Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

M.Sc Lila Oubraham

born in: Algiers, Algeria

Oral-examination: 11.05.2016

Nuclease-mediated gene manipulation of factors implicated in zebrafish neurogenesis

Referees: Prof. Dr. Uwe Strähle
                 Prof. Dr. Nicholas S. Foulkes

Summary

Complex and differential gene expression programs give rise to several cell types that constitute the different parts of the organism. This cell fate determination is controlled by a group of proteins, named transcription regulators (TRs). Our group investigates the molecular mechanisms underlying neurogenesis using the zebrafish as a model system. To that purpose, a genome-wide analysis of TR gene expression was performed in our laboratory, and hundreds of these regulators were identified. On the basis of these initial studies, a number of TRs were selected for further characterization. For this project, two model systems for zebrafish neurogenesis were chosen: The embryonic spinal cord and the adult telencephalon. The spinal cord is considered as relatively simple and is used to understand the neural differentiation and function in vertebrates during development. Based on morpholinos knockdown experiments, two closely related genes *sox1a* and *sox1b* encoding transcription factors, were shown to play a role in the specification of a newly observed sub-type of interneurons, named V2c, in the ventral spinal cord of the zebrafish. Nevertheless, the epistatic relationship between these genes has still to be investigated. On the other hand, in the adult brain, new neurons are continuously generated from neural progenitor cells. The zebrafish brain contains many more progenitor zones compared to mammals, reflecting its great capacity of neuronal proliferation and regeneration. Focusing on the telencephalic zones, because of its similarities with the mammalian brain, several factors were identified with a potential role in adult neurogenesis. Among those regulators, Id1 that acts as a dominant negative factor for basic helix-loop-helix (bHLH) transcription factors was chosen for further investigations because of promising previous observations using morpholino knockdown and protein overexpression methods. However, in order to perform loss-of-function studies and to deepen our understanding about the role of chosen TR genes and their molecular mechanisms controlling neurogenesis and adult brain regeneration, newly developed tools for gene manipulation were used: The transcription activator-like effector nucleases (TALENs) and the Type II clustered regulatory interspaced short palindromic repeats (CRISPR). Both methods use the nuclease activity to create double-stranded-breaks (DSBs) that are repaired via an error-prone pathway, non-homologous end-joining (NHEJ), creating mismatches in the DNA, or via a homologous recombination (HR) pathway allowing incorporation of a DNA sequence in a specific location of the genome.

After implementation of these two methods shortly after their development, I could show the efficiency of both in inducing mutations in the zebrafish genome, thus by using the gene *no-tail* (*ntl*) as a proof of principle. I then created with TALENs heritable mutations in the gene *id1* and assessed three generations of progeny in order to select a knockout line for the cited gene. Mutations in *id3* were also induced with both TALENS and CRISPR and other *id* genes were selected in order to continue the study concerning their possible redundant activities and their implication in adult neurogenesis and brain regeneration in the zebrafish. In parallel, I used the CRISPR system to simultaneously mutate the genes *sox1a* and *sox1b* in order to confirm their role in the specification of interneurons in the zebrafish spinal cord. Considering the long-lasting lack of site-specific mutagenesis in zebrafish and the controversies in the last few years about the specificity of the morpholino antisense technology, the new genetic tools characterized and applied in this PhD work open the door to a wide range of investigations.

Zusammenfassung

Alle Zellen eines Organismus enthalten die gleichen Erbanlagen. Die DNA und die darin enthaltenen Gene sind folglich in allen Zellen identisch. Erst die differenzierte Expression dieser Gene führt zur Bildung von unterschiedlichen Zelltypen, die die Grundlagen für die verschiedenen Gewebe eines Organismus darstellen. Diese komplexe Expressionskontrolle beruht auf einer Gruppe von Proteinen die man Transkriptionsregulatoren (TR) nennt. Um die Rolle von verschiedenen TR in der Neurogenese des Zebrafisches zu untersuchen, wurden für die vorliegende Arbeit zwei Modelsysteme herangezogen: das embryonale Rückenmark und das adulte Telencephalon. Das Rückenmark wird als relative einfache Struktur angesehen und wurde verwendet um die neuronale Differenzierung in Vertebraten während der embryonalen Entwicklung verstehen zu lernen. Für zwei eng verwandte Gene, sox1a und sox1b, die für Transkriptionsfaktoren kodieren, konnte anhand von Morpholino Experimenten gezeigt werden, dass sie eine Rolle bei der Spezifizierung eines neuen interneuronalen Subtypus, V2c, spielen, der im ventralen Rückenmark des Zebrafisches lokalisiert ist. Jedoch muss die epistatische Beziehung der beiden Gene untereinander noch untersucht werden. Im adulten Gehirn werden Neuronen kontinuierlich aus neuronalen Stammzellen gebildet. Das Gehirn des Zebrafisches enthält, verglichen mit dem von Säugetieren, erheblich mehr proliferierende Zonen, was sein enormes Vermögen in der neuronalen Proliferation und Regeneration aufzeigt. Da das Gehirn des Zebrafisches Ähnlichkeiten mit dem Säugergehirn aufweist, hat unsere Arbeitsgruppe den Blick auf Regionen des Telencephalons gerichtet und mehrere Faktoren isoliert, die möglicherweise eine Rolle in der Neurogenese spielen. Unter diesen Faktoren befindet sich das auf Transkriptionsfaktoren dominant-negativ agierende *id1*. Da *id1* spezifisch in der ventrikulären Zone des Telencephalons exprimiert wird (eine proliferierende Zone im Gehirn des Zebrafisches) und die Tatsache das *id1* im verletzten Gehirn hochreguliert wird, wurde der Faktor für eine eingehendere Untersuchung ausgewählt. Um ein tiefergehendes Verständnis zur Funktion und den molekularen Mechanismen welche die Regeneration im adulten Gehirn steuern, wurden *loss-of-function* Studien durchgeführt, die auf neu entwickelten Werkzeugen zur Genmanipulation basieren: *transcription activator-like effector nucleases* (TALENs) und *Type II clustered regulatory interspaced short palindromic repeats* (CRISPR). Beide Methoden verwenden die Aktivität einer Nuklease um Doppelstrangbrüche zu erzeugen, die entweder durch das fehleranfällige *non-homologous end-joining* (NHEJ) repariert werden, bei dem es zu Fehlpaarungen kommen kann, oder durch

homologe Rekombination (HR), welche den Einbau von DNA Sequenzen an spezifischen Stellen im Genom erlaubt.

Nach der Etablierung dieser zwei Methoden schon kurz nach deren Entwicklung, konnte ich am Beispiel des *notail* (*ntl*) Lokus die Effizienz beider Techniken bei der Induktion von Mutationen im Zebrafischgenom demonstrieren. Im Anschluss habe ich in die Keimbahn reichende Mutationen im id1 Gen eingefügt und die Nachkommen der drei folgenden Generationen nach einem erfolgreichen *knock-out* von id1 selektioniert. Im Folgenden wurde, um die Rolle weiterer Mitglieder aus der id-Familie zu untersuchen und deren mögliche Redundanz zu ergründen, Mutationen in id3 mit Hilfe von TALENs und CRISPR eingefügt. Parallel dazu verwand ich das CRISPR System um die Gene *sox1a* und *sox1b* zeitgleich zu mutieren. Damit sollte ihre Rolle bei der Spezifikation von Interneuronen im Rückenmark des Zebrafisches bestätigt werden, die die vorläufigen Morpholino Experimente angedeutet hatten. Bedenkt man, dass es lange Zeit nicht möglich war gerichtete Mutagenese im Zebrafisch durchzuführen und die Kontroversen der letzten Jahre, ob der Spezifität der Morpholino antisense Technologie, öffnen die in dieser Arbeit charakterisierten und angewandten Werkzeuge die Tür für eine große Bandbreite an wissenschaftlichen Untersuchungen.

ACKNOWLEDGEMENTS

Table of contents

ABREVIATIONS

aa   amino acids

AD   activation domain

bHLH   basis helix-loop-helix

BLBP   brain lipid-binding protein

bp   base pair

Cas   CRISPR-associated

CMV   cauliflower mosaic virus

CNS   central nervous system

CoDA   context-dependant assembly

CRISPR   clustered regulatory interspaced short palindromic repeats

crRNA   CRISPR-RNA

DG   dentate gyrus

DNA   deoxyribonucleic acid

Dpf   days post-fertilization

Dpl   days post-lesion

DSB   double-strand break

dsDNA   double-stranded DNA

DV   dorsal-ventral

ENU   N-ethyl-N-nitrosourea

ER   estrogen receptor

FLASH   fast ligation-based automatable solid-phase high-throughput

FP   floor plate

GABA   gamma-aminobutyric acid

GFAP   glial acidic fibrillar protein

GFP   green fluorescent protein

gRNA	guide RNA
HD	homeodomain
HDR	homology-directed repair
HLH	helix-loop-helix
HMG	high mobility group
Hpf	hours post-fertilization
HR	homologous recombination
ICA	iterative cap assembly
Id1	DNA-binding protein inhibitor 1
KA	Kolmer-Agduhr
Kb	kilo-base pair
LFP	lateral floor plate
LIC	ligation-independent cloning
MA	modular assembly
MFP	medial floor plate
MO	morpholino oligonucleotide
NC	notochord
NHEJ	non-homologous-end-joining
NLS	nuclear localisation signal
NPC	neural progenitor cell
nt	nucleotides
Ntl	no-tail
NSC	neural stem cell
OPEN	oligomerized pool engineering
PAM	protospacer adjacent motif
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction

polyA	polyadenylation
REAL	restriction enzyme and ligation
RMS	rostral migratory stream
RNA	ribonucleic acid
RNAseq	RNA sequencing
RVD	repeat variable di-residue
S100β	S100 calcium binding protein β
sgRNA	single guide RNA
SGZ	subgranular zone
Shh	sonic hedgehog
ssDNA	single stranded DNA
SVZ	subventricular zone
TALE	transcription activator-like effector
TALEN	transcription activator-like effector nuclease
TF	transcription factor
TILLING	targeting induced local lesions in genomes
TR	transcription regulator
tracrRNA	transactivating CRISPR-RNA
wt	wild type
ZFN	zinc finger nuclease

Chapter I: INTRODUCTION

Evolving from a single cell, corresponding to the fertilized egg, to complex multicellular organisms represented at all times an intriguing part of life. Clearly, several processes that lead to the development of the complex structures of the adult body, called embryogenesis, involve a plethora of genes. First of all, the egg undergoes a series of cell divisions, known as cleavage. Then the pattern formation process gives to the cells an organized activity that can define the body plan (e.g. antero-posterior axis) and the germ layers (ectoderm, mesoderm and endoderm) with distinct features that will give rise to different tissues. The third developmental process, or morphogenesis allows the embryo to change form, mostly during gastrulation (outer embryonic cells are internalized). Finally, cell differentiation is characterized by the gradual formation of distinct cell types (Wolpert et al. 2006).

Vertebrate embryonic development depends broadly on diverse proteins differently expressed in distinct types of cells and at different time points. This differential expression of genes explains the emergence of distinct cell types with the same initial genomic information (Gilbert 2010). Particularly during development, the transcriptional regulatory machinery is of great importance. Indeed, the key actors of cell fate determination are transcriptional regulators (TRs) which activate or repress the transcription of specific genes, thus controlling gene expression in a spatial and temporal manner. Among those regulators, transcription factors (TFs) act by binding to DNA at specific regions known as gene regulatory elements, such as promoters, enhancers or repressors. Specific combinations of TR activities lead to the generation of different cell types with distinct functions (Davidson 2010). Moreover, transcription is also affected by chromatin structure as it controls the accessibility of the DNA binding factors to their specific regions in the genome. Chromatin modifiers are key drivers of this process and act in line with the TRs to regulate gene expression (Alberts et al. 2013).

Although most general principles of development apply to all live multicellular organisms, their diversity makes biological investigations highly complex. Therefore, focus was placed on a small number of organisms with easy manipulation and ability to use in diverse experimental ways: the model organisms. Among them zebrafish, or *Danio rerio*, is a relatively small freshwater fish that is native to rivers in India. This teleost fish can be easily maintained in aquariums in large numbers in laboratories. The potential of zebrafish as a research tool was highlighted in the 1980 by George Streisinger and colleagues (Streisinger *et*

al, 1981). A vertebrate organism with a number of advantages as an experimental system: a large number of offsprings (200-300 eggs per mating) with a short generation time (3-4 months to get adult fish). An external fertilization and a transparent embryo allowing the accessibility of all developmental stages (which some are showed in Figure 1). A fully sequenced genome and the possibility to create transgenic and homozygous lines. In addition to that, the zebrafish is a vertebrate like humans and is, in opposition to mouse or rat, easier and cheaper to handle. During more than 20 years, the zebrafish confirmed its significant value as a model for human development and diseases.
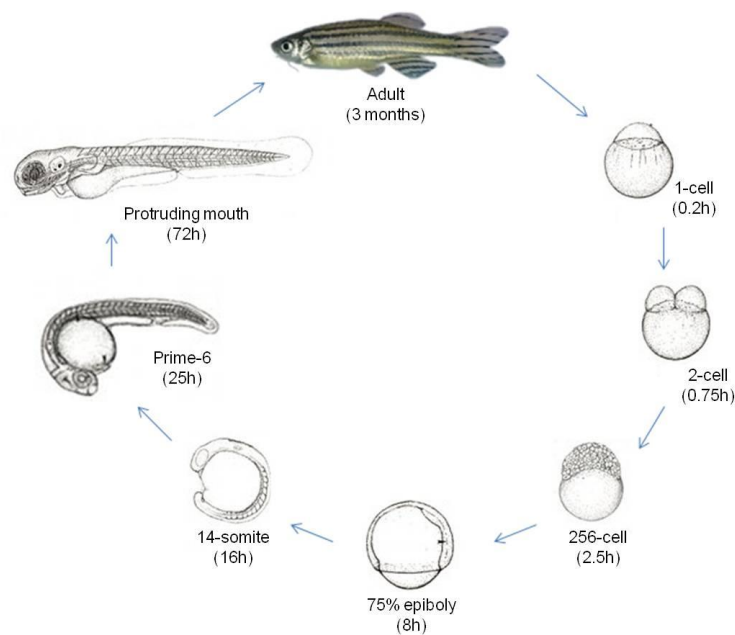
Figure 1: A schematic representation of the fast development of the zebrafish. Adult zebrafish of 3 months old lay between 200 and 300 eggs. The zygote period (1-cell) precedes the first cleavage that occurs at 0.75 hours (2-cells). The blastula stage is represented with a 256-cell at 2.5 hours and the gastrula at 8 hours with a 75% epiboly embryo. During segmentation, the somites are formed (e.g. 14-somite stage at 16 hours). At 25 hours, the heart is already beating and the larval stage starts after 72 hours with a protruding mouth. During the larval period, the fish starts swimming and feeding then reaches reproductive maturity in 3 months.

1.1. Neurogenesis in zebrafish

One of the most complex organ systems is the nervous system, which contain multiple cell types forming a large neural network. The process leading to the generation of mature and functional neurons from undifferentiated neural progenitors is named neurogenesis (Detrich et al. 2010). In neurogenesis, different steps allow the formation and the maintenance of the central nervous system (CNS): induction of neural progenitors, amplification of the progenitor

pool by division, specification and determination of cell fate and finally differentiation to post-mitotic neurons (Wilson & Edlund 2001). Each of these processes is meticulously orchestrated in a spatial and temporal manner, by specific signaling pathways and by the expression of different TFs, in order to generate the multiple cell types forming the CNS (Bally-Cuif & Hammerschmidt 2003).

In the zebrafish embryo, the formation of the neural plate at the dorsal part during gastrulation represents the first step of neural induction, as the cells of the ectoderm are specified to form the neuroectoderm containing neural precursor cells. It has been shown in other organisms (frog and chick) that neural induction is triggered by different extrinsic signaling factors: the bone morphogenic proteins (BMPs), wingless-integrated (Wnt) and fibroblast growth factor (Fgf), plus intrinsic transcription factors, specifically by members of the SoxB1 family (Wilson & Edlund 2001; Rogers et al. 2009).

The neural plate represents the basic outline for the nervous system with rough coordinates and subdivisions. Its folding during neurulation gives rise to the neural tube that will form the brain anteriorly and the spinal cord posteriorly (Gilbert, 2010). After neural tube formation, neuroepithelial cells transform into radial glial cells, which are cells with a cell body in the ventricular zone and long radial fibers extending through the neural tube. Radial glial cells divide both symmetrically and asymmetrically, in this last case giving rise to one differentiating neuron and one radial glial cell, thus maintaining the progenitor pool. The Notch signaling pathway was suggested to play a role during this process as high Notch activity was observed in the self-renewing daughter cells (Schmidt et al. 2013a).

At late gastrulation, another process allows the formation of new neurons to occur with the maintenance of the progenitor pool. Indeed, at this stage, "proneural genes" are expressed in progenitors of the "proneural cluster" which can initiate neurogenesis (Appel & Chitnis 2002). However, a limited number of the progenitors undergo the exit of the cell cycle and differentiation into neurons (Figure 2). Thus, a selection process called "lateral inhibition" occurs in which one of the cells of the proneural cluster expressing high level of proneural genes will form a neuroblast, while its neighbor-cell down regulates the expression of proneural genes and remains as a progenitor (Wolpert et al. 2006). The process of lateral inhibition is initiated by the induction of a Notch ligand and its expression in the future progenitors, plus the activation of the Notch signaling pathway in the neighboring cells which

lead to the expression of repressors that downregulate proneural gene expression (*hes* genes) (Bertrand et al. 2002).
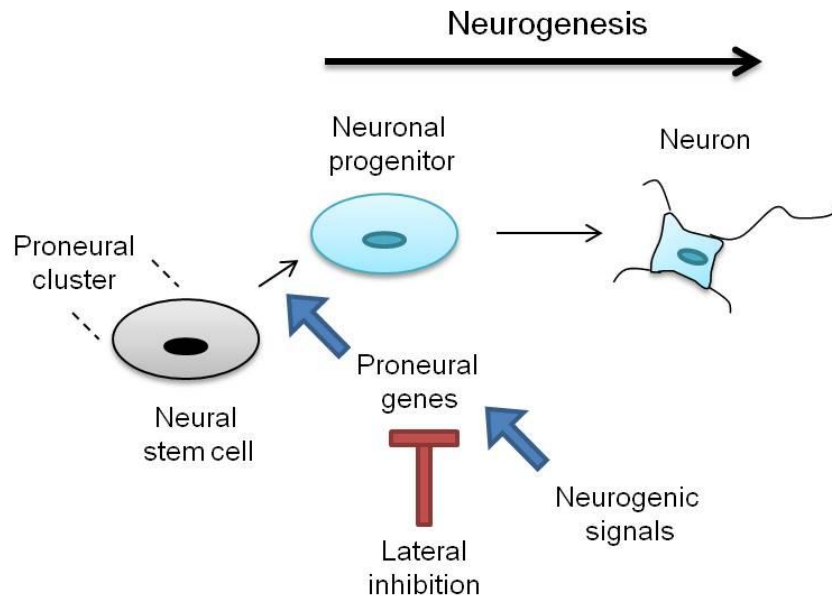
Figure 2: Regulation of neurogenesis in early development. The expression of proneural genes is induced by neurogenic signals and the progenitors accumulating high levels initiate differentiation into neurons. Meantime, proneural genes are down-regulated in other neiboughing cells through lateral inhibition, thus staying in a progenitor state. Adapted from (Bertrand et al. 2002).

The first expressed proneural genes in zebrafish encode transcription regulators (e.g. the bHLH transcription factor neurogenin1). The sequential activation of TFs, in a precise spatial and temporal manner, leads to the generation of distinct types of neurons and glia from neural progenitors. The induction of the neurogenic cascade and the neural specification is under the control of different extracellular signaling pathways that regulate the activity of these diverse TFs (Guillemot 2007). Thus, during neurogenesis, the expressed regulators can be used as markers to distinguish between the different cell-types that coexist in the CNS.

Our group undertook a systematic analysis of TR gene expression in zebrafish in order to elucidate the transcriptional regulatory networks during the development of the embryonic CNS and in the adult telencephalon. First by microarray analysis, a number of TRs are profiled at different stages of embryonic development and in some adult body parts. Then, a genome-wide analysis of gene expression states were compiled by RNA sequencing (RNAseq). 3302 TR genes were detected in the zebrafish genome with possible affiliation to the transcriptional machinery. It was observed that most of these TR genes were expressed at

24 hours post-fertilization (hpf) as it represents a critical time for neurogenesis and organogenesis. By *in situ* hybridization, an atlas of gene expression for 1711 genes at 24 hpf was provided (Armant et al. 2013).

1.1.1. Neurogenesis in the zebrafish spinal cord

In order to understand the neural specification and function in vertebrates during development, the spinal cord represents a relatively simple model (Figure 3A). Combinations of TFs are activated by signaling pathways in a spatial and temporal manner, defining different progenitor domains along the dorsal-ventral (DV) axis of the neural tube. Each domain generates one or more neuronal subtypes (Dessaud et al. 2008).

Actually, distinct cell types populate defined domains along the DV axis of the neural tube. Floor plate cells are formed in the ventral midline of the spinal cord with three stripes of cells: one medial floor plate (MFP) flanked by cells of lateral floor plate (LFP) on both sides (Strähle et al. 2004). Several experiments suggested that the main signal inducing specification of the LFP in zebrafish is the secreted Sonic hedgehog (Shh) protein. A gradient of this morphogen, released by the notochord and the floor plate, controls the specification of motorneurons and different subtypes of interneurons in the ventral spinal cord (Appel & Eisen 1998; Jessell 2000). A combination of TFs from the homeodomain (HD) and basis helix-loop-helix (bHLH) families are regulated by the Shh signaling pathway and thus define distinct progenitor domains that will give rise to different neuronal subtypes (Dessaud et al. 2008) (Figure 3B).

The LFP progenitors generate V3 and Kolmer-Agduhr" (KA") interneurons under the control of the homeobox TFs Nkx2.9, Nkx2.2a and Nkx2.2b. Gata2 and Tal2 are two TFs that regulate the specification of the KA". More dorsally, the motoneuron progenitors expressing *olig2* give rise to KA' interneurons, whereas the V2 domain produces V2a and V2b interneurons (Yang et al. 2010).

The p2 progenitor domain from which V2 interneurons are specified is known in mouse to also generate a third subtype, V2c interneurons (Panayi et al. 2010). In zebrafish, the p2 domain indeed gives rise to V2a interneurons, characterized by *vsx1* and *vsx2* expression, and V2b interneurones, characterized by the expression of *gata2*, *gata3*, *tal1* and *tal2*. However, no clear evidence showed the existence of V2c interneurons in zebrafish. Only few years ago,

a study performed in our laboratory by Yang *et al*. suggested the presence of a new cell subtype in the spinal cord and was referred to as V2c interneurons (Yang et al., *unpublished data*). Experiments showed that the specification of V2c interneurons in zebrafish was under control of *sox1a* and *sox1b* genes, which are detected in the systematic expression screen of TRs in a 24 hpf zebrafish embryo and expressed in distinct cells of the ventral spinal cord (Armant et al., 2013).
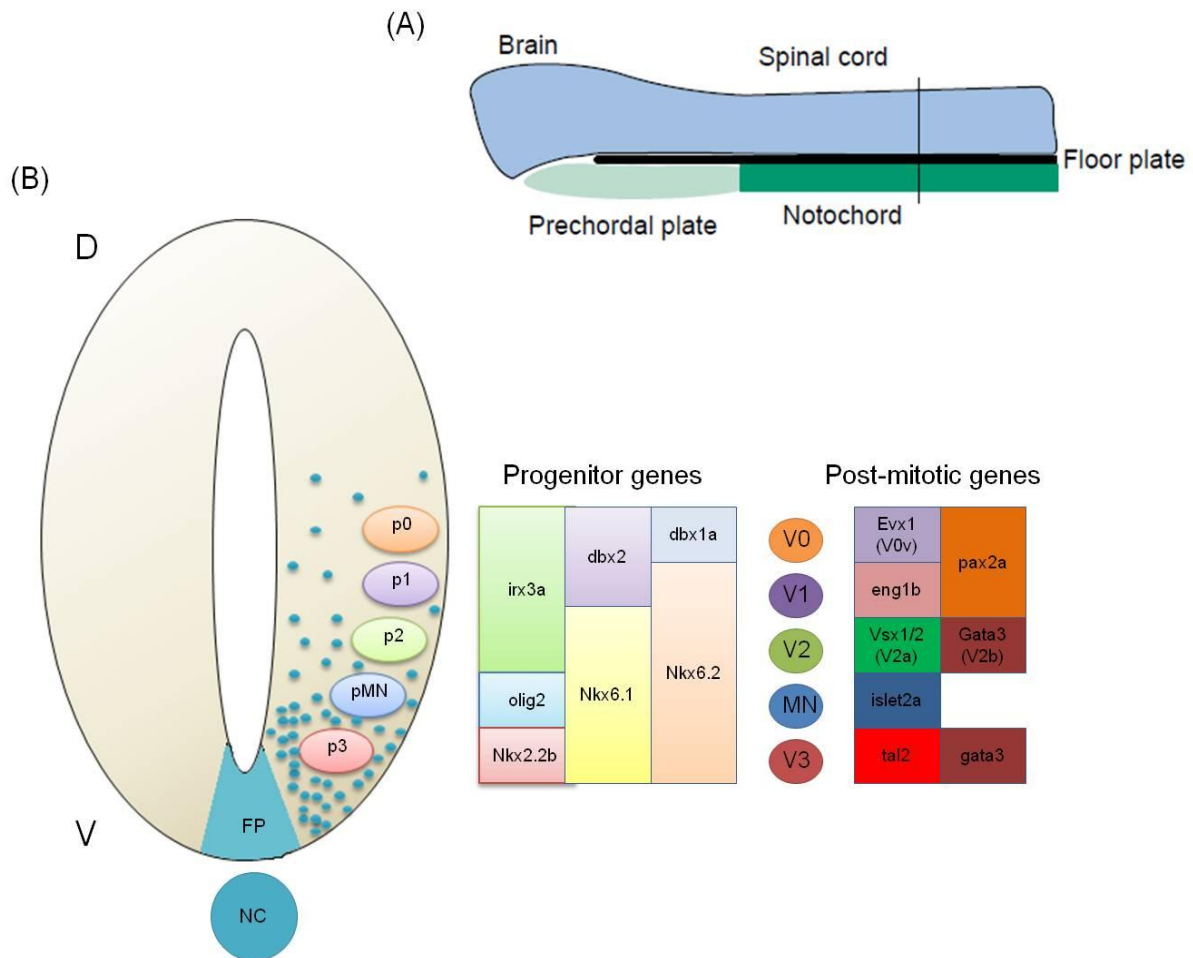
Figure 3: Organization of the spinal cord. (A) Relative position of the brain, spinal cord, notochord and floor plate. (B) A schematic representation of a cross section through the ventral spinal cord. Shh molecules (blue dots) are released from the floor plate (FP) and the notochord (NC) and spread in a gradient manner. The different progenitor domains (p3 to p0) express a combination of TF genes regulated by Shh signaling pathway. Distinct neuronal subtypes (V3 to V0) are generated from the progenitor domains and express a set of post-mitotic genes that can be used to characterize them. Adapted from Rowitch 2004 and England et al. 2011.

Sox1 a and *b* are closely related genes that encode TFs, belonging to the SRY-box containing genes B1 (SoxB1) family. They contain a homologous sequence of a DNA-binding domain: the HMG-box. Both genes are expressed in the GABAergic KA (KA' and KA") and in V2 interneurons in the spinal cord. Functional analysis using morpholino knock-down were

previously performed in order to down-regulate the *sox1* genes. However, the high shared identity of the two genes made it impossible to target *sox1 a* and *sox1b* individually (Yang et al., *unpublished data*).

It was shown that the double knockdown of *sox1a/b* did not affect the differentiation of the KA interneurons subtypes. On the other hand, the knockdown of both genes seems to increase the number of cells expressing *tal2*, *gata2*, *gata3* and *gad67*, which are markers of V2b neurons (Batista et al. 2008), suggesting the increase of this subtype. Nevertheless, the expression of V2a neuronal markers did not change (*vsx1/2*). Taking into account that *sox1a/b* are only partially co-expressed with the markers of V2b subtype, it was suggested that a proportion of the *sox1a/b* expressing cells might differ from V2b and characterize a new subtype of V2 interneurons: V2c (Yang et al., *unpublished data*).

The study of Yang and colleagues also showed a decrease in the newly identified V2c neurons when Sox1a/b protein translation was down-regulated. Indeed, a clear reduction of cells expressing *sox1a* or *sox1b* mRNA was observed (*sox1+/gad67-* or *sox1b+/gad67-*), while V2b cells expressing both mRNAs (*sox1a+/gad67+* or *sox1b+/gad67+*) increase. These observations suggest that V2b and V2c interneurons derive from the same precursor, and most likely, *sox1a/b* have a role in the specification of V2c cells.

1.1.2. Neurogenesis in the zebrafish adult brain

For a long time, the nervous system was considered as incapable of generating new neurons, until Joseph Altman suggested the possibility of adult neurogenesis in the mammalian brain, more than 50 years ago (Altman 1962). Further studies established that newborn neurons are continuously generated in the adult brain from constitutively active progenitor regions. In mammals, only two main zones of the telencephalon are known for their neurogenic activity: the subventricular zone (SVZ) lining the lateral ventricle, and the subgranular zone (SGZ) of the dentate gyrus (DG) in the hippocampus (Reynolds & Weiss, 1992). Unlike mammals, the adult brain of teleost fish has a widespread cell proliferation in all subdivisions along the rostrocaudal axis with high regenerative capacity. Presently, 16 regions were identified in the entire brain of zebrafish with proliferative activity (Grandel et al. 2006; Kaslin et al. 2008). Among these areas, the progenitor zones in the telencephalon of the zebrafish are highly interesting because of similar aspects in the mammalian brain, making it an excellent model system for comparative studies.

Proliferation zones have the particularity to harbor stem cell niches formed by different neural stem cells that give rise to diverse cell types (Kizil et al. 2011; Merkle & Alvarez-Buylla 2006). The progenitor cells populating the ventricular zone (VZ) of the zebrafish telencephalon express distinct sets of markers that can be used to distinguish and classify the different subtypes (März et al. 2010). In the adult zebrafish telencephalon, radial glial cells have a critical role in generating new neurons. In contrast, radial glial cells in mammals have a role in the embryonic development of the telencephalon but then lose their properties and turn into astrocytes (Kizil et al. 2012).

We will focus in this part of the study on progenitors populating the VZ of the adult zebrafish telencephalon, according to the classification that was suggested by März et al. and depicted in figure 4. Type I cells are non-dividing radial glia, mainly quiescent, are present in the ventral part of the VZ and express glial markers such as Nestin, S100ß, the glial acidic fibrillar protein (GFAP) and the brain lipid-binding protein (BLBP). Type II cells represent slowly-dividing cells that express, additionally to the previous markers, the proliferating cell nuclear antigen (PCNA). These cells constitute the majority of proliferating cells in the medial region of the VZ. Type II cells give rise to type III cells (neuroblasts). Type II cells can be separated into two groups: Type IIIa cells that express the neuronal and oligodendrocyte progenitor marker PSA-NCAM, but still co-express PCNA, Nestin and other glial markers, and the type IIIb cells which lack the expression of Nestin and all glial markers. All described types of progenitors also express Sox2, which is a stem cell marker (März et al. 2010; Schmidt et al. 2013b).

The type II stem cells have the ability to divide symmetrically and asymmetrically to self-renew and to generate type II cells. The balance between quiescent and proliferating cells has to be controlled in the adult brain in order to maintain the stem cell pool. The Notch signaling pathways seems to be implicated in this process, similarly to lateral inhibition in the embryo. Indeed, the activation of the Notch pathway drives the stem cells into quiescence, while blocking Notch restores stem cell division and formation of neurons (Chapouton et al. 2010; Rothenaigner et al. 2011).
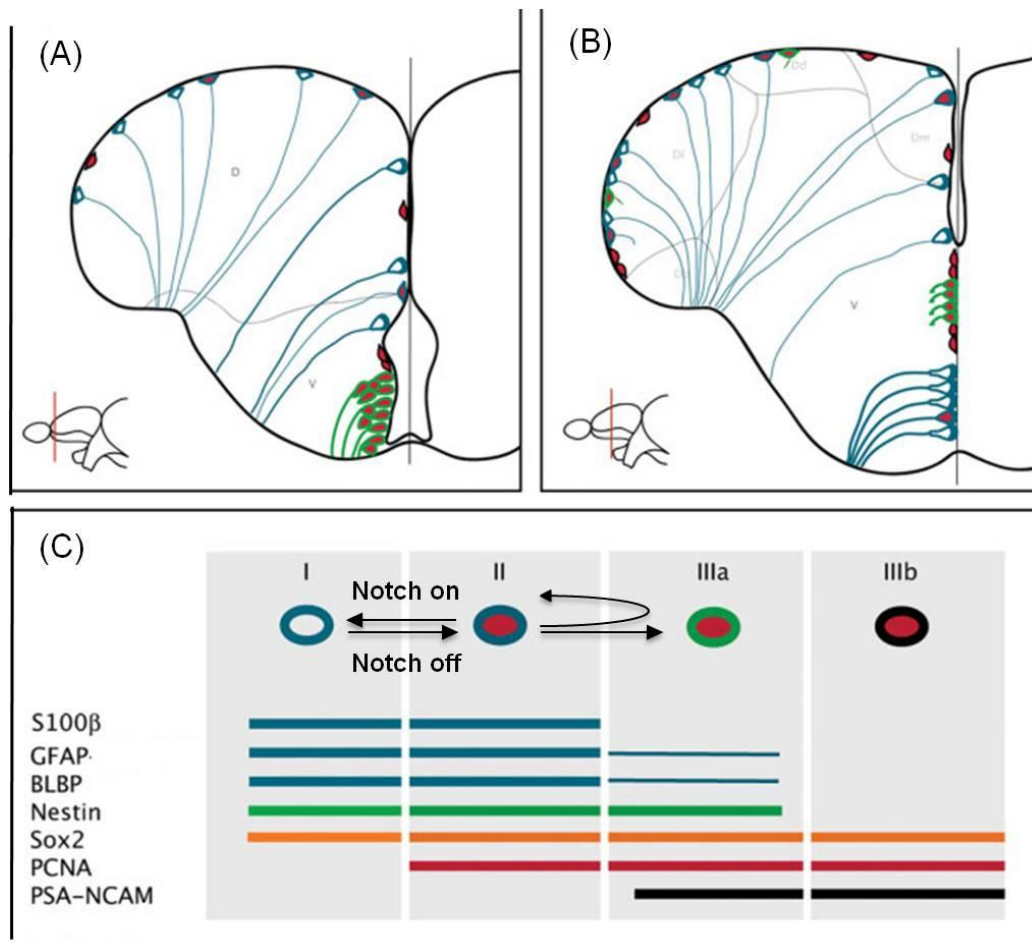
Figure 4: Distinct cell types at the ventricular zone of the adult zebrafish telencephalon with different sets of expressed markers. (A) and (B) represent a schematization of the cells on cross-sections at anterior and medial levels of the telencephalon, respectively. (C) Expression of the different radial glia, progenitor and neuroblast markers. Type I and II cells show strong expression of radial glia markers, while in Type III cells these markers are low or absent. The regulation of the balance between quiescent and proliferating cells by the Notch signaling pathway is represented. Modified from (März et al. 2010).

1.2. Regeneration of the zebrafish adult brain

The ability to replace lost neurons, due to injury or degenerative diseases, by newly born ones has always been of a great interest. However, the ability of neuronal regeneration is highly related to adult neurogenesis potential. As mentioned above, this ability in mammals is very limited compared to the one of zebrafish. Indeed, studies in teleost fish demonstrated an extensive regeneration potential in different organs (e.g. spinal cord, retina) and also in the CNS (Becker & Becker 2008).

In order to understand the regulatory network controlling the regenerative capacity of the zebrafish brain, a screen to identify TR genes that play a role in this process was performed in

our laboratory. First of all, a stab-injury assay was established (März et al. 2011). A wound is generated by inserting a needle through the skull into one telencephalic hemisphere, while the opposite hemisphere is kept intact for control. After 5 days post-lesion (dpl), as the maximal proliferative time (high expression of PCNA), the brains were prepared and in-depth RNAseq was used to assess the expression of genes in both injured and control hemispheres. Focusing on misregulated genes with TR features (InterPro protein domain prediction), 279 regulatory genes were identified. The majority of these genes are up-regulated after injury and the most represented signaling pathways are immune response, apoptosis and Wnt (Viales et al. 2015).

Additionally, using *in situ* hybridization, TR genes expressed specifically at the VZ of the telencephalon were investigated and 60 of them were shown to be up-regulated upon injury, thus suggesting an implication in the regulation of neuronal regeneration. Among them, *id1*, one of the id genes supposed to have a role in maintaining the neural stem cell pool in the mouse adult brain (Niola et al. 2012; Jung et al. 2010).

The DNA-binding protein inhibitor 1, Id1, belongs to a sub-family of helix-loop-helix (HLH) proteins that lack the basic DNA binding domain. Thus, by interacting with its partners, members of the basic HLH (bHLH) family of transcription factors, it inhibits their DNA binding and transcriptional regulation abilities (Norton, 2000). Thus, they act as dominant negative regulators for bHLH TFs and play a role in cell proliferation and differentiation during development and at adult stage (Wong et al. 2004; Norton 2000; Lasorella et al. 2014). For instance, Id proteins bind to the HLH domain of the ubiquitous E proteins (class of bHLH TFs), preventing the formation of heterodimers with tissue-specific bHLH factors and by that blocking formation of functional transcription complexes. Since it was shown that TFs from the bHLH family regulate cell fate determination and differentiation, we can consider that Id proteins can control these processes in a variety of tissues (Ruzinova & Benezra 2003; Jung et al. 2010).

The previous study of Viales et al. showed that *Id1* is expressed in all the ventricular domains of the zebrafish telencephalon, more abundantly in the dorsomedial and the dorsolateral VZ of the pallium, except for the rostral migratory stream (RMS) (Figure 5A). The expression of the gene is highly up-regulated at 5 dpl (Figure 5B) then decrease gradually at 7 dpl. Immunohistochemistry experiments using the stem cell markers S100ß and PCNA indicated that *id1* is predominantly expressed in quiescent type I cells. Further gain and loss-of-function experiments for *id1* were performed. Actually, lipofection of an *id1* expression construct in

the ventricle suggested a role of this regulator in conferring quiescence to radial glial cells (lower proportion of PCNA+ proliferating cells). Whereas, the knock-down of *id1* by vivo-morpholino showed an increase in proliferating cells (PCNA+ cells). The obtained results of this previous study suggested a specific role of *id1* in sustaining the balance between quiescence and neurogenesis and maintaining a neural stem cell pool in the adult brain.
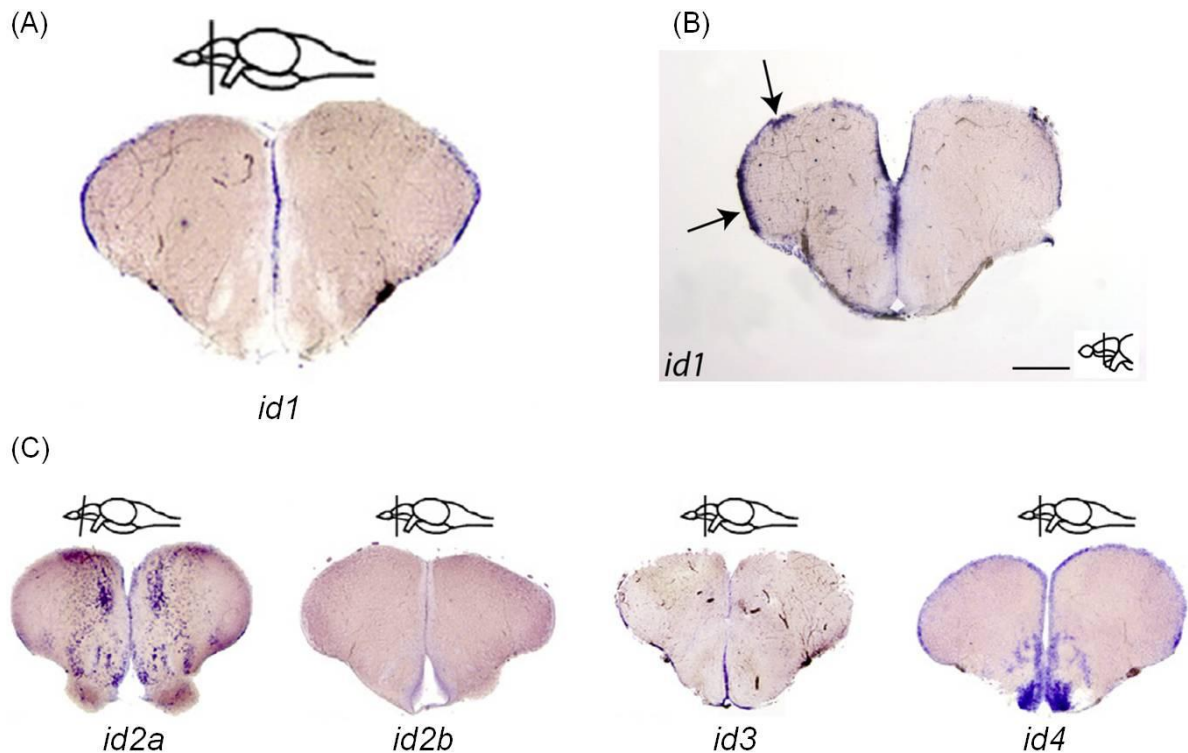
Figure 5: *In situ* hybridization on sections of the adult zebrafish telencephalon. (A) *id1* expression in the ventricular zone of the adult zebrafish telencephalon. (B) id1 expression shows an upregulation (arrows) 5 days after brain injury (left hemisphere) compared to the uninjured control (right hemisphere). Adapted from Viales et al. 2015. (C) Expression of the members of the Id family in the adult zebrafish telencephalon. Id2a, id3 and id4 are expressed in different part of the adut brain, while id2b is not detected. Adapted from Diotel et al. 2015.

However, the molecular mechanism behind *id1* function has still to be clarified in the zebrafish telencephalon. Moreover, other members of the Id protein family can be involved. In the mouse, it was shown that a conditional triple knockout of *id1*, *id2* and *id3* lead to the loss of stemness of NSCs that undergo premature differentiation and to their detachment of the stem cell niches. Thus, suggesting a role of Id proteins in synchronizing stemness and anchorage of NSCs (Niola et al. 2012). In zebrafish, four other *id* genes have been identified in addition to *id1*: *id2a*, *id2b*, *id3* and *id4*. It was already shown in the zebrafish embryo that the *id* genes have partially overlapping expression patterns. However, the expression of these

genes in the adult zebrafish telencephalon was just recently documented. Indeed, our colleagues Diotel et al. showed, by *in situ* hybridization, a wide expression of the *id* genes in the entire adult brain of zebrafish, except for *id2b* that was barely detectable (Figure 5C) (Diotel, Beil, et al. 2015). Whether other members of the Id family contribute to the regulation of neurogenesis in the adult brain has still to be clarified.

1.3. Reverse genetic technologies in zebrafish

Using forward genetics, meaning to go from the phenotype to the gene, large scale mutational screens were performed in zebrafish and provided a lot of understanding of vertebrate development (Tübingen and Boston screens, Nüsslein-Volhard 2012). On the other hand, reverse genetic tools would provide mutations in any genes of interest to evaluate its functions, but they were for a long time unavailable for zebrafish studies. This state has changed already with the development of strategies for loss-of-function studies, such as mutagenesis screens (e.g. TILLING), or the antisense morpholino oligonucleotides. Additionally, in the last few years, engineered endonucleases have emerged and enable to achieve directed genome modifications in zebrafish (Figure 6) (Huang et al. 2012).
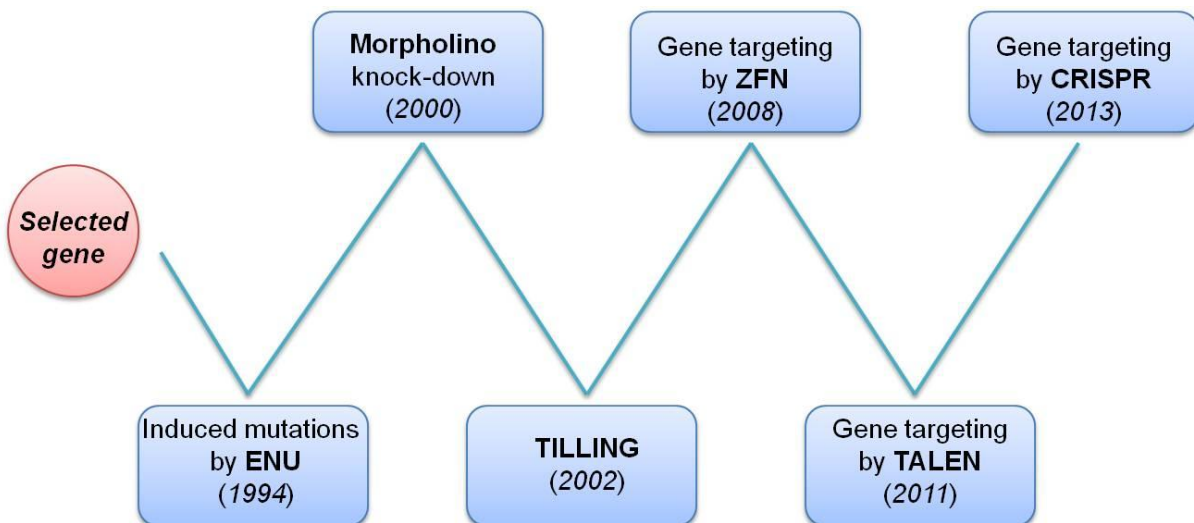
Figure 6: Chronoligical evolution of reverse genetic approaches used in zebrafish starting from a terget gene. ENU, N-ethyl-N-nitrosourea; TILLING, Targeting Induced Local Lesions IN Genomes; ZFN, zinc finger nuclease; TALEN, transcription activator-like effector nuclease; CRISPR, clustered regularly interspaced short palindromic repeats.

In reverse genetics, the first main goal is to select a specific gene in order to modify it or its expression and to investigate its function. The used approaches for reverse genetics can be non-targeted (e.g. TILLING) or targeted (e.g. site-specific nucleases) (Tierney & Lamour

2005). With the fully sequenced genome of zebrafish, many novel and functional genes were identified by bioinformatics analysis or through screening by *in situ* hybridization for tissue-specific gene expression (Huang et al. 2012).

1.3.1. Morpholinos knockdown

For a long time, morpholino oligonucleotides (MOs) were the most used knockdown tools in zebrafish even though, strictly speaking, they do not represent a proper reverse genetic approach as they do not disrupt the genetic information itself. Morpholino antisense oligomers were developed in 1997 as a method to inhibit RNA translation *in vivo* (Summerton et al. 1997). MOs, shown in figure 7A, are chemically modified oligonucleotides of commonly 25 nucleic acid bases bound to morpholine rings (instead of ribose or desoxyribose) and linked to uncharged phosphorodiamidate groups (in place of the phosphodiester groups) (Summerton & Weller 1997). MOs are resistant to nucleases and can bind via base-pairing to mRNAs with high affinity and prevent their translation via steric interference (Figure 7B). The translational blocking is the consequence of MO binding to the 5' untranslated region (UTR) which obstructs the initiation complex of translation (Sumanas & Larson 2002). MOs can also interfere with the splicing process by blocking sites involved in the pre-mRNA processing, leading to the expression of altered transcripts; premature stops and non-sense-mediated decay (Draper et al. 2001). The splicing blocking can be used for gene knock down, but also to study the function of mRNA isoforms resulting from alternative splicing.

In zebrafish, MOs are usually injected in early embryos (1-8 cell-stage) to achieve ubiquitous delivery (Bill et al. 2009). The MO-injected embryos are named "morphants". Hundreds of genes were targeted in this manner and distinct phenotypes were observed. However, this approach has several limitations: the limited time of MO effect as the embryonic development lead to their dilution (up to 5 dpf). The non heritable genome modifications restrain the use of MOs to transient studies. The occurrence of off-target effects which complicate the evaluation of specific impacts. It has been shown that MOs activate the p53 pathway, leading to cell death and embryonic defects in a non-specific manner (Robu et al. 2007). Therefore, control experiments are essential in order to draw proper conclusions using MO gene knock down. For instance, the use of control MOs with several mismatches (5 nt mismatches) which should have no activity, rescue experiments by co-injecting mRNA coding for the target protein, several MOs targeting the same gene should have similar phenotypes, and injecting a p53 MO or using a p53 mutant in order to avoid unspecific effects (Eisen & Smith 2008).
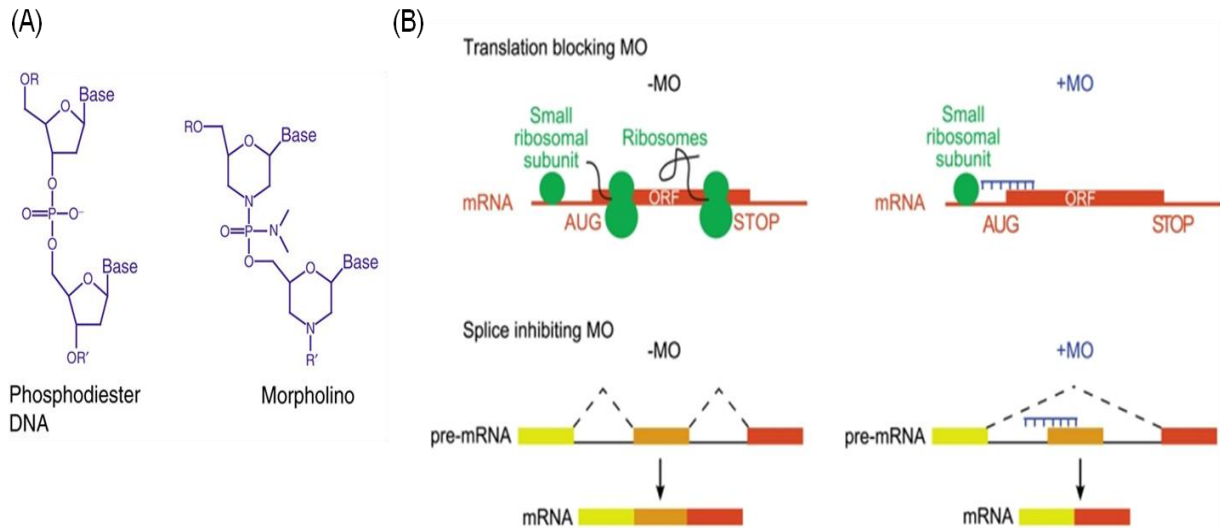
Figure 7: Morpholino anti-sense. (A) Schematic representation of morpholino oligonucleotide structure compared to the DNA. The desoxyriboses is replaced by morpholine rings, while the anionic phosphodiester linker is replaced by a non-ionic phosphorodiamidate group. Adapted from Corey & Abrams 2001 (B) The mode of action of MOs. The translation blocking MO targeting the 5' UTR region prevent translation by blocking the small ribosomal subunit of the initiation complex. The splice inhibiting MOs interfere with important sites of the splicing process of the pre-mRNA giving rise to truncated transcripts. Adapted from Hardy et al. 2010.

Furthermore, very recent studies raised a subject of preoccupation concerning the specificity of MO knock-down. In fact, comparative analyses between morphants and corresponding mutants (obtained with site-specific nucleases) showed discrepancies in 80% of the cases (Kok et al. 2014). A large number of morphant phenotypes could be due to off-target effects of the MOs, according to the authors and to other scientists working on the zebrafish model (Stainier et al. 2015). Indeed, in the study of Kok et al., as an example, they could recapitulate an established morphant phenotype (for the long non-coding RNA *megamind*) even though the embryos lacked the target site of the MO, while the corresponding mutant did not show any resembling phenotype. The conclusion of these worrisome observations is that the MO technology cannot lead to definitive assertions concerning gene function in zebrafish or other models.

To overcome these issues, the suggestion was made to use, more broadly, proper genetic approaches to generate mutants, such as site-specific nuclease technologies that are displayed below.

14

1.3.2. Targeting Induced Local Lesions in Genomes

Targeting Induced Local Lesions in Genomes, or TILLING, is a mutagenesis screening strategy used to identify mutations in a gene of interest from a population of chemically mutagenized fish (Stemple 2004). The mutagen N-ethyl-N-nitrosourea or ENU, used in zebrafish, introduces point mutations in the genome. Sample DNAs of individuals from the library of mutagenized fish are used as a template for screening. After PCR amplification of specific regions of the gene of interest, usually the exons, the products are melted and re-annealed. If a point mutation is present, a heteroduplex will be formed and recognized by the endonuclease Cel-I which cleaves and generates smaller fragments. The detection of mutations is then done by gel electrophoresis and the potentially mutated fragments are subjected to sequencing for confirmation (*Fish Physiology: Zebrafish, 1st Edition,* 2010). Another approach for detection of the mutations is by direct sequencing of the amplified fragment. Large-scale screenings for point mutations were facilitated by the development of next-generation sequencing and whole exome enrichment technologies (Huang et al., 2012).

The large mutant libraries and the large sequencing capacities required to identify a specific mutation in a gene of interest make the TILLING strategy labor-intensive and hard to accomplish in any laboratory. To counter this, a zebrafish mutation project by TILLING has been organized by the Sanger Institute in order to enable researchers to obtain required information about their genes of interest (http://www.sanger.ac.uk/Projects/D_rerio/zmp/).

1.3.3. Zinc-Finger Nucleases

Despite the broad use of the approaches cited above to study the function of important genes in zebrafish, the limit in this model was still the site-specific genome modification for gene targeting, which was restricted for a long time to the mouse model (available embryonic stem cell lines). In 2008, this has changed with two publications that demonstrated the effective generation of targeted mutations in zebrafish by using Zinc-Finger Nucleases, or ZFNs (Doyon et al. 2008; Meng et al. 2008).

ZFNs are chimeric proteins consisting of a zinc-finger DNA-binding domains fused to the catalytic domain of the endonuclease FokI (Kim et al. 1996). The type C2H2 zinc-finger protein used to create the ZFN comprises 30 amino acids and is characterized by a zinc ion coordinated by two cystein residues in two anti-parallel β- sheets and two histidine residues in

an α-helix. It binds to a specific DNA sequence of 3 bp via its alpha helix domain (Ochiai & Yamamoto 2015).
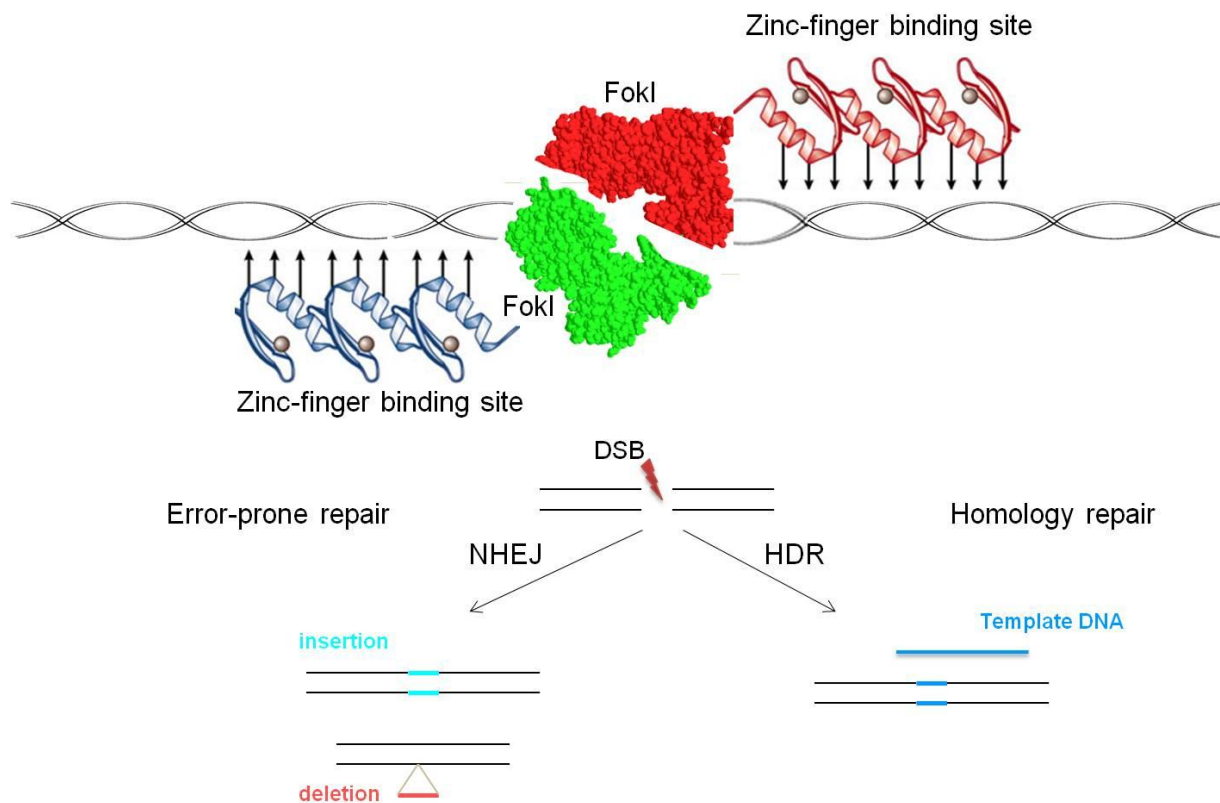
Figure 8: Schematic representation of the zinc finger nucleases (ZFNs). Two opposing zinc-finger binding sites recognize a specific sequence of the DNA. Each finger binds to a DNA triplet. The binding site is fused to the catalytic domain of the endonuclease FokI, which dimerize in order to create double-strand-breaks (DSBs) in the DNA. DSBs are repaired via the error-prone pathway: Non-homologous end-joining (NHEJ) in which the DNA is altered by insertions or deletions (indels), or via Homology repair using a template DNA.

Typically, the DNA-binding domain of a ZFN contains three zinc-finger motifs recognizing a sequence of 9 bp. Furthermore, since the non-specific FokI cleavage domain is active as a dimer, an opposing pair of ZFN is needed to cleave a DNA target sequence, in the spacer region of about 6 bp, between the two binding domains (Bitinaite et al. 1998; Huang et al. 2012). Consequently, double-strand breaks (DSBs) are generated into the DNA and have to be repaired otherwise lethal. The DSBs are repaired by two naturally occurring mechanisms: Non-Homologous End-Joining (NHEJ) or Homologous Recombination (HR). In NHEJ repair, the break ends are directly joined with the introduction of small insertions or deletions (indels) that can alter the genomic sequence as it is an error-prone repair pathway. On the other hand, the HR can be used to edit a gene as it is based on the replacement of the DNA

break using a repair template, either a genomic one or a supplied one in a form of a single strand oligonucleotide (Carroll 2014) (Figure 8).

In zebrafish, the work of two groups, Doyen *et al*. and Meng *et al*. showed the efficiency of ZFN to generate target-specific mutations in this model organism. Using proof-of-principle genes with well known phenotypes, both demonstrated the disruption of these loci in a specific manner by NHEJ repair pathway (Doyon et al. 2008; Meng et al. 2008). Since then, ZFNs as a first genome targeting method in the zebrafish have been used to generate mutations in a large number of genes. Indeed, these customizable proteins can be designed to target any DNA sequence of interest. Several methods are used to design ZFNs, including the modular assembly (MA) which makes use of the direct ligation of preselected zinc-finger modules recognizing DNA triplets and collected in libraries (Amacher 2008). However, the zinc-finger proteins designed with this method showed high failure rates because of the context dependency of zinc-fingers (for other zinc-fingers and for the DNA) (Ramirez et al. 2008). To counter this difficulty, methods to select individual fingers with well-validated *in vivo* activity have been developed. The Oligomerized Pool Engineering (OPEN) method uses the bacterial two-hybrid system to identify new modules from randomized libraries, in a context-dependant manner (Maeder et al. 2008). The Context-Dependant Assembly (CoDA) combines both cited methods to ligate characterized context-dependant zinc-finger modules, already archived, by modular assembly (Sander, Dahlborg, et al. 2011). Also, in order for researchers to avoid the relatively laborious and time consuming construction of zinc-finger proteins, commercially available ZFNs, even though pricey, can be acquired from Sangamo Boscience (CompoZr ZFNs, Sigma-Aldrich).

ZFNs are an effective tool to generate targeted mutations, however, the synthesis of ZFNs with high specificity to the target is expensive, labor intensive and time consuming. Moreover, recently the development of alternative technologies such as transcription activator-like effector nucleases (TALENs) and CRISPR (Clustered regularly interspaced short sequences)/Cas (CRISPR-associated) have emerged as faster and cheaper alternatives.

1.3.4. Transcription Activator-Like Effectors Nucleases

1.3.4.1. Transcription Activator-Like Effectors origins

Transcription Activator-Like (TAL) effectors represent a large family of type-III effectors, predominantly found in *Xanthomonas*, a gram negative bacterial plant pathogen (Boch & Bonas 2010). Type-III denomination comes from the type of protein secretion system used by these effectors. The type-III secretion system is essential for *Xanthomonas* pathogenicity. Indeed, it separates the bacterial membranes and a so called "needle complex" penetrates the host plant cell (Galán & Wolf-Watz 2006). The bacterial proteins are then directly secreted into the host cytoplasm. Different plant pathways are targeted in order to suppress the plant defense and to facilitate the bacterial virulence (Figure 9).
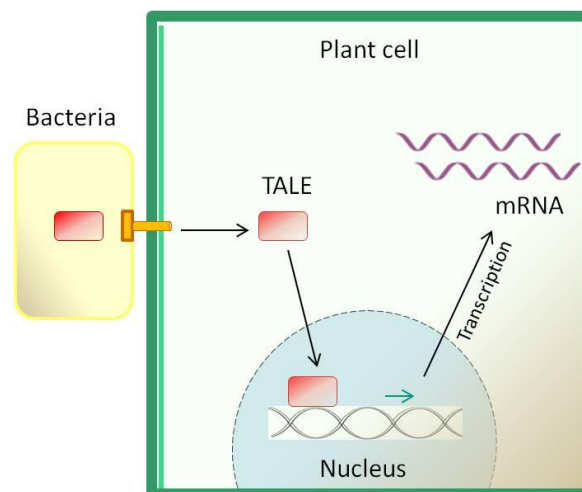
Figure 9: Representation of TAL effector functions in a plant cell. The TALE is secreted by a bacterial plant pathogen via the needle complex of the type III secretion system. The TALE translocates into the nucleus where it binds to the DNA and activate the transcription of host genes to facilitate spread and colonization.

The first member of this family to be isolated from the pepper and tomato pathogen *Xanthomonas campestris* was the *avrBs3* gene. The DNA sequence analysis (by deletion analysis) of this gene showed a number of direct-repeats present 17.5 times, each one consisting of 34 amino acids (aa) (Bonas et al. 1989). The last repeat of AvrBs3 is only similar to the other repeats in its first 20 aa and was therefore called "half-repeat". Several homologs were discovered from other *Xanthomonas* species, and strikingly with highly conserved tendem-repeats. The difference lied on the number, the order and the length of the

repeats. Comparison of the repeats uncovered hypervariable residues at the position 12 and 13 (Hopkins et al. n.d.).

The AvrBs3 family effectors contain a N-terminal region required for the type III secretion, an essential nuclear-localisation signal (NLS) and an acidic activation domain (AD) at the C-terminal region (Figure 10A) (Bogdanove et al. 2010). These motifs were also shown to be conserved in the *avrBs3* homologs.
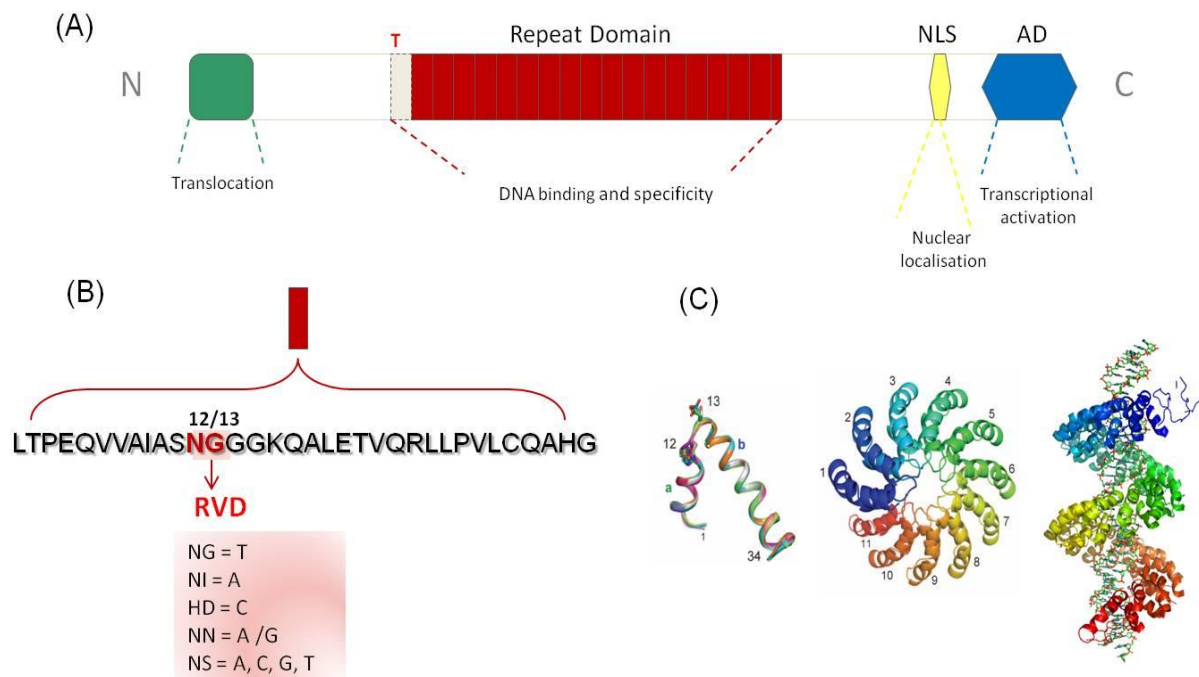
Figure 10: TAL effector organization and structure. (A) The TALE contains a translocation domain (green) at the N-terminal region. The central region consists of the repeat domain (red) which binds DNA, preceded by a thymine at position 0. A NLS (yellow) permits nuclear transport and a transactivation domain AD (blue) which activates transcription at the C-terminal region. (B) A closer view into one repeat which contains 34 amino acids with a variable di-residue (RVD) at position 12 and 13, having one-to-one correspondence to a single nucleotide (red text box). (C) Structural representation of the TALE repeats. From left to right, one repeat harboring the RVD, a TAL effector of 11.5 repeats and its interaction with DNA. The structure representation is adapted from Deng et al. 2012.

The functionality of the AD domain of TAL effectors was first shown in the protein of the rice pathogen *Xanthomonas oryzae*, but the activity of TALEs as transcription factors was revealed with AvrBs3 and its targets named UPA (upregulated by AvrBs3). Indeed, this TAL effector directly binds to an UPA box (a conserved promoter element) to activate the transcription of the UPA genes which will facilitate the colonization and spread of the pathogen (Bogdanove et al. 2010).

The specificity of TAL effectors was already shown more than ten years ago to be determined by its repeat domain (Herbers et al. 1992). Several years later, another discovery increased interest in TAL effectors, the capacity of AvrBs3 to interact directly with DNA (Römer et al. 2007; Kay et al. 2007). This novel DNA-binding domain was further studied in order to identify the base pair recognition specificities. In 2009, two different studies cracked indeed the code of the DNA binding specificity of TAL effectors (Moscou & Bogdanove 2009; Boch et al. 2009).

Moscou and Bogdanove solved the cipher *in silico* by alignment of known TAL effector target promoter sequences and the hypervariable residues, named repeat-variable diresidue (RVD), with minimal entropy. The consistence of RVD-nucleotide associations seems to concord to a one-to-one correspondence between an RVD and a nucleotide in the target site of TAL effectors (Figure 10B). This cipher, partially degenerate, has no evident context dependence, except for a strictly conserved T at position zero (Moscou & Bogdanove 2009).

Boch *et al.* used the UPA box on which AvrBs3 directly binds to investigate the specificity of the TAL effector binding. The size of the UPA box corresponds roughly to the number of repeats in AvrBs3. It was observed indeed that the RVD of each repeat projected onto the UPA box correlates with a single nucleotide in the target DNA. To confirm this model of recognition, DNA target sequences were predicted for three different TAL effectors (Hax2, Hax3, Hax4) using a GUS assay (β-glucuronidase). The promoter boxes were cloned into a vector driven by a minimal promoter (tomato Bs4) and in front of the β-glucoronidase reporter gene. After co-transfection with the effectors driven by the cauliflower mosaic virus 35S-promoter (CMV), the induction of only the corresponding promoters was observed. Also in this study, a conserved T at the 5' end of the target site was noticed and its importance was demonstrated (Boch et al. 2009).

These two studies represent a breakthrough in gene targeting with breaking a simple code that can be used to predict the DNA targets of TAL effectors, but also and mostly to generate DNA binding domains targeting any DNA sequence. Both works identified the TAL effector recognition specificity as follow: HD = C; NG = T; NI = A; NN = A or G; NS = A, C, G or T (single-letter codes for the amino acids 12 and 13).

Structural bases of the TAL effector organization and the TAL-DNA interaction were uncovered in 2012 by two studies: one based on a naturally occurring TAL of 23.5 repeats

from the rice pathogen *Xanthomonas oryzae* (Mak et al. 2012), and the second on an engineered TAL dHax3 of 11.5 repeats (Deng et al. 2012) (Figure 10C). Both studies reached the conclusions that the repeats are connected with each other and form a right-handed superhelix that surrounds the major groove of DNA. The two RVDs are exposed to DNA and just the second residue at position 13 interacts specifically with a nucleotide in the DNA, while the other residue 12 stabilizes the interaction between the protein and the DNA.

The modular DNA-binding domain of TAL effectors quickly attracted attention for biotechnology. Like zink-finger domains, the TALE DNA-binding specificities can be used to design molecules which recognize any target DNA sequence and fused to nucleases, activator domains or repressor domains, can generate highly specific tools to manipulate gene expression.

1.3.4.2.   TAL Effector Nucleases design and assembly – different approaches

Similar in a certain way to zinc-finger domains, because of their natural and modular binding specificities, TAL effectors were perfect candidates for genome engineering. Thus, TAL Effector Nucleases (TALENs), like Zink-Finger Nucleases (ZFNs) in the past, were created by fusing the TALE DNA-binding domain to the non-specific catalytic domain of the endonuclease FokI (Christian et al. 2010) (Li et al. 2011) (Miller et al. 2011). Unlike ZFNs that recognize nucleotide triplets in the DNA, TALENs have the capacity to recognize single nucleotides. This feature makes TALENs more specific and more flexible than ZFNs. The fusion with FokI permits the use of TALENs for targeted gene modification. Indeed, the FokI domain, in its active dimerized form, cleaves both strands of DNA and creates double-strand breaks (DSBs) in the genome. This DNA damage can be then repaired via either the error-prone non-homologous end-joining (NHEJ) machinery, which results in small insertions or deletions (indels) that can lead to a frameshift or to the introduction of a premature stop codon, or via homologous recombination (HR) or homology-directed repair (HDR), in the presence of a DNA sequence used as a donor (Figure 11).

With the enthusiasm that grew in the last few years for the TALEN technology, different methods have been established to design and create/construct an effector against a specific DNA target sequence.
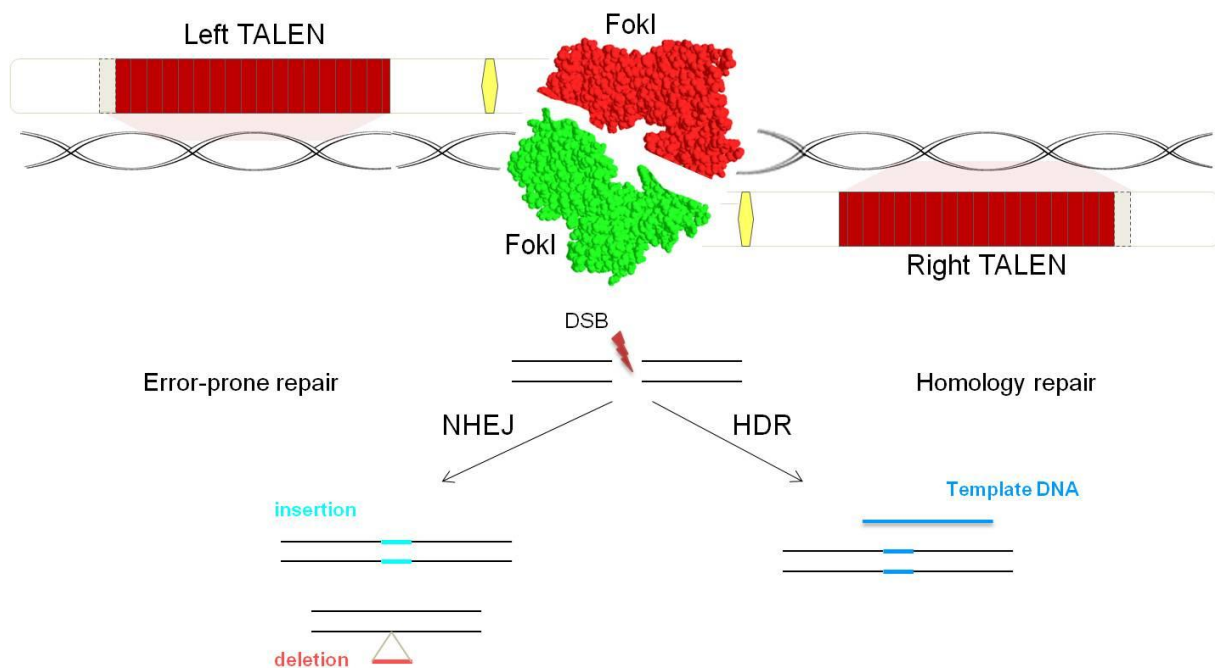
Figure 11: A schematic representation of TAL effector fusion with the catalytic domain of the endonuclease FokI (TALEN). The short distance between opposite positioned TALENs allows the dimerization of FokI catalytic domains (red and green) which then activates the cleavage of the DNA in both sides, creating double-strand breaks (DSB). DSBs are repaired via error-prone repair with the non-homologous end-joining machinery (NHEJ) which leads to insertions or deletions in the target DNA, or via homology repair when a template DNA is available.

The first step to get a working TAL effector binding site is to design a repeat array which recognizes the desired target sequence. To that purpose, several tools have been developed and are available online or as downloadable softwares. The study of Cermack et al. provided with the first online tool for TALEN-targeting design (TAL Effector Nucleotide Targeter, TALENT; https://tale-nt.cac.cornell.edu/) , based on guidelines inspired by the natural occurring effectors, such as the presence of a T upstream of the TALEN binding site for which the length is comprised between 15 and 20 bases. As FokI acts as a dimer, a pair of binding sites is needed and the spacer region between the opposing domains, in which the cleavage will take place, would have a size of 15 to 30 bp (Doyle et al. 2012). In the following years, this tool and other ones with similar criteria have been optimized and harbored additional options. The Target Finder, part of the TALENT suite, allows the identification of the best scoring site with less off-target activity. Like Mojo Hand (http://talendesign.org/, (Neff et al. 2013)), CHOPCHOP (https://chopchop.rc.fas.harvard.edu/index.php (Montague et al. 2014a)) or other web-based

tools, prediction of off-targets is included, as well as selection of the RVD targeting Guanine. In fact, the RVD NN, previously mentioned, recognizes G but also A. Another RVD targeting exclusively G was brought into light recently/subsequently, NH. But it was shown in a study that using NH instead of NN increases the specificity for G but lowers the activity of the overall TALEN (Streubel et al. 2012).

From cut-ligation to bead-dependant, whether it is manual or high-throughput, many strategies have emerged in order to construct TALENs. The first successful assembly of custom TAL effector nucleases was shown by Christian et al. in 2010. The modular assembly was the first way used to construct *de novo* TALENs (Morbitzer et al. 2011; Cermak et al. 2011). The most accomplished method at the time was well described by Cermack et al. and uses the Golden Gate cloning, allowing multiple DNA fragments to be cut and ligated in an ordered manner and in a single reaction. Indeed, it is based on the ability of type IIS restriction endonucleases to cleave outside their recognition sites and to create unique 4 bp overhangs which can be ligated in an ordered fashion as it eliminates the original restriction site (Engler et al. 2009).

Other cut-ligation methods have emerged: Zhang lab used the same basic Golden Gate method but replacing the overnight bacterial culture by PCR amplification, which make the construction time shorter (3 days compared to 5 days for the normal Golden Gate assembly). Hierarchical digestion/ligation strategies, like the "unit assembly" which relies on isocaudomer; restriction enzymes with different recognition sites but generate identical ends upon cleavage (Huang et al. 2011a), or the restriction enzyme and ligation (REAL) method that uses both type I and type II restriction enzymes for series of cut-ligation steps (D Reyon et al. 2012). Besides, a ligation-independent cloning method (LIC) was reported, based on a library of DNA fragments which anneal to each other via long ssDNA overhangs (10-30 bp) (Schmid-Burgk et al. 2013).

Naturally, high-throughput and automated methods to construct TALENs have emerged. The fast ligation-based automatable solid-phase high-throughput (FLASH) method, as the first large-scale method for TALEN assembly, relies on a library of one, two, three or four TALE repeats in different possible combinations of RVDs. It uses solid-phase magnetic beads, allowing automation of the different steps of digestion, purification and ligation of the diverse/various repeat fragments that will be assembled in an iterative manner. This method permits the simultaneous construction of up to 96 different TALENs in a 96-well plate

(Deepak Reyon et al. 2012). The iterative cap assembly (ICA), similarly to the FLASH method, uses solid-state beads on which, this time, DNA repeat monomers are assembled sequentially. Another property of this method is the use of capping oligonucleotides (short hairpins) to inactivate incomplete constructed chains. ICA has the advantage of being applicable to microarray printing technology in order to construct a large amount of TALENs in a short time (Briggs et al. 2012).

Depending on the need, on the capability and on the TAL scaffold to be used, TALENs can be constructed with any of these methods. For our part, the Golden Gate assembly was a suitable method at the time because of its simplicity and its availability.

1.3.5. Clustered Regularly Interspaced Short Palindromic Repeats
1.3.5.1. A bacterial immune system

Bacteria and archaea had evolved their protection against invaders, such as bacteriophages or plasmids, to the point of benefiting from a complex and effective adaptive and heritable immune system. In 1987, Ishino and colleagues observed an unusual structure while analyzing a DNA fragment of *Escherichia coli* from the K12 strain. Direct repeats arranged with 29 nt homologous sequences, separated by 32 nt spacer regions (Ishino et al., 1987). At the time, the biological relevance of these first observed sequences was not known but they were recognized few years later as a family of repetitive DNA regions particular to prokaryotes (Mojica et al. 2000). In the early 2000s, with the expansion of genomic studies, the Clustered Regularly Interspaced Short Palindromic Repeats, or CRISPR, were characterized *in silico* (Jansen et al. 2002). The CRISPR elements were shown to comprise the CRISPR locus which consists of an array of short nearly identical direct repeats separated by polymorphic sequences, or spacers, and a set of CRISPR-associated (*cas*) genes (Jansen et al. 2002).

In 2005, three groups simultaneously reported homologies between the polymorphic sequences of the CRISPR loci and DNA sequences of viruses or plasmids (Bolotin et al. 2005; Mojica et al. 2005; Pourcel et al. 2005). It is known now that the spacers are actually derived from foreign DNA invading bacteria and that they are incorporated to the genome to provide a memory in order to resist against the same invaders (Barrangou et al. 2007).

There are three types of CRISPR systems according to a unified classification following multiple criteria (Makarova et al. 2011). The simplest and well studied system is the type II from *Streptococcus thermophilus* which uses the Cas9 endonuclease in its defense against bacteriophages and plasmids. The process of this immune system is built on three different steps (Bhaya et al. 2011) (Figure 12):
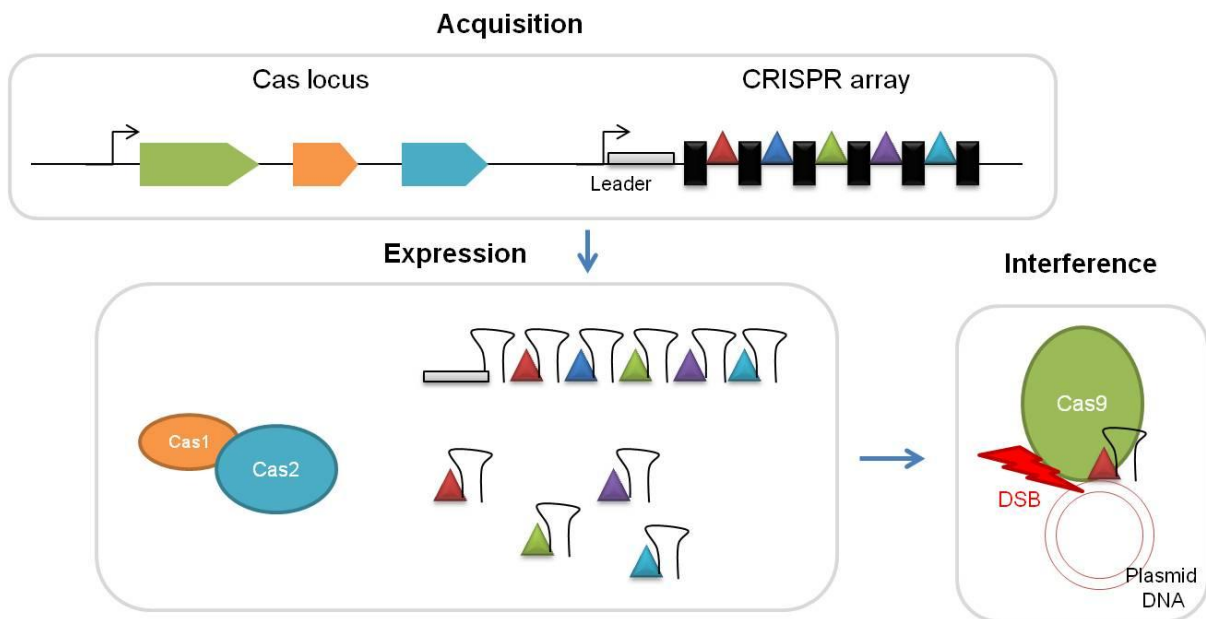
Figure 12: Schematic representation of the CRISPR type II immune system. *Acquisition*: the first step in which the hypervariable spacer containing the PAM (colored triangle) is integrated between direct repeats (black rectangle) within the CRISPR array preceded by the leader sequence. The Cas locus in the vicinity contains the genes encoding Cas proteins. *Expression*: the CRISPR array is transcribed and processed into CRISPR RNAs (crRNAs) containing the spacer region, with the help of the Cas proteins (Cas1, Cas2). *Interference*: during invasion of a plasmid DNA, the crRNA recognizes and binds to the complementary sequence of the spacer and the recruited Cas9 cleaves the target DNA and creates double strand breaks (DSB).

1) The spacer *acquisition* is the first to happen, in which the spacers are integrated in between the repeats and at the leader region. The acquisition of the spacer requires a recognition motif which is at the end of the sequence, named protospacer adjacent motif (PAM). The proteins Cas1 and Cas2 which are involved in this stage are expressed from the *Cas* locus located in the neighboring of the CRISPR array;

2) Second, *expression* from the CRISPR locus. Pre-CRISPR RNAs (pre-crRNAs) (primary transcripts) are transcribed and processed/cleaved into small CRISPR RNAs (crRNAs) or the guide RNAs. This step requires a *trans*-activating small CRISPR RNA (tracrRNA) which forms a duplex with the CRISPR repeats;

3) The final *interference* step requires the crRNA to recognize and to bind the complementary sequence of a foreign DNA, providing a memory of past invaders. Thus, the cleavage of this complex is initiated by Cas proteins.

During the studies of the CRISPR system, the important parts that allow the interference to be achieved were identified: the leader sequence next to the CRISPR array for the acquisition of new spacers (Karginov & Hannon 2010). The complex tracrRNA-crRNA which recognizes the target DNA sequence (Bhaya et al. 2011). The trinucleotide NGG motif PAM that is present at the proximal end of the target which is important for the recognition and for the recruitment of the endonuclease Cas9, avoiding then self-targeting (Sternberg et al. 2014). It was also shown that the cleavage activity of Cas9 depends on a perfect complementarity of a 12 bp "seed region" which is proximal to the PAM in the sequence of the guide RNA, whereas mismatches in the 8 bp most distant from the PAM are tolerated (up to 6) (Jinek et al. 2012).
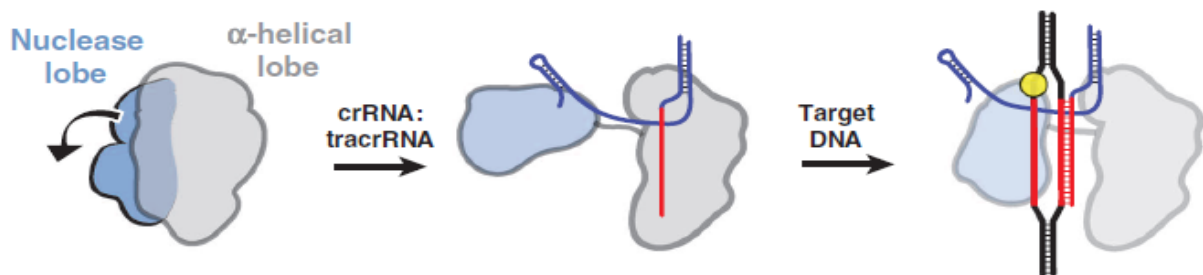
Figure 13: Structure of the Cas9 endonuclease. Representation of the structure of the Cas9 with its two lobes (grey and blue on the left) which change conformation upon interaction with the crRNA:tracrRNA (middle) and reorient to form a central channel to bind the target DNA (right). Adapted from Jinek et al. 2014.

Cas9 is a multi-functional protein organized in two lobes: an α-helical recognition lobe which binds the gRNA and the DNA and a nuclease lobe (Figure 13). The nuclease lobe contains two domains: HNH-nuclease domain which cleaves the target DNA strand, and RuvC-like nuclease domain which cut the non-target strand (Jinek et al. 2014a). The interaction with the PAM is governed by the carboxyl-terminal domain of the nuclease lobe (Nishimasu et al. 2014). The architecture of Cas9 protein, that is autoinhibited in absence of interactions with nucleic acids, undergoes a conformational change, leading to an active conformation, after binding to the guide RNA and the target DNA (Jinek et al. 2014; Jiang & Doudna 2015).

1.3.5.2. Programmable targeting with small RNAs

With the properties of the CRISPR system in mind, numerous investigations were performed in order to use this relatively easy procedure for gene targeting (Figure 14). A synthetic chimera of tracrRNA-crRNA was created as a single guide RNA (sgRNA) to simplify the targeting (Jinek et al. 2012) and different variants of Cas9 proteins were used in several model organisms: wild-type Cas9 that cleaves dsDNA causing DSBs, which will be repaired vie NHEJ repair or HR if a template is available (Mali et al. 2013a; Hwang et al. 2013; Yasue et al. 2014); a mutated Cas9, the Cas9 nickase (Cas9n) that introduces site specific single-strand nicks (Cong et al. 2013; Ran et al. 2013); and a catalytically inactive Cas9, the dead Cas9 (dCas9) that can be fused to various domains, for instance transcriptional activators or repressors (Gilbert et al. 2013; Maeder et al. 2013).

The CRISPR/Cas9 technology opened the door for wide range applications: genome editing, transcriptional control, DNA labeling. And all these possibilities are facilitated by the easy use and do-it-yourself features of this method. In addition, the targeting efficiency of the CRISPR system can be favorably compared to other methods such as ZFNs and TALENs.

It is preferable to target near the N-terminus of the coding region of a gene because a resulting frameshift is more likely to create a null allele the more upstream it is. Several online tools to design the gRNAs are available, with more or less the same criteria. The most important signature is indeed the presence of the PAM motif (NGG) at the direct vicinity of the target sequence. The 5' end of the gRNA may require specific dinucleotides depending on the polymerase used for the RNA synthesis. For example, in our case, 5'GG- is required for the T7 polymerase. Determination of off-targets became rapidly a common goal for most of the tools. Indeed, the Cas9 endonuclease can tolerate up to 5 base mismatches if adjacent to the PAM sequence and also a single base variation in the PAM itself. Among the available tools, ZiFiT Targeter version 4.2 (http://zifit.partners.org/ZiFiT/Disclaimer.aspx), an updated version from the Zinc Finger Consortium was mostly used. A target is identified from a query sequence and the criterion according to the selected promoter can be added (Sander et al. 2007). The output is given in the form of two oligonucleotides which can be annealed and cloned into an expression vector. Another web-based tool, CHOPCHOP (https://chopchop.rc.fas.harvard.edu/index.php) can be used to identify gRNA targets from species-specific gene IDs (Montague et al. 2014b).
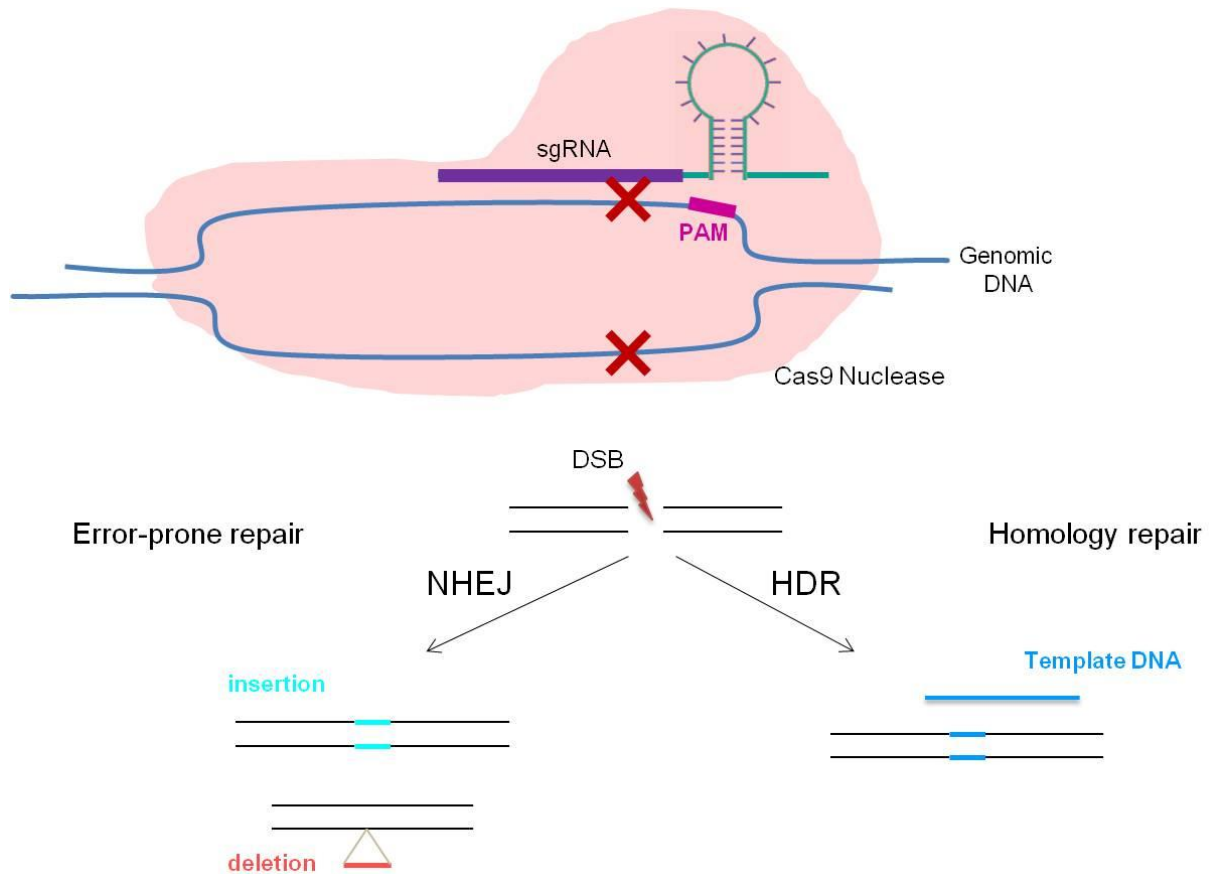
Figure 14: A schematic representation of the CRISPR system. The single guide RNA (sgRNA) targets a sequence of the DNA containing a protospacer adjacent motif (PAM), required for target recognition, and recruits the Cas9 endonuclease that cleaves both strands. DSBs are subsequently generated and repaired via error-prone repair (NHEJ), creating insertions or deletions in the target DNA (indels), or via homology repair using a template DNA.

A number of Cas9 proteins were published from different groups to work in zebrafish. Most of them derive from *Streptococcus pyogenes*, either codon optimized for zebrafish or synthetically produced (Hwang et al. 2013; Jao et al. 2013a; Ablain et al. 2015).

Aim

A better understanding of the molecular mechanisms involved in embryonic and adult neurogenesis is necessary to open the door for new alternatives against neurodegenerative diseases, such as Alzheimer's or Parkinson's diseases. The main characters of these mechanisms are TFs, proteins that regulate gene expression by binding to promoter or enhancer regions of the DNA. This transcriptional control plays an important role in formation and differentiation of functional neurons. In order to better understand the molecular networks controlling the processes of neurogenesis and regeneration of the brain, the zebrafish CNS is used as a model for its easy handling and comparable neuronal mechanisms to other vertebrates comprising mammals/humans.

The tools to target specific genes in the zebrafish genome were greatly optimized in the last few years. Thus, in this study, the two newly developed methods, TALENs and CRISPR, were used in order to perform loss-of-function studies to determine the role of interesting TR genes and their molecular mechanisms controlling neurogenesis and adult brain regeneration, in two model systems, namely the embryonic spinal cord and the adult telencephalon.

Chapter II: MATERIAL AND METHODS

2.1. Equipment and materials

Bacteria incubators	Heraeus, Hanau, Germany
Cellstar cellculture dish	Greiner, Nürtingen, Germany
Cool centrifuge J2-HS	Beckman, Stuttgart, Germany
Cryostatic vial	Eppendorf, Hamburg, Germany
Dissection forceps	Fine Tip No.5 (Dumont)
Eppendorf microcentrifuge tubes	Eppendorf, Hamburg, Germany
Falkon tubes	Greiner, Nürtingen, Germany
FemtoJet microinjector	Eppendorf, Hamburg, Germany
Flaming-Brown Needle puller	Sutter Instruments, USA
Gas microinjector	Tritech research inc., L.A., USA
Glass needle	1.0 mm outer diameter, 0.58 mm inner diameter, with filament (TW100, WPI Inc.)
Incubator for fish embryos	Heraeus, Hanau, Germany
LED pannels 660 / 740	Roither , Switzerland
Magnetic thermomixer	Heidolph, Rosenfeld, Germany
Microcentrifuge 5417 R and C	Eppendorf, Hamburg, Germany
Microcentrifuge Biofuge pico	Heraeus, Hanau, Germany
Microfiltration columns	Pall, Ann Arbor, USA
Microloader tips	Eppendorf, Hamburg, Germany
Microscopes	Leica DMI6000 SD Stereomicroscope SMZ645 Fluorescent stereomicroscopeMZ FLI- II
Multitank recirculation system	Schwarz Ltd, Germany
NanoDrop ND-1000	Peqlab, Erlangen, Germany
Omnifix®-F 0.01-1ml Syringe	Braun, Melsungen AG, Germany
PCR-Thermocycler	Biozym Scientific GmbH, HessischOldendorf, Germany
Petri dishes	Greiner, Nürtingen, Germany
Pipette tips	Corning, Corning, UK

Rotilabo® -syringe filters 0.22 µm	Roth, Karlsruhe, Germany
Spectrophotometer	Eppendorf, Hamburg, Germany
Spin-X-Filter	Costar, Corning, USA
Sterile filters	Renner, Darmstadt, Germany
Thermomixer	Eppendorf, Hamburg, Germany
TOP seal A	Perkin Elmer, Massachusetts, USA
Vortex	Bender &Hohbein, Karlsruhe
Water bath	Kötterman, Uetze, Germany

2.2.   Chemicals

1-phenyl-2Tricaine	Sigma-Aldrich, Taufkirchen, Germany
4',6-diamidino-2-phenylindole (DAPI)	Sigma-Aldrich, Taufkirchen, Germany
Acetic acid	Merck, Darmstadt, Germany
Agarose	Peqlab, Erlangen, Germany
Ammoniumacetate	Merck, Darmstadt, Germany
Ampicillin	Roth, Karlsruhe, Germany
Bacto Agar	Roth, Karlsruhe, Germany
Calcium acetate	Roth, Karlsruhe, Germany
Calciumchloride	Sigma-Aldrich, Taufkirchen, Germany
Chloroform	Sigma-Aldrich, Taufkirchen, Germany
Citric acid	Roth, Karlsruhe, Germany
Dimethylsulfoxide (DMSO)	Fluka, Neu-Ulm, Germany
Disodiumhydrogen phosphate	Roth, Karlsruhe, Germany
Dithiothreitol (DTT)	Roth, Karlsruhe, Germany
DNA-Ladder (1 kb)	New England Biolabs, Frankfurt a.M, Germany
DNA-Ladder (Mix)	Peqlab, Erlangen, Germany
DNAseI	Ambion Ltd, Warrington, UK
dNTP	Promega, Mannheim, Germany
Ethanol (EtOH)	Roth, Karlsruhe, Germany
Ethidiumbromide	Roth, Karlsruhe, Germany
Ethylene diaminetetra acetic acid (EDTA)	Roth, Karlsruhe, Germany

Formaldehyde	Merck, Darmstadt, Germany
Gentamicin	Sigma,Taufkirchen, Germany
Glycerol	Roth, Karlsruhe, Germany
GoTaq DNA polymerase	Promega, Mannheim, Germany
Hydrochloric acid (HCl)	Merck, Darmstadt, Germany
Isoamyl alcohol	Roth, Karlsruhe, Germany
Isopropanol	Roth, Karlsruhe, Germany
Kaliumchloride	Sigma-Aldrich, Taufkirchen, Germany
Kanamycin	Sigma-Aldrich, Taufkirchen, Germany
Leibovitz's L-15	Gibco, Karlsruhe, Germany
Low melting agarose	Roth, Karlsruhe, Germany
Luciferin	Biosynth AG, Staad, Schweiz
Magnesium chloride hexahydrate	Roth, Karlsruhe, Germany
Magnesium sulphate ($MgSO_4$)	Sigma-Aldrich, Taufkirchen, Germany
Methanol (MeOH)	Roth, Karlsruhe, Germany
Methylene blue	Interpret, Germany
Nicotinamide adenine dinucleotide	New England Biolabs, Frankfurt a.M, Germany
Nuclease free water	Ambion,Huntigdon,UK
Oligonucleotides	Metabion, Planegg, Germany
Paraformaldehyde	Merck, Darmstadt, Germany
Phosphate Buffered Saline (PBS)	Life Technologies GmbH, Darmstadt, Germany
PEG 8000	Promega, Mannheim, Germany
Penicilin/Streptomycin	Invitrogen, Karlsruhe, Germany
Pfu-DNA polymerase	Promega, Mannheim, Germany
pGEMT-easy	Promega, Mannheim, Germany
Phenol	Roth, Karlsruhe, Germany
Phenol red	Roth, Karlsruhe, Germany
Phosphate buffered saline w/o $CaCl_2$ and $MgCl_2$ ($PBS^{-/-}$)	Invitrogen, Karlsruhe, Germany
Plasmid safe nuclease	Biozym Scientific GmbH, HessischOldendorf, Germany
Proteinase K	Sigma-Aldrich, Taufkirchen, Germany
Quick ligase	New England Biolabs, Frankfurt a.M, Germany

Restriction endonucleases	Promega, Mannheim, Germany
Restriction enzymes	New England Biolabs, Frankfurt a.M, Germany
Reverse transcriptase	Promega, Mannheim, Germany
Sodium acetate (NaAc)	Roth, Karlsruhe, Germany
Sodium chloride (NaCl)	Roth, Karlsruhe, Germany
Sodium citrate tribasic dihydrate	Sigma-Aldrich, Taufkirchen, Germany
Sodium Fluoride (NaF)	Roth, Karlsruhe, Germany
Sodium hydrogen carbonate (NaHCO$_3$)	Roth, Karlsruhe, Germany
Sodium hydroxide (NaOH)	Roth, Karlsruhe, Germany
Sodiumdodecylsulphate (SDS)	Roth, Karlsruhe, Germany
Sp6 RNA polymerase	Life Technologies GmbH, Darmstadt, Germany
T4 DNA Ligase	New England Biolabs, Frankfurt a.M, Germany
T4 DNA Polymerase	New England Biolabs, Frankfurt a.M, Germany
T5 Exonuclease	Biozym Scientific GmbH, HessischOldendorf, Germany
T7 RNA polymerase	Life Technologies GmbH, Darmstadt, Germany
T7 endonuclease I	New England Biolabs, Frankfurt a.M, Germany
Taq DNA Ligase	New England Biolabs, Frankfurt a.M, Germany
Taq-Polymerase	Promega, Mannheim, Germany
Tetracycline	Roth, Karlsruhe, Germany
Tris-base	Roth, Karlsruhe, Germany
Tris-HCl	Roth, Karlsruhe, Germany
Triton-X-100	Roth, Karlsruhe, Germany
Trypsin 0.25% (w/v)-EDTA	Gibco/Invitrogen, Karlsruhe, Germany
TSAP	Promega, Mannheim, Germany
Tween 20	Roth, Karlsruhe, Germany

2.3. Kits

MEGAshortscript™ T7 Kit	Life Technologies GmbH, Darmstadt, Germany
mMESSAGEmMACHINE Transcription Kits	Ambion, Darmstadt, Germany
Nucleospin DNA purification kit	Macherey-Nagel, Düren, Germany
PCR purification kit	Macherey-Nagel, Düren, Germany
peqGold Gel extraction kit	Peqlab, Erlangen, Germany
QIAGEN Plasmid Midi purification kit	Qiagen, Hilden, Germany
QIAGEN Plasmid Maxi Kit	Qiagen, Hilden, Germany
QIAquick Gel Extraction Kit	Qiagen, Hilden, Germany
QIAquick PCR Purification Kit	Qiagen, Hilden, Germany

2.4. Software

Ensembl Genome Browser Genome *databases* for vertebrates and other eukaryotic species	http://www.ensembl.org/index.html
UCSC Genome Browser Reference sequence and working draft assemblies for a large collection of genomes	http://genome.ucsc.edu/
Primer3web version 4.0.0 Online tool to pick primers from DNA sequence	http://bioinfo.ut.ee/primer3/
PCR Primer Stats Online tool to evaluate potential PCR primers	http://www.bioinformatics.org/sms2/pcr_primer_stats.html
OligoCalc Online oligonucleotide properties calculator	http://www.basic.northwestern.edu/biotools/oligocalc.html
ApE- A plasmid Editor v2.0.47 Sequence editor using a modified GenBank format	*Davis M Wayne* of University of Utah http://biologylabs.utah.edu/jorgensen/wayned/ape/
The Basic Local Alignment Search Tool (BLAST)	http://blast.ncbi.nlm.nih.gov/
ClustalW2 DNA or protein multiple sequence alignment program	http://www.ebi.ac.uk/Tools/msa/clustalw2/

Tal Effector Nucleotide Targter 2.0 Online tool to design custom TAL effectors	https://tale-nt.cac.cornell.edu/
Mojo Hand Online tool for genome editing applications	http://talendesign.org/
ZiFiT Targeter version 4.2 Online tool for genome editing applications	http://zifit.partners.org/ZiFiT/Disclaimer.aspx
CHOPCHOP Online tool for genome editing applications	https://chopchop.rc.fas.harvard.edu/index.php
GT-Scan Identifying Unique Genomic Targets Online tool for predicting optimal target sites	http://gt-scan.braembl.org.au/gt-scan/
Pfam database A collection of protein families	http://pfam.xfam.org/

2.5.   Oligonucleotides

All oligonucleotides have been designed using the software "primer3" version 0.4.0 (Rozen & Skaletsky 2000), evaluated using PCR Primer Stat and purchased from Metabion GmbH, Planegg/Steinkirchen, Germany.

Primers for PCR amplification

gene	Primer Fw	Primer Rev
ntl	AATGGGCCACTTTGTACGTC	CACCAATGAAATGATTGTCACC
		CTGGATGAAAGCACCCGTAT
id1	GCACCTCGCTTCAGCTATTC	GATTCTCCAGCATGTCATCG
Id3	CTTCCATCGTTCAGTGAGCA	AAACACGCCTGACATTTTCC
Id2a	AACCAGCGAGTGAACAAACC	ACATCCAGAACACCCCTGAC
Id4	TGCGTTCACACTCAGAGAGG	CAGAGCAGGTAAGCAGGACC
sox1a	ACTGGCTACAGGAGCGAAAA	TACAGACCGAGGCGCAAAAC
sox1b	ACCGATCTCCTGTTCCAGTG	GGTTAGAGAGTCCGGTGGGT

2.6. TALEN target sites and binding sequences

gene	Target (Spacer) + <u>Restriction site</u>	Binding site left	Binding site right	Enzyme
ntl	TATGTCTG<u>CCTC</u>AAG	T TTATTTGATCGGAAA	TCCCGACCAGCGCCTGG	MnlI
id1	CAA<u>GATCC</u>CGCTGCT	T TGGCCATCTCCAAATG	GGACGAGCAGATGACC	AlwI
Id3	CCT<u>CGCGA</u>TCAGCCG	T GTATCTCGGAGCAGAG	CTGCAAGAGTCCTTCCG	NruI
sox1a	<u>GATGG</u>AAACGGACCT	T GAATGTATAGCATGAT	TCATTCCCCGGGACCCC	BccI
sox1b	GATG<u>GAGACG</u>GACTT	T GAATGTATAGCATGAT	GCACTCTCCAGGAGCCC	Esp3I

2.7. CRISPR target site and gRNA oligonucleotides

gene	Target + <u>PAM</u>	Oligo1	Oligo2	NB
ntl	GGATCATCTCCTTAGCGCCG<u>TGG</u>	TAGGATCATCTCCTTAGCGCCG	AAACCGGCGCTAAGGAGATGAT	fw
ntl'	GGTGACACTGGCTCTGAGCA<u>GCC</u>	TAGGTGACACTGGCTCTGAGCA	AAACTGCTCAGAGCCAGTGTCA	rev
id3	GGACTGATCGCCTTCATTCT<u>CCC</u>	TAGGACTGATCGCCTTCATTCT	AAACAGAATGAAGGCGATCAGT	rev
id3'	GGACTTGTTCTGCGGGATGC<u>ACC</u>	TAGGACTTGTTCTGCGGGATGC	AAACGCATCCCGCAGAACAAGT	rev
id2a	GGCTGAGAGGATCGTCCACG<u>CCC</u>	TAGGCTGAGAGGATCGTCCACG	AAACCGTGGACGATCCTCTCAG	rev
id4	GGAAGCTTATGAGGGCGAAC<u>GCC</u>	TAGGAAGCTTATGAGGGCGAAC	AAACGTTCGCCCTCATAAGCTT	rev
sox1a	GGAAACGGACCTTCATTCCC<u>CGG</u>	TAGGAAACGGACCTTCATTCCC	AAACGGGAATGAAGGTCCGTTT	fw
sox1a'	GGGGCAAACGGGTCCAAATT<u>CGG</u>	TAGGGGCAAACGGGTCCAAATT	AAACAATTTGGACCCGTTTGCC	fw
sox1b	GGTCACACGGGACCCAACAG<u>CGG</u>	TAGGTCACACGGGACCCAACAG	AAACCTGTTGGGTCCCGTGTGA	fw
sox1b'	GGAGTGGAAACTCATGTCCG<u>AGG</u>	TAGGAGTGGAAACTCATGTCCG	AAACCGGACATGAGTTTCCACT	fw

2.8. Fish lines

ll zebrafish embryos used in this study were of the wildtype Tübingen AB line.

2.9. Bacterial strain

E.coli, Nova Blue®   Chemical competent cells (Invitrogen)

Genotype:K-12 strain

2.10. Methods

Preparation of Luria Bertani (LB)-medium: For 1L: 10 g tryptone, 5 g yeast extract, and 10 g NaCl were dissolved in 950 mL deionized water. The pH of the medium was adjusted to 7.0 using 1N NaOH and brought to a volume of 1 L. The solution was autoclaved on liquid cycle for 20 min at 15 psi and then stored at +4°C. If necessary, antibiotics were added before use (50 μg/ml ampiciline, kanamycine, spectinomycine; 10 μg/ml tetracycline). If tetracycline was used solution was stored in the dark.

LB agar-plates: LB medium was prepared as mentioned above, but 15 g/L agar was added before autoclaving. After autoclaving it has been cooled to approximately 55°C. Antibiotics (if needed) were added and the solution was poured into Petri dishes. After hardening, plates were inverted and stored at +4°C (in case of tetracycline the storage is in the dark).

Preparation of glycerol stock: To prepare a glycerol stock for the TAL Effector library of plasmids (see supplementary information, S.1), 5ml LB media was inoculated with a single colony and grown to mid log phase (carefully not allowing saturation). Then, 800 μl of the E-coli culture was transferred in ao-ring tube and filled with 200μl autoclaved glycerol (80%). Tubes were carefully mixed by vortexing and stored at -80°C.

Plasmid DNA extraction: Liquid cultures were prepared by inoculating 3 ml or 100 ml aliquots of LB medium (containing the appropriate antibiotic), for miniprep or maxiprep respectively, using single bacterial colonies. Plasmid DNA was extracted from saturated overnight cultures (at 37°C) using a column-based extraction system according to the manufacturer's instructions, for small yield (NucleoSpin® Plasmid, Macherey-Nagel) or high yield (Qiagen Plasmid Maxi kit) of DNA preparation. The pellets were dissolved in 50 μl to 100 μl Millipore water. Yields of DNA were calculated based on optical density measurements of dilutions made from the stocks. The optical density was measured at both 260 nm and 280 nm and used to assess DNA purity as well as concentration (NanoDrop, Peqlab).

DNA Digestion: DNA is digested as follows: 10 μg total DNA

10 μl 10x buffer

2 μl restriction enzyme (10 U/μl)

Adjusting water to 100 μl

The reaction is incubated at 37°C (or depending on the enzyme requirements). 1 µl of the digestion is loaded on a 1.5% agarose gel for verification (adding 4 µl water and 1 µl blue loading dye).

Phenol/chloroform purification: For phenol/chloroform purification of plasmid DNA, 100 µl of plasmid containing solution was mixed with 100 µl of phenol/chloroform (1:1 v/v) and mixed by vortexing for 10 sec before centrifugation at 14000 g for 2 min at room temperature (RT). Following the centrifugation, the supernatant was carefully removed (avoiding interphase) and transferred into a clean 1.5 ml Eppendorf tube. 100 µl of chloroform were added to the supernatant and again vortexed and centrifuged under the same conditions. The new supernatant was transferred into a fresh Eppendorf tube and 100 µl isopropanol (100%) was added and the mix was stored for at least 30 min at -80°C before centrifugation at 13000 g for 25 min at 4°C. The resulting DNA pellet was washed with 200 µl of ethanol (75%) and re-centrifuged for 10 min at 4°C. Supernatant was carefully removed and the DNA pellet was air dried at RT for about 15 min before resolving in 16 µl of RNase free water by pipetting up and down several times. The DNA was then stored at -20°C.

Agarose Gel Electrophoresis: For size and quality check, DNA fragments are separated by electrophoresis. Depending on the expected size of the DNA fragments, agarose gels with a concentration varying from 1.0% to 2.0% were prepared. For the visualization of the DNA on a UV-transilluminator, Ethidium Bromide at final concentration of 0,5 µg/ml is added. Before loading, the samples are supplemented with loading dye (NEB) and the electrophoresis is carried out in 1xTAE electrophoresis buffer with 3 -5 V/cm intensity of electric field. An appropriate DNA marker (DNA ladder, NEB) was loaded in parallel for determining the size and the approximate quantity of the DNA samples.

Isolation of DNA from agarose gel: For cloning purposes or microinjections, DNA fragments resulting from a restriction digest or PCR were separated by agarose gel electrophoresis. The band containing the desired DNA fragment was cut out from the gel and purified with the peqGOLD gel extraction kit was used following tightly the manufacturer's protocol.

Polymerase Chain reaction (PCR): The amplification of DNA fragments from genomic DNA or plasmids is performed by PCR. Two enzyme systems were used depending on the purpose. Ordinary Taq polymerase was used when proof reading activity was not needed (for

example colony PCR tests). Proof-reading polymerases were utilized for the amplification of DNA fragments for cloning purposes. The PCR was performed according to the user manuals provided with the enzymes, with adjustment of the annealing temperature and elongation time according to the used primers and the size of the amplified fragments. All PCRs were performed on a thermocycler (Biozym Scientific).

Colony PCR: Colony PCR was used to quickly screen and identify the correct size of plasmid fragments present in *E. coli* colonies.

A reaction was set up as follows:

PCR mix:

> 5 µl 10x buffer (500 mM KCL, 100 mM Tris-HCL(pH9.0), 1,0% Triton X 100)
> 1 µl dNTPs (10 mM each dATP, dTTP, dGTP, dCTP)
> 1 µl 20 µM forward primer
> 1 µM 20 µM reverse primer
> 0.2-1µl Taq polymerase (GO, Pfu (for proofreading))
> Adjusted water to 50 µl total reaction volume

All reactions were set up on ice and the enzyme was added last. To each pre-chilled tube containing the PCR mix, 2 µl of bacterial colony was added. The tubes were placed into a thermocycler for PCR reaction. PCR conditions varied within the experiments regarding the requirements for the desired fragments.

Ligation: Ligation was performed within 5 min at room temperature using the Quick ligation kit (NEB) according to the manufactures protocol.

Bacterial transformation: For transformation 1 µl DNA was added to 50 µl of competent bacterial cells. After incubation on ice for 20 min. a heat shock was performed for 45 sec. at 42°C. The transformations were kept on ice for another 10 min. before adding 500 µl of LB-medium. Then cells were shaken for 15-30 min at 37°C before plated onto respective LB and antibiotics containing agar plates for colony growth over night at 37°C.

Zebrafish care: The zebrafish work was conducted along the guidelines of the European Zebrafish Research Center (EZRC-KIT). Wildtype zebrafish *danio rerio* (AB line, University of Oregon; Eugene, USA) were raised, bred and crossed according to standard methods,

maintained on a 14 h/10 h light-dark cycle at 28.5°C in recirculation systems (Schwarz Ltd Germany, Müller and Pfleger Ltd Germany) and fed commercial food and in-house hatched brine shrimp as described (Westerfield, 2007). All zebrafish husbandry and experimental procedures were performed in accordance with the German animal protection standards and were approved by the Government of Baden-Württemberg, Regierungspräsidium Karlsruhe, Germany. Fertilized eggs were collected within 2 h of laying and transferred into petri dishes (10 cm diameter) containing E3 medium and the fungicide methylene blue (1 mg/ml). E3 medium was changed regularly.

TALEN target site design: The TAL Effector Targeter (TALEN-T) software (https://tale-nt.cac.cornell.edu/) (Doyle et al. 2012) was initially used to design TALEN candidates, according to three criteria: a Thymine start, the TALEN binding site length ranged from 15-25 bases, the spacer region between 15 and 30 base pairs (bp), and a possible restriction site within the spacer to simplify downstream analysis. However, the use of the GoldyTALEN backbone (available from Addgene, 38142) restricted the length of the spacer region to 15-16 bp. Another, simple, open access software, Mojo Hand (http://talendesign.org/), was used (Neff et al. 2013).

TALEN synthesis: The assembly of the TALEN binding domain, with the repeats containing the proper RVDs was performed using the Golden Gate approach (Cermak et al. 2011). Each TALE binding site (left or right arm) is designed to target 15 to 17 bp DNA of interest. The first assembly is performed with cut-ligation method using the endonuclease BsaI (New England Biolabs). The primers pCR8_F1 and pCR8_R1 (supplementary information, S.2) are used to sequence the intermediate products. The positive arrays are purified (NucleoSpin® Plasmid, Macherey-Nagel) and introduced with cut-ligation procedure, using the endonuclease Esp3I (Life Technologies), into the GoldyTALEN expression plasmid containing the wild-type FokI catalytic domain (available from Addgene, 38142). All obtained TALEN expression vectors were verified by sequencing using the primers: SeqTALEN_5-1 and TAL_F2 (supplementary information, S.2). DNA purification for high yield of the final DNA constructs is performed (Qiagen Plasmid Maxi kit) and the concentration is assessed (Nanodrop).

TALEN mRNA synthesis and microinjection: *In vitro* transcription was performed after linearization of the plasmid with SacI endonuclease (New England Biolabs) at 37°C overnight digestion. 5'-capped and tailed mRNAs were generated (T3 mMessage Machine kit, Ambion;

Poly(A) tailing kit, Life Technologies) and purified by phenol/chloroform extraction. The mRNA was injected into one-cell stage zebrafish embryos at a concentration of 100 ng/µl (approximately 200 pg per target).

DNA isolation from embryos: Genomic DNA was extracted at 72 hours post-fertilization (hpf) from 30-50 individual larvae, by incubating in 20 mg/ml Proteinase K mixed with TE buffer (10 mM of Tris-HCl pH 8.0, 1 mM of EDTA), at 55°C for 240 min, followed by heating at 94°C for 10 min before cooling down to 4°C.

Analysis of somatic mutations with TALENs: Targeted genomic regions were amplified from the isolated genomic DNA by PCR, using primers which anneal between 200 and 800 bp upstream and downstream from the expected cut site. The purified PCR products were suggested to restriction digestion with enzymes recognizing the regions to be cleaved (New England Biolabs or Life Technologies). The fragments are assessed by gel electrophoresis and the positive ones are purified and cloned into the pGEM-T easy vector (Promega). After transformation of the ligated vectors, single colonies were re-inoculated for plasmid DNA isolation (NucleoSpin® Plasmid, Macherey-Nagel) and sequencing (using M13 primers). All primers used for DNA amplification are shown in Table 3.

CRISPR target site design and construction: Potential targets for the sgRNA-Cas system were identified according to the ZiFiT Targeter website (http://zifit.partners.org/), and following the criteria: 5'-GG-$(N)_{18}$-NGG-3'. The appropriately designed oligonucleotides (Table 2) were annealed and introduced into the expression plasmid pDR274 (available from Addgene, 42250) digested with BsaI restriction enzyme (New England Biolabs). All obtained sgRNA expression vectors were verified by sequencing (using M13 primer). After DraI (New England Biolabs) digestion of the expression vector, the sgRNAs were transcribed *in vitro* (T7 mMessage Machine kit, Ambion). The vector encoding the Cas9 nuclease (available from Addgene, 46757) was linearized with XbaI restriction enzyme (New England Biolabs) and transcribed *in vitro* (T3 mMessage Machine ultra kit, Ambion). Both the sgRNA and the Cas9 encoding mRNAs were then purified by phenol/chlorophorm extraction and concentrations were assessed (Nanodrop). The mRNAs were co-injected into one-cell stage zebrafish embryos. The mixed solution contained 25 pg of sgRNA and 150 pg of Cas9 mRNA per target, approximately.

T7 endonuclease assay: After 3 days post injection, injected and un-injected embryos are collected and the genomic DNA is extracted from single embryos as described before. Targeted genomic regions were amplified using primers designed to anneal upstream and downstream of the expected cut site (Table 3, Annex). PCR products were purified according to the manufacturer's instructions (Nucleospin® Gel and PCR clean-up, *Macherey-Nagel*). T7 Endonuclease I (New England Biolabs) assay was performed following published protocols (Jao *et al.*, 2013). 2 µl of 10xNEB 2 buffer (New England Biolabs) was added to 200 ng total PCR product. The DNA sample was re-annealed for formation of heteroduplex, as follows: 95°C for 5 min, 95 to 85°C decreasing by -2°C/s, then ramping until reaching 25°C at -0.1°C/s. 10U of T7 Endonuclease I was added to the DNA and incubated at 37°C for 15 min. The digested products were loaded on an agarose gel for electrophoresis.

Genotyping by fin biopsy: Fish are anesthetized by immersion in 0.004% MS-222 (Tricaine, Sigma) until gill movement is slowed. The anesthetized fish is placed on a clean surface where the fin is clipped with a sterilized scalpel. The amputated fin is transferred into a sterile 1,5µl eppendorf tube containing 100 µl of self made tail buffer (1M Tris-HCl pH 8.5, 5M NaCl, 0,5M EDTA, 10% SDS, 10mg/ml Proteinase K). After overnight incubation at 62°C, the genomic DNA is extracted with Isopropanol, washed with 70% Ethanol and re suspended in nuclease-free water. The samples are stored at 4°C.

pGEM T easy: The PCR products from targeted sequences are cloned into pGEM T easy vector using the T4 DNA ligase. After incubation of 1h at room temperature, bacterial transformation is performed and followed by selection of transformants on LB/Ampicillin plates.

DNA Sequencing: Purified pGEM T positive clones and purified PCR products are sequenced according to the company requirements by M13 primer and primers used for amplification, respectively. All sequencing in this study is done by Microsynth AG (Sanger Sequencing Division). Single insertions, deletions or substitutions were considered as PCR or sequencing artefacts. All primers used for sequencing are shown in Table 2 (Annex). DNA sequence alignments were performed using Fish Factor Database global alignment (http://141.52.98.21/ffdb/gloal.php) and ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Chapter III: RESULTS

In the whole-genome screen performed in our laboratory few years ago, 3302 potential TR genes were identified in the zebrafish embryo. The cDNAs of 2149 TRs were cloned and the expression pattern of 1711 TR genes was obtained by *in situ* hybridization at 24 hpf (Armant et al. 2013). Among the identified TRs, two interesting regulators: Sox1a and Sox1b were expressed in the brain and in distinct cells of the spinal cord. Our group was interested in the cells expressing *sox1a* and *sox1b* in the V2 domain of the spinal cord, in order to investigate regulators of the specification of V2 interneurons (Yang et al., *unpublished data*).

Another systematic screen was performed to identify TRs implicated in adult neurogenesis and brain regeneration of the zebrafish. The expression analyses were fulfilled by *in situ* hybridization in the intact and injured adult zebrafish telencephalon. From this study, an expression map of more than 1200 TRs was established in the adult zebrafish telencephalon and 279 TR genes were identified to be misregulated after brain injury (Diotel et al. 2015; Viales et al. 2015). Among them, *id1* drew attention as it was shown to be expressed in the VZ of the adult telencephalon and to be upregulated after injury.

In spite of previously performed studies on these target genes using morpholino knockdown and overexpression (cf. Introduction), many grey areas persist in the comprehension of the molecular mechanisms in which they are implicated. For this reason, we had established in our laboratory genome editing technology for genetic mutagenesis and I created heritable mutations in several target genes. Hence, zebrafish knockout lines can be obtained in order to be used as a null background for further molecular investigations.

The overall aim of the current PhD project has been to use loss-of-function strategies to study the molecular mechanisms involved in neurogenesis and adult brain regeneration, focusing on our preferred genes: *id* and *sox1*. To that purpose, I first had to establish in the laboratory newly developed gene-targeting methods; Transcription Activator-Like Effector Nucleases (TALENs) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR).

3.1. Launch of TAL effector technology for use in zebrafish

At the beginning of this project, the TAL effector technology just came out on the field of genome editing and little was known about the different methods to create TALENs and their efficiency in zebrafish (Sander et al. 2011; Huang et al. 2011). As the Golden Gate assembly

was well documented and simple to use in an average laboratory, it was a suitable method for the start of our study (Cermak et al. 2011).

3.1.1. Golden Gate Assembly of TALENs

As previously mentioned (cf. Introduction), the Golden Gate cloning is based on type IIS restriction enzymes, which digest DNA outside of their recognition site and leave 4 nucleotides overhangs. These sequences can be designed in a way to generate unique complementary ends between several fragments that can be ligated in an ordered fashion (Figure 15) (Engler et al. 2009).

(A)

(B)
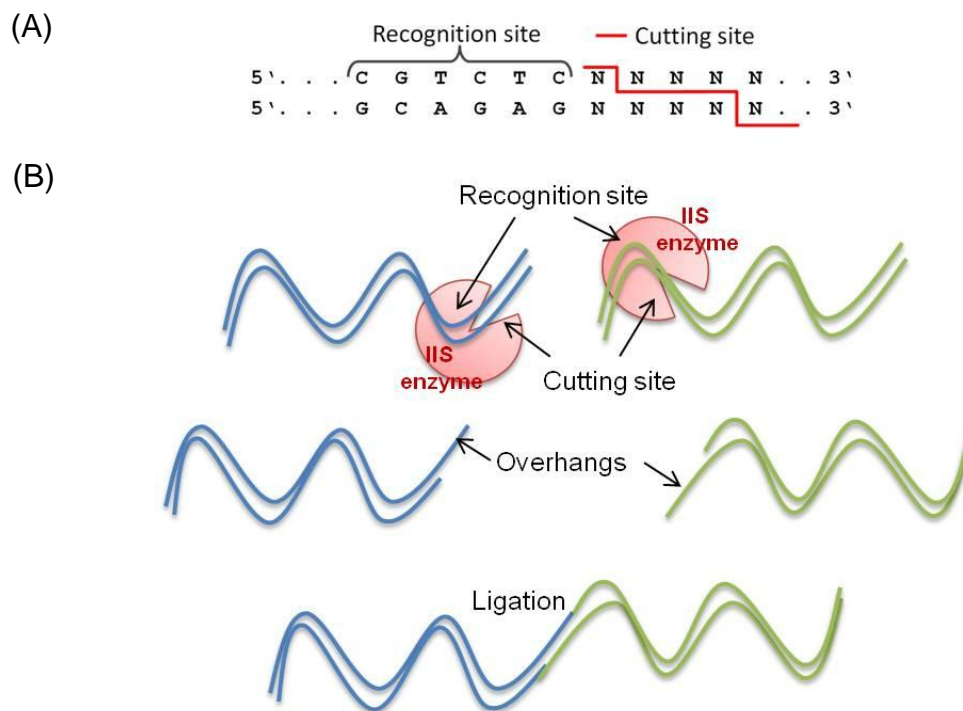
Figure 15: Type IIS endonuclease mode of function for Golden Gate cloning. Type II restriction enzymes recognize a short sequence of 6 bp and cleave close to their recognition site leaving 5'-overhangs of 4 bases. The example of Esp3I (or BsmBI) is shown in (A). (B) Two DNA fragments can be designed with chosen ends flanked by a type IIs restriction site. The digestion of the fragments generates ends with complementary 4 nt overhangs while removing the recognition site. These ends can be then ligated in an ordered fashion.

The Golden Gate assembly of TALENs consists of two steps:

- A first reaction to assemble the modules into intermediate arrays of 1-10 repeats;
- A second reaction to obtain the final construct, by introducing the intermediate arrays with the last half-repeat into an appropriate backbone.

A library of 72 plasmids was available from Cermak et al. (Golden Gate kit, supplementary information S.1), containing:

- Repeat modules for 10 successive positions (designed overhangs) with the various RVDs: pHD (binding to C), pNG (binding to T), pNI (binding to A), pNN (binding to G or A) and pNK (binding to A, C, T or G).
- Last half-repeat plasmids for the five different RVDs: pLR-RVD.
- Array plasmids for the intermediary assembly: pFUS_A (which can contain 10 modules), pFUS_B1 – pFUS_B10 (which can contain 1 to 10 modules).
- Backbone plasmids: pTAL1 – 4 (yeast expression vectors).

After obtaining a bacterial stock from each item (from a collaborative laboratory), DNA purification by maxi-preparation was performed for each and every plasmid and a glycerol stock was produced.

The designed TALEN is constructed as follows according to the Voytas protocol (Cermak et al. 2015):

1) Ist Assembly, in which the first 10 RVD modules are cut with the type IIS enzyme BsaI and ligated in the same reaction to each other and into the pFUS_A intermediate array. The remaining number of modules according to the final length of the TALE (from 1 to 9) is assembled with the same procedure into the appropriate intermediate array (pFUS_B1 – pFUS_B10). The reactions are incubated in a thermocycler for the fragments to be digested and ligated (Figure 16, (A/B));

2) Treatment with Plasmid Safe DNase to eliminate unligated linear fragments (incomplete products or linearized plasmids);

3) Transformation into *Escherichia coli* (*E.coli*), then selection of colonies for overnight bacterial culture in order to purify the DNA (the intermediate arrays contain Spectinomycine resistance). The isolated plasmids are sequenced to identify the correct arrays;

4) IInd Assembly, the 10 containing intermediate arrays are joined with the following 1 to 10 second arrays and with the appropriate last half-repeat into the desired backbone, by cut-ligation reaction. The digestion in this mixture is performed by the type IIS enzyme Esp3I (Figure 16, (C));

5) After transformation into *E.coli* and overnight culture (the backbones contain Ampicillin resistance), the final construct containing the TAL array fused to FokI is purified and characterized by DNA sequencing (Figure 16, (D)).
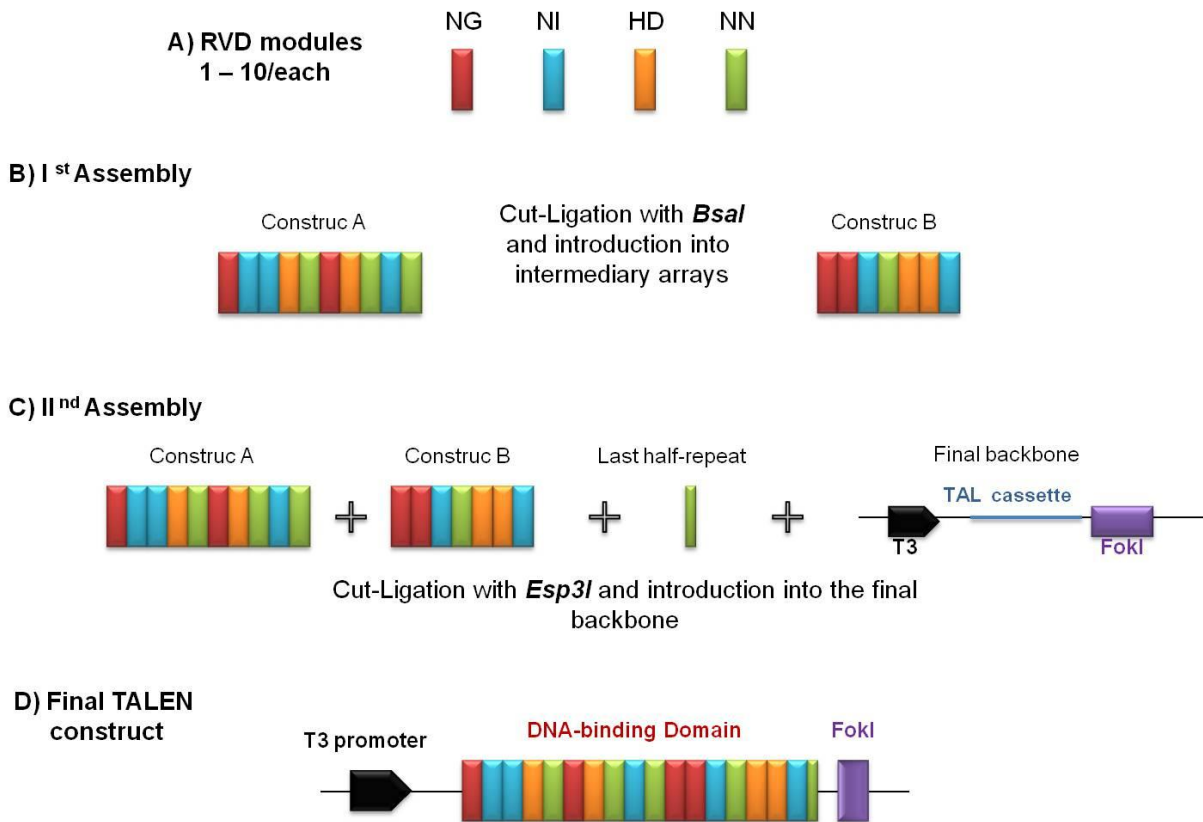

Figure 16: Representation of the outlines of the Golden Gate assembly of TALENs. (A) Representation of the four modules coding for the main RVDs with different colors (NG, NI, HD, NN). (B) The Ist Assembly consists of using the endonuclease BsaI which cuts outside its recognition site and leaves 4 bp overhangs. The cleaved modules are then ligated to each other in an ordered manner, into intermediary arrays, and the constructs A (of 10 RVD modules) and B (of 7 RVD modules) are obtained. (C) In the II nd Assembly, the constructs A and B are ligated using the same principle with the endonuclease Esp3I, with the addition of the last half-repeat, into the final expression backbone containing a T3 promoter driving the TAL cassette and the catalytic domain of the endonuclease FokI. (D) The final TALEN construct is represented with the complete DNA-binding domain, fused to the FokI protein and driven by the T3 promoter (simplified pGoldyTALEN). The used RVDs: NG = T; NI = A; HD = C; NN = G or A.

3.1.2.   First steps for the assembly of customized TALENS

As this study began, the efficiency of the TALEN technology was not as well documented as nowadays. In the first attempt to implement this kind of strategies, a trustworthy control to assess the efficiency of our designed TALENs was therefore elaborated. The *no-tail (ntl)* gene

was interesting for us to be used as a positive control because of its clear phenotype. The zebrafish *ntl* gene is the homologue of the mouse *T* gene, or *Brachyury,* early expressed and required for development of the mesoderm and the establishment of the body axis in vertebrates (Schulte-Merker et al. 1994). The mutant phenotype is similar to the one described in mouse, with the lack of a differentiated notochord and all the posterior structures, but the anterior structures seem to develop normally. The zebrafish *ntl* gene was also shown to encode a transcription factor (Kispert & Herrmann, 1993).

First of all, TALENs against the *ntl* gene were designed with the TALEN Targeter online tool (cf. Introduction) and assembled using the Golden Gate method. Unfortunately, the first attempts to construct the TAL intermediary arrays were unsuccessful. We argued that the issue might come from the modules themselves. Indeed, control assemblies with random TAL repeats were performed and showed anomalies in some of the plasmids defining the different positions of the polymorphic residues. Thus, indicating a confusion of constructs in the given library. In order for us to identify the mixed constructs, we started to assemble into the pFUS_A intermediate array, each of the 10 RVDs recognizing the same bases. For instance, the modules NG 1 to 10 were assembled and when one of the RVDs was not at its right place, the assembly wouldn't work. By that, the type of the wrong RVD was identified and then we tested individually each of them by random construction using already assured modules and their integrity was confirmed by sequencing. The whole library was therefore tested and then the right order of the RVDs was restored.

As mentioned above, the Ist assembly was the first to be achieved: The first 10 modules were mixed in a solution containing the pFus_A intermediate array, the BsaI type IIs endonuclease and the T4 ligase. The last remaining modules, here 6, were mixed in a similar solution except for the intermediate array that was pFus_B6. After the thermocycling procedure and the bacterial culture, the resulting fragment was checked by colony PCR using a supplied primer set (pCR8_F1 and pCR8_R1, supplementary information S.2) and then visualized on an electrophoresis gel. The results are shown in figure 17 for three different clones: The first array "pFus_A" can be observed with a prominent band at around 1200 bp of length, while the second array "pFus_B6" is represented with a smaller band of about 800 bp. Smearing and a "ladder" of bands is also observed every 100 bp resulting from the presence of repeats and indicating a positive clone (Figure 17A-left panel).
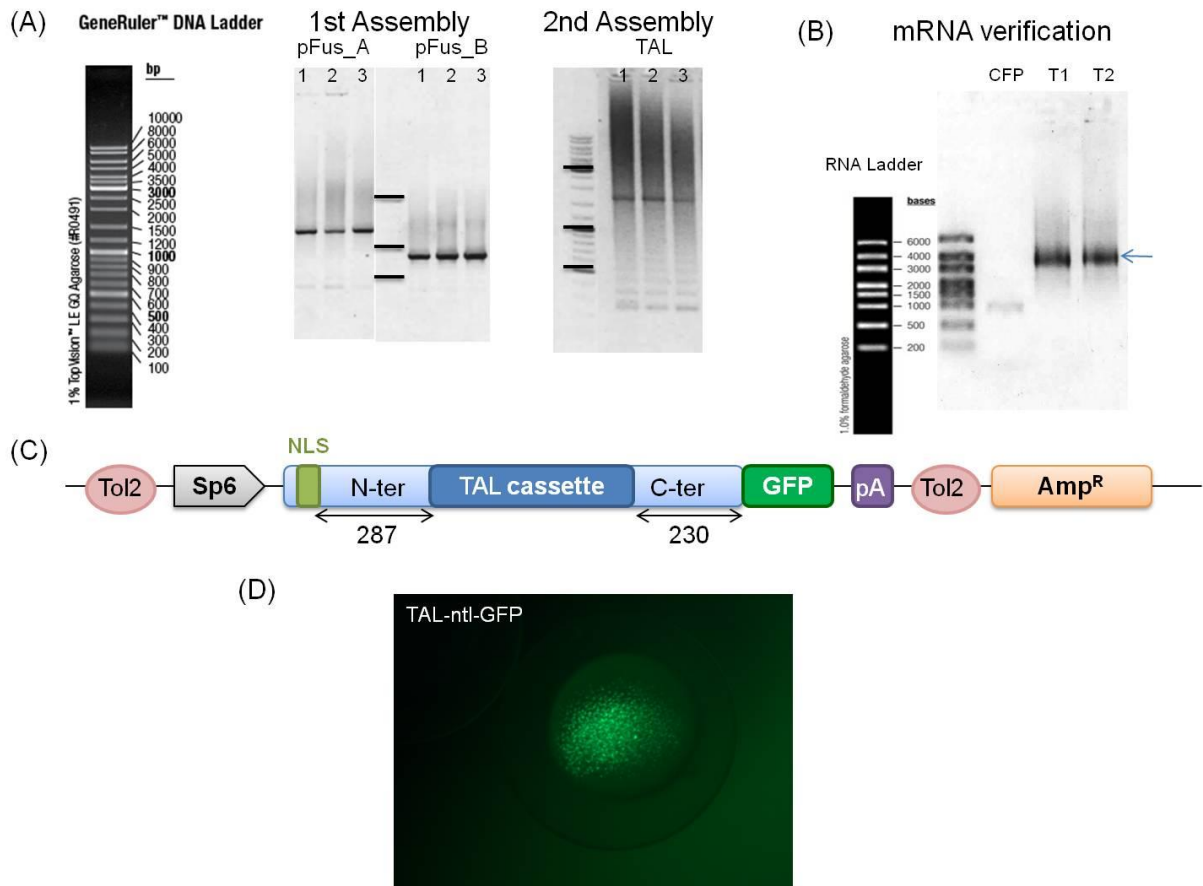
Figure 17: Construction and injection of TALEs against the *no-tail* gene. (A) Gel electrophoresis representing the 1st and the 2nd assemblies of TALENs against *ntl* with the DNA ladder used for all experiments. The 1st assembly comprises intermediate arrays containing 10 repeats (pFus_A ~ 1.2 Kb in 3 clones) and the second array containing 6 repeats (pFus_B ~ 0.8 Kb in 3 clones) (left panel). The 2nd assembly represents the ligation of pFus_A and pFus_B with the last half-repeat in the vector backbone giving a band of ~ 1.8 Kb (right panel). (B) Gel electrophoresis showing the *in vitro* transcribed mRNA with the corresponding size of both left and right TALENs (T1 and T2 ~ 3.5 Kb), with the mRNA of CFP as a control (1 Kb). An RNA ladder is used to assess the size of the RNA. (C) Representation of the Sp6:TAL-GFP used as a vector backbone containing the TAL cassette fused to GFP with an NLS and a polyadenylation signal (pA) and under the Sp6 promoter. The vector also comprises Tol2 arms and resistance to ampicillin (Amp[R]). (D) Visualisation of the expression of the destination vector containing GFP fused to TALE against *ntl* at 6 hpf.

For the IInd assembly, the first used backbone pTAL1 from the library, was derived from a precursor to TALEN yeast expression vector (Cermak et al. 2011). The TAL cassette consists of a fragment of a natural occurring TAL effector gene used as a context for custom repeat arrays. During the assembly step, the same disorder issue occurred and all the last half-repeats had to be tested individually and indeed some were altered. They were controlled by sequencing and their integrity was ensured. Thanks to that, *de novo* designed TAL effectors were successfully assembled and cloned into the final backbone. Similarly to the I[st] assembly, the final assembled constructs were produced by mixing the two intermediate repeats (A and

B6) with the backbone and the corresponding last half-repeat in a solution containing the second type IIs endnuclease Esp3I with the T4 ligase. The resulting clones were then tested by colony PCR using a supplied primer set (pTAL_F1 and pTAL_R2, supplementary information S.2) and visualized on an electrophoresis gel. The positive clones show a prominent band around 1800 bp with the smearing and the "ladder" effect (Figure 17A-right panel).

The pTAL1 vector, like the other available backbones from the kit, was optimized for use in yeast, and thus do not have the requirements for expression in zebrafish. For this reason, an expression vector containing the Sp6 bacteriophage promoter with a polyadenylation (polyA) signal for transcription termination and a Kozak sequence to ensure the eukaryotic translation of a functional TALEN was designed and created. Tol2 arms were also added to facilitate genome integration that can be used in further investigations. The TAL cassette that will receive the designed binding site was derived from the available pTAL3 vector containing also the Fok1 protein and a nuclear-localization signal (NLS). The resulting vector was named Sp6:TAL-FokI. Another alternative vector was created by replacing the FokI protein with GFP; Sp6:TAL-GFP (Figure 17C) (M. Ferg, *unpublished data*).

The injected embryos with both constructs (Sp6:TAL-GFP and Sp6:TAL-FokI) against *ntl* were observed under a stereo microscope. At 6 hours post-fertilization (hpf), GFP was detected mosaically spread in the embryos (Figure 17D). Alas, no phenotype was observed with the FokI version of the construct, also at later stages. However, the DNA had to be assessed in order to identify possible mutations. To achieve that, the genomic DNA of a pool of 10 injected embryos and the DNA of wild-type embryos was extracted, following the published protocol (Sander, Cade, et al. 2011). A fragment of about 200 bp upstream and downstream of the target site was then amplified by PCR and the amplicon was sub-cloned into the pGEM T easy vector system (Promega). A number of clones was subsequently sequenced without success in detecting mutations.

When a truncated version of the TALEN was shown to be more efficient in zebrafish (Bedell et al. 2012), it immediately caught my interest and the vector was ordered from Addgene (plasmid #38142) (Carlson et al. 2012). This newly developed TALEN destination vector was obtained by truncating the pTAL scaffold by 152 amino acids at the N-terminus (a region that has only a function for transport into plant cells), with +63 amino acids at the C-terminal region (originally 278 residues) (Miller et al. 2011). This TALEN scaffold was then adapted

to the Golden Gate assembly method and introduced into a messenger RNA expression vector driven by the T3 promoter (based on pT3Ts vector), resulting into the pGoldyTALEN backbone (Figure 18A) (Carlson et al. 2012).

Additionally, to optimize and facilitate downstream analyses of the TALEN effects, the spacer region would be designed in a sequence containing a centrally located restriction site. Actually, when the TALEN mutates the DNA in this region, the recognition site of the enzyme would be disrupted preventing the cleavage. This strategy can be easily used as a read-out for presence of mutations at the target site; only by subjecting the amplified region of interest to restriction digestion and the result can be observed on an electrophoresis gel (the presence of an uncut band determines the presence of mutations). The amplified fragment is then cloned into the pGEM T easy vector in order to determine the type of mutations occurring at the DNA level.

3.1.3 Successful obtaining of a *ntl* phenotype using the TALEN technology

After the development of the GoldyTALEN plasmid backbone, a new TALEN pair against the gene *ntl* was designed in the first exon of the gene, around the start codon ATG, according to new criteria using the Mojo Hand online tool. Indeed, as the C-terminal region which is fused to the FokI domain is truncated, the spacer needed to be shorter in order to grant the dimerization of FokI. The binding sites of the left and the right TALEN proteins had lengths of 14 and 16 bases, respectively, with a spacer region of 15 bp. The recognition site of the enzyme MnlI was present in the middle of the spacer sequence. The *in vitro* transcription using the T3 promoter and the addition of the polyA signal (as the vector did not contain one) were performed. Then the mRNAs of each TALEN (left and right) were verified by electrophoresis, and equal amounts of the two TALEN parts were co-injected into one-cell-stage embryos of zebrafish. First of all, different concentrations were tested (60, 100, 300, 500 and 800 ng/μl).

After 24 hpf, injected embryos with ntl-TALENs showed a similar phenotype to the *ntl* one, characterized by the absence of the posterior structures and a differentiated notochord, compared to the wild type uninjected embryos (Figure 18B, left panels). After 72 hpf, an even more clear phenotype was observed with the lack of the posterior part (Figure 18B, middle and right panels).
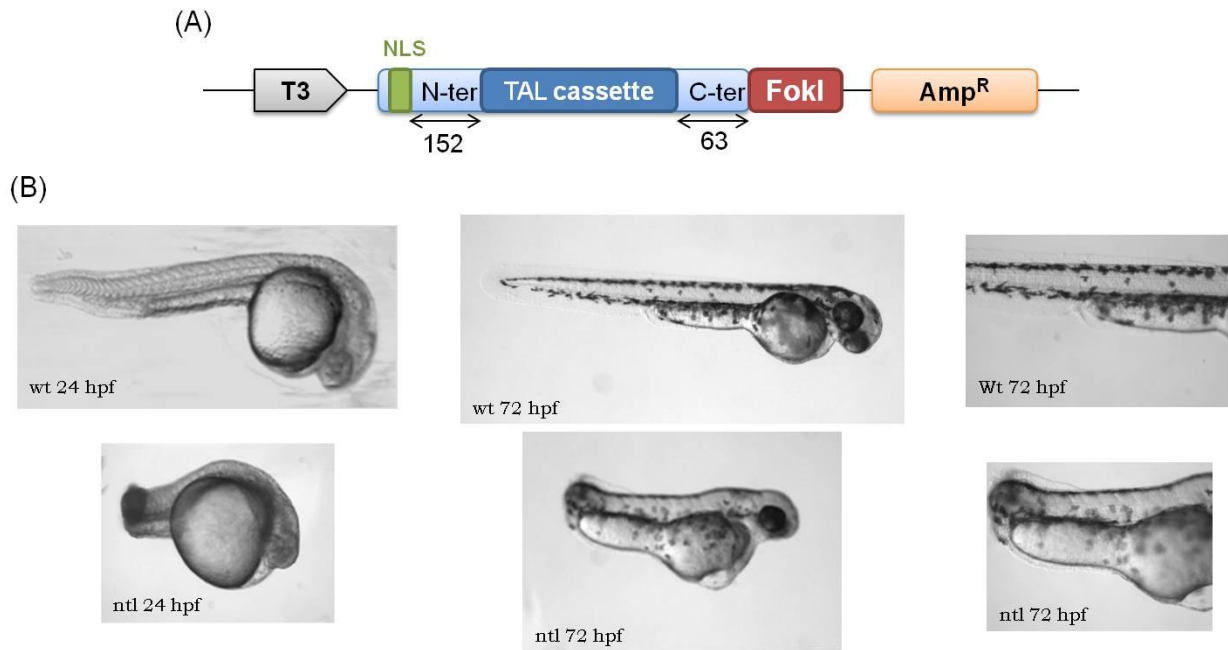
Figure 18: The pGoldyTALEN backbone and its successful use in obtaining a *ntl* phenotype. (A) Representation of the pGoldyTALEN with truncations on the N- and C-terminal parts of the TAL cassette that is fused to the FokI catalytic domain and under the T3 promoter. It also contains an NLS and resistance to ampicillin (Amp^R). (B) ntl phenotype observation after TALEN injection. The upper panel shows a wild-type zebrafish embryo at 24 hpf and at 72 hpf with a closer look at the trunk (from left to right). The bottom part shows the same stages for an embryo injected with both left and right *ntl*-TALENs and the clear lack of the posterior part is observed, while the overall shape of the anterior part seems normal.

The *ntl* phenotype was observed at all tested concentrations in this first experiment, compared to uninjected wt fish. Just by counting the number of fish showing different effects of the injections, some tendencies could be observed. For instance, the phenotype seems more specific at concentrations within a range from 60 to 300 ng/μl, while increasing the concentration results in severe deformations of the embryos bodies, including heart oedema and head malformation, most probably caused by higher toxicity with also an increase of death (Figure 19).
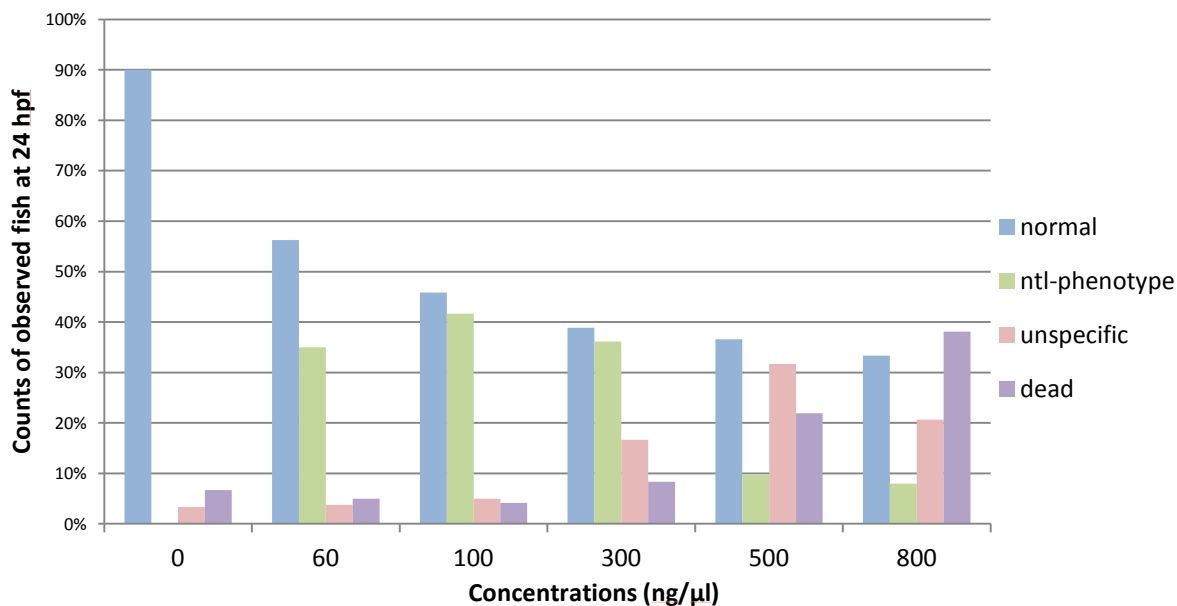
Figure 19: First tested concentrations of *ntl*-TALENs and their effect on zebrafish embryos. At 48 hpf, the injected embryos were observed under stereomicroscope and counted according to the observed effect (here converted into percentages). The *ntl* phenotype was observed at all concentrations except in uninjected embryos used as a negative control (green). A high rate of the *ntl* phenotype was observed at 100 ng/μl. Unspecific effects depict embryos malformations that do not correspond to the known *ntl* phenotype (red). The number of dead and normal looking embryos was also counted (purple and blue, respectively).

In order to analyze the DNA for loss of restriction enzyme recognition site that present at the expected cleaved region, the genomic DNA was extracted from about 30 single embryos at 3 days post-fertilization (dpf): 2x from wt, 18x from 100 ng/μl injected embryos and 10x from 300 ng/μl injected embryos as they showed the best rate of observable phenotype. The fragments of interest were then amplified by PCR, and digested with the appropriate enzyme MnlI. 20-30% of the tested embryos showed mutations at the restriction site on gel electrophoresis: the amplified fragment from a wild type embryo with a size of 426 bp (uncut) was submitted to digestion with MnlI and two bands of 270 bp and 155 bp resulted from it (cut) (Figure 20A, wt-embryo). The amplification and the digestion of the fragment from injected embryos with TALENS gave also the same bands, with the difference that an uncut band was still present in addition of the cut bands (Figure 20A, mut-embryo). Thus, suggesting indeed a disruption of the recognition site of the restriction enzyme MnlI in some cells of the zebrafish embryo. A statistical analysis for three different experiments placed the mutation rate to 35% of the tested embryos (Figure 20B). A positive fragment was

subsequently cloned into pGEM T easy vector and the clones were subjected to sequence analysis using the M13 standard primers. About 30% of the tested clones confirmed the presence of insertions or deletions of varying lengths (e.g. deletion of 4 bp in Mut1, insertion of 14 bp in Mut2) compared to the wild type sequence (WT) (Figure 20C).
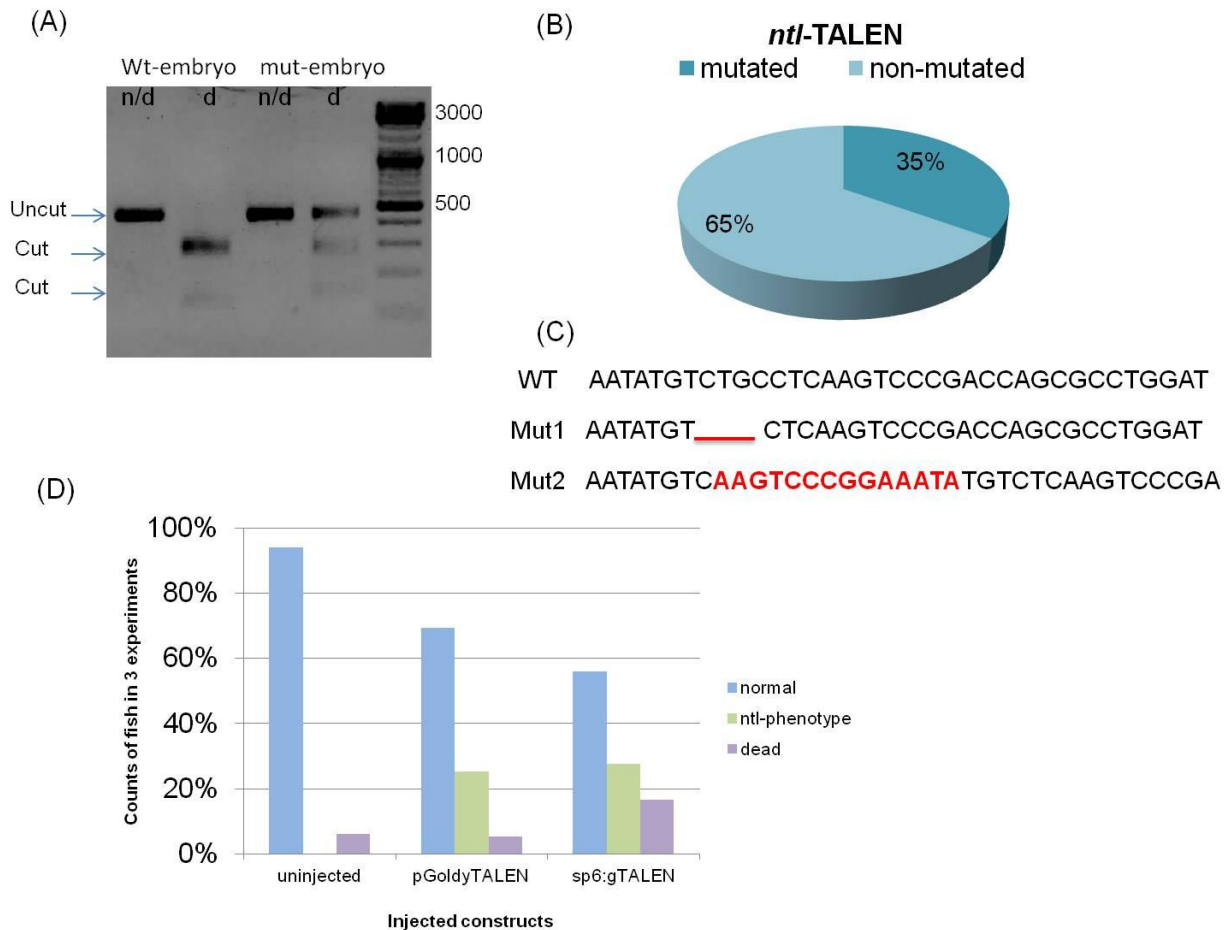
Figure 20: Identification of mutations induced by TALENs in the *ntl* gene. (A) Gel electrophoresis showing the targeted fragment of the gene *ntl* of 426 bp (n/d for undigested) submitted to the restriction enzyme MnlI (d for digested). The genomic DNA extracted from a wild type embryo was used as a negative control (Wt-embryo) and the digested fragment gave rise to 2 bands of about 270 bp and 155 bp (cut). The DNA fragment from an injected embyo (mut-embryo) showed the same 2 cut bands with the presence of a remaining uncut band. (B) Statistical representation of the number of mutated embryos injected with the *ntl*-TALEN from 3 different experiments. 35% of the embryos showed mutations at the target site. (C) Examples of mutations at the target site compared to a WT control sequence. Mut1 shows a deletion of 4 nt and Mut2 shows an insertion of 14 nt (red). (D) Embryos injected with pGoldyTALEN and sp6:gTALEN against the *ntl* gene were observed under stereomicroscope at 24 hpf. The *ntl* phenotype was observed for both constructs with comparable rates of about 25% (green) while no ntl phenotype was observed in uninjected embryos. The number of normal looking and dead embryos are also represented (blue and purple, respectively).

To keep a high mutation probability with low toxicity, the tested concentration of 100 ng/μl was chosen to be used in the following experiments of this study as it showed best results in mutations and a clear *ntl* phenotype with low unspecific and toxic effects. Following experiments using the *ntl*-TALENs were performed in order to test and compare the efficiency of a the created vector Sp6:TAL-FokI after replacing the TAL cassette with the the truncated one used for the pGoldyTALEN (Ferg, *unpublished data*). Injected embryos with pGoldyTALENs and Sp6:gTAL-FokI against the *ntl* gene were counted from three independent experiments and as shown in figure 20D, the percentage of embryos with an observable *ntl* phenotype are comparable for both constructs. However, a slightly higher death effect was observed with the Sp6:gTAL-FokI. These experiments also allowed us to note the efficiency rate of the *ntl*-TALENs which lies between 25% and 30%.

The specific phenotype and the presence of mutations at the target site of the *ntl* gene suggest that *de novo* customized TALENs according to the aforementioned procedure are quite efficient to induce gene modifications in zebrafish.

3.2. Launch of the CRISPR/Cas method for use in zebrafish

During the realization of the current study, another new genome targeting method emerged and was developed for the use on the zebrafish model: the clustered regularly interspaced short palindromic repeats, or CRISPR (Hwang et al. 2013). Due to the simplicity and the fast development of this technique, it was in our interest to establish it and to use it in an ordinary basis. Indeed, the targeting using the guide RNAs (gRNAs) is as easy as to design an oligonucleotide and the costs are much lower compared to the TALEN construction and even to the morpholino oligonucleotides. Furthermore, the method showed good results for multiplex gene targeting in zebrafish (Jao et al. 2013a) allowing to study several genes simultaneously in this model organism (e.g. *id* genes, *sox1* genes).

3.2.1. First steps in using the CRISPR technology

Like for the TALENs at the beginning of this study, we started using the CRISPR technology right after its coming out on the market. For this, the use of a proper control was also essential. Taken the fact that the *no-tail* gene was profitable to test the TALEN method, the same target was used as a proof of principle for the first experiments using the CRISPR/Cas system.

First of all, a gRNA was designed using the ZiFiT Targeter software (cf. Introduction) to target a 20 nucleotides sequence downstream of the ATG start codon, in the first exon of the *ntl* gene (Figure 21A). The gRNA was constructed by annealing two complementary oligonucleotides of 24 nt, with selected overhangs in both ends. The expression vector pDR274 carrying a T7 promoter from Joung's lab (Hwang et al. 2013), tested in zebrafish and available from Addgene (plasmid #42250), was used to clone the double stranded gRNA. BsaI restriction enzyme from the type IIs endonucleases family was used to cut and to create the corresponding overhangs for the fragment to be introduced into the vector (Figure 21B).

In the same publication from Hwang et al., the Cas9 nuclease used to create mutations in the zebrafish genome was derived from *Streptococcus pyogenes,* with the addition of a T7 promoter 5' site and a nuclear-localization signal (NLS) at the C-terminal region. Following the published protocol, both gRNA and Cas9 mRNAs were injected into one-cell stage zebrafish embryos, with concentrations of 12.5 ng/µl and 300 ng/µl, respectively. After 24 hpf, the injected embryos were observed under stereo microscope, but no specific *ntl*

phenotype was detected. On the other hand, many other unspecific malformations, including heart oedema, were observed in a number of embryos. This would be explained by the large concentration of the injected Cas9 mRNA that would lead to higher toxicity. However, reduction of this concentration in following injections did not give more promising results. We first argued that the Cas9 mRNA was not optimally transcribed, thus the obtained mRNA was run on gel electrophoresis before each injection.
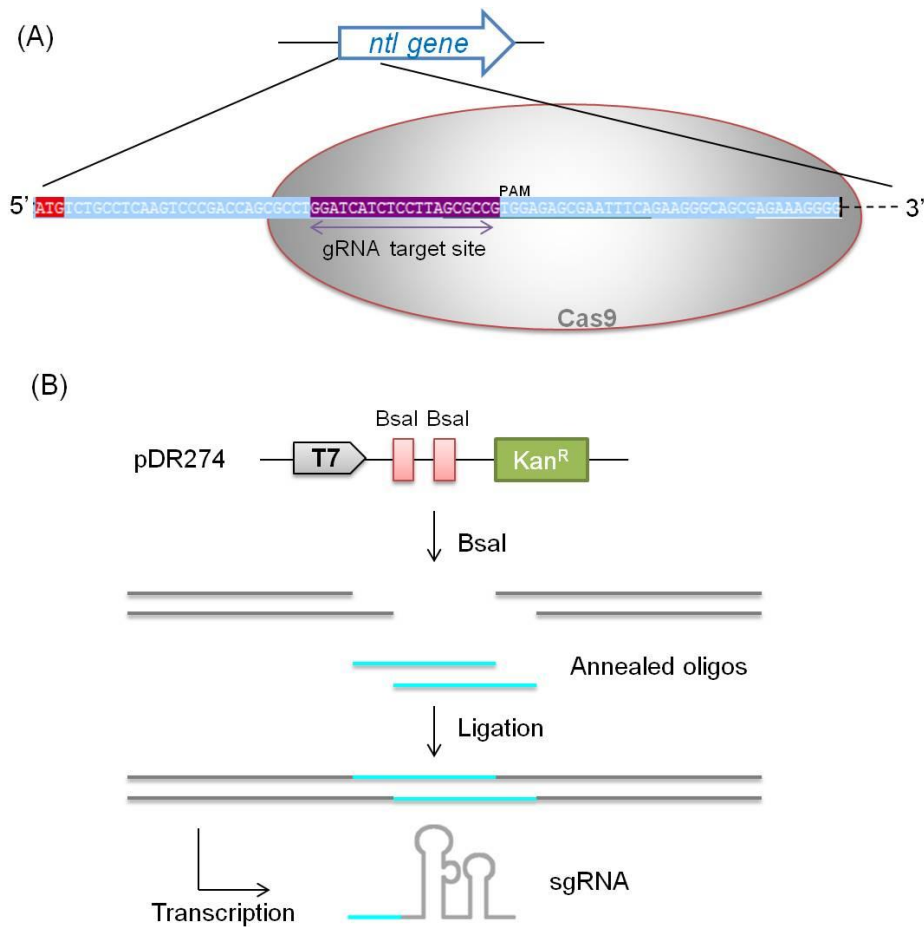
Figure 21: The CRISPR/Cas system method targeting the *ntl* gene. (A) Representation of the target sequence of the gene *ntl* (gRNA in purple) located in the exon 1 after the ATG (red) and directly followed by the PAM sequence (TGG). The Cas9 protein is shown in grey. (B) A scheme showing the pDR274 expression vector containing the T7 promoter and the kanamycine resistance (Kan[R]) with 2 recognition sites of the endonuclease BsaI. After digestion with BsaI, the annealed oligos are ligated into the vector and the sgRNA is transcribed *in vitro*.

In spite of the preceding observations, a deeper analysis at the DNA level had to be done to confirm our suspicions. In the case of the CRISPR method, unfortunately no restriction site analysis, or at least rarely, could be done because of the restricted target site design which does not give the freedom of including a restriction site at the supposed cut region. To

overcome this difficulty, the T7 Endonuclease I (T7EI) assay that was also used by Hwang *et al*., was applied to detect mutations at the targeted sequence. First, the genomic DNA was amplified with primers flanking the target sequence. The resulting PCR products were then denatured and re-annealed, giving rise to different structures: homoduplexes with or without modifications and heteroduplexes where the mutations are revealed as mismatches, which will be recognized and cleaved by the T7 endonuclease. The fragments would then be separated by electrophoresis and the targeting efficiency analyzed on the gel.

The DNA targets of 10 single embryos from three different injections and two wild-type embryos (uninjected) were tested with this method, but no mutations were detected. Few of them were still sub-cloned into pGEM T easy vector and sequenced to definitely confirm the absence of any mutation at the target DNA sequence. Indeed, none was observed, suggesting that no mutations occurred at the target region of the *ntl* gene. A second designed gRNA against *ntl* was also tested, leading to the same observations.

Considering that other laboratories had also difficulties to obtain positive results with the same Cas9, another protein, from Hruscha *et al*., was obtained and tried. The difference with the previously used protein was that the wild type Cas9 of *S. pyogenes* was modified by adding two NLS at both ends and recombined into a pCS2 vector harboring the SP6 promoter (Hruscha et al. 2013). We considered that two NLS at the beginning and the end of the protein would probably optimize its nuclear transport and thereby its efficiency.

The same gRNA targeting the *ntl* gene was injected into one cell stage embryos together with the newly used Cas9 mRNA, at 2.4 µg/µl and 0.5 µg/µl respective concentrations, according to the proposed protocol (Hruscha et al. 2013). As to be expected from the high injected amount of mRNAs, a lot of malformed embryos were observed, but no specific *ntl* phenotype could be detected. The T7 endonuclease assay and sequencing analysis were performed but unfortunately without any success.

Afterwards, a synthetically developed Cas9 from Chen's group was successfully used in zebrafish and we decided to test its efficiency in our laboratory (Jao et al. 2013b). The coding sequence of the Cas9 of *S. pyogenes* was synthesized with codon optimized usage for zebrafish and two SV40 large T-antigen NLS were added at the N-terminal and the C-terminal regions. This sequence was cloned into pT3TS vector carrying the T3 promoter and consequently named pT3TS-nCas9n (Addgene, plasmid #46757). Following the given

protocol, the *in vitro* transcribed mRNA of both T3-Cas9 and the gRNA against the *ntl* gene were injected into one cell stage zebrafish embryos with the following concentrations: 12,5 ng/µl (~25 pg) of the gRNA and 75 ng/µl (~150 pg) of the Cas9 mRNA.
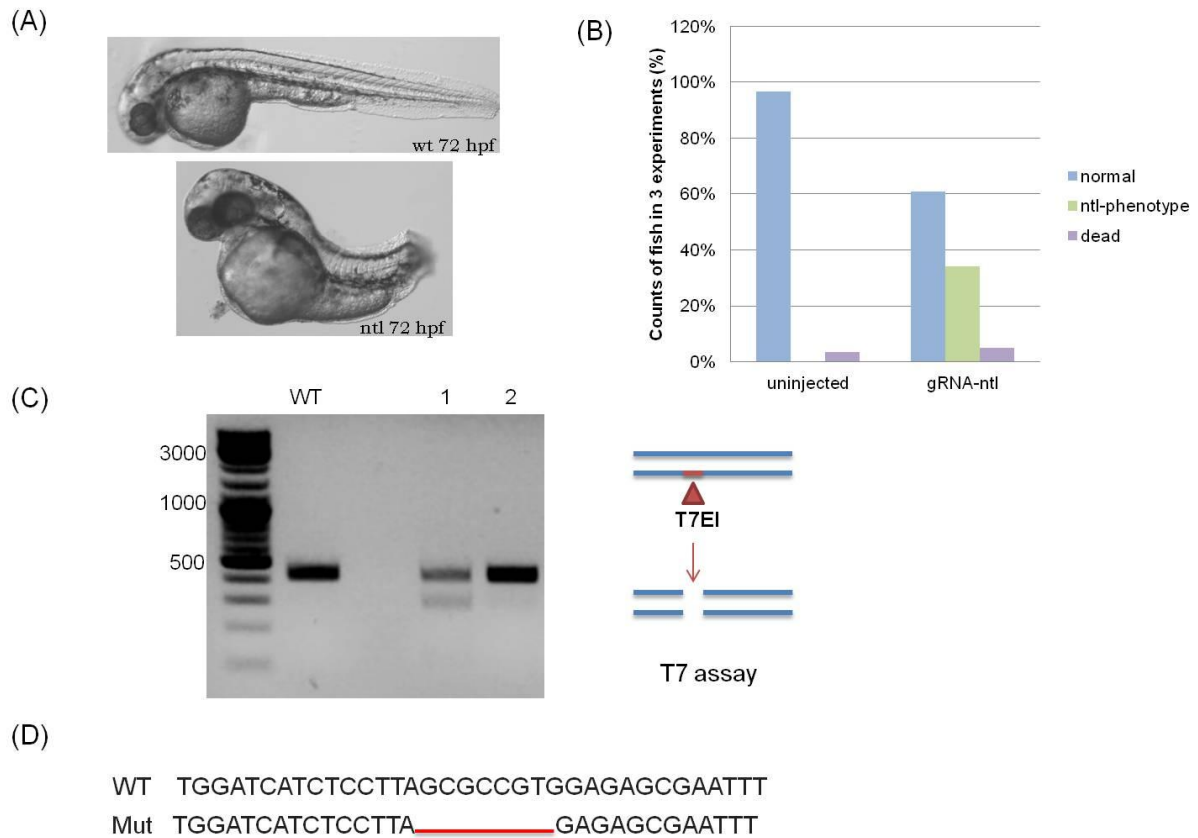

Figure 22: Disruption of the *ntl* gene with the CRISPR system. (A) At 72 hpf, a *ntl* phenotype is shown in the lower panel with the lack of the posterior part, compared to a wild type uninjected embryo (wt). (B) Representation of the percentage of counted embryos injected with gRNA against the *ntl* gene compared to uninjected embryos. Observations at 24 hpf of the embryos under stereomicroscope in 3 independent experiments indicate the presence of a *ntl* phenotype in about 35% of the fish (green). The counts of normal looking and dead embryos are also represented (blue and purple, respectively). (C) Gel electrophoresis showing the amplified target sequence of 426 bp from wild type and injected single embryos, subjected to the T7 assay. A cut band representing an indel recognized by the T7EI is observed in the DNA of the embryo 1, whereas just the whole amplified sequence is observed in the wt uninjected embryo and a non mutated embryo 2. (D) An example of a sequencing result observed at the target site with a 7 bp deletion (red), compared to a wt control.

Observations after 24 hpf showed the long-awaited specific *ntl* phenotype with the disruption of the posterior part, as previously described. Observations in three independent experiments showed that the *ntl* phenotype was detected in approximately 35% of the injected embryos (Figure 22B). Further analyses were performed on amplified fragments from the genomic DNA of single embryos using the T7 endonuclease assay. The results showed indeed

additional bands corresponding to a cleavage at the target site (Figure 22C). The PCR products from positively cleaved DNA fragments were sub-cloned into pGEM T easy vector and sequenced to examine the mutation types. Indeed, insertions and deletions were observed at the target region (Example of a deletion of 8 nt in Figure 22D compared to the wt).

After trying different protocols and Cas9 proteins, the CRISPR system was finally successful in our hands and could be used easily to target several genes of interest.

3.3. Targeted-modification of genes implicated in zebrafish neurogenesis
3.3.1. Id1 disruption using the TALEN technology

A stab wound assay was established few years ago in our laboratory in which zebrafish adult brains were stabbed in one hemisphere using a needle, while the opposite hemisphere was kept intact as a control (März et al. 2011). After the identification by RNA-seq of 279 TRs that are misregulated upon injury, another expression screen by *in situ* hybridization was performed in order to identify the regulators that are specifically expressed in the VZ of the telencephalon (for possible regulation of the NSCs). More than 60 genes encoding TRs were upregulated in the VZ following brain injury and among them the HLH DNA-binding inhibitor: Id1 (Viales et al. 2015). Loss and gain of function studies, with *vivo* morpholino and DNA lipofection in the telencephalon were previously performed and suggested a potential role of *id1* in neurogenesis and brain regeneration of the zebrafish by maintaining the quiescence of the NSCs (Viales et al. 2015). An in-depth study on the function of this gene could reveal a precise understanding of the molecular mechanisms involved in the processes of adult neurogenesis. However, further confirmations and dissections are needed to draw definitive conclusions about the role of *id1*. Against this background, the adoption of TALENs as a new gene-targeting tool to be used in zebrafish for creating a knock-out mutant for *id1* naturally emerged as a viable option.

TALENs against *id1* were designed in the exon 1 (Figure 23A) with the Mojo Hand online tool and constructed according to the Golden Gate method to finally be cloned into the pGoldyTALEN backbone. Both left and right mRNA TALENs were transcribed *in vitro* under the T3 promoter and injected with an equal amount (100 ng/μl) into one-cell-stage zebrafish embryos. TALENs against the *ntl* gene were injected in parallel as a positive control. The embryos were examined at 24 hpf, but no observable phenotype was detected in

the *Id1*-TALENs injected fish, even at later points of development, while the *ntl* phenotype was clearly observed already at 24 hpf.
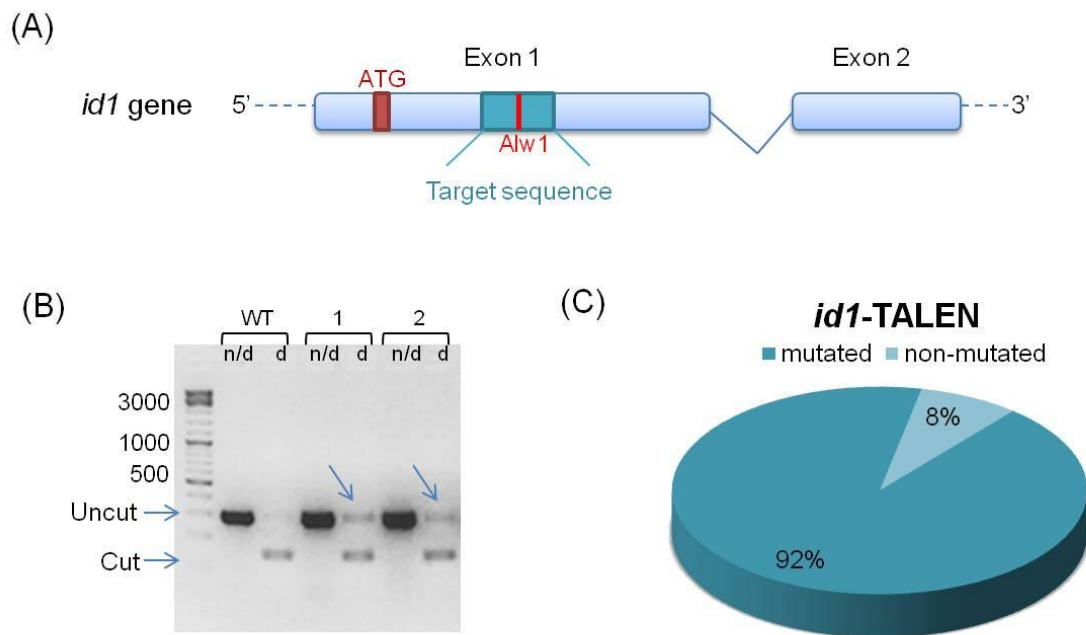

Figure 23: *Id1* targeting with TALENs. (A) Schematic representation of the *id1* gene with the target sequence for TALENs in the exon 1 and the restriction enzyme AlwI at the expected cut site. (B) Loss of restriction site analysis shown on gel electrophoresis after digestion with the enzyme AlwI of the amplified fragment of 300 bp from the *Id1* sequence (n/d). Two overlapping bands of 150 bp (cut) result from the digestion (d) and in TALEN injected embryos, an uncut band remains at 300 bp (1 and 2, uncut). (C) Statistical analysis of the rate of mutated embryos in three different experiments, with 92% positively mutated fragments.

To assess the efficiency of the *Id1*-TALENs, the loss of restriction site analysis was as well performed, as the recognition site of the enzyme AlwI was present at the spacer region. Genomic DNA was then extracted from 10 single embryos at 3 dpf and the target region was PCR amplified with the designed primers. After submitting the fragment to AlwI digestion, the amplified fragment of 300 bp was cleaved resulting in two overlapping fragments of 150 bp (Figure 23B). In wild type embryos, the cut bands were clearly observed while in injected embryos with *Id1*-TALENs a remaining uncut band was detected (Figure 23B, 1 and 2). The loss of restriction site was observed by gel electrophoresis in more than 90% of the tested embryos, suggesting the presence of mutations at the recognition site of the enzyme (Figure 23C).

In order to confirm this result, positive fragments were sub-cloned into pGEM T easy vector in order to sequence the DNA of interest, and more than 50% of the sequenced clones (n=14)

exhibited indels at the desired sequence (insertions and deletions of different sizes). An example is shown in Figure 24A with a deletion of 13 bp in the mutated sequence compared to the wild type. The line followed to predict the loss of function of a mutated Id1 protein was to focus on the HLH motif and check if the mutations would disrupt or truncate the activity domain of Id1. This was done *in silico*: i.e. firstly by translating the mutated sequence of the gene, variable numbers of amino acids are obtained for each different mutation. Then, the peptide sequence was subjected to the pfam database (http://pfam.xfam.org/) from the European Bioinformatics Institute (EBI) (Finn et al. 2014) in order to predict the loss of the active domain in the mutated protein. The best obtained mutation created an early stop codon in the *id1* sequence and gave rise to a short peptide of 40 aa, which should lack the HLH domain, thus suggesting loss of activity of the Id1 protein (Figure 24B).
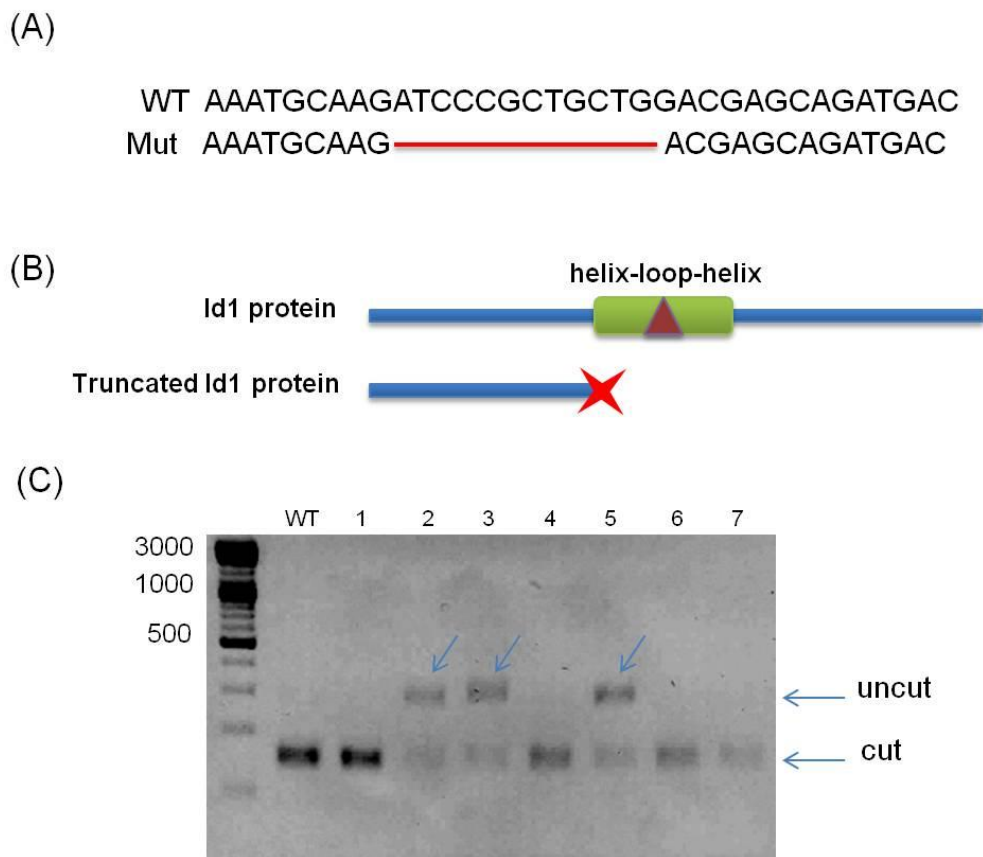
Figure 24: Mutation analysis at the *id1* target site and evidence of germline transmission. (A) An example of a sequenced pGEM T clone containing a positively mutated fragment, with a deletion of 13 bp (Mut), compared to a wild type control (WT). (B) Prediction of the consequence of mutations in the *id1* gene with the loss of the HLH active domain. (C) Loss of restriction site analysis of fragments amplified from gDNA of fish from the F1 generation. Compared to the wt with only the overlapping cut bands of 150 bp, the positively mutated fish contain an uncut band of 300 bp (2, 3, 5).

As there were evidences of successful germline transmission of mutations caused by TALENs in zebrafish (Huang et al. 2011), the founders F0 fish injected with *Id1*-TALENs were raised. The grown-up fish were outcrossed with wild-type so that the mutations are isolated. The genomic DNA was then extracted from the progeny at 3 dpf (single embryos) and analyzed with the same loss of restriction procedure previously used. While in wild type embryos the amplified fragment was completely digested by the enzyme AlwI, in the progeny of the mutated *Id1* fish the remaining uncut band was still observable in 3 out of 7 tested embryos (Figure 24C). Mutations were indeed identified in a number of batches of this F1 generation, suggesting that the mutations were efficiently transmitted to the germline in 7 out of 10 founders. The progeny was raised according to the appropriate conditions (cf. Material and Methods).
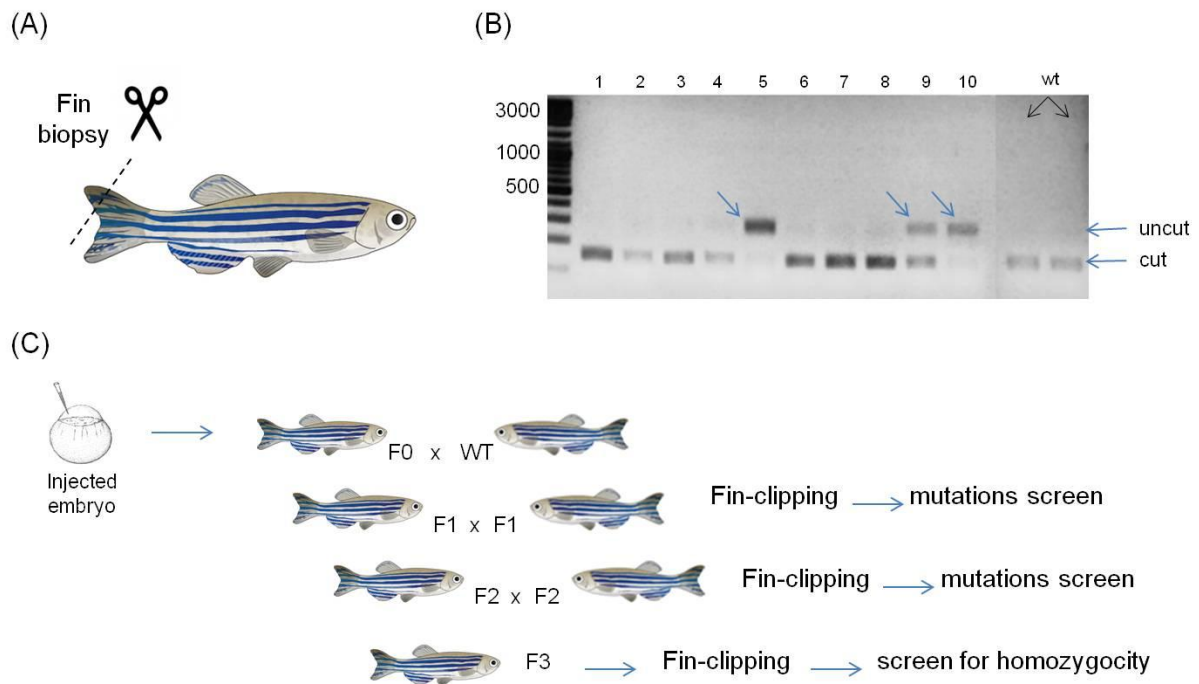
Figure 25: Adult fish genotyping. (A) Fin biopsy of the zebrafish consists of removing a part of the fish fin in order to test the genomic DNA of individuals. (B) Gel electrophoresis showing digestion of DNA from adult F1 generation fish. The cut bands of 150 bp represent the digestion with the enzyme AlwI of the amplified fragment. The uncut band of 300 bp represents the remaining band with loss of restriction site (5, 9, 10). (C) Fish crossing scheme since injection of the embryos with Id1-TALENs until obtaining of the F3 generation carrying homozygous fish.

In order to identify the F1 fish that inherited mutations caused by TALENs at the *id1* gene, genotyping was performed using genomic DNA extracted from fin clips (Figure 25A). This procedure consists in cutting a slice of the fish fin, according to the ethical laws for

anaesthesia and fish handling, and extracting the genomic DNA from it in order to discriminate the mutation carriers. The DNA from each fin was isolated and purified, and the target sequence was then amplified by PCR. Digestion with the restriction enzyme AlwI was performed and the results were observed on gel electrophoresis: the resulting bands are shown in Figure 25B. Similarly to the previously described cases, in the wild type situation (2x wt) all fragments were cut, whereas in some mutated fish a distinct uncut band was observed (mutants 5, 9, 10), reflecting the presence of mutations at the target site.

In this way, the mutated fish were separated from the non mutated ones. They were then in-crossed in order to obtain an F2 generation carrying mutations (Figure 25C). When the embryos reach adulthood, their genome was again tested by fin biopsy and the type of the mutation was assessed by sequencing (after sub-cloning into pGEM T vector), in order to verify the type of mutations leading to the inactivation of the protein Id1.

Following the scheme used for the crossing strategy (Figure 25C), the mutants of the F2 generation were again in-crossed and an F3 generation was obtained to get with high probability homozygous fish. To identify these knockout fish, the same fin biopsy procedure has to be performed on the adult fish in order to sequence the amplified fragments and obtain the mutation profile in all the cells. On the other hand, my observations of the progeny at early stages did not clearly show lethality as an effect of the lack of *id1*. At the present time, these ongoing experiments did not yet give the expected results, however more fish have to be tested to draw any kind of conclusion.

3.3.2. Id3 disruption using TALENs and CRISPR/Cas9 methods

Id3 is, like Id1, a member of the HLH family of TRs. Former studies have shown compensatory functions of genes of the *id* family, especially the role of *id1* and *id3* is flagrantly redundant during mouse brain development. According to the study performed in 1999 by Lyden et al. in which *id1* and *id3* were mutated, only the double knockout mice died. Indeed, the presence of at least one copy of *id1* or *id3* prevented embryonic lethality and no distinguishable phenotype was observed in this case (Lyden et al. 1999). Later on, Niola et al. performed studies on *id* genes by generating a conditional triple knockout of *id1*, *id2* and *id3* that led to neonatal lethality. The loss of the Id proteins also disrupted proliferative capacity and self-renewal of the NSCs resulting in premature differentiation (Niola et al. 2012). Moreover, in a recent work of our group on differential expression of the *id* genes in the adult

brain, it was shown that four out of the five genes are expressed in the adult telencephalon: *id1*, *id2a*, *id3* and *id4* (Diotel et al. 2015). In addition, the zebrafish *id* genes were also shown in the embryo to have partially overlapping expressions, which can be correlated with the absence of an observable phenotype in the *id1* mutated fish. It could be thus explained by a compensation of the function of *id1* during development from the other *id* genes.

Firstly, a pair of TALENs targeting the first exon of *id3* was designed using the Mojo Hand online tool (Figure 26A). The RVDs corresponding to left and right binding arms of 16 and 17 nt, respectively, were assembled. The spacer region of 15 nt contained the recognition site of the NruI restriction enzyme to be used for the identification of mutations. After *in vitro* transcription, the mRNAs of both TALENs were injected into one-cell stage embryos.
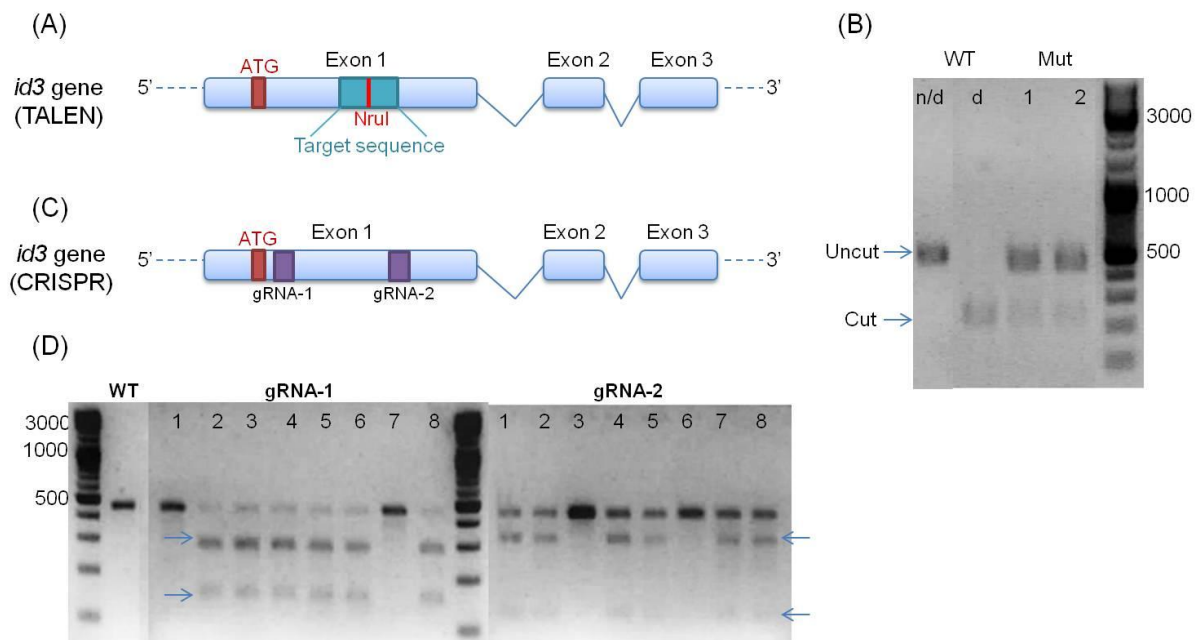
Figure 26: *id3* targeting with TALENs and CRISPR. (A) Schematic representation of the *id3* gene with the target sequence for TALENs containing the NruI restriction enzyme. (B) Loss of restriction site analysis shown on gel electrophoresis with the amplified fragment of 440 bp (n/d) digested with NruI (d), giving rise to two approximately overlapping bands of about 230 and 210 bp. The positively mutated fish 1 and 2 contain additionally an uncut band. (C) Schematic representation of the *id3* gene with the target sequences for two gRNAs (gRNA-1 and -2). (D) T7 assay on the amplified DNA fragments. For both gRNAs-1 and -2 cut bands are observed at the right size, compared to the wt control that has no cut fragments.

At 3 dpf, the genomic DNA was extracted from single embryos and a targeted fragment of 440 bp was amplified by PCR. The fragment was then suggested to restriction digestion with NruI and positive uncut bands were observed on gel electrophoresis, in addition to the two cut

bands of approximately 230 bp and 210 that are almost overlapping (Figure 26B, 1 and 2), compared to the wild type only cut bands. Thus, suggesting the efficiency of TALENs to create mutations in the *id3* gene.

In parallel, two gRNAs against *id3* were designed in the exon 1 (Figure 26C, gRNA-1 and -2). The transcribed mRNAs were injected individually, together with the T3-Cas9 mRNA, into one-cell stage embryos. The genomic DNA from single injected embryos was extracted and the fragment of 440 bp containing the targeted sites was amplified using the same pair of primers that was used for the TALEN experiment. In order to assess the efficiency of the gRNAs, the T7 endonuclease I assay was performed. The results shown in figure 26D suggest the really good efficiency of both gRNAs. Indeed, compared to the wild type control (Figure 26D, WT) in which one band of the aforementioned size was observed, digested bands are present in the case of the injected embryos with the gRNA-1 and the gRNA-2 against *id3*. The first targeting gRNA-1 was expected to give rise to two fragments of approximately 320 bp and 120 bp in size after digestion with the T7EI endonuclease that cuts at mutations sites. Our observations on the gel electrophoresis correspond to these expectations for 6 out of 8 tested embryos (Figure 26D, gRNA-1). In a similar way, the expected sizes of the fragments resulting from the targeting gRNA-2 and after digestion with T7EI would be of 155 bp and 285 bp approximately. Also in this case, the observed cut bands on gel electrophoresis are in line with the expected resulting fragments, and this for 6 out of 8 embryos (Figure 26D, gRNA-2).

In order to confirm the presence of mutations at the target site and to find an alternative to laborious screening for the CRISPR mutations, another method was tested. The amplified PCR product was directly sequenced in order to seek for mutations simply by investigating the chromatogram of the resulting DNA sequence. In practical terms, the sequence profile (AB1 format) of a wild type fragment was compared to the sequence of possibly mutated fragments. A good sequence is generally characterized by evenly-spaced peaks of four different colors corresponding to the nt and the lack of baseline "noise" (see figure 27A and C, representing the wt fragment). In the case of a mutated fragment, the sequencing chromatogram starts changing at a point; the noise increases and the peaks are disordered (see figure 27B and D). Naturally, the most important observation is that the change in the sequencing profile appears in the vicinity of the PAM in which the Cas9 is expected to cut the DNA (Figure 27A' and C'). This method was tested for different fragments and it was

validated as a good alternative to assess the efficiency of the CRISPR/Cas9 system in a relatively quick manner. In addition, the use of the T7 assay showed some discrepancies from an experiment to another, due most likely to the enzyme itself. However, in order to identify the type of mutations occurring at the targeted site, sub-cloning into pGEM T easy vector cannot be avoided. In this context, one fragment for each targeting gRNA was cloned and after colony PCR, some positive clones were sequenced and deletions of different sizes were observed in both fragments (Figure 27B' and D').
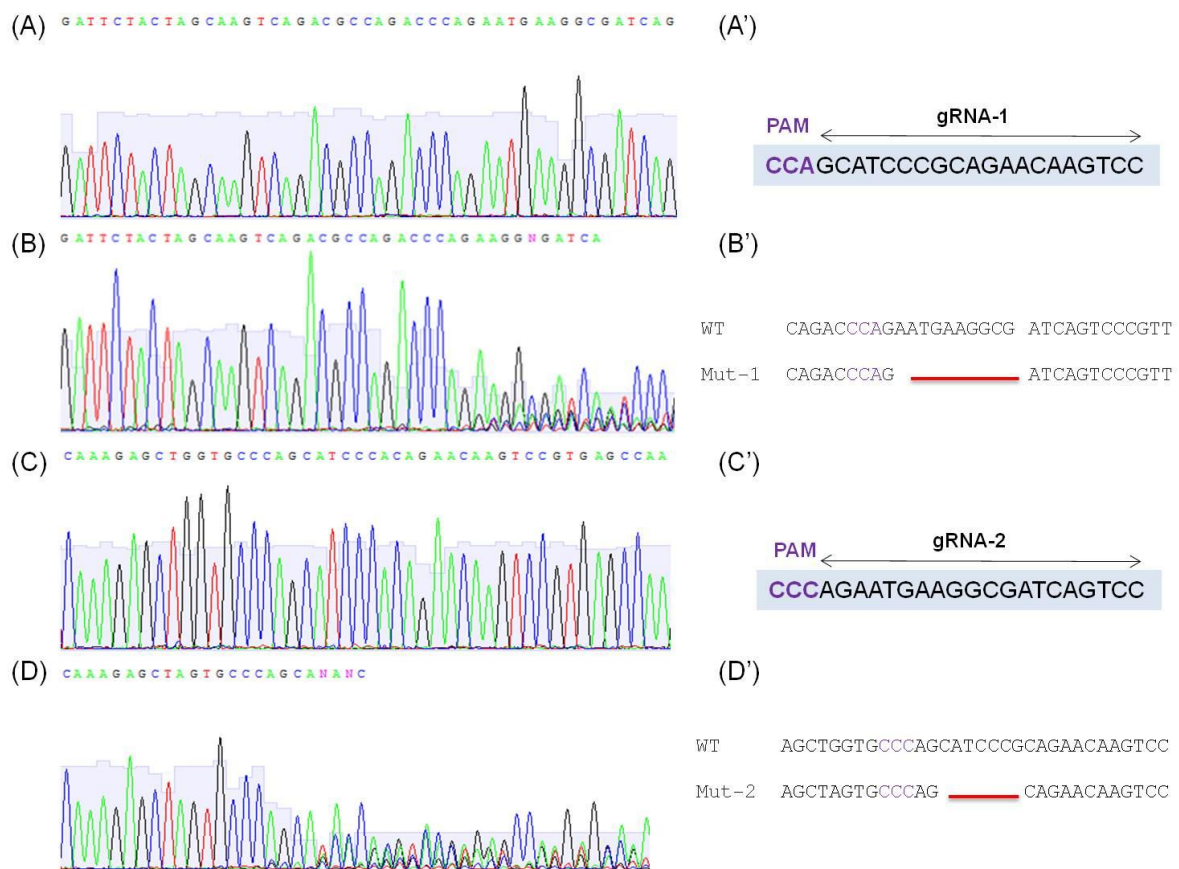
Figure 27: Identification of CRISPR mutations by sequence chromatograms. (A) Sequence chromatogram of the wt fragment at the target site of gRNA-1 shown in A'. (B) Sequence chromatogram of a mutated sequence with higher noise around the expected cut site around the PAM. (B') Sequencing result of a cloned fragment with a deletion of 10 bp at the target site, compared to a wt sequence. (C) Sequence chromatogram of the wt fragment at the target site of gRNA-2 shown in C'. (D) Sequence chromatogram of a mutated sequence with higher noise around the expected cut site around the PAM. (D') Sequencing result of a cloned fragment with a deletion of 7 bp at the target site, compared to a wt sequence.

Our experiments indicate a really good targeting of the *id3* gene with two different methods, TALENs and CRISPR, even though the efficiency and the time needed shifts the balance towards the CRISPR system for more simplicity and timeliness.

3.3.3. Sox1 gene disruptions using the CRISPR/Cas9 technology in a multiplexed manner

The genes *sox1a* and *sox1b* are closely related genes that might characterize a new subtype of interneurons in zebrafish: V2c in the ventral spinal cord, which was identified in mouse in 2009 by Panayi et al.. Previously performed expression analyzes by Yang and colleagues showed that a proportion of *sox1a/b* expressing cells are different from V2b interneurons as they just partially co-express the same markers. Also, an increase in V2b cells was observed when *sox1a* and *b* are downregulated by MOs knockdown, suggesting that these two genes might be cell fate regulators of the V2c interneurons. Hence, further genetic studies have still to be performed in order to confirm these observations and to draw conclusions about cell fate determination in the ventral spinal cord and the molecular mechanisms implicated in the specification of V2c interneurons.

To that purpose, two gRNAs fulfilling all the criteria were designed in the only exon of each gene, *sox1a* and *sox1b*: gRNA-A1, gRNA-A2, gRNA-B1, gRNA-B2 (Figure 28A). Each one of the gRNAs was tested individually by injection into one-cell stage embryos together with the T3-Cas9. No striking phenotype was observed in any of the injected embryos. However, the T7 endonuclease assay analyzes were performed for each of the injected gRNAs. For each gene, one gRNA was identified as functional, the gRNA-A2 and the gRNA-B2, for sox1a and sox1b respectively (Figure 28A).

The two efficient gRNAs were then co-injected with the T3-Cas9 mRNA in order to target both genes simultaneously. After extraction of genomic DNA from single embryos, both fragments from *sox1a* and *sox1b* were amplified and the T7 assay was performed. Observations on the gel electrophoresis showed the presence of cut bands in a number of injected embryos, compared to the wild type and two embryos injected only with the T3-Cas9 mRNA. Moreover, mutations were indeed detected in either target sequences and in the same tested embryo (Figure 28B, 5). Both fragments from the positive embryo were sub-cloned into pGEM T easy vector to be sequenced. Sequence analysis confirmed the presence of mutations

at the targeted region for both genes: 32 nt deletion in *sox1a* and 4 nt deletion in *sox1b*, both leading to shortened proteins. Actually, our observations conclude that the multiplex gene targeting is possible in zebrafish with two different gRNAs targeting two different genes.
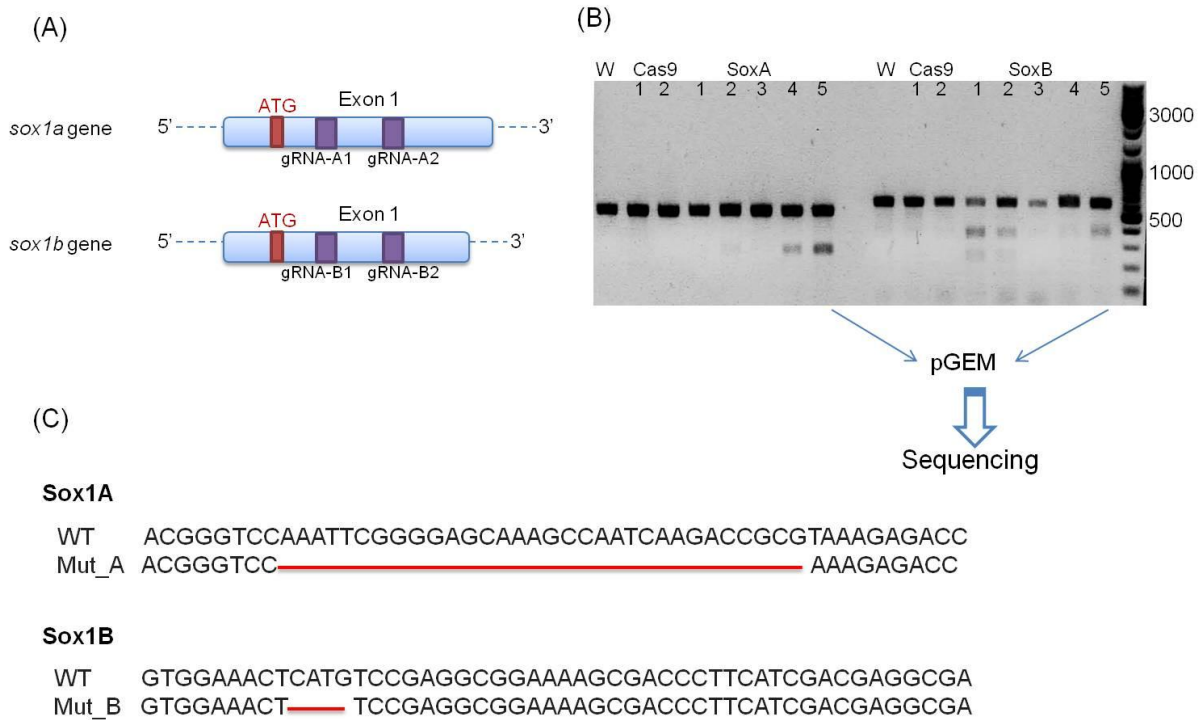
Figure 28: Targeting sox1a and sox1b with CRISPR. (A) Schematic representation of the *sox1* genes with the target sequences for two gRNAs for each (gRNA-A1 and -A2, gRNA-B1 and gRNA-B2). (B) T7 assay on the amplified DNA fragments from embryos co-injected with both gRNA-A2 and gRNA-B2. Cas9-1 and 2 represent embryos injected only with T3-Cas9 mRNA. For both *sox1a* and *sox1b* genes, cut bands are observed at the right size, compared to the wild type control that has no cut fragments (W) and to the Cas9 negative controls. The injected embryo 5 shows mutations for both genes and the fragments derived from this embryo were sub-cloned into pGEM T easy vector to be sequenced. Examples of sequencing results show a 32 bp deletion in *sox1a* and 4 bp deletion in *sox1b* (red).

The injected embryos with both gRNAs were raised under controlled conditions and at adult stage, fin biopsies were performed in order to discriminate the positively mutated fish. After extracting the genomic DNA from the pieces of each fin, the T7 assay was performed and a number of positively mutated fish in both genes were identified. An example is shown in figure 29A, in which the fish 5 and 6 show indeed cut bands in both *sox1a* and *sox1b* fragments, expressing the presence of mutations at the target sites. The positive identified fish were isolated and some of them were outcrossed with wild type fish in order to obtain the F1 generation.
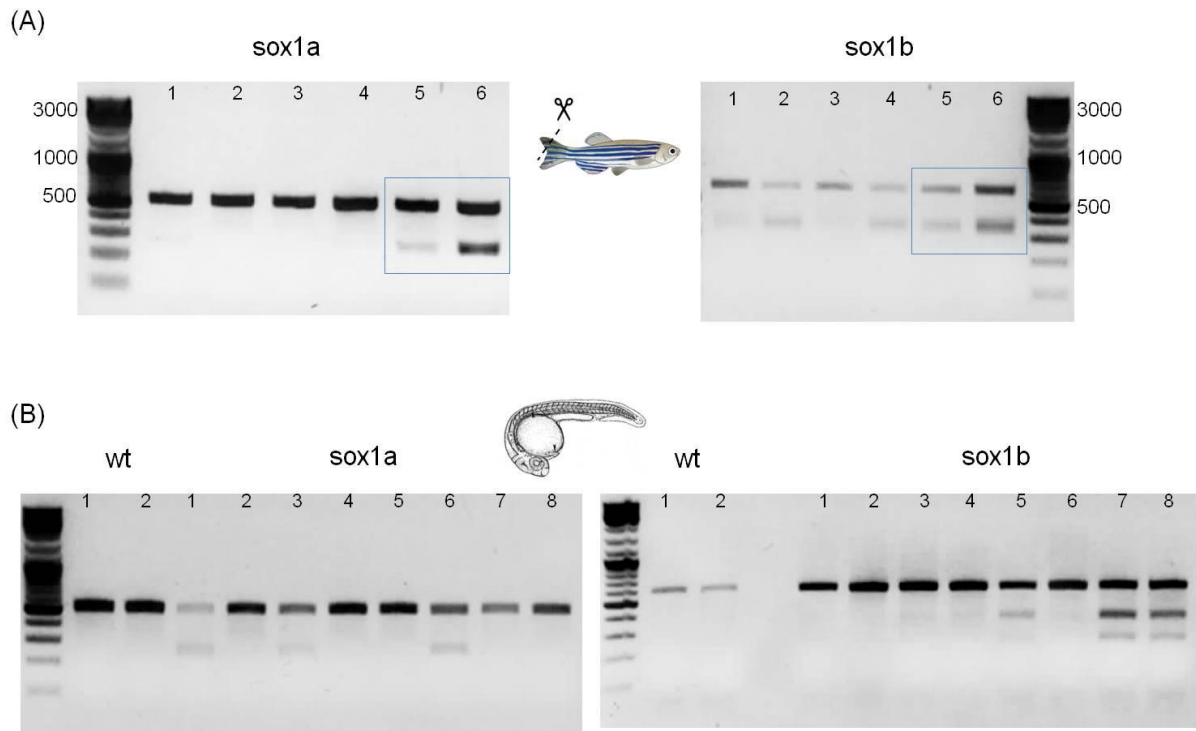
Figure 29: *sox1a* and *sox1b* mutations in adults and germline. (A) T7 assay performed after fin biopsy on adult fish raised after injection of both gRNAs against the sox1 genes. Both 5 and 6 fish show the presence of mutations in both genes (blue frame). (B) T7 assay performed on embryos from the F1 generation derived from outcrossing. Cut bands suggesting mutations are observed in both genes (sox1a 1, 3 6 and sox1b 5, 7, 8), compared to the wild type controls (wt, 1 and 2).

To the purpose of verifying germline transmission of the mutations in the *sox1* genes, embryos from the F1 generation were tested by T7 assay. Indeed, cut bands were observed in some of the embryos compared to the wild type fragments, suggesting the effective transmission of mutations (Figure 29B: 1 and 6 for sox1a, 7 and 8 for sox1b). However, at that point no tested embryo contained mutations in both genes.

The obtained results on the multiplexed targeting of both *sox1a* and *sox1b* genes open the door to expanded investigations on related genes that could have redundant functions. The knockout of these genes could give the opportunity to follow their effect on the development of neurons in the spinal cord, especially the different sub types of interneurons that we are interested in.

Chapter IV: DISCUSSION

Reverse genetics in zebrafish revolutionized both the way of performing functional studies on genes and the understanding of their implication in different molecular mechanisms in the cell. First with TILLING and antisense morpholinos, a great number of genes were identified for their role in vertebrate development (Moens et al. 2008; Knowlton et al. 2008). Afterwards, the development of gene targeting approaches in zebrafish opened the door for further specific studies, as almost any sequence of the zebrafish genome can be modified with TALENs or CRISPR methods (Hisano et al. 2014).

4.1. Efficient implementation of the TALEN technology for the disruption of targeted genes implicated in neurogenesis in zebrafish

TAL effector nucleases are fusion proteins containing the DNA binding domain of TAL effectors derived from a bacterial plant pathogen and the catalytic domain of the endonuclease FokI. The flexible design of TAL binding domains allows the targeting of any desired sequence in the DNA that is then cut by the FokI protein creating DSBs. Hence, the DSBs are consequently repaired via an error-prone pathway, i.e. NHEJ creating indels at the target sites or by HR if a template is available.

At the beginning of this project, less was known about the routine use of the TALEN technology. First of all, I successfully set up and tested the method within our laboratory. A library of modules containing the different RVDs and backbones developed by the Voytas group was used (Cermak et al. 2011) and the assembly of designed TALENs was performed. The Golden Gate assembly method was used to construct the custom TALENs (Engler et al. 2009). As for my experience, some issues can occur during the procedure, especially concerning the library of modules that has to be carefully manipulated and organized. A good positive control was necessary in order to test *de novo* TALENs and to that purpose the *ntl* gene with its clear phenotype was a perfect candidate. In fact, my first success with the TALEN method required the correct assembly of desired RVD modules using the Golden Gate method to target the *ntl* gene and the use of the pGoldyTALEN backbone containing a truncated version of the TAL cassette (Bedell et al. 2012). Approximately one week is needed to obtain a ready-to-use TALEN. After assessing the mutations at the target site using loss of restriction site analyses and sequencing, I could attest the efficiency of my custom TALENs.

After demonstrating that the TALEN technology can be successfully used to disrupt chosen zebrafish genes, my first preferred gene that was targeted was *id1*. Id1, the DNA-binding protein inhibitor 1, is part of the helix-loop-helix family of TRs that are dominant negative regulators of basic HLH TFs. It was shown to be involved in adult neurogenesis and brain regeneration, most likely by maintaining the neural stem cell pool by regulating the balance between quiescence and proliferation of NSCs, joining some findings in the adult mouse (Viales et al. 2015; Jung et al. 2010; Niola et al. 2012). TALENs against *id1* successfully induced mutations at the target sequence resulting in several types of indels (deletions of 4 to 13bp, insertions of 14 or 27 bp). The efficiency rate of mutations that I obtained reached up to 90% of tested embryos, which is within the range of published results using TALENs in zebrafish (Dahlem et al. 2012; Bedell et al. 2012; Ma et al. 2013). However, the TALEN-mediated mosaic mutations make it difficult to predict an effect on a given gene. Moreover, our results suggest that the efficiency may vary from one locus to another and possibly the different indel sizes can also have an effect, as for non-trinucleotide mutations that lead to frameshifting. I could as well assume that individual TALENs could have different activities (architecture, context dependency…).

The mutations are effectively inheritable, thus different transgenic lines were obtained by mating the TALEN-injected founder fish with wild type animals. Indeed, 70% of those mated couples gave rise to a positively mutated F1 progeny, attesting of the genetic inheritance of the alterations. After identification of positive mutation carriers, F1 fish were in-crossed in order to obtain the F2 generation of mutated fish. DNA sequencing and domain prediction platforms were used to identify the type of mutations that can lead to truncated proteins that lack the HLH domain that is the active protein-protein interaction domain of the Id proteins. In this context, the last obtained F3 generation of fish should most probably contain mutants with homozygous mutations.

Obtaining a knockout *id1* mutant will allow further studies to be done in order to understand the molecular mechanisms in which *id1* is involved. For instance, the observed results after *vivo*-morpholino injection can be strongly confirmed in a mutant. In addition, the stab wound assay can be performed in adult fish brains lacking Id1 to observe the effect on the neuronal regeneration occurring in the zebrafish telencephalon.

During the current study, I did not observe any phenotype in the zebrafish embryos mutated with *id1*-TALENs. A reasonable explanation was either a maternal mRNA contribution that

was lightly shown for this gene in zebrafish (Harvey et al. 2013), or a possible redundancy of other *id* genes that can compensate the role of *id1* during development. In mouse, it was indeed shown that knocking out *id1* or *id3* separately has no observable effect on the brain and no significant developmental defects, while the double knockout of both *id1* and *id3* genes leads to a smaller brain size and embryonic lethality (Lyden et al. 1999). Furthermore, other studies using conditional triple knockout mice for *id1*, *id2* and *id3* led to neonatal lethality and showed a defect in proliferation and self-renewal of NSCs (Niola et al. 2012). Also, in order to determine possible interferences between the *id* genes, other members of the Id family were added to our list of target genes in this study. Hence, I used TALENs against the *id3* gene and attested of their efficiency according to the presence of mutations at the target site. I also successfully demonstrated the germline transmission of mutations in the *id3* gene. Future plans include also obtaining a knockout fish line for the *id3* gene, making it possible to cross both *id1* and *id3* mutant lines in order to get double knockout fish allowing the study of possible effects that can have both genes on their respective functions. Meanwhile, a recent study of our group was performed by *in situ* hybridization and showed the expression of other *id* genes in the adult zebrafish telencephalon (Diotel, Beil, et al. 2015). It was determined that, beside *id1*, *id2a* is also expressed in type 1 and type 2 neural progenitors in addition to its expression in type 3 cells. Thus, *id2a* makes a promising future candidate in studying the role of the *id* genes in adult neurogenesis and brain regeneration in the zebrafish.

Furthermore, a possible lethality caused by the knockout of *id1* could not yet be ruled out even if it was not observed at embryonic stages. Therefore, more statistical analyses should help to determine if the complete lack of id1 can indeed cause the premature death of the fish.

4.2. Efficient implementation of the CRISPR technology for use in zebrafish and disruption of targeted genes implicated in neurogenesis

During the progress of using the TALEN method within my project, another DNA editing technology has emerged: the CRSPR/Cas9. This method is based on the design of short single guide RNAs that target a specific sequence in the DNA and recruit the endonuclease Cas9 which induces DSBs. Like for the TALENs, the DSBs will be repaired by NHEJ thus creating indels in the target sequence or by HR if a DNA template is present. This attractive technology has fast aroused a great interest, especially among the zebrafish community.

Indeed, taking into account the simplicity and the rapidity of designing and constructing a sgRNA compared to a TALEN, it was clearly a good opportunity to be taken. In this sense, we decided to establish this method in our laboratory, as it could allow the disruption of a great number of targets in a faster way and the investigation of a large number of DNA sequence variants when using different gRNAs to target a single gene. At first, I successfully tested the CRISPR method using the *ntl* gene as a control considering that it was conveniently used for TALENs.

Our interest in the CRISPR technology had grown even more with the possibility to use the cited method in a multiplexed manner in order to knockout several genes in the same time (Jao et al. 2013a). Having in mind the disruption of the different *id* genes, several gRNAs were designed in order to be tested. First tested individually, gRNAs against *id3* were successfully used with an efficiency up to 80% of the analyzed embryos. However, the preliminary results concerning *id2a* and *id4* were inconclusive and broader experimental expertise is needed in order to select efficient gRNAs against these genes.

In parallel, I was interested in two related genes, *sox1a* and *sox1b*, that have a role in the specification of interneurons in the spinal cord. Indeed, *sox1a* and *b* were detected in the systematic screen for TRs that was performed in 24 hpf zebrafish embryos (Armant et al. 2013). Both genes are expressed in cells of the ventral spinal cord and experimental evidences with morpholinos suggest their implication in the specification of a newly discovered sub-type of interneurons in the zebrafish: V2c (Yang et al., *unpublished data*). These data are also consistent with the mouse model in which the Sox1 TF was shown to be required for the V2c subtype derived from the p2 domain in the ventral spinal cord (Panayi et al. 2010). In order to target both genes simultaneously, the CRISPR method seemed to be the most appropriate. Two efficient gRNAs were used and I could without contest prove the efficiency of the CRISPR method for multi-targeting. The mutations carrier fish for both genes were identified and their progeny will allow us to obtain a double knockout of *sox1a* and *sox1b*, making it possible to assess more deeply the function of these genes. The observations that were previously made with knockdown studies, including the increase of V2b cells supposedly a substitution for non formed V2c cells, can be confirmed and the molecular mechanisms inducing the specification of these interneurons can be then determined in a stronger way with mutants.

73

My results suggest indeed that the CRISPR technology is the most suitable to use in a casual manner because of its straightforward and fast application. Moreover, the overall mutation efficiency obtained with the CRISPR was higher than for TALENs according to my observations. Nevertheless, the efficiency rate, like for TALENs, may also be locus specific or related to the used gRNA. Indeed, a number of gRNAs did not give any positive results while other gRNAs targeting the same genes were really efficient.

4.3. TALEN vs. CRISPR: comparison between the two used site-specific nuclease technologies

After setting up both DNA editing methods in our laboratory, pro and cons of three crucial aspects were compared:

- **Design and construction**

The CRISPR gRNAs target sites have to be in the direct vicinity of the PAM (NGG) site and in addition a 5'-GG is required in order to successfully transcribe the gRNA *in vitro* (using the T7 polymerase). Other restrictions can be added as it was shown in the publication of Gagnon et al., in which more than 100 gRNAs were tested. They indeed observed that a guanine adjacent to the PAM and 50% of GC-content improve the efficiency for the target site. They also affirm that using the Sp6 polymerase for *in vitro* transcription allows more flexibility as it is equally efficient to transcribe 5'GG- or 5'AG gRNAs (Gagnon et al. 2014).

Contrarily to the CRISPR, the TALEN design requires two binding proteins separated by a defined spacer, of 15 to 16 bases, with a thymine at position 0 of the target sequence. However, it was shown that the 5'-T may not be required for artificially designed TAL effectors in contrast to the naturally occurring ones (Meckler et al. 2013; Jankele & Svoboda 2014).

Regarding construction, the TALEN method is based on the modular assembly of varying sets of RVDs that is performed in two steps (1st assembly and 2nd assembly) for each TAL binding site; considering that a pair of TALENs is needed to allow the FokI endonuclease to dimerize and create DSBs in the DNA. The modules are part of a large library of more than 70 elements. For an average laboratory without high-throughput competence, approximately one

week is needed to obtain the final pair of TALENs, without including the time needed to sequence the constructs.

On the other hand, creating a gRNA of 20 bp is far less complex. Two overlapping oligonucleotides are required with an expression vector and a Cas9-containing plasmid. Hence, less than 3 days are needed to obtain a ready-to transcribe gRNA.

- **Specificity**

High target site specificity can be achieved with TALENs or CRISPR allowing precise gene editing. For the CRISPR, a gRNA recognizes a target sequence in the DNA directly followed by the PAM and guides the Cas9 endonucleases to cut both strands. Even if high efficiencies were reported with this method, there is a concern about a high rate of off-target sites as gRNAs can tolerate up to five mismatches (Fu et al. 2013; Kuscu et al. 2014). In order to counteract this issue, strategies were developed to significantly improve the specificity of the CRISPR: One method uses a Cas9 nickase that mutates a single-strand in the DNA and allows more specificity if used in a pair wise manner in which two opposing gRNAs are required to induce mutations at the target site located in between. Even though the nickase can still recognize off-targets, but the result is less mutagenic than DSBs (Mali et al. 2013b; Ran et al. 2013). Another strategy consists of using a pair of gRNAs to recruit an inactive Cas9 fused to the FokI endonuclease (RFN) which considerably reduces off-target mutations (Tsai et al. 2014). Fu et al. also advise to use shorter gRNAs of about 17 nt to reduce off-target effects (Fu et al. 2014).

Concerning TALENs, the off-target effect is less of a concern, as it consists of DNA-binding proteins composed of a succession of RVDs that recognize an average of 17 nt. Thus, a pair of TALENs flanking a spacer region of 15 nt represent a rather long binding region that is unlikely to be recognized more than once in the genome (approximately 34 nt).

Very often, off-targeting activity correlates with cytotoxicity since it leads to unwanted genomic modifications and thus it can affect the awaited results. Therefore, the specificity of engineered endonucleases is of a great concern when developing these strategies. The issue of off-targeting was addressed in several studies using different approaches to identify off-target mutagenesis using engineered nucleases. For instance, a method was applied by Wang et al., in which Integrase Defective Lentiviral Vectors (IDLV) were used to mark the resulting DSBs

from TALENs and CRISPR (Wang et al. 2015). They observed no off-target effects for TALENs and dCas9, while half CRISPR/Cas9 showed off-targets in a dose-dependent manner. Otherwise, ChIP-Seq analyses were also used to identify the bound sequences by the CRISPR (Kuscu et al. 2014; Wu et al. 2014). Deep sequencing and whole-genome sequencing were greatly used to collect data concerning off-target effects of these methods. Moreover, potential misrecognized sites can be predicted *in silico* using suitable algorithms that are used in most online tools, such as CHOPCHOP that can predict off-targets for both CRISPR and TALENs (Montague et al. 2014b).

Nonetheless, the problem of off-targeting is also influenced by the delivery approach. Indeed, using mRNA or purified proteins reduces the off-target effects, compared to transfection of plasmids usually used in cells, probably because of the time of expression and the shorter activity duration of the nucleases (Maggio & Gonçalves 2015; Kim et al. 2014). In zebrafish, a recent publication showed a low risk of off-targeting activity *in vivo* using the CRISPR method (Cas9 mRNA) (Varshney & Burgess 2014). It was also shown that mistargeting is correlated with the dosage of the CRISPR/Cas9. Also, lowering the amount of injected mRNA is of great help in order to minimize off-targets (Fu et al. 2013; Ansai & Kinoshita 2014; Wang et al. 2015).

- **Efficiency**

Both TALENs and CRISPR efficiently target and modify the genome. Evaluating this ability is still hardly feasible. Actually, the efficiency is measured according to the mutation events generated by NHEJ or HR, thus highly depending on the used cell system beside the different mode of action of each nuclease. More accurate analyses should be performed in a high-throughput manner with different engineered endonucleases and within the same system (Hendel et al. 2014). Yet a rough comparison between the large library of publications shows a higher efficiency of the CRISPR method to induce mutations compared to the TALENs (Hwang et al. 2013; Dahlem et al. 2012; Chen et al. 2014), although both can exceed a rate of 50% in zebrafish (Cade et al. 2012; Hwang et al. 2013). However, but unsurprisingly, the modified CRISPR systems optimizing specificity (e.g. Nickases and RFN) show reduced mutation efficiency.

Another difference that can affect efficiency is that TALENs, unlike CRISPR, are sensitive to cytosine methylation, which sometimes leads to a lower or no activity at all (Bultmann et al.

2012; Valton et al. 2012; Hsu et al. 2013; Wu et al. 2014). Nevertheless, in some cases, genome editing can be hardly achieved by both methods and the reason is most probably the sensitivity to chromatin structure as the target sequences have to be accessible to the nucleases.

On the other hand, the CRISPR/Cas9 can be efficiently and easily used in a multiplexed manner in order to target several genes simultaneously (Cong et al. 2013; Jao et al. 2013b; Wang et al. 2013). As for TALENs, two pairs can also be used to delete large genomic fragments, but the large size of the proteins can be of a concern (Ma et al. 2013; Xiao et al. 2013). Moreover, the construction and the testing of multiple pairs of TALENs can be complicated and laborious to achieve.

Method Features	TALEN	CRISPR
Type of recognition	Protein-DNA	RNA-DNA
Recognition site	Typically 15 to 20 nt per TALEN monomer	20 nt + PAM
Targeting design	Flexible with 5'-T at position 0	PAM requirement and 5'-GG for polymerase
Engineering	Moderate	Easy
Methylation sensitive	Sensitive	Not sensitive
Off-target effects	Low observed off-target effects	Potentially higher off-target effects
Multiplexing	Rarely used	Capable

Table 1: Different aspects of comparison between the TALEN and the CRISPR technologies.

In conclusion, both TALENs and CRISPR methods are valuable to create permanent DNA modifications with relatively high but variable efficiencies. TALENs can be designed to target any sequence of the genome without major constrains and with high specificity. However, TALEN construction is more complicated and more costly than obtaining a ready-to-use

gRNA. Choosing one method over the other depends on the attempted application and on the sequence that is to be targeted (Table 1).

4.4. Programmable site-specific nucleases revolutionize reverse genetics in zebrafish

For a long time, the morpholino anti-sense technology monopolized the area of gene function studies in the zebrafish. It was indeed an accessible method, widely used for gene knockdown in the zebrafish community. But the unexpectedly fast development of efficient genome editing technologies raised questions about the effectiveness of morpholino knockdowns. It was already clear that MOs can bind to non-specific targets leading to artifacts and false phenotypes. Even with the application of precautionary measures, such as mRNA rescue or suppression of p53 activity, the distinction between specific and unspecific phenotypes was still difficult (Schulte-Merker & Stainier 2014). On top of this, recent studies claim that a great number of phenotypes caused by MOs did not correspond to the mutant ones (Kok et al. 2014; Law & Sargent 2014; Stainier et al. 2015). Indeed, Kok et al. generated mutant lines for 24 genes using ZFNs, TALENs and CRISPR and only a small number had caused a phenotype. Moreover, after comparison of available data from ZFIN and the Sanger Zebrafish Mutation Project (ZMP), they observed discrepancies between morphants and mutants in approximately 80% of the cases (Kok et al. 2014). It is likely that the pseudophenotypes generated by MOs are caused by off-target effects and even rescue experiments cannot guarantee their specificity. However, the MO antisense technology can still be used in parallel, preferably in addition to a genetic mutant (e.g. double mutants) while keeping a critical eye when characterizing a new phenotype.

In closing, emergence of site-specific nuclease technologies opens a wide range of applications for studying complex gene functions in animals that were so far not accessible for targeted gene knockout technology. Additionally, both TALENs and CRISPR methods cleared the road for using knockin technologies in zebrafish. Indeed, several publications report nowadays successful DNA integration by HR-mediated genome editing or via homology independent repair (Zu et al. 2013; Chang et al. 2013; Auer et al. 2014; and for review see Auer & Del Bene 2014).

4.5. Conclusion and perspectives

The fast development of site-specific mutagenesis in the last few years caught great interest, especially from the zebrafish community that was in need of more accessible targeted genome editing tools. TALENs and CRISPR methods were implemented and their efficiency in zebrafish was showed in this work. Having in mind the understanding of the molecular mechanisms that control neurogenesis and regeneration in the zebrafish brain, TRs with promising preliminary results were selected for investigation. Thus, I targeted several *id* genes that are involved in adult neurogenesis: *id1* was mutated with TALENs and three generations of fish were tested for germline transmission. Mutations in *id3* were demonstrated with both strategies and heritable mutations were proven. The *id2a* and the *id4* genes were also targeted with the CRISPR method without yet finding efficient gRNAs. Having in hand mutants for each of the *id* genes, deeper investigations can be performed in the future to understand their role in adult neurogeneis and regeneration of the zebrafish brain. Hence, by using expression studies and the stab wound assay that was successfully employed in our laboratory, the effects of the different knockouts on these processes can be monitored. Moreover, creating combination lines will allow the study of possible functional redundancies for these genes.

On the other hand, previous studies in the ventral spinal cord of the zebrafish led to the highlight of two related genes, *sox1a* and *sox1b*. They were suggested to play a role in the specification of a newly observed sub-type of interneurons in the zebrafish, named V2c. This class of interneurons that was already identified in mouse has still to be investigated in zebrafish as little is known about its function. Taking advantage of the multiplex targeting of the CRISPR method, both genes were efficiently mutated simultaneously. A null background for the two genes will be used to confirm previous observations and to measure the role of *sox1a* and *sox1b* in the fate determination of the V2c interneurons.

REFERENCES

Ablain, J. et al., 2015. A CRISPR/Cas9 Vector System for Tissue-Specific Gene Disruption in Zebrafish. *Developmental Cell*, 32(6), pp.756–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25752963 [Accessed March 9, 2015].

Alberts, B. et al., 2013. *Essential Cell Biology, Fourth Edition*, Garland Science. Available at: https://books.google.com/books?id=Cg4WAgAAQBAJ&pgis=1 [Accessed July 28, 2015].

Altman, J., 1962. Are New Neurons Formed in the Brains of Adult Mammals? *Science*, 135(3509), pp.1127–1128. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.135.3509.1127.

Amacher, S.L., 2008. Emerging gene knockout technology in zebrafish: zinc-finger nucleases. *Briefings in functional genomics & proteomics*, 7(6), pp.460–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2722258&tool=pmcentrez&rendertype=abstract [Accessed November 25, 2012].

Anon, 2010. *Fish Physiology: Zebrafish, 1st Edition | Steve Perry, Marc Ekker, Anthony Farrell, Colin Brauner | ISBN 9780123749833*, Available at: http://store.elsevier.com/product.jsp?isbn=9780123749833 [Accessed June 8, 2015].

Ansai, S. & Kinoshita, M., 2014. Targeted mutagenesis using CRISPR/Cas system in medaka. *Biology Open*, 3(5), pp.362–371. Available at: http://bio.biologists.org/cgi/doi/10.1242/bio.20148177.

Appel, B. & Chitnis, A., 2002. Neurogenesis and specification of neuronal identity. *Results and problems in cell differentiation*, 40, pp.237–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12353479 [Accessed August 10, 2015].

Appel, B. & Eisen, J.S., 1998. Regulation of neuronal specification in the zebrafish spinal cord by Delta function. *Development (Cambridge, England)*, 125, pp.371–380.

Armant, O. et al., 2013. Genome-wide, whole mount in situ analysis of transcriptional regulators in zebrafish embryos. *Developmental biology*, 380(2), pp.351–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23684812 [Accessed November 12, 2013].

Auer, T.O. et al., 2014. homology-independent DNA repair Highly efficient CRISPR / Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. , pp.142–153.

Auer, T.O. & Del Bene, F., 2014. CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. *Methods*, 69(2), pp.142–150. Available at: http://dx.doi.org/10.1016/j.ymeth.2014.03.027.

Bally-Cuif, L. & Hammerschmidt, M., 2003. Induction and patterning of neuronal development, and its connection to cell cycle control. *Current Opinion in Neurobiology*, 13(1), pp.16–25. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0959438803000151 [Accessed July 3, 2013].

Barrangou, R. et al., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(5819), pp.1709–1712.

Batista, M.F., Jacobstein, J. & Lewis, K.E., 2008. Zebrafish V2 cells develop into excitatory CiD and Notch signalling dependent inhibitory VeLD interneurons. *Developmental biology*, 322(2), pp.263–75. Available at: http://www.sciencedirect.com/science/article/pii/S0012160608010695 [Accessed August 4, 2015].

Becker, C.G. & Becker, T., 2008. Adult zebrafish as a model for successful central nervous system regeneration. *Restorative neurology and neuroscience*, 26, pp.71–80.

Bedell, V.M. et al., 2012. In vivo genome editing using a high-efficiency TALEN system. *Nature*, 490(7422), pp.114–118. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23000899 [Accessed November 1, 2012].

Bertrand, N., Castro, D.S. & Guillemot, F., 2002. Proneural genes and the specification of neural cell types. *Nature reviews. Neuroscience*, 3(7), pp.517–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12094208 [Accessed November 5, 2012].

Bhaya, D., Davison, M. & Barrangou, R., 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics*, 45, pp.273–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22060043 [Accessed May 21, 2013].

Bill, B.R. et al., 2009. A primer for morpholino use in zebrafish. *Zebrafish*, 6(1), pp.69–77. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2776066&tool=pmcentrez&rendertype=abstract.

Bitinaite, J. et al., 1998. FokI dimerization is required for DNA cleavage. *Proceedings of the National Academy of Sciences*, 95(18), pp.10570–10575. Available at: http://www.pnas.org/content/95/18/10570 [Accessed May 19, 2015].

Boch, J. et al., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, 326(5959), pp.1509–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19933107 [Accessed October 27, 2012].

Boch, J. & Bonas, U., 2010. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology*, 48, pp.419–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19400638 [Accessed October 27, 2012].

Bogdanove, A.J., Schornack, S. & Lahaye, T., 2010. TAL effectors: finding plant genes for disease and defense. *Current opinion in plant biology*, 13(4), pp.394–401. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20570209 [Accessed October 30, 2012].

Bolotin, A. et al., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)*, 151(Pt 8), pp.2551–61. Available at: http://mic.sgmjournals.org/content/151/8/2551.long [Accessed July 14, 2014].

Bonas, U., Stall, R.E. & Staskawicz, B., 1989. Genetic and structural characterization of the avirulence gene avrBs3 from Xanthomonas campestris pv. vesicatoria. *Molecular & general genetics : MGG*, 218, pp.127–136.

Briggs, A.W. et al., 2012. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic acids research*, 40(15), p.e117. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3424587&tool=pmcentrez&rendertype=abstract [Accessed November 1, 2012].

Bultmann, S. et al., 2012. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic acids research*, 40(12), pp.5368–77. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3384321&tool=pmcentrez&rendertype=abstract [Accessed November 9, 2012].

Cade, L. et al., 2012. Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic acids research*, 40(16), pp.8001–10. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3439908&tool=pmcentrez&rendertype=abstract [Accessed October 30, 2012].

Carlson, D.F. et al., 2012. Efficient TALEN-mediated gene knockout in livestock. *Proceedings of the National Academy of Sciences of the United States of America*, 109(43), pp.17382–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23027955 [Accessed October 29, 2012].

Carroll, D., 2014. Genome engineering with targetable nucleases. *Annual review of biochemistry*, 83, pp.409–39. Available at: http://www.annualreviews.org/doi/full/10.1146/annurev-biochem-060713-035418?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3dpubmed [Accessed May 26, 2015].

Cermak, T. et al., 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research*, 39(12), p.e82. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3130291&tool=pmcentrez&rendertype=abstract [Accessed October 26, 2012].

Cermak, T., Starker, C.G. & Voytas, D.F., 2015. Efficient design and assembly of custom TALENs using the Golden Gate platform. *Methods in molecular biology (Clifton, N.J.)*, 1239, pp.133–59. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25408404 [Accessed October 28, 2015].

Chang, N. et al., 2013. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell research*, 23(4), pp.465–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3616424&tool=pmcentrez&rendertype=abstract [Accessed May 21, 2013].

Chapouton, P. et al., 2010. Notch activity levels control the balance between quiescence and recruitment of adult neural stem cells. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(23), pp.7961–7974.

Chen, L. et al., 2014. Advances in genome editing technology and its promising application in evolutionary and ecological studies. *GigaScience*, 3(1), p.24. Available at: http://www.gigasciencejournal.com/content/3/1/24.

Christian, M. et al., 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186(2), pp.757–61. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2942870&tool=pmcentrez&rendertype=abstract [Accessed November 2, 2012].

Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR / Cas Systems. *Science*, 819(2013). Available at: http://www.sciencemag.org/content/339/6121/819.short.

Corey, D.R. & Abrams, J.M., 2001. Morpholino antisense oligonucleotides: tools for investigating vertebrate development. *Genome biology*, 2(5), p.REVIEWS1015. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=138935&tool=pmcentrez&rendertype=abstract [Accessed July 8, 2015].

Dahlem, T.J. et al., 2012. Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS genetics*, 8(8), p.e1002861. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3420959&tool=pmcentrez&rendertype=abstract [Accessed November 7, 2012].

Davidson, E.H., 2010. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*, Academic Press. Available at: https://books.google.com/books?id=F2ibJj1LHGEC&pgis=1 [Accessed July 22, 2015].

Deng, D. et al., 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science (New York, N.Y.)*, 335(6069), pp.720–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22223738 [Accessed November 1, 2012].

Dessaud, E., McMahon, A.P. & Briscoe, J., 2008. Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development (Cambridge, England)*, 135(15), pp.2489–503. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18621990 [Accessed May 21, 2013].

Detrich, H.W., Westerfield, M. & Zon, L.I., 2010. *The Zebrafish: Cellular and Developmental Biology, Part A: Cellular and Developmental Biology*, Elsevier Science. Available at: http://books.google.de/books?id=rgD7ABqAt2MC.

Diotel, N., Rebecca, R.V., et al., 2015. Comprehensive expression map of transcription regulators in the adult zebrafish telencephalon reveals distinct neurogenic niches. *Journal of Comparative Neurology*, 1221, p.n/a–n/a. Available at: http://doi.wiley.com/10.1002/cne.23733.

Diotel, N., Beil, T., et al., 2015. Differential expression of id genes and their potential regulator znf238 in zebrafish adult neural progenitor cells and neurons suggests distinct functions in adult neurogenesis. *Gene expression patterns : GEP*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26107416 [Accessed August 10, 2015].

Doyle, E.L. et al., 2012. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic acids research*, 40(Web Server issue), pp.W117–22. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3394250&tool=pmcentrez&rendertype=abstract [Accessed November 7, 2012].

Doyon, Y. et al., 2008. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nature biotechnology*, 26(6), pp.702–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2674762&tool=pmcentrez&rendertype=abstract [Accessed March 5, 2015].

Draper, B.W., Morcos, P.A. & Kimmel, C.B., 2001. Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. *Genesis (New York, N.Y. : 2000)*, 30(3), pp.154–6. Available at: http://doi.wiley.com/10.1002/gene.1053 [Accessed May 5, 2015].

Eisen, J.S. & Smith, J.C., 2008. Controlling morpholino experiments: don't stop making antisense. *Development (Cambridge, England)*, 135(10), pp.1735–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18403413 [Accessed June 26, 2015].

England, S. et al., 2011. Roles of Hedgehog pathway components and retinoic acid signalling in specifying zebrafish ventral spinal cord neurons. *Development*, 138, pp.5121–5134.

Engler, C. et al., 2009. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PloS one*, 4(5), p.e5553. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2677662&tool=pmcentrez&rendertype=abstract [Accessed October 26, 2012].

Finn, R.D. et al., 2014. Pfam: The protein families database. *Nucleic Acids Research*, 42(D1), pp.222–230.

Fu, Y. et al., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, (June), pp.1–6. Available at: http://www.nature.com/doifinder/10.1038/nbt.2623 [Accessed June 24, 2013].

Fu, Y. et al., 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology*, 32(November 2013), pp.279–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24463574.

Gagnon, J. a. et al., 2014. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE*, 9(5), pp.5–12.

Galán, J.E. & Wolf-Watz, H., 2006. Protein delivery into eukaryotic cells by type III secretion machines. *Nature*, 444(7119), pp.567–573.

Gilbert, L.A. et al., 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2), pp.442–51. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3770145&tool=pmcentrez&rendertype=abstract [Accessed July 11, 2014].

Gilbert, S.F., 2010. *Developmental Biology*, Sinauer Associates. Available at: https://books.google.com/books?id=aLzVQwAACAAJ&pgis=1 [Accessed July 28, 2015].

Grandel, H. et al., 2006. Neural stem cells and neurogenesis in the adult zebrafish brain: origin, proliferation dynamics, migration and cell fate. *Developmental biology*, 295(1), pp.263–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16682018 [Accessed November 8, 2012].

Guillemot, F., 2007. Cell fate specification in the mammalian telencephalon. *Progress in neurobiology*, 83(1), pp.37–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17517461 [Accessed June 18, 2013].

Hardy, S. et al., 2010. Reverse genetics in eukaryotes. *Biology of the cell / under the auspices of the European Cell Biology Organization*, 102(10), pp.561–80. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3017359&tool=pmcentrez&rendertype=abstract [Accessed July 8, 2015].

Harvey, S.A. et al., 2013. Identification of the zebrafish maternal and paternal transcriptomes. *Development (Cambridge, England)*, 140(13), pp.2703–10. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3678340&tool=pmcentrez&rendertype=abstract [Accessed October 27, 2015].

Hendel, A. et al., 2014. Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell reports*, 7(1), pp.293–305. Available at: http://www.sciencedirect.com/science/article/pii/S2211124714001545 [Accessed July 9, 2015].

Herbers, K., Conrads-Strauch, J. & Bonas, U., 1992. Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. *Nature*, 356(6365), pp.172–174.

Hisano, Y., Ota, S. & Kawahara, A., 2014. Genome editing using artificial site-specific nucleases in zebrafish. *Development Growth and Differentiation*, 56(1), pp.26–33.

Hopkins, C.M. et al., Identification of a family of avirulence genes from Xanthomonas oryzae pv. oryzae. *Molecular plant-microbe interactions : MPMI*, 5(6), pp.451–459.

Hruscha, A. et al., 2013. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development (Cambridge, England)*, 140(24), pp.4982–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24257628.

Hsu, P.D. et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9), pp.827–32. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3969858&tool=pmcentrez&rendertype=abstract.

Huang, P. et al., 2011a. Heritable gene targeting in zebrafish using customized TALENs. *Nature biotechnology*, 29(8), pp.699–700. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21822242 [Accessed October 26, 2012].

Huang, P. et al., 2011b. Heritable gene targeting in zebrafish using customized TALENs. *Nature biotechnology*, 29(8), pp.699–700. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21822242 [Accessed October 26, 2012].

Huang, P. et al., 2012. Reverse Genetic Approaches in Zebrafish. *Journal of Genetics and Genomics*, 39(9), pp.421–433. Available at: http://dx.doi.org/10.1016/j.jgg.2012.07.004.

Hwang, W.Y. et al., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology*, 31(3), pp.227–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23360964 [Accessed May 21, 2013].

Ishino, Y. et al., 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, 169(12), pp.5429–5433.

Jankele, R. & Svoboda, P., 2014. TAL effectors: tools for DNA Targeting. *Briefings in Functional Genomics*, 13(5), pp.409–419. Available at: http://bfgp.oxfordjournals.org/cgi/doi/10.1093/bfgp/elu013.

Jansen, R. et al., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6), pp.1565–1575.

Jao, L.-E., Wente, S.R. & Chen, W., 2013a. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences of the United States of America*, 110, pp.13904–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3752207&tool=pmcentrez&rendertype=abstract.

Jao, L.-E., Wente, S.R. & Chen, W., 2013b. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences of the United States of America*, 110(34), pp.13904–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3752207&tool=pmcentrez&rendertype=abstract [Accessed March 19, 2014].

Jessell, T.M., 2000. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature reviews. Genetics*, 1(October), pp.20–29.

Jiang, F. & Doudna, J. a, 2015. The structural biology of CRISPR-Cas systems. *Current Opinion in Structural Biology*, 30, pp.100–111. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0959440X15000111.

Jinek, M. et al., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive
    bacterial immunity. *Science (New York, N.Y.)*, 337(6096), pp.816–21. Available at:
    http://www.ncbi.nlm.nih.gov/pubmed/22745249 [Accessed May 21, 2013].

Jinek, M. et al., 2014a. Structures of Cas9 endonucleases reveal RNA-mediated
    conformational activation. *Science (New York, N.Y.)*, 343(February), p.1247997.
    Available at: http://www.ncbi.nlm.nih.gov/pubmed/24505130.

Jinek, M. et al., 2014b. Structures of Cas9 endonucleases reveal RNA-mediated
    conformational activation. *Science (New York, N.Y.)*, 343, p.1247997. Available at:
    http://www.ncbi.nlm.nih.gov/pubmed/24505130.

Jung, S. et al., 2010. Id proteins facilitate self-renewal and proliferation of neural stem cells.
    Stem cells and development, 19(6), pp.831–841.

Karginov, F. V. & Hannon, G.J., 2010. The CRISPR System: Small RNA-Guided Defense in
    Bacteria and Archaea. *Molecular Cell*, 37(1), pp.7–19. Available at:
    http://dx.doi.org/10.1016/j.molcel.2009.12.033.

Kaslin, J., Ganz, J. & Brand, M., 2008. Proliferation, neurogenesis and regeneration in the
    non-mammalian vertebrate brain. *Philosophical transactions of the Royal Society of
    London. Series B, Biological sciences*, 363(1489), pp.101–22. Available at:
    http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2605489&tool=pmcentrez&r
    endertype=abstract [Accessed June 26, 2015].

Kay, S. et al., 2007. A bacterial effector acts as a plant transcription factor and induces a cell
    size regulator. *Science (New York, N.Y.)*, 318(5850), pp.648–651.

Kim, S. et al., 2014. Highly efficient RNA-guided genome editing in human cells via delivery
    of purified Cas9 ribonucleoproteins. *Genome research*, 24(6), pp.1012–9. Available at:
    http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4032847&tool=pmcentrez&r
    endertype=abstract [Accessed July 9, 2014].

Kim, Y.G., Cha, J. & Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger
    fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of
    the United States of America*, 93(3), pp.1156–60. Available at:
    http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=40048&tool=pmcentrez&ren
    dertype=abstract [Accessed November 27, 2014].

Kispert, A. & Herrmann, B.G., 1993. The Brachyury gene encodes a novel DNA binding
    protein. *The EMBO journal*, 12(8), pp.3211–3220.

Kizil, C. et al., 2011. Adult neurogenesis and brain regeneration in zebrafish. *Developmental
    neurobiology*, 72(3), pp.429–61. Available at:
    http://www.ncbi.nlm.nih.gov/pubmed/21595047 [Accessed November 8, 2012].

Kizil, C. et al., 2012. Adult neurogenesis and brain regeneration in zebrafish. *Developmental
    neurobiology*, 72(3), pp.429–61. Available at:
    http://www.ncbi.nlm.nih.gov/pubmed/21595047 [Accessed November 8, 2012].

Knowlton, M.N. et al., 2008. A PATO-compliant zebrafish screening database (MODB): management of morpholino knockdown screen information. *BMC bioinformatics*, 9, p.7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2221974&tool=pmcentrez&rendertype=abstract [Accessed November 12, 2015].

Kok, F.O. et al., 2014. Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish. *Developmental Cell*, 32(1), pp.97–108. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25533206 [Accessed December 22, 2014].

Kuscu, C. et al., 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature biotechnology*, 32(May), pp.677–683. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24837660.

Lasorella, A., Benezra, R. & Iavarone, A., 2014. The ID proteins: master regulators of cancer stem cells and tumour aggressiveness. *Nature Reviews Cancer*, 14(2), pp.77–91. Available at: http://dx.doi.org/10.1038/nrc3638 [Accessed June 30, 2015].

Law, S.H.W. & Sargent, T.D., 2014. The serine-threonine protein kinase PAK4 is dispensable in zebrafish: identification of a morpholino-generated pseudophenotype. *PloS one*, 9(6), p.e100268. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0100268 [Accessed October 27, 2015].

Li, T. et al., 2011. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic acids research*, 39(1), pp.359–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3017587&tool=pmcentrez&rendertype=abstract [Accessed November 1, 2012].

Lyden, D. et al., 1999. Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature*, 401(6754), pp.670–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10537105.

Ma, A.C. et al., 2013. High Efficiency In Vivo Genome Engineering with a Simplified 15-RVD GoldyTALEN Design. *PLoS ONE*, 8(5), pp.1–8.

Maeder, M.L. et al., 2013. CRISPR RNA-guided activation of endogenous human genes. *Nature methods*, 10(10), pp.977–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3794058&tool=pmcentrez&rendertype=abstract [Accessed July 22, 2015].

Maeder, M.L. et al., 2008. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell*, 31(2), pp.294–301. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2535758&tool=pmcentrez&rendertype=abstract [Accessed April 24, 2015].

Maggio, I. & Gonçalves, M. a F. V, 2015. Genome editing at the crossroads of delivery, specificity, and fidelity. *Trends in biotechnology*, 33(5), pp.280–291. Available at: http://www.sciencedirect.com/science/article/pii/S0167779915000499.

Mak, A.N.-S. et al., 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science (New York, N.Y.)*, 335(6069), pp.716–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3427646&tool=pmcentrez&rendertype=abstract [Accessed October 30, 2012].

Makarova, K.S. et al., 2011. Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology*, 9(6), pp.467–77. Available at: http://dx.doi.org/10.1038/nrmicro2577 [Accessed July 9, 2014].

Mali, P. et al., 2013a. RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, 339(6121), pp.823–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23287722 [Accessed May 21, 2013].

Mali, P. et al., 2013b. RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, 339(6121), pp.823–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3712628&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

März, M. et al., 2010. Heterogeneity in progenitor cell subtypes in the ventricular zone of the zebrafish adult telencephalon. *Glia*, 58(7), pp.870–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20155821 [Accessed November 6, 2012].

März, M. et al., 2011. Regenerative response following stab injury in the adult zebrafish telencephalon. *Developmental dynamics : an official publication of the American Association of Anatomists*, 240(9), pp.2221–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22016188 [Accessed December 4, 2012].

Meckler, J.F. et al., 2013. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Research*, 41, pp.4118–4128.

Meng, X. et al., 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature biotechnology*, 26(6), pp.695–701. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2502069&tool=pmcentrez&rendertype=abstract [Accessed June 29, 2015].

Merkle, F.T. & Alvarez-Buylla, A., 2006. Neural stem cells in mammalian development. *Current opinion in cell biology*, 18(6), pp.704–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17046226 [Accessed October 26, 2012].

Miller, J.C. et al., 2011. A TALE nuclease architecture for efficient genome editing. *Nature biotechnology*, 29(2), pp.143–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21179091 [Accessed November 2, 2012].

Moens, C.B. et al., 2008. Reverse genetics in zebrafish by TILLING. *Briefings in functional genomics & proteomics*, 7(6), pp.454–9. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2899843&tool=pmcentrez&r endertype=abstract [Accessed November 6, 2012].

Mojica, F.J.M. et al., 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology*, 36(1), pp.244–246. Available at: http://doi.wiley.com/10.1046/j.1365-2958.2000.01838.x [Accessed February 12, 2015].

Mojica, F.J.M. et al., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, 60(2), pp.174–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15791728 [Accessed August 25, 2014].

Montague, T.G. et al., 2014a. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, 42(W1), pp.401–407.

Montague, T.G. et al., 2014b. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, 42, pp.1–7.

Morbitzer, R. et al., 2011. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic acids research*, 39(13), pp.5790–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3141260&tool=pmcentrez&r endertype=abstract [Accessed October 29, 2012].

Moscou, M.J. & Bogdanove, A.J., 2009. A simple cipher governs DNA recognition by TAL effectors. *Science (New York, N.Y.)*, 326(5959), p.1501.

Neff, K.L. et al., 2013. Mojo Hand, a TALEN design tool for genome editing applications. *BMC bioinformatics*, 14, p.1. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3575288&tool=pmcentrez&r endertype=abstract [Accessed May 23, 2013].

Niola, F. et al., 2012. Id proteins synchronize stemness and anchorage to the niche of neural stem cells. *Nature Cell Biology*, 14(5), pp.477–487. Available at: http://dx.doi.org/10.1038/ncb2490.

Nishimasu, H. et al., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156(5), pp.935–49. Available at: http://www.cell.com/article/S0092867414001561/fulltext [Accessed July 10, 2014].

Norton, J.D., 2000. COMMENTARY ID helix-loop-helix proteins in cell growth , differentiation and tumorigenesis. , 3905, pp.3897–3905.

Nüsslein-Volhard, C., 2012. The zebrafish issue of Development. *Development (Cambridge, England)*, 139(22), pp.4099–103. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23093421 [Accessed July 8, 2015].

Ochiai, H. & Yamamoto, T., 2015. *Targeted Genome Editing Using Site-Specific Nucleases: ZFNs, TALENs, and the CRISPR/Cas9 System* T. Yamamoto, ed., Springer. Available at:

https://books.google.com/books?id=8PASBgAAQBAJ&pgis=1 [Accessed June 30, 2015].

Panayi, H. et al., 2010. Sox1 is required for the specification of a novel p2-derived interneuron subtype in the mouse ventral spinal cord. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(37), pp.12274–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20844123 [Accessed May 22, 2013].

Pourcel, C., Salvignol, G. & Vergnaud, G., 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, England)*, 151(Pt 3), pp.653–63. Available at: http://mic.sgmjournals.org/content/151/3/653.long [Accessed August 11, 2014].

Ramirez, C.L. et al., 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature methods*, 5(5), pp.374–5. Available at: http://dx.doi.org/10.1038/nmeth0508-374 [Accessed April 27, 2015].

Ran, F.A. et al., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6), pp.1380–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3856256&tool=pmcentrez&rendertype=abstract [Accessed July 10, 2014].

Reynolds, B. a & Weiss, S., 1992. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science (New York, N.Y.)*, 255(5052), pp.1707–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1553558.

Reyon, D. et al., 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnology*, 30(5), pp.460–465. Available at: http://dx.doi.org/10.1038/nbt.2170.

Reyon, D., Khayter, C. & Regan, M., 2012. Current Protocols in Molecular Biology Engineering Designer Transcription Activator-Like Effector Nucleases (TALENs). *... Protocols in Molecular ...*. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3753116/.

Robu, M.E. et al., 2007. p53 activation by knockdown technologies. *PLoS genetics*, 3(5), p.e78. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1877875&tool=pmcentrez&rendertype=abstract [Accessed June 26, 2015].

Rogers, C.D., Moody, S.A. & Casey, E.S., 2009. Neural induction and factors that stabilize a neural fate. *Birth defects research. Part C, Embryo today : reviews*, 87(3), pp.249–62. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2756055&tool=pmcentrez&rendertype=abstract [Accessed July 28, 2015].

Römer, P. et al., 2007. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science (New York, N.Y.)*, 318(5850), pp.645–648.

Rothenaigner, I. et al., 2011. Clonal analysis by distinct viral vectors identifies bona fide neural stem cells in the adult zebrafish telencephalon and characterizes their division properties and fate. *Development (Cambridge, England)*, 138(8), pp.1459–69. Available at: http://dev.biologists.org/content/138/8/1459.long [Accessed July 24, 2015].

Rowitch, D.H., 2004. Glial specification in the vertebrate neural tube. *Nature reviews. Neuroscience*, 5(5), pp.409–19. Available at: http://dx.doi.org/10.1038/nrn1389 [Accessed July 20, 2015].

Rozen, S. & Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology (Clifton, N.J.)*, 132, pp.365–86.

Ruzinova, M.B. & Benezra, R., 2003. Id proteins in development, cell cycle and cancer. *Trends in Cell Biology*, 13(8), pp.410–418. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0962892403001478 [Accessed November 23, 2012].

Sander, J.D., Dahlborg, E.J., et al., 2011. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature methods*, 8(1), pp.67–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3018472&tool=pmcentrez&rendertype=abstract [Accessed July 3, 2015].

Sander, J.D., Cade, L., et al., 2011. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature biotechnology*, 29(8), pp.697–698.

Sander, J.D. et al., 2007. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic acids research*, 35(Web Server issue), pp.W599–605. Available at: http://nar.oxfordjournals.org/content/35/suppl_2/W599.full [Accessed July 22, 2015].

Schmid-Burgk, J.L. et al., 2013. A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes. *Nature biotechnology*, 31(1), pp.76–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23242165 [Accessed May 23, 2013].

Schmidt, R., Strähle, U. & Scholpp, S., 2013a. Neurogenesis in zebrafish - from embryo to adult. *Neural development*, 8, p.3. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3598338&tool=pmcentrez&rendertype=abstract [Accessed May 27, 2013].

Schmidt, R., Strähle, U. & Scholpp, S., 2013b. Neurogenesis in zebrafish - from embryo to adult. *Neural development*, 8, p.3. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3598338&tool=pmcentrez&rendertype=abstract.

Schulte-Merker, S. et al., 1994. no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development (Cambridge, England)*, 120, pp.1009–1015.

Schulte-Merker, S. & Stainier, D.Y.R., 2014. Out with the old, in with the new: reassessing morpholino knockdowns in light of genome editing technology. *Development*

(Cambridge, England), 141(16), pp.3103–4. Available at:
http://dev.biologists.org/content/141/16/3103.abstract [Accessed August 23, 2015].

Stainier, D.Y.R., Kontarakis, Z. & Rossi, A., 2015. Making Sense of Anti-Sense Data.
Developmental Cell, 32(1), pp.7–8. Available at:
http://www.ncbi.nlm.nih.gov/pubmed/25584794 [Accessed January 12, 2015].

Stemple, D.L., 2004. TILLING--a high-throughput harvest for functional genomics. *Nature reviews. Genetics*, 5(2), pp.145–150.

Sternberg, S.H. et al., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507, pp.62–7. Available at:
http://www.ncbi.nlm.nih.gov/pubmed/24476820.

Strähle, U. et al., 2004. Vertebrate floor-plate specification: Variations on common themes.
Trends in Genetics, 20(3), pp.155–162.

Streisinger, G. et al., 1981. Production of clones of homozygous diploid zebra fish (Brachydanio rerio). *Nature*, 291(5813), pp.293–296.

Streubel, J. et al., 2012. TAL effector RVD specificities and efficiencies. *Nature biotechnology*, 30(7), pp.593–5. Available at:
http://www.ncbi.nlm.nih.gov/pubmed/22781676 [Accessed October 31, 2012].

Sumanas, S. & Larson, J.D., 2002. Morpholino phosphorodiamidate oligonucleotides in zebrafish: A recipe for functional genomics? *Briefings in Functional Genomics and Proteomics*, 1(3), pp.239–256.

Summerton, J. et al., 1997. Morpholino and phosphorothioate antisense oligomers compared in cell-free and in-cell systems. *Antisense & nucleic acid drug development*, 7(2), pp.63–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9149841 [Accessed June 26, 2015].

Summerton, J. & Weller, D., 1997. Morpholino antisense oligomers: design, preparation, and properties. *Antisense & nucleic acid drug development*, 7(3), pp.187–95. Available at:
http://www.ncbi.nlm.nih.gov/pubmed/9212909 [Accessed June 9, 2015].

Tierney, M.. & Lamour, K.H., 2005. An Introduction to Reverse Genetic Tools for Investigating Gene Function. *The Plant Health Instructor*. Available at:
http://www.apsnet.org/edcenter/advanced/topics/Pages/ReverseGeneticTools.aspx [Accessed July 8, 2015].

Tsai, S.Q. et al., 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology*, 32(April), pp.569–76. Available at:
http://www.ncbi.nlm.nih.gov/pubmed/24770325.

Valton, J. et al., 2012. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *Journal of Biological Chemistry*, 287, pp.38427–38432.

Varshney, G.K. & Burgess, S.M., 2014. Mutagenesis and phenotyping resources in zebrafish for studying development and human disease. *Briefings in functional genomics*, 13(2), pp.82–94. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3954039&tool=pmcentrez&rendertype=abstract [Accessed June 30, 2015].

Viales, R.R. et al., 2015. The Helix-Loop-Helix Protein Id1 Controls Stem Cell Proliferation During Regenerative Neurogenesis in the Adult Zebrafish Telencephalon. *Stem Cells*, 33, pp.892–903. Available at: http://doi.wiley.com/10.1002/stem.1883.

Wang, H. et al., 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell*, 153(4), pp.910–918. Available at: http://dx.doi.org/10.1016/j.cell.2013.04.025.

Wang, X. et al., 2015. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nature Biotechnology*, 33(2), pp.175–178. Available at: http://dx.doi.org/10.1038/nbt.3127 [Accessed January 19, 2015].

Wilson, S.I. & Edlund, T., 2001. Neural induction: toward a unifying mechanism. *Nature neuroscience*, 4 Suppl, pp.1161–1168.

Wolpert, L. et al., 2006. *Principles of Development* Third. Oxford University Press, ed., Available at: http://www.amazon.de/Principles-Development-published-University-Hardcover/dp/B00ES253K8 [Accessed July 27, 2015].

Wong, Y.-C., Wang, X. & Ling, M.-T., 2004. Id-1 expression and cell survival. *Apoptosis : an international journal on programmed cell death*, 9(3), pp.279–89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15258459.

Wu, X. et al., 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology*, 32(7), pp.670–676. Available at: http://dx.doi.org/10.1038/nbt.2889 [Accessed July 9, 2014].

Xiao, A. et al., 2013. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic acids research*, 41(14), p.e141. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3737551&tool=pmcentrez&rendertype=abstract [Accessed June 8, 2015].

Yang, L., Rastegar, S. & Strähle, U., 2010. Regulatory interactions specifying Kolmer-Agduhr interneurons. *Development (Cambridge, England)*, 137(16), pp.2713–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20610488 [Accessed May 31, 2013].

Yasue, A. et al., 2014. Highly efficient targeted mutagenesis in one-cell mouse embryos mediated by the TALEN and CRISPR/Cas systems. *Scientific reports*, 4, p.5705. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25027812.

Zu, Y. et al., 2013. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nature methods*, 10(4), pp.329–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23435258 [Accessed May 23, 2013].

Supplementary information

The Golden Gate Kit

S.1: RVD modules, intermediate arrays and final backbones from the Voytas group (Cermak et al. 2011).
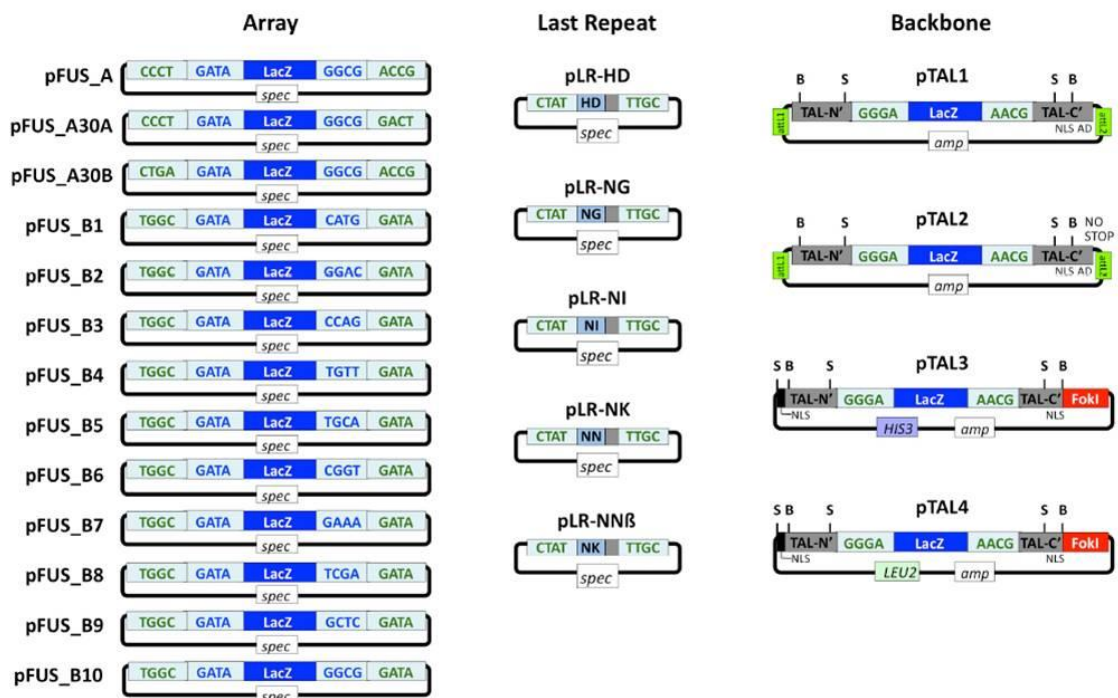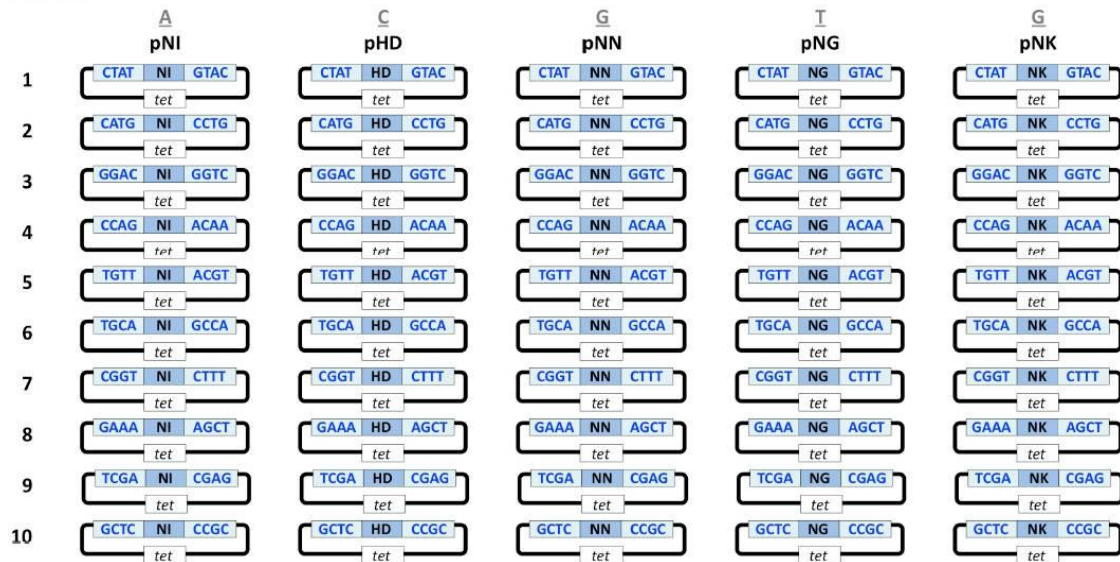
S.2: Primers for PCR and sequencing

Primer name	Primer sequence
pCR8_F1	TTGATGCCTGGCAGTTCCCT
pCR8_R1	CGAACCGAACAGGCTTATGT
TAL_F1	TTGGCGTCGGCAAACAGTGG
TAL_R2	GGCGACGAGGTGGTCGTTGG
SeqTALEN_5-1	CATCGCGCAATGCACTGAC