UPPSALA
UNIVERSITET

CRISPR-Cas mediated knockout of *pdgfrß* and *foxc1a* in zebrafish

Marta Bastos de Oliveira

Degree project in biology, Master of science (2 years), 2014
Examensarbete i biologi 45 hp till masterexamen, 2014
Biology Education Centre and Dept. Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University
Supervisor: Dr. Lwaki Ebarasi
External opponent: Dr. Tatjana Haitina

ABSTRACT

CRISP Repeats are components of an immune system which protects many bacteria and archea against foreign genetic elements. They function by targeting these elements in a sequence specific fashion, guiding the nuclease Cas9 to degrade them. The CRISPR-Cas system from *S. pyogenes* has been adapted and is used as an *in vivo* genome editing tool in a variety of organisms. In this project, we established and optimized the CRISPR-Cas gene editing tool and use it to knock out (KO) genes of interest in zebrafish. This was achieved by co-injecting a gene-specific single guide RNA (sgRNA) and Cas9 nuclease mRNA in the early zebrafish embryos. The cleaved DNA is repaired by Non Homologous End Joining (NHEJ) repair mechanism, generating unpredictable indel mutations. The induced mutations were detected by High Resolution Melting Analysis (HRMA) technique. The targeted genes were *pdgfrß, foxc1a*, *fabp3a*, *fabp11a* and *s1pr1*, in which a requirement for correct vascular development and function has been demonstrated by a morpholino-mediated knock down (KD) strategy. The *pdgfrß* and *foxc1a* genes were successfully mutated in the injected zebrafish and heterozygous carriers for the mutations are currently being raised. We anticipate that the mutant offspring of the carrier founders will be useful in loss of function analyses that will increase our understanding of the roles of these genes in kidney and vasculature biology and pathology. Despite the need to further investigate the specificity as well as germline transmission efficiency, the CRISPR-Cas9 system is a rapid and efficient way of modifying the zebrafish genome.

TABLE OF CONTENTS

LIST OF ABREVIATIONS 1

INTRODUCTION 2

CRISPR-Cas system 2

CRISPR-Cas mediated immunity 3

Genome editing 5

CRISPR-Cas as a genome editing technique 6

CRISPR-Cas applications 7

Use of zebrafish in functional studies 8

Genes of interest 9

pdgfrß 9

foxc1a 9

fabp11a and *fabp3* 10

s1pr1 10

RELEVANCE AND AIMS 11

MATERIAL AND METHODS 12

Identification of the genes of interest in zebrafish genome 12

Identification of zebrafish gene sequence to target with the CRISPR-Cas9 12

Cas9 Nuclease and Single Guide RNA (sgRNA) constructs 12

Identification of the correct clones 13

RNA synthesis 14

Microinjection of zebrafish embryos 14

Genomic DNA extraction 15

Detection of CRISPR-Cas induced mutations 15

Zebrafish culture and breeding 16

RESULTS 17

Finding the target sequences 17

Constructs and sgRNA generation 18

Phenotyping 18

Detection of mutants – Genotyping 21

DISCUSSION 24

Efficient monoallelic mutation 24

Toxicity and Off-target effects 25

Limitations of the project 27

An ongoing project 27

CONCLUSION 29

ACKNOWLEDGEMENTS 30

REFERENCES 31

APPENDIX I 36

LIST OF ABREVIATIONS

CRISPR- Clustered Regularly Interspaced Short Palindromic Repeats

PAM- Protospacer Adjacent Motif

Cas- crispr associated

Bp- base pairs

E. Coli- Escherichia Coli

S. Thermophilus - Streptococcus thermophilus

CRISPR RNA- crRNA

tracrRNA- *trans*-activating crRNA

sgRNA- single guide RNA

TALE- Transcription activator-like effector

ZF- Zink finger

TALEN- Transcription activator-like effector nuclease

ZFN- Zink finger nuclease

DSB- Double Strand Breaks

HDR-Homology Directed Repair

NHEJ- Non Homologous End Joining

KI- Knockin

KO- Knockout

Indels- insertions and deletions

WT- Wild type

dsDNA- double strand DNA

MO- Morpholinos

Hpf- hours post fertilization

Dpf- days post fertilization

INTRODUCTION

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR associated (*Cas*) genes are found in a large number of prokaryotes. Their primary function is to recognize and destroy foreign nucleic acids, for instance plasmids and viruses, providing the host with an adaptive immune system [1–3]. In recent years, this system has been manipulated to provide the scientific community with a simple gene editing tool, based on a single-RNA guided nuclease[4]. The CRISPR-Cas technique can therefore be used to edit genes of interest in several organisms, therefore improving knowledge about their function in health and disease.

CRISPR-Cas system

CRISPR regions were first observed in the K12 strain of *Escherichia Coli* (*E. Coli*) by Ishino and colleagues and later detected in other microorganisms[5,6]. CRISPR regions were subsequently acknowledged as an independent family of DNA repetitive regions by Mojica and colleagues and later on named CRISPR[6,7] Among other criteria, the definition of CRISPR regions relates to the existence of similar-sized non-repetitive (spacer) DNA interspaced by short repetitive regions, associated with *Cas* genes[6]. In fact, CRISPR loci are a series of repeats of approximately 20 to 50 base pairs separated by specific non-repeat spacer sequences (Figure 1), in which the number of repeat-spacer units varies among organisms[8]. Upstream of the repeat-spacer regions of the CRISPR loci, there are two types of elements (Figure 1). Firstly, the leader sequence, which is an AT rich non-coding element composed of 300-500 base pairs (bp) and the promoter of this system[6]. Secondly, the *Cas* genes that are only present in organisms containing the CRISPR loci. More than 45 families of *Cas* genes have been identified, however only a subset is present in each organism and only *cas1* and *cas2* seem to be universal[9].

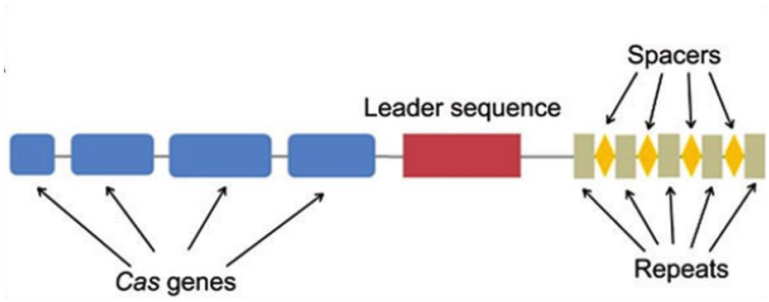

Figure 1. Typical and simplified version of CRISPR loci. From F. Zhang et al, 2014[75]

The CRISPR repetitive regions, some *Cas* genes as well as the leader sequence were found to be homologous between many prokaryote organisms[6]. However, the uniqueness of the spacer sequence to each organism was the most intriguing aspect of this locus. When the homology between the spacer sequences and viral or plasmid DNA elements (foreign DNA) was independently detected by three research groups, a possible biological function as an immune system was hypothesized[10–12]. In 2007, Deveau and Horvath provided the first experimental evidence of the integration of a DNA region of the foreign nucleic acid, a protospacer, from the prokaryote *Streptococcus thermophilus (S. Thermophilus)* in its CRISPR region, after being exposed to a phage challenge[1]. In addition, they showed that *S. Thermophilus* acquired a resistance trait against the same virus, or a genetic memory of infection against that invader, similar to the adaptive immunity observed in higher eukaryotes, validating the previous hypothesis.

The broad existence of this system, present in 46% of the bacteria and 86% of the archaea, illustrates its importance for the total fitness of the prokaryotes[13]. However, it is important to mention that acquisition of foreign DNA may occur by horizontal transfer, for instance in the acquisition of an antibiotic resistance-bearing plasmid. In these situations, the CRISPR-Cas system is non-adaptive and inactive[14].

CRISPR-Cas mediated immunity

There are three stages to the CRISPR-Cas mediated immunity process. The integration of a small piece of DNA from the invading nucleic acids at one end of the CRISPR loci is the first one. In this stage, viral or plasmid invasion triggers the insertion of a 30bp single resistance-conferring spacer, at the leader side of the CRISPR locus. The viral sequence to be integrated is determined by the presence of a short sequence known as Protospacer Adjacent Motif (PAM), which differs between variants of the CRISPR-Cas systems and can either code for a gene or intragenic region[15]. It is known that the PAM sequence is essential to the CRISPR-Cas mediated immunity because a single point mutation in this sequence allows viruses to overcome this immune response[15]. Each integration event is followed by duplication of the CRISPR repeat, maintaining the CRISPR locus architecture[8]. Although the mechanism of protospacer integration and duplication is not yet well understood, it seems to involve several Cas proteins[6].

The second phase consists of the biogenesis of the CRISPR RNA (crRNA). During this process, the transcript pre-crRNA is produced from the CRISPR locus and then cleaved resulting in a library of short crRNAs, each containing a unique spacer sequence flanked by the adjacent repeats[17] There are different pathways that ensure this cleavage or processing. In some organisms, it is catalyzed by endoribonucleases operating as a subunit in a larger complex, the Type I CRISPR–Cas systems [9]. In others, by one sole endoribonuclease which binds to the crRNA in a structure and sequence-specific

way, the Type III CRISPR-Cas systems[9,18]. Finally, the processing can be ensured by a housekeeping, double-stranded RNA-specific RNAseIII, in the presence of the Cas9 nuclease and guided by a *trans*-activating small RNA (tracrRNA), the Type II CRISPR-Cas systems[9,19]. In the Type II CRISPR-Cas system, the tracrRNA forms a duplex structure with part of the repeat in the pre-crRNA, which anneals with the target DNA and guides the Cas9 endonuclease to cleave the complex, silencing the invader's DNA (Figure 2).

In all cases, the generated mature 40 to 60bp crRNA has a 30bp spacer region flanked by partial repeat sequences in the 5' and 3'end[8,17].

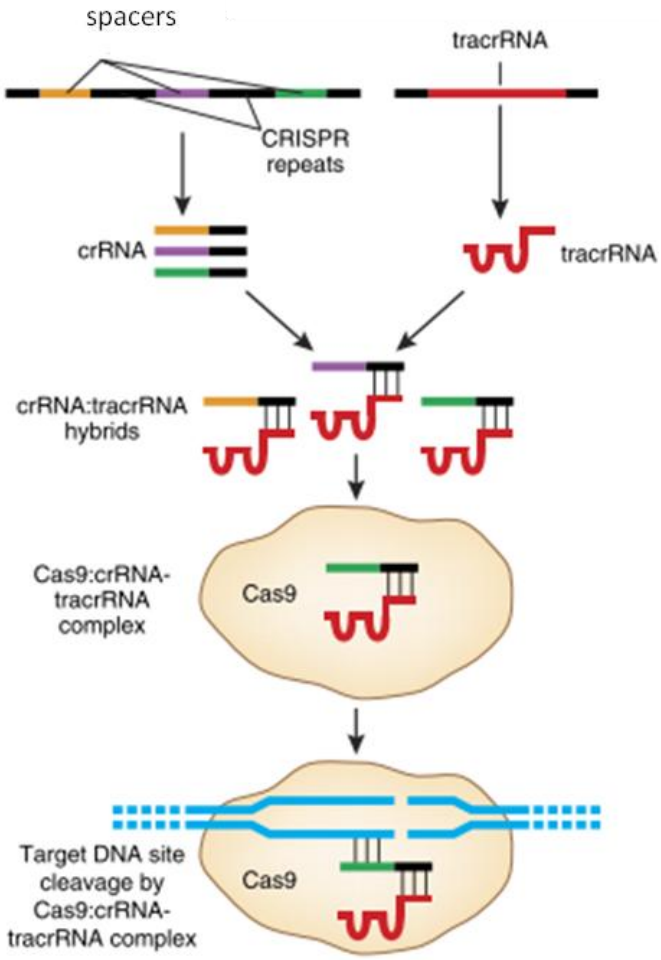

Figure 2. Type II CRISPR-Cas system. Biogenesis of tracrRNA and crRNA encoded by the CRISP loci containing the spacer region complementary to the foreign DNA. The crRNAs hybridizes with tracrRNAs and this RNA duplex associates with the Cas9 nuclease. Cas9 enzyme recognizes the crRNA-tracrRNA-DNA complex and cleaves foreign DNA bearing the protospacer sequence. Modified from J. Sander et al[20]

The third and last phase of the CRISPR-mediated immunity is the target interference, in which the mature crRNA associated with different Cas proteins recognize and cleave the target nucleic acid.

The foreign nucleic acids are identified by direct Watson-Crick base pairing between their sequence and the associated crRNA spacer sequence[1,16,21–23]. Garneau and colleagues showed that both strands of the targeted region of DNA are cleaved, efficiently promoting DNA silencing, without the risk of self DNA destruction [24]. The self-recognition is made by the repeat sequence of the crRNA, which is complementary to the self CRISPR locus, recognizing and sparing the host's DNA from interference[25]. In other words, in addition to complementarity between protospacer and spacer region in the crRNA, non-complementarity between the repeat region of the crRNA and the target protospacer is essential for the process of interference. However, it has been shown that host sequences are occasionally integrated in the CRISPR locus as spacers, as a result of autoimmunity rather than gene regulation[26,27]. The authors further propose that accidental integration of host DNA can lead to a fitness cost, leading to inactivation of the CRISPR-Cas system in many organisms[26].

Genome editing

Genome editing refers to the artificial modification of a desired region of the genome in a diverse range of organisms, based on the use of engineered nucleases[28]. By creating double strand breaks (DSBs), the nucleases stimulate the activation of poorly understood cellular pathways of DNA repair, such as Homology Directed Repair (HDR) or Non Homologous End Joining (NHEJ)[29]. HDR acts by mediating a custom repair DNA strand as a template for the repair of the DSB. Thus, researchers can introduce desired sequence changes by delivering an engineered donor template DNA with homologous flanking regions to the DSB region, leading to the generation of precise gene alterations or knockins (KI)[30] (Figure 3). NHEJ is the cell's default DNA repair mechanism. The repair process of this error prone pathway can induce mutations in the original sequence, inserting or deleting small amounts of nucleotides (indels) (Figure 3), possibly producing a knockout (KO) in the original sequence via frameshift mutation[30].

Besides the nuclease's action, it is necessary to guide this enzyme to the target gene in order to induce specific mutagenesis. For that, DNA-Binding proteins such as Zinc Fingers (ZF) and Transcription activator like effectors (TALEs) are used. ZF domains are DNA-binding protein domains and are widely distributed among eukaryotes, in which each ZF motif binds to three bases[28]. TALEs are another group of DNA-binding proteins, naturally occurring in plant pathogenic bacteria genus *Xanthomonas*, in which each domain of the protein recognizes a single base pair[28]. Like ZFs, TALEs can be customized and assembled in modules and therefore used to recognize desired DNA sequences. Hence, the techniques ZF nucleases (ZFN) and TALE nucleases (TALENs) stand on the engineering of chimeric proteins, with both a DNA specific-binding motif (ZF or TALE) and a nonspecific nuclease - FokI cleavage domain, in this case - which cleaves the desired, recognized strand[28]

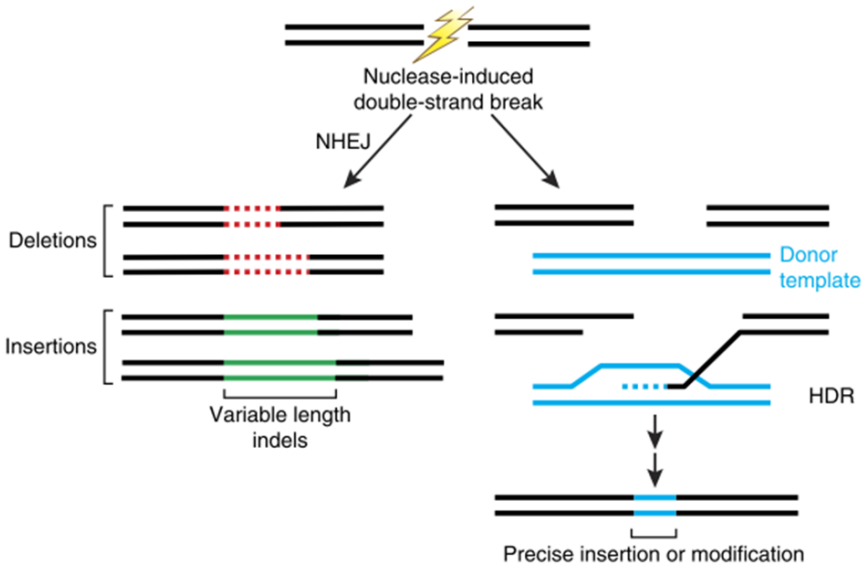

Figure 3- Nuclease-mediated genome editing. DSBs induced by the nuclease are repaired by NHEJ or HDR pathways. Error prone NHEJ repair can produce indel mutations of variable length at the site of the DSB while HDR can introduce precise point mutations or insertions from a customized DNA template From J. Sander et al[20]

CRISPR-Cas as a genome editing technique

The simplicity of the bacterial CRISPR-Cas system from *Streptococcus pyogenes* made it possible to adapt it for use as a gene editing tool. By using RNA-guided instead of protein-guided nucleases, it is easier to assemble in comparison to ZFNs and TALENs. Cas9, the hallmark of type II CRISPR-Cas system and the nuclease of this system, was purified from *S. pyogenes* and its requirements for site-specific cleavage unravelled by Jinek and colleagues[4]. The authors proposed that the tracrRNA is important to target DNA binding, possibly by rearranging the conformation of the molecules[4]. Additionally, the reason for the persistent detection of a PAM sequence (5′-NGG for *S. pyogenes*) in the target DNA was investigated, leading to the suggestion that this sequence might be important for the targeting of the protospacer (target) sequence, either to allow double strand denaturation, strand invasion or R-loop formation, necessary for the pair binding. The first 10 nucleotides in the 5' extremity of the crRNA, the "seed sequence", was revealed to be crucial for the interaction between crRNA and target strand, and therefore, for the site-specific cleavage. However, it has been shown that a simplified, truncated version of the original crRNA:tracrRNA is enough to induce Cas9-mediated cleavage, if the "seed sequence" is left intact. By doing so, Jinek et al developed the first engineered chimeric crRNA:tracrRNA duplex – single guide RNA (sgRNA), greatly simplifying the system and enabling its use in genome editing[4]. The customizable sgRNA construct is composed of a 5' region with 20 nucleotide, the target recognition sequence, and a 3' region composed of the

crRNA fused to the 5' extremity of tracrRNA producing a hairpin structure that mimics the RNA-RNA duplex,recognizable by the Cas9 nuclease (Figure 4). Effective *in vitro* trials were performed for the first time by the same research group, showing the potential and versatility of this system[4]. The DNA strand complementary to the target-binding sequence of crRNA is cleaved 3 nucleotides upstream of the PAM, by the HNH domain of the Ca9 and the non-complementary strand is cleaved 3 to 8 nucleotides upstream by the RuvC-like domain[4]. With this system, it is possible to direct modification to virtually all sequences of the form N_{20}-NGG, by cloning the 20-nucleotide target sequence into the core sgRNA.

Some research groups have developed improved systems by adding nuclear localization signals and/or altering the Cas9 sequence to codon-optimized versions. Moreover, several groups provided web-based software which facilitates the identification of potential CRISPR targets in sequences provided by the users[31,32]. The mentioned initiatives have contributed much in increasing the efficiency and expanded use the CRISPR-Cas system. Despite the efforts of the scientific community to better comprehend the biochemistry and mode of action of the nuclease Cas9 and sgRNA, the occurrence of off target effects is yet to be fully understood[33]. For instance, some studies show that some mismatches between the sgRNA and the target can be tolerated, raising doubts about the specificity of the CRISPR-Cas system[34].

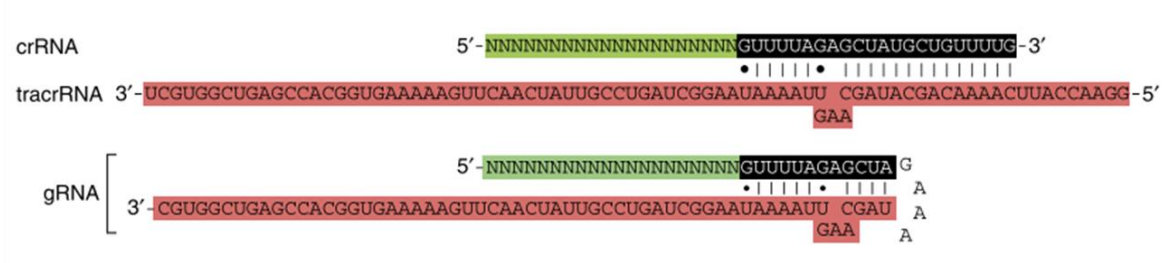

Figure 4- Comparison between sequences of crRNA:tracrRNA duplex and sgRNA. From Sander and Joung, 2014[20]

CRISPR-Cas applications

Following the ground-breaking discoveries mentioned above, successful proof-of-principle studies were conducted in bacteria, human cell lines, zebrafish, and in a number of other living organisms[31,35–40], using the CRISPR system for a variety of goals.

Efficient biallelic KO in mice and zebrafish allowing phenotyping in F0 has been demonstrated[38,41]. Moreover, it was shown that delivery of multiple sgRNAs, multiplexing, can be used to introduce mutations at multiple desired sites *in vivo*[41] or *in vitro*[38], being particularly useful to KO or KI parallel pathways or redundant genes. Multiple sgRNAs can also be introduced in the organism to

stimulate deletion or inversion of large regions located between the targeted cleavage sites[42,43]. In addition to relying on NHEJ, simultaneous delivery of a single or dsDNA template can be performed to stimulate HDR and therefore insert KI[44], restore a mutated disease causing allele into a WT one[45] or create new animal disease models by introducing or substituting specific modifications in a gene[46]. Another possibility is to use sgRNA libraries to perform forward genetic screens in the same way of short hairpin RNA, but with gene KO instead of gene knock downs (KD), such as the GeCKO library in human cells[47].

Use of zebrafish in functional studies

Zebrafish is a powerful model organism used in biomedical research, for a number of reasons. Firstly, zebrafish adults can be kept in aquaria at a relatively high density with both low maintenance needs and cost. Secondly, a zebrafish lays over 100 eggs per clutch, and the embryos develop quite rapidly, having all vertebrate organs formed by 48 hours post fertilization (hpf). These characteristics make this organism very suitable to perform large scale chemical, therapeutically or mutagenic screens, since phenotypes can be rapidly analysed in a large number of individuals[48]. In fact, the zebrafish larvae are the only vertebrates in which these types of experiments are possible. Thirdly, zebrafish fertilization and embryo development occurs *ex utero*, and the embryo is completely transparent until 24 hpf. This enables *in situ* and *in vivo* observation of healthy and mutant embryonic development as well as an easy manipulation of the embryos through microinjection[48]. Also for these reasons, the zebrafish embryo and larvae are considered an outstanding organism for imaging, at single cell resolution. Approximately 70% of the zebrafish genes have an orthologous human gene, as many structures and organs are structurally and physiologically similar to those of humans'[49,50]. This enables the possibility of using zebrafish as a vertebrate model of disease. Lastly, as knowledge about its genome increases, many reverse genetic approaches are available such as mutagenesis and genetic mapping and transgenesis[50].

Functional biology refers to the field of biology which attempts to elucidate gene functions and interactions, in other words, tries to understand the connection between genotype and phenotype. The main workflow involves "the perturbation of gene function to investigate the consequence on the function of other genes in a genome"[51]. For many years, the main approach for performing functional studies in zebrafish was to use morpholinos (MO), i.e. synthetic and stable oligonucleotides, to KD the expression of specific genes[52]. However, there are some caveats to this method, such as the high possibility of toxicity and off-target effects and the knock down not necessarily leading to a total "loss-of-function" of the targeted gene, but only a titration, due to the fact that the MO targets the transcripts and not the gene itself[52]. Therefore, techniques based on site-specific nucleases make it possible to verify earlier MO-derived findings.

Although protein-guided nuclease tools for gene editing have been successfully implemented in zebrafish, the simplicity of CRISPR-Cas mediated editing, efficiency and lower costs make it an attractive, robust and easy approach to modifying the zebrafish genome[53,54]. Achieved for the first time in J. Keith Joungs's lab[31], CRISPR-Cas mediated mutations in zebrafish have already been shown by the same group to pass through to the germline at high rates, with relevant probability of biallelic mutations[46]. Furthermore, introduction of precise genome modifications (point mutations) and efficient multiplexing in zebrafish was for the first time accomplished using this technique[41,46]. The CRISPR-Cas9 technique has been used to promote chromosomal deletions and inversions as well as to make zebrafish reporter lines by Homology Independent DNA Repair (HIR) induced KI[43,55].

Hence, the implementation of CRISPR-Cas in the zebrafish opens the door for a cost effective approach to rapidly modifying its genome, making it possible to not only generate new mutant lines but to also further explore and validate previously reported KD phenotypes and consequent analysis of loss-of-function phenotypes that might cause or participate in human diseases. This model system is of medical relevance as it offers a quick screen for phenotypes, an easy way of reporting them by imaging and a possible comparison with human biology.

Genes of interest

pdgfrß

PDGFRß is the tyrosinase kinase receptor of the Platelet derived growth factor b (PDGFB), a mitogen for many cell types of mesenchymal and neuroectodermal origin[56]. The PDGFB-PDGFRß signaling axis is essential to the recruitment of pericytes necessary for the maturation, function and organisation of microvessels. Absence of pericytes and PDGFB-PDGFRß signaling leads to immature, leaky and dysfunctional microvessels as well as kidney glomerular aneurysms and a compromised BBB[57–59]. Homozygous *Pdgfrß* null mice die shortly before birth, making functional studies hard to carry out[58]. To overcome this problem, mice chimeras as well as mice lacking PDGFRß retention motif have been generated, as models of reduced number of pericytes[60,61]. However, conclusions from studies are limited. One option to better understand function of Pdgfrß, is to study a loss-of-function phenotype in zebrafish.

foxc1a

The *FOXC2* transcription factor is expressed in the murine podocytes and is essential for correct glomerular development. *Foxc2* -/- mice lack essential structural features of the glomerular filter structure such as podocyte slit diaphragms and foot processes, endothelial cell fenestrations and a mature Glomerular Basement Membrane (GBM). This results in abnormally shaped glomeruli which contain few, abnormally dilated, capillary loops[62]. Foxc1a is the functional equivalent of the

mammalian FOXC2 in the zebrafish and KD morphants exhibit a reduced number of podocytes. However, zebrafish null mutants are required to confirm these results and better elucidate this putative function.

fabp11a and *fabp3*

Mammalian Fatty acid-binding protein (FABP) 4, also known as adipocyte-type FABP (A-FABP) and adipocyte lipid-binding protein (aP2), is a component of lipid-binding proteins (LBP) family found in the adipose tissue[63]. Functional analysis of *Fabp4* null mice suggests a participation in intracellular fatty acid trafficking[64]. In zebrafish, two proteins seem to be orthologous to mouse and human FABP4: Fabp3 and Fabp11a[65], the former being expressed mostly in larval central nervous system, retina, pancreas and liver while the latter is expressed in the developing eyes, liver and brain[66]. However, it is not certain which is the functional equivalent of the mammal FABP4 and more functional studies to clarify this are desired.

s1pr1

S1PR1 is the receptor of the lipid sphingosine-1-phosphate (S1P). Loss of function studies in mice and KD in zebrafish revealed its importance for inhibiting angiogenesis by stimulating cell-to-cell adhesion, necessary for the acquisition of vascular stability[67]. Angiogenesis, particularly important during development, is the process by which the already established primitive vascular network develops new vessels by migration of endothelial tip cells from pre-existing vessels, a process stimulated by Vascular Endothelial Growth Factor A (VEGFA). As the developmental stage advances, it further requires stabilization and cessation of sprouting and activation of S1PR1/S1P seems to be required at that stage. In accordance to these results, MO studies in zebrafish reveal reduced blood flow, pericardial and brain edema and excess of vascular sprouting in the hindbrain. However, zebrafish null mutants are required to validate these results.

RELEVANCE AND AIMS

In the present project we employ the CRISPR-Cas9 mediated cleavage tool to set the stage for future functional studies on a set of target genes related to vascular and kidney biology using zebrafish as our model organism. We attempt to generate gene-specific mutations by using the CRISPR-Cas technique to yield both heterozygous and biallelic mutants in a select set of genes. These mutants will be used in future studies aimed at promoting our understanding of the functional relevance of these genes during vascular and kidney development and function. Knowledge gained in these studies is fundamental to the elucidation of pathways and functional relationships important for the development of vascular and kidney disease as well as possible therapeutic approaches.

Accordingly, we defined two main objectives for the present thesis:

1) Establish and optimize the CRISPR-Cas9 technique and use it to KO *pdgfrß, fabp3, fabp11a* and *s1pr1* in zebrafish as well as test a simple method for detecting the induced mutations.
2) Analyse the phenotype(s), if any, of the mutants generated, and compare them with previously described phenotypes from KD experiments.

MATERIAL AND METHODS

Identification of the genes of interest in zebrafish genome

A BLAST search of mouse proteins PDGFRß, FOXC2, FABP4 and S1PR1 a against zebrafish proteins on both ENSEMBL and NCBI databases led to the identification of zebrafish homologs of the mouse proteins. This was performed due to the fact that the zebrafish genome is not fully curated, i.e. its genes are not yet fully annotated and validated and therefore names may be misleading.

Identification of zebrafish gene sequence to target with the CRISPR-Cas9

The exons chosen to be targets were: exon 1 and 2 of *pdgfrb*, exon 1 of *fox1a* in two different early locations, exons 1 and 2 of *fabp3*, exon 2 of *fabp11a* and exon 1 and 2 of *s1pr1*. The early exons were the chosen targets to increase the chance of full loss-of-function occurrence. Blast analysis of the chosen targets was performed against the zebrafish genome to ensure specificity. Up to 1000bp sequence of each exon was input into the online software CRISPR Design Tool[32] (http://crispr.mit.edu/), to search for protospacer target sequences with the form $5'(N)_{20}$-NGG-3′, NGG being the PAM sequence necessary for the Cas9 to cut the protospacer DNA. The output included several 20bp target options, with different specificity values, based on a statistical logarithm of off target hits. The option with best specificity was chosen, in order to minimize the chances of off target binding of the sgRNA and therefore cutting by the nuclease Cas9.

Cas9 Nuclease and Single Guide RNA (sgRNA) constructs

The plasmid constructs used, T7sgRNA and pCS2-nCas9n (Plasmids 46759 and 47929 respectively) were purchased from Addgene. Both vectors were developed by Chen and Wente labs, optimized for zebrafish genome engineering, by having nuclear localization signals (nls) at both termini and codon optimized nucleotides. The core sgRNA vector harbors the crRNA:tracrRNA backbone, preceded by a multiple cloning site, to make it possible to insert the specific spacer for each target region, downstream of a T7 promoter.

The specific protospacer sequence was generated by annealing short oligonucleotides which were designed to be complementary to the gene targets in the zebrafish genome. The design of the primers consisted in leaving overhangs to enable cloning into the vector (Forward: TA N18, Reverse: AAACN18). The annealing of the primers occurred under the following conditions: incubation at

95ºC for 5 minutes, ramping to 50ºC at 0,1º/sec, incubating at 50ºC for 10 minutes and chilling to 4ºC at speed 1ºC/sec.

The cloning of the oligonucleotides in the vector consisted of a one-step digestion and ligation reaction. The core T7sgRNA vector was digested with BsmBI, BglII and SalI (New England Biosciences) and its extremities ligated with the annealed oligonucleotides by T4 DNA ligase (ThermoFisher Scientific), with the program: 3 cycles of 20 minutes at 37ºC and 15 minutes at 16ºC, followed by 10 minutes at 37ºC, 15 minutes at 55ºC and 15 minutes at 80ºC. The construct was transformed into chemically competent DH5alpha cells by using 2ul of ligation product and plating 10 ul in an LB-agarose plate with ampicilin for selection.

Identification of the correct clones

5 colonies of each construct were chosen and preceded to a pre-culture to extract the plasmid DNA by using QIAGEN MiniPrep according to the manufacturer's instructions, followed by incubation with the SalI enzyme. The restriction site of SalI is in the DNA fragment excluded of the vector in the cloning step. So, the vectors that did not incorporate the new 20bp fragment (annealed oligonucleotides), and still had the SalI restriction site, were linearized. The electrophoresis was used as a screening for a correct clone, since it is possible to see a difference between a linearized and non-linearized vector (Figure 5). A non-linearized vector from each construct was sequenced with the M13F primer to confirm the presence of the insert. The correct clone was then extracted using QIAGEN MidiPrep, according to the manufacturer's instructions, to a minimum concentration of115 ng/ul.

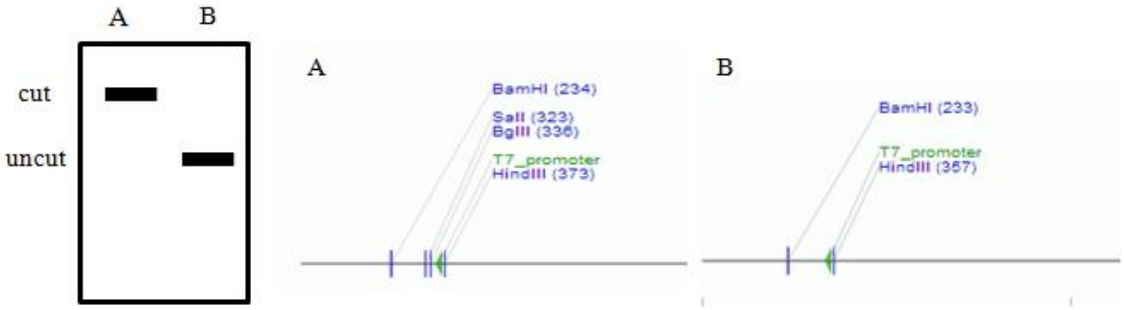

Figure 5. Schematic of screening process to detect the correct clone. The screen is based on the elimination of a restriction site (SalI) when correct ligation of annealed oligonucleotides is accomplished (case B), comparing with other vectors which self-annealed (case A).

RNA synthesis

The T7sgRNA vector was linearized with BamHI, and treated with Proteinase K (100ug/ml) and SDS (0,5%), at 50ºC for 20 minutes. The linearized plasmid was purified using the PCR purification kit by QIAGEN. The *in vitro* transcription was performed using the MEGAshortranscript T7 kit (Ambion/Invitrogen) with 1 ug of purified linearized vector, according to the manufacturer's instructions. The RNA was purified via alcohol precipitation and re-dissolved in RNAse-free water.
The Pcs2-nls- zCas9-nls vector was linearized with NotI, purified with the QIAprep column, transcribed using the kit mMESSAGE mMACHINE SP6, purified via Lithium Chloride precipitation and re-dissolved in RNAse free water.

Microinjection of zebrafish embryos

SgRNA and Cas9 RNA were co-injected into 1-cell stage embryos, to ensure a high efficiency delivery of the injected mRNA to the embryo as well as to reduce mosaicism. The injected volume was ~1 nl of solution, which was measured with a millimetric ruler, in mineral oil. All the injection solutions had 5% of phenol red (Sigma) as a tracer and 160mM of the isotonic solution Potassium Chloride (KCl). The medium in which the embryos grew was E3 water, with Penicillin and Streptomycin (P0781, Sigma) at concentration 1:1000. The embryos were dechorionated with pronase (Roche), by adding it in a concentration of 0,012mg/ml, for 1 hour. The injected embryos were phenotyped for morphological abnormalities under a stereomicroscope over the first 5 days of development. When analysed, embryos and larvae were anesthetized by placing them in water with 0.003% tricaine methanesulfonate (Sigma, St. Louis, MO). Embryos that had a putative phenotype were removed and lysed for further genotyping analysis. 5 dpf WT-like embryos/larvae were raised to adulthood. For each different sgRNA, the starting injected concentration was 100ng/ul as per Chen and Wente guidelines (Table 1). Later, the concentrations were adjusted by evaluating the efficiency, as well as possible injection toxicity risks like mortality.

Table 1. Scheme of initial injection design. Groups 1, 2 and 3 are part of the same zebrafish clutch. Amount injected expressed in picogram.

Injection group	1	2	3
Cas9 RNA	150	150	0
Specific sgRNA	100	0	0
Unspecific sgRNA	0	100	0

The groups 2 and 3 from table 1 are the controls of the experiment, necessary for the interpretation of results. Control group 3 serves as a clutch control, since the early mortality in zebrafish embryos can be naturally high, as well as direct comparison between WT and possible mutant while phenotyping is needed. Group 2 is an injection control, necessary to determine whether the results in group 1 are specific of the sgRNA injected or unspecific toxicity of the injections.
To better assess mortality and toxicity effects of the sgRNA and Cas9 RNA, a control injection was preformed following the design described (Table 2). The embryos were followed and classified as dead, deformed and WT-like, i.e., healthy.

Table 2. Design of toxicity control injection. Groups (N ~100) from the same zebrafish clutch. Amount injected expressed in picograms.

Group	1	2	3	4	5	6	7	8	9	10	11	12
Unspecific sgRNA	0	0	10	20	40	80	100	0	10	20	80	100
Cas9 RNA	0	150	150	150	150	150	150	300	300	300	300	300

Genomic DNA extraction

To extract the genomic DNA of embyos and larvae, they were incubated for 2 hours at 55ºC in 50 ul of lysis buffer (10mM Tris-HCl (pH 8.0), 1 mM EDTA, 50 mM KCl, 0.3% Tween-20, 0.3% NP-40 and 500 ug/ml proteinase K). The proteinase K reaction was terminated by incubating the embryos at 95ºC for 5 minutes. DNA concentration was measured by Synergy™ HT microplate reader (BioTek).

Detection of CRISPR-Cas induced mutations

The technique High Resolution Melting Analysis (HRMA) was used to detect mutations. HRMA is the quantitative analysis of the DNA fragment melting curve by generating a profile that is both specific and sensitive enough to detect single-base differences between the strands. This method is able to detect mixtures of induced DNA polymorphisms present at a targeted region of a single embryo. For this, amplicons of 100-150 bp including the target region were generated in the PCR reaction containing 60ng of embryonic DNA, 1X Precision melt supermix (containing the LC Green Plus dye, enzymes and dNTPs, Bio-Rad) and 200 nM each primer. Conditions of the qPCR reaction were: denaturation at 95ºC for 3 min.; 45 cycles of 95ºC, 30 seconds; 56ºC, 1 min.; denaturation at 95ºC, 30 seconds; quick anneal at 56ºC for 10 seconds and temperature increase (0,2ºC increment) to 95ºC. The HRMA was performed in CFX96 Real-Time System, C1000 touch Thermal Cycler (Bio-

Rad Laboratories, Inc.) and its output analyzed using *Precision Melt* Analysis™ *Software* (Bio-Rad Laboratories, Inc.).

Zebrafish culture and breeding

Adult zebrafish (*Danio rerio*) were grown in the fish facility with a 14 hours light/10 hours dark cycle. Zebrafish embryos used for injections were obtained from natural spawning of two lines, wildtype AB and Tg*(kdrl:EGFP,gata:dsred)*. The double transgenic line Tg(kdrl:EGFP;gata-1a:dsred) allows for the visualization of endothelial and blood cells. Embryos and fish were raised and maintained as described in standard protocols[68].

RESULTS

Finding the target sequences

All target sequences chosen, amongst the ones presented by the software CRISPR Design Tool, are listed in table 3, as well as their scores of specificity, or "inverse likelihood of off targets binding"[32]. The scoring algorithm compares the 20bp sequence desired to target against the zebrafish genome and identifies the possible off target locations, integrating and quantifying contributions of mismatch locations, densities and identity according to specificity rules of the CRISPR-Cas9 system. For instance, mismatches in the PAM proximal region are less tolerated then in the PAM distal region and the maximal number of consecutive mismatches should be less than 4[32]. Moreover, it provides additional information from the predicted off target locations, their sequences as well as a probability of each off target event. Between the 20bp target options for each input, the ones with higher specificity (or "high quality" qualified) were chosen, in order to minimize the chances of off target binding of the sgRNA and therefore cleavage evens by the nuclease Cas9. The oligonucleotides designed are stated in table 3.

Table 3. Target sequences and respective specificity scores and oligonucleotides. "High quality" target sequences identified by the software CRISPR Design Tool, with a PAM sequence (fully underlined), and their score of specificity, i.e. "100% minus a weighted sum of off-target binding probability"[32]. For each sequence there are the corresponding designed oligonucleotides in the form TA N_{18} (forward) and AAAC N_{18} (reverse), N_{18} being the target sequence without the PAM sequence.

Target name	Target sequence (5' to 3')	Score	Oligonucleotides (5' to 3')
pdgfrß **[exon1]**	TAGGAAGTGGATGCGGCTGATGG	92%	TATAGGAAGTGGATGCGGCTGA AAACTCAGCCGCATCCACTTCC
pdgfrß **[exon2]**	CTCGTCCCGCCTGAAGCGCCAGG	98%	TACTCGTCCCGCCTGAAGCGCC AAACGGCGCTTCAGGCGGGACG
foxc1a **[exon1.1]**	CGGCACAACTCCCAGCGAGTTGG	99%	TACGGCACAACTCCCAGCGAGT AAACACTCGCTGGGAGTTGTGC
foxc1a **[exon1.2]**	CCGGGTTTCTTATCATCGCGTGG	98%	TACCGGGTTTCTTATCATCGCG AAACCGCGATGATAAGAAACCC
fabp3 **[exon1]**	TATCGGCACGTGGAACTTGAAGG	96%	TATATCGGCACGTGGAACTTGA AAACTCAAGTTCCACGTGCCGA
fabp3 **[exon2]**	CATATTGGCAACTTGGCGCGTGG	97%	TACATATTGGCAACTTGGCGCG AAACCGCGCCAAGTTGCCAATA
fabp11 [exon 2]	GGTCTTTCTGTCATCCGCCGTGG	80%	TAGGTCTTTCTGTCATCCGCCG AAACCGGCGGATGACAGAAAGA
s1pr1 **[exon 2.1]**	CCCAGTGAAGTTGTAGTGTCTGG	96%	TACCCAGTGAAGTTGTAGTGTC AAACGACACTACAACTTCACTG
s1pr1 **[exon 2.2]**	ACAAGTTGACCCCAACGCAATGG	97%	TAACAAGTTGACCCCAACGCAA AAACTTGCGTTGGGGTCAACTT

Constructs and sgRNA generation

It was possible to clone all the previously annealed oligonucleotides (Figure 6a), except for the ones for the target *s1pr1* [exon 2.2]. This was confirmed by an initial screen based on elimination of the SalI restriction site when the annealed oligonucleotides are correctly ligated (Figure 6b). Cloning efficiencies varied from 100% (*pdgfrß* [exon 2], *foxc1a* [exon1.2] and *s1pr1* [exon 2.1]), 60% (*pdgfrß* [exon 1], *fabp11a* [exon2] and [*foxc1a* exon 1.1], 20% (*fabp3* [exon1] and [exon2]) and 0% (*s1pr1* [exon 2.2]).

In vitro transcription of the several sgRNA was achieved, after linearization of the vectors, with the exception of *fabp3* [exon 1] (fig 6 c). In all examples it is possible to see the elimination or reduction (in the *fabp3* [exon 1] case) of the DNA template (linearized pT7) after 15 minute treatment with DNAse.

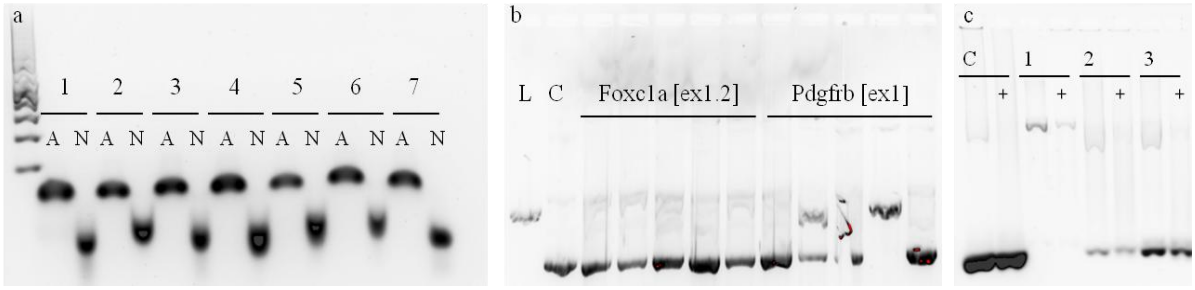

Figure 6. Agarose gels of the annealing, cloning and transcription processes. **a)** Successfully annealed 20bp oligonucleotides (A) compared with their non annealed version (N) as control, (1) *pdgfrß* [exon1], (2) *pdgfrß* [exon2], (3) foxc1a [exon1.1], (4) foxc1a [exon 1.2], (5) *fabp3* [exon 1], (6) *fabp3* [exon 2], (7) *fabp11a* [exon 2]. **b)** SalI restriction site existence screen. pT7sgRNA plasmids from 5 different colonies for each construct were digested with SalI enzyme, to screen for the presence of its restriction site, eliminated after correct ligation of the annealed oligonucleotides in the pT7sgRNA. In this case, all colonies from pT7sgRNA_*foxc1a* [exon1.2] were positive for the insert, as well as 3/5 of the pT7sgRNA_*pdgfrß*[exon1]. Linearized (L) pT7sgRNA backbone and Circular (C) original pT7sgRNA backbone used as controls. **c)** RNA after *in vitro* transcription of control template (C), *fabp3* [exon 1] (1), *fabp3* [exon 2] (2) and *s1pr1* [exon 2]. Treatment with DNAse (+) eliminated the vector template in all cases but in *fabp3* [exon 1].

Phenotyping

The majority of the injected embryos were WT-like. However, in all different groups injected with sgRNA, including the control unspecific sgRNA, there was a prevalence of certain aberrant

phenotypes such as curved tail, perturbed anterior-posterior axis, underdeveloped eyes, pericardial edema, yolk edema and blood accumulations (Figure 7). These aberrant phenotypes started were observed as early as 1dpf or later at 3-4dpf.

There was no evidence for significant high mortality in any of the specific sgRNA injections, in comparison with the uninjected control group (Table 4). However, the difference between number of dead embryos in both *pdgfrß* sgRNA injected groups and control group is the highest of all the injection groups. Moreover, the amount of sgRNA did not correlate with the number of deaths or deformities in any of the sgRNA injected groups (Table 4). Generally, the group injected with the unspecific sgRNA had a higher mortality and higher number of deformed embryos than the group injected with the specific sgRNAs (Table 4). All healthy embryos were transferred to tanks and left to grow in the fish facility. Some of these may be heterozygous and therefore carry a mutation. Outcrosses will be done, as well as genotyping, to identify the heterozygous founders.

Analysis of the number of deaths and deformed embryos on each group of the toxicity control experiment led to observation of further toxicity-derived phenotypes. Furthermore, it was possible to observe a tendency for a decreased WT-like phenotype, as the concentration of injected RNA increases (Figure 8). It is also possible to see an increase of deformed embryos as the concentration of the sgRNA increases and the Cas9 RNA injected amount is kept at 150ng. However, at higher concentrations, when the Cas9 RNA injected amount is 300pg, the number of dead embryos increases drastically (Figure 8).

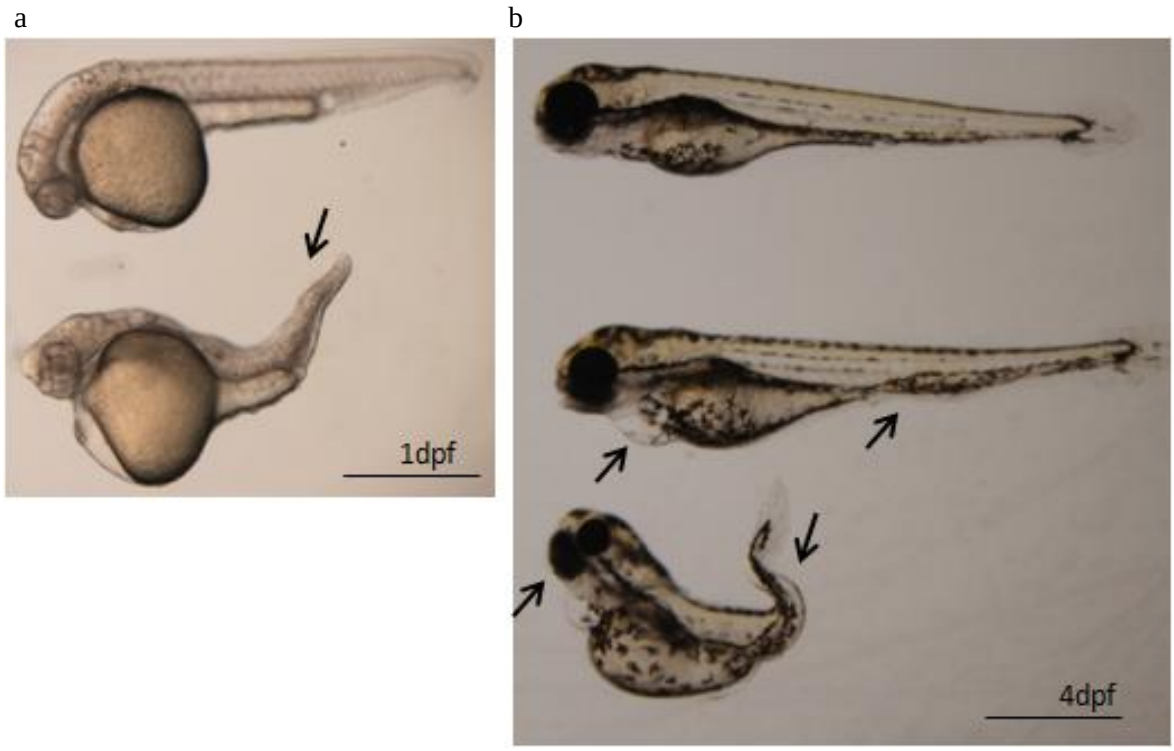

Figure 7. Uninjected and injected zebrafish embryos. WT phenotypes on top and aberrant phenotypes beneath. Phenotypes include curved tail, pericardial edema, underdeveloped eyes and yolk extension malformation. Scale bar 0,5 mm in a) and 0,3mm in b)

Table 4. Difference between percentage of dead and deformed embryos in the groups injected with specific sgRNA (for each target) and uninjected control and injected control groups of the same clutch. Negative percentages correspond to cases where controls had higher number of dead and/or deformed embryos in comparison to the specific sgRNA injections.

	Uninjected control		**Injected control**	
Target	Δ dead	Δ deformed	Δ dead	Δ deformed
***pdgfrß* [exon1]**				
35pg	19,1	9,1	-19,4	-35,5
100pg	23,0	10,6	7,6	-31,6
***pdgfrß* [exon2]**				
35pg	15,7	17,9	-22,9	-26,7
50pg	27,3	-1,0	20,8	-12,5
70pg	13,8	11,5	-32,1	5,8
100pg	15,4	-3,9	8,9	-15,4
***foxc1a* [exon 1.1]**				
35pg	29,0	12,2	-9,6	-32,3
100pg	-16,9	4,6	-32,3	-37,7
***foxc1a* [exon1.2]**				
35pg	-3,1	6,7	-	-
70pg	-5,8	0,5	-46,7	-5,2
100pg	0,9	3,2	-14,4	-39,1
***fabp3* [exon2]**				
70 pg	6,1	8,3	8,7	-18,4
***fabp11a* [exon 2]**				
70pg	11,3	1,3	13,9	-25,4
***s1pr1* [exon 2.1]**				
70 pg	3,2	2,3	5,8	-24,4

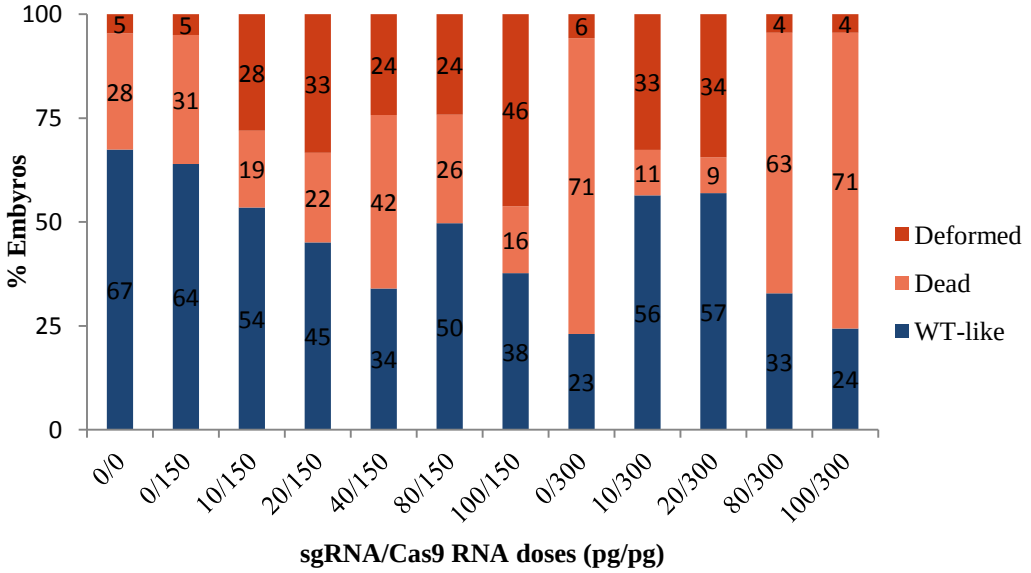

Figure 8. Percentage of dead, deformed and healthy embryos, counted from1 to5dpf. All groups were part of the same clutch of zebrafish. Group number from 70 to 150.

Detection of mutants – Genotyping

In each HRMA performed, uninjected WT fish were included, so as to compare natural WT variation (e.g. naturally occurring polymorphisms) in the amplicon to variation caused by sgRNA and Cas9 mutagenesis. After background subtraction and normalization of the fluorescent signal the software clusters each sequence variants, which exhibit similar melting profiles. All clusters were analysed with and without temperature shifting, i.e. correcting for minor temperature variation which increases the ability to distinguish heterozygous from homozygous, and all samples clustered together were listed. It was considered that there was no successful mutagenesis in the injected embryos which cluster together with the WT embryos. Samples in clusters with only injected embryos and no WT embryos were considered potential mutants. A conservative approach was taken while interpreting the clusters. Possible mutant clusters located between two WT clusters were considered negative for the existence of mutation. These criteria were used to generate an efficiency table, for each concentration injected for each target.

It was possible to generate the proposed HRM analysis in 6 out of the 7 targets. The PCR of the CRISPR-targeted *fabp3* [exon2] was not successfully performed on the injected samples, amplifying DNA only from WT embryos (not shown). This can indicate a potential positive result, as it could be explained by non-annealing of the primers in all the injected samples because of a mutation in the primer annealing region. However, these results are inconclusive.

The HRMA of the sgRNA targeting *pdgfrb* [exon 1] exhibited some variance in the melting profiles of the 11 uninjected WT samples (Figure 8a), however these WT samples were all clustered together by the software, together with 15 injected samples, in which no mutation occurred. The software clustered together 8 injected samples (4 injected with 100pg and 4 injected with 50pg) (Table 5), which are considered to be a different sequence variant from the uninjected WT samples, and therefore positive for sgRNA and Cas9 derived mutation. For the sgRNA targeted *pdgfrß* [exon 2], two clusters were found. However, when analyzed, WT samples were distributed between them (4WT in cluster 1 and 8 WT in cluster 2), which might be explained by a natural variation in the WT samples, such as a SNP polymorphism. This was therefore accounted as 0% efficiency (Table 5).

The sgRNAs targeting *foxc1a* [exon 1.1] and [exon1.2] showed successful mutagenesis, with a variety of efficiencies, since the melting profiles of the uninjected samples are clustered (Figure 8c and 8d). Cluster analysis showed that the lowest efficiencies were achieved in the *foxc1a* [exon1.1] target (Table 5). The melting profiles of the targeted gene *foxc1a* [exon 1.2] exhibit a clear division between uninjected samples (black), samples injected with 70 and 100pg (orange) and samples injected with 35pg (yellow), the last ones showing the biggest changes in comparison with the non-injected samples. In fact, analysis of the clusters show total mutagenesis on the samples injected with 35pg, with 6 different sequence variants from the WT cluster.

In the case of the targeted *fabp11a*, the WT samples' melting profiles are equally distributed in two different clusters (4 samples in each), hence no major difference was detected between the melting curves of the injected and uninjected WT samples (Figure 8e). Lastly, the HRMA for detection of unknown mutations in the *s1pr1*target site resulted in three different clusters, all containing uninjected samples. The last two cases are therefore negative for targeted mutagenesis, according to the criteria applied.

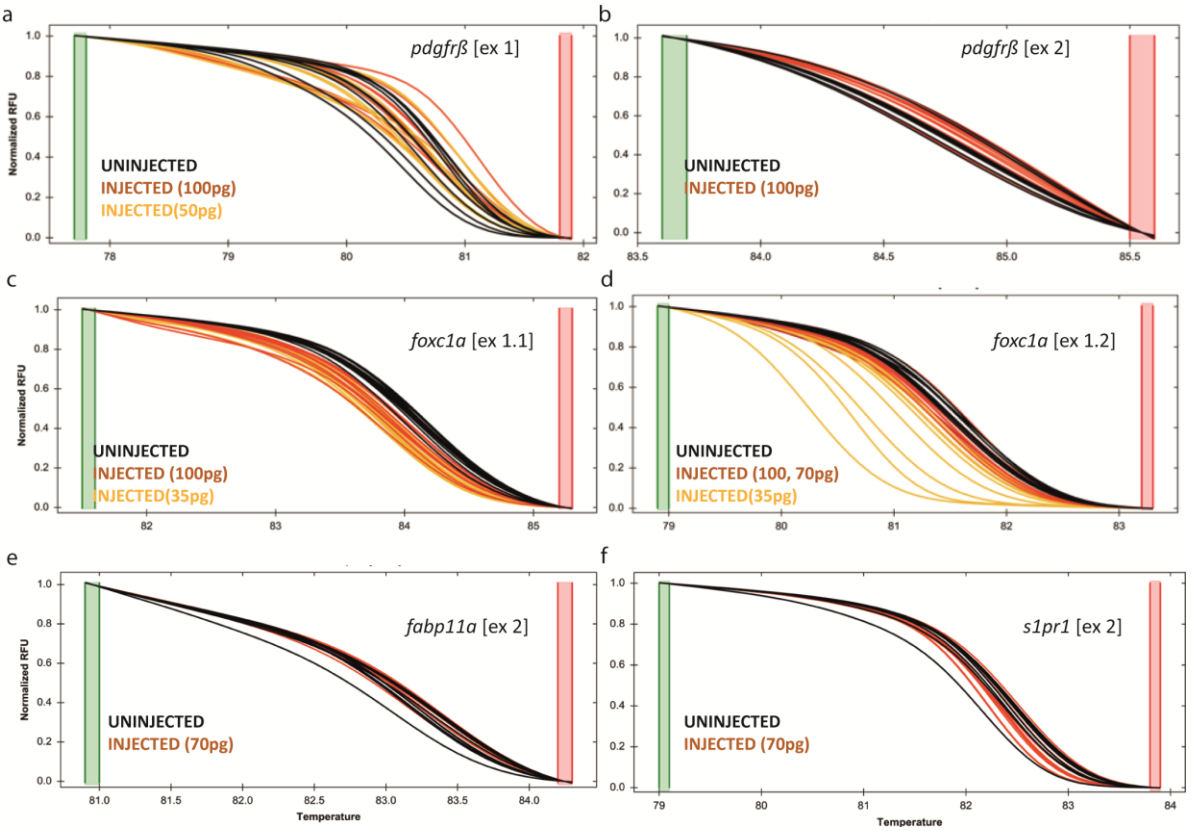

Figure 8. Normalizing melting curves for 6 different targets. Comparison between uninjected and injected samples targeting a) *pdgfrß* [exon1] b) *pdgfrß* [exon2] c) *foxc1a* [exon 1.1] d) *foxc1a* [exon 1.2] e) f*abp11a* [exon 2] f) *s1pr1* [exon2.1] Each curve represents an embryo. RFU: relative fluorescent unit.

Table 5. Efficiencies of mutagenesis for each target. In all cases, the amount of Cas9 injected was 150pg. All efficiencies were calculated based on HRMA clusters data.

Target	Amount sgRNA injected	# embryos	# embryos with mutation	%
pdgfrß **[exon 1]**	35pg	8	7	87,5
	50pg	12	4	33,3
	100pg	11	4	36,4
pdgfrß **[exon 2]**	35pg	11	0	0
	50pg	10	0	0
	70pg	16	0	0
	100pg	9	0	0
foxc1a **[exon 1.1]**	35pg	22	3	13,6
	100pg	10	2	20
foxc1a **[exon 1.2]**	35pg	11	11	100
	70pg	10	4	40,0
	100pg	10	9	90
fabp3 **[exon 2]**	70pg	12	0	0
fabp11a **[exon2]**	70pg	12	0	0
s1pr1 **[exon 1]**	70pg	12	0	0

DISCUSSION

The CRISPR-Cas genome editing methodology was successfully applied to target 3 out of 7 genome locations within the zebrafish, generating KO mutants for two different genes, *pdgfrß* and *foxc1a*. Therefore, it was demonstrated in this project that customized sgRNA-Cas9 system can be used to induce mutations in zebrafish. Generating mutants instead of purchasing them and genome-scale KO production are now not only possible but also feasible through the CRISPR-Cas technique. We anticipate that the use of this technique will continue to expand and develop as more laboratories and model organisms embrace its use. We believe a wider use of this robust and uncomplicated tool, if applied with caution, is therefore encouraged.

Efficient monoallelic mutation

MOs directed against our genes of interest led to specific phenotypes, such as hydrocephalus and kidney cysts in *pdgfrb* MO (in preparation); glomerular vasculature underdevelopment in *foxc1a*[69]; hydrocephalus, trunk vascular patterning defects in *fabp1* MO (unpublished); and brain edema and excess of vascular sprouting in the hindbrain in *s1pr1* MO[67]. It is expected that KOs phenocopy the KDs. However, slight differences are anticipated such as a stronger KO phenotype, due to the fact that the KD phenotype reflects only a titration out of protein function and the KO may reflect a total loss of protein function or a milder KO phenotype, as there could be some toxicity in the MO which causes a part of the KD phenotype. Nevertheless, the MO phenotype is to be used as a guideline for the detection of a specific phenotype due to loss of protein and therefore comparable to a biallelic recessive mutation. In this project, none of the specific phenotypes were found on the injected embryos, most likely reflecting the inexistence or very low amount of injected embryos with biallelic mutations caused by the sgRNA-Cas9 RNA induced mutagenesis.

Despite the fact that biallelic mutants were not identified through phenotyping for any of the targets, it was possible to detect embryos carrying mutation(s) by HRMA. It is important to mention that these mutants might be carrying several types of gene modifications. This might occur due to delays in either the translation of the Cas9 mRNA or in the activation of the Cas9, both leading to the occurrence of the mutagenesis event after the first cell division and most likely, different acquired mutations in the daughter cells, each of which may arise in a mosaic way during development. Hence, each melting profile reflects the complex combination of genotypes of the founder embryos, the combination of the indel size and a the percentage of mutated cells[70]. Melting curves of biallelic mutants would, most likely, behave as heterozygous in the analysis, due to their intrinsic genetic complexity, disabling the possibility of distinguishing them from the monoallelic mutants.

The results shown point to an achievement of sgRNA-Cas9 mediated mutagenesis in two genes out of the 5, with a variety of efficiencies (from 14 to 90%). There are two possible reasons why it is easier to induce mutations in some genes than in others. One possibility is that some genes are more easily reached by the sgRNA, which might depend on epigenetic factors of the genomic region such as the chromatin state but not the presence of methylation[32,71]. Alternatively, some sgRNA might be more efficient than others, without having any biochemical differences between them (personal communication). There seem to exist not yet identified, extra rules of activity of this complex which would explain why certain combinations of nucleotides of the sgRNA are more efficient than others, and therefore why some genes are easier to mutate than others, even if having the same epigenetic state. Accordingly, efforts on learning about the specificity of this system, and the biochemical interaction between the complex sgRNA-target DNA-Cas9, as well as analysis of efficiencies of each sgRNA large scale mutagenesis are fundamental for the continuation of the practice of CRISPR-Cas derived mutagenesis.

The amount of sgRNA injected does not seem to increase the efficiency in this project, against previously reported work[31]. Instead, efficiency differences might be explained by the injection into differentially developed embryos, differences in the distribution of the mRNAs following the injections and differences in translation efficiencies. It is important to state that these observations lack replications and therefore statistical support.

Toxicity and Off-target effects

Toxicity can be caused by the interaction of a compound, in this case the sgRNA and Cas9 RNA, with a target that is distinct from the intended[72], in this case the target DNA. Toxicity is considered as an off-target effect, because it is a result of the collateral damage caused by the RNA injected in an organ, structure or cell, which leads to a phenotypic defect or death. Effects observed generally, independently of the type of sgRNA injected, such as early death, pericardial edema and curved tail, were considered toxic effects in this project. In the current literature there is an ongoing debate regarding the possibility of occurrence of CRISPR-Cas caused specific off-target effects, which occur when the sgRNA and Cas9 target DNA locations other than the intended one. It was shown before that extensive, but not necessarily total homology with the sgRNA guide element is required for targeting with mismatch tolerance depending on the number and position of sgRNA-DNA mismatches and that most off-target cleavage occurs in locations in the genome predicted by sofware[4,22,36]. By taking this into account, the choice of a 20bp target with the highest specificity amongst the options of the software's output allowed us to minimize the risk of off-target events. The mentioned specific off-target mutagenesis are major concerns in this project, since mutagenesis in on-target and off-target genome locations could lead to phenotypes which would not only be a

consequence of the intended mutated gene, but a complex synergistic result of several mutated locations in the genome. If such an event occurs, it would be hard to detect it without an extensive genomic analysis, since the phenotypes would seem specific for the intended mutation.

In our experience, increase of amount of sgRNA and Cas9 RNA injected increases the amount of dead and deformed embryos, as it was shown by the toxicity control experiment. One way of designing similar projects with an increased specificity would be to use a "double nicking strategy" which makes use of a mutated Cas9 (nickase D10) that generates single-stranded nicks, not repaired by NHEJ. Thus, in order to have a DSB, and therefore a mutated gene, two sgRNA need to recognize sequences flanking the region of interest to modify in the same DNA molecule, which effectively would double the number of bases that need to be recognized at the target site and increase specificity of genome editing[73]. Additionally, to guarantee the absence of off-target events, one could genotype the most likely off-target locations, predicted by the software.

Although we believe the off-target effect risks were minimized in this project, there is still a need for a systematic and inclusive analysis of cleavage specificity generated from directed mutagenesis of similar target sites, which would help to gain insights into targeting specificity and the extent of potential off-target effects, as well as to improve the use of this complex system as a molecular tool and in therapeutics. Moreover, further work analysing kinetics and stability of sgRNA-DNA complex as well as research on Cas9 mutants or orthologs could lead to finding of variants with higher specificity, which would yield to an additional power for off-target predictability and an improvement of the specificity of the technique.

The explanation for the injected control with an unspecific sgRNA having led to more off-target effects than the specific injected sgRNA may reside in their fundamental difference: the absence of a sequence complementary to a target in the zebrafish genome. By having only the hairpins characteristic of the common sgRNA and lacking the other component of the sgRNA, it might be more active within the zebrafish genome, interacting randomly, and therefore generating more toxicity or off-target effects. A sgRNA with a 20bp sequence, not complementary to any region in the zebrafish genome would have been a more accurate control. However, it is important to highlight the fact that the group of control embryos injected with the control sgRNA was only necessary for analysis of the presence of embryos with biallelic mutations, as a validation of the specificity of their phenotypes. In the scope of this project, it was of extreme importance to compare the phenotypes of the embryos injected with specific sgRNA with the phenotypes of the embryos injected with control unspecific sgRNA to conclude if phenotypes achieved were due to general unspecific toxicity effects, common to all injected groups.

Limitations of the project

The possibility to direct Cas9 to target specific sites by simply designing a sgRNA complementary to the region to be targeted holds an enormous advantage in functional biology projects. However, it is important that we better understand this system in order to minimize the risk of off-target effects when making mutants, and as a result draw the correct conclusions on protein functions while interpreting the phenotypes of mutants. The fact that the effectiveness of mutagenesis in certain regions cannot be predicted is also a limitation, with a range of effectiveness from zero, as in the examples *fabp3*[exon 2], *fabp11a* [exon 2] and *s1pr1* [exon 2], to biallelic mutations in the injected founders reported[41], using the same Cas9.

The necessary comparison between the previously generated MO phenotypes and the possible KO phenotypes can be helpful when analysing the mutants, but can also result in a biased approach. If one is expecting a particular phenotype one could miss certain details on unexpected phenotypes, especially in the case of this work in which the efficiency of biallelic mutants in the injected embryos was not known. However, efforts were made to overcome this problem while analysing the mutants, by having more than one person phenotyping and discussing possible phenotypes. Moreover, It is not possible to state with total certainty that biallelic KO mutants were not generated by the technique, solely based on the phenotypes we were expecting to see.

Genotyping is an extremely important step for detecting the effectiveness of the technique, and to identify the type of mutations. All the options for genotyping mosaic individuals with unknown mutations, if any, have caveats. For example, the used technique, HRMA is a useful, rapid and cost friendly technique but it does not allow to distinguish injected mosaic biallelic mutated individuals from monoallelic mutated ones nor to specify which modifications were introduced in the genome by the NHEJ repair mechanism, after the Cas9 cleavage. Other techniques such as Fragment Length Analysis, another PCR based technique, are able to detect the presence of insertions or deletions, their size and how many copies of the gene are mutated, but at the expense of detection of small modifications such as 1bp indels or substitutions (detected by the HRMA)[74]. Another option would be to introduce a template oligonucleotide designed with the desired modification and a restriction site, inducing HDR and causing a predictable modification in a desired location. By doing so, the genotyping would be easier to achieve, as it would be based on the presence/absence of the known restriction site.

An ongoing project

To ultimately confirm the presence of a specific mutation in the injected embryos positively identified by HRMA, TOPO cloning needs to be performed to ensure the sequencing of only one of the several types of amplicons, which will be as many as the number of NHEJ events in the injected founders. Furthermore, the sgRNA-Cas9 injections that did not result in mutagenesis should be

repeated and, if necessary, sgRNA redesigned to ensure the generation of the *fabp3*, *fabp11a* and *s1pr1* mutants.

Since there were no biallelic mutants, phenotypic analysis is still to be done. For that, the mosaic founders should be outcrossed in order to generate F1 embryos in which each individual is heterozygous for one mutation only. Hence, it is necessary to genotype and sequence the F1 individuals, by fin clipping them. In case of absence of mutation, it might because of i) low heritability, due to absence of mutagenesis in the germline or ii) no mutagenesis in the parent founder. Ultimately, the zebrafish population desired to become a mutant carrier heterozygous line is the F1, which possesses only one characterized mutation present in their genome. F1 siblings with the mutations at the same loci are to be crossed, to obtain mutants in a quarter of the F2 offspring which would be further characterized by phenotyping as well as by performing *in situ* hybridizations of angiogenesis and kidney development key genes and vascular and glomerular functional integrity test (dextran injections). It would also be necessary to confirm if the mutants generated yield full loss-of-function phenotypes. This could be easily done by performing a Western blot to detect the presence or absence of protein. It will be interesting to see whether a mutation in a different region of the gene will generate different levels of protein, or different forms, and therefore different phenotypes.

CONCLUSION

It was possible to achieve the first but not the second aim of the project, since the sgRNA-Cas9 derived heterozygous mutants were generated and detected. By doing so, the present project was fundamental to establish this technique in the working group. However, by not generating any one-step biallelic mutants, it was not possible to achieve the phenotypic analysis in the F0, the second aim of the project. . This study supports the previous reported cases of high occurrence of Cas9-derived mutagenesis for some of the aimed genes, which illustrates our incomplete understanding of how such an effectively evolved system should be applied to achieve the fidelity nature does. Although it is becoming a very useful and accepted technique for generating mutants *in vivo*, it is important to highlight that this is a very recent technique and much still remains to fully understand the mode of action of the sgRNA-DNA-Cas9 complex. Therefore, more research in the biochemistry and mode of action of the CRISPR-Cas immune system in the bacteria are still necessary to refine this genome editing tool. Despite these problems, we are confident that the CRISPR-Cas derived genome editing technique will enable the performance of functional studies in mutants not available so far, and help us understand the role of important genes such as *pdgfrß* and *foxc1a* and the mechanisms underlying their physiological and pathophysiological effects. Effectively, because they are key genes in the development of kidney and vasculature, it is essential to better understand their functions and their possible relation with certain human diseases.

ACKNOWLEDGEMENTS

I'm very thankful to my supervisor Lwaki Ebarasi for guidance, support, patience and for introducing me to such exciting fields of the genome editing and kidney and vascular biology.
I am also extremely grateful to everyone in the Christer Betsholtz group, for making me feel like part of the group and for the help on the day-to-day life as well as many laughs and support.
I would like to thank Linnea Gunnarsson for all the help, particularly with the injections, and Katarina Garpenstrand, for the understanding and availability with the logistics of the fish facility.
I also want to thank my family and friends, without whom I could not have accomplished this project.
Lastly, I would like to truly thank Gonçalo not only for all the support, but also for the fruitful discussions about my work, which helped me see so much. I feel very lucky for having you by my side during this journey.

REFERENCES

1. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–12 (2007).

2. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327,** 167–70 (2010).

3. Terns, M. P. & Terns, R. M. CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **14,** 321–7 (2011).

4. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–21 (2012).

5. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, a. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169,** 5429–33 (1987).

6. Jansen, R., Embden, J. D. a Van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–75 (2002).

7. Mojica, F., Díez-Villaseñor, C. & Soria, E., Juez, L. Biological significance of a family of regularly spaced repeats in the genomes of Archea, Bacteria and mitochondria. *Mol. Microbiol.* **36,** 244–246 (2000).

8. Wiedenheft, B., Sternberg, S. H. & Doudna, J. a. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482,** 331–8 (2012).

9. Makarova, K. S., Brouns, S. J. J., Horvath, P., Sas, D. F. & Wolf, Y. I. Evolution and classification of the CRISPR-Cas systems. *NIH Public Access* **9,** 467–477 (2012).

10. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–61 (2005).

11. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151,** 653–63 (2005).

12. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–82 (2005).

13. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.* **190,** 1390–400 (2008).

14. Bondy-Denomy, J. & Davidson, A. R. To acquire or resist: the complex biological effects of CRISPR–Cas systems. *Trends Microbiol.* 1–8 (2014).

15. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–40 (2009).

16. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321,** 960–4 (2008).

17. Hale, C. R. *et al.* RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139,** 945–56 (2009).

18. Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M. & Macmillan, A. M. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat. Struct. Mol. Biol.* **18,** 688–92 (2011).

19. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–7 (2011).

20. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* (2014).

21. Luciano A. Marraffini and Erik J. Sontheimer. CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *NIH Public Access* **322,** 1843–1845 (2009).

22. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 10098–103 (2011).

23. Gudbergsdottir, S. *et al.* Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79,** 35–49 (2011).

24. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).

25. Marraffini, L. A. & Sontheimer, E. J. Self vs. non-self discrimination during CRISPR RNA-directed immunity. *NIH Public Access* **463,** 568–571 (2010).

26. Adi Stern, Leeat Keren, Omri Wurtzel, Gil Amitai, and R. S. Self-targeting by CRISPR: gene regulation or autoimmunity? *NIH Public Access* **26,** 335–340 (2011).

27. Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* **6,** 181–6 (2008).

28. Gaj, T., Gersbach, C. a & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31,** 397–405 (2013).

29. Sakuma, T. & Woltjen, K. Nuclease-mediated genome editing: At the front-line of functional genomics technology. *Dev. Growth Differ.* **56,** 2–13 (2014).

30. Carroll, D. & Beumer, K. J. Genome engineering with TALENs and ZFNs: Repair pathways and donor design. *Methods* 1–5 (2014).

31. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31,** 227–9 (2013).

32. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–32 (2013).

33. Krasteva, P. V. Biochemistry: CRISPR snapshots of a gene-editing tool. *Nat. Methods* **11,** 365–365 (2014).

34. Carroll, D. Staying on target with CRISPR-Cas. *Nat. Biotechnol.* **31,** 807–9 (2013).

35. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2,** e00471 (2013).

36. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–23 (2013).

37. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. a. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–9 (2013).

38. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153,** 910–8 (2013).

39. Yu, Z. *et al.* Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila. *Genetics* **195,** 289–91 (2013).

40. Nakayama, T. *et al.* Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in Xenopus tropicalis. *Genesis* **51,** 835–43 (2013).

41. Jao, L.-E., Wente, S. R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 13904–9 (2013).

42. Fujii, W., Kawasaki, K., Sugiura, K. & Naito, K. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. *Nucleic Acids Res.* **41,** e187 (2013).

43. Xiao, A. *et al.* Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res.* **41,** e141 (2013).

44. Auer, T. O. & Del Bene, F. CRISPR/Cas9 and TALEN-mediated knock-in approaches in zebrafish. *Methods* (2014).

45. Wu, Y. *et al.* Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* **13,** 659–62 (2013).

46. Hwang, W. Y. *et al.* Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One* **8,** e68708 (2013).

47. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–7 (2014).

48. Stemple, D. L. & Driever, W. Zebrafish: tools for investigating cellular differentiation. *Curr. Opin. Cell Biol.* **8,** 858–64 (1996).

49. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496,** 498–503 (2013).

50. Lieschke, G. J. & Currie, P. D. Animal models of human disease: zebrafish swim into view. *Nat. Rev. Genet.* **8,** 353–67 (2007).

51. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326,** 1509–12 (2009).

52. Eisen, J. S. & Smith, J. C. Controlling morpholino experiments: don't stop making antisense. *Development* **135,** 1735–43 (2008).

53. Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. & Wolfe, S. a. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* **26,** 695–701 (2008).

54. Huang, P. *et al.* Heritable gene targeting in zebrafish using customized TALENs. *Nat. Biotechnol.* **29,** 699–700 (2011).

55. Auer, T. O., Duroure, K., De Cian, A., Concordet, J.-P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24,** 142–53 (2014).

56. Fredriksson, L., Li, H. & Eriksson, U. The PDGF family: four gene products form five dimeric isoforms. *Cytokine Growth Factor Rev.* **15,** 197–204 (2004).

57. Hellström, M., Kalén, M., Lindahl, P., Abramsson, a & Betsholtz, C. Role of PDGF-B and PDGFR-beta in recruitment of vascular smooth muscle cells and pericytes during embryonic blood vessel formation in the mouse. *Development* **126,** 3047–55 (1999).

58. Soriano, P. Abnormal kidney development and hematological disorders in PDGF beta-receptor mutant mice. *Genes Dev.* **8,** 1888–1896 (1994).

59. Armulik, A. *et al.* Pericytes regulate the blood-brain barrier. *Nature* **468,** 557–61 (2010).

60. Lindblom, P. *et al.* Endothelial PDGF-B retention is required for proper investment of pericytes in the microvessel wall. *Genes Dev.* **17,** 1835–40 (2003).

61. Lindahl, P. *et al.* Paracrine PDGF-B/PDGF-Rbeta signaling controls mesangial cell development in kidney glomeruli. *Development* **125,** 3313–22 (1998).

62. Takemoto, M. *et al.* Large-scale identification of genes implicated in kidney glomerulus development and function. *EMBO J.* **25,** 1160–74 (2006).

63. Flynn, E. J., Trent, C. M. & Rawls, J. F. Ontogeny and nutritional control of adipogenesis in zebrafish (Danio rerio). *J. Lipid Res.* **50,** 1641–52 (2009).

64. Coe, N. R., Simpson, M. a & Bernlohr, D. a. Targeted disruption of the adipocyte lipid-binding protein (aP2 protein) gene impairs fat cell lipolysis and increases cellular fatty acid levels. *J. Lipid Res.* **40,** 967–72 (1999).

65. Karanth, S., Denovan-Wright, E. M., Thisse, C., Thisse, B. & Wright, J. M. The evolutionary relationship between the duplicated copies of the zebrafish fabp11 gene and the tetrapod FABP4, FABP5, FABP8 and FABP9 genes. *FEBS J.* **275,** 3031–40 (2008).

66. Liu, R.-Z. *et al.* The fabp4 gene of zebrafish (Danio rerio)--genomic homology with the mammalian FABP4 and divergence from the zebrafish fabp3 in developmental expression. *FEBS J.* **274,** 1621–33 (2007).

67. Gaengel, K. *et al.* Article The Sphingosine-1-Phosphate Receptor S1PR1 Restricts Sprouting Angiogenesis by Regulating the Interplay between VE-Cadherin and VEGFR2. 587–599 (2012).

68. Saxon, L. & Sariola, H. Pediatric Nephrology Early organogenesis of the kidney. 385–392 (1987).

69. Brien, L. L. O. *et al.* Wt1a, Foxc1a, and the Notch mediator Rbpj Physically Interact and Regulate the Formation of Podocytes in Zebrafish. *Dev. Biol.* **358,** 318–330 (2012).

70. Dahlem, T. J. *et al.* Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet.* **8,** e1002861 (2012).

71. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* (2014).

72. Rubinstein, A. L. Zebrafish assays for drug toxicity screening. *Ashley Publ.* 231–240 (2006).

73. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154,** 1380–9 (2013).

74. Parant, J. M., George, S. a, Pryor, R., Wittwer, C. T. & Yost, H. J. A rapid and efficient method of genotyping zebrafish mutants. *Dev. Dyn.* **238,** 3168–74 (2009).

75. Zhang, F., Wen, Y. & Guo, X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum. Mol. Genet.* 1–7 (2014).

APPENDIX I

Normalized melting curves of 6 different targets, in a cluster view, different colors representing different clusters and therefore, different sequence variants of a) *pdgfrß* [exon1] b) *pdgfrß* [exon2] c) *foxc1a* [exon 1.1] d) *foxc1a* [exon 1.2] e) f*abp11a* [exon 2] f) *s1pr1* [exon2.1]. Each curve represents an embryo.