6-mediated mRNA synthesis and microinjection into WIK embryos at 1-4 cell stage.

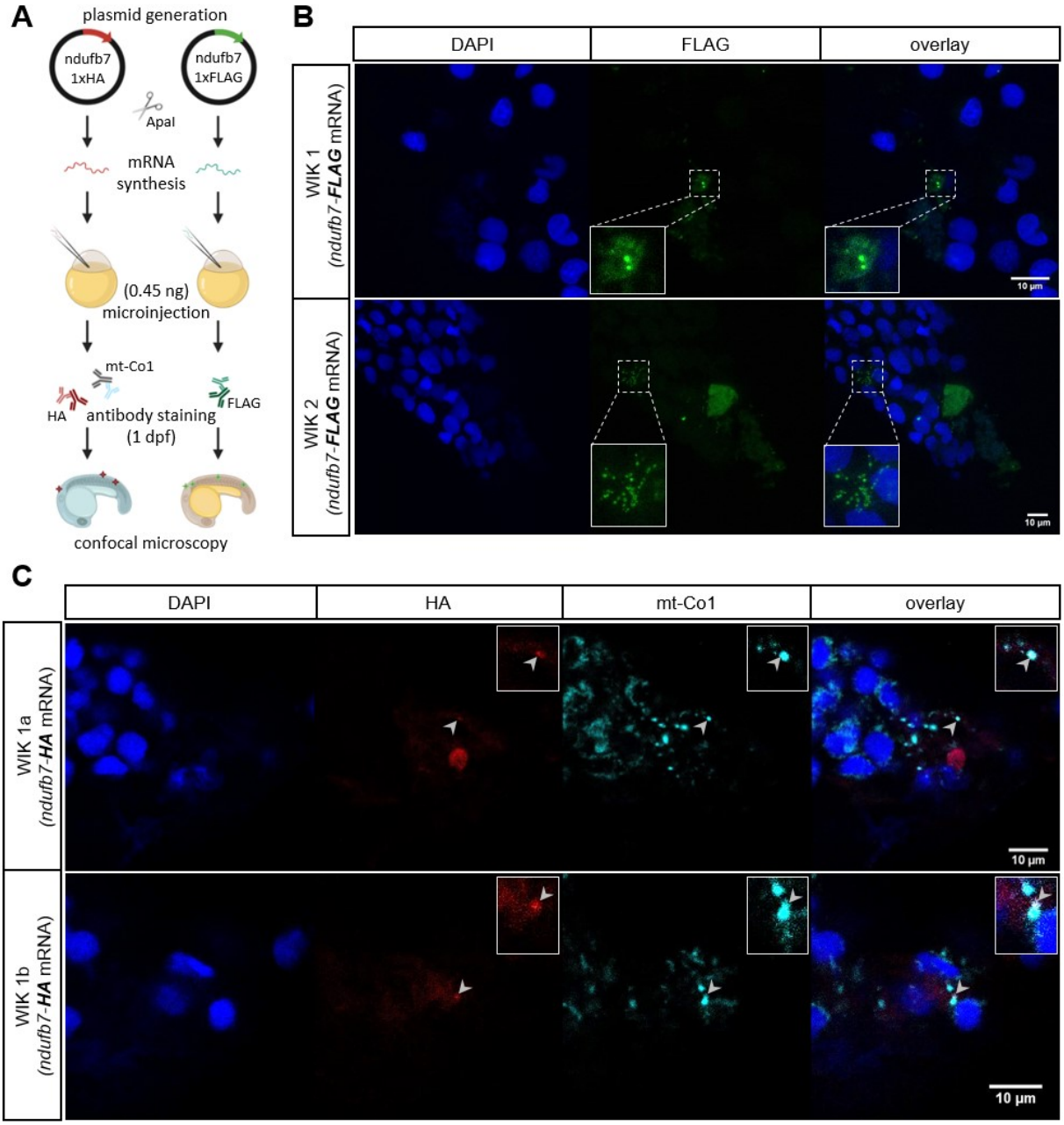


Figure 16. *ndufb7^{H28}-HA* and *ndufb7^{N28}-FLAG* mRNA overexpression in WIK strain embryos. A Schematic depiction of the experimental procedure generated with BioRender. **B** Confocal imaging of two *ndufb7^{N28}-FLAG* mRNA injected, immunostained embryos at 1 dpf. Imaged region: embryonic surface, maximum projection. **C** Confocal imaging of an *ndufb7^{H28}-HA* mRNA injected, immunostained embryo at 1 dpf. Grey arrows mark overlapping Ndufb7-HA and mt-Co1 signals, shown as zoom-ins in grey boxes. Imaged region: embryonic surface. Blue: DAPI (nuclei), green: FLAG (Ndufb7-FLAG), red: HA (Ndufb7-HA), turquoise: mt-Co1 (mitochondria), scale bar: 10 µm. *Confocal images (B, C) were taken by I. Piragyte-Langa.*

Injection of 0.45 ng heated and snap-cooled mRNA per embryo (which approximately corresponds to 1.30 ng *ndufb7-L2-mRuby2* mRNA in terms of number of mRNA molecules), neither had a noticeable negative impact on embryo survival nor caused an abnormal phenotype.

Per tag-type, 100 embryos were collected at 1 dpf and stained for either HA in combination with the mitochondrial marker mt-Co1 or for FLAG alone. Due to incompatibility of the anti-FLAG antibody with the anti-mt-Co1 antibody, these two antibodies could not be used in combination. The experimental approach is outlined in **Figure 16A**. Dr. Piragyte-Langa and I then pre-selected fluorescent embryos for confocal imaging by whole-embryo fluorescence microscopy. However, the amount of fluorescent signal was very low: with both, the fluorescence stereoscope and the confocal microscope only a few fluorescent speckles were detectable at the embryonic surface. Two such examples are shown in **Figure 16B** (FLAG) and **Figure 16C** (HA and mt-Co1). While we observed some overlap between HA and mt-Co1 signals, as marked with grey arrows in **Figure 16C**, the observed signal largely does not resemble the expected mitochondrial expression pattern.

At this point, it was unclear whether attachment of a single HA- or FLAG-tag is insufficient for visualization of tagged Ndufb7 in general (the small tag may simply become inaccessible after protein folding) or whether technical problems may have occurred during mRNA synthesis, injection or the immunostaining procedure.

4.3.2 Stable overexpression of *ndufb7^{H28}-HA* and *ndufb7^{N28}-FLAG* by Tol2-mediated transgenesis

Because mRNA overexpression turned out to be unsuitable for validation of the short-tag approach, I next designed a stable overexpression model under Dr. Piragyte-Langas supervision which enables: 1) pre-selection of *ndufb7^{H28}-HA* or *ndufb7^{N28}-FLAG* expressing embryos prior to immunostaining thanks to expression of a selection marker, 2) a larger time window for experimental analysis, and 3) generation of sufficient material for analysis of OXPHOS complex assembly by Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE).

I cloned the coding sequence for *ndufb7^{H28}-HA* or *ndufb7^{N28}-FLAG* under the control of a Ubiquitin-B promoter element (*ubb*) into a *Tol2* transposon vector, followed by injection into 1-cell stage WIK embryos in combination with *tol2* mRNA (see **Figure 17A**, (Kawakami, 2007)). The vector contains the coding sequence for the fluorophore mCerulean under the control of an α-crystallin (*cryaa*) promoter. Following early Tol2-mediated random integration into the genome, *cryaa* induces expression of the fluorophore in the embryonic lens from 2 dpf onwards, thus making it a suitable selection marker (Zou *et al.*, 2015). The *ubb* promoter reportedly induces expression in all major zebrafish organs from as early as 2.75 hpf onwards (Mosimann *et al.*, 2011). Following injections at 1-cell stage, 153 *Tg(ubb:ndufb7-HA, cryaa:mCerulean)* and 96

Tg(ubb:ndufb7-FLAG, cryaa:mCerulean) embryos were selected according to mCerulean expression in the lens. Per tag-type, 20 mCerulean-positive embryos were deyolked at 2 dpf, fixed and immunostained for HA + mt-Co1 or FLAG respectively, together with plasmid only injected or uninjected embryos as controls. The residual 133 *Tg(ubb:ndufb7-HA, cryaa:mCerulean)* and 76 *Tg(ubb:ndufb7-FLAG, cryaa:mCerulean)* mCerulean-positive embryos were raised to adulthood.

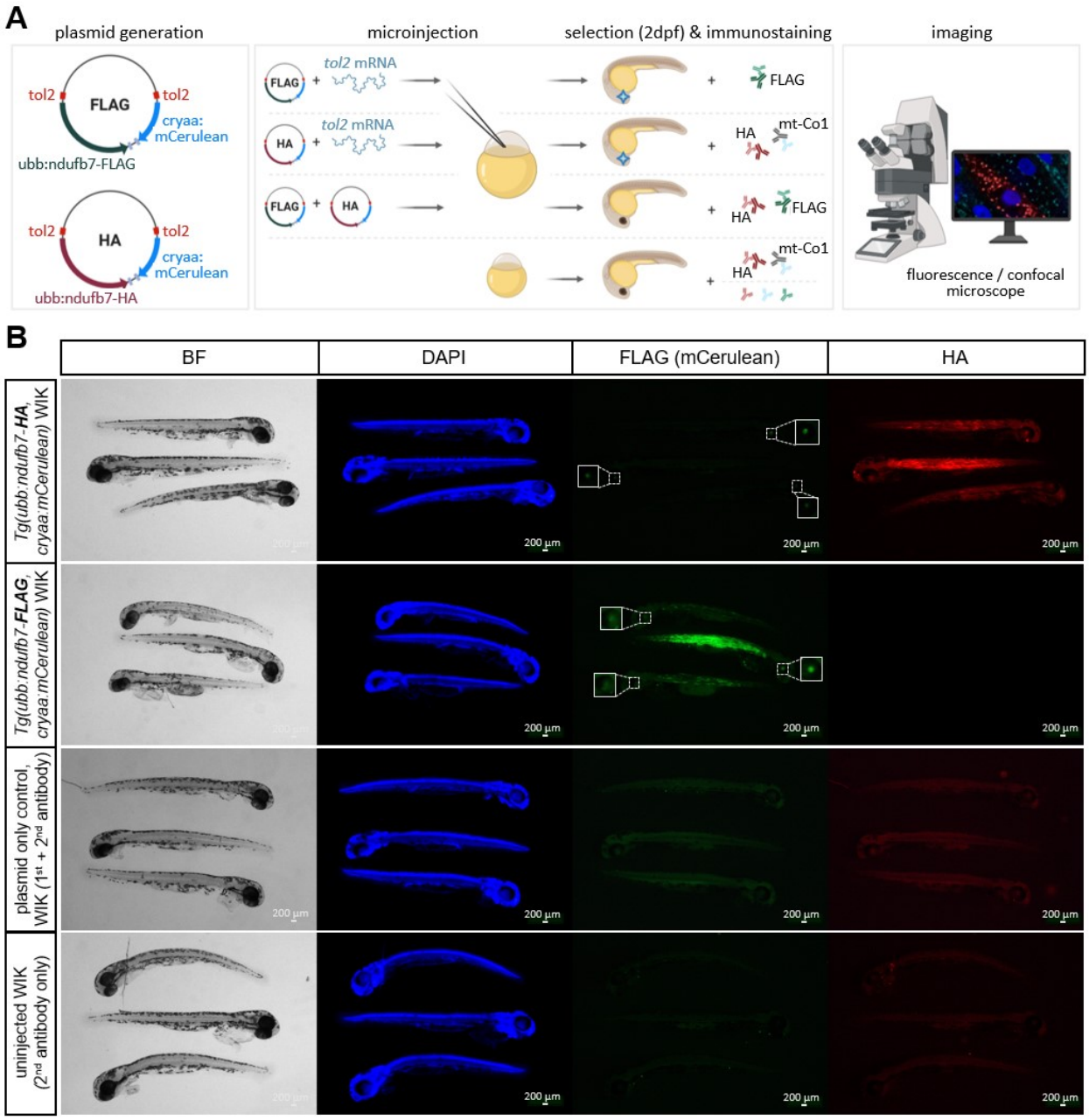


Figure 17. Generation and immunostaining of transgenic WIK lines stably overexpressing *ubb:ndufb7^{H28}-HA* or *ubb:ndufb7^{N28}-FLAG* in the presence of a *cryaa:mCerulean* selection marker. A Schematic depiction of the experimental procedure generated with BioRender. **B** Deyolked *Tg(ubb:ndufb7-HA, cryaa:mCerulean)* and *Tg(ubb:ndufb7-FLAG, cryaa:mCerulean)* embryos at 2 dpf, immunostained with antibodies against either HA or FLAG (or both, in the case of controls). The green dots in the green channel in row 1 represent mCerulean expression in the lens (unstained). BF: bright field, blue: DAPI (nuclei), green: FLAG (Ndufb7-FLAG) and mCerulean, red: HA (Ndufb7-HA). Scale bar: 200 µm.

The three immunostained example embryos per sample type, depicted in **Figure 17B,** show that expression levels of the *ubb:ndufb7^{H28}-HA* and *ubb:ndufb7^{N28}-FLAG* transgenes vary among the injected embryos. At 2 dpf, transgene expression is most prominent in tail muscle tissue in all cases. Expression of the selection marker mCerulean in the lens can be seen in the green channel (FLAG, mCerulean) in **Figure 17B row 1** and **2**. Confocal imaging of the embryonic tail muscle revealed that the majority of Ndufb7^{H28}-HA signal overlaps with mt-Co1, thus confirming mitochondrial localization and detectability of the HA-tagged nOXPHOS subunit (**Figure 18**).

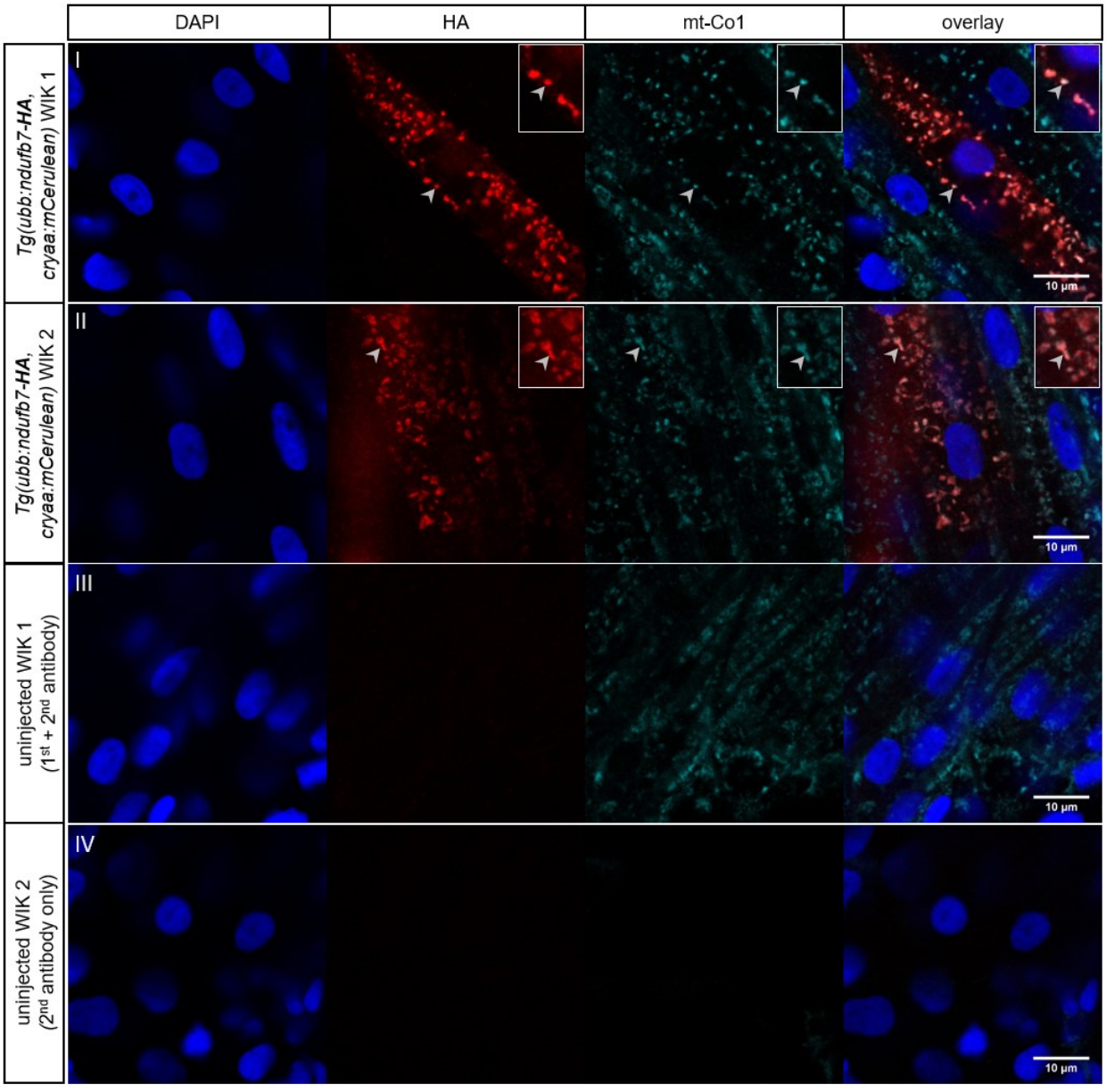


Figure 18. Confocal imaging of *Tg(ubb:ndufb7-HA, cryaa:mCerulean)* WIK tail muscle cells following immunostaining for HA (red) and the mitochondrial marker mt-Co1 (turquoise) at 2 dpf. Grey arrows point to examples of overlapping Ndufb7^{H28}-HA and mt-Co1 signal. Panels III, IV: controls, DAPI: blue (nuclei), scale bar: 10 µm.

While mitochondrial localization of Ndufb7^{N28}-FLAG could not be definitely confirmed by confocal imaging (**Figure 19**) due to incompatibility of the anti-FLAG and anti-mt-Co1 antibodies, the expression pattern closely resembles that of Ndufb7^{H28}-HA (**Figure 18**), suggesting

mitochondrial localization of the tagged nOXPHOS subunit. Furthermore, the confocal imaging results prove that both, a single HA-tag as well as a single FLAG-tag are sufficient for visualizing the two allelic variants of Ndufb7 by immunostaining.

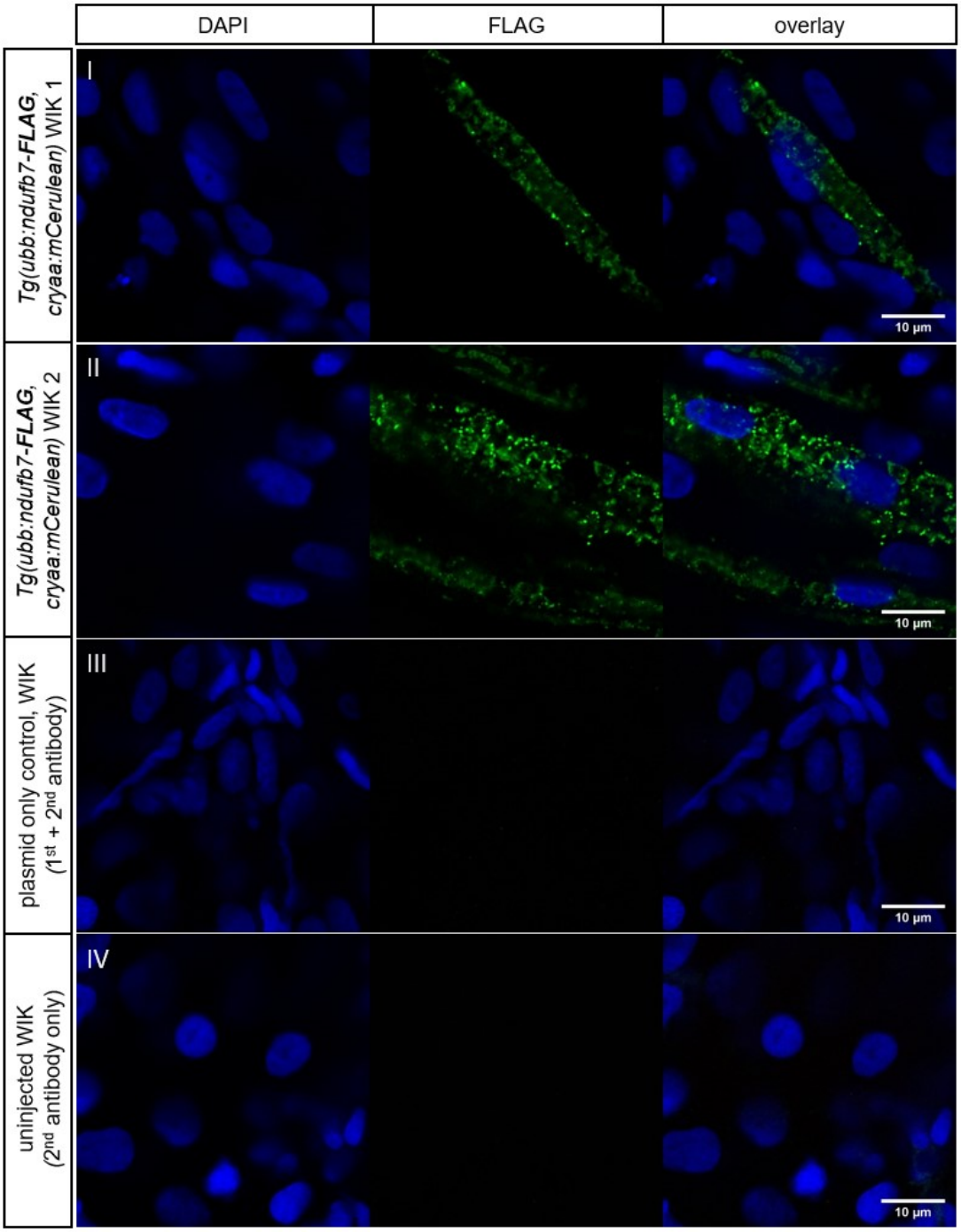


Figure 19. Confocal imaging of *Tg(ubb:ndufb7-FLAG, cryaa:mCerulean)* WIK tail muscle cells following immunostaining for FLAG (green) at 2 dpf. Panels III, IV: controls, DAPI: blue (nuclei), scale bar: 10 µm.

Having validated detection and mitochondrial localization of (very likely) both tagged Ndufb7 variants, the transgenic plasmids were injected once more, followed by collection of mCerulean-positive embryos, mitochondria extraction, BN-PAGE and Western Blotting (data not reported). By doing so, I hoped to confirm the integration of the tagged Ndufb7 variants into complex I. However, it quickly became clear that a single injection round does not yield sufficient material

for BN-PAGE. Hence, this experiment will be repeated with injected F0 adults or in the F1 generation of the *Tg(ubb:ndufb7-HA, cryaa:mCerulean)* and *Tg(ubb:ndufb7-FLAG, cryaa:mCerulean)* mCerulean-positive embryos that are currently being raised to adulthood.

4.4 Strategy III: Differentiation of a wildtype (K84) *ndufb7* allele and a disease variant (W84) *ndufb7* allele by CRISPR/Cas9-mediated mutagenesis

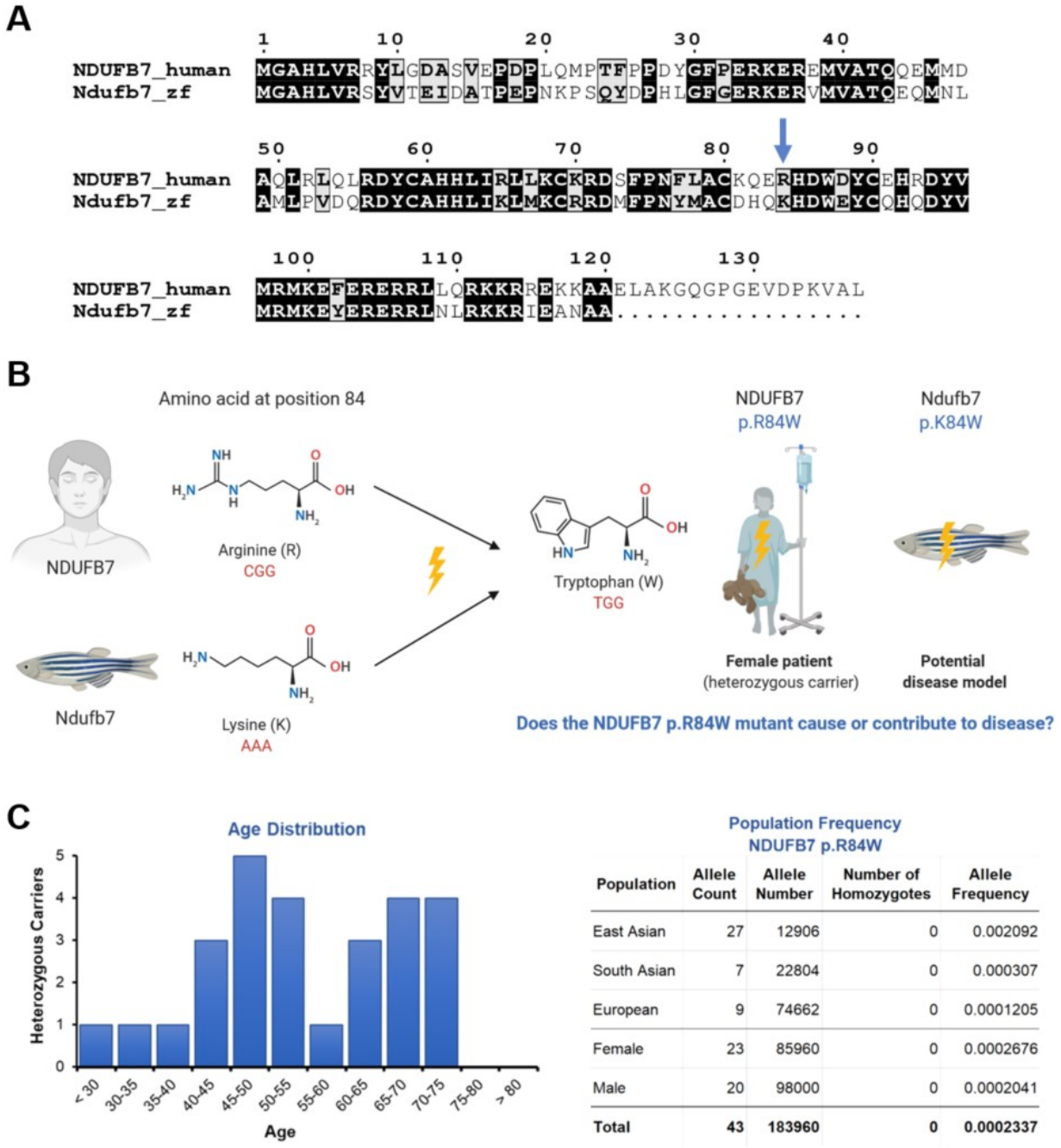
Strategies I and II bring several advantages in terms of experimental design. However, their major strength is also a potential disadvantage, the consequence of which can only be evaluated over the lifetime of the final, established animal model(s): a tag introduced at the genomic locus of *ndufb7*, however big or small, is artificial and intrusive by nature. The uncertainty whether the tagged Ndufb7 subunit remains functional or whether it may potentially disturb the OXPHOS system are valid concerns. Furthermore, it is unclear if the identified naturally occurring SNP in *ndufb7* (H28>N) disturbs or enhances the OXPHOS system enough to trigger a potentially existing allelic selection mechanism. To address these uncertainties, we devised a third strategy that allows approaching the research question from a different angle and is meant to complement strategies I and II: instead of differentiation between two tagged, naturally occurring *ndufb7* alleles in wildtype animals, a wildtype allele is compared to a (likely) disease causing allelic variant with a single amino acid-difference introduced by CRISPR/Cas9-mediated mutagenesis.

Mutagenesis does not allow for antibody-based detection, live imaging or, most probably, sequence-specific PCR amplification, which complicates the analysis significantly. However, in contrast to the animal models resulting from strategies I and II, the mutagenesis model system is non-intrusive thanks to the absence of an artificial tag, thereby avoiding any unexpected, potentially negative impact a tag may have on the OXPHOS system. In addition, the mutagenesis model is potentially useful for studying human disease and thus could facilitate a broader use of the established animal model in the future.

4.4.1 Modeling the potentially disease-causing human SNP variant NDUFB7^{R84W} in zebrafish to mimic an unpublished patient case

According to a general search of the Genome Aggregation Database (gnomAD), 99 missense and predicted loss of function (pLoF) SNPs have been identified in the coding sequence of human *NDUFB7* by genome and exome sequencing of various human populations (Karczewski *et al.*, 2020). Out of these 99 SNPs leading to an amino acid change in NDUFB7, only 3 were found in homozygous condition: p.R93H (5 homozygotes detected), p.K118T and p.K118E (4 homozygotes detected each). However, apart from a recently described disease causing intronic

SNP (Correia *et al.*, 2021), there is no record of pathological SNPs in *NDUFB7* according to literature and databases such as ClinVar, OMIM and the Human polymorphisms and disease mutations index (UniProt).



Population Frequency
NDUFB7 p.R84W

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
East Asian	27	12906	0	0.002092
South Asian	7	22804	0	0.000307
European	9	74662	0	0.0001205
Female	23	85960	0	0.0002676
Male	20	98000	0	0.0002041
Total	**43**	**183960**	**0**	**0.0002337**

Figure 20. Comparison of human and zebrafish Ndufb7 protein sequence in the context of modeling the NDUFB7^{R84W} patient mutation in the zebrafish. A protein alignment of human and zebrafish Ndufb7 orthologues. The blue arrow marks amino acid position 84. Black highlighted letters: perfect alignment, grey highlighted letters: amino acids with similar characteristics. **B** Schematic depiction of the human and zebrafish 84W mutation in the context of establishing a potential disease model. Generated with BioRender. **C** Age distribution and population frequency of the NDUFB7^{R84W} mutation according to the public gnomAD database.

I therefore submitted a search request for *NDUFB7* in GeneMatcher, a search engine designed to connect researchers and clinicians with an interest in the same gene (Sobreira *et al.*, 2015). Shortly thereafter, I was contacted by a senior research specialist based in the United States,

who reported a female patient with a heterozygous missense SNP in exon 2 of *NDUFB7* which induces an amino acid change at position 84 in the peptide sequence (NM_004146:exon2:c.C250T:p.R84W).

The patient was described to have features of atypical Angelman or Rett syndrome, including a number of neurological symptoms: failure to thrive, microcephaly, cortical blindness, mild hearing loss, sleep apnea, asthma, poor feeding, gastroesophageal reflux disease (GERD), nephrotic syndrome, contractures, scoliosis, seizures, profound intellectual disability, aggressive, self-injurious behavior, and eosinophilic esophagitis. However, it is presently unclear whether, and to what extent, the SNP causing the R84W amino acid change in NDUFB7 contributes to the disease phenotype.

According to gnomAD, this particular SNP has been recorded 43 times (number of homozygotes: 0), most frequently in the Asian population. The corresponding age distribution of heterozygous carriers and allelic frequencies obtained via the gnomAD database are shown in **Figure 20C**. To investigate the potential of modeling the NDUFB7^{R84W} patient mutation in zebrafish, I aligned the human NDUFB7 protein sequence with its zebrafish orthologue, revealing that the amino acid at position 84 differs between the two organisms: Arginine (R) in human NDUFB7 is replaced by Lysine (K) at position 84 in zebrafish Ndufb7 (see **Figure 20A**). R and K share similar properties: both are positively charged basic amino acids involved in maintaining protein stability (Barlow and Thornton, 1983). We hypothesize that the drastic change in amino acid properties from the hydrophilic R / K to the highly hydrophobic Trypthophan (W) may have a comparable, potentially destabilizing effect on protein structure in both organisms (Palego *et al.*, 2016). The three amino acids are depicted in the disease model scheme in **Figure 20B**.

The hypothesis that Ndufb7^{K84W} (from here on referred to as W84) may lead to destabilization of Ndufb7 in the zebrafish, was corroborated by protein modeling of this particular mutation in combination with the two allelic variants present in the WIK strain, Ndufb7^{H28} (H28) and Ndufb7^{N28} (N28). The following results are provided with permission of J. L. Cabrera:

Since there was no protein structure available for zebrafish Ndufb7 in the protein data base PDB, a model for the wild type variant of Ndufb7 (H28, K84) was generated using Rosetta comparative modeling (Song *et al.*, 2013). Next, the impact of single amino acid changes on protein stability was determined with Rosetta ddG monomer (Kellogg, Leaver-Fay and Baker, 2011) as described in materials and methods. Using the “double” wild type variant (**H28 + K84**) as **baseline** / reference model, the following ddG values and the corresponding interpreted effects were determined per amino acid combination:

1) **N28:** the natural **WIK SNP** results in a ddG of 0.888 kcal/mol when compared to the reference model. Amino acid changes causing positive ddG values > 0.5 are interpreted

as destabilizing. However, the median (|ddG| = 3.15 kcal/mol) and interquartile range (IQR = [2.4-12.96]) of the absolute ddG value obtained for the 19 possible amino acid changes at position H28 suggest that the WIK SNP N28 may finally not have a high impact on protein stability, considering that the ddG value for N28 is below the first quartile value (0.888 < 2.4).

2) **W84:** comparison of the reference model with the **disease mutation** alone resulted in a ddG value of 1.593 kcal/mol. With a median |ddG| of 1.329 kcal/mol and an IQR of [0.765-13.2] for this position, the ddG value for W84 is above the first quartile (1.593 > 0.765). This result suggests that the W84 mutation may have a considerable impact on Ndufb7 protein stability.

3) **N28 + W84:** unsurprisingly, comparison of the reference model to Ndufb7 containing a combination of both amino acid changes (**WIK SNP and disease mutation**) results in the strongest predicted destabilization effect on Ndufb7, with a ddG value of 2.238 kcal/mol.

All four models and their corresponding ddG values are depicted in **Figure 21**.

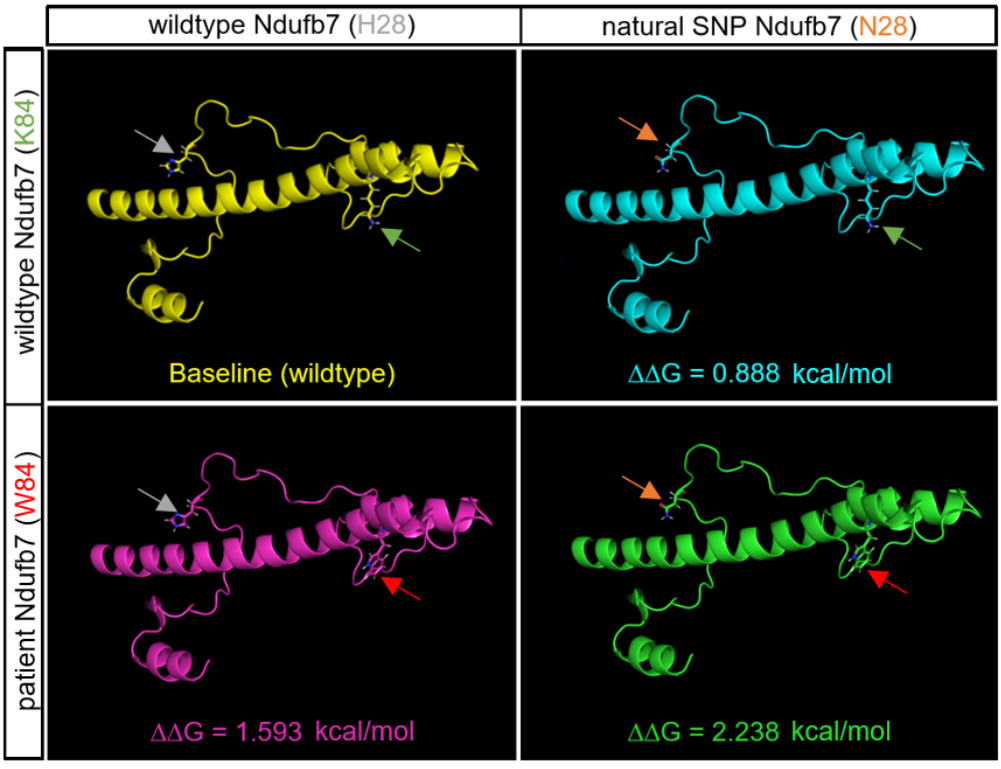


Figure 21. Structural protein modeling of the four zebrafish Ndufb7 variants: wildtype (K84, green arrow, top row) versus patient mutant (W84, red arrow, bottom row) in combination with the two WIK allelic variants wildtype (H28, grey arrow, column 1) or SNP (N28, orange arrow, column 2). Differences in protein stability (ddG) were calculated in relation to the wildtype (H28, K84) Ndufb7 protein variant and are stated underneath each protein model. *Protein models are a courtesy of J. L. Cabrera.*

J. L. Cabreras above described protein modeling results suggest that the W84 patient mutation may indeed cause a dysfunction in the OXPHOS machinery by destabilizing Ndufb7. On the

other hand, the modeling results suggest that the naturally occurring WIK SNP variant Ndufb7^{N28} may not have a strong enough effect on protein stability to trigger an allelic selection mechanism in a heterozygous environment, if such mechanism exists.

We therefore conclude that mimicking the human NDUFB7^{R84W} patient variant by CRISPR-Cas9 mediated mutagenesis in the zebrafish WIK strain may be useful for two reasons. First, to answer the original research question regarding allele-specific selection in the presence of two Ndufb7 variants, and second, to investigate whether the W84 mutation causes a neurologic disease phenotype similar to that of the human patient.

4.4.2 Comparison of transient *ndufb7^{H28}-L2-mRuby2* wild type and mutant variant expression by mRNA overexpression

Introduction of the likely destabilizing and potentially disease causing W84 mutation in zebrafish Ndufb7 raises several questions:

1) Does W84 influence the mitochondrial localization and functionality of Ndufb7?
2) Is the W84 Ndufb7 variant detected and degraded in the presence of wild type Ndufb7?
3) Does expression of W84 Ndufb7 cause phenotypic abnormalities in the presence of a wild type variant, for example by exerting a dominant-negative effect?
4) Is CRISPR-Cas9 mediated mutagenesis of *ndufb7* exon 2 feasible?

Addressing questions one and two in the absence of a tag is experimentally challenging. I therefore repurposed the transient overexpression system previously generated for Strategy I by mutagenizing the *ndufb7-L2-mRuby2* plasmid, followed by plasmid linearization with either NotI or ApaI enzyme and mRNA synthesis. In total, I compared four different fluorophore-tagged Ndufb7 variants by parallel mRNA injections: (a) the originally created wild type variant *ndufb7^{H28,K84}-L2-mRuby2*, from here on referred to as **K84 wt,** and its derivatives (b) **W84**: the patient variant *ndufb7^{H28,W84}-L2-mRuby2*, (c) **R84**: the human variant *ndufb7^{H28,R84}-L2-mRuby2* as first control, and (d) **r.m K84 wt**: a reverse mutated version of the wild type variant, *ndufb7^{H28,W84K}-L2-mRuby2*, generated from the W84 mutant plasmid as second (technical) control.

As depicted in **Figure 22A**, I injected 1.05 ng of (NotI linearized) mRNA per WIK or AB embryo and recorded the number of resulting fluorescent embryos at 11 hpf and 1 dpf, together with survival and abnormal phenotype. Interestingly, I did not observe a pronounced difference in the number of fluorescent embryos or the level of fluorescence per embryo when comparing embryos overexpressing the Ndufb7 W84 patient variant to those overexpressing wildtype/human control variants (see **Figure 22B**, red channel images). Furthermore, I did not observe any distinct negative effects on embryo survival and physiology in association with the

W84 patient variant (see **Figure 22B**, bright field images). Percentages of embryonic fluorescence, survival and phenotype counts per strain and mRNA variant/type are shown in **Figure 26**. The absence of an abnormal physical phenotype suggests that the fluorophore-tagged Ndufb7 variant containing the W84 patient mutation does not cause a dominant-negative effect upon overexpression.

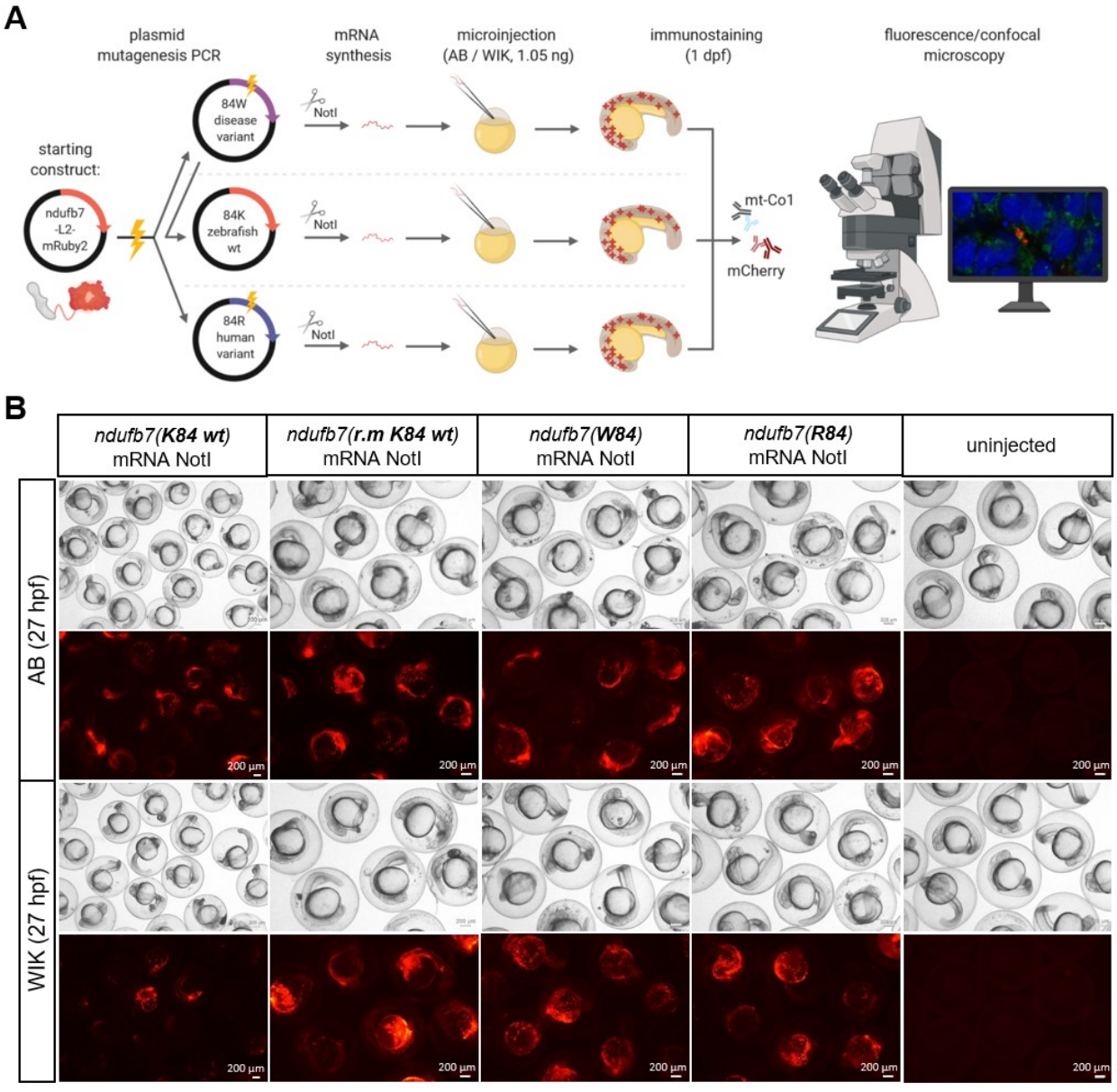


Figure 22. NotI-linearized mRNA overexpression of the four different *ndufb7-L2-mRuby2* variants in AB and WIK embryos followed by whole embryo live imaging at 27 hpf. A Schematic depiction of the experimental procedure, generated with BioRender. **B** Fluorescence stereoscope images of mRNA injected embryos compared to uninjected control one day post injection. Column 1 and 2: wildtype Ndufb7(K84)-L2-mRuby2 overexpression (original vs reverse mutated), Column 3: patient Ndufb7(W84)-L2-mRuby2 overexpression, Column 4: Ndufb7(R84)-L2-mRuby2 overexpression, Column 5: uninjected control. Scale bar: 200 µm.

While technical constraints of transient mRNA overexpression should be kept in mind (short-term expression in the presence of a big fluorophore tag), whole embryo fluorescence microscopy of the injected embryos suggests that there may be no difference in protein regulation upon overexpression of the mRuby2-tagged W84 mutant in comparison to the control

variants (r.m K84, R84). To investigate potential differences in expression patterns on tissue/cellular level, I next assessed Ndufb7-L2-mRuby2 (W84, R84, K84) localization by confocal imaging.

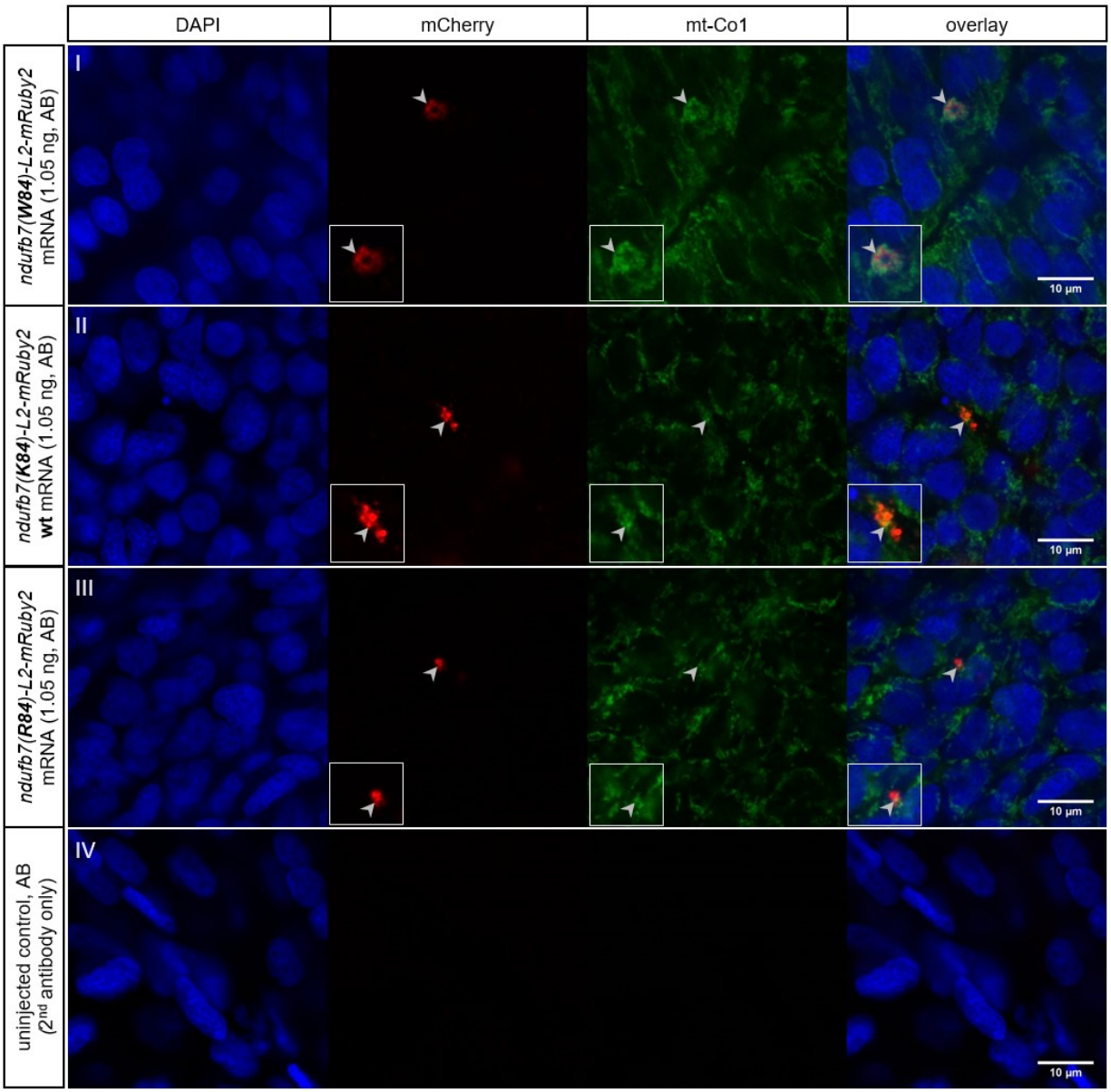


Figure 23. Confocal imaging of the three Ndufb7-L2-mRuby2 variants K84, W84 and R84 following overexpression of Not-I linearized mRNA in AB strain embryos (1 dpf). Co-staining of mRuby2 with the mitochondrial marker mt-Co1 suggests partial mitochondrial localization of the tagged protein variants in the embryonic tail. Grey arrows point to examples for overlapping signals, zoom-ins are shown in grey boxes. Panel IV: control, blue: DAPI (nuclei), red: anti-mCherry antibody (Ndufb7-L2-mRuby2), green: anti-mt-Co1 antibody (mitochondria). Scale bar: 10 µm.

In total, 30 fluorescent embryos were collected per sample type one day after injection with mRNA generated from NotI-linearized plasmids. The deyolked embryos were fixed and immunostained to visualize nuclei, mitochondria (anti-mt-Co1) and the overexpressed Ndufb7-L2-mRuby2 variants (anti-mCherry antibody). My confocal imaging results from injected AB and WIK strain embryos are in line with the whole embryo fluorescence microscopy results: I did not observe a distinct difference in expression of the tagged W84 Ndufb7 mutant compared to the

wild type/human variants K84 and R84. As shown in **Figure 23** for the AB strain and **Figure 24** for the WIK strain, I observed a partial overlap of tagged Ndufb7 with the mitochondrial marker mt-Co1 in the case of all three Ndufb7 variants (W84, K84, R84). These results suggest that the W84 patient mutation may not have a negative influence on mitochondrial localization of Ndufb7.

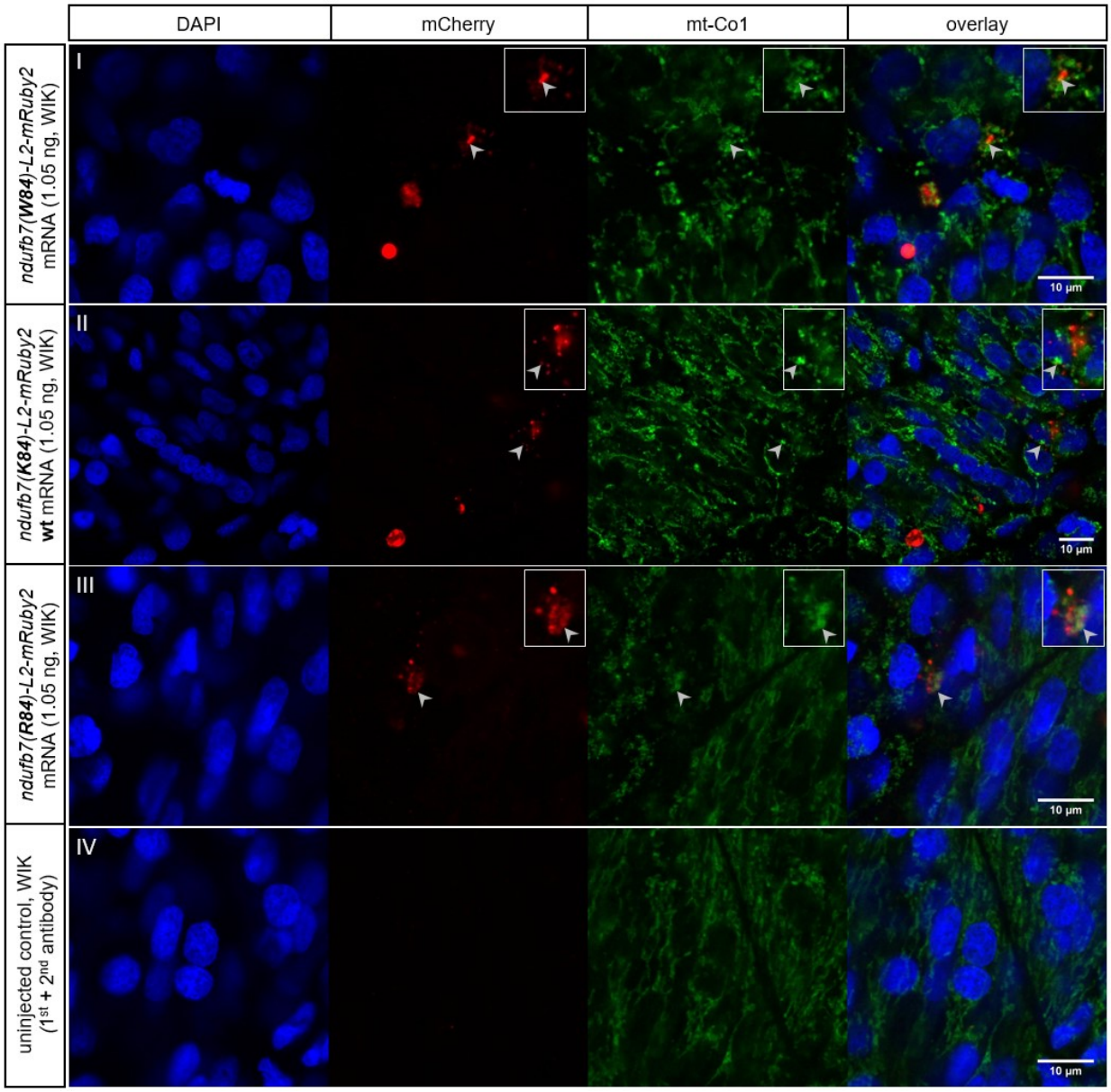


Figure 24. Confocal imaging of the three Ndufb7-L2-mRuby2 variants K84, W84 and R84 following overexpression of Not-I linearized mRNA in WIK strain embryos (1 dpf). Co-staining of mRuby2 with the mitochondrial marker mt-Co1 suggests partial mitochondrial localization of the tagged protein variants in the embryonic tail. Grey arrows point to examples for overlapping signals, corresponding zoom-ins are shown in grey boxes. Panel IV: control, blue: DAPI (nuclei), red: anti-mCherry antibody (Ndufb7-L-mRuby2), green: anti-mt-Co1 antibody (mitochondria). Scale bar: 10 µm.

Interestingly, while I did not observe any pronounced differences between fluorophore expression levels and in terms of phenotype when comparing the four variants of NotI-linearized *ndufb7-L2-mRuby2* mRNA with each other (*K84 wt, r.m K84 wt, W84* and *R84*) in the AB and WIK strain, the outcome changed drastically when switching to a different restriction enzyme for plasmid linearization: ApaI. As depicted in **Figure 25**, both the AB and WIK strain express NotI-

linearized mRNA at high levels with little effect on survival or phenotype. While WIKs also express the four variants of ApaI-linearized mRNA, there is no fluorescence in ABs (compare **Figures 22B and 25B**). Despite being non-fluorescent, many AB embryos develop severe phenotypical abnormalities following injection of ApaI-linearized mRNA, regardless of the injected mRNA variant. Examples for such phenotypical abnormalities can be seen in row 1 of **Figure 25B**.

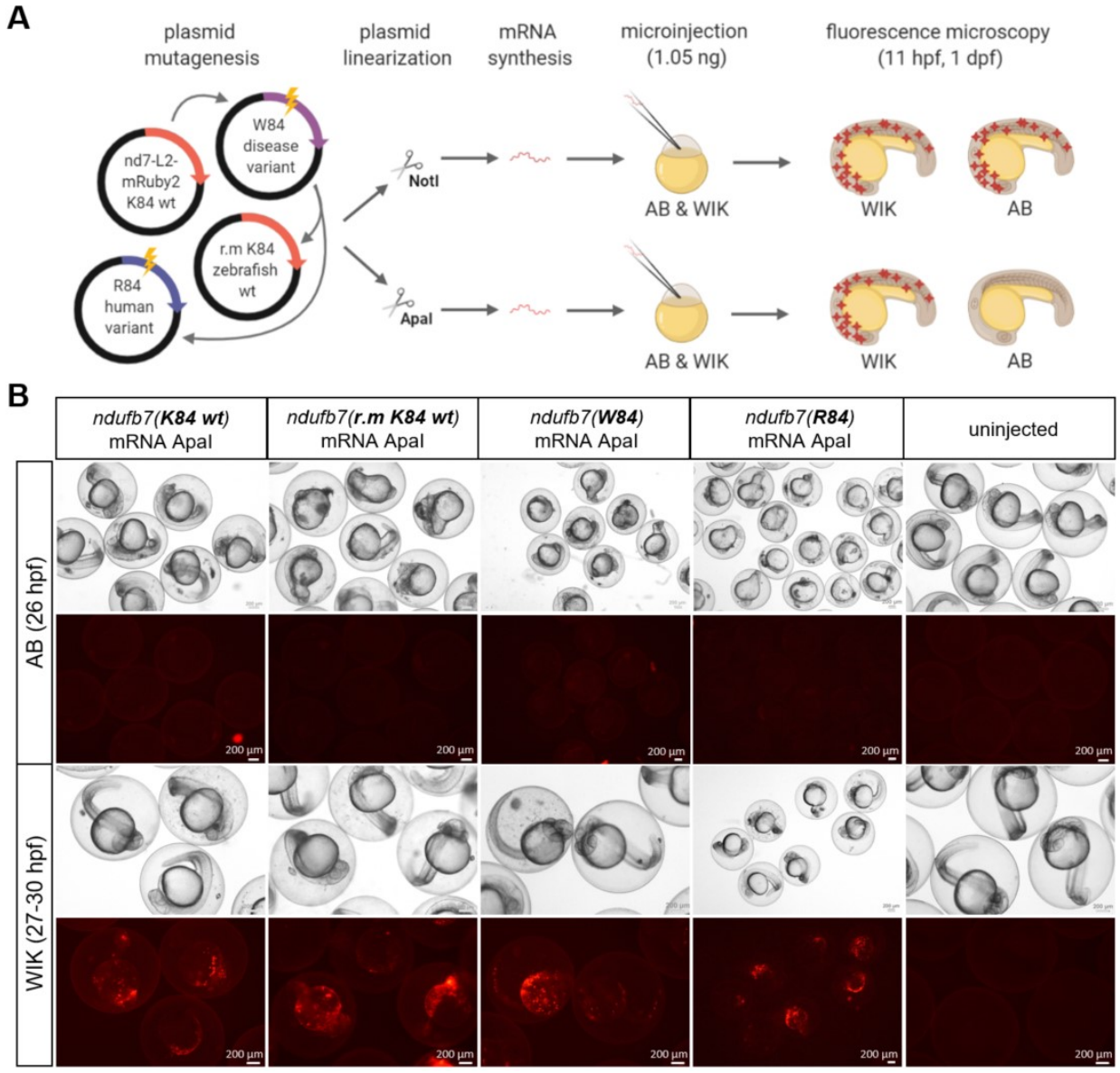


Figure 25. ApaI-linearized mRNA overexpression of the four different *ndufb7-L2-mRuby2* variants in AB and WIK embryos followed by whole embryo live imaging at 1 dpf. A Schematic depiction of the different outcomes in fluorescence per strain (AB/WIK), depending on the restriction enzymes used for plasmid linearization: NotI or ApaI. Scheme generated with BioRender. **B** Fluorescence stereoscope images of ApaI-linearized mRNA injected embryos compared to uninjected control at 1 dpf. Column 1 and 2: wildtype Ndufb7(K84)-L2-mRuby2 overexpression (original vs reverse mutated), Column 3: patient Ndufb7(W84)-L2-mRuby2 overexpression, Column 4: Ndufb7(R84)-L2-mRuby2 overexpression, Colum 5: uninjected control. Scale bar: 200 µm.

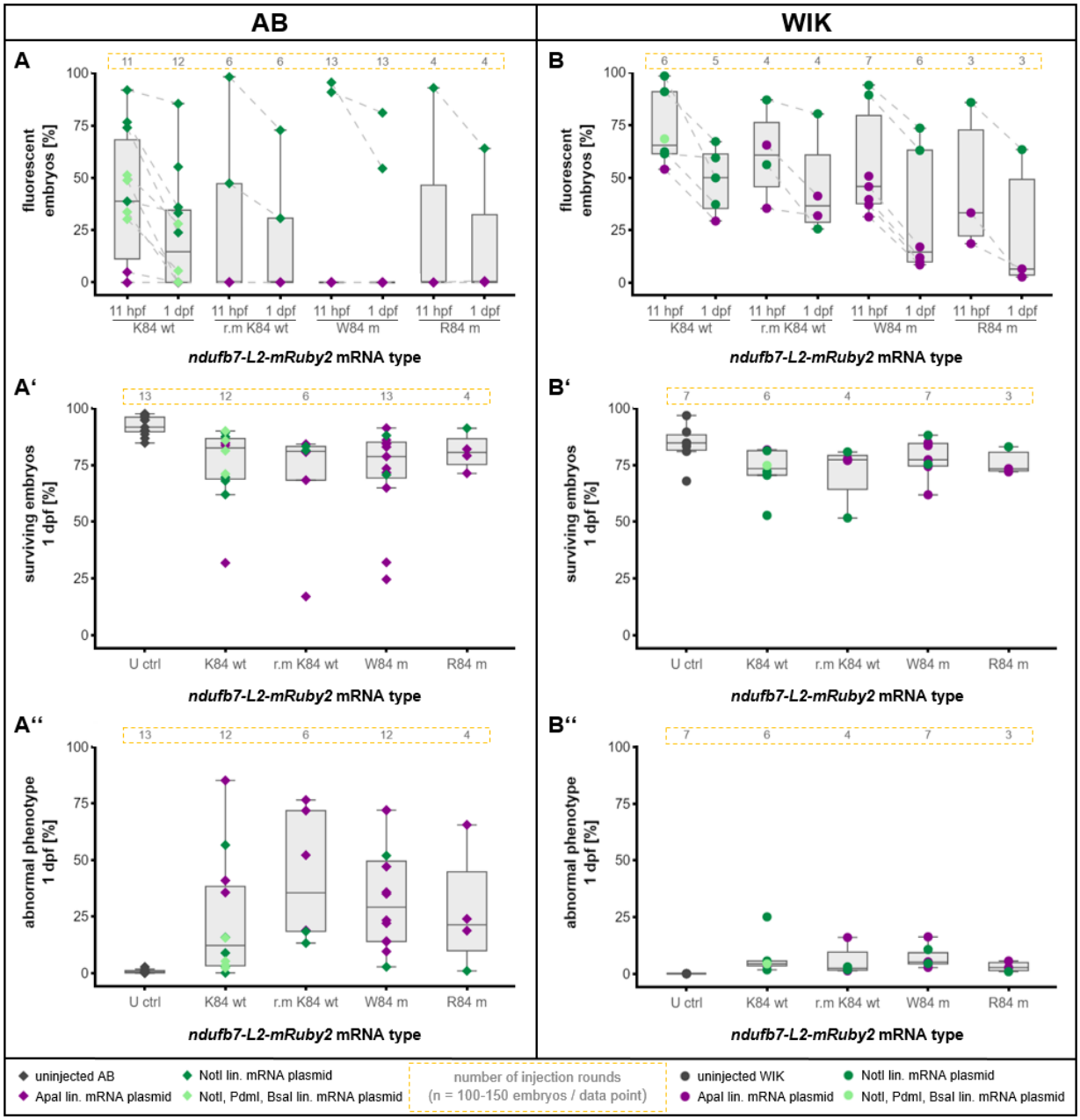


Figure 26. Summarized manual counts of AB and WIK embryos injected with different types of ApaI- or NotI-linearized *ndufb7-L2-mRuby2* mRNA (original and reverse mutagenized K84 wild type; W84 and R84 mutant). **A, B** percentage of fluorescent survivors at 11 hpf and 1 dpf (AB and WIK strain respectively). The decrease in fluorescence between the two time points of one injection round is shown as grey dash line. **A', B'** percentage of survivors at 1 dpf (AB and WIK strain respectively) **A'', B''** percentage of survivors with an abnormal phenotype at 1 dpf (AB and WIK strain respectively). Injections of ApaI-linearized mRNA are shown as purple data points. Injections of NotI(-PdmI-BsaI)-linearized mRNA are shown as green data points. Uninjected control (U ctrl): dark grey. Each data point represents one round of injection (n = 100-150 embryos/injection). The total number of injection rounds per mRNA type is stated above each boxplot.

Upon repeated, parallel injections of ApaI- versus NotI-linearized mRNA (1.05 ng/embryo) of any of the four *ndufb7-L2-mRuby2* variants into ABs and WIKs, the following pattern emerged:

- **WIK strain:** embryos express both, NotI-linearized and ApaI-linearized mRNA at 11 hpf and 1 dpf, with a slight preference towards NotI-linearized mRNA (**Figure 26B**). In terms

of overall survival and abnormal phenotype, embryos from the WIK strain tolerate NotI- and ApaI-linearized mRNA equally well (**Figure 26B', B''**).

- **AB strain:** judging by overall fluorescence, embryos from the AB strain express NotI-linearized mRNA at high levels at 11 hpf and 1 dpf, while largely failing to express ApaI-linearized mRNA at both time points (**Figure 26A**). In terms of survival, embryos from the AB strain tolerate NotI- and ApaI-linearized mRNA at levels comparable to WIKs, with a slightly better outcome following injection of NotI-linearized mRNA (**Figure 26A'**). However, embryos from the AB strain develop severe abnormalities following injection of ApaI-linearized mRNA, while NotI-linearized mRNA appears to be well tolerated (**Figure 26A''**).

By injecting embryos from the AB and WIK strains in parallel (one after another, in varying order) for each type of *ndufb7-L2-mRuby2* mRNA, I ensured that the injection conditions were kept constant to facilitate comparison between the two strains.

As depicted in the orange box in **Figure 27A**, the NotI and ApaI restriction enzyme cutting sites are separated by 25 bp on the pCS2 plasmid backbone, with the NotI site located closer to the SV40 poly A signal. According to the manufacturer of the mRNA synthesis kit, SP6 RNA polymerase is highly processive and transcription does not necessarily terminate at the poly A site. Thus, plasmid linearization prior to mRNA synthesis is recommended.

During plasmid digest, the NotI restriction enzyme produces 5' overhangs (cut site: 5′-GC^GGCCGC-3′). According to the manufacturer, 1 µL NotI (fast digest) is capable of processing 1 µg plasmid DNA in 30 minutes. The ApaI restriction enzyme, on the other hand, induces a 3' overhang upon digest, at a processing speed of 1 µg plasmid DNA with 1 µL enzyme in 5 minutes (cut site: 5′-GGGCC^C-3′). Neither enzyme is supposed to have star activity within the used time frame. Although supposedly all types of restriction enzymes can be used for plasmid linearization, there has been one report according to which 3' overhangs can cause extension of mRNA synthesis by the RNA polymerase on the opposite strand of the linearized template, thereby producing elongated transcripts which include the sequence of the complementary strand (Schenborn and Mierendorf, 1985). The hypothesis that ApaI linearization leads to generation of elongated transcripts, is conceptually visualized in **Figure 27A**. To test whether ApaI restriction is followed by aberrant mRNA synthesis resulting in elongated transcripts, I therefore analyzed ApaI-mRNA and NotI-mRNA generated in parallel from two *ndufb7-L2-mRuby2* plasmid variants (reverse mutagenized K84 wildtype and R84 mutant) on a reducing agarose gel. Albeit at low levels, I was indeed able to confirm the presence of elongated transcripts in the ApaI-mRNA but not the NotI-mRNA samples (**Figure 27C**, marked by blue arrows).

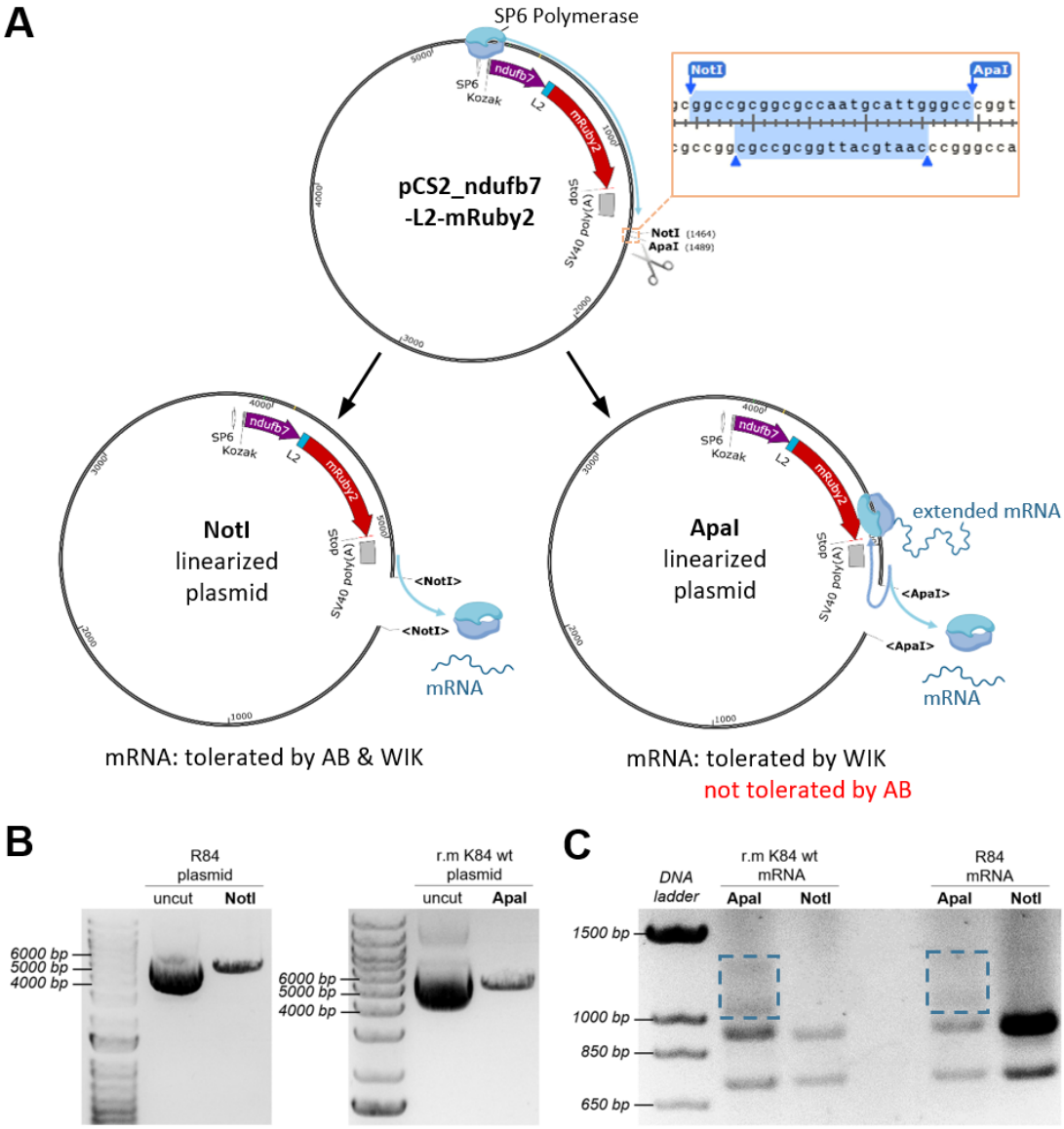


Figure 27. mRNA generation from ApaI-linearized plasmids leads to partially elongated, aberrant transcription during *in vitro* mRNA synthesis with SP6 RNA polymerase. A theoretical model of how SP6 RNA polymerase may transcribe NotI-linearized (left) versus ApaI-linearized (right) mRNA plasmids. The exact NotI and ApaI cutting sites are shown in the orange box. The scheme was generated using SnapGene and BioRender. **B** linearized versus undigested *ndufb7-L2-mRuby2* plasmids (reverse mutagenized wildtype K84 versus human variant R84). **C** ApaI-RNA versus NotI-mRNA on a reducing agarose gel. DNA ladder was loaded as vague reference only. Blue boxes mark aberrant transcripts.

While it does not seem to represent a major problem in the WIK strain, the abnormal phenotype and lack of fluorophore expression following overexpression of the (elongated) ApaI-mRNA transcripts in ABs may at least be partially attributable to this unexpected technical issue.

4.4.3 K84W mutagenesis in exon 2 of *ndufb7* by CRISPR/Cas9 + ssODN-mediated knock-in

My (NotI-linearized) mRNA overexpression results suggest that the patient variant Ndufb7^{W84} does not have a major negative impact on protein localization and overall fitness of the organism. However, simultaneous endogenous expression of the two naturally present Ndufb7 variants (H28/N28, K84) may outcompete the W84 mutant and thus cancel out potential negative effects of the overexpressed patient variant. Furthermore, potential adverse effects caused by the presence of the fluorophore tag cannot be excluded at this point.

Thus, we next asked whether introduction of the patient mutation by CRISPR/Cas 9-mediated mutagenesis of endogenous *ndufb7* may negatively influence the fitness of the organism. In the human patient, the R84W conversion was caused by a SNP. However, introducing the corresponding K84W mutation in the zebrafish involves a three base pair exchange from AAA to TGG. To determine whether the genomic region surrounding the mutagenesis site in *ndufb7* exon 2 can be targeted using CRISPR/Cas9, I tested two different sgRNAs: sgRNA 134, which induces a cut within the mutagenesis site, and sgRNA 147, which induces a cut three base pairs after AAA (**Figure 28B**).

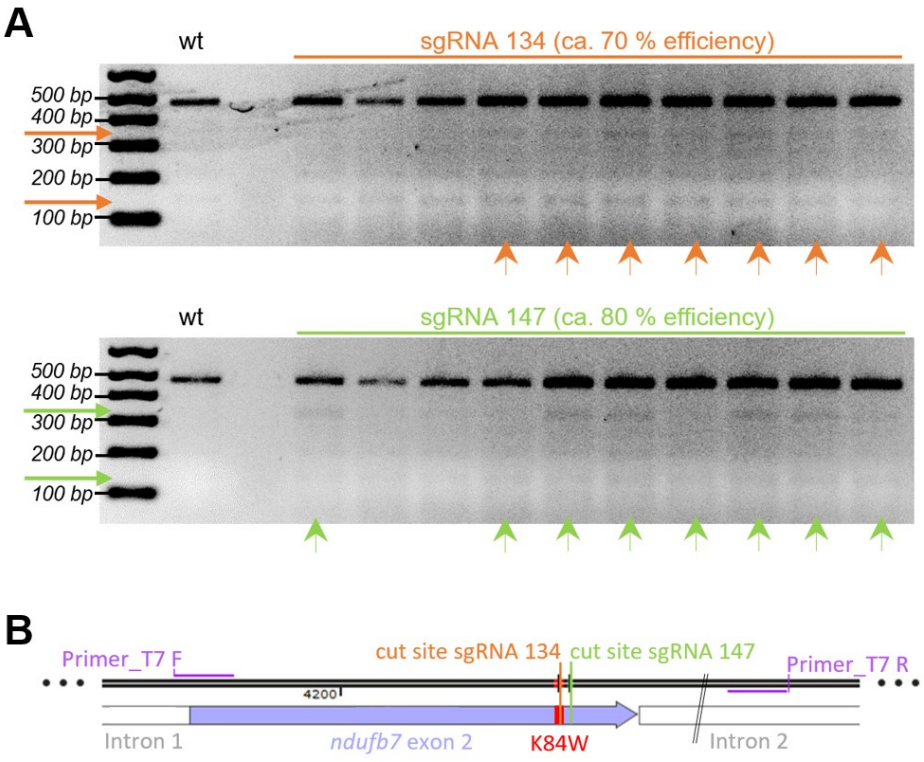


Figure 28. sgRNA efficiency test for CRISPR/Cas9-mediated mutagenesis of *ndufb7 exon 2* by T7 Endonuclease 1 (T7E1) assay. A agarose gel of the unmodified wildtype (wt) versus CRISPR/Cas9 modified target region. Loaded: T7E1 digested PCR product from single embryos injected with sgRNA+Cas9 at 1-cell stage. Horizontal arrows show expected band sizes following sgRNA/Cas9 cut. Vertical arrows mark samples considered positive for the intended cut. Shown are T7E1 assay results for sgRNA 134 (70 % cutting efficiency, expected bands at 145 + 342 bp) and sgRNA 147 (80 % cutting efficiency, expected bands at 149 + 338 bp); unmodified wildtype band: 487 bp. **B** genomic locus of zebrafish *ndufb7* exon 2, including cutting sites of the tested sgRNAs 147 and 134, mutagenesis site and PCR primer binding sites used for amplification of the target region.

Following injection of the individual sgRNAs complexed with Cas9 into 1-cell stage embryos, I isolated gDNA from ten single embryos and amplified the target region with the primers indicated in **Figure 28B**. T7 assay revealed a cutting efficiency of approximately 70 % for sgRNA 134 and approximately 80 % cutting efficiency in the case of sgRNA 147. The expected band sizes are: 487 bp in the case of unmodified wild type, 145 + 342 bp upon successful targeting with sgRNA 134, and 149 + 338 bp when cut with Cas9 and sgRNA 147 (**Figure 28A and C**).

To facilitate the AAA>TGG base pair exchange in *ndufb7* exon 2, Dr. Piragyte-Langa ordered an asymmetric, 90 nucleotide single-stranded oligo DNA nucleotide (ssODN) repair template through IDT. The + stranded repair template is homologous to the 90 bp region surrounding the target site (left homology arm: 54 bp, right homology arm: 33 bp) and contains the desired TGG mutation. I then first injected the + stranded ssODN into the yolk, followed by injection of either sgRNA 147 or sgRNA 134 with Cas9-RNP into WIK embryos at 1-cell stage. A schematic depiction of the experimental procedure including the intended HDR-mediated insertion of the repair template is shown in **Figure 29A**.

Out of 200 injected embryos, I selected 12 healthy looking embryos injected with the mutagenesis mixture, and collected them at 5 dpf to analyze mutagenesis success by T7 assay. In addition, I collected pools of 5 uninjected wild type and 4 ssODN RT-only injected embryos as negative controls, and four pools of 5 CRISPR/Cas9-only injected embryos as positive control. As shown in **Figure 29B**, I observed the desired band after T7E1 digest in the single embryos N2, N3, N4 and N5. I gel-purified the PCR-amplified genomic region from these four embryos and sent the PCR products for sequencing. The Sanger sequencing results revealed that one out of the four sequenced embryos contained the desired AAA>TGG in-frame mutation in exon 2 of *ndufb7* (embryo N2), thereby confirming the feasibility of CRISPR/Cas9-mediated mutagenesis (**Figure 29C and D**).

Having validated the *in vivo* mutagenesis approach, I next reinjected the mutagenesis mixture with either sgRNA 134 or sgRNA 147 (both showed similar targeting efficiency in T7 assay; **Figure 28**) into WIK embryos to raise them to adulthood, followed by determination of KI success and germline transmission in F1 generation. In total, 160 embryos injected with ssODN + sgRNA 147/Cas9 are currently being raised, together with 102 embryos injected with the ssODN + sgRNA 134/Cas9 injection mixture.

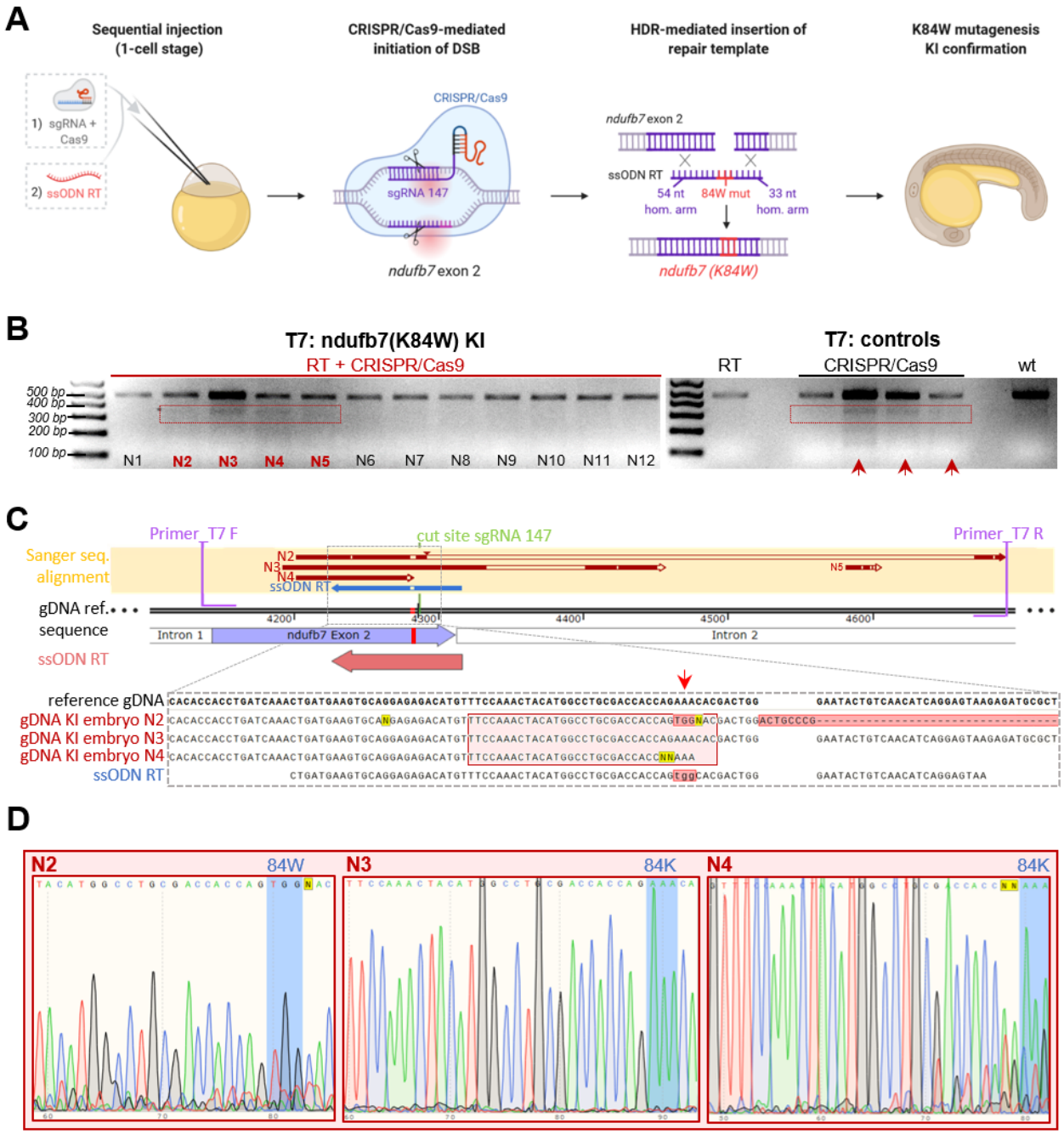


Figure 29. gDNA-level confirmation of *ndufb7*^{K84W} KI success following sequential injection of ssODN repair template and sgRNA 147/Cas9-RNP in WIK embryos. A Schematic representation of the *in vivo* mutagenesis process: sequential injections followed by induction of a double strand break (DSB) by CRISPR/Cas9 and insertion of the ssODN repair template sequence by homology directed repair (HDR). Generated with BioRender. **B** Determination of KI success by T7 assay of single embryos at 5 dpf. Gel 1 shows the resulting bands for 12 single embryos injected with the mutagenesis mixture (N1-12). Gel 2 shows two negative controls (RT: repair template only and wt: uninjected wildtype) and four positive controls (CRISPR/Cas9 only). Samples positive for T7E1 digest are marked in red. **C** sequence validation results of the genomic *ndufb7* exon 2 region after *in vivo* mutagenesis, amplified from the gDNA of single embryos N2-N5. Aligned Sanger sequences are shown as red arrows above the gDNA reference sequence and in detail within the grey-dotted box. Primer binding sites used for amplification from gDNA and sgRNA 147 cutting sites are shown. ssODN RT: repair template sequence. The red arrow marks the AAA>TGG mutation site. Alignments were performed in SnapGene. **D** Close up of the sequenced bases in the mutagenesis region (mutagenesis site is highlighted in blue) as detected by Sanger sequencing. Peak color per base: A green, T red, C blue, G black. N2, N3 and N4 represent the corresponding single embryos from C and D.

5 Discussion and Conclusions

As outlined in the introduction, the critical protein complexes of the eukaryotic OXPHOS system rely on the proper interaction between nuclear encoded and mitochondrial encoded OXPHOS subunits. While OXPHOS is not the only cellular process in eukaryotic cells that utilizes bigenomically encoded gene products, it is sensitive to variation and thus represents a suitable model system to study the interaction between the two genomes (Koopman *et al.*, 2013; Kotrys and Szczesny, 2019). A growing body of evidence suggests that mtDNA haplotypes, and hence mtOXPHOS variants, are heavily subjected to positive selection mechanisms during early embryogenesis (Sharpley *et al.*, 2012; Burgstaller *et al.*, 2014; Wei *et al.*, 2019). However, it remains unclear whether there is selection for or against a nOXPHOS subunit, when two different allelic variants are present.

The WIK zebrafish strain in our laboratory is heterozygous for *ndufb7*: one alle encodes the wildtype Ndufb7^{H28} protein whereas the other allele contains a missense SNP resulting in the Ndufb7^{N28} protein variant. Interestingly, RNA-seq data generated from this strain suggests a slight bias in transcription towards the wildtype allele (67 %). Aside from this allelic bias in *ndufb7* mRNA expression levels, Dr. Piragyte-Langa and I chose Ndufb7 as nOXPHOS target subunit for three main reasons: first, Ndufb7 seems to be critical for OXPHOS complex I function (Stroud *et al.*, 2016). Selecting the “best fitting” protein variant in the heterozygous WIK strain may thus represent an evolutionary advantage. Second, structural models of yeast and bovine complex I suggest that Ndufb7 is located on the surface of complex I (Wirth *et al.*, 2016). This may allow enough room for incorporation of an artificial tag without disturbing protein function. Third, considering that mitochondrial import likely occurs via the post translationally attached myristol group at the N-terminus, we hypothesize that addition of a C-terminal tag may not interfere with mitochondrial localization.

The main objective of this thesis was to find and validate a suitable strategy for differential visualization of the two *ndufb7* alleles present in the heterozygous zebrafish WIK strain. The resulting animal model should in the future enable the study of potential allele-specific bias in *ndufb7* gene expression in a tissue/cell type-specific manner throughout embryonic development and during adulthood.

5.1 Comparative assessment of the three tested strategies for establishment of a zebrafish model that enables allele-specific analysis of *ndufb7* gene expression

The assembly of the OXPHOS machinery is a highly orchestrated process. Therefore, our main concern during animal model design was potential interference with OXPHOS function. To increase the chances of success, I therefore designed and followed three experimental strategies in parallel under the supervision of Dr. Piragyte-Langa. Feasibility, advantages, and limitations of each strategy are discussed in the following subsections and summarized in **Figure 30.**

5.1.1 Strategy I: Differential fluorophore-tagging of two naturally occurring *ndufb7* allelic variants by CRISPR/Cas9-mediated knock-in of *mEGFP* and *mRuby2*

As a first strategy, we chose to visualize expression of the two *ndufb7* variants by attachment of two differently colored fluorophore tags (one color per allele). The use of fluorophores brings several advantages (**Figure 30, strategy 1.**). Specifically, fluorophore tags enable: (a) tag-specific sequence amplification at the DNA and mRNA level by simple PCR screening to determine KI success and expression levels, (b) antibody-based differentiation between the two tagged allelic variants at the protein level *in vivo* and *in vitro*, and (c) a direct readout of tissue- or cell type-specific allelic expression tendencies by live imaging throughout the development, thanks to intrinsic fluorescence of the differentially tagged allelic variants. The fluorophore strategy therefore fulfils two main requirements of the animal model: it enables differentiation between the two *ndufb7* variants at DNA, mRNA and protein level, as well as straightforward visualization of the two alleles in a tissue/cell type specific manner without a necessity of antibody staining. Given the complexity of the OXPHOS system, modifying an OXPHOS subunit with a large fluorophore tag is a challenging task. To my knowledge, fluorophore-tagging of nOXPHOS genes has mainly been reported in bacteria (Erhardt *et al.*, 2014), yeast (Stoldt *et al.*, 2018), and cell culture (Vogel *et al.*, 2007; Wilkens, Kohl and Busch, 2013; Rieger *et al.*, 2017) to date, but not in higher eukaryotic organisms such as the zebrafish – especially not in the context of allele specific gene expression analysis.

In an attempt to develop such an allele-specific, fluorophore-tagged zebrafish model, I first tested six different intrinsically fluorescent proteins (mRuby2, Clover, mCitrine, mTurquoise2, eGFP and mEGFP) and three connecting linker sequences (L1: GGG, L2: $(GGGGS)_3$, L3: $(EAAAK)_3$) by mRNA overexpression. Out of the three tested linker sequences, I found that flexible linkers of varying length, L1 and L2, work best in combination with Ndufb7^{H28} and mRuby2. In

comparison, the tested long and rigid linker L3 yielded only a low amount of fluorescent signal, suggesting rapid degradation of the *ndufb7-L3-mRuby2* construct and/or the resulting fusion protein *in vivo*. For the knock-in, I finally settled on L2, because this construct consistently yielded the highest amount of fluorescence per embryo in mRNA overexpression experiments.

While the red fluorophore mRuby2 was generally well tolerated by both zebrafish strains used for fluorophore testing (AB and WIK), mRNA overexpression of the synthetic green fluorophore candidates revealed a concerning pattern: injection of the generated *ndufb7-L1-Clover/-Citrine/-Turquoise/-eGFP* or *-mEGFP* mRNA yielded a much lower amount of fluorescent signal by 1 dpf particularly in the AB strain, where at most 2 % of the injected embryos were fluorescent. Despite being largely non-fluorescent, up to 95 % of injected AB embryos developed an abnormal phenotype at 1 dpf, suggesting that the green fluorophores may not be tolerated in combination with Ndufb7. However, the observed differences between expression patterns of the red and green fluorophores following mRNA injection can likely be explained by technical issues attributed to mRNA plasmid linearization (ApaI vs. NotI), as pointed out in **Figure 27**. In support of this observation, fluorophore expression levels increased upon injection of the *ndufb7-L1-mEGFP* mRNA plasmid. Immunostaining in combination with confocal microcopy finally confirmed that both fusion proteins, mEGFP- and mRuby2-tagged Ndufb7^{H28}, colocalize with the mitochondrial marker mt-Co1. Hence, the transient overexpression experiments suggest that mEGFP and mRuby2 are suitable for tagging of the two *ndufb7* alleles, in combination with a flexible linker (L1 or L2).

Under the guidance of Dr. Piragyte-Langa, I used a knock-in approach established by Hoshijima and colleagues, according to whom a repair template containing perfect long homology arms (1000-1200 bp) facilitates homology directed repair (HDR) by homologous recombination upon CRISPR-Cas9 mediated double-strand break induction in the target region. The authors achieved a germline transmission rate of approximately 6 % with this approach (K. Hoshijima, Jurynec and Grunwald, 2016). To avoid repeated cutting by CRISPR/Cas9 following integration of the repair template, I introduced two single base pair mutations in the sgRNA target sites of the repair template. Using this approach, I was able to confirm successful knock-in of *L2-mRuby2* at the intended *ndufb7* locus on genomic and mRNA level by Sanger sequencing, thus validating the feasibility of the knock-in approach. Upon closer inspection of the insertion site, I noticed that the repair machinery successfully incorporated the fluorophore sequence at the STOP signal site of *ndufb7.* I observed imperfect repair at the sgRNA 1/Cas9 target site, as well as the absence of the two single base pair mutations deliberately introduced into the repair template to prevent repeated cutting by Cas9. This indicates that the 3' recombination event happened directly at the sgRNA target site in the 3' UTR of *ndufb7* in the case of embryo 1 and 4 (**Figure 14B** and **Suppl. Information 1**). I used a combination of two sgRNAs for fluorophore KI because it was postulated to marginally enhance targeting efficiency (Li *et al.*, 2014), and to increase the

likelihood of DSB induction at the target site: sgRNA 1 targeting the 3' UTR and sgRNA 2, which induces a cut in *ndufb7* exon 3. For the 95 currently growing *mRuby2* and/or *mEGFP* knock-in animals, it will be necessary to carefully evaluate the integrity of both sgRNA target sites in *ndufb7* exon 3 and the 3' UTR following confirmation of germline transmission. While the integrity of the *ndufb7* CDS and its 3' UTR may represent a valid concern, I was able to confirm the presence of tagged *ndufb7* mRNA transcripts in four out of eight fluorescent single embryos collected at 1 dpf, which suggests that functional knock-in of the fluorophore is achievable.

Besides successful validation of the KI strategy, I encountered one persistent problem during all KI-injection attempts: knock-in embryos for both *ndufb7-L2-mRuby2* and *ndufb7-L2-mEGFP* that were fluorescent at 1 dpf lost all fluorescent signal within 5 days. Notably, despite the loss of fluorescence, I was able to confirm integration at the genomic *ndufb7* locus by gDNA PCR amplicon sequencing, as described above. However, this was only the case for a small number of previously fluorescent embryos. I was not able to confirm the presence of *L2-mRuby2* tag in the majority of (previously) fluorescent embryos collected for gDNA extraction at 1 dpf or 5 dpf, indicating that successful integration happened only rarely. Aberrant fluorescent signal could potentially be a result of several mechanisms:

1) it is possible that most of the observed fluorescent signal stems from random integration of the repair template in euchromatic regions as previously observed in cell lines (Koch *et al.*, 2018). This hypothesis is supported by the fact that I observed similar levels of fluorescence in embryos injected only with the repair templates (in absence of CRISPR/Cas9; see **Figure 12**).
2) while not intended, it is possible that one or both sgRNAs target not only the desired genomic *ndufb7* locus, but induce CRISPR/Cas9-mediated integration elsewhere in the genome in a region that is later silenced. The 5' homology sequence of the repair template contains several ATGs in frame with the fluorophore coding sequence, which may enable expression outside the intended genomic locus.
3) although technically unexpected, fluorophore expression may happen directly from the (non-integrated) repair template itself. This in turn could explain why fluorescence is lost over time, as the injected repair template is degraded. However, this is unlikely, because the repair template was cloned into a pKHR4 backbone specifically established for zebrafish genome editing purposes which does not contain a promoter element for *in vivo* expression (Hoshijima et al., 2016a).
4) apart from potential oligomerization of fluorophores, the speckly fluorescence pattern suggests that CRISPR/Cas9-mediated fluorophore tagging of *ndufb7* happens after 1-cell stage, thus yielding mosaic expression (Costantini *et al.*, 2012; Cranfill *et al.*, 2016). It is plausible that expression of fluorophore tagged Ndufb7 is not tolerated well during embryonic development and that cells expressing the tagged variant simply die

(Popgeorgiev *et al.*, 2018). Alternatively, intolerance of the fluorophore tagged Ndufb7 variant may trigger silencing of gene expression in a heterozygous environment.

Potential silencing of the fluorophore-tagged allele might be beneficial, considering that the intended purpose of the animal model is to facilitate the study of nOXPHOS gene selection mechanisms, albeit investigation of SNP-related effects would not remain possible in this context. However, all the other scenarios described above indicate obstacles that would require a substantial amount of time and effort to overcome. There is a considerable size difference between the two fluorophores (~27 kDa) and the target protein Ndufb7 (~15 kDa). The speckly fluorescent pattern I observed both in mRNA overexpression and KI animals may indicate that the fluorescent tags form oligomers with each other (Toseland, 2013). In support of this hypothesis, we did not observe similarly bright fluorescent speckles upon overexpression of other tagged nOXPHOS subunits (Sdha, Ndufb10; results not shown). While mitochondrial import remains functional according to my immunostaining results, I was not yet able to confirm integration of the fluorophore-tagged subunit into OXPHOS complex I. Overexpression of the tagged subunit by mRNA injection and subsequent mitochondria extraction at 1 dpf did not yield a sufficient amount of material for BN-PAGE and Western Blotting (tested with > 1500 injected embryos; data not shown). Together with Dr. Piragyte-Langa, I aim to analyze the maturing KI animals for germline transmission and KI success in the future. In case of positive candidates, Dr. Piragyte-Langa and I will further evaluate the functionality of Ndufb7 in the context of OXPHOS complex I assembly and function.

5.1.2 Strategy II: Differential short epitope-tagging of two naturally occurring *ndufb7* variants with single HA and FLAG tags

As a second approach, I explored differential tagging of the two *ndufb7* variants by attachment of short epitope tags. This method solves some of the issues related to fluorophore tagging, while maintaining most of the advantages associated with fluorophore tagging (**Figure 30, strategy 2.**). The following methods remain possible: (a) tag-specific sequence amplification at the DNA and mRNA level by simple PCR screening to determine KI success and expression levels, and (b) antibody-based differentiation between the two tagged allelic variants at the protein level *in vivo* and *in vitro*. While live imaging is not an option with short epitope tags, they are much smaller in size and thus less likely to interfere with protein function. Furthermore, in contrast to fluorophore tags, oligomer formation and ROS production are not expected to be an issue with this approach. Besides enabling purification of fusion proteins, short epitope tags may increase yield, solubility and folding of the target protein they are attached to (Walls and Loughran, 2011; Zhao, Li and Liang, 2013).

I selected the short epitope tags HA and FLAG which have both been successfully used for visualization of OXPHOS proteins in the past (Artika, 2010; Protasoni *et al.*, 2020; Zhang *et al.*, 2020). HA is a strong immunoreactive epitope tag that was originally obtained from human influenza hemagglutinin (HA) in 1987 (Field *et al.*, 1988). The HA tag consists of nine amino acids (YPYDVPDYA) resulting in a molecular weight of 1.10 kDa. It can be fused to a target protein at either the N- or C-terminus and its detection with monoclonal anti-HA antibodies is highly specific (Schembri *et al.*, 2007; Moon, Kim and Rhim, 2012; Zhao, Li and Liang, 2013). FLAG is a hydrophilic epitope tag which was introduced to purify fusion proteins in 1988 (Hopp *et al.*, 1988). Due to its hydrophilic properties, FLAG is usually situated on the surface of a recombinant protein, where it is easily accessible to highly specific anti-FLAG monoclonal antibodies. The eight amino acid peptide (DYKDDDDK) has a molecular weight of 1.01 kDa.

To validate that a single HA or FLAG tag is sufficient for detection of Ndufb7, I first tested the fusion proteins using the above-described transient mRNA overexpression system. However, because overexpression of Ndufb7^{H28}-HA and Ndufb7^{N28}-FLAG mRNA does not enable pre-selection of embryos prior to immunostaining, I selected injected embryos blindly. This approach finally did not yield enough HA/FLAG signal to confirm mitochondrial localization of the tagged constructs following immunostaining with the corresponding antibodies.

I therefore generated transgenic lines stably overexpressing either *ndufb7^{H28}-HA* or *ndufb7^{N28}-FLAG* under the control of a ubiquitin promoter element in combination with the selection marker *cryaa:mCerulean.* The α-crystallin promoter induces lens-specific expression of mCerulean which first becomes visible at 2 dpf, thus enabling selection of embryos harboring the transgene from this time point onwards (Posner *et al.*, 2017). Following immunostaining of mCerulean positive embryos, I detected varying levels of Ndufb7^{H28}-HA or Ndufb7^{N28}-FLAG expression, predominantly in the embryonic tail. Co-staining for the mitochondrial marker mt-Co1 revealed that the majority of Ndufb7^{H28}-HA localizes to mitochondria. A similar expression pattern was observed for Ndufb^{N28}-FLAG. The transgene expressing embryos appeared morphologically healthy, suggesting that the short epitope tags are well tolerated in combination with Ndufb7.

It remains to be evaluated whether epitope-tagged Ndufb7 is successfully integrated into OXPHOS complex I – F0 transgenic animals *Tg(ubb:ndufb7-FLAG, cryaa:mCerulean)* and *Tg(ubb:ndufb7-HA, cryaa:mCerulean)* are currently being raised to adulthood. Incorporation of Ndufb7^{H28}-HA and Ndufb7^{N28}-FLAG into complex I will be assessed by BN-PAGE in mitochondrial extracts from F1 embryos in the future. While the transgenic lines generally appear healthy, it is currently unknown whether the short epitope tags influence protein function. It is possible that the tagged subunit is not incorporated into OXPHOS complex I in the overexpression model in physiologically healthy animals, given that the tagged variant is in direct competition with the two endogenously expressed alleles. Considering the promising

immunostaining results, Dr. Piragyte-Langa and I are therefore planning to knock in both C-terminal tags at the genomic *ndufb7* locus in the WIK strain using short ssODN repair templates and CRISPR/Cas9 tools (same approach as in 5.2.3).

5.1.3 Strategy III: Differentiation of a wildtype (K84) *ndufb7* allele and a disease variant (W84) *ndufb7* allele by CRISPR/Cas9-mediated mutagenesis

Strategies I and II have several advantages in terms of versatility and simplicity of experimental design. However, both strategies rely on artificial tags for differentiation between the two allelic variants of *ndufb7,* which may interfere with protein function or OXPHOS complex I assembly. Furthermore, there is no guarantee that the identified naturally occurring SNP in *ndufb7* in the WIK strain disturbs the OXPHOS system enough to trigger a potentially existing allelic selection mechanism. In fact, the Ndufb7 protein modeling results suggest that the SNP-induced amino acid change (H28N) may not have a significant impact on protein stability. Thus, despite the suspected allelic bias in SNP versus wildtype mRNA transcript levels, studying the impact of the H28N amino acid change may finally not reveal potential selective mechanisms, because the effect size of the amino acid change may simply be too small to trigger selection.

For this reason, Dr. Piragyte-Langa and I devised a third strategy that allows approaching the research question from a different angle: instead of differentiation between two tagged, naturally occurring *ndufb7* alleles in wildtype WIK, a wildtype allele is compared to a (likely) disease causing allelic variant with a single amino acid-difference in position 84 of the Ndufb7 peptide chain (K84>W). The NDUFB7^{R84W} variant was recently identified in a female human patient (heterozygous carrier) with a severe neuropathological disease phenotype (unpublished patient case). However, it is unclear if the W84 mutation contributes to disease in this patient. Notably, while this specific mutation has been detected in the human population 43 times according to gnomAD, there are only heterozygous carriers on record (Karczewski *et al.*, 2020).

While we do not know the exact genetic profile of the patient heterozygous for *NDUFB7^{W84}*, the described symptoms suggest high similarities to Leigh syndrome (LS), a severe, early-onset neurological disease caused by OXPHOS dysfunction. Overlapping symptoms include: sleep apnea, poor feeding, failure to thrive, asthma, seizures, contractures, scoliosis, (mild) hearing loss, nephrotic syndrome, cortical blindness and difficulty feeding (LEIGH, 1951; López *et al.*, 2006; Baertling *et al.*, 2014; Bonfante *et al.*, 2016; Lee *et al.*, 2016; Schubert and Vilarinho, 2020). According to the clinician who described the case, the patient phenotype partially resembles Angelman or Rett syndrome. Both are complex neurologic disorders which have been associated with mitochondrial abnormalities in the past (Su *et al.*, 2011; Shulyakova *et al.*, 2017).

Interestingly, several of the subunits which form the P_D-b module of OXPHOS complex I together with NDUFB7 have been associated with Leigh Syndrome, namely MT-ND5, NDUFB3 and

NDUFB8 (Alston *et al.*, 2016; Ng *et al.*, 2018; Piekutowska-Abramczuk *et al.*, 2018). NDUFB9 was found to promote breast cancer progression by mediating mitochondrial metabolism (Li *et al.*, 2015). A recently published case report furthermore associated an intronic biallelic mutation in the *NDUFB7* gene with severe congenital lactic acidosis, encephalopathy, hypertrophic cardiomyopathy, and a severe complex I defect. The authors report a decrease in protein levels of the P_D-b module subunits NDUFB7 and NDUFB8, as well as the Q module subunits NDUFS3 and NDUFS2 in patient fibroblasts (Correia *et al.*, 2021). If the W84 mutation negatively affects the functionality of NDUFB7, it is plausible that the mutation contributes to the disease phenotype observed in the above-described patient.

Indeed, computational modeling of the Ndufb^{W84} mutant variant in the zebrafish WIK strain suggests that, in contrast to the Ndufb7^{N28} SNP variant, the patient mutation likely destabilizes the protein. Ndufb7 protein folding involves disulfide bond formation (s-s) between the cysteines of the two Cx_9C domains, specifically at amino acids 59s-s90 and 69s-s80. Abnormal protein folding results in protein degradation (Reinhardt *et al.*, 2020). The K84 mutagenesis site is situated within the second of the two Cx_9C domains: CDHQ**K**HDWEYC. Given the drastic change in amino acid properties from the positively charged, hydrophilic K to the highly hydrophobic W, I asked whether the Ndufb7^{W84} mutant variant can be detected in mitochondria. To investigate this, I mutagenized *ndufb7(K84)-L2-mRuby2* mRNA plasmids previously generated for strategy I, synthesized *ndufb7(W84)-L2-mRuby2* mutant mRNA and transiently overexpressed it by mRNA injection into WIK and AB embryos. I hypothesized that in case disulfide bond formation is disrupted or destabilized by the W84 mutation, the mutant variant may be degraded faster than the wildtype, thus yielding lower amounts of fluorescent signal in comparison to the wildtype variant. However, as shown in **Figures 22B, 23, 24 and 26** I did not observe a notable difference in overall signal amount, mitochondrial localization and tolerance when comparing expression of wildtype Ndufb7^{K84}-L2-mRuby2, mutant Ndufb7^{W84}-L2-mRuby2 and two additional positive controls (human wildtype variant Ndufb7^{R84}-L2-mRuby2 and Ndufb7^{W84K}-L2-mRuby2 reverse mutagenized mRNA) in the embryonic zebrafish. This indifference in mRNA tolerance and fluorescent signal may be due to several reasons:

- Folding efficiency, protein stability and/or degradation of the mutant Ndufb7^{W84} variant may be similar to Ndufb7^{K84}. However, my mRNA overexpression experiments did not provide information on the folding status of Ndufb7.
- The fluorophore tag may trap both Ndufb7 variants in the mitochondrial IMS regardless of the folding status of Ndufb7. Potential fluorophore oligomerization may further exacerbate this effect.
- The absence of a distinctive phenotype suggests that Ndufb7^{W84} may not have a dominant-negative effect. However, any such effect may potentially be masked due to interference by the fluorophore tag.

Considering the uncertainties associated with fluorophore tags, a combined mutagenesis and short epitope-tag approach may provide more precise insights into mutant tolerance and intracellular localization.

It should be noted that the fluorescent pattern observed following overexpression of all types of *ndufb7-L2-mRuby2* mRNA (wildtype, human and patient variant) did not perfectly co-localize with the mitochondrial marker. Instead, I observed a similar amount of partially overlapping signals for the different mRNA types (examples shown in **Figures 23 and 24**).

To be able to investigate whether the Ndufb7^{W84} mutant variant has a negative effect on assembly of OXPHOS complex I when expressed at physiological levels throughout the entire organism (homozygous and/or heterozygous), I next generated mutant knock-in WIK embryos by injecting CRISPR/Cas9 and a 90-nucleotide long ssODN repair template encoding the AAA>TGG mutation. I followed a previously published mutagenesis approach (Burg *et al.*, 2018). As demonstrated in **Figure 29** I successfully validated the feasibility of this knock-in approach by confirming mutagenesis at the correct locus by Sanger sequencing. Knock-in animals (F0) are currently being raised to adulthood. Dr. Piragyte-Langa and I aim to assess correct mutagenesis and germline transmission efficiency in the F1 generation in the future.

In vivo mutagenesis does not allow for antibody-based detection, live imaging or, probably, sequence-specific PCR amplification, which complicates the analysis significantly. However, considering the predicted destabilizing effect of the W84 mutation, this type of model system may not only be useful for investigation of the research question, but also to study a previously undescribed human disease case (see **Figure 30, 3.** for strengths and weaknesses of this strategy).

5.1.4 Strengths and weaknesses of strategies I, II and III in direct comparison

The main strengths and weaknesses of the three strategies are listed in **Figure 30** and briefly summarized below.

The following two characteristics of the zebrafish model stand out in comparison to the mouse and cell culture models used by our collaborators: 1) external, rapid development and 2) transparency of zebrafish embryos, which readily enables differentiation between organs as they develop. Thanks to these characteristics, the unique strength of the approach based on fluorophore tags (strategy I) is that it permits investigation of protein expression and localization on tissue and organ level over time by live imaging, thus enabling minimally invasive analysis of allelic expression patterns by fluorescence microscopy. This type of analysis is not possible with the short epitope tags HA and FLAG (strategy II) and the mutagenesis approach (strategy III). The benefits shared by strategies I and II include the ability to use antibody-based detection

methods on protein level *in vivo* and *ex vivo* (e.g. immunofluorescence microscopy and SDS-PAGE or BN-PAGE coupled with Western Blot). Furthermore, both strategies allow for tag-specific amplification at DNA and mRNA level which enables a relatively simple and quick form of screening by PCR. Due to the absence of a detectable tag, these methods are not readily applicable in the case of strategy III.

While methodologically versatile, strategies I and II are intrusive in nature, because the introduction of a tag disrupts the natural genomic *ndufb7* locus. Artificial tags may introduce an unanticipated experimental bias by interfering with mRNA or protein folding, processing or assembly and function of the resulting protein. Considering the carefully orchestrated assembly of all subunits into OXPHOS complex I, the risk of interference with protein function likely increases with increasing tag size. This presumably puts Ndufb7 tagged with the peptides HA or FLAG at an advantage over fluorophore tags. In contrast, the unique strength of strategy III is the non-intrusive nature of the single amino-acid change in *ndufb7* induced by mutagenesis. The mutagenesis strategy thus provides a complementary approach to strategies I and II that deliberately avoids introduction of a potential tag-related artificial bias. An additional advantage of strategy III is the similarity of the amino acid change (in terms of amino acid properties) to an uncharacterized mutation identified in a human patient, thanks to which the mutagenesis model may provide valuable insights into a previously uncharacterized human disease case in the future. Interestingly, due to absence of an observable phenotype, my Ndufb7^{W84} overexpression results suggest that the mutation does not have a dominant-negative effect over the wildtype protein variant (indicated by the red half-cross in **Figure 30.3.**).

As shown in this thesis, I was able to demonstrate the technical feasibility of all three strategies. Importantly, my immunofluorescence imaging results confirm that tagged Ndufb7 subunit variants localize to mitochondria. However, the observed patterns of fluorescence suggest that where tagging of Ndufb7 is concerned, short epitope tags are preferable over the much larger fluorophore tags. The next steps for the three strategies will be: 1) determination of germline-transmission success, 2) subsequent confirmation of Ndufb7 integration into OXPHOS complex I in the offspring of knock-in animals with germline transmission and 3) crossbreeding of the established animal lines to investigate allelic expression patterns in heterozygous offspring.

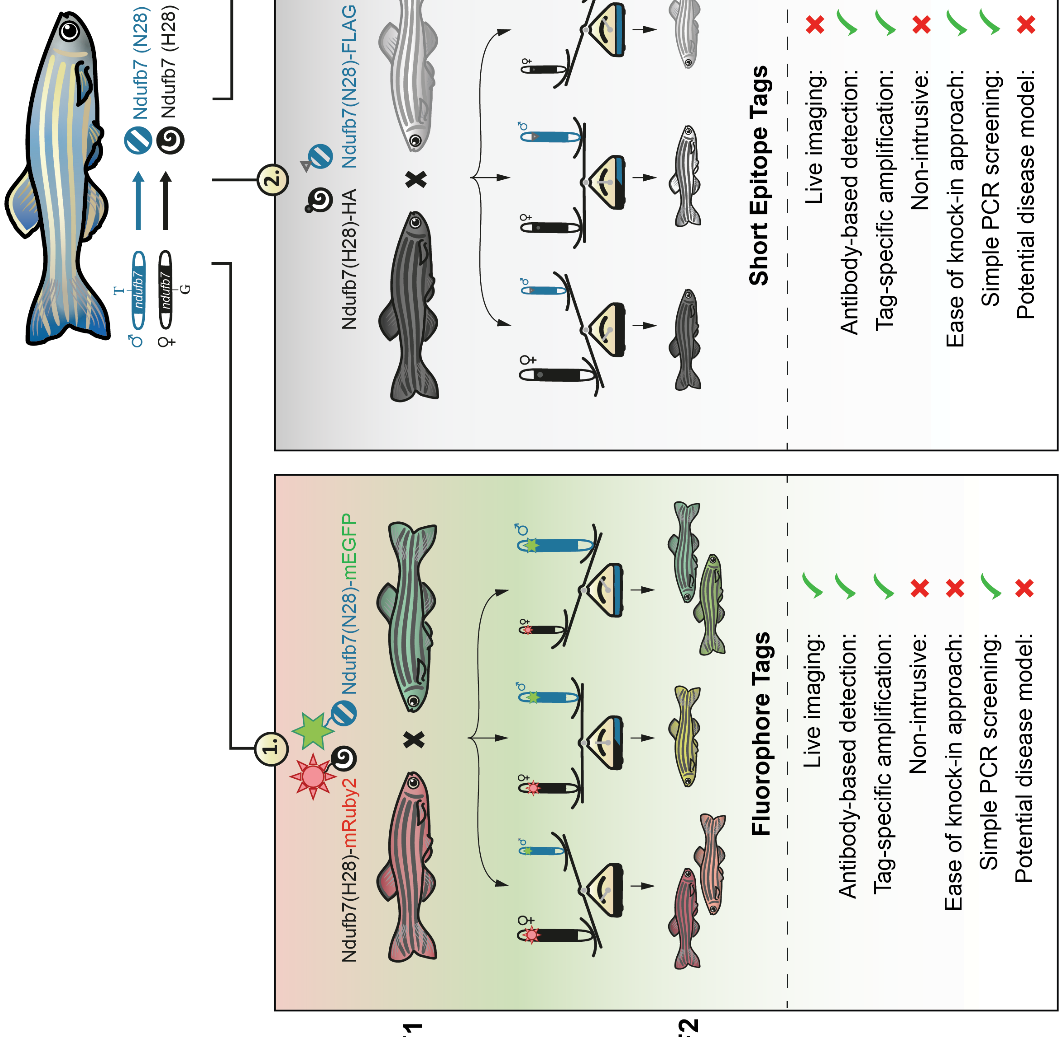


Figure 30. Methodologic strengths and weaknesses of the three strategies tested for establishment of a zebrafish model to study allelic expression of the nOXPHOS subunit ndufb7. Strategy I: differential tagging of the two naturally occurring ndufb7 alleles with fluorophores, strategy II: differential tagging of the two naturally occurring ndufb7 alleles with short epitope tags, and strategy III: differentiation of a wildtype (K84) allele and a disease variant (W84) allele. The colors blue and black represent the paternal and maternal allele, or the disease and wildtype allele, respectively. The methodologic strengths (green ticks) and weaknesses (red crosses) of each approach are listed. Because overexpression of the disease variant (W84, strategy III) did not cause an observable phenotype, a dominant-negative effect of the mutant over the wildtype variant is unlikely.

5.2 Where to go from here: Methods for investigation of allelic bias in nOXPHOS gene expression

Mitonuclear incompatibilities have been found to impact overall fitness in a plethora of eukaryotic organisms (reviewed in Lane, 2011; Nguyen *et al.*, 2020). However, the extent of SNP-dependent allelic selection of nuclear encoded OXPHOS genes remains undescribed. To address this topic, the zebrafish models described in this thesis will be used to investigate if the nuclear encoded variants of Ndufb7 (Ndufb7^{H28}, Ndufb7^{N28}, Ndufb7^{K84} and Ndufb7^{W84}) and their incorporation into OXPHOS complex I is regulated in an allele-specific manner, once germline-transmission has been confirmed. Should one of the Ndufb7 (SNP-) variant alleles indeed be preferentially utilized over the other in a heterozygous environment, as suggested by the RNA-seq results described in **section 4.1**, it will be necessary to investigate potential mechanisms of gene expression control at DNA, mRNA and / or protein level to identify the responsible cellular mechanism(s). The following paragraphs place the zebrafish models designed in this thesis project back in the context of the HFSP collaboration (introduced in **section 2**): First, I shortly describe preliminary results from our collaborators that point towards allele-specific expression of some nOXPHOS genes and name some continuative questions resulting from these early results. Second, I briefly discuss several methods irrespective of the model organism that may be useful for investigation of allele-specific gene expression control mechanisms in the future.

To look for evidence of monoallelic expression tendencies of nuclear OXPHOS genes on the level of transcription, our collaborators are currently: 1) generating RNA-seq data from mouse neurons, and 2) re-analyzing publicly available single cell RNA-seq datasets generated specifically for the purpose of investigating allele specific expression (such as Reinius *et al.*, 2016). Preliminary results suggest that a small number of nOXPHOS genes containing SNPs (for example *Cox6c, Cox5b, Cox7a2, Atp5g1* and *Atp5md1*; unpublished data) show a strong bias towards expression of one allele over the other in embryonic mouse liver, neurons, and adult fibroblasts. However, in most cases the two allelic variants seem to be represented at similar mRNA transcript levels. As a next step, these findings will be validated with complementary experiments in human cell lines.

While these preliminary results support a certain extent of the hypothesized allelic bias in transcription of SNP-containing nOXPHOS genes, it is unclear whether the detected mRNA transcript bias is reflected on the protein and/or DNA level. The same is the case for the larger number of biallelically transcribed nOXPHOS genes:

- Are there epigenetic mechanisms in place that determine allelic bias in nOXPHOS gene expression?
- Are bi- or monoallelic mRNA transcripts translated into corresponding protein amounts?
- In the case of biallelic transcripts, are both resulting protein variants finally imported into mitochondria and included in the assembly of the OXPHOS machinery with the same efficiency?
- If both variants are incorporated into the OXPHOS machinery, is there a difference in protein turnover between the two allelic variants of a nOXPHOS subunit?

All these questions cannot be addressed with RNA-seq data alone. Over the past years, epigenetic approaches have been explored, such as analysis of allele specific DNA methylation patterns by combining SNP identification via whole genome sequencing with bisulfite sequencing (Gertz *et al.*, 2011; Fang *et al.*, 2012; Abante *et al.*, 2020). However, this kind of data collection is not yet readily available for analysis of potential epigenetic allele specific control mechanisms in different tissues and/or throughout development. Similarly, Dr. Piragyte-Langa and I did not find any evidence for epigenetic control of nOXPHOS genes at the DNA level when comparing datasets on histone modifications available through the WashU Epigenome Browser (Zhou *et al.*, 2013). Therefore, we designed the herein described zebrafish models to enable targeted investigation of the research question in a bottom-up approach: from the protein level – starting from allele specific identification of nOXPHOS subunit integration into complex I (BN-PAGE), upwards to potential gene expression control mechanisms at the DNA level.

Due to rapid progress in RNA sequencing technologies, most efforts for detection of differences in the expression of allelic variants have focused on the mRNA level so far. Allele specific transcripts of nuclear genes have for example been investigated by utilizing SNP-arrays (Gimelbrant *et al.*, 2007; Jeffries *et al.*, 2012; Zwemer *et al.*, 2012), RNA-seq (Li *et al.*, 2012; Eckersley-Maslin *et al.*, 2014), RNA FISH (Hansen and Van Oudenaarden, 2013; Ke *et al.*, 2013; Levesque *et al.*, 2013), and, more recently, a central dogma tagging MS2 approach which enables allele specific visualization of mRNA and protein products in single cells (Sheinberger *et al.*, 2017). Furthermore, a recent review on combining genetics and proteomics analyses for large scale population studies suggests that RNA-seq may provide evidence for SNP-dependent allelic bias also on the protein level (Suhre, McCarthy and Schwenk, 2021). However, several proteomics studies raised the concern that allelic bias on the mRNA level is not always reflected on the protein level, likely due to differences in posttranslational processing of two allelic variants (Ohtsuki *et al.*, 2012; Vogel and Marcotte, 2012; Battle *et al.*, 2015; Kühl *et al.*, 2017; Shi *et al.*, 2018; Her *et al.*, 2020).

Both anterograde (nucleus to mitochondria) and retrograde (mitochondria to nucleus) signaling mechanisms may contribute to gene expression control of mitochondrial and nuclear OXPHOS genes. Known anterograde and retrograde control mechanisms include for example epigenetic mechanisms and signaling via transcription factors, small molecules, ROS and metabolites (Whelan and Zuckerbraun, 2013; Fang *et al.*, 2016; English *et al.*, 2020). In addition to well-described control mechanisms, a study in the Manila clam recently suggested that mitochondrial short non-coding (nc) RNAs may contribute to nuclear gene regulation by RNA interference (Passamonti *et al.*, 2020). While the function of mitochondrially encoded ncRNAs remains largely undescribed, a study in alveolar epithelial type II cells from mouse lungs found evidence that retrograde signaling by the mitochondrially encoded ncRNA mito-ncR-805 influences mRNA levels of several nOXPHOS genes, including *Ndufb7* (Blumental-Perry *et al.*, 2020). Together with better-characterized nuclear encoded ncRNAs, mitochondrially encoded ncRNAs may therefore represent a potential mechanism for transcriptional and translational control of nOXPHOS variants (Srijyothi *et al.*, 2018).

Proteomics approaches (reviewed in Timp and Timp, 2020) may be a promising tool for snap-shot analysis of SNP-dependent variation in protein expression. While time-, cost- and data-intensive proteomics methods do not readily allow detailed study of protein functionality in a tissue/cell-type specific manner throughout development of an organism, they may be useful in combination with our mutagenesis approach (strategy III) in the future. However, given our interest in allele-specific selection mechanisms of nOXPHOS genes across cell-types and developmental phases, I consider a combined molecular biology and imaging-based approach of the question – whether nOXPHOS subunits may be regulated in an allele-specific manner in diploid organisms – most versatile and convenient. While strictly computational, our collaborators protein modeling results suggest that the single amino acid change H28>N in Ndufb7 caused by a naturally occurring SNP, might finally not destabilize the resulting protein enough to trigger a selective mechanism. The identified disease mutant Ndufb7^{W84} is, however, predicted to have a potentially significant destabilizing impact on protein function. To harness the strengths of both strategies, provided epitope-tagging of Ndufb7 does not cause disturbances to the OXPHOS system following knock-in, a combination of strategies II and III may therefore be the best way forward.

6 References

Abante, J. *et al.* (2020) 'Detection of haplotype-dependent allele-specific DNA methylation in WGBS data', *Nature Communications*. doi: 10.1038/s41467-020-19077-1.

Acín-Pérez, R. *et al.* (2008) 'Respiratory Active Mitochondrial Supercomplexes', *Molecular Cell*. doi: 10.1016/j.molcel.2008.10.021.

Agip, A. N. A. *et al.* (2018) 'Cryo-em structures of complex i from mouse heart mitochondria in two biochemically defined states', *Nature Structural and Molecular Biology*. doi: 10.1038/s41594-018-0073-1.

Alberts, B. *et al.* (2014) 'Energy Conversion: Mitochondria and Chloroplasts', in *Molecular Biology of the Cell*. 6th edn. doi: 10.1201/9780203833445-14.

Alston, C. L. *et al.* (2016) 'A recurrent mitochondrial p.Trp22Arg NDUFB3 variant causes a distinctive facial appearance, short stature and a mild biochemical and clinical phenotype', *Journal of Medical Genetics*. doi: 10.1136/jmedgenet-2015-103576.

Anderson, S. *et al.* (1981) 'Sequence and organization of the human mitochondrial genome', *Nature*. doi: 10.1038/290457a0.

Argos, P. (1990) 'An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion', *Journal of Molecular Biology*. doi: 10.1016/0022-2836(90)90085-Z.

Arribat, Y. *et al.* (2019) 'Mitochondria in Embryogenesis: An Organellogenesis Perspective', *Frontiers in Cell and Developmental Biology*. doi: 10.3389/fcell.2019.00282.

ARTIKA, I. M. (2010) 'Bioenergetic Consequences of FLAG Tag Addition to the C-Terminus of Subunit 8 of Yeast Saccharomyces cerevisiae Mitochondrial ATP Synthase', *HAYATI Journal of Biosciences*. doi: 10.4308/hjb.17.3.151.

Aurora, R. *et al.* (1997) 'Local interactions in protein folding: Lessons from the α-helix', *Journal of Biological Chemistry*. doi: 10.1074/jbc.272.3.1413.

Auton, A. *et al.* (2015) 'A global reference for human genetic variation', *Nature*. doi: 10.1038/nature15393.

Baertling, F. *et al.* (2014) 'A guide to diagnosis and treatment of Leigh syndrome', *Journal of Neurology, Neurosurgery and Psychiatry*. doi: 10.1136/jnnp-2012-304426.

Balik-Meisner, M. *et al.* (2018) 'Population genetic diversity in zebrafish lines', *Mammalian Genome*. doi: 10.1007/s00335-018-9735-x.

Bar-Yaacov, D., Blumberg, A. and Mishmar, D. (2012) 'Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. doi: 10.1016/j.bbagrm.2011.10.008.

Barel, O. *et al.* (2008) 'Mitochondrial Complex III Deficiency Associated with a Homozygous Mutation in UQCRQ', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2008.03.020.

Barlow, D. J. and Thornton, J. M. (1983) 'Ion-pairs in proteins', *Journal of Molecular Biology*. doi: 10.1016/S0022-2836(83)80079-5.

Barreto, F. S. *et al.* (2018) 'Genomic signatures of mitonuclear coevolution across populations of Tigriopus californicus', *Nature Ecology and Evolution*. doi: 10.1038/s41559-018-0588-1.

Barreto, F. S. and Burton, R. S. (2013) 'Evidence for compensatory evolution of ribosomal

proteins in response to rapid divergence of mitochondrial rRNA', *Molecular Biology and Evolution*. doi: 10.1093/molbev/mss228.

Battle, A. *et al.* (2015) 'Impact of regulatory variation from RNA to protein', *Science*. doi: 10.1126/science.1260793.

Birch-Machin, M. A. *et al.* (2000) 'Late-onset optic atrophy, ataxia, and myopathy associated with a mutation of a complex II gene', *Annals of Neurology*. doi: 10.1002/1531-8249(200009)48:3<330::AID-ANA7>3.0.CO;2-A.

Biterge, B. and Schneider, R. (2014) 'Histone variants: Key players of chromatin', *Cell and Tissue Research*, pp. 457–466. doi: 10.1007/s00441-014-1862-4.

Blumental-Perry, A. *et al.* (2020) 'Retrograde signaling by a mtDNA-encoded non-coding RNA preserves mitochondrial bioenergetics', *Communications Biology*. doi: 10.1038/s42003-020-01322-4.

Bonfante, E. *et al.* (2016) 'The neuroimaging of Leigh syndrome: case series and review of the literature', *Pediatric Radiology*. doi: 10.1007/s00247-015-3523-5.

Bourgeron, T. *et al.* (1995) 'Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency', *Nature Genetics*. doi: 10.1038/ng1095-144.

Broncel, M. *et al.* (2015) 'Multifunctional reagents for quantitative proteome-wide analysis of protein modification in human cells and dynamic profiling of protein lipidation during vertebrate development', *Angewandte Chemie - International Edition*. doi: 10.1002/anie.201500342.

Brown, W. M., George, M. and Wilson, A. C. (1979) 'Rapid evolution of animal mitochondrial DNA', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.76.4.1967.

Burg, L. *et al.* (2018) 'Conditional mutagenesis by oligonucleotide-mediated integration of loxP sites in zebrafish', *PLoS Genetics*. doi: 10.1371/journal.pgen.1007754.

Burgstaller, J. P. *et al.* (2014) 'MtDNA Segregation in Heteroplasmic Tissues Is Common InVivo and Modulated by Haplotype Differences and Developmental Stage', *Cell Reports*. doi: 10.1016/j.celrep.2014.05.020.

Burke, E. (2016) 'Why Use Zebrafish to Study Human Disease?', *I am Intramural Blog*.

Burr, S. P., Pezet, M. and Chinnery, P. F. (2018) 'Mitochondrial DNA Heteroplasmy and Purifying Selection in the Mammalian Female Germ Line', *Development Growth and Differentiation*. doi: 10.1111/dgd.12420.

Burton, R. S., Ellison, C. K. and Harrison, J. S. (2006) 'The sorry state of F2 hybrids: Consequences of rapid mitochondrial DNA evolution in allopatric populations', in *American Naturalist*. doi: 10.1086/509046.

Campbell, P. D. and Marlow, F. L. (2013) 'Temporal and tissue specific gene expression patterns of the zebrafish kinesin-1 heavy chain family, kif5s, during development', *Gene Expression Patterns*. doi: 10.1016/j.gep.2013.05.002.

Carroll, J. *et al.* (2006) 'Bovine complex I is a complex of 45 different subunits', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M607135200.

Chaban, Y., Boekema, E. J. and Dudkina, N. V. (2014) 'Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation', *Biochimica et Biophysica Acta - Bioenergetics*. doi: 10.1016/j.bbabio.2013.10.004.

Chen, X., Zaro, J. L. and Shen, W. C. (2013) 'Fusion protein linkers: Property, design and

functionality', *Advanced Drug Delivery Reviews*. doi: 10.1016/j.addr.2012.09.039.

Cokus, S. J. *et al.* (2019) 'Tissue-specific transcriptomes reveal gene expression trajectories in two maturing skin epithelial layers in zebrafish embryos', *G3: Genes, Genomes, Genetics*. doi: 10.1534/g3.119.400402.

Correia, S. P. *et al.* (2021) 'Severe congenital lactic acidosis and hypertrophic cardiomyopathy caused by an intronic variant in NDUFB7.', *Human mutation*. United States. doi: 10.1002/humu.24173.

Costantini, L. M. *et al.* (2012) 'Assessing the Tendency of Fluorescent Proteins to Oligomerize Under Physiologic Conditions', *Traffic*. doi: 10.1111/j.1600-0854.2012.01336.x.

Cranfill, P. J. *et al.* (2016) 'Quantitative assessment of fluorescent proteins', *Nature Methods*. doi: 10.1038/nmeth.3891.

Dobrucki, J. W., Feret, D. and Noatynska, A. (2007) 'Scattering of exciting light by live cells in fluorescence confocal imaging: Phototoxic effects and relevance for FRAP studies', *Biophysical Journal*. doi: 10.1529/biophysj.106.096636.

Dröse, S. *et al.* (2011) 'Functional dissection of the proton pumping modules of mitochondrial complex I', *PLoS Biology*. doi: 10.1371/journal.pbio.1001128.

Dudkina, N. V. *et al.* (2010) 'Structure and function of mitochondrial supercomplexes', *Biochimica et Biophysica Acta - Bioenergetics*. doi: 10.1016/j.bbabio.2009.12.013.

Eckersley-Maslin, M. A. *et al.* (2014) 'Random monoallelic gene expression increases upon embryonic stem cell differentiation', *Developmental Cell*. doi: 10.1016/j.devcel.2014.01.017.

English, J. *et al.* (2020) 'Decoding the rosetta stone of mitonuclear communication', *Pharmacological Research*. doi: 10.1016/j.phrs.2020.105161.

Erhardt, H. *et al.* (2014) 'Organization of the Escherichia coli aerobic enzyme complexes of oxidative phosphorylation in dynamic domains within the cytoplasmic membrane', *MicrobiologyOpen*. doi: 10.1002/mbo3.163.

Fang, E. F. *et al.* (2016) 'Nuclear DNA damage signalling to mitochondria in ageing', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm.2016.14.

Fang, F. *et al.* (2012) 'Genomic landscape of human allele-specific DNA methylation', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1201310109.

Field, J. *et al.* (1988) 'Purification of a RAS-responsive adenylyl cyclase complex from Saccharomyces cerevisiae by use of an epitope addition method.', *Molecular and Cellular Biology*. doi: 10.1128/mcb.8.5.2159.

Floros, V. I. *et al.* (2018) 'Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos', *Nature Cell Biology*. doi: 10.1038/s41556-017-0017-8.

Formosa, L. E. *et al.* (2018) 'Building a complex complex: Assembly of mitochondrial respiratory chain complex I', *Seminars in Cell and Developmental Biology*. doi: 10.1016/j.semcdb.2017.08.011.

Formosa, L. E. and Ryan, M. T. (2018) 'Mitochondrial OXPHOS complex assembly lines', *Nature Cell Biology*. doi: 10.1038/s41556-018-0098-z.

Friederich, M. W. *et al.* (2017) 'Mutations in the accessory subunit NDUFB10 result in isolated complex I deficiency and illustrate the critical role of intermembrane space import for complex I

holoenzyme assembly', *Human Molecular Genetics*. doi: 10.1093/hmg/ddw431.

Gemberling, M. *et al.* (2013) 'The zebrafish as a model for complex tissue regeneration', *Trends in Genetics*. doi: 10.1016/j.tig.2013.07.003.

George, R. A. and Heringa, J. (2002) 'An analysis of protein domain linkers: Their classification and role in protein folding', *Protein Engineering*. doi: 10.1093/protein/15.11.871.

Gertz, J. *et al.* (2011) 'Analysis of dna methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation', *PLoS Genetics*. doi: 10.1371/journal.pgen.1002228.

Ghezzi, D. and Zeviani, M. (2012) 'Assembly factors of human mitochondrial respiratory chain complexes: Physiology and pathophysiology', *Advances in Experimental Medicine and Biology*. doi: 10.1007/978-1-4614-3573-0_4.

Gimelbrant, A. *et al.* (2007) 'Widespread monoallelic expression on human autosomes', *Science*. doi: 10.1126/science.1148910.

Goedhart, J. *et al.* (2012) 'Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%', *Nature Communications*. doi: 10.1038/ncomms1738.

González-Rosa, J. M. *et al.* (2011) 'Extensive scar formation and regression during heart regeneration after cryoinjury in zebrafish', *Development*. doi: 10.1242/dev.060897.

Griesbeck, O. *et al.* (2001) 'Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M102815200.

Grunwald, D. J. and Eisen, J. S. (2002) 'Headwaters of the zebrafish - emergence of a new model vertebrate', *Nature Reviews Genetics*. doi: 10.1038/nrg892.

Guerrero-Castillo, S. *et al.* (2017) 'The Assembly Pathway of Mitochondrial Respiratory Chain Complex I', *Cell Metabolism*. doi: 10.1016/j.cmet.2016.09.002.

Guo, R. *et al.* (2017) 'Architecture of Human Mitochondrial Respiratory Megacomplex I2III2IV2', *Cell*. doi: 10.1016/j.cell.2017.07.050.

Guryev, V. *et al.* (2006) 'Genetic variation in the zebrafish', *Genome Research*. doi: 10.1101/gr.4791006.

Haack, T. B. *et al.* (2012) 'Mutation screening of 75 candidate genes in 152 complex I deficiency cases identifies pathogenic variants in 16 genes including NDUFB9', *Journal of Medical Genetics*. doi: 10.1136/jmedgenet-2011-100577.

Hansen, C. H. and Van Oudenaarden, A. (2013) 'Allele-specific detection of single mRNA molecules in situ', *Nature Methods*. doi: 10.1038/nmeth.2601.

Havird, J. C. and Sloan, D. B. (2016) 'The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msw185.

Hayashi, Y. *et al.* (2017) 'Distinct requirements for energy metabolism in mouse primordial germ cells and their reprogramming to embryonic germ cells', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1620915114.

Heppert, J. K. *et al.* (2016) 'Comparative assessment of fluorescent proteins for in vivo imaging in an animal model system', *Molecular Biology of the Cell*. doi: 10.1091/mbc.E16-01-0063.

Her, L. *et al.* (2020) 'Targeted proteomics analysis revealed the imbalanced allele-specific protein expression of carboxylesterase1 in human livers', *The FASEB Journal*. doi: 10.1096/fasebj.2020.34.s1.09320.

Hill, G. E. *et al.* (2019) 'Assessing the fitness consequences of mitonuclear interactions in natural populations', *Biological Reviews*. doi: 10.1111/brv.12493.

Hopp, T. P. *et al.* (1988) 'A short polypeptide marker sequence useful for recombinant protein identification and purification', *Bio/Technology*. doi: 10.1038/nbt1088-1204.

Hoshijima, K., Jurynec, M. J. and Grunwald, D. J. (2016) 'Precise Editing of the Zebrafish Genome Made Simple and Efficient', *Developmental Cell*. doi: 10.1016/j.devcel.2016.02.015.

Hoshijima, K., Jurynec, M. J. and Grunwald, D. J. (2016) 'Precise genome editing by homologous recombination', *Methods in Cell Biology*. doi: 10.1016/bs.mcb.2016.04.008.

Van Houten, B., Hunter, S. E. and Meyer, J. N. (2016) 'Mitochondrial DNA damage induced autophagy, cell death, and disease', *Frontiers in Bioscience - Landmark*. doi: 10.2741/4375.

Howe, K. *et al.* (2013) 'The zebrafish reference genome sequence and its relationship to the human genome', *Nature*. doi: 10.1038/nature12111.

Hunte, C., Zickermann, V. and Brandt, U. (2010) 'Functional modules and structural basis of conformational coupling basis of conformational coupling in mitochondrial complex I', *Science*. doi: 10.1126/science.1191046.

Jao, L. E., Wente, S. R. and Chen, W. (2013) 'Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1308335110.

Jeffries, A. R. *et al.* (2012) 'Stochastic choice of allelic expression in human neural stem cells', *Stem Cells*. doi: 10.1002/stem.1155.

Johnson, A. B., Lewis, A. J. and Et Al. (2002) 'An Overview of Gene Control', *Molecular Biology of the Cell*.

Karczewski, K. J. *et al.* (2020) 'The mutational constraint spectrum quantified from variation in 141,456 humans', *Nature*, 581(7809), pp. 434–443. doi: 10.1038/s41586-020-2308-7.

Kasashima, K., Nagao, Y. and Endo, H. (2014) 'Dynamic regulation of mitochondrial genome maintenance in germ cells', *Reproductive Medicine and Biology*. doi: 10.1007/s12522-013-0162-0.

Kasiviswanathan, R. *et al.* (2012) 'Human mitochondrial DNA polymerase γ exhibits potential for bypass and mutagenesis at UV-induced cyclobutane thymine dimers', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M111.306852.

Kawakami, K. (2007) 'Tol2: A versatile gene transfer vector in vertebrates', *Genome Biology*. doi: 10.1186/gb-2007-8-s1-s7.

Kazak, L., Reyes, A. and Holt, I. J. (2012) 'Minimizing the damage: Repair pathways keep mitochondrial DNA intact', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm3439.

Ke, R. *et al.* (2013) 'In situ sequencing for RNA analysis in preserved tissue and cells', *Nature Methods*. doi: 10.1038/nmeth.2563.

Kellogg, E. H., Leaver-Fay, A. and Baker, D. (2011) 'Role of conformational sampling in computing mutation-induced changes in protein structure and stability', *Proteins: Structure, Function and Bioinformatics*. doi: 10.1002/prot.22921.

Koch, B. *et al.* (2018) 'Generation and validation of homozygous fluorescent knock-in cells using CRISPR-Cas9 genome editing', *Nature Protocols*. doi: 10.1038/nprot.2018.042.

Koopman, W. J. H. *et al.* (2013) 'OXPHOS mutations and neurodegeneration', *EMBO Journal*. doi: 10.1038/emboj.2012.300.

Kotrys, A. V. and Szczesny, R. J. (2019) ‘Mitochondrial Gene Expression and Beyond-Novel Aspects of Cellular Physiology’, *Cells*. doi: 10.3390/cells9010017.

Kühl, I. *et al.* (2017) ‘Transcriptomic and proteomic landscape of mitochondrial dysfunction reveals secondary coenzyme Q deficiency in mammals’, *eLife*. doi: 10.7554/eLife.30952.

Kühlbrandt, W. (2019) ‘Structure and Mechanisms of F-Type ATP Synthases’, *Annual Review of Biochemistry*. doi: 10.1146/annurev-biochem-013118-110903.

Lam, A. J. *et al.* (2012) ‘Improving FRET dynamic range with bright green and red fluorescent proteins’, *Nature Methods*. doi: 10.1038/nmeth.2171.

Lane, N. and Martin, W. (2010) ‘The energetics of genome complexity’, *Nature*. doi: 10.1038/nature09486.

Latorre-Pellicer, A. *et al.* (2016) ‘Mitochondrial and nuclear DNA matching shapes metabolism and healthy ageing’, *Nature*. doi: 10.1038/nature18618.

Lechuga-Vieco, A. V. *et al.* (2020) ‘Cell identity and nucleo-mitochondrial genetic context modulate OXPHOS performance and determine somatic heteroplasmy dynamics’, *Science Advances*. doi: 10.1126/sciadv.aba5345.

Lee, J. S. *et al.* (2016) ‘Leigh syndrome in childhood: Neurologic progression and functional outcome’, *Journal of Clinical Neurology (Korea)*. doi: 10.3988/jcn.2016.12.2.181.

LEIGH, D. (1951) ‘Subacute necrotizing encephalomyelopathy in an infant.’, *Journal of neurology, neurosurgery, and psychiatry*. doi: 10.1136/jnnp.14.3.216.

Letts, J. A. and Sazanov, L. A. (2017) ‘Clarifying the supercomplex: The higher-order organization of the mitochondrial electron transport chain’, *Nature Structural and Molecular Biology*. doi: 10.1038/nsmb.3460.

Levesque, M. J. *et al.* (2013) ‘Visualizing SNVs to quantify allele-specific expression in single cells’, *Nature Methods*. doi: 10.1038/nmeth.2589.

Li, K. *et al.* (2014) ‘Optimization of genome engineering approaches with the CRISPR/Cas9 system’, *PLoS ONE*. doi: 10.1371/journal.pone.0105779.

Li, L. D. *et al.* (2015) ‘Down-regulation of NDUFB9 promotes breast cancer cell proliferation, metastasis by mediating mitochondrial metabolism’, *PLoS ONE*. doi: 10.1371/journal.pone.0144441.

Li, S. M. *et al.* (2012) ‘Transcriptome-wide survey of mouse CNS-Derived cells reveals monoallelic expression within novel gene families’, *PLoS ONE*. doi: 10.1371/journal.pone.0031751.

Lin, S. *et al.* (2014) ‘Comparison of the transcriptional landscapes between human and mouse tissues’, *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1413624111.

López, L. C. *et al.* (2006) ‘Leigh syndrome with nephropathy and CoQ10 deficiency due to decaprenyl diphosphote synthase subunit 2 (PDSS2) mutations’, *American Journal of Human Genetics*. doi: 10.1086/510023.

Lynch, M. (1997) ‘Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes’, *Molecular Biology and Evolution*. doi: 10.1093/oxfordjournals.molbev.a025834.

Marques, I. J., Lupi, E. and Mercader, N. (2019) ‘Model systems for regeneration: Zebrafish’, *Development (Cambridge)*. doi: 10.1242/dev.167692.

Mashal, R. D., Koontz, J. and Sklar, J. (1995) ‘Detection of mutations by cleavage of DNA

heteroduplexes with bacteriophage resolvases', *Nature Genetics*. doi: 10.1038/ng0295-177.

Meeker, N. D. *et al.* (2007) 'Method for isolation of PCR-ready genomic DNA from zebrafish tissues', *BioTechniques*. doi: 10.2144/000112619.

Mella-Alvarado, V. *et al.* (2013) 'Tissue and cell-specific transcriptional activity of the human cytomegalovirus immediate early gene promoter (UL123) in zebrafish', *Gene Expression Patterns*. doi: 10.1016/j.gep.2013.01.003.

Mitchell, P. (1961) 'Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism', *Nature*. doi: 10.1038/191144a0.

Moon, J. M., Kim, G. Y. and Rhim, H. (2012) 'A new idea for simple and rapid monitoring of gene expression: requirement of nucleotide sequences encoding an N-terminal HA tag in the T7 promoter-driven expression in E. coli.', *Biotechnology letters*. doi: 10.1007/s10529-012-0966-8.

Mosimann, C. *et al.* (2011) 'Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish', *Development*. doi: 10.1242/dev.059345.

Neiman, M. and Taylor, D. R. (2009) 'The causes of mutation accumulation in mitochondrial genomes', *Proceedings of the Royal Society B: Biological Sciences*. doi: 10.1098/rspb.2008.1758.

Ng, Y. S. *et al.* (2018) 'MT-ND5 Mutation Exhibits Highly Variable Neurological Manifestations at Low Mutant Load', *EBioMedicine*. doi: 10.1016/j.ebiom.2018.02.010.

Nieto-Arellano, R. and Sánchez-Iranzo, H. (2019) 'ZfRegeneration: A database for gene expression profiling during regeneration', *Bioinformatics*. doi: 10.1093/bioinformatics/bty659.

Ohtsuki, S. *et al.* (2012) 'Simultaneous absolute protein quantification of transporters, cytochromes P450, and UDP-glucuronosyltransferases as a novel approach for the characterization of individual human liver: Comparison with mRNA levels and activities', *Drug Metabolism and Disposition*. doi: 10.1124/dmd.111.042259.

Osada, N. and Akashi, H. (2012) 'Mitochondrial-nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome c oxidase complex', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msr211.

Palego, L. *et al.* (2016) 'Tryptophan biochemistry: Structural, nutritional, metabolic, and medical aspects in humans', *Journal of Amino Acids*. doi: 10.1155/2016/8952520.

Passamonti, M. *et al.* (2020) 'Clues of in vivo nuclear gene regulation by mitochondrial short non-coding RNAs', *Scientific Reports*. doi: 10.1038/s41598-020-65084-z.

Pfanner, N., Warscheid, B. and Wiedemann, N. (2019) 'Mitochondrial proteins: from biogenesis to functional networks', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/s41580-018-0092-0.

Piekutowska-Abramczuk, D. *et al.* (2018) 'NDUFB8 Mutations Cause Mitochondrial Complex I Deficiency in Individuals with Leigh-like Encephalomyopathy', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2018.01.008.

Popgeorgiev, N. *et al.* (2018) 'Control of Programmed Cell Death During Zebrafish Embryonic Development', in *Recent Advances in Zebrafish Researches*. doi: 10.5772/intechopen.74494.

Posner, M. *et al.* (2017) 'The zebrafish as a model system for analyzing mammalian and native α-crystallin promoter function', *PeerJ*. doi: 10.7717/peerj.4093.

Protasoni, M. *et al.* (2020) ' Respiratory supercomplexes act as a platform for complex III -mediated maturation of human mitochondrial complexes I and IV ', *The EMBO Journal*. doi:

10.15252/embj.2019102817.

Raj, A. *et al.* (2010) 'Variability in gene expression underlies incomplete penetrance', *Nature*. doi: 10.1038/nature08781.

Rauch, G.-J., Granato, M. and Haffter, P. (1997) 'A polymorphic zebrafish line for genetic mapping using SSLPs on high-percentage agarose gels', *Technical Tips Online*. doi: 10.1016/s1366-2120(08)70068-0.

Reinhardt, C. *et al.* (2020) 'AIF meets the CHCHD4/Mia40-dependent mitochondrial import pathway', *Biochimica et Biophysica Acta - Molecular Basis of Disease*. doi: 10.1016/j.bbadis.2020.165746.

Reinius, B. *et al.* (2016) 'Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq', *Nature Genetics*. doi: 10.1038/ng.3678.

Rich, P. (2003) 'Chemiosmotic coupling: The cost of living', *Nature*. doi: 10.1038/421583a.

Rieger, B. *et al.* (2017) 'Lifetime imaging of GFP at CoxVIIIa reports respiratory supercomplex assembly in live cells', *Scientific Reports*. doi: 10.1038/srep46055.

Rutter, J., Winge, D. R. and Schiffman, J. D. (2010) 'Succinate dehydrogenase - Assembly, regulation and role in human disease', *Mitochondrion*. doi: 10.1016/j.mito.2010.03.001.

Ruzicka, L. *et al.* (2019) 'The Zebrafish Information Network: New support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources', *Nucleic Acids Research*. doi: 10.1093/nar/gky1090.

Schembri, L. *et al.* (2007) 'The HA tag is cleaved and loses immunoreactivity during apoptosis', *Nature Methods*. doi: 10.1038/nmeth0207-107.

Schenborn, E. T. and Mierendorf, R. C. (1985) 'A novel transcription property of SP6 and T7 RNA polymerases: Dependence on template structure', *Nucleic Acids Research*. doi: 10.1093/nar/13.17.6223.

Schmidt, T. R. *et al.* (2001) 'Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome c oxidase', *Molecular Biology and Evolution*. doi: 10.1093/oxfordjournals.molbev.a003836.

Schubert, M. B. and Vilarinho, L. (2020) 'Molecular basis of Leigh syndrome: A current look', *Orphanet Journal of Rare Diseases*. doi: 10.1186/s13023-020-1297-9.

Sharpley, M. S. *et al.* (2012) 'Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition', *Cell*. doi: 10.1016/j.cell.2012.09.004.

Sheinberger, J. *et al.* (2017) 'CD-tagging-MS2: Detecting allelic expression of endogenous mRNAs and their protein products in single cells', *Biology Methods and Protocols*. doi: 10.1093/biomethods/bpx004.

Shi, J. *et al.* (2018) 'Determining Allele-Specific Protein Expression (ASPE) Using a Novel Quantitative Concatamer Based Proteomics Method', *Journal of Proteome Research*. doi: 10.1021/acs.jproteome.8b00620.

Shulyakova, N. *et al.* (2017) 'Mitochondrial dysfunction in the pathogenesis of rett syndrome: Implications for mitochondria-targeted therapies', *Frontiers in Cellular Neuroscience*. doi: 10.3389/fncel.2017.00058.

Signes, A. and Fernandez-Vizarra, E. (2018) 'Assembly of mammalian oxidative phosphorylation complexes I–V and supercomplexes', *Essays in Biochemistry*. doi: 10.1042/EBC20170098.

Sloan, D. B. *et al.* (2014) 'Cytonuclear interactions and relaxed selection accelerate sequence

evolution in organelle ribosomes', *Molecular Biology and Evolution*. doi: 10.1093/molbev/mst259.

Van Der Sluis, E. O. *et al.* (2015) 'Parallel structural evolution of mitochondrial ribosomes and OXPHOS complexes', *Genome Biology and Evolution*. doi: 10.1093/gbe/evv061.

Sobreira, N. *et al.* (2015) 'GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene', *Human Mutation*. doi: 10.1002/humu.22844.

Song, Y. *et al.* (2013) 'High-resolution comparative modeling with RosettaCM', *Structure*. doi: 10.1016/j.str.2013.08.005.

Srijyothi, L. *et al.* (2018) 'Roles of Non-Coding RNAs in Transcriptional Regulation', in *Transcriptional and Post-transcriptional Regulation*. doi: 10.5772/intechopen.76125.

Stock, D. *et al.* (2000) 'The rotary mechanism of ATP synthase', *Current Opinion in Structural Biology*. doi: 10.1016/S0959-440X(00)00147-0.

Stoldt, S. *et al.* (2018) 'Spatial orchestration of mitochondrial translation and OXPHOS complex assembly', *Nature Cell Biology*. doi: 10.1038/s41556-018-0090-7.

Streisinger, G. *et al.* (1981) 'Production of clones of homozygous diploid zebra fish (Brachydanio rerio)', *Nature*. doi: 10.1038/291293a0.

Stroud, D. A. *et al.* (2016) 'Accessory subunits are integral for assembly and function of human mitochondrial complex i', *Nature*. doi: 10.1038/nature19754.

Su, H. *et al.* (2011) 'Mitochondrial dysfunction in CA1 hippocampal neurons of the UBE3A deficient mouse model for Angelman syndrome', *Neuroscience Letters*. doi: 10.1016/j.neulet.2009.06.079.

Suhre, K., McCarthy, M. I. and Schwenk, J. M. (2021) 'Genetics meets proteomics: perspectives for large population-based studies', *Nature Reviews Genetics*. doi: 10.1038/s41576-020-0268-2.

Sun, F. *et al.* (2005) 'Crystal structure of mitochondrial respiratory membrane protein Complex II', *Cell*. doi: 10.1016/j.cell.2005.05.025.

Szklarczyk, R. *et al.* (2011) 'NDUFB7 and NDUFA8 are located at the intermembrane surface of complex i', *FEBS Letters*. doi: 10.1016/j.febslet.2011.01.046.

Timp, W. and Timp, G. (2020) 'Beyond mass spectrometry, the next step in proteomics', *Science Advances*. doi: 10.1126/sciadv.aax8978.

Toseland, C. P. (2013) 'Fluorescent labeling and modification of proteins', *Journal of Chemical Biology*. doi: 10.1007/s12154-013-0094-5.

Turelli, M. and Orr, H. A. (2000) 'Dominance, epistasis and the genetics of postzygotic isolation', *Genetics*.

Udenwobele, D. I. *et al.* (2017) 'Myristoylation: An important protein modification in the immune response', *Frontiers in Immunology*. doi: 10.3389/fimmu.2017.00751.

Ueda, E. *et al.* (2019) 'Myristoyl group-aided protein import into the mitochondrial intermembrane space', *Scientific Reports*. doi: 10.1038/s41598-018-38016-1.

Uhlén, M. *et al.* (2015) 'Tissue-based map of the human proteome', *Science*. doi: 10.1126/science.1260419.

Vogel, C. and Marcotte, E. M. (2012) 'Insights into the regulation of protein abundance from proteomic and transcriptomic analyses', *Nature Reviews Genetics*. doi: 10.1038/nrg3185.

Vogel, R. O. *et al.* (2007) 'Identification of mitochondrial complex I assembly intermediates by tracing tagged NDUFS3 demonstrates the entry point of mitochondrial subunits', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M609410200.

Walker, J. E. *et al.* (1992) 'Sequences of 20 subunits of NADH: Ubiquinone oxidoreductase from bovine heart mitochondria. Application of a novel strategy for sequencing proteins using the polymerase chain reaction', *Journal of Molecular Biology*. doi: 10.1016/0022-2836(92)91052-Q.

Walls, D. and Loughran, S. T. (2011) 'Tagging recombinant proteins to enhance solubility and aid purification.', *Methods in molecular biology (Clifton, N.J.)*. doi: 10.1007/978-1-60761-913-0_9.

Wei, W. *et al.* (2019) 'Germline selection shapes human mitochondrial DNA diversity', *Science*. doi: 10.1126/science.aau6520.

Whelan, S. P. and Zuckerbraun, B. S. (2013) 'Mitochondrial signaling: Forwards, backwards, and in between', *Oxidative Medicine and Cellular Longevity*. doi: 10.1155/2013/351613.

Wilkens, V., Kohl, W. and Busch, K. (2013) 'Restricted diffusion of OXPHOS complexes in dynamic mitochondria delays their exchange between cristae and engenders a transitory mosaic distribution', *Journal of Cell Science*. doi: 10.1242/jcs.108852.

Wirth, C. *et al.* (2016) 'Structure and function of mitochondrial complex i', *Biochimica et Biophysica Acta - Bioenergetics*. doi: 10.1016/j.bbabio.2016.02.013.

Wittig, I., Braun, H. P. and Schägger, H. (2006) 'Blue native PAGE', *Nature Protocols*. doi: 10.1038/nprot.2006.62.

Wright, A. F. (2005) 'Genetic Variation: Polymorphisms and Mutations', in *Encyclopedia of Life Sciences*. doi: 10.1038/npg.els.0005005.

Xia, D. *et al.* (1997) 'Crystal structure of the cytochrome bc1 complex from bovine heart mitochondria', *Science*. doi: 10.1126/science.277.5322.60.

Xu, X., Lai, Y. and Hua, Z. C. (2019) 'Apoptosis and apoptotic body: Disease message and therapeutic target potentials', *Bioscience Reports*. doi: 10.1042/BSR20180992.

Yan, Z., Ye, G. and Werren, J. (2018) 'Evolutionary rate coevolution between mitochondria and mitochondria-associated nuclear-encoded proteins in insects', *bioRxiv*. doi: 10.1101/288456.

Yuan, S. and Sun, Z. (2009) 'Microinjection of mRNA and morpholino antisense oligonucleotides in zebrafish embryos', *Journal of Visualized Experiments*. doi: 10.3791/1113.

Zacharias, D. A. *et al.* (2002) 'Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells', *Science*. doi: 10.1126/science.1068539.

Zhang, S. *et al.* (2020) 'Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly', *Nature Communications*. doi: 10.1038/s41467-020-14999-2.

Zhao, X., Li, G. and Liang, S. (2013) 'Several affinity tags commonly used in chromatographic purification', *Journal of Analytical Methods in Chemistry*. doi: 10.1155/2013/581093.

Zheng, Q. *et al.* (2014) 'The contribution of reactive oxygen species to the photobleaching of organic fluorophores', *Photochemistry and Photobiology*. doi: 10.1111/php.12204.

Zhou, X. *et al.* (2013) 'Exploring long-range genome interactions using the WashU Epigenome Browser', *Nature Methods*. doi: 10.1038/nmeth.2440.

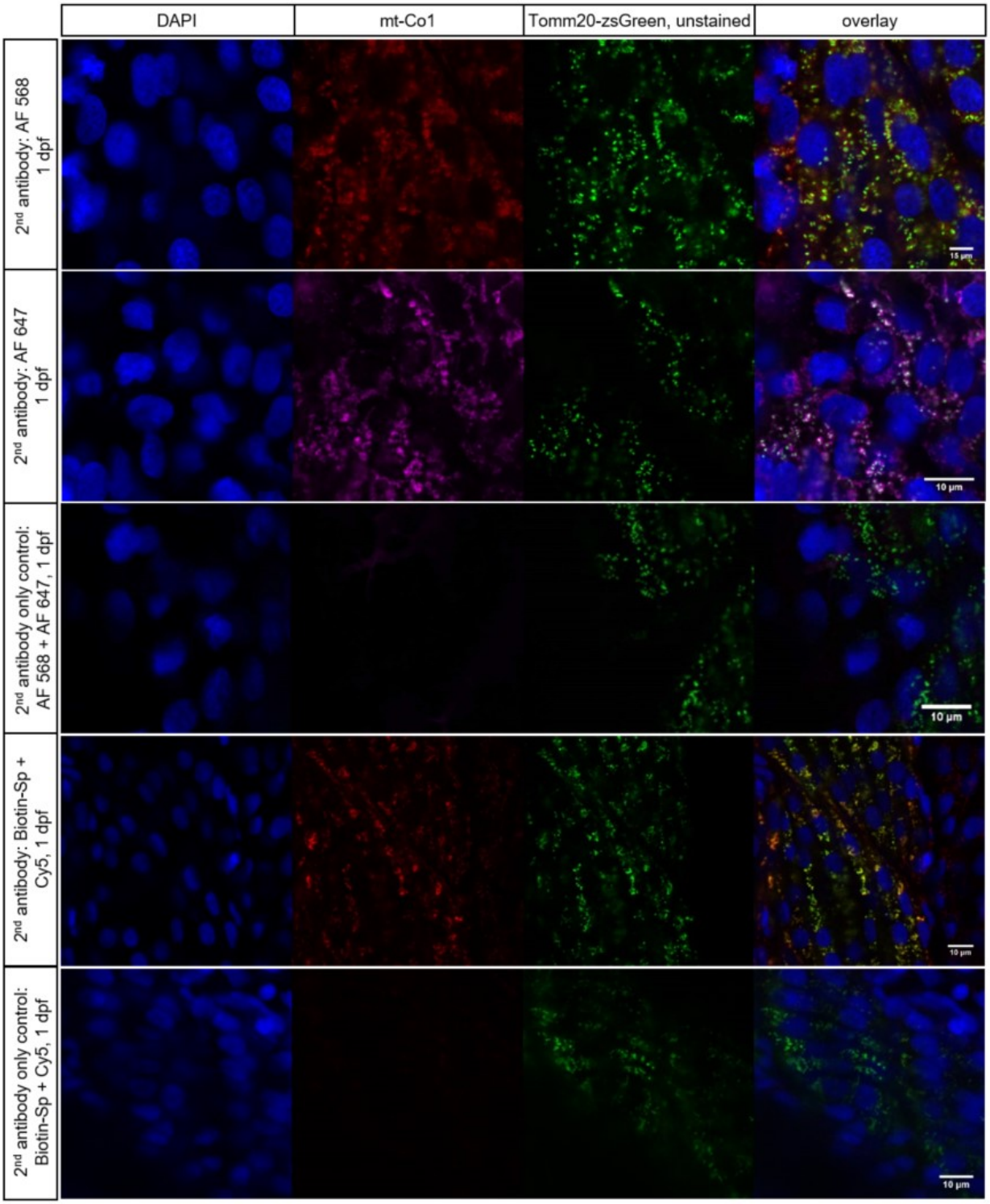
Zhu, J. *et al.* (2015) 'Structure of subcomplex Iβ of mammalian respiratory complex I leads to new supernumerary subunit assignments', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1510577112.

Zong, S. *et al.* (2018) ‘Structure of the intact 14-subunit human cytochrome c oxidase’, *Cell Research*. doi: 10.1038/s41422-018-0071-1.

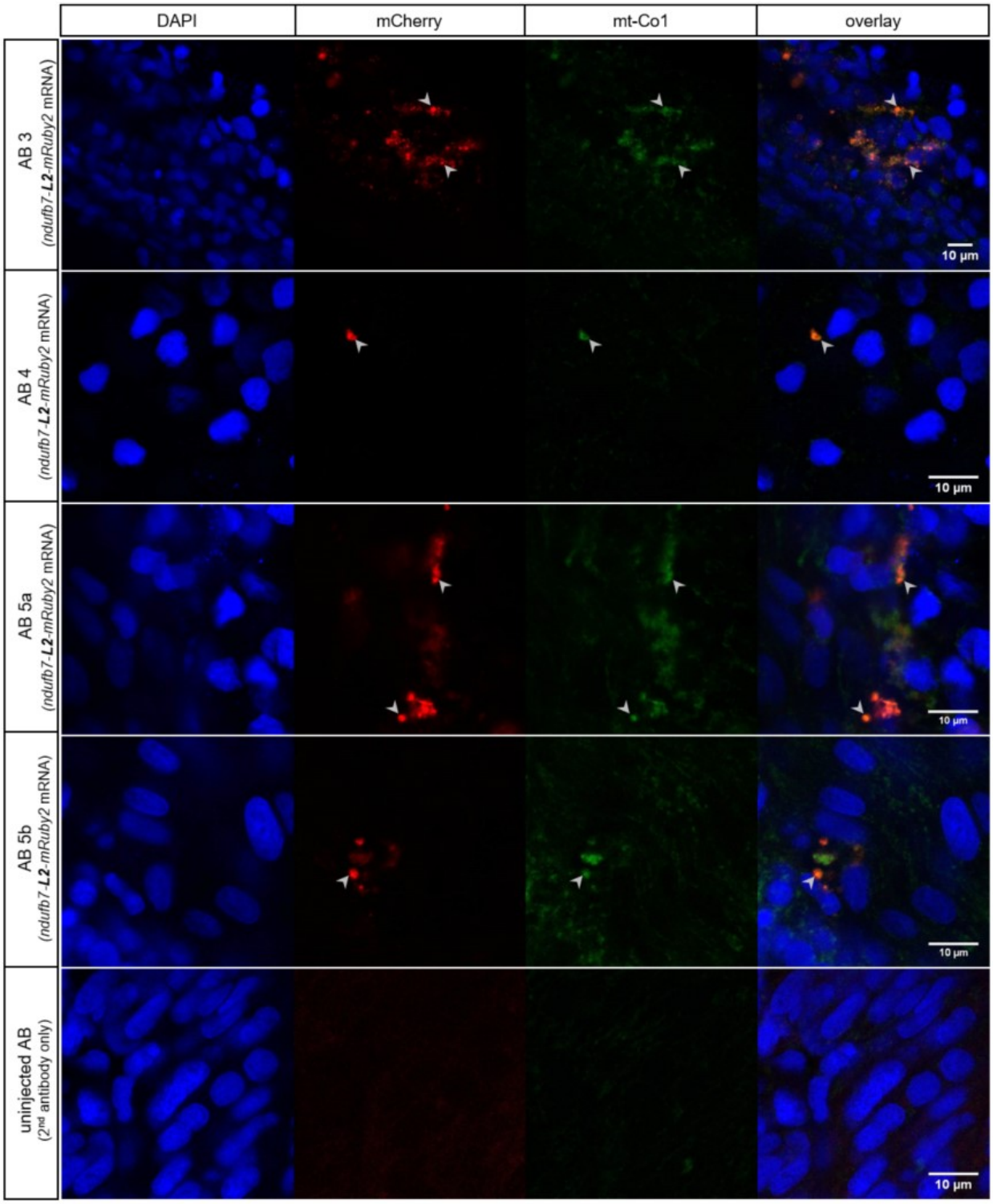
Zou, P. *et al.* (2015) ‘A conserved role of αA-crystallin in the development of the zebrafish embryonic lens’, *Experimental Eye Research*. doi: 10.1016/j.exer.2015.07.001.

Zwemer, L. M. *et al.* (2012) ‘Autosomal monoallelic expression in the mouse’, *Genome Biology*. doi: 10.1186/gb-2012-13-2-r10.

7 Supplementary Information



Suppl. Figure 1. Validation of mitochondrial staining via mt-Co1 antibody and secondary staining with Alexa Fluor 568, Alexa Fluor 647 or Biotin-Sp and Cy5 in *Tg(actc1b:tomm20-ZsGreen)* embryos. Imaged region: embryonic tail muscle. Blue: DAPI (nuclei), red/magenta: stained mt-Co1 (mitochondria), green: endogenous Tomm20-ZsGreen (unstained), scale bar: 10 µm.



Suppl. Figure 2. Validation of confocal imaging results for immunostaining of tagged Ndufb7 following *ndufb7-L2-mRuby2* mRNA overexpression in AB strain embryos at 1 dpf. Co-staining of mRuby2 with the mitochondrial marker mt-Co1 suggests mitochondrial localization of the tagged protein in the embryonic tail. Grey arrows point to examples for overlapping signals. Blue: DAPI (nuclei), red: anti-mCherry antibody (Ndufb7-L-mRuby2), green: anti-mt-Co1 antibody + Biotin-Cy5 (mitochondria). Scale bar: 10 µm.

Suppl. Information 1

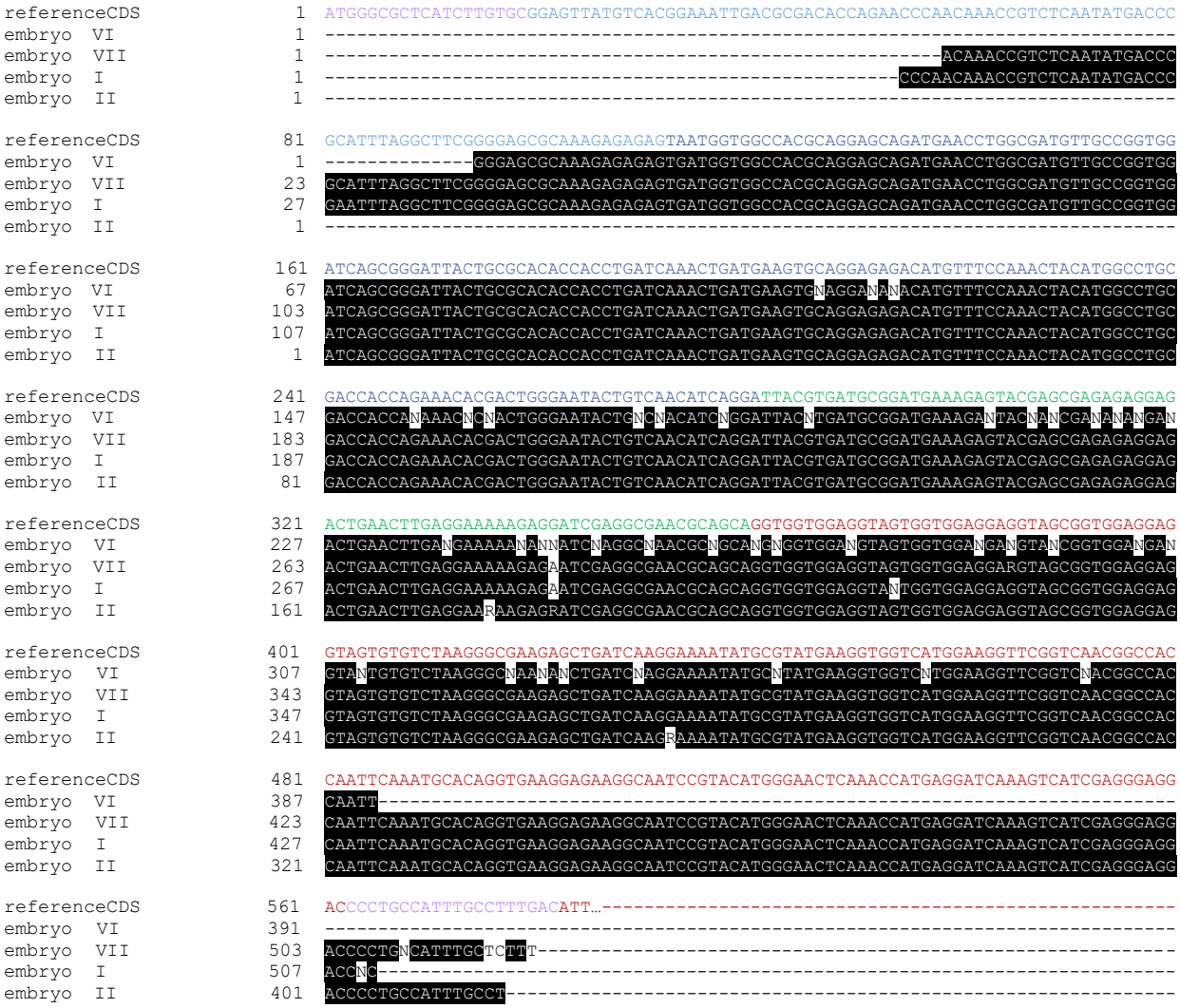
Sanger sequencing alignment: *L2-mRuby2* KI confirmation on gDNA level

Primer binding sites
part of repair template: *mRuby2*, STOP, 3' homology arm, sgRNA1 binding site
outside of repair template: 3' downstream gDNA sequence
correct alignment

Alignment tools: 1) M-Coffee (T-COFFEE, Version_11.00); 2) BOXSHADE (version 3.21)

Suppl. Information 2

Sanger sequencing alignment: *L2-mRuby2* KI confirmation on mRNA level (cDNA)



Primer binding sites
outside of repair template: *ndufb7* exon 1, *ndufb7* exon 2
part of repair template: *ndufb7* exon 3, *L2-mRuby2* insert
correct alignment

Alignment Tools: 1) M-Coffee (T-COFFEE, Version_11.00); 2) BOXSHADE (version 3.21)

8 Rights and Permissions

SPRINGER NATURE LICENSE

License Number:	5024970472428
License date:	Mar 09, 2021
Licensed Content Publisher:	Springer Nature
Licensed Content Publication:	Nature Structural & Molecular Biology
Licensed Content Title:	Clarifying the supercomplex: the higher-order organization of the mitochondrial electron transport chain
Licensed Content Author:	James A Letts et al
Licensed Content Date:	Oct 5, 2017
Type of Use:	Thesis/Dissertation
Requestor type:	academic/university or research institute
Format:	print and electronic
Portion:	figures/tables/illustrations
Number of figures/tables/illustrations:	1
High-res required:	no
Will you be translating?	no
Circulation/distribution:	1 - 29
Author of this Springer Nature content:	no
Title:	Development of a model system to study nuclear OXPHOS gene regulation in the zebrafish
Institution name:	Universität Wien
Expected presentation date:	Apr 2021
Portions:	Figure 1
Requestor Location:	M. F. Wien Austria
Total:	0.00 USD